To appear in Physical Review E 2002

On the storage capacity of attractor neural networks with depressing synapses*

Joaquín J. Torres†, L. Pantic‡ and Hilbert J. Kappen‡
† Institute "Carlos I" for Theoretical and Computational Physics and
Department of Electromagnetism and Matterial Physics,
University of Granada, E-18071-Granada, Spain.
‡ Department of Biophysics, University of Nijmegen,
Geert Grooteplein 21, 6525 EZ Nijmegen,The Netherlands.

(Dated: October 24, 2002)

Abstract

We compute the capacity of a binary neural network with dynamic depressing synapses to store and retrieve an infinite number of patterns. We use a biologically motivated model of synaptic depression and a standard mean-field approach. We find that at $T = 0$ the critical storage capacity decreases with the degree of the depression. We confirm the validity of our main mean field results with numerical simulations.

PACS: 87.10.+e, 87.15.Aa, 05.45.Ra, 75.10.Nr

---

*Corresponding author: Joaquín J. Torres, *Departamento de Electromagnetismo y Física de la Materia, Universidad de Granada, E-18071 Granada, Spain*; e-mail: jtorres@onsager.ugr.es

I. INTRODUCTION

One of the interesting questions that arise in neural network modelling is how biological processes in real neurons at the cellular and subcellular level influence the network behavior and its ability to process information.

In most neural network studies sofar, the synaptic connection between neurons has been modelled as a constant strength [1, 2], possibly subject to learning on a slow time scale. However, recently it has been reported that synaptic strength is a dynamic quantity that strongly depends on the presynaptic neural activity [3, 4]. Synaptic strength can decrease (depression) or increase (facilitation), depending on the type of synapse [3]. After learning, most synapses are depressing with a typical time constant of 10-20 ms. The synaptic strength recovers on the order of seconds. The molecular mechanism underlying this dynamics is the depletion of neurotransmitter's vesicles due to presynaptic neural firing and their restoration on a larger time scale [4]. These synapses are found throughout the cortex as well as in the hippocampus [5] [15].

There have been various studies of the effect of dynamic synapses on the information transfer in feed-forward neural network behaviour, such as filtering of redundant noise [6], extraction of temporal patterns for speech recognition [7] and robust coincidence detection [8].

Only few studies, for instance [9, 10], have focused on the role of dynamical synapses in recurrent neural networks, and in particular on associative memory [1]. In such networks, long-term storage of the memory patterns is produced by adjusting the strength of the synapses according to the Hebb rule. With such static synapses the network dynamics has fixed points (attractors) corresponding to the stored memory patterns [11]. The question one may ask is how the retrieval properties and the fixed points of the Hopfield network are affected by dynamical synapses. Recently, it has been shown numerically that synaptic depression allows for fast switching among stored patterns by reducing the stability of such attractors. These results are observed not only in networks of binary neurons, but also of more biologically plausible integrate-and-fire neurons [10]. These numerical results were confirmed analytically for binary neurons by a mean field analysis in the limit of a small number of stored patterns: $\alpha \equiv P/N \to 0$.

In this paper, we compute the storage capacity of an attractor neural network with dynamical depressing synapses when we store an infinite number of patterns, $P = \alpha N$. In the study presented in [9] the particular situation of strong depression and fast recovery has been analyzed. However, it is known that a large variety of values for the depression and recovery time constants are encountered in the brain. In this paper, we show that only the ratio of these time constants affects the storage capacity and compute the dependence of the storage capacity on this ratio. Our main conclusion, supported by both analytical computations and Monte Carlo simulations is that the critical storage capacity for retrieval of patterns as fixed points is strongly reduced by the degree of depression of the synapses, even at temperature $T = 0$.

