


Recognition capabilities of a Hopfield model with auxiliary hidden neuronsMarco Benedetti,¹ Victor Dotsenko,² Giulia Fischetti,¹ Enzo Marinari^{1,3} ,^{1,3} and Gleb Oshanin²¹*Università di Roma La Sapienza, Piazzale Aldo Moro 5, I-00185 Rome, Italy*²*Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée (UMR 7600),
4 Place Jussieu, F-75252 Paris Cedex 05, France*³*CNR-Nanotec and INFN, Sezione di Roma 1, I-00185 Rome, Italy*

(Received 18 January 2021; accepted 24 May 2021; published 11 June 2021)

We study the recognition capabilities of the Hopfield model with auxiliary hidden layers, which emerge naturally upon a Hubbard-Stratonovich transformation. We show that the recognition capabilities of such a model at zero temperature outperform those of the original Hopfield model, due to a substantial increase of the storage capacity and the lack of a naturally defined basin of attraction. The modified model does not fall abruptly into the regime of complete confusion when memory load exceeds a sharp threshold. This latter circumstance, together with an increase of the storage capacity, renders such a modified Hopfield model a promising candidate for further research, with possible diverse applications.

DOI: [10.1103/PhysRevE.103.L060401](https://doi.org/10.1103/PhysRevE.103.L060401)**I. INTRODUCTION**

Starting from the seminal paper by Little [1], it was realized that disordered spin systems can store information, working as content-addressable memories. In this context, the model proposed by Hopfield [2] (\mathcal{H} model from now on) has often served as a reference. Its phase diagram [3–5] contains a *retrieval phase*, where one can use a system composed of N neurons to “store” patterns containing N symbols. By storage one means that patterns can be recovered: starting from the exact pattern we have stored or from a damaged pattern, where a fraction η of the spins do not coincide with the configuration we want to retrieve, we end up close enough to it. Despite being robust in many respects, this model has one essential shortcoming: If $\alpha \equiv P/N$ is larger than a critical value $\alpha_c \approx 0.138$, it is impossible to store more than P different uncorrelated patterns. When $\alpha > \alpha_c$, the network enters a state of complete confusion, every memory is abruptly forgotten and no pattern can be retrieved. This phenomenon is known as *blackout catastrophe*. During previous decades, several approaches have been proposed to remedy this and other related issues (see, for example, [6–21]). In particular, in Ref. [22] it was demonstrated that there is always a kind of trade-off between an increase of the learning and retrieval capacity and a certain deterioration of the functionality of a learning algorithm. For example, α_c can be made equal to 1 upon implementing a pseudoinverse rule, instead of the Hebb training algorithm. A penalty one has to pay is that the former, in contrast to the Hebb rule, is never incremental and local only if not immediate. In Ref. [22] a novel local, incremental, and immediate algorithm has been suggested, which permitted one to increase α_c by a factor of $\sqrt{2}$. In Ref. [23] a different strategy has been employed in which the standard quadratic function of the overlap between the patterns and the spin configuration is replaced by higher order polynomials. In doing

so, one increases quite substantially the value of α_c but has to keep track of the ensuing multispin interactions.

In this paper, we tackle the problems of a low storage capacity and of an abrupt transition into a regime of complete confusion using a different approach, based on a hidden-layer representation of \mathcal{H} . The paper is outlined as follows: In Sec. II we introduce basic definitions and describe the dynamical learning algorithm. In Sec. III we present our main results. Lastly, in Sec. IV we conclude with a brief recapitulation of our findings.

II. MODELS AND TECHNIQUES

The \mathcal{H} model [2] describes a system of N binary neurons $\sigma_i = \pm 1$, $i \in \{1, \dots, N\}$, with a long-range spin-glass-like Hamiltonian [24],

$$H[J, \sigma] = -1/2 \sum_{i,j=1}^N J_{ij} \sigma_i \sigma_j, \quad J_{ij} = \frac{1}{N} \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu.$$

