Storage capacity of neural networks storing spatially correlated patterns

Mathias Schlüter and Friedrich Wagner
Institut für Theoretische Physik der Christian Albrechts Universität zu Kiel, Olshausenstraße 40, 24098 Kiel, Germany
(Received 8 July 1993)

The storage capacity of Ising spin networks storing spatially correlated patterns with explicit learning rules is investigated. The correlations are introduced by using equilibrium configurations of the Ising model at temperature $1/\beta$. Using the Hebb rule, the storage capacity decreases strongly with increasing β. This can be avoided by a modification of the Hebb rule which includes the inverse equilibrium correlation matrix of the Ising model. The theoretically derived expressions for the storage capacity are in good agreement with the numerical simulation. In order to demonstrate the working mechanism of the alternative learning rule the storage of simple linear patterns is discussed in detail.

PACS number(s): 87.10.+e, 05.50.+q

I. INTRODUCTION

Models of Ising spin networks used as an *associative memory* are in general formulated in spin variables $s_x = \pm 1$ $(x = 1, \ldots, N)$ describing the activity of N neurons (firing or not firing).

The dynamic of the system describes the change of the variables in time. It is established by calculating the new state $s_x^{(n+1)}$ of neuron x as a nonlinear function of the previous network state $s^{(n)}$. In the simple case of two-state neurons the nonlinear function is given by the "sgn" function

$$s_x^{(n+1)} = \mathrm{sgn}\left(\sum_{y=1}^{N} w_{xy} s_y^{(n)}\right) \quad (w_{xx} = 0) \,. \tag{1}$$

A learning rule specifies how to store a given set of p patterns $\xi_x^\mu = \pm 1$ $(\mu = 1, \ldots, p)$ in the synaptic matrix w_{xy} in order to enable the network to reproduce the pattern associated to a presented initial (noisy) pattern. Thus the stored patterns should be attractors of dynamic (1). In the Hopfield model [1] a particular simple learning rule is given by the Hebb rule [2]:

$$w_{xy}^H = \frac{1}{N} \sum_{\mu=1}^{p} (\xi_x^\mu \xi_y^\mu - \delta_{xy}). \tag{2}$$

It is motivated by Hebb's postulate about learning [3].

Symmetry of the synaptic matrix ensures the existence of an energy functional which increases in time if only one neuron is updated per time step according to (1) (asynchronous dynamics). Since the energy functional is bounded from below, convergence of (1) will be guaranteed and fixed points are local minima of the energy functional.

For patterns whose spins are set independently of μ **and** x with equal probability to ± 1 (we call them random patterns) the Hopfield model can be solved by methods known from spin glasses [4,5]. By defining the parameter $\alpha = p/N$ (the number of patterns per neuron) the critical value of the storage capacity is calculated to $\alpha_c \simeq 0.144$. If α exceeds α_c, the fixed points of (1) will have no similarity with any of the stored pattern.

Because of the statistical independence of different spins inside the random patterns the pattern space cannot have any spatially topology while realistic patterns possess spatial structure. In the present paper we want to investigate the storage capacity of explicit learning rules for more realistic spatially correlated patterns. The spatial correlations are introduced statistically in order to get quantitative statements about the storage capacity when explicit learning rules are used. This concept of spatial correlations in pattern space is of current interest in physics of neural networks [6,7].

As an example we use the d-dimensional Ising model as a generator of spatially correlated patterns. This means that independent equilibrium configurations of the Ising model at temperature $T = 1/\beta$ are stored as correlated patterns. The next-neighbor interaction in the Ising model induces spatial correlation. Its amount is controlled by the inverse temperature or coupling strength β and can be described by the correlation matrix or two-point function (Sec. II). For inverse pattern temperature $\beta = 0$ the usual random patterns are recovered and the correlation between different spins of a pattern increases with increasing β. For dimension $d > 1$ there exists a finite critical β_c where patterns with correlations over all length scales are present. Our analytical investigations are valid for $\beta < \beta_c$. In Sec. II we give some properties of the Ising equilibrium configurations used as correlated patterns.

In Sec. III first we will see that the storage capacity $\alpha_c(\beta)$ decreases with increasing β when the Hebb rule (2) is used to store these patterns. This can be heuristically understood by the fact that for distances $|x - y| < l(\beta)$ [$l(\beta)$ is the correlation length] the Hebb rule (2) measures for large N and p the equilibrium correlation function of the Ising model. It bears no information about particular patterns, which suggests a decrease of $\alpha_c(\beta)$ with increasing $l(\beta)$.

The storage behavior can be improved when a linear

1063-651X/94/49(2)/1690(9)/$06.00

transformation on the spin variables decorrelating the patterns is applied. If the symmetric Hebb rule (2) is formulated for the decorrelated patterns the transformation back to the spin variables leads to an alternative asymmetric learning rule. Below the critical region its storage capacity is almost equal to that of random patterns. We call this alternative learning rule "pseudosymmetric" because it corresponds to the symmetric Hebb rule for the decorrelated patterns.

Of course, the capacity is still far from the optimal storage capacity ($\alpha_c = 2$ for random patterns) [8,9] but learning rules reaching larger storage capacities have the form of iterative algorithms [9,10–12] resulting in a tedious learning process. In contrast to this the pseudosymmetric learning rule is explicit with negligible learning time in comparison to iterative learning rules. Besides, properties of the storage mechanism can be understood comparatively easily because the synaptic matrix is given explicitly.

In Sec. IV the theoretical predictions for $\alpha_c(\beta)$ are compared with numerical simulations for both the original Hebb rule and its pseudosymmetric modification derived in Sec. III. Section V is devoted to the discussion of how a network can implement this new learning rule.

