Associative memory in an analog iterated-map neural network

C. M. Marcus, F. R. Waugh, and R. M. Westervelt

Division of Applied Sciences and Department of Physics, Harvard University, Cambridge, Massachusetts 02138

The behavior of an analog neural network with parallel dynamics is studied analytically and numerically for two associative-memory learning algorithms, the Hebb rule and the pseudoinverse rule. Phase diagrams in the parameter space of analog gain β and storage ratio α are presented. For both learning rules, the networks have large "recall" phases in which retrieval states exist and convergence to a fixed point is guaranteed by a global stability criterion. We also demonstrate numerically that using a reduced analog gain increases the probability of recall starting from a random initial state. This phenomenon is comparable to thermal annealing used to escape local minima but has the advantage of being deterministic, and therefore easily implemented in electronic hardware. Similarities and differences between analog neural networks and networks with two-state neurons at finite temperature are also discussed.

I. INTRODUCTION

The design of artificial neural networks is the inverse of a standard problem in modern nonlinear dynamics. Usually, the goal is to describe the set of attractors of a given nonlinear dynamical system such as a set of differential equations or an iterated map. In neural networks, the starting point is a set of desired attractors, and the problem is to find a dynamical system which possesses these attractors, and if possible, no other "spurious" attractors. Models of neural networks using two-state neurons are outside the realm of standard nonlinear dynamics, where analytical techniques frequently assume a continuous state space. On the other hand, networks of two-state neurons are well treated within spin-glass theory, and this avenue has recently lead to many important analytical results, particularly on the problem of associative memory.[1-3]

Using these new results to make fast computing devices requires *parallel* algorithms and architectures; sequential dynamics—where neuron states are updated one at a time—implemented in software on a conventional computer is simply too slow for any large neural network application. Unfortunately, parallel computation is plagued with stability problems not found in serial dynamics. Stability is also a central problem in the design of very large scale integrated (VLSI) electronics, especially when extensive feedback is present.[4-6]

In this paper we show that stability problems associated with parallel dynamics in associative-memory networks can be eliminated by using analog neurons. Specifically, we present phase diagrams for analog associative memories using the Hebb rule and the pseudoinverse rule which show that over a large range of neuron gain—defined as the maximum slope of the neuron transfer function—these networks can be updated in parallel while maintaining good recall and *guaranteed convergence to a fixed point*. This feature distinguishes analog networks from Ising-spin networks (with or without temperature) which must be updated serially to prevent oscillation. We will also discuss a second important advantage of analog associative memories, which is that lowering the neuron gain can greatly increase the chances that an initial state far from all memories will correctly flow to a recall state without getting trapped in a spurious attractor.

In comparing the deterministic dynamics of analog networks to the thermodynamics of Ising networks, there is a strong, though imperfect, analogy between temperature and analog gain. This analogy is strengthened by the frequent appearance of the hyperbolic tangent function both in mean-field statistical mechanics of the Ising model and in the circuit equations describing electronic and biological neural networks. The analogy between temperature and gain is more than just formal: the power of stochastic dynamics to "anneal" a neural network into a good (low-energy) solution is also seen in analog neural networks using smooth nonlinear neuron transfer functions and completely deterministic dynamics. The usefulness of analog annealing is supported by analytical and numerical results (to be reported elsewhere[7]) showing that the number of spurious attractors for a symmetric analog neural network can be greatly reduced by lowering the neuron gain. The use of analog annealing and the gain-temperature analogy has been discussed by Hopfield and Tank in the context of the traveling-salesman problem,[8] and similar ideas have been used in the field of artificial vision.[9]

The rest of the paper is organized as follows: In Sec. II we define the iterated-map neural network and show that for a broad range of transfer functions and symmetric connections, the only attractors are period-two limit cycles and fixed points. We also show that all limit cycles can be eliminated by lowering the neuron gain (maximum slope) below a critical value. Some of the results in this section were presented in Ref. 10. In Sec. III we investigate analog associative memories with the Hebb[11] and pseudoinverse[12,13] learning rules and present phase diagrams in the plane of neuron gain β and memory storage ratio α. In Sec. IV numerical results for the associative-

memory networks are presented. These results agree well with the analytical results of Sec. III. The numerical results in Sec. IV also show that the probability of retrieval is increased at low analog gain, suggesting the use of analog annealing to enhance recall. Finally, some applications of these results and conclusions are presented in Sec. V.

II. MODEL AND STABILITY CRITERION

A. The iterated-map network

The dynamical system we study is an iterated-map neural network in which all neurons have continuous input-output transfer functions and updating is done in parallel.[10] The network is defined by the set of coupled nonlinear equations,

$$x_i(t+1) = F_i\left[\sum_j T_{ij}x_j(t) + I_i\right], \tag{1}$$

where the real variables $x_i(t)$, $i = 1, \ldots, N$ describe the state of the system at time t. The interconnection matrix T_{ij} is assumed real and symmetric. The external biases I_i and the nonlinear neuron transfer functions F_i are also real valued and may be different for each i.

The continuous-time version of Eq. (1), given by

$$\frac{dx_i(t)}{dt} = -x_i(t) + F_i\left[\sum_j T_{ij}x_j(t) + I_i\right], \tag{2}$$

can be cast in the form of the electronic circuit equations described by Hopfield[14] by the change of variables $u_i(t)/R_i \equiv \sum_j T_{ij}x_j(t) + I_i$, $t' \equiv R_i C_i t$, $f_i(z) \equiv F_i(z/R_i)$ under the assumption that the time constants $R_i C_i$ are equal for all i. These substitutions give the following equations for the neuron input voltages $u_i(t')$:

$$C_i\frac{du_i(t')}{dt'} = -u_i(t')/R_i + \sum_j T_{ij}f_j(u_j(t')) + I_i. \tag{3}$$

In these circuit equations, R_i is interpreted as the total fan-in resistance at the input of neuron i and C_i is the neuron input capacitance.

