# Categorization in the symmetrically dilute Hopfield network

P. R. Krebs
*Instituto de Física e Matemática, Universidade Federal de Pelotas, Caixa Postal 354, 96010-900 Pelotas, RS, Brazil*

W. K. Theumann
*Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051, 91501-970 Porto Alegre, RS, Brazil*
(Received 18 March 1999)

A symmetrically dilute Hopfield model with a Hebbian learning rule is used to study the effects of gradual dilution and of synaptic noise on the categorization ability of an attractor neural network with hierarchically correlated patterns in a two-level structure of ancestors and descendants. Categorization consists in recognizing the ancestors when the network has been trained exclusively with the descendants. We consider a macroscopic number of ancestors, each with a finite number of descendants, and take into account the stochastic noise produced by the former in an equilibrium study of the network, by means of replica-symmetric mean-field theory. Phase diagrams are obtained that exhibit a categorization, a spin-glass, and a paramagnetic phase, as well as the dependence of the order parameters on the relevant quantities. The de Almeida–Thouless lines that limit the validity of the replica-symmetric results are also obtained. It is shown that gradual dilution increases considerably the region where a stable categorization phase may be found. [S1063-651X(99)00310-4]

PACS number(s): 87.10.+e, 64.60.Cn

## I. INTRODUCTION

There has been much interest in understanding the properties and predicting the behavior of large attractor neural networks. Of primary concern are the storage capacity, the quality of the retrieval overlaps, and the ability to retrieve a set of learned patterns when a network starts to evolve from an arbitrary initial state [1,2]. A further, relevant, issue is the categorization (or generalization) ability of a network [3,4]. This is the property of recognizing patterns in a high level of a hierarchical structure when a network is only exposed to patterns in a lower level during the training stage.

The presence of an exponentially large number of unwanted spin-glass-like states may limit severely the performance of a network. Indeed, except for a low storage ratio, a network is very likely to be trapped in these states, preventing the occurrence of finite overlaps with the patterns of interest. To overcome this problem, a dilute Hopfield model with low *symmetric* connectivity has been considered some time ago to study the retrieval problem [5,6]. This problem consists in the search for retrieval states with zero overlap except with one pattern. The storage capacity of the network in the extremely dilute limit was found to be considerably enhanced, when compared either with that of the standard symmetric model of full connectivity [1,2] or with the storage capacity of the extremely dilute asymmetrical model [7], particularly at finite temperature $T$, which is the rounding-off parameter in the neuron response function. The effect of the gradual dilution on the phase diagram of the random symmetrically dilute network is to reduce the stability of the spin-glass states below the critical storage capacity $\alpha_c$, enhancing thereby the retrieval states. In the limit of vanishingly small connectivity, stable spin-glass states are excluded up to a ratio $\alpha_c=1$, for $T\leqslant 1$, according to exact results on the formally equivalent Sherrington-Kirkpatrick (SK) spin-glass model [8]. It has also been pointed out in Ref. [6] that the retrieval performance of the network increases with a decrease of the connectivity in states within the memory phase, although the performance is impaired by the dilution on the phase boundary.

The categorization problem has been studied extensively in networks of various architectures, with either binary or multistate units, and different learning rules [4,9–21]. The categorization problem has, apparently, not been studied in the symmetrically dilute Hopfield model. This is an interesting model with partial connectivity between neurons, that has an energy function and, in contrast to the extremely dilute asymmetric network, has a nontrivial dynamics and a more complex behavior.

A prototype categorization problem with a set of hierarchically correlated patterns in two levels is the recognition of concepts, or ancestors, from the extraction of common features among the descendants (or examples of the concepts) presented to the network in the training stage. These features may be characterized by symmetric overlaps between the state of the network and the training patterns, which can be constructed for any network, whether the training rule is symmetric or not. Symmetric overlaps represent, usually, unwanted spurious mixture states for the retrieval problem which are only destabilized at low $T$ for uncorrelated patterns [22], but they have a crucial role in the categorization problem [4]. Indeed, the correlation parameter that characterizes the hierarchical structure of patterns stabilizes the symmetric mixture states up to relatively high $T$. This leads to a large categorization phase in the phase diagram for $\alpha$ vs $T$ [9], where single concepts may be recognized with small error when an appropriate number of examples is presented to the network. However, as in the retrieval problem, the categorization phase for the fully connected network has to compete with a spin-glass phase in the ordered region of the phase diagram.

The purpose of this paper is to study the effects of a gradual dilution of the synaptic connections on the categorization ability in a symmetrically dilute Hopfield neural network model with binary units and patterns in a two-level

 

hierarchy. We are particularly interested in the case of low or vanishing connectivity and our aim is to investigate to what extent the influence of spin-glass states can be reduced, leading to an improvement of the categorization ability of the network. It is also interesting to investigate the dependence of the categorization performance of the network on the temperature $T$. Indeed, in an earlier work on the fully connected Hopfield network we found that a small-to-moderate $T$ may be useful to reduce the categorization error in the case of a finite number of concepts [9].

In the present work we do not consider the dynamics but, instead, we study the equilibrium statistical mechanics of the symmetrically dilute Hopfield model. The outline of the paper is the following. In Sec. II we introduce the model and the relevant order parameter for the problem that gives the categorization error of the network. The free-energy density and other order parameters that are built into it are obtained in Sec. III, in a replica-symmetric mean-field theory. The limit of validity of that theory is also specified there. The results are presented and discussed in Sec. IV, and we end with a summary and conclusions in Sec. V.

## II. THE MODEL

We consider a random dilute Hopfield model of a neural network with $N$ binary units $S_i=\pm 1$; $i=1,\ldots,N$, described by the Hamiltonian

$$H=-\frac{1}{2}\sum_{i,j} J_{ij}^{d} S_i S_j , \tag{1}$$

where the sum is over all $i$ and $j$, and the dilute *symmetric* synaptic connections, $J_{ij}^{d}=J_{ji}^{d}$, are specified by the appropriate learning rule that involves the probability distribution of the random dilution. Before specifying the rule, we note that $\Sigma_j J_{ij}^{d} S_j$ is the local field on unit $i$ due to the activity of the other units. An increase in the local field produces, in general, an alignment of the component of the state of the network with an example, improving the retrieval performance.

