Hopfield-like network with complementary encodings of memories

Louis Kang [1,2,*] and Taro Toyoizumi [3,4]

[1]*Neural Circuits and Computations Unit, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan*
[2]*Graduate School of Informatics, Kyoto University, 36-1 Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
[3]*Laboratory for Neural Computation and Adaptation, RIKEN Center for Brain Science, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan*
[4]*Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*

We present a Hopfield-like autoassociative network for memories representing examples of concepts. Each memory is encoded by two activity patterns with complementary properties. The first is dense and correlated across examples within concepts, and the second is sparse and exhibits no correlation among examples. The network stores each memory as a linear combination of its encodings. During retrieval, the network recovers sparse or dense patterns with a high or low activity threshold, respectively. As more memories are stored, the dense representation at low threshold shifts from examples to concepts, which are learned from accumulating common example features. Meanwhile, the sparse representation at high threshold maintains distinctions between examples due to the high capacity of sparse, decorrelated patterns. Thus, a single network can retrieve memories at both example and concept scales and perform heteroassociation between them. We obtain our results by deriving macroscopic mean-field equations that yield capacity formulas for sparse examples, dense examples, and dense concepts. We also perform simulations that verify our theoretical results and explicitly demonstrate the capabilities of the network.

I. INTRODUCTION

Autoassociation is the ability for a network to store patterns of activity and to retrieve complete patterns when presented with incomplete cues. Autoassociative networks are widely used as models for neural phenomena, such as episodic memory [1–3], and also have applications in machine learning [4,5]. It is well known that properties of the stored patterns can influence the computational capabilities of the network. Sparse patterns, in which a small fraction of the neurons is active, can be stored at higher capacity compared to dense patterns [6–12]. Correlated patterns can be merged by the network to represent shared attributes [13–15]. Previous autoassociation models have largely considered the storage of patterns with a single set of statistics, which requires trade-offs among computational features. For example, the ability to learn categories with correlated patterns may be desired, but correlations decrease the capacity for retrieving patterns distinctly.

We consider the possibility that a network can store two types of patterns with different properties and, thus, different computational roles. This idea is inspired by the architecture of the hippocampus in mammalian brains [16]. The hippocampal subfield CA3 is the presumptive autoassociative network that stores memories of our daily experiences [1,6], and it receives sensory information from two parallel pathways with complementary properties [17]. The mossy fibers present sparser, decorrelated patterns to CA3 for storage, and

the perforant path presents denser, correlated patterns. Both pathways originate from the same upstream region, the entorhinal cortex, so they presumably encode the same sensory experiences. However, based on theoretical studies described above, their computational capabilities may differ. We expect the denser, correlated patterns to build representations of *concepts* through the accumulation of *examples*; meanwhile, we expect the sparser, decorrelated patterns to represent distinct examples at high capacity. We wish to explore whether an autoassociative network can store and retrieve memory encodings from each pathway. Doing so could enable information representation at different scales, enabling the network to simultaneously discriminate between examples and generalize across them.

To address this aim, we implement a Hopfield-like network [18] that stores memories, each of which is an example ν of a concept μ. Each example is encoded as both a sparse pattern $\boldsymbol{\xi}_{\mu\nu}$ and a dense pattern $\boldsymbol{\psi}_{\mu\nu}$. The former is generated independently and exhibits no correlation with other sparsely encoded examples. The latter is generated from a dense encoding $\boldsymbol{\psi}_\mu$ of the concept μ with correlations among examples within the same concept. The model is defined in Sec. II, along with an outline of the derivation of its mean-field equations.

In Sec. III we present our major results regarding pattern retrieval. We can use a high or low activity threshold to retrieve sparse or dense patterns, respectively. The network has a high capacity for sparse examples $\boldsymbol{\xi}_{\mu\nu}$ and a low capacity for dense examples $\boldsymbol{\psi}_{\mu\nu}$. As the number of examples stored increases beyond the dense example capacity, a critical load is reached above which the network instead retrieves dense concepts $\boldsymbol{\psi}_\mu$. This critical load can be smaller than the sparse

*Corresponding author: louis.kang@riken.jp

example capacity, which means that the network can recover both $\boldsymbol{\xi}_{\mu\nu}$'s as distinct memories and $\boldsymbol{\psi}_\mu$'s as generalizations across them.

In Sec. IV we show that the network can perform heteroassociation between sparse and dense encodings of the same memory. Their respective energies can predict regimes in which heteroassociation is possible. We discuss our results and their significance in Sec. V. Mean-field equations governing network behavior are derived in Appendix A, and capacity formulas for $\boldsymbol{\xi}_{\mu\nu}$, $\boldsymbol{\psi}_{\mu\nu}$, and $\boldsymbol{\psi}_\mu$ are derived in Appendixes B, C, and D.

II. THE MODEL

A. Patterns and architecture

We consider a Hopfield network with neurons $i = 1, \ldots, N$ that are either inactive ($S_i = 0$) or active ($S_i = 1$). The network stores $\nu = 1, \ldots, s$ examples for each of $\mu = 1, \ldots, p$ concepts. The concept load per neuron is $\alpha = p/N$. Examples are encoded both sparsely as $\boldsymbol{\xi}_{\mu\nu}$ and densely as $\boldsymbol{\psi}_{\mu\nu}$. Following Ref. [7], sparse examples are generated independently with density a:

$$\xi^i_{\mu\nu} = \begin{cases} 0 & \text{with probability } 1 - a \\ 1 & \text{with probability } a. \end{cases} \quad (1)$$

While the term *sparsity* has also been used in the literature, we use *density* for a because higher a implies lower sparsity. Following Ref. [13], dense examples within a concept are correlated in the following way. Each concept corresponds to a dense pattern $\boldsymbol{\psi}_\mu$, generated independently with density $\frac{1}{2}$:

$$\psi^i_\mu = \begin{cases} 0 & \text{with probability } \frac{1}{2} \\ 1 & \text{with probability } \frac{1}{2}. \end{cases} \quad (2)$$

Dense examples are then generated from these concepts, with the correlation parameter $c > 0$ controlling the likelihood that example patterns match their concept:

$$\psi^i_{\mu\nu} = \begin{cases} \psi^i_\mu & \text{with probability } \frac{1+c}{2} \\ 1 - \psi^i_\mu & \text{with probability } \frac{1-c}{2}. \end{cases} \quad (3)$$

The average Pearson correlation coefficient between $\boldsymbol{\psi}_{\mu\nu}$ and $\boldsymbol{\psi}_\mu$ is c, and that between $\boldsymbol{\psi}_{\mu\nu}$ and $\boldsymbol{\psi}_{\mu\omega}$ for $\nu \neq \omega$ is c^2. The average overlaps are

$$\langle \psi^i_{\mu\nu} \psi^i_\mu \rangle = \frac{1}{4} + \frac{c}{4},$$
$$\langle \psi^i_{\mu\nu} \psi^i_{\mu\omega} \rangle = \frac{1}{4} + \frac{c^2}{4}, \quad (4)$$

where angle brackets indicate averaging over patterns.

During storage, the parameter 2γ sets the relative strength of dense encodings compared to sparse encodings. The factor

of 2 is for theoretical convenience. Linear combinations of $\boldsymbol{\xi}_{\mu\nu}$ and $\boldsymbol{\psi}_{\mu\nu}$ are stored in a Hopfield-like fashion with symmetric synaptic weights

$$
\begin{aligned}
J_{ij} &= \frac{1}{N} \sum_{\mu\nu} \left[(1 - 2\gamma)(\xi^i_{\mu\nu} - a) + 2\gamma \left(\psi^i_{\mu\nu} - \frac{1}{2} \right) \right] \\
&\quad \times \left[(1 - 2\gamma)(\xi^j_{\mu\nu} - a) + 2\gamma \left(\psi^j_{\mu\nu} - \frac{1}{2} \right) \right] \\
&= \frac{1}{N} \sum_{\mu\nu} \left(\eta^i_{\mu\nu} + \zeta^i_{\mu\nu} \right) \left(\eta^j_{\mu\nu} + \zeta^j_{\mu\nu} \right) \quad (5)
\end{aligned}
$$

for $i \neq j$, and $J_{ii} = 0$. The second expression uses rescaled sparse and dense patterns

$$
\begin{aligned}
\eta^i_{\mu\nu} &\equiv (1 - 2\gamma)(\xi^i_{\mu\nu} - a), \\
\zeta^i_{\mu\nu} &\equiv 2\gamma \left(\psi^i_{\mu\nu} - \frac{1}{2} \right). \quad (6)
\end{aligned}
$$

After initializing the network with a cue, neurons are asynchronously and stochastically updated via Glauber dynamics [19]. That is, at each time step t, one neuron i is randomly selected, and the probability that it becomes active is given by

$$P[S_i(t+1) = 1] = \frac{1}{1 + \exp\left\{ -\beta \left[\sum_j J_{ij} S_j(t) - \theta \right] \right\}}. \quad (7)$$

Thus, activation likely occurs when the total synaptic input $\sum_j J_{ij} S_j(t)$ is greater than the activity threshold θ. The inverse temperature $\beta = 1/T$ sets the width of the threshold, with $\beta \to 0$ corresponding to chance-level activation and $\beta \to \infty$ corresponding to a strict, deterministic threshold. We shall see that θ plays a key role in selecting between sparse and dense patterns; a higher θ suppresses activity and favors recovery of sparse patterns, and vice versa for lower θ and dense patterns.

B. Overview of mean-field equations

Network behavior in the mean-field limit is governed by a set of equations relating macroscopic order parameters to one another. Their complete derivation following Refs. [7,13,19,20] is provided in Appendix A, but we will outline our approach here. The first task is calculating the replica partition function $\langle Z^n \rangle$, where the angle brackets indicate averaging over rescaled patterns $\boldsymbol{\eta}_{\mu\nu}$ and $\boldsymbol{\zeta}_{\mu\nu}$ and n is the number of replica systems. By introducing auxiliary fields via Hubbard-Stratonovich transformations and integrating over interactions with off-target patterns, we obtain

$$\langle Z^n \rangle \propto \int \left[\prod_{\nu\rho} dm^\rho_{1\nu} \left(\frac{\beta N}{2\pi} \right)^{\frac{1}{2}} \right] \left[\prod_{\rho\sigma} dq^{\rho\sigma} dr^{\rho\sigma} \right] \exp(-\beta N f), \quad (8)$$

where ρ and σ are replica indices, $m^\rho_{1\nu}$, $r^{\rho\sigma}$, and $q^{\rho\sigma}$ are order parameters, and

$$
\begin{aligned}
f = {}& \frac{1}{2} \sum_{\nu\rho} \left(m^\rho_{1\nu} \right)^2 + \frac{\alpha}{2\beta} \mathrm{Tr} \log\{ \delta_{\nu\omega} \delta^{\rho\sigma} - \beta\Gamma^2 (1 - \kappa^2) \delta_{\nu\omega} + \kappa^2] q^{\rho\sigma} \} + \frac{\beta\alpha}{2} \sum_{\rho\sigma} q^{\rho\sigma} r^{\rho\sigma} \\
& - \frac{1}{\beta} \left\langle \log \mathrm{Tr}_S \exp\left\{ \beta \left[\sum_{\nu\rho} m^\rho_{1\nu} \chi_{1\nu} S^\rho - \left(\theta + \frac{\alpha s \Gamma^2}{2} \right) \sum_\rho S^\rho + \frac{\beta\alpha}{2} \sum_{\rho\sigma} r^{\rho\sigma} S^\rho S^\sigma \right] \right\} \right\rangle. \quad (9)
\end{aligned}
$$

δ is the Kronecker delta and

$$\Gamma^2 \equiv (1 - 2\gamma)^2 a(1 - a) + \gamma^2,$$

$$\kappa^2 \equiv \frac{\gamma^2 c^2}{(1 - 2\gamma)^2 a(1 - a) + \gamma^2}. \tag{10}$$

Equation (9) assumes a successful retrieval regime in which the network overlaps significantly with either one sparse example η_{11} or dense, correlated examples $\zeta_{1\nu}$ of one concept. We capture these two possibilities by introducing $\chi_{1\nu}$, where $\chi_{1\nu}^i = \eta_{11}^i \delta_{1\nu}$ or $\zeta_{1\nu}^i$ respectively for retrieval of sparse or dense patterns. Through self-averaging, we have replaced averages over neurons i with averages over entries $\chi_{1\nu}$ at a single neuron. Thus, the index i no longer appears in Eq. (9).

Then, we use the replica symmetry ansatz and saddle-point method to obtain the following mean-field equations in terms of the replica-symmetric order parameters $m_{1\nu}$, r, and Q:

$$m_{1\nu} = \langle\!\langle \chi_{1\nu} \, \text{sig}[\beta h] \rangle\!\rangle,$$

$$r = s\Gamma^4 \frac{[1 - Q(1 - \kappa^2)(1 + s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1 - \kappa^2)]^2[1 - Q(1 + s_0\kappa^2)]^2} \langle\!\langle \text{sig}[\beta h]^2 \rangle\!\rangle,$$

$$Q = \beta\Gamma^2 \langle\!\langle \text{sig}[\beta h]^2 - \text{sig}[\beta h] \rangle\!\rangle, \tag{11}$$

where the double angle brackets indicate averages over $\chi_{1\nu}$ and z, an auxiliary random field with a standard normal distribution. Meanwhile, $s_0 \equiv s - 1$, the sigmoid or logistic function $\text{sig}(x) \equiv 1/(1 + e^{-x})$,

$$h \equiv \sum_\nu m_{1\nu}\chi_{1\nu} - \phi + \sqrt{\alpha r}z,$$

$$\phi \equiv \theta - \frac{Q\alpha s\Gamma^2}{2} \frac{1 + s_0\kappa^4 - Q(1 - \kappa^2)(1 + s_0\kappa^2)}{[1 - Q(1 - \kappa^2)][1 - Q(1 + s_0\kappa^2)]}. \tag{12}$$

As derived in Appendix A, $m_{1\nu}$'s are network overlaps with the target pattern and other patterns correlated with it, r represents noise due to overlap with off-target patterns, and Q is related to the overall neural activity. h is the local field in the mean-field limit, which encapsulates the mean network interaction experienced by each neuron. ϕ is the shifted threshold, which is empirically very similar to the original threshold θ.

Equation (11) applies to all target pattern types $\chi_{1\nu}$ that we wish to recover. We now simplify the mean-field equations for either sparse targets with $\chi_{1\nu} = \eta_{11}\delta_{1\nu}$ or dense patterns with $\chi_{1\nu} = \zeta_{1\nu}$. In the latter case, we will perform further simplifications corresponding to recovery of either one dense example ζ_{11} or one dense concept ζ_1, in which case the network overlaps equally with all dense examples $\zeta_{1\nu}$ belonging to it. We also take the $T \to 0$ limit, which implies a strict threshold without stochastic activation. The full derivations are provided in Appendixes A, B, C, and D, but the results for each target type are provided below.

(1) Sparse example η_{11}: Eq. (11) becomes

$$m_{11} = \frac{(1 - 2\gamma)a}{2}\left[\text{erf}\,\frac{\phi}{\sqrt{2\alpha r}} + \text{erf}\,\frac{(1 - 2\gamma)m_{11} - \phi}{\sqrt{2\alpha r}}\right],$$

$$r = \frac{s(1 + s_0\kappa^4)\Gamma^4}{2}$$

$$\times\left[1 - \text{erf}\,\frac{\phi}{\sqrt{2\alpha r}} + a\,\text{erf}\,\frac{(1 - 2\gamma)m_{11} - \phi}{\sqrt{2\alpha r}}\right]. \tag{13}$$

(2) Dense example ζ_{11}: If we call $m_0 \equiv m_{1\nu}$ the overlap with other dense examples $\nu > 1$ of the same concept, Eq. (11) becomes

$$m_{11} = \frac{\gamma}{2}\left[\frac{1 + c}{4}(\text{erf}\,Y_{++} + \text{erf}\,Y_{+-}) + \frac{1 - c}{4}(\text{erf}\,Y_{-+} + \text{erf}\,Y_{--})\right],$$

$$m_0 = \frac{\gamma c}{2[1 - Q\frac{\gamma^2}{\Gamma^2}(1 - c^2)]}\left[\frac{1 + c}{4}(\text{erf}\,Y_{++} + \text{erf}\,Y_{+-}) - \frac{1 - c}{4}(\text{erf}\,Y_{-+} + \text{erf}\,Y_{--})\right],$$

$$r = \frac{s\Gamma^4}{2}\frac{[1 - Q(1 - \kappa^2)(1 + s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1 - \kappa^2)]^2[1 - Q(1 + s_0\kappa^2)]^2}\left[1 - \frac{1 + c}{4}(\text{erf}\,Y_{++} - \text{erf}\,Y_{+-}) - \frac{1 - c}{4}(\text{erf}\,Y_{-+} - \text{erf}\,Y_{--})\right],$$

$$Q = \frac{\Gamma^2}{\sqrt{2\pi}\sigma_0}\left[\frac{1 + c}{4}\left(e^{-Y_{++}^2} + e^{-Y_{+-}^2}\right) + \frac{1 - c}{4}\left(e^{-Y_{-+}^2} + e^{-Y_{--}^2}\right)\right], \tag{14}$$

where

$$\sigma_0^2 \equiv s_0\gamma^2(1 - c^2)m_0^2 + \alpha r,$$

$$Y_{\pm\pm} \equiv \frac{\gamma m_{11} \pm s_0\gamma cm_0 \pm \phi}{\sqrt{2}\sigma_0}. \tag{15}$$

Sign choices in $Y_{\pm\pm}$ correspond to respective signs on the right-hand side of the equation.

(3) Dense concept ζ_1: If we call m_1 the overlap with the target dense concept and $m_s \equiv m_{1\nu}$ the overlap with all of its dense examples ν, Eq. (11) becomes

$$m_1 = \frac{\gamma}{4}(\text{erf}\,Y_+ + \text{erf}\,Y_-),$$

$$m_s = \frac{\gamma c}{4[1 - Q\frac{\gamma^2}{\Gamma^2}(1 - c^2)]}(\text{erf}\,Y_+ + \text{erf}\,Y_-),$$

$$r = \frac{s\Gamma^4}{2}\frac{[1 - Q(1 - \kappa^2)(1 + s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1 - \kappa^2)]^2[1 - Q(1 + s_0\kappa^2)]^2}$$

$$\times\left[1 - \frac{1}{2}(\text{erf}\,Y_+ - \text{erf}\,Y_-)\right],$$

$$Q = \frac{\Gamma^2}{\sqrt{8\pi}\sigma_s}\left(e^{-Y_+^2} + e^{-Y_-^2}\right), \tag{16}$$

where

$$\sigma_s^2 \equiv s\gamma^2(1-c^2)m_s^2 + \alpha r$$

$$Y_{\pm} \equiv \frac{s\gamma c m_s \pm \phi}{\sqrt{2}\sigma_s}. \tag{17}$$

The sign choice in $Y_{\pm}$ corresponds to the sign on the right-hand side of the equation.

III. $T = 0$ CAPACITIES

A. Retrieval regimes

Large values for the overlaps m_{11} and m_1 in Eqs. (13), (14), and (16) signal that retrieval of target patterns is possible. To be more precise, we derive in Appendix A that for $T = 0$,

$$m_{1\nu} = \langle \chi_{1\nu} S \rangle, \tag{18}$$

where $\chi_{1\nu}$ and S are, respectively, the pattern entry and activity for a single neuron and angle brackets indicate an average over $\chi_{1\nu}$. Again, the neuron index i does not appear due to self-averaging. Successful retrieval means that the network activity $\mathbf{S}$ is similar to the original, unscaled patterns $\boldsymbol{\xi}_{11}$, $\boldsymbol{\psi}_{11}$, and $\boldsymbol{\psi}_1$ with 0/1 entries. With the rescalings in Eq. (6), this condition implies $m_{11} \sim (1-2\gamma)a(1-a)$ for sparse example targets, $m_{11} \sim \gamma/2$ for dense example targets, and $m_1 \sim \gamma/2$ for dense concept targets. For ease of comparison, we define a rescaled overlap

$$m' = \begin{cases} m_{11}/(1-2\gamma)a(1-a) & \text{sparse example,} \\ m_{11}/(\gamma/2) & \text{dense example,} \\ m_1/(\gamma/2) & \text{dense concept,} \end{cases} \tag{19}$$

so $m' \sim 1$ corresponds to the retrieval phase, as an order-of-magnitude estimate.