The quenched coupling matrix J_{ij} is defined according to Hebb’s learning rule [25], where $\xi_i^\mu = \pm 1$, $\mu \in \{1, \dots, P\}$, $i \in \{1, \dots, N\}$ are the P configurations (*patterns*) that we want to be able to retrieve. The partition function reads

$$Z = \sum_{\{\sigma\}} \exp \left\{ \frac{\beta}{2N} \sum_{i,j=1}^N \sum_{\mu=1}^P \xi_i^\mu \xi_j^\mu \sigma_i \sigma_j \right\}, \quad (1)$$

where β is the inverse temperature of the system. Recently in [26] it has been shown that \mathcal{H} can be thought of as the result of a Hubbard-Stratonovich transformation,

$$Z = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \prod_{\mu=1}^P dX_\mu \sum_{\sigma} \exp \{-\beta \tilde{H}[\xi, X, \sigma]\}, \quad (2)$$

where the X_μ are P Gaussian auxiliary variables and

$$\tilde{H}[\xi, X, \sigma] \equiv N/2 \sum_{\mu=1}^P X_\mu^2 + \sum_{\mu=1}^P \sum_{i=1}^N \sigma_i \xi_i^\mu X_\mu. \quad (3)$$

Note that \tilde{H} is extensive since the X_μ variables are typically of order $N^{-\frac{1}{2}}$ (see [26]).

Integrating over the X_μ in Eq. (2) leads back to Eq. (1), and corresponds to the assumption of complete thermalization of the X_μ variables. In this work, we follow a different strategy: We regard the model defined by Eq. (2) as fundamental (X model from now on), considering the continuous auxiliary variables as hidden neurons in our system. The hidden neurons enter the learning dynamics on the same footing as the two-state σ variables, which constitute the input layer of the network. At $T = 0$ energy barriers can and do break the equivalence among the two models, making joint thermalization of the X_μ and σ_i variables impossible. Numerical simulations convincingly demonstrate that this has drastic consequences on the retrieval properties of the system.

At $T = 0$, the recognition process in the X model is led by the steepest descent procedure, with sequential updating. One sweep is composed of two steps. First one fixes all the X_μ variables to minimize the energy given σ_i : The optimal value is known analytically and it is $X_\mu^{\text{opt}} = -(1/N) \sum_i \sigma_i \xi_i^\mu$. We fix all the X_μ to this value. Then one updates all the σ_i for fixed $\{X_\mu\}$. On each site σ_i is set to ± 1 , such to minimize the energy. The two steps are repeated until a local energy minimum is reached: This is signaled by the fact that no X_μ and no σ_i change.

The biological interpretation of the X model differs markedly from that of \mathcal{H} : The structure of the interactions in X is represented by a bipartite graph, since there is no direct interaction between hidden neurons nor between two-state neurons. Moreover, the strength of the synapses connecting the two populations is not continuous, but can take only two values. It is indeed plausible that discreteness in synaptic strength plays a role in biological systems, since quantized synapses should be more stable against noise. Such discrete synaptic strength has been experimentally measured [27] and explored in many theoretical works (see, for example, [28–30]). Another crucial difference between X and \mathcal{H} is that the number of synaptic connections and neurons used in \mathcal{H} is entirely determined by the number N of bits of information contained in each memory. On the other hand, in the X model the size of the memories only determines the size of the input layer, while the number of synaptic connections and the number of hidden neurons needed to memorize P patterns depends on P .

In what follows we are going to compare the recognition performance of the two models on the same task, namely memorizing $P = \alpha N$ patterns, each composed of N bits. This seems a natural choice, since the two models are identical at $T \neq 0$ under this condition. Since the two networks are so different, one could in principle compare them on other grounds, for example, demanding that they store the same number of patterns per neuron. Even when we include in the counting the X_μ degrees of freedom and we redefine α for the X model accordingly, the comparison is still strongly in favor of the X model for $\eta = 0$, although the advantage shrinks as η is increased.

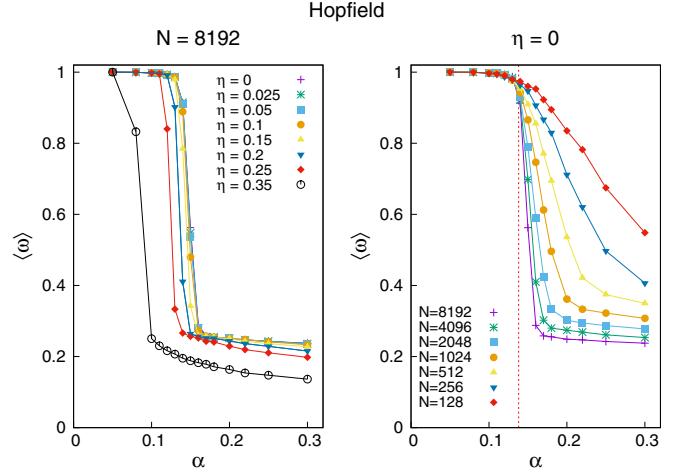


FIG. 1. (Left) The average overlap $\langle \omega \rangle$ as a function of α for different η values at $N = 8192$, for \mathcal{H} . (Right) $\langle \omega \rangle$ as a function of α for different N values at $\eta = 0$, for \mathcal{H} .