B. Global stability criterion

In Ref. 10 we prove two properties of the iterated-map neural network, Eq. (1), for symmetric T_{ij} and all F_i monotonic single-valued functions which change in magnitude slower than linear at large argument. Under these conditions, (i) all attractors of Eq. (1) are either fixed points or period-two limit cycles, and (ii) all limit cycles can be eliminated, leaving only fixed-point attractors, by lowering all neuron gains β_i to satisfy the stability criterion

$$\frac{1}{\beta_i} > -\lambda_{\min} \quad \text{for all } i, \tag{4}$$

where $\beta_i (>0)$ is the maximum slope of F_i (the F_i are taken to be monotonically *increasing* without loss of generality) and $\lambda_{\min}$ is the minimum eigenvalue of the connection

matrix T_{ij}. When T_{ij} has negative eigenvalues, $\lambda_{\min}$ is the most negative eigenvalue. The requirements for the functions F_i are rather liberal, as illustrated in Fig. 1. Notice that the F_i do not need to be bounded at large argument, can be concave-up or concave-down at any finite argument, and can be a different function for each neuron. Figure 1 also shows the maximum slope β_i for a particular F_i.

$$L(t) = \frac{1}{2}\sum_{i,j}T_{ij}x_i(t)x_j(t) - \sum_i I_i x_i(t) + \sum_i G_i(x_i(t)), \tag{5a}$$

$$G_i(x_i) \equiv \int_0^{x_i} F_i^{-1}(z)dz, \tag{5b}$$

is a Liapunov function when the stability criterion is obeyed. That is, $L(t)$ decreases at each discrete time step and is bounded below and so must ultimately reach a minimum. The minima of $L(t)$ are at the fixed points of Eq. (1),

$$x_i^*(t) = x_i^*(t+1) = F_i\left[\sum_j T_{ij}x_j^*(t)\right], \quad i = 1, \ldots, N. \tag{6}$$

The function $L(t)$ is similar to the Liapunov functions for continuous-time (differential-equation) analog systems given by Hopfield[14] and, in more general form, by Cohen and Grossberg.[15] An important distinction is that continuous-time analog networks with symmetric con-
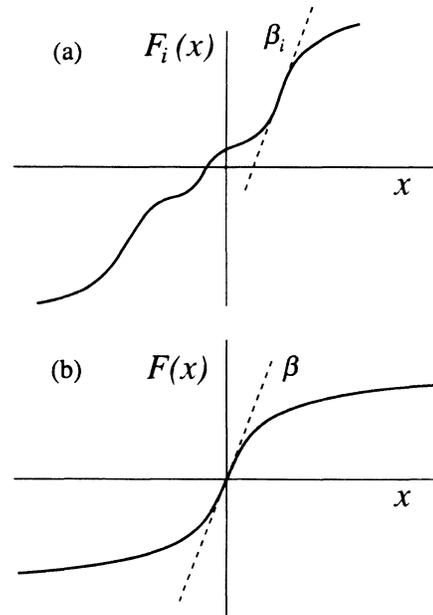
FIG. 1. (a) An example of a nonlinear neuron transfer function which meets the conditions for the dynamic properties given in Sec. II B. Those conditions are the following: Each function must be single valued and monotonic and must grow in magnitude slower than linear in the limit of large positive or negative argument. The maximum slope β_i which appears in the stability criterion (4) is also indicated. (b) An example of a nonlinear function F (identical for all i) which meets the less general conditions assumed for the associative-memory phase diagrams, Figs. 2 and 3. These conditions are given at the beginning of Sec. III.

nections will converge to a fixed point for any set of single-valued, monotonic nonlinearities, whereas the iterated-map network must also satisfy Eq. (4) to guarantee convergence.

C. Analog gain versus temperature and the TAP equations

For the particular choice of nonlinear transfer function $F_i(x) = \tanh(\beta x)$, the fixed points of (1) correspond to the solutions of the "naive" mean-field theory for the equilibrium magnetization of an Ising model with spin coupling matrix T_{ij} at temperature $1/\beta$. Naive refers to the fact that a proper thermodynamic treatment of the Ising model must also take into account the so-called reaction field due to the effect of a spin's own field on itself.[16,17] The mean-field theory for the Ising spin glass[18] (random symmetric T_{ij})—*including* the reaction field—was analyzed by Thouless *et al.*[17] (TAP), who showed that in equilibrium, the average magnetization x_i at site i is given by the solution to the set of equations

$$x_i = \tanh\left[\beta \left(h_i + \sum_j T_{ij}x_j - \beta x_i \sum_j T_{ij}^2(1-x_j^2)\right)\right], \quad (7)$$

where the third term in the large parentheses is the reaction field and h_i is a local externally applied magnetic field.

Attempts to solve the TAP equations by iteration have shown that convergence is rare for either parallel or serial updating, usually leading to a period-two limit cycle for parallel dynamics and a periodic or chaotic attractor for sequential dynamics.[19,20] Using better techniques than straightforward iteration to find TAP solutions is also problematic, because of the many shallow saddle points throughout the energy landscape.[21] Thus, while the TAP equations give the correct thermodynamic description of the mean-field Ising model, they are ill suited as an analog dynamical system on which to base a fast, stable analog neural network.

In the limit $\beta \to \infty$ ($T = 0$) the reaction field vanishes and the TAP equations agree with the naive mean-field theory.[17] However, in the $\beta \to \infty$ limit, T_{ij} must be positive definite to insure stability by Eq. (4). Unfortunately, this requirement is inconsistent with most currently used learning algorithms. For example, the usual practice of setting $T_{ii} = 0$ for all i yields matrices which cannot be positive definite.

Iteration of the naive mean-field equations for finding the ground state and other properties of spin glasses has been previously investigated by Soukoulis *et al.*[22] These authors found that the naive mean-field equations gave surprisingly good quantitative results, comparable to finite-temperature Monte Carlo techniques. A stability criterion based on local stability analysis was also presented for their iterative method.[23] Later, Reger *et al.*[24] argued that for spin glasses, the results obtained by iterating and annealing the naive mean-field equations are inferior to those found by Monte Carlo methods and are only useful for studying qualitative effects. However, the comparisons of Reger *et al.* were made specifically for the finite-range spin-glass problem and do not necessarily apply to neural networks. In general, using an ana-

log approach rather than a thermal approach for finding good solutions in a complicated landscape is better suited to neural networks than spin glasses. This is because there are many fewer metastable (spurious) states in a neural network than in a spin glass[7,25] and also because the basins of attraction for recall in a neural network are very large—much larger than the basins of the spurious states—and therefore should be robust to the distortions of the landscape caused by annealing (see Sec. IV B).

