PHYSICAL REVIEW A VOLUME 37, NUMBER 2 JANUARY 15, 1988

Layered feed-forward neural network with exactly soluble dynamics

Ronny Meir and Eytan Domany
Department of Electronics, Weizmann Institute of Science, 76 100 Rehovot, Israel
(Received 29 June 1987)

A model layered feed-forward neural network is studied and solved exactly in the thermodynamic limit. Layer-to-layer recursion relations are found and analyzed as a function of the relevant external parameters. Stochasticity is introduced by a "temperature" variable. A region of good recall is found, separated from a region of no recall by a first-order line terminating at a critical point. The exact time evolution of mixtures of patterns is given as well.

I. INTRODUCTION

There has recently been much interest among physicists in the field of neural networks.[1-11] One reason for this is the availability of powerful tools that had been developed independently for the study of the related but different subject of spin glasses.[12] The problem in spin glasses is usually phrased in the following form: Given a probability distribution for a set of quenched random variables (for instance, the couplings between spins), find the stable (or low-lying) states of the system and from it the thermodynamics of the model. In the case of neural networks that model a memory the problem is reversed:[2] Given a desired set of stable states, one seeks those parameters which will indeed make these states stable. The models studied usually consist of a highly interconnected system of spins, whose couplings are constructed ("learned") in such a way that a set of given states become local minima of a Hamiltonian, or more generally, the attractors of the dynamics. These "memories" are *associative;* that is, a learned pattern is "recalled" on the basis of incomplete or erroneous information. The associative nature of such memories is due simply to the fact that many initial states which are "close" to one of the stored configurations will flow to it or to a nearby pattern. One of the main problems of these models has been, however, the existence of many metastable states which are uncorrelated with the patterns.[4] In fact, the number of such spurious states is usually exponentially large. On the other hand, these memories are robust: Destroying even a finite fraction of the learned couplings does not destroy the network's ability to recall. This robustness is due to the fact that the stored patterns are spread over the whole network, and do not occupy a prescribed site as in the digital computer; memory is distributed.

The various neural network models that have been proposed over the past few years can be divided into several broad classes.

The *first* class of models was spurred by the work of Hopfield[2] which, in turn, draws on earlier work of Little.[1] The Little model was formulated in a dynamical language, i.e., the dynamics of the model was specified without reference to an underlying Hamiltonian. The Hopfield model, in contrast, is based on the existence of a Hamiltonian, which implies a relaxational type of dynamics. This in turn implies that the couplings between the spins must be symmetric. The problem addressed here is that of using the stable states of a spin-glass-like system for storing and retrieving information. Recall of an embedded pattern is equivalent to relaxation to a nearby stable state. This system was shown to possess desirable attributes in its capacity as a content-addressable memory. The thermodynamics of the Hopfield model has been solved by Amit *et al.*,[4] and the complete phase diagram has been obtained. The structure of the metastable states of this model has been analyzed by Gardner.[13] The nature of the dynamics of the model, which is a more difficult issue, has recently been addressed by Gardner *et al.*,[14] who studied the zero-temperature limit. They were able, however, to obtain an exact solution only for the first few time steps of the dynamics. It should be noted that studying the dynamics is usually a more complicated problem than that of statics. For example, the statics of the Sherrington-Kirkpatrick (SK) model, within the Parisi solution,[15] is well understood,[12] while the answer to dynamical questions such as remanence is still open.

Various modifications to the Hopfield model have been proposed since it was first suggested. These include, among others, modification of the learning rules[5,10,11] which eliminate some of its undesirable features, such as the catastrophic deterioration of performance above a critical number of patterns[5] and the constraint of random patterns.[8,10]

The *second* class of models deviates from the original one by introduction of nonlocal learning rules.[6,7] This work was spurred by earlier work of computer scientists on the mathematical construct called the pseudoinverse (see a detailed description in Ref. 16). By local rule one usually means that during the learning phase of the network, the change in the coupling J_{ij} depends only on the activity of spins i and j. Here, the locality of the rule is sacrificed with the bonus of being able to store correlated patterns, something which is impossible in the original Hopfield model. These models may be of interest in the fabrication of neural networks but seem implausible from a biological point of view. More recently, Dieder-

ich and Opper[8] have introduced a local learning algorithm which is capable of embedding correlated patterns in a network. The learning procedure is of the perceptron type.[17,18]

A *third* class of models eliminates the restriction of symmetric bonds. This makes sense biologically, but makes the problem much harder mathematically. The reason for this is that once the bonds are made asymmetric, there no longer exists a Lyapunov (monotonically decreasing) function which governs the dynamics. The sophisticated tools developed by workers in the field as spin glasses cannot deal with such non-Hamiltonian systems. Results concerning the behavior of such systems have been obtained recently by Hertz *et al.*[19] (using Langevin equation methods), by Sompolinsky and Kanter[20] (using such networks for generating and recognizing time sequences), and by Derrida *et al.*[21] (for a diluted version of the Hopfield model). Interestingly enough this last model is actually solved rather easily *because* of its asymmetry. Note that the bonds in these models are asymmetric but the network is *functionally* symmetric.

A *fourth* class of models introduces layered architectures. This type of model, on which we will focus in what follows, has been studied extensively by computer scientists over the past few decades, but little progress has been made concerning analytic results.[18] The main feature of this class which distinguishes it from the previous classes is the existence of "hidden units." These systems usually[22] consist of an input unit, an output unit, and intermediate hidden units that do the processing. Contact with the external world is made only via the input and output units. Such a system confers the immediate advantage of being able to use the hidden units to one's benefit. No external constraints are placed on these units, and they are used in order to construct good "internal representations" of the environment. Recently, Linsker[23] has studied the self-organization of feature detectors in such feed-forward networks.

The protype of these models was the perceptron proposed by Rosenblatt[17] many years ago. This system was shown by Minsky and Papert[18] to be of limited value. However, in recent years there has been much work on multilayered systems (as opposed to the single-layered perceptron), and ingenious methods have been proposed to circumvent some of the limitations of the perceptron. Notable among these systems is the Boltzmann machine[24] and various elaborations on it, which use ideas from statistical mechanics to define their operation. These systems are capable of performing a multitude of tasks such as learning to read aloud,[25] recognizing symmetry groups,[26] learning the past tense,[27] and more. The performance of these systems has been studied mainly by computer simulation. The first appearance of related, multilayered models in the recent physics literature is through the work of Hogg and Huberman,[28] followed by a number of other groups.[29,30]

We have recently introduced a class of multilayer neural networks.[29] Here we concentrate on the detailed analytic solution of the version of the model that we previously called "simple." The main results of this paper were presented previously[31,32] without detailed derivation.

The network consists of layers of linear threshold units, similar to the perceptron. We impose full connectivity between adjacent layers, but no connections exist within each layer (see Fig. 1). Thus the network is of the feed-forward type. The couplings between units in different layers will be given by a Hebbian rule (see below). The novelty in this model is that it allows self-organization of the internal (hidden) *and* output representations. This represents the fact that only the first layer of a network is in direct contact with the "external world," and hence only on the first (input) layer are the representations of the key patterns externally dictated. On all subsequent layers the system is free to "choose" an internal representation of any key pattern. For example, in the context of pattern recognition this translates to realizating that even though the input pattern may correspond to an actual physical image on the retina or on the reading unit of the network, there is no such meaning for the "image" produced in any internal (or the final) stage of the network. If final visual display of the recognized pattern is required this can be achieved by a simple switching mechanism that reads the final pattern, and if it corresponds to a key pattern, the desired display is generated. Moreover, the system can use its freedom to redefine the key patterns on subsequent layers. With proper self-organization of internal and output representation, obtained in an iterated learning procedure, we have shown[29] that the network is capable of perfect recall of key patterns and excellent recognition of noisy input patterns. No such iterated learning is allowed in the network considered here. We focus on a version where the representations on all layers are fixed. This is the case which we have solved exactly. By

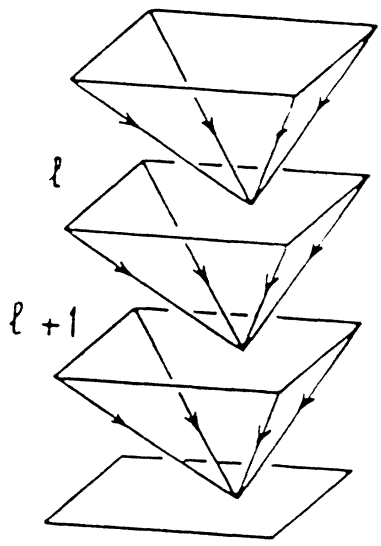

FIG. 1. Network with layered architecture. State of cell i in layer $l+1$ is determined by the states of all cells of layer l. Input patterns are presented to the first layer, output is read out from the last. Each cell contains a binary variable $S_i^l = \pm 1$ that defines its state.

exact solution we mean that given any initial condition we can follow the evolution of the state on each layer exactly for any number of layers. In particular, we obtain the asymptotic (number of layers goes to infinity) behavior of various parameters, which will be defined below. Our results can be summarized as follows.