We study both the X model and \mathcal{H} for different values of α , η , and N , with $\alpha = 0.05, 0.08, 0.1–0.18$ with increments of $0.01, 0.2, 0.22, 0.25, 0.3$ (for the X model we have also added simulations at higher α values, both below and above one). We have used $\eta = 0, 0.01, 0.025, 0.05, 0.1, 0.15, 0.2, 0.25$, and 0.35 . For both systems we have studied 10^5 samples for $N = 128$ and $256, 210^4$ samples for $N = 512, 10^3$ samples for $N = 1024, 10^2$ samples for $N = 2048, 20$ samples for $N = 4096$, and a small variable number of samples (normally of order 10) for $N = 8192$.

III. MAIN RESULTS

(a) *Finite T Monte Carlo.* We first analyzed the finite temperature structure of both the \mathcal{H} and the X model, and verified that at $T > 0$ they give the same results. We have implemented an annealing protocol to make this in an effective and controlled way.

(b) *The average overlap $\langle \omega \rangle$.* Next we work at $T = 0$, with the steepest descent procedure described above. Here the X model is allowed to behave differently from \mathcal{H} . We begin by measuring the average overlap $\langle \omega \rangle := \langle N^{-1} \sum_i \xi_i^\mu \sigma_i \rangle$ between the memorized pattern ξ^μ in the neighbourhood of which the recognition process starts and the stable spin configuration $\{\sigma_i\}_{i=1}^N$ where energy minimization ends. The average is taken over both the many samples collected and the P memories within a sample.

Our results for \mathcal{H} are shown in Fig. 1. On the left, we use the largest available value of N and plot the values of $\langle \omega \rangle$ for different values of η . On the right, we select $\eta = 0$ and show the N dependence of $\langle \omega \rangle$. The \mathcal{H} model behaves as expected (see [3]): Increasing N at $\eta = 0$ the well-known transition forms close to $\alpha \sim 0.136$, where a vertical dashed line is drawn. Upon increasing η , the transition moves to lower values of α , but stays very similar in nature and shape. Even at our highest value of $\eta = 0.35$ (where the overlap of the starting point with the original pattern is as low as 0.3), at low $\alpha \lesssim 0.05$ the system is still in the recognition phase. Having accurate data for large systems, we are also able to use finite

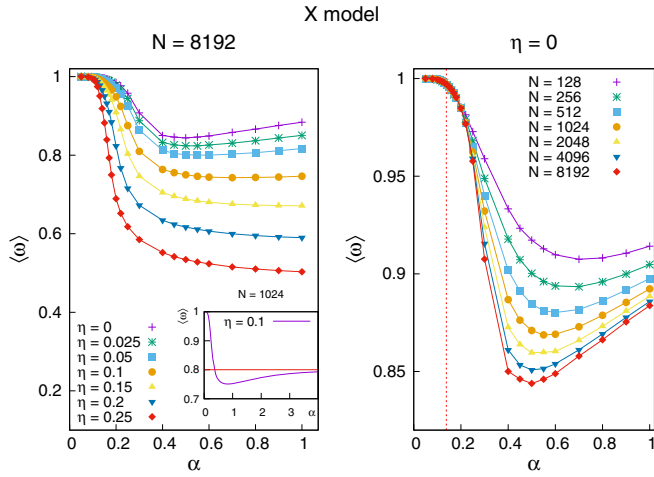


FIG. 2. As in Fig. 1, but for X.

size scaling for a quantitative analysis of these effects (these precision measurements could not fit in this article, and will be reported in a subsequent publication [31]).