To determine the extent of retrieval phase, we numerically solve the mean-field equations for a given set of network parameters. Phase boundaries are found by adjusting the number of examples stored per concept s and looking for the appearance or disappearance of nontrivial solutions. These boundaries will change as a function of the number of concepts per neuron α, the sparse pattern density a, the dense pattern correlation c, and the relative dense storage strength γ. We treat the shifted activity threshold ϕ as a free parameter that can be adjusted to maximize m_{11} and m_1.

Figure 1(a) shows that for a given concept load α, the network can retrieve sparse and dense examples below critical example loads s_c, which we call the capacities. Above the capacities, catastrophic interference between the target and off-target patterns prevents successful retrieval. Figure 1(b) shows that the network can retrieve dense concepts above a critical s_c. Thus, it builds concepts, which are not directly stored, through accumulating shared features among dense examples. With greater correlation c, fewer examples are required to appreciate commonalities, so s_c is lower. Note that for low enough α, the network can recover both sparse examples and dense concepts at intermediate values of s. Thus, our network is capable of retrieving both example and concept representations of the same memories by tuning an activity threshold.

Optimal retrieval of dense patterns occurs at threshold $\phi = 0$ and of sparse patterns at $\phi/(1-2\gamma)^2 a \approx 0.6$. These values which match results for classic Hopfield networks that
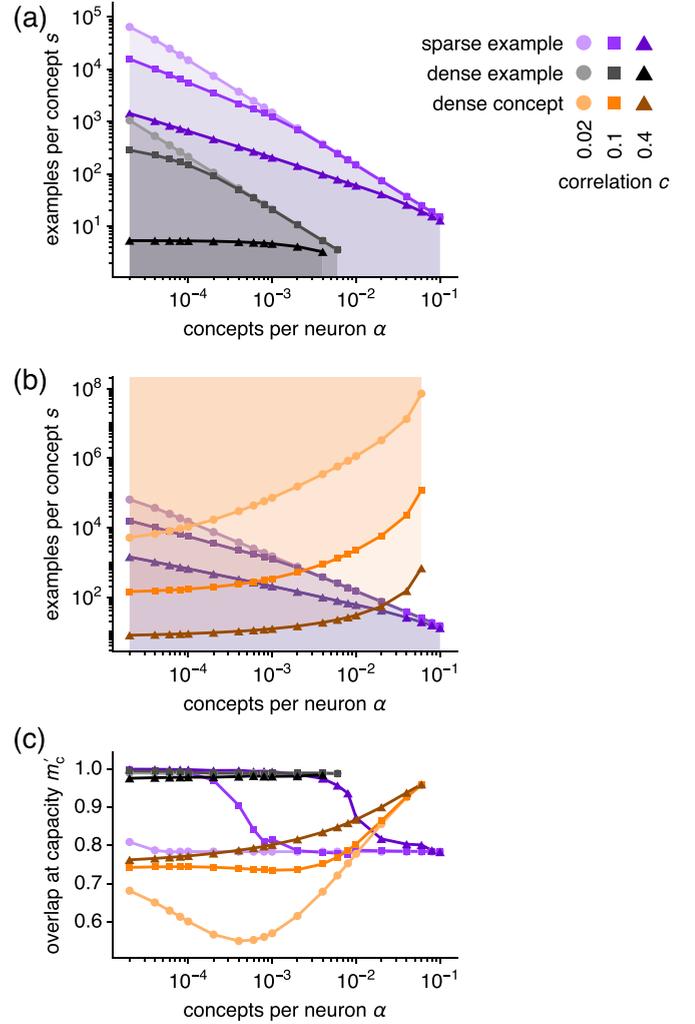
FIG. 1. Retrieval properties for sparse examples, dense examples, and dense concepts. (a), (b) Retrieval regimes (shaded regions) obtained by numerically solving the mean-field equations. Their boundaries correspond to capacities s_c. Sparse patterns are recovered at high threshold and dense patterns at low threshold. (a) More examples can be retrieved sparsely than they can be densely. (b) For small enough concept loads α and intermediate example loads s, both sparse examples and the dense concepts can be retrieved. (c) Network overlap with target patterns at capacity [Eq. (19)]. The curves for the three dense example conditions closely follow one another. Sparse patterns have density $a = 0.01$ and the dense storage strength is $\gamma = 0.1$.

store only dense or only sparse patterns [7,21]. At s_c, the rescaled overlap m'_c takes values above 0.5 over the parameters explored [Fig. 1(c)] before jumping discontinuously to a much lower value immediately outside the retrieval regime. Such a first-order transition has also been observed in classic Hopfield networks [7,13,19].

B. Overview of capacity formulas

We then seek to obtain mathematical formulas for the capacity, or critical example load, s_c of each type of pattern. Not only would these formulas provide a direct way of determining whether pattern retrieval is possible for a given
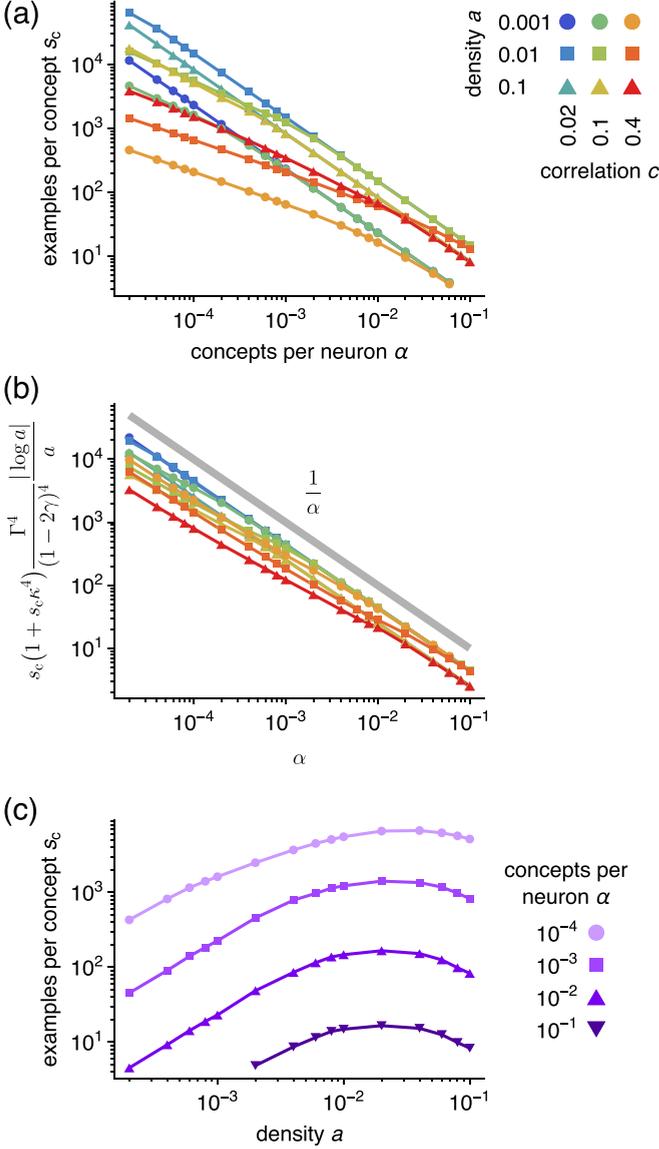
FIG. 2. (a) Capacity s_c for sparse examples. Connected points indicate numerical analysis of Eq. (13). (b) Collapse of s_c curves under rescaled variables. Gray line indicates theoretical formula Eq. (20). (c) s_c is maximized at intermediate values of density a. Dense patterns have correlation $c = 0.1$. The dense storage strength is $\gamma = 0.1$.

set of network parameters, they would offer mathematical insight into network behavior. As detailed in Appendixes B, C, and D, we apply various approximations to the mean-field equations Eqs. (13), (14), and (16) to derive the following formulas for s_c, which match well with numerical solutions over a wide range of parameters (Figs. 2, 3, and 4).

(1) Sparse example $\boldsymbol{\eta}_{11}$ (Fig. 2): The capacity is

$$\frac{1}{\alpha} \sim s_c(1 + s_c\kappa^4)\frac{\Gamma^4}{(1-2\gamma)^4}\frac{|\log a|}{a}, \tag{20}$$

which means that

$$s_c \sim \sqrt{\frac{1}{4\kappa^8} + \frac{(1-2\gamma)^4}{\gamma^4 c^4}\frac{a}{|\log a|}\frac{1}{\alpha}} - \frac{1}{2\kappa^4}. \tag{21}$$
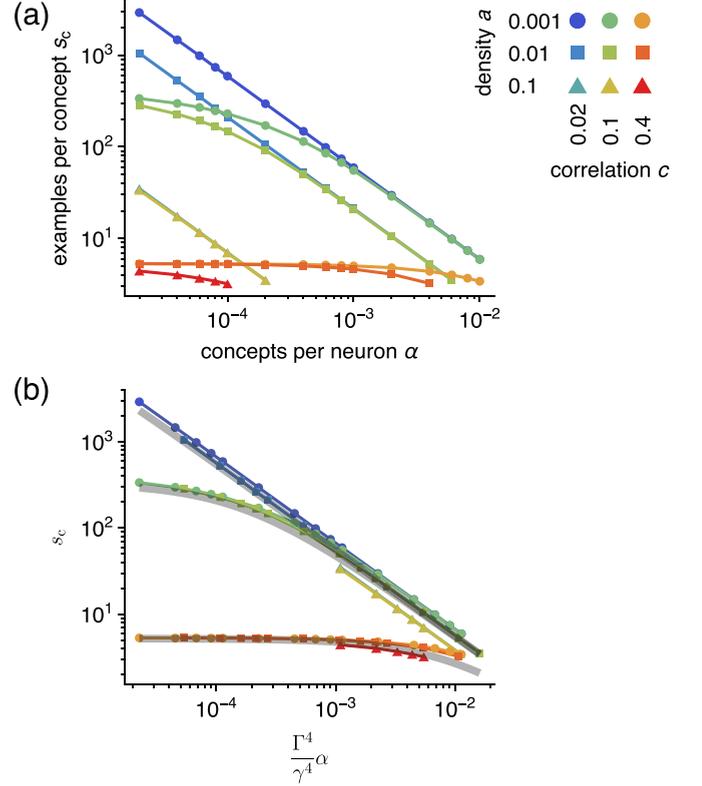
FIG. 3. (a) Capacity s_c for dense examples. Connected points indicate numerical analysis of Eq. (14). (b) Collapse of s_c curves under rescaled variables. Gray lines indicate theoretical formula (23). The dense storage strength is $\gamma = 0.1$.

In sparse Hopfield networks without dense patterns, the capacity always increases for sparser patterns [7]. In contrast, our capacity for sparse examples peaks at intermediate densities a [Fig. 2(c)]. While sparser patterns interfere less with one another, their smaller basins of attraction are more easily overwhelmed by those of dense patterns, whose density is always 0.5. We can quantitatively understand the tradeoff between these two factors in the $c^2 \to 0$ limit, where Eq. (20) becomes

$$\alpha s_c \sim \frac{a}{(a + a_d)^2|\log a|} \tag{22}$$

for $a_d \equiv \gamma^2/(1 - 2\gamma)^2$. a_d represents interference from dense patterns and acts as the crossover point in the tradeoff. For $a \gg a_d$, $\alpha s_c \sim 1/a|\log a|$, recovering the classic sparse Hopfield scaling in which sparser patterns exhibit higher capacity [7]. However, for $a \ll a_d$, a_d dominates the denominator and $\alpha s_c \sim a/|\log a|$, disfavoring sparser patterns. If we ignore the slowly varying logarithm in Eq. (22), s_c is exactly maximized at $a = a_d$. Using the value $\gamma = 0.1$ in Fig. 2(c), $a_d \approx 0.016$, which agrees well with the numerically obtained maxima.

(2) Dense example $\boldsymbol{\zeta}_{11}$ (Fig. 3): The capacity is

$$s_c \sim \frac{1}{3c^3 + 18\frac{\Gamma^4}{\gamma^4}\alpha}. \tag{23}$$

At large α, this critical number of examples per concept s_c is inversely proportional to the number of concepts per neuron α, indicating that the total number of examples stored per
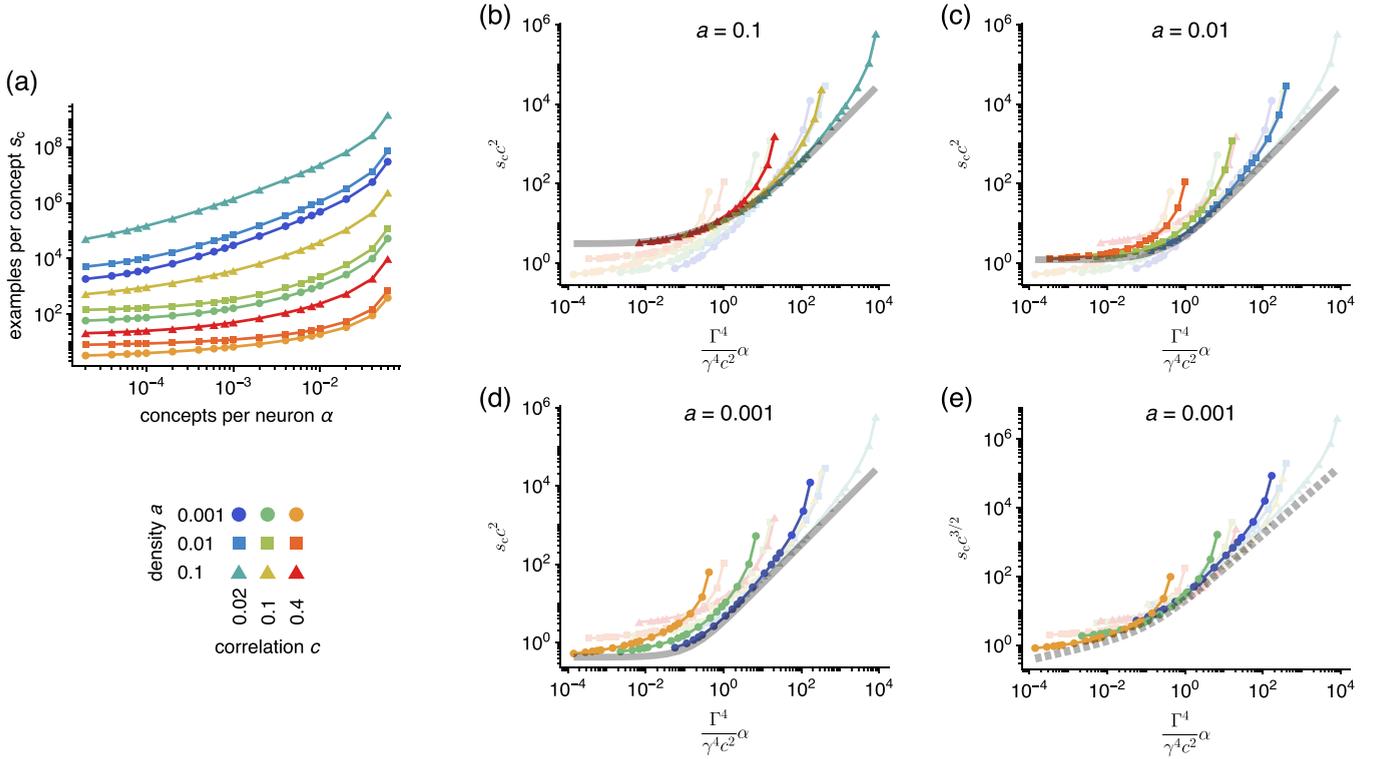
FIG. 4. (a) Capacity, or critical example load, s_c for dense concepts. Connected points indicate numerical analysis of Eq. (16). (b)–(d) Approximate collapse of s_c curves under rescaled variables. Gray solid lines indicate theoretical formula Eq. (24). (e) For the sparsest patterns, s_c curves exhibit better collapse under the rescaling $s_c c^{3/2}$ compared to $s_c c^2$ in (d). Gray dotted line indicates the theoretical formula (25), which better matches the numerical results. It exhibits weak dependence on dense correlation c, and we show its behavior only for $c = 0.02$. The dense storage strength is $\gamma = 0.1$.

neuron αs_c saturates at a constant value. When examples are distributed into many concepts, concept identity becomes insignificant, so only the total number of stored patterns matters. At small α, s_c itself saturates at a constant value determined by the dense correlation c. When concepts are few, interference with other concepts becomes less important than interference within the same concept, so only the number of stored patterns per concept matters.

(3) Dense concept $\boldsymbol{\zeta}_1$ (Fig. 4): There are two cases. For larger densities a, the critical example load approximately collapses as a function of $s_c c^2$ [Figs. 4(b) and 4(c)]. This function can be obtained by numerically inverting the following first equation for y and substituting it into the second:

$$\frac{2\Gamma^4}{\gamma^4 c^2}\alpha \approx \frac{(1-2\gamma)^2 a(1-a)}{\gamma^2} \frac{\left(\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^3}{y^2\left(\text{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)}$$

$$- \left(\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2,$$

$$s_c c^2 \approx \frac{(1-2\gamma)^2 a(1-a)}{\gamma^2} \frac{\sqrt{\frac{2}{\pi}}ye^{-y^2/2}}{\text{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}}. \quad (24)$$

The solution is unique for any parameter values because the right-hand side of the first equation always monotonically decreases as a function of y over its positive range. For smaller

a, the critical example load does not collapse so tightly as a function of $s_c c^2$ for different values of c [Fig. 4(d)]. We calculate that it instead approximately collapses as a function of $s_c c^{3/2}$ [Fig. 4(e)]:

$$s_c c^{3/2} \approx 3\left(\frac{3\pi}{4}\right)^{1/4}\left(\frac{\Gamma^4}{\gamma^4 c^2}\alpha\right)^{1/4} + \frac{3\pi}{4}c^{-1/2}\left(\frac{\Gamma^4}{\gamma^4 c^2}\alpha\right). \quad (25)$$

The second term contains a factor of $c^{-1/2}$, which changes relatively slowly compared to the other powers of c found in the rescaled concept load $\alpha\Gamma^4/\gamma^4 c^2$. The two terms capture the behavior of s_c at low and high rescaled concept load, respectively. Nevertheless, more universal scaling relationships have yet to be found for the dense concept s_c, indicating that many network features may independently govern concept building.

C. Capacities of simulated networks

We perform simulations to verify our capacity calculations. For each simulation condition, we construct replicate networks that store different randomly generated patterns. When generating sparse patterns of density a, we fix the number of active neurons to Na to reduce finite-size effects. Neural dynamics proceed asynchronously in cycles wherein every neuron is updated once in random order. We use $N = 10000$ neurons and dense strength $\gamma = 0.1$, unless otherwise noted. Retrieval is assessed by the following definition of overlap

between network activity **S** and the unscaled target pattern $\boldsymbol{\omega}$, which is a sparse example $\boldsymbol{\xi}_{\mu\nu}$, a dense concept $\boldsymbol{\psi}_{\mu\nu}$, or a dense concept $\boldsymbol{\psi}_{\mu}$:

$$\hat{m} = \frac{1}{Na_\omega(1-a_\omega)} \sum_i (\omega_i - a_\omega)S_i, \qquad (26)$$

where $a_\omega = a$ for sparse patterns and $a_\omega = 1/2$ for dense patterns. Based on Eqs. (1), (2), and (3), we expect $\hat{m} \approx 1$ to indicate successful retrieval. For random activity, $\hat{m} \approx 0$. This overlap $\hat{m}$ is similar to m' in Eq. (19), which concerned the *scaled* target patterns $\boldsymbol{\chi} = \boldsymbol{\eta}_{11}, \boldsymbol{\zeta}_{11}$, and $\boldsymbol{\zeta}_1$.

Capacities are assessed by using the true target patterns as cues; in other words, our simulations probe the stability of the target patterns. For sparse examples, we optimize over the threshold θ by numerical search. For dense patterns, we use $\theta = 0$. We use $\beta \to \infty$ in Eq. (7) because our theoretical calculations were performed for $T \to 0$. We define successful retrieval as $\hat{m} > (1 + \hat{m}_0)/2$, where $\hat{m}_0$ is the overlap expected for off-target patterns within the same concept. Using Eq. (4), $\hat{m}_0 = 0$ for sparse examples, $\hat{m}_0 = c^2$ for dense examples, and $\hat{m}_0 = c$ for dense concepts.

Figure 5 reveals good agreement between simulations and numerical analysis of the mean-field equations for capacities of all target types. This supports the validity of our derivations and the simplifications we invoked to perform them.