III. ANALOG ASSOCIATIVE MEMORY

We now apply the iterated-map neural network to the problem of associative memory. In this section we assume a less general form for the iterated-map network, where $I_i = 0$ for all i and the nonlinear functions F_i are odd, single-valued functions and identical for all i. We also assume the function F (dropping the index i) has its maximum slope at zero input $F'(0) = \beta$, and that the slope of F is a nonincreasing function of the magnitude of the argument. The maximum slope β will be referred to as the analog gain of the neurons. Possible forms for F include, but are not limited to, tanh-like functions. As in Sec. II B, we do not require that F saturate at large argument though it must increase in magnitude slower than linear at large positive or negative argument. Without loss of generality, we take F to be monotone increasing and normalized such that the size scale of the accessible state space is $O(1)$, that is, a nonzero solution of $m^* = F(m^*)$ is $O(1)$. Figure 1(b) shows a function which meets the conditions assumed in this section. Under these assumptions, the associative-memory network is given by the set of equations

$$x_i(t+1) = F\left[\sum_j T_{ij}x_j(t)\right], \quad i = 1, \ldots, N. \quad (8)$$

We consider connection matrices T_{ij} for two learning rules, the Hebb rule[11] and the pseudoinverse rule,[12,13] for random unbiased memory patterns $\xi_i^\mu = \pm 1$. For the Hebb rule,

$$T_{ij} = \frac{1}{N}\sum_{\mu=1}^{\alpha N}\xi_i^\mu\xi_j^\mu, \quad T_{ii} = 0 \quad (9)$$

where αN is the number of stored memory patterns. For the pseudoinverse rule,

$$T_{ij} = \frac{1}{N}\sum_{\mu,\nu=1}^{\alpha N}\xi_i^\mu(C^{-1})_{\mu\nu}\xi_j^\nu, \quad T_{ii} = 0 \quad (10a)$$

where C^{-1} is the inverse of the correlation matrix

$$C_{\mu\nu} = \frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu\xi_i^\nu. \quad (10b)$$

Notice that we are considering the modified pseudoinverse rule with $T_{ii} = 0$ studied by Kanter and Sompolinsky.[13] These authors showed that this modification increases the basins of attraction for the memories without sacrificing error-free recall. The analysis in the Secs. III A and III B assumes $\beta > 0$, $0 < \alpha < 1$, and $N \gg 1$.

A. Hebb rule

A phase diagram for the Hebb rule showing four distinct regions in the parameter space of analog gain β and the storage ratio α is shown in Fig. 2. The four regions are characterized as follows: In the region marked "origin" a fixed point at the origin, $x_i = 0$ for all i, is the unique global attractor. In the region marked "spin glass" the origin is no longer an attractor, but neither are the memory recall states. In this region the network converges to a fixed point with small $[O(N^{-1/2})]$ overlap with all memories. In the region marked "recall" fixed points having large overlaps with memory patterns exist and have large basins of attraction. In the recall region the iterated map works well as an associative memory. The boundary separating recall from spin glass is shown in Fig. 2 for the special choice $F(z) = \tanh(\beta z)$. For this choice of nonlinearity, this boundary agrees with the ferromagnetic transition curve found by Amit et al.[26] for the Ising model at finite temperature. However, the analysis leading to this curve, presented in Appendix A, is not restricted to case $F(z) = \tanh(\beta z)$. In the region marked "oscillation" the stability criterion (4) is no longer obeyed and convergence to a fixed point is not guaranteed. Numerically, we find that limit cycles are quite prevalent in this region, especially for larger values of β and α (see Sec. IV).

The stability of the origin can be determined by linearizing Eq. (8) about the point $x_i = 0$, which gives N decoupled linear iterated maps: $\phi_i(t+1) = \beta\lambda_i\phi_i(t)$ for evolution along the ith eigenvector of the matrix T_{ij}, with associated eigenvalue λ_i. For $|\beta\lambda_i| < 1$ for all i, the origin is stable, and because of the form of the function F, it is also the unique attractor of Eq. (8).[27] Notice that for $\lambda_i = \lambda_{min}$, this condition is identical to the global stability criterion (4).

The minimum and maximum eigenvalues for the Hebb matrix (9) with $\alpha < 1$ in the large-N limit are[28]

$$\lambda_{min} = -\alpha \quad [N(1-\alpha)\text{- fold degenerate}] \tag{11a}$$

$$\lambda_{max} = 1 + 2\sqrt{\alpha} \quad \text{(edge of continuous distribution)} . \tag{11b}$$

Thus for $\alpha < 1$ the boundary where the origin loses stability is defined by the condition $\beta = 1/(1+2\sqrt{\alpha})$. From the value of λ_{min}, we can also identify the border of the oscillatory region as $\beta = 1/\alpha$. Crossing the origin–spin-glass line corresponds to a forward pitchfork bifurcation of the origin, analogous to a second-order transition in thermodynamics. Note that this transition occurs along a different curve from the corresponding paramagnet–spin-glass transition in the Ising model associative memory.[26]

Crossing the border from the recall region into the spin-glass region marks the disappearance of a fixed point having a large overlap with a single memory. As in the case of the Ising model network, this transition is due to the random overlaps of the state of the network with patterns other than the one being recalled. These overlaps generate an effective noise source which destabilizes the fixed point near the recalled pattern. Because our system has no reaction field (by design), the analysis is somewhat simpler than the replica[18] or cavity[2,29] approaches used to analyze the thermodynamic Ising-model network. In Appendix A we derive a set of four self-consistent equations assuming random, unbiased memory patterns:

$$m^1 = \frac{1}{\sqrt{2\pi}} \int dy \, \exp(-y^2/2) F(\sigma y + m^1) , \tag{12a}$$

$$C = \frac{1}{\sqrt{2\pi}} \int dy \, \exp(-y^2/2) F'(\sigma y + m^1) , \tag{12b}$$

$$q = \frac{1}{\sqrt{2\pi}} \int dy \, \exp(-y^2/2) F^2(\sigma y + m^1) , \tag{12c}$$

$$\sigma = \frac{\sqrt{\alpha q}}{1 - C} , \tag{12d}$$

where $F'(z) \equiv dF(z)/dz$. The quantity m^1 in Eq. (12) is the overlap of the network state vector with a single memory pattern, arbitrarily chosen to be pattern 1. In the recall state, these equations have a self-consistent solution with $m^1 \sim 1$. For the particular choice $F(z) = \tanh(\beta z)$, the quantities C and q obey the usual relation $C = \beta(1-q)$.

