Spin-glass models of a neural network

J. L. van Hemmen
Sonderforschungsbereich 123 der Universität Heidelberg, D-6900 Heidelberg 1, Federal Republic of Germany
(Received 1 November 1985)

A general theory of spin-glass-like neural networks with a Monte Carlo dynamics and finitely many attractors (stored patterns) is presented. The long-time behavior of these models is determined by the equilibrium statistical mechanics of certain infinite-range Ising spin glasses, whose thermodynamic stability is analyzed in detail. As special cases we consider the Hopfield and the Little model and show that the free energy of the latter is twice that of the former because of a *duplication* of spin variables which occurs in the Little model. It is also indicated how metastable states can be partly suppressed or even completely avoided.

I. INTRODUCTION

There are deep connections between the behavior of complex biological systems and the physics of spin glasses.[1] Intriguing properties such as learning and unlearning,[2-4] fault tolerance with respect to internal failures and input-data errors,[5] and information storage and retrieval[6-8] have been related to the existence of attractive sets (equilibrium states) in the phase space of an Ising spin glass. More specifically, many of these neural-network models have been shown to behave like autoassociative memories.[9] They can be used in the context of pattern recognition and in the design of VLSI (very-large-scale integration) chips. The memory mechanism can be succinctly explained as follows.

A large, neural network consisting of many nonlinear elements is simulated by an Ising spin glass with Monte Carlo (MC) dynamics. For the sake of definiteness we assume that the temperature T equals zero. Then a spin is flipped only if energy is gained, the attractors are associated with the ground states, the basins of attraction with the energy valleys, and the spurious states with *local* energy minima, i.e., metastable states.

Noise can be simulated[1] by taking $T>0$. The previous arguments still hold, with energy replaced by free energy, and attractive fixed points now become attractive sets where the equilibrium states (ergodic components) live. Therefore, the collective long-time behavior of the neural network is governed by the equilibrium statistical mechanics of the underlying Ising spin glass.

This paper addresses the problem of providing the exact free energy and explicit stability criteria for a large class of model Hamiltonians (cost functions). The method of solution is quite general and not restricted to Ising spins, and may be of interest in other contexts, too. It is also indicated how metastable states can be partly suppressed or even completely avoided.

II. NEURAL NETWORKS AND SPIN GLASSES

A neuron may be modeled[10] by a two-state, threshold element with several inputs (synaptic junctions) and one output (the axon). Each neutron is located at a specific site, say i, and its state may be described by an Ising spin variable $S(i)$ which assumes the values ± 1. The input signal to the site i is written $\sum_j J_{ij} S(j)$ and the J_{ij} are called bonds.

The phase space Ω of N neurons (=spins) is the set of all 2^N Ising spin configurations $\{S(i)=\pm 1; 1\le i \le N\}$. In the Hopfield model[5] one assumes threshold zero and takes p *random* configurations $\xi_\alpha=\{\xi_{i\alpha},\ 1\le i\le N\}$, $1\le\alpha\le p$, where the $\xi_{i\alpha}$ are independent, identically distributed random variables which assume the values ± 1 with equal probability. In general, the ξ_α belong to a probability space. Only in the Hopfield case they may be identified with an Ising configuration $\{S(i)=\xi_{i\alpha},\ 1\le i\le N\}$ in Ω. Note that they have *fixed* values, randomly chosen according to their distribution.

The p random configurations ξ_α are then stored in the bonds,

$$J_{ij}=JN^{-1}\sum_{\alpha=1}^{p}\xi_{i\alpha}\xi_{j\alpha},\quad 1\le i,j\le N\ , \tag{1}$$

and a $T=0$ MC spin-flip dynamics is introduced in Ω by requiring that a spin be flipped only if energy is gained. Here the Hamiltonian (energy or cost function) is given by

$$H_N=-\tfrac{1}{2}\sum_{i,j=1}^{N}J_{ij}S(i)S(j)\ . \tag{2}$$

$H_N=H_N(\mathbf{S})$ depends on the specific spin configuration $\mathbf{S}$ and, therefore, is a function on Ω. One starts with a certain configuration $\mathbf{S}_0$, a so-called *key pattern,* which somewhat resembles one of the ξ_α, and tries to retrieve the original pattern, say ξ_γ, by following $\mathbf{S}_t$ as it converges under the system's dynamics to one of the attractive fixed points in phase space. The hoped-for fixed point is ξ_γ but there are also spurious states or ghost patterns[4] and other ξ_α to which the system may converge. If $\mathbf{S}_0$ is not in the right basin of attraction, there is no hope for retrieval of the original pattern.

Another complication one has to take care of is the noise. A noisy system may be described by a finite $-T$ ($T>0$) MC dynamics.[1] Whatever the temperature, the MC dynamics always converges to one of the equilibrium or metastable states of the Hamiltonian (2). Their stability is determined by the free energy and, hence, by the equilibrium statistical mechanics of the underlying Ising-spin system. In case (1), the J_{ij} have mean zero ($i \neq j$) and, accordingly, one might argue that (2) strongly resembles a spin-glass Hamiltonian. For $p=2$, it does not contain frustration, however.[11]

As was already pointed out by Peretto,[1] there is a close analogy between the Hopfield model[3] and a certain spin-glass model.[12–14] Both were proposed at the same time (early 1982) and independently of each other. The spin-glass model is the simplest one that contains both randomness and frustration and whose predictions have been verified not only by experiment but also by more recent, large-scale MC simulations[15] of the three-dimensional $\pm J$ model. I took[12–14]

$$J_{ij}=JN^{-1}(\xi_i\eta_j+\xi_j\eta_i) , \quad (3)$$

where the ξ's and η's are independent, identically distributed random variables which only have to satisfy the requirement that they have mean zero and variance one. The independence is convenient but not necessary. Ergodicity suffices.[16,17] It could be shown (see Ref. 13 for full details) that the stable and metastable states of this model correspond one-to-one to attractive fixed points in a low-dimensional order-parameter space. Moreover, the phase diagram and, consequently, also the stability of the ground states *depend on the probability distribution*.[18] I, therefore, think that this "lack of universality" does occur in memory models specified by Eqs. (1) and (2).

Throughout what follows p in (1) is a fixed but finite positive integer. As compared to Hopfield,[3] a scaling by N^{-1} has been added so as to make the thermodynamics of the model well defined as $N \to \infty$. We now rewrite (2) by using (1),

$$H_N=-\tfrac{1}{2}JN\sum_{\alpha=1}^{p}\left[N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}S(i)\right]^2 , \quad (4)$$

and define the order parameters ($1 \le \alpha \le p$),

$$m_{\alpha,N}=N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}S(i) . \quad (5)$$

It has to be constantly borne in mind that, as $N\to\infty$, for a fixed p-vector $\boldsymbol{\mu}$ there may be several, sometimes very many, spin configurations $\mathbf{S}$ such that (5) gives $\mathbf{m}(\mathbf{S})=\boldsymbol{\mu}$. For instance, for "most" spin configurations $\mathbf{m}=0$. These dominate the high-T behavior. As the temperature is lowered, special configurations take over; cf. Eq. (8) below.

The Hamiltonian (2) is nothing but a quadratic form in the order parameters,

$$H_N=-\tfrac{1}{2}NJ\sum_{\alpha=1}^{p}m_{\alpha,N}^2 . \quad (6)$$

H_N is extensive ($\propto N$) whereas the $m_{\alpha,N}$ are intensive [$O(1)$]. Since[19]

$$N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}\xi_{i\gamma}=O(N^{-1/2}),\quad \alpha\neq\gamma \quad (7)$$

and 1 for $\alpha=\gamma$, one easily verifies that the p stored patterns ξ_α are ground states of (6), i.e., the p-spin configurations

$$\{S(i)=\xi_{i\alpha},\ 1\le i\le N\},\ \ 1\le\alpha\le p \quad (8)$$

minimize H_N. Plainly, if ξ_α is a ground state then $-\xi_\alpha$ is another one.

In this paper a general theory is developed to describe the behavior of spin-glass-like neural networks with Hamilton function

$$H_N=-NF(m_{1N},m_{2N},\ldots,m_{pN})\equiv-NF(\mathbf{m}_N) , \quad (9)$$

where $F(\mathbf{m})=F(m_1,\ldots,m_p)$ is a convex function of the variables m_α. A special case is (6). The convexity requirement is not really necessary but greatly facilitates the stability analysis. The $\xi_{i\alpha}$ are random variables whose probability distribution is assumed to be even around zero. This is convenient but by no means necessary. The method works equally well for more general distributions with nonzero mean. Without loss of generality we may take the variance one. Since the system is endowed with an MC dynamics, our first task is to evaluate the free energy exactly. This will be done in Sec. III A. A general stability analysis is presented in Sec. III B. In Sec. IV we study a family of models that comprises the Hopfield model and show how to some extent one can eliminate the spurious states. The relation between the Little[2] and the Hopfield model is clarified in Sec. V. Finally, a discussion of the results is given in Sec. VI. Though neurons are suitably described by Ising spins, the present considerations are, up to some trivial modifications, equally valid for Heisenberg spins and other, more general n-vector models.

