Spin-glass models of neural networks

Daniel J. Amit and Hanoch Gutfreund

Racah Institute of Physics, Hebrew University, 91904 Jerusalem, Israel

H. Sompolinsky

Department of Physics, Bar-Ilan University, 52100 Ramat-Gan, Israel

(Received 22 March 1985)

Two dynamical models, proposed by Hopfield and Little to account for the collective behavior of neural networks, are analyzed. The long-time behavior of these models is governed by the statistical mechanics of infinite-range Ising spin-glass Hamiltonians. Certain configurations of the spin system, chosen at random, which serve as memories, are stored in the quenched random couplings. The present analysis is restricted to the case of a finite number p of memorized spin configurations, in the thermodynamic limit. We show that the long-time behavior of the two models is identical, for all temperatures below a transition temperature T_c. The structure of the stable and metastable states is displayed. Below T_c, these systems have $2p$ ground states of the Mattis type: Each one of them is fully correlated with one of the stored patterns. Below $T \sim 0.46T_c$, additional dynamically stable states appear. These metastable states correspond to specific mixings of the embedded patterns. The thermodynamic and dynamic properties of the system in the cases of more general distributions of random memories are discussed.

I. INTRODUCTION

A. Models of neural networks

Recently, a number of models have been proposed which view the human memory as a collective property of large interconnected neural networks. Two such models, one proposed recently by Hopfield[1] and a closely related model, proposed some ten years ago by Little,[2] are the focus of this paper. We briefly review the physiological background of these models. For more details the reader is referred to Refs. 1 and 2 and to a recent study of these models by Peretto.[3] In both models each neuron is viewed as an Ising spin with two possible states: an "up" position or a "down" position depending on whether the neuron has, or has not, fired an electrochemical signal (within an interval of time of the order of a millisecond). The state of the network of N such neurons at time t is defined as the instantaneous configuration of all the spin variables at time t:

$$|\alpha,t\rangle = |S_1^\alpha, S_2^\alpha, \ldots, S_N^\alpha; t\rangle . \tag{1.1}$$

The dynamic evolution of these states, in the phase space of 2^N states, is determined by the interactions among the neurons. The neurons are interconnected by synaptic junctions of strength J_{ij}, which determine the contribution of a signal fired by the jth neuron to the postsynaptic potential which acts on the ith neuron. This contribution can be either positive (excitatory synapse) or negative (inhibitory synapse). The potential V_i on each neuron is the sum of all postsynaptic potentials delivered to it in an integrating period of time, of the order of a few milliseconds, i.e.,

$$V_i = \sum_j J_{ij}(S_j + 1) . \tag{1.2}$$

In the absence of noise, or external perturbation, each neuron fires a signal if its potential V_i exceeds a threshold value U_i. Thus the stable states of the network will be those configurations in which each of the spin variables S_i is aligned with its molecular field $h_i = V_i - U_i$, i.e.,

$$S_i h_i = S_i(V_i - U_i) > 0 . \tag{1.3}$$

It will be assumed throughout the paper that the J_{ij}'s are symmetric, i.e., $J_{ij} = J_{ji}$. In such a case Eq. (1.3) is equivalent to the requirement that the configurations $\{S_i\}$ be local minima (i.e., stable to all single-spin flips) of the Hamiltonian

$$H = -\frac{1}{2}\sum_i h_i S_i = -\frac{1}{2}\sum_{i,j} J_{ij} S_i S_j , \tag{1.4}$$

where it is usually assumed that the threshold potentials satisfy $U_i \simeq \sum_j J_{ij}$. Thus there is no external field term in H. In the presence of noise there is a finite probability of having configurations other than those given by Eq. (1.3). This can be taken into account by introducing an effective temperature $1/\beta$, characterizing the level of noise in the system,[4] as will be described below.

For the network to have a capacity for learning and memory its stable configurations must be correlated with certain configurations, which are determined by the learning process. This is achieved by choosing the interactions J_{ij} to be given by

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu, \quad i \neq j . \tag{1.5}$$

The p sets of $\{\xi_i^\mu\}$ are certain configurations of the network which were fixed by the learning process. The ξ_i^μ are taken to be quenched random variables, assuming the values $+1$ and -1 with equal probabilities. Note that according to Eq. (1.5) every pair of neurons is connected.

The model (1.3)–(1.5) will have the capacity of storage and retrieval of information if indeed the emergent dynamically stable configurations $\{S_i\}$ are correlated with the "learned memories" $\{\xi_i^\mu\}$. This question is at the center of the present study. However, in order to complete the definition of the model one has to prescribe a dynamic mechanism, by which the network evolves from an arbitrary initial condition.

B. The generalized Hopfield model and the Little model

Hopfield's dynamic model is essentially a $T=0$ Monte Carlo[5] (or Glauber) dynamics. Starting from an arbitrary initial configuration the system evolves by a sequence of single-spin flips, involving spins which are misaligned with their instantaneous molecular fields. This process monotonically decreases the value of H, (1.4), and leads eventually to steady states, which are the local minima of (1.6). A natural generalization of this model to a system with noise is to adopt Glauber single-spin dynamics at a finite temperature $1/\beta$. The distribution of configurations (1.1) relaxes, in this case to a Gibbs distribution

$$P\{S\} \propto \exp(-\beta H\{S\}) , \tag{1.6}$$

with H of (1.4). We refer to this finite-temperature model as the generalized Hopfield model. Note that stability of a state to all single-spin flips is not sufficient for dynamic stability at finite temperatures.

In Little's model the probability that the ith spin be in a state S_i' at time $t+\delta t$, given a configuration $\{S_i\}$ of the network at time t, is proportional to

$$P(S_i') = \frac{\exp(-\beta S_i' h_i\{S_i\})}{\exp(-\beta S_i' h_i\{S_i\}) + \exp(+\beta S_i' h_i\{S_i\})} , \tag{1.7}$$

where $h_i = V_i - U_i = \sum_j J_{ij} S_j$, as before. The matrix W, of transition probabilities from the state $|\alpha, t\rangle$ to $|\beta, t+\delta t\rangle$, is just a product of the probabilities (1.7), i.e.,

$$\langle \beta | W | \alpha \rangle = \prod_{i=1}^{N} [\tfrac{1}{2} e^{-\beta h_i^\alpha S_i^\beta} \operatorname{sech}(\beta h_i^\alpha S_i^\beta)] . \tag{1.8}$$

Thus, at each time step all the spins check simultaneously their states against their molecular field. Each step may consist of many, even N, spin flips.

It has been shown by Peretto[3] that as long as J_{ij} is symmetric, the master equation, based on the transition rates given by (1.8), obeys detailed balance and hence leads to a stationary Gibbs distribution of states, $\exp(-\beta \bar{H})$ with the effective Hamiltonian,

$$\bar{H} = -\frac{1}{\beta} \sum_i \ln \left[2 \cosh \left(\beta \sum_j J_{ij} S_j \right) \right] . \tag{1.9}$$

The synchronous dynamics of this model seems at first glance to lead to a very different collective behavior from the asynchronous mechanism of Hopfield. The dynamics of real systems is, most probably, in between. Thus it is important to investigate to what extent this difference is relevant.

C. Relationship to models of random magnets

The study of the models described above is interesting not only in the context of models of memory but also in the context of the statistical mechanics of disordered magnetic systems. The Hamiltonian defined in (1.4) and (1.5) is a special case of infinite-range spin glasses[6] where every pair of spins is interacting via a quenched random exchange J_{ij}. In the canonical infinite-ranged spin-glass model, introduced by Sherrington and Kirkpatrick[7] (SK), each J_{ij} is an independent random variable. In this model the disorder leads to the appearance of an infinite number of ground states (in the $N \to \infty$ limit) with static and dynamic properties, which are very different from the usual ferromagnetic case.

The other extreme is the case of (1.5) with $p = 1$, which is an infinite-range Mattis model.[8] Here the disorder can be gauged away, hence it is irrelevant thermodynamically. There are two ground states ($\{S_i\} = \pm\{\xi_i\}$) with no frustration: Each "bond" $S_i S_j J_{ij}$ in the ground state is positive. The model (1.4) and (1.5) with $p > 1$ represents an intermediate case. There is always a finite fraction of frustrated bonds. Nevertheless, the correlation between the bonds may be sufficiently strong to yield a structure of broken symmetry phase considerably simpler than that of the SK model. Van Hemmen[9] introduced and solved a related model with $p = 2$. His mean-field equation has been generalized to arbitrary p by Provost and Vallee.[10] However, the structure and the properties of the mean-field solutions for general p have not been investigated, nor have the possible existence and the properties of metastable states. These issues are the main focus of this paper.