II. PROPERTIES OF THE PATTERNS

The patterns we want to store are statistically independent. Thus their probability distribution factorizes:

$$P(\xi^\mu, \xi^\nu) = P(\xi^\mu)P(\xi^\nu) \quad \text{for } \mu \neq \nu. \tag{3}$$

The generalization of our pattern statistic consists in the fact that single spins inside the *same* pattern are correlated. For simplicity we use equilibrium configurations of a d-dimensional Ising model with periodic boundary conditions as spatially correlated patterns. For the present we keep the inverse temperature β well below the critical point β_c. The probability $P(\xi)$ of a configuration to occur is given by the Gibbs distribution

$$P(\xi) = \frac{1}{Z} e^{-\beta E(\xi)}, \tag{4}$$

with

$$Z = \sum_{\{\xi\}} e^{-\beta E(\xi)} \tag{5}$$

and

$$E(\xi) = \tfrac{1}{2} \sum_{x,y} \xi_x \xi_y \delta_{|x-y|,1} , \tag{6}$$

where $\delta_{|x-y|,1}$ restricts the summation to next-neighbor pairs. The patterns can be characterized by the thermal averages of $\xi_{x_1} \cdots \xi_{x_n}$ (n-point functions). The correlation matrix or two-point function

$$g_{xy} = \langle \xi_x \xi_y \rangle = \sum_{\{\xi\}} \xi_x \xi_y P(\xi) \tag{7}$$

depends only on the distance $|x - y|$ and becomes exponentially small for large $|x - y|$:

$$g_{xy} \sim e^{-|x-y|/l(\beta)}, \quad \text{with } |x - y| \gg l(\beta) , \tag{8}$$

where $l(\beta)$ denotes the correlation length.

Because of translational invariance g can be inverted by Fourier transformation. While the off-diagonal elements of g are in general different from zero the inverse correlation matrix (called proper vertex function) is to good approximation a tridiagonal matrix:

$$g^{-1}_{xy} = A(t)\delta_{xy} - B(t)\delta_{|x-y|,1}, \tag{9}$$

with

$$A(t) \geq 1\,, \; B(t) \geq 0\,, \; t = \tanh\beta\,.$$

This is the "lattice version" of a Mexican hat. If one introduces the lattice Laplacian

$$\Delta_{xy} = \delta_{|x-y|,1} - 2d\delta_{xy}, \tag{10}$$

g^{-1} reads as

$$g^{-1}_{xy} = \frac{1}{\chi(t)} \left[\delta_{xy} - C(t)\Delta_{xy}\right] . \tag{11}$$

$\chi = \frac{1}{N} \sum_{xy} g_{xy}$ is the equilibrium susceptibility and C a positive function of t. (9) is exact for $d = 1$ with

$$A = \frac{1+t^2}{1-t^2}\,, \quad B = \frac{t}{1-t^2} \tag{12}$$

if finite-size effects are neglected ($t^N \simeq 0$). For $d = 2$ we checked numerically that g^{-1} has the form (9) below the critical region ($\beta < \beta_c \simeq 0.44$). For our purpose a high-temperature expansion up to $O(t^5)$ is sufficient. It reproduces (9) with

$$A = 1 + 4t^2 + 12t^4\,, \quad B = t(1+t^2)\,. \tag{13}$$

The computation of the storage capacity requires knowledge about the four-point function $g^{(4)}_{x_1 \ldots x_4}$ describing higher correlations. Separating $g^{(4)}$ into a connected and a disconnected part

$$g^{(4)}_{x_1 \ldots x_4} = g^{c(4)}_{x_1 \ldots x_4} + g_{x_1x_2}g_{x_3x_4} + g_{x_1x_3}g_{x_2x_4} + g_{x_1x_4}g_{x_2x_3} \tag{14}$$

the connected part $g^{c(4)}$ becomes exponentially small outside the critical region for any large separation of the arguments x_i ($i = 1, \ldots, 4$).

An important property of the Ising patterns is self-averaging for large N. Let S_{xy} be a matrix depending only on the distance $|x - y|$ with exponentially small elements for $|x - y| \gg l(\beta)$. Then the formula

$$\sum_{x,y} \xi_x \xi_y S_{xy} \simeq \mathrm{Tr}(gS) \tag{15}$$

holds for $l^d(\beta) \ll N$. (15) can easily be proved by dividing the lattice in p *independent* sublattices of size $[\lambda l(\beta)]^d$

with $\lambda > 1$. Then the left-hand side in (15) becomes a measurement of the observable appearing on the right-hand side. The relative error made in (15) is of order $p^{-1/2} = \{[\lambda l(\beta)]^d/N\}^{1/2}$ which becomes small for large N.

The storage capacity of the Hebb rule will involve

$$\chi_{SG} = \frac{1}{N}\mathrm{Tr}(g^2) . \tag{16}$$

In case $d = 1$ neglecting finite-size corrections

$$\chi_{SG} = \frac{1+t^2}{1-t^2} \tag{17}$$

and the high-temperature expansion of $1/\chi_{SG}$ for $d = 2$ reads as

$$\frac{1}{\chi_{SG}} = 1 - 4(t^2 + 5t^4 - 6t^6) + O(t^8). \tag{18}$$

χ_{SG} corresponds to the Edwards-Anderson spin glass susceptibility investigated in spin glass models [13].

Our formalism can be generalized to other pattern statistics, if the two-point function satisfies (8) and n-point functions have the cluster property expressed by (14). Self-averaging (15) is a consequence of (8) and (14). The modification of the Hebb rule which we work out in the following section is no longer a simple one as for our patterns if the inverse of g does not satisfy (9).

III. THE PSEUDOSYMMETRIC LEARNING RULE

If we define the postsynaptic potential

$$h_x = \sum_{y=1}^{N} w_{xy}s_y, \quad \text{with } w_{xx} = 0 \tag{19}$$

of neuron x we immediately observe for the Hebb rule (2) that the postsynaptic potential of neuron x can be written as

$$h_x(s) = \sum_{\mu} \xi_x^\mu m_\mu(s) - \alpha s_x , \tag{20}$$

where we have introduced the overlap with pattern ξ^μ:

$$m_\mu(s) = \frac{1}{N}\xi^{\mu T} s = \sum_x \xi_x^\mu s_x . \tag{21}$$

Now we assume that the network state s is close to one pattern, say ξ^ν, and the overlaps can be approximated by

$$m_\mu(s) \simeq \begin{cases} 1, & \mu = \nu \\ m_{\mu\nu}, & \mu \neq \nu \end{cases} . \tag{22}$$

$m_{\mu\nu}$ is the overlap between pattern μ and ν. Now we can separate h_x into a signal and a noise term:

$$h_x^\nu(s) \simeq \xi_x^\nu + \sum_{\substack{\mu \\ \mu\neq\nu}} \xi_x^\mu m_{\mu\nu} - \alpha s_x . \tag{23}$$

The average of $m_{\mu\nu}$ over our pattern statistic is zero for different patterns. If the patterns are attractors of the dynamic the application of the nonlinear "sgn" function will suppress the noise term. Thus we see that the performance of the associative memory is satisfactory if the variance of $m_{\mu\nu}$ becomes small. So we expect for the storage capacity

$$\frac{\alpha_c^H(\beta)}{\alpha_c(0)} \simeq \frac{1}{N\langle m_{\mu\nu}^2(\beta)\rangle} . \tag{24}$$

Carrying out the average over the pattern statistic the Hebb rule's storage capacity becomes

$$\frac{\alpha_c^H(\beta)}{\alpha_c(0)} \simeq \frac{1}{\chi_{SG}(\beta)} . \tag{25}$$

A more detailed derivation of this formula can be found in the Appendix. Since the spin glass susceptibility χ_{SG} increases strongly with β capacity drops from the known value $\alpha_c(0) = 0.144$ for random patterns to zero, if β reaches the critical point ($\beta = \infty$ in $d = 1$).

