Neural networks with high-order connections

Jeferson J. Arenzon* and Rita M. C. de Almeida
Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, 91501-970 Porto Alegre, Rio Grande do Sul, Brazil
(Received 11 June 1993)

We present results for two different kinds of high-order connections between neurons acting as corrections to the Hopfield model. Equilibrium properties are analyzed using the replica mean-field theory and compared with numerical simulations. An optimal learning algorithm for fourth-order connections is given that improves the storage capacity without increasing the weight of the higher-order term. While the behavior of one of the models qualitatively resembles the original Hopfield one, the other presents a new and very rich behavior: depending on the strength of the fourth-order connections and the temperature, the system presents two distinct retrieval regions separated by a gap, as well as several phase transitions. Also, the spin-glass states seems to disappear above a certain value of the load parameter α, α_g.

PACS number(s): 87.10.+e, 75.10.Hk

I. INTRODUCTION

Synapses connecting more than two neurons have been introduced in an attempt to both improve the storage capacity of existing models [1–6] and to be a simulacrum of synapses existing in real brains (see [1] and references therein). Biologically, the idea of multisynapses has a strong motivation [1]: axon-axon-dendrite connections, for instance, are relatively common in real nervous systems and can be described as third-order synapses and even more intricate connections, involving more than two axons, may also exist in the brain. However, since second-order synapses are highly dominant, higher-order terms should be considered as corrections. As stressed in Ref. [1], this feature may play an essential role in the functioning of central nervous systems of superior vertebrate organisms. Moreover, when some pairwise connections are close enough they may interact somehow, and that can also be considered as high-order synapses (although in this case there are only interactions of even order).

Networks with N infinite range interacting Ising spins ($S_i = \pm 1$) associated with the state of the neurons (active or inactive) are considered to describe learning, storage, and retrieval of information. Possible configurations of the network are represented by N-dimensional vectors $\mathbf{S} = (S_1, \ldots, S_N)$ and the stored information (memories) is associated with P of these states, denoted by the vectors $\boldsymbol{\xi}^\mu$, $\mu = 1, \ldots, P$. The network load is measured by the parameter α, usually P/N, and the performance of a model for attractor networks can be measured by its storage capacity and its ability in recalling the stored patterns, in particular, the maximum allowed noise in an initial configuration and the time needed by the network to evolve and stabilize at, or near, one of the P memories.

Several works introduced multispin interactions by generalizing the Hopfield model [7] and Hebb learning rule by a monomial of degree $k > 2$ in the Ising spins [2,3]. These models have been investigated both analytically and through computer simulations. Alternatively, a recently introduced model [4] simultaneously considers several orders of interactions besides the second-order Hopfield term. In this paper we present a truncated version (hereafter called the truncated model) of that model and study the effect of the (weighted) first correction to the Hopfield term. We also investigate the effect of a Hopfield-like correction [hereafter called the generalized Hopfield (GH) model] and compare both prescriptions.

The paper is organized as follows: Sec. II defines the models and in Sec. III analytical and numerical results are presented. In Sec. IV we summarize and present our conclusions.

II. THE MODELS

A. The generalized Hopfield model

In order to compare performances, we contemplate a straightforward generalization of the Hopfield model (GH) by a polynomial of degree M that considers multispin interactions [2,3], namely

$$E = -\frac{N}{2} \sum_{\ell=2}^{M} \varepsilon_\ell \sum_\mu m_\mu^\ell \quad , \tag{1}$$

where M is an integer ($M > 2$), ε_ℓ are real constants, and the overlap m_μ between the state of the network $\mathbf{S}$ and the pattern $\boldsymbol{\xi}^\mu$ is given by

$$m_\mu = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu S_i \quad . \tag{2}$$

Here we have high-order terms as corrections (that do not need to be small) to the original second-order Hopfield

 ©1993 The American Physical Society

model. This system has a completely different static and dynamical behavior from the one that will be defined in Sec. II B [6]. Here the higher-order corrections do not qualitatively change the $T = 0$ behavior of the Hopfield network, although the (α, T) phase diagram presents some new features. The Hopfield network has its performance determined by the nature of the local field on a given neuron S_i. This local field has two competing components, a signal term that tends to align the spin S_i with a given pattern and a noise term that has a random orientation. A correction to the second-order Hopfield term may then act in two different ways: it enhances the signal term and/or it decreases the noise one. Also, the joint analysis of both terms yields an estimate of the critical capacity of the net: the lowest-order interaction ($\ell_{\min}$) in the energy function is the most relevant contribution for the cross talk noise from the high patterns implying that the maximum number of patterns that can be embedded is $O(N^{\ell_{\min}-1})$. Thus, the presence of the second-order term implies that the maximum number of patterns that can be stored is proportional to N [6,8].

The learning rule, for any $\ell \geq 2$, is

$$J^H_{i_1\cdots i_\ell}(P+1) = J^H_{i_1\cdots i_\ell}(P) + \frac{\varepsilon_\ell}{N^{\ell-1}}\xi^{P+1}_{i_1}\xi^{P+1}_{i_2}\cdots\xi^{P+1}_{i_\ell} \quad . \tag{3}$$

Here one has the full symmetry of indices and all connections are symmetric. Note that in a more general case, the weights could be considered different for each pattern, $\varepsilon_\ell = \varepsilon_\ell(\mu)$, generalizing the model studied by Viana [9].

This model presents an overall behavior qualitatively similar to the standard Hopfield model at $T = 0$. For $\alpha < \alpha_c$, the retrieval quality is good ($m \simeq 1$) and the size of the basins of attraction decreases with α. Also, if the initial state is out of the basin of attraction, the mean convergence time grows as the network increases, while inside the basins just one or two steps are enough [6,10]. Underlying this similarity is the fact that the loss of retrieval abilities due to an overload of the network originates in the same mechanism: the noise term overwhelms the signal one when $\alpha > \alpha_c$.

B. The truncated model

The complete energy function of a model previously proposed [4], with all orders of interactions (up to $2P$), for a network storing P patterns is

$$E = N\prod_{\mu=1}^{P}\left(1 - m_\mu^2\right) \quad . \tag{4}$$

This energy function is proportional to the product of the Hamming distances between the network state $\mathbf{S}$ and the patterns $\boldsymbol{\xi}^\mu$ and their inverses $-\boldsymbol{\xi}^\mu$. From Eq. (4) it is clear that $E(\mathbf{S}) \geq 0$ (if $\mathbf{S} = \boldsymbol{\xi}^\mu$, for any μ, the equality holds). It means that, no matter how large α is, the patterns are always global minima of E. A complete discussion of the phase space landscape in the $\alpha \to 0$ limit as well as simulation results for $\alpha \neq 0$ can be found in Refs. [4,5].