The learning rule consists, for our purpose, in presenting to the network a finite set of $s$ examples $\{\xi_i^{\mu\nu}\}$, $\nu=1,\ldots,s$, of each of a macroscopic number of concepts, $p=\alpha c N$, with finite $\alpha=O(1)$, within the set $\{\xi_i^{\mu}\}$, $\mu=1,\ldots,p$, according to the generalized Hebb rule,

$$J_{ij}^{d}=\frac{c_{ij}}{cN}\sum_{\mu=1}^{p}\sum_{\nu=1}^{s}\xi_i^{\mu\nu}\xi_j^{\mu\nu} \tag{2}$$

in which $c_{ij}=c_{ji}$ is 1 with probability $c$ and 0 with probability $1-c$, where $c$ is the connectivity of the network, while $c_{ii}=0$. Thus, the synapses which are built exclusively from examples are cut symmetrically at random so that on the average each unit remains connected to $cN$ other units, and $\alpha$ is the ratio of concepts to be recognized. Following Sompolinsky [23], we restrict ourselves in this work to a *dense* network in which $c$ is of $O(1)$ when $N\to\infty$, meaning that each unit remains connected to $O(N)$ other units. It should be noted, however, that $c$ may become arbitrarily small after the thermodynamic limit. We come back to this point in the next section. In contrast, in the case of sparse networks, $c=O(1/N)$ and each unit is just connected to a finite number of other units. When $c=1$ we have the generalized Hebbian rule for the standard categorization problem in a fully connected network [4,9], while the limit $c\to 0$ corresponds to the extreme low-connectivity network.

The components of the concepts, which are assumed to be binary patterns, $\xi_i^{\mu}=\pm 1$, are taken to be statistically independent and equally distributed unbiased random variables. Each concept generates a finite set of examples $\{\xi_i^{\mu\nu}=\pm 1\}$ which are assumed to be statistically independent and equally distributed random variables chosen according to the probability distribution

$$P(\xi_i^{\mu\nu})=\frac{1}{2}(1+b\xi_i^{\mu})\delta(\xi_i^{\mu\nu}-1)+\frac{1}{2}(1-b\xi_i^{\mu})\delta(\xi_i^{\mu\nu}+1), \tag{3}$$

with the Kronecker $\delta$ and $0\leqslant b\leqslant 1$. We see that $b\xi^{\mu}$ is the bias that an example of the concept $\xi^{\mu}$ may be $+1$ and this will depend on the value taken by the concept. Thus, Eq. (3) implies a correlation $\langle \xi_i^{\lambda}\xi_j^{\mu\nu}\rangle=b\,\delta_{i,j}\delta_{\lambda,\mu}$ between a given concept and its examples and a correlation $\langle \xi_i^{\lambda\rho}\xi_j^{\mu\nu}\rangle=b^2\delta_{i,j}\delta_{\lambda,\mu}$, for $\rho\neq\nu$, among different examples of the same concept. Here, the brackets $\langle\ \rangle$ denote configurational averages over the examples and over the concepts, in this order.

In resemblance with Sompolinsky's work for a symmetrically dilute random network [2,23], the synaptic connections may be written as $J_{ij}^{d}=J_{ij}+\delta J_{ij}$, in which

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\sum_{\nu=1}^{s}\xi_i^{\mu\nu}\xi_j^{\mu\nu} \tag{4}$$

are the interactions of the fully connected network trained with examples [3,4] and the dilution term $\delta J_{ij}=(1-c_{ij}/c)J_{ij}$ may be interpreted, in the $N\to\infty$ limit, as a synaptic Gaussian noise effectively independent of the training examples, of mean configurational average $\langle \delta J_{ij}\rangle_c=0$ and variance $\langle(\delta J_{ij})^2\rangle_c=\Delta^2/N$. Here, the brackets $\langle\ \rangle_c$ denote configurational averages over the patterns and the $c_{ij}$, in which $\Delta^2\equiv a^2(1-c)/c$ with $a^2=\alpha c s[1+(s-1)b^4]$. Indeed, $\langle \xi_i^{\lambda\rho}\delta J_{ij}\rangle_c=0$ because the average of $1-(c_{ij}/c)$ is zero and $\langle(\xi_i^{\lambda\rho})^2\xi_j^{\lambda\rho}\rangle=0$ for *unbiased* examples distributed according to Eq. (3). It should be noted that the Gaussian synaptic noise is of a different nature than the synaptic noise of temperature $T$. Thus, with the separation of the synaptic connections into two parts, Eq. (1) becomes a sum of a Hopfield Hamiltonian for a fully connected neural network model trained with examples according to a generalized Hebbian learning rule [4] and a spin-glass Hamiltonian with random Gaussian interactions [8] the width of which depends now on the connectivity of the dilute network.

It is known that the categorization ability in a fully connected Hopfield model is enhanced either by an increase of the number of examples $s$ presented to the network in the training stage or by a larger correlation parameter $b$. This can be understood from the increase in the local field at a unit due to a larger width $a/\sqrt{N}$ of the random connection $J_{ij}$. It can be seen that a decrease in the connectivity $c$, in the case of a symmetrically dilute network, should similarly enhance the categorization ability. This will be demonstrated

by the results presented in Sec. III.

The order parameter that describes the recognition of a concept $\xi^\mu$ in a noiseless network, with $T=0$, as the main characteristic of the categorization problem, is the overlap

$$m^\mu=\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu S_i, \tag{5}$$

for $\mu=1,\ldots,p$, between the state $\{S_i\}$ of the network and the concept $\xi^\mu$. The state may be near a true minimum or a metastable state of the Hamiltonian $H$. We consider the half space in which $0\leqslant m^\mu\leqslant 1$ and the value of $m^\mu$ is a measure of the success in recognizing a concept. Due to the sum over sites, this overlap may be seen as the configurational average over the probability distribution of that concept. As a sum of a large number (in the limit $N\to\infty$) of random variables, it should not depend on a particular realization of these variables and becomes a self-averaging quantity.