III. GENERAL THEORY

To evaluate the free energy of the Hamilton function (9) we have to perform the trace, a sum over all 2^N Ising spin configurations. We would like to make a change of variables to new coordinates $m_{1N},\ldots,m_{pN}$, and to this end we determine something like a Jacobian[12–14] (Sec. III A). The final answer is obtained through the solution of a fixed-point equation in the p-dimensional order-parameter space. We introduce a dynamics in this space and show that the attracting fixed points can be put in one-to-one correspondence with the stable and metastable stationary states of the free-energy functional. So we have two dynamics: An MC dynamics in the phase space and another, much simpler one in the order-parameter space. As a corollary we obtain a simple stability criterion (Sec. III B). Then the previous results are applied to a simple example (Sec. III C).

A. The free energy

The free energy per spin $f(\beta)$ at the inverse temperature β is defined by

$$-\beta f(\beta)=\lim_{N\to\infty}[N^{-1}\ln \operatorname{Tr}\exp(-\beta H_N)] . \quad (10)$$

For the sake of convenience we divide the trace Tr ( ) by 2^N and consider henceforth the normalized trace, denoted by tr ( ). Then the Ising spins are independent, stochastic variables with mean zero and the normalized trace is their expectation value. By (9) we get

$$-\beta f(\beta)=\lim_{N\to\infty}\{N^{-1}\ln \mathrm{tr} \times\exp[NF(m_{1N},m_{2N},\ldots,m_{pN})]\}\ , \tag{11}$$

where for the moment we have absorbed β in F. In the Appendix of Ref. 14 it was proven that the "Jacobian" we are looking for is

$$\mathrm{Prob}(\mathbf{m}\le\mathbf{m}_N\le\mathbf{m}+d\mathbf{m})\sim\exp[-Nc^*(\mathbf{m})]d\mathbf{m}\ , \tag{12}$$

where $c^*(\mathbf{m})$ is the Legendre transform[20]

$$c^*(\mathbf{m})=\sup_{\mathbf{t}}[\mathbf{m}\cdot\mathbf{t}-c(\mathbf{t})] \tag{13}$$

of the c function

$$c(\mathbf{t})=\lim_{N\to\infty}\left\{N^{-1}\ln \mathrm{tr}\exp\left[\sum_{\alpha=1}^{p}t_\alpha\left[\sum_{i=1}^{N}\xi_{i\alpha}S(i)\right]\right]\right\} =\left\langle\ln\left[\cosh\left[\sum_{\alpha=1}^{p}t_\alpha\xi_\alpha\right]\right]\right\rangle . \tag{14}$$

The function $c(\mathbf{t})$ is convex and so is its Legendre transform $c^*(\mathbf{m})$. The second equality in (14) is obtained by exploiting the ergodicity[21] of the $\xi_{i\alpha}$ so as to reduce an expression with pN random variables, the *fixed* $\xi_{i\alpha}$, to a single average over p random variables ξ_α, $1\le\alpha\le p$. This average is denoted by angular brackets.

Combining (11)–(14) we find, using the Laplace method,

$$-\beta f(\beta)=\lim_{N\to\infty}\left[N^{-1}\ln\int_{\mathbb{R}^p}d\mathbf{m}\exp N[F(\mathbf{m})-c^*(\mathbf{m})]\right] =\sup_{\mathbf{m}}[F(\mathbf{m})-c^*(\mathbf{m})]\ . \tag{15}$$

To get simple analytic expressions we assume that F is continuously differentiable and, for the stability analysis (Sec. III B), that F also has continuous second partial derivatives.

The stationary points of the free-energy functional (15) are those $\mathbf{m}$ which satisfy the equation

$$\nabla F(\mathbf{m})=\nabla c^*(\mathbf{m})\ . \tag{16}$$

By strict convexity[20] $[\partial c^*=(\partial c)^{-1}]$, which we henceforth assume, we may rewrite (16) as

$$\mathbf{m}=\nabla c[\nabla F(\mathbf{m})]\equiv\phi(\mathbf{m})\ . \tag{17}$$

This is a *fixed-point equation* in the p-dimensional order-parameter space. It implicitly depends on the inverse temperature β. Among the solutions $\boldsymbol{\mu}$ of (17) we have to choose the one(s) that maximize(s) the right-hand side of (15). These are called stable. Solutions $\boldsymbol{\mu}$ that only give rise to a *local* maximum are called metastable. Otherwise $\boldsymbol{\mu}$ is an unstable state.

Until now we have not used the specific form (14) of $c(\mathbf{t})$. If we do this and use vector notation, we get instead of (17),

$$\mathbf{m}=\langle\xi\tanh[\xi\cdot\nabla F(\mathbf{m})]\rangle\ . \tag{18}$$

Suppose $\boldsymbol{\mu}$ is a solution of (18). Then

$$\begin{aligned}\mu^2&=\langle\boldsymbol{\mu}\cdot\xi\tanh(\ )\rangle\\ &\le\langle(\boldsymbol{\mu}\cdot\xi)^2\rangle^{1/2}\langle\tanh^2(\ )\rangle^{1/2}\\ &\le\langle(\boldsymbol{\mu}\cdot\xi)^2\rangle^{1/2}=\left[\sum_{\alpha=1}^{p}\mu_\alpha^2\right]^{1/2}=||\boldsymbol{\mu}||\ ,\end{aligned} \tag{19}$$

so that $||\boldsymbol{\mu}||^2\le||\boldsymbol{\mu}||$ and $\boldsymbol{\mu}$ is always in the unit sphere of $\mathbb{R}^p$. If one looks for solutions of a fixed-point equation via an iterative search procedure (see below), it is good to know where they can be found.

We now want to evaluate the free energy (15). Given $\mathbf{m}=\boldsymbol{\mu}$, the variable $\mathbf{t}$ in (13) has to be such that $\boldsymbol{\mu}=\nabla c(\mathbf{t})$. [In fact, this $\mathbf{t}$ is unique if $c(\mathbf{t})$ is strictly convex.[20]] But a solution $\boldsymbol{\mu}$ of the fixed-point equation (17) is of the form $\boldsymbol{\mu}=\nabla c[\nabla F(\boldsymbol{\mu})]$. Hence $\mathbf{t}=\nabla F(\boldsymbol{\mu})$ and, by (13),

$$c^*(\boldsymbol{\mu})=\boldsymbol{\mu}\cdot\nabla F(\boldsymbol{\mu})-c[\nabla F(\boldsymbol{\mu})]\ . \tag{20}$$

Combining (15)–(17) and (20) we obtain

$$-\beta f(\beta)=F(\boldsymbol{\mu})-\boldsymbol{\mu}\cdot\nabla F(\boldsymbol{\mu})+c[\nabla F(\boldsymbol{\mu})]\ , \tag{21}$$

where $\boldsymbol{\mu}$ satisfies (17) and is such that it maximizes the right-hand side of (21). The energy per spin $u(\beta)$ is given by $u(\beta)=-\beta^{-1}F(\boldsymbol{\mu})$ or, returning to the notation of Eq. (9) and bringing in the β again, we get $u(\beta)=-F(\boldsymbol{\mu})$. The entropy per spin $s(\beta)$ now readily follows from $s(\beta)=-\beta f(\beta)+\beta u(\beta)$. One has to add ln2 to get the usual entropy.

The ergodic components[22] of the model are labeled by the vectors $\boldsymbol{\mu}$ which are the stable solutions of (17). Each ergodic component is a free-energy valley separated from neighboring valleys (if any) by a free-energy barrier of height N. Metastable states also correspond to free-energy valleys, but with a higher free energy per spin. As $N\to\infty$, the barriers become infinitely high and the ergodic components become "truly" ergodic. However, for large but finite N there is a negligible probability that once the system is caught in one of the valleys the MC dynamics will let it mount a barrier and wander into another ergodic component.

B. Stability

How do we find the fixed points $\boldsymbol{\mu}=\phi(\boldsymbol{\mu})$ of Eq. (17)? In general, there is no hope for solving (17) analytically, so one must have recourse to numerical iteration. Starting with a well-chosen $\mathbf{m}_0$, one generates a sequence

$$\mathbf{m}_1=\phi(\mathbf{m}_0),\ \ \mathbf{m}_2=\phi(\mathbf{m}_1),\ \ \mathbf{m}_3=\phi(\mathbf{m}_2),\ \ldots\ . \tag{22}$$

In this way one introduces a discrete *dynamics* in the order-parameter space that maps each $\mathbf{m}$ onto $\phi(\mathbf{m})$. The question then is whether $\mathbf{m}_n$ converges to a fixed point $\boldsymbol{\mu}$ and whether $\boldsymbol{\mu}$ gives rise to a (local) maximum of the free-energy functional (15). We define $\boldsymbol{\mu}$ to be a *stable fixed point* if there exists a neighborhood U of $\boldsymbol{\mu}$ such that any sequence (22) which starts in U converges to $\boldsymbol{\mu}$. Otherwise $\boldsymbol{\mu}$ is called unstable.

We now prove the following *theorem:* $\boldsymbol{\mu}$ is a stable fixed point if and only if it gives rise to a local or global maximum of the free-energy functional (15). In other words, μ is an unstable fixed point if and only if it corresponds to an unstable state of (15). A numerical iteration gives physically relevant (stable and metastable) states only.