The multi-interaction nature of Eq. (4) becomes evident when it is displayed as

$$E = N\left[1 - \sum_{\mu_1} m^2_{\mu_1} + \sum_{\mu_1<\mu_2} m^2_{\mu_1}m^2_{\mu_2} + \cdots + (-1)^P \sum_{\mu_1<\cdots<\mu_P} m^2_{\mu_1}\cdots m^2_{\mu_P}\right] \quad . \tag{5}$$

Notice that, although the first nontrivial term is the Hopfield energy function, the higher-order terms are different from any previous model because they contain mixed memory terms. Also, since we have the Hopfield term along with higher-order ones, we expect that the number of patterns that can be stored is $O(N)$. We now define an energy function by neglecting the constant zeroth-order term and considering the next M terms ($M < P$). This energy function, after introducing weights and renormalizing it by a factor 1/2, reads

$$E = \frac{N}{2}\sum_{\ell=1}^{M}(-1)^\ell\varepsilon_{2\ell}\sum_{\mu_1<\cdots<\mu_\ell} m^2_{\mu_1}m^2_{\mu_2}\cdots m^2_{\mu_\ell} \quad . \tag{6}$$

Equation (6) defines a model that can be regarded as the Hopfield model ($\ell = 1$) plus correction terms. Different from the previous case, here one cannot have $\varepsilon_2 = 0$ because the remaining terms are mixtures and no pattern can be retrieved only with them. In what follows we consider only the first correction to the Hopfield term ($M = 2$). Equation (6) can then be rewritten as ($\varepsilon_{2\ell} = \delta_{1\ell} + \varepsilon\delta_{2\ell}$)

$$E = -\frac{1}{2}\sum_{i,j} J_{ij}S_iS_j + \frac{\varepsilon}{2}\sum_{i,j,k,l} J^T_{ijkl}S_iS_jS_kS_l \quad . \tag{7}$$

The learning rule for the second-order couplings J_{ij} is the Hebb prescription [11], Eq. (3). The fourth-order synapses J^T_{ijkl}, on the other hand, may be implemented through the following learning rule:

$$J^T_{ijkl} = \frac{1}{2N^3}\sum_{\substack{\mu,\nu\\ \mu\neq\nu}} \xi^\mu_i\xi^\mu_j\xi^\nu_k\xi^\nu_l \tag{8}$$

and when a new pattern is learned,

$$J^T_{ijkl}(P+1) = J^T_{ijkl}(P) + \frac{\varepsilon}{N^2}J_{ij}(P)\xi^{P+1}_k\xi^{P+1}_l \quad , \tag{9}$$

where $\boldsymbol{\xi}^{P+1}$ is the $(P+1)$th pattern to be taught to the net. Equation (9) may be regarded as the multisynapses

described in the Introduction, with the last term being the action of two axons upon a binary synapse. Notice that for $P = 1$ this model is equivalent to the standard Hopfield model since the fourth-order synapses do not exist. Note that while the pairwise connections are symmetric, i.e., $J_{ij} = J_{ji}$, the fourth-order connections do not present full symmetry under all possible indices interchange (only $i \leftrightarrow j$ and $k \leftrightarrow l$). These couplings can be symmetrized if they are rewritten as

$$J'_{ijkl} = \frac{1}{3}\left(J_{ijkl} + J_{ljki} + J_{kjil}\right) \tag{10}$$

and hence the energy (7) is a Lyapunov function for the dynamics

$$S_i(t+1) = \mathrm{sgn}\Bigg(\sum_j J_{ij} S_j(t) - \varepsilon \sum_{j,k,l} J'_{ijkl} S_j(t) S_k(t) S_l(t)\Bigg) \tag{11}$$

and the use of the statistical mechanics tools is allowed. One should also notice the importance of the self-couplings here. For instance, the couplings J_{iikl} and J_{ijkk} give rise to contributions to the energy of the same order as the couplings J_{ij}, which does not happen in the generalized Hopfield model. This can be easily seen if one rewrites the couplings as

$$J^T_{ijkl} = \frac{1}{2N} J_{ij} J_{kl} - \frac{1}{2} J^H_{ijkl} \quad . \tag{12}$$

Then, for the self-couplings mentioned above,

$$J^T_{iikl} = \frac{\alpha}{2N} J_{kl} - \frac{1}{2N^2} J_{kl} \simeq \frac{\alpha}{2N} J_{kl} \quad , \tag{13}$$

where the factor N^{-1} is compensated by a sum over sites (i) in (7). Notice that the contribution from J^H_{iikl} is negligible.

A previous numerical simulation [6] for $\varepsilon = 1$ shows a continuous transition from the retrieval phase to a nonretrieval one at $T = 0$. The basins of attraction seem to be large and α independent: m_0^c (the minimum initial overlap that allows retrieval in the thermodynamic limit) is ~ 0.1 for all $\alpha < \alpha_c$. The mean convergence time $\langle T \rangle$ (the average number of whole network updatings required to reach a stable state) increases with α and, for nonsmall values of α, does not depend on the initial overlap m_0. Long convergence times and large dispersions on its average values are often related to the irregularity of phase space around the memories [10] (existence of spurious states). However, this interpretation is only valid when the stored patterns are at, or very near, the bottom of the basins of attraction ($m \simeq 1$). Since the transition here is continuous, for large values of α this is no longer true: initial states with overlap m_0 with the chosen pattern lay on the surface of a hypersphere of radius proportional to $1 - m_0$ centered at the pattern. On average, the distance in the phase space from this surface to the bottom of the basin of attraction is the same as the distance from the memory to the energy minimum, independently of m_0. Consequently, the convergence time does not depend on m_0. For small values of α, on the other hand, the retrieval quality is good ($m \simeq 1$) and a small decrease in the convergence time is observed as m_0 increases [6].

III. MEAN-FIELD THEORY

A. The free energy and saddle-point equations

The mean-field analysis is performed by means of the standard techniques introduced by Amit, Gutfreund, and Sompolinsky [12]. In the GH model, the cross talk noise is governed by the second-order term while the fourth-order term contributes only to the signal term. On the other hand, in the truncated model there is a contribution from the higher-order terms (for instance, from the self-couplings mentioned in the preceding section) to the overall noise due to microscopically overlapping patterns. Up to the fourth order ($\varepsilon_{2\ell} = \delta_{1\ell} + \varepsilon\delta_{2\ell}$), the truncated energy function can be rewritten as

$$E = -\frac{N}{2}\sum_\mu m_\mu^2 - \frac{N\varepsilon}{4}\sum_\mu m_\mu^4 + \frac{N\varepsilon}{4}\left(\sum_\mu m_\mu^2\right)^2 \quad , \tag{14}$$

and the free energy per neuron can be obtained using the replica trick. After assuming that the replicas are symmetric and taking the limit of zero replicas we get

$$\begin{aligned} f_T = & -\frac{1}{2}(1-\varepsilon y)\sum_\mu m_\mu^2 - \frac{\varepsilon}{4}\sum_\mu m_\mu^4 - \frac{\varepsilon}{4}y^2 + \sum_\mu t_\mu m_\mu + \frac{\alpha}{2\beta}\ln[1-\beta(1-\varepsilon y)(1-q)] - \frac{1}{2}\frac{\alpha q(1-\varepsilon y)}{1-\beta(1-\varepsilon y)(1-q)} \\ & + \frac{1}{2}\alpha\beta r(1-q) - \frac{1}{\beta}\left\langle\!\!\left\langle \ln 2\cosh[\beta\left(\mathbf{t}\cdot\boldsymbol{\xi} + \sqrt{\alpha r}z\right)]\right\rangle\!\!\right\rangle \quad , \end{aligned} \tag{15}$$

where y is introduced to linearize the last term in Eq. (14) and the variables t_μ, q, and r are usually introduced to linearize the nonlinear terms. The symbol $\langle\!\langle \quad \rangle\!\rangle$ stands for two averages: over the finite number of patterns which may condense and over the Gaussian variable z, related to the infinite number of microscopically overlapping memories. The saddle-point equations are

$$\mathbf{m} = \langle\langle \boldsymbol{\xi} \tanh[\beta(\boldsymbol{\xi}\cdot\mathbf{t} + \sqrt{\alpha r} z)]\rangle\rangle \,, \tag{16}$$

$$t_\mu = (1-\varepsilon y) m_\mu + \varepsilon m_\mu^3, \tag{17}$$

$$q = \langle\langle \tanh^2[\beta\,(\mathbf{t}\cdot\boldsymbol{\xi} + \sqrt{\alpha r} z)]\rangle\rangle \,, \tag{18}$$

$$r = q\left[\frac{1-\varepsilon y}{1-\beta(1-\varepsilon y)(1-q)}\right]^2 \,, \tag{19}$$

$$y = \sum_\mu m_\mu^2 + \alpha \frac{1-\beta(1-\varepsilon y)(1-q)^2}{[1-\beta(1-\varepsilon y)(1-q)]^2} \quad . \tag{20}$$