A similar model was also studied in the context of the mean-field theory of random-axis ferromagnets.[11] In the limit of strong local anisotropy the system is mapped onto a spin-glass Ising model with $J_{ij} \propto \hat{n}_i \cdot \hat{n}_j$ where $\hat{n}_i$ is the direction of the local easy axis. These J_{ij} are similar to Eq. (1.5) but with ξ_i^μ which are the Cartesian components of random unit vectors, rather than independent and discrete random variables. This raises the issue of the sensitivity of the properties of the models to the form of the distribution of ξ_i^μ.

D. Outline and summary of results

In this paper a statistical mechanical study of the Hopfield and the Little models is presented. This study is explicitly restricted to the limit $N \to \infty$ and finite p. In Sec. II the solutions of the mean-field theory of the Hopfield model are studied. It is shown that at all $T < T_c = 1$ the free energy ground states are all Mattis states: Each one of them is correlated with one of the p memories, $\{\xi_i^\mu\}$. At $T = 1$, additional mean-field solutions with higher free energy appear. These are symmetric states which have equal overlap with several memories.

Section III presents the stability analysis of these solutions. It is shown that as the temperature is decreased below

$$T \simeq 0.461 \tag{1.10}$$

some of these solutions become local minima. At $T=0$ all symmetric saddle points which overlap with an odd number of memories become local minima. These states are truly metastable: They are separated by free energy barriers proportional to N.

In Sec. IV mean-field *asymmetric* solutions which have unequal overlaps on some memories are discussed. They appear only below $T \sim 0.57$ and none of them are stable at temperatures higher than that of (1.10).

The properties of the Little model are studied in Sec. V. We show that the model has the same thermodynamic properties as the Hopfield model, including the properties of the metastable states.

It has been remarked[3] that in the Little model the system may in certain cases get trapped in indefinite oscillations between several configurations, unable to reach the aligned states given by (1.3). We show that with the J_{ij}'s given by (1.5) this does not happen.

In Sec. VI we consider distributions of $\{\xi_i^\mu\}$ other than ± 1. It is shown that the low-temperature properties of the models depend strongly on the details of the distribution of $\{\xi_i^\mu\}$. If the probability density at $\xi_i^\mu = 0$ is sufficiently large, the "mixed" states, rather than the Mattis states, become the ground states of the system. On the other hand, certain continuous distributions may eliminate the "mixed" states, leaving the Mattis states as the only dynamically stable states of the system.

II. THE GENERALIZED HOPFIELD MODEL

A. Mean-field theory

We now turn to the investigation of the thermodynamics of the Hamiltonian

$$H = -\frac{1}{2} \sum_{\substack{i,j \\ i \neq j}} \left[\left[\frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \right] S_i S_j \right] , \qquad (2.1)$$

where ξ_i^μ are independent random variables with zero mean. This system will be studied in the limit $N \to \infty$ and finite p. The ensemble averaged free energy density is given by

$$-\beta f(\beta) = \lim_{N \to \infty} [N^{-1} \langle\!\langle \ln \underset{s}{\mathrm{Tr}} \exp(-\beta H) \rangle\!\rangle] , \qquad (2.2)$$

where $\beta \equiv 1/T$ (with units in which $k_B = 1$). The notation $\langle\!\langle \cdots \rangle\!\rangle$ stands for the average over the distribution of $\{\xi_i^\mu\}$. The partition function is rewritten, for a given realization of the ξ's, as

$$Z = \underset{s}{\mathrm{Tr}} \exp(-\beta H) = \exp(-\beta p/2) \underset{s}{\mathrm{Tr}} \exp\left[(\beta/2N) \sum_\mu \left[\sum_i S_i \xi_i^\mu \right]^2 \right]$$

$$= (N\beta)^{p/2} e^{-\beta p/2} \int \prod_\mu \frac{dm^\mu}{\sqrt{2\pi}} \exp\left[\frac{-N\beta \mathbf{m}^2}{2} + \sum_i \ln[2\cosh(\beta \mathbf{m} \cdot \boldsymbol{\xi}_i)] \right] , \qquad (2.3)$$

where a vector notation for the p components of ξ_i^μ and m^μ has been introduced. As long as p remains finite, the integral over m is dominated by its saddle-point value,

$$-\frac{1}{N\beta} \ln Z = \frac{1}{2} \mathbf{m}^2 - \frac{1}{N\beta} \sum_i \ln[2\cosh(\beta \mathbf{m} \cdot \boldsymbol{\xi}_i)] . \qquad (2.4)$$

The order parameter $\mathbf{m}$ is determined by the saddle-point equations $\partial \ln Z / \partial m^\mu = 0$,

$$\mathbf{m} = \frac{1}{N} \sum_i \boldsymbol{\xi}_i \tanh(\beta \mathbf{m} \cdot \boldsymbol{\xi}_i) . \qquad (2.5)$$

At any finite N the right-hand sides of Eqs. (2.4) and (2.5) depend on the particular realization of $\{\xi_i^\mu\}$. However, in the limit $N \to \infty$ the random fluctuations are suppressed and both $\ln Z$ and $\mathbf{m}$ are self-averaged, as discussed in Refs. 9 and 10. The sums $(1/N)\sum_i$ are, therefore, replaced by averages over $\{\xi_i\}$, leading to the mean-field equations[10]

$$f(\beta) = \frac{1}{2}\mathbf{m}^2 - \frac{1}{\beta} \langle\!\langle \ln[2\cosh(\beta \mathbf{m} \cdot \boldsymbol{\xi})] \rangle\!\rangle , \qquad (2.6)$$

$$\mathbf{m} = \langle\!\langle \boldsymbol{\xi} \tanh(\beta \mathbf{m} \cdot \boldsymbol{\xi}) \rangle\!\rangle . \qquad (2.7)$$

To interpret the order parameter $\mathbf{m}$ one adds an external source conjugate to $\xi_i^\mu S_i$, to find that $\mathbf{m}$ is just the average overlap between the local magnetization and the ξ's. Explicitly, one has

$$m^\mu = \langle\!\langle \xi_i^\mu \langle S_i \rangle \rangle\!\rangle , \qquad (2.8)$$

where

$$\langle S_i \rangle = \tanh(\beta \mathbf{m} \cdot \boldsymbol{\xi}_i) \qquad (2.9)$$

is the *thermal* average of the spin at site i.

The detailed structure of the solutions of Eq. (2.7) is essential for the determination of the correlations of the spin states $\{\langle S_i \rangle\}$ with each of the p quenched "memories" $\{\xi_i^\mu\}$.

B. The Mattis states

We will restrict ourselves in this subsection to the case in which the distribution of ξ_i is given by

$$P\{\xi_i^\mu\} = \prod_{\mu,i} p(\xi_i^\mu) ,$$

$$p(\xi_i^\mu) = \frac{1}{2}\delta(\xi_i^\mu - 1) + \frac{1}{2}\delta(\xi_i^\mu + 1) . \qquad (2.10)$$

Expanding Eqs. (2.6) and (2.7) in powers of $\mathbf{m}$ one obtains

$$f = -T\ln 2 + \frac{1}{2}(1 - \beta)\mathbf{m}^2 + O(m^4) , \qquad (2.11)$$

$$m^\mu = \beta m^\mu + \frac{2}{3}\beta^3(m^\mu)^3 - \beta^3 m^\mu \mathbf{m}^2 + O(m^4) ,$$

$$\mu = 1, 2, \ldots, p \qquad (2.12)$$

from which it is seen that above $T = 1$ the only solution is

the paramagnetic state $\mathbf{m}=0$, with $f=-T\ln 2$. This solution becomes unstable below $T_c=1$ where solutions with nonzero m appear. *We will denote by n the dimensionality of $\mathbf{m}$, i.e., the number of nonzero components of $\mathbf{m}$ in a particular solution below T_c.* It is clear from Eqs. (2.7) and the form (2.10) that permuting the m^μ's or changing the signs of each of the n nonzero components independently generates entirely equivalent solutions. Hence without loss of generality we can restrict ourselves to solutions in which *the first n components* (i.e., m^μ with $\mu=1,2,\ldots,n$) *are positive;* the rest of them are zero.