Now we want to improve the storage behavior by a modification of the Hebb rule based on a decorrelation of the Ising patterns ξ. The decorrelation can be achieved by the following transformation:

$$\hat{\xi} \stackrel{\text{def}}{=} g^{-1/2}\xi \quad \text{with } g^{-1/2}g^{-1/2} = g^{-1} . \tag{26}$$

The transformed patterns have no two-point correlation

$$\left\langle \hat{\xi}_x \hat{\xi}_y \right\rangle = (g^{-1/2} g g^{-1/2})_{xy} = \delta_{xy} \tag{27}$$

and they are normalized because of the self-averaging property (15):

$$\frac{1}{N}\hat{\xi}^T\hat{\xi} = \frac{1}{N}(g^{-1/2}\xi)^T(g^{-1/2}\xi) = \frac{1}{N}\xi^T(g^{-1}\xi) \simeq \frac{1}{N}\mathrm{Tr}(g^{-1}g) = 1 . \tag{28}$$

The overlap for the decorrelated patterns

$$\hat{m}_{\mu\nu} = \frac{1}{N}\hat{\xi}^{\mu T}\hat{\xi}^\nu \tag{29}$$

has zero mean for different patterns and its variance is independent of β and equal to that of random patterns:

$$\langle \hat{m}_{\mu\nu}^2 \rangle = \frac{1}{N} \quad \text{for } \mu \neq \nu . \tag{30}$$

So the decorrelated patterns have the same properties as random patterns if only two-point correlations are considered. This suggests that the *decorrelated* patterns should be stored by the *symmetric* Hebb rule:

$$\hat{w}_{xy}^H = \frac{1}{N}\sum_\mu \hat{\xi}_x^\mu \hat{\xi}_y^\mu - \alpha\delta_{xy} . \tag{31}$$

It stores the patterns as well separated local minima (different patterns possess small overlaps) of an existing energy functional. The dynamical variables are transformed in analogy to the pattern transformation:

$$\hat{s}_x = \sum_y g_{xy}^{-1/2} s_y \ . \tag{32}$$

Unfortunately a dynamic in this transformed configuration space which finds the local minima of the energy functional is difficult to formulate because neighboring configurations cannot be described as simply as for spin configurations (neighboring configurations simply differ in one spin flip). But if we define the postsynaptic potential in analogy to (19),

$$\hat{h}_x(\hat{s}) = \sum_y \hat{w}_{xy}^H \hat{s}_y, \tag{33}$$

and assume that the network state is close to $\hat{\xi}^\nu$ we can separate a signal term as in (23) by taking normalization of the patterns (28) into account:

$$\hat{h}_x^\nu(\hat{s}) \simeq \hat{\xi}_x^\nu + \sum_{\substack{\mu \\ \mu\neq\nu}} \hat{\xi}_x^\mu \hat{m}_{\mu\nu} - \alpha \hat{s}_x \ . \tag{34}$$

Now it is obvious that a proper dynamic should suppress the noise term. If we transform the signal term back to an Ising spin

$$\sum_y g_{xy}^{1/2} \hat{h}_y^\nu(\hat{s}) \simeq \xi_x^\nu + \sum_{\substack{\mu \\ \mu\neq\nu}} \xi_x^\mu \hat{m}_{\mu\nu} - \alpha s_x \tag{35}$$

one immediately observes that the "sign" function applied on $\sum_y g_{xy}^{1/2} \hat{h}_y$ is a dynamic in the *spin* variables with the desired property:

$$s_x^{(n+1)} = \text{sgn}\left(\sum_y g_{xy}^{1/2} \hat{h}_y (g^{-1/2} s^{(n)}) \right) \ .$$

Inserting (33) we obtain the pseudosymmetric learning rule:

$$s_x^{(n+1)} = \text{sgn}\left(\sum_y w_{xy}^{\text{PS}} s_y^{(n)} \right), \tag{36}$$

with

$$w^{\text{PS}} = g^{1/2} \hat{w}^H g^{-1/2}$$

and, more explicitly,

$$w_{xy}^{\text{PS}} = \frac{1}{N} \sum_\mu \xi_x^\mu \sum_{y'} g_{yy'}^{-1} \xi_{y'}^\mu - \alpha \delta_{xy} \ . \tag{37}$$

We call the learning rule "pseudosymmetric" because it results from a nonorthogonal transformation of the symmetric Hebb rule formulated in the decorrelated pattern space.

The right hand side of (35) is formally the same as that of (23) with $m_{\mu\nu}$ replaced by $\hat{m}_{\mu\nu}$. Thus by following the argumentation leading to the Hebb rule's storage capacity we expect for the pseudosymmetric learning rule

$$\frac{\alpha_c^{\text{PS}}(\beta)}{\alpha_c(0)} \simeq \frac{1}{N \langle \hat{m}_{\mu\nu}^2(\beta) \rangle} \ . \tag{38}$$

Because of (30) the pseudosymmetric learning rule's storage capacity should be β independent (see also Appendix):

$$\alpha_c^{\text{PS}}(\beta) \simeq \alpha_c(0), \tag{39}$$

which means a considerable improvement of the storage capacity. (39) only holds below the critical region since the decorrelation only considers the two-point function. The last term in (37) shall suppress self-excitations ($w_{xx} \simeq 0$). Numerical simulations have proven that it makes no difference on the storage capacity whether one uses (37) or prohibits self-excitations rigorously ($w_{xx} = 0$). However, numerical simulations are made without any self-excitations but calculations in the Appendix will be more comfortable if (37) is used.