Analogously, for the GH model the free energy is

$$f_H = -\frac{1}{2}\sum_{\ell=2}^{M}\sum_\mu \varepsilon_\ell m_\mu^\ell + \sum_\mu t_\mu m_\mu + \frac{\alpha}{2\beta}\ln[1-\beta(1-q)] - \frac{1}{2}\frac{\alpha q}{1-\beta(1-q)}$$
$$+ \frac{1}{2}\alpha\beta r(1-q) - \frac{1}{\beta}\,\langle\langle \ln 2\cosh[\beta\,(\mathbf{t}\cdot\boldsymbol{\xi} + \sqrt{\alpha r} z)]\,\rangle\rangle \quad , \quad \varepsilon_2 \neq 0 \quad , \tag{21}$$

and the equations to be solved are the same as in the original Hopfield model [12], except for t_μ. The equations for $\mathbf{m}$ and q are the same as Eqs.(16) and (18) and t_μ and r read (here we do not have y)

$$t_\mu = \frac{1}{2}\sum_{\ell=2}^{M}\varepsilon_\ell \ell m_\mu^{\ell-1} \,, \tag{22}$$

$$r = \frac{q}{[1-\beta(1-q)]^2} \quad , \quad \varepsilon_2 \neq 0 \,. \tag{23}$$

The sets of coupled nonlinear equations given by Eqs. (16)–(20) and (16), (18), (22), and (23) are numerically solved in Secs. III B and III C in the case where the network presents a macroscopic overlap m with one of the memories ($m_\mu = m\delta_{1\mu}$). Since we are mainly interested in the properties of the truncated model, results for the GH model will be given when they differ from the original Hopfield model or when comparing both models.

B. The $T = 0$ limit

The $T = 0$ ($\beta \to \infty$) limit of the saddle-point equations of the truncated model is

$$m = \mathrm{erf}\left(\frac{t}{\sqrt{2\alpha r}}\right) \,, \tag{24}$$

$$t = (1-\varepsilon y)m + \varepsilon m^3 \,, \tag{25}$$

$$r = \left[\frac{1-\varepsilon y}{1-C(1-\varepsilon y)}\right]^2 \,, \tag{26}$$

$$y = m^2 + \frac{\alpha r}{(1-\varepsilon y)^2} \,, \tag{27}$$

$$C = \sqrt{\frac{2}{\alpha\pi r}}\exp\left(-\frac{t^2}{2\alpha r}\right) \,, \tag{28}$$

where $q \to 1$ and $C \equiv \beta(1-q)$. These equations are numerically solved for several values of α and ε and the results are compared with the simulation whose details are presented later. We found essentially two different regimes, depending on the value of ε. For large ε (~ 0.5) the overlap m decreases with α, going monotonically from 1 down to zero at $\alpha_c^+(\varepsilon)$, signaling a second-order phase transition. As ε decreases (~ 0.36), the overlap presents a local minimum, before finally going to zero at $\alpha_c^+(\varepsilon)$. If $\varepsilon < \varepsilon_c \simeq 0.3587$, the minimum yields a gap separating two retrieval regions. These results are illustrated in Figs. 1 and 2 and summarized in the $T = 0$ phase diagram of Fig. 3. The critical values $\alpha_c^\pm(\varepsilon)$ are associated with second-order transitions at $T = 0$. The gap is delimited by the lines $\alpha_c^-(\varepsilon)$ and $\alpha_c'(\varepsilon)$, which meet at an endpoint nearby (0.3543,0.7784) (see Fig. 4): the left border is always given by $\alpha_c'(\varepsilon)$ while the right border is defined by $\alpha_c^-(\varepsilon)$ (second order) for $\varepsilon < 0.3543$ and by $\alpha_c'(\varepsilon)$ (first order) for $0.3543 < \varepsilon < 0.3587$. As ε approaches zero, the gap width between the first and second retrieval region $\Delta \equiv \alpha_c^- - \alpha_c'$ goes to infinity and for negative values of ε only the first-order transition associated with $\alpha_c'(\varepsilon)$ is present (see Fig. 3). Near ε_c the behavior of the model

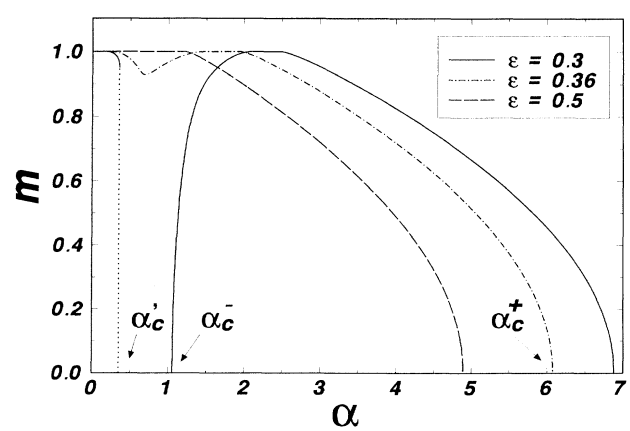

FIG. 1. Overlap versus α for the truncated model and $\varepsilon = 0.3, 0.36$, and 0.5. When $\varepsilon \to 0^+$, then $\alpha_c' \to 0.138$ and $\alpha_c^\pm \to \infty$, recovering the original Hopfield model.

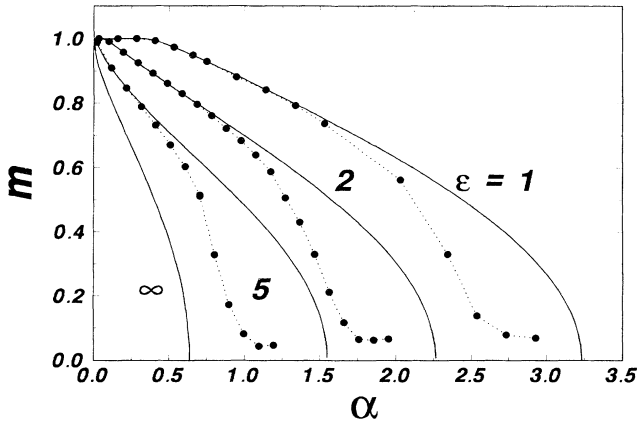

FIG. 2. Overlap versus α for the truncated model for $\varepsilon = 1, 2, 5$, and ∞. The retrieval quality decreases for increasing values of ε and as $\varepsilon \to \infty$, $\alpha_c \to 2/\pi$. The circles are the results for the numerical simulation using $N = 512$ (see text).

