Inference and chaos by a network of nonmonotonic neurons

David R. C. Dominguez*
Theoretical Physics Department C-XI, Universidad Autonóma de Madrid, Cantoblanco, 28049 Madrid, Spain
(Received 7 June 1996)

The generalization properties of an attractive network of nonmonotonic neurons that infers concepts from samples are studied. The macroscopic dynamics for the overlap between the neuron states and concepts, as well as the activity of the neurons, is obtained and numerically studied. Complex behavior leading from fixed points to chaos through a cascade of bifurcation is found when we increase the correlation between samples, decrease the activity of the samples and the load of concepts, or tune the threshold of fatigue of the neurons. Both the information dimension and the Liapunov exponent are given and a phase diagram is built. [S1063-651X(96)06110-7]

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

I. INTRODUCTION

There are two sources for building more sophisticated models of brain behavior as associative memory other than the original Hopfield model for neural networks. One is the closeness to realistic facts observed in neural systems and the other is the trial to attain more complex learning abilities. Among the successful attempts for the former are the multistate neuron models [1], which include three-state, analog, and nonmonotonic neurons. The capability of generalization, the inference of rules from examples, is an example of the latter [2,3]. The categorization or capability to retrieve patterns of activity in different levels of a hierarchical classification is another instance [4]. Here we work out a connection between the multistate neural networks and the categorization networks, which leads to a different kind of generalization, as a property of such neural devices to infer a full concept from small samples of that concept. While in most neural models of learning (see Ref. [5] and references therein) the generalization function measures the ability of the network to give correct answers to each question, after being trained with samples of question-answer pairs, in the present model the samples are patterns that carry information about the concepts, which can be identified with the answers.

The multistate neuron model was introduced to account for some degrees of ignorance of pieces of the full pattern. It differs qualitatively from the two-state model because, in the absence of part of the information, fewer bits are required to represent the so-called *small pattern*, as one picked up information from the active sites, keeping the inactive sites off. Several models of multistate neurons were studied with the Hebbian learning algorithm. The behavior of the analog neural network was studied first in the case of binary memorized patterns [6] and yields a phase diagram similar to that of stochastic binary neurons, replacing the temperature T for the inverse of the gain parameter (the slope at the origin of the transfer function). The three-state neural network in the presence of three-state uncorrelated patterns was studied within the extremely diluted synapse scheme, showing an enhancement of the storage capacity with an adequate control of its firing threshold. This is more notable when the pattern *activity* (the rate of nonvanishing states per sites of the pattern) is small [7]. Nonmonotonic neural networks, which take into account the fatigue of each neuron after being exposed to a large post-synaptic potential, was studied by means of a signal-noise analysis [8]. This network exhibits an interesting superretrieval phase, with vanishing error even for a large number of learned patterns. If the neurons are able to change their states to the opposite of the signal of its local field, the capacity of the network becomes even larger than that of three-state neurons [9].

For all these cases a parallel deterministic dynamics was assumed, given by the set of equations

$$\sigma_{it+1}=F_\theta(h_{it}), \quad i=1,\ldots,N, \tag{1}$$

where σ_{it} is the neuron state of site i at time t, θ is the threshold parameter that represents deviation of the signal function, and, as usual, only odd bounded input-output (IO) F_θ functions are considered. The local field of site i at time t is

$$h_{it}=\sum_{i\ (\neq j)}^{N} J_{ij}\sigma_{jt}, \tag{2}$$

J_{ij} being the elements of the synaptic matrix. In the case of three-state neurons and patterns, the existence of a threshold for which the retrieval is optimized was also found by statistical mechanical techniques within the replica symmetric approximation [10], but it is not useful for the nonmonotonic network since it does not have an energy function [11].

