Information storage in neural networks with low levels of activity

Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky
Racah Institute of Physics, Hebrew University, Jerusalem, Israel
(Received 16 July 1986)

The Hopfield model of a neural network is extended to allow for the storage and retrieval of biased patterns, $\{\xi_i^\mu\}$, where $N^{-1}\sum_i^N \xi_i^\mu = a$ is arbitrary. Such patterns represent levels of activity (i.e., percentage of firing neurons) equal to $\frac{1}{2}(1+a)$, $-1<a<1$. If the coupling constants (synaptic efficacies) are constructed as in the original Hopfield model, the system can retrieve at most a very small number of patterns ($p<1+a^{-2}$). This is due to the finite correlations (overlaps) between the patterns. The model is modified by subtracting the bias a from each pattern as it enters into the couplings. This modification restores the ability of the model to store a macroscopic number of patterns. Yet spurious states are found to plague the dynamics. It is then argued that the dynamics of the network should be consistent with the levels of activity of the stored patterns. This is implemented by adding a global constraint, which restricts the configuration space to states whose mean activity is in the neighborhood of $\frac{1}{2}(1+a)$. The consequences of the restricted dynamics are analyzed in the replica symmetric mean-field theory. The global constraint suppresses spurious states and leads to the unexpected result that the storage capacity is higher than that of the unbiased network, up to very high values of the bias ($|a|\simeq 0.99$). However, the information content of such networks is shown to be a monotonically decreasing function of a.

I. INTRODUCTION

A fair amount of analytical control over the Hopfield-Little[1,2] model of a neural network has been achieved.[3-5] So much so that some of the underlying simplifications initially introduced in the model have begun to be lifted. In particular, the properties of the network as an associative memory have been shown to be robust under the introduction of thermal noise and the dilution of synapses, as well as under the clipping of synaptic efficacies.[3,6]

Here we address another restriction—that of 50% mean neural activity. It has been a standing feature of the model since Hopfield that the N-neuron network stores p patterns ξ_i^μ ($\mu=1,\ldots,p$; $i=1,\ldots,N$), where $\xi_i^\mu=\pm 1$ with equal probability. Hence, in each stored pattern 50% of the neurons are active ($+1$) and 50% are passive (-1). Consequently, in the process of retrieval 50% of the neurons are active, on the average. Moreover, the patterns are uncorrelated; namely, in a large network

$$\frac{1}{N}\sum_i \xi_i^\mu \xi_i^\nu = 0 , \tag{1.1}$$

if $\mu\neq\nu$.

This situation is unsatisfactory on several counts.

1. Neurophysiological evidence[7] indicates that mean firing rates are significantly lower than 50%.

2. The Hopfield dynamics is symmetric with respect to the interchange of firing by nonfiring ($S_i\rightarrow -S_i$). Hence, with every stored pattern the network stores the reversed pattern as well.[3] Usually, this type of symmetry, when undesired, is lifted by an external field (threshold). But here, if the learned patterns are of the Hopfield random type, then near the stored patterns (in configuration space) the "magnetization" of the states vanishes. Therefore, it is difficult to find a natural way of suppressing the doubling of the stored patterns.

3. Pattern recognition usually deals with contexts in which the background presents a much larger area than the foreground. Hence, if the coding of such contexts on a neural network maps larger areas on larger groups of active (or passive) neurons, then the levels of activity must be allowed to differ significantly from 50%.

4. Recent treatments of neural networks focused in large part on totally uncorrelated patterns as described by (1.1). More realistic networks have to confront the presence of correlated patterns in models of associative memory. Several models which deal with correlated patterns have been proposed.[8-13] These models involve, however, a much more complicated dependence of J_{ij} on the learned patterns. In order to learn and retrieve correlated patterns, one has to introduce nonlocality, either in the learning process or in the dynamics.

To deal with these aspects we have studied associative memory of random patterns whose mean activities differ from 50%. We refer to such patterns as *biased patterns*. For instance, every component ξ_i^μ in a learned pattern can be chosen independently with probability $P(\xi)$,

$$P(\xi)=\tfrac{1}{2}(1+a)\delta(\xi-1)+\tfrac{1}{2}(1-a)\delta(\xi+1) . \tag{1.2}$$

The average of each ξ is a and the mean activity in each of the stored patterns is

$$\int d\xi\, P(\xi)\tfrac{1}{2}(\xi+1)=\langle\langle \tfrac{1}{2}(\xi+1)\rangle\rangle=\tfrac{1}{2}(1+a) . \tag{1.3}$$

Since $-1\leq a\leq 1$, one has an arbitrary level of activity, as well as an arbitrary ratio of background to foreground.

With such a distribution of stored patterns, memories are necessarily correlated, though in a rather simple way. One has

$$\langle\langle \xi_i^\mu \xi_i^\nu \rangle\rangle = \delta^{\mu\nu} + a^2(1-\delta^{\mu\nu}) \ . \tag{1.4}$$

Thus, if for example one chooses to represent two-dimensional figures directly on a two-dimensional arrangement of the N neurons (see, e.g., Refs. 7 and 11), the considerations of point 3 above imply that such correlations have to be allowed.

Furthermore, in the language of magnetism, each stored biased pattern has a net magnetization (equal to a per spin) a uniform "external field"—neuronal threshold—will break the symmetry between firing and nonfiring, in a natural way, which is independent of the particular patterns stored. If that field is large enough, it will eliminate the reversed patterns mentioned in point 2, above, even as metastable states.

The introduction of asymmetric, biased patterns raises the following questions.

(1) Should the dynamics be modified?

(2) With plausible dynamics, what is the quality of retrieval?

(3) What is the effect on spurious states?

(4) How is the storage capacity modified?

(5) What is the information content of the network?

In Sec. II we show that a naive application of Hopfield's dynamics is catastrophic in that even at small values of the bias parameter a, the stored patterns become unstable at very low storage levels. This is due to the fact that the noise generated by the other patterns[14,15] in the retrieval of each pattern does not average to zero.

In Sec. III we propose one modification, namely, the replacement of the synaptic efficacies[2]

$$J_{ij} = \frac{1}{N} \sum_\mu \xi_i^\mu \xi_j^\mu$$

by

$$\bar{J}_{ij} = \frac{1}{N} \sum_\mu (\xi_i^\mu - a)(\xi_j^\mu - a) \ . \tag{1.5}$$

This modification has a number of attractive features. Firstly, it removes the catastrophe by shifting the noise back to have a zero mean. Considered from a biological point of view, (1.5) can be interpreted as learning by modification of synaptic efficacies due to neural activity. If the network persists in some activity state $S_i = \xi_i^\nu$, then

$$\Delta \bar{J}_{ij} = c(\xi_i^\nu - a)(\xi_j^\nu - a) \ . \tag{1.6}$$

Since

$$\sum_i \xi_i^\nu = Na \ ,$$

we have

$$\sum_j \Delta \bar{J}_{ij} = 0 \ , \tag{1.7}$$

which can be read as follows: *The total modification of synapses on a given neuron (i) is unchanged during learning.* This is, of course, also a property of the original Hopfield description,[1] if the patterns are uncorrelated. This property fits in nicely with the hypothesis that the modification of synapses takes place by a local redistribution of receptors, whose total number (per neuron) is conserved.[8] Equation (1.5) is an expression of the nonlocality in our learning scheme. In getting modified each synapse has to be aware of the mean, allowed activity rate of the entire network.