For the proof we need some results from the theory of functions of several real variables.[23]

(i) Let f: $\mathbb{R}^n \to \mathbb{R}^m$ be a continuously differentiable function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then the derivative $(Df)(a)$ of f at the point a is the $m \times n$ matrix $(\partial f^i/\partial x_j)(a) \equiv (\partial_j f^i)(a)$, with $1 \le i \le m$ and $1 \le j \le n$. Here the components of f have been denoted by f^i.

(ii) Chain rule: For $\mathbb{R}^n \overset{f}{\to} \mathbb{R}^m \overset{g}{\to} \mathbb{R}^p$ we have

$$D(g \circ f)(a) = Dg(f(a))Df(a) . \tag{23}$$

(iii) Inverse-function theorem: Let f: $\mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable with $\det[Df(a)] \neq 0$. Then f^{-1} exists and

$$D(f^{-1})(y) = [Df(f^{-1}(y))]^{-1} . \tag{24}$$

Henceforth we will assume that also $c(\mathbf{t})$ is twice continuously differentiable. This is certainly true for (14).

Now $\mu = \phi(\mu)$ is a stable fixed point if and only if the eigenvalues of $D\phi(\mu)$ are in absolute value less than one. We have to prove that this fact implies and is implied by μ being a (local) maximum of the free-energy functional (15).

We first show that the eigenvalues of $D\phi(\mu)$ are positive (and thus real). Let us denote the components of ϕ by ϕ^i, $1 \le i \le p$. The chain rule (23) gives

$$\begin{aligned} \partial_j \phi^i(\mu) &= \sum_k (\partial_i \partial_k c)(\nabla F(\mu)) \cdot (\partial_k \partial_j F)(\mu) \\ &\equiv \sum_k A_{ik} B_{kj} . \end{aligned} \tag{25}$$

That is, $D\phi(\mu)$ is the product of two matrices, A and B. Both A and B are positive, i.e., are Hermitean and have positive eigenvalues since c and F are convex,[20] c by definition and F by assumption. Only here we exploit the convexity of F. Quite frequently, as in (14), $c(\mathbf{t})$ is *strongly* convex[20] with A such that its eigenvalues do not vanish $(A > 0)$.

Let $\sigma(X)$ denote the set of eigenvalues of a square matrix. Then

$$\sigma(D\phi(\mu)) = \sigma(AB) = \sigma(A^{1/2}BA^{1/2}) \subseteq \mathbb{R}^+ , \tag{26}$$

since $A^{1/2}BA^{1/2} \ge 0$ and for any two square matrices X and Y we always have[24] $\sigma(XY) = \sigma(YX)$. So the eigenvalues of $D\phi(\mu)$ are real and positive. We now show that they are less than one if and only if μ is a (local) maximum of the free-energy functional (15).

Let us consider (15) more carefully and note that μ gives rise to a (local) maximum if and only if (16) holds and in addition

$$(\partial_i \partial_j F)(\mu) < (\partial_i \partial_j c^*)(\mu) . \tag{27}$$

Equation (27) should not be read elementwise but interpreted as a matrix inequality. It is a transcription of the requirement that the second derivative of (15) be strictly negative-definite.

In agreement with general usage[20] we write $\nabla c^* = \partial c^*$. Then ∂c^* maps $\mathbb{R}^p$ into $\mathbb{R}^p$ and $\partial c^* = (\partial c)^{-1}$. Moreover, the matrix elements of $D(\partial c^*)$ are just the $\partial_i \partial_j c^* = \partial_j \partial_i c^*$ which occur in (27). Combining the inverse-function theorem (24) and the relation $\partial c^* = (\partial c)^{-1}$ we get, using (16),

$$\begin{aligned} D(\partial c^*)(\mu) &= D((\partial c)^{-1})(\mu) \\ &= [D(\partial c)(\partial c^{-1}(\mu))]^{-1} \\ &= [D(\partial c)(\partial c^*(\mu))]^{-1} \\ &= [D(\partial c)(\nabla F(\mu))]^{-1} . \end{aligned} \tag{28}$$

Now the matrix elements of $D(\partial c)$ are $\partial_i \partial_j c$ so that (27) may be rewritten

$$B < A^{-1} , \tag{29}$$

where B and A where defined by (25). But $B < A^{-1}$ is equivalent to $A^{1/2}BA^{1/2} < \mathbb{1}$, whence $\sigma(AB) = \sigma(A^{1/2}BA^{1/2}) < 1$ and this is what the proof should prove.

As a corollary we note that the *stability criterion* which follows from (25), (27), and (29),

$$\sigma(AB) < 1 , \tag{30}$$

is quite convenient. The matrices A and B are readily calculated and in practice it frequently happens that $F(\mathbf{m}) = \sum_\alpha g(m_\alpha)$ with g a convex function of one real variable. Then $B = (\partial_i \partial_j F)(\mu)$ is a diagonal matrix. The criterion (30) also holds when F is not convex. Since $\sigma(AB) = \sigma(A^{1/2}BA^{1/2})$, the eigenvalues of AB are all real.

The above theorem has an interesting application to pattern recognition. One specifies a key pattern at $t = 0$ that is a somewhat-blurred version of one of the stored patterns. Then one allows the system to relax in phase space to an attracting fixed point of the MC dynamics *or* in the p-dimensional order-parameter space by determining the order parameters (5) at $t = 0$ and iterating the mapping (22), i.e., ϕ. As yet it is not clear whether the basin of attraction of a fixed point μ in the order-parameter space corresponds one-to-one to sets of configurations in phase space that relax via the MC dynamics to the support of an ergodic component (= bottom of a free energy valley) corresponding to μ.

C. A simple example

We finish this section by studying a simple example,

$$H_N = -NJx^{-1} \left| N^{-1} \sum_{i=1}^{N} S(i) \right|^x , \tag{31}$$

which does not contain any randomness. We assume $x > 1$. The case $x = 1$ will be treated in Sec. IV. Here $p = 1$ and $F(m) = \beta J x^{-1} |m|^x$, which is convex. We may immediately apply our general theory since we nowhere used the requirement that the $\xi_{i\alpha}$ have zero mean. So we take $\xi_{i1} \equiv 1$ in (5) and (14), find $c(t) = \ln[\cosh(t)]$, and see that (21) reads

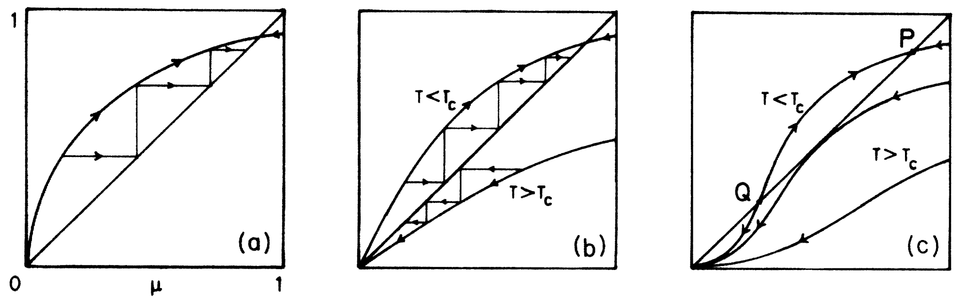

FIG. 1. Graphical solution of the fixed-point equation (33). There are three cases: (a) $1<x<2$; μ^{x-1} has an infinite derivative at $\mu=0^+$ and, hence, whatever the temperature, one always gets a nonzero, stable fixed point (P). (b) $x=2$; the Curie-Weiss model. Through a simple geometric construction[26] one checks the (in)stability of the fixed points, $\mu=0$ and P if $T>T_c$. By the theorem (Sec. II B) we then know immediately which fixed point maximizes the free-energy functional (32). The transition at $T_c(\beta_c J=1)$ is second order. (c) For $x>2$, the function μ^{x-1} has a vanishing derivative at $\mu=0^+$ and, hence, $\tanh(\beta J\mu^{x-1})$ has an S shape. At T_c (middle curve) the system undergoes a first-order phase transition. For $T<T_c$, there are three fixed points, $\mu=0$, P and Q. Whereas Q is unstable, the other two are stable. An evaluation of the free energy (32) shows that $\mu=0$ gives rise to a metastable state while P corresponds to a global maximum.

$$-\beta f(\beta)=\beta J(x^{-1}-1)\,|\mu|^x+\ln[\cosh(\beta J\,|\mu|^{x-1})]\ , \tag{32}$$

while μ satisfies the fixed-point equation

$$\mu=\tanh[\beta J \operatorname{sgn}(\mu)\,|\mu|^{x-1}]\ ,$$

i.e.,

$$|\mu|=\tanh[\beta J\,|\mu|^{x-1}]\ . \tag{33}$$

If μ is a solution to (33), then $-\mu$ is another solution. So we may assume $\mu\geq 0$. We now analyze Eq. (33) for the three cases (a) $1<x<2$, (b) $x=2$, and (c) $x>2$. See also Fig. 1.

Case (a) does not exhibit any transition. Whatever the temperature, we always get a stable, nonzero fixed point since the derivative of $\tanh[\beta J\mu^{x-1}]$ diverges at $x=0^+$.

Case (b) is well known. It is the Curie-Weiss model and the only one that can be solved by elementary means.[25] If $\beta J<1$, there is a unique solution $\mu=0$ whereas for $\beta J>1$ the solution $\mu=0$ has lost its stability and a new, stable $\mu>0$ bifurcates away from zero. Hence the transition at T_c is second order.