We first discuss solutions with $n=1$. Assuming $m^\mu=0$ for all $\mu>1$,

$$f=\tfrac{1}{2}(m^1)^2-\frac{1}{\beta}\ln[2\cosh(\beta m^1)] , \qquad (2.13)$$

$$m^1=\tanh(\beta m^1) . \qquad (2.14)$$

These are the usual mean-field equations of Ising ferromagnets. Indeed, this solution corresponds to a state in which all the local magnetizations (up to a negligible fraction as $N\to\infty$) are equal to

$$\langle S_i\rangle=\xi_i^1\tanh(\beta m^1) . \qquad (2.15)$$

This state is thermodynamically equivalent (via a Mattis transformation[8]) to the ferromagnetic state. There are $2p$ equivalent states of this form corresponding to different μ's and different signs of m. We refer to these states as Mattis states.

The Mattis states are the global minima of the free energy both near $T=1$ and $T=0$ and most probably at all $T<1$. Near $T=1$, we show explicitly in the next subsection that the Mattis free energy is the lowest of all other saddle points [see Eq. (2.29)]. At $T=0$, Eqs. (2.13) and (2.14) read

$$\mathbf{m}(T=0)=(1,0,0,\ldots,0) , \qquad (2.16)$$

$$E(T=0)=-\tfrac{1}{2} . \qquad (2.17)$$

To show that $-\tfrac{1}{2}$ is the ground-state energy at $T=0$, note that the general mean-field equations [Eqs. (2.6) and (2.7)] yields as $T\to 0$ the following:

$$E=-\tfrac{1}{2}\mathbf{m}^2 , \qquad (2.18)$$

$$\mathbf{m}=\langle\langle\,\xi\,\mathrm{sgn}(\mathbf{m}\cdot\xi)\rangle\rangle . \qquad (2.19)$$

We have used here the limits

$$\tanh(\beta\mathbf{m}\cdot\xi)\to\mathrm{sgn}(\mathbf{m}\cdot\xi) ,$$
$$T\ln[2\cosh(\beta\mathbf{m}\cdot\xi)]\to|\mathbf{m}\cdot\xi| \qquad (2.20)$$

which, by defining $\mathrm{sgn}(0)=0$, apply also to the case where $\mathbf{m}\cdot\xi$ may take the value zero, as long as the distribution of ξ is *discrete*. Each component m^μ is, of course, bounded from above by 1. However, $\mathbf{m}$ obeys a stronger bound which is

$$\mathbf{m}^2\le 1 \qquad (2.21)$$

with the equality being satisfied only for a one-component $\mathbf{m}$. The bound (2.21) can be derived using Eq. (2.19) and the Schwartz inequality,

$$\mathbf{m}^2=\langle\langle\,|\,\xi\cdot\mathbf{m}\,|\,\rangle\rangle$$
$$\le[\langle\langle(\xi\cdot\mathbf{m})^2\rangle\rangle]^{1/2}$$
$$=\left[\sum_{\mu,\nu=1}^n m^\mu m^\nu\langle\langle\xi^\mu\xi^\nu\rangle\rangle\right]^{1/2}=(\mathbf{m}^2)^{1/2} ,$$

which implies that $\mathbf{m}^2$ is less than unity for $n>1$ and equals unity for $n=1$.[12]

From the point of view of storage and retrieval of memory these states are ideal, since each of them is fully correlated with one of the "quenched" memories. However, although the Mattis states are the only states which contribute to the thermodynamics of the system, solutions with $n>1$ may also be important for the dynamics, if they are local minima of the free energy.

C. Symmetric solutions

A particularly simple class of solutions of Eqs. (2.7) consists of those in which all n nonzero components of $\mathbf{m}$ are equal in magnitude, i.e.,

$$\mathbf{m}=m_n(1,1,\ldots,1,0,0,\ldots,0) , \qquad (2.22)$$

where the first n components are unity and the remaining $p-n$ zeros. For a given n there are $\binom{p}{n}2^n$ solutions which are equivalent to that of (2.22). These symmetric solutions are important because they are the only solutions that exist throughout the whole temperature range $T<1$. The transition temperatures for the appearance of asymmetric solutions, in which some of the n components have different magnitudes, are all lower than 1. To see this note that dividing each of the first n equations of (2.12) by m^μ results in $\tfrac{2}{3}(m^\mu)^2\simeq T-1+\mathbf{m}^2$ *which is independent of μ*. It is straightforward to see that the equality of all nonzero $(m^\mu)^2$ holds order by order in perturbation theory about $T=1$ to all orders. Thus the breaking of the symmetry among $(m^\mu)^2$ occurs below a critical temperature which is less than 1. In fact, we will show in Sec. IV that the critical temperatures for the appearance of the asymmetric solutions are all lower than $T\sim 0.57$.

The mean-field equations for the symmetric states (2.22) are

$$f_n=\frac{n}{2}m_n^2-\frac{1}{\beta}\langle\langle\ln[2\cosh(\beta m_n z_n)]\rangle\rangle , \qquad (2.23)$$

$$m_n=(1/n)\langle\langle z_n\tanh(\beta m_n z_n)\rangle\rangle , \qquad (2.24)$$

where

$$z_n^i=\sum_{\mu=1}^n\xi_i^\mu .$$

The distribution of z_n^i is given, according to Eq. (2.10), by

$$p(z_n)=2^{-n}\begin{bmatrix}n\\k\end{bmatrix} , \qquad (2.25)$$

where

$$k=(z_n+n)/2$$

is the number of positive ξ_i^μ's contributing to z_n^i. Incidentally, we note that Eq. (2.25) is the distribution of a ran-

dom walk on a one-dimensional lattice. These solutions correspond to states in which the local magnetization is induced by a molecular field

$$h_i = m_n z_n^i ,$$

i.e.,

$$\langle S_i \rangle = \tanh(\beta m_n z_n^i) . \tag{2.26}$$

In other words, the symmetric solutions with $n > 1$ represent states which are equal mixtures of several memories.

To evaluate the solutions near T_c we expand Eqs. (2.23) and (2.24) in powers of $\mathbf{m}$ and obtain

$$\bar{f}_n \equiv \beta f_n - \ln 2$$
$$\simeq \frac{n}{2}(T-1)(\beta m_n)^2 + \frac{1}{12}(\beta m_n)^4 \langle\!\langle z_n^4 \rangle\!\rangle , \tag{2.27}$$

$$m_n \simeq \beta m_n - \frac{\langle\!\langle z_n^4 \rangle\!\rangle}{3n}(\beta m_n)^3 . \tag{2.28}$$

Using the equality $\langle\!\langle z_n^4 \rangle\!\rangle = n(3n-2)$ one obtains the final results

$$\bar{f}_n \simeq -\frac{3nt^2}{4(3n-2)} , \tag{2.29}$$

$$m_n^2 \simeq \frac{3t}{3n-2} , \tag{2.30}$$

where $t \equiv 1-T$. Thus, $T=1$ is the critical temperature for the appearance of all symmetric solutions. Equations (2.29) imply that near $T=1$ the free energy is *monotonically increasing with* n; the lowest free energy state is $n=1$, namely the Mattis states with

$$m_1^2 \simeq 3t \quad \text{and} \quad \bar{f}_1 \simeq -\frac{3t^2}{4} .$$

To study the symmetric solutions near $T=0$, we use Eqs. (2.18) and (2.19) to obtain

$$m_n(T=0) = \frac{1}{n}\langle\!\langle |z_n| \rangle\!\rangle , \tag{2.31}$$

$$f_n(T=0) = -\frac{1}{2}nm_n^2 . \tag{2.32}$$

Using (2.25) one arrives, for even n, at

$$m_{2k} = \frac{1}{2^{2k}}\begin{bmatrix} 2k \\ k \end{bmatrix} ,$$
$$f_{2k} = -\frac{2k}{2^{4k+1}}\begin{bmatrix} 2k \\ k \end{bmatrix}^2 , \quad k=1,2,\ldots \tag{2.33}$$

and for odd n

$$m_{2k+1} = \frac{1}{2^{2k}}\begin{bmatrix} 2k \\ k \end{bmatrix} ,$$
$$f_{2k+1} = -\frac{2k+1}{2^{4k+1}}\begin{bmatrix} 2k \\ k \end{bmatrix}^2 , \quad k=0,1,\ldots . \tag{2.34}$$

This sequence of f_n is bounded from below by the ground-state energy $f_1 = -0.5$ and from above by

$f_2 = -0.25$. Moreover, the sequence (2.33) is monotonically decreasing with k, while the sequence (2.34) is monotonically increasing. Both have a common limit $(= -1/\pi)$ as $k \to \infty$. This limit coincides with the ground-state energy per spin for a Gaussian distribution of ξ^μ (see Sec. VI). Details of the derivations of these results are left to Appendix A.