(37) is a generalization of a learning rule found in [6] for feedforward perceptrons by modifying the Hebb rule. Here the pseudosymmetric learning rule is derived for fully connected associative neural nets with a decorrelation formalism which directly suggests an improvement of the storage capacity.

Our alternative learning rule is asymmetric and therefore convergence of dynamic (1) is not guaranteed. In fact, non-convergence occurred in some cases in our numerical simulations. Nevertheless investigation of the *Hamming distance's* time development

$$d_H(\xi^\nu, s^{(n)}) = \frac{1}{2N} \sum_x (1 - \xi_x^\nu s_x^{(n)}), \ n = 0, 1, 2, 3, \ldots$$

[$\xi^\nu = s^{(0)}$ is the inital configuration (see Sec. IV)] shows that the Hamming distance of nonconvergent sequences varies very *slightly* and *periodically* after a certain number of iterations. This indicates that the network's state cycles in configuration space with neglectable "radius." This sharp localization of the network's state in configuration space justifies the breakoff after a finite number of iterations.

Because of the structure of g^{-1} [see (9)] summation over y' contains only the $y = y'$ term and next-neighbor terms. Therefore the new learning rule is "almost" local and of the form

$$w_{xy}^{\text{PS}} = \frac{1}{N} \sum_\mu \left(A \xi_x^\mu \xi_y^\mu - B \xi_x^\mu \sum_{y'} \delta_{|y-y'|,1} \xi_{y'}^\mu \right) - \alpha \delta_{xy}, \tag{40}$$

with

$$A \geq 1, \ B \geq 0 \ .$$

Since the y' summation extends only over next neighbors of y, (40) shows the remarkable effect that the original Hebb rule (2) which is represented by the first term on the right hand side of (40) is modified in such a way that the neighboring neurons of y influence the synaptic strength w_{xy}^{PS} between postsynaptic neuron x and presynaptic neuron y *inhibitorily*. Thus we recover the inhibitory influence on neighboring neurons also observed in biological neural nets [14].

Several learning rules have been developed which are able to store correlated patterns and reach high storage capacities. Nevertheless they have the form of iterative algorithms [9,10–12] or they include a matrix inversion for every *particular* set of patterns to be learned (pseudoinverse method [15–17]). So for *every set* of new presented patterns one has to go through a tedious learning process. In our *explicit* learning rule the matrix g^{-1} has to be calculated only *once* for a *specific pattern statistic* and then every set of correlated patterns generated with the same statistic can be stored with very little effort.

Since the nontrivial part of g^{-1} is the lattice Laplacian [see (11)] the learning rule (37) implements *a priori* knowledge about dimension and number of next neighbors. The network also has to know the pattern temperature. In Sec. V, however, we will show a simple mechanism for how the network can learn the temperature.

IV. NUMERICAL SIMULATIONS

In this Secon we present our numerical simulations of the storage capacities $\alpha_c(\beta)$ using the Hebb rule (2) and the pseudosymmetric learning rule (37) to store correlated patterns generated at temperature $1/\beta$ by a Monte Carlo algorithm. In our simulations the network size is $N = 900$. Since only the ratio $\alpha_c(\beta)/\alpha_c(0)$ needs to be determined, results are rather insensitive to details of the criterion $\alpha_c(\beta)$ is determined from. Following [4] one starts dynamic (1) from a pattern ξ^ν. Dynamic (1) is iterated until convergence is reached. The final fixed point ω is compared with the initial configuration ξ^ν. If the number of wrong bits does not exceed $f_{\max}N$ with $f_{\max} = 0.025$, a "success counter" is incremented. The average success $\rho_0(\alpha)$ is determined by repeating this procedure T times ($T \simeq 400$).

$$\rho_0(\alpha) = [\Theta(f_{\max} - d_H(\xi^\nu, \omega))]_\xi , \qquad (41)$$

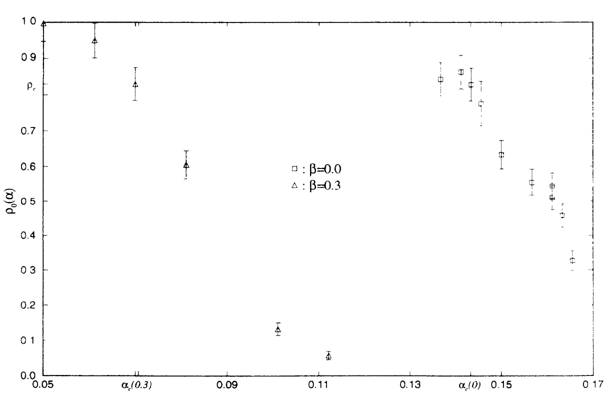

FIG. 1. Average success $\rho_0(\alpha)$ for random patterns ($\beta = 0.0$) and $d = 2$ Ising patterns at inverse temperature $\beta = 0.3$. The patterns are stored with the Hebb rule (2).

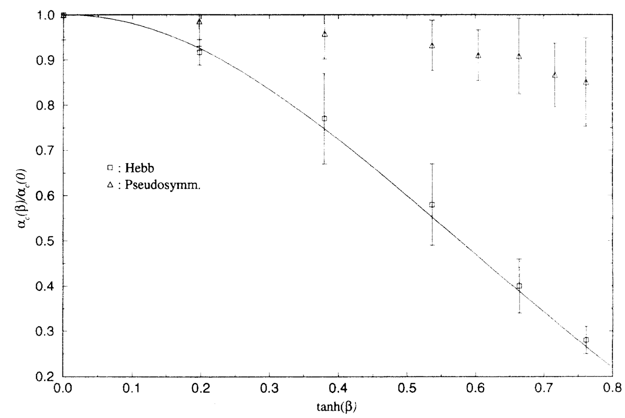

FIG. 2. Normalized storage capacities $\alpha_c(\beta)/\alpha_c(0)$ for $d = 1$ Ising patterns stored with the Hebb rule (2) and with the pseudosymmetric learning rule (37). The solid line represents $1/\chi_{SG}$ from Eq. (17) confirming the theoretical prediction (25).

with

$$d_H(\xi^\nu, \omega) = \frac{1}{2N} \sum_x (1 - \xi_x^\nu \omega_x) .$$

Θ is the Heaviside function and $[\,]_\xi$ means average over T patterns. In the region of α_c, $\rho_0(\alpha)$ decreases drastically. This sharp transition defines $\alpha_c(\beta)$ sufficiently accurately, if one chooses $\alpha_c(\beta) = \rho_0^{-1}(\rho_c)$ with $\rho_c = 0.83$. In Fig. 1 we show an example for $d = 2$ Ising patterns at $\beta = 0$ and $\beta = 0.3$ stored by the Hebb rule. $\alpha_c(0) = 0.143(8)$ is obtained in this way, which agrees with results in [4,5,18].