The consequences of the traditional dynamics[1,3] with the couplings (1.5) are studied in Sec. III for finite p and are found to be unsatisfactory: As a increases, spurious states start dominating the dynamics, *even though the stored patterns remain locally stable.* The number of spurious states is found to increase with a, and they become the absolute minima of the energy as a increases.

In Sec. IV this system is studied near saturation—$p = \alpha N$—and the difficulties persist. Here, again, spurious states are found to dominate the energy landscape. For example, already at rather low values of the bias a, the critical storage level α_c of the spurious states becomes much higher than that of the stored patterns.

The results of Secs. III and IV indicate that to have a network which can effectively store and retrieve biased patterns, it does not suffice to modify the couplings. In fact, from a biological perspective, there are rather compelling reasons why the dynamical process should be modified as well. The fact that the network has learned patterns of mean bias a implies that the activity of the network is constrained so that it wanders mostly among states that have the preferred mean activity. In other words, there must be some global control on the dynamics of the network which prevents too high or too low activity. If neural activities are much lower than 50%, this must be so whether or not the network is retrieving. The control restricts the regions in state space in which a healthy neural network can move. This is why it learns a restricted class of patterns, and it is in these regions that it should be trying to retrieve.

Section V is devoted to the study of neural networks in which the dynamics is constrained to be consistent with the bias in the patterns. First we show that if the dynamics is restricted rigidly to states with a given value for the "magnetization"

$$M \equiv \sum_i S_i = Na$$

[neural activity $\frac{1}{2}N(1+a)$], then the learned patterns are stable, and spurious states, with macroscopic overlaps with small numbers of patterns, do not appear. Moreover, it is shown that such a network has a higher storage capacity α_c than that of the unbiased network—$\alpha_c(a) > \alpha_c(0)$ for $|a| < 0.99$. This, however, does not imply higher information content because the space of memorized patterns has been restricted. This aspect is discussed in Sec. VI.

A biological system is not expected to impose such a global constraint rigidly. Hence Sec. V goes on to study the functioning of a network whose dynamics is a Monte Carlo walk with

$$H = -\tfrac{1}{2} \sum_{i,j} J_{ij} S_i S_j + \frac{g}{2N}(M - Na)^2 \ , \tag{1.8}$$

where J_{ij} is given by (1.5), and g is a parameter measuring the stiffness of the constraint.

What the quadratic term in (1.8) implies is that every

neuron feels an extra *uniform* contribution to its threshold (local field), proportional to the deviation of the level of activity from the normal level. For very large values of g one returns to the rigid constraint with reduced numbers of spurious states, high α_c, and excellent retrieval quality. The dependence of these properties on g is computed analytically and is compared with simulations.

In Sec. VII we make some concluding remarks about the class of constraints to which the conclusions of the paper apply. Finally, in the Appendix it is shown that the control of the mean activity can be imposed globally, rather than locally, as in Eq. (1.2).

II. NAIVE HOPFIELD DYNAMICS

If one follows in the footsteps of Ref. 1, with the ξ^μ's chosen according to the asymmetric distribution (1.2), then the dynamics of the network is governed by the Hamiltonian

$$H=-\tfrac{1}{2}\sum_{i,j}J_{ij}S_iS_j \tag{2.1}$$

with

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu \ . \tag{2.2}$$

The inadequacy of such a network is already apparent when the number of stored patterns, p, remains finite as $N\to\infty$. The free energy and the equations for the stable configurations are derived precisely as in Ref. 3 (to be referred to as I). They are, respectively,

$$f=\tfrac{1}{2}\sum_\mu (m^\mu)^2-\frac{1}{\beta}\langle\langle \ln 2\cosh(\beta\mathbf{m}\cdot\boldsymbol{\xi})\rangle\rangle \ , \tag{2.3}$$

$$m^\mu=\langle\langle \xi^\mu\tanh(\beta\mathbf{m}\cdot\boldsymbol{\xi})\rangle\rangle \ , \tag{2.4}$$

where β is the inverse temperature (synaptic noise) and

$$m^\mu=\frac{1}{N}\sum_i\langle\xi_i^\mu S_i\rangle \ , \tag{2.5}$$

The single angular brackets denote a thermal average and the double ones stand for an average over the quenched distribution of the ξ's.

At $T=0$ Eq. (2.4) becomes

$$m^\mu=\langle\langle \xi^\mu\mathrm{sgn}(\mathbf{m}\cdot\boldsymbol{\xi})\rangle\rangle \ . \tag{2.6}$$

It is easily verified that (2.6) has no Mattis-type solutions, in which only a single component of $\mathbf{m}$ is nonzero. The reason is that if $m^1\neq 0$ while $m^i=0$ for $i>1$, then (2.6) gives

$$m^i=\langle\langle \xi^i\mathrm{sgn}\xi^1\rangle\rangle=\langle\langle \xi^i\xi^1\rangle\rangle=a^2\neq 0 \ ,$$

which is the uniform correlation of two biased, random patterns, mentioned in (1.4).

At sufficiently low p the original stored patterns—$S_i=\xi_i^\nu$—are still stable. However, they become destabilized above p which satisfies

$$(p-1)a^2=1 \ . \tag{2.7}$$

This is due to the fact that the local field acting on the spins[1,15] has a contribution from the other patterns, which does not average to zero. In the state $S_i=\xi_i^\nu$,

$$h_i=\sum_{\substack{j\\(j\neq i)}}J_{ij}S_j=\frac{1}{N}\sum_{\substack{j\\(j\neq i)}}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu\xi_j^\nu\cong\xi_i^\nu(1+\delta_i) \tag{2.8}$$

and the noise is

$$\begin{aligned}\delta_i&=\frac{1}{N}\sum_{\substack{j\\(j\neq i)}}\sum_{\substack{\mu,\nu\\(\mu\neq\nu)}}\xi_i^\mu\xi_j^\mu\xi_i^\nu\xi_j^\nu\\&=\langle\langle \xi^\mu\xi^\nu\rangle\rangle\sum_{\substack{\mu,\nu\\(\mu\neq\nu)}}\xi_i^\mu\xi_i^\nu\\&=a^2\sum_{\substack{\mu,\nu\\(\mu\neq\nu)}}\xi_i^\mu\xi_i^\nu \ ,\end{aligned} \tag{2.9}$$

which implies that (2.7) is the limit of stability.

III. DYNAMICS WITH MODIFIED HEBB RULE

The first corrective option is to replace the acquired synaptic efficacies J_{ij} of (2.2) by $\bar{J}_{ij}$, Eq. (1.5). The system will then follow the dynamics of the Hamiltonian

$$H=-\frac{1}{2N}\sum_{i,j}\sum_\mu(\xi_i^\mu-a)(\xi_j^\mu-a)S_iS_j \ . \tag{3.1}$$

The consequences of this modification will be described below.