As shown in Fig. 2, the behavior of the X model is very different. In the $\eta = 0$ case the sharp transition of \mathcal{H} contrasts with a smooth, nonmonotonous behavior of the X model. At low α we still get a recognition phase, which persists to higher values of α than in \mathcal{H} (and we will attribute to this effect some importance). Increasing α we see that $\langle \omega \rangle$ smoothly decreases, and reaches a minimum close to $\alpha = 0.05$. Here, for $N = 8192$ we have $\langle \omega \rangle \sim 0.84$, but finite size effects are strong. Increasing α further, $\langle \omega \rangle$ starts to grow.

This second, high- α , regime does not correspond to a recognition phase. To gain some intuition of this, notice that the number of hidden neurons X_μ in our model is equal to P . Hence, it is clear that for very large α the X_μ are numerous enough to satisfy, by themselves, all the constraints of the problem. In turn, this implies that, as $\alpha \rightarrow \infty$, any configuration σ can be accommodated in an energy minimum simply by relaxing the hidden neurons to their optimal value, making the dynamic ineffective. The behavior of the system when $\eta > 0$ is very telling. As in \mathcal{H} , we still have a recognition region at low α , which shrinks for increasing η . We still have a smooth decrease of $\langle \omega \rangle$ for increasing α , and an asymptotic slow increase that slows down for increasing η . The asymptotic value for $\alpha \rightarrow \infty$ is exactly the initial overlap $1 - 2\eta$ (inset of Fig. 2). This clearly confirms that the large α regime is not a recognition regime, but rather a regime of ineffective dynamics. To get a deeper insight about this asymptotic behavior, we can use the same technique adopted in [32], and look into what happens under a step of the steepest descent dynamics. Consider any binary neuron configuration σ . By plugging the expression for X_μ^{opt} into the Hamiltonian Eq. (3), one sees that the change in energy upon flipping σ_j , after the X_μ variables have thermalized to their optimal value given σ and the patterns ξ^μ , is

$$\Delta_j \tilde{H} = \frac{2}{N} \sum_{\mu=1}^P \left(1 + \sigma_j \xi_j^\mu \sum_{h \neq j} \sigma_h \xi_h^\mu \right). \quad (4)$$

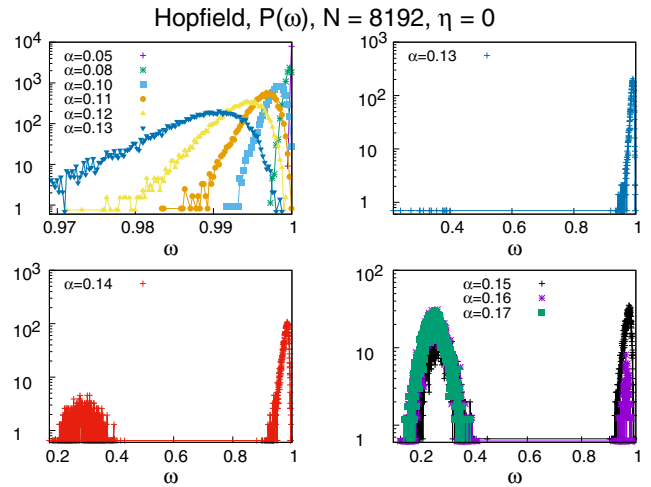
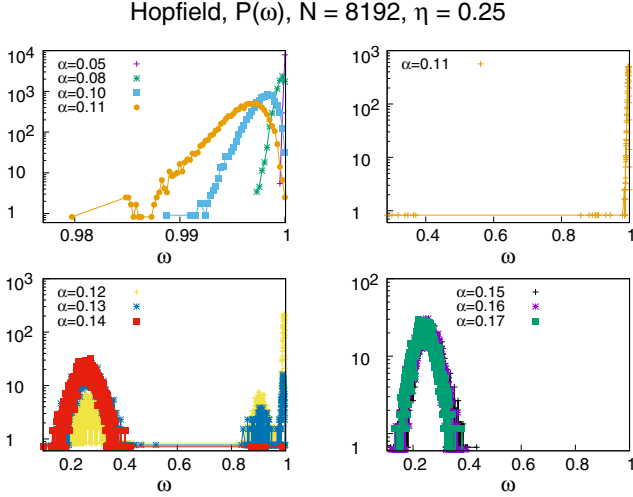


FIG. 3. The overlap probability distribution $P(\omega)$ for different α values at $\eta = 0$ and $N = 8192$, for \mathcal{H} .