IV. HETEROASSOCIATION

A. Performance of simulated networks

Our network stores linear combinations of sparse and dense patterns, and its connectivity matrix contains interactions between the two [Eq. (5)]. Thus, we suspect that in addition to autoassociation for each target type, it can perform heteroassociation between them. We run simulations to test this ability. We use $p = 10$ concepts and store either $s = 20$ examples per concept during retrieval of sparse examples $\boldsymbol{\xi}_{\mu\nu}$ and dense concepts $\boldsymbol{\psi}_{\mu}$ or $s = 3$ during retrieval of dense examples $\boldsymbol{\psi}_{\mu\nu}$. Sparse patterns have density $a = 0.01$ and dense patterns have correlation parameter $c = 0.4$. We initialize the network state to a noisy version of a sparse example, dense example, or dense concept, and attempt to retrieve each type as the target pattern. We create these noisy cues by randomly flipping a fraction 0.01 of the cue pattern between inactive and active. We then asynchronously evolve the network similarly as in the previous section. With theoretical motivation in Appendix B, we define the rescaled parameters

$$\theta' = \theta/(1-2\gamma)^2 a \quad \text{and} \quad \beta' = \beta(1-2\gamma)^2 a, \qquad (27)$$

with rescaled temperature $T' = 1/\beta'$. To retrieve sparse examples, we apply a threshold $\theta' = 0.6$, and to retrieve dense examples and concepts, we apply $\theta' = 0$; these thresholds are immediately applied from the start. We use inverse temperature $\beta' = 50$. Finally, we assess the overlap between the final network activity and the target pattern. If concepts are used as cues and examples are desired as targets, the highest overlap with any example within the cued concept is reported. Successful retrieval is defined via the overlap $\hat{m}$ as described above [Eq. (26)].
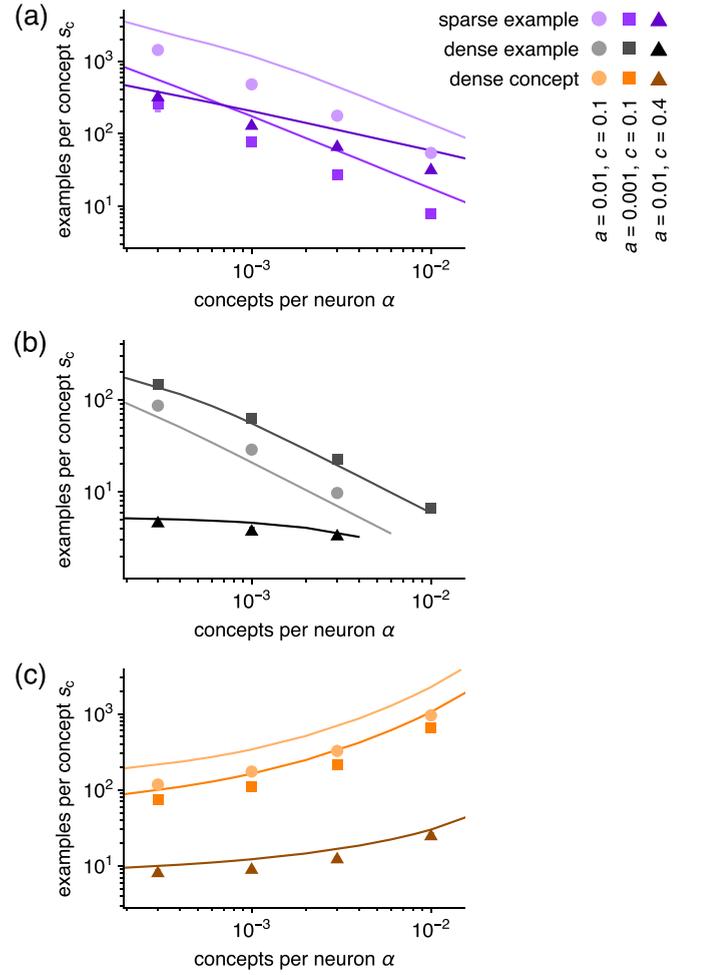
FIG. 5. Capacities s_c for (a) sparse examples, (b) dense examples, and (c) dense concepts obtained by numerical calculations (lines) and simulations (points). Lines indicate analysis of the mean-field equations (13), (14), and (16). Points indicate means over eight replicate simulated networks, and vertical bars indicate standard deviations which are often obscured by the points. In each replicate network, 20 cues are tested with simulations lasting 10 update cycles. The dense storage strength is $\gamma = 0.1$.

Figure 6(a) shows that the network is generally capable of heteroassociation using the parameters described above, which define the baseline condition. By increasing the number of concepts, heteroassociative performance is largely preserved, but note that the retrieval of dense concepts from sparse examples is impaired [Fig. 6(b)]. We next amplify noise by either raising the temperature, which introduces more randomness during retrieval, or randomly flipping more neurons during cue generation. Sparse example targets are more robust than dense example targets with respect to higher temperature [Fig. 6(c)]; meanwhile, dense example cues are more robust than sparse example cues with respect to cue corruption [Fig. 6(d)]. These observations encompass autoassociation as well as heteroassociation. Thus, the dual encoding of memories not only allows for retrieval of both examples and concepts, as noted in Fig. 1(b), but it also mitigates the impact of noise since sparse and dense patterns are more robust to retrieval and cue noise, respectively.
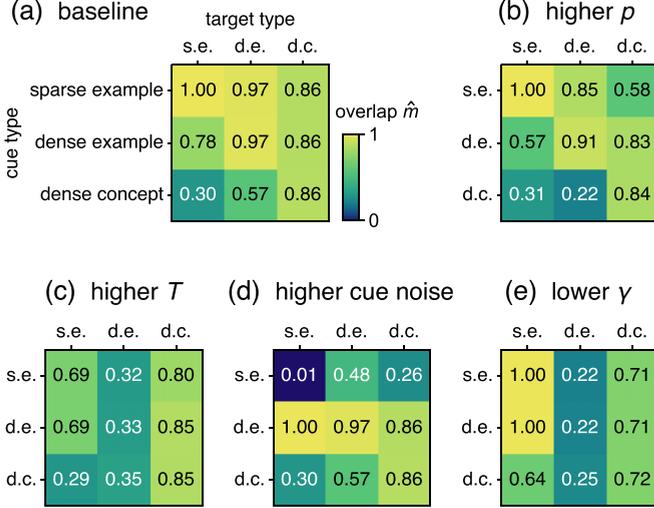
FIG. 6. Auto- and heteroassociation among sparse and dense patterns demonstrated by network simulations. (a) Baseline to which various conditions are compared. (b) The number of concepts is increased from $p = 5$ to 30. (c) The rescaled temperature is increased from $T' = 1/50$ to $1/5$. (d) The fraction of the cue pattern flipped is increased from 0.01 to 0.2. (e) The dense pattern storage strength is decreased from $\gamma = 0.1$ to 0.055. For dense example and concept targets, we use rescaled threshold $\theta' = 0$. For sparse example targets, we use $\theta' = 0.6$. Overlaps $\hat{m}$ reported are averages over eight replicate networks, with one corresponding to perfect retrieval and 0 corresponding to random activity. In each, 20 cues are tested with simulations lasting 20 update cycles.

B. Bidirectional heteroassociation and γ

Notice in Fig. 6(a) that while dense concept targets can be retrieved from sparse example cues, the reverse is not possible. The ability to perform bidirectional heteroassociation between a concept and its examples is of computational significance, so we seek to find network parameters that achieve it. Intuitively, lowering the storage strength of dense patterns γ should bias the network towards retrieving sparse patterns. Indeed, doing so improves retrieval of sparse examples from dense concepts [Fig. 6(e)]. Moreover, the network is still capable of the reverse process, albeit with some decrease in performance.

The value of γ appears critical to the ability to retrieve sparse examples from dense concepts. We hypothesize that this connection is mediated by the relative energy of different pattern types. As described in Appendix A, the Hamiltonian of our network is

$$H = -\frac{1}{2N} \sum_{\mu\nu} \sum_{i \neq j} \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i \right)\left(\eta_{\mu\nu}^j + \zeta_{\mu\nu}^j \right) S_i S_j + \theta \sum_i S_i,$$

(28)

where, again, $\boldsymbol{\eta}_{\mu\nu}$ and $\boldsymbol{\zeta}_{\mu\nu}$ are rescalings of sparse examples $\boldsymbol{\xi}_{\mu\nu}$ and dense examples $\boldsymbol{\psi}_{\mu\nu}$ [Eq. (6)]. We set the network activity $\mathbf{S}$ to a sparse example $\boldsymbol{\xi}_{\mu\nu}$ or dense concept $\boldsymbol{\psi}_\mu$ and calculate the average over patterns $\langle H \rangle$. Using Eq. (4), we
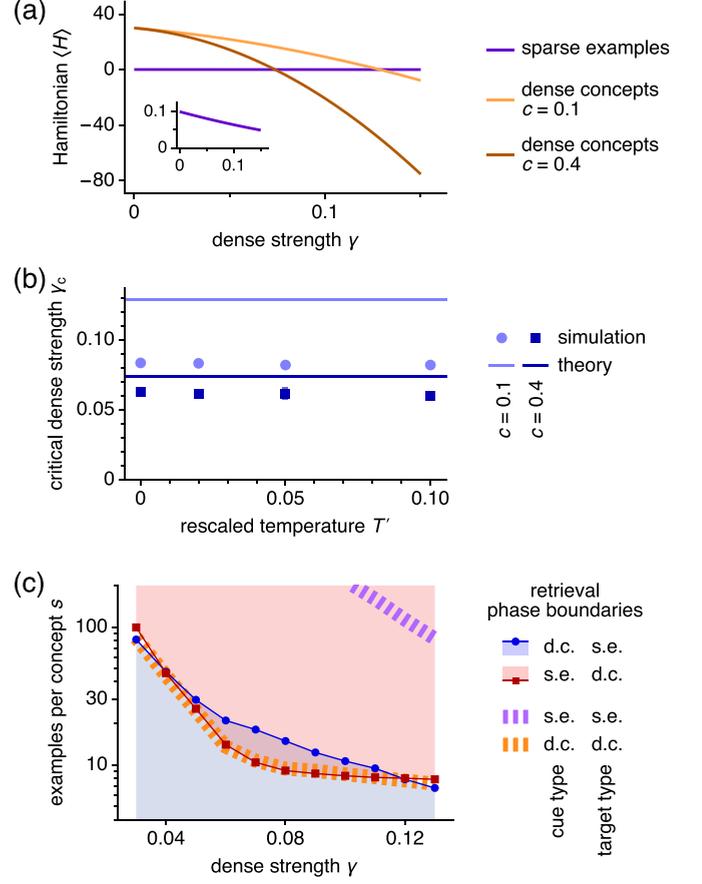
FIG. 7. The dense pattern storage strength γ controls the ability to retrieve sparse examples from dense concepts by changing their relative energies. (a) Hamiltonian energies for rescaled threshold $\theta' = 0.6$ and network size $N = 10000$ [Eq. (29)]. For $c = 0.1$, we store $s = 80$ patterns per concept, and for $c = 0.4$, $s = 20$. Inset shows sparse example energy in detail. (b) Critical dense strength γ_c below which sparse examples can be retrieved by dense concept cues. Theoretical predictions are the locations of energy crossovers in (a). (c) Phase diagram for auto- and heteroassociation among sparse examples and dense concepts in simulated networks. Dense patterns have correlation parameter $c = 0.4$, and the temperature is $T = 0$. Blue and red shaded regions exhibit unidirectional heteroassociation, and the doubly shaded region exhibits bidirectional heteroassociation. Autoassociation occurs below the purple dashed line and above the orange dashed line; for clarity, these regions are not shaded. We use $p = 5$ concepts, and sparse patterns have density $a = 0.01$. Simulations are performed without cue noise. For dense concept targets, we use $\theta' = 0$, and for sparse example targets, we use $\theta' = 0.6$. Points indicate means over eight replicate networks, and vertical bars indicate standard deviations which are often obscured by the points. In each replicate network, 20 cues are tested with simulations lasting 20 update cycles.

obtain

$$\frac{\langle H \rangle}{N} \approx \begin{cases} -\frac{(1-2\gamma)^2 a^2(1-a)^2}{2} + \theta a & \text{sparse example,} \\ -\frac{s\gamma^2 c^2}{8} + \frac{\theta}{2} & \text{dense concept.} \end{cases}$$

(29)

Figure 7(a) shows Eq. (29) calculated in the retrieval regime for sparse examples with $\theta' = 0.6$. The Hamiltonian for dense concepts decreases with γ and eventually crosses the value

for sparse examples, which remains relatively constant. To connect these results with heteroassociative performance, first consider $c = 0.4$, which is the correlation value used in the simulations in Fig. 6. Recall that baseline networks experience difficulty in retrieving sparse examples from dense concepts [Fig. 6(a)]. These networks have $\gamma = 0.1$, for which dense concepts exhibit lower energy than sparse examples do [Fig. 7(a)], even with the high threshold $\theta' = 0.6$ intended to retrieve the latter. The increase in energy required to proceed from cue to target may explain the failure to perform this heteroassociation. It can be performed for $\gamma = 0.055$ [Fig. 6(e)], and here the energy of dense concepts at high threshold increases above that of sparse examples [Fig. 7(a)]. Thus, the progression from cue to target is energetically favored.

The crossover point γ_c between the high-threshold energies of dense concepts and sparse examples appears to define the phase boundary for heteroassociation from the former to the latter. To test this prediction, we evaluate simulated networks at varying values of γ. Successful retrieval of sparse examples is assessed through the overlap $\hat{m}$ with the same cutoff values as described above [Eq. (26)]. Figure 7(b) demonstrates that the energy crossover indeed predicts γ_c for $c = 0.4$. The $c = 0.1$ case shows lower quantitative agreement between simulation and theory, although the qualitative observation of a higher γ_c is captured. Finite-size effects, higher energies of intermediate states along possible transition paths, and trapping in local energy minima may account for the discrepancy. For $T > 0$, the disregard of entropic contributions in our Hamiltonian analysis may also contribute to the disparity, although the lack of significant temperature dependence in our simulations makes this consideration less important [Fig. 7(b)].

For the $c = 0.4$ and $T = 0$ case, we construct a heteroassociation phase diagram by simulating networks with various dense strengths γ and example loads s [Fig. 7(c)]. At intermediate values of γ and s, there is a regime for successful bidirectional heteroassociation between sparse examples and dense concepts. At lower values of either γ or s, only unidirectional heteroassociation from dense concept cues to sparse example targets is possible, and at higher values, only the reverse unidirectional heteroassociation is possible. For comparison, autoassociation capacities for sparse examples and dense concepts are also shown. The phase boundary for retrieving sparse examples is much higher with identical cues than with dense concept cues, reflecting our observations that even below capacity, this heteroassociation direction is granted only for certain γ. In contrast, the phase boundary for retrieving dense concepts is similar with either type of cue, indicating an easier heteroassociation direction.

Due to the importance of γ, we present additional mean-field capacity results in which it is systematically varied (Fig. 8). For low density a, there is a range of intermediate γ and s in which both sparse examples and dense concepts are stable [Fig. 8(b)]. Figures 8(c)–8(h) illustrate that our theoretical capacity formulas are still valid as functions over γ.

V. DISCUSSION

In summary, we present a Hopfield-like network that stores memories as both sparse patterns with low correlation and dense patterns with high correlation. By adjusting the activity threshold, the network can retrieve patterns of either sparsity. The capacity for sparse patterns is large, so many distinct memories can be retrieved. In contrast, as more dense patterns are stored, they merge according to their correlation structure such that concepts are built through the accumulation of examples. We derive mean-field equations that govern the retrieval of sparse examples, dense examples, and dense concepts, and we calculate capacity formulas for each type of retrieved pattern. We observe that the network can retrieve one type of target pattern from its corresponding cue of a different type, and we explain that regimes of successful heteroassociation can be predicted by the relative energies of cue and target patterns.

Our network offers an alternative paradigm for building memory hierarchies in autoassociative networks. Ultrametric networks have been previously explored as an architecture for storing and retrieving memories at different scales [22–27]. Their structure resembles a tree spanning multiple levels. Each pattern at one level serves as a concept-like trunk from which correlated branches are generated to form the next, more example-like level. While these models are insightful and influential, they possess certain disadvantages that our network can address. They typically use an activity threshold or, equivalently, an external field to move between levels, which is also the case in our work. In one ultrametric model, the field is inhomogeneous and proportional to the pattern retrieved [26]. Our activity threshold is homogeneous and does not require memory of the pattern retrieved, though implementing such a feature may improve retrieval performance. In another hierarchical model, coarser representations are stored more sparsely and retrieved at higher threshold [27]. This arrangement prevents the network from leveraging the higher capacity of sparser patterns to store finer representations, which are more numerous. Moreover, ultrametric Hopfield networks often require complex storage procedures that require *a priori* knowledge of concepts or other examples [23,24,26,27]. They do not permit the unsupervised learning of concepts through the accumulation of examples over time, which is achieved by our simple Hebbian learning rule and strengthens the biological significance of our model. Meanwhile, our model's requirement for sparse, decorrelated patterns in addition to dense, correlated patterns can be implemented by neural circuits which are thought to naturally perform decorrelation through sparsification [6,16,28–33].

While the two pattern types are linearly summed in our model to facilitate mathematical derivations, it is possible to implement nonlinear summation, which may better reflect how inputs are combined in biological neurons [34–36]. In Ref. [16] we have shown through simulations that the central capabilities of this model can be preserved under sublinear and superlinear summation. Reference [36] explores multiplicative integration with the storage of only one pattern per memory formed by the neurons commonly activated through both pathways. Returning to biological motivation for our model, our results offer a mechanistic explanation for how complementary pathways within the hippocampus can underlie its observed ability to recall memories at different resolutions. The hippocampus has long been known to mediate episodic memory, the ability to recall specific, personally
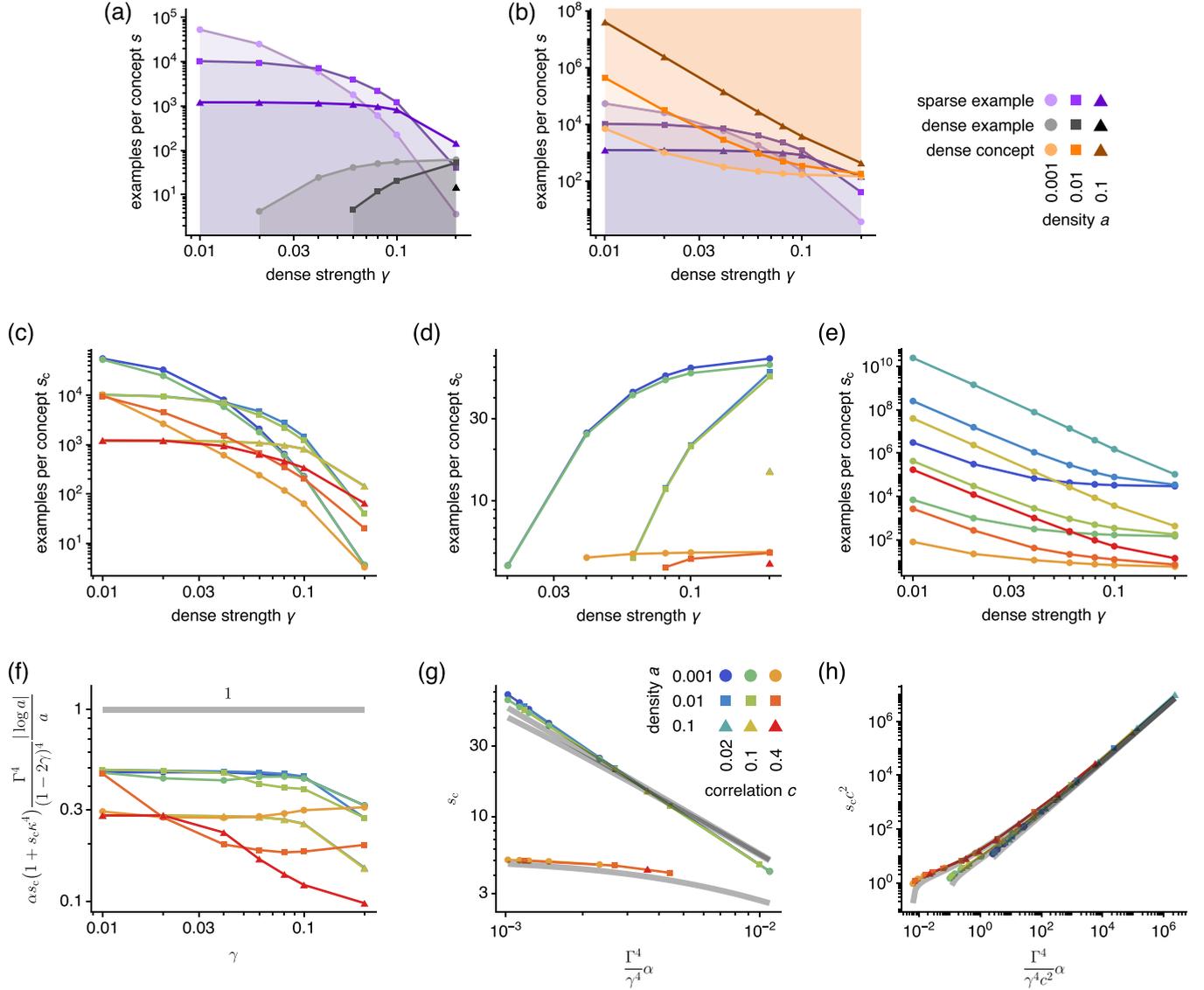
FIG. 8. Capacities s_c as a function of dense pattern storage strength γ. (a), (b) Retrieval regimes (shaded regions) for sparse examples, dense examples, and dense concepts obtained by numerically solving the mean-field equations. Their boundaries correspond to capacities s_c. Dense patterns have correlation parameter $c = 0.1$. Capacities s_c for (c) sparse examples, (d) dense examples, and (e) dense concepts. Collapse of s_c curves for (f) sparse examples, (g) dense examples, and (h) dense concepts under rescaled variables. Gray lines indicate theoretical formulas (20), (23), and (24), respectively. The concept load is $\alpha = 0.001$ concepts per neuron.