B. Pseudoinverse rule

The pseudoinverse learning rule, Eq. (10), offers several advantages over the Hebb rule, chiefly a greater storage capacity, error-free recall states, and the ability to store correlated patterns.[12,13] Its primary disadvantage is that it is nonlocal, meaning that a given element of the connection matrix, T_{ij}, cannot be determined from the ith and jth elements of the memory patterns, but depends on all components of all memories. However, iterative learning algorithms have been described which are local and converge to the pseudoinverse rule.[30]

A phase diagram for the pseudoinverse rule showing three distinct regions depending on analog gain β and storage ratio α is shown in Fig. 3. The phase diagram
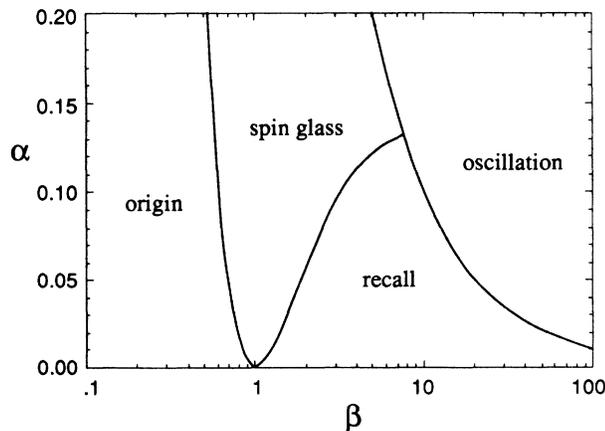
FIG. 2. Phase diagram for the Hebb rule associative memory with neuron transfer function $F(z) = \tanh(\beta z)$. The parameter β is the neuron gain, and α is the number of stored patterns divided by the number of neurons N. All borders separating the regions are based on analysis at large N, as described in the text.
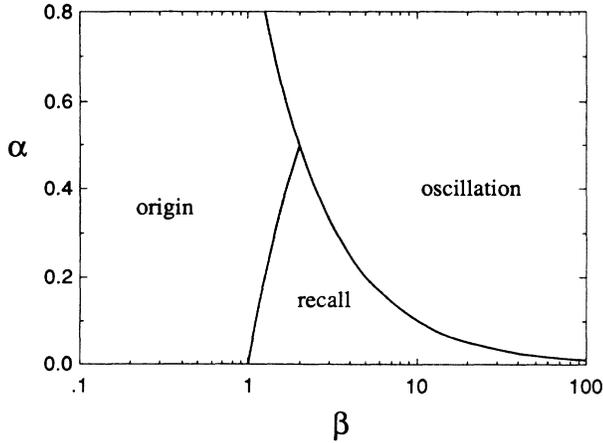
FIG. 3. Phase diagram for the pseudoinverse rule (diagonal elements equal zero) with sigmoidal neuron transfer function as described at the beginning of Sec. III. The parameter β is the neuron gain, and α is the number of stored patterns divided by the number of neurons N. All borders separating the regions are based on analysis at large N, as described in the text. Note that the pseudoinverse rule does not possess a spin-glass phase for $\alpha < 1$.

differs from that of the Hebb rule in three distinctive ways: First, there is no spin-glass phase. This does not imply that the pseudoinverse rule does not possess spurious attractors; just as for the Hebb rule, there are many spurious fixed-point attractors within the recall and oscillatory regions which have small overlap with all memories. Unlike the Hebb rule, however, there is no region of the pseudoinverse phase diagram where *only* spurious fixed-point attractors are found. The second difference is that the recall region is much larger, extending to $\alpha = 0.5$ for $\beta = 2$. Above this point, and for higher gain, recall states still exist, but convergence to a fixed point is not guaranteed. The third distinctive feature is the adjacency of the origin and oscillation regions at larger values of α. Crossing the border between these two regions, say, by increasing β, constitutes a multiple flip bifurcation[31] in which $N(1-\alpha)$ eigendirections about the origin simultaneously lose stability, giving rise to a period-two limit cycle in the subspace orthogonal to all memories.

As in Sec. III A, the region marked origin satisfies $|\beta\lambda_i| < 1$ for all i, where λ_i are the N eigenvalues of the pseudoinverse matrix (10). For $T_{ii} = 0$, the extremal eigenvalues in the limit of large N are[13]

$$\lambda_{min} = -\alpha \quad [N(1-\alpha)\text{- fold degenerate}] \tag{13a}$$

$$\lambda_{max} = 1-\alpha \quad (N\alpha\text{- fold degenerate}) . \tag{13b}$$

Below $\alpha = 0.5$ the origin loses stability at gain $\beta = 1/(1-\alpha)$. This condition defines the border between the regions marked origin and recall. In Appendix B we show that stable recall states appear as soon as this bifurcation occurs. From the stability criterion Eq. (4), and Eq. (13a), convergence to a fixed point is not guaranteed for $\beta > 1/\alpha$, which defines the region marked oscillation in Fig. 3.
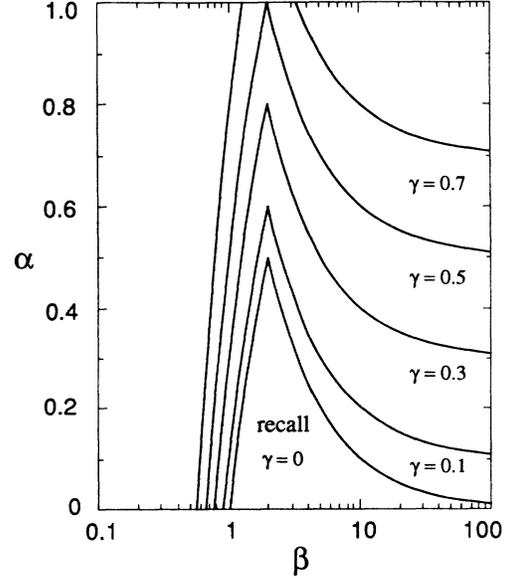
FIG. 4. The recall region for the pseudoinverse rule for various values of diagonal element γ. Note that the maximum capacity in the recall region is for analog gain $\beta = 2$, regardless of γ. Although the recall region is expanded for positive diagonal element, too large a diagonal will greatly reduce the basins of attraction for the recall states, as discussed by Kanter and Sompolinsky in Ref. 13.

Adding a positive diagonal element $T_{ii} = \gamma > 0$ to the connection matrix shifts the eigenvalues to $\lambda_{min} = -\alpha + \gamma$ and $\lambda_{max} = 1 - \alpha + \gamma$ and increases the maximum storage capacity in the recall region to $\alpha_{max} = \frac{1}{2} + \gamma$. The recall region for several values of positive self-coupling are shown in Fig. 4. Note that the maximum always occurs at $\beta = 2$. Recently, Krauth et al.[32] have demonstrated that using a small positive diagonal element with the pseudoinverse rule in an Ising network (at zero temperature) increases the radius of attraction for the recall states. For example, they find numerically that for $\alpha = 0.5$, using a diagonal term of ~ 0.075 instead of zero increases the basins of attraction by about 50%. However, too large a diagonal term greatly reduces the basins of attraction for the recall states.[13,32]