(i) There exists a region in the α, T plane where the system performs well as a content-addressable memory. Here $\alpha=p/N$, where p is the number of patterns to be stored and N is the number of sites in each layer. T is a temperature variable; for $T=0$ the dynamics of the network is deterministic, with the degree of stochasticity increasing with T. The value of the critical α above which no recall is possible[31] is given at $T=0$ by $\alpha_c=0.27$. In addition, we find[32] that recall is not possible above $T=1$, as in the Hopfield model.[4] The transition between the region of no recall and that of good recall is first order (see below). The full phase diagram of the model is calculated, and is given in Fig. 3.

(ii) The relaxation time to the asymptotic behavior τ defined below, is seen to diverge near the transition line. Divergence of τ is usually characteristic of a second-order transition. We find that at $T=0$, $\tau\sim(\alpha_c-\alpha)^{-1/2}$ as α_c is approached.[31]

(iii) The response of the system to an initial state which is correlated with a finite number of stored memories is studied. The system is found to be dynamically unstable to a mixture of two patterns. That is, starting with an initial state that has a finite overlap with two patterns, the final state has a vanishing overlap with the state of smaller initial overlap. Above a certain α the final state has zero overlap with either of the patterns.

The paper is organized as follows. In Sec. II we define our model and explain its operation (i.e., dynamics). The analytic solution of the model is given in Sec. III in the form of layer-to-layer recursion relations. These recursion relations are analyzed in Sec. IV which also contains comparison of the analytic results (valid for infinite systems) with computer simulations. The reader, if not interested in the technical details of the solution, may go directly to Sec. IV. Our findings, as well as future work are discussed in Sec. V.

II. DEFINITION OF THE MODEL

The model we studied is the following. Consider L layers; each contains N cells (spins), with a binary variable $S_i^l=\pm1$ associated with cell i of layer l. Each cell is *connected to all* cells of the neighboring layers (see Fig. 1). The bonds are, however, *unidirectional;* the state of layer $l+1$ is determined by the state (at the previous time step) of layer l according to a probabilistic rule. The dynamic process is one which sets the layers sequentially; input corresponds to setting the first layer in an initial state S_i^1. At the next time step the second layer is set in state S_i^2 and so on. The probability that the ith spin in the $(l+1)$th layer has the value S_i^{l+1}, given that on the previous layer l the cells are in state S_i^l is taken to be

$$P(S_i^{l+1})=\frac{e^{\beta S_i^{l+1}h_i^{l+1}}}{2\cosh(\beta S_i^{l+1}h_i^{l+1})}\ , \tag{1}$$

where

$$h_i^{l+1}=\sum_{j=1}^{N}J_{ij}^l S_j^l \tag{2}$$

is the field produced by the spins of layer l at site i of layer $l+1$. The parameter

$$\beta=1/T$$

governs the stochasticity of the dynamics, which is deterministic for $T\to0$ (or $\beta\to\infty$) and becomes more stochastic as T decreases. The couplings or bonds J_{ij}^l are chosen by the popular prescription,[1,2]

$$J_{ij}^l=\frac{1}{N}\sum_{\nu=1}^{\alpha N}\xi_{i,\nu}^{l+1}\xi_{j,\nu}^{l}\ , \tag{3}$$

where $\xi_{i,\nu}^l$ with $\nu=1,2,\ldots,\alpha N$ are the stored key patterns.

It should be noted that each key pattern carries a layer index. This is a central feature that characterizes the class of model neural networks studied in Ref. 29; it has conceptual as well as technical significance. The conceptual novel aspects are discussed in the Introduction. The main point is that while the input representation of the key pattern ν, i.e., $\xi_{i,\nu}^1$ is externally dictated, the network is free to choose the internal as well as output representations $\xi_{i,\nu}^l$, $l>1$.

In all our studies we assume that the input, internal, and output representations $\xi_{i,\nu}^l$ of the key patterns are randomly chosen; all $\xi_{i,\nu}^l=\pm1$ with equal probability. It is precisely this fact, of the independent choice of representations on different layers, that technically allows analytic solution of our model. Note that since $\xi_{i,\nu}^l$ depends on l, the couplings J_{ij}^l also vary with layer index. Since the layer index also corresponds to (discrete) time steps, our layered model can be viewed as one with a single layer of cells, but time-dependent couplings.

By exact solution of our model we mean that given an initial state with overlap m_1^1 with a key pattern (say pattern number 1), we have a recursive formula that yields the average overlap on any subsequent layer or time step, averaged over all key patterns ξ. The overlap m_μ^l is defined by

$$m_\mu^l=\frac{1}{N}\sum_{i=1}^{N}\xi_{i,\mu}^l S_i^l\ . \tag{4}$$

As an example of an important question answered by our solution, consider an initial state that has a finite overlap m_1^1 with key pattern $\nu=1$ but vanishing $[O(1/\sqrt{N})]$ overlaps with all others. Dynamics resulting from such an initial state produces average overlaps m_1^l on subsequent layers. If, as $l\to\infty$, $m_1^l\to m^*>0$, the network "remembers," and we have finite recall. However, if $m^*=O(1/\sqrt{N})$ there is no recall. From our solution one can calculate how m_1^l depends on l, the relaxation time to m^*, as well the dependence of m^* on the parameters α,T and the initial overlap m_1^1. We derive the solution using the same methods as applied to

the Hopfield models' stable states,[13] and to dynamics of the Sherrington-Kirkpatrick and Little models.[14]

III. EXACT SOLUTION

Consider a random assignment of $\nu=1,2,\ldots,\alpha N$ key patterns $\xi^l_{i,\nu}$ on each of L layers of the network. Choose an initial state on the first layer S^l_i such that its overlap with one key pattern (say, $\nu=1$) is $m^1_1=O(1)$, and with the other patterns $\nu\neq 1$, $m^1_\nu=O(1/\sqrt{N})$. The question we ask is the following: What is the probability $P(m^L_1 \mid m^1_1)$ that the dynamic rules (1) and (2) produce on layer L a state S^L that has overlap $m^L=O(1)$ with key pattern $\xi^L_{i,1}$ [and $O(1/\sqrt{N})$ with the others]? Note that we must average both over the random assignment of the ξ's and over the probability distribution given in Eq. (1). The conditional probability to get a configuration

$$\mathbf{S}^{l+1}=(S^{l+1}_1,S^{l+1}_2,\ldots,S^{l+1}_N)$$

on layer $l+1$, given the configuration $\mathbf{S}^l$ on the previous layer, is obtained by taking the product of Eq. (1) over all sites,

$$P_\xi(\mathbf{S}^{l+1}\mid\mathbf{S}^l)=\prod_{i=1}^{N}\left[\frac{e^{\beta S^{l+1}_i h^{l+1}_i}}{2\cosh(\beta S^{l+1}_i h^{l+1}_i)}\right], \tag{5}$$

where h^{l+1}_i is given in Eq. (2). The subscript ξ denotes the dependence of P_ξ on all the key patterns $\xi^1_{i,\nu}$. A sequence of configurations $\mathbf{S}^1,\ldots,\mathbf{S}^L$ will be generated by our dynamic rules with the probability,

$$P_\xi(\mathbf{S}^1,\mathbf{S}^2,\ldots,\mathbf{S}^L)=P_\xi(\mathbf{S}^L\mid\mathbf{S}^{L-1})\cdots P_\xi(\mathbf{S}^2\mid\mathbf{S}^1)\ . \tag{6}$$