In conclusion, we obtain the following ordering of the energies of the symmetric saddle points at $T=0$:

$$f_1 < f_3 < f_5 < \cdots f_\infty \cdots < f_6 < f_4 < f_2 . \tag{2.35}$$

Note that in the "even" solutions there is a finite probability of $z_n = 0$. Hence, a finite fraction of the spins remain disordered at all temperatures. This difference between the odd and even solutions manifests itself in the low-temperature value of the Edwards-Anderson order parameter,[13]

$$q_n = \langle\!\langle \langle S_i \rangle^2 \rangle\!\rangle = \langle\!\langle \tanh^2(\beta m_n z_n) \rangle\!\rangle . \tag{2.36}$$

In the case of odd n the minimum value of $|z|$ is 1. Hence,

$$q_n \simeq 1 - 2p(z_n = 1)\exp(-2\beta m_n) \to 1$$
$$\text{as } \beta \to \infty \text{ (or } T \to 0), \tag{2.37}$$

whereas for even n one has

$$q_n = 1 - p(z_n = 0) \quad \text{at } T=0 . \tag{2.38}$$

In Sec. III we study the stability properties of the various symmetric saddle points in order to determine their importance to the dynamics. The asymmetric solutions will be studied in Sec. IV.

III. METASTABILITY IN THE GENERALIZED HOPFIELD MODEL

A. The stability matrix of the symmetric solutions

The local stability of the saddle points of f, Eq. (2.6), is determined by the eigenvalues of the matrix $\underline{A}$,

$$A^{\mu\nu} = \frac{\partial^2 f}{\partial m^\mu \partial m^\nu} = \delta^{\mu\nu} - \beta(\delta^{\mu\nu} - Q^{\mu\nu}) , \tag{3.1}$$

with

$$Q^{\mu\nu} \equiv \langle\!\langle \xi^\mu \xi^\nu \tanh^2(\beta \mathbf{m} \cdot \xi) \rangle\!\rangle . \tag{3.2}$$

Solutions of Eq. (2.7) are locally stable if all the eigenvalues of A are positive.

The general form of $\underline{A}$ in the case of the symmetric solutions is quite simple. Its diagonal elements are all

$$A^{\mu\mu} = 1 - \beta(1-q) ,$$

where $q = Q^{\mu\mu}$ is given by Eq. (2.36). The off-diagonal elements *with* $\mu,\nu \leq n$ are all equal to

$$\beta Q = \beta \langle\!\langle \xi^1 \xi^2 \tanh^2(\beta m_n z_n) \rangle\!\rangle \tag{3.3}$$

and all other elements vanish. Recall that we have chosen a solution which has the form (2.22).

The matrix $\underline{A}$ has three groups of eigenvalues: (1) a nondegenerate eigenvalue

$$\lambda_1 = 1 - \beta(1-q) + \beta(n-1)Q \qquad (3.4)$$

corresponding to "longitudinal" fluctuations in the amplitude m_n; (2) an eigenvalue of degeneracy $p - n$,

$$\lambda_2 = 1 - \beta(1-q) , \qquad (3.5)$$

which corresponds to fluctuations in directions which mix more memories; and (3) an eigenvalue of degeneracy $n - 1$,

$$\lambda_3 = 1 - \beta(1-q) - \beta Q , \qquad (3.6)$$

which is associated with fluctuations of anisotropy in the space of the n "occupied" memories. Since Q is positive for all $T < 1$ (see Appendix B), the lowest eigenvalue of $\underline{A}$ is λ_3. It is this eigenvalue which determines the stability of the solutions. This is the case for all n except $n = 1$, for which Q does not exist. In this case the only eigenvalue is $\lambda_1 = 1 - \beta(1-q)$. In the following these eigenvalues are calculated near T_c and near $T = 0$.

B. Stability near T_c and near $T = 0$

Expanding (2.36) and (3.3) in powers of $t = 1 - T$ one obtains

$$q \simeq 3nt/(3n-2), \quad Q \simeq 2q/n ,$$

from which it follows that

$$\lambda_1 \simeq -t + q + (n-1)Q \simeq 2t > 0 , \qquad (3.7)$$

$$\lambda_2 \simeq -t + q \simeq \frac{2t}{3n-2} > 0 , \qquad (3.8)$$

but

$$\lambda_3 \simeq -t + q - Q \simeq \frac{-4t}{3n-2} < 0 \qquad (3.9)$$

for all $n > 1$. Hence near $T = 1$, only the solution with $n = 1$ is locally stable; all other solutions are saddle points.

The situation is quite different at lower temperatures. Near $T = 0$ the odd-n symmetry solutions order fully, with at most exponentially small deviations [see Eq. (2.37)]. Thus, as $T \to 0$, $q = 1$ and $Q = 0$ (up to exponentially small corrections) and all eigenvalues equal unity. The solutions are all locally stable. On the other hand, in the even-n solutions the system resists full order, even at $T = 0$ [Eq. (2.38)]. Consequently, both λ_2 and λ_3 are proportional to $-\beta$, whereas $\lambda_1 \sim 1$. These results imply that while the even-n symmetric solutions are unstable for all T, the odd-n ones become locally stable below a certain temperature, $0 < T_n < 1$, given by the vanishing of λ_3, i.e.,

$$-T_n = (q - Q)$$

$$= \langle\langle (1 - \xi^1\xi^2)\tanh^2(\beta m_n z_n) \rangle\rangle . \qquad (3.10)$$

Numerical solution of this equation yields $T_3 = 0.461$, $T_5 = 0.385$, and $T_7 = 0.345$. It can be seen from Eq. (2.37) that the finite-T corrections, at low T, are exponentially small, as long as $T \ll m_n$. As $n \to \infty$, $m_n \propto 1/\sqrt{n}$ (see, e.g., Appendix A). Thus, for large n, T_n is expected to scale as $1/\sqrt{n}$, which implies that only the odd-n symmetric solutions with $n < T^{-2}$ are stable. This is corroborated by numerical solutions of Eq. (3.10) for large n.

C. Dynamic stability

It has been shown above that the Mattis states are the only stable mean-field solutions near $T = 1$. Below $T = 0.461$ some of the odd-n symmetric solutions become local minima as well, whereas the even-n ones remain unstable at all T. In addition, at low temperature some of the asymmetric solutions become locally stable as will be discussed in Sec. IV. The significance of the locally stable solutions to the dynamic evolution of the system depends on the basin of attraction in phase space of each of these states and on the energy barriers that separate them from each other or from the ground states. It is quite hard to estimate the basins of attraction, though one generally expects that states higher in free energy have significantly smaller basins of attraction than those of the ground states. The barriers are easier to estimate. Since fluctuations of the free energy per spin about these states are finite, the free energy barriers separating them are all proportional to the size of the system, N. Hence, all local minima of the mean-field free energy are true metastable states. The lifetime of such a metastable state is proportional to $\exp(N \Delta f)$ where Δf is the difference between the free energy per spin of the metastable state and that of the lowest saddle point above it. For instance, the lowest energy path from the $n = 3$ state $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$ to the $n = 1$ state $(1, 0, \ldots, 0)$ passes through the $n = 2$ saddle point $(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$, yielding an energy barrier per spin

$$\Delta f = f_2(T = 0) - f_3(T = 0) = 0.175$$

[see Eqs. (2.33) and (2.34)].

IV. ASYMMETRIC SOLUTIONS OF THE GENERALIZED HOPFIELD MODEL

In Sec. II it was shown that solutions which appear continuously at $T = 1$ are symmetric, namely all nonzero components of $\mathbf{m}$ are equal in magnitude. At low temperatures, however, additional saddle points appear[14] which are asymmetric. The appearance of these additional solutions becomes apparent by following the change in stability of the symmetric saddle points as the temperature is reduced. When a particular saddle point changes stability in a certain direction, it does not usually exchange stability with another existing symmetric saddle point, which lies in that direction. Instead, a new, asymmetric saddle point between the two existing saddle points appears. The highest temperature where such a change in stability occurs is when the $n = 2$ symmetric solution

$$\mathbf{m} = (m, m, 0, 0, \ldots, 0) \qquad (4.1)$$

becomes unstable to the mixing of more memories. The eigenvalue that controls this stability is, according to Sec. III,

$$\lambda_2 = 1 - \beta(1-q) . \qquad (4.2)$$

This eigenvalue is positive near $T = 1$ [see Eq. (3.8)], is negative for that solution at $T = 0$, and changes sign at

$$T_2 \simeq 0.575 . \qquad (4.3)$$

Other symmetric saddle points do not change stabilities in

any direction at that temperature. Instead, a new set of saddle points of the form

$$\mathbf{m} = (m, m, \epsilon, \epsilon, \dots, \epsilon, 0, 0, \dots, 0) \ , \quad (4.4)$$

appears, where there are k entries of ϵ and $p - k - 2$ entries of zeros. The magnitude of the new components, ϵ, vanishes continuously as T approaches T_2 from below.