II. THE MODEL

Our starting point consists of a network of N binary neurons $s_i = 1, 0$. Each neuron follows a probabilistic dynamics which is given by:

$$\mathrm{Prob}\{s_i(t+1) = 1\} = \frac{1}{2}[1 + \tanh\{2\beta h_i(t)\}], \tag{1}$$

where $h_i = \sum_j \omega_{ij} x_j s_j - \theta_i$ is the local field associated to neuron i, θ_i is the threshold for neuron i to fire and $\beta = \frac{1}{T}$ represents the level of noise due to the stochastic synaptic activity. The first term of h_i is the total synaptic current arriving at neuron i with ω_{ij} representing the static synaptic connection strength or *weights* between neurons j and i and x_j the depression variable [10]. The static weights are defined according to the standard covariance rule, that is:

$$\omega_{ij} = \frac{1}{Nf(1-f)} \sum_{\nu=1}^{P} (\xi_i^\nu - f)(\xi_j^\nu - f), \tag{2}$$

where ξ^ν's ($\nu = 1, \ldots, P$) are P memorized patterns that are stored in the network. The depression variable x_j depends on time and models the dynamic properties of the synapse [10] and is a simplification of the original model by [4] in the limit of $\tau_{\text{in}} = 0$:

$$x_i(t+1) = x_i(t) + \frac{1 - x_i(t)}{\tau_{rec}} - U x_i(t) s_i(t) \ \forall i = 1, \ldots, N. \tag{3}$$

For $\tau_{rec} = 0$ we recover the static synapse ($x_i = 1$).

In the following we shall consider random patterns with average activity $\langle \xi_i^\mu \rangle = f = 1/2$. In order to recover the standard static Hopfield model with $\theta_i = 0$ in the $\{+1, -1\}$ code, we have to choose the neuron threshold in $\{1, 0\}$ code equal to

$$\theta_i(\boldsymbol{\xi}) = \frac{1}{2Nf(1-f)} \sum_j \sum_\mu (\xi_i^\mu - f)(\xi_j^\mu - f). \tag{4}$$

For each stored pattern ν, we define the average of x_j for the active and inactive units in the pattern, that is $x_+^\nu \equiv \frac{1}{Nf} \sum_{j \in Act(\nu)} x_j$, and $x_-^\nu \equiv \frac{1}{N(1-f)} \sum_{j \notin Act(\nu)} x_j$. These sums extend over the units that are active or inactive in the pattern ν, respectively. In a similar way, we define the mean activity of the network for the units which are active and inactive in pattern ν, that is $m_+^\nu \equiv \frac{1}{Nf} \sum_{j \in Act(\nu)} s_j$ and $m_-^\nu \equiv \frac{1}{N(1-f)} \sum_{j \notin Act(\nu)} s_j$. Then, the overlap of the network activity with a stored pattern ν becomes

$$m^\nu \equiv \frac{1}{Nf(1-f)} \sum_i (\xi_i^\nu - f) s_i = m_+^\nu - m_-^\nu, \ \ \forall \nu, \tag{5}$$

and the corresponding local fields

$$h_i = \sum_{\nu=1}^{P} (\xi_i^\nu - f) \left[x_+^\nu m_+^\nu - x_-^\nu m_-^\nu \right]. \tag{6}$$

III. MEAN-FIELD ANALYSIS

From equations (1) and (3) and within a standard mean-field approximation, the network dynamics is driven by the coupled map