Case (c) is quite interesting. Since μ^{x-1} has a vanishing derivative at $\mu=0^+$, the right side of (33) represents an S-shaped curve. At $T=T_c$ this curve touches the straight line which represents the left side of (33). For $T<T_c$ we get in addition to $\mu=0$ two more fixed points of which P is stable and Q unstable. The system undergoes a *first-order* transition to the state whose order parameter is determined by P. In the phase space we now get three free-energy valleys corresponding to the two ergodic components with $\mu=\mu_+>0$ and $\mu=-\mu_+$ and one metastable state with $\mu=0$. As $\beta\to\infty$, $\mu=0$ loses its stability and the order-parameter space is divided into two basins of attraction, one for μ_+ and the other one for $-\mu_+$ with $\mu=0$ as their common boundary. Moreover, they are in one-to-one correspondence to the basins of attraction of the MC dynamics in phase space.

But what about a finite β? The above theorem guarantees the correspondence only locally, for a neighborhood of the fixed points. Why then is $\mu=0$ attractive? There are overwhelmingly many spin states satisfying the condition $m_N(\mathbf{S})=N^{-1}\sum_{i=1}^N S(i)\approx 0$. All these contribute to the entropy $s(\beta)$ and count as soon as $\beta^{-1}=T$ is positive because $f(\beta)=u(\beta)-Ts(\beta)$.

IV. AUTOASSOCIATIVE MEMORY MODELS

The importance of autoassociative memory models to pattern recognition is well established.[1–9] One stores several patterns and if a certain key pattern, which somewhat resembles one of these, is received the problem is to retrieve the original version. There is, however, a serious difficulty that hinders an efficient retrieval: Not only are the original patterns available as attracting fixed points of a suitable dynamics (here of the MC type) but also other images, ghost patterns or spurious states,[4] appear. Plainly, the latter are not wanted. They are usually associated with metastable states. How can we avoid or, at least, suppress them? To get some insight into this question we study a class of models that comprises the Hopfield model. Particular attention is paid to the low-noise (low-T) behavior and the corresponding microscopic ground states. In Sec. IV A we introduce the model and calculate its free energy, in Sec. IV B and IV C some special solutions to the fixed-point equation are obtained and their stability is studied and finally, in Sec. IV D, we analyze an interesting special case without any metastable state.

A. The model and its free energy

We want to study the following family of models:

$$H_N=-NJ\sum_{\alpha=1}^{p}x^{-1}\left|N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}S(i)\right|^x\ , \tag{34}$$

labeled by $x\geq 1$. For the moment we assume $x>1$ since the case $x=1$ will be considered in Sec. IV D. Throughout this section p is a finite positive integer and the $\xi_{i\alpha}$ have *fixed* values, randomly chosen according to their distribution (even; mean zero and variance one). For $x=2$ and $\xi_{i\alpha}=\pm 1$, the model is the one proposed by Hopfield.[3]

Comparing (34) with (11) we see that we have to put $F(\mathbf{m})=\beta J\sum_{\alpha=1}^{p} x^{-1}|m_\alpha|^x$. For the values of x chosen each of the $|m_\alpha|^x$ is convex and continuously differentiable and so is their sum $F(\mathbf{m})$. Using Eqs. (14), (18), and (21) we find that the free energy is

$$-\beta f(\beta)=\beta J\sum_{\alpha=1}^{p}(x^{-1}-1)\mu_\alpha^x + \left\langle \ln\left[\cosh\left[\beta J\sum_{\alpha=1}^{p}\mu_\alpha^{x-1}\xi_\alpha\right]\right]\right\rangle , \qquad (35)$$

with fixed point-equation

$$\boldsymbol{\mu}=\left\langle \boldsymbol{\xi}\tanh\left[\beta J\sum_{\alpha=1}^{p}\xi_\alpha\mu_\alpha^{x-1}\right]\right\rangle \equiv \langle \boldsymbol{\xi}\tanh[\beta J(\boldsymbol{\xi}\cdot\boldsymbol{\mu}^{x-1})]\rangle . \qquad (36)$$

In (35) and (36) we have used the fact that the distribution of the ξ_α's is even to reduce everything to the positive orthant $\{\mu_\alpha\geq 0;\ 1\leq\alpha\leq p\}$. In general, one has to replace $\boldsymbol{\mu}$ by $|\boldsymbol{\mu}|=(|\mu_1|,|\mu_2|,\ldots,|\mu_p|)$. As the stability is determined by (30), the problem is in principle solved. We will, however, study some special solutions of (36).

B. Special solutions

For low-enough temperature T (high enough $\beta=1/k_BT$) there are quite a few solutions to (36). See, for instance, Ref. 8 for the case $x=2$ and the ξ's ± 1 or Gaussian. Here we concentrate on some typical examples without specifying the distribution of the ξ's.

(a) We first turn to the simplest possible solution. Suppose that only *one* component of $\boldsymbol{\mu}$, say μ_α, is nonzero. Then, by (36),

$$\mu_\alpha=\langle \xi_\alpha\tanh(\beta J\mu_\alpha^{x-1}\xi_\alpha)\rangle , \qquad (37)$$

and the other components μ_γ, $\gamma\neq\alpha$, vanish identically—in agreement with our ansatz. The right side of (37) does not depend on α and we, therefore, may drop it. As $\beta\to\infty$ ($T\downarrow 0$) and $\mu_\alpha\neq 0$ we end up with

$$\mu_\alpha=\langle|\xi|\rangle\leq\langle\xi^2\rangle^{1/2}=1 , \qquad (38)$$

in agreement with (19). Here, as in (19), the inequality follows from the Cauchy-Schwarz inequality. As we already noted, $\mu_\gamma=0$ when $\gamma\neq\alpha$.

What is the microscopic state corresponding to (38), i.e., which state gives $\mu_\gamma=\langle|\xi|\rangle\delta_{\alpha\gamma}$? It is

$$S(i)=\mathrm{sgn}(\xi_{i\alpha}),\quad 1\leq i\leq N \qquad (39)$$

where $\mathrm{sgn}(z)=1$ if $z>0$, -1 if $z<0$, and 0 if $z=0$. To check that (39) gives the correct values for the order parameters, we use (5) in tandem with the (strong) law of large numbers,[21]

$$m_\gamma=N^{-1}\sum_{i=1}^{N}\xi_{i\gamma}S(i)=N^{-1}\sum_{i=1}^{N}\xi_{i\gamma}\mathrm{sgn}(\xi_{i\alpha}) =\langle|\xi|\rangle\delta_{\alpha\gamma} .$$

These so-called Mattis states[8] are precisely the ones we are interested in. They contain the stored information (patterns) and are to be retrieved.

(b) Let us suppose, more generally, that $\boldsymbol{\mu}$ contains n nonzero, equal components μ $(1\leq n\leq p)$. Such a $\boldsymbol{\mu}$ is called an *n-symmetric state.* By permutation invariance we may relabel the α and assume that only the first n components of $\boldsymbol{\mu}$ are nonzero. Then $\boldsymbol{\mu}=(\mu,\ldots,\mu,0,\ldots,0)\equiv\mu\mathbf{1}_n$ and, by (36),

$$\mu=\left\langle\xi_\alpha\tanh\left[\beta J\mu^{x-1}\sum_{\gamma=1}^{n}\xi_\gamma\right]\right\rangle,\quad 1\leq\alpha\leq n \qquad (40)$$

while the other components identically vanish. Note that, once again by permutation invariance, the right side of (40) *does not depend on* α. By adding the first n components of $\boldsymbol{\mu}$ and dividing by n we get

$$\mu=n^{-1}\langle(\mathbf{1}_n\cdot\boldsymbol{\xi})\tanh[\beta J\mu^{x-1}(\mathbf{1}_n\cdot\boldsymbol{\xi})]\rangle . \qquad (41)$$

$\mu=0$ is always a solution of (40) and (41). As $\beta\to\infty$ and $\mu\neq 0$ we find

$$\mu=n^{-1}\left\langle\left|\sum_{\gamma=1}^{n}\xi_\gamma\right|\right\rangle\leq n^{-1}\left\langle\left[\sum_{\gamma=1}^{n}\xi_\gamma\right]^2\right\rangle^{1/2}=n^{-1/2} , \qquad (42)$$

in agreement with (19).

Returning to (40) we take the limit $\beta\to\infty$ once again and obtain

$$\mu=\left\langle\xi_\alpha\mathrm{sgn}\left[\sum_{\gamma=1}^{n}\xi_\gamma\right]\right\rangle,\quad 1\leq\alpha\leq n \qquad (43)$$

with $\mu_\alpha=0$ for $\alpha>n$. Adding (43) for $1\leq\alpha\leq n$ and dividing by n one regains (42). Moreover, one easily verifies[21] that

$$S(i)=\mathrm{sgn}\left[\sum_{\gamma=1}^{n}\xi_{i\gamma}\right] \qquad (44)$$

is the microscopic state corresponding to (43). Here we assume that $\sum_{\gamma=1}^{n}\xi_{i\gamma}$ does not vanish with probability one. According to (14), the n patterns ξ_γ, $1\leq\gamma\leq n$, have been *mixed* $(n>1)$, which is to be avoided. We will see shortly how this can be done.