A. Finite number of patterns

The order parameters are now the modified overlaps

$$m^\mu=\frac{1}{N}\sum_i\langle(\xi_i^\mu-a)S_i\rangle \ , \tag{3.2}$$

which vary in the interval $0\leq m^\mu<(1-a^2)$. One finds for the free energy and for the saddle-point equations

$$f=\tfrac{1}{2}\sum_\mu(m^\mu)^2-\frac{1}{\beta}\left\langle\!\!\left\langle \ln 2\cosh\left[\beta\sum_\mu m^\mu(\xi^\mu-a)\right]\right\rangle\!\!\right\rangle \ , \tag{3.3}$$

$$m^\mu=\left\langle\!\!\left\langle (\xi^\mu-a)\tanh\left[\beta\sum_\mu m^\mu(\xi^\mu-a)\right]\right\rangle\!\!\right\rangle \ . \tag{3.4}$$

Since

$$\langle\langle(\xi^\mu-a)(\xi^\nu-a)\rangle\rangle=0 \ , \tag{3.5}$$

it follows that there are always retrieval state solutions

$$m^\mu=m\delta^{\mu,\nu}$$

for which

$$m=\tfrac{1}{2}(1-a^2)[\tanh\beta m(1-a)+\tanh\beta m(1+a)] \ . \tag{3.6}$$

In the limit $T\to 0$

$$m=1-a^2 \ , \tag{3.7}$$

which represents a perfect overlap with the νth stored pattern, namely,

$$\bar{m}=\frac{1}{N}\sum_i \langle \xi_i^\mu S_i\rangle = 1$$

in this state.

One can show, following I, that all symmetric mixture states are solutions of Eq. (3.4). These solutions are classified by the number of components of m, which are nonzero and have equal magnitude.

The symmetric mixtures are

$$m=(m_n,\ldots,m_n,0,\ldots,0) \tag{3.8}$$

with

$$nm_n=\left\langle\!\!\left\langle \sum_{\mu=1}^{n}(\xi^\mu-a)\tanh\left[\beta m\sum_{\mu=1}^{n}(\xi^\mu-a)\right]\right\rangle\!\!\right\rangle$$
$$\underset{T\to 0}{\to}\left\langle\!\!\left\langle \sum_{\mu=1}^{n}(\xi^\mu-a)\,\mathrm{sgn}\left[\sum_{\mu=1}^{n}(\xi^\mu-a)\right]\right\rangle\!\!\right\rangle \tag{3.9}$$

and

$$m^{n+1}=\left\langle\!\!\left\langle(\xi^{n+1}-a)\tanh\left[\beta m_n\sum_{\mu=1}^{n}(\xi^\mu-a)\right]\right\rangle\!\!\right\rangle$$
$$\times\langle\!\langle(\xi^{n+1}-a)\rangle\!\rangle$$
$$\times\left\langle\!\!\left\langle\tanh\left[\beta m_n\sum_{\mu=1}^{n}(\xi^\mu-a)\right]\right\rangle\!\!\right\rangle=0\ . \tag{3.10}$$

Undesired features appear when the stability of the mixture states is examined. In the unbiased case[3] all these saddle points appeared, but only the ones with an odd number of mixed patterns turned out to be stable. The even mixtures were found to be unstable. In fact, they were found to be unstable not only at $T=0$, but for all T.

The instability of the even mixtures can be traced to the appearance of many sites at which the local field

$$h_i=m_n\sum_\mu \xi_i^\mu \equiv m_n z_{ni} \tag{3.11}$$

vanishes. In fact, there are

$$\frac{1}{2N}\begin{bmatrix}N\\N/2\end{bmatrix}$$

such sites. But when the patterns are biased, the local field is

$$h_i=m_n(z_{ni}-na)\ , \tag{3.12}$$

which will vanish only for special values of a, and for those only for special values of n.

Thus at $T=0$ the number of stable symmetric spurious states doubles. Moreover, as $|a|$ increases the energies of the states begin to cross, and as $|a|$ becomes greater than $\sqrt{2}-1$, mixture states become the global minima of f. The beginnings of the sequence of crossings are shown in Fig. 1, in which the energy of the first few states is plotted against the value of the bias.

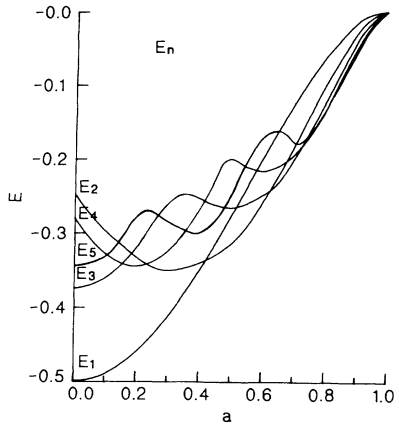

FIG. 1. The energies of the first five symmetric mixture states at $T=0$, for a network with a finite number of stored patterns, vs the bias. The network has modified Hebb synapses and unconstrained dynamics.

B. Finite temperature

The spurious states in the *unbiased* network are controlled by the synaptic noise (temperature). For $0.46<T<1$ only the pure patterns are stable. Considering the biased network at finite temperature, one finds that there is a critical temperature

$$T_c(a)=1-a^2\ . \tag{3.13}$$

Slightly below $T_c(a)$ one finds, on expanding (3.9), that

$$m_1^2=\frac{3T_c}{1+3a^2}(T_c-T) \tag{3.14}$$

and the eigenvalue measuring the stability against the mixing of a second pattern is

$$\lambda=\frac{2-6a^2}{1+3a^2}\frac{T-T_c}{T_c}\ . \tag{3.15}$$

Thus the retrieval (Mattis) state becomes unstable near T_c for $a^2>\frac{1}{3}$. As the temperature is lowered, one first encounters stable spurious states. Only at lower temperature do the stored patterns become stable, at which point the spurious states (the even ones) are already lower in energy.