$$\begin{aligned}
x_+^\nu(t+1) &= x_+^\nu(t) + \frac{1-x_+^\nu(t)}{\tau_{rec}} - U x_+^\nu(t) m_+^\nu(t), \\
x_-^\nu(t+1) &= x_-^\nu(t) + \frac{1-x_-^\nu(t)}{\tau_{rec}} - U x_-^\nu(t) m_-^\nu(t), \\
m_+^\nu(t+1) &= \frac{1}{N} \sum_{i \in Act}^{(\nu)} \Bigg[1 + \tanh \Bigg\{ \beta \Bigg(\Big[x_+^\nu(t) m_+^\nu(t) - x_-^\nu(t) m_-^\nu(t) \Big] \\
&\quad + \sum_{\mu \neq \nu} (2\xi_i^\mu - 1) \left[x_+^\mu(t) m_+^\mu(t) - x_-^\mu(t) m_-^\mu(t) \right] \Bigg) \Bigg\} \Bigg], \\
m_-^\nu(t+1) &= \frac{1}{N} \sum_{i \notin Act}^{(\nu)} \Bigg[1 - \tanh \Bigg\{ \beta \Bigg(\Big[x_+^\nu(t) m_+^\nu(t) - x_-^\nu(t) m_-^\nu(t) \Big] \\
&\quad - \sum_{\mu \neq \nu} (2\xi_i^\mu - 1) \left[x_+^\mu(t) m_+^\mu(t) - x_-^\mu(t) m_-^\mu(t) \right] \Bigg) \Bigg\} \Bigg], \\
m^\nu(t+1) &= \frac{1}{N} \sum_i (2\xi_i^\nu - 1) \left[1 + \tanh \left\{ \beta \sum_\mu (2\xi_i^\mu - 1) \left[x_+^\mu(t) m_+^\mu(t) - x_-^\mu(t) m_-^\mu(t) \right] \right\} \right].
\end{aligned} \tag{7}$$

The last three equations of (7) are not independent and are related by Eq. (5). In the thermodynamic limit ($N \to \infty$) the fixed point equations for the coupled dynamics of neurons and synapses (7) is given as

$$\begin{aligned}
x_+^\nu &= \frac{1}{\gamma m_+^\nu + 1} \\
x_-^\nu &= \frac{1}{\gamma m_-^\nu + 1} \\
m_+^\nu &= \frac{1}{2} \left[1 + \frac{2}{N} \sum_{i \in Act}^{(\nu)} \tanh \left\{ \beta \left(\left[x_+^\nu m_+^\nu - x_-^\nu m_-^\nu \right] + \sum_{\mu \neq \nu} \epsilon_i^\mu \left[x_+^\mu m_+^\mu - x_-^\mu m_-^\mu \right] \right) \right\} \right] \\
m_-^\nu &= \frac{1}{2} \left[1 - \frac{2}{N} \sum_{i \notin Act}^{(\nu)} \tanh \left\{ \beta \left(\left[x_+^\nu m_+^\nu - x_-^\nu m_-^\nu \right] + \sum_{\mu \neq \nu} (-\epsilon_i^\mu) \left[x_+^\mu m_+^\mu - x_-^\mu m_-^\mu \right] \right) \right\} \right] \\
m^\nu &= \frac{1}{N} \sum_i \epsilon_i^\nu \tanh \left\{ \beta \sum_\mu \epsilon_i^\mu \left[x_+^\mu m_+^\mu - x_-^\mu m_-^\mu \right] \right\},
\end{aligned} \tag{8}$$

where $\gamma \equiv \tau_{rec} U$ and $\epsilon_i^\mu \equiv 2\xi_i^\mu - 1$.

The sum in the last equation of (8) can be seen as an average over the random patterns (quenched disorder). We follow the standard approach of [2] (AGS) and assume that the network has a macroscopic overlap with one pattern (*condensed patterns*) in the steady state ($\nu = 1$), and the remaining $P - 1$ overlaps are $\mathcal{O}(\frac{1}{\sqrt{N}})$. Then, it is straightforward to see that both m_+^μ and m_-^μ are $\mathcal{O}(\frac{1}{\sqrt{N}})$, and x_+^μ, $x_-^\mu = 1 - \mathcal{O}(\frac{1}{\sqrt{N}})$ for $\mu \neq 1$. Introducing new overlap variables $M^\mu \equiv x_+^\mu m_+^\mu - x_-^\mu m_-^\mu$, and using the previous reasoning we have that $M^\mu \sim \mathcal{O}(\frac{1}{\sqrt{N}}) \forall \mu \neq 1$. In the thermodynamic limit $M^\mu \approx m^\mu \forall \mu \neq 1$, and the small order parameters M^μ can be considered as independent random variables with mean zero and variance $\alpha r/P$ where $r \equiv \frac{1}{\alpha} \sum_{\mu \neq 1} M^{\mu 2}$.

