Journal of Statistical Physics, Vol. 50, Nos. 1/2, 1988

Nonlinear Neural Networks. II. Information Processing

J. L. van Hemmen[1], D. Grensing[2], A. Huber[2], and R. Kühn[2,3]

Received April 14, 1987; revision received August 7, 1987

Information processing in nonlinear neural networks with a finite number q of stored patterns is studied. Each network is characterized completely by its synaptic kernel Q. At low temperatures, the nonlinearity typically results in $2^{q-2}-q$ metastable, pure states in addition to the q retrieval states that are associated with the q stored patterns. These spurious states start appearing at a temperature $\tilde{T}_q$, which depends on q. We give sufficient conditions to guarantee that the retrieval states bifurcate first at a critical temperature T_c and that $\tilde{T}_q/T_c \to 0$ as $q \to \infty$. Hence, there is a large temperature range where *only* the retrieval states and certain symmetric mixtures thereof exist. The latter are unstable, as they appear at T_c. For clipped synapses, the bifurcation and stability structure is analyzed in detail and shown to approach that of the (linear) Hopfield model as $q \to \infty$. We also investigate memories that forget and indicate how forgetfulness can be explained in terms of the eigenvalue spectrum of the synaptic kernel Q.

KEY WORDS: Nonlinear neural networks; bifurcation and stability; information processing; logical operations; forgetfulness.

1. INTRODUCTION

The equilibrium statistical mechanics of a nonlinear neural network with finitely many, say q, stored patterns allows an exact solution.[1,2] As was shown in a previous paper,[1] to be referred to as GT, the free energy may be obtained by maximizing a functional with respect to all solutions of a fixed-point equation. The stable and metastable states constitute free

[1] Sonderforschungsbereich 123, Universität Heidelberg, D-6900 Heidelberg, West Germany.

[2] Institut für Theoretische Physik und Sternwarte der Universität Kiel, D-2300 Kiel, West Germany.

[3] Present address: Sonderforschungsbereich 123, Universität Heidelberg, D-6900 Heidelberg, West Germany.

energy valleys in phase space and therefore are basins of attraction for any dynamics of Monte Carlo type. They are fully determined by a set of order parameters $m(\mathbf{x})$ with $\mathbf{x} \in \mathscr{C}^q = \{-1, 1\}^q$. The $m(\mathbf{x})$ solve the fixed-point equation associated with the model under consideration.

The neurons are modeled as Ising spins $S(i) = \pm 1$ and the data (the "patterns") are stored in the synaptic efficacies (the "coupling constants"), which we will denote here by J_{ij}. Throughout what follows we assume $J_{ij} = J_{ji}$ and require locality via some generalized Hebbian rule so that, given q patterns, J_{ij} may be assumed to be determined by the *local* information ξ_i and ξ_j available to neurons i and j only, where $\xi_i = \{\xi_{i\alpha}, 1 \leqslant \alpha \leqslant q\}$. So we may write $J_{ij} = Q(\xi_i; \xi_j)$ for some *synaptic kernel* Q, which is symmetric by assumption. That is, $Q(\mathbf{x}; \mathbf{y}) = Q(\mathbf{y}; \mathbf{x})$ for all $\mathbf{x}$ and $\mathbf{y}$ in $\mathscr{C}^q$ or, more generally, $\mathbb{R}^q$. Under a rather weak invariance condition [GT (3.1)], which is satisfied by nearly all neural network models, a complete spectral theory for Q can be given and hence[1] all the states that bifurcate from the high-temperature phase can be determined. What these states look like is one of the main subjects of the present work.

Given the above conceptual structure, we will study in this paper how information is processed. We know how the data are stored, through the synaptic kernel Q, and the question is how they can be retrieved. In the present context, retrieval means downhill motion in a free energy valley or ergodic component, and thus a nontrivial set of order parameters $\{m(\mathbf{x}), \mathbf{x} \in \mathscr{C}^q\}$ which should have bifurcated from zero, i.e., from the high-temperature phase $m \equiv 0$.

In Section 2 the solutions of the fixed-point equation are analyzed in detail. The roles of symmetry and parity are clarified, the bifurcation from $m \equiv 0$ is related to the spectrum of Q, the stability of bifurcating solutions is studied, and it is indicated under what condition the stability matrix attains a much simpler, block-diagonal structure.

The eigenvectors $v_\rho(\mathbf{x})$ of Q, where (cf. GT, Section 3) the label ρ denotes one of the 2^q subsets of $\{1,..., q\}$, can be associated with states of the neural network model that has Q as synaptic kernel. As will be shown in Section 3, these states, which are called *pure states*, are stable or metastable at sufficiently low temperatures if and only if λ_ρ, the eigenvalue belonging to v_ρ, is positive. The retrieval states constitute a special but very important example. They are characterized uniquely by the fact that $|\rho|$, the size of the set ρ, equals one. There are q retrieval states corresponding to q stored patterns. For the Hopfield model, which is linear, these are the only pure states, but for a typical *non*linear neural network there are about 2^{q-2} pure states belonging to positive eigenvalues. At low temperatures, all these states are stable or metastable and hence constitute basins of attraction. Most of them are not wanted, however.

For a large class of models (of the inner-product type) it turns out that the retrieval states bifurcate first and that they are the *only* ones that are stable just below T_c; see Section 3, Eqs. (3.5) and (3.6). A natural question is, however, whether there are other states that bifurcate from zero at T_c. This is indeed the case. As shown in Section 4, there are 2^q *symmetric* states that bifurcate from $m \equiv 0$ at T_c. It is proven that, except for the retrieval states, they are all unstable when they first appear. We present a general method to analze this type of bifurcation problem, which occurs at a q-fold *degenerate* eigenvalue and is to be analyzed in a 2^q-dimensional space. (For a bifurcation at a lower temperature, the degeneracy is even higher.)

Specializing to clipped synapses, we show in Section 5 that, as q becomes large, the bifurcation and stability structure in the temperature range $\tilde{T}_q < T < T_c$ *reduces to that of the linear Hopfield model* with the same number of patterns. Here $\tilde{T}_q = 2^{-q}\tilde{\lambda}$, where $\tilde{\lambda}$ is the second largest eigenvalue of the synaptic kernel. For the Hopfield model, $\tilde{\lambda}$ vanishes, but for a general nonlinear neural network model $\tilde{\lambda} > 0$. Below $\tilde{T}_q$ we enter the temperature range where the nonlinearity becomes important. Fortunately, $\tilde{T}_q/T_c \to 0$ as $q \to \infty$ [see GT (3.36) and (4.18)].

The spectral theory developed previously[1] also applies to forgetful memories. In Section 6 we briefly sketch how the forgetfulness may be explained.

Finally, in Section 7, the results are discussed and it is indicated how the disadvantages of the nonlinearity may be eliminated while keeping the benefits.

2. GENERAL PROPERTIES OF THE FIXED-POINT EQUATION

For unbiased binary input data the equilibrium statistical mechanics of a neural network model with synaptic kernel Q is governed by real solutions to the fixed-point equation

$$m(\mathbf{x}) = \tanh[\beta 2^{-q}(Qm)(\mathbf{x})], \qquad \mathbf{x} \in \{-1, 1\}^q \tag{2.1}$$

which is GT (2.22). Now Q operates as a $2^q \times 2^q$ matrix Q with elements $Q(\mathbf{x}; \mathbf{y})$. Below we present, for easy reference and by way of preparation for the sections to follow, a number of general properties of the (real) solutions to (2.1). Q will only be subject to the restrictions of being real, symmetric, and satisfying the invariance condition GT (3.1).

2.1. Covariance

Let m be a solution of the fixed-point equation (2.1) and g an arbitrary element of the group of inversions $\mathcal{G}_q$ introduced in GT, Section 3.1. Then the 2^q-vector gm with components

$$(gm)(\mathbf{x}) = (m)(g\mathbf{x}) \tag{2.2}$$

also satisfies (2.1). To wit, let g operate on both sides of (2.1) in the manner specified by (2.2). Then one obtains

$$\begin{aligned}(gm)(\mathbf{x}) &= \tanh\left[\beta 2^{-q} \sum_{\mathbf{y}} Q(g\mathbf{x}; \mathbf{y})\, m(\mathbf{y})\right] \\ &= \tanh\left[\beta 2^{-q} \sum_{\mathbf{y}} Q(\mathbf{x}; g\mathbf{y})\, m(\mathbf{y})\right] \\ &= \tanh\left[\beta 2^{-q} \sum_{\mathbf{y}} Q(\mathbf{x}; \mathbf{y})\, m(g\mathbf{y})\right] \\ &= \tanh[\beta 2^{-q}(Q(gm))(\mathbf{x})] \end{aligned} \tag{2.3}$$

Here, we have used the invariance property GT (3.2) in the second, and the inversion property $g^{-1} = g$ in the third step. By a similar argument, the covariance (2.3) is seen to hold for our general fixed-point equation GT (2.20),

$$m(\mathbf{x}) = \tanh\left[\beta \int d\mu(\mathbf{y})\, Q(\mathbf{x}; \mathbf{y})\, m(\mathbf{y})\right] \tag{2.4}$$

provided the measure μ is invariant under $\mathcal{G}_q$.

In the same manner, covariance as in (2.3) follows for the set of solutions of (2.1) or (2.4) with respect to *any* group $\mathcal{G}$ of transformations that map the range of $\mathbf{x}$ onto itself and leave both Q and μ invariant, such as the group of rotations in the Gaussian case treated in GT, Section 4.

Due to the covariance property (2.3), the solutions of (2.1), and more generally of (2.4), always come in maximal sets of solutions that are transformed into each other by suitable elements of $\mathcal{G}_q$. These sets form *equivalence classes* under $\mathcal{G}_q$. Each equivalence class consists of solutions invariant under one of the subgroups of $\mathcal{G}_q$. For the purposes of the bifurcation and stability analysis in this and the subsequent sections, it will generally suffice to restrict ourselves to one element of each equivalence class, and we shall as a rule make no special mention of solutions equivalent to the ones considered.

2.2. Parity

In this subsection we will use without further ado the properties GT (3.5)–(3.10) of the eigenvectors v_ρ and the eigenvalues λ_ρ of the synaptic kernel $Q(\mathbf{x}; \mathbf{y})$.

Since the eigenvectors v_ρ form a complete set, any solution m of (2.1) can be expanded in terms of them,

$$m(\mathbf{x}) = \sum_\rho \alpha_\rho v_\rho(\mathbf{x}) \tag{2.5}$$

Inserting (2.5) into (2.1) and using the orthogonality of the v_ρ, one obtains for the coefficients α_ρ the system of equations

$$\alpha_\rho = 2^{-q} \sum_{\mathbf{x}} v_\rho(\mathbf{x}) \tanh \left[\beta 2^{-q} \sum_\sigma \lambda_\sigma \alpha_\sigma v_\sigma(\mathbf{x}) \right] \tag{2.6}$$

where ρ and σ range through all the subsets of $\{1,..., q\}$. The system (2.6) is equivalent to (2.1).

If, now, the synaptic kernel Q has definite parity in the sense of GT (3.10), then we have

$$\begin{array}{llll} Q \text{ odd:} & \alpha_\rho = 0 & \text{whenever} & |\rho| \text{ is even} \\ Q \text{ even:} & \alpha_\rho = 0 & \text{whenever} & |\rho| \text{ is odd} \end{array} \tag{2.7}$$

This is seen immediately from (2.6), since $v_\rho(\mathbf{x})$ is even (odd) under $\mathbf{x} \to -\mathbf{x}$ if the cardinality $|\rho|$ is even (odd), and λ_ρ vanishes for all v_ρ that have a parity opposite to that of Q.

In practical work, Q always has a definite parity. For instance, in the case of clipped synapses (Section 5) and memories that forget (Section 6), Q is odd.

