Neural networks with dynamical thresholds

D. Horn and M. Usher
School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel
(Received 27 June 1988; revised manuscript received 29 December 1988)

We incorporate local threshold functions into the dynamics of the Hopfield model. These functions depend on the history of the individual spin (= neuron). They reach a maximal height if the spin remains constant. The resulting one-pattern model has ferromagnetic, paramagnetic, and periodic phases. This model is solved by a master equation and approximated by simplified systems of equations that are substantiated by numerical simulations. When several patterns are included as memories in the model, it exhibits transitions—as well as oscillations—between them. The latter can be excluded by known methods. By introducing threshold functions which affect only spins which remain positive, thus mimicking fatigue of the individual neurons, one can obtain open-ended movement in pattern space. Using couplings which form pointers from one pattern to another, our system leads to self-driven temporal sequences of patterns, resembling the process of associative thinking.

I. INTRODUCTION

Neural-network models for associative memory are dynamical systems with attractors that represent cognitive events such as memory contents. A well-known example is the Hopfield model[1] which is based on the Hamiltonian

$$H=-\tfrac{1}{2}\sum_{i,j}J_{ij}S_iS_j \tag{1.1}$$

whose dynamical degrees of freedom are N classical spin variables $S_i=\pm 1$ which interact with one another through the couplings J_{ij}. The latter are constructed by the factorized Hebbian rule

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu \ . \tag{1.2}$$

The binary vectors ξ_i^μ are the input patterns (memories) of the model which, under appropriate conditions,[2] form the fixed points into which the spin variables flow under the dynamical equations

$$S_i(t+1)=F_T\left[\sum_j{}'J_{ij}S_j(t)\right] , \tag{1.3}$$

where the prime designates the fact that $j\neq i$. Here F_T is the conventional statistical choice for temperature T

$$F_T(x)=\pm 1 \text{ with probability } (1+e^{\mp 2x/T})^{-1} \tag{1.4}$$

and the updating is performed either sequentially or in parallel. The dynamical flow process is identified with memory retrieval. Generally one should also view the couplings J_{ij} as dynamical degrees of freedom which evolve on a much longer time scale representing the procedure of learning. Many algorithms for their construction were suggested, deviating from the simple factorized form of Eq. (1.2). They may also be asymmetric, in which case an energy function cannot be defined but the equations of motion (1.3) are still applicable. Throughout this paper we will work with equations of motion rather than with a Hamiltonian.

Clearly this paradigm has to be enlarged in order to account for the richness of cognitive processes which evolve continuously and do not stop once a memory pattern is reached. One possibility for such a generalization is to add a set of pointers to the couplings:

$$J_{ij}\to J_{ij}+\lambda K_{ij}, \quad K_{ij}=\frac{1}{N}\sum_{\mu,\nu=1}^{p}d_{\mu\nu}\xi_i^\mu\xi_j^\nu \ . \tag{1.5}$$

In particular, if the couplings $d_{\mu\nu}$ have only one nonvanishing element for each ν and they are chosen to have a built-in time delay[3] such as in

$$S_i(t+1)=F_T\left[\sum_j J_{ij}S_j(t)+\lambda\sum_j K_{ij}S_j(t-\tau)\right] , \tag{1.6}$$

they can drive the system in a predetermined temporal sequence moving from one attractor to another. Using such structures one can account for counting processes[4] and recognition of temporal sequences.[5] Similar behavior can be obtained without time delay by relying on an external oscillating field[6] or internal noise of the system.[7]

We propose a different generalization in which pointers are not used as the cause for moving out of an attractor. Pointers may be present but the destabilizing effect is due to new degrees of freedom, threshold parameters. The importance of dynamic threshold parameters in regulating cognitive functions was pointed out by Braitenberg.[8] We will assume that the local threshold parameter θ_i changes with time in a fashion which depends on the history of the spin variable S_i at the same location. The equations of motion of our system are

$$S_i(t+1)=F_T\left[\sum_j J_{ij}S_j(t)-\theta_i(t)\right] . \tag{1.7}$$

As we will see, they generate motion in pattern space.

There are two choices of threshold functions which we consider. Both depend on an accumulated spin variable

defined iteratively by

$$R_i(t+1)=R_i(t)/c+S_i(t+1)\ . \tag{1.8}$$

This is an effective integration of the spin variable over time which saturates at the value $\pm c/(c-1)$ if the spin stays constant at ± 1. c will be chosen to be slightly larger than 1. If the threshold parameter is chosen proportional to R, it destabilizes the tendency of the system to stay in a fixed point and may lead to oscillatory behavior. This can be easily demonstrated by Monte Carlo runs in which we choose for simplicity the factorized form (1.2) for the J_{ij}. Using a single-pattern scheme we solve the problem in Sec. II in terms of a master equation and several simplified equations of motion for the overlap and the threshold function. Similar sets of equations are known to describe dilute models.[9] All compare well with one another and agree with simulation calculations exhibiting the existence of a periodic phase in this problem.