Figure 2 shows $\alpha_c(\beta)/\alpha_c(0)$ for $d = 1$ as a function of $\tanh\beta$. Using the Hebb rule (squares) α_c decreases in agreement with the theoretical prediction (25) (solid line). The pseudosymmetric learning rule (triangles)

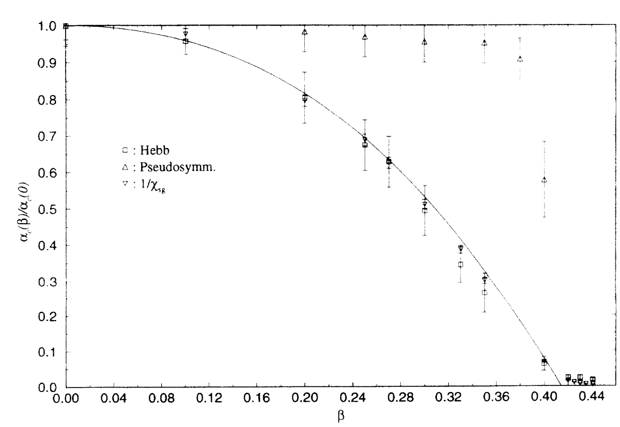

FIG. 3. Normalized storage capacities $\alpha_c(\beta)/\alpha_c(0)$ for $d = 2$ Ising patterns stored with the Hebb rule (2) and with the pseudosymmetric learning rule (37). Here the high-temperature expansion (18) (solid line) and Monte Carlo simulations (inverse triangles) of $1/\chi_{SG}$ in $d = 2$ confirm the theoretical prediction (25).

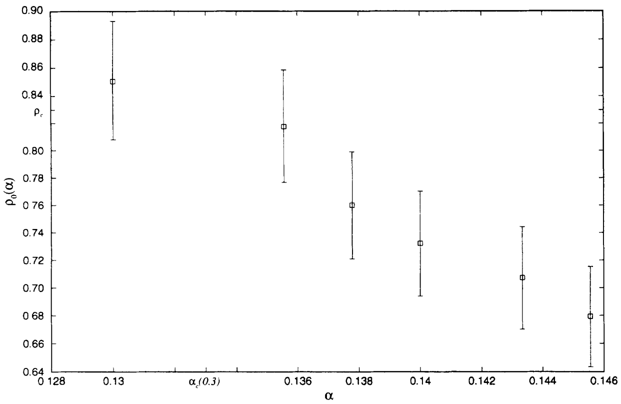

FIG. 4. Average success $\rho_0(\alpha)$ when $d = 2$ Ising patterns at inverse temperature $\beta = 0.3$ are stored with the pseudosymmetric learning rule. A storage capacity of $\alpha_c(0.3) = 0.133(8)$ is obtained if the noise level is chosen as $r = 5\%$.

leads obviously to a much better behavior of the storage capacity. The slight decrease with increasing $\tanh\beta$ indicates that the pseudosymmetric learning rule (37) includes only the two-point vertex function. The same is shown for $d = 2$ in Fig. 3. The solid line is the high-temperature expansion of $1/\chi_{\rm SG}$. Supplementary values of $1/\chi_{\rm SG}$ obtained by Monte Carlo simulations analogous to those used in the $\pm J$ Ising spin glass model [19] are shown (inverse triangles). Here $\alpha_c^{\rm PS}(\beta) \simeq \alpha_c(0)$ until the critical point at $\beta_c \simeq 0.44$ is reached.

Figure 4 shows that a neural network really works as an associative memory when its synaptic matrix is established by the pseudosymmetric learning rule. We stored $d = 2$ Ising patterns at inverse temperature $\beta = 0.3$ and started dynamic (1) with an initial pattern ζ which is a noisy variant of a stored pattern, say ξ^ν. The noise level $r = d_H(\xi^\nu, \zeta)$ should exceed $f_{\rm max}$. So we choose $r = 0.05 > f_{\rm max}$. With the criterion described above we obtain for the storage capacity $\alpha_c(0.3) = 0.133(8)$ demonstrating the net's robustness against noise.

In addition to the neglected four-point function the global magnetization of the patterns causes the bad performance of the associative memory for $\beta \geq \beta_c$. The nonvanishing magnetization is responsible for $\alpha_c(\beta) = 0$ in this region as easily can be seen by a signal-to-noise analysis (see Appendix, [20]). To remove disturbing influence of the magnetization we apply a learning rule similar to that suggested in [21] which is able to store patterns where the spins are set independently but with asymmetric probability for their orientation. Therefore the patterns possess by definition a nonvanishing magnetization. We modify the usual Hebb rule in the following way:

$$\tilde{w}_{xy}^H = \frac{1}{N}\left[1 + \sum_{\mu=1}^{p} \frac{1}{1-\bar{m}_\mu^2}\delta\xi_x^\mu \delta\xi_y^\mu\right] \quad , \tag{42}$$

with

$$\bar{m}_\mu \stackrel{\rm def}{=} \frac{1}{N}\sum_x \xi_x^\mu ,$$

$$\delta\xi_x^\mu \stackrel{\rm def}{=} \xi_x^\mu - \bar{m}_\mu \, .$$

If magnetization $\bar{m}_\mu$ is measured for a particular pattern ξ^μ, the network does not need to know pattern temperature *a priori*. Besides, there is no symmetry breaking in a finite lattice and by measuring the magnetization we are sure that the subtracted magnetization in $\delta\xi_x^\mu$ has the right sign. Figure 5 shows the storage capacity of the modified Hebb rule (42). As expected in the high-temperature phase storage capacity is equal to the Hebb rule's storage capacity but for $\beta > \beta_c$ an improvement is obvious. The storage capacity of (42) reaches its minimum at the *pseudocritical* point $\beta_c(N = 900) \simeq 0.42$.