2.3. Bifurcation

Obviously, the *trivial solution* $m(\mathbf{x}) \equiv 0$ satisfies (2.1) for all β. It is the *only* (real) solution for temperatures above or equal to the critical temperature ($k_B = 1$)

$$T_c = 2^{-q} \lambda_{\max} \tag{2.8}$$

where $\lambda_{\max}$ denotes the largest positive eigenvalue of the synaptic kernel Q. To see this, we consider an arbitrary solution $m \not\equiv 0$ of (2.1) and note that its norm squared satisfies the inequality

$$\begin{aligned}(m, m) &\equiv \sum_{\mathbf{x}} |m(\mathbf{x})|^2 \\ &= \sum_{\mathbf{x}} |\mathrm{m}(\mathbf{x})|\, |\tanh[(\beta/2^q)(Qm)(\mathbf{x})]| \\ &< \beta 2^{-q} \sum_{\mathbf{x}} |m(\mathbf{x})|\, |(Qm)(\mathbf{x})| \end{aligned} \tag{2.9}$$

In obtaining the *strict* inequality in (2.9) we have used the fact that $|\tanh(u)| < |u|$ for all $u \in \mathbb{R}$ except $u = 0$, and that the argument of the tanh in (2.9) must be nonzero for some $\mathbf{x}$ if m is nontrivial. Now, for each nonzero term in the sums in (2.9), the quantities $m(\mathbf{x})$, $\tanh[\beta 2^{-q}(Qm)(\mathbf{x})]$, and $(Qm)(\mathbf{x})$ all have the same sign. [This is by virtue of (2.1) and the fact that the hyperbolic tangent is odd.] Hence we conclude

$$\begin{aligned}(m, m) &< \beta 2^{-q} \sum_{\mathbf{x}} m(\mathbf{x})(Qm)(\mathbf{x}) \\ &= \beta 2^{-q}(m, Qm) \\ &\leqslant \beta 2^{-q} \lambda_{\max}(m, m)\end{aligned} \tag{2.10}$$

This implies $\beta > 2^q/\lambda_{\max}$ when $m \not\equiv 0$. Thus, for all temperatures down to and including the critical temperature (2.8), the zero solution is the only (real) solution of (2.1).

Nontrivial solutions that *bifurcate from zero* can do so only at temperatures where $2^q T$ equals one of the positive eigenvalues of Q. This is most easily understood by writing (2.1) as

$$F_{\mathbf{x}}(m, \beta) = 0, \qquad \mathbf{x} \in \{-1, 1\}^q \tag{2.11}$$

with the functions $F_{\mathbf{x}}$ defined by

$$F_{\mathbf{x}}(m, \beta) := m(\mathbf{x}) - \tanh[\beta 2^{-q}(Qm)(\mathbf{x})] \tag{2.12}$$

By the implicit function theorem[(3),4] there exist, for every pair $(m^{(0)}, \beta^{(0)})$ that satisfy (2.11), a neighborhood B of $\beta^{(0)}$ and a neighborhood $\mathscr{M}$ of $m^{(0)}$ such that, for every $\beta \in B$, Eq. (2.11) has a *unique* solution $m(\beta)$ in $\mathscr{M}$, provided that the matrix of partial derivatives D with elements

$$D_{\mathbf{x},\mathbf{y}} := \partial F_{\mathbf{x}}/\partial m(\mathbf{y}) \tag{2.13}$$

has nonzero determinant at $(m^{(0)}, \beta^{(0)})$. So, as long as $\det(D)$ does not vanish, $m(\beta)$ is locally unique, and therefore a bifurcation is not possible. According to (2.12), the matrix elements of D are

$$D_{\mathbf{x},\mathbf{y}} = \delta_{\mathbf{x},\mathbf{y}} - \beta 2^{-q} Q(\mathbf{x}; \mathbf{y})[1 - \tanh^2(\beta 2^{-q}(Qm)(\mathbf{x}))] \tag{2.14}$$

[4] See Ref. 4, p. 35, and Chapter IV on bifurcation at a multiple eigenvalue (pp. 70–85).

so that along the trivial solution

$$\det D|_{(\mathbf{m}=0;\beta)} = \det(\mathbb{1} - \beta 2^{-q}Q) = \prod_{\rho \subseteq \{1,\dots,q\}} (1 - \beta 2^{-q}\lambda_\rho)$$

It follows that the only temperatures where solutions of (2.1) can bifurcate from zero are given by the condition ($k_B = 1$)

$$T_\rho = 2^{-q}\lambda_\rho \tag{2.15}$$

where λ_ρ ranges through the *positive* eigenvalues of the synaptic kernel Q.

Conversely, nontrivial solutions do indeed bifurcate from zero at every temperature T_ρ. To see this, one need only try in (2.1) the Ansatz

$$m(\mathbf{x}) = \alpha_\rho v_\rho(\mathbf{x}), \qquad \mathbf{x} \in \{-1, 1\}^q \tag{2.16}$$

where v_ρ is an eigenvector of Q with positive eigenvalue λ_ρ and α_ρ a corresponding amplitude, which is to be determined. Upon inserting (2.16) into (2.1), one finds, due to the eigenvector property of v_ρ that α_ρ has to satisfy

$$\alpha_\rho v_\rho(\mathbf{x}) = \tanh[(\beta 2^{-q}\lambda_\rho)\,\alpha_\rho v_\rho(\mathbf{x})], \qquad \mathbf{x} \in \{-1, 1\}^q \tag{2.17}$$

Since $v_\rho(\mathbf{x})$ only assumes the values ±1, Eq. (2.17) reduces to the single equation

$$\alpha_\rho = \tanh[(\beta 2^{-q}\lambda_\rho)\alpha_\rho] \tag{2.18}$$

For each *positive* eigenvalue λ_ρ of Q, Eq. (2.18) has a nontrivial positive solution for all β's larger than a critical β_ρ determined by the condition

$$\beta_\rho 2^{-q}\lambda_\rho = 1 \tag{2.19}$$

i.e., for *all* T below the bifurcation temperature T_ρ given by (2.15).

The nontrivial solutions of type (2.16) are *unique* in the sense that there is—up to the choice of the sign of α_ρ—exactly one of these for each subset ρ of $\{1,\dots, q\}$ with $\lambda_\rho > 0$. These solutions of the fixed-point equation will be referred to as the *pure states*. Their physical interpretation, their stability behavior, and their relevance for the retrieval of stored information are discussed in Section 3.

Let us finally remark that the first inequality (2.10) contains further valuable information concerning both the global structure of the nontrivial solutions of (4.1) and their local behavior near bifurcation points. This information can be extracted by inserting, into both sides of (2.10), the eigenvector expansion (2.5). Using the orthogonality of the v_ρ, one obtains from (2.10) the inequality

$$\sum_\rho (\beta 2^{-q}\lambda_\rho - 1)\alpha_\rho^2 > 0 \tag{2.20}$$

for any nontrivial solution of the fixed-point equation (2.1). The conclusions to be drawn from (2.20) are most conveniently formulated by extending the definition $T_\rho = 2^{-q}\lambda_\rho$ of the bifurcation temperatures so as to be valid for $\lambda_\rho \leqslant 0$ also. In so doing we can rewrite (2.20) in the form

$$\sum_\rho \frac{T_\rho - T}{T}\alpha_\rho^2 > 0 \tag{2.21}$$

This then implies that any *non*trivial (real) solution to (2.1) contains *at least one* $\alpha_\rho \neq 0$ with ρ such that $T_\rho > T$, since otherwise (2.21) would be violated. As a special case, we recover our earlier result that for $T \geqslant T_c = \max_\rho T_\rho$ the only real solution to the fixed-point equations is the trivial one.

2.4. Stability

Ergodic components are labeled by solutions $m(\mathbf{x})$ of the fixed-point equation (2.1). As was shown in GT, Section 2, their thermodynamic stability is determined by a $2^q \times 2^q$ matrix, the *stability matrix* $\mathscr{S}$, whose elements are given by GT (2.18),

$$\mathscr{S}_{\mathbf{x},\mathbf{y}} = \beta 2^{-q} Q(\mathbf{x};\mathbf{y}) - [1 - m^2(\mathbf{x})]^{-1}\,\delta_{\mathbf{x},\mathbf{y}} \tag{2.22}$$

Here we have used the fact that $p_\gamma = 2^{-q}$ and $m_\gamma \leftrightarrow m(\mathbf{x})$ for $\mathbf{x}$ in $\mathscr{C}^q = \{-1, 1\}^q$. A phase with order parameters $m(\mathbf{x})$ is stable if $\mathscr{S}$ has negative eigenvalues only.

We now perform a change of basis, taking the eigenvectors v_ρ of Q as new basis vectors. Then Q becomes a diagonal matrix and $\mathscr{S}$ reappears in the form

$$\mathscr{S}_{\sigma\sigma'} = \beta(2^{-q}\lambda_\sigma)\,\delta_{\sigma\sigma'} - 2^{-q}\sum_{\mathbf{x}} \frac{v_\sigma(\mathbf{x})v_{\sigma'}(\mathbf{x})}{1 - m^2(\mathbf{x})} \tag{2.23}$$

The indices σ and σ' run through the 2^q subsets of $\{1,..., q\}$. For future work it is convenient to put

$$2^{-q}\lambda_\sigma = \Lambda_\sigma \tag{2.24}$$

Before proceeding, we present two simple examples of (2.23).

First, for the trivial solution $m(\mathbf{x}) \equiv 0$, $\mathscr{S}$ is a diagonal matrix with elements

$$\mathscr{S}_{\sigma\sigma'} = \{\beta A_\sigma - 1\}\, \delta_{\sigma\sigma'} \tag{2.25}$$

Second, the stability of a pure state corresponding to $m(\mathbf{x}) = a_\rho v_\rho(\mathbf{x})$ where a_ρ satisfies (2.18) is also determined by a diagonal matrix, this time of the form

$$\mathscr{S}_{\sigma\sigma'} = \{\beta A_\sigma - (1 - a_\rho^2)^{-1}\}\, \delta_{\sigma\sigma'} \tag{2.26}$$

In both cases stability leads to the simple requirement that $\mathscr{S}_{\sigma\sigma}$ be negative for all σ.

Suppose now that R is a subset of $\{1,..., q\}$ with cardinality $|R| > 1$ and let

$$m(\mathbf{x}) = m^{(R)}(\mathbf{x}) = \sum_{\rho \subseteq R} a_\rho v_\rho(\mathbf{x}) \tag{2.27}$$

be a solution to (2.1). This case is a generalization of the previous, second, example, since here ρ may range through *several* subsets of R. We want to prove the following:

The stability matrix corresponding to (2.27) is *block*-diagonal with blocks of maximal size $2^{|R|} \times 2^{|R|}$. If the synaptic kernel has definite (odd or even) parity, as in GT (3.10), each of the blocks is again block-diagonal and the subblocks have maximal dimension $2^{|R|-1} \times 2^{|R|-1}$.

For the proof, we start by noting that, given R, every set $\sigma \subseteq \{1,..., q\}$ can be decomposed into a part in R and a remainder in the complement of R,

$$\sigma = \alpha \cup \tau, \qquad \text{with} \quad \alpha \subseteq R, \quad \tau \cap R = \varnothing \tag{2.28}$$

so that the stability matrix (2.23) can be rewritten

$$\begin{aligned} \mathscr{S}_{\sigma\sigma'} &= \mathscr{S}_{\alpha\tau,\alpha'\tau'} \\ &= \beta A_{\alpha\tau}\, \delta_{\alpha\alpha'}\, \delta_{\tau\tau'} - 2^{-q} \sum_{\mathbf{x}} \frac{v_\alpha(\mathbf{x})\, v_{\alpha'}(\mathbf{x})\, v_\tau(\mathbf{x})\, v_{\tau'}(\mathbf{x})}{1 - [m^{(R)}(\mathbf{x})]^2} \end{aligned} \tag{2.29}$$

Here we have used GT (3.5) so as to factorize the $v_\sigma(\mathbf{x})$ into $v_\alpha(\mathbf{x})\, v_\tau(\mathbf{x})$. We note that α and α' may occur in $m^{(R)}(\mathbf{x})$, but τ and τ' do not. Suppose $\tau \neq \tau'$. Then there is at least one single index ι, which is either in τ or in τ' but not in both. Take an arbitrary $\mathbf{x}$ and flip the component ι, i.e., $x_\iota \to -x_\iota$. In this way the $\mathbf{x}$'s can be paired, the terms of the sum in (2.29) cancel pairwise,

$$\mathscr{S}_{\alpha\tau,\alpha'\tau'} = \left[A_{\alpha\tau}\, \delta_{\alpha\alpha'} - 2^{-q} \sum_{\mathbf{x}} \frac{v_\alpha(\mathbf{x})\, v_{\alpha'}(\mathbf{x})}{1 - [m^{(R)}(\mathbf{x})]^2} \right] \delta_{\tau\tau'} \tag{2.30}$$

and the block-diagonal nature of $\mathscr{S}$ has been established.

Furthermore, since $m^{(R)}(\mathbf{x})$ is supposed to be a solution to the fixed-point equation and the synaptic kernel Q has definite parity, we know from Section 2.2 that $m^{(R)}(\mathbf{x})$ is mapped onto $\pm m^{(R)}(\mathbf{x})$ through the inversion $\mathbf{x} \to -\mathbf{x}$, so that $[m_R(\mathbf{x})]^2$ is invariant. Hence the sum in (2.30) is nonzero only if $|\alpha|$ and $|\alpha'|$ are either both event or both odd. This proves the second half of the assertion.

3. PURE STATES

The pure states $\alpha_\rho v_\rho(\mathbf{x})$, $\rho \subseteq \{1,..., q\}$, have already been introduced in the previous section. For neural network models whose synaptic kernel $Q(\mathbf{x}; \mathbf{y})$ has the invariance property GT (3.1), they are the simplest nontrivial solutions of the transcendental fixed-point equations (2.1). Here we study the physical properties of these states and discuss their relevance to the retrieval of the information embedded in the network.