experienced events [37,38]. It is thought to be capable of pattern separation, a process that accentuates differences between similar memories [39,40]. Meanwhile, more recent research has uncovered that the hippocampus is also involved in generalizing over episodes through statistical learning [41–45]. The observation of neurons that respond to many representations of a single celebrity or personal acquaintance is one striking case of learning concepts through the accumulation of individual experiences [46,47]. Because of a difference in sparsity, memory types in our model are retrieved at different activity thresholds, which may correspond biologically to different levels of inhibition in CA3. In many mammals, including rodents and primates, the hippocampus exhibits a *theta* oscillation, in which inhibition is modulated with subsecond periodicity [48]. By analyzing neural encoding properties as a function of theta phase in Ref. [16], we indeed find experi-

mental support for the theta oscillation to serve as the activity threshold in our model. The selection between example-like and concept-like representations by theta inhibition has certain computational advantages. A downstream network that serves to integrate information across the two representational scales can access both over subsecond timescales. Meanwhile, the activation of only one encoding at a time may prevent interference or overshadowing between them. In addition, certain tasks may better performed with either more example-like or more concept-like representations. Such preferential recall could be accomplished by adjusting the average inhibitory tone. It is conceivable that our model may apply to brain regions other than the hippocampus if they receive converging inputs with different sparsities and correlation structures. In particular, the *Drosophila* olfactory system and the mammalian cerebellum contain circuits that start at a common

upstream region; branch into two pathways, one of which undergoes decorrelation through sparsification; and converge at a common downstream region. In the former system, the antennal lobe is the upstream region, Kenyon cells perform decorrelation, and the lateral horn is the downstream region [49,50]. In the latter, the three components are mossy fibers (different from the ones in hippocampus), granule cells (also different from the ones in hippocampus), and deep cerebellar nuclei [49,51]. Note that these circuits have an additional population between the decorrelation and downstream regions that is believed to perform rich computations through highly plastic synapses: the mushroom body output neurons in *Drosophila* and the Purkinje cells in the cerebellum [51,52]. More investigation is required to assess the applicability of our model to these systems.

APPENDIX A: MEAN-FIELD EQUATIONS

1. Replica partition function

This derivation of mean-field equations governing the macroscopic behavior of our network is strongly influenced by Refs. [7,13,20]. All of our calculations will be performed in the thermodynamic limit where the network size $N \to \infty$. Our network, presented in Sec. II, is described by a Hamiltonian

$$H = -\frac{1}{2N} \sum_{\mu\nu} \sum_{i \neq j} \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right)\left(\eta_{\mu\nu}^j + \zeta_{\mu\nu}^j\right) S_i S_j + \theta \sum_i S_i$$

$$= -\frac{1}{2N} \sum_{\mu\nu} \left[\sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right) S_i\right]^2$$

$$+ \frac{1}{2N} \sum_{\mu\nu} \sum_i \left[\left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right) S_i\right]^2 + \theta \sum_i S_i. \quad (A1)$$

To reiterate, **S** is the network activity and θ is the activity threshold. $\boldsymbol{\eta}_{\mu\nu}$ and $\boldsymbol{\zeta}_{\mu\nu}$ are rescaled sparse and dense patterns, respectively, for $\nu = 1, \ldots, s$ examples in each of $\mu = 1, \ldots, \alpha N$ concepts [Eq. (6)]. Each rescaled pattern entry is randomly generated as follows:

$$\eta_{\mu\nu}^i = \begin{cases} (1-2\gamma)(1-a) & \text{with probability } a \\ -(1-2\gamma)a & \text{with probability } 1-a \end{cases},$$

$$\zeta_{\mu\nu}^i = \begin{cases} \zeta_\mu^i & \text{with probability } \frac{1+c}{2} \\ -\zeta_\mu^i & \text{with probability } \frac{1-c}{2} \end{cases},$$

$$\zeta_\mu^i = \begin{cases} \gamma & \text{with probability } \frac{1}{2} \\ -\gamma & \text{with probability } \frac{1}{2} \end{cases} \quad (A2)$$

for sparse pattern density a, dense pattern correlation c, and dense pattern storage strength 2γ. Their average values are 0, and the average overlaps between them are also 0 except for

$$\langle \zeta_{\mu\nu}^i \zeta_\mu^i \rangle = \gamma^2 c,$$

$$\langle \zeta_{\mu\nu}^i \zeta_{\mu\omega}^i \rangle = \gamma^2 c^2. \quad (A3)$$

We will forgo introducing external fields. By averaging over examples and concepts,

$$\frac{1}{2N} \sum_{\mu\nu} \sum_i \left[\left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right) S_i\right]^2$$

$$\approx \frac{\alpha s}{2}[(1-2\gamma)^2 a(1-a) + \gamma^2] \sum_i S_i, \quad (A4)$$

If we define

$$\Gamma^2 \equiv (1-2\gamma)^2 a(1-a) + \gamma^2, \quad (A5)$$

we obtain

$$H = -\frac{1}{2N} \sum_{\mu\nu} \left[\sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right) S_i\right]^2 + \left(\theta + \frac{\alpha s \Gamma^2}{2}\right) \sum_i S_i. \quad (A6)$$

To understand this system, we would like to calculate its free energy averaged over instantiations of the patterns: $F = -(1/\beta)\langle \log Z \rangle$. Here Z is the partition function, $\beta = 1/T$ is inverse temperature, and angle brackets indicate averages over $\eta_{\mu\nu}^i$ and $\zeta_{\mu\nu}^i$. Since we cannot directly average over the logarithm of the partition function Z, we use the replica trick by writing formally:

$$\frac{F}{N} = -\frac{1}{\beta N} \langle \log Z \rangle$$

$$= -\frac{1}{\beta N} \lim_{n \to 0} \frac{\langle Z^n \rangle - 1}{n}$$

$$= -\frac{1}{\beta N} \lim_{n \to 0} \frac{1}{n} \log \langle Z^n \rangle. \quad (A7)$$

We interpret Z^n as a partition function for a set of replica networks $\rho = 1, \ldots, n$ with the same parameter values and stored patterns, but the neural activities S_i^ρ may vary across replicas. The Hamiltonian of each replica is

$$H^\rho = -\frac{1}{2N} \sum_{\mu\nu} \left[\sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right) S_i^\rho\right]^2 + \left(\theta + \frac{\alpha s \Gamma^2}{2}\right) \sum_i S_i^\rho, \quad (A8)$$

and the replica partition function, averaged over patterns, is

$$\langle Z^n \rangle = \left\langle \text{Tr}_S \prod_\rho \exp[-\beta H^\rho] \right\rangle. \quad (A9)$$

The trace is evaluated over all neurons i and replicas ρ. We invoke the standard Gaussian integral identity

$$\int dm \, e^{-Am^2 + Bm} = \sqrt{\frac{\pi}{A}} e^{B^2/4A} \quad (A10)$$

to obtain

$$\langle Z^n \rangle = \left\langle \text{Tr}_S \prod_\rho \left\{ \exp\left[-\beta\left(\theta + \frac{\alpha s \Gamma^2}{2} \right) \sum_i S_i^\rho \right] \prod_{\mu\nu} \int dm_{\mu\nu}^\rho \left(\frac{\beta N}{2\pi} \right)^{\frac{1}{2}} \exp\left[-\frac{\beta N}{2}\left(m_{\mu\nu}^\rho \right)^2 + \beta m_{\mu\nu}^\rho \sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i \right) S_i^\rho \right] \right\} \right\rangle. \tag{A11}$$

2. Uncondensed patterns

We search for a retrieval regime in which the network successfully recovers a sparse example $\boldsymbol{\eta}_{11}$, a dense example $\boldsymbol{\zeta}_{11}$, or a dense concept $\boldsymbol{\zeta}_1$. All stored patterns in other concepts $\mu > 1$ are called *uncondensed* and will not significantly overlap with the network activity. We seek to expand in these small overlaps and integrate over them. First,

$$\left\langle \prod_{\substack{\mu>1 \\ \nu\rho}} \exp\left[\beta m_{\mu\nu}^\rho \sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i \right) S_i^\rho \right] \right\rangle = \prod_{\substack{i \\ \mu>1}} \left\langle \prod_\nu \exp\left[\beta\left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i \right) \sum_\rho m_{\mu\nu}^\rho S_i^\rho \right] \right\rangle. \tag{A12}$$

Using

$$Y_\nu \equiv \beta \sum_\rho m_{\mu\nu}^\rho S_i^\rho, \tag{A13}$$

where we have suppressed dependence on μ and i for convenience, we can write

$$\left\langle \prod_\nu \exp\left[(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i) Y_\nu \right] \right\rangle = \left[\prod_\nu \langle \exp\left[\eta_{\mu\nu}^i Y_\nu \right] \rangle \right] \left\langle \prod_\nu \exp\left[\zeta_{\mu\nu}^i Y_\nu \right] \right\rangle. \tag{A14}$$

For uncondensed patterns $\mu > 1$, $m_{\mu\nu}^\rho \ll 1$ because, as we will derive later, it is the overlap between $\mathbf{S}^\rho$ and $\boldsymbol{\eta}_{\mu\nu} + \boldsymbol{\zeta}_{\mu\nu}$. Thus, we can crucially expand in $Y_\nu \ll 1$ and average over the uncondensed patterns:

$$\langle \exp\left[\eta_{\mu\nu}^i Y_\nu \right] \rangle \approx 1 + \frac{1}{2}(1 - 2\gamma)^2 a(1-a) Y_\nu^2 \approx \exp\left[\frac{1}{2}(1-2\gamma)^2 a(1-a) Y_\nu^2 \right]. \tag{A15}$$

Continuing,

$$\left\langle \prod_\nu \exp\left[\zeta_{\mu\nu}^i Y_\nu \right] \right\rangle \approx \frac{1}{2} \prod_\nu \left(1 + \gamma c Y_\nu + \frac{\gamma^2}{2} Y_\nu^2 \right) + \frac{1}{2} \prod_\nu \left(1 - \gamma c Y_\nu + \frac{\gamma^2}{2} Y_\nu^2 \right)$$

$$= 1 + \frac{1}{2} \sum_{\nu\omega} [\gamma^2(1-c^2)\delta_{\nu\omega} + \gamma^2 c^2] Y_\nu Y_\omega$$

$$\approx \exp\left\{ \frac{1}{2} \sum_{\nu\omega} [\gamma^2(1-c^2)\delta_{\nu\omega} + \gamma^2 c^2] Y_\nu Y_\omega \right\}. \tag{A16}$$

Averaging is performed first over ν, then over μ [Eq. (A3)]. Combining the equations above, we obtain

$$\left\langle \prod_{\substack{\mu>1 \\ \nu\rho}} \exp\left[\beta m_{\mu\nu}^\rho \sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i \right) S_i^\rho \right] \right\rangle = \prod_{\mu>1} \exp\left\{ \frac{\beta N}{2} \beta \Gamma^2 \sum_{\nu\omega\rho\sigma} [(1-\kappa^2)\delta_{\nu\omega} + \kappa^2] q^{\rho\sigma} m_{\mu\nu}^\rho m_{\mu\omega}^\sigma \right\} \tag{A17}$$

if we define

$$\kappa^2 \equiv \frac{\gamma^2}{\Gamma^2} c^2 \tag{A18}$$

and enforce

$$q^{\rho\sigma} = \frac{1}{N} \sum_i S_i^\rho S_i^\sigma. \tag{A19}$$

We will do so by introducing the following integrals over delta-function representations:

$$\prod_{\rho\sigma} \int dq^{\rho\sigma} \, \delta\left(q^{\rho\sigma} - \frac{1}{N} \sum_i S_i^\rho S_i^\sigma \right) \propto \int \left(\prod_{\rho\sigma} dq^{\rho\sigma} \, dr^{\rho\sigma} \right) \exp\left(-\frac{\beta^2 \alpha N}{2} \sum_{\rho\sigma} q^{\rho\sigma} r^{\rho\sigma} + \frac{\beta^2 \alpha}{2} \sum_{i\rho\sigma} r^{\rho\sigma} S_i^\rho S_i^\sigma \right), \tag{A20}$$

where $r^{\rho\sigma}$ are additional auxiliary variables whose integration limits extend from $-i\infty$ to $i\infty$, and the factor of $\beta^2 \alpha N/2$ is introduced for later convenience.

We can now integrate over the uncondensed overlaps $m_{\mu\nu}^\rho$:

$$
\left\langle \prod_{\substack{\mu>1 \\ \nu\rho}} \int dm_{\mu\nu}^\rho \left(\frac{\beta N}{2\pi}\right)^{\frac{1}{2}} \exp\left[-\frac{\beta N}{2}\left(m_{\mu\nu}^\rho\right)^2 + \beta m_{\mu\nu}^\rho \sum_i \left(\eta_{\mu\nu}^i + \zeta_{\mu\nu}^i\right)S_i^\rho\right]\right\rangle
$$

$$
\propto \prod_{\mu>1} \int \left[\prod_{\nu\rho} dm_{\mu\nu}^\rho \left(\frac{\beta N}{2\pi}\right)^{\frac{1}{2}}\right] \exp\left(-\frac{\beta N}{2}\sum_{\nu\omega\rho\sigma}\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2]q^{\rho\sigma}\}m_{\mu\nu}^\rho m_{\mu\omega}^\sigma\right)
$$

$$
= \left(\det\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2]q^{\rho\sigma}\}^{-\frac{1}{2}}\right)^{\alpha N - 1}
$$

$$
\approx \exp\left(-\frac{\alpha N}{2}\text{Tr}\log\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2]q^{\rho\sigma}\}\right), \tag{A21}
$$

where $\alpha N \gg 1$ is the total number of concepts.

Thus, so far, our partition function is

$$
\langle Z^n\rangle \propto \int \left[\prod_{\nu\rho} dm_{1\nu}^\rho \left(\frac{\beta N}{2\pi}\right)^{\frac{1}{2}}\right]\left(\prod_{\rho\sigma} dq^{\rho\sigma}\, dr^{\rho\sigma}\right)
$$

$$
\times \exp\left[-\beta N\left(\frac{1}{2}\sum_{\nu\rho}\left(m_{1\nu}^\rho\right)^2 + \frac{\alpha}{2\beta}\text{Tr}\log\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2]q^{\rho\sigma}\} + \frac{\beta\alpha}{2}\sum_{\rho\sigma}q^{\rho\sigma}r^{\rho\sigma}\right)\right]
$$

$$
\times \left\langle\text{Tr}_S \exp\left[\beta\sum_{i\nu\rho}m_{1\nu}^\rho(\eta_{1\nu}^i + \zeta_{1\nu}^i)S_i^\rho - \beta\left(\theta + \frac{\alpha s\Gamma^2}{2}\right)\sum_{i\rho}S_i^\rho + \frac{\beta^2\alpha}{2}\sum_{i\rho\sigma}r^{\rho\sigma}S_i^\rho S_i^\sigma\right]\right\rangle. \tag{A22}
$$

3. Condensed patterns

Now we consider the target patterns, whose large overlaps cannot be expanded into Gaussians and integrated away. When retrieving sparse examples, the network overlaps significantly with one stored pattern $\boldsymbol{\eta}_{11}$, but not $\boldsymbol{\eta}_{1\nu}$ for $\nu > 1$ and $\boldsymbol{\zeta}_{1\nu}$, which are nearly orthogonal to $\boldsymbol{\eta}_{11}$. When retrieving dense examples or concepts, the network overlaps significantly with all stored examples $\boldsymbol{\zeta}_{1\nu}$ within the target concept because they are correlated, but not $\boldsymbol{\eta}_{1\nu}$. Thus, either $\sum_i \eta_{11}^i S_i^\rho$ or $\sum_i \zeta_{1\nu}^i S_i^\rho$ is much larger than the other terms in $\sum_i(\eta_{1\nu}^i + \zeta_{1\nu}^i)S_i^\rho$, so we replace

$$
\sum_{i\nu\rho}m_{1\nu}^\rho\left(\eta_{1\nu}^i + \zeta_{1\nu}^i\right)S_i^\rho \approx \sum_{i\nu\rho}m_{1\nu}^\rho\chi_{1\nu}^i S_i^\rho, \tag{A23}
$$

where $\chi_{1\nu}^i = \eta_{11}^i\delta_{1\nu}$ or $\zeta_{1\nu}^i$ depending on whether we are considering recovery of sparse or dense patterns. These patterns with significant overlaps are called *condensed* patterns.

We now invoke self-averaging over the i indices. For any function $G(\chi^i, S_i)$,

$$
\text{Tr}_S \exp\left[\sum_i G(\chi^i, S_i)\right] = \prod_i \text{Tr}_{S_i} \exp G(\chi^i, S_i) = \exp\left[\sum_i \log \text{Tr}_{S_i} \exp G(\chi^i, S_i)\right] = \exp\left[N\langle\log\text{Tr}_S \exp G(\chi, S)\rangle\right]. \tag{A24}
$$

Now χ and S represent the pattern entry and activity of a single neuron. This single neuron is representative of the entire network because pattern entries are generated independently for each neuron, so we can replace the average over neurons i with an average over possible pattern entries χ. In doing so, we no longer need to pattern-average the trace of the exponential in Eq. (A22); critically, that average has been subsumed by a pattern average inside the exponential, which allows us to write

$$
\langle Z^n\rangle \propto \int \left[\prod_{\nu\rho} dm_{1\nu}^\rho \left(\frac{\beta N}{2\pi}\right)^{\frac{1}{2}}\right]\left(\prod_{\rho\sigma} dq^{\rho\sigma}\, dr^{\rho\sigma}\right)\exp(-\beta N f), \tag{A25}
$$

where

$$
f = \frac{1}{2}\sum_{\nu\rho}\left(m_{1\nu}^\rho\right)^2 + \frac{\alpha}{2\beta}\text{Tr}\log\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2]q^{\rho\sigma}\} + \frac{\beta\alpha}{2}\sum_{\rho\sigma}q^{\rho\sigma}r^{\rho\sigma}
$$

$$
- \frac{1}{\beta}\left\langle\log\text{Tr}_S \exp\left\{\beta\left[\sum_{\nu\rho}m_{1\nu}^\rho\chi_{1\nu}S^\rho - \left(\theta + \frac{\alpha s\Gamma^2}{2}\right)\sum_\rho S^\rho + \frac{\beta\alpha}{2}\sum_{\rho\sigma}r^{\rho\sigma}S^\rho S^\sigma\right]\right\}\right\rangle. \tag{A26}
$$

The replica partition function is now written in a form amenable to the saddle-point approximation. That is, in the $N \to \infty$ limit, we can replace integrals in Eq. (A25) with the integrand evaluated where derivatives of f with respect to the variables of integration equal 0.

4. Saddle-point equations for interpretation

Before proceeding with further simplifying f by invoking replica symmetry, we seek to obtain physical interpretations for m, q, and r, which will serve as the order parameters of our system. To do so, we must recall several previously derived forms of the replica partition function and apply the saddle-point conditions to them.