The second contribution in the brackets is what one gets for \mathcal{H} (it shows that σ_j is pulled by all the memories ξ^μ , with a strength proportional to the overlap between the memory ξ^μ and σ), while the additional 1 comes because of the X variable interaction. It is a memory independent constant price that one has to pay, due to the fact that we are flipping a spin “against the will” of the X_μ , which were optimal for σ . This effectively introduces a threshold in the dynamics of σ , and a stabilizing effect for the configuration of the binary neurons. This stabilizing effect dominates the dynamics in the high α region, making every configuration stable, and the network useless. On the contrary, and we take this as one of our important findings, the presence of the X_μ degrees of freedom helps the learning in the small α regime, enlarging the recognition region as compared to \mathcal{H} , and, what is maybe even more important, eliminating complete confusion for α larger than a sharp threshold.

(c) *The probability distribution $P(\omega)$.* Even if $\langle \omega \rangle$ is giving us a good amount of information, it is appropriate to analyze the behavior of the full probability distribution $P(\omega)$. In Fig. 3 we show $P(\omega)$ at $\eta = 0$ for different values of α , for \mathcal{H} . From left to right and from top to bottom we plot $P(\omega)$ for increasing values of α . The horizontal scales of the four frames are very different. Plots are in the linear-log scale. We first show (top left) results for the low values of α , where $P(\omega)$ is concentrated close to $\omega = 1$. Going right from there we plot again $\alpha = 0.13$, to show that, just below the critical point, a few points are already at low overlap. The number of these points decreases as N increases. In the bottom left frame we have $\alpha = 0.14$ where, as expected, we have a bi-modality. The peak close to 1 is, at this value of N , still leading (remember that the y scale is a log scale), but a peak at $\omega \sim 0.30$ has appeared. On the right we have the high values of α . Now the low ω peaks start to dominate. Their location is very stable, and only shifts very lightly.

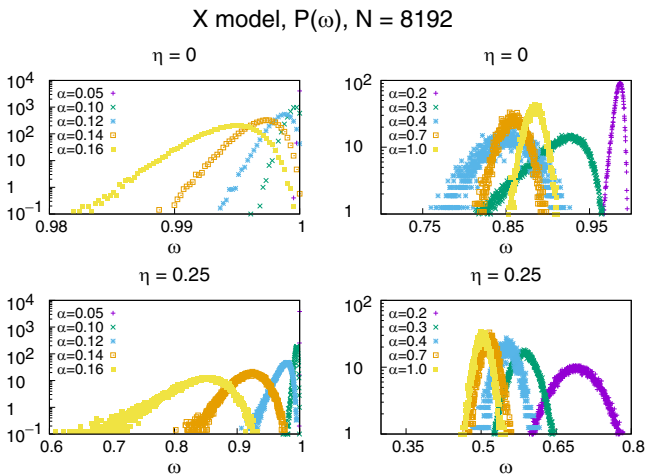
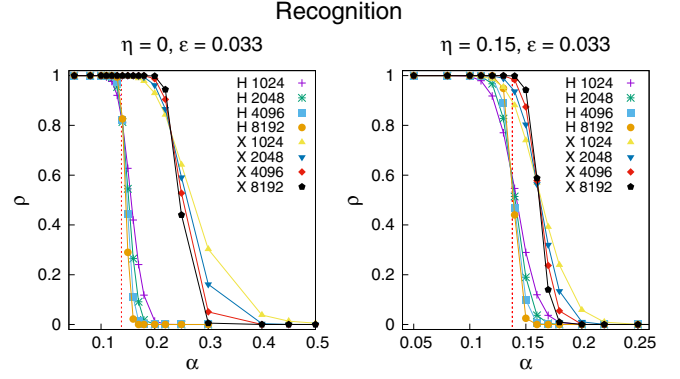
In Fig. 4 we analyze the case $\eta = 0.25$ for \mathcal{H} . After a rescaling of the value of α everything is analogous to $\eta = 0$. The position of the low ω peak is again remarkably constant. The only clear difference is that here at high-intermediate


 FIG. 4. As in Fig. 3, but for $\eta = 0.25$.

α value a three peaks structure is visible (there is a clear peak at high $\omega < 1$). All together, we find for the structure of \mathcal{H} exactly what we expected.