Given a pure state $\alpha_\rho v_\rho(\mathbf{x})$, the value of the amplitude α_ρ is determined by

$$\alpha_\rho = \tanh(\beta \Lambda_\rho \alpha_\rho) \tag{3.1}$$

[cf. Eq. (2.17)]. As in Eq. (2.24), we have put $\Lambda_\rho = 2^{-q}\lambda_\rho$. According to the bifurcation analysis of Section 2.3, nontrivial solutions of (3.1) only exist for $\lambda_\rho > 0$ and in the temperature range

$$T < T_\rho = \Lambda_\rho \tag{3.2}$$

By virtue of the observation following Eq. (2.26), the pure state $\alpha_\rho v_\rho(\mathbf{x})$ is stable when

$$\beta \Lambda_\sigma - (1 - \alpha_\rho^2)^{-1} < 0 \qquad \text{for all} \quad \sigma \subseteq \{1,..., q\} \tag{3.3}$$

Since the first term on the left-hand side is linear in β and the second term grows exponentially with β, the left-hand side of (3.3) is negative and stability of $\alpha_\rho v_\rho(\mathbf{x})$ is established at sufficiently low temperatures. The onset of stability occurs when

$$\beta \Lambda_{\max} = (1 - \alpha_\rho^2)^{-1} \tag{3.4}$$

where $\Lambda_{\max} = 2^{-q}\lambda_{\max}$ and $\lambda_{\max}$ is the largest (positive) eigenvalue of Q. This equation has to be solved together with the transcendental equation (3.1), leading to a temperature T_ρ^* below which the pure state $\alpha_\rho v_\rho(\mathbf{x})$ is stable. T_ρ^* should not be confused with the temperature T_ρ where the pure state $\alpha_\rho v_\rho(\mathbf{x})$ bifurcates from $m(\mathbf{x}) \equiv 0$.

Only in the special case where $\lambda_\rho = \lambda_{\max}$ and therefore $T_\rho = T_c$ does one find that the two temperatures T_ρ and T_ρ^* coincide. This is easily seen by expanding (3.1) in powers of $t_\rho = (T_\rho - T)/T_\rho$. For $0 < t_\rho \ll 1$ one gets

$$\alpha_\rho^2 = 3t_\rho + \mathcal{O}(t_\rho^2) \tag{3.5}$$

while the stability criterion (3.3) requires

$$(1 + t_\rho)\frac{T_c}{T_\rho} - (1 + 3t_\rho) + \mathcal{O}(t_\rho^2) < 0 \tag{3.6}$$

If $\lambda_\rho = \lambda_{\max}$ or, equivalently, $T_\rho = T_c$, then (3.6) is true for all t_ρ with $0 < t_\rho \ll 1$. On the other hand, if $0 < \lambda_\rho < \lambda_{\max}$ and therefore $T_c/T_\rho > 1$, then (3.6) is violated for small t_ρ, so that the pure states belonging to λ_ρ with $0 < \lambda_\rho < \lambda_{\max}$ are indeed unstable when they first appear.

To interpret the pure states in physical terms, we recall that $m(\mathbf{x})$ represents a local magnetization on the sublattice

$$I(\mathbf{x}) = \{i\colon \xi_i = \mathbf{x}\} \tag{3.7}$$

[cf. GT (2.11)]. Thus, for the pure state $\alpha_\rho v_\rho(\mathbf{x})$ we have

$$m(\mathbf{x}) = \langle S_i \rangle = \alpha_\rho \prod_{\mu \in \rho} x_\mu = \alpha_\rho \prod_{\mu \in \rho} \xi_{i\mu}; \qquad i \in I(\mathbf{x}) \tag{3.8}$$

Here $\langle \cdot \rangle$ denotes a thermal average with respect to the ergodic component associated with the pure state under consideration.

Performing a gauge (Mattis) transformation defined by

$$S_i \to S_i' = \prod_{\mu \in \rho} \xi_{i\mu} S_i \tag{3.9}$$

one finds that the sublattice magnetizations of the transformed pure states are given by

$$\begin{aligned} m'(\mathbf{x}) = \langle S_i' \rangle &= \prod_{\mu \in \rho} x_\mu \langle S_i \rangle \\ &= \alpha_\rho, \qquad i \in I(\mathbf{x}) \end{aligned} \tag{3.10}$$

This reveals that the transformed pure state is ferromagnetic and homogeneous. Since the transformation (3.9) leaves the partition function invariant, pure states are "Mattis states."

It follows from (3.8) that for the pure states $\alpha_\rho v_\rho(\mathbf{x})$ associated with *singletons*, i.e., subsets ρ of $\{1,..., q\}$ with cardinality $|\rho| = 1$, the local magnetization is

$$\langle S_i \rangle = \alpha_\rho \xi_{i\rho} \tag{3.11}$$

Each of these states is correlated with one of the q patterns that have been embedded in the system. Adopting the convention of Amit *et al.*[5] we refer to these states as *retrieval states.* Their retrieval quality is determined by α_ρ and thus by (2.18). The nearer α_ρ is to one, the better is the retrieval quality. As $T \to 0$, α_ρ approaches one at an exponential rate.

The pure states corresponding to eigenvectors v_ρ of Q with $|\rho| > 1$ are not correlated with single patterns. Instead, according to (3.8), they are constructed out of products of several patterns embedded in the network. These states are an *exclusive property of nonlinear models*, since, because of GT (3.13), they are absent in linear neural networks of the Hopfield type.[6-8]

The states with $|\rho| > 1$ do not improve the retrieval of the stored patterns. Quite to the contrary, once they have become stable at low temperatures, they *do* form basins of attraction for the dynamics of the network. At first glance they therefore seem to spoil the usefulness of *nonlinear* models. Fortunately, as q becomes large, it frequently happens that the temperatures T_ρ where these unwanted states appear are shifted arbitrarily close to zero. For instance, the model with clipped synapses gives

$$T_\rho/T_c = \lambda_\rho/\lambda_{\max} \propto \begin{cases} q^{-2} & \text{for } \ q = 4k \qquad \text{or } q = 4k+1 \\ q^{-1} & \text{for } \ q = 4k+2 \quad \text{or } q = 4k+3 \end{cases} \tag{3.12}$$

for the highest temperature T_ρ where an unwanted pure state can bifurcate from zero. Here we have ignored the case $|\rho| = q = 4k+1$ where a state correlated with the product of *all* patterns bifurcates from zero at T_c. At $T = 0$, this state is characterized by

$$S_i = \prod_{\alpha=1}^{q} \xi_{i\alpha} \tag{3.13}$$

It strongly deviates from the original patterns.

On the other hand, one can also regard the states $\alpha_\rho v_\rho(\mathbf{x})$ with $|\rho| > 1$ as virtues of a nonlinear model because they can be associated with nontrivial computational capabilities.[9] Consider, for instance, a network with synaptic kernel of the form

$$\begin{gathered} Q(\xi_i; \xi_j) = A_0 + A_1 \sum_\mu \xi_{i\mu}\xi_{j\mu} + A_2 \sum_{|\rho|=2} v_\rho(\xi_i)\, v_\rho(\xi_j) \\ A_0 = A_2 > 0, \qquad A_1 > 0 \end{gathered} \tag{3.14}$$

For Q given by (3.14), the fixed-point equation (2.1) allows the following types of solution involving *at most* two patterns:

$$\begin{aligned}
&\text{(i)} \quad m(\mathbf{x}) = \alpha_0 \sigma_0 && \text{with} \quad \alpha_0 = \tanh(\beta A_0 \alpha_0) \\
&\text{(ii)} \quad m(\mathbf{x}) = \alpha_\mu \sigma_\mu x_\mu && \text{with} \quad \alpha_\mu = \tanh(\beta A_1 \alpha_\mu) \\
&\text{(iii)} \quad m(\mathbf{x}) = \alpha_{\mu\nu} \sigma_{\mu\nu} x_\mu x_\nu && \text{with} \quad \alpha_{\mu\nu} = \tanh(\beta A_2 \alpha_{\mu\nu}) \\
&\text{(iv)} \quad m(\mathbf{x}) = \alpha_0 \sigma_0 + \alpha_1(\sigma_\mu x_\mu + \sigma_\nu x_\nu) - \alpha_2 \sigma_0 \sigma_\mu \sigma_\nu x_\mu x_\nu \\
&\qquad \text{with} \quad \alpha_0 = \alpha_2 = a, \quad a = \tfrac{1}{2}\tanh(2\beta A_0 a) \\
&\qquad \qquad \alpha_1 = b, \quad b = \tfrac{1}{2}\tanh(2\beta A_1 b)
\end{aligned} \tag{3.15}$$

Here the $\sigma_0, \sigma_\mu, \sigma_\nu$, and $\sigma_{\mu\nu}$ are ± 1 and can be chosen freely. Thus, for each pair (μ, ν) of patterns, Eq. (3.15) stands for 16 different solutions to (2.1). At sufficiently low temperatures these solutions represent truly metastable states of the system and together they implement at $T=0$ the complete set of 16 logical operations that can be performed on the pair (μ, ν). See Table I.

Table I. Logical Operations on the Pair (μ, ν), Represented As Zero-Temperature Solutions of (3.15)[a]

		(i)	(ii)		(iii)	(iv) $\sigma_0\,\sigma_\mu\,\sigma_\nu$			
x_μ	x_ν	$\sigma_0 = +$	$\sigma_\mu = +$	$\sigma_\nu = +$	$\sigma_{\mu\nu} = +$	$+++$	$++-$	$+-+$	$-++$
+	+	+	+	+	+	+	+	+	+
+	−	+	+	−	−	+	+	−	−
−	+	+	−	+	−	+	−	+	−
−	−	+	−	−	+	−	+	+	−
					$\Leftrightarrow$	$\vee$		$\Rightarrow$	$\wedge$

[a] Plus stands for +1 or TRUE, minus for −1 or FALSE. The operations are selected by specifying the type of solution and the values of $\sigma_0, \sigma_\mu, \sigma_\nu$, and $\sigma_{\mu\nu}$ as appropriate. The bottom row presents conventional symbols for some of the operations.

4. SYMMETRIC STATES

In this section we give a complete characterization of *all* solutions of the fixed-point equations that bifurcate from $m \equiv 0$ at the critical temperature $T_c = 2^{-q}\lambda_{\max}$. For nondegenerate eigenvalue $\lambda_{\max}$, the analysis is simple. Only the pure state $\alpha_\rho v_\rho$ with $\lambda_\rho = \lambda_{\max}$ branches off from zero. However, a much richer structure appears when $\lambda_{\max}$ is degenerate, as, for instance, in the case of clipped synapses and the Hopfield model. In fact, in the context of information retrieval in neural network models it is imperative that the stored patterns bifurcate first. For the sake of definiteness we therefore restrict our attention to the case where $\lambda_{\max}$ is q-fold *degenerate* and $\lambda_\rho = \lambda_{\max}$ for all ρ with $|\rho| = 1$, i.e., for all the retrieval states. For this type of model we show that any solution of (2.1) that bifurcates from $m \equiv 0$ at T_c can (up to equivalence) be written as

$$m(\mathbf{x}) = \sum_{\rho \subseteq \mathcal{N}} \alpha_\rho v_\rho(\mathbf{x}) \tag{4.1}$$

where ρ ranges through all the subsets of $\mathcal{N} \subseteq \{1,..., q\}$ and

$$\alpha_\rho = \begin{cases} \alpha_{|\rho|}, & \rho \subseteq \mathcal{N} \\ 0, & \text{otherwise} \end{cases} \tag{4.2}$$

The corresponding states are called *symmetric states* because the right-hand side of (4.1) is invariant under all permutations of the lements of $\mathcal{N}$. The symmetric states of Amit *et al.*[10] constitute a special case of what is considered here.

In passing we note that it is implicitly assumed throughout what follows that the synaptic kernel Q is odd [GT (3.10)] so that α_ρ has to vanish for all ρ with even cardinality $|\rho|$.

Our analysis of the bifurcation phenomena at the *multiple* eigenvalue $\lambda_{\max}$ takes advantage of the Liapunov–Schmidt procedure, a description of which can be found, e.g., in Sattinger.[4] For a first view, the reader may also consult Golubitsky and Schaeffer (Ref. 11, Chapter I, §3, for the method itself, and §4, for stability).