The task of generalization by a neural network can be realized in a manifold of contexts. One kind is the categorization, which takes place if we use an alternative Hebbian learning algorithm that stores s examples having correlation b with one hierarchical ancestor for each of the p concepts. For the connected model, in the context of an attractor neural network, the following modified Hebbian learning algorithm has been studied [4]:

$$J_{ij}=\frac{1}{N}\sum_{\mu}^{p}\sum_{\rho}^{s}\eta_i^{\mu\rho}\eta_j^{\mu\rho}. \tag{3}$$

*Electronic address: david@tfdeci.fys.kuleuven.ac.be

The correlation of the learning example $\eta^{\mu\rho}$ with one concept of the set $\{\xi^{\mu}\}$ is $\langle \eta_i^{\mu\rho}\xi_j^{\nu}\rangle = b\,\delta_{\mu\nu}\delta_{ij}$. The phase transition from a disordered to a generalization phase, where the neurons retrieve one concept, was found to be discontinuous with b for a fully connected network [12] and smooth for a diluted network [13]. After sufficiently increasing s or b and decreasing $\alpha \equiv p/N$, the error in the generalization became small enough to consider such a task successful.

Another interesting kind of generalization is inference. The coherence between the learned patterns with activity $a \ll 1$ allows many patterns to be simultaneously retrieved [14]. Then, by learning small patterns, we can infer the existence of a whole pattern, with activity $a \sim 1$. Enlarging the effective size of the pattern, we can extract much more information than the original patterns contain. For instance, we would see wood where before we had seen only trees. To obtain such an inferential property, however, a more sophisticated algorithm is required. Fortunately, it comes from a modified version of the Hebbian algorithm in Eq. (3). Nevertheless, it requires a mathematically difficult effort to make a connection between generalization and multistate neurons. A unique investigation treating the generalization with analog neurons [15] uses binary examples. Then it is worth analyzing such models in their simpler, extremely diluted version, which yields an exactly soluble dynamics and is biologically relevant at the same time [16]. In this version, a network of three-state monotonic neurons shows a clear improvement of the performance as a generalization device if small activity examples are learned [17].

We describe the model of a network of nonmonotonic neurons in the next section. After obtaining the recursion relations for the inferential properties in Sec. III, we present our conclusions in Sec. IV, drawing the curves of generalization with special attention to the nonsteady solutions.

II. MODEL

We adopt the dynamics given in Eqs. (1) and (2) and start by defining an IO function. Although most works employ stairlike (modeling by the q Ising network) or other monotonic functions F_θ, we will avoid this restriction and choose instead

$$F_\theta(x) \equiv \begin{cases} \mathrm{sgn}(x), & |x| < \theta \\ 0, & |x| \geq \theta. \end{cases} \tag{4}$$

Thus the IO function tells us the way in which the network updates each neuron, which becomes fatigued outside the interval $|h_{it}| < \theta$, according to Eq. (1).

For the synaptic interactions we will assume the Hebbian algorithm in Eq. (3), but the examples to be learned will be three-state variables, like the neuron state itself. In order to preserve the odd symmetry of the neurons, those patterns are uniformly distributed around the zero state. Thus the examples $\eta_i^{\mu\rho}$ are independent random variables built from the *concepts* ξ_i^{μ} through the stochastic process

$$\eta_i^{\mu\rho} = \xi_i^{\mu}\lambda_i^{\mu\rho}, \quad \langle \lambda_i^{\mu\rho}\rangle \equiv b, \quad \langle (\lambda_i^{\mu\rho})^2\rangle \equiv a, \tag{5}$$

where $\xi_i^{\mu} = \pm 1$ with equal probability. The random variables introduced here $\lambda_i^{\mu\rho}$ are characterized by their mean b and their square mean a for all examples $\eta^{\mu\rho}$.