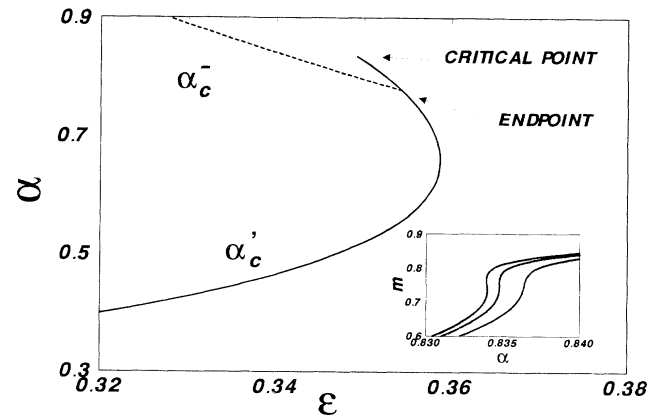

FIG. 4. Region near the point where the gap appears. The dashed line is second order while the solid one stands for first order. The line α_c' ends in a critical point at $\varepsilon \simeq 0.3492$ and both lines cross at an end point at $\varepsilon \simeq 0.3543$. Inset: overlap versus α near the critical point showing the second first-order transition for $\varepsilon = 0.3493, 0.3492$, and 0.3491 (from left to right).

is very complex near the gap and the effects of replica symmetry breaking (RSB) should be taken into account in order to decide what kind of critical points actually exists. Apparently (in the replica symmetry approximation) there is a critical point at $(\varepsilon, \alpha) \simeq (0.3492, 0.833)$, where the first-order line ceases to exist, and an end point nearby (0.3543,0.7784) where both lines (first and second order) cross (see Fig. 4). It must be emphasized that the retrieval solution is unstable in this limit ($T = 0$), as can be seen by the negative value of the entropy

$$S_0 \equiv \beta^2 \left.\frac{\partial f}{\partial \beta}\right|_{T=0} = -\frac{\alpha}{2}\ln[1 - C(1-\varepsilon y)] - \frac{\alpha}{2}\frac{C(1-\varepsilon y)}{1 - C(1-\varepsilon y)} \quad . \tag{29}$$

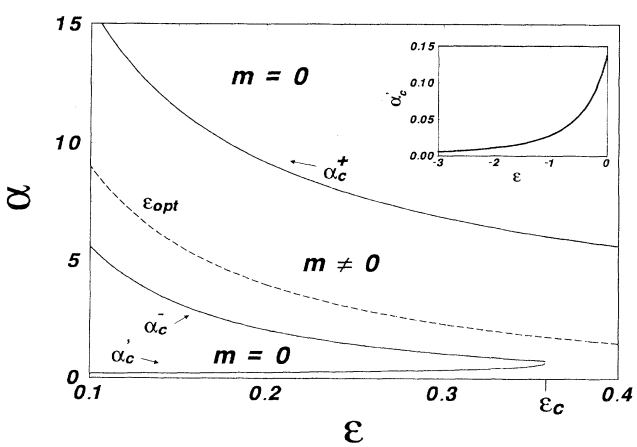

FIG. 3. The $T = 0$ phase diagram for the truncated model. For $\varepsilon > \varepsilon_c \simeq 0.3587$ the gap $\Delta = \alpha_c^- - \alpha_c'$ disappears while for $\varepsilon \to 0$, the gap $\Delta \to \infty$ as ε^{-1}. The lines $\alpha_c^{\pm}(\varepsilon)$ are second order while $\alpha_c'(\varepsilon)$ is first order. The dashed line ($\varepsilon_{\rm opt}$) is the value of $\alpha \neq 0$ where the peak ($m = 1$) occurs. Inset: the values of α at the first-order transition (α_c') for $\varepsilon < 0$.

The numerical values of the entropy are larger than in the Hopfield model, possibly indicating that the effects of RSB are stronger here. There is a range of ε, $0.3492 < \varepsilon < 0.3543$, where the system suffers up to four phase transitions (two first and two second order) since, for a fixed value of ε, as α is increased it crosses twice the line α_c'. The richness of this phase diagram deserves a separate study including the analysis of the stability of the solutions as well the effects of RSB, mainly in this low temperature region, but it is beyond the scope of this paper. The peak in the second retrieval region occurs because the noise in the Hopfield term of the local field in Eq. (11) is completely compensated by the noise generated by self-couplings in the fourth-order term when $\varepsilon = y^{-1}$ (see discussion below). For negative values this never happens since $1 + |\varepsilon| y$ is always positive (y is positive defined).

The points where m continuously approaches zero $[m \sim |\alpha - \alpha_c^{\pm}(\varepsilon)|^{1/2}$ as $\alpha \to \alpha_c^{\pm}(\varepsilon)]$, $\alpha_c^{\pm}(\varepsilon)$, are obtained by expanding Eqs. (24)–(28) for small m:

$$\alpha_c^{\pm}(\varepsilon) = \left(\frac{1}{\sqrt{\varepsilon}} \pm \sqrt{\frac{2}{\pi}}\right)^2 \quad . \tag{30}$$

For $\varepsilon > \varepsilon_c$ there is only one transition at $\alpha_c^+(\varepsilon)$, as can be seen in the $T = 0$ phase diagram (Fig. 3), and the critical value of α is a decreasing function of ε. Two different retrieval phases appear for $\varepsilon < \varepsilon_c$ and when the right border is second order, given by α_c^-, the width of the second retrieval region is

$$\Delta' \equiv \alpha_c^+(\varepsilon) - \alpha_c^-(\varepsilon) = 4\sqrt{\frac{2}{\varepsilon\pi}} \quad . \tag{31}$$

As $\varepsilon \to 0$, Δ goes to infinity as ε^{-1} and Δ' as $\varepsilon^{-1/2}$. Thus, when the truncated model recovers the Hopfield

model ($\varepsilon \to 0$), the location of the second retrieval region in the α axis goes to infinity and $\alpha_c' \to 0.138$.

When $\varepsilon \to \infty$, the solution for m can be obtained as a function of α and it reads

$$m = \mathrm{erf}\left(\sqrt{\frac{1}{2}\ln\frac{2}{\alpha\pi}}\right) \tag{32}$$

and

$$\alpha_c^+(\infty) = \frac{2}{\pi} \quad . \tag{33}$$

In this limit case, the system behaves suitably as an associative device ($m \simeq 1$) only if the number of embedded patterns is finite ($\alpha = 0$).

It is also possible to obtain the location of the maximum value of m in the second retrieval region: the values of α that allow $m = 1$ in (24)–(28) are $\alpha = 0$ and

$$\alpha = \frac{1-\varepsilon}{\varepsilon} \quad . \tag{34}$$

At the peak the noise from the high patterns (measured by r) goes to zero because $\varepsilon = y^{-1}$ (and since y is positive defined, this only happens for positive ε). Thus, the contribution from the second-order term does not contribute in the peak and one shall take into account the next (fourth order) term. This allows us to introduce an optimal learning rule by choosing the weight ε as the value that satisfies (34).