We begin by dividing the set of 2^q fixed-point equations

$$G_\rho(\beta, \boldsymbol{\alpha}) = \alpha_\rho - 2^{-q} \sum_{\mathbf{x}} v_\rho(\mathbf{x}) \tanh\left[\beta \sum_{\sigma} A_\sigma \alpha_\sigma v_\sigma(\mathbf{x})\right] = 0 \tag{4.3}$$

into two parts, one related to the *retrieval amplitudes* α_ρ with $|\rho| = 1$ (to be denoted by a_μ), and the other to the remaining *product-state amplitudes* α_ρ with $|\rho| \geqslant 3$ (to be denoted by b_σ), so that (4.3) reads

$$G_\mu(\beta, \mathbf{a}, \mathbf{b}) = a_\mu - 2^{-q} \sum_{\mathbf{x}} x_\mu \tanh\left[\beta A_1 \sum_{\mu'} a_{\mu'} x_{\mu'} + \beta \sum_{\sigma'} A_{\sigma'} b_{\sigma'} v_{\sigma'}(\mathbf{x})\right] = 0 \tag{4.4a}$$

$$G_\sigma(\beta, \mathbf{a}, \mathbf{b}) = b_\sigma - 2^{-q} \sum_{\mathbf{x}} v_\sigma(\mathbf{x}) \tanh\left[\beta A_1 \sum_{\mu'} a_{\mu'} x_{\mu'} + \beta \sum_{\sigma'} A_{\sigma'} b_{\sigma'} v_{\sigma'}(\mathbf{x})\right] = 0 \tag{4.4b}$$

We look for solutions of (4.4) that become small as $T \to T_c = \Lambda_1$ from below (or equivalently $\beta\Lambda_1 - 1 \to 0$ from above). The Liapunov–Schmidt procedure[4] consists of three main steps.

First, consider (4.4b) and define D to be the matrix with elements

$$D_{\sigma\sigma'} = \left.\frac{\partial G_\sigma(\beta, \mathbf{a}, \mathbf{b})}{\partial b_{\sigma'}}\right|_{\mathbf{a} = \mathbf{b} = \mathbf{0}} = \delta_{\sigma\sigma'}(1 - \beta\Lambda_\sigma)$$

Near $\beta = \beta_c$, the matrix D is nonsingular, since $\beta_c \Lambda_\sigma \neq 1$ for all σ with $|\sigma| \neq 1$. By the implicit function theorem, (4.4b) defines a function $\mathbf{b}(\beta, \mathbf{a})$ which is locally unique and analytic in β and $\mathbf{a}$ in a neighborhood of $\beta\Lambda_1 - 1 = 0$ and $\mathbf{a} = \mathbf{0}$. Moreover, by using the fact that $G_\rho(\beta, \boldsymbol{\alpha} = \mathbf{0}) = 0$ for all β and that $G_\rho(\beta_c, \boldsymbol{\alpha})$ is not identically zero in a neighborhood of $\boldsymbol{\alpha} = \mathbf{0}$, general theory (Ref. 4, Lemma 4.1) tells us that the zeroth- and first-order terms in a series expansion of $\mathbf{b}(\beta, \mathbf{a})$ in powers of $\beta\Lambda_1 - 1$ and $\mathbf{a}$ vanish. Therefore, the expansion of $\mathbf{b}(\beta, \mathbf{a})$ can be written

$$\mathbf{b}(\beta, \mathbf{a}) = \sum_{k=2}^{\infty} \sum_{i+j=k} \mathbf{b}_{ij}(\mathbf{a})(\beta\Lambda_1 - 1)^j \tag{4.5}$$

where $\mathbf{b}_{ij}$ is homogeneous of degree i, i.e., $\mathbf{b}_{ij}(\tau\mathbf{a}) = \tau^i \mathbf{b}_{ij}(\mathbf{a})$.

Second, inserting the functional relation $\mathbf{b} = \mathbf{b}(\beta, \mathbf{a})$ into (4.2a), one obtains

$$\begin{aligned} F_\mu(\beta, \mathbf{a}) &:= G_\mu(\beta, \mathbf{a}, \mathbf{b}(\beta, \mathbf{a})) \\ &= a_\mu - 2^{-q} \sum_{\mathbf{x}} x_\mu \tanh\left[\beta\Lambda_1 \sum_{\mu'} a_{\mu'} x_{\mu'} + \beta \sum_{\sigma} \Lambda_\sigma b_\sigma(\beta, \mathbf{a})\, v_\sigma(\mathbf{x})\right] \end{aligned} \tag{4.6}$$

The equations

$$F_\mu(\beta, \mathbf{a}) = 0, \qquad 1 \leqslant \mu \leqslant q \tag{4.7}$$

constitute a set of q fixed-point equations for the retrieval amplitudes alone. In this way, one has reduced the 2^q-dimensional problem (4.3) to a q-dimensional one, namely (4.7).

The third step consists in solving (4.7) to lowest order in $\beta\Lambda_1 - 1$ by means of a so-called *reduced bifurcation equation.*[4] To obtain this equation, one expands (4.6) in a power series in $\beta\Lambda_1 - 1$ and $\mathbf{a}$, which can be written

$$F_\mu(\beta, \mathbf{a}) = \sum_{k=1}^{\infty} \sum_{i+j=k} F_{\mu ij}(\mathbf{a})(\beta\Lambda_1 - 1)^j \tag{4.8}$$

where $F_{\mu ij}$ is homogeneous of degree i. In the present case, first expanding the hyperbolic tangent in (4.6), one obtains

$$\begin{aligned} F_\mu(\beta, \mathbf{a}) = & -a_\mu(\beta A_1 - 1) + \tfrac{1}{3}(\beta A_1)^3 \sum_{\mu_1\mu_2\mu_3} a_{\mu_1} a_{\mu_2} a_{\mu_3}\, \delta_{\mu,\mu_1\circ\mu_2\circ\mu_3} \\ & + \tfrac{1}{3}\beta(\beta A_1)^2 \sum_{\mu_1\mu_2\sigma} a_{\mu_1} a_{\mu_2} A_\sigma b_\sigma\, \delta_{\mu,\mu_1\circ\mu_2\circ\sigma} \\ & + \tfrac{1}{3}\beta^2(\beta A_1) \sum_{\mu_1\sigma_1\sigma_2} a_{\mu_1} A_{\sigma_1} b_{\sigma_1} A_{\sigma_2} b_{\sigma_2} \delta_{\mu,\mu_1\circ\sigma_1\circ\sigma_2} \\ & + \tfrac{1}{3}\beta^3 \sum_{\sigma_1\sigma_2\sigma_3} A_{\sigma_1} b_{\sigma_1} A_{\sigma_2} b_{\sigma_2} A_{\sigma_3} b_{\sigma_3} \delta_{\mu,\sigma_1\circ\sigma_2\circ\sigma_3} + \cdots \end{aligned} \tag{4.9}$$

Here we have used the orthogonality of the $v_\rho(\mathbf{x})$ [see GT (3.7)] and the fact that

$$v_{\sigma_1}(\mathbf{x})\, v_{\sigma_2}(\mathbf{x}) \cdots v_{\sigma_k}(\mathbf{x}) = v_{\sigma_1\circ\sigma_2\circ\cdots\circ\sigma_k}(\mathbf{x}) \tag{4.10}$$

where $\sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_k$ denotes the set of those elements of $\sigma_1,\ldots, \sigma_k$ that occur an *odd* number of times in the list $\sigma_1,\ldots, \sigma_k$. One now inserts the identity $\beta A_\sigma = [1 + (\beta A_1 - 1)]\, A_\sigma / A_1$ and the power series expansion (4.5) of $\mathbf{b}(\beta, \mathbf{a})$ into (4.9) and collects the lowest order terms of (4.8), which are

$$\begin{aligned} F_{\mu 11}(\mathbf{a}) &= -a_\mu \\ F_{\mu 30}(\mathbf{a}) &= \frac{1}{3} \sum_{\mu_1\mu_2\mu_3} a_{\mu_1} a_{\mu_2} a_{\mu_3}\, \delta_{\mu,\mu_1\circ\mu_2\circ\mu_3} \\ &= \frac{1}{3}\left(a_\mu^3 + 3a_\mu \sum_{\mu' \neq \mu} a_{\mu'}^2\right) \end{aligned} \tag{4.11}$$

By means of a device called the *Newton diagram*,[4] one introduces a new scaling of the variables $t := \beta A_1 - 1$ and $\mathbf{a}$ in (4.8) as follows. One plots every point (i, j) on a lattice of nonnegative integers for which $F_{\mu ij}$ in the expansion (4.8) does *not* vanish. From (4.11), two of these points are $(3, 0)$ and $(1, 1)$. Moreover, it follows from (4.5) and (4.9) that all other such points, as seen from the origin $(0, 0)$, lie beyond the line through the points $(3, 0)$ and $(1, 1)$; the equation describing this line is

$$li + mj = r \qquad \text{with} \quad l = 1, m = 2, r = 3 \tag{4.12}$$

This corresponds precisely to Fig. 4.1 of Sattinger.[4] The expansion (4.8) may therefore be rewritten

$$F_\mu(\beta, \mathbf{a}) = \sum_{k=0}^{\infty} \sum_{li+mj=r+k} F_{\mu ij}(\mathbf{a})(\beta A_1 - 1)^j \tag{4.13}$$

with l, m, and r given by (4.12). We now rescale the variables $\mathbf{a}$ and t by introducing a parameter ε, which is supposed to become small, and setting

$$\mathbf{a} = \varepsilon^l \boldsymbol{\zeta}, \qquad t = \beta A_1 - 1 = \varepsilon^m \tau \tag{4.14}$$

so that

$$\begin{aligned} F_\mu(\varepsilon^m \tau, \varepsilon^l \boldsymbol{\zeta}) &= \sum_{k=0}^{\infty} \varepsilon^{r+k} \sum_{li+mj=r+k} F_{\mu ij}(\boldsymbol{\zeta}) \tau^j \\ &= \varepsilon^r R_\mu(\boldsymbol{\zeta}, \tau) + \varepsilon^{r+1} g_\mu(\boldsymbol{\zeta}, \tau, \varepsilon) \end{aligned} \tag{4.15}$$

The *reduced bifurcation equations* are then obtained from (4.7) and (4.15) by dividing by ε^r, taking the limit $\varepsilon \to 0$, and using (4.11),

$$R_\mu(\boldsymbol{\zeta}, \tau) = -\zeta_\mu \tau + \frac{1}{3}\left(\zeta_\mu^3 + 3\zeta_\mu \sum_{\mu' \neq \mu} \zeta_{\mu'}^2\right) = 0 \tag{4.16}$$

An immediate consequence of Eq. (4.16) is

$$\tfrac{2}{3}\zeta_\mu^2 = \zeta^2 - \tau \tag{4.17}$$

for every nonzero ζ_μ, so that the nonzero ζ_μ^2 *are all equal.* If one denotes by n the number of nonzero ζ_μ in a solution of (4.17), then one finds

$$\zeta_\mu^2 = \tau/(n - 2/3) \tag{4.18}$$

In view of (4.14) and (4.18) we can put τ equal to one, since, up to the sign, τ may be absorbed into the parameter ε. Obviously, the indices and the signs of the nonzero components of $\boldsymbol{\zeta}$ are at our disposal. Given n, the free choice of the indices reflects the invariance of the fixed-point equations with respect to relabeling the components of $\mathbf{x} \in \mathscr{C}^q$. The free choice of the signs is a manifestation of the covariance property presented in Section 2.1.

For small but nonzero ε, Eq. (4.18) provides a solution of (4.7) to the lowest order in ε. Using (4.14), we can express this solution in terms of the original variables a_μ and t as

$$\begin{aligned} a_\mu &= \varepsilon^l \zeta_\mu = t^{l/m} \zeta_\mu \\ &= \left(\frac{t}{n - 2/3}\right)^{1/2} \end{aligned} \tag{4.19}$$

Here we have used our freedom of choosing positive signs in (4.19). Furthermore, we can assume without loss of generality that $\mu \in \mathscr{N} = \{1, ..., n\}$, with $n \leqslant q$. We note that the solution (4.19) is invariant under all permutations of the elements of $\mathscr{N}$ in that a_μ does not depend on μ.

One easily verifies that the Ansatz (4.1)–(4.2) is consistent with the fixed-point equation (4.3). Conversely, starting with (4.19), one can expand the $b_\sigma(\beta, \mathbf{a})$ in (4.9) in an ascending power series with respect to $\mathbf{a}$ and $(\beta A_1 - 1)$, use the ideas[4] associated with (4.14), and compare the terms order by order so as to show that in any order the permutation symmetry is not broken. Some details are given in the Appendix. This then proves (4.1)–(4.2).

Once we know that only symmetric states can bifurcate from $m \equiv 0$ at $T = T_c$, the next question is which of these states is stable or at least metastable. We will show that just below T_c the retrieval states are the *only* states that are stable. All the other symmetric states are *un*stable. For the proof we will exploit the information contained in the reduced bifurcation equation (4.16) and the fact (Ref. 4, Section 4.3, particularly Theorem 4.3) that the stability of a bifurcating solution is determined by the Jacobian of $R(\zeta, \tau)$. To understand the ensuing arguments, we must make a small detour, however.