C. Higher storage levels—signal-to-noise considerations

When the number of stored patterns increases with N as

$$p=\alpha N\ ,$$

the embedded patterns in the unbiased network are no longer stable.[4] Nevertheless, for $\alpha<0.138$ (in the replica symmetric phase) the unbiased network retrieves with high fidelity the individual stored patterns. Thus one is tempted to ask what would be the consequences of approaching saturation in the biased network. As a first

step we repeat the storage capacity estimate based on the requirement of small noise-to-signal ratio in the stored patterns—insuring their stability.[1,15] The local field at neuron i in pattern 1 is

$$h_i=\sum_{\substack{j\\(j\neq i)}} J_{ij}S_j=\frac{1}{N}\sum_{\substack{j\\(j\neq i)}}\sum_{\mu=1}^{p}(\xi_i^\mu-a)(\xi_j^\mu-a)\xi_j^1$$
$$=\xi_i^1\left[(1-a^2)(1-\xi_i^1 a)+\frac{1}{N}\sum_{\substack{j\\(j\neq i)}}\sum_{\mu>1}^{p}(\xi_i^\mu-a)(\xi_j^\mu-a)\xi_j^1\right]. \tag{3.16}$$

Since in the noise term $j\neq i$ and $\mu\neq 1$, its mean vanishes. The mean of the square of the noise term is

$$R^2=\frac{(N-1)(p-1)}{N^2}(1-a^2)^2\simeq\frac{p}{N}(1-a^2)^2 . \tag{3.17}$$

On the other hand, the signal S, the first term in the large parentheses in (3.16), has a minimum at

$$S_0=(1-a^2)(1-|a|), \tag{3.18}$$

hence

$$\frac{R}{S_0}\simeq\sqrt{p/N}\,\frac{1}{1-|a|} . \tag{3.19}$$

Thus, in contrast to the naive introduction of biased patterns discussed in Sec. II, if $p/N\ll(1-|a|)^2$, the original patterns are stable at $T=0$. On the other hand, we have here an indication, even on this simple level, that the storage capacity decrease as a increases.

A more careful treatment of the probability distribution of the noise term[15] would lead to the conclusion that patterns will be retrieved with no errors if

$$\alpha<\alpha_c=\frac{(1-|a|)^2}{2\ln N}=(1-|a|)^2\alpha_c(0) .$$

IV. BIASED NETWORK NEAR SATURATION

A. The average free energy

When $\alpha=p/N$ is finite, the calculation follows in the footsteps of Ref. 4 (to be referred to as II), and the quenched averaging over the ξ's is performed using the replica method. Thus the free energy is computed from the average

$$\langle\langle Z^n\rangle\rangle=\left\langle\!\!\left\langle \mathrm{Tr}_{S^\rho}\exp\left[\frac{\beta}{2N}\sum_{i,j,\mu,\rho}(\xi_i^\mu-a)S_i^\rho(\xi_j^\mu-a)S_j^\rho-\tfrac{1}{2}\beta pn(1-a^2)+\beta\sum_\nu h^\nu\sum_{i,\rho}(\xi_i^\nu-a)S_i^\rho\right]\right\rangle\!\!\right\rangle , \tag{4.1}$$

where ν $(=1,\ldots,s)$ denotes the patterns which are candidates for condensation.

Linearizing the quadratic term in ξ, for the uncondensed patterns $(\mu>s)$, by a Gaussian transformation and averaging over the high ξ's, one finds

$$\langle\langle Z^n\rangle\rangle=\exp[-\tfrac{1}{2}\beta pn(1-a^2)]$$
$$\times\left\langle\!\!\left\langle \mathrm{Tr}_{S^\rho}\int\prod_{\mu,\rho}(dm_\rho^\mu/\sqrt{2\pi})\exp\left[-\tfrac{1}{2}\sum_{\mu,\rho}(m_\rho^\mu)^2+\sum_{i,\mu}\ln\left[\cosh\sqrt{\beta/N}\sum_\rho m_\rho^\mu S_i^\rho+a\sinh\sqrt{\beta/N}\sum_\rho m_\rho^\mu S_i^\rho\right]\right]\right.$$
$$\left.\times\exp\left[-a\sqrt{\beta/N}\sum_{\substack{i,\rho,\mu\\(\mu>s)}}m_\rho^\mu S_i^\rho\right]\exp\beta N\left[-\tfrac{1}{2}\sum_{\nu,\rho}(m_\rho^\nu)^2+\sum_{\nu,\rho}(m_\rho^\nu+h^\nu)\frac{1}{N}\sum_i(\xi_i^\nu-a)S_i^\rho\right]\right\rangle\!\!\right\rangle . \tag{4.2}$$

If $m_\rho^\mu=O(1)$, for $\mu>s$, one can expand the ln in (4.2) and keep only the quadratic terms. Then integrating over m_ρ^μ one finds

$$\langle\langle Z^n\rangle\rangle=\exp-\tfrac{1}{2}\beta pn(1-a^2)\mathrm{Tr}_{S^\rho}\exp-\tfrac{1}{2}p\,\mathrm{Tr}\ln[1-\beta(1-a^2)\underline{Q}]$$
$$\times\left\langle\!\!\left\langle\int\prod_{\nu,\rho}(dm_\rho^\nu/\sqrt{2\pi})\exp\beta N\left[-\tfrac{1}{2}\sum_{\nu,\rho}(m_\rho^\nu)^2+\sum_{\nu,\rho}(m_\rho^\nu+h^\nu)\frac{1}{N}\sum_i(\xi_i^\nu-a)S_i^\rho\right]\right\rangle\!\!\right\rangle , \tag{4.3}$$

where, as in II,

$$Q_{\rho\sigma}=\frac{1}{N}\sum_i S_i^\rho S_i^\sigma,\quad Q_{\rho\rho}=1 . \tag{4.4}$$

Introducing $r_{\rho\sigma}$ as Lagrange multipliers for the nondiagonal elements of $\underline{Q}$, $q_{\rho\sigma}$, one finds for the extensive part of the free energy

$$\beta f=\tfrac{1}{2}\alpha\bar\beta+(\alpha/2n)\mathrm{Tr}\ln[(1-\bar\beta)-\bar\beta q]+(\beta/2n)\sum_{\nu,\rho}(m_\rho^\nu)^2+(\alpha\beta^2/2n)\sum_{\substack{\rho,\sigma\\(\rho\neq\nu)}}r_{\rho\sigma}q_{\rho\sigma}$$
$$-(1/n)\left\langle\!\!\left\langle\ln\mathrm{Tr}_S\exp\left[\tfrac{1}{2}\alpha\beta^2\sum_{\substack{\rho,\sigma\\(\rho\neq\sigma)}}r_{\rho\sigma}S^\rho S^\sigma+\beta\sum_{\nu,\rho}(m_\rho^\nu+h^\nu)(\xi^\nu-a)S^\rho\right]\right\rangle\!\!\right\rangle , \tag{4.5}$$

with $\bar\beta=(1-a^2)\beta$.

B. Replica symmetric theory

In the replica symmetric phase (4.5) becomes

$$\beta f=\tfrac{1}{2}\alpha\bar{\beta}+\tfrac{1}{2}\sum_{\nu}(m^{\nu})^2+\tfrac{1}{2}\alpha\left[\ln(1-\bar{\beta}+\bar{\beta}q)-\frac{\bar{\beta}q}{1-\bar{\beta}(1-q)}\right]+\tfrac{1}{2}\alpha\beta^2 r(1-q)$$
$$-\left\langle\!\left\langle \ln 2\cosh\beta\left[\sqrt{\alpha r}z+\sum_{\nu}(m^{\nu}+h^{\nu})(\xi^{\nu}-a)\right]\right\rangle\!\right\rangle , \tag{4.6}$$

where in (4.6) the double angular brackets indicate a Gaussian average over z as well as an average over the discrete ξ^{ν}, with the ξ's distributed according to (1.2).