An analogous modification of the *pseudosymmetric learning rule* (37) for patterns with nonvanishing magnetization leads to the following prescription:

$$\tilde{w}_{xy}^{\rm PS} = \frac{1}{N}\left[1 + \sum_{\mu=1}^{p} \delta\xi_x^\mu \sum_{y'} {g^c}^{-1}{}_{yy'}\delta\xi_y^\mu\right] . \tag{43}$$

${g^c}^{-1}$ is the inverse matrix of g^c whose elements are given by $g_{yy'}^c = g_{yy'} - m^2$ where m is the equilibrium magnetization per spin.

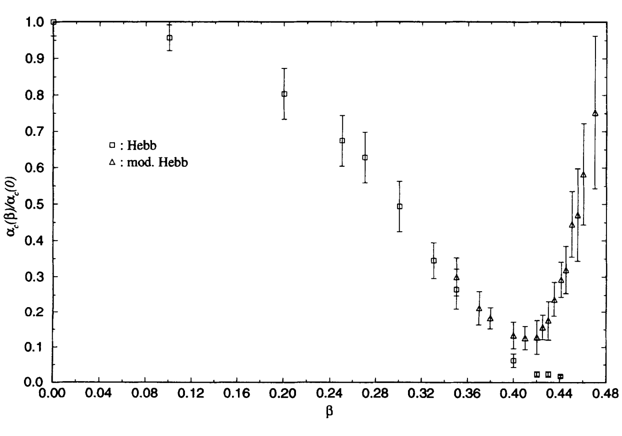

FIG. 5. Normalized storage capacities $\alpha_c(\beta)/\alpha_c(0)$ for $d = 2$ Ising patterns stored with the Hebb rule (2) and with the modified Hebb rule (42).

V. FURTHER DISCUSSION ON THE PSEUDOSYMMETRIC LEARNING RULE

Pattern correlation enters into the pseudosymmetric learning rule via the inverse correlation matrix g^{-1} of the Ising model used. Since it has to be known prior to the learning phase, the question remains of how a network can learn g^{-1}. First we show that the network has to know only topology of the pattern space but *not* pattern temperature. In the second place we demonstrate with a special pattern how the pseudosymmetric learning rule works.

Since the form of g^{-1} is known, only the actual value of β has to be learned. This can be achieved by adding a new system of neurons counting the total activity defined as

$$a = \frac{1}{N} \sum_{x,y} \xi_x^\mu \xi_y^\mu g_{xy}^{-1}(\beta_0) \ . \tag{44}$$

Self-averaging leads to

$$a = \frac{1}{N} \mathrm{Tr} \left[g(\beta) g^{-1}(\beta_0) \right] \ . \tag{45}$$

If activity a is measured by (44) and β_0 is the present value, the unknown inverse pattern temperature β is fixed by (45). The new system of neurons has to contact the other pattern-recognizing system closely in order to induce the learned temperature (more generally the amount of correlation) into the pseudosymmetric learning rule.

The matrix $g^{-1}(\beta_0)$ in (44) is to good approximation given by (9) and all nondiagonal elements vanish except next-neighbor indices. Therefore the new system consists of neurons with short range nerve fibers and establishes a next-neighbor relationship or a kind of "metric structure" in the neural network. Furthermore, since the amount of correlation does not change rapidly in time, the new system of neurons needs only information local in time (one pattern is sufficient to learn the temperature). In contrast to that, the other system of neurons learning *particular* patterns needs information nonlocal in space and time (particular patterns change rapidly). These neurons have long range nerve fibers. A similar situation is observed also in the cerebral cortex where two very different populations of neurons are in close contact to each other having the same properties as the two systems described above [22,23].

For $d = 1$, g^{-1} has the form

$$g_{xy}^{-1} = \frac{1}{\sqrt{1-u^2}} \left(\delta_{xy} - \frac{u}{2} \delta_{|x-y|,1} \right) , \tag{46}$$

with

$$u = \tanh(2\beta) \ .$$

It follows directly from (12) by applying addition theorems. Evaluating the trace in (45) by inserting (46) we obtain a relation between a and u:

$$u = \frac{2u_0 \left(1 - a\sqrt{1-u_0^2} \right)}{\left(1 - a\sqrt{1-u_0^2} \right)^2 + u_0^2} , \tag{47}$$

with

$$u_0 = \tanh(2\beta_0) \ .$$

Unless the network does not start from the extreme values $u_0 = 0$ or $u_0 = 1$ a new temperature can be learned when one pattern is presented. Thus only topology of the pattern space has to be implemented in the network *a priori*.

Now we choose an extreme pattern for a linear $d = 1$ network as an illustrative example to demonstrate the efficiency of the pseudosymmetric learning rule. Let ξ be a linear pattern divided in l chain pieces in such a way that each chain piece consists of spins with equal orientation. Furthermore, let $l \ll N$ in such a way that $O((l/N)^2)$ is neglected. Calculating (44) in $d = 1$ we obtain

$$a = \frac{1}{\sqrt{1-u_0^2}} \left[1 - \frac{u_0}{N} \sum_x \xi_x \xi_{x+1} \right]$$

and with $\sum_x \xi_x \xi_{x+1} = N - 2l$

$$a = \frac{1}{\sqrt{1-u_0^2}} \left[1 - u_0 \left(1 - 2\frac{l}{N} \right) \right] \ .$$

Inserting this into (47) u is obtained as a function of l/N only:

$$u = \frac{1 - 2l/N}{1 - 2l/N + 2(l/N)^2} = 1 + O((l/N)^2) \ .$$

So up to $O((l/N)^2) g^{-1}$ becomes proportional to the lattice Laplacian (10)

$$g_{xy}^{-1} \propto -\Delta_{xy} \ .$$

With $w_{xy}^H = \xi_x \xi_y$ (in the following we do not consider possible self-excitations for more convenience) this leads to

$$w_{xy}^{\mathrm{PS}} \propto w_{xy}^H \left[1 - \xi_y \left(\xi_{y+1} + \xi_{y-1} \right) \right] \tag{48}$$

and more explicitly

$$w_{xy}^{\mathrm{PS}} \propto \begin{cases} 0 & \text{for} \quad \xi_y = \xi_{y+1} = \xi_{y-1} \\ 2w_{xy}^H & \text{for} \quad \xi_y = -\xi_{y+1} = -\xi_{y-1} \\ w_{xy}^H & \text{otherwise} \ . \end{cases} \tag{49}$$

This formula demonstrates that no synaptic connection is necessary between presynaptic neuron y and postsynaptic neuron x, if the presynaptic neuron y is inside a chain piece (this means spin y *and* its neighbors have the same sign). Whereas the Hebb rule requires $O(N^2)$ synapses, the pseudosymmetric rule stores only interesting parts (the borders between different chain pieces) requiring $O(N)$ synapses. From the *a priori* knowledge about the patterns (topology and "amount of correlation") the network is able to recognize those parts of the pattern bearing important information about the *particular* pattern. The other parts (spins inside the chain pieces) are regarded as the uninteresting "normal case" which need not be stored. Thus the synaptic matrix is diluted in a way that information already given by the pattern statistic and being the same for all patterns will be suppressed when a pattern is learned. This dilution leads to a better storage behavior in comparison with the Hebb rule which suppresses important information about particular patterns by measuring uninteresting statistical properties (see Sec. I).