The above results can be qualitatively understood through a signal to noise analysis. The local field acting upon the ith neuron when the system is recalling the first pattern is

$$h_i \simeq (1-\alpha\varepsilon)\xi_i^1 + (1-\alpha\varepsilon-\varepsilon)\frac{1}{N}\sum_{\mu>1}\sum_{j\,(\neq i)}\xi_i^\mu\xi_j^\mu\xi_j^1 \quad , \tag{35}$$

where we considered the contributions from both two and four neuron couplings in Eq. (7). The increase in ε has then two effects: it acts both on the signal and noise terms. Depending on the value of ε, an increase in α may either suppress the signal or the noise term. When ε is small enough, the sum in Eq. (35) may overwhelm the signal term and we have a Hopfield-like mechanism of suppressing retrieval abilities. In this case the overall behavior is qualitatively similar to the second-order Hopfield model, as in the GH model. This also corresponds to the first retrieval region for $\varepsilon < \varepsilon_c$. On the other hand, when $\varepsilon > \varepsilon_c$, together with a decrease in the noise term, one can also observe a detectable decrease in the signal term with increasing α when in the second retrieval region. Now the mechanism of loss of retrieval ability is not due to an overwhelming noise term, but to the annihilation of the signal one: the energy barriers around the minima are decreased. A similar effect is observed in the complete model [4]. This explains the qualitatively diverse behavior of the network in this region. Also, the second (noise) term has zero mean and variance equal to $\sigma^2 = \alpha(1-\alpha\varepsilon-\varepsilon)^2$. An important result is that the variance is zero if $\alpha = 0$ or $\alpha = (1-\varepsilon)/\varepsilon$. These are the points where the retrieval is perfect ($m = 1$); higher-order noise terms, which we did not write, have dispersions at least $O(N^{-1})$ times smaller. Also, as for increasing values of ε the coefficient of the signal term $1-\alpha\varepsilon$ changes signal for decreasing values of α, explaining why the critical value of α decreases.

To verify these results we performed zero temperature simulations with network sizes up to 512 neurons. The steps are the following: one of the embedded memories is chosen as the initial state and a spin is (sequentially) flipped whenever this lowers the system energy. This procedure is repeated until a stable fix point is reached and the final overlap m with the chosen memory is measured. The averages were taken over five different sets of patterns and the number of runs in each set were 200 and the sizes were $N = 128, 256$, and 512 neurons. In Fig. 5 we can see the final overlap m versus α for $\varepsilon = 0.3$, clearly showing the existence of the gap. The results for $\varepsilon = 1$ can be found in Ref. [6]. As in other models, when compared with the mean-field calculations, the simulation yields some discrepancies (mainly near the transitions), which are in part due to the replica symmetry instability at $T = 0$ and supported by the negative entropy at zero temperature obtained with the replica symmetry ansatz. Also note that the remanent magnetization above α_c^+ for $\varepsilon = 0.3$, $m \sim 0.1$ is lower than the one found for the Hopfield and GH models ($m \sim 0.2$) and, as can be seen in Fig. 2, is a decreasing function of ε.

A final remark concerning the free energy of the truncated model at $T = 0$ is that the retrieval states are global minima for all values of α below α_c^+ in the truncated model if $\varepsilon > \varepsilon_c$. For $\varepsilon < \varepsilon_c$ there is a range of α in the first retrieval region where they are local minima (metastable): $\alpha_M < \alpha < \alpha_c'$ (in the second retrieval region they are always global minima). For $\varepsilon \to 0$, $\alpha_M \to 0.05$, as expected [12].

On the other hand, in the GH model whose equations at $T = 0$ are [the equations for m and C are the same as

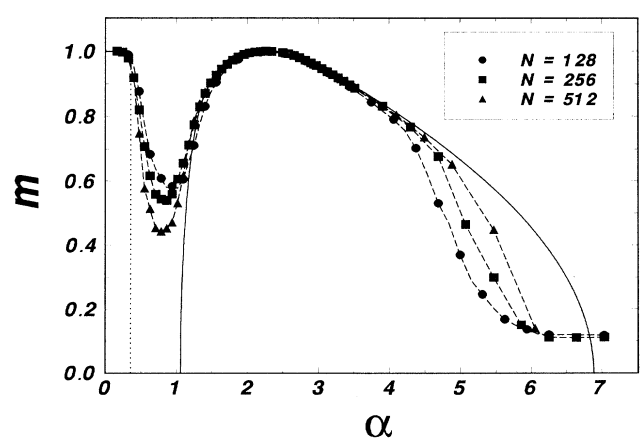

FIG. 5. Overlap versus α for the truncated model and $\varepsilon = 0.3$. The full curve is the $T = 0$ solution of Eqs. (24)–(28) while the points are obtained through numerical simulation (see text). Notice the perfect retrieval at $\alpha = 0$ and $\alpha = (1-\varepsilon)/\varepsilon \simeq 2.33$.

(24) and (28), respectively]

$$t = \frac{1}{2}\sum_{\ell=2}^{M} \varepsilon_\ell \ell m^{\ell-1} \ , \tag{36}$$

$$r = (1-C)^{-2} \ , \tag{37}$$

the storage capacity is a monotonically crescent function of both M and ε_ℓ. Figure 6 shows α_c versus ε in the case $\varepsilon_\ell = \delta_{2\ell} + \varepsilon\delta_{4\ell}$, that is, when only second- and fourth-order terms are considered. The line α_M where the retrieval states become global minima and the value of m at the criticality $m_c(\varepsilon)$ are also plotted. When $\varepsilon \to \infty$, the asymptotic value of $m_c(\varepsilon)$ goes to $m_c(\infty) = 0.918$ and the critical value of α grows as $\alpha_c \sim \varepsilon^2$, which can be understood as follows. The larger ε is, the more important the fourth-order term and the system capacity tends to be of order $O(N^3)$ (attained in the absence of the second-order term [2]), that is, the asymptotic behavior of the maximum number of storable patterns goes as $P_c \sim \varepsilon^2 N$. The value of α_M also goes as ε^2 as $\varepsilon \to \infty$. For negative values of ε there is a cutoff: below $\varepsilon^{\rm cut}$ there is no retrieval [t, given by Eq. (36), is null]. For instance, for the case $\varepsilon_\ell = \delta_{2\ell} + \varepsilon\delta_{k\ell}$, $\varepsilon^{\rm cut} = -2/k$. As $\varepsilon \to \varepsilon^{\rm cut}$ (from above), $\alpha_c \to 0$ and $m_c \to 1$. In the general case we have

$$\sum_{\ell=2}^{M} \varepsilon_\ell^{\rm cut} \ell = 0 \ , \tag{38}$$

defining a hyperplane in the space of the nonzero ε_ℓ's. There is also a cutoff in α_M below which the memories are never global minima of the free energy. For instance, for the same case $k = 4$, $\varepsilon_M^{\rm cut} = 2/\pi - 1 \simeq -0.363$.

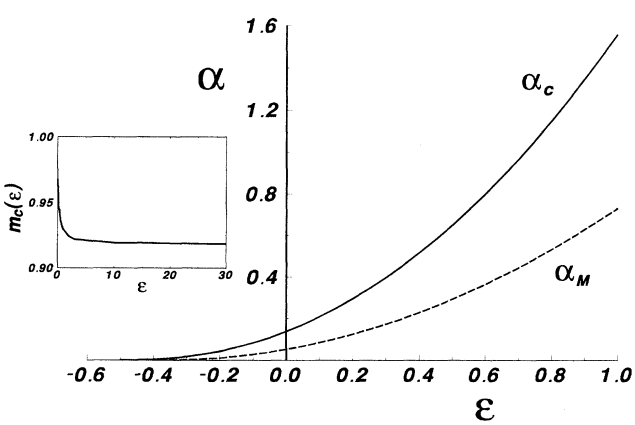

FIG. 6. Critical values of α versus ε at $T = 0$ for the GH model. The dashed line is the line below which the memories are global minima of the free energy α_M. The overlap at the criticality $m_c(\varepsilon)$ is shown in the inset and its asymptotic value $m_c(\infty)$ is 0.918. There is a cutoff for negative values of ε (in this particular case $\varepsilon^{\rm cut} = -0.5$), where $\alpha_c \to 0$ and $m_c \to 1$ as $\varepsilon \to \varepsilon^{\rm cut}$ from above. There is another cutoff below which the memories are never global minima of the free energy: $\varepsilon_M^{\rm cut} \simeq -0.363$.