To perform the average over the random patterns and obtain the final mean-field equations we followed the procedure described in [12]. This consists of splitting the sum appearing in the argument of the tanh in (8) in three parts (one for the condensed pattern $\nu = 1$, one for the particular non-condensed pattern $\mu = \nu$, and a third for the rest of the non-condensed patterns), and performing a Taylor expansion of the tanh around the small overlap M^ν. After neglecting terms involving $(M^\nu)^2$, we obtain

$$m^\nu = \left[\frac{1}{1-\beta(1-q)}\right] \frac{1}{N} \sum_i \epsilon_i^\nu \epsilon_i^1 \tanh\left[\beta\left(M^1 + \sum_{\mu \neq 1,\nu} \epsilon_i^\mu \epsilon_i^1 M^\mu\right)\right] \tag{9}$$

for a non-condensed pattern ν, with

$$q \equiv \frac{1}{N} \sum_i \tanh^2 \left\{\beta \sum_\mu \epsilon_i^\mu M^\mu\right\} \tag{10}$$

the *spin-glass* order parameter [2]. Then, we can compute r by evaluating the square of (9) and obtaining the well known relation between r and q, that is:

$$r = \frac{q}{[1-\beta(1-q)]^2}. \tag{11}$$

Moreover, for the condensed pattern 1 we easily obtain

$$\begin{aligned}
x_+^1 &= \frac{1}{\gamma m_+^1 + 1} \\
x_-^1 &= \frac{1}{\gamma m_-^1 + 1} \\
m^1 &= \frac{1}{N} \sum_i \tanh\left[\beta\left(M^1 + \sum_{\mu \neq 1} \epsilon_i^\mu \epsilon_i^1 M^\mu\right)\right] \\
m_+^1 &= \frac{1}{2}\left\{1 + \frac{2}{N} \sum_{i \in Act}^{(1)} \tanh\left[\beta\left(M^1 + \sum_{\mu \neq 1} \epsilon_i^\mu M^\mu\right)\right]\right\} \\
m_-^1 &= \frac{1}{2}\left\{1 - \frac{2}{N} \sum_{i \notin Act}^{(1)} \tanh\left[\beta\left(M^1 + \sum_{\mu \neq 1} (-\epsilon_i^\mu) M^\mu\right)\right]\right\}.
\end{aligned} \tag{12}$$

Taking into account that $\epsilon_i^\mu \epsilon_i^1$ is random and independent of M^μ and therefore $z \equiv (1/\sqrt{\alpha r}) \sum_{\mu \neq 1} \epsilon_i^\mu \epsilon_i^1 M^\mu$ is a Gaussian noise with zero mean and variance 1, the average $1/N \sum_i$ reduces to an average over the Gaussian noise z. Similarly, ϵ_i^μ is random and independent of M^μ and $z_1 \equiv (1/\sqrt{\alpha r}) \sum_{\mu \neq 1} \epsilon_i^\mu M^\mu$ and $z_2 \equiv (1/\sqrt{\alpha r}) \sum_{\mu \neq 1} (-\epsilon_i^\mu) M^\mu$ are also Gaussian noises with zero mean and variance 1. Then, the sums appearing in the steady-state conditions for $m_\pm^1$, Eq. (12), become averages over these Gaussian noises.

The resulting system of equations becomes:

$$\begin{aligned} u &= \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh\left[\beta\left(\frac{4u}{\gamma^2(1-u^2)+4\gamma+4} + \sqrt{\alpha r} z\right)\right], \\ q &= \int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh^2\left[\beta\left(\frac{4u}{\gamma^2(1-u^2)+4\gamma+4} + \sqrt{\alpha r} z\right)\right], \\ r &= \frac{q}{[1-\beta(1-q)]^2}. \end{aligned} \tag{13}$$

where $u \equiv 2m_+^1 - 1$. To obtain (13) we used the steady state conditions (12), and the fact that in the steady-state the relation $m_+^1 + m_-^1 = 1$ holds. The system (13) is a direct extension of the AGS equations in the standard Hopfield model [2] and reduces to the AGS equations in the limit of $\gamma \to 0$, that is, in the case of non-depressed or *static* synapses.