Then the parameter a is the *activity* of the examples themselves, while b is the *correlation* between examples and their respective concepts. On the one hand, we can recover the pure generalization model [4] by setting $\lambda_i^{\mu\rho} = \pm 1$ $(a=1)$ with a bias b for the positive value and threshold $\theta \to \infty$. In this simple limit the neurons are thought of as being submitted to background noise, perhaps due to some dirtiness on the pattern. On the other hand, the pure multistate model can also be obtained by taking the number of examples $s=1$ in Eq. (3) and correlation $b=1$. A low activity $a \ll 1$ indicates that at many sites the patterns are not active $|\eta_i^{\mu\rho}| \neq 1$, with the effective size of the learned patterns being $N_e = aN$. So, when the activity a is not close to 1, we can speak of a *small pattern* [1]. In our model the alternative viewpoint is the following: the small examples are *samples* of the full activity concepts to be inferred.

The task of generalization (inference) is successful if the distance between the state of the neuron and the concept ξ^{μ}, defined as $E_t^{\mu} \equiv (1/N)\Sigma_i |\xi_i^{\mu} - \sigma_{it}|$, becomes small after some time t. This is the so-called Hamming distance, which in this context is called the *generalization error*. In order to measure the quality of the retrieval of the small patterns [7] one needs to consider a Euclidean quadratic distance instead of the Hamming distance, but we are interested exclusively in the capacity of the network to infer a larger concept of full activity from the samples, in which case E_t^{μ} suffices.

Here we should remark that since E^{μ} is μ dependent it looks like a training error with respect to one pattern [5]. However, it is not dependent on the examples $\eta^{\mu\rho}$, being indeed a generalization error, which is p degenerate for all concepts, and a particular state σ near to ξ^1 can be chosen.

The relevant order parameters for the dynamics during some specified time t when the state of network is given by $\{\sigma_{it}\}$ are the *retrieval overlaps*

$$m_{Nt}^{\mu\rho} \equiv \frac{1}{aN}\sum_{j}^{N} \eta_j^{\mu\rho}\sigma_{jt} \tag{6}$$

of the αth example of the μth concept. They are normalized parameters within the interval $[-1,1]$, which attain the extreme value $m_N^{\mu\rho} = 1$ whenever $\eta_j^{\mu\rho} = \sigma_j$, by virtue of Eq. (5). Using this definition, with the synaptic interaction in Eq. (3), the local field in Eq. (2) becomes

$$h_{it} = a\sum_{\mu}^{p}\sum_{\rho}^{s} \eta_i^{\mu\rho} m_{Nt}^{\mu\rho}. \tag{7}$$

Next we need to analyze the evolution of the ps coupled equations (6) instead of the N original equations (1).

Because we are interested in the generalizing property of our network, we take an initial configuration whose retrieval overlaps are only macroscopic of $O(1)$ for the s examples of a given concept, let us say the first one, and symmetric (equal for all ρ). We write $m_{Nt=1}^{1s} = \Sigma_{\rho}^{s} m_{Nt=1}^{1\rho}$ for the *symmetric* overlap. In the thermodynamic limit, the retrieval overlaps $m_{Nt=1}^{1\rho}$ in Eq. (6) are infinite sums of independent random variables (IRVs), whose fluctuations around its mean

value $\langle\langle m^{1\rho}_{Nt=1}\rangle\rangle$ can be neglected. Then the law of large numbers (LLNs) applies to get

$$m_{t=1}\equiv\lim_{N\to\infty} m^{1s}_{Nt=1}=\langle\langle x_s F_\theta(\Lambda_{t=0})\rangle_{x_s}\rangle_{\omega_0}, \tag{8}$$

which is i-site independent. Here we have defined the variable of the field $\Lambda_{t=0}\equiv\xi^1 h_{t=0}=m_{t=0}sa^2x_s+\omega_0$, where $m_{t=0}$ is the initial symmetric retrieval overlap, $x_s\equiv(1/as)\Sigma^s_\rho\lambda^{1\rho}$, and ω_0 is the noise produced by the $p-1$ residual concepts in Eq. (6). The averages in the angular brackets are over both x_s and ω_0 terms in the field. We have used the odd property of F_θ and have written the argument in F_θ here as a sum of two different kinds of terms. The first one favors the ordering in the direction of the first concept, while the second ω_0 introduces an additional noise to the original mistakes represented for those sites where $\lambda^{1\rho}_i=-1$.