In order to obtain the probability for a configuration $\mathbf{S}^L$ on layer L, given the initial state $\mathbf{S}^1$, we must sum this over all intermediate layers;

$$P_\xi(\mathbf{S}^L\mid\mathbf{S}^1)=\sum_{\mathbf{S}^2,\ldots,\mathbf{S}^{L-1}}P_\xi(\mathbf{S}^L\mid\mathbf{S}^{L-1})\cdots P_\xi(\mathbf{S}^2\mid\mathbf{S}^1)\ . \tag{7}$$

Finally, averaging this quantity over the probability distribution of the random variables ξ, we obtain the probability P for $\mathbf{S}^L$ given $\mathbf{S}^1$ for a random realization of the ξ's,

$$P(\mathbf{S}^L\mid\mathbf{S}^1)=\langle\langle P_\xi(\mathbf{S}^L\mid\mathbf{S}^1)\rangle\rangle\ . \tag{8}$$

The double averaging sign indicates an average over the ξ's. Anticipating that this quantity depends only on the initial and final overlaps m^1 and m^L, respectively, we express the probability as a function of these overlaps,

$$P(m^L_1\mid m^1_1)=e^{Ns(m^L_1)}P(\mathbf{S}^L\mid\mathbf{S}^1)\ . \tag{9}$$

Here $\exp[Ns(m^L_1)]$ is the number of states $\mathbf{S}^L$ that have overlap m^L_1 with ξ^L_ν. This number is given by[13]

$$\exp[Ns(m)]=\begin{bmatrix}N\\ \frac{1}{2}N(1-m)\end{bmatrix}\ .$$

For large N one has

$$s(m)=-\tfrac{1}{2}(1-m)\ln\tfrac{1}{2}(1-m)-\tfrac{1}{2}(1+m)\ln\tfrac{1}{2}(1+m)\ .$$

In order to evaluate $P(\mathbf{S}^L\mid\mathbf{S}^1)$, given by

$$P(\mathbf{S}^L\mid\mathbf{S}^1)=\left\langle\!\!\left\langle\sum_{\mathbf{S}^2,\ldots,\mathbf{S}^{L-1}}\prod_{i=1}^{N}\prod_{l=1}^{L-1}\frac{\exp\left[\frac{\beta}{N}S^{l+1}_i\sum_\mu\xi^{l+1}_{i,\mu}\sum_j\xi^l_{j,\mu}S^l_j\right]}{2\cosh\left[\frac{\beta}{N}S^{l+1}_i\sum_\mu\xi^{l+1}_{i,\mu}\sum_j\xi^l_{j,\mu}S^l_j\right]}\right\rangle\!\!\right\rangle, \tag{10}$$

we first have to decouple the sums over the patterns and over the sites. To do that we introduce a set of variables m^l_μ for $\mu=1,\ldots,\alpha N$ and $l=1,\ldots,L-1$ through the following relation:

$$1=\prod_{\mu,l}^{L-1}\int_{-\infty}^{\infty}dm^l_\mu\,\delta\left[m^l_\mu-\frac{1}{N}\sum_i S^l_i\xi^l_{i,\mu}\right]=\prod_{\mu,l}^{L-1}\int_{-\infty}^{\infty}\frac{dm^l_\mu\,d\hat{m}^l_\mu}{2\pi/N}\exp\left[iNm^l_\mu\hat{m}^l_\mu-\hat{m}^l_\mu\sum_i S^l_i\xi^l_{i,\mu}\right]. \tag{11}$$

Note that the variable m^1_1 (corresponding to $\mu=l=1$) is *not* included here; i.e., it is not an integration variable. With this definition the equation for P becomes

$$P(\mathbf{S}^L\mid\mathbf{S}^1)=\int_{-\infty}^{\infty}\prod_{\mu,l}^{L-1}\frac{dm^l_\mu\,d\hat{m}^l_\mu}{2\pi/N}e^{iN\sum_{\mu,l}m^l_\mu\hat{m}^l_\mu}\left\langle\!\!\left\langle\sum_{\mathbf{S}^2,\ldots,\mathbf{S}^{L-1}}e^{-i\sum_{\mu,l}\hat{m}^l_\mu\sum_j\xi^l_{j,\mu}S^l_j}\prod_{i,l}\frac{\exp\left[\beta\sum_\mu S^{l+1}_i\xi^{l+1}_{i,\mu}m^l_\mu\right]}{2\cosh\left[\beta\sum_\mu S^{l+1}_i\xi^{l+1}_{i,\mu}m^l_\mu\right]}\right\rangle\!\!\right\rangle. \tag{12}$$

In order to perform the average over the ξ's it is convenient to introduce additional variables which bring them to an exponential form. These are introduced through the following definition:

$$1=\prod_{i,l}^{L-1}\int_{-\infty}^{\infty}d\varphi^{l+1}_i\,\delta\left[\varphi^{l+1}_i-\sum_\mu S^{l+1}_i\xi^{l+1}_{i,\mu}m^l_\mu\right]=\prod_{i,l}^{L-1}\int_{-\infty}^{\infty}\frac{d\varphi^{l+1}_i\,d\hat{\varphi}^{l+1}_i}{2\pi}\exp\left[i\varphi^{l+1}_i\hat{\varphi}^{l+1}_i-i\hat{\varphi}^{l+1}_i\sum_\mu S^{l+1}_i\xi^{l+1}_{i,\mu}m^l_\mu\right]. \tag{13}$$

With this relation inserted in (12) we obtain

$$P(\mathbf{S}^L \mid \mathbf{S}^1)=\int_{-\infty}^{\infty}\prod_{\mu,l}^{L-1}\frac{dm_\mu^l\, d\hat{m}_\mu^l}{2\pi/N}\prod_{i,l}^{L-1}\frac{d\varphi_i^{l+1}d\hat{\varphi}_i^{l+1}}{2\pi}\exp\left[iN\sum_{\mu,l}m_\mu^l\hat{m}_\mu^l+i\sum_{i,l}\varphi_i^{l+1}\hat{\varphi}_i^{l+1}\right]$$
$$\times\sum_{\mathbf{S}^2,\ldots,\mathbf{S}^{L-1}}\left\langle\!\!\left\langle\exp\left[-i\sum_{j,l=2}^{L-1}\hat{m}_1^l\xi_{j,1}^l S_j^l-i\sum_{j,l=1}^{L-1}\hat{\varphi}_j^{l+1}S_j^{l+1}\xi_{j,1}^{l+1}m_1^l\right]\right\rangle\!\!\right\rangle_{\xi_{i,1}^l}$$
$$\times Y(m_\mu^l,\hat{m}_\mu^l,\hat{\varphi}_i^l,\mathbf{S}^l)\prod_{i=1}^{N}\prod_{l=1}^{L-1}\frac{e^{\beta\varphi_i^{l+1}}}{2\cosh(\beta\varphi_i^{l+1})}\,. \tag{14}$$

In the expression above we have separated the terms with $\mu>1$ from the term with $\mu=1$. In this equation and in the remainder of the paper whenever μ appears it will only take values $\mu>1$. Y is given by

$$Y(\mathbf{S}^l)=\left\langle\!\!\left\langle\exp\left[-i\sum_{\mu,l,j}\hat{m}_\mu^l\xi_{j,\mu}^l S_j^l-i\sum_{\mu,l,j}\hat{\varphi}_j^{l+1}S_j^{l+1}\xi_{j,\mu}^{l+1}m_\mu^l\right]\right\rangle\!\!\right\rangle_{\xi_{i,\mu}^l}$$
$$=\exp\left[\sum_{\mu,j,l}\ln\cosh(i\hat{m}_\mu^l+i\hat{\varphi}_j^l m_\mu^{l-1})+\sum_{\mu,j}\ln\cosh(i\hat{m}_\mu^1)+\sum_{\mu,j}\ln\cosh(i\hat{\varphi}_j^L m_\mu^{L-1})\right]. \tag{15}$$

The second equality is obtained by a simple gauge transformation $\xi_{i,\mu}^l S_i^l\to\sigma_{i,\mu}^l$. Technically, the reason why the model can be solved exactly while the Little model cannot[14] is that here the patterns carry a layer index and so the above sum can be simply done. In the Little model, the ξ's do not carry a layer index (i.e., one has $\xi_{i,\mu}^l=\xi_{i,\mu}^{l+1}$ if $l\to t$) and so the sums over ξ and S do not decouple. In our model the variables $\sigma_{i,\mu}^l$ are independent; hence, the summation over $\xi_{i,\mu}^l$, implied in Y can be performed separately from the remaining sums (over $\mathbf{S}$) in (14).