According to (2.22), a phase corresponding to a solution $m(\mathbf{x})$ of the fixed-point equation (2.1) is stable if the matrix

$$\mathscr{S} = \beta 2^{-q} Q - \operatorname{diag}\{[1 - m^2(\mathbf{x})]^{-1}\} \tag{4.20}$$

has negative eigenvalues only. Here $\operatorname{diag}[d(\mathbf{x})]$ is the diagonal matrix with elements $d(\mathbf{x})$, the $\mathbf{x}$ labeling the 2^q corners of the hypercube $[-1, 1]^q$. We now want to relate $\mathscr{S}$ to the fixed-point equation. To this end, we rewrite (2.1),

$$G(\beta, m(\mathbf{x})) = m(\mathbf{x}) - \tanh\left[\beta 2^{-q} \sum_{\mathbf{y}} Q(\mathbf{x}; \mathbf{y})\, m(\mathbf{y})\right] = 0 \tag{4.21}$$

and show[12] that $\mathscr{S}$ is negative-definite if and only if $DG(\beta, m(\mathbf{x}))$, the derivative of G at $(\beta, m(\mathbf{x}))$, has *positive* eigenvalues only. Plainly, both $\mathscr{S}$ and DG are $2^q \times 2^q$ matrices and

$$DG(\beta, m(\mathbf{x})) = \mathbb{1} - \operatorname{diag}[1 - m^2(\mathbf{x})]\, \beta 2^{-q} Q \tag{4.22}$$

Let Y be a diagonal matrix with nonzero elements and let X be an arbitrary, real, symmetric matrix. Then the congruent matrices[13] X and YXY have the same number of positive eigenvalues and the same number of negative eigenvalues. For X we choose $DG(\beta, m(\mathbf{x}))$, put $Y = \operatorname{diag}\{[1 - m^2(\mathbf{x})]^{-1/2}\}$, and note that YXY and $Y^2 X$ have the same eigenvalues. Hence DG and $\operatorname{diag}\{[1 - m^2(\mathbf{x})]^{-1}\} DG = -\mathscr{S}$ have the same number of positive eigenvalues and the same number of negative eigenvalues. Q.E.D.

How to exploit the above observation, which leads from $\mathcal{S}$ to $DG(\beta, m(\mathbf{x}))$? In the mathematical literature on bifurcation—and that is after all what we are interested in here—one frequently encounters[4,11,14] the notion of stability, which means the following. One wants to determine the asymptotic behavior of the solutions of a system of ordinary differential equations, $\dot{\mathbf{x}} + G(\lambda, \mathbf{x}) = 0$, where $\mathbf{x}$ is a vector in $\mathbb{R}^p$, say, λ is a bifurcation parameter, and G maps $\mathbb{R} \times \mathbb{R}^p$ into $\mathbb{R}^p$. An "equilibrium point" $\mathbf{x}$ satisfies $G(\lambda, \mathbf{x}) = 0$. It is stable if $DG(\lambda, \mathbf{x})$ has positive eigenvalues only and it is unstable if $DG(\lambda, \mathbf{x})$ has at least one negative eigenvalue (principle of linearized stability). In the present context, $p = 2^q$, $\mathbf{x}$ has to be replaced by $m(\mathbf{x})$, Eq. (4.21) tells us that $G(\lambda, \mathbf{0}) = 0$ whatever $\lambda = \beta \Lambda_1 - 1$, and we want to determine the stability of a solution that bifurcates from $m \equiv 0$. We have seen that this kind of stability coincides with the thermodynamic stability of the phase associated with $m(\mathbf{x})$. Now it is shown in Sattinger (Ref. 4, Section 4.3) that the stability of a bifurcating solution is determined by the eigenvalues of the Jacobian of the *reduced* bifurcation equation (4.16), i.e., the Jacobian of $R(\zeta)$, the dependence upon τ being suppressed. This remarkable result greatly simplifies the stability analysis, since the dimensionality of (4.16) is at most q, whereas (2.1) refers to 2^q equations.

According to the above discussion, thermodynamic stability is lost as soon as one of the eigenvalues of $DR(\zeta)$ is negative. By virtue of (4.17) we know that $\frac{2}{3}\zeta_\mu^2 = \zeta^2 - 1$ and a simple calculation then shows

$$(DR)_{\mu\mu} = \tfrac{2}{3}\zeta_\mu^2; \qquad (DR)_{\mu\nu} = 2\zeta_\mu \zeta_\nu = 2\zeta_\mu^2, \qquad \mu \neq \nu \tag{4.23}$$

where, by (4.18), $\zeta_\mu^2 = (n - \frac{2}{3})^{-1}$ does not depend on μ. We may rewrite (4.23)

$$DR(\zeta) = \zeta_\mu^2 [2 \cdot \mathbf{1} - \tfrac{4}{3} \cdot \mathbb{1}] \tag{4.24}$$

where $\mathbb{1}$ is the unit matrix and $\mathbf{1}$ has all matrix elements equal to one. Given n, the latter has $(n-1)$ eigenvalues zero and a simple eigenvalue n. If $n = 1$, $DR(\zeta)$ has a single, positive eigenvalue and the bifurcating solution is stable. This confirms the result of the previous section that the retrieval states are stable. However, for $n > 1$ the bifurcating solutions are all unstable, at least near T_c, since $DR(\zeta)$ has $n-1$ negative eigenvalues $-\frac{4}{3}\zeta_\mu^2 = -4/(3n-2)$. As will be shown in the next section, this result does not preclude that some of the symmetric states may become (meta)stable at a lower temperature.

The method we have presented in this section also works for clipped synapses with $q = 4k + 1$. Then the largest eigenvalue $\lambda_{\max}$ is $(q+1)$-fold degenerate and one has to take care of an extra product state; cf. (3.13). The method works equally well for all bifurcations at temperatures $T_\rho = 2^{-q}\lambda_\rho < T_c$, with $|\rho| \neq 1$. Since we are mainly interested in the retrieval states, we will not touch upon this type of problem, however.

5. CLIPPED SYNAPSES: STABILITY ANALYSIS OF SYMMETRIC STATES

As they appear, n-symmetric states with $n > 1$ are unstable. One may wonder, though, whether they ever become stable, and, if so, at what temperature. Taking advantage of the general properties of nonlinear neural networks derived in Sections 2–4, we now focus our attention on the special case of the synaptic function $\phi(x) = \text{sgn}(x)$. The corresponding neural network is known in the literature as a model of "clipped synapses," where the synaptic strengths are only allowed to take the values ± 1 and 0, so that this model is fully digitized and thus may be implemented more easily in a silicon version. It will be shown that, as q becomes large, the bifurcation and stability structure of the symmetric states reduces to that of the (linear) Hopfield model with the same number of patterns. Since the general proof is rather laborious, we will concentrate on $2 \leqslant n \leqslant 5$. At the end of this section a general argument is provided that n-symmetric states with n even are *always* unstable, whenever ϕ is odd.

We investigate the stability of symmetric solutions to the fixed-point equation (2.1), i.e., symmetrically built mixture states of the form

$$m_n(\mathbf{x}) = \sum_{\rho \subseteq \{1,\ldots,n\}} a_{|\rho|} v_\rho(\mathbf{x}) \tag{5.1}$$

with $n \leqslant q$. Here we have restricted ourselves, without any loss of generality, to solutions in which only the first n components of $\mathbf{x}$ are involved.

Inserting (5.1) into the fixed-point equation (2.1) and projecting onto the ρth eigenvector $v_\rho(\mathbf{x})$, one gets

$$a_{|\rho|} = 2^{-q} \sum_{\mathbf{x}} v_\rho(\mathbf{x}) \tanh\left[\beta \sum_{\alpha} \Lambda_\alpha a_{|\alpha|} v_\alpha(\mathbf{x})\right] \tag{5.2}$$

where Λ_α is defined by (2.24), i.e., $\Lambda_\alpha = 2^{-q}\lambda_\alpha$. Throughout what follows, it is to be noted that by the general results of Section 3.2 the Λ_α or, in the usual notation, Λ_ρ only depend on the *size* $|\rho|$ of the set ρ. Below we therefore write Λ_i with $i = |\rho|$. By virtue of the general theorem in Section 2.4 the stability matrix $\mathscr{S}$ is block-diagonal with blocks of size 2^{n-1}, so that we are faced with the problem of singling out the largest eigenvalue from 2^{q-n+1} blocks, each of dimension 2^{n-1}, and to determine its change of sign (if any) as the temperature decreases. As soon as the largest eigenvalue of $\mathscr{S}$ is negative, the symmetric state corresponding to $\{a_\rho\}$ is metastable.

To illustrate what symmetric solutions look like, we start by giving two simple examples:

$$m_2(\mathbf{x}) = a_1(x_1+x_2) + a_2 x_1 x_2 \tag{5.3}$$

$$m_3(\mathbf{x}) = a_1(x_1+x_2+x_3) + a_2 \sum_{1 \leqslant i_1 < i_2 \leqslant 3} x_{i_1} x_{i_2} + a_3 x_1 x_2 x_3 \tag{5.4}$$

The corresponding fixed-point equations for the amplitudes are, for $n=2$,

$$a_1 = \tfrac{1}{2}\tanh(2\beta a_1 A_1), \qquad a_2 = 0 \tag{5.5}$$

and for $n=3$,

$$\begin{aligned} 3a_1 + a_3 &= \tanh[\beta(3a_1 A_1 + a_3 A_3)] \\ a_1 - a_3 &= \tanh[\beta(a_1 A_1 - a_3 A_3)] \\ a_2 &= 0 \end{aligned} \tag{5.6}$$

The vanishing of the amplitudes a_2 in both cases is not surprising. Quite generally, it was shown in Section 2.2 that $a_{|\rho|}=0$ for $|\rho|$ even.

Sections 5.1–5.3 are devoted to a stability analysis of n-symmetric solutions with $2 \leqslant n \leqslant 5$. In Section 5.4 it is shown that n-symmetric solutions with n even are always unstable, whenever the synaptic function ϕ is odd.

5.1. Two-Symmetric States

By the theorem of Section 2.4, the stability matrix is block-diagonal with blocks of dimension 2, so that its eigenvalues μ_ρ are solutions of quadratic equations. We only quote the results:

$$\begin{aligned} \mu_{1,2} = \tfrac{1}{2}\{&\beta A_{|\tau|} + A_{2+|\tau|}) - (c+1) \\ &\pm [\beta^2(A_{|\tau|} - A_{2+|\tau|})^2 + (c-1)^2]^{1/2}\} \end{aligned} \tag{5.7a}$$

$$\mu_3 = \beta A_{1+|\tau|} - 1, \qquad \mu_4 = \beta A_{1+|\tau|} - c \tag{5.7b}$$

Here $\tau \subseteq \{3, 4,..., q\}$ and $c = 1/(1-4a_1^2)$. Stability is ruined by the existence of the third eigenvalue μ_3, which for the choice $\tau = \varnothing$ is positive for all temperatures $T \leqslant T_1 = T_c$; cf. (2.15). Hence, the 2-symmetric solution is always *un*stable. But it should be noted that *secondary* bifurcations are possible if one of the remaining eigenvalues vanishes. This follows from

Section 2.3 and the stability analysis of Section 4. The highest temperature where such a bifurcation can occur is defined by $\mu_1(|\tau|=1)=0$, leading to a temperature $T_2^*(q)$ that approaches the Hopfield value[10] $T_2=0.574419\ldots$ in the limit $q\to\infty$, as is illustrated in Fig. 1. To obtain this number, we have rescaled the temperature so as to get $T_c=1$.

5.2. Three-Symmetric States

For the 3-symmetric solutions of (5.2) the block-diagonal stability matrix consists of 8×8 matrices of the form

$$B_{|\tau|}=$$

$$\left[\begin{array}{cccc|cccc}
A_{|\tau|}+P_0 & P_2 & P_2 & P_2 & & & & \\
P_2 & A_{|\tau|+2}+P_0 & P_2 & P_2 & & 0 & & \\
P_2 & P_2 & A_{|\tau|+2}+P_0 & P_2 & & & & \\
P_2 & P_2 & P_2 & A_{|\tau|+2}+P_0 & & & & \\
\hline
 & & & & A_{|\tau|+1}+P_0 & P_2 & P_2 & P_2 \\
 & & & & P_2 & A_{|\tau|+1}+P_0 & P_2 & P_2 \\
 & 0 & & & P_2 & P_2 & A_{|\tau|+1}+P_0 & P_2 \\
 & & & & P_2 & P_2 & P_2 & A_{|\tau|+3}+P_0
\end{array}\right] \tag{5.8}$$

Here τ runs through all subsets of $\{4, 5,\ldots, q\}$ and the quantities P_0 and P_2 denote the following two types of sum over all vectors $\mathbf{x}$:

$$P_0=-2^{-q}\sum_{\mathbf{x}}\frac{1}{1-m_3(\mathbf{x})^2} \tag{5.9a}$$

$$P_2=-2^{-q}\sum_{\mathbf{x}}\frac{x_i x_j}{1-m_3(\mathbf{x})^2},\qquad 1\leqslant i<j\leqslant 3 \tag{5.9b}$$

with $m_3(\mathbf{x})$ being defined by Eq. (5.4). The $|\tau|$th block $B_{|\tau|}$ occurs

$$\binom{q-3}{|\tau|}$$

times, so that the $B_{|\tau|}$'s constitute the full matrix of second derivatives $(\mathscr{S}_{\sigma\sigma'})$, which is 2^q-dimensional.