At still lower temperatures when other saddle points change stability, additional sets of asymmetric solutions appear. For instance, at each temperature T_{2k} where λ_2 of an $n = 2k$ symmetric solution becomes negative, a set of saddle points similar to (4.4) appears:

$$\mathbf{m} = (m, m, \dots, m, \epsilon, \epsilon, \dots, \epsilon, 0, 0, \dots, 0) \ , \quad (4.5)$$

where the first $2k$ components are m, the next l components are ϵ, and the last $p - 2k - l$ components are zeros. The temperatures $T_4 \simeq 0.465$ and $T_6 \simeq 0.408$ and T_{2k} decreases as $1/\sqrt{k}$ as $k \to \infty$. Of course, other saddle points with even more complicated asymmetry also develop. As $T \to 0$, some of these asymmetric saddle points merge with the symmetric ones. For instance, the $(m, m, \epsilon, 0, 0, \dots, 0)$ which appears below T_2 approaches very rapidly the $n = 3$ symmetric state $(m, m, m, 0, 0, \dots, 0)$. On the other hand, many of them do remain distinct even at $T = 0$, with energies which are always higher than the $n = 3$ symmetric value -0.375. Two examples with $n = 5$ and 6 are

$$\mathbf{m} = (\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{4}, 0, 0, \dots, 0), \quad E = -0.344 \quad (4.6)$$

$$\mathbf{m} = (\tfrac{3}{8}, \tfrac{3}{8}, \tfrac{3}{8}, \tfrac{3}{8}, \tfrac{3}{16}, 0, 0, \dots, 0), \quad E = -0.334 \quad (4.7)$$

which satisfy Eq. (2.19), as can be checked explicitly.

Although most of the asymmetric solutions seem to follow the scenario described above of a continuous appearance below a critical temperature, we have also encountered a few solutions which appear *discontinuously*, for instance, the solution

$$\mathbf{m} = (\tfrac{3}{8}, \tfrac{3}{8}, \tfrac{1}{4}, \tfrac{1}{4}, \tfrac{1}{2}, 0, 0, \dots, 0) \quad (4.8)$$

below $T \simeq 0.085$. However, this phenomenon is apparently restricted to very low T. Also, none of the discontinuous solutions that we found were stable.

The stability analysis of the various solutions at finite T is increasingly difficult as the symmetry of the solution decreases. In general, they are unstable as they first appear. At lower temperatures some of them do become metastable, as in the case of the odd-n symmetric solutions. The highest temperature where a metastable asymmetric solution was found was

$$T \sim 0.452$$

just below the temperature $T = 0.461$ at which the $n = 3$ symmetric state becomes metastable. At $T = 0$ the criterion of stability of an arbitrary solution is rather simple. Following the same reasoning as in the case of the symmetric solutions (Sec. III) we note that the solutions are stable if their molecular fields are always finite, i.e., that

$$|\mathbf{m} \cdot \boldsymbol{\xi}| > 0 \quad (4.9)$$

holds for all realizations of $\boldsymbol{\xi}$. In such a case all off-diagonal elements [Eq. (3.3)] of the stability matrix are exponentially small as $T \to 0$ and all diagonal elements approach 1 . Such states are surrounded by energy barriers proportional to N. Saddle points in which there is a finite probability that $\mathbf{m} \cdot \boldsymbol{\xi} = 0$ represent states which do not fully order even at $T = 0$. They have some eigenvalues which are proportional to $-\beta$. For instance, the state (4.6) is metastable at $T = 0$ because its molecular field is bounded below by $\tfrac{1}{4}$ and becomes unstable at $T \simeq 0.18$, whereas the solutions (4.7) and (4.8) are unstable even at $T = 0$.

A rather central question is whether the system possesses other, "spurious" states, i.e., states which are not separated by barriers of order N, but are nonetheless long lived, at low T. Such states would surely be stable to single-spin flips at $T = 0$. The answer is that, in the limit of $N \to \infty$ and p finite, the only states which are stable to all single-spin flips are the true metastable states, namely solutions of the mean-field equations which satisfy (4.9). The argument proceeds as follows. One notes that the condition for the stability of a state $\{S_i\}$ to all single-spin flips is that each spin S_i be aligned with its molecular field, namely that

$$
\begin{aligned}
S_i &= \mathrm{sgn} \left[(1/N) \sum_{i(\neq j)} \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu} S_j \right] \\
&= \mathrm{sgn} \left[\mathbf{m} \cdot \boldsymbol{\xi}_i - \frac{p}{N} S_i \right] \ , \quad (4.10)
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{m} &= (1/N) \sum_j \boldsymbol{\xi}_j S_j \\
&= (1/N) \sum_j \boldsymbol{\xi}_j \, \mathrm{sgn} \left[\mathbf{m} \cdot \boldsymbol{\xi}_j - \frac{p}{N} S_j \right] \ . \quad (4.11)
\end{aligned}
$$

As long as $\mathbf{m} \cdot \boldsymbol{\xi}_i$ has a nonzero lower bound, the term $(p/N) S_i$ can be neglected and Eqs. (4.10) and (4.11) become precisely the saddle-point equations of the mean-field theory [Eqs. (2.9) and (2.7)] at $T = 0$ and have the same set of fluctuations. On the other hand, states which have a finite fraction of sites with $\mathbf{m} \cdot \boldsymbol{\xi}_i = 0$ are rendered unstable by the term $-(p/N) S_i$.

V. THERMODYNAMICS OF THE LITTLE MODEL

The synchronous dynamic process introduced by Little leads to a stationary Gibbs distribution of states with the effective Hamiltonian

$$\overline{H} = -\frac{1}{\beta} \sum_i \ln \left[2 \cosh \left[\beta \sum_j J_{ij} S_j \right] \right] \quad (5.1)$$

as was shown by Peretto.[3] In the limit of $T = 0$, Eq. (5.1) reads

$$\overline{H} = -\sum_i \left| \sum_j J_{ij} S_j \right| . \qquad (5.2)$$

We study the thermodynamic properties of the model

(5.1), with the same J_{ij} as in the Hopfield model, i.e., $J_{ij} = N^{-1} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu$ for $i \neq j$, and $\{\xi_i\}$ is distributed according to Eq. (2.10). The partition function of the Hamiltonian (5.1) can be written as

$$
\begin{aligned}
Z &= \operatorname*{Tr}_s \exp(-\beta \overline{H}) \\
&= \operatorname*{Tr}_s \int \left[\prod_\mu dm^\mu \right] \exp\left[\sum_i \ln[2\cosh(\beta \xi_i \cdot \mathbf{m})] \right] \delta\left(\mathbf{m} - N^{-1} \sum_i \xi_i S_i \right) \\
&= \left[\frac{\beta N}{2\pi} \right]^p \int \left[\prod_\mu dm^\mu dt^\mu \right] \exp\left[-N\beta \mathbf{t} \cdot \mathbf{m} + \sum_i \ln[2\cosh(\beta \xi_i \cdot \mathbf{m})] + \sum_i \ln[2\cosh(\beta \xi_i \cdot \mathbf{t})] \right] .
\end{aligned} \qquad (5.3)
$$

The contours of integration in the complex planes of m^μ and t^μ are understood to be analytically deformed, so that they pass the saddle point. Using again the self-averaging property of the free energy, the saddle-point equations reduce to

$$\mathbf{m} = \langle\!\langle \xi \tanh(\beta \mathbf{t} \cdot \xi) \rangle\!\rangle , \qquad (5.4)$$

$$\mathbf{t} = \langle\!\langle \xi \tanh(\beta \mathbf{m} \cdot \xi) \rangle\!\rangle , \qquad (5.5)$$

and the free energy density at the saddle point is

$$
\begin{aligned}
f(\beta) = \mathbf{t} \cdot \mathbf{m} &- \frac{1}{\beta} \langle\!\langle \ln[2\cosh(\beta \xi \cdot \mathbf{m})] \rangle\!\rangle \\
&- \frac{1}{\beta} \langle\!\langle \ln[2\cosh(\beta \xi \cdot \mathbf{t})] \rangle\!\rangle .
\end{aligned} \qquad (5.6)
$$