Recall Eqs. (A17) and (A20) obtained after introducing q and r but before integrating over the uncondensed patterns. Using those expressions in the partition function and performing self-averaging similarly to above, we can obtain

$$\langle Z^n \rangle \propto \int \left[\prod_{\mu\nu\rho} dm_{\mu\nu}^\rho \left(\frac{\beta N}{2\pi} \right)^{\frac{1}{2}} \right] \left(\prod_{\rho\sigma} dq^{\rho\sigma} \, dr^{\rho\sigma} \right) \exp(-\beta N f), \tag{A27}$$

where

$$
f = \frac{1}{2} \sum_{\mu\nu\rho} \left(m_{\mu\nu}^\rho \right)^2 - \frac{\beta \Gamma^2}{2} \sum_{\substack{\mu > 1 \\ \nu\omega\rho\sigma}} [(1-\kappa^2)\delta_{\nu\omega} + \kappa^2] q^{\rho\sigma} m_{\mu\nu}^\rho m_{\mu\omega}^\sigma + \frac{\beta\alpha}{2} \sum_{\rho\sigma} q^{\rho\sigma} r^{\rho\sigma}
$$

$$
- \frac{1}{\beta} \left\langle \log \mathrm{Tr}_S \exp \left\{ \beta \left[\sum_{\nu\rho} m_{1\nu}^\rho \chi_{1\nu} S^\rho - \left(\theta + \frac{\alpha s \Gamma^2}{2} \right) \sum_\rho S^\rho + \frac{\beta\alpha}{2} \sum_{\rho\sigma} r^{\rho\sigma} S^\rho S^\sigma \right] \right\} \right\rangle
$$

$$
= \frac{1}{2} \sum_{\mu\nu\rho} \left(m_{\mu\nu}^\rho \right)^2 - \frac{\beta \Gamma^2}{2} \sum_{\substack{\mu > 1 \\ \nu\omega\rho\sigma}} [(1-\kappa^2)\delta_{\nu\omega} + \kappa^2] q^{\rho\sigma} m_{\mu\nu}^\rho m_{\mu\omega}^\sigma + \frac{\beta\alpha}{2} \sum_{\rho\sigma} q^{\rho\sigma} r^{\rho\sigma} - \frac{1}{\beta} \langle \log \mathrm{Tr}_S \exp[-\beta \mathcal{H}] \rangle \tag{A28}
$$

and

$$
\mathcal{H} \equiv -\sum_{\nu\rho} m_{1\nu}^\rho \chi_{1\nu} S^\rho + \left(\theta + \frac{\alpha s \Gamma^2}{2} \right) \sum_\rho S^\rho - \frac{\beta\alpha}{2} \sum_{\rho\sigma} r^{\rho\sigma} S^\rho S^\sigma \tag{A29}
$$

is the effective single-neuron Hamiltonian across replicas.

At the saddle point, derivatives of f with respect to variables of integration are 0, so

$$
0 = \frac{\partial f}{\partial m_{1\nu}^\rho} = m_{1\nu}^\rho - \left\langle \frac{\mathrm{Tr}_S \chi_{1\nu} S^\rho \exp(-\beta\mathcal{H})}{\mathrm{Tr}_S \exp(-\beta\mathcal{H})} \right\rangle
$$

$$
\Rightarrow m_{1\nu}^\rho = \langle \chi_{1\nu} \overline{S^\rho} \rangle, \tag{A30}
$$

$$
0 = \frac{\partial f}{\partial r^{\rho\sigma}} = \frac{\beta\alpha}{2} q^{\rho\sigma} - \frac{\beta\alpha}{2} \left\langle \frac{\mathrm{Tr}_S S^\rho S^\sigma \exp(-\beta\mathcal{H})}{\mathrm{Tr}_S \exp(-\beta\mathcal{H})} \right\rangle
$$

$$
\Rightarrow q^{\rho\sigma} = \overline{\langle S^\rho S^\sigma \rangle} = \begin{cases} \langle \overline{S^\rho S^\sigma} \rangle & \rho \neq \sigma \\ \langle \overline{S^\rho} \rangle & \rho = \sigma \end{cases}, \tag{A31}
$$

$$
0 = \frac{\partial f}{\partial q^{\rho\sigma}} = \frac{\beta\alpha}{2} r^{\rho\sigma} - \frac{\beta}{2} \sum_{\substack{\mu > 1 \\ \nu\omega}} \Gamma^2 [(1-\kappa^2)\delta_{\nu\omega} + \kappa^2] m_{\mu\nu}^\rho m_{\mu\omega}^\sigma
$$

$$
\Rightarrow r^{\rho\sigma} = \frac{1}{\alpha} \sum_{\substack{\mu > 1 \\ \nu\omega}} \Gamma^2 [(1-\kappa^2)\delta_{\nu\omega} + \kappa^2] m_{\mu\nu}^\rho m_{\mu\omega}^\sigma. \tag{A32}
$$

Bars over variables represent the thermodynamic ensemble average. Thus, $m_{1\nu}^\rho$ is the overlap of the network with the condensed pattern to be recovered, $q^{\rho\sigma}$ is the Edwards-Anderson order parameter reflecting the overall neural activity, and $r^{\rho\sigma}$ represents interference from network overlap with uncondensed patterns $m_{\mu\nu}^\rho$.

To explicitly see that $m_{\mu\nu}^\rho$ describes the overlap of the network with uncondensed patterns for $\mu > 1$, recall Eq. (A11) obtained before introducing q and r. By introducing χ and performing self-averaging similarly to above, we can obtain

$$\langle Z^n \rangle = \int \left[\prod_{\mu\nu\rho} dm_{\mu\nu}^\rho \left(\frac{\beta N}{2\pi} \right)^{\frac{1}{2}} \right] \exp(-\beta N f), \tag{A33}$$

where

$$
f = \frac{1}{2} \sum_{\mu\nu\rho} \left(m_{\mu\nu}^\rho\right)^2 - \frac{1}{\beta} \left\langle \log \mathrm{Tr}_S \exp \left\{ \beta \left[\sum_{\nu\rho} m_{1\nu}^\rho \chi_{1\nu} S^\rho - \left(\theta + \frac{\alpha s \Gamma^2}{2} \right) \sum_\rho S^\rho + \sum_{\substack{\mu > 1 \\ \nu\rho}} m_{\mu\nu}^\rho (\eta_{\mu\nu} + \zeta_{\mu\nu}) S^\rho \right] \right\} \right\rangle
$$

$$
= \frac{1}{2} \sum_{\mu\nu\rho} \left(m_{\mu\nu}^\rho\right)^2 - \frac{1}{\beta} \langle \log \mathrm{Tr}_S \exp[-\beta \mathcal{H}] \rangle, \tag{A34}
$$

and

$$
\mathcal{H} \equiv -\sum_{\nu\rho} m_{1\nu}^\rho \chi_{1\nu} S^\rho + \left(\theta + \frac{\alpha s \Gamma^2}{2} \right) \sum_\rho S^\rho - \sum_{\substack{\mu > 1 \\ \nu\rho}} m_{\mu\nu}^\rho (\eta_{\mu\nu} + \zeta_{\mu\nu}) S^\rho \tag{A35}
$$

is the effective single-neuron Hamiltonian. At the saddle point, this Hamiltonian is equivalent to the form in Eq. (A29) due to Eqs. (A17) and (A32). The saddle-point condition applied to Eqs. (A33) and (A34) yields

$$
0 = \frac{\partial f}{\partial m_{\mu\nu}^\rho} = m_{\mu\nu}^\rho - \left\langle \frac{\mathrm{Tr}_S \, (\eta_{\mu\nu} + \zeta_{\mu\nu}) S^\rho \exp(-\beta\mathcal{H})}{\mathrm{Tr}_S \exp(-\beta\mathcal{H})} \right\rangle
$$

$$
\Rightarrow m_{\mu\nu}^\rho = \langle (\eta_{\mu\nu} + \zeta_{\mu\nu}) \overline{S^\rho} \rangle \quad \text{for } \mu > 1. \tag{A36}
$$

Thus $m_{\mu\nu}^\rho$ is indeed the network overlap with $\eta_{\mu\nu} + \zeta_{\mu\nu}$ for $\mu > 1$. As asserted *ex ante* to derive Eq. (A15), we expect it to be small.

5. Replica-symmetry ansatz

We are now finished with seeking physical interpretations for order parameters, and we return to the primary task of calculating the free energy Eq. (A7) using Eqs. (A25) and (A26). To do so, we assume replica symmetry:

$$
m_{\mu\nu}^\rho = m_{\mu\nu}, \quad q^{\rho\sigma} = q, \quad q^{\rho\rho} = q_0, \quad r^{\rho\sigma} = r, \quad r^{\rho\rho} = r_0. \tag{A37}
$$

Our expression for f then becomes

$$
f = \frac{1}{2} n \sum_\nu (m_{1\nu})^2 + \frac{\alpha}{2\beta} \mathrm{Tr} \log\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2][(q_0 - q)\delta^{\rho\sigma} + q]\} + \frac{\beta\alpha n}{2} q_0 r_0 + \frac{\beta\alpha n(n-1)}{2} qr
$$

$$
- \frac{1}{\beta} \left\langle \log \mathrm{Tr}_S \exp \left\{ \beta \left[\left(\sum_\nu m_{1\nu}\chi_{1\nu} - \theta - \frac{\alpha s \Gamma^2}{2} + \frac{\beta\alpha}{2}(r_0 - r) \right) \sum_\rho S^\rho + \frac{\beta\alpha}{2} r \left(\sum_\rho S^\rho \right)^2 \right] \right\} \right\rangle. \tag{A38}
$$

The eigenvalues of a constant $n \times n$ matrix with entries A are nA with multiplicity 1 and 0 with multiplicity $n - 1$. Thus, the second term in Eq. (A38) under the limit in Eq. (A7) becomes

$$
\lim_{n\to 0} \frac{1}{n} \mathrm{Tr} \log\{\delta_{\nu\omega}\delta^{\rho\sigma} - \beta\Gamma^2[(1-\kappa^2)\delta_{\nu\omega} + \kappa^2][(q_0 - q)\delta^{\rho\sigma} + q]\}
$$

$$
= \lim_{n\to 0} \frac{1}{n} \{\log[1 - \beta\Gamma^2(1 + s\kappa^2 - \kappa^2)(q_0 - q + nq)] + (n-1)\log[1 - \beta\Gamma^2(1 + s\kappa^2 - \kappa^2)(q_0 - q)]
$$

$$
+ (s-1)\log[1 - \beta\Gamma^2(1-\kappa^2)(q_0 - q + nq)] + (s-1)(n-1)\log[1 - \beta\Gamma^2(1-\kappa^2)(q_0 - q)]\}
$$

$$
= (s-1)\left\{ \log[1 - Q(1-\kappa^2)] - \frac{\beta q \Gamma^2(1-\kappa^2)}{1 - Q(1-\kappa^2)} \right\} + \log[1 - Q(1 + s\kappa^2 - \kappa^2)] - \frac{\beta q \Gamma^2(1 + s\kappa^2 - \kappa^2)}{1 - Q(1 + s\kappa^2 - \kappa^2)},
$$

$$
\equiv \Lambda[q, q_0], \tag{A39}
$$

where

$$
Q \equiv \beta(q_0 - q)\Gamma^2. \tag{A40}
$$

To evaluate the last term in Eq. (A38), we can use another Gaussian integral [Eq. (A10)] to perform the trace over S in the limit $n \to 0$:

$$
\left\langle \log \mathrm{Tr}_S \exp \left\{ \beta \left[\left(\sum_\nu m_{1\nu} \chi_{1\nu} - \theta - \frac{\alpha s \Gamma^2}{2} + \frac{\beta \alpha}{2}(r_0 - r) \right) \sum_\rho S^\rho + \frac{\beta \alpha}{2} r \left(\sum_\rho S^\rho \right)^2 \right] \right\} \right\rangle
$$

$$
= \left\langle \log \mathrm{Tr}_S \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \exp \left\{ \beta \left[\left(\sum_\nu m_{1\nu} \chi_{1\nu} - \theta + \frac{\beta \alpha}{2}(r_0 - r) - \frac{\alpha s \Gamma^2}{2} + \sqrt{\alpha r} z \right) \sum_\rho S^\rho \right] \right\} \right\rangle
$$

$$
= \left\langle \log \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left\{ 1 + \exp \left[\beta \left(\sum_\nu m_{1\nu} \chi_{1\nu} - \theta + \frac{\beta \alpha}{2}(r_0 - r) - \frac{\alpha s \Gamma^2}{2} + \sqrt{\alpha r} z \right) \right] \right\}^n \right\rangle
$$

$$
\approx \left\langle \log \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \left(1 + n \log \left\{ 1 + \exp \left[\beta \left(\sum_\nu m_{1\nu} \chi_{1\nu} - \theta + \frac{\beta \alpha}{2}(r_0 - r) - \frac{\alpha s \Gamma^2}{2} + \sqrt{\alpha r} z \right) \right] \right\} \right) \right\rangle
$$

$$
\approx n \left\langle \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \log \left\{ 1 + \exp \left[\beta \left(\sum_\nu m_{1\nu} \chi_{1\nu} - \theta + \frac{\beta \alpha}{2}(r_0 - r) - \frac{\alpha s \Gamma^2}{2} + \sqrt{\alpha r} z \right) \right] \right\} \right\rangle. \tag{A41}
$$

The free energy Eq. (A7) under replica symmetry becomes

$$
\frac{F}{N} = \frac{1}{2} \sum_\nu (m_{1\nu})^2 + \frac{\alpha}{2\beta} \Lambda[q, q_0] + \frac{\beta \alpha}{2}(q_0 r_0 - qr)
$$

$$
- \frac{1}{\beta} \left\langle\!\!\left\langle \log \left\{ 1 + \exp \left[\beta \left(\sum_\nu m_{1\nu} \chi_{1\nu} - \theta + \frac{\beta \alpha}{2}(r_0 - r) - \frac{\alpha s \Gamma^2}{2} + \sqrt{\alpha r} z \right) \right] \right\} \right\rangle\!\!\right\rangle, \tag{A42}
$$

where now the double angle brackets indicate an average over $\chi_{1\nu}$ as well as the Gaussian variable z.

6. Mean-field equations

We can now minimize this free energy over the order parameters by setting derivatives of F to zero, which yields the mean-field equations. This step is equivalent to applying the saddle-point approximation to replica-symmetric f in the $n \to 0$ limit. We first note that

$$
\frac{\partial \Lambda}{\partial q} = (s - 1) \frac{\beta^2 q \Gamma^4 (1 - \kappa^2)^2}{[1 - Q(1 - \kappa^2)]^2} + \frac{\beta^2 q \Gamma^4 (1 + s\kappa^2 - \kappa^2)^2}{[1 - Q(1 + s\kappa^2 - \kappa^2)]^2}. \tag{A43}
$$

The combined fraction has numerator $\beta^2 q \Gamma^4$ multiplied by

$$
(s - 1)\{(1 - \kappa^2)[1 - Q(1 + s\kappa^2 - \kappa^2)]\}^2 + \{(1 + s\kappa^2 - \kappa^2)[1 - Q(1 - \kappa^2)]\}^2
$$

$$
= s[1 - Q(1 - \kappa^2)(1 + s\kappa^2 - \kappa^2)]^2 + s(s - 1)\kappa^4. \tag{A44}
$$

Meanwhile,

$$
\frac{\partial \Lambda}{\partial q_0} = (s - 1) \left[-\frac{\beta \Gamma^2 (1 - \kappa^2)}{1 - Q(1 - \kappa^2)} - \frac{\beta^2 q \Gamma^4 (1 - \kappa^2)^2}{[1 - Q(1 - \kappa^2)]^2} \right] - \frac{\beta (1 + s\kappa^2 - \kappa^2)}{1 - Q(1 + s\kappa^2 - \kappa^2)} - \frac{\beta^2 q \Gamma^4 (1 + s\kappa^2 - \kappa^2)^2}{[1 - Q(1 + s\kappa^2 - \kappa^2)]^2}
$$

$$
= -\frac{\partial \Lambda}{\partial q} - \beta \Gamma^2 \left[\frac{(s - 1)(1 - \kappa^2)}{1 - Q(1 - \kappa^2)} + \frac{1 + s\kappa^2 - \kappa^2}{1 - Q(1 + s\kappa^2 - \kappa^2)} \right]. \tag{A45}
$$

The combined fraction inside the square brackets has numerator

$$
(s - 1)(1 - \kappa^2)[1 - Q(1 + s\kappa^2 - \kappa^2)] + (1 + s\kappa^2 - \kappa^2)[1 - Q(1 - \kappa^2)] = s[1 - Q(1 - \kappa^2)(1 + s\kappa^2 - \kappa^2)]. \tag{A46}
$$

Thus, derivatives of F with respect to the order parameters are

$$
0 = \frac{\partial F}{\partial q} = \frac{\alpha}{2\beta} \left\{ \beta^2 q s \Gamma^4 \frac{[1 - Q(1 - \kappa^2)(1 + s\kappa^2 - \kappa^2)]^2 + (s - 1)\kappa^4}{[1 - Q(1 - \kappa^2)]^2 [1 - Q(1 + s\kappa^2 - \kappa^2)]^2} \right\} - \frac{\beta \alpha}{2} r
$$

$$
\Rightarrow r = q s \Gamma^4 \frac{[1 - Q(1 - \kappa^2)(1 + s_0 \kappa^2)]^2 + s_0 \kappa^4}{[1 - Q(1 - \kappa^2)]^2 [1 - Q(1 + s_0 \kappa^2)]^2}, \tag{A47}
$$

$$0 = \frac{\partial F}{\partial q_0} = \frac{\alpha}{2\beta}\left\{-\frac{\partial \Lambda}{\partial q} - \beta s\Gamma^2 \frac{1 - Q(1-\kappa^2)(1+s\kappa^2-\kappa^2)}{[1-Q(1-\kappa^2)][1-Q(1+s\kappa^2-\kappa^2)]}\right\} + \frac{\beta\alpha}{2}r_0$$

$$\Rightarrow r_0 = r + \frac{s\Gamma^2}{\beta}\frac{1 - Q(1-\kappa^2)(1+s_0\kappa^2)}{[1-Q(1-\kappa^2)][1-Q(1+s_0\kappa^2)]}, \tag{A48}$$

$$0 = \frac{\partial F}{\partial m_{1\nu}} = m_{1\nu} - \langle\langle \chi_{1\nu}\, \text{sig}[\beta h]\rangle\rangle$$

$$\Rightarrow m_{1\nu} = \langle\langle \chi_{1\nu}\, \text{sig}[\beta h]\rangle\rangle, \tag{A49}$$

where $\text{sig}(x) \equiv 1/(1+e^{-x})$ and

$$s_0 \equiv s - 1, \quad h \equiv \sum_\nu m_{1\nu}\chi_{1\nu} - \theta + \frac{\beta\alpha}{2}(r_0 - r) - \frac{\alpha s\Gamma^2}{2} + \sqrt{\alpha r}z. \tag{A50}$$

h is the local field under the mean-field approximation. We can simplify it via

$$\frac{\beta\alpha}{2}(r_0 - r) - \frac{\alpha s\Gamma^2}{2} = \frac{\alpha s\Gamma^2}{2}\frac{1 - Q(1-\kappa^2)(1+s_0\kappa^2) - [1-Q(1-\kappa^2)][1-Q(1+s_0\kappa^2)]}{[1-Q(1-\kappa^2)][1-Q(1+s_0\kappa^2)]}$$

$$= \frac{Q\alpha s\Gamma^2}{2}\frac{1 + s_0\kappa^4 - Q(1-\kappa^2)(1+s_0\kappa^2)}{[1-Q(1-\kappa^2)][1-Q(1+s_0\kappa^2)]}. \tag{A51}$$

Thus,

$$h = \sum_\nu m_{1\nu}\chi_{1\nu} - \phi + \sqrt{\alpha r}z, \quad \phi \equiv \theta - \frac{Q\alpha s\Gamma^2}{2}\frac{1 + s_0\kappa^4 - Q(1-\kappa^2)(1+s_0\kappa^2)}{[1-Q(1-\kappa^2)][1-Q(1+s_0\kappa^2)]} \tag{A52}$$

where ϕ is the shifted threshold; we shall see that in retrieval regimes, it is almost identical to θ.

Continuing, and using the identities $\int dz\, e^{-z^2/2}z f(z) = \int dz\, e^{-z^2/2}\, df(z)/dz$ and $d\,\text{sig}(x)/dx = \text{sig}(x) - \text{sig}(x)^2$,

$$0 = \frac{\partial F}{\partial r} = -\frac{\beta\alpha}{2}q - \frac{\sqrt{\alpha}}{2\sqrt{r}}\langle\langle z\,\text{sig}[\beta h]\rangle\rangle + \frac{\beta\alpha}{2}\langle\langle \text{sig}[\beta h]\rangle\rangle$$

$$\Rightarrow q = \langle\langle \text{sig}[\beta h]^2\rangle\rangle, \tag{A53}$$

$$0 = \frac{\partial F}{\partial r_0} = \frac{\beta\alpha}{2}q_0 - \frac{\beta\alpha}{2}\langle\langle \text{sig}[\beta h]\rangle\rangle$$

$$\Rightarrow q_0 = \langle\langle \text{sig}[\beta h]\rangle\rangle. \tag{A54}$$

Thus, we recover the mean-field equations presented in Eq. (11).