The eigenvalues of the matrices $B_{|\tau|}$ are easily calculated if one uses the fact that the determinant of the matrix[15]

$$M=\begin{pmatrix}
\lambda_1 & & & \mu \\
 & \lambda_2 & & \\
 & & \lambda_3 & \\
\mu & & & \lambda_4
\end{pmatrix} \tag{5.10}$$

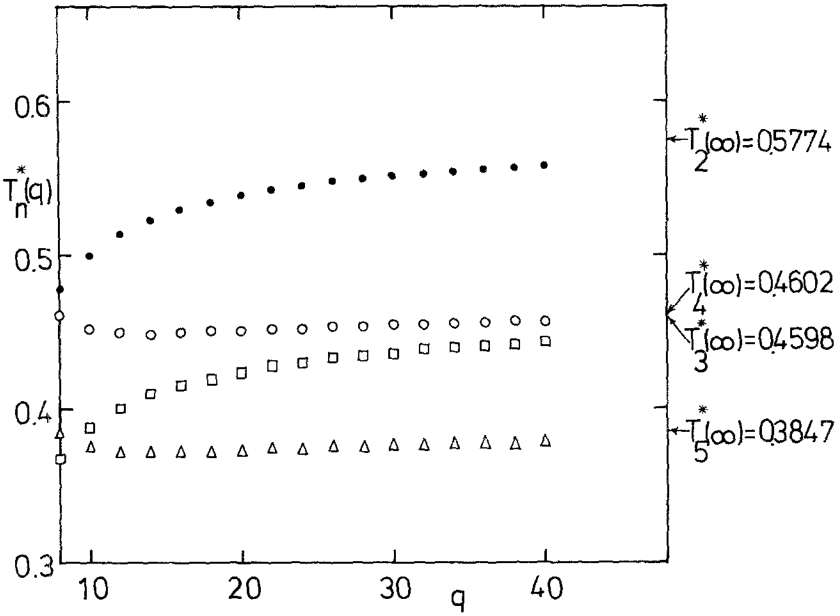

Fig. 1. Bifurcation temperatures for n-symmetric states in a network of clipped synapses, with $2 \leqslant n \leqslant 5$, as a function of the number q of embedded patterns. Through secondary bifurcations, the 3- and 5-symmetric states acquire metastability below $T^*_3(q)$ and $T^*_5(q)$, marked by open squares and open triangles, respectively. The temperatures $T^*_2(q)$ and $T^*_4(q)$, marked by full and open circles, denote the highest temperature where secondary bifurcations from the 2- and 4-symmetric states occur. These states, however, remain unstable as T is lowered through $T^*_2(q)$ and $T^*_4(q)$, respectively.

is simply given by

$$\det M = f(\mu) - \mu f'(\mu) \tag{5.11}$$

where

$$f(\mu) = \prod_{k=1}^{4} (\lambda_k - \mu) \tag{5.12}$$

Some simple algebra, facilitated by the use of (5.11), and a straightforward but tedious procedure of counting then leads to the following three groups of eigenvalues (which are labeled in an obvious way).

(a)

$$\mu(\pm) = P_0 + (1 \pm 2)P_2 \tag{5.13a}$$

with multiplicity $(2-s)2^{q-3}$, $s = \pm 1$.

(b)

$$\mu(i)=\beta A_i+P_0-P_2, \qquad i=1,3,\ldots,q-2 \tag{5.13b}$$

with multiplicity

$$2\binom{q-2}{i-1}$$

(c)

$$\begin{aligned}\mu(i;s_1,s_2)={}&P_0+P_2+\beta(A_i+A_{i+2})/2\\&+s_1[4P_2^2+s_2\beta P_2(A_i-A_{i+2})+\beta^2(A_i-A_{i+2})^2/4]^{1/2}\\&i=1,3,\ldots,q-3+s_2\end{aligned} \tag{5.13c}$$

where $s_1, s_2=\pm 1$ and the (s_2-dependent) multiplicity is

$$\binom{q-3}{i-(s_2+1)/2}$$

Substitution of the 3-symmetric solution (5.4) into (5.9) yields for the linear combinations of P_0 and P_2 appearing in Eq. (5.13) the following expressions:

$$\begin{aligned}P_0+P_2&=-\tfrac{1}{2}(\cosh^2 A+\cosh^2 B)\leqslant 0\\P_0-P_2&=-\cosh^2 B\leqslant 0\\P_0+3P_2&=-\cosh^2 A\leqslant 0\end{aligned} \tag{5.14}$$

where A and B stand for

$$A=3\beta A_1 a_1+\beta A_3 a_3, \qquad B=\beta A_1 a_1-\beta A_3 a_3 \tag{5.15}$$

From (5.14) it follows at once that $\mu(\pm)$ is less than zero for all temperatures and all q, so that the dangerous eigenvalues of the stability matrix are those in the groups (b) and (c). Furthermore, since $P_1-P_2\leqslant 0$ and $A_i<0$ for $i=3, 7,\ldots, 4k-1,\ldots$, we have $\mu(1)\geqslant\mu(i)$ for all i. As will be shown in the discussion below, $\mu(1)$ is even larger than any eigenvalue of group (c), so that $\mu(1)$ is the most dangerous eigenvalue. It therefore governs the stability of a 3-symmetric solution.

The proof can be decomposed into a finite number of easy steps. First we show that $P_2\leqslant 0$ for all temperatures $T\leqslant T_1=T_c$. This is easily seen if one uses (5.14) to establish

$$P_2=-\tfrac{1}{4}(\cosh^2 A-\cosh^2 B) \tag{5.16}$$

where A and B are defined by (5.15). Since $A_3 < 0$ and the amplitudes a_1 and a_3 are given by $a_1 = -a_3 = \frac{1}{2}$ at zero temperature, which can be proved with the help of Eq. (5.6), we have $A \geqslant B$ at least at $T = 0$. But this last inequality even holds for all nonzero temperatures $T < T_1$, since $a_3 = 0$ would imply $a_1 = 0$, which contradicts the fact that the zero solution has been found to be unstable for $T < T_1$. Then, combining (5.16) and the fact that $A \geqslant B$ for all $T \leqslant T_1$, we finally obtain $P_2 \leqslant 0$, as advertised.

Now the two cases (i) $A_i - A_{i+2} < 0$ and (ii) $A_i - A_{i+2} \geqslant 0$ must be studied separately. Let us begin with case (i), where the signs of $(A_i - A_{i+2})$ and P_2 lead to the obvious inequality $\mu(i; ++) \geqslant \mu(i; +-) \geqslant \mu(i; -s_2)$ $(s_2 = \pm 1)$, so that the eigenvalues of group (c), which will enter the competition with the only candidate of group (b), namely $\mu(1)$, are the eigenvalues $\mu(i; ++)$ with i odd. Calculating the difference

$$\begin{aligned}\mu(1) - \mu(i; ++) = {} & \beta(A_1 - \tfrac{1}{2}A_{i+2}) - [\tfrac{1}{2}\beta(A_i - A_{i+2}) + 2P_2] \\ & - \{[\tfrac{1}{2}\beta(A_i - A_{i+2}) + 2P_2]^2 - P_2\beta(A_i - A_{i+2})\}^{1/2}\end{aligned} \tag{5.17}$$

we find that the right-hand side of this equation is positive, so that $\mu(1) \geqslant \mu(i; ++)$. Case (ii), $A_i - A_{i+2} \geqslant 0$, is easier to handle due to the obvious relation

$$\mu(i) \geqslant \mu(i; +-) \geqslant \mu(i; ++) \geqslant \mu(i; -s_2), \qquad s_2 = \pm 1 \tag{5.18}$$

from which we can deduce immediately that $\mu(1)$ is the largest eigenvalue, since $\mu(1) \geqslant \mu(i)$ for all i.

Having singled out the largest eigenvalue of the stability matrix, we find that stability of the symmetric combination $m_3(\mathbf{x})$ is guaranteed if $\mu(1) < 0$. The condition for the onset of stability then is

$$\beta A_1\{1 - \tanh^2[\beta(A_1 a_1 - A_3 a_3)]\} = 1 \tag{5.19}$$

which has to be solved together with the transcendental equations (5.6) for the amplitudes a_1 and a_3. The numerical solution of these equations yields $T_3^*(q)$; see Fig. 1. Note that $T_3^*(q)$ approaches $T_3^{\text{Hopfield}} = 0.4598$ as $q \to \infty$. This is not surprising, since in the limit of large q the condition (5.19) reduces to

$$\beta^{-1} = 1 - \tanh^2(\beta A_1 a_1)$$

which is just the result of Amit *et al.*[10] for the Hopfield case, except for a trivial rescaling of the temperature so as to put $T_c = 1$.

5.3. Four- and Five-Symmetric Solutions

The analysis for the 4-, 5- (and higher) symmetric solutions proceeds similarly. Exploiting the block-diagonal nature of the stability matrix $\mathscr{S}$ established in Section 2.4, we find the following.

The 4-symmetric solution is always unstable, since $\mathscr{S}$ has a positive eigenvalue in its odd $|\alpha|$, $|\tau|=0$ block [cf. Eqs. (2.28), (2.29)] for all $T \leqslant T_c$. *Secondary* bifurcations are possible where the 4-symmetric solution changes its relative stability (though it remains unstable). The highest temperature where this occurs is related to an eigenvalue in an even $|\alpha|$, $|\tau|=1$ block of $\mathscr{S}$ changing sign from negative to positive as T is lowered through $T_4^*(q)$, which approaches $T_4^{\text{Hopfield}}=0.4602^{(10)}$ as $q \to \infty$. This is illustrated in Fig. 1.

The 5-symmetric solution is unstable just below T_c. It acquires stability at $T_5^*(q)$, where the largest eigenvalue (in an even $|\alpha|$, $|\tau|=1$ block) of $\mathscr{S}$ becomes negative. Again, as is illustrated in Fig. 1, $T_5^*(q)$ approaches $T_5^{\text{Hopfield}}=0.3847^{(10)}$ as $q \to \infty$. In both cases the temperature has been rescaled so as to get $T_c=1$.

5.4. Instability of Symmetric States Which Are Even

In Sections 5.2 and 5.3 the 2- and 4-symmetric states were shown to be unstable at any temperature, a result stronger than that of Section 4, where we studied the stability just below T_c. Here we show that, whenever ϕ is odd, *all* n-symmetric solutions with *n even* are unstable at low temperatures (and thus, in view of the result of Section 4, presumably throughout the whole temperature regime $0 \leqslant T \leqslant T_c$). To this end, we return to the stability matrix $\mathscr{S}$ in its original form (2.22),

$$\mathscr{S}_{\mathbf{x},\mathbf{y}}=\beta 2^{-q}\phi(\mathbf{x}\cdot\mathbf{y})-\delta_{\mathbf{x},\mathbf{y}}[1-m^2(\mathbf{x})]^{-1} \tag{5.20}$$

which, for metastable solutions $m(\mathbf{x})$, is required to be negative-definite. We focus our attention on the diagonal elements ($\mathbf{x}=\mathbf{y}$):

$$\beta 2^{-q}\phi(q)-[1-m^2(\mathbf{x})]^{-1} \tag{5.21}$$

and ask whether there exists an $\mathbf{x}$ with $m(\mathbf{x})=0$ for all T, that is, we look for sublattices $I(\mathbf{x})=\{i\colon \xi_i=\mathbf{x}\}$ *that do not order at any temperature.* Once we have established the existence of an $\mathbf{x}$ with $m(\mathbf{x})=0$ for all T, we find that the corresponding diagonal element in the stability matrix is *positive* for all β obeying the inequality

$$\beta\phi(q)>2^q \tag{5.22}$$

This contradicts the stability requirement that $\mathscr{S}$ be negative-definite, and hence $\mathscr{S}_{\mathbf{x},\mathbf{x}} < 0$.