The most interesting feature for us is the generalizing property of our network. It is characterized by the overlap of the neural state with the first concept, given in the first time step by

$$M_{t=1}\equiv\lim_{N\to\infty}\frac{1}{N}\sum_i\ \xi^1_i\sigma_{it=1}=\langle\langle F_\theta(\Lambda_{t=0})\rangle_{x_s}\rangle_{\omega_0}, \tag{9}$$

which is related to the generalization error (the Hamming distance) by $E^1_{t=1}=1-M_{t=1}$. For multistate neurons it is useful to define the *dynamical activity* order parameter, given in the first time step by

$$Q_{t=1}\equiv\lim_{N\to\infty}\frac{1}{N}\sum_i\ (\sigma_{it=1})^2=\langle\langle[F_\theta(\Lambda_{t=0})]^2\rangle_{x_s}\rangle_{\omega_0}. \tag{10}$$

This accounts for the active neurons and plays a role similar to the *spin-glass* parameter of the thermodynamic equilibrium approach for binary neurons since it allows one to measure the degree of order even when there is no retrieval at all [18,19]. In Eqs. (9) and (10) we have used the LLNs for a sum of IRVs, with vanishingly fluctuations, in the thermodynamic limit.

III. DILUTED DYNAMICS

Although it is easy to solve the single time Eq. (8) and to obtain the generalization error E_t, the recursion relations for any time t are not easily solved. We then use the extremely diluted synapse approximation, for which the first time step gives exact results for any number of time steps. In this limiting situation the synaptic interactions take a vanishing value for almost all pairs of neurons $\{ij\}$ and are of the form given in Eq. (3) only for a small fraction $C/N\ll 1$ of them. Equations (8)–(10) are then reproducible for any t, with the following simple distribution ($\doteq$) of the noise caused by the examples of the $p-1$ residual concepts: $\omega_t\doteq z_p\sqrt{\alpha Q_t r}$, where $\alpha=p/C$, $r=s[a^2+(s-1)b^4]$, and $z_p\doteq N(0,1)$ is a Gaussian random variable with mean $\langle z_p\rangle=0$ and unit variance. Q_t is the *dynamical activity* at time t.

We will also use an approximation for the case of many examples ($s>10$): $x_s\doteq b/a+z_s\sqrt{(a-b^2)/(sa^2)}$ with $z_s\doteq N(0,1)$ independent of z_p. With these remarks, after some algebra with both Gaussian z_s and z_p we can write the arbitrary time-step dynamics for the macroscopic parameters, with the IO function given by Eq. (4).

The dynamical activity is

$$Q_{t+1}=\frac{1}{2}[\mathrm{erf}(A_+)-\mathrm{erf}(A_-)],\quad A_\pm\equiv\frac{m_t sab\pm\theta}{\sqrt{v_t}}, \tag{11}$$

with $v_t\equiv sa^2(a-b^2)(m_t)^2+\alpha r Q_t$, and the symmetric retrieval overlap is

$$m_{t+1}=\frac{b}{a}M_{t+1}+m_t(a-b^2)C_{t+1}, \tag{12}$$

where we have defined $\mathrm{erf}(x)\equiv\int_0^x dy\,\varphi(y)$, $\varphi(y)\equiv\exp(-y^2/2)/\sqrt{2\pi}$. Here

$$M_{t+1}=\mathrm{erf}\left(\frac{sabm_t}{\sqrt{v_t}}\right)-\frac{1}{2}[\mathrm{erf}(A_+)+\mathrm{erf}(A_-)] \tag{13}$$

is the overlap of generalization and

$$C_{t+1}\equiv\langle F'_\theta(\Lambda_t)\rangle_z=\frac{1}{\sqrt{v_t}}\left[2\varphi\left(\frac{sabm_t}{\sqrt{v_t}}\right)-\varphi(A_+)-\varphi(A_-)\right]. \tag{14}$$

We will make no restrictions about the values the parameters b and a can assume within the (0,1) interval, except that they must satisfy $a\geqslant b^2$ (the equality corresponding to constant microscopic activities $\lambda\equiv b$).