As we show in Fig. 5, things are again very different for the X model. Here we only need two frames for each value of η to clearly show our data. In the X model we do not see any trace of bi-modality, but only a smooth behavior. For both $\eta = 0$ and $\eta = 0.25$ at low α we see that the mass of the distribution is centered close to 1. When α increases the distribution first shifts to lower α values, and eventually to larger ones, its $N \rightarrow \infty$ limit developing a narrow peak and being centered at $1 - 2\eta$.

(d) *The recognition rate.* In order to get further insight, we define the *recognition rate* ρ as the probability that a minimization run ends with $\omega \geq 0.967$ [3]: The threshold for recognition is $\epsilon = 1 - 0.967 = 0.033$. In \mathcal{H} , that undergoes a sharp transition, selecting a different threshold would give the same asymptotic result. We show in Fig. 6 ρ as a function of α , both for the \mathcal{H} and X model, for $\eta = 0$ in the left frame and for $\eta = 0.15$ in the right one. Even if, as we have seen in


 FIG. 5. The probability distribution $P(\omega)$ for different α values at $\eta = 0$ and $\eta = 0.25$, for X.

 FIG. 6. Comparison between the recognition rate ρ as a function of α in \mathcal{H} and X, for different N values, at $\eta = 0$ and $\eta = 0.15$. The recognition threshold is $\epsilon = 0.033$.

detail, the two models work very differently, the plots for the \mathcal{H} and X model are similar, both at $\eta = 0$ and at $\eta > 0$. The X model has a wider learning phase. We can say that there is a very low α regime where the new X variables are irrelevant since they are not needed for recognition, and a very large α regime where they fix the system on the observed pattern, but cannot lead to recognition. Only in the region where α is slightly larger than α_c , they are put to good use, and help in the memorizing. Also, when α increases, they avoid complete confusion: The X memory becomes less efficient if too many patterns are shown, but the *blackout catastrophe* is avoided.

(e) *The dynamical exponent.* We have also analyzed how the computational cost of the recall process scales with N . We assume that the number of sweeps \mathcal{S} needed to reach the stable state scales asymptotically as N^ζ , plus N -dependent subleading corrections [33–40]. In a sweep we include, for the X model, both the cost of putting the X_μ in their optimal position and the cost of updating every σ_i once. In the absence of any slowing down we expect to find $\zeta = 0$. We define an effective exponent dependent on two values of N as

$$\zeta(N_1, N_2) \equiv \log(\mathcal{S}^{(N_1)}/\mathcal{S}^{(N_2)})/\log(N_1/N_2). \quad (5)$$

The effective exponent $\zeta(N_1, N_2)$ is a finite size estimate of ζ and asymptotically for large N_1 and N_2 one has $\zeta(N_1, N_2) \rightarrow \zeta$. In Fig. 7 we show $\zeta(N_1, N_2)$ as a function of α for \mathcal{H} . On the right $\eta = 0$ and different couples of N are considered; on the left we consider different η values, using $N_1 = 4096$ and $N_2 = 8192$. The effective exponent for $\eta = 0$ is small at small α , and develops a narrow peak of value close to one at α_c . The peak is becoming narrower as the lattice size increases, and is already very thin when we consider our largest systems, with $N_1 = 4096$ and $N_2 = 8192$, suggesting that its asymptotic width is zero. For $\alpha > \alpha_c$ the exponent plateaus, at a value close to 0.6. As expected \mathcal{H} is critically close to α_c . The situation for $\eta > 0$ is similar, but that at $\eta = 0.35$ where a sizable critical peak cannot be detected anymore.

In Fig. 8 we have the same plot for the X model and, again, here the situation is different. There is still a peak at low α (although larger than α_c), but the N dependence does not suggest an asymptotically null width (even though one would need very large values of N to make sharp claims about