To proceed we make an ansatz concerning the variables $m_\mu^l,\hat{m}_\mu^l$. Since the initial condition was a state $\mathbf{S}^1$ that has a finite overlap m_1^1 only with pattern number 1, we assume that for all times (layers) the only finite overlap is with pattern 1, m_1^l, and for all the other patterns, $\mu\neq1$, $m_\mu^l\sim O(1/\sqrt{N})$. So, we rescale the integration variables for $\mu>1$ as follows:

$$m_\mu^l=\frac{1}{\sqrt{N}}\lambda_\mu^l\,,\qquad \hat{m}_\mu^l=\frac{1}{\sqrt{N}}\hat{\lambda}_\mu^l\,. \tag{16}$$

Expanding the expression for Y to lowest order in N gives

$$Y=\exp\left[-\tfrac{1}{2}\sum_{\mu,l}(\hat{\lambda}_\mu^l)^2+\frac{1}{N}\sum_l\left[\sum_\mu i\hat{\lambda}_\mu^l\lambda_\mu^{l-1}\right]\left[\sum_i i\hat{\varphi}_i^l\right]\right.$$
$$\left.-\frac{1}{2N}\sum_l\left[\sum_\mu(\lambda_\mu^l)^2\right]\left[\sum_i(\hat{\varphi}_i^{l+1})^2\right]\right]. \tag{17}$$

In order to separate variables λ_μ^l that carry a pattern index, from the $\hat{\varphi}_i^l$ that carry a site index, we need to introduce additional variables using the following identities:

$$1=\prod_{l=2}^{L-1}\int_{-\infty}^{\infty}dp^l\,\delta\left[p^l-\frac{1}{\alpha N}\sum_{\mu>1}\hat{\lambda}_\mu^l\lambda_\mu^{l-1}\right]$$
$$=\prod_{l=2}^{L-1}\int_{-\infty}^{\infty}\frac{dp^l\,d\hat{p}^l}{2\pi/\alpha N}\exp\left[i\alpha N p^l\hat{p}^l-i\hat{p}^l\sum_{\mu>1}i\hat{\lambda}_\mu^l\lambda_\mu^{l-1}\right], \tag{18}$$

$$1=\prod_{l=1}^{L-1}\int_{-\infty}^{\infty}dq^l\,\delta\left[q^l-\frac{1}{\alpha N}\sum_{\mu>1}(\lambda_\mu^l)^2\right]$$
$$=\prod_{l=1}^{L-1}\int_{-\infty}^{\infty}\frac{dq^l\,d\hat{q}^l}{2\pi/\alpha N}\exp\left[i\alpha N q^l\hat{q}^l-i\hat{q}\sum_{\mu>1}(\lambda_\mu^l)^2\right].$$

The order parameter q^l introduced above is the analogue of the variable r in the solution of Amit *et al.*,[4] which measures the mean-square, random overlap of a configuration with the patterns $\mu>1$.

At this stage we go back to the expression (14) for the probability distribution. We still have to calculate the average over the patterns $\xi_{i,1}^l$, and the sum over $\mathbf{S}^2,\ldots,\mathbf{S}^{L-1}$, keeping two constraints in mind: On the first and last layers neither $\xi_{i,1}^1$, $\xi_{i,1}^L$ nor states $\mathbf{S}^1,\mathbf{S}^L$ are summed over. We assume that the corresponding overlaps are fixed: m_1^1 and m_1^L with $\xi_{i,1}^1$ and $\xi_{i,1}^L$, respectively. Doing this we get the contribution of the first pattern,

$$\exp\left[\sum_{i,l}\ln[2\cosh(i\hat{m}_1^l+i\hat{\varphi}_i^l m_l^{l-1})]-i\sum_i\theta_i^L\hat{\varphi}_i^L m_1^{L-1}\right],$$

where θ_i^L is $+1$ for $\frac{1}{2}(1+m_1^L)N$ sites and -1 for the rest.

Combining all the above manipulations into a single expression we obtain

$$P(\mathbf{S}^L \mid \mathbf{S}^1)=\int_{-\infty}^{\infty}\prod_{l=2}^{L-1}\frac{dm_1^l\,d\hat{m}_1^l}{2\pi/N}\prod_{l=1}^{L-1}\frac{dq^l\,d\hat{q}^l}{2\pi/\alpha N}\prod_{l=2}^{L-1}\frac{dp^l\,d\hat{p}^l}{2\pi/\alpha N}\prod_{i,l}^{L-1}\frac{d\varphi_i^{l+1}\,d\hat{\varphi}_i^{l+1}l}{2\pi}\prod_{\mu,l}^{L-1}\frac{d\lambda_\mu^l\,d\hat{\lambda}_\mu^l}{2\pi}$$

$$\times\exp\left[iN\sum_l\hat{m}_1^l m_1^l+\sum_{\mu,l}i\hat{\lambda}_\mu^l\lambda_\mu^l\right]\exp\left[i\alpha N\sum_l\hat{p}^l p^l+i\alpha N\sum_l\hat{q}^l q^l+\sum_{i,l}i\hat{\varphi}_i^{l+1}\varphi_i^{l+1}\right]$$

$$\times\exp\left[-\sum_{l,\mu}i\hat{p}^l i\hat{\lambda}_\mu^l\lambda_\mu^{l-1}-\sum_{l,\mu}i\hat{q}^l(\lambda_\mu^l)^2-\tfrac{1}{2}\sum_{l,\mu}(\hat{\lambda}_\mu^l)^2\right]$$

$$\times\exp\left[\alpha\sum_{i,l}p^{l+1}i\hat{\varphi}_i^{l+1}-\frac{\alpha}{2}\sum_{i,l}q^l(\hat{\varphi}_i^{l+1})^2\right]\exp\left[\sum_{i,l}\ln[2\cosh(i\hat{m}_1^l+i\hat{\varphi}_i^l m_1^{l-1})]-i\sum_i\theta_i^L\hat{\varphi}_i^L m_1^{L-1}\right]$$

$$\times\prod_{i=1}^{N}\prod_{l=1}^{L-1}\frac{e^{\beta\varphi_i^{l+1}}}{2\cosh(\beta\varphi_i^{l+1})}\,. \tag{19}$$

The integral over the variables $\hat{m}_1^l$ can be done. Suppressing the index 1 on m_1^l we finally obtain

$$P(m^L \mid m^1)=C\int\prod_{l=1}^{L-1}dq^l\,d\hat{q}^l\int\prod_{l=2}^{L-1}dp^l\,d\hat{p}^l\,dm^l\,e^{NF}\,, \tag{20}$$

where

$$F=i\alpha\left[\sum_{l=1}^{L-1}q^l\hat{q}^l+\sum_{l=2}^{L-1}\hat{p}^l p^l\right]$$

$$+\alpha\ln Z(\hat{q}^1,\ldots,\hat{q}^{L-1},\hat{p}^2,\ldots,\hat{p}^{L-1})$$

$$+\sum_{l=2}^{L}[s(m^l)+f^l] \tag{21}$$

and

$$f^l=\tfrac{1}{2}(1-m^l)\ln I_+^l+\tfrac{1}{2}(1+m^l)\ln I_-^l\,, \tag{22}$$

$$I_\pm^l=\int_{-\infty}^{\infty}\frac{dy}{\sqrt{2\pi}}e^{-y^2/2}\times\left[1+\exp\{-2\beta[(\alpha q^{l-1})^{1/2}y\mp m^{l-1}-\alpha p^l]\}\right]^{-1}. \tag{23}$$