7. Zero-temperature limit

From now on, we consider only the $T = 0$ limit with $\beta \to \infty$. In this limit,

$$\int \frac{dz}{\sqrt{2\pi}}e^{-z^2/2}\,\text{sig}[\beta(Az+B)] \approx \int \frac{dz}{\sqrt{2\pi}}e^{-z^2/2}\,\Theta[Az+B] = \frac{1}{2}\left(1 + \text{erf}\frac{B}{\sqrt{2}A}\right), \tag{A55}$$

where Θ is the Heaviside step function and erf is the error function. Thus, Eqs. (A49), (A53), and (A47) become

$$m_{1\nu} \approx \left\langle\left\langle \chi_{1\nu}\,\Theta\left[\sum_\nu m_{1\nu}\chi_{1\nu} - \phi + \sqrt{\alpha r}z\right]\right\rangle\right\rangle = \frac{1}{2}\left\langle\chi_{1\nu}\,\text{erf}\frac{\sum_\nu m_{1\nu}\chi_{1\nu} - \phi}{\sqrt{2\alpha r}}\right\rangle, \tag{A56}$$

$$q \approx \left\langle\left\langle \Theta\left[\sum_\nu m_{1\nu}\chi_{1\nu} - \phi + \sqrt{\alpha r}z\right]^2\right\rangle\right\rangle = \frac{1}{2}\left[1 + \left\langle\text{erf}\frac{\sum_\nu m_{1\nu}\chi_{1\nu} - \phi}{\sqrt{2\alpha r}}\right\rangle\right], \tag{A57}$$

$$r \approx \frac{s\Gamma^4}{2}\frac{[1 - Q(1-\kappa^2)(1+s_0\kappa^2)]^2 + s_0\kappa^4}{[1-Q(1-\kappa^2)]^2[1-Q(1+s_0\kappa^2)]^2}\left[1 + \left\langle\text{erf}\frac{\sum_\nu m_{1\nu}\chi_{1\nu} - \phi}{\sqrt{2\alpha r}}\right\rangle\right]. \tag{A58}$$

Here single angle brackets again indicate an average over $\chi_{1\nu}$, with the average over z performed. The formula for $m_{1\nu}$ was obtained using $\langle\chi_{1\nu}\rangle = 0$ for both sparse and dense patterns.

Also when $\beta \to \infty$,

$$\int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \{\mathrm{sig}[\beta(Az+B)] - \mathrm{sig}[\beta(Az+B)]^2\}$$

$$= \int \frac{dz}{\sqrt{2\pi}\beta A} e^{-z^2/2} \frac{\partial}{\partial z} \mathrm{sig}[\beta(Az+B)]$$

$$\approx \int \frac{dz}{\sqrt{2\pi}\beta A} e^{-z^2/2} \frac{\partial}{\partial z} \Theta[Az+B] = \int \frac{dz}{\sqrt{2\pi}\beta} e^{-z^2/2} \delta[Az+B] = \frac{1}{\sqrt{2\pi}\beta|A|} e^{-B^2/2A^2}. \tag{A59}$$

We use this identity to simplify Eq. (A40) via Eqs. (A53) and (A54):

$$Q \approx \Gamma^2 \left\langle\!\!\left\langle \delta\left[\sum_\nu m_{1\nu}\chi_{1\nu} - \phi + \sqrt{\alpha r}z\right]\right\rangle\!\!\right\rangle = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}} \left\langle \exp\left[-\frac{\left(\sum_\nu m_{1\nu}\chi_{1\nu} - \phi\right)^2}{2\alpha r}\right]\right\rangle. \tag{A60}$$

Eqs. (A56), (A58), and (A60) are the zero-temperature mean-field equations connecting the order parameters $m_{1\nu}$, r, and Q (we no longer need q, q_0, and r_0). All further derivations will start with these equations.

APPENDIX B: CAPACITY FOR SPARSE EXAMPLES

1. Sparse mean-field equations

The mean-field equations Eqs. (A56), (A58), and (A60) involve a generic target pattern $\chi_{1\nu}$. We now consider the case where the network recovers a sparse example, so $\chi_{1\nu} = \eta_{11}\delta_{1\nu}$. Using this expression, we can simplify the mean-field equations and find the critical example load s_c above which sparse examples can no longer be retrieved. In this section, we take the sparse limit with $a \ll 1$.

For convenience, we rename $m \equiv m_{11}$ and $\eta \equiv \eta_{11}$. From Eq. (A2), we have

$$\eta = \begin{cases} (1-2\gamma)(1-a) & \text{with probability } a \\ -(1-2\gamma)a & \text{with probability } 1-a \end{cases} \approx \begin{cases} 1-2\gamma & \text{with probability } a \\ 0 & \text{with probability } 1 \end{cases}. \tag{B1}$$

Then, Eqs. (A56), (A58), and (A60) become

$$m = \frac{1}{2}\left\langle \eta \,\mathrm{erf}\, \frac{m\eta - \phi}{2\alpha r}\right\rangle = \frac{(1-2\gamma)a}{2}\left[\mathrm{erf}\, \frac{\phi}{\sqrt{2\alpha r}} + \mathrm{erf}\, \frac{(1-2\gamma)m - \phi}{\sqrt{2\alpha r}}\right], \tag{B2}$$

$$r = \frac{s\Gamma^4}{2} \frac{[1 - Q(1-\kappa^2)(1+s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1-\kappa^2)]^2[1 - Q(1+s_0\kappa^2)]^2}\left[1 + \left\langle\mathrm{erf}\, \frac{m\eta - \phi}{\sqrt{2\alpha r}}\right\rangle\right]$$

$$= \frac{s\Gamma^4}{2} \frac{[1 - Q(1-\kappa^2)(1+s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1-\kappa^2)]^2[1 - Q(1+s_0\kappa^2)]^2}\left[1 - \mathrm{erf}\, \frac{\phi}{\sqrt{2\alpha r}} + a\,\mathrm{erf}\, \frac{(1-2\gamma)m - \phi}{\sqrt{2\alpha r}}\right], \tag{B3}$$

$$Q = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}}\left\langle\exp\left[-\frac{(m\eta - \phi)^2}{2\alpha r}\right]\right\rangle = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}}\left\{\exp\left[-\frac{\phi^2}{2\alpha r}\right] + a\exp\left[-\frac{[(1-2\gamma)m - \phi]^2}{2\alpha r}\right]\right\}. \tag{B4}$$

We will soon see that these equations yield $Q \ll 1$ in the retrieval regime. In that case,

$$m = \frac{(1-2\gamma)a}{2}\left\{\mathrm{erf}\, \frac{\phi}{\sqrt{2\alpha r}} + \mathrm{erf}\, \frac{(1-2\gamma)m - \phi}{\sqrt{2\alpha r}}\right\}, \tag{B5}$$

$$r = \frac{s(1+s_0\kappa^4)\Gamma^4}{2}\left\{1 - \mathrm{erf}\, \frac{\phi}{\sqrt{2\alpha r}} + a\,\mathrm{erf}\, \frac{(1-2\gamma)m - \phi}{\sqrt{2\alpha r}}\right\}. \tag{B6}$$

These mean-field equations for sparse examples are presented in Eq. (13) with m replaced by its original name m_{11}. They can be numerically solved to find regimes of successful retrieval, but we will analyze them further in search of formulas for the capacity s_c.

In the limit that the network only stores sparse patterns with $2\gamma = 0$, these mean-field equations simplify to those of the sparse Hopfield network [7]. Note that their error function obeys $\mathrm{erf}\, x \to -1$ as $x \to \infty$, which is commonly called the complementary error function. To match our equations to theirs, make the replacements $\gamma \to 0$, $\Gamma^2 \to a(1-a)$, $s \to 1$, and $\mathrm{erf}\, x \to 1 - \mathrm{erf}\, x$ in our equations, and eliminate higher orders of $a \ll 1$.

Instead of invoking this limit to match the equations exactly, we can rewrite the mean-field equations Eqs. (B5) and (B6) in the form of Ref. [7] with the rescalings

$$m = (1-2\gamma)am',$$

$$\phi = (1-2\gamma)^2 a\theta',$$

$$r = s(1+s_0\kappa^4)\Gamma^4 r',$$

$$\alpha = \frac{(1-2\gamma)^4 a^2}{s(1+s_0\kappa^4)\Gamma^4}\alpha'. \tag{B7}$$

Then

$$m' = \frac{1}{2}\left(\mathrm{erf}\,\frac{\theta'}{\sqrt{2\alpha'r'}} + \mathrm{erf}\,\frac{m'-\theta'}{\sqrt{2\alpha'r'}}\right), \qquad \text{(B8)}$$

$$r' = \frac{1}{2}\left(1 - \mathrm{erf}\,\frac{\theta'}{\sqrt{2\alpha'r'}} + a\,\mathrm{erf}\,\frac{m'-\theta'}{\sqrt{2\alpha'r'}}\right). \qquad \text{(B9)}$$

Successful retrieval means that $m' \approx 1$, which requires $\theta'/\sqrt{2\alpha'r'} \gg 1$ and $(m'-\theta')/\sqrt{2\alpha'r'} \gg 1$. Under these limits, $0 < \theta' < 1$ and $Q \ll 1$, which validates our previous assumption. We can use asymptotic forms of the error function to obtain

$$m' = 1 - \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{\theta'}e^{-\theta'^2/2\alpha'r'}$$

$$- \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{m'-\theta'}e^{-(m'-\theta')^2/2\alpha'r'}, \qquad \text{(B10)}$$

$$r' = \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{\theta'}e^{-\theta'^2/2\alpha'r'} + \frac{a}{2}$$

$$- \frac{a}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{m'-\theta'}e^{-(m'-\theta')^2/2\alpha'r'}. \qquad \text{(B11)}$$

2. Capacity formula for $\theta' \to 0$

To derive capacity formulas, we need to make further assumptions about θ'. First, we consider small θ'. Because we still require $m' \approx 1$, the third term in Eq. (B11) becomes much smaller than the first, so

$$r' \approx \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{\theta'}e^{-\theta'^2/2\alpha'r'} + \frac{a}{2}. \qquad \text{(B12)}$$

This equation no longer depends on m'. If we take $y \equiv \theta'/\sqrt{\alpha'r'} \gg 1$, it becomes

$$\frac{\theta'^2}{\alpha'} = \frac{1}{\sqrt{2\pi}}ye^{-y^2/2} + \frac{a}{2}y^2. \qquad \text{(B13)}$$

The capacity is the maximum example load s for which this equation still admits a solution. Note that s is proportional to α' according to Eq. (B7). Thus, we maximize α' by minimizing the right-hand side of Eq. (B13) over y:

$$0 = \frac{1}{\sqrt{2\pi}}(1-y^2)e^{-y^2/2} + ay,$$

$$ye^{-y^2/2} \approx \sqrt{2\pi}a,$$

$$y = \sqrt{-W_{-1}(-2\pi a^2)} \approx \sqrt{2|\log a|}, \qquad \text{(B14)}$$

where W_{-1} is the negative branch of the Lambert W function, which is also known as the product logarithm. Substituting Eq. (B14) back into Eq. (B13), we obtain the maximal value

$$\alpha'_c \sim \frac{\theta'^2}{a|\log a|}. \qquad \text{(B15)}$$

This expression implicitly defines the capacity s_c for $\theta' \to 0$.

We can use this expression to obtain critical values for m'_c and r'_c:

$$m'_c \approx 1 - \frac{1}{2\sqrt{\pi|\log a|}}\frac{\theta'}{1-\theta'}a^{(1-\theta')^2/\theta'^2}, \qquad \text{(B16)}$$

$$r'_c \approx \frac{a}{2}. \qquad \text{(B17)}$$

Note that $m'_c \approx 1$, which confirms that our solution is self-consistent.

3. Capacity formula for $\theta' \to 1$

Next, we derive a capacity formula for large θ'. In this case, Eqs. (B10) and (B11) become

$$m' = 1 - \frac{1}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{m'-\theta'}e^{-(m'-\theta')^2/2\alpha'r'} \qquad \text{(B18)}$$

$$r' = \frac{a}{2} - \frac{a}{\sqrt{2\pi}}\frac{\sqrt{\alpha'r'}}{m'-\theta'}e^{-(m'-\theta')^2/2\alpha'r'}, \qquad \text{(B19)}$$

which yields

$$r' = \frac{a}{2} - a(1-m') \approx \frac{a}{2}. \qquad \text{(B20)}$$

If we define $y \equiv (m'-\theta')/\sqrt{\alpha'r'} \gg 1$ and use Eq. (B20), we can write Eq. (B18) as

$$\sqrt{\frac{a\alpha'}{2}} = \frac{1-\theta'}{y} - \frac{1}{\sqrt{2\pi}y^2}e^{-y^2/2}. \qquad \text{(B21)}$$

Again, the example load s is proportional to α' [Eq. (B7)], so we maximize α' by maximizing the right-hand size of Eq. (B21) with respect to y:

$$0 = -\frac{1-\theta'}{y^2} + \frac{1}{\sqrt{2\pi}}\left(\frac{2}{y^3} + \frac{1}{y}\right)e^{-y^2/2},$$

$$ye^{-y^2/2} \approx \sqrt{2\pi}(1-\theta'),$$

$$y \approx \sqrt{-W_{-1}[-2\pi(1-\theta')^2]} \approx \sqrt{2|\log(1-\theta')|}. \qquad \text{(B22)}$$

Substituting Eq. (B22) into Eq. (B21), we obtain

$$\alpha'_c \sim \frac{(1-\theta')^2}{a|\log(1-\theta')|}. \qquad \text{(B23)}$$

This expression implicitly defines the capacity s_c for $\theta' \to 1$.

Similarly to before, we use this expression to obtain the critical value for m'_c:

$$m'_c \approx 1 - \frac{1-\theta'}{y^2} = 1 - \frac{1-\theta'}{2|\log(1-\theta')|}. \qquad \text{(B24)}$$

$m'_c \approx 1$, which confirms that our solution was obtained self-consistently.

4. Maximizing capacity over θ'

We have derived two expressions for α_c', which is proportional to the capacity s_c, in different regimes of the rescaled threshold θ':

$$\alpha_c' \sim \begin{cases} \dfrac{\theta'^2}{a|\log a|} & \theta' \to 0, \\ \dfrac{(1-\theta')^2}{a|\log(1-\theta')|} & \theta' \to 1. \end{cases} \tag{B25}$$

We now take θ' to be a free parameter and maximize the capacity over it. The first expression for α_c' grows from 0 as θ' increases from 0, and the second one grows from 0 as θ' decreases from 1. Thus, the optimum value should lie somewhere in between, and we estimate its location by finding the crossover point where the two expressions meet.

We assume that θ' is sufficiently far from 1 such that $|\log(1-\theta')| \sim 1$. Then the crossover point is given by

$$\frac{\theta'^2}{a|\log a|} \approx \frac{(1-\theta')^2}{a},$$

$$\theta' = \frac{\sqrt{|\log a|}}{1 + \sqrt{|\log a|}}. \tag{B26}$$

Substituting this optimal threshold back into Eq. (B25), we obtain

$$\alpha_c' \sim \frac{1}{a|\log a|}. \tag{B27}$$

By converting α' back to α with Eq. (B7), we recover the capacity formula Eq. (20) at optimal threshold.

APPENDIX C: CAPACITY FOR DENSE EXAMPLES

1. Dense asymmetric mean-field equations

We return to the generic mean-field Eqs. (A56), (A58), and (A60) and consider the case where the network recovers a dense example ζ_{11}. Due to correlations, the network will overlap with all dense patterns $\zeta_{1\nu}$, so $\chi_{1\nu} = \zeta_{1\nu}$. Using this expression, we can simplify the mean-field equations and find the critical example load s_c above which dense examples can no longer be retrieved.

Recall from Eq. (A2) that

$$\zeta_{1\nu} = \begin{cases} \zeta_1 & \text{with probability } \frac{1+c}{2} \\ -\zeta_1 & \text{with probability } \frac{1-c}{2} \end{cases},$$

$$\zeta_1 = \begin{cases} \gamma & \text{with probability } \frac{1}{2} \\ -\gamma & \text{with probability } \frac{1}{2} \end{cases}. \tag{C1}$$

To help us in our calculations, we note the integrals

$$\int_{-\infty}^{\infty} dx\, e^{-(x-A)^2/\rho^2} e^{-(x-B)^2/\sigma^2} = \sqrt{\frac{\pi}{\rho^{-2} + \sigma^{-2}}} \exp\left[-\frac{(A-B)^2}{\rho^2 + \sigma^2}\right],$$

$$\int_{-\infty}^{\infty} dx\, e^{-(x-A)^2/\rho^2} \operatorname{erf}\left[\frac{x-B}{\sigma}\right] = \sqrt{\pi}\rho \operatorname{erf}\left[\frac{A-B}{\sqrt{\rho^2 + \sigma^2}}\right],$$

$$\int_{-\infty}^{\infty} dx\, e^{-(x-A)^2/\rho^2} x \operatorname{erf}\left[\frac{x-B}{\sigma}\right] = \rho\left\{\frac{\rho^2}{\sqrt{\rho^2 + \sigma^2}} \exp\left[-\frac{(A-B)^2}{\rho^2 + \sigma^2}\right] + \sqrt{\pi}A \operatorname{erf}\left[\frac{A-B}{\sqrt{\rho^2 + \sigma^2}}\right]\right\}. \tag{C2}$$

During successful retrieval, the network overlaps strongly with the target pattern ζ_{11}. It will also overlap with other examples $\zeta_{1\nu}$ for $\nu > 1$ to a degree governed by the correlation parameter c [Eq. (A3)]. As $N \to \infty$, these other overlaps converge towards one another due to the law of large numbers; we call this asymptotic value $m_0 \equiv m_{1\nu}$ for $\nu > 1$. Thus, we can write

$$\sum_\nu m_{1\nu}\zeta_{1\nu} = m_{11}\zeta_{11} + m_0 \sum_{\nu>1} \zeta_{1\nu} = m\zeta + s_0 m_0 x_0. \tag{C3}$$

We rename $m \equiv m_{11}$ and $\zeta \equiv \zeta_{11}$ for convenience. x_0 is the average over the $s_0 = s - 1$ other examples in concept 1, and it follows a binomial distribution with mean $c\zeta_1$ and variance $\gamma^2(1-c^2)/s_0$ according to Eq. (C1). In the large s limit, it can be approximated by a Gaussian random variable with the same central moments. We also introduce m_1, which is the network overlap with the concept pattern ζ_1.

With these considerations, Eqs. (A56), (A58), and (A60) yield

$$m = \frac{1}{2}\left\langle\!\!\left\langle \zeta \operatorname{erf} \frac{m\zeta + s_0 m_0 x_0 - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle, \tag{C4}$$

$$m_0 = \frac{1}{2}\left\langle\!\!\left\langle x_0 \operatorname{erf} \frac{m\zeta + s_0 m_0 x_0 - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle, \tag{C5}$$

$$m_1 = \frac{1}{2}\left\langle\!\!\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta + s_0 m_0 x_0 - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle, \tag{C6}$$

$$r = \frac{s\Gamma^4}{2} \frac{[1 - Q(1 - \kappa^2)(1 + s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1 - \kappa^2)]^2[1 - Q(1 + s_0\kappa^2)]^2} \left[1 + \left\langle\!\!\left\langle \operatorname{erf} \frac{m\zeta + s_0 m_0 x_0 - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle \right], \tag{C7}$$

$$Q = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}} \left\langle\!\!\left\langle \exp\left[-\frac{(m\zeta + s_0 m_0 x_0 - \phi)^2}{2\alpha r} \right] \right\rangle\!\!\right\rangle. \tag{C8}$$

The double angle brackets indicate averages over ζ and x_0, which is a Gaussian random variable with mean and variance listed above. We define the following variables:

$$\sigma_0^2 \equiv s_0 \gamma^2 (1 - c^2) m_0^2 + \alpha r,$$
$$Y_{\pm\pm} \equiv \frac{\gamma m \pm s_0 \gamma c m_0 \pm \phi}{\sqrt{2}\sigma_0}, \tag{C9}$$

with choices for $+$ and $-$ in $Y_{\pm\pm}$ corresponding to signs in the right-hand side. Now we come to the task of performing the averages in Eqs. (C4)–(C8). For each variable, we average successively over ζ, x_0, and ζ_1.