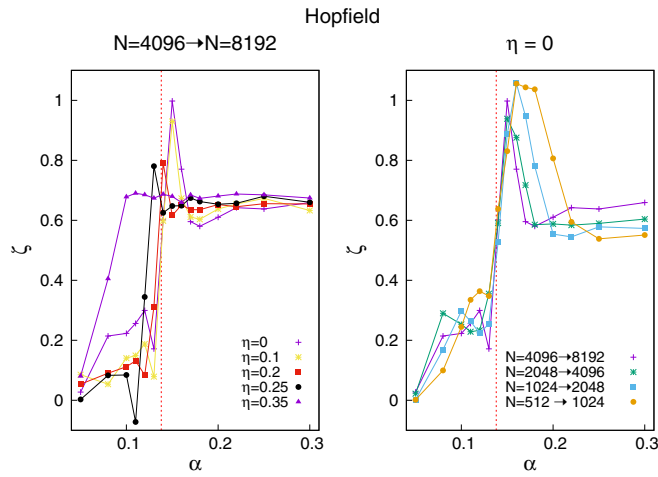


FIG. 7. (Left) The effective exponent $\zeta(4096, 8192)$ as a function of α for different η values, for \mathcal{H} . (Right) The effective exponent $\zeta(N_1, N_2)$ as a function of α for different values of N_1 and N_2 at $\eta = 0$, for \mathcal{H} .

this). Also the peak at $\eta > 0$ is very different from \mathcal{H} , and the effective exponent has a smooth slow decay for large α .

IV. CONCLUSIONS

In this letter we have discussed how the introduction of hidden layers in the Hopfield model leads to interesting new features in the zero-temperature associative memory performance. In our model, the probability distribution of the overlap as well as its average value differ markedly from the ones in the Hopfield model. As a consequence, the recognition performance is improved. What we find really crucial is that the interaction between visible and hidden neurons has

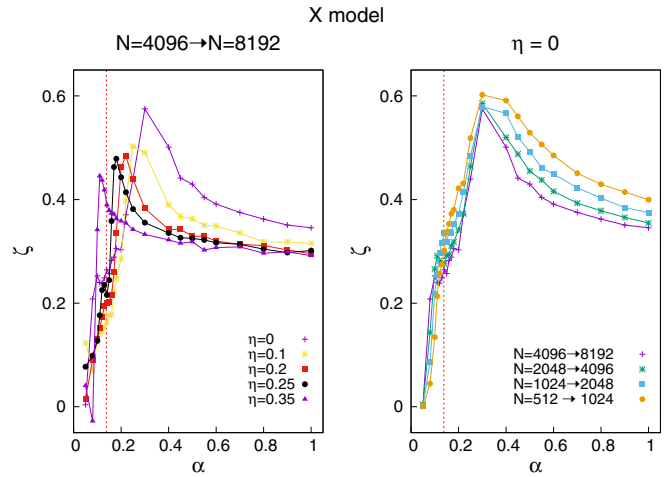


FIG. 8. As in Fig. 7, but for X.

a stabilizing effect on the zero-temperature dynamics, which prevents the *blackout catastrophe*: The system does not forget everything above a sharp threshold α_c . This, together with the smaller value of the dynamical scaling exponent, implying a faster recognition process, suggests that our atypical hidden layers may considerably improve the functioning of Hopfield-like neural systems. This opens an interesting perspective for further research in the field of associative memory.

ACKNOWLEDGMENTS

We are very grateful to Stefano Fusi and Marc Mézard for precious conversations. We have been supported by funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant No. 694925-LoTgGlasSy).