C. The $T \neq 0$ case

The GH model presents some universal features that do not depend on the particular values of ε_ℓ and M. For instance, the line below which the SG states exist is

$$T_g = 1 + \sqrt{\alpha} \qquad \forall\, \varepsilon_\ell, M \quad , \tag{39}$$

and $\varepsilon_2 \neq 0$. The phase diagram for $\varepsilon_\ell = \delta_{2\ell} + \delta_{4\ell}$ is presented in Fig. 7 ($\varepsilon = 1$ since the qualitative features do not change with ε). There are three relevant lines: T_g, given by Eq. (39), signaling the appearance of SG (spin-glass) states; T_M, where the retrieval states ($m \neq 0$) first appear, and finally T_c, where the retrieval states become global minima of the free energy. At $\alpha = 0$, T_g and T_M do not meet, implying that there is a transition between the retrieval and paramagnetic phase also for small values of α. This phase diagram is similar to that for the Q-state Potts neural network model [14], due to a formal equivalence between multistate neurons and binary spins with diluted multispin interactions [15].

The SG phase is reentrant and hence the maximum possible value of α is not at $T = 0$ ($\alpha_c \simeq 1.556$) but at a nonzero value of T: $\alpha_c^{\rm max} \simeq 1.566$ for $T \simeq 0.126$. The degree of reentrance depends on ε: in the Hopfield limit ($\varepsilon_\ell = \delta_{2\ell}$) it is very small [13]. As a consequence, a small amount of noise improves the storage capacity of the system. This can also be observed in the behavior of m with T near the reentrant region: the overlap first increases before decreasing, indicating a small improvement with thermal noise. These effects may be an artifact of the replica symmetry (supposed to be stronger near the reentrant region): $\alpha_c^{\rm max}$ is believed to be a lower bound for the actual $T = 0$ critical capacity obtained when the replica symmetry is broken (the reentrant phase would then disappear). In other words, $\alpha_c^{\rm RSB} \geq \alpha_c^{\rm max}$, which is supported by numerical simulations in the case $\varepsilon = 1$ [6].

The phase diagram for the truncated model with the lines T_M and T_c for $\varepsilon = 1$ is shown in Fig. 8. In Fig. 9

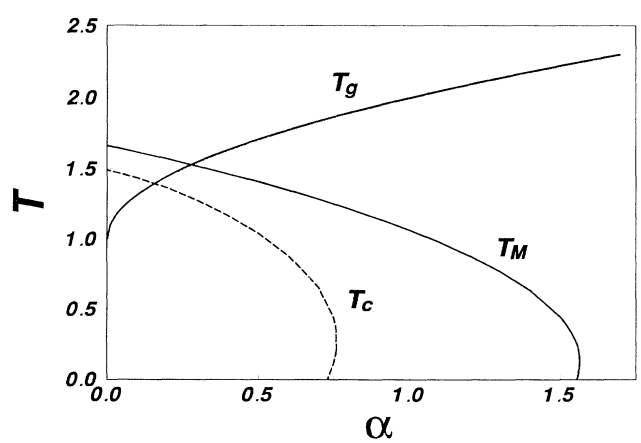

FIG. 7. Phase diagram for the $\varepsilon = 1$ GH model. Note the strong reentrant behavior for both T_c and T_M. The line T_g is the same for all values of ε and M. Notice that in this case there is a transition between the retrieval phase and the paramagnetic one for small α.

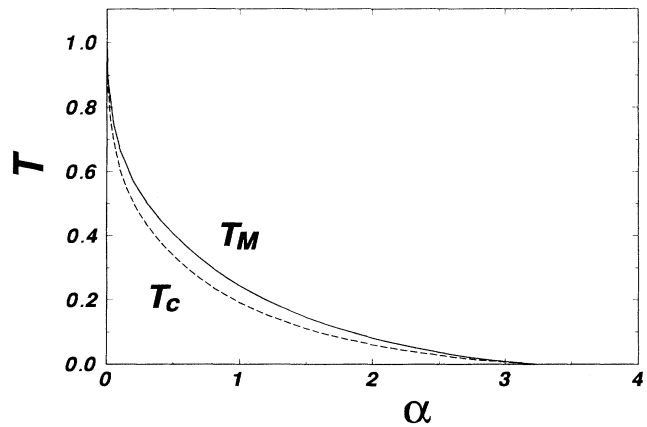

FIG. 8. Phase diagram for the truncated model with $\varepsilon = 1$. Below the line T_M we have $m \neq 0$ solutions (retrieval states) and these states become global minima of the free energy below T_c (dashed line). Notice that there is only one retrieval region and no reentrant phase.

the line T_M is shown for two values of ε: 0.36 and 0.3. For $\varepsilon > \varepsilon_c$, the retrieval states are always global minima of the free energy for low temperatures, although for $\varepsilon < \varepsilon_c$ they may become local minima in the first retrieval region. For all values of ε and $\alpha = 0$, $T_M = T_g = T_c = 1$. Nevertheless, for $\alpha \neq 0$, the qualitative features of the phase diagram depend on ε, different from the GH model. When $\varepsilon = 1$ (Fig. 8), T_M decreases monotonically and there is no reentrant phase. As ε decreases, the T_M line develops a minimum, implying that for some range of temperatures there are two retrieval regions (e.g., for $\varepsilon = 0.36$, $0.086 < T < 0.202$), as shown in Fig. 9. For even smaller ε, the minimum becomes a gap (Fig. 9) and the first retrieval region is reentrant. Also, the introduction of thermal noise in the system decreases the value of α at which m has a maximum in the second retrieval region. At $T = 0$ there are three phase transition lines in the ε-α plane: one first order α_c' and two second order $\alpha_c^{\pm}$ (Fig. 3). For $T \neq 0$, all three phase transition lines are first order.

The temperature at which the SG solution ($m = 0, q \neq 0$) continuously disappears ($q \to 0$) is

$$T_g(\alpha, \varepsilon) = \frac{1+\sqrt{\alpha}}{\sqrt{\alpha}} \left[(1-\alpha\varepsilon)(1+\sqrt{\alpha}) - 1\right] \quad . \tag{40}$$

These lines are shown in Fig. 10 for several values of ε. From the above equation we can see that $T_g(0, \varepsilon) = 1$ for all values of ε and for $\varepsilon = 0$ one recovers the Hopfield line $T_g(\alpha, 0) = 1 + \sqrt{\alpha}$. The line T_g, for $\varepsilon \neq 0$, goes to zero at (see inset of Fig. 10)

$$\alpha_g = \frac{1}{4}\left(1 - \sqrt{1 + \frac{4}{\varepsilon}}\right)^2 \quad . \tag{41}$$

The same happens in the pseudoinverse model [16], although in that case the transition is discontinuous and $\alpha_g \simeq 0.363$. Although there is a solution with $q = 0$ for the SG phase along the line $\alpha = \varepsilon^{-1}$, the critical line α_g, where the SG phase disappears, is such that $\alpha_g \leq \varepsilon^{-1}$ (see Fig. 10, inset).