How many n-symmetric states can we get? There are $\binom{p}{n}$ ways of selecting n elements out of p and by permutation invariance there are equally many solutions to (36), provided $\mu\neq 0$. In addition, if $\boldsymbol{\mu}=(\mu_\alpha)\neq 0$ is a solution, then $\boldsymbol{\mu}'=(\pm\mu_\alpha)$ is another solution. It is reasonable to incorporate these also. Altogether, there are $2^n\binom{p}{n}$ nonzero, n-symmetric states.

Apparently we now have found quite a few spurious states. Are they stable against the Mattis states (39)? To answer this question we first check the ground-state energy of both states, where $\beta=+\infty$,

$$\text{Eq. (39)} \Longrightarrow u(\infty)/(Jx^{-1}) = -(\langle|\xi|\rangle)^x \geq -1 \ , \tag{45a}$$

$$\text{Eq. (44)} \Longrightarrow u(\infty)/(Jx^{-1}) = -n\left[n^{-1}\left\langle\left|\sum_{\gamma=1}^{n}\xi_\gamma\right|\right\rangle\right]^x \geq -n^{1-x/2} \ . \tag{45b}$$

By the central limit theorem,

$$n^{-1}\left\langle\left|\sum_{\gamma=1}^{n}\xi_\gamma\right|\right\rangle = n^{-1/2}\left\langle\left|\frac{1}{\sqrt{n}}\sum_{\gamma=1}^{n}\xi_\gamma\right|\right\rangle = O(n^{-1/2}) \tag{46}$$

as n becomes large and then the *in*equality in (45b) becomes an equality, apart from the multiplicative constant $(\sqrt{2/\pi})^x$. Only in the Gaussian case we have exact equality for all n.

If $1<x<2$, so that $1-\frac{1}{2}x>0$, the spurious n-symmetric states are expected to be stable. This expectation will be nicely confirmed by the case $x=1$, to be treated in Sec. IV D. The larger n, the more stable the spurious states become. If $x=2$, Eq. (45) leaves the stability problem undecided and one has to perform a detailed calculation of the middle terms, which may be nasty.[27] However, if $x>2$, then $1-\frac{1}{2}x<0$ and the larger n, the more advantageous the Mattis states (39) are. In fact, if "deep valleys are also broad,"[4] which seems reasonable, we would be done. Moreover, in Sec. III C we have seen that if we increase x, then T_c is lowered and at low T only *energy* arguments are important.

(c) Before turning to a stability analysis for finite temperatures ($T>0$) we quickly discuss another type of solution to the fixed-point equation (36). We now make the ansatz that $\boldsymbol{\mu}=(\mu_1,\ldots,\mu_1,\mu_2,\ldots,\mu_2,0,\ldots,0)\equiv\mu_1\mathbf{1}_m+\mu_2\mathbf{1}_n$ with m times a μ_1 and n times a μ_2. The μ_1 and μ_2 satisfy a coupled set of nonlinear equations,

$$\begin{aligned}\mu_1 &= m^{-1}\langle(\mathbf{1}_m\cdot\boldsymbol{\xi})\tanh[\beta J(\mu_1^{x-1}\mathbf{1}_m\cdot\boldsymbol{\xi}+\mu_2^{x-1}\mathbf{1}_n\cdot\boldsymbol{\xi})]\rangle \ ,\\ \mu_2 &= n^{-1}\langle(\mathbf{1}_n\cdot\boldsymbol{\xi})\tanh[\beta J(\mu_1^{x-1}\mathbf{1}_m\cdot\boldsymbol{\xi}+\mu_2^{x-1}\mathbf{1}_n\cdot\boldsymbol{\xi})]\rangle \ .\end{aligned} \tag{47}$$

The components μ_α with $m+n<\alpha\leq p$ vanish identically, which is consistent with the ansatz. If $x=2$, the case $\mu_1\neq\mu_2$ can indeed be realized,[8] and so on. All in all, one should be prepared to find a huge amount of solutions to (36). Whether they are stable is another question, to which we now turn.

C. Stability

According to Sec. III B, the stability of a fixed point $\boldsymbol{\mu}=\nabla c(\nabla F(\boldsymbol{\mu}))$ is determined by the eigenvalues of the matrix $H(\boldsymbol{\mu})=B^{1/2}AB^{1/2}$ where

$$A_{\alpha\gamma}(\boldsymbol{\mu})=(\partial_\alpha\partial_\gamma c)(\nabla F(\boldsymbol{\mu})), \quad B_{\alpha\gamma}=\partial_\alpha\partial_\gamma F(\boldsymbol{\mu}) \ . \tag{48}$$

$\boldsymbol{\mu}$ is stable if and only if all the eigenvalues of $H(\boldsymbol{\mu})$ are less than one; cf. Eq. (30). Since $F(\boldsymbol{\mu})=\sum_{\alpha=1}^{p}\beta Jx^{-1}|m_\alpha|^x$ and $c(\mathbf{t})=\langle\ln[\cosh(\mathbf{t}\cdot\boldsymbol{\xi})]\rangle$, a simple calculation gives for $\boldsymbol{\mu}$ in the positive orthant

$$H_{\alpha\gamma}(\boldsymbol{\mu})=\beta J(x-1)(\mu_\alpha\mu_\gamma)^{x/2-1} \times\langle\xi_\alpha\xi_\gamma\cosh^{-2}[\beta J(\boldsymbol{\xi}\cdot\boldsymbol{\mu}^{x-1})]\rangle \ . \tag{49}$$

Let us first consider the case $\boldsymbol{\mu}=0$. Then, whatever β, H is a diagonal matrix whose elements are either $+\infty$ if $1<x<2$, so $\mu=0$ is unstable, or they all vanish for $x>2$ and $\boldsymbol{\mu}=0$ is stable. For $x=2$, $\boldsymbol{\mu}=0$ is stable only if $\beta J<1$. All this is in agreement with the qualitative picture suggested by Fig. 1. One may object, however, that the function F has continuous, second-order partial derivatives at $\boldsymbol{\mu}=0$ only if $x\geq2$ so that the above argument is not completely correct. Its conclusion is correct, though. In spite of that, we restrict our discussion to $x\geq2$ and relegate the case $1<x<2$ to the Appendix.

For a Mattis state $\boldsymbol{\mu}=(\mu,0,\ldots,0)$ we find that $H(\boldsymbol{\mu})$ is a diagonal matrix with $(p-1)$ elements equal to zero ($x>2$), which is certainly less than the required upper bound (one), or $\beta J\langle\cosh^{-2}\{(\beta J\mu)\xi\}\rangle$ ($x=2$), while in the one-dimensional fixed-point space we get ($x\geq2$)

$$\lambda_1=\beta J(x-1)\mu^{x-2}\langle\xi^2\cosh^{-2}[(\beta J\mu^{x-1})\xi]\rangle \ . \tag{50}$$

The stability of an n-symmetric state $\boldsymbol{\mu}=(\mu,\ldots,\mu,0,\ldots,0)$, which lives in an n-dimensional fixed-point space ($n>1$), is determined by a block-diagonal matrix with two blocks of size $n\times n$ and $(p-n)\times(p-n)$. The diagonal elements of the $n\times n$ matrix are

$$H_{\alpha\alpha}(\boldsymbol{\mu})=\beta J(x-1)\mu^{x-2}\langle\xi_\alpha^2\cosh^{-2}[\beta J(\boldsymbol{\mu}^{x-1}\cdot\boldsymbol{\xi})]\rangle\equiv a \ , \tag{51a}$$

$1\leq\alpha\leq n$. They are all equal, and so are the off-diagonal elements

$$H_{\alpha\gamma}(\boldsymbol{\mu})=\beta J(x-1)\mu^{x-2}\langle\xi_\alpha\xi_\gamma\cosh^{-2}[\beta J(\boldsymbol{\mu}^{x-1}\cdot\boldsymbol{\xi})]\rangle \equiv b \ , \tag{51b}$$

$1\leq\alpha\neq\gamma\leq n$; in general $|b|<a$. So for $H(\boldsymbol{\mu})$ restricted to the n-dimensional fixed-point space we may write

$$H=(a-b)\mathbf{1}+nb\left|\frac{1}{\sqrt{n}}\mathbf{1}_n\right\rangle\left\langle\frac{1}{\sqrt{n}}\mathbf{1}_n\right| \ , \tag{52}$$

where $\mathbf{1}$ is the unit matrix and $\mathbf{1}_n=(1,1,\ldots,1)\in\mathbf{R}^n$. We then get one nondegenerate eigenvalue

$$\begin{aligned}\lambda_1&=(a-b)+nb\\ &=\beta J(x-1)\mu^{x-2}\\ &\quad\times\langle[\xi_\alpha^2+(n-1)\xi_\alpha\xi_\gamma]\cosh^{-2}(\beta J\boldsymbol{\mu}^{x-1}\cdot\boldsymbol{\xi})\rangle \ ,\end{aligned} \tag{53a}$$

and another, $(n-1)$-degenerate eigenvalue

$$\lambda_2=a-b=\beta J(x-1)\mu^{x-2} \times\langle(\xi_\alpha^2-\xi_\alpha\xi_\gamma)\cosh^{-2}(\beta J\boldsymbol{\mu}^{x-1}\cdot\boldsymbol{\xi})\rangle \ . \tag{53b}$$

Turning to the $(p-n)$-dimensional, orthogonal complement of the fixed-point space we find a diagonal matrix with a single $(p-n)$-degenerate eigenvalue

$$\lambda_3=\beta J\langle\cosh^{-2}[\beta J(\boldsymbol{\mu}\cdot\boldsymbol{\xi})]\rangle, \ \text{if } x=2 \tag{53c}$$

and

$$\lambda_3=0, \ \text{if } x>2 \ . \tag{53d}$$

Note the decreasing complexity in going from λ_1 to λ_3. For $n=1$, λ_2 does not exist and λ_1 and λ_3 reproduce our previous results. For $n>1$, the stability analysis is a delicate affair. Suppose, for instance, that the ξ's have a discrete probability distribution not including zero, that $x=2$ and n is even, then $\mathrm{Prob}(\sum_{\alpha=1}^{n}\xi_\alpha=0)\neq 0$ and, thus, $\lambda_3\to\infty$ as $\beta\to\infty$. Hence, the n-symmetric state $\boldsymbol{\mu}\neq 0$ is unstable at low temperatures. However, for n odd it may well happen that $\mathrm{Prob}(\sum_{\alpha=1}^{n}\xi_\alpha=0)=0$. Then $\lambda_3\to 0$ as $\beta\to\infty$ and we have to study λ_1 and λ_2.