It should be stressed that the recognition of concepts, i.e., the existence of finite overlaps with the states of the network, must emerge as a spontaneous feature of the network trained with examples. The network is neither exposed to the concepts in the training stage nor to concept dependent external fields. A measure of the failure in recognizing a concept is given by the categorization error, defined as the Hamming distance

$$\epsilon^\mu=\frac{1}{2}(1-m^\mu) \tag{6}$$

between the state $\{S_i\}$ and the concept $\xi^\mu$, where $\mu=1,\ldots,p$. Our aim is to find out the dependence of the categorization error on the number of examples $s$, the so-called categorization curves, and how these depend on the connectivity $c$ and on the ratio $\alpha$ of recognized concepts. Since all concepts are equivalent, we concentrate on one of them, say $\mu=1$.

## III. MEAN-FIELD THEORY

We consider now the mean-field theory for finite $\alpha$. For the purpose of practical calculations of the properties of the network, it is convenient to introduce a field- ($h^\mu$) dependent term for each concept into the Hamiltonian, adding altogether $\Sigma_{i\mu}h^\mu\xi_i^\mu S_i$, and taking the fields to be zero at the end. The averaged free-energy density, per connected site, is given by

$$f=-\lim_{N\to\infty}\frac{1}{cN\beta}\langle\ln Z\rangle_c, \tag{7}$$

where $Z=\mathrm{Tr}\exp(-\beta H)$ is the partition function of the model, in which $H$ is the sum of a generalized Hopfield Hamiltonian and a spin-glass Hamiltonian with random Gaussian interaction $\delta J_{ij}$, as discussed in the preceding section, while $\beta=T^{-1}$. Here, and in the following, we first take the thermodynamic limit $N\to\infty$, keeping the connectivity $c$ fixed in accordance with the model introduced in the preceding section, and thereafter we vary $c$, which may eventually become vanishingly small. The overlap with a given concept is then obtained as

$$m^\mu=df/dh^\mu|_{h^\mu=0}, \tag{8}$$

where at the end the field is set to zero. In the case of a noisy network, with temperature $T$, the components of the state in Eq. (5) are to be replaced by their thermal averages $\langle S_i\rangle_T$ with $Z$.

Proceeding in a now standard way, for a macroscopic number of concepts $p$ and hence, for nonzero $\alpha=p/cN$, we make use of the replica method to write for a specific concept, say $\mu=1$,

$$\langle\ln Z\rangle_c=\lim_{n\to 0}\frac{\langle Z^n\rangle_c-1}{n}, \tag{9}$$

in which $\langle Z^n\rangle_c$ is the configurational average of the replicated partition function. Combining well-known procedures to deal with the generalized Hopfield Hamiltonian, on one hand, and with the explicit spin-glass part on the other [8,24], we find

$$\langle Z^n\rangle_c=e^{-\beta npS/2}e^{nN\beta^2\Delta^2/4}\int\prod_{(\nu,\rho)}dm_\rho^{1\nu}\int\prod_{(\rho\sigma)}dq_{\rho\sigma}\,dr_{\rho\sigma}\times\exp(-N\beta f), \tag{10}$$

where

$$f=\frac{1}{2}\sum_{\nu,\rho}(m_\rho^{1\nu})^2+\frac{\alpha c\beta}{2}\sum_{(\rho,\sigma)}q_{\rho\sigma}r_{\rho\sigma}+\frac{\beta\Delta^2}{4}\sum_{(\rho,\sigma)}(q_{\rho\sigma})^2-\frac{\alpha c}{\beta}\ln G(q_{\rho\sigma})-\frac{1}{\beta}\langle\mathrm{Tr}_{S^\rho}\exp(-\beta H_\xi)\rangle \tag{11}$$

in which $\rho$ and $\sigma$ are replica indices and $(\rho,\sigma)$ denotes different pairs of replicas. Here, $\mathrm{Tr}_{S_\rho}$ means the trace over the states, and

$$m_\rho^{1\nu}=\frac{1}{N}\sum_i\xi_i^{1\nu}S_i^\rho \tag{12}$$

is the overlap of the replicated state with an example $\xi_i^{1\nu}$ of the specified concept, while

$$q_{\rho\sigma}=\sum_i S_i^\rho S_i^\sigma \tag{13}$$

is the spin-glass order parameter for $\rho\neq\sigma$ and $r_{\rho\sigma}$ is the usual auxiliary parameter that will be interpreted below. Also,

$$G(q_{\rho\sigma})=\int_{-\infty}^{+\infty}\prod_{\nu,\rho}[dx_{\nu,\rho}/\sqrt{2\pi}]\exp\left(-\frac{1}{2}\sum_{\nu,\rho}x_{\rho\nu}^2+\frac{\beta}{2}\sum_{\rho,\sigma}\sum_{\lambda,\nu}x_{\rho\sigma}x_{\lambda\nu}B_{\lambda\nu}Q_{\rho\sigma}\right) \tag{14}$$

in which

$$B_{\lambda\nu}=b^2+(1-b^2)\delta_{\lambda,\nu},\qquad Q_{\rho\sigma}=q_{\rho\sigma}+(1-q_{\rho\sigma})\delta_{\rho,\sigma}, \tag{15}$$

and, finally,

$$H_{\xi}=\sum_{\nu,\rho} m_{\rho}^{1\nu}\xi^{1\nu}S^{\rho}+\frac{1}{2}\sum_{(\rho,\sigma)}(\alpha c r_{\rho\sigma}+\beta\Delta^{2}q_{\rho\sigma})$$

$$\times S^{\rho}S^{\sigma}-h^{1}\xi^{1}\sum_{\rho}S^{\rho}. \tag{16}$$

So far we have the averaged replicated partition function prior to any assumption in replica space. It is interesting to note that, in the limit $c\to 0$, the dependence on $r_{\rho\sigma}$ and part of that in $q_{\rho\sigma}$ drop out and, with a rescaling of $\alpha$ and $\beta$ that involves the number of examples $s$, $\langle Z^{n}\rangle_{c}$ becomes formally similar, for general $s$, to the expression obtained for the memorization problem in this limit by Watkin and Sherrington [5] and is identical to their result, as it should be, when $s=1$. In this case it coincides with the expression for the SK spin-glass model [8], after a local gauge transformation in which a state $\{S_i\}$ is replaced by $\{\xi_i^{\mu}S_i\}$, for all $i=1,\ldots,N$. A crucial step in this transformation is that $(\xi_i^{\mu})^2=1$. In contrast, in our case we have an effective magnetization given by the symmetric overlap $m_s=m^{1\nu}$, for $\nu=1,\ldots,s$, which characterizes the categorization phase, as discussed below, and that is associated with an average sum $\Sigma_{\nu}\xi_i^{1\nu}/s$ that cannot be used in any simple gauge transformation, unless $s$ (or $b$)$=1$. This observation has an important consequence on an argument concerning the phase boundary to the ordered state in the limit of vanishing connectivity, which will be discussed below.