The saddle-point equations for the order parameters m^{ν}, q, and r are

$$m^{\nu}=\left\langle\!\left\langle(\xi^{\nu}-a)\tanh\beta\left[\sqrt{\alpha r}z+\sum_{\nu}(m^{\nu}+h^{\nu})(\xi^{\nu}-a)\right]\right\rangle\!\right\rangle , \tag{4.7a}$$

$$q=\left\langle\!\left\langle\tanh^2\beta\left[\sqrt{\alpha r}z+\sum_{\nu}(m^{\nu}+h^{\nu})(\xi^{\nu}-a)\right]\right\rangle\!\right\rangle , \tag{4.7b}$$

$$r=q(1-a^2)^2/[1-\bar{\beta}(1-q)]^2 . \tag{4.7c}$$

C. Solutions at $T=0$—critical storage level

We now choose $h=0$ and investigate the retrieval states, $\mathbf{m}=(m,0,\ldots,0)$ in the limit $\beta\to\infty$. Since

$$\int\frac{dz}{\sqrt{2\pi}}\exp(-\tfrac{1}{2}z^2)\tanh\beta(\sqrt{\alpha r}z+x)\underset{\beta\to\infty}{\to}\operatorname{erf}(x/\sqrt{2\alpha r}) , \tag{4.8}$$

Eq. (4.7a) becomes

$$m=\tfrac{1}{2}(1-a^2)\{\operatorname{erf}[m(1+a)/\sqrt{2\alpha r}]+\operatorname{erf}[m(1-a)/\sqrt{2\alpha r}]\} . \tag{4.9}$$

The order parameter $q\to 1$, but

$$C\equiv\beta(1-q)\underset{T\to 0}{\to}\sqrt{2/\pi\alpha r}\,\langle\!\langle\exp[-m^2(\xi-a)^2/2\alpha r]\rangle\!\rangle$$
$$=\sqrt{1/2\pi\alpha r}\,\{(1+a)\exp[-m^2(1+a)^2/2\alpha r]+(1-a)\exp[-m^2(1-a)^2/2\alpha r)]\} \tag{4.10}$$

and Eq. (4.7c) becomes

$$r=\frac{(1-a^2)^2}{[1-(1-a^2)C]^2} . \tag{4.11}$$

Equations (4.9)–(4.11) are solved numerically. They exhibit a sharp transition at $\alpha=\alpha_c(a)$. Below α_c there exists a dynamically stable retrieval state, which is macroscopically stable—with $m>m_c(a)$. Above $\alpha_c(a)$ the only macroscopically stable state is $m=0$—the spin-glass state. In Fig. 2 we present α_c for the retrieval state versus a. For comparison we plot in this figure $\alpha_c(0)(1-|a|)^2$, which would have been suggested by the signal-to-noise analysis of Sec. III C above. In Fig. 3 we present the values of m_c as a function of a. The dominance of the spurious mixture states prevails at finite α as well. The maximum value of α for which spurious states with symmetric mixtures of two and three patterns are stable [α_c,

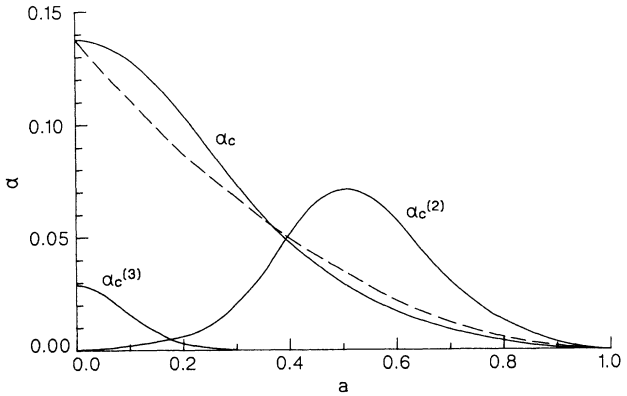

FIG. 2. Critical storage ratio α_c vs bias for retrieval states and symmetric mixtures of two and three patterns. The unmarked curve is $\alpha(0)(1-|a|)^2$.

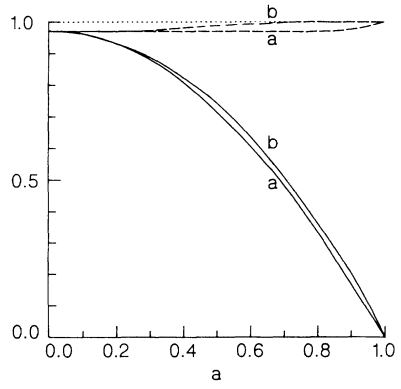

FIG. 3. Overlap vs bias at $\alpha_c(a)$. Solid curves represent the order parameter m [Eq. (3.2)], dashed curves represent the total overlap $\overline{m}$: a, constrained dynamics (with rigid constraint, see Sec. V); b, unconstrained dynamics.

(Refs. 2 and 3)] are plotted in Fig. 2. It should be pointed out that for $a>0.4$ there is a range of α above α_c in which the mixture of two patterns remains stable, while the memorized patterns can no longer be retrieved. Note that the full overlap of the states with a pattern is $m+a^2$. See Eq. (3.2).

D. The entropy and breaking of replica symmetry

One indication for the importance of the breaking of replica symmetry (RSB) in the retrieval states is the magnitude of the negative entropy in these states at $T=0$ relative to that of the spin-glass state, or in the symmetric retrieval states. We therefore calculate

$$S=-\frac{\partial f}{\partial T}(T=0)=-\tfrac{1}{2}\alpha[\ln(1-\bar{C})+\bar{C}/(1-\bar{C})] , \tag{4.12}$$

where

$$\bar{C}=(1-a^2)\lim_{T\to 0} C$$

and $\lim C$ is given by the right-hand side (rhs) of Eq. (4.10).

The value of S is clearly negative. Its value at α_c decreases with increasing a. For example, at $a=0$, $S=-1.4\times10^{-3}$,[4] at $a=0.3$, $S=-7.7\times10^{-4}$, and at $a=0.8$, $S=-1.6\times10^{-6}$. We conclude that even though the negative value of the entropy indicates that replica symmetry must be broken in the retrieval states, the small absolute value of the entropy implies that the effect cannot be very significant (see, e.g., II).

V. BIASED NETWORK WITH CONSTRAINED DYNAMICS

A. Rigid constraint

We consider here a network whose phase space is restricted to all the states $\{S_i\}$ which obey the constraint

$$\frac{1}{N}\sum_i^N S_i=a . \tag{5.1}$$

Thus the mean activity of the network is fixed during the dynamic evolution and has the same value as the mean activity of the embedded patterns. The Hamiltonian which governs the dynamics is given by Eq. (3.1). The partition function can be represented by

$$Z=\int_{-i\infty}^{i\infty}\frac{dh}{2\pi}e^{-Nah}\mathrm{Tr}_S\exp\left[-\beta H+\beta h\sum S_i\right] . \tag{5.2}$$

Since the constraint is global, the integral over h is dominated by the saddle-point value h_0. This amounts to adding an external field h_0, whose magnitude guarantees the constraint equation (5.1).