IV. NUMERICAL RESULTS

A. Numerical verification of the phase diagrams

In this section the phase diagrams for the Hebb rule and pseudoinverse rule are investigated numerically for networks of size $N = 100$ with $F(z) = \tanh(\beta z)$ and random, unbiased memory patterns. The data in Figs. 5 and 6 show, as a function of analog gain β, the fraction of randomly chosen initial states which converged to a particular type of attractor—either the origin, a memory pattern (or its inverse), a spurious fixed point, or a period-two limit cycle. These attractor types are the only possibilities. Each panel in these figures is for a fixed value of α, so each represents a horizontal slice through
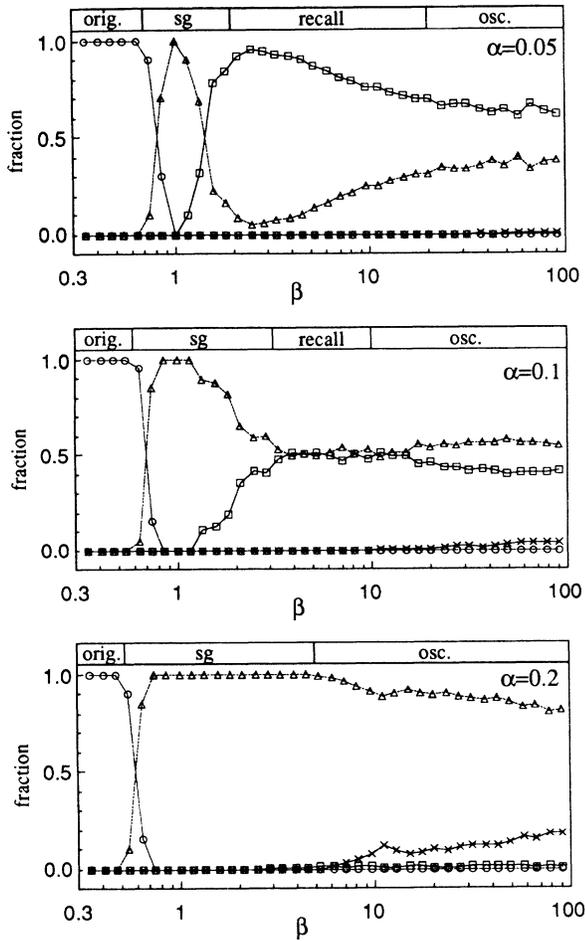
FIG. 5. Numerical data for the Hebb rule showing the fraction of random initial states which lead to the four types of attractors: the origin (circle), a memory pattern or its inverse (square), a spurious fixed point (triangle), or a period-two limit cycle (cross), as a function of neuron gain β. Each data point represents a total of 1000 initial states from 20 matrices constructed from random, unbiased memory patterns with $N = 100$. The three panels are for $\alpha N = 5$, 10, and 20 patterns, and the strip along the top indicates the regions of the phase diagram, Fig. 2, for that value of α.

FIG. 6. Numerical data for the pseudoinverse rule showing the fraction of random initial states which lead to the four types of attractors: the origin (circle), a memory pattern or its inverse (square), a spurious fixed point (triangle), or a period-two limit cycle (cross), as a function of neuron gain β. Each data point represents a total of 1000 initial states from 20 matrices constructed from random, unbiased memory patterns with $N = 100$. The three panels are for $\alpha N = 10$, 25, and 70 patterns, and the strip along the top indicates the regions of the phase diagram, Fig. 3, for that value of α.

the phase diagrams for the Hebb rule (Fig. 5) or the pseudoinverse rule (Fig. 6).

The data in each panel were generated as follows: For each of 38 values of β, ranging from $\beta \sim 0.3$ to $\beta \sim 90$, 20 T_{ij} matrices were generated using random, unbiased patterns $\xi_i^\mu = \pm 1$. For each matrix, 50 initial states located at random corners of the state space [$x_i(0) = \pm 1$, $i = 1, \ldots, 100$] were chosen and the attractor for each was found by iterating the map, Eq. (8). The condition for convergence was $\|\mathbf{x}(t) - \mathbf{x}(t-2)\| < 10^{-6}$, where distances are defined $\|\mathbf{z}\| \equiv (\frac{1}{2}N) \sum_i |z_i|$. Though the initial states were located at the corners of the hypercubic state space, all attractors were real-valued N vectors located away from the corners of the state space. Plotted in each panel are the fractions of the $20 \times 50 = 1000$ runs at each β which converged to each of the four attractor types. A

fixed point $\mathbf{x}^*$ was counted as a recall state if, for any μ, $\|\mathrm{sgn}(\mathbf{x}^*) \pm \xi^\mu\| < 0.05$; similar criteria were used to recognize the other attractor types.

Along the top of each panel in Figs. 5 and 6 is a strip marked "orig.," "recall," etc. These strips show the regions of the theoretical phase diagram (from Figs. 2 and 3) for the particular value of α in that panel. The appearance of the various attractor types corresponds very closely to the theoretical regions in these slices, giving strong numerical support to the phase diagrams. Furthermore, the data indicate that the basins of attraction for limit cycles in the oscillation region do occupy a significant part of state space as soon as the stability criterion is violated. That is, the oscillation region is more that just the region where convergence to a fixed point is not guaranteed by the stability criterion; it is in fact the region where oscillatory modes are quite abundant.

B. Analog annealing: Improved recall at low gain

Figures 5 and 6 show that the probability of recall is greater at lower values of analog gain in the recall region. This phenomenon suggests a potentially powerful technique for annealing a deterministic analog neural network to a good (low-energy) solution.[8] Annealing by varying the analog gain is not only useful as a fast numerical technique, but can be easily implemented in analog electronics, eliminating the need for electronic noise generators to perform stochastic annealing.

As in the case of standard simulated annealing,[33] convergence times at reduced gain can be quite long. To speed convergence, the gain should follow an annealing schedule, starting at the low-gain border of the recall phase and ending at the high-gain border. The phase diagrams, Figs. 2 and 3, can be used to find the range of gains for the annealing schedule for a given storage ratio α. Further numerical work, as well as analytical results showing that the average number of spurious attractors is dramatically reduced at low analog gain, is planned to be presented in a subsequent paper.[7]

V. DISCUSSION

We have studied the properties of an analog iterated-map neural network configured as an associative memory, with an emphasis on stability and the differences between this system and the corresponding thermodynamic Ising-model neural networks. For the special case of unbiased random patterns and identical nonlinear transfer functions, we have presented novel phase diagrams for two important learning algorithms, the Hebb rule and the pseudoinverse rule. For each rule there is a well-defined region with good recall properties and guaranteed convergence to a fixed point. These diagrams are confirmed numerically, and it is demonstrated numerically that using a lower gain increases the probability of good recall, allowing deterministic analog annealing.