IV. ATTRACTORS AND CONCLUSIONS

Two fixed-point ordered phases can appear: namely, the *generalization* phase $\{G{:}M>0,Q>0\}$ and the *self-sustained activity* $\{S{:}M=0,Q>0\}$ (or microscopic chaotic [20]) phase. However, the most interesting attractors are the nonsteady macroscopic phases. Although Eqs. (11)–(14) are deterministic, averaged over the stochasticity induced by the extensive load $p=\alpha C$, some complex behavior remains present in the large-time dynamics. A *doubling* of period generalization phase $\{D{:}M_t>0,Q_t>0\}$ appears, without a fixed point, where cyclic or chaotic attractors arise. It can be viewed in the curves of generalization showed in the figures below.

In Fig. 1 (bottom) we see the generalization error E_t dependence on the sample correlation b and activity a, in which we took $a=b$. Fixed values of the number of examples, load rate, and threshold of fatigue are used. When b is increased until $b_1\sim 0.19$, the generalization error has a fixed-point behavior. It initially falls until an optimal value $E_t\sim 0.07$ at $b_{\mathrm{op}}\sim 0.15$. Then it reaches a first bifurcation, beyond which it oscillates between two values, exhibiting a periodic behavior. A cycle-four is found after a second bifurcation at $b_2\sim 0.31$, and this doubling of period follows until a quasiperiodic behavior takes place at $b_\infty\sim 0.35$. For $b_\infty<b<b_S$, regions of chaos intercalate with windows of periodicity. After $b_S\sim 0.6$, although the correlation is large, the activity is large too and destroys the capacity of generalization, so that $E_t=1$. The same behavior was qualitatively found as a function of activity a (b), keeping fixed b (a).

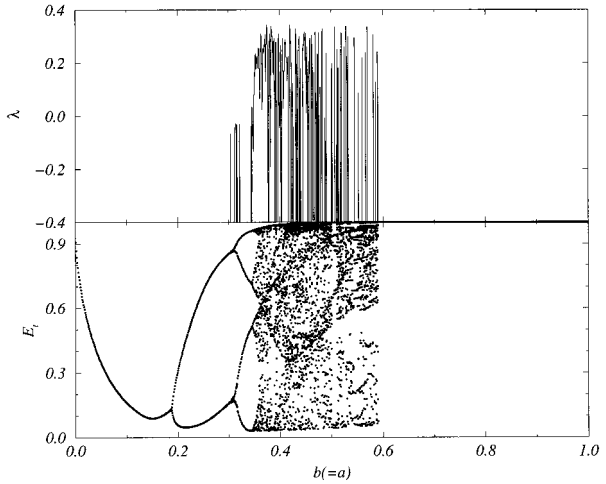

FIG. 1. Bottom: generalization error E_t as a function of the pattern correlation and activity $a=b$, the number of examples $s=20$, load rate $\alpha=0.01$, and threshold of fatigue $\theta=1$. Top: Liapunov exponent for the attractor of the bottom figure.

For sufficiently low (high) activity (correlation), E_t oscillates aperiodically, eventually closer to each chosen initial value but never equal to it.

In order to measure the degree of the nonregular behavior we calculated the Liapunov exponent in the region of $a=b$ above. It was estimated as [21] $\lambda_L \sim (1/T)\ln[\delta m_T/\delta m_0]$ for $T \gg 1$, where δm_t is the distance between two trajectories initially near each other. It gives positive values within the interval $b_\infty < b < b_S$, attaining the value $\lambda_L \sim 0.34$ at $b_C \sim 0.41$, as we can see in Fig. 1 (top). This indicates how chaotic the oscillation of E_t in this attractor is, which shows sensitivity to initial conditions. We also calculated the information dimension of the attractor, estimated by [21] $d_H \sim \ln(N_r)/|\ln(r)|$, $r \ll 1$, where N_r is the number of balls with radius r necessary to cover all points E_t. For the point b_C we got $d_H = 0.81$. The noninteger value of d_H shows that such an attractor is a fractal.