Assuming now replica symmetry, we write

$$\begin{aligned} m_{\rho}^{1\nu}&=m^{1\nu}\ \text{ for all }\ \rho,\\ q_{\rho\sigma}&=q,\quad \rho\neq\sigma\\ r_{\rho\sigma}&=r,\quad \rho\neq\sigma \end{aligned} \tag{17}$$

and find, up to constant terms,

$$f=\frac{1}{2}\sum_{\nu}(m^{1\nu})^{2}+\frac{C\alpha r c}{2}+\frac{\beta(q\Delta)^{2}}{4}-\frac{\alpha c}{\beta}$$

$$\times\ln G(q)-\frac{1}{\beta}\int_{-\infty}^{+\infty}Dz\left\langle \ln\left[2\cosh\beta\right.\right.$$

$$\left.\left.\times\left(\sqrt{\alpha r c}z+\sum_{\nu}m^{1\nu}\xi^{1\nu}-h^{1}\xi^{1}\right)\right]\right\rangle \tag{18}$$

in which $Dz\equiv\exp(-z^2)dz/\sqrt{2\pi}$ is the usual Gaussian measure, $C=\beta(1-q)$, $\Delta$ specifies the width of the random Gaussian interaction, as discussed in Sec. I, while

$$\ln G(q)=-\frac{1}{2}\left((s-1)\ln(1-C\theta_1)+\ln(1-C\theta_2)\right.$$

$$\left.-\beta q s\frac{1-C\theta_1\theta_2}{(1-C\theta_1)(1-C\theta_2)}\right). \tag{19}$$

The replica-symmetric order parameters are given by the saddle-point equations in zero external field,

$$m^{1\nu}=\left\langle\left\langle \xi^{1\nu}\tanh\beta\left(\sqrt{\alpha r c}z+\sum_{\nu}m^{1\nu}\xi^{1\nu}\right)\right\rangle\right\rangle_{z},$$

$$q=\left\langle\left\langle \tanh^{2}\beta\left(\sqrt{\alpha r c}z+\sum_{\nu}m^{1\nu}\xi^{1\nu}\right)\right\rangle\right\rangle_{z}, \tag{20}$$

$$r=sq\frac{(1-C\theta_1\theta_2)^2+(s-1)b^4}{(1-C\theta_1)^2(1-C\theta_2)^2}+q\frac{\Delta^2}{\alpha c},$$

where $\langle\ \rangle_z$ denotes the integral over the Gaussian measure, $\theta_1=1-b^2$, and $\theta_2=1+(s-1)b^2$. In the noiseless limit, where $\beta\to\infty$, we have that $q\to 1$ and $C$ remains finite.

The order parameters may be interpreted as follows; $m^{\mu\nu}=\langle\xi^{\mu\nu}\langle S\rangle_T\rangle$ is the overlap between the state of the network and an example of the concept to be recognized, here $\mu=1$; $q=\langle\langle S\rangle_T^2\rangle$ is the spin-glass order parameter, and

$$r=\frac{1}{\alpha}\sum_{\mu\geqslant 2}\sum_{\nu}\langle(m^{\mu\nu})^2\rangle \tag{21}$$

is the contribution of the *uncondensed* overlaps in this problem [24]. The averaged free-energy density and the order parameters for the symmetrically dilute Hopfield model are recovered in the limit $s=1=b$ [6]. Also, in the limit $c\to 0$, we recover the equations of Watkin and Sherrington [5] for the extremely dilute network. As one would expect, there is no need for the parameter $r$ in this limit. Making use of Eqs. (8) and (18) one obtains, in the replica-symmetric theory, the overlap

$$m^{1}=\left\langle\left\langle \xi^{1}\tanh\beta\left(\sqrt{\alpha r c}z+\sum_{\nu}m^{1\nu}\xi^{1\nu}\right)\right\rangle\right\rangle_{z}, \tag{22}$$

with the particular concept.

Next, we have to make a choice for the overlap $m^{1\nu}$ with the examples of concept $\mu=1$. One way of doing this is taking [4]

$$m^{1\nu}=\delta_{1\nu}(m^{11}-m_{s-1})+m_{s-1}, \tag{23}$$

where $m_{s-1}$ is the symmetric overlap with $s-1$ examples. This is the appropriate choice when there is a bias for storing single examples by means of a given learning rule, as in the present problem. The categorization problem with competing symmetric and retrieval states in nondilute networks has been studied elsewhere [4,9,19]. Alternatively, one may directly consider the symmetric overlap with $s$ examples, $m^{1\nu}=m_s$, $\nu=1,\ldots,s$. This enables us to write

$$\sum_{\nu}m^{1\nu}\xi^{1\nu}=m_s x_s, \tag{24}$$

in terms of the symmetric sum of $s$ examples, $x_s=\Sigma_{\nu}\xi^{1\nu}$, which is a random variable that follows a binomial distribution dependent on the concept $\xi^1$ [9]. However, the emergence of other features, such as the recognition of concepts, must then be a spontaneous property of the network which should not depend on the particular choice of Eq. (23). The symmetric overlap with $s$ examples, given by

$$m_s=\frac{1}{sN}\sum_{i}\sum_{\nu}\xi_i^{1\nu}\langle S_i\rangle_T, \tag{25}$$

describes the simultaneous recognition of a set of examples by the state of the network. This becomes, in replica-symmetric mean-field theory,

$$m_s=\frac{1}{s}\langle\langle x_s \tanh\beta(\sqrt{\alpha r c}z+m_s x_s)\rangle\rangle_z, \tag{26}$$

where the inner brackets denote an average over the probability distribution of $x_s$ that includes an average over $\xi^1$.