D. The case $x=1$

Let us first consider a very simple example that is devoid of any randomness,

$$H_N=-NJ\left|N^{-1}\sum_{i=1}^{N}S(i)\right| . \tag{54}$$

$F(m)=\beta J\,|\,m\,|$ is convex but *not* continuously differentiable. Since both $c(t)=\ln[\cosh(t)]$ and $c^*(m)$ are convex and $c^{**}(t)=c(t)$,[20] we use the analog of Fenchel's duality theorem[28] so as to transform

$$-\beta f(\beta)=\sup_m[F(m)-c^*(m)] \tag{55a}$$

into

$$-\beta f(\beta)=\sup_t[c(t)-F^*(t)] . \tag{55b}$$

Now

$$F^*(t)=\sup_m[mt-\beta J\,|\,m\,|\,]=0,\ \text{ if } |\,t\,|\leqq\beta J \tag{56}$$

and $+\infty$ elsewhere. Hence

$$-\beta f(\beta)=\sup_{|t|\leqq\beta J}[c(t)]=\ln[\cosh(\beta J)] , \tag{57}$$

since a convex function attains a global maximum on a closed convex domain D at one of the extreme points of D.[20]

We now turn to (34) and take $x=1$,

$$H_N=-NJ\sum_{\alpha=1}^{p}\left|N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}S(i)\right| . \tag{58}$$

The Legrende transform of the energy function $F(\mathbf{m})=\sum_{\alpha=1}^{p}\beta J\,|\,m_\alpha\,|$ is readily obtained as the multidimensional analog of (56),

$$F^*(\mathbf{t})=\sum_{\alpha=1}^{p}(\beta J\,|\,m_\alpha\,|\,)^*(t_\alpha)=0 \tag{59}$$

if $|\,t_\alpha\,|\leq\beta J$ for all α, and $+\infty$ elsewhere. Hence

$$-\beta f(\beta)=\sup_{\substack{|t_\alpha|\leq\beta J\\ 1\leq\alpha\leq p}}[c(\mathbf{t})] . \tag{60}$$

Since $c(\mathbf{t})=\langle\ln[\cosh(\mathbf{t}\cdot\boldsymbol{\xi})]\rangle$ is a convex function and the cube $\{\,|\,t_\alpha\,|\leq\beta J;\ 1\leq\alpha\leq p\}$ is a convex, polyhedral domain, $c(\mathbf{t})$ attains a global maximum at one of the extreme points of the cube, i.e., at one of *the corners;* cf. Fig. 2. No metastable state can exist and, by symmetry, all the corners give rise to the same maximum. Moreover, they all favor a complete mixing of the original patterns. As $T\to 0$ ($\beta\to\infty$) the corresponding microscopic states are

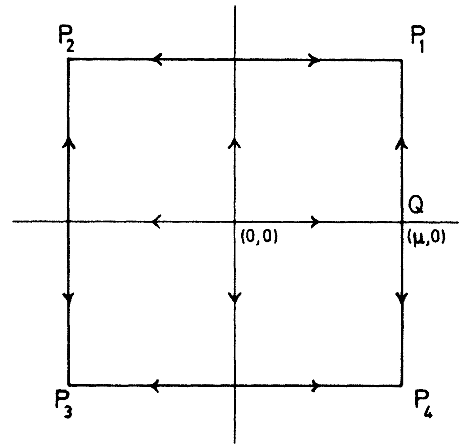

FIG. 2. The free-energy functional (60) assumes its maxima at the corners P_1,P_2,P_3,P_4 of a square (in general, a hypercube) because of convexity. The flow lines of the dynamics (22) have been indicated. The Mattis state $Q=(\mu,0)$ is unstable and so are all the other Mattis states.

$$\mathbf{R}^p\ni\mathbf{x}\to\left\{S(i)=\mathrm{sgn}\left[\sum_{\alpha=1}^{p}x_\alpha\xi_{i\alpha}\right];\ 1\leq i\leq N\right\} \tag{61}$$

where $\mathbf{x}$ is one of the corners of the cube $[-1,+1]^p$. If p is odd, there are 2^p of these states.

The previous considerations also apply if p grows logarithmically with N. Then the number of the states (61) is *extensive,* i.e., proportional to N. Their prescription has interesting applications to coding.

Summarizing: All information is immediately mixed up and forgotten. Therefore, the generalized Hopfield model with $x=1$ (and also $1<x<2$) describes an extremely forgetful being. If, however, one is willing to consider the 2^p fully mixed patterns (61) as the information one wants to retrieve, then the $x=1$ case is a model *without* metastable states. At $T=0$, one may flip up to half of the spins before leaving the basin of attraction of a certain (fully mixed) pattern (61).

V. THE LITTLE MODEL

The statistical mechanics of the Little model[2] may be described by an effective Hamiltonian[1] $\overline{H}_N$,

$$-\beta\overline{H}_N=\sum_{i=1}^{N}\ln\left[\cosh\left[\beta J\sum_j J_{ij}S(j)\right]\right] . \tag{62}$$

As compared to Ref. 1, a trivial $N\ln 2$ has been dropped. The $S(i)$ are Ising spins whose normalized trace will be denoted by $\mathrm{tr}_S(\ )$. Since the J_{ij} are as in the Hopfield case (1), we may define order parameters as in (5), i.e.,

$$m_{\alpha,N}^{(1)}=N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}S(i),\ \ 1\leq\alpha\leq p . \tag{63}$$

To each site i we also assign a duplicate spin $\sigma(i)=\pm 1$, with normalized trace $\mathrm{tr}_\sigma(\ )$ and order parameters

$$m_{\alpha,N}^{(2)}=N^{-1}\sum_{i=1}^{N}\xi_{i\alpha}\sigma(i),\ \ 1\leq\alpha\leq p . \tag{64}$$

Using the duplicate spins we now rewrite the Hamiltonian (62),

$$-\beta\overline{H}_N=\ln\left\{2^{-N}\prod_{i=1}^{N}\left[\exp\left[\beta J\sum_{\alpha=1}^{p}\xi_{i\alpha}m_{\alpha,N}^{(1)}\right]+\exp\left[-\beta J\sum_{\alpha=1}^{p}\xi_{i\alpha}m_{\alpha,N}^{(1)}\right]\right]\right\}$$
$$=\ln\left[\mathrm{tr}_\sigma\prod_{i=1}^{N}\exp\left[\beta J\sigma(i)\sum_{\alpha=1}^{p}\xi_{i\alpha}m_{\alpha,N}^{(1)}\right]\right]$$
$$=\ln\left\{\mathrm{tr}_\sigma\exp\left[N\beta J\left[\sum_{\alpha=1}^{p}m_{\alpha,N}^{(1)}m_{\alpha,N}^{(2)}\right]\right]\right\}, \tag{65}$$

so that

$$-\beta f(\beta)=\lim_{N\to\infty}\{N^{-1}\ln[\mathrm{tr}_S\exp(-\beta\overline{H}_N)]\}$$
$$=\lim_{N\to\infty}N^{-1}\ln\{\mathrm{tr}_S\,\mathrm{tr}_\sigma\exp[N\beta J(\mathbf{m}_N^{(1)}\cdot\mathbf{m}_N^{(2)})]\}\ . \tag{66}$$

For $\mathbf{m}=(\mathbf{m}^{(1)},\ \mathbf{m}^{(2)})\in\mathbf{R}^{2p}$, the energy function is

$$F(\mathbf{m})=\sum_{\alpha=1}^{p}\beta Jm_\alpha^{(1)}m_\alpha^{(2)}=\beta J\mathbf{m}^{(1)}\cdot\mathbf{m}^{(2)}\ . \tag{67}$$

$F(\mathbf{m})$ is *not* convex. Furthermore, the c function which is relevant to (66) has the argument $\mathbf{t}=(\mathbf{t}_1,\mathbf{t}_2)$ in $\mathbf{R}^{2p}=\mathbf{R}^p\times\mathbf{R}^p$ and is defined by

$$c(\mathbf{t})=\lim_{N\to\infty}N^{-1}\ln[\mathrm{tr}_S\,\mathrm{tr}_\sigma\exp N(\mathbf{t}_1\cdot\mathbf{m}_N^{(1)}+\mathbf{t}_2\cdot\mathbf{m}_N^{(2)})]$$
$$=\lim_{N\to\infty}\{N^{-1}\ln[\mathrm{tr}_S\exp N(\mathbf{t}_1\cdot\mathbf{m}_N^{(1)})]+N^{-1}\ln[\mathrm{tr}_\sigma\exp N(\mathbf{t}_2\cdot\mathbf{m}_N^{(2)})]\}$$
$$=c(\mathbf{t}_1)+c(\mathbf{t}_2)\ . \tag{68}$$

Here, by abuse of notation, the $c(\mathbf{t}_i)$ are the Hopfield c functions (14) with argument in $\mathbf{R}^p$. With *probability one*,[21] they do not depend on the fixed set of ξ's we started with.