Note that the integration variables m^l are the overlap on the intermediate layers. The function Z is given by

$$Z=\int\prod_{l=1}^{L-1}\frac{d\lambda^l\,d\hat{\lambda}^l}{2\pi}\exp\left[i\sum_{l=1}^{L-1}\lambda^l\hat{\lambda}^l-i\sum_{l=1}^{L-1}\hat{q}^l(\lambda^l)^2-\tfrac{1}{2}\sum_{l=1}^{L-1}(\hat{\lambda}^l)^2-i\sum_{l=2}^{L-1}\hat{p}^l i\hat{\lambda}^l\lambda^{l-1}\right]. \tag{24}$$

In the limit $N\to\infty$ we can calculate the integral (20) using the saddle-point method. This corresponds to evaluating the derivatives of the form $\partial F/\partial x=0$, where x stands for any one of the integration variables. We supplement these equations with $\partial F/\partial m^L=0$ corresponding to determining the value of m^L at which $P(m^L \mid m^1)$ is maximal. The saddle-point equations thus obtained are

$$\begin{aligned}
q^l&=\langle(\lambda^l)^2\rangle_Z\,,\\
p^l&=\langle i\hat{\lambda}^l\lambda^{l-1}\rangle_Z\,,\\
i\alpha\hat{q}^l&=-\frac{\partial f^{l+1}}{\partial q^l}\,,\\
i\alpha\hat{p}^l&=-\frac{\partial f^l}{\partial p^l}\,,\\
m^l&=\frac{I_-^l-I_+^l}{I_-^l+I_+^l}\,.
\end{aligned} \tag{25}$$

In these equations the symbol $\langle\cdots\rangle_Z$ stands for averaging with respect to the weight function Z. In the Appendix we show that these equations have a solution with $i\hat{q}^l=0, p^l=0$ for all l. Using the results given in the Appendix we finally obtain the solution for the overlaps m^l in the form of the following recursion relations:[32]

$$\begin{aligned}
m^{l+1}&=\int_{-\infty}^{\infty}\frac{dy}{\sqrt{2\pi}}e^{-y^2/2}\tanh\{\beta[(\alpha q^l)^{1/2}y+m^l]\}\,,\\
q^{l+1}&=1+q^l\left[\int_{-\infty}^{\infty}\frac{dy}{\sqrt{2\pi}}e^{-y^2/2}\beta\cosh^{-2}\times\beta[(\alpha q^l)^{1/2}y+m^l]\}\right]^2.
\end{aligned} \tag{26}$$

These recursion relations, to be used with $q^1=1$ (see the Appendix) and the initial overlap $m^1=m_1^1$, constitute the solution of our dynamical problem.

At zero temperature these equations reduce to

$$\begin{aligned}
m^{l+1}&=\mathrm{erf}[m^l/(2\alpha q^l)^{1/2}]\,,\\
q^{l+1}&=1+\frac{2}{\alpha\pi}\exp[-(m^l)^2/\alpha q^l]\,.
\end{aligned} \tag{27}$$

IV. ANALYSIS OF THE SOLUTION

The recursions presented in Eq. (26) constitute the solution of the dynamics of our network. The first (input) layer is set in some state that has a finite overlap m^1 with one key pattern, say, $\xi^1_{i,1}$. The resulting dynamics, given by Eq. (1), produces a sequence of states $\mathbf{S}^l$ on the subsequent layers. These states will have overlap m^l with $\xi^l_{i,1}$ the "image" of ξ^1_1 on layer l. Our solution describes how m^l depends on l for various values of the parameters α and T, and the initial overlap m^1.

The most important question one wants to answer concerns the asymptotic value of the overlap. That is, whether for long times (large l) the sequence of overlaps m^l converges to a nonzero limiting value m^*. If so, the system is capable of recall; if, however, $m^*=0$, all traces of the initially finite overlap with key pattern 1 are lost. The limiting value to which the dynamic process described by (26) converges is determined by the stable fixed points of the recursion, and by their domains of attraction.

Analysis of the fixed point equations [obtained from (26) by setting $m^*=m^l=m^{l+1}$ and the same for q^l], yields the following results. We find that $m^*=0$ is a stable fixed point for all α and T. For $T<1$ and $\alpha<\alpha_c(T)$ additional fixed points exist. In Fig. 2 we plot the fixed points $m^*(\alpha,T)$ as a function of α for various temperatures in the range $0\le T\le 1$. Each line of fixed points has a stable branch (solid curve), which may serve as the limiting value m^* of the dynamics, and an unstable branch (dashed curve). The two branches merge at $\alpha_c(T)$. Observing that on the stable branch $m^*\sim 1$ for low enough T, we note that the system is characterized by high recall in a range of temperatures around zero.

For a fixed value of $T<1$ and $\alpha<\alpha_c(T)$ the system converges to one of the stable fixed points, either with $m^*>0$ or $m^*=0$. To which one it actually goes depends on the value of the initial overlap m^1; there exist critical values $m_c^1(\alpha,T)$ such that for $m^1>m_c^1$ the limiting overlap is finite, while for $m^1<m_c^1$ the system flows to $m^*=0$ This defines a surface in the three-dimensional phase space (α,T,m^1) that separates a region of finite recall from one of vanishing limiting overlap. This surface is depicted in Fig. 3.

Cuts of this surface at fixed temperatures T yield Fig. 4, which gives the critical value of m^1 below which no recall is possible. That is, for $m^1>m_c^1(\alpha,T)$ the limiting overlap $m^*\neq 0$, while for $m^1<m_c^1(\alpha,T)$, $m^*=0$. As the curve $m_c^1(\alpha,T)$ is crossed, m^* jumps discontinuously to zero; the transition is first order. For example, as we decrease m^1 from 1 (for fixed $\alpha<\alpha_c$ and T), eventually the limiting m^* jumps to zero. Such behavior was seen in another model recently.[7]

Such discontinuous change of the limiting overlap appears to be a fairly common feature of various neural network models. In many instances it is not easy to see this in numerical simulations. One may find that for a finite-sized system the average m^* (averaged over the various patterns ξ, for example) is a smooth function of some variable. However, as the size is increased, the function may (slowly) become steeper. This behavior can be seen in Fig. 5(a). In such cases it is more revealing to consider the *histograms* of m^*, as demonstrated in Fig. 5(b). As can be clearly seen, especially for $N=100$, even though $\langle m^\infty\rangle$, the *average* limiting overlap, is a smooth function of m^1, the historgrams reveal a pronounced bimodal distribution, with the weights of the two peaks varying relatively slowly with m^1. This, however, is a finite-size effect; as the system size increases, the "jump" from the distribution centered on $m^*=0$ to that near 1 becomes sharper. Finite-size effects are relatively unimportant *away* from the transition region; in this regime (see Fig. 6), excellent agreement with the exact solution (valid for $N\to\infty$) is obtained, even for N as low as 200. As discussed above, finite-size effects become important as the phase boundary is approached. This can be seen to some extent in Fig. 6; while the

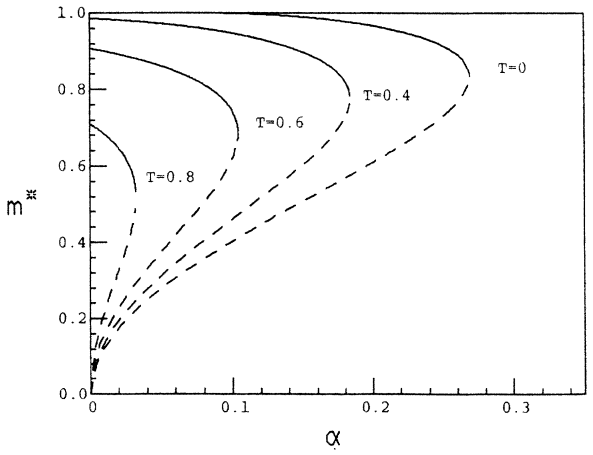

FIG. 2. Fixed points m^* of the dynamic recursions Eq. (26), vs α, for various temperatures T. The upper branches and the $m^*=0$ line (both heavy lines) are stable, the lower branches (dashed) are unstable fixed points. For each $T<1$, the two branches merge at $\alpha=\alpha_c(T)$, which is plotted in Fig. 1.