The relevance of the symmetric mixture states, with finite $m_s$, is that they characterize the categorization phase for a given number of examples $s$, with nonzero overlap $m^1$ with a concept, by taking into account the common features of the examples. This can be seen most easily by considering the noiseless limit, $\beta\to\infty$, of $m^1$ in Eq. (22). Indeed, this yields

$$m^1=\langle\xi^1 \operatorname{erf}(m_s x_s/\sqrt{2\alpha r c})\rangle, \tag{27}$$

which only is finite for $m_s\neq 0$.

The mean-field equations in the replica-symmetric theory for the symmetrically dilute Hopfield model for the retrieval problem [6], as well as those for the SK spin-glass problem [8], are not valid below a de Almeida–Thouless (AT) line [25] in a large part of the relevant phase diagram, here of $T$ vs $\alpha$. In contrast, the breakdown of the replica-symmetric mean-field equations is far less important in the fully connected Hopfield model [24]. Thus, it is important to determine that line, particularly where it meets the phase boundary between the categorization phase and the pure spin-glass phase. Extending the calculation of the AT line to the present case of a model with hierarchically correlated patterns, we find that it is given by the joint solution of the equation

$$\frac{\beta^2\alpha r c}{q}\left\langle\left\langle \operatorname{sech}^4\beta\left(\sqrt{\alpha r c}z+\sum_\nu m^{1\nu}\xi^{1\nu}\right)\right\rangle\right\rangle_z=1, \tag{28}$$

together with the saddle-point equations for the order parameters.

The numerical solution of the mean-field equations in the replica-symmetric theory and the AT line yield the results for the phase diagrams and the order parameters discussed in the next section.

## IV. RESULTS

The choice in Eq. (24) illustrates the kind of stable solutions we concentrate on in this work. The retrieval of a single example, all $s$ examples being equivalent, is given by $m^{11}\neq 0$. The categorization phase ($C$), in which the network performs a generalization task for which it has not been specifically trained, like the recognition of concepts, is specified by the fully symmetric overlap $m^{1\nu}=m_s\neq 0$, for $\nu=1,\ldots,s$, since we do not consider the retrieval of single examples. A finite overlap $m^1$ with a concept appears in that phase and there is, in general, a nonzero spin-glass order parameter $q$. Furthermore, there is a spin-glass phase (SG) described by $m_s=0$ and $m^1=0$, while $q\neq 0$, and a disordered paramagnetic phase ($P$), where $m_s=0$, $m^1=0$, and $q=0$. There are, of course, other solutions to the saddle-point mean-field equations, in which we are not specifically interested in this work.

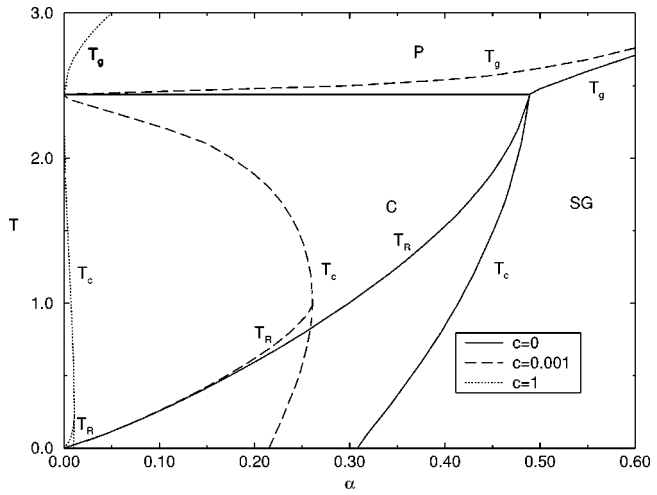


FIG. 1. Phase diagram for the synaptic noise $T$ as a function of the ratio $\alpha$ of recognized concepts, for a correlation parameter $b=0.4$, $s=10$ examples, and connectivity $c=1$, 0.001, and $0(c\to 0)$. The categorization phase ($C$) appears to the left of the lines $T_c$, as indicated explicitly for $c=0$, and the spin-glass phase (SG) to the right of $T_c$. The paramagnetic phase ($P$) appears above $T_g$ and $T_R$ are the de Almeida–Thouless lines.

First we consider the phase diagram for the temperature against the ratio $\alpha$ of recognized concepts, shown in Fig. 1, for various values of the connectivity $c$ when the overlap parameter between an example and the corresponding concept is $b=0.4$ for $s=10$ examples. Similar diagrams are obtained for other values of these parameters. Increasing either $b$ or $s$, keeping the other one fixed, allows one to reach larger values of $\alpha$, that is, to recognize a larger number of concepts and to support a bigger noise level $T$. The categorization phase appears to the left of the phase boundary $T_c$, while spin-glass states appear everywhere below the $P$-SG phase boundary $T_g$, except when $c=0$. Indeed, the spin-glass states turn out to be unstable within the categorization phase in the extremely dilute limit. In the region within the boundary $T_c$, when stable categorization and spin-glass states coexist for finite $c$, the former are more stable for small $\alpha$, while the latter become more stable for larger $\alpha$. This is similar to the competition between retrieval and spin-glass states in the symmetrically dilute network for the memorization problem [6]. Retrieval states of examples are expected to appear at lower values of $\alpha$.

Note that the categorization phase is considerably enhanced by synaptic dilution of the network, and the globally stable spin-glass states are correspondingly reduced, in particular in the limit $c\to 0$. It is also worth comparing our phase diagram for the symmetrically dilute network, in the low connectivity limit, with the phase diagram found for the categorization problem in the extremely *asymmetric* dilute network [12]. The phase boundary for the latter, not shown in the figure, starts at the place where the $C$-$P$ phase boundary meets the $T$ axis and it decreases continuously down to the $\alpha$ axis where it meets the low-$T$ end point of the $C$-SG phase boundary for the extremely dilute symmetric network. Thus, the critical ratio $\alpha_c$, at $T=0$, is the same in the two models. Moreover, it can be argued that the behavior in either of the limits $T=0$ or $\alpha=0$ has to be the same. The reason for this is that disorder due to temperature, at $\alpha=0$,