By Sec. III A we immediately get

$$-\beta f(\beta)=\sup_{\mathbf{m}}[\beta J\mathbf{m}^{(1)}\cdot\mathbf{m}^{(2)}-c^*(\mathbf{m})]\ . \tag{69}$$

As we will show shortly, the symmetry between the two components $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$ of $\mathbf{m}=(\mathbf{m}^{(1)},\mathbf{m}^{(2)})$ is not broken. That is, if $\mathbf{m}$ maximizes the free-energy functional, then $\mathbf{m}^{(1)}=\mathbf{m}^{(2)}=\boldsymbol{\mu}$.

Due to the above observation the fixed-point equation (18), which reads

$$\begin{bmatrix}\mathbf{m}^{(1)}\\ \mathbf{m}^{(2)}\end{bmatrix}=\left\langle\begin{bmatrix}\xi\tanh[\beta J(\mathbf{m}^{(2)}\cdot\xi)]\\ \xi\tanh[\beta J(\mathbf{m}^{(1)}\cdot\xi)]\end{bmatrix}\right\rangle, \tag{70}$$

may be reduced to

$$\boldsymbol{\mu}=\langle\xi\tanh[\beta J(\boldsymbol{\mu}\cdot\xi)]\rangle\ , \tag{71}$$

which is (36) with $x=2$, i.e., the generalized Hopfield model. Furthermore, by (21) the free energy directly follows:

$$-\beta f(\beta)=[F(\mathbf{m})-\mathbf{m}\cdot\nabla F(\mathbf{m})+c(\nabla F(\mathbf{m}))]$$
$$=\beta J(\boldsymbol{\mu}^2-2\boldsymbol{\mu}^2)+2c(\beta J\boldsymbol{\mu})$$
$$=2[-\tfrac{1}{2}\beta J\boldsymbol{\mu}^2+c(\beta J\boldsymbol{\mu})]\ , \tag{72}$$

and thus

$$f_{\text{Little}}(\beta)=2f_{\text{Hopfield}}(\beta)\ . \tag{73}$$

This result was obtained in a completely different way in Ref. 8.

We now show[29] that in (69) $\mathbf{m}^{(1)}$ and $\mathbf{m}^{(2)}$ must be equal without using the special form of the fixed-point equation (70). Without loss of generality we may assume that $\beta J=1$ and that the supremum is attained at $\mathbf{m}'=(\mathbf{m}_1',\mathbf{m}_2')$. Then

$$\sup_{\mathbf{m}}[\beta J\mathbf{m}^{(1)}\cdot\mathbf{m}^{(2)}-c^*(\mathbf{m})]=\mathbf{m}_1'\cdot\mathbf{m}_2'+[-c^*(\mathbf{m}_1',\mathbf{m}_2')]\ . \tag{74}$$

The function $c(\mathbf{t}_1,\mathbf{t}_2)$ is symmetric in its two arguments and so is its Legendre transform $c^*(\mathbf{m}_1,\mathbf{m}_2)$. Moreover, $-c^*(\mathbf{m}_1,\mathbf{m}_2)$ is concave and thus

$$-c^*(\mathbf{m}_1',\mathbf{m}_2')=\tfrac{1}{2}[-c^*(\mathbf{m}_1',\mathbf{m}_2')]+\tfrac{1}{2}[-c^*(\mathbf{m}_2',\mathbf{m}_1')]$$
$$\le -c^*\left[\frac{\mathbf{m}_1'+\mathbf{m}_2'}{2},\frac{\mathbf{m}_1'+\mathbf{m}_2'}{2}\right], \tag{75}$$

while

$$\mathbf{m}_1'\cdot\mathbf{m}_2'\le\left[\frac{\mathbf{m}_1'+\mathbf{m}_2'}{2}\right]^2 \tag{76}$$

with equality if and only if $\mathbf{m}_1'=\mathbf{m}_2'$. Combining (74)–(76) we get for the right-hand side of (74)

$$\mathbf{m}_1'\cdot\mathbf{m}_2'+[-c(\mathbf{m}_1',\mathbf{m}_2')]$$
$$\le\left[\frac{\mathbf{m}_1'+\mathbf{m}_2'}{2}\right]^2-c^*\left[\frac{\mathbf{m}_1'+\mathbf{m}_2'}{2},\frac{\mathbf{m}_1'+\mathbf{m}_2'}{2}\right], \tag{77}$$

and the inequality is *strict* if $\mathbf{m}_1' \neq \mathbf{m}_2'$, as claimed.

Turning to the stability problem we note that, as $N \to \infty$, both the original spin configurations and their duplicates give rise to the same stationary points $\boldsymbol{\mu}$. These are determined by (71) and the requirement that they maximize the free-energy functional (69), i.e.,

$$-\beta f(\beta) = \sup_{\boldsymbol{\mu}} [\boldsymbol{\mu}^2 - c^*(\boldsymbol{\mu},\boldsymbol{\mu})] \; . \tag{78}$$

The function $\boldsymbol{\mu}^2$ is convex and so is $d(\boldsymbol{\mu}) \equiv c^*(\boldsymbol{\mu},\boldsymbol{\mu})$. We, therefore, can apply the analog of Fenchel's duality theorem[28] so as to get

$$-\beta f(\beta) = \sup_{\mathbf{t} \in \mathbb{R}^p} [d^*(\mathbf{t}) - \tfrac{1}{4}\mathbf{t}^2] \tag{79}$$

and using the relation

$$\begin{aligned} d^*(\mathbf{t}) &= \sup_{\boldsymbol{\mu}} [\boldsymbol{\mu}\cdot\mathbf{t} - c^*(\boldsymbol{\mu},\boldsymbol{\mu})] \\ &= \sup_{(\boldsymbol{\mu}_1,\boldsymbol{\mu}_2)} [\boldsymbol{\mu}_1\cdot(\tfrac{1}{2}\mathbf{t}) + \boldsymbol{\mu}_2(\tfrac{1}{2}\mathbf{t}) - c^*(\boldsymbol{\mu},\boldsymbol{\mu})] \\ &= c^{**}(\tfrac{1}{2}\mathbf{t},\tfrac{1}{2}\mathbf{t}) = c(\tfrac{1}{2}\mathbf{t},\tfrac{1}{2}\mathbf{t}) \\ &= 2c(\tfrac{1}{2}\mathbf{t}) \; , \end{aligned} \tag{80}$$

where the last equality follows from (68), we find

$$-\beta f(\beta) = \sup_{\mathbf{t}} [2c(\tfrac{1}{2}\mathbf{t}) - (\tfrac{1}{2}\mathbf{t})^2] \; , \tag{81}$$

i.e., putting $\frac{1}{2}\mathbf{t} = \mathbf{u}$,

$$-\beta f_{\text{Little}}(\beta) = 2 \sup_{\mathbf{u}} [c(\mathbf{u}) - \tfrac{1}{2}\mathbf{u}^2] \; , \tag{82}$$

which is (73) once again. But more can be said. Equation (82) shows that *the stability analysis of the Hopfield model directly applies to the Little model.* Since we have duplicated the spin variables we only need to do half of the usual amount of work.

VI. DISCUSSION

A flexible formalism has been presented to determine the equilibrium statistical mechanics, in particular the free energy and the stable states, of a spin-glass-like neural network. The main results, (21) and (30), have been obtained without any supposition on the energy function F—except for convexity, which greatly simplifies the stability analysis. As we have shown elsewhere, through the notion of *extremal set*,[30] even the ground-state analysis may depend critically on the chosen probability distribution. We, therefore, have concentrated on a general framework and refrained from giving too many details.

One may wonder, though, where we used the randomness of the $\xi_{i\alpha}$'s. The answer is: To obtain thermodynamic stability *as* $N \to \infty$ through their ergodicity.[21] If the $\xi_{i\alpha}$'s are arbitrary, one may use the same formalism but the c function is not readily evaluated and the ensuing analysis is, at least analytically, intractable.

The family of Hamiltonians (34) which are labeled by $x \geq 1$ comprises the Hopfield model ($x=2$) and shows a wide spectrum of bifurcation phenomena and stability problems. Whereas for $x=1$ metastable states do not exist and the original patterns are fully mixed these models favor the stability of single patterns whenever $x > 2$. The Hopfield case $x=2$ is particular in that the Hessian of F is just a nonzero multiple of the unit matrix. Moreover, a simple duplication of the spin variables enables a direct solution of the Little model in terms of the Hopfield model.