The ensemble average free energy of Eq. (5.2) can be studied by the same methods as in Secs. III and IV. We first discuss the effect of the constraint in the finite-p case.

1. The finite-p case

The free energy per spin is

$$f=\tfrac{1}{2}\sum_\mu (m^\mu)^2 - \frac{1}{\beta}\left\langle\!\left\langle \ln 2\cosh\beta\left[\sum_\mu m^\mu(\xi^\mu-a)+h_0\right]\right\rangle\!\right\rangle . \tag{5.3}$$

The parameters m^μ and h_0 are determined by

$$m^\mu=\frac{1}{N}\sum_i^N\langle S_i\rangle(\xi_i^\mu-a)$$

$$=\left\langle\!\left\langle(\xi^\mu-a)\tanh\beta\left[\sum_\mu m^\mu(\xi^\mu-a)+h_0\right]\right\rangle\!\right\rangle \tag{5.4}$$

and

$$\frac{1}{N}\sum_i^N\langle S_i\rangle=\left\langle\!\left\langle\tanh\beta\left[\sum_\mu m^\mu(\xi^\mu-a)+h_0\right]\right\rangle\!\right\rangle=a . \tag{5.5}$$

At high temperature the thermodynamic state is uncorrelated with the patterns: $m^\mu=0$ for all μ, and h_0 is the field required to induce the magnetization

$$\langle S_i\rangle=\tanh\beta h_0=a .$$

Below T_c, p ordered states appear, characterized by $m^\nu=\delta^{\mu\nu}m$,

$$m=\left\langle\!\left\langle(\xi^\mu-a)\tanh\beta\left[\sum_\mu m^\mu(\xi^\mu-a)+h_0\right]\right\rangle\!\right\rangle . \tag{5.6}$$

As T decreases, the correlation m with the pattern increases and the value of h_0 decreases. At $T=0$, h_0 vanishes and the state reduces to $S_i=\xi_i^\mu$ which automatically guarantees the constraint (5.5). Note that the constraint breaks the symmetry of the system under the global inversion $S_i\to-S_i$. In particular the reversed patterns $S_i=-\xi_i^\mu$ are no longer stable (in fact, they are outside the allowed phase space).

In addition to the suppression of the reversed states, h_0 suppresses considerably the appearance of spurious metastable mixture states, especially those which mix a small number of patterns. For instance, it can be easily checked that states which mix two or three patterns are not solutions of the equations.

2. The finite-α limit

The mean-field theory of Secs. IV A—IV C is easily extended to incorporate the constraint. The only modification is the inclusion of a field h_0 which guarantees that Eq. (5.1) is obeyed. Equations (4.7) turn into

$$m^{\nu}=\left\langle\!\left\langle(\xi^{\nu}-a)\tanh\beta\left[\sqrt{r\alpha}z+\sum_{\mu}^{s}m^{\mu}(\xi^{\mu}-a)+h_0\right]\right\rangle\!\right\rangle , \tag{5.7a}$$

$$q=\left\langle\!\left\langle\tanh^2\beta\left[\sqrt{r\alpha}z+\sum_{\mu}^{s}m^{\mu}(\xi^{\mu}-a)+h_0\right]\right\rangle\!\right\rangle , \tag{5.7b}$$

$$r=q(1-a^2)^2/[1-\beta(1-a^2)(1-q)]^2 \tag{5.7c}$$

with the additional equation

$$a=\left\langle\!\left\langle\tanh\beta\left[\sqrt{r\alpha}z+\sum_{\mu}^{s}m^{\mu}(\xi^{\mu}-a)+h_0\right]\right\rangle\!\right\rangle . \tag{5.8}$$

The retrieval state at $T=0$ is determined by

$$m=\tfrac{1}{2}(1-a)^2\left[\operatorname{erf}\left[\frac{m(1-a)+h_0}{\sqrt{2\alpha r}}\right]+\operatorname{erf}\left[\frac{m(1+a)-h_0}{\sqrt{2\alpha r}}\right]\right] . \tag{5.9}$$

$$C=\frac{1/\sqrt{r}-1}{1-a^2}$$

$$=(2\pi ar)^{-1/2}\left[(1+a)\exp\left[-\frac{[m(1-a)+h_0]^2}{2\alpha r}\right]+(1-a)\exp\left[-\frac{[m(1+a)-h_0]^2}{2\alpha r}\right]\right] , \tag{5.10}$$

$$a=\tfrac{1}{2}(1+a)\left[\operatorname{erf}\left[\frac{m(1-a)+h_0}{\sqrt{2\alpha r}}\right]-\tfrac{1}{2}(1-a)\operatorname{erf}\left[\frac{m(1+a)-h_0}{\sqrt{2\alpha r}}\right]\right] . \tag{5.11}$$

The largest value of α for which a solution with $m\neq 0$ exists is plotted in Fig. 4. Note the remarkable increase in $\alpha_c(a)$ compared to its value when the constraint (5.1) is lifted (Fig. 2). The origin of this increase is simple: In the unrestricted phase space the retrieval state is only a saddle point for α greater than $\alpha_c(a)$ of Fig. 2. However, the instabilities are in directions which violate the constraint (5.1). Thus, in the constrained phase space, the retrieval state is stable until α becomes greater than $\alpha_c(a)$ of Fig. 4. The retrieval quality at α_c is plotted, as a function of a, in Fig. 3.

It is also interesting that $\alpha_c(a)$ increases with a for almost all values a. It has a maximum value $\alpha_c=0.18$ at $a=0.925$. This increase fits nicely with the finite p limit,

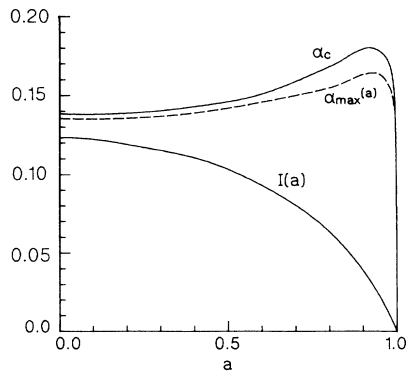

FIG. 4. Curves of the critical α and the α for which the information content is maximal vs bias. $I(a)$ is the actual information content at saturation.

where as a increases the number of spurious states decreases considerably, as discussed in Sec. V A 1 above. Of course, ultimately, as a approaches unity, the value of m which is restricted by definition to $m<1-a^2$ [see Eqs. (3.2) and (5.1)], decreases to zero, and $\alpha_c\to 0$. This, however, occurs only very close to $a=1$. In fact, studying Eqs. (5.9)–(5.11), in the $a\to 1$ limit, we find that $1/\alpha_c$ diverges only logarithmically at $a=1$. Denoting $\alpha_0=\alpha\,|\ln(1-a)|$ the asymptotic behavior near $a=1$ is

$$m\simeq A(1-a),\quad h_0\simeq b(1-a) ,$$
$$r\simeq 4(1-a)^2,\quad C\to 0 , \tag{5.12}$$

where $A=\sqrt{\alpha_0/2}$ and $b=\sqrt{2\alpha_0}$. In this regime

$$\alpha_c^*(a)=\frac{\alpha^*}{|\ln(1-a)|} \tag{5.13}$$

as $a\to 1$, with α^* of $O(1)$.