We performed a numerical simulation to verify whether there is or is not a value of α above which the number of spurious states suffers a sudden decreasing. Different from the simulation presented in Sec. III, here the initial state is chosen at random and, after the system reached a stable state, one searches for the memory with the maximum overlap with that state. The idea is that if the system starts from random positions, the final state is either one of the embedded memories or some spurious state, if any. In order to quantify the results we define the following quantity:

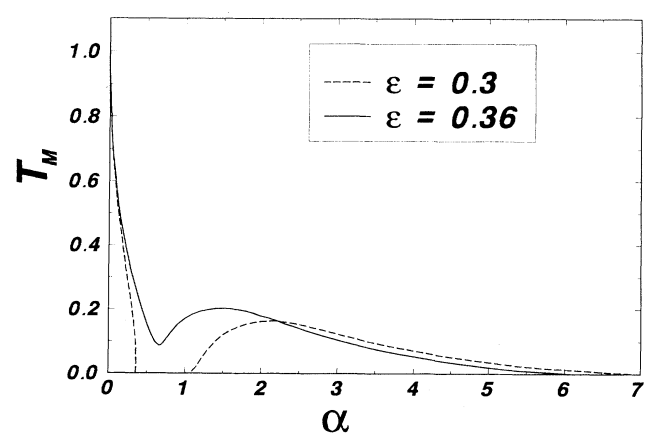

FIG. 9. The line T_M for the truncated model with $\varepsilon = 0.3$ and 0.36. In the latter case, it already shows the structure of two retrieval regions, although they are still connected. There is a range of temperature in which we have two retrieval regions: $0.086 < T < 0.202$. For $\varepsilon = 0.3$ both regions separate and the first retrieval region is reentrant.

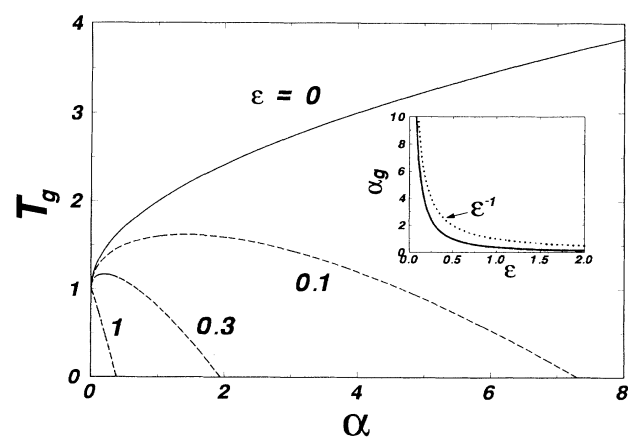

FIG. 10. Transition temperature for the SG solutions T_g for several values of ε in the truncated model. As $\varepsilon \to 0$, one recovers the Hopfield line $T_g = 1 + \sqrt{\alpha}$. In the inset we show the points α_g where $T_g = 0$, as well as the line ε^{-1} (see text).

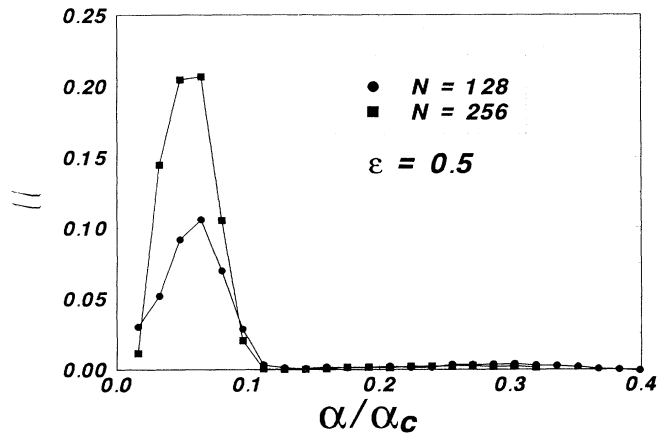

FIG. 11. Simulation results for $\mathcal{M} \equiv \langle\langle m_{m_0=1} - m_{ris} \rangle\rangle$ versus α/α_c for the truncated model and $\varepsilon = 0.5$. The maximum increases with the size N of the network. For $0.5 < \alpha < \alpha_c \simeq 4.893$, $\mathcal{M} = 0$ signals the low occupancy of the phase space by the spurious states.

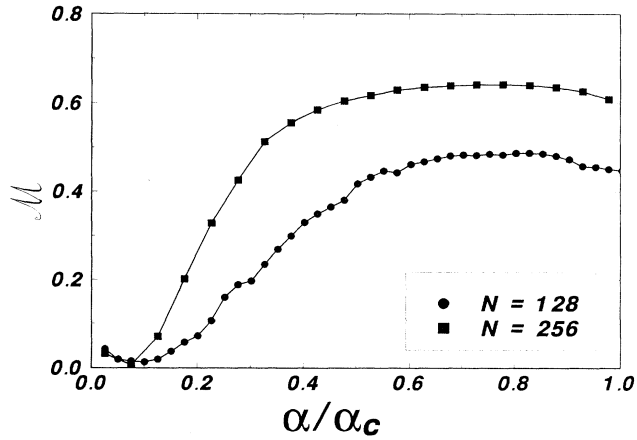

FIG. 13. Simulation results for $\mathcal{M} \equiv \langle\langle m_{m_0=1} - m_{ris} \rangle\rangle$ versus α/α_c for the GH model and $\varepsilon = 1$. As in the original Hopfield model, the phase space occupation by the spurious states grows with both α and N, although for small values of α it seems to have a minimum.

$$\mathcal{M} \equiv \langle\langle m_{m_0=1} - m_{ris} \rangle\rangle \quad , \tag{42}$$

where m_{ris} is the mean final overlap when the initial state is random and $m_{m_0=1}$ is the mean final overlap when the initial state is one of the memories. The symbol $\langle\langle \quad \rangle\rangle$ stands for average over several sets of patterns and initial states. Notice that this quantity is proportional to the fractional "occupation" of the phase space by the basins of spurious states. Also, if $\alpha > \alpha_c$, there is no memory retrieval and this quantity yields no relevant information. The simulation results for the truncated model with $\varepsilon = 0.5$ are presented in Fig. 11: $\mathcal{M} = 0$ for $\alpha = 0$ and for α roughly above 0.5. This result supports the analytical prediction of the disappearance of the SG states above a given value of α. However, the agreement is only qualitative since the predicted value of α_g/α_c for $\varepsilon = 0.5$ is 0.2. We expect this difference to be an artifact of replica symmetry instability at low temperatures. For the sake of comparison, $\mathcal{M}$ is also shown for the Hopfield and GH models, Figs. 12 and 13, respectively. In both cases, below α_c, $\mathcal{M}$ seems to increase both with α and N. For the Hopfield model, Fig. 12, the $\mathcal{M}$ dependence on α is linear, the exponent depending on N. For $\varepsilon = 1$, $\mathcal{M}$ attains a minimum for small α and grows for increasing values of both α and N. Thus, except for the truncated model, the spurious states dominate the phase space landscape either by its increasing number or by increasing basins of attraction.