A similar mechanism is used in electronic data processing when a sequence of bits (0,1) is wanted to be stored. Only the *number* of zeros will be stored and not the zeros themselves, if the sequence consists almost of zeros. This is more efficient as long as the number of 0-1 changes is small compared with the total number of bits.

VI. CONCLUDING REMARKS

The conception of *statistically* correlated patterns in this paper can easily be generalized in such a way that other pattern statistics can be investigated. In the present work we use the Ising model as a generator of correlated patterns because numerical and analytical investigation is relatively comfortable. However, our developed concepts should also be applicable to classes of natural patterns consisting of many (may be *a priori* unknown) patterns with common (well known) properties. Then these common properties can be summarized by a pattern statistic $P_c(\xi)$ belonging to the special pattern class c. If the pattern class fulfills the demands mentioned in this paper (cluster properties), particular patterns should be stored satisfactorily with the pseudosymmetric learning rule.

Other interesting phenomena may be observed if neural networks with "critical" patterns (e.g., Ising patterns at $\beta = \beta_c$) are investigated. We believe that explicit learning rules which improve the storage of such patterns require new network architectures.

Recently we received a paper also dealing with spatially correlated patterns [27] in autoassociative memories. There the optimal storage capacity for weak spatial correlations in one dimension is investigated analytically. Furthermore, the author determines statistical properties of the synaptic matrix resulting from an *iterative learning algorithms* [12]. Our investigations and independently from [27] developed concepts for spatially correlated patterns lead to an *explicit* learning rule whose properties can be directly investigated as demonstrated in the present work. Maybe these different ideas can be combined in order to optimize storage mechanisms of spatially correlated patterns with respect to storage capacity and learning expense.

ACKNOWLEDGMENTS

We would like to thank Professor Dr. H. G. Schuster and Dirk Lukas for their interest and valuable discussions.

APPENDIX

In this appendix we give a derivation of (25) and (39). Of course we are not able to solve the introduced neural network models exactly for any given network temperature as made in [4] for random patterns, but since we are only interested in the quotient $\alpha_c(\beta)/\alpha_c(0)$ which compares storage capacities for different pattern statistics, the following calculation should be sufficient to describe the storage behavior.

The Hebb rule (2) and the pseudosymmetric learning rule (37) can be treated in parallel

$$w(k)_{xy} = \frac{1}{N}\sum_{\mu} \xi_x^{\mu} k(R)_{yy'} \xi_{y'}^{\mu} - \alpha\delta_{xy}, \tag{A1}$$

with

$$k(R) = \begin{cases} 1 & \text{for} \quad R = H \\ g^{-1} & \text{for} \quad R = P \end{cases} .$$

We use Einstein's summation convention except for the special index x and the pattern indices μ, ν. So for $R = H$ (A1) reduces to the Hebb rule and for $R = P$ (A1) becomes the pseudosymmetric learning rule. Now let the network configuration be ξ^{ν}. Then the local field acting on spin x is given by

$$h_x^{\nu} = \sum_y w(k)_{xy}\xi_y^{\nu} - \alpha\xi_x^{\nu} .$$

The storage capacity will be determined by investigating the probability that spin ξ_x^{ν} is stable:

$$P(m_x^{\nu} > 0), \quad \text{with} \quad m_x^{\nu} = \xi_x^{\nu} h_x^{\nu} .$$

m_x^{ν} can be written as

$$m_x^{\nu} = \sum_{\mu} X^{\mu} - \alpha, \tag{A2}$$

with

$$X^{\mu} = \frac{1}{N}\xi_x^{\mu}\xi_{y'}^{\mu} k(R)_{yy'}\xi_y^{\nu}\xi_x^{\nu} . \tag{A3}$$

We *assume* that m_x^{ν} is the sum of the *independent* random variables X^{μ} [24]. By the central limit theorem $(p \gg 1)$ m_x^{ν} is Gauss distributed with

$$\bar{m} \stackrel{\text{def}}{=} \langle m_x^{\nu}\rangle = \sum_{\mu} \langle X^{\mu}\rangle - \alpha, \tag{A4}$$

$$\sigma^2 \stackrel{\text{def}}{=} \Delta^2 m_x^{\nu} = \sum_{\mu} \Delta^2 X^{\mu}, \tag{A5}$$

with

$$\Delta^2 X^{\mu} = \left\langle X^{\mu 2}\right\rangle - \langle X^{\mu}\rangle^2 .$$

Now we have to calculate first and second momentum of X^{μ}. If one considers independence of the configurations ξ^{μ} and ξ^{ν} for $\mu \neq \nu$, the following formulas are derived easily.