B. Soft constraint

The rigid constraint (5.1) may be relaxed by imposing a finite energy cost on fluctuations away from the optimal activity. The simplest way of imposing such a soft constraint is to add to the energy function H a term

$$\frac{g}{2N}\left[\sum_i S_i-Na\right]^2 , \tag{5.14}$$

where g is positive. Such a term represents a negative background $-g/N$ in the efficacy of all synapses, together with a constant magnetic field (neuronal threshold) which equals ag.

Modification of the mean-field equations due to the addition (5.14) is straightforward. The new equations for

the replica symmetric saddle point are the same as Eqs. (5.7). Equation (5.8) is replaced by

$$a-\frac{h_0}{g}=\left\langle\!\left\langle \tanh\beta\left[\sqrt{r\alpha}z+\sum_{\mu}^{s}m^{\mu}(\xi^{\mu}-a)+h_0\right]\right\rangle\!\right\rangle \tag{5.15}$$

which reduces to Eq. (5.8) in the limit $g\to\infty$. Likewise, in the $T\to 0$ limit the lhs of Eq. (5.11) has to be replaced by $a-h_0/g$.

Note that when $a=0$, the solution of Eq. (5.11) is $h_0=0$, and consequently the term (5.14) does not affect the system. The reason for this is that in this case ($a=0$) the stable states of the system have zero net magnetization, even at $g=0$. Hence, as long as the initial states do not have a finite magnetization, the system remains in the region of phase space which has zero magnetization, and will not be affected by a term of the type $g(\sum_i S_i)^2/N$.

Results for α_c as a function of the strength of the constraint, g, are shown in Fig. 5 for different values of a. The curves interpolate between the unconstrained results ($g=0$) of Sec. IV and the rigid constraint ($g=\infty$). The value of g at which α_0 reaches approximate saturation depends on a. At $a\simeq 0.5$, $\alpha_c(g)$ levels off around

$$g\simeq 10\ .$$

At low and high values of a, $\alpha_c(g)$ levels off already at lower values of g.

VI. THE CONTENT OF INFORMATION IN THE NETWORK

In previous studies it has been convenient to measure the capacity of the network by the maximum number of patterns $p_c=\alpha_c N$ that can be stored. It should be pointed out that p_c alone does not determine the maximum amount of *information* that can be stored. The difference between the two quantities is particularly pronounced in the biased network. In Sec. V it has been shown that by constraining the dynamics of the network, α_c can be made to increase with the bias a. On the other hand it is intuitively clear that the amount of information stored in patterns which are alike is much less than in uncorrelated patterns.

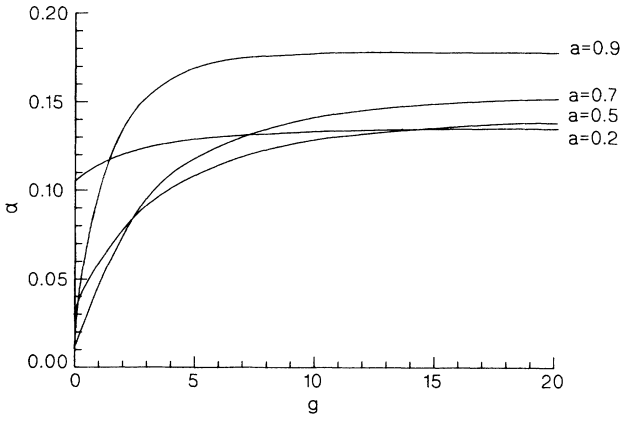

FIG. 5. Curves of the critical storage α vs strength of constraint, for several values of the bias. Note that for low and high bias the constraint saturates at relatively low g.

To quantify the information content of the network we will adopt here the rigid constraint (5.1) in which case both the stored and retrieved configurations lie in the same restricted phase space. The measure of information must take into account two factors: (1) the amount of the information stored in the embedded pattern and (2) the reduction of information due to errors in the retrieval of the patterns. The amount of information stored in each pattern depends on the size of its configuration space. In our case this information is just the entropy associated with the space ensemble of random patterns ξ_i^{μ} subject to the constraint that the total magnetization is Na. This yields

$$S(a)=-\tfrac{1}{2}(1+a)\ln[\tfrac{1}{2}(1+a)]-\tfrac{1}{2}(1-a)\ln[\tfrac{1}{2}(1-a)] \tag{6.1}$$

for the information per spin stored in *each pattern*. Suppose the retrieved pattern has an overlap per spin $\overline{m}=\langle\!\langle S\xi^{\mu}\rangle\!\rangle$ with the stored pattern ξ^{μ}. The *missing information* is the entropy associated with all possible configurations which have total magnetization Na and an overlap $N\overline{m}$ with a given pattern. This entropy (per spin) is

$$\begin{aligned}S(\overline{m},a)=&-\tfrac{1}{4}(1+2a+\overline{m})\ln(1+2a+\overline{m})\\&-\tfrac{1}{4}(1-2a+\overline{m})\ln(1-2a+\overline{m})\\&-\tfrac{1}{2}(1-\overline{m})\ln(1-\overline{m})+\tfrac{1}{2}(1+a)\ln(1+a)\\&+\tfrac{1}{2}(1-a)\ln(1-a)+\ln 2\ .\end{aligned} \tag{6.2}$$

Combining the results (6.1) and (6.2) we obtain for the total information provided by the network

$$\begin{aligned}I(\alpha,a)&=pN[S(a)-S(\overline{m},a)]/\ln 2\\&=\frac{\alpha N^2 S(a)}{\ln 2}\left[1-\frac{S(\overline{m},a)}{S(a)}\right]\ .\end{aligned} \tag{6.3}$$

Note that $I(\alpha,a)$ has been normalized so that the information content of one binary bit is unity. For a random retrieved state, $\overline{m}=a^2$ and

$$S(\overline{m},a)=S(a^2,a)=S(a)\ ,$$

yielding $I=0$ as expected.

Actually the value $\overline{m}$ is determined by the mean-field theory as a function of a and α as discussed in Sec. IV. The maximum capacity of information is achieved not by maximizing α but rather by maximizing $I(\alpha,a)$. This may lead to an optimal value of α which is slightly less than α_c. In fact, that is the case even in the unbiased network, for which Eq. (6.3) reduces to

$$I(\alpha,0)=\alpha N^2[\tfrac{1}{2}(1+\overline{m})\ln(1+\overline{m})+\tfrac{1}{2}(1-\overline{m})\ln(1-\overline{m})]\ , \tag{6.4}$$

using for $\overline{m}$ the result of the replica symmetric mean-field theory (MFT) leads to a maximum of Eq. (6.4) at

$\alpha_{\max}=0.134$ whereas $\alpha_c=0.138$. We have calculated $I(\alpha,a)$ for all $\alpha<\alpha_c$ and $0<a<1$, using the results of Sec. V for $\overline{m}(\alpha,a)$. The dashed line of Fig. 4 shows the value of α which maximizes I (for a given a). It is slightly below α_c for all a. The maximum value of I is also plotted in Fig. 4 as a function a. Even as α_c (or $\alpha_{\max}$) is increasing, the maximum information capacity is decreasing with a, mainly due to the factor $S(a)$ [see Eq. (6.3)].