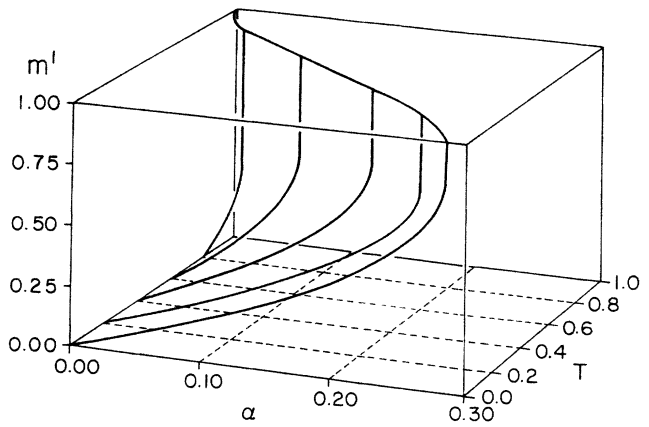

FIG. 3. Three-dimensional phase diagram. The surface shown separates the region of finite asymptotic overlap (above the surface) from that of vanishing final overlap (below the surface).

upper two curves (corresponding to initial overlaps m^1 well within the $m^* \neq 0$ phase) exhibit perfect agreement with simulations, development from $m^1 = 0.2$ does deviate slightly from the exact solution. This is due to the fact that near m_c^1 some members of the simulated ensemble flow to the "wrong" phase. However, as N increases the relative weight of these "errors" decreases. The lower curve of Fig. 6 shows another interesting effect. Even though the final overlap is 0, initially the overlap increases. Similar increase was found for the first time steps of the Little model.[14]

Another representation of the phase boundary is given in Fig. 7 which is the projection of the transition surface onto the (α, T) plane (see Fig. 3). It presents a curve $\alpha_c(T)$ that can also be obtained from Fig. 2 by plotting for each T the highest possible value of α for which a fixed point with $m^* \neq 0$ can be found.

Even though the fixed point equations derived from (26) are solved numerically, some information can be obtained analytically. For example, in order to calculate the shape of the phase boundary near the critical point we have expanded Eq. (26) (taken at the fixed point) for small $m, \alpha, \alpha q$ and $t = 1 - T$. We obtain the phase boundary equation $T \sim 1 - (\sqrt{8}/3)\sqrt{\alpha}$. Similar square-root behavior was obtained by Amit *et al.*[4] for the Hopfield model.

In a similar way we may express the deviation of $m^*(\alpha, T)$ for small α from its value at $\alpha = 0$. For $T > 0$ the value of $\alpha = 0$ is given by the solution of the equation $m_0 = \tanh(\beta m_0)$. The solution for small α deviates linearly from this, i.e., $m = m_0 - F(m_0, \beta)\alpha$. Hence, the $m^*(\alpha)$ curves of Fig. 2 have finite slopes (for $T > 0$) at $\alpha = 0$. At zero temperature, however, the deviation from $m^* = 1$ is exponentially small;

$$m^* \simeq 1 - \sqrt{2\alpha/\pi}\exp(-1/2\alpha) .$$

It is interesting to note that even though the transition is first order, the model exhibits "critical slowing down." Relaxation to the limiting value of m^* is exponential,

$$m^l - m^* \sim \exp(-l/\tau) .$$

The relaxation rate τ is determined by the recursion relations (26), linearized near m^*. Since as $\alpha \to \alpha_c$ two branches merge, one stable and one unstable, the fixed point at α_c must be marginally stable, and hence, τ must diverge. Indeed, we find that at $T = 0$,

$$\tau \sim (\alpha_c - \alpha)^{-1/2} .$$

For $T = 0$ we have extended the calculations to states which are a mixture of several patterns. That is, the ini-

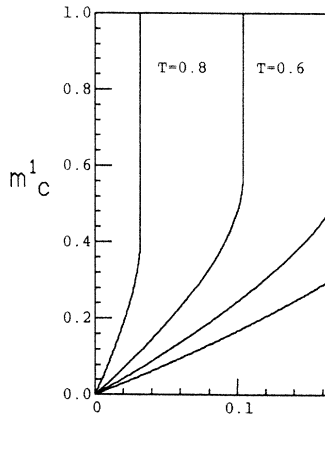

FIG. 4. The two phases, one with high remanence (memory) and one with $m^* = 0$ are separated by a first-order line. For initial overlap $m^1 < m_c^1(\alpha, T)$ vanishing limiting m^* is obtained, even for $\alpha < \alpha_c$. In the phase with $m^* > 0$ the limiting overlap is given by the upper branch of Fig. 2. We plot $m_c^1(\alpha)$ for various values of T.

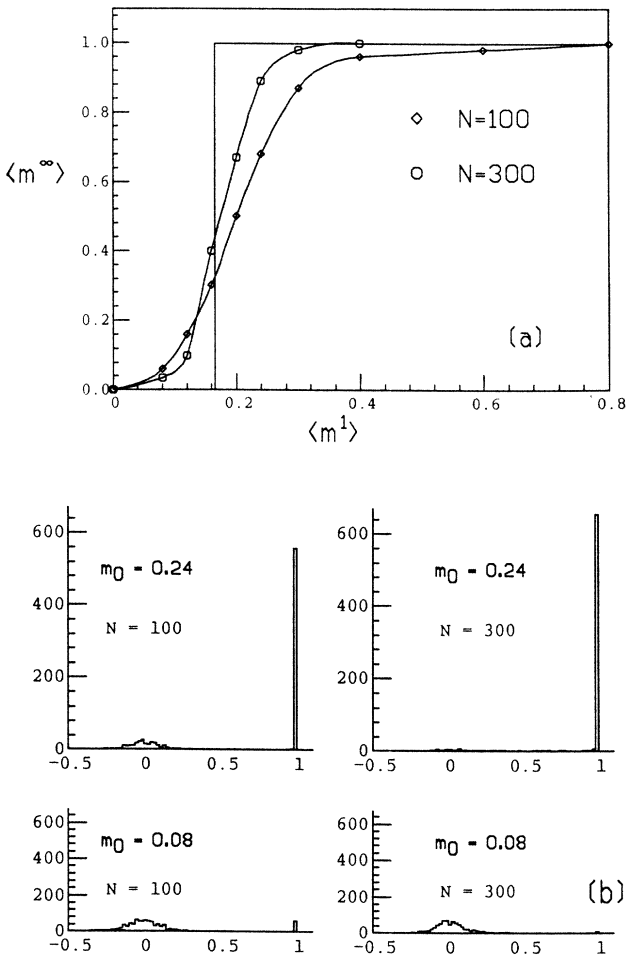

FIG. 5. (a) Average asymptotic overlap vs initial overlap. The exact solution predicts the step function shown, while numerical simulations yield the continuous curves plotted. We note that as the size of the system increases from $N = 100$ to $N = 300$ the curves approaches the step function predicted from the theory ($\alpha = 0.10$). (b) Histograms of the number of members of the simulated ensemble which end up with a given asymptotic overlap. For an initial overlap of 0.24 (larger than m_c^1 of Fig. 4) most members of the ensemble flow to final overlap near 1, and this number increases as the size of the system increases. For initial overlap of 0.08 (less than m_c^1) the "false" peak at $m \sim 1$ shrinks as the system size increases ($\alpha = 0.10$).