The behavior as a function of θ is shown in Fig. 2 (bottom), where the effect of the fatigue is singled out. When the threshold is small enough the generalization is poor because the local fields almost everywhere exceed θ, which lead the neurons to their fatigue phase. After $\theta_{1-} \sim 1.3$ the probability of the local field being lower than θ becomes relevant and then a periodic regime start. A chaotic regime happens for $3.8 < \theta < 6$, when the local fields fluctuate around θ. An atypical exit from the chaotic regime occurs when θ is so large that the local fields gradually leave the nonsigmoidal phase until at $\theta_{1+} \sim 15$ a new fixed point regime sets in, but now with a good generalization.

A bifurcation diagram was also found as a function of the load rate of concepts α. The noise induced by the saturation of concepts caused a large fluctuation for the local fields. Thus the chaotic behavior, which implies a very sensitive flow of the neural states with their previous states, is lost for large α. A phase diagram of the model is shown in the Fig. 2 (center) for fixed values of a,b,s. For small values of α, a transition from a S phase to a D phase occurs whenever the threshold of fatigue crosses the solid curve. For larger values of α, the solid curve separates the S phase from the G phase.

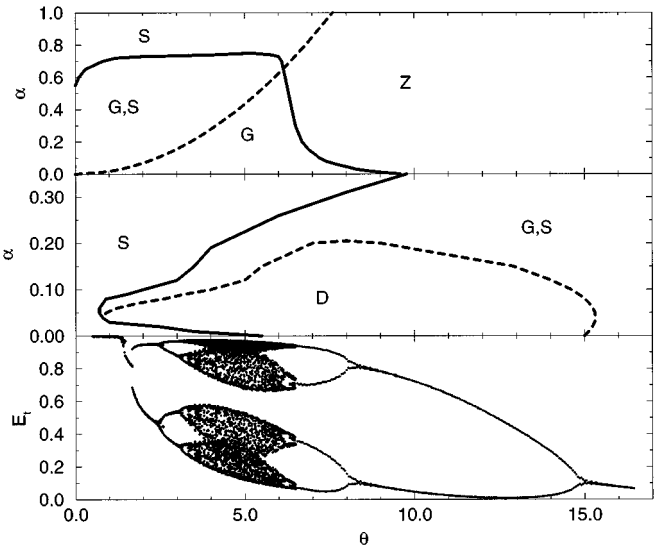

FIG. 2. Bottom: generalization error E_t as a function of θ with $b=0.5=a$, $s=50$, and $\alpha=0.05$. Center: phase diagram $\alpha(\theta)$, with $b=0.5=a$ and $s=50$. The dashed curve separates the D phase from the G phase, while the solid curve separates the S phase from the G or the D phase. Top: phase diagram $\alpha(\theta)$, with the same parameters as the center figure, but with monotonic three-state neurons.

The G phase is separated from the D phase by the dashed curve. Different from the phase diagram obtained in [23], here no phase $\{Z:M=0,Q=0\}$ can be reached, as can be seen from Eq. (11), with $m_t=0$, which reads $Q_{t+1}=\mathrm{erf}(\theta/\sqrt{\alpha r Q_t})$. For $Q_t \to 0$ we get $Q_{t+1} \to 1$. The S phase competes in one region with the G phase, but the latter is more stable overall in this region.

In order to compare with the monotonic case, for which the IO function $F_\theta(x) \equiv \mathrm{sgn}(x), |x| > \theta (\equiv 0, |x| \leqslant \theta)$ can be taken, we built the phase diagram $\alpha(\theta)$ of the Fig. 2 (top). The parameter θ here represents a threshold of fire for the neurons. There is no D phase for this case, but instead a Z phase can appear for large enough values of θ.