Although the theory contains two order parameters $\mathbf{t}$ and $\mathbf{m}$, all mean-field solutions obey

$$\mathbf{t} = \mathbf{m} . \qquad (5.7)$$

To prove this, we subtract Eqs. (5.5) from (5.4) to obtain

$$\mathbf{m} - \mathbf{t} = \langle\!\langle \xi[\tanh(\beta \xi \cdot \mathbf{t}) - \tanh(\beta \xi \cdot \mathbf{m})] \rangle\!\rangle ,$$

from which it follows that

$$
\begin{aligned}
\sum_\mu (m^\mu - t^\mu)^2 = \langle\!\langle (\xi \cdot \mathbf{m} - \xi \cdot \mathbf{t})[\tanh(\beta \xi \cdot \mathbf{t}) \\
- \tanh(\beta \xi \cdot \mathbf{m})] \rangle\!\rangle .
\end{aligned}
$$

The right-hand side is obviously *nonpositive*, hence the two sides must be zero, implying the equality (5.7). Substituting Eq. (5.7) in Eqs. (5.4)—(5.6) reduces them to

$$f(\beta) = \mathbf{m}^2 - \frac{2}{\beta} \langle\!\langle \ln[2\cosh(\beta \xi \cdot \mathbf{m})] \rangle\!\rangle , \qquad (5.8)$$

$$\mathbf{m} = \langle\!\langle \xi \tanh(\beta \xi \cdot \mathbf{m}) \rangle\!\rangle . \qquad (5.9)$$

Comparison of Eqs. (5.8) and (5.9) with Eqs. (2.6) and (2.7) reveals that *the Little free energy is exactly twice the free energy of the Hopfield model, at all T*. Both have the same mean-field equations for $\mathbf{m}$. Thus, below $T = 1$ the Little model has the same ground states (i.e., the Mattis states) and saddle points as the generalized Hopfield model. Moreover, as the stability analysis below will show, also the metastability of the saddle points is the same in both models.

In order to perform a stability analysis of the saddle points of (5.3) we have to use the variables appropriate to the rotated contours of integration. These variables are x_μ and y_μ defined via

$$\delta t_\mu = \frac{1}{\sqrt{2}}(x_\mu + iy_\mu), \quad \delta m_\mu = \frac{1}{\sqrt{2}}(x_\mu - iy_\mu) , \qquad (5.10)$$

where δt_μ and δm_μ are the deviations of t_μ and m_μ along their contours, from their saddle-point values. In terms of these variables the stability $2p \times 2p$ matrix consists of the following two $p \times p$ blocks:

$$
\begin{aligned}
(\underline{A}_x)^{\mu\nu} &= \frac{\partial^2 f}{\partial x_\mu \partial x_\nu} \\
&= \delta^{\mu\nu}(1 - \beta) + \beta \langle\!\langle \xi^\mu \xi^\nu \tanh^2(\beta \xi \cdot \mathbf{m}) \rangle\!\rangle
\end{aligned} \qquad (5.11)
$$

and

$$
\begin{aligned}
(\underline{A}_y)^{\mu\nu} &= \frac{\partial^2 f}{\partial y_\mu \partial y_\nu} \\
&= \delta^{\mu\nu}(1 + \beta) - \beta \langle\!\langle \xi^\mu \xi^\nu \tanh^2(\beta \xi \cdot \mathbf{m}) \rangle\!\rangle .
\end{aligned} \qquad (5.12)
$$

The mixed derivatives $\partial^2 f / \partial x_\mu \partial y_\nu$ are zero.

Comparison of Eqs. (5.11) and (5.12) with the stability matrix $\underline{A}$ (3.1) of the generalized Hopfield model reveals that $\underline{A}_x$ is identical to $\underline{A}$ and $\underline{A}_y$ to $2\underline{I} - \underline{A}$. Since the eigenvalues of A are bounded from above by 1, $\underline{A}_y$ is positive definite and the stability of the saddle points is determined by the same stability matrix as in the Hopfield model. Hence, the analysis of Secs. III and IV applies here as well. In particular, the Mattis states are the only locally stable states near $T_c = 1$. Additional local minima appear only below $T = 0.461$.

It has been pointed out[3] that the synchronous dynamics of Little may lead, at $T = 0$, to indefinite cycles of transitions between some of the ground states of the effective Hamiltonian, which occur in one or a small number of time steps. This phenomenon is absent in the single-spin dynamics of Hopfield. As an example, consider a d-dimensional hypercubic lattice with nearest-neighbor ferromagnetic interaction, $J > 0$. In a single-spin flip dynamics the system will spend most of its time, at low temperatures, in one of the two ferromagnetic ground

states, with a very small probability of hopping from one of these states to the other, in a finite time. On the other hand, the Hamiltonian (5.1) [or (5.2)] has four ground states: the two ferromagnetic states and the two antiferromagnetic states. In each of the antiferromagnetic states every spin is antiparallel to its molecular field $-\sum_j J_{ij} S_j$ and therefore, according to the dynamics implied by Eq. (1.8), wants to flip. Consequently, starting from the antiferromagnetic states, the system will make a transition in a single time step to the time-reversed (antiferromagnetic) state and vice versa.

Could such cycles appear in our case? The answer is no. A necessary condition for these cycles to occur is that some of the ground states of the Hamiltonian $\bar{H}$ contain spins which are antiparallel to their molecular field. However, we have argued here that in the $N \to \infty$ limit the ground states (as well as the metastable states) of the Hamiltonian (5.1) are also local minima of the Hopfield Hamiltonian

$$H = -\tfrac{1}{2} \sum_i \left[S_i \left[\sum_j J_{ij} S_j \right] \right] .$$

This necessarily implies that each spin S_i in these states is aligned with $\sum_j J_{ij} S_j = \mathbf{m} \cdot \boldsymbol{\xi}$, as was discussed in Sec. IV. Thus, not only the thermodynamic properties of the two models but also their long-time behavior is the same.

VI. GENERAL DISTRIBUTION OF MEMORIES

A. General distribution of $\{\xi_i^\mu\}$

We consider here a general distribution $P(\xi_i)$ which is invariant under reflections $\xi_i^\mu \to -\xi_i^\mu$ (for each μ separately) and permutations of the components ξ_i^μ. For convenience we will normalize the variance by $\langle\!\langle (\xi_i^\mu)^2 \rangle\!\rangle = 1$. The mean-field equations (2.6) and (2.7) hold, of course, for an arbitrary distribution and imply a phase transition to a broken symmetry state at $T=1$. The Mattis solutions as well as the class of symmetric solutions, of the form (2.22), always exist at all $T<1$. Moreover, expanding Eqs. (2.7) order by order in perturbation about $T=1$ shows that for almost all distributions of the form

$$P(\xi_i) = \prod_\mu p(\xi_i^\mu) , \tag{6.1}$$

asymmetric solutions do not exist near $T=1$. The only exception is the Gaussian distribution which will be discussed in Sec. VI B. Nevertheless, the stability of the various solutions as well as the appearance at low T of the asymmetric solutions depend on the form of the probability $p(\xi_i^\mu)$.

We first determine the conditions under which the *Mattis states become unstable* near $T=1$ and 0 for a general distribution of the form (6.1). The mean-field equation for the Mattis state, $m^\mu = 0$, for all $\mu > 1$ is

$$m^1 \equiv m = \langle\!\langle \xi^1 \tanh(\beta m \xi^1) \rangle\!\rangle . \tag{6.2}$$

Generalizing the stability analysis of Sec. III one obtains for the symmetric solutions with a general distribution the following three eigenvalues:

$$\lambda_1 = 1 - \beta(1 - \bar{q}) + (n-1)\beta Q , \tag{6.3}$$

$$\lambda_2 = 1 - \beta(1 - q) , \tag{6.4}$$

$$\lambda_3 = 1 - \beta(1 - \bar{q}) - \beta Q , \tag{6.5}$$

where $q = \langle\!\langle \tanh^2(\beta z_n) \rangle\!\rangle$, $Q = \langle\!\langle \xi^1 \xi^2 \tanh^2(\beta z_n) \rangle\!\rangle$, and $\bar{q} \equiv \langle\!\langle (\xi^1)^2 \tanh^2(\beta z_n) \rangle\!\rangle$. The variable z_n is $\sum_{\mu=1}^n \xi^\mu$. In the Mattis case $(n=1)$ only the eigenvalues $\lambda_1 = 1 - \beta(1 - \bar{q})$ and $\lambda_2 = 1 - \beta(1 - q)$ exist. A negative λ_2 implies that when $p > 1$ the Mattis state is unstable to mixing of more memories. Expanding Eq. (6.2) in powers of $t = 1 - T_c$ yields $q \cong 3t / \langle\!\langle (\xi^\mu)^4 \rangle\!\rangle$, $1 - \beta(1-q) \cong t[-1 + 3/\langle\!\langle (\xi^\mu)^4 \rangle\!\rangle]$, which implies the instability of the Mattis states near T_c if