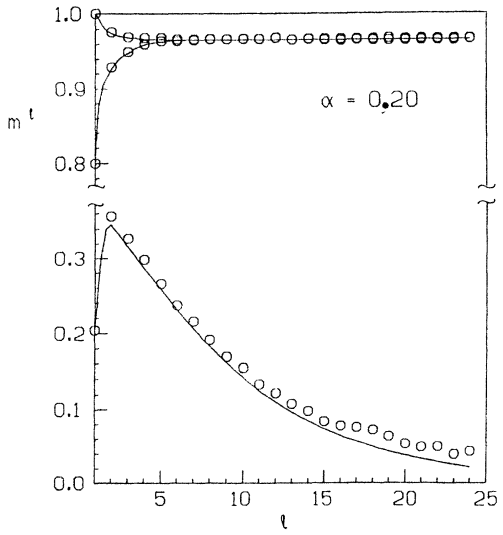

FIG. 6. Overlap m^l as function of layer index (or time) l (for $T=0$). The initial overlap for the two upper curves is above m_c^1 of Fig. 4; the circles represent simulations (with $N=200$) that agree perfectly with the analytic curves. The lower curve starts at an initial overlap of 0.2, below m_c^1. Deviations of simulations from the analytic curve are due to finite-size effects.

tial configuration has finite overlap with $\nu=1,2,\ldots,K$ patterns and an overlap of order $1/\sqrt{N}$ with the others. For example, consider a state with initial overlaps $(m_1^1,m_2^1,\ldots,m_K^1,0,\ldots 0)$. In order to present our solution for m_ν^{l+1}, we first define 2^K quantities M_k^l, with $k=1,2,\ldots,2^K$ given by

$$M_k^l=\sum_{\nu=1}^{K}\eta_{\nu,k}^l m_\nu^l \ . \tag{28}$$

Here the coefficients $\eta_{\nu,k}^l=\pm 1$, with k denoting one of the 2^K possible assignments of ± 1 as coefficients of m_ν^l. With this notation, using similar techniques to those described above, we obtain the recursion relations,

$$\begin{aligned} m_\nu^{l+1}&=\frac{1}{2^{K-1}}\sum_k \delta(\eta_{\nu,k}^l,1)\,\mathrm{erf}[M_k^l/(2\alpha q^l)^{1/2}] \ , \\ q^{l+1}&=1+\frac{1}{2^{2K}}\frac{2}{\alpha\pi}\left[\sum_k \exp[-(M_k^l)^2/2\alpha q^l]\right]^2 \ . \end{aligned} \tag{29}$$

Here δ is the Kronecker δ. For the case of finite initial overlap with *two* key patterns, i.e., $K=2$, it is convenient to define new variables: $m_+^l=m_1^l+m_2^l$ and $m_-^l=m_1^l-m_2^l$, and Eq. (29) take the form

$$\begin{aligned} m_+^{l+1}&=\mathrm{erf}[m_+^l/(2\alpha q^l)^{1/2}] \ , \\ m_-^{l+1}&=\mathrm{erf}[m_-^l/(2\alpha q^l)] \ , \\ q^{l+1}&=1+\frac{1}{2\pi\alpha}(e^{-(m_+^l)^2/2\alpha q^l}+e^{-(m_-^l)^2/2\alpha q^l})^2 \ . \end{aligned} \tag{30}$$

The result of these recursions are presented in Fig. 8. For a system whose initial state has finite overlap m_1^1 and m_2^1 with two kep patterns we find that it either (a)

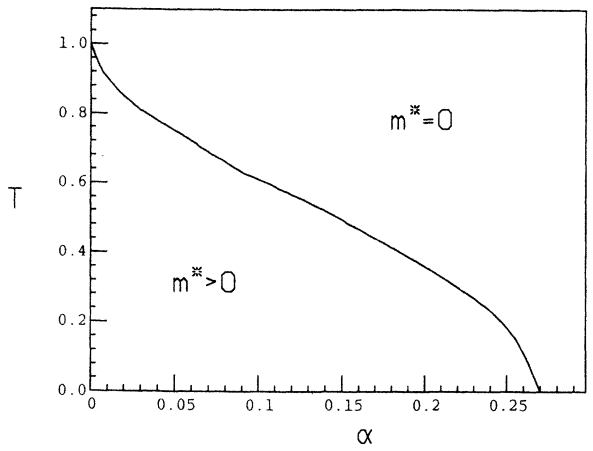

FIG. 7. Projection of the transition surface plotted in Fig. 3 onto the (α,T) plane. $m^*=0$ (no recall) for (α,T) above the transition line, and $m^*>0$ below it.

flows to no recall ($m_1^*=m_2^*=0$) or (b) the larger initial overlap wins; in this case, if $m_1^1>m_2^1$, we get $m_1^*>0$ and $m_2^*=0$ (and vice versa). Which of these limiting behaviors is realized depends on α and on the initial overlaps. Thus, states having a finite overlap with two key patterns are dynamically unstable in our network. Note that as α increases towards α_c, the region of no recall grows. We have also solved the recursion relations (29) for initial states which have finite overlaps with three or more patterns. The results we obtain in these cases are very similar to those obtained by Amit *et al.*[4] in the Hopfield model. (We are indebted to Professor H. Gut-

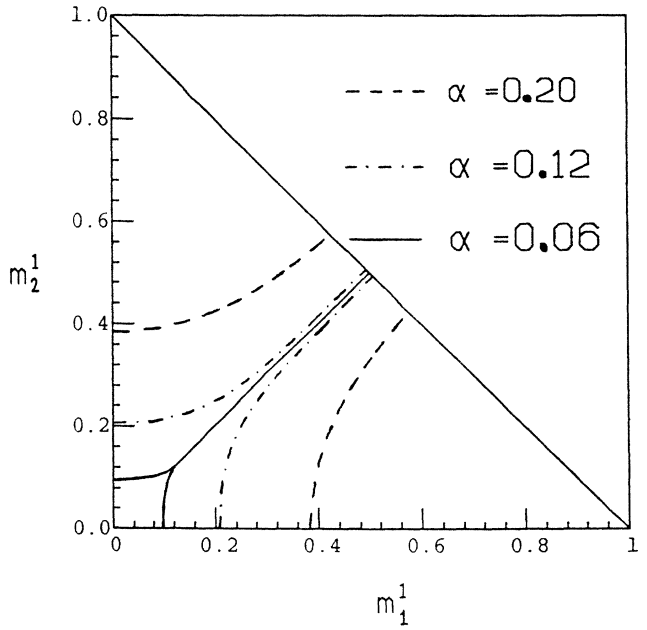

FIG. 8. An initial state with finite overlaps with *two* key patterns (m_1^1,m_2^1) will "flow" to one of three "sinks," (0,0), $(0,m^*)$, or $(m^*,0)$, depending on the initial conditions. The boundaries of the three regions are indicated for $\alpha=0.06$, 0.12, 0.20.

freund for drawing our attention to this point.) Symmetric mixture states (i.e., states which have the same finite overlap with each of the K patterns) are stable for odd K and unstable for even K.

V. DISCUSSION

In this paper we have considered a layered feed-forward neural network, with full connectivity between neighboring layers. The case of random key patterns on each layer, and corresponding Hebbian couplings, is exactly soluble. Such networks do not have an underlying Hamiltonian (or energy-Lyapunov) function. Operation of the network is defined in terms of the dynamic rule that governs the response of the network to the various inputs presented to it.

By solving the model we mean that in the limit of large N (cells per layer), given the overlap of the input with one key pattern (or more) we can calculate the overlap on all subsequent layers. In fact, we calculated the average of these quantities over all random key pattern assignments. However, since the dynamics are self-averaging, we expect (and this expectation is confirmed by numerical simulations) that the results for a single large system will agree with the average value.

One should note that so far we have solved the dynamics of the *operation* of the network. There is no *dynamic learning* (i.e., change of the couplings during the learning phase) in the present model. We plan to present results on a version, in which the dynamic learning does occur, in a future publication.[33]

One can view our model as a cellular automaton with probabilistic evolution rules. These rules themselves are random functions of "time." Viewing the Little model in a similar way we note that the difference between the two models is that the random rules in the Little model are fixed in time. Another way of viewing the difference is to say that our model is, in some sense, an annealed version of the quenched Little model. The equivalence between certain properties of annealed and quenched cellular automata has been recently demonstrated[34] for the case of the Kauffman model.[35] We do *not* expect this equivalence to hold in our case because of the infinite connectivity (see Ref. 34). Another model whose dynamics has recently been exactly solved is a diluted version of the Hopfield model[20] and extensions to other learning rules.[36] In these models it was found that the transition from the region of good recall to the region of no recall is second order. In our model (as in the Hopfield model[4]) the transition is found to be first order.

On the technical side we note that using entirely different methods, we have obtained rather similar equations to those obtained by Amit *et al.*[4] Since our model is dynamic in nature we obtain an additional order parameter p which is related to linear response.