The usefulness of analog dynamics goes beyond the stability and enhanced recall properties studied here. By taking advantage of the generality of the stability results of Sec. II B, one can design stable networks of neurons having nonsigmoidal transfer functions with computationally useful properties. As an example, the stability results apply to three-state ($+1$, 0, -1) neurons,[34] generalized to a smooth "staircase" analog transfer function. Networks of three-state analog neurons bear a strong resemblance to the mean-field spin-1 Ising model at finite temperature,[35] with regions of parameter space where both the origin and recall states are locally stable. Such systems might be used to allow an "I don't know" state of the network, such that initial states with insufficient overlap with any pattern will converge to the origin. In a preliminary investigation of this system, we found that the attractors of this system include not only the recall states and the origin, but new mixture states in which a pattern was partially recalled, with some neurons converging to the zero-output state.

Another generalization of the iterated-map associative memory is the deliberate inclusion of limit cycles as recall states. Several techniques for storing and recalling limit cycles have been explored in both continuous-time systems with delay[36] and parallel-update networks.[37] Because these models use asymmetric connections, very little is known analytically about their stability or the types of attractors they can produce. On the other hand, it is possible to store two-cycle attractors in the iterated-map network using a symmetric connection matrix. This can be done most easily with a generalized Hebb rule in which a weighted Hebb matrix of the desired oscillatory modes ζ^ν is subtracted from a Hebb matrix for the fixed-point patterns ξ^μ:

$$T_{ij} = \frac{1}{N} \left[\sum_{\mu=1}^{M_{\mathrm{fp}}} \xi_i^\mu \xi_j^\mu - \Lambda \sum_{\nu=1}^{M_{\mathrm{osc}}} \zeta_i^\nu \zeta_j^\nu \right] . \tag{14}$$

The weighting factor Λ can be used to cause fixed-point patterns and two-cycle patterns to appear at different values of analog gain. A detailed analysis of such an analog network, yielding, for example, the combined storage capacity of limit cycles as well as fixed points, remains an open problem.

APPENDIX A: STORAGE CAPACITY FOR THE HEBB RULE

In this appendix we find the border separating the spin-glass region from the recall region in the phase diagram for the Hebb rule, (Fig. 2). The derivation is a slight generalization of a cavity method approach presented in Ref. 29 but is somewhat simpler because of the absence of the reaction field. The form assumed for the nonlinear function F (taken to be identical for all i) is described at the beginning of Sec. III. We also assume all memory patterns, $\xi_i^\mu = \pm 1$, to be unbiased, and we set $I_i = 0$ for all i. For the special choice $F(z) = \tanh(\beta z)$, the border we obtain is the same as that obtained for the Ising-model network at temperature $1/\beta$.[2,26,29] Throughout this appendix and Appendix B, sums over roman indices $(i,j,k,\dots)$ run from 1 to N; sums over greek indices $(\mu,\nu,\rho,\dots)$ run from 1 to αN.

A recall state is characterized by the existence of a fixed point of the iterated map, which satisfies

$$x_i = F\left[\sum_j T_{ij} x_j \right], \quad i=1,\dots,N , \tag{A1}$$

and which has a large $[O(1)]$ overlap with a single memory pattern, where the overlaps m^μ are defined

$$m^\mu = \frac{1}{N} \sum_i \xi_i^\mu x_i \; . \tag{A2}$$

For the Hebb matrix, Eq. (9), the input h_i to neuron i can be written in terms of the m^μ as

$$h_i = \sum_j T_{ij} x_j = \sum_\mu \xi_i^\mu m^\mu \; , \tag{A3}$$

which gives a set of αN fixed-point equations for the overlaps

$$m^\mu = \frac{1}{N} \sum_i \xi_i^\mu F(h_i), \quad \mu = 1, \ldots, \alpha N \; . \tag{A4}$$

For F odd and $\xi_i^\mu = \pm 1$, these equations can be written

$$m^\mu = \frac{1}{N} \sum_i F(\xi_i^\mu h_i) = \frac{1}{N} \sum_i F(H_i^\mu) = \langle F(H^\mu) \rangle \; , \tag{A5}$$

where $H_i^\mu \equiv \xi_i^\mu h_i$. Borrowing spin-glass terminology, H_i^μ will be referred to as a local field for memory μ. The brackets in (A5) denote an average over the index i: $\langle z \rangle \equiv (1/N) \sum_i z_i$. In the large-$N$ limit, this average can be written as an integral over the distribution of local fields $P(H^\mu)$:

$$m^\mu = \int dH^\mu P(H^\mu) F(H^\mu) \; . \tag{A6}$$

We now seek a self-consistent expression for the distribution function $P(H^1)$ when $m^1 \sim 1$ and $m^\mu \sim O(N^{-1/2})$ for $\mu > 1$. The local field for pattern 1,

$$H_i^1 = \xi_i^1 \sum_\nu \xi_i^\nu m^\nu \; , \tag{A7}$$

can be split into two parts,

$$H_i^1 = m^1 + \xi_i^1 \sum_{\nu > 1} \xi_i^\nu m^\nu \; . \tag{A8}$$

For $\alpha \sim O(1)$, the second term on the right side of (A8) acts as a noise term which we take to be Gaussian distributed with zero mean and variance σ^2 given by

$$\sigma^2 = \sum_{\nu > 1} (m^\nu)^2 \; . \tag{A9}$$

To evaluate the sum of squares in (A9), we first write the overlaps m^ν with the uncondensed patterns using (A3) and (A4):

$$m^\nu = \frac{1}{N} \sum_i \xi_i^\nu F\left[\sum_\rho \xi_i^\rho m^\rho \right] \; . \tag{A10}$$

Notice that the right side of (A10) is of the form $\sum_i A_i B_i$. A sum of this form with *uncorrelated* random variables A_i and B_i has an expected square of $\sum_i A_i^2 B_i^2$. In (A10), however, the two factors in the sum over i are *correlated* through the $\rho = \nu$ term in the argument of F, and this term must be treated separately before squaring. Writing the correlated terms separately,

$$m^\nu = \frac{1}{N} \sum_i \xi_i^\nu F\left[\sum_{\rho \neq \nu} \xi_i^\rho m^\rho + \xi_i^\nu m^\nu \right] \; , \tag{A11}$$

and noting that the single term $(\rho = \nu)$ is small compared to the sum over all the rest $(\rho \neq \nu)$, we expand F to first

order in m^ν giving

$$m^\nu = \frac{1}{N} \sum_i \xi_i^\nu \left[F\left[\sum_{\rho \neq \nu} \xi_i^\rho m^\rho \right] + \xi_i^\nu m^\nu F'\left[\sum_{\rho \neq \nu} \xi_i^\rho m^\rho \right] \right] \; , \tag{A12}$$

where F' is the derivative of the function F. The missing $\rho = \nu$ term in the argument of F' only affects the value of F' to order $O(1/N)$ which we neglect by taking the argument to be the whole h_i. We now define the quantity C,

$$C \equiv \langle F'(h) \rangle = \frac{1}{N} \sum_i F'(h_i) \; , \tag{A13}$$

and write (A12) as

$$m^\nu(1 - C) = \frac{1}{N} \sum_i \xi_i^\nu F\left[\sum_{\rho \neq \nu} \xi_i^\rho m^\rho \right] \; . \tag{A14}$$