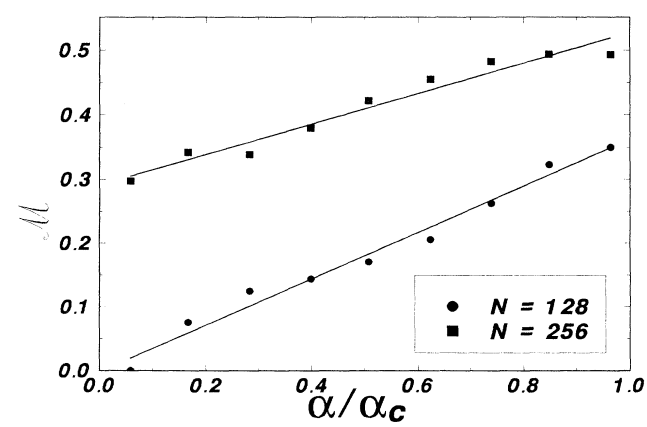

FIG. 12. Simulation results for $\mathcal{M} \equiv \langle\langle m_{m_0=1} - m_{ris} \rangle\rangle$ versus α/α_c for the Hopfield model. Notice the linear behavior below α_c.

IV. CONCLUSIONS

We compared the effect of two different fourth-order corrections to the standard Hopfield model by considering two learning rules and investigated their behavior calculating the retrieval capabilities with and without thermal noise. The phase diagrams for both models were presented, and the strikingly different behaviors presented by them come from the nature of the fourth-order connections, which may or may not present mixed memory terms.

The original nontruncated model [4,5] showed an improved performance due to a strong reduction of spurious states, with the consequent enhancing of the load capacity. The limit of the storage capacity of the network originates in the lowering of the energy barriers between memories and not in the dislocation of the energy minima from the patterns; the retrieval is then always perfect at $T = 0$. However, the order of the couplings (and their number) increases with P such that the ratio of information per synapse decreases.

Here we introduced a truncated model that shows a very rich behavior, summarized by the many phase diagrams presented in preceding sections. Besides several phase transitions, either first or second order, the sys-

tem may present a gap separating two distinct retrieval phases. The first one recovers the behavior of the standard Hopfield model in the $\varepsilon \to 0$ limit. On the other hand, in the second retrieval region an increase in the load of the network acts lowering the signal term. Thus, the limit in the capacity in this region is then due to the effect of lowering the energy barriers, different from the Hopfield model. Possibly the SG phase disappearing for $\alpha > \alpha_g$ is another consequence of suppressing the noise: there are no detectable spurious states, which is supported by numerical simulations. When the gap is not present, there is only one retrieval region that is more Hopfield-like for small α but presents a $T = 0$ second-order transition when losing its retrieval abilities. The second region, when it exists, presents a maximum in the curve m versus α that allows the definition of an optimal learning rule for which the system works in the minimum noise region [see Eq. (35)]: $\varepsilon_{\rm opt} = (1+\alpha)^{-1}$. The retrieval quality is maximum and the capacity of the network is greatly enhanced [although it is still $O(N)$ to guarantee that the signal term is not zero]. From the biological point of view, the synapses of order higher than two are corrections to the second-order ones, which is indeed the case here, since the role played for the fourth-order corrections depends on the value of ε and the network presents larger load capacities with decreasing ε. These results are valid for $T = 0$, and the optimal behavior is expected to hold at low temperatures. It could be interesting to check what happens for higher values of T. Also other quantities, for instance, the size of the basins of attraction and/or the convergence time, still deserve further investigation.

The effect of the sixth-order term in the expansion Eq. (6) may also be considered. However, we do not expect new effects to appear, except for an increase in the storage capacity (or maybe more than one gap), and there may exist at least one pair $(\varepsilon_4^{\rm opt}, \varepsilon_6^{\rm opt})$ that would improve the capacity of the model. Furthermore, the situation when only high-order connections are diluted is interesting from the biological point of view, but the dynamics is difficult to treat mathematically [3,17]. A version of the model in which dilution is present in both terms is being presently studied [18].

The stability of the solutions together with RSB effects should be studied, mainly in the $T = 0$ limit, which surprisingly showed a very rich phase diagram. An interesting problem is to investigate whether the critical point and the end point merge or not in a tricritical one when the replica symmetry is broken and then to obtain the critical exponents near those points.

At last, we should point out that these results still have an unclear biological relevancy, although they are interesting *per se*. Maybe other fields that use spin Hamiltonians may be benefited by using this model. As an example, one can use it as a fitness function in theoretical population genetics [19], where the gap might stand for some constraints that cannot be satisfied by any species in that environment or for some forbidden genetical traits.

ACKNOWLEDGMENTS

We acknowledge useful discussions with A.T. Bernardes, J.R. Iglesias, N. Lemke, J.C.M. Mombach, P.M.C. de Oliveira, T.J.P. Penna, and F.A. Tamarit. Also, J.J.A. thanks M.C. Barbosa for fruitful discussions on critical points. This work was partially supported by the Brazilian agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico, Financiadora de Estudos e Projetos, and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul.

* Electronic address: Arenzon@if1.ufrgs.br

[1] P. Peretto and J. J. Niez, Biol. Cybern. **54**, 53 (1986).
[2] E. Gardner, J. Phys. A **20**, 3453 (1987); L. F. Abbott and Y. Arian, Phys. Rev. **A36**, 5091 (1987); D. Horn and M. Usher, J. Phys. (Paris) **49**, 389 (1988).
[3] I. Kanter, Phys. Rev. A **38**, 5972 (1988); F. A. Tamarit, D. A. Stariolo, and E. M. F. Curado, Phys. Rev. A **43**, 7083 (1991); L. Wang and J. Ross, Phys. Rev. A **44**, R2259 (1991).
[4] R. M. C. de Almeida and J. R. Iglesias, Phys. Lett. A **146**, 239 (1990); J. J. Arenzon, R. M. C. de Almeida, and J. R. Iglesias, J. Stat. Phys. **69**, 385 (1992).
[5] J. J. Arenzon, R. M. C. de Almeida, J. R. Iglesias, T. J. P. Penna, and P. M. C. de Oliveira, J. Phys. (Paris) I **1**, 55 (1992); T. J. P. Penna, P. M. C. de Oliveira, J. J. Arenzon, R. M. C. de Almeida, and J. R. Iglesias, Int. J. Mod. Phys. C **2**, 711 (1991).
[6] J. J. Arenzon, R. M. C. de Almeida, J. R. Iglesias, T. J. P. Penna, and P. M. C. de Oliveira, Physica A **197**, 1 (1993).
[7] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[8] V. Deshpande and C. Dasgupta, J. Stat. Phys. **64**, 755 (1991).
[9] L. Viana, J. Phys. (Paris) **49**, 167 (1988).
[10] G. Kohring, J. Phys. A **23**, 2237 (1990).
[11] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, MA, 1991).
[12] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).
[13] J. P. Naef and A. Canning, J. Phys. (Paris) I **2**, 247 (1992).
[14] D. Bollé, P. Dupont, and J. Huyghebaert, Phys. Rev. A **45**, 4194 (1992).
[15] I. Kanter, J. Phys. **C20**, L257 (1987); Phys. Rev. A **37**, 2739 (1988).
[16] I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).
[17] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[18] N. Lemke, J. J. Arenzon, and F. A. Tamarit (unpublished).
[19] S. Kaufmann and S. Levin, J. Theor. Biol. **128**, 11 (1987); B. Derrida and L. Pelliti, Bull. Math. Biol. **53**, 355 (1991); H. Flyvbjerg and B. Lautrup, Phys. Rev. A **46**, 6714 (1992).