It is not too surprising that the motion of the neuron states themselves can be over a chaotic trajectory, where the memory of the initial configuration is not preserved. But in this case the macroscopic parameter measuring the retrieval of one pattern is almost always $M_t=0$ because the motion is ergodic over the trajectory, running equally over all possible states, the huge majority of which have vanishingly overlapped with that pattern. This is the case of the S phase. In the present model, however, the chaos appears on the less complex macroscopic trajectories for the overlap in such a manner that almost always $M_t > 0$. Then we can conjecture that in the nonsteady regimes, the network preserves a memory of what concept was used as a seed in the initial configuration. Thus it cannot be related to the properties of sequential generalization [22], for which a set of concepts can be retrieved consecutively. Because the vector of overlaps $\vec{M}_t$ can be roughly orthogonal to its previous state, many other directions M_t^μ become macroscopic each time. Only one concept, however, is persistently retrieved, at varying magnitude.

A similar result was recently found for the pure multistate model for the retrieval of patterns, but using continuous non-

monotonic neurons instead of our discrete neurons [23]. This shows that the present complex behavior is a consequence of the nonmonotonicity rather than a characteristic of the generalization model.

The diagrams in Figs. 1 and 2 demonstrate how a network of nonmonotonic neurons can exhibit a complex behavior. The coherent retrieval of samples leads to the ability to infer a large activity concept, even for a large load ratio. The periodicity of the generalization can be controlled by the activity of the samples, their correlation with each other, and the gain parameter of the neurons. We hope it is worth verifying such a behavior of the inferential properties with other learning algorithms and higher levels of hierarchy.

ACKNOWLEDGMENT

This work was financially supported by Cnpq, Brazil.

[1] J. S. Yedidia, J. Phys. A **22**, 2265 (1989).
[2] H. Watkin, A. Rau, and M. Biehl, Rev. Mod. Phys. **65**, 499 (1993).
[3] M. Opper and D. Haussler, Phys. Rev. Lett. **75**, 3772 (1995).
[4] J. F. Fontanari and R. Meir, Phys. Rev. A **40**, 2806 (1989).
[5] H. Seung, H. Sompolinsky, and N. Tishby, Phys. Rev. A **45**, 6056 (1992).
[6] C. M. Marcus, F. R. Waugh, and R. M. Westervelt, Phys. Rev. A **41**, 3355 (1990).
[7] D. Bollé, B. Vinck, and V. A. Zagrebnov, J. Stat. Phys. **70**, 1099 (1993).
[8] M. Shiino and T. Fukai, J. Phys. A **26**, L831 (1993).
[9] K. Kobayashi, Network **2**, 237 (1991).
[10] D. Bollé, H. Rieger, and G. M. Shim, J. Phys. A **27**, 3411 (1994).
[11] I. Opris, Phys. Rev. E **51**, 2619 (1995).
[12] E. Miranda, J. Phys. (France) I **1**, 999 (1991).
[13] R. Crisogono, A. Tamarit, N. Lemke, J. Arenzon, and E. Curado, J. Phys. A **28**, 1593 (1995).
[14] C. Meunier, D. Hansel, and A. Verga, J. Stat. Phys. **55**, 859 (1989).
[15] D. A. Stariolo and F. A. Tamarit, Phys. Rev. A **46**, 5249 (1992).
[16] B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[17] D. R. C. Dominguez and W. K. Theumann, J. Phys. A **29**, 749 (1996).
[18] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[19] *Neural Networks and Spin Glasses*, edited by W. K. Theumann and R. Köberle (World Scientific, Singapore, 1990).
[20] M. Bouten and A. Engel, Phys. Rev. E **47**, 139 (1993).
[21] J. P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).
[22] D. Bollé and B. Vinck, Physica A **223**, 293 (1996).
[23] C. N. Laughton and A. C. C. Coolen, J. Phys. A **27**, 8011 (1994).