We have generalized our model in two directions. (a) Learning of biased patterns[10] and (b) different learning schemes.[5] We have obtained exact results in both these cases.[37]

ACKNOWLEDGMENTS

We thank H. Orland, S. Levit, and B. Derrida for most helpful discussions, and B. Derrida for communicating his results prior to publication. This research was supported by the U.S.–Israel Binational Science Foundation, the Israel Academy of Sciences, and the Minerva Foundation.

APPENDIX

In this appendix we present a solution to the saddle-point equations (25). To do this we look for a solution with

$$i\hat{q}^{\,l}=0 \quad \text{for all } l \ . \tag{A1}$$

We will proceed with this assumption and show finally that it is self-consistent. Now, the equation for $p^{\,l_0}$ as given by (25) is $p^{\,l_0}=\langle i\hat{\lambda}^{\,l_0}\lambda^{l_0-1}\rangle_Z$, which is given by

$$p^{\,l_0}=\frac{1}{Z}\int\prod_{l=1}^{L-1}\frac{d\lambda^l d\hat{\lambda}^{\,l}}{2\pi}i\hat{\lambda}^{\,l_0}\lambda^{l_0-1}\exp\left[i\sum_{l=1}^{L-1}\lambda^l\hat{\lambda}^{\,l}-i\sum_{l=1}^{L-1}\hat{q}^{\,l}(\lambda^l)^2-\tfrac{1}{2}\sum_{l=1}^{L-1}(\hat{\lambda}^{\,l})^2-i\sum_{l=2}^{L-1}\hat{p}^{\,l}i\hat{\lambda}^{\,l}\lambda^{l-1}\right] . \tag{A2}$$

Z in the above expression is given in Eq. (24). Assuming $\hat{q}^{\,L-1}=0$ and integrating over λ^{L-1} gives $\delta(\hat{\lambda}^{\,L-1})$. In the last term of (A2) $\hat{\lambda}^{\,L-1}$ multiplies λ^{L-2}; hence, if $\hat{\lambda}^{\,L-1}=0$ and $\hat{q}^{\,L-2}=0$, there remains only one term with λ^{L-2}, and the integral over λ^{L-2} also yields $\delta(\hat{\lambda}^{\,L-2})$, and so on, until $l=l_0$ is reached, for which one gets $\int\hat{\lambda}^{\,l_0}\delta(\hat{\lambda}^{\,l_0})d\hat{\lambda}^{\,l_0}=0$. Thus $\hat{q}^{\,l}=0$ yields

$$p^{\,l}=0 \quad \text{for all } l \ . \tag{A3}$$

However, for $p^{\,l}=0$ it is easy to see that $I^l_{\pm}$ of Eq. (23) satisfies the relation

$$I^l_{+}+I^l_{-}=1 \ . \tag{A4}$$

The equation for q^l is obtained from calculating the average $\langle(\lambda^l)^2\rangle_Z$ appearing in Eq. (25). A straightforward evaluation of this integral, similar to that of Eq. (A1), yields

$$q^{\,l}=1++q^{\,l-1}(i\hat{p}^{\,l})^2 \ . \tag{A5}$$

For $l=1$ the above calculation yields $q^{\,1}=1$ which serves as the initial condition for the cariable q^l.

Now we must evaluate $i\hat{p}^{\,l}$ from Eq. (25). Using the expression for f given in (22) together with the last of Eqs. (25) and (A2) we get

$$i\alpha\hat{p}^{\,l}=-\frac{\partial}{\partial p^{\,l}}[I^l_{+}+I^l_{-}] \ . \tag{A6}$$

It should be noted that (A4) holds for $p^l=0$; in (A6) we must first take the derivative and then set $p^l=0$. From (23) it is easy to see that

$$\left.\frac{\partial I_+^l}{\partial p^l}\right|_{p^l=0}=\left.\frac{\partial I_-^l}{\partial p^l}\right|_{p^l=0},$$

and we find

$$i\alpha\hat{p}^l=-2\left.\frac{\partial I_+^l}{\partial p^l}\right|_{p^l=0}. \tag{A7}$$

Substituting this into (A5) yields the recursion relation for q^l, the second equation of (26). As to the recursion for m^l, we use (A4) and the last of Eq. (25) to get

$$m^l=1-2I_+^l. \tag{A8}$$

Using Eq. (23) for I_+^l (with $p^l=0$) yields the first of Eqs. (26).

Finally, in order to check self-consistency of the solution we must demand that indeed $\hat{q}^l=0$. To show this we need to evaluate $\partial f^{l+1}/\partial q^l$ which can be shown to yield

$$i\alpha\hat{q}^l=-\frac{\partial}{\partial q^l}(I_+^{l+1}+I_-^{l+1}). \tag{A9}$$

Using this result it is simple to check that indeed $\hat{q}^l=0$, and our solution is consistent with the starting assumption that led to it.

[1]W. A. Little, Math. Biosci. **19**, 101 (1975); W. A. Little and G. L. Shaw, *ibid.* **39**, 281 (1978).

[2]J. J. Hopfield, Proc. Nat. Acad. Sci. U.S.A. **79**, 2554 (1982).

[3]W. Kinzel, Z. Phys. B **60**, 205 (1985).

[4]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. **173**, 30 (1987).

[5]M. Mezard, J. P. Nadal, and G. Toulouse, J. Phys. (Paris) **47**, 1457 (1986).

[6]L. Personnaz, I. Guyon, and G. Dreyfus, J. Phys. (Paris) Lett. **46**, L-359 (1985).

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

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

[9]H. Sompolinsky, Phys. Rev. A **34**, 2571 (1986).

[10]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).

[11]G. Parisi, J. Phys. A **19**, L675 (1986).

[12]K. Binder and A. P. Young, Rev. Mod. Phys. **58**, 801 (1986).

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

[14]E. Gardner, B. Derrida, and P. Mottishaw, J. Phys. (Paris) **48**, 741 (1987).

[15]G. Parisi, J. Phys. A **13**, 1101 (1980); **13**, 1887 (1980); **13**, L115 (1980).

[16]T. Kohonen, *Self-Organization and Associative Memory* (Springer-Verlag, Berlin, 1984).

[17]F. Rosenblatt, *Principles of Neurodynamics* (Spartan, Washington D.C., 1961).

[18]M. Minsky and S. Papert, *Perceptrons* (MIT Press, Cambridge, MA, 1969).

[19]J. Hertz, G. Grinstein, and S. Solla, in *Glassy Dynamics*, edited by L. Van Hemmen and I. Morgenstern (Springer-Verlag, Berlin, 1987).

[20]H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 1861 (1986).

[21]B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).

[22]J. L. McClelland and D. E. Rummelhart, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (MIT Press, Cambridge, MA, 1986).

[23]R. Linsker, Proc. Nat. Acad. Sci. U.S.A. **83**, 7508 (1986); **83**, 8390 (1986); **83**, 8779 (1986).

[24]D. Ackley, G. Hinton, and T. Sejnowski, Cognitive Sci. **9**, 147 (1985).

[25]T. J. Sejnowski and C. R. Rosenberg, Complex Systems **1**, 145 (1987).

[26]T. J. Sejnowski, P. K. Kienker, and G. E. Hinton, Physica (Utrecht) **22D**, 260 (1986).

[27]D. E. Rummelhart and J. L. McClelland in Ref. 22, Vol. 2, p. 216.

[28]T. Hogg and B. Huberman, J. Stat. Phys. **41**, 115 (1985); Phys. Rev. Lett. **52**, 1024 (1984).

[29]E. Domany, R. Meir, and W. Kinzel, Europhys. Lett. **2**, 175 (1986).

[30]R. Scalletar and A. Zee (unpublished).

[31]R. Meir and E. Domany, Phys. Rev. Lett. **59**, 359 (1987).

[32]R. Meir and E. Domany, Europhys. Lett. **4**, 645 (1987).

[33]R. Meir and E. Domany (unpublished).

[34]B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986); B. Derrida and G. Weisbuch, J. Phys. (Paris) **47**, 1297 (1986).

[35]S. A. Kauffman, Physica (Utrecht) **10D**, 145 (1984).

[36]B. Derrida and J. P. Nadal (unpublished).

[37]R. Meir, J. Phys. (Paris) (to be published).