$$\langle\!\langle (\xi^\mu)^4 \rangle\!\rangle > 3 . \tag{6.6}$$

The stability at low temperatures depends on the behavior of $p(\xi)$ near the origin. As long as $p(0) = 0$, $1 - \beta(1-q) \sim 1$ and the Mattis states are stable. This is, however, not necessarily so when $P(0) \neq 0$, in which case we have

$$1 - q = \int_{-\infty}^{\infty} d\xi\, p(\xi) \operatorname{sech}^2(\beta m \xi)$$
$$= \frac{T}{m} \int_{-\infty}^{\infty} d\xi\, p(T\xi/m) \operatorname{sech}^2 \xi \simeq \frac{2Tp(0)}{m} , \tag{6.7}$$

where

$$m = m(T=0) = \int d\xi\, p(\xi) |\xi| . \tag{6.8}$$

Thus, the Mattis states are unstable to mixing of more memories at low T, if

$$2p(0) > \langle\!\langle |\xi| \rangle\!\rangle . \tag{6.9}$$

As an example, consider the following distribution:

$$p(\xi) = \frac{a}{\sqrt{2}} e^{-|\xi|/\sqrt{2}} + \frac{1-a}{2} [\delta(\xi - 1) + \delta(\xi + 1)] .$$

Evaluating the inequalities (6.7) and (6.9) one finds that for $a < 0.4$, the Mattis states are stable at all $T < 1$; for $a > (1 + 1/\sqrt{2})^{-1} \simeq 0.6$, the Mattis states are unstable at all T; and for $0.4 < a < (1 + 1/\sqrt{2})^{-1}$, they are unstable near T_c and stable at low temperatures.

Using a continuous distribution of $\{\xi_i^\mu\}$ may have a similar effect as increasing temperature. It may smooth the free energy surface and eliminate the rich structure of metastable states and saddle points that exist in the ± 1 case at low temperature, as was described above. We have demonstrated this effect by studying the mean-field equations with a rectangular distribution,

$$p(\xi_i^\mu) = \frac{1}{l}, \quad \frac{-l}{2} \leq \xi_i^\mu \leq \frac{l}{2} . \tag{6.10}$$

Investigating the symmetric solutions with $n \leq 3$, we have found that the only stable solutions at all T are the $n=1$ states. Furthermore, the $n=2$ and 3 solutions do not change stability in any direction and their free energies do not cross, at all T below 1. These last properties imply that there are no topological constraints which would force the generation at lower temperatures of additional, asymmetric saddle points. Thus, it is quite possible that

in this case, or with other continuous distributions, the only dynamically stable states are the Mattis states at all $T < 1$ and p.[15]

B. Rotationally invariant distributions

In certain circumstances the appropriate distribution of the random vectors ξ_i is invariant under arbitrary $O(p)$ rotations of ξ_i. In the case of random axis ferromagnets the ξ_i is the local direction of the easy axis. In the absence of bulk anisotropy these directions are uniformly distributed on the three-dimensional unit sphere.

In the context of models of memory, rotational invariance emerges in the case of a *Gaussian* distribution,

$$P(\xi_i) = \prod_\mu p(\xi_i^\mu) ,$$

$$p(\xi) = \frac{1}{\sqrt{2\pi}} \exp(-\xi^2/2) . \tag{6.11}$$

The ξ_i^μ will have a Gaussian distribution if, e.g., each one of them is itself a sum of many independent random variables. Rotationally invariant distributions lead to thermodynamic behavior which is qualitatively very different from that of Eq. (2.10). The free energy (2.6) depends in this case only on the amplitude

$$m^2 = \frac{1}{p} \sum_{\mu=1}^p (m_\mu)^2$$

which is determined by Eq. (2.7), whereas the direction of m is left arbitrary. Thus, in the limit $N \to \infty$ the manifold of ground states has a continuous degeneracy, similar to $O(p)$ uniform models.

As an example we work out explicitly the case of a Gaussian distribution, Eq. (6.11). Rotating the μ axes, so that the direction of m coincides with one of the axes, Eqs. (2.6) and (2.7) read

$$f = (p/2)m^2 - \frac{1}{\beta} \int_{-\infty}^\infty \frac{d\xi}{\sqrt{2\pi}} e^{-\xi^2/2} \ln[2\cosh(\beta m\sqrt{p}\,\xi)] ,$$

$$\tag{6.12}$$

$$m = \frac{1}{\sqrt{p}} \int_{-\infty}^\infty \frac{d\xi}{\sqrt{2\pi}} e^{-\xi^2/2}\xi \tanh(\beta m\sqrt{p}\,\xi) . \tag{6.13}$$

Integrating Eq. (6.13) by parts leads to the relation

$$1 = \beta \int_{-\infty}^\infty \frac{d\xi}{\sqrt{2\pi}} e^{-\xi^2/2}\mathrm{sech}^2(\beta m\sqrt{p}\,\xi) = \beta(1-q) \tag{6.14}$$

which is an implicit equation for m. It is interesting that in this case the local susceptibility [which is $\beta(1-q)$] is constant below T_c, just as in the infinite-range SK model.

At $T = 0$, $m = \sqrt{2/\pi p}$ and $f(T=0) = -1/\pi$. As for fluctuations, the eigenvalue $\lambda_1 = 1 - \beta(1-\bar{q})$ corresponds to amplitude fluctuations and is positive at all $T < 1$. On the other hand, there are $p-1$ degenerate modes, with eigenvalue $1 - \beta(1-q)$ which, according to Eq. (6.14), is identically zero in the ordered phase. These modes correspond to *transverse* fluctuations (changing the direction of

m) and hence are marginal.

The continuous symmetry in these models has been studied in some detail in the context of random axis models.[11] It should be noted that, unlike $O(p)$ uniform models, the continuous degeneracy of the ground state in our case is valid only in the thermodynamic limit. In any finite system there will be N possible directions of m, one of which will be singled out by the fluctuations in ξ_i as the ground state of the system. This state will have, in general, projections on all the original memories.

VII. DISCUSSION

One of the main results of this work is that despite the difference in the dynamic mechanisms of the Little and Hopfield models, the long-time behavior, so essential for the retrieval of memories, is identical in the two models. While the dynamic properties of these models have not been analyzed in this paper, extensive numerical simulations of the two processes have been carried out. They confirm their overall similarity.

Previous numerical studies of these models have noticed the existence of "spurious states," namely stable configurations of the network which deviate significantly from the original embedded patterns. Here we have shown that in the limit of large networks all these spurious states are not random but correspond to well-defined mixtures of several patterns.

We have shown that these metastable states appear only at low temperatures, whereas in the temperature range $0.46 < T < 1$ only the states which are correlated with single memories are stable. This suggests that thermal noise plays an important role in enhancing the efficiency of these systems. This efficiency also depends rather strongly on the distribution, of the learned information. Using a continuous distribution, rather than a discrete one, may increase or destroy the dynamic stability of the embedded patterns, depending on the details of that distribution. So far only uncorrelated patterns have been discussed. Introducing correlations between them may, of course, change significantly the properties of the system.

Next we address the question of the storage capacity of these model networks. Throughout this work we have assumed the limit of an infinite size (N) network, with a finite number (p) of stored patterns. *In this limit* the structure of the low-lying states, as well as the barriers between them, are not affected by the increase of p. However, this limit represents a rather modest storage capacity. It is important to know whether this capacity can still increase as p becomes of the order of N^x for some positive x. We have studied the case $x = 1$,[17] $p = \alpha N$, and found that in this limit the system becomes a spin glass. Yet, for small enough values of α the system preserves its quality as an associative memory.

Finally, we mention a few of the issues that deserve further investigation:

(1) The crossover from the finite p behavior to the spin-glass behavior when p becomes of order N;[17]

(2) the detailed dynamic properties of the Little and Hopfield models, e.g., the rate of relaxation and the sizes of the basins of attraction of the various stable states;

(3) the possible exploitation, in the information theoretic sense, of the long-lived mixed states;

(4) the effect of introducing correlations among the embedded patterns;

(5) the effect of modifying the form of the connections J_{ij}. In particular, adding an asymmetric part to J_{ij} may lead to interesting new dynamic behavior.