With the $\rho = \nu$ term removed from the argument of F, the two factors in the sum over i on the right side of (A14) are now uncorrelated and can be squared to yield an expected value of

$$[(1 - C)m^\nu]^2 = \frac{1}{N} \sum_i F^2\left[\sum_{\rho \neq \nu} \xi_i^\rho m^\rho \right] \approx \frac{1}{N} \sum_i F^2(h_i) \; , \tag{A15}$$

where, again, the $O(1/N)$ error in the value of F^2 from the $\rho = \nu$ term is ignored. Next, we define the quantity q in analogy with the Edwards-Anderson order parameter,

$$q \equiv \langle F^2(h) \rangle = \frac{1}{N} \sum_i F^2(h_i) \; , \tag{A16}$$

and write (A15) as

$$(m^\nu)^2 = q/N(1 - C)^2 \; . \tag{A17}$$

From (A9) and (A17), the variance σ^2 of the local-field distribution is given in terms of the quantities C and q by

$$\sigma^2 = \alpha q / (1 - C)^2 \; . \tag{A18}$$

Because F' and F^2 are both even functions, we can multiply their arguments by ± 1 without changing their values. This allows us to write the averages in Eqs. (A13) and (A16) in terms of H_i^1 rather than h_i, and finally as integrals over the distribution of local fields $P(H^1)$, given by the normalized Gaussian distribution

$$P(H^1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(H^1 - m^1)^2}{2\sigma^2} \right] \; , \tag{A19}$$

where the variance σ^2 is given by (A18). Together with Eq. (A6), the self-consistent equations for quantities m^1, C, and q are given by the following integrals:

$$m^1 = \int dH^1 P(H^1) F(H^1) \; , \tag{A20a}$$

$$C = \int dH^1 P(H^1) F'(H^1) \; , \tag{A20b}$$

$$q = \int dH^1 P(H^1) F^2(H^1) \; . \tag{A20c}$$

After a change of variables, $y \equiv (H^1 - m^1)/\sigma$, Eqs. (A18)–(A20) yield the self-consistent set of equations (12a)–(12d) in Sec. III A.

APPENDIX B: RECALL STATES
FOR THE PSEUDOINVERSE RULE

In this appendix we show that for the pseudoinverse learning rule, stable recall states exist whenever $\alpha < 1$ and $\beta > 1/(1-\alpha)$. This implies that there is no spin-glass phase for the pseudoinverse rule in the iterated-map network, in contrast to the thermodynamic Ising-spin network with the same learning rule.[13] The analysis below follows Kanter and Sompolinsky.[13]

As described in Appendix A, a recall state is defined as a fixed point which has a large overlap with a single pattern (again, taken to be pattern 1). For large N, the single large overlap m^1 can be written as an integral over the distribution of local fields,

$$m^1 = \frac{1}{N} \sum_i F(H_i^1) \xrightarrow{N \to \infty} \int dH^1 P(H^1) F(H^1) , \quad \text{(B1)}$$

where $P(H^1)$ is a Gaussian distribution whose mean and variance must be found self-consistently. The local field for memory pattern 1,

$$H_i^1 = \xi_i^1 \sum_{j \neq i} T_{ij} x_j , \quad \text{(B2)}$$

with the pseudoinverse matrix

$$T_{ij} = \frac{1}{N} \sum_{\mu,\nu} \xi_i^\mu (C^{-1})_{\mu\nu} \xi_j^\nu , \quad \text{(B3)}$$

$$C_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu , \quad \text{(B4)}$$

is given by

$$H_i^1 = \xi_i^1 \left[\sum_{\mu,\nu} \xi_i^\mu [(C^{-1})_{\mu\nu} m^\nu - \alpha x_i] \right] . \quad \text{(B5)}$$

The $-\alpha x_i$ term explicitly takes care of setting the diagonals to zero since the T_{ii} as defined by (B3) are narrowly peaked around α at large N. The state vector x_i, $i = 1, \ldots, N$, can be written as a weighted sum of the pattern vectors, with real-valued weights a^μ, plus a vector χ_i, $i = 1, \ldots, N$, which is perpendicular to the subspace spanned by the patterns

$$x_i = \sum_\mu a^\mu \xi_i^\mu + \chi_i . \quad \text{(B6)}$$

From (A2), (B4), and (B6), the weights a^μ are related to the overlaps m^μ through the inverse correlation matrix

$$a^\mu = \sum_\nu (C^{-1})_{\mu\nu} m^\nu . \quad \text{(B7)}$$

Writing the local field H_i^1 in terms of the a^μ,

$$H_i^1 = (1-\alpha)a^1 + \xi_i^1(1-\alpha) \left[\sum_{\mu > 1} \xi_i^\mu a^\mu \right] - \alpha \chi_i , \quad \text{(B8)}$$

reveals a similar structure to the Hebb rule [compare (B8) to (A8)], with a "signal" term proportional to a^1 and a "noise" term due to the other patterns. The third term on the right causes the state to relax towards the subspace spanned by the memories and does not add any destabilizing noise. Comparing Eqs. (B8) and (A8) also reveals why the pseudoinverse rule allows perfect recall with an extensive number of patterns and the Hebb rule does not: for the pseudoinverse rule, the variance of the Gaussian noise due to the other patterns is given by

$$\sigma_{PI}^2 = (1-\alpha)^2 \sum_{\mu > 1} (a^\mu)^2 , \quad \text{(B9)}$$

whereas for the Hebb rule, the variance is

$$\sigma_H^2 = \sum_{\mu > 1} (m^\mu)^2 . \quad \text{(B10)}$$

When the state of the network is fully aligned with, say, pattern 1, then all a^μ, $\mu > 1$, vanish. On the other hand, the overlaps m^μ, $\mu > 1$, do not vanish, even when the state is perfectly aligned with a pattern, unless all memories are orthogonal. Therefore the noise term for the Hebb rule is always nonzero.