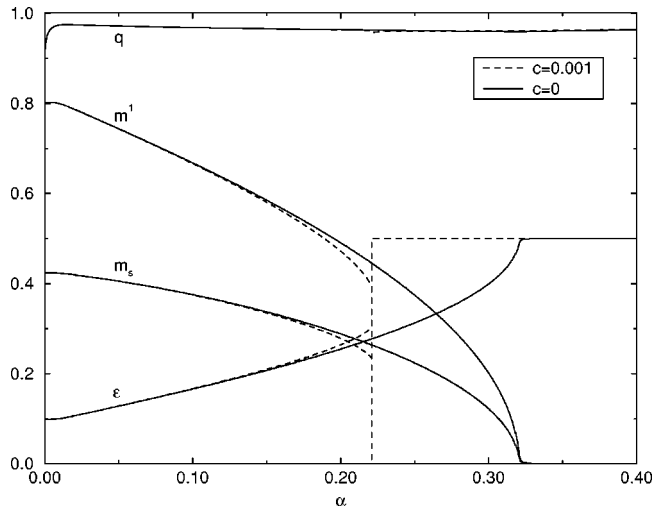


FIG. 2. Overlap with a concept $m^1$, symmetric overlap with $s$ examples $m_s$, spin-glass order parameter $q$, and categorization error $\epsilon(s)$ for $T=0.1$, $b=0.4$, $s=10$, and connectivity $c=0$ or $c=0.001$.

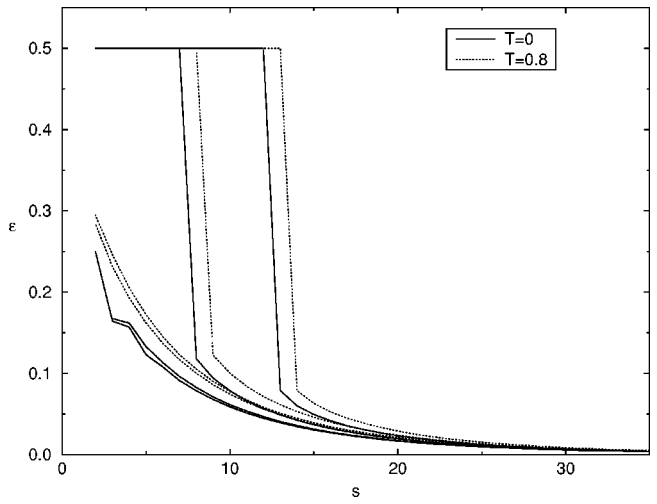


FIG. 3. Categorization curves $\epsilon(s)$ for $c=0$ (the same as $c=0.001$), $c=0.1$, 0.5, and 1, from left to right, with $\alpha=0.03125$, $b=0.5$ for $T=0$ and $T=0.8$.

should have the same effect on both models. On the other hand, at $T=0$ we are left with stochastic noise from the macroscopic number of concepts, which is assumed to be Gaussian to start with in the asymmetrically dilute network, while it becomes Gaussian in our case through the assumption of replica symmetry. As a result it turns out that there is a much larger categorization phase in the symmetrically dilute network.

The AT lines $T_R$, below which the replica-symmetric solutions for the order parameters become unstable to replica symmetry-breaking perturbations, are also shown in Fig. 1. Within the accuracy of our calculations, these lines meet the corresponding replica-symmetric phase boundaries where the slope of the latter change sign. This is similar to what has been found before for the memorization problem [6,26]. There are reasons to believe that the phase diagram and the performance of the network, discussed here and below, provide lower limits to the exact results for these features, as will be argued in the next section.

To judge to what extent a larger categorization phase is more useful with stronger dilution, we consider now the performance of the network. Indeed, the performance for a given $\alpha/\alpha_0$ (where $\alpha_0=2/\pi$ is the critical storage ratio for the extremely dilute Hopfield model) within the categorization phase is improved by dilution, as demonstrated by the order parameters, shown in Fig. 2, for $T=0.1$, $b=0.4$, $s=10$, and either for $c=0$ or $c=0.001$. In the case of vanishing connectivity, the symmetric overlap with the examples and, hence, the overlap with a concept, vanishes continuously on approach to the phase boundary from below, while for finite connectivity the overlaps drop discontinuously to zero on the phase boundaries, as in the case of the fully connected network. Note that the discontinuities decrease with dilution. As one would expect, however, the overlaps with a concept or of a mixture state decrease in both cases with an increasing ratio $\alpha$, while the categorization error increases. Also, since the $C$-SG phase boundaries have an increasing critical $\alpha_c(T)$ for decreasing $c$, the overlap with the concepts and, hence, the categorization ability of the network should decrease on the phase boundary with decreasing connectivity. Indeed, we checked that this is the case.

The improvement of the categorization ability of the network with dilution can best be seen from the categorization curves shown in Fig. 3 for the error $\epsilon(s)$ in terms of the number of examples, for connectivity $c=0$ (the same as $c=0.001$), $c=0.1$, 0.5, and 1, and for fixed $\alpha$ and $b$, either for $T=0$ or $T=0.8$. For a number $s$ of examples smaller than a critical $s_c$, there is an error of 0.5 due to a vanishing overlap with the concepts and the presence of a spin-glass state. We remind the reader of this almost always present state for low $T$. On the other hand, there is a rapid drop in the categorization error at a critical number of examples, $s_c$, as found before for the fully connected network, and this number is lower the smaller the connectivity in the present model. Thus, an extremely dilute network tends to categorize for a smaller number of examples than a fully connected network. It should be pointed out that the categorization curves have a nonmonotonic dependence on $T$, for values below $T=0.8$, not shown for clarity in the figure. This will be discussed below.

The effects of stochastic noise on the performance of the network within the categorization phase, due to the presence of a macroscopic number of concepts, at $T=0$, are shown for $b=0.2$ and $c=0$ by the categorization curves in Fig. 4, for various $\alpha/\alpha_0$. The starting point for $s=1$ corresponds to the retrieval of a single example. Note that the categorization error first decreases monotonically with an increase in the number of examples $s$, for small $\alpha$, while for larger values it first increases until an appropriate $s$ has been reached, starting to decrease thereafter. The reason for this is the competition between symmetric mixture states that favor categorization and the presence of spin-glass states that tend to destroy it. Eventually, when $\alpha/\alpha_0$ is close to 0.15, a spin-glass state with $\epsilon=0.5$ is reached continuously, reflecting the nature of the $C$-SG phase boundary. With an increase in the number of examples, however, the state of the network may again get into the categorization phase starting to recognize the concepts. Similar results are obtained for other values of the correlation parameter $b$. For larger values, the trend towards categorization starts for a lower $s$.