IV. RESULTS

The system (13) must be solved numerically for finite temperature. In the limit of $T \to 0$ it is easy to compute the critical value of P/N, namely α_c, at which nontrivial ferromagnetic solutions (*Mattis states*) appear. α_c gives the maximun number of patterns that the system is able to retrieve. For $\beta \to \infty$, we use the standard approximations $\int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2}(1 - \tanh^2 \beta[az+b]) \approx \sqrt{\frac{2}{\pi}} \frac{1}{a\beta} e^{-b^2/2a^2}$ and $\int \frac{dz}{\sqrt{2\pi}} e^{-z^2/2} \tanh \beta[az+b]) \approx \operatorname{erf}\left(\frac{b}{\sqrt{2}a}\right)$. Then, introducing a new variable $y = f(u,\gamma)/\sqrt{2\alpha r}$ with $f(u,\gamma) \equiv \frac{4u}{\gamma^2(1-u^2)+4\gamma+4}$, the system (13) reduces to

$$y\left(\sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-y^2}\right) = f[\operatorname{erf}(y), \gamma]. \tag{14}$$

One can easily check that $u = 0$ ($y = 0$) is a solution of (14) for all α. The appearance of an additional non-zero solution indicates the presence of a stable ferromagnetic state. The largest α at which these non-zero solution appears defines the critical storage capacity α_c

at $T = 0$ as a function of the degree of synaptic depression γ. In Fig. 1 we plot the line $\alpha_c(\gamma)$ (solid line). The figure predicts a decrease of the critical storage capacity as the degree of depression is increased. For $\gamma \rightarrow 0$ (non-depressed synapses), we recover the well know critical storage capacity for static synapses, that is $\alpha_c \approx 0.138$. For $\gamma \rightarrow \infty$ our mean-field theory predicts $\alpha_c \rightarrow 0$.

To check the validity of the mean field results, we also performed a series of numerical simulations for $T = 0$. We used the numerical protocol introduced in [13]. This consist of simulating a number N_r of retrieval experiments for each value of the loading parameter α. For a particular α, each retrieval experiment has been performed with a different set of random patterns and starting in a pattern (namely pattern 1) with its pixels disordered with probability 0.1. At each time step, we updated in parallel all neurons according to the rule $s_i = \Theta(h_i)\ \forall i$, where $\Theta(x)$ is the step function, and the corresponding fraction of synaptic resources x_i according to (3). After 200 iterations, we recorded the final state of the network and its overlap with pattern 1 for each particular retrieval experiment. Finally, we averaged those overlaps over the N_r retrieval experiments. For example, Fig. 2 shows the final averaged overlap $\langle m^1 \rangle$ as a function of α for the particular value of $\gamma = 1$ and four different network sizes. Following [13], we assume that retrieval of the original pattern is not effective when $\langle m^1 \rangle < 0.75$. This criterium gives the critical value of α for effective retrieval, that is $\alpha_c(\gamma)$, and for a particular network size N. To obtain the value $\alpha_c(\gamma)$ for $N \rightarrow \infty$, we fitted all $\alpha_c(\gamma)$ obtained for different network sizes to a straight line as a function of $1/N$ and extrapolated to $1/N = 0$ in Fig. 3. The resulting $N \rightarrow \infty$ estimates for different values of γ are plotted on Fig. 1. For $\gamma = 0$, we obtained $\alpha_c = 0.146 \pm 0.002$ which is in agreement with the estimate of α_c reported in [2] and [13].

In general, we observe that the value of $\alpha_c(\gamma)$ obtained in numerical simulations is slightly higher than the mean-field prediction. This discrepancy must be attributed to several factors: the small network sizes used in our numerical studies; possible non-linear finite size scaling effects at larger N; and the criterium we use to define $\alpha_c(\gamma)$ for the destabilization of memory pattern. In addition, our mean field approach uses the replica symmetric ansatz, which is well-known to underestimate the exact value of α_c for static synapses [2].