In a recall state (for pattern 1), $a^1 = m^1$ and $a^\mu = 0$ for $\mu > 1$, giving a δ-function distribution for the local fields,

$$P(H^1) = \delta(H^1 - (1-\alpha)m^1) . \quad \text{(B11)}$$

Inserting this distribution into (B1) gives the self-consistent solution for the overlap with pattern 1,

$$m^1 = F((1-\alpha)m^1) . \quad \text{(B12)}$$

When the function F is tanh-like with maximum slope β, there is nonzero m^1 given by (B12) whenever $\alpha < 1$ and $\beta > 1/(1-\alpha)$. The value of m^1 grows continuously from zero at the transition. In analogy with thermodynamics, the appearance of recall states is therefore a second-order transition. As mentioned above, the behavior of the analog network with the pseudoinverse rule for the particular choice $F(z) = \tanh(\beta z)$ is not the same as the corresponding Ising-spin network at finite temperature $1/\beta$: as shown by Kanter and Sompolinsky,[13] the recall states for the Ising model appear at a value of β significantly above $1/(1-\alpha)$ and that the transition to the recall state is first order. These differences can be attributed to the absence of a reaction field in our system.

[1] Proceedings of the Heidelberg Colloquium on Glassy Dynamics, Vol. 275 of Lecture Notes in Physics, edited by J. L. van Hemmen and I. Morgenstern (Springer, Berlin, 1987).

[2] M. Mezard, G. Parisi, and M. A. Virasoro, Spin Glass Theory and Beyond (World Scientific, Singapore, 1987).

[3] A recent collection of papers on the spin-glass approach to neural networks appears in J. Phys. A 22, 1953 (1989).

[4] C. A. Mead, Analog VLSI and Neural Systems (Addison-Wesley, Reading, MA, 1989).

[5] J. L. Wyatt Jr. and D. L. Stanley, in Neural Information Processing Systems, Denver, Colorado, 1987, edited by D. Z. Anderson (AIP, New York, 1988), p. 860.

[6] C. M. Marcus and R. M. Westervelt, Phys. Rev. A 39, 347 (1989); in Advances in Neural Information Processing, Denver, Colorado, 1988, edited by D. S. Touretzky (Morgan Kaufmann, San Mateo, 1989), p. 568.

[7] F. R. Waugh, C. M. Marcus, and R. M. Westervelt (unpublished).

[8]J. J. Hopfield and D. W. Tank, Biol. Cybern. **52**, 141 (1985).

[9]C. Koch, J. Marroquin, and A. Yuille, Proc. Nat. Acad. Sci. U.S.A. **83**, 4263 (1986); A. Blake and A. Zisserman, *Visual Reconstruction* (MIT Press, Cambridge, MA, 1987).

[10]C. M. Marcus and R. M. Westervelt, Phys. Rev. A **40**, 501 (1989).

[11]D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).

[12]L. Personnaz, I. Guyon, and G. Dreyfus, J. Phys. Lett. (Paris) **46**, L359 (1985).

[13]I. Kanter and H. Sompolinsky, Phys. Rev. A **35**, 380 (1987).

[14]J. J. Hopfield, Proc. Nat. Acad. Sci. U.S.A. **81**, 3008 (1984).

[15]M. A. Cohen and S. Grossberg, IEEE Trans. Syst. Man Cybern. **SMC-13**, 815 (1983).

[16]R. Brout and H. Thomas, Physics (N.Y.) **3**, 417 (1967).

[17]D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philos. Mag. **35**, 593 (1977).

[18]D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975); S. Kirkpatrick and D. Sherrington, Phys. Rev. B **17**, 4384 (1978).

[19]A. J. Bray and M. A. Moore, J. Phys. C **12**, L441 (1979).

[20]M. Y. Choi and B. A. Huberman, Phys. Rev. B **28**, 2547 (1983).

[21]K. Nemoto and H. Takayama, J. Phys. C **18**, L529 (1985).

[22]C. M. Soukoulis, K. Levin, and G. S. Grest, Phys. Rev. Lett. **48**, 1756 (1982); Phys. Rev. B **28**, 1495 (1983).

[23]D. D. Ling, D. R. Bowman, and K. Levin, Phys. Rev. B **28**, 262 (1983).

[24]J. D. Reger, K. Binder, and W. Kinzel, Phys. Rev. B **30**, 4028 (1984).

[25]E. J. Gardner, J. Phys. A **19**, L1047 (1986).

[26]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N. Y.) **173**, 30 (1987).

[27]See, for example, theorem 12.1.2 in J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables* (Academic, New York, 1970).

[28]A. Crisanti and H. Sompolinsky, Phys. Rev. A **36**, 4922 (1987).

[29]E. Domany, W. Kinzel, and R. Meir, J. Phys. A **22**, 2081 (1989).

[30]S. Diederich and M. Opper, Phys. Rev. Lett. **58**, 949 (1987).

[31]J. Guckenheimer and P. Holmes, *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer, New York, 1983).

[32]W. Krauth, M. Mezard and J. P. Nadal, Complex Systems **2**, 387 (1988). Note that their definition of γ differs from ours: By our definition, the diagonal element of the connection matrix is γ; by their definition, the diagonal element is $\gamma \sum_{j \ (\neq i)} T_{ij} \xi_i^\mu \xi_j^\mu$, or, in the large-$N$ limit, $\gamma(1-\alpha)$.

[33]S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Science **220**, 671 (1983); S. Kirkpatrick, J. Stat. Phys. **34**, 975 (1984).

[34]J. S. Yedidia, J. Phys. A **22**, 2265 (1989); C. Meunier, D. Hansel, and A. Verga, J. Stat. Phys. **55**, 859 (1989).

[35]M. Blume, Phys. Rev. **141**, 517 (1966); W. Capel, Physica **32**, 966 (1966).

[36]S. Grossberg, Stud. Appl. Math. **44**, 135 (1970). D. Kleinfeld, Proc. Nat. Acad. Sci. U.S.A. **83**, 9469 (1986); H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861 (1986); H. Gutfreund and M. Mezard, *ibid.* **61**, 235 (1988); U. Riedel, R. Kühn, and J. L. van Hemmen, Phys. Rev. A **38**, 1105 (1988).

[37]S. Dehaene, J. P. Changeux, and J. P. Nadal, Proc. Nat. Sci. U.S.A. **84**, 2727 (1987); I. Guyon, L. Personnaz, J. P. Nadal, and G. Dreyfus, Phys. Rev. A **38**, 6365 (1988); Y. Mori, P. Davis, and S. Nara, J. Phys. A **22**, L525 (1989).