A similar situation occurs at $T=0$ for small but finite

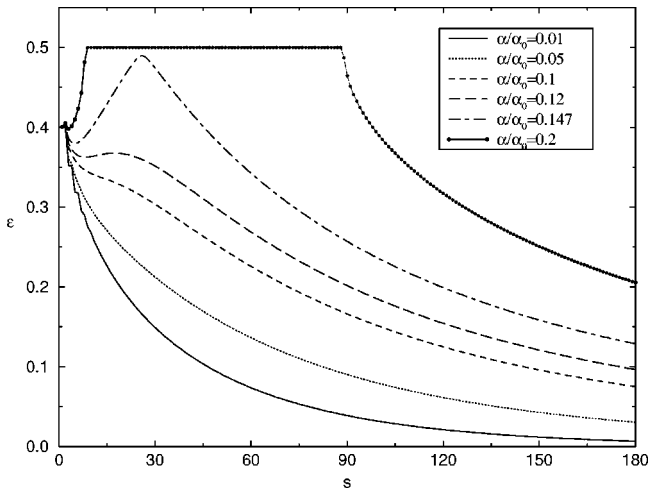


FIG. 4. Categorization curves $\epsilon(s)$ at $T=0$, for $b=0.2$, $c=0$, and several values of $\alpha/\alpha_0$, as indicated, where $\alpha_0=2/\pi$ is the storage ratio of the extremely dilute Hopfield model.

connectivity and small $\alpha$. However, for $c=0.001$, say, when a critical $\alpha/\alpha_0$ of 0.0936 is reached the categorization error jumps *discontinuously* on the phase boundary to a spin-glass value of 0.5, in accordance with the nature of the transition discussed in the context of Fig. 1. The results in either case clearly indicate that an increase in the stochastic noise level, due to a macroscopic number of concepts, always seems to deteriorate the performance of the network. Except for very small $\alpha$, where the stochastic noise simply slows down the recognition of concepts with the number of examples presented to the network, there is a monotonic destabilization of the symmetric mixture states for a *given* concept by the interference of the random overlaps with the examples of *all* other concepts. These are effects that only appear for a sizable number of examples.

A somewhat different behavior of the network in the categorization phase is obtained when the effects of finite noise level $T$ are taken into account. At relatively high stochastic noise $\alpha$, this is shown in Fig. 5 for $\alpha/\alpha_0=0.3$, with $b=0.3$, and $c=0$. It can be seen that the effect is, first, to improve the categorization ability of the network for low $T\leqslant 0.4$, while a further increase in the noise level deteriorates

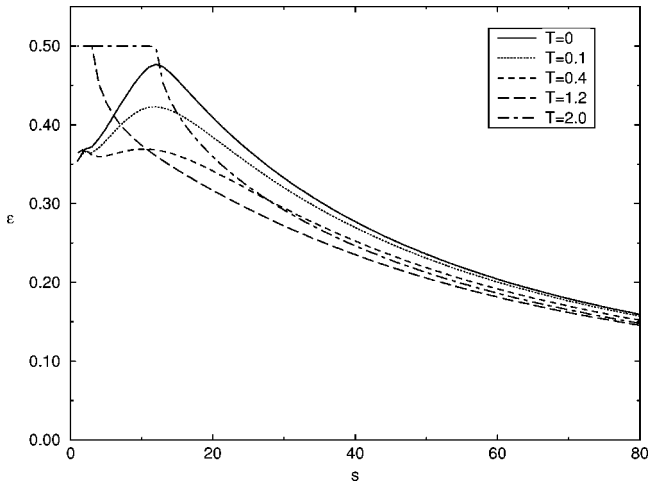


FIG. 5. Categorization curves $\epsilon(s)$ at $\alpha/\alpha_0=0.3$, $b=0.3$, and $c=0$ for several temperatures $T$ as indicated.

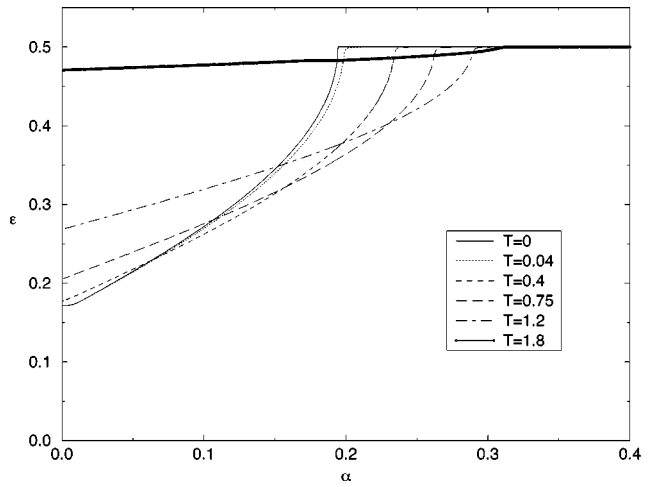


FIG. 6. Dependence of the categorization error $\epsilon$ on $\alpha$ for varying $T$, with $b=0.3$, $s=10$, and $c=0$.

the categorization ability. This non-monotonic behavior in $T$ is also present at lower $\alpha$ and it is a feature of the network that is only present for more than one example. Indeed, we found that for all values of $T$ for which the network is in the categorization phase, the categorization curves start at the same error $\epsilon(s)=0.35$ for $s=1$, as shown in Fig. 5. We also found nonmonotonic behavior in $T$ near the phase boundary $C$-SG for very small $c$, but it may not be easy to see for moderate to large $c$.

The different dependence of the categorization performance of the network on both types of noise can be seen best in the evolution of the categorization error with an increase in either $\alpha$ or $T$, keeping the other one fixed, shown in Fig. 6 for $c=0$, $b=0.3$, and $s=10$ examples. In contrast to the dependence on $\alpha$, the dependence on $T$ is clearly nonmonotonic apparently down to asymptotically small $\alpha$. As a result, the effects of an increase in stochastic noise level can be compensated, up to a certain extent, by an increase in synaptic noise $T$.