V. CONCLUSION

We conclude, that depressing synapses have a negative effect on the storage capacity of attractor neural networks. The effect is quite dramatic: the value of α_c is reduced by a factor 10 for $\gamma \approx 2$. Since γ is the ratio of recovery and depression time constants, this means that the stored patterns can only be retrieved when recovery is fast compared to depression. This is in sharp contrast with experimental findings where τ_{rec} is of the order of 1 sec and U^{-1} is of the order of miliseconds.

One must realize, that the classical definition of storage capacity as used in this paper considers the asymptotic stability of memory patterns as fixed points of the dynamics. We have shown previously [10] that attractor neural networks with dynamical synapses display in addition to a ferro-magnetic phase an oscillatory phase, where the network activity switches rapidly between stored memories. None of these memories are fixed points, but rather meta-stable states. This behaviour is also observed at high memory loading. Therefore, the disappearance of memories as fixed points does not imply the disappearance of memory from these networks. The storage of memories as metastable states will require a different definition of storage capacity. How such a new storage capacity depends on the dynamics of the synapses is an open question.

VI. ACKNOWLEDGEMENTS

This work has been supported by “Ministerio de Ciencia y Tecnología” and FEDER (“Ramón y Cajal” contract and project no BFM2001-2841) and the Dutch Foundation for Neural Networks (SNN).

REFERENCES

[1] J. Hopfield, Proceedings Nat. Acad. Sci. USA **79**, 2554 (1982).

[2] D. Amit, H. Gutfreund, and H. Sompolinsky, Ann. of Phys. **173**, 30 (1987).

[3] L. Abbott, J. Varela, K. Sen, and S. Nelson, Science **275**, 220 (1997).

[4] M. Tsodyks and H. Markram, Proc. Natl. Acad. Sci. USA **94**, 719 (1997).

[5] L. Dobrunz and C. Stevens, Neuron **18**, 995 (1997).

[6] T. Natschlger, W. Maass, and A. Zador, Network: Comput. Neural Syst. **12**, 75 (2001).

[7] J. S. Liaw and T. W. Berger, Hippocampus **6**, 591 (1996).

[8] L. Pantic, J. Torres, and H. Kappen, Neurocomputing **38-40**, 285 (2001).

[9] D. Bibitchkov, J. M. Herrmann, and T. Geisel, Network: Comput. Neural Syst. **13**, 115 (2002).

[10] L. Pantic, J. Torres, H. Kappen, and S. Gielen, Neural Computation, in press (2002).

[11] D. Amit, *Modeling brain function* (Cambridge University Press, Cambridge, 1989).

[12] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the theory of neural computation* (Addison-Wesley, Reedwood City, 1991).

[13] D. Volk, Inter. J. Modern Phys. **9**, 693 (1998).

[14] M. Giugliano, M. Bove, and M. Grattarola, IEEE Trans. Biomed. Eng. **47**, 611 (2000).

[15] A similar mechanism has been used to explain the regulation of insulin secretion in pancreatic β-cells [14].

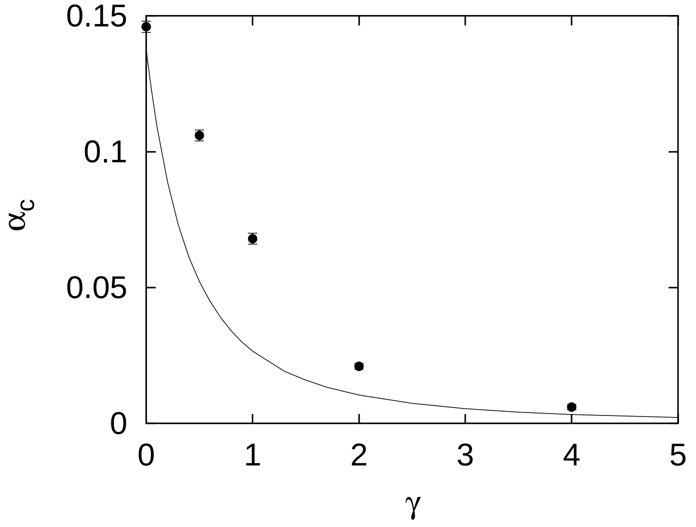

FIG. 1: Critical storage capacity α_c vs $\gamma = \tau_{rec} U$ in a neural network with depressing synapses: mean-field result (solid line), numerical data points corresponding to the finite-size scaling analysis of the Monte Carlo simulations (data points) (cf. Fig. 3).