First,

$$Q = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}} \left\{ \frac{1+c}{2} \left\langle\!\!\left\langle \exp\left[-\frac{(m\zeta_1 + s_0 m_0 x_0 - \phi)^2}{2\alpha r} \right] \right\rangle\!\!\right\rangle + \frac{1-c}{2} \left\langle\!\!\left\langle \exp\left[-\frac{(m\zeta_1 - s_0 m_0 x_0 + \phi)^2}{2\alpha r} \right] \right\rangle\!\!\right\rangle \right\}. \tag{C10}$$

Then,

$$\frac{1}{\sqrt{2\pi\alpha r}} \left\langle\!\!\left\langle \exp\left[-\frac{(m\zeta_1 + s_0 m_0 x_0 - \phi)^2}{2\alpha r} \right] \right\rangle\!\!\right\rangle$$
$$= \frac{1}{\sqrt{2\pi\alpha r}} \sqrt{\frac{s_0}{2\pi \gamma^2 (1 - c^2)}} \left\langle \int dx_0 \, e^{-s_0(x_0 - c\zeta_1)^2 / 2\gamma^2 (1-c^2)} e^{-s_0^2 m_0^2 [x_0 + (m\zeta_1 - \phi)/s_0 m_0]^2 / 2\alpha r} \right\rangle$$
$$= \frac{1}{\sqrt{2\pi \left[s_0 \gamma^2 (1 - c^2) m_0^2 + \alpha r \right]}} \left\langle \exp\left[-\frac{(m\zeta_1 + s_0 c m_0 \zeta_1 - \phi)^2}{2\left[s_0 \gamma^2 (1 - c^2) m_0^2 + \alpha r \right]} \right] \right\rangle$$
$$= \frac{1}{\sqrt{2\pi}\sigma_0} \frac{1}{2} \left(e^{-Y_{++}^2} + e^{-Y_{+-}^2} \right). \tag{C11}$$

Thus,

$$Q = \frac{\Gamma^2}{\sqrt{2\pi}\sigma_0} \left[\frac{1+c}{4} \left(e^{-Y_{++}^2} + e^{-Y_{+-}^2} \right) + \frac{1-c}{4} \left(e^{-Y_{-+}^2} + e^{-Y_{--}^2} \right) \right]. \tag{C12}$$

Next,

$$\left\langle\!\!\left\langle \operatorname{erf} \frac{m\zeta + s_0 m_0 x_0 - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle = \left\{ \frac{1+c}{2} \left\langle\!\!\left\langle \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{2\alpha r} \right\rangle\!\!\right\rangle - \frac{1-c}{2} \left\langle\!\!\left\langle \operatorname{erf} \frac{m\zeta_1 - s_0 m_0 x_0 + \phi}{2\alpha r} \right\rangle\!\!\right\rangle \right\}. \tag{C13}$$

Then,

$$\left\langle\!\!\left\langle \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{2\alpha r} \right\rangle\!\!\right\rangle = \sqrt{\frac{s_0}{2\pi \gamma^2 (1 - c^2)}} \left\langle \int dx_0 \, e^{-s_0(x_0 - c\zeta_1)^2 / 2\gamma^2 (1-c^2)} \operatorname{erf}\left[\frac{s_0 m_0}{\sqrt{2\alpha r}} \left(x_0 + \frac{m\zeta_1 - \phi}{s_0 m_0} \right) \right] \right\rangle$$
$$= \left\langle \operatorname{erf} \frac{m\zeta_1 + s_0 c m_0 \zeta_1 - \phi}{\sqrt{2\left(s_0 \gamma^2 (1 - c^2) m_0^2 + \alpha r \right)}} \right\rangle = -\frac{1}{2} \left(\operatorname{erf} Y_{++} - \operatorname{erf} Y_{+-} \right). \tag{C14}$$

Thus,

$$r = \frac{s\Gamma^4}{2} \frac{[1 - Q(1 - \kappa^2)(1 + s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1 - \kappa^2)]^2[1 - Q(1 + s_0\kappa^2)]^2} \left[1 - \frac{1+c}{4} (\operatorname{erf} Y_{++} - \operatorname{erf} Y_{+-}) - \frac{1-c}{4} (\operatorname{erf} Y_{-+} - \operatorname{erf} Y_{--}) \right]. \tag{C15}$$

Next,

$$m = \frac{1}{2} \left\{ \frac{1+c}{2} \left\langle\!\!\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{2\alpha r} \right\rangle\!\!\right\rangle + \frac{1-c}{2} \left\langle\!\!\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 - s_0 m_0 x_0 + \phi}{2\alpha r} \right\rangle\!\!\right\rangle \right\}. \tag{C16}$$

Then,

$$\left\langle\!\!\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{2\alpha r} \right\rangle\!\!\right\rangle = \sqrt{\frac{s_0}{2\pi\gamma^2(1-c^2)}} \left\langle \zeta_1 \int dx_0\, e^{-s_0(x_0 - c\zeta_1)^2/2\gamma^2(1-c^2)} \operatorname{erf}\left[\frac{s_0 m_0}{\sqrt{2\alpha r}}\left(x_0 + \frac{m\zeta_1 - \phi}{s_0 m_0}\right)\right] \right\rangle$$

$$= \left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 + s_0 c m_0 \zeta_1 - \phi}{\sqrt{2(s_0\gamma^2(1-c^2)m_0^2 + \alpha r)}} \right\rangle = \frac{\gamma}{2}\left(\operatorname{erf} Y_{++} + \operatorname{erf} Y_{+-}\right). \tag{C17}$$

Thus,

$$m = \frac{\gamma}{2}\left[\frac{1+c}{4}\left(\operatorname{erf} Y_{++} + \operatorname{erf} Y_{+-}\right) + \frac{1-c}{4}\left(\operatorname{erf} Y_{-+} + \operatorname{erf} Y_{--}\right)\right]. \tag{C18}$$

Similarly,

$$m_1 = \frac{1}{2}\left[\frac{1+c}{2}\left\langle\!\!\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{2\alpha r} \right\rangle\!\!\right\rangle - \frac{1-c}{2}\left\langle\!\!\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 - s_0 m_0 x_0 + \phi}{2\alpha r} \right\rangle\!\!\right\rangle\right]$$

$$= \frac{\gamma}{2}\left[\frac{1+c}{4}\left(\operatorname{erf} Y_{++} + \operatorname{erf} Y_{+-}\right) - \frac{1-c}{4}\left(\operatorname{erf} Y_{-+} + \operatorname{erf} Y_{--}\right)\right]. \tag{C19}$$

Finally,

$$m_0 = \frac{1}{2}\left[\frac{1+c}{2}\left\langle\!\!\left\langle x_0 \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{2\alpha r} \right\rangle\!\!\right\rangle - \frac{1-c}{2}\left\langle\!\!\left\langle x_0 \operatorname{erf} \frac{m\zeta_1 - s_0 m_0 x_0 + \phi}{2\alpha r} \right\rangle\!\!\right\rangle\right]. \tag{C20}$$

Then,

$$\left\langle\!\!\left\langle x_0 \operatorname{erf} \frac{m\zeta_1 + s_0 m_0 x_0 - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle = \sqrt{\frac{s_0}{2\pi\gamma^2(1-c^2)}}\left\langle\int dx_0\, e^{-s_0(x_0 - c\zeta_1)^2/2\gamma^2(1-c^2)} x_0 \operatorname{erf}\left[\frac{s_0 m_0}{\sqrt{2\alpha r}}\left(x_0 + \frac{m\zeta_1 - \phi}{s_0 m_0}\right)\right]\right\rangle$$

$$= \gamma^2(1-c^2)m_0\sqrt{\frac{2}{\pi[s_0\gamma^2(1-c^2)m_0^2 + \alpha r]}}\left\langle\exp\left[-\frac{(m\zeta_1 + s_0 c m_0\zeta_1 - \phi)^2}{2[s_0\gamma^2(1-c^2)m_0^2 + \alpha r]}\right]\right\rangle$$

$$+ c\left\langle \zeta_1 \operatorname{erf} \frac{m\zeta_1 + s_0 c m_0\zeta_1 - \phi}{\sqrt{2[s_0\gamma^2(1-c^2)m_0^2 + \alpha r]}}\right\rangle$$

$$= \frac{\gamma^2(1-c^2)m_0}{\sqrt{2\pi}\sigma_0}\left(e^{-Y_{++}^2} + e^{-Y_{+-}^2}\right) + \frac{\gamma c}{2}\left(\operatorname{erf} Y_{++} + \operatorname{erf} Y_{+-}\right). \tag{C21}$$

Thus,

$$m_0 = \frac{\gamma c}{2}\left[\frac{1+c}{4}\left(\operatorname{erf} Y_{++} + \operatorname{erf} Y_{+-}\right) - \frac{1-c}{4}\left(\operatorname{erf} Y_{-+} + \operatorname{erf} Y_{--}\right)\right] + Q\frac{\gamma^2}{\Gamma^2}(1-c^2)m_0$$

$$= \frac{\gamma c}{2[1 - Q\frac{\gamma^2}{\Gamma^2}(1-c^2)]}\left[\frac{1+c}{4}\left(\operatorname{erf} Y_{++} + \operatorname{erf} Y_{+-}\right) - \frac{1-c}{4}\left(\operatorname{erf} Y_{-+} + \operatorname{erf} Y_{--}\right)\right]. \tag{C22}$$

These mean-field equations are presented in Eq. (14) with m replaced by its original name m_{11}. They can be numerically solved to find regimes of successful retrieval, but we will analyze them further in search of a formula for the capacity s_c.

In the limit that the network only stores dense, uncorrelated patterns with $2\gamma = 1$ and $c = 0$, these mean-field equations simplify to those of the Hopfield network with 0/1 neurons [53], which has half the capacity of the original Hopfield network with $+1/-1$ neurons. Note that the pattern storage strength in Ref. [53] is twice that of ours; in other words, their connectivity weights are scaled by a factor of 4 in comparison to ours. To match our equations to theirs, make

the replacements $\gamma \to 1$, $\Gamma \to 1$, $s \to 1$, $c \to 0$, and $\kappa \to 0$ in our equations, and recall Eqs. (A47)–(A52).

2. Simplified mean-field equations

To derive a capacity formula, we make three further assumptions. First, we assume $c^2 \ll 1$, which implies $\kappa^2 \ll 1$ as well. Second, we assume that the rescaled threshold $\phi = 0$. This assumption is justified empirically, for we find that the capacity is maximized at $|\phi| < 10^{-6}$ over all parameter ranges in Fig. 3. It is also justified theoretically, since we will derive that $Q \ll 1$, which means $\phi \approx \theta$. For dense patterns in classic Hopfield network, retrieval is maximized at threshold $\theta = 0$

[21]. Finally, we assume $s \gg 1$, so $s_0 = s$; this is not necessary, but it makes the expressions simpler.

We rescale the order parameters with

$$m = \frac{\gamma}{2}m',$$

$$m_0 = \frac{\gamma}{2}m_0',$$

$$r = \frac{s\Gamma^4}{2}r',$$

$$\alpha = \frac{\gamma^4}{2\Gamma^4}\alpha'. \tag{C23}$$

The mean-field equations then become

$$m' = \frac{1+c}{2}\operatorname{erf} Y_+' + \frac{1-c}{2}\operatorname{erf} Y_-', \tag{C24}$$

$$m_0' = \frac{c}{1 - Q\frac{\gamma^2}{\Gamma^2}}\left(\frac{1+c}{2}\operatorname{erf} Y_+' - \frac{1-c}{2}\operatorname{erf} Y_-'\right), \tag{C25}$$

$$r' = \frac{1}{(1-Q)^2}, \tag{C26}$$

$$Q = \sqrt{\frac{2}{\pi}}\frac{\Gamma^2}{\sigma_0'\gamma^2}\left(\frac{1+c}{2}e^{-(Y_+')^2} + \frac{1-c}{2}e^{-(Y_-')^2}\right), \tag{C27}$$

where

$$\sigma_0'^2 \equiv s\left(m_0'^2 + \alpha'r'\right),$$

$$Y_\pm' \equiv \frac{m'}{\sqrt{2}\sigma_0'}\left(1 \pm \frac{scm_0'}{m'}\right). \tag{C28}$$

Successful retrieval means that $m' \approx 1$, which requires $Y_\pm' \gg 1$. This condition in turn yields $m_0' \approx c^2$ and $Q \ll 1$ through Eqs. (C25) and (C27), which confirms our previous assumption. Thus, $Y_\pm' \approx (y/\sqrt{2})(1 \pm sc^3)$, where

$$y \equiv \frac{m'}{\sigma_0'}. \tag{C29}$$

For $Y_\pm' \gg 1$, we need $sc^3 \lesssim 1$ and $y \gg 1$, which we use to boldly simplify Eqs. (C24)–(C27):

$$m' = 1 - \frac{1}{y^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}, \tag{C30}$$

$$m_0' = \frac{c^2}{1 - Q\frac{\gamma^2}{\Gamma^2}}\left[1 - \left(\frac{1}{y^2} - sc^2\right)\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right], \tag{C31}$$

$$\alpha' = \left(\frac{m'^2}{sy^2} - m_0'^2\right)(1-Q)^2, \tag{C32}$$

$$Q = \frac{\Gamma^2}{\gamma^2 m'}\left(1 - sc^4y^2 + \frac{1}{2}s^2c^6y^4\right)\sqrt{\frac{2}{\pi}}ye^{-y^2/2}. \tag{C33}$$

For mathematical tractability, we have expanded in sc^3 and $1/y$, even though the former is not strictly small and the latter can be empirically close to 1.

3. Capacity formula

In Eqs. (C30)–(C33), we substitute formulas for m', m_0', and Q into the equation for α' and keep only leading terms in

$1/y$ and c. After much simplification, we obtain

$$s(\alpha' + c^4) \approx \frac{1}{y^2} - \frac{\Gamma^2}{\gamma^2}\left(\frac{1}{y} + \frac{1}{2}s^2c^6y^3\right)\sqrt{\frac{8}{\pi}}e^{-y^2/2}. \tag{C34}$$

At the critical value of s above which Eq. (C34) equation cannot be satisfied by any y, derivatives with respect to y on both sides of the equation must be equal. In other words, we expect the critical s_c to be a saddle-node bifurcation point. For mathematical tractability, we ignore the term proportional to s^2. This simplification is rather arbitrary, but it can be empirically justified by comparing the resulting formula with numerical analysis of the full mean-field equations (Fig. 3). We also eliminate higher orders in $1/y$ to obtain

$$0 \approx -\frac{2}{y^3} + \frac{\Gamma^2}{\gamma^2}\sqrt{\frac{8}{\pi}}e^{-y^2/2}. \tag{C35}$$

Solving for y, we obtain

$$y = \sqrt{-3W_{-1}\left[-\frac{1}{3}\left(\frac{\pi}{2}\right)^{1/3}\left(\frac{\gamma}{\Gamma}\right)^{4/3}\right]}$$

$$\approx \sqrt{3\log\left[3\left(\frac{2}{\pi}\right)^{1/3}\left(\frac{\Gamma}{\gamma}\right)^{4/3}\right]}, \tag{C36}$$

where W_{-1} is the negative branch of the Lambert W function. Since this function involves a logarithm, it varies very slowly as a function of γ/Γ. For $\gamma = 0.1$ and a between 0.001 and 0.1, this expression for y ranges from 1.7 and 3.3. Within this range, $m' > 0.88$ according to Eq. (C30), which confirms that our earlier simplifications using $m' \approx 1$ yield self-consistent results.

We can use Eq. (C35) to simplify Eq. (C34) to leading order in $1/y$:

$$s_c(\alpha' + c^4) \approx \frac{1}{y^2} - s_c^2c^6. \tag{C37}$$

Solving for s_c,

$$s_c = \frac{\sqrt{(\alpha' + c^4)^2 + \frac{4}{y^2}c^6} - (\alpha' + c^4)}{2c^6}. \tag{C38}$$

To heuristically obtain a simpler equation, we note that $s_c \to 1/yc^3$ when $\alpha' \to 0$ and $s_c \to 1/y^2\alpha'$ when $\alpha' \to \infty$. We simply capture both these behaviors with

$$s_c \sim \frac{1}{yc^3 + y^2\alpha'}. \tag{C39}$$

Again, y varies slowly within its range, so we simplify this equation by simply setting $y \sim 3$. After converting α' back to α with Eq. (C23), we obtain Eq. (23).

APPENDIX D: CRITICAL LOAD FOR DENSE CONCEPTS

1. Dense symmetric mean-field equations

We return to the generic mean-field equations Eqs. (A56), (A58), and (A60) and consider the case where the network recovers a dense concept ζ_1. Due to correlations, the network will overlap with all dense patterns $\zeta_{1\nu}$, so $\chi_{1\nu} = \zeta_{1\nu}$. Using this expression, we can simplify the mean-field equations and find the critical example load s_c below which dense concepts

cannot be retrieved. Recall the dense pattern statistics Eq. (C1) and Gaussian integrals Eq. (C2), which will aid us in our derivations.

Successful retrieval means that the network overlaps strongly with the target concept ζ_1. The correlation parameter c produces overlaps with all example patterns $\zeta_{1\nu}$ [Eq. (A3)], which converge to an asymptotic value $m_s \equiv m_{1\nu}$ as $N \to \infty$. The "s" signifies "symmetric," i.e., equal overlap with all examples in concept 1. Thus, we can write

$$\sum_\nu m_{1\nu}\zeta_{1\nu} = m_s \sum_\nu \zeta_{1\nu} = sm_s x_s. \tag{D1}$$

x_s is the average over the s examples in concept 1, and it follows a binomial distribution with mean $c\zeta_1$ and variance $\gamma^2(1-c^2)/s$ according to Eq. (C1). In the large s limit, it can be approximated by a Gaussian random variable with the same central moments. We explicitly introduce m_1, which is the network overlap with the target concept ζ_1.

With these considerations, Eqs. (A56), (A58), and (A60) yield

$$m_s = \frac{1}{2}\left\langle\!\!\left\langle x_s \, \mathrm{erf}\, \frac{sm_s x_s - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle, \tag{D2}$$

$$m_1 = \frac{1}{2}\left\langle\!\!\left\langle \zeta_1 \, \mathrm{erf}\, \frac{sm_s x_s - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle, \tag{D3}$$

$$r = \frac{s\Gamma^4}{2} \frac{[1 - Q(1-\kappa^2)(1+s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1-\kappa^2)]^2[1 - Q(1+s_0\kappa^2)]^2}$$
$$\times \left[1 + \left\langle\!\!\left\langle \mathrm{erf}\, \frac{sm_s x_s - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle\right], \tag{D4}$$

$$Q = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}}\left\langle\!\!\left\langle \exp\left[-\frac{(sm_s x_s - \phi)^2}{2\alpha r}\right] \right\rangle\!\!\right\rangle. \tag{D5}$$

The double angle brackets indicate averages over ζ and x_s, which is a Gaussian random variable with mean and variance listed above. We define the following variables:

$$\sigma_s^2 \equiv s\gamma^2(1-c^2)m_s^2 + \alpha r,$$
$$Y_\pm \equiv \frac{s\gamma c m_s \pm \phi}{\sqrt{2}\sigma_s}, \tag{D6}$$

with choices for $+$ and $-$ in $Y_\pm$ corresponding the sign in the right-hand side. Now we come to the task of performing the averages in Eqs. (D2)–(D5). For each variable, we average successively over x_s and ζ_1.