[1] W. A. Little, *Mathematical Biosciences* **19**, 101 (1974).
 [2] J. Hopfield, *Proc. Natl. Acad. Sci.* **79**, 2554 (1982).
 [3] D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys.* **173**, 30 (1987).
 [4] J. L. van Hemmen and R. Kühn, *Phys. Rev. Lett.* **57**, 913 (1986).
 [5] B. M. Forrest, *J. Phys. A: Math. Gen.* **21**, 245 (1988).
 [6] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Phys. Rev. E* **84**, 066106 (2011).
 [7] E. Gardner and B. Derrida, *J. Phys. A: Math. Gen.* **21**, 271 (1988).
 [8] E. Rolls and A. Treves, *Neural Networks and Brain Function* (Oxford University Press, Oxford, 1997).
 [9] S. Fusi, P. J. Drew, and L. F. Abbott, *Neuron* **45**, 599 (2005).
 [10] S. Fusi and L. F. Abbott, *Nature Neuroscience* **10**, 485 (2007).
 [11] A. Barra, A. Bernacchia, E. Santucci, and P. Contucci, *Neural Networks* **34**, 1 (2012).
 [12] E. Agliari, A. Barra, A. De Antoni, and A. Galluzzi, *Neural Networks* **38**, 52 (2013).
 [13] E. Agliari, A. Barra, A. Galluzzi, F. Guerra, D. Tantari, and F. Tavani, *Phys. Rev. Lett.* **114**, 028103 (2015).
 [14] C. Baldassi, F. Gerace, H. J. Kappen, C. Lucibello, L. Saglietti, E. Tartaglione, and R. Zecchina, *Phys. Rev. Lett.* **120**, 268103 (2018).
 [15] C. Baldassi, F. Pittorino, and R. Zecchina, *Proc. Natl. Acad. Sci.* **117**, 161 (2020).
 [16] F. Schönsberg, Y. Roudi, and A. Treves, *Phys. Rev. Lett.* **126**, 018301 (2021).
 [17] G. Parisi, *J. Phys. A: Math. Gen.* **19**, L617 (1986).
 [18] E. Marinari, *Neural Comput.* **31**, 503 (2019).
 [19] J. L. Van Hemmen, G. Keller, and R. Kühn, *Europhysics Letters (EPL)* **5**, 663 (1988).
 [20] J. L. Van Hemmen, L. B. Ioffe, R. Kühn, and M. Vaas, *Physica A* **163**, 386 (1990).
 [21] H. Ramsauer, B. Schäfl, J. Lehner, P. Seidl, M. Widrich, L. Gruber, M. Holzleitner, T. Adler, D. Kreil, M. K. Kopp, G. Klambauer, J. Brandstetter, and S. Hochreiter, Hopfield networks is all you need, in *International Conference on Learning Representations (ICLR)*, (2021).
 [22] A. Storkey, in *Artificial Neural Networks—ICANN’97* (Springer, Berlin/Heidelberg, 1997), pp. 451–456.
 [23] D. Krotov and J. J. Hopfield, in *Advances in Neural Information Processing Systems*, Vol. 29, edited by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Curran Associates, Red Hook, 2016).
 [24] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1986).

- [25] D. O. Hebb, *Journal of Clinical Psychology* **6**, 307 (1950).
- [26] M. Mézard, *Phys. Rev. E* **95**, 022117 (2017).
- [27] C. C. H. Petersen, R. C. Malenka, R. A. Nicoll, and J. J. Hopfield, *Proc. Natl. Acad. Sci.* **95**, 4732 (1998).
- [28] D. J. Amit and S. Fusi, *Netw., Comput. Neural Syst.* **3**, 443 (1992).
- [29] C. Baldassi, A. Braunstein, N. Brunel, and R. Zecchina, *Proc. Natl. Acad. Sci.* **104**, 11079 (2007).
- [30] H. Sompolinsky, in *Heidelberg Colloquium on Glassy Dynamics*, Vol. 275 (Springer, Berlin/Heidelberg, 1987), pp. 485–527.
- [31] M. Benedetti, V. Dotsenko, G. Fischetti, E. Marinari and Gleb Oshanin (unpublished).
- [32] V. Folli, M. Leonetti, and G. Ruocco, *Front. Comput. Neurosci.* **10**, 144 (2017).
- [33] M. E. Fisher, in *Critical Phenomena*, International School of Physics ‘Enrico Fermi’ No. 51, edited by M. S. Green and Società Italiana di Fisica (Academic Press, New York, 1971).
- [34] M. E. Fisher and M. N. Barber, *Phys. Rev. Lett.* **28**, 1516 (1972).
- [35] Y. Imry and D. Bergman, *Phys. Rev. A* **3**, 1416 (1971).
- [36] E. Brézin, *Journal de Physique* **43**, 15 (1982).
- [37] S. Caracciolo, R. G. Edwards, S. J. Ferreira, A. Pelissetto, and A. D. Sokal, *Phys. Rev. Lett.* **74**, 2969 (1995).
- [38] H. G. Ballesteros, L. A. Fernández, V. Martín-Mayor, A. Muñoz Sudupe, G. Parisi, and J. J. Ruiz-Lorenzo, *J. Phys. A* **32**, 1 (1999).
- [39] V. Privman and M. E. Fisher, *J. Stat. Phys.* **33**, 385 (1983).
- [40] M. Gubinelli, Ph.D thesis, Pisa University, 2002.