VII. DISCUSSION

In this work we have modified Hopfield's model to incorporate storage and retrieval of patterns with fixed bias. The motivation to consider such patterns comes from biological as well as practical considerations of pattern recognition. Other models of neural networks have been constructed for the storage and retrieval of correlated patterns.[8–13] The main virtue of our model is that it retains the simplicity of the learning rules. In particular, the synaptic updatings due to the learning of new patterns are still local, except for their dependence on the global bias.

Our simple model is suited to handle the minimal correlations induced by their constant bias. It can be further generalized to patterns with a *distribution* of levels of activities,

$$\frac{1}{N}\sum_{i=1}^{N}\xi_i^\mu=a_\mu,\quad \mu=1,\ldots,p$$

where a_μ are distributed between a minimum value a_0 and a maximum value of a_1. The modified synaptic efficacies would be in this case

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}(\xi_i^\mu-a_\mu)(\xi_j^\mu-a_\mu)\ .$$

One possible dynamical constraint would be to restrict the motion in phase space to states whose total magnetization is bounded below and above by a_0 and a_1, respectively. It would be interesting to see whether this approach can be generalized to more structured sets of correlated patterns (e.g., hierarchical patterns).

ACKNOWLEDGMENTS

We are grateful to Roni Agranat for having impressed upon us the need for background-frequency asymmetry in pattern recognition and to Dr. E. Ve'adia and Dr. A. Arieli, of the Hebrew University Medical School, who demonstrated to us that the mean activity level is much lower than 50%. The work of D.J.A. and H.S. has been supported in part by a grant from the Fund for Basic Research of the Israel Academy of Science and Humanities.

APPENDIX

In this appendix we show that the ensemble of patterns which are random except for the global constraints

$$\sum_{i}^{N}\xi_i^\mu=Na,\quad \mu=1,\ldots,p \tag{A1}$$

is equivalent to the ensemble of independent random variables ξ_i^μ with local bias [Eq. (1.2)]. This equivalence holds only in the limit $N\to\infty$.

The ensemble with the global constraints can be represented by the following distribution:

$$P\{\xi\}=C\prod_{i,\mu}[\tfrac{1}{2}\delta(\xi_i^\mu-1)+\tfrac{1}{2}\delta(\xi_i^\mu+1)]\prod_{\mu}\tfrac{1}{2}\delta\left[\sum\xi_i^\mu-Na\right], \tag{A2}$$

where C is a normalization constant. Let $g\{\xi\}$ be an arbitrary function of the patterns. The average of $\exp(g\{\xi\})$ is given by

$$\langle\langle\exp(g\{\xi\})\rangle\rangle=C\int_{-i\infty}^{i\infty}\prod_{\mu}\frac{dz^\mu}{2\pi}\exp\left[-Na\sum_m z^\mu\right]\mathrm{Tr}_\xi\exp\left[\sum_\mu z^\mu\sum_i\xi_i^\mu+g\{\xi\}\right]$$
$$=C\int_{-i\infty}^{i\infty}\prod_{\mu}\frac{dz^\mu}{2\pi}\exp\left[-Na\sum_m z^\mu+N\sum_\mu\ln\cosh z^\mu+\Delta\right], \tag{A3}$$

where

$$\mathrm{Tr}_\xi(\cdots)\equiv 2^{-Np}\sum_{(\xi_i^\mu=\pm1)}(\cdots),$$
$$\Delta=\ln\mathrm{Tr}_\xi\exp\left[\sum_\mu z^\mu\sum_i\xi_i^\mu+g\{\xi\}\right]. \tag{A4}$$

The integrals over z^μ are calculated by their saddle point. If e^g is *not* of $O(e^{Np})$, the saddle-point equation is not affected by g and reduces simply to

$$z_0^\mu=\tanh^{-1}a=\tfrac{1}{2}\ln\left[\frac{1+a}{1-a}\right]. \tag{A5}$$

This implies that the average of e^g over ξ, Eq. (A3), is effectively an average with the probability distribution $P\{\xi\}=\prod_{i,\mu}P(\xi_i^\mu)$, where

$$P\{\xi\}=C'\prod_{i,\mu}[\tfrac{1}{2}\delta(\xi_i^\mu-1)\exp(z_0^\mu)+\tfrac{1}{2}\delta(\xi_i^\mu+1)\exp(-z_0^\mu)], \tag{A6}$$

which is equivalent to Eq. (1.2).

It remains to show that in our case e^g is *not* of $O(e^{Np})$. This is certainly correct with regard to the averaging over the finite number of macroscopically condensed patterns $\{\xi^\nu\}$, $\nu=1,\ldots,s$. The free energy is self-averaging with respect to them. Therefore, the averaging can be performed on *local* terms separately, i.e., on $f(\xi_i^1,\ldots,\xi_i^\nu)=O(1)$. The averaging over the rest $N\alpha-s$ patterns is performed on

$$e^g=\exp\left[\sqrt{\beta/N}\sum_{\mu,\rho}m_\rho^\mu\sum_i S_i^\rho\xi_i^\mu\right].$$

See Eq. (4.2). Since m_ρ^μ is of $O(1/\sqrt{N})$, g is only of $O(N)$ and not of $O(N^2\alpha)$ and Eq. (A6) is therefore valid.

[1]J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982); **81**, 3088 (1984); J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, Nature **304**, 158 (1983).

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

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

[4]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985); Ann. Phys. (N.Y.) (to be published); A. Crisanti, D. J. Amit, and H. Gutfreund, Europhys. Lett. **2**, 337 (1986).

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

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

[7]M. Abelles, in *Studies of Brain Function* (Springer-Verlag, New York, 1982); J. J. Hopfield, in *Modelling and Analysis in Biomedicine,* edited by C. Nicollini (World Scientific, New York, 1984).

[8]T. Kohonen, *Self Organization and Associative Memory* (Springer-Verlag, New York, 1984).

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

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

[11]M. Virasoro, in *Disordered Systems and Biological Organization,* edited by E. Bienenstock (Springer-Verlag, Berlin, 1986); N. Parga and M. Virasoro, J. Phys. (Paris) (to be published).

[12]Vik. S. Dotsenko, J. Phys. C **18**, L1017 (1985).

[13]J. Denker (unpublished); J. M. Lapedes and R. M. Farber (unpublished).

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

[15]G. Weisbuch and F. Fogelman-Soulie', J. Phys. (Paris) Lett. **46**, L-623 (1985); R. J. McEliece, E. C. Posner, E. R. Rodemich, and S. S. Venkatesh (unpublished).