First,

$$Q = \frac{\Gamma^2}{\sqrt{2\pi\alpha r}}\left\langle\!\!\left\langle \exp\left[-\frac{(sm_s x_s - \phi)^2}{2\alpha r}\right] \right\rangle\!\!\right\rangle$$

$$= \frac{\Gamma^2}{\sqrt{2\pi\alpha r}}\sqrt{\frac{s}{2\pi\gamma^2(1-c^2)}}\left\langle \int dx_s \, e^{-s(x_s - c\zeta_1)^2/2\gamma^2(1-c^2)} e^{-s^2 m_s^2(x_s - \phi/sm_s)^2/2\alpha r} \right\rangle$$

$$= \frac{\Gamma^2}{\sqrt{2\pi[s\gamma^2(1-c^2)m_s^2 + \alpha r]}}\left\langle \exp\left\{-\frac{(scm_s\zeta_1 - \phi)^2}{2[s\gamma^2(1-c^2)m_s^2 + \alpha r]}\right\} \right\rangle$$

$$= \frac{\Gamma^2}{\sqrt{8\pi}\sigma_s}\left(e^{-Y_+^2} + e^{-Y_-^2}\right). \tag{D7}$$

Next,

$$\left\langle\!\!\left\langle \mathrm{erf}\, \frac{sm_s x_s - \phi}{\sqrt{2\alpha r}} \right\rangle\!\!\right\rangle = \sqrt{\frac{s}{2\pi\gamma^2(1-c^2)}}\left\langle \int dx_s \, e^{-s(x_s - c\zeta_1)^2/2\gamma^2(1-c^2)} \, \mathrm{erf}\left[\frac{sm_s}{\sqrt{2\alpha r}}\left(x_s - \frac{\phi}{sm_s}\right)\right] \right\rangle$$

$$= \left\langle \mathrm{erf}\, \frac{scm_s\zeta_1 - \phi}{\sqrt{2[s\gamma^2(1-c^2)m_s^2 + \alpha r]}} \right\rangle = -\frac{1}{2}\left(\mathrm{erf}\, Y_+ - \mathrm{erf}\, Y_-\right). \tag{D8}$$

Thus,

$$r = \frac{s\Gamma^4}{2}\frac{[1 - Q(1-\kappa^2)(1+s_0\kappa^2)]^2 + s_0\kappa^4}{[1 - Q(1-\kappa^2)]^2[1 - Q(1+s_0\kappa^2)]^2}\left[1 - \frac{1}{2}\left(\mathrm{erf}\, Y_+ - \mathrm{erf}\, Y_-\right)\right]. \tag{D9}$$

Next,

$$m_1 = \frac{1}{2}\left\langle\!\!\left\langle \zeta_1 \, \mathrm{erf}\, \frac{sm_s x_s - \phi}{2\alpha r} \right\rangle\!\!\right\rangle$$

$$= \frac{1}{2}\sqrt{\frac{s}{2\pi\gamma^2(1-c^2)}}\left\langle \zeta_1 \int dx_s \, e^{-s(x_s - c\zeta_1)^2/2\gamma^2(1-c^2)} \, \mathrm{erf}\left[\frac{sm_s}{\sqrt{2\alpha r}}\left(x_s - \frac{\phi}{sm_s}\right)\right] \right\rangle$$

$$= \frac{1}{2}\left\langle \zeta_1 \, \mathrm{erf}\, \frac{scm_s\zeta_1 - \phi}{\sqrt{2[s\gamma^2(1-c^2)m_s^2 + \alpha r]}} \right\rangle$$

$$= \frac{\gamma}{4}\left(\mathrm{erf}\, Y_+ + \mathrm{erf}\, Y_-\right). \tag{D10}$$

FIG. 9. (a) Critical example load s_c for dense concepts obtained through numerical analysis of Eq. (16). We either set $\phi = 0$ (dark, thin lines) or maximize over ϕ (light, thick lines). (b) Right-hand side of Eq. (D17) and its terms plotted separately. We use $a = 0.1$, $c = 0.1$, $\alpha = 10^{-4}$, and $s = s_c = 929$. (c) y as a function of α for density $a = 0$ obtained by numerically solving Eq. (D20). For all results, $\gamma = 0.1$.

Finally,

$$
\begin{aligned}
m_s &= \frac{1}{2}\left\langle\!\!\left\langle x_s \operatorname{erf} \frac{sm_s x_s - \phi}{2\alpha r} \right\rangle\!\!\right\rangle \\
&= \frac{1}{2}\sqrt{\frac{s}{2\pi\gamma^2(1-c^2)}}\left\langle \int dx_s\, e^{-s(x_s - c\zeta_1)^2/2\gamma^2(1-c^2)} x_s \operatorname{erf}\left[\frac{sm_s}{\sqrt{2\alpha r}}\left(x_s - \frac{\phi}{sm_s}\right)\right]\right\rangle \\
&= \frac{\gamma^2(1-c^2)m_s}{2}\sqrt{\frac{2}{\pi\left[s\gamma^2(1-c^2)m_s^2 + \alpha r\right]}}\left\langle \exp\left\{-\frac{(scm_s\zeta_1 - \phi)^2}{2\left[s\gamma^2(1-c^2)m_s^2 + \alpha r\right]}\right\}\right\rangle + \frac{c}{2}\left\langle \zeta_1 \operatorname{erf} \frac{scm_s\zeta_1 - \phi}{\sqrt{2\left[s\gamma^2(1-c^2)m_s^2 + \alpha r\right]}}\right\rangle \\
&= Q\frac{\gamma^2}{\Gamma^2}(1-c^2)m_s + \frac{\gamma c}{4}\left(\operatorname{erf} Y_+ + \operatorname{erf} Y_-\right) \\
&= \frac{\gamma c}{4\left[1 - Q\frac{\gamma^2}{\Gamma^2}(1-c^2)\right]}\left(\operatorname{erf} Y_+ + \operatorname{erf} Y_-\right).
\end{aligned}
\tag{D11}
$$

These mean-field equations are presented in Eq. (16).

2. Simplified mean-field equations

To derive a formula for the critical example load s_c, we make three further assumptions. First, we assume $c^2 \ll 1$, which implies $\kappa^2 \ll 1$ as well. Second, we assume that rescaled threshold $\phi = 0$. This assumption is justified empirically. We find that s_c is minimized at $|\phi| < 0.5$ over all parameter ranges in Fig. 4; moreover, these values are very close to that obtained by enforcing $\phi = 0$ [Fig. 9(a)]. Finally, we assume $s \gg 1$, so $s_0 = s$; this is not necessary, but it makes the expressions simpler.

We rescale the order parameters with

$$
m_s = \frac{\gamma c}{2}m_s', \quad r = \frac{s\Gamma^4}{2}(m_s')^2 r', \quad \alpha = \frac{\gamma^4 c^2}{2\Gamma^4}\alpha'.
\tag{D12}
$$

We also define

$$
y \equiv \sqrt{\frac{sc^2}{1 + \alpha' r'}},
\tag{D13}
$$

so $Y_\pm \approx y/\sqrt{2}$. The mean-field equations then become

$$
m_s' - Qm_s'\frac{\gamma^2}{\Gamma^2} = \operatorname{erf}\frac{y}{\sqrt{2}},
\tag{D14}
$$

$$
\frac{1}{\alpha'}\left(\frac{sc^2}{y^2} - 1\right) = \frac{[m_s' - Qm_s'(1 + s\kappa^2)]^2 + (m_s')^2 s\kappa^4}{(m_s' - Qm_s')^2[m_s' - Qm_s'(1 + s\kappa^2)]^2},
\tag{D15}
$$

$$
Qm_s' = \frac{\Gamma^2}{\gamma^2 sc^2}\sqrt{\frac{2}{\pi}}y e^{-y^2/2}.
\tag{D16}
$$

We can now substitute expressions for m'_s and Qm'_s into Eq. (D15) to obtain

$$0 = \frac{1}{\alpha'}\left(\frac{sc^2}{y^2} - 1\right)\left[\mathrm{erf}\,\frac{y}{\sqrt{2}} - \frac{(1-2\gamma)^2 a(1-a)}{\gamma^2 sc^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right]^2 - 1 - s\kappa^4\frac{\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} + \frac{1}{sc^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2}{\left[\mathrm{erf}\,\frac{y}{\sqrt{2}} - \left(\frac{(1-2\gamma)^2 a(1-a)}{\gamma^2 sc^2} + 1\right)\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right]^2}. \quad \text{(D17)}$$

At the critical value of s above which Eq. (D17) equation cannot be satisfied by any y, its derivative with respect to y must be 0. In other words, we expect the critical s_c to be a saddle-node bifurcation point.

3. Critical load relations for $a \gtrsim \gamma^2$

To derive formulas for s_c, we need to make further assumptions about a. First, we consider the case where a is not too small. In Fig. 9(b) we plot the right-hand side (r.h.s.) of Eq. (D17), along with its first two terms and third term separately. The first two terms generally capture the behavior of the r.h.s. The third term contributes a pole, whose location approximately sets the position of the local maximum of the r.h.s. where its derivative equals 0. Thus, we use the first two terms to satisfy Eq. (D17) and the denominator of the third term to satisfy its derivative:

$$\alpha' \approx \left(\frac{s_c c^2}{y^2} - 1\right)\left[\mathrm{erf}\,\frac{y}{\sqrt{2}} - \frac{(1-2\gamma)^2 a(1-a)}{\gamma^2 s_c c^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right]^2,$$

$$0 \approx \mathrm{erf}\,\frac{y}{\sqrt{2}} - \left[\frac{(1-2\gamma)^2 a(1-a)}{\gamma^2 s_c c^2} + 1\right]\sqrt{\frac{2}{\pi}}ye^{-y^2/2}. \quad \text{(D18)}$$

We can manipulate these equations to obtain Eq. (24) if we convert α' back to α with Eq. (D12).

4. Critical load formula for $a \ll \gamma^2$

Next we consider $a \to 0$. In this case, the second equation of Eq. (D18) implies $y \to 0$, which does not correspond to a retrieval solution $m'_s \approx 1$ according to Eq. (D14). Thus the pole location cannot be used to satisfy the derivative of Eq. (D17). To proceed, we instead set $a = 0$ in Eq. (D17) and obtain

$$0 = \frac{1}{\alpha'}\left(\frac{sc^2}{y^2} - 1\right)\left(\mathrm{erf}\,\frac{y}{\sqrt{2}}\right)^2\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2 - \left(\mathrm{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2 - sc^4\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} + \frac{1}{sc^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2. \quad \text{(D19)}$$

We then directly calculate its derivative with respect to y. Along with Eq. (D19) without taking the derivative, this gives

$$\alpha' = \frac{\left(\frac{s_c c^2}{y^2} - 1\right)\left(\mathrm{erf}\,\frac{y}{\sqrt{2}}\right)^2\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2}{\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2 + s_c c^4\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} + \frac{1}{s_c c^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^2},$$

$$\alpha' c^2 = \frac{\mathrm{erf}\,\frac{y}{\sqrt{2}}\left[\frac{s_c c^2}{y^2}\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right) + \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right]\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} - \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)^3}{(s_c c^2 + 1)\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\left(\mathrm{erf}\,\frac{y}{\sqrt{2}} + \frac{1}{s_c c^2}\sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right)\left[(y^2 - 1)\mathrm{erf}\,\frac{y}{\sqrt{2}} + \sqrt{\frac{2}{\pi}}ye^{-y^2/2}\right]}. \quad \text{(D20)}$$

To find a formula for s_c, we boldly expand these equations in leading powers of y while preserving extra powers of c^4. By solving Eq. (D20) numerically, we see that $y \sim 1$, so this simplification is not strictly valid [Fig. 9(b)]; nevertheless, our ultimately derived formula matches reasonably well with numerical results [Fig. 4(e)]. The equations become

$$\alpha' \approx \frac{2s_c c^2 y^4[3s_c c^2 - (3 + s_c c^2)y^2]}{3\pi[s_c c^2 y^4 + 9c^2(1 + s_c c^2)^2]}, \quad \alpha' c^2 \approx \frac{s_c c^2(3 + s_c c^2)y^6}{27\pi(1 + s_c c^2)^2}. \quad \text{(D21)}$$

Equating these two expressions for α', we get

$$s_c c^2(3 + s_c c^2)y^6 + 27(1 + s_c c^2)^2(3 + s_c c^2)c^2 y^2 = 54s_c c^2(1 + s_c c^2)^2 c^2. \quad \text{(D22)}$$

We can solve this equation for y using the cubic formula to obtain

$$y^2 = (\sqrt{A^3 + B^2} + B)^{1/3} - (\sqrt{A^3 + B^2} - B)^{1/3}, \quad \text{where} \quad A = \frac{9(1 + s_c c^2)^2 c^2}{s_c c^2}, \quad B = \frac{27(1 + s_c c^2)^2 c^2}{3 + s_c c^2}. \quad \text{(D23)}$$

Substituting this expression into Eq. (D21), we find an equation for α' in terms of s_c:

$$\alpha' = \frac{s_c c^2}{\pi B}\{2B - 3A[(\sqrt{A^3 + B^2} + B)^{1/3} - (\sqrt{A^3 + B^2} - B)^{1/3}]\}. \quad \text{(D24)}$$

Finally, we can solve for s_c as a series in α'. We keep only the leading term in α' and the leading term in c to obtain

$$s_c \approx 3 \left(\frac{3\pi}{8} \right)^{1/4} \frac{(\alpha')^{1/4}}{c^{3/2}} + \frac{3\pi}{8} \frac{\alpha'}{c^2}. \tag{D25}$$

This yields Eq. (25) if we convert α' back to α with Eq. (D12).

[1] B. McNaughton and R. Morris, Hippocampal synaptic enhancement and information storage within a distributed memory system, Trends Neurosci. **10**, 408 (1987).

[2] R. C. O'Reilly and J. W. Rudy, Conjunctive representations in learning and memory: Principles of cortical and hippocampal function, Psychol. Rev. **108**, 311 (2001).

[3] E. T. Rolls and R. P. Kesner, A computational theory of hippocampal function, and empirical tests of the theory, Prog. Neurobiol. **79**, 1 (2006).

[4] J. J. Hopfield and D. W. Tank, "Neural" computation of decisions in optimization problems, Biol. Cybern. **52**, 141 (1985).

[5] A. Barra, A. Bernacchia, E. Santucci, and P. Contucci, On the equivalence of Hopfield networks and Boltzmann machines, Neural Netw. **34**, 1 (2012).

[6] D. Marr, Simple memory: A theory for archicortex, Philos. Trans. R. Soc. B **262**, 23 (1971).

[7] M. V. Tsodyks and M. V. Feigel'man, The enhanced storage capacity in neural networks with low activity level, Europhys. Lett. **6**, 101 (1988).

[8] P. Kanerva, *Sparse Distributed Memory* (MIT Press, Cambridge, MA, 1988).

[9] J.-P. Nadal and G. Toulouse, Information storage in sparsely coded memory nets, Netw. Comput. Neural Syst. **1**, 61 (1990).

[10] E. T. Rolls and A. Treves, The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain, Netw. Comput. Neural Syst. **1**, 407 (1990).

[11] A. Treves and E. T. Rolls, What determines the capacity of autoassociative memories in the brain? Netw. Comput. Neural Syst. **2**, 371 (1991).

[12] G. Palm, Neural associative memories and sparse coding, Neural Netw. **37**, 165 (2013).

[13] J. F. Fontanari, Generalization in a Hopfield network, J. Phys. **51**, 2421 (1990).

[14] D. A. Stariolo and F. A. Tamarit, Generalization in an analog neural network, Phys. Rev. A **46**, 5249 (1992).

[15] David Renato Carreta Dominguez, Information capacity of a hierarchical neural network, Phys. Rev. E **58**, 4811 (1998).

[16] L. Kang and T. Toyoizumi, Distinguishing examples while building concepts in hippocampal and artificial networks, bioRxiv:2023.02.21.529365 (2023).

[17] D. Amaral and L. Pierre, Hippocampal neuroanatomy, in *The Hippocampus Book*, edited by P. Andersen, R. Morris, D. Amaral, T. Bliss, and J. O'Keefe (Oxford University Press, Oxford, 2006), pp. 37–114.

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

[19] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Spin-glass models of neural networks, Phys. Rev. A **32**, 1007 (1985).

[20] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation*, Santa Fe Institute Studies in the Sciences of Complexity: Lecture Notes No. 1 (CRC Press, Boca Raton, FL, 2018).

[21] G. Weisbuch and F. Fogelman-Soulie, Scaling laws for the attractors of Hopfield networks, J. Phys. Lett. **46**, 623 (1985).

[22] M. Mézard and M. A. Virasoro, The microstructure of ultrametricity, J. Phys. **46**, 1293 (1985).

[23] V. S. Dotsenko, 'Ordered' spin glass: A hierarchical memory machine, J. Phys. C: Solid State Phys. **18**, L1017 (1985).

[24] C. Cortes, A. Krogh, and J. A. Hertz, Hierarchical associative networks, J. Phys. A: Math. Gen. **20**, 4449 (1987).

[25] M. A. Virasoro, The effect of synapses destruction on categorization by neural networks, Europhys. Lett. **7**, 293 (1988).

[26] H. Gutfreund, Neural networks with hierarchically correlated patterns, Phys. Rev. A **37**, 570 (1988).

[27] A. Krogh and J. A. Hertz, Mean-field analysis of hierarchical associative networks with 'magnetisation', J. Phys. A: Math. Gen. **21**, 2211 (1988).

[28] A. Treves and E. T. Rolls, Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network, Hippocampus **2**, 189 (1992).

[29] R. C. O'Reilly and J. L. McClelland, Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off, Hippocampus **4**, 661 (1994).

[30] W. E. Vinje and J. L. Gallant, Sparse coding and decorrelation in primary visual cortex during natural vision, Science **287**, 1273 (2000).

[31] M. T. Wiechert, B. Judkewitz, H. Riecke, and R. W. Friedrich, Mechanisms of pattern decorrelation by recurrent neuronal circuits, Nat. Neurosci. **13**, 1003 (2010).

[32] X. Pitkow and M. Meister, Decorrelation and efficient coding by retinal ganglion cells, Nat. Neurosci. **15**, 628 (2012).

[33] N. A. Cayco-Gajic, C. Clopath, and R. A. Silver, Sparse synaptic connectivity is required for decorrelation and pattern separation in feedforward networks, Nat. Commun. **8**, 1116 (2017).

[34] S. Kim, S. J. Guzman, H. Hu, and P. Jonas, Active dendrites support efficient initiation of dendritic spikes in hippocampal CA3 pyramidal neurons, Nat. Neurosci. **15**, 600 (2012).

[35] J. Makara and J. Magee, Variable dendritic integration in hippocampal CA3 pyramidal neurons, Neuron **80**, 1438 (2013).

[36] P. Kaifosh and A. Losonczy, Mnemonic functions for nonlinear dendritic integration in hippocampal pyramidal circuits, Neuron **90**, 622 (2016).

[37] W. B. Scoville and B. Milner, Loss of recent memory after bilateral hippocampal lesions, J. Neurol. Neurosurg. Psych. **20**, 11 (1957).

[38] L. R. Squire, Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans, Psychol. Rev. **99**, 195 (1992).

[39] J. K. Leutgeb, S. Leutgeb, M.-B. Moser, and E. I. Moser, Pattern separation in the dentate gyrus and CA3 of the hippocampus, Science **315**, 961 (2007).

[40] J. B. Aimone, W. Deng, and F. H. Gage, Resolving new memories: A critical look at the dentate gyrus, adult neurogenesis, and pattern separation, Neuron **70**, 589 (2011).

[41] B. J. Knowlton and L. R. Squire, The learning of categories: Parallel brain systems for item memory and category knowledge, Science **262**, 1747 (1993).

[42] D. Zeithamova, W. T. Maddox, and D. M. Schnyer, Dissociable prototype learning systems: Evidence from brain imaging and behavior, J. Neurosci. **28**, 13194 (2008).

[43] A. C. Schapiro, E. Gregory, B. Landau, M. McCloskey, and N. B. Turk-Browne, The necessity of the medial temporal lobe for statistical learning, J. Cognit. Neurosci. **26**, 1736 (2014).

[44] M. L. Mack, B. C. Love, and A. R. Preston, Dynamic updating of hippocampal object representations reflects new conceptual knowledge, Proc. Natl. Acad. Sci. USA **113**, 13203 (2016).

[45] N. V. Covington, S. Brown-Schmidt, and M. C. Duff, The necessity of the hippocampus for statistical learning, J. Cognit. Neurosci. **30**, 680 (2018).

[46] R. Quian Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, Invariant visual representation by single neurons in the human brain., Nature (London) **435**, 1102 (2005).

[47] R. Quian Quiroga, A. Kraskov, C. Koch, and I. Fried, Explicit encoding of multimodal percepts by single neurons in the human brain, Curr. Biol. **19**, 1308 (2009).

[48] G. Buzsáki, Theta oscillations in the hippocampus, Neuron **33**, 325 (2002).

[49] A. Litwin-Kumar, K. D. Harris, R. Axel, H. Sompolinsky, and L. F. Abbott, Optimal degrees of synaptic connectivity, Neuron **93**, 1153 (2017).

[50] J. M. Jeanne, M. Fişek, and R. I. Wilson, The organization of projections from olfactory glomeruli onto higher-order neurons, Neuron **98**, 1198 (2018).

[51] J. F. Medina, J. C. Repa, M. D. Mauk, and J. E. LeDoux, Parallels between cerebellum- and amygdala-dependent conditioning, Nat. Rev. Neurosci. **3**, 122 (2002).

[52] T. Hige, Y. Aso, G. M. Rubin, and G. C. Turner, Plasticity-driven individualization of olfactory coding in mushroom body output neurons, Nature (London) **526**, 258 (2015).

[53] A. D. Bruce, E. J. Gardner, and D. J. Wallace, Dynamics and statistical mechanics of the Hopfield model, J. Phys. A: Math. Gen. **20**, 2909 (1987).