APPENDIX A

To compute the explicit value of m_n, at $T=0$, Eq. (2.31), we proceed as follows:

$$\langle\langle \,|z|\,\rangle\rangle = \left\langle\!\!\left\langle z\left[1+\frac{1}{\pi i}\int_{-\infty}^{\infty}\frac{d\theta}{\theta}e^{i\theta z}\right]\right\rangle\!\!\right\rangle \tag{A1}$$

$$=\frac{2^{-n}}{\pi}\int\frac{d\theta}{\theta}\left[-\frac{d}{d\theta}\right]\sum_{k=0}^{n}\begin{bmatrix}n\\k\end{bmatrix}e^{i\theta(2k-n)}=-\frac{2^{-n}}{\pi}\int\frac{d\theta}{\theta}\frac{d}{d\theta}e^{-i\theta n}(1+e^{2i\theta})^{n}=\frac{n}{\pi}\int\frac{d\theta}{\theta}\sin\theta\cos^{n-1}\theta \tag{A2}$$

$$=\frac{n}{2^{2k}}\begin{bmatrix}2k\\k\end{bmatrix}, $$

where $k=n/2$ for even n and $k=(n-1)/2$ for odd n. (See Ref. 16, Eq. 3.832.34.)

Next we show that f_{2k}, Eq. (2.33), decreases monotonically with k and f_{2k+1}, Eq. (2.34), increases monotonically with k. To this end we use the identity

$$\begin{bmatrix}2k+2\\k+1\end{bmatrix}=\frac{2(2k+1)}{k+1}\begin{bmatrix}2k\\k\end{bmatrix}. \tag{A3}$$

One can write

$$f_{2k+3}-f_{2k+1}=\frac{2k+1}{2^{4k+1}}\begin{bmatrix}2k\\k\end{bmatrix}\left[1-\frac{(2k+3)(2k+1)}{4(k+1)^2}\right]>0, \tag{A4}$$

$$f_{2k+2}-f_{2k}=\frac{2k}{2^{4k+1}}\begin{bmatrix}2k\\k\end{bmatrix}\left[1-\frac{(2k+1)^2}{4k(k+1)}\right]<0 \tag{A5}$$

which proves our claim.

It is clear that the two sequences, Eqs. (2.33) and (2.34), have a common limit as $k\to\infty$. To calculate this limit we use the Stirling formula

$$n!\sim\sqrt{2\pi}(n+1)^{n+1/2}e^{-(n+1)} \tag{A6}$$

giving, for the asymptotic form of m_n, with even n,

$$m_n=2^{-n}\begin{bmatrix}n\\n/2\end{bmatrix}=(2/n\pi)^{1/2}. \tag{A7}$$

The limiting value for f_n is obtained by substituting (A7) in Eq. (2.32). We find

$$\lim_{n\to\infty}f_n=-1/\pi. \tag{A8}$$

APPENDIX B

In the following we show that the off-diagonal elements of the stability matrix $\underline{A}$, Eqs. (3.1) and (3.2), are non-negative, at all T, for all the saddle points given by Eqs. (2.22). Specifically, we will show that

$$B=\langle\langle \xi^1\xi^2\tanh^2(xz_n)\rangle\rangle \tag{B1}$$

is non-negative for all x.

For definiteness we choose $x>0$. Since ξ^μ takes on the values ±1, z_n is distributed according to

$$p(z_n)=2^{-n}\begin{bmatrix}n\\k\end{bmatrix}, \quad k=(z_n+n)/2$$

as in (2.25). B can be written as

$$B=2^{-(n-1)}\sum_{k=0}^{n-2}\begin{bmatrix}n-2\\k\end{bmatrix}[\tanh^2(2k-n+4)x$$

$$-\tanh^2(2k-n+2)]. \tag{B2}$$

Note that $\tanh^2 y$ is an increasing function of y for $y>0$ and decreasing for $y<0$. In general, some of the terms on the right-hand side of (B2) are negative, but they are outweighed by the positive terms.

To see this, the sum in (B2) is divided in the following way. (i) Remove the term with $k=n-2$. It is manifestly positive. (ii) For odd n the term with $k=(n-3)/2$ vanishes. (iii) The rest of the sum, comprising an even number of terms, for all n, is split into two sums, each with $[(n-2)/2]$ terms, namely

$$D_1=\sum_{k=0}^{[(n-2)/2]-1}\begin{bmatrix}n-2\\k\end{bmatrix}\{\tanh^2[(2k-n+4)x]$$

$$-\tanh^2[(2k-n+2)x]\}$$

and

$$D_2 = \sum_{k=[(n-1)/2]}^{n-3} \begin{bmatrix} n-2 \\ k \end{bmatrix} \{ \tanh^2[(2k-n+4)x]$$

$$-\tanh^2[(2k-n+2)x] \} \ .$$

Substituting $n-3-k$ for k in the second sum, it becomes

$$\sum_{k=0}^{[(n-2)/2]-1} \begin{bmatrix} n-2 \\ k+1 \end{bmatrix} \{ \tanh^2[(2k-n+2)x]$$

$$-\tanh^2[(2k-n+4)x] \}$$

after use has been made of the identities

$$n-3-[(n-1)/2]=[(n-2)/2]-1$$

and

$$\begin{bmatrix} n-2 \\ n-3-k \end{bmatrix} = \begin{bmatrix} n-2 \\ k+1 \end{bmatrix}$$

and of the fact that the square of the hyperbolic function is an even function. B can now be rewritten as

$$B = 2^{-(n-1)} \left\{ \sum_{k=0}^{[(n-2)/2]-1} \left[\begin{bmatrix} n-2 \\ k+1 \end{bmatrix} - \begin{bmatrix} n-2 \\ k \end{bmatrix} \right] (\{ \tanh^2[(2k-n+2)x] - \tanh^2[(2k-n+4)x] \} \right.$$

$$\left. + \{ \tanh^2(nx) - \tanh^2[(n-2)x] \}) \right\} \ . \tag{B3}$$

Each term in the sum (B3) is positive, since both small curly brackets are positive, in the range of variation of k. The last term in the large curly brackets is also positive. Hence $B \geq 0$ for all x.

Next we show that all the eigenvalues of $\underline{A}$ are bounded from above by 1. Since we have shown above that Q is always positive, the largest eigenvalue is

$$\lambda_2 = 1 - \beta + \beta[q+(n-1)Q] \ . \tag{B4}$$

[See Eqs. (3.4)–(3.6).] But

$$\langle\!\langle z_n^2 \tanh^2(m_n z_n) \rangle\!\rangle / n = q + (n-1)Q$$

and hence

$$q + (n-1)Q < \langle\!\langle z_n^2 \rangle\!\rangle / n = 1 \ ,$$

leading to

$$\beta[q+(n-1)Q] < \beta$$

and consequently $\lambda_i (i=1,2) \leq \lambda_3 \leq 1$.

[1]J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982); Proc. Natl. Acad. Sci. USA B **1**, 3088 (1984).

[2]W. A. Little, Math. Biosci. **19**, 101 (1974); W. A. Little and G. L. Shaw, Behav. Biol. **14**, 115 (1975); Math. Biosci. **39**, 281 (1978).

[3]P. Peretto, Biol. Cybern. **50**, 51 (1984).

[4]G. L. Shaw and R. Vasudevan, Math. Biosci. **21**, 207 (1974).

[5]R. J. Glauber, J. Math. Phys. **4**, 294 (1963); K. Binder, in *Fundamental Problems in Statistical Mechanics V,* edited by E. G. D. Cohen (North-Holland, Amsterdam, 1980), p. 21.

[6]A collection of works on spin glasses can be found in *Heidelberg Colloquium on Spin Glasses,* Vol. 192 of *Lecture Notes in Physics,* edited by J. L. Van Hemmen and I. Morgenstern, (Springer, New York, 1983).

[7]S. Kirkpatrick and D. Sherrington, Phys. Rev. B **17**, 4384 (1978).

[8]D. C. Mattis, Phys. Lett. **56A**, 421 (1976).

[9]J. L. Van Hemmen, Phys. Rev. Lett. **49**, 409 (1982); in Ref. 6, p. 203.

[10]J. P. Provost and G. Vallee, Phys. Rev. Lett. **50**, 598 (1983).

[11]C. Jayaprakash and S. Kirkpatrick, Phys. Rev. B **21**, 4072 (1980).

[12]In Ref. 3 the same conclusion has been reached, but with the use of the bound $\sum_{\mu=1}^{p} |m^\mu| \leq 1$. This bound, however, is incorrect and, in fact, is not obeyed by most of the saddle points.

[13]S. F. Edwards and P. W. Anderson, J. Phys. F **5**, 965 (1975).

[14]We are grateful to M. Virasoro and N. Parga for drawing our attention to the existence of asymmetric solutions.

[15]Details will be published elsewhere.

[16]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).

[17]D. J. Amit, H. Gutfreund, and H. Sompolinsky (unpublished).