Though the present analysis was restricted to only a finite number p of stored patterns, it clearly shows that pattern recognition through spin-glass-like neural networks offers a surprising richness of structure which deserves detailed attention and further research.

ACKNOWLEDGMENTS

It is a pleasure to thank Aernout van Enter for a critical reading of the manuscript. This work has been supported by the Deutsche Forschungsgemeinschaft (Bonn, Germany), under Sonderforschungsbereich 123 (Stochastische mathematische Modelle).

APPENDIX

In Sec. IV we have seen that the parameter x of the Hamiltonian (34) plays an important role. However, for $1 < x < 2$ we could not apply our stability analysis directly because the energy function $F(\mathbf{m})$ was not twice continuously differentiable at $\mathbf{m}=\mathbf{0}$. Section IV D suggests, however, that here also we should take advantage of the analog of Fenchel's duality theorem[28]

$$\begin{aligned} -\beta f(\beta) &= \sup_{\mathbf{m}} [F(\mathbf{m}) - c^*(\mathbf{m})] \\ &= \sup_{\mathbf{t}} [c(\mathbf{t}) - F^*(\mathbf{t})] \; . \end{aligned} \tag{A1}$$

Since $F(\mathbf{m}) = \sum_{\alpha=1}^{p} \beta J x^{-1} |m_\alpha|^x$, we have to determine the Legendre transform

$$\begin{aligned} (\beta J x^{-1} |m|^x)^*(t) &= \sup_{m} (mt - \beta J x^{-1} |m|^x) \\ &= \beta J y^{-1} |t/(\beta J)|^y \; , \end{aligned} \tag{A2}$$

where $x^{-1} + y^{-1} = 1$. Note that then $(x-1)(y-1)=1$. Using (A2) we obtain

$$\begin{aligned} -\beta f(\beta) = \sup_{\mathbf{t}} \Big\{ &\langle \ln[\cosh(\mathbf{t}\cdot\boldsymbol{\xi})] \rangle \\ &- \sum_{\alpha=1}^{p} \beta J y^{-1} |t_\alpha/(\beta J)|^y \Big\} \; . \end{aligned} \tag{A3}$$

What have we gained? Since $x^{-1}+y^{-1}=1$, the inequality $1 < x < 2$ implies $y > 2$ whereas $x > 2$ leads to $1 < y < 2$. Hence, everything in sight in (A3) is everywhere twice continuously differentiable if we assume $1 < x < 2$. Let us do so and define $|\mathbf{t}| = (|t_1|, |t_2|, \ldots, |t_p|)$. An extremum of the expression between the curly brackets in (A3) is found if

$$\langle \xi_\alpha \tanh(|\mathbf{t}|\cdot\boldsymbol{\xi}) \rangle = |t_\alpha/(\beta J)|^{y-1}, \quad 1 \leq \alpha \leq p \tag{A4}$$

and a maximum is attained at $\mathbf{t}$ if the Hessian,

$$\langle \xi_\alpha \xi_\gamma \cosh^{-2}(\mathbf{t}\cdot\boldsymbol{\xi}) \rangle - \frac{y-1}{\beta J} |t_\alpha/(\beta J)|^{y-2} \delta_{\alpha\gamma} \; , \tag{A5}$$

is negative definite. If we put

$$|t_\alpha| = \beta J |\mu_\alpha|^{x-1} \tag{A6}$$

and use the relation $(x-1)(y-1)=1$, then (A4) is readily transformed into the (general version of the) fixed-point equation (36).

Let us check the stability of $\boldsymbol{\mu}=0$ in the case $1<x<2$. By (A6), $\boldsymbol{\mu}=0$ corresponds to $\mathbf{t}=0$, so (A5) is reduced to the unit matrix, which is never negative, and $\boldsymbol{\mu}=0$ is bound to be unstable, whatever the temperature. This is in agreement with our preliminary argument of Sec. IV C.

What are the stable states then? If $\boldsymbol{\mu}=(\mu,\ldots,\mu,0,\ldots,0)$ is an n-symmetric state (Sec. IV B) with $n<p$, then the restriction of the Hessian (A5) to the $(p-n)$-dimensional, orthogonal complement of the fixed-point space is the positive diagonal matrix $\langle \cosh^{-2}[\beta J(|\boldsymbol{\mu}|^{x-1}\cdot\boldsymbol{\xi})]\rangle\delta_{\alpha\gamma}$. That is, all directions in the orthogonal complement are repulsive. This phenomenon is nicely illustrated by Fig. 2, where $x=1$ and $p=2$. If $0<n<p$, we cannot but choose $n=1$. Consider the point $Q=(\mu,0)$. In the fixed-point space, along the coordinate axis, Q is attracting and, thus, stable. But in the orthogonal complement, along the line P_1P_4 perpendicular to the x axis, the natural flow is expanding and Q is unstable. We also see which points are stable: the corners, like $P_1=(\mu,\mu)$. So we have shown that among the symmetric states only the p-symmetric ones have a chance to be stable if $1<x<2$.

---

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

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

[3]J. J. Hopfield, Proc. Nat. Acad. Sci. USA **79**, 2554 (1982); *ibid.* **81**, 3088 (1984).

[4]J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature (London) **304**, 158 (1983).

[5]B. A. Huberman and T. Hogg, Phys. Rev. Lett. **52**, 1048 (1984); T. Hogg and B. A. Huberman, Proc. Nat. Acad. Sci. USA **81**, 6871 (1984).

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

[7]G. Weisbuch and F. Fogelman-Soulié, J. Phys. (Paris) Lett. **46**, L623 (1985).

[8]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).

[9]T. Kohonen, *Self-Organization and Associative Memory* (Springer, Heidelberg, 1984).

[10]W. C. McCulloch and W. Pitts, Bull. Math. Biophys. **5**, 115 (1943).

[11]This easily follows from the blue and red picture of Ref. 12 and Sec. V of Ref. 13.

[12]J. L. van Hemmen, Phys. Rev. Lett. **49**, 409 (1982).

[13]J. L. van Hemmen, A. C. D. van Enter, and J. Canisius, Z. Phys. B **50**, 311 (1983).

[14]J. L. van Hemmen, in *Heidelberg Colloquium on Spin Glasses*, edited by J. L. van Hemmen and I. Morgenstern, Vol. 192 of *Lecture Notes in Physics* (Springer-Verlag, Heidelberg, 1983), pp. 203–233.

[15]A. T. Ogielski and I. Morgenstern, Phys. Rev. Lett. **54**, 928 (1985); J. Appl. Phys. **57**, 3382 (1985).

[16]J. L. van Hemmen, Phys. Rep. **65**, 43 (1980).

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

[18]See, for instance, Figs. 1, 7, 10, 11, and 13 of Ref. 13 and Figs. 1 and 2 of Ref. 14.

[19]By the strong law of large numbers and the law of the iterated logarithm. See J. Lamperti, *Probability* (Benjamin, New York, 1966), Secs. 7 and 11.

[20]A. W. Roberts and D. E. Varberg, *Convex Functions* (Academic, New York, 1973), pp. 30, 102, 103, 110, and 115, Theorems 51.E, and 15.D, and problem 43.D.

[21]In the present case the strong law of large numbers (Ref. 19, Sec. 7) suffices.

[22]A. C. D. van Enter and J. L. van Hemmen, Phys. Rev. A **29**, 355 (1984).

[23]M. Spivak, *Calculus on Manifolds* (Benjamin, New York, 1965). The relevant theorems are (i) 2-7 and 2-8, (ii) 2-2, and (iii) 2-11.

[24]If $A>0$, the proof becomes extremely simple. Let $\lambda\in\sigma(AB)$. Then $AB\mathbf{x}=\lambda\mathbf{x}$, and $\mathbf{y}$ defined by $\mathbf{x}=A^{1/2}\mathbf{y}$ satisfies the equation $A^{1/2}BA^{1/2}\mathbf{y}=\lambda\mathbf{y}$. Alternatively, one can use $\sigma(XY)=\sigma(YX)$ by taking $x=A^{1/2}$ and $y=A^{1/2}B$.

[25]M. Kac, in *Statistical Physics, Phase Transitions and Superfluidity*, edited by M. Chrétien, E. P. Gross, and S. Deser (Gordon and Breach, New York, 1968), Vol. 2, pp. 248–249; H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1971), Sec. 6.5.

[26]A. N. Kolmogorov and S. V. Fomin, *Introductory Real Analysis* (Dover, New York, 1975), p. 68 of Sec. 8.1.

[27]See, e.g., the extensive discussion of the Hopfield model with $\xi=\pm1$ in Ref. 8. As in the spin-glass case,[18] the final result may depend on the probability distribution.

[28]J. L. van Hemmen, A. C. D. van Enter, and J. Canisius, Ref. 13, the Appendix. Here some printing errors have escaped our attention. After the first inequality in (A4) $\mathbf{x}^{**}$ ranges through X^{**} (not X). And the second inequality is obtained by using (A2) [not (A1)].

[29]Cf. J. L. van Hemmen, A. C. D. van Enter, and J. Canisius, Ref. 13, Sec. VI.

[30]J. L. van Hemmen, A. C. D. van Enter, and J. Canisius, Ref. 13, Sec. VII. See in particular Figs. 10, 11, and 13.