To show that such an $\mathbf{x}$ exists, it is sufficient to prove that the system of equations

$$\sum_{\substack{\rho \subseteq \{1,\ldots,n\} \\ |\rho| = k}} \prod_{\nu \in \rho} x_\nu = 0, \qquad k = 1, 3,\ldots, n-1 \tag{5.23}$$

has a solution in $\mathscr{C}^q = \{-1, 1\}^q$ since, irrespective of the values of the symmetric amplitudes $a_\rho = a_{|\rho|} = a_k$, $k = 1, 3,\ldots, n-1$, the magnetization

$$m(\mathbf{x}) = \sum_{k=1,3,\ldots,n-1} a_k \sum_{\substack{\rho \subseteq \{1,\ldots,n\} \\ |\rho| = k}} \prod_{\nu \in \rho} x_\nu \tag{5.24}$$

vanishes on all sublattices $I(\mathbf{x})$ for which the vector $\mathbf{x} \in \mathscr{C}^q$ satisfies Eqs. (5.23). For $k = 1$, we infer that the number of x_i with $x_i = +1$ must equal the number of x_j with $x_j = -1$ (which can be accomplished only if n is even). Moreover, we note that a permutation of the first n components of $\mathbf{x}$ leaves Eq. (5.23) unaltered. Both facts can be exploited to perform a permutation P with $x_{P(i)} = -x_i$ for all i, $1 \leqslant i \leqslant n$, with the effect that the left-hand side of (5.23) changes its sign for $k = 3, 5,\ldots, n-1$. This completes our proof that every even-symmetric solution $m(\mathbf{x})$ is identically zero on all sublattices $I(\mathbf{x})$ with $x_1 + \cdots + x_n = 0$. According to (5.22) and (5.23), it is therefore unstable at low temperatures.

6. FORGETFUL MEMORIES

In this section we briefly consider neural network models that, unlike the Hopfield model, are capable of ever acquiring new information, albeit at the expense of gradually forgetting previously stored data. These models are usually characterized by learning rules that involve an iterative definition of the synaptic couplings J_{ij}; see GT (1.9) and Refs. 6, 7, and 16–19. To be specific, if one denotes by $J_{ij}(\alpha)$ the value of the coupling between neurons i and j after α patterns have been embedded in the system, then one has

$$J_{ij}(\alpha) = N^{-1}\phi(\mu_\alpha \xi_{i\alpha} \xi_{j\alpha} + N J_{ij}(\alpha - 1)), \qquad 1 \leqslant \alpha \leqslant q \tag{6.1}$$

with

$$J_{ij}(0) = 0$$

for some suitably defined function ϕ: $\mathbb{R} \to \mathbb{R}$. This prescription includes, e.g., the nonlinear learning within bounds algorithms of Hopfield,[6,16] Toulouse *et al.*,[17] Nadal *et al.*,[18] and Parisi,[19] and also the linear palimpsestic schemes studied by Mezard *et al.*[7]

In this section we are not, however, going to investigate the extensive spin-glass limit where q, i.e., the number of stored patterns, becomes of the order of the system size N, but rather study the essentially finite-q behavior ($q \ll \log N$) of models described by (6.1), to which the general theory developed in GT applied. It turns out that some of the salient features of forgetfulness are already present in the finite-q limit and that the *mechanisms* of forgetfulness can be explicitly identified.

Putting $J_{ij} = J_{ij}(q)$, with $J_{ij}(q)$ given by (6.1), the J_{ij} so defined are of the general form

$$J_{ij} = N^{-1} Q(\xi_i; \xi_j) \tag{6.2}$$

with synaptic kernel Q satisfying the invariance property GT (3.1). The spectral theory of Section 3 of GT therefore applies to the synaptic kernels Q associated with forgetful memories described by (6.1), no matter what ϕ, and, interestingly, mechanisms of forgetfulness can be discovered in the *spectrum* of Q. The reason for this is related to a probabilistic interpretation of the eigenvalues λ_ρ, $\rho \subseteq \{1,..., q\}$, of Q as *embedding strengths* of the stored patterns (or of products of stored patterns, as the case may be).

To understand this, consider $v_\rho(\mathbf{x})$, $\mathbf{x} \in \mathscr{C}^q$, i.e., an eigenvector of Q corresponding to the eigenvalue λ_ρ of Q. The product pattern

$$v_\rho(\xi_i) = \prod_{\alpha \in \rho} \xi_{i\alpha} \tag{6.3}$$

is associated with $v_\rho(\mathbf{x})$, and its embedding strength may be defined as

$$e_\rho = N^{-1} \sum_{i,j} v_\rho(\xi_i)\, v_\rho(\xi_j) J_{ij} \tag{6.4}$$

The quantity e_ρ is an overlap between the pattern associated with v_ρ and the couplings J_{ij} of the network and it may be interpreted as a measure of the trace left by the pattern v_ρ in the network. Using (6.2), one can compute the embedding strength e_ρ to give

$$\begin{aligned} e_\rho &= N^{-2} \sum_{i,j} v_\rho(\xi_i)\, v_\rho(\xi_j)\, Q(\xi_i; \xi_j) \\ &= N^{-2} \sum_{\mathbf{x},\mathbf{y}} \left[\sum_{i \in I(\mathbf{x})} \sum_{j \in I(\mathbf{y})} v_\rho(\xi_i)\, v_\rho(\xi_j)\, Q(\xi_i; \xi_j) \right] \\ &= 2^{-2q} \sum_{\mathbf{x},\mathbf{y}} v_\rho(\mathbf{x})\, v_\rho(\mathbf{y})\, Q(\mathbf{x}; \mathbf{y}) \\ &= 2^{-q} \lambda_\rho \end{aligned} \tag{6.5}$$

Here $I(\mathbf{x})$ and $I(\mathbf{y})$ have already been defined in (3.7). We have used the fact that the index set $\{1,..., N\}$ may be written as a disjoint union of the $I(\mathbf{x})$, $\mathbf{x} \in \mathscr{C}^q$; see GT (2.11) and (2.12). Therefore, e_ρ is proportional to the eigenvalue λ_ρ of Q. Note that $-\frac{1}{2}e_\rho$ equals the ground-state energy of the pure state (6.3) associated with v_ρ.

Given the above interpretation of the eigenvalues λ_ρ as embedding strengths, there are two possible mechanisms of forgetfulness.

First, the embedding stengths of the stored patterns decay so rapidly as a function of storage ancestry that except for the very last ones they disappear in the noise created by thermal motion or by themselves.[20]

Second, the embedding strengths of the retrieval states decay faster than those of certain product states and are therefore swamped by them.

Both scenarios may be observed in the finite-q limit for appropriate choices of the function ϕ and the weights μ_α in (6.1). Ultimately, however, a pattern will only be forgotten if its embedding strength becomes smaller than some level of (thermal or static) noise in the system. In the finite-q

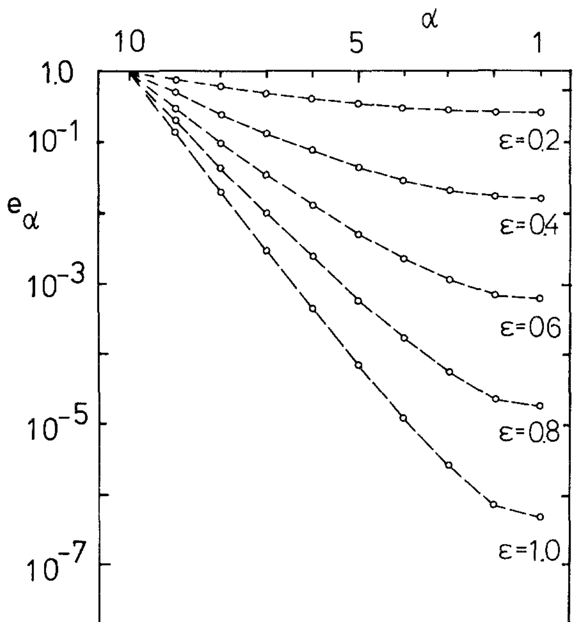

Fig. 2. Decay of embedding strengths e_α of the patterns $\{\xi_{i\alpha}\}$ as a function of storage ancestry α in model (i), for various values of the parameter ε. The embedding strength of the most recently stored pattern $\alpha = q = 10$ has been normalized to 1.

limit this can of course only be thermal noise, since finitely many patterns do not generate any static synaptic noise. Infinitely many, however, do.(20)

To illustrate our general ideas, we have studied the spectrum of the synaptic kernel Q associated with (6.1) for the following two choices of the function ϕ:

$$\text{(i)}\quad \phi(x) = \tanh(x)$$

$$\text{(ii)}\quad \phi(x) = \begin{cases} \operatorname{sgn}(x), & |x| > 1 \\ x, & |x| \leqslant 1 \end{cases} \tag{6.6}$$

and the weights

$$\mu_\alpha = \varepsilon/\sqrt{q}, \qquad 1 \leqslant \alpha \leqslant q \tag{6.7}$$

in (6.1).

Both models have in common that in the limit $\varepsilon \to \infty$ they will only memorize the most recently stored pattern and that, taking $\varepsilon^{-1}\phi(x)$ instead

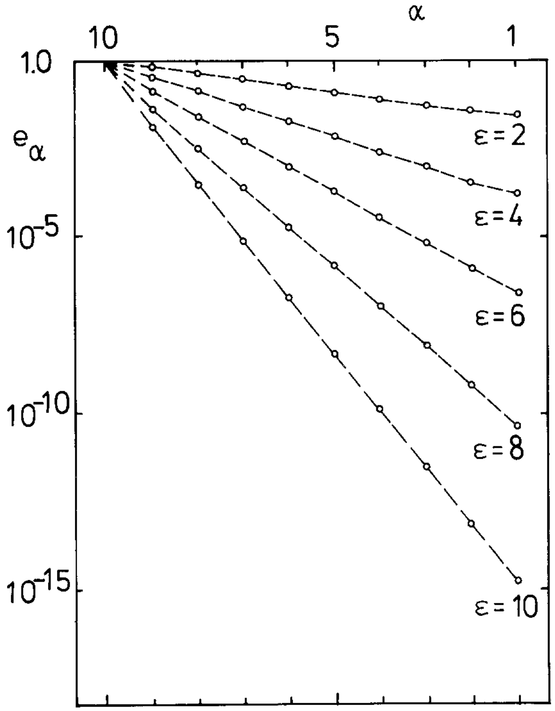

Fig. 3. Same as Fig. 2, for a set of larger values of ε. Here the decay of the e_α is approximately exponential.

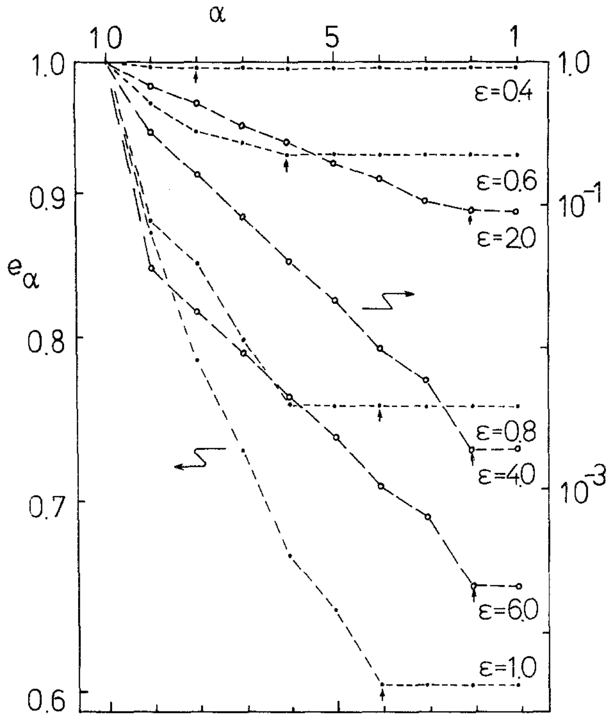

Fig. 4. Decay of the embedding strengths e_α as a function of storage ancestry α in model (ii). For $\varepsilon \leqslant 1/\sqrt{q}$ (≈ 0.316 for $q = 10$), this model reduces to the linear Hopfield model because the nonlinearity of ϕ does not become operative. The e_α for $1 \leqslant \alpha \leqslant n(q, \varepsilon) := 1 + \text{entier}(\sqrt{q}/\varepsilon)$ are degenerate, so that the e_α curves level off. This should be contrasted with Figs. 2 and 3. The numbers $n(q, \varepsilon)$ for $q = 10$ are marked by arrows in the figure. For $\varepsilon > 2\sqrt{q}$ (≈ 6.325 for $q = 10$) this model remembers only the most recently stored pattern. Note the different scales for $\varepsilon \leqslant 1$ and $\varepsilon \geqslant 2$.