$\mu \neq \nu$:

$$\langle X^{\mu}\rangle = \frac{1}{N^2}\mathrm{Tr}(gkg) = \frac{\tilde{\chi}(R)}{N},$$

$$\left\langle X^{\mu 2}\right\rangle = \frac{1}{N^2}\mathrm{Tr}(gkgk) = \frac{\tilde{\chi}(R)}{N}$$

$$\Rightarrow \Delta^2 X^{\mu} = \frac{\tilde{\chi}(R)}{N} - \frac{\tilde{\chi}(R)^2}{N^2},$$

with

$$\tilde{\chi}(R) = \begin{cases} \chi_{\mathrm{SG}} & \text{for} \quad R = H \\ 1 & \text{for} \quad R = P. \end{cases} \tag{A6}$$

$\mu = \nu$:

$$\langle X^\nu \rangle = \frac{1}{N}\mathrm{Tr}(gk) = 1,$$

$$\left\langle X^{\nu 2} \right\rangle = \begin{cases} 1 & \text{for} \quad R = H \\ \frac{1}{N^2} g_{yy'zz'} g^{-1}_{yy'} g^{-1}_{zz'} & \text{for} \quad R = P \end{cases}$$

$$\Rightarrow \sigma_1^2 \overset{\text{def}}{=} \Delta^2 X^\nu = \begin{cases} 0 & \text{for} \quad R = H \\ \frac{1}{N^2} g_{yy'zz'} g^{-1}_{yy'} g^{-1}_{zz'} - 1 & \text{for} \quad R = P. \end{cases}$$

Inserting this into (A4) and (A5) we obtain up to $O(1/N)$

$$\bar{m} = \langle m_x^\nu \rangle = 1 + \alpha[\tilde{\chi}(R) - 1] + O(1/N), \tag{A7}$$

$$\sigma^2 = \Delta^2 m_x^\nu = \sigma_1^2 + \alpha\tilde{\chi}(R) + O(1/N) \ . \tag{A8}$$

And the probability of $m_x^\nu > 0$ is given by

$$P(m_x^\nu > 0) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_0^\infty dm\, e^{-\frac{(m-\bar{m})^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \int_{-\gamma(\beta)}^\infty dz\, e^{-z^2/2} = f(\gamma(\beta)), \tag{A9}$$

with

$$\gamma(\beta) = \frac{\bar{m}}{\sigma} = \frac{1 + \alpha[\tilde{\chi}(R) - 1]}{\sqrt{\sigma_1^2 + \alpha\tilde{\chi}(R)}} \quad . \tag{A10}$$

For $\alpha < [\tilde{\chi}(R) - 1]^{-1}$ γ decreases monotonously with increasing α and the storage capacity $\alpha_c(\beta)$ is determined by the condition that the probability $P(m_x^\nu > 0)$ should be the same for *all temperatures*:

$$f(\gamma_c(\beta)) = f(\gamma_c(0)) \ .$$

γ_c means γ at the point $\alpha = \alpha_c$. The function f is monotonously increasing and therefore equality of the arguments follows:

$$\gamma_c(\beta) = \gamma_c(0) \ .$$

This equation determines $\alpha_c(\beta)$. It reads explicitly as

$$\frac{1 + \alpha_c(\beta)[\tilde{\chi}(R) - 1]}{\sqrt{\sigma_1^2 + \alpha_c(\beta)\tilde{\chi}(R)}} = \frac{1}{\sqrt{\alpha_c(0)}} \tag{A11}$$

leading for the Hebb rule to

$$\frac{[1 + \alpha_c^H(\beta)(\chi_{\mathrm{SG}} - 1)]^2}{\alpha_c^H(\beta)\chi_{\mathrm{SG}}} = \frac{1}{\alpha_c(0)} \quad . \tag{A12}$$

Since $\alpha_c(\beta)(\chi_{\mathrm{SG}} - 1) \ll 1$, (25) follows to good approximation:

$$\frac{\alpha_c(\beta)}{\alpha_c^H(0)} \simeq \frac{1}{\chi_{\mathrm{SG}}} \ .$$

For the pseudosymmetric learning rule we obtain from (A11)

$$\sigma_1^2 + \alpha_c^{\mathrm{PS}}(\beta) = \alpha_c(0) \ . \tag{A13}$$

The smallness of σ_1^2 can be seen by inserting (14) in the equation for σ_1^2 leading to

$$\sigma_1^2 = \frac{1}{N^2} g^c_{yy'zz'} g^{-1}_{yy'} g^{-1}_{zz'} + \frac{2}{N} \ .$$

Outside the critical region the term including g^c is small because of the cluster property of g^c. Thus $\sigma_1^2 \ll 1$ and the pseudosymmetric learning rule's storage capacity is almost independent of β:

$$\alpha_c^{\mathrm{PS}}(\beta) \simeq \alpha_c(0) \ .$$

For magnetized patterns ($\beta > \beta_c$, $\langle \xi^\mu \rangle = m$) α_c^H becomes zero because of $\chi_{\mathrm{SG}} > Nm^4 \gg 1$. Besides, σ_1^2 becomes comparable with $\alpha_c(0)$ and (A13) leads to $\alpha_c^{\mathrm{PS}} \simeq 0$ as for the Hebb rule. So above the critical point ($\beta > \beta_c$) the learning rules (2), (37) have to be modified according to (42), (43).

[1] J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).
[2] W. A. Little, Math. Biosci. **19**, 101 (1974); W. A. Little and G. L. Shaw, *ibid.* **39**, 281 (1978).
[3] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[4] D. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).
[5] R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh, IEEE Trans. Inf. Theory **33**, 461 (1987).
[6] R. Monasson, J. Phys. A **25**, 3701 (1992).
[7] R. Monasson, J. Phys. (Paris) I **3**, 1141 (1993).
[8] T. M. Cover, IEEE Trans. Electron Comput. **EC-14**, 326 (1965).
[9] E. Gardner, J. Phys. A **21**, 257 (1988).
[10] S. Diederich and M. Opper, Phys. Rev. Lett. **58**, 949 (1987).
[11] G. Pöppel and U. Krey, Europhys. Lett. **4**, 979 (1987).
[12] W. Krauth and M. Mézard, J. Phys. A **20**, L745 (1987).
[13] R. R. P. Singh and S. Chakravarty, Phys. Rev. B **36**, 546 (1987).
[14] D. L. Alkon, *Memory Traces in the Brain* (Cambridge University Press, Cambridge, England, 1988).
[15] L. Personnaz, I. Guyon, and G. Dreyfus, J. Phys. (Paris) Lett. **46**, 359 (1985).
[16] I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).
[17] T. Kohonen, IEEE Trans. Comput. **C-23**, 444 (1974).
[18] G. A. Kohring, J. Stat. Phys. **59**, 1077 (1990).
[19] J. S. Wang and R. H. Swendsen, Phys. Rev. B **38**, 4840 (1988).
[20] D. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).
[21] M. V. Feigel'man and L. B. Ioffe, Int. J. Mod. Phys. B **1**, 51 (1987).
[22] V. Braitenberg, in *Theoretical Approaches to Complex Systems*, edited by R. Heim and G. Palm (Springer, Berlin, 1978).
[23] R. Miller, Psychobiology **15**, 241 (1987).
[24] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, MA, 1991), p. 37.