## V. SUMMARY AND CONCLUDING REMARKS

We studied in this work the effects of stochastic and synaptic noise on the ability to recognize the ancestors of a hierarchical two-level structure of patterns in a symmetrically dilute network trained only with the descendants in that structure. This is a network that has a nontrivial dynamics and, consequently, it may have rich and interesting equilibrium behavior, even in the low-connectivity limit, as demonstrated here. Due to the presence of a correlation between an ancestor and its descendants, there appears a correlation between the latter that stabilizes the symmetric mixture states with the descendants. These are states that generate finite overlaps with the ancestors. Although the phase diagrams look similar to those for the memorization (also called the retrieval) problem in the symmetrically dilute network, the interpretation of the ordered phase is quite different.

The categorization ability of the symmetrically dilute network trained with a generalized Hebbian rule was studied in this work for varying connectivity, with particular interest in the extremely dilute limit for which there is a considerable enhancement of the categorization phase when compared

with that for the fully connected network, as can be seen by a comparison of our present result and that of a previous work [9]. As a result there is a reentrant categorization to spin-glass phase boundary with a considerably increased critical ratio $\alpha_c$. For $\alpha$ below, the categorization curves for the error $\epsilon(s)$ are found to have a nonmonotonic dependence on the synaptic noise level $T$, for low to moderate values. In contrast, the asymmetrically dilute network has only monotonic behavior. We also find that the transition from the categorization to the spin-glass phase boundary is a discontinuous one for low but finite connectivity, as in the case of the fully connected network, and that the transition becomes a continuous one in the vanishing connectivity limit.

The explicit results presented in this work were obtained in replica-symmetric mean-field theory. We also determined the de Almeida–Thouless line that limits the stability of our results to replica symmetry-breaking perturbations, for each connectivity. As a consequence, the reentrant parts of the $C$-SG phase boundaries are not stable to these perturbations. In particular, one may worry about the low-connectivity limit where the whole phase boundary is a reentrant one. However, as we showed in Sec. III, the average replicated partition function, $\langle Z^n \rangle_c$, for our problem, prior to the assumption of replica symmetry, is formally similar but not identical, to that of the SK spin-glass model. This is in contrast to the result of Watkin and Sherrington for the memorization problem in the extremely dilute symmetric network, for which there is a formal identification [5]. On the basis of a plausible assumption for this model, it has been argued that the exact spin-glass to ferromagnetic phase boundary should be a straight line parallel to the $T$ axis [27]. In a replica mean-field theory this means a result to all orders in replica symmetry-breaking perturbations [28]. The argument cannot, strictly, be used in our case since there does not seem to be a simple gauge transformation that can formally identify our problem to the SK spin-glass problem, as discussed in Sec. III. On the other hand, it is known that results for the order parameter, in the memorization problem within replica-symmetric mean-field theory, give lower bounds to results obtained by numerical simulations on large networks. One may argue for our problem that the shape of the true categorization to spin-glass phase boundary, in the low-connectivity limit, should not be very different from that of the SK model.

The reason for considering a dilute network is that it is a more economical architecture than the fully connected network and, in the present case of symmetric dilution, it has also a better performance. Indeed, particularly in the low-connectivity limit, the network is fairly robust to synaptic noise for a moderate $\alpha$, since at least the upper part of the categorization phase shown in Fig. 1 is correctly given by the replica-symmetric solution. Although a symmetric dilution is not appealing on biological grounds, it may be a more suitable architecture for hardware implementations.

## ACKNOWLEDGMENTS

We are indebted to Alba Theumann for showing us how to calculate the de Almeida–Thouless line in mean-field theory for a network with hierarchically correlated patterns, and thank R. Erichsen, Jr. for stimulating discussions. This work was supported by FAPERGS (Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, Brazil), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil), and FINEP (Financiadora de Estudos e Projetos, Brazil).

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[2] D. J. Amit, *Modeling Brain Function* (Cambridge University Press, Cambridge, England, 1989).
[3] J. F. Fontanari and R. Meir, Phys. Rev. A **40**, 2806 (1989).
[4] J. F. Fontanari, J. Phys. (France) **51**, 2421 (1990).
[5] T. L. H. Watkin and D. Sherrington, Europhys. Lett. **14**, 791 (1991).
[6] A. Canning and J.-P. Naef, J. Phys. I **2**, 1791 (1992).
[7] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[8] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **32**, 1792 (1975).
[9] P. R. Krebs and W. K. Theumann, J. Phys. A **26**, 3983 (1993).
[10] E. N. Miranda, J. Phys. I **1**, 999 (1991).
[11] M. C. Branchtein and J. J. Arenzon, J. Phys. I **2**, 2019 (1992).
[12] C. R. da Silva, F. A. Tamarit, N. Lemke, J. J. Arenzon, and E. M. F. Curado, J. Phys. A **28**, 1593 (1995).
[13] D. R. C. Dominguez and W. K. Theumann, J. Phys. A **29**, 749 (1996).
[14] D. R. C. Dominguez, Phys. Rev. E **54**, 4066 (1996).
[15] D. R. C. Dominguez and W. K. Theumann, J. Phys. A **30**, 1403 (1997).
[16] D. R. C. Dominguez and D. Bollé, Phys. Rev. E **56**, 7306 (1997).
[17] D. R. C. Dominguez, Phys. Rev. E **58**, 4811 (1998).
[18] C. Rodriguez Neto and J. F. Fontanari, J. Phys. A **31**, 531 (1998).
[19] J. A. Martins and W. K. Theumann, Physica A **253**, 38 (1998).
[20] R. Lima Costa and A. Theumann, Physica A (to be published).
[21] R. Erichsen, Jr., W. K. Theumann, and D. R. C. Dominguez (unpublished).
[22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[23] H. Sompolinsky, Phys. Rev. A **34**, 2571 (1986); in *Heidelberg Colloquium on Glassy Dynamics, Heidelberg, 1986*, edited by J. L. van Hemmen and I. Morgenstern (Springer-Verlag, Berlin, 1987), p 485.
[24] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).
[25] J. R. L. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).
[26] J.-P. Naef and A. Canning, J. Phys. I **2**, 247 (1992).
[27] G. Toulouse, J. Phys. (France) Lett. **41**, L447 (1980).
[28] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).