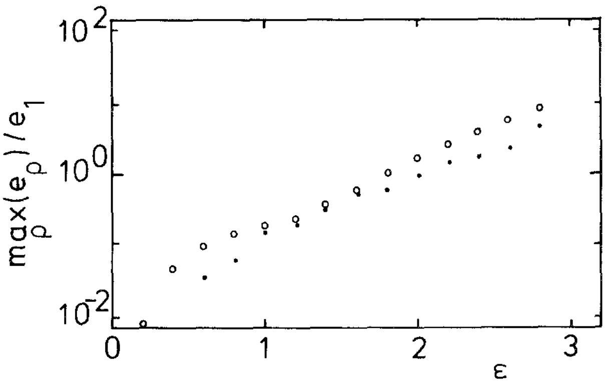

Fig. 5. Ratio of embedding strengths of the most strongly embedded product state and the most weakly embedded (first stored) pattern as a function of the parameter ε for models (i) and (ii) (marked by open circles and by dots, respectively).

of ϕ, in the limit $\varepsilon \to 0$ they reduce to the nonforgetting linear Hopfield model. Intermediate cases of forgetfulness occur between these limits. Our results are summarized in Figs. 2–5. In obtaining them, no approximation was made. Further details about forgetful memories, including the case of extensively many stored patterns, will be reported elsewhere.(21)

7. DISCUSSION

In the present paper and the previous one(1) we have presented a complete analysis of neural network models with finitely many patterns. The general theory (GT, Section 2) allows an *arbitrary* synaptic kernel Q and hence arbitrary nonlinearity. Furthermore, the distribution of the $\xi_{i\alpha}$ was taken to be *arbitrary as well.* Under a weak invariance condition [GT (3.1)] a complete spectral theory has been derived for the case where the $\xi_{i\alpha}$ assume the values ± 1 with equal probability. The Gaussian distribution has been considered in detail also. Combined with the results of Sections 4–6 of the present paper this allows a deeper understanding of the way in which the solution for extensively many patterns(20) joins onto the one for a finite but *large* number of patterns.

A large subclass of neural network models is provided by the inner product models where

$$Q(\mathbf{x}; \mathbf{y}) = \phi(\mathbf{x} \cdot \mathbf{y}) \tag{7.1}$$

for some synaptic function ϕ. For this type of model the states associated with the q stored patterns bifurcate first, at $T_c = 2^{-q}\lambda_1$, where λ_1 is the largest eigenvalue of the synaptic kernel Q. The retrieval quality α_ρ, with $|\rho| = 1$, is determined by

$$\alpha_\rho = \tanh[(T_c/T)\, \alpha_\rho] \tag{7.2}$$

which is Eq. (3.1). The lower the temperature, the closer α_ρ is to one, and the better the retrieval. For instance, if $T_c/T = 5$, then $\alpha_\rho \approx 1 - 10^{-4}$ and since the error percentage is determined by $\frac{1}{2}(1 - \alpha_\rho)$, we find that only one out of every 20,000 spins has the wrong sign. At present the largest simulation samples contain 4000 spins and the above accuracy more than suffices.

The nonlinearity becomes noticeable if $T < \tilde{T}_q = 2^{-q}\tilde{\lambda}$, where $\tilde{\lambda}$ is the *second* largest eigenvalue of Q. For the Hopfield model, which is linear, $\tilde{\lambda}$ vanishes, but for all other known models $\tilde{\lambda} > 0$. Below $\tilde{T}_q$ we enter the temperature regime where abundantly many (typically 2^{q-2}) pure product states appear, which all acquire metastability as $T \to 0$. They are a consequence of the nonlinearity, but are usually unwanted. So the important question is: Can we choose the temperature T so low that the retrieval quality of the stored patterns is acceptable while at the same time T still exceeds $\tilde{T}_q$? Indeed we can. The point is that $\tilde{T}_q/T_c \propto q^{-1}$ effectively turns out to be an upper bound, which is easily reduced to 1/5, say.

The symmetric states bifurcate at T_c, are fortunately unstable as they appear, but cannot be dispensed with. In fact, they exist for linear and nonlinear inner product models alike.

In summary, for a given temperature $T>0$ the benefits of a nonlinear model can be kept, while the disadvantages as compared to the Hopfield model are gradually eliminated as $q \to \infty$.

APPENDIX

In this Appendix we establish that, if to lowest order in t the solution of (4.7) is given by (4.19), then the full solution $\boldsymbol{\alpha}$ of (4.3) is symmetric in the sense that, as $t \to 0^+$,

$$\alpha_\rho = \begin{cases} \alpha_{|\rho|} \sim t^{|\rho|/2}, & \rho \subseteq \mathcal{N} \\ 0, & \rho \cap \mathcal{N}^c = \rho \cap (\{1,..., q\} \backslash \mathcal{N}) \neq \varnothing \end{cases} \tag{A.1}$$

where $\mathcal{N}$ is the subset of $\{1,..., q\}$ defined by $\mathcal{N} = \{\mu\colon a_\mu \neq 0 \text{ in } (4.7)\}$.

To this end, we expand the right-hand side of (4.3), writing, as before, $\alpha_\rho = a_\mu$ if $\rho = \{\mu\}$, and $\alpha_\rho = b_\rho$ if $|\rho| \geqslant 3$. We then obtain [cf. (4.5)]

$$\begin{aligned} 0 = G_\rho = {} & \alpha_\rho(1 - \beta\Lambda_\rho) + \frac{(\beta\Lambda_1)^3}{3} \sum_{\mu_1\mu_2\mu_3} a_{\mu_1} a_{\mu_2} a_{\mu_3} \delta_{\rho,\mu_1 \circ \mu_2 \circ \mu_3} \\ & + \beta \frac{(\beta\Lambda_1)^2}{3} \sum_{\mu_1\mu_2\sigma} a_{\mu_1} a_{\mu_2} \Lambda_\sigma b_\sigma \delta_{\rho,\mu_1 \circ \mu_2 \circ \sigma} \\ & + \beta^2 \frac{\beta\Lambda_1}{3} \sum_{\mu_1\sigma_1\sigma_2} a_{\mu_1} \Lambda_{\sigma_1} b_{\sigma_1} \Lambda_{\sigma_2} b_{\sigma_2} \delta_{\rho,\mu_1 \circ \sigma_1 \circ \sigma_2} \\ & + \frac{\beta^3}{3} \sum_{\sigma_1\sigma_2\sigma_3} \Lambda_{\sigma_1} b_{\sigma_1} \Lambda_{\sigma_2} b_{\sigma_2} \Lambda_{\sigma_3} b_{\sigma_3} \delta_{\rho,\sigma_1 \circ \sigma_2 \circ \sigma_3} + \cdots \end{aligned} \tag{A.2}$$

We first consider the equations $G_\sigma = 0$ with $\sigma \cap \mathcal{N}^c \neq \varnothing$. Denote the corresponding α_σ by c_σ. Because of the Kronecker δ's in (A.2) and the definition of the $\circ$ product in (4.10), every term in G_σ other than the linear term $c_\sigma(1 - \beta\Lambda_0)$ must also contain amplitudes of "c type," so that G_σ is of the form

$$G_\sigma = c_\sigma(1 - \beta\Lambda_\sigma) + \sum_{\substack{\sigma' \\ \sigma' \cap \mathcal{N}^c \neq \varnothing}} c_{\sigma'} g_{\sigma'}(\{a_\mu\}, \{b_\rho\}, \{c_\sigma\}) \tag{A.3}$$

Since the sum over σ' in (A.3) exclusively collects *non*linear terms of G_σ, each $g_{\sigma'}$ must contain at least two other amplitudes of a, b, or c type. Since we know beforehand that the b_ρ and c_σ are at least of order t [cf. (4.5) and Ref. 4, Lemma 4.1], we conclude from (A.3) that *all* c_σ are at least of order t^2. Repeating this argument indefinitely, we conclude that the c_σ are zero to *all* orders of t^n, $n \in \mathbb{N}$, as $t \to 0^+$, and hence can be ignored in what follows.

We now consider the b_ρ with $\rho \subseteq \mathcal{N}$. We show that, as $t \to 0^+$,

$$b_\rho = b_{|\rho|} \sim t^{|\rho|/2} \tag{A.4}$$

We recall that $\Lambda_\sigma = \Lambda_{|\sigma|}$ and that to lowest order in t the a_μ are, according to (4.19), symmetric and of order $t^{1/2}$. Further, using $\beta\Lambda_\sigma = (1+t)\,\Lambda_{|\sigma|}/\Lambda_1$, we find that, as $t \to 0^+$, the amplitudes $b_{\{\nu_1,\nu_2,\nu_3\}}$, $\nu_i \in \mathcal{N}$ $(i = 1, 2, 3)$, are determined by

$$0 = b_{\{\nu_1,\nu_2,\nu_3\}}(1 - \Lambda_3/\Lambda_1) + 2a_{\nu_1}a_{\nu_2}a_{\nu_3} \tag{A.5}$$

from (A.2). Given the symmetry of the a_μ, we conclude that to lowest order in t,

$$b_\rho = b_{|\rho|} \sim t^{|\rho|/2} \qquad \text{for} \quad |\rho| = 3 \tag{A.6}$$

Suppose now that we had established (A.4) for all b_ρ such that $|\rho| \leqslant |\rho_0| - 2$. Then to lowest order in t the equation for b_{ρ_0} reads

$$0 = b_{\rho_0}(1 - \Lambda_{|\rho_0|}/\Lambda_1) + \sum_{\{\mu\},\{\rho\}} g_{\rho_0}(\{a_\mu\}, \{b_\rho\}) \tag{A.7}$$

Here g_{ρ_0} is a product of a_μ and b_ρ with $\mu \in \{\mu\}$ and $\rho \in \{\rho\}$ such that the $\circ$ product gives ρ_0,

$$\prod_{\mu \in \{\mu\}}^{\circ} \mu \circ \prod_{\rho \in \{\rho\}}^{\circ} \rho = \rho_0 \tag{A.8}$$

Since all a_μ and b_ρ carry t factors, the sum in (A.7) includes only terms where $\{\mu\}$, $\{\rho\}$ constitutes a *disjoint decomposition* of ρ_0 in the series (A.2). These, however, involve only those amplitudes whose symmetry to lowest order in t is already established by assumption. Since the number of disjoint decompositions of ρ_0 with a given number of 1-, 3-, 5-,..., $(|\rho_0| - 2)$-element sets depends only on the *size* of ρ_0, the symmetry of b_{ρ_0} and

$$b_{\rho_0} = b_{|\rho_0|} \sim t^{|\rho_0|/2} \tag{A.9}$$

follow. This then completes the proof of (A.1).

REFERENCES

1. J. L. van Hemmen, D. Grensing, A. Huber and R. Kühn, *J. Stat. Phys.*, this issue, preceding paper.
2. J. L. van Hemmen and R. Kühn, *Phys. Rev. Lett.* **57**:913 (1986).
3. M. H. Protter and C. B. Morrey, *A First Course in Real Analysis* (Springer, New York, 1977), Chapter 14.
4. D. H. Sattinger, *Group Theoretic Methods in Bifurcation Theory* (Springer, Berlin, 1979).
5. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. Lett.* **55**:1530 (1985); *Ann. Phys.* **173**:30 (1987).
6. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
7. M. Mézard, J. P. Nadal, and G. Toulouse, *J. Phys. (Paris)* **47**:1457 (1986).
8. J. L. van Hemmen and V. A. Zagrebnov, *J. Phys. A: Math. Gen.* **20**:3989 (1987).
9. H. Sompolinsky, *Phys. Rev. A* **34**:2571 (1986).
10. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **32**:1007 (1985).
11. M. G. Golubitsky and D. G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. 1 (Springer, Berlin, 1985).
12. J. L. van Hemmen, *Phys. Rev. A* **34**:3435 (1986), Section III.
13. F. R. Gantmacher, *The Theory of Matrices*, Vol. 1 (Chelsea, New York, 1977), Sections X.1 and 2.
14. D. H. Sattinger, *Bull. Am. Math. Soc.* **3**:779–819 (1980) and references therein.
15. A. C. Aitken, *Determinants and Matrices* (Oliver & Boyd, London, 1958), p. 135.
16. J. J. Hopfield, in *Modelling in Analysis and Biomedicine*, C. Nicolini, ed. (World Scientific, Singapore, 1984), pp. 369–389, in particular p. 381.
17. G. Toulouse, S. Dehaene, and J.-P. Changeux, *Proc. Natl. Acad. Sci. USA* **83**:1695 (1986).
18. J. P. Nadal, G. Toulouse, J.-P. Changeux, and S. Dehaene, *Europhys. Lett.* **1**:535 (1986).
19. G. Parisi, *J. Phys. A: Math. Gen.* **19**:L617 (1986).
20. J. L. van Hemmen, *Phys. Rev. A* **36**:1959 (1987).
21. J. L. van Hemmen, G. Keller, and R. Kühn, SFB Preprint No. 436 (Heidelberg, 1987).