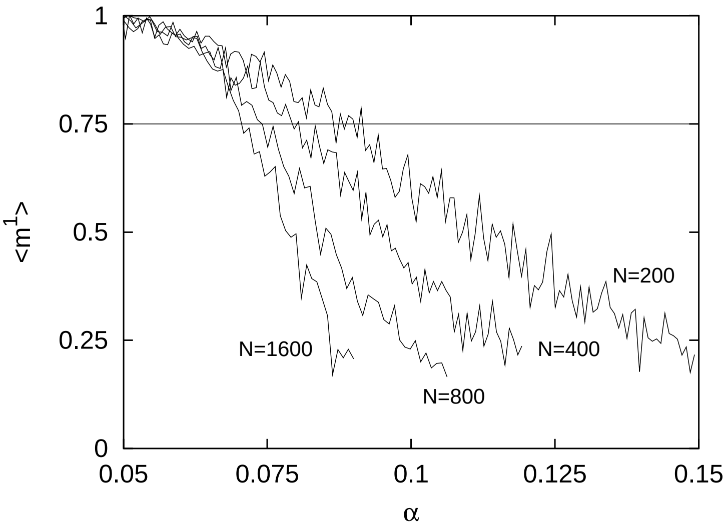

FIG. 2: Retrieval of a pattern in a neural network with depressing synapses at $T = 0$. The figure shows the average overlap of the final state of the network with the starting pattern as a function of the loading parameter α and for different network sizes. Here, we have considered for illustrative purposes the particular situation for $\gamma = 1$. The lines represent the average overlap obtained in $N_r = 150$ retrieval experiments for each value of α. The value $\langle m^1 \rangle = 0.75$, is used as a criterium to define α_c.

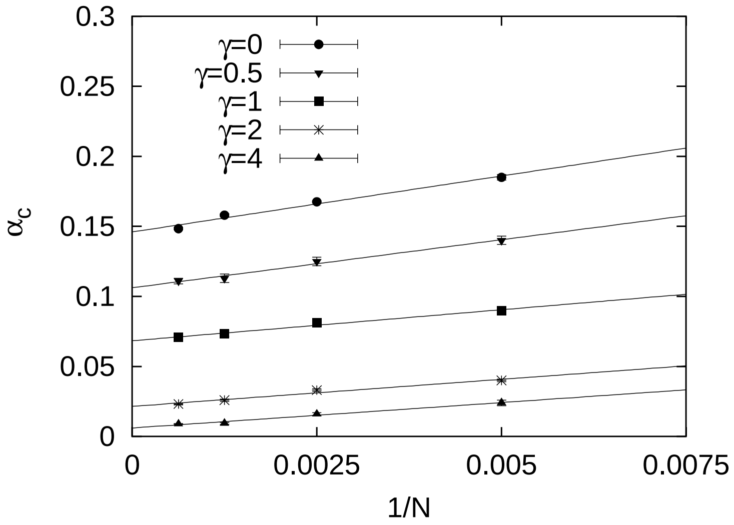

FIG. 3: Finite-size scaling of the Monte Carlo simulations of a neural network with depressing synapses for different values of the depressing factor $\gamma = \tau_{rec}U$. The data points for each value of γ are the critical storage capacity found for different network sizes ($N = 200,\ 400,\ 800$ and 1600). The solid lines corresponds to the linear fitting of the data points. Extrapolation of those lines to $N \to \infty$ gives the data-points represented in Fig. 1.