In Sec. III we discuss the dynamics of models which contain several memories. This is studied by employing a generalization of the simplified equations of motion developed in Sec. II and compared with the results of simulation calculations. The problem of two patterns exhibits continuous interchanges between them with or without sign flip of the overlap. This type of behavior is observed also in bigger sets of input patterns. Our simplified analytic models provide an understanding of this behavior.

We consider it interesting to investigate another threshold function which vanishes for all negative R. This version mimics the effects of fatigue in a system in which $S_i(t)=1$ is interpreted as neuron number i firing at time t. The more it fires the higher is the threshold for firing again at the next round. This version is studied in Sec. IV. Here we observe how, when the attractor is destabilized by the threshold function, the system flows into another temporary fixed point. The inversion of a pattern, which plays a dominant role in the periodic behavior discussed in Sec. II, may be avoided by using patterns with negative activity, i.e., negative average magnetization. We present simulation calculations of such systems and discuss the variation of the active memory duration (the relative time the system spends in the input patterns) with the number of input patterns. The phenomenon of motion from one pattern to another can be assisted by adding pointers of Eq. (1.5) into the couplings. We let the parameters $d_{\mu\nu}$ connect various patterns thus defining relations between memories and forming a variety of routes which the system may choose to wander in. We demonstrate the creation of families of patterns by this procedure. Some implications of our results are discussed in Sec. V.

II. PERIODIC BEHAVIOR IN THE SINGLE-PATTERN MODEL

We investigate the dynamical system defined by

$$S_i(t+1)=F_T(h_i(t)-bR_i(t))\ ,\quad h_i=\sum_\mu m^\mu \xi_i^\mu\ , \tag{2.1}$$

where m^μ designates the overlap of the spin configuration with pattern μ. h_i is the local field acting on spin i. This represents the modification of the factorized Hopfield model by a dynamical threshold proportional to the accumulated spin variable R defined in Eq. (1.8). We will concentrate first on the case of a single pattern because it lends itself to an algebraic investigation which unravels the important characteristics of the new phenomena. With a single input pattern we have a ferromagnetic model modified by the threshold term. Using the notation $m=m^1$ we obtain the equation

$$m(t+1)=\frac{1}{N}\sum_i \xi_i^1 \tanh\frac{h_i(t)-bR_i(t)}{T} \tag{2.2}$$

from (2.1) through multiplication by ξ^1 and statistical averaging. The factor ξ_i^1, being ± 1, can be moved into the argument of the tanh on the right-hand side. Let us denote the value of $\xi_i^1 R_i$ at the location i by r and assume that its value is described by a distribution function $P(r,t)$. This allows us to turn Eq. (2.2) into the integral equation

$$m(t+1)=\int dr\, P(r,t)\tanh\frac{m(t)-br}{T}\ . \tag{2.3}$$

The value of r at location i changes in one iteration to $r/c\pm 1$ with probabilities

$$\pi^{\pm}(m(t),r)=(1+e^{\mp 2[m(t)-br]/T})^{-1}\ . \tag{2.4}$$

Therefore the probability obeys the recursion relation

$$\begin{aligned}P(r,t+1)=&c\pi^+(m(t),cr-c)P(cr-c,t)\\&+c\pi^-(m(t),cr+c)P(cr+c,t)\ .\end{aligned} \tag{2.5}$$

The factor c guarantees the proper normalization. Thus we obtain a master equation describing the one-pattern case.

We will present below numerical solutions of the master equation as well as results of simulations of such systems. We find it useful to study also simplified approximate versions of the system of equations (2.3) and (2.5) in order to develop an intuitive understanding of the mechanism through which the dynamics evolve. Since the most important features of a probability distribution are its average ρ and standard deviation σ, let us replace $P(r,t)$ by an expression which has these values and is easy to manipulate,

$$P(r,t)=\tfrac{1}{2}\delta(r-\rho(t)-\sigma(t))+\tfrac{1}{2}\delta(r-\rho(t)+\sigma(t))\ . \tag{2.6}$$

Replacing (2.5) (which cannot be satisfied with this choice) by its first two moments, i.e., expectation values of r and r^2, we are led to a closed set of three equations

$$\begin{aligned}m(t+1)=&\tfrac{1}{2}\tanh\frac{m(t)-b\rho(t)-b\sigma(t)}{T}\\&+\tfrac{1}{2}\tanh\frac{m(t)+b\rho(t)+b\sigma(t)}{T}\ ,\end{aligned} \tag{2.7}$$

$$\rho(t+1)=\frac{\rho(t)}{c}+m(t+1)\ , \tag{2.8}$$

$$\sigma^2(t+1)=\frac{\sigma^2(t)}{c^2}+\frac{\sigma(t)}{c}\left[\tanh\frac{m(t)-b\rho(t)-b\sigma(t)}{T}+\tanh\frac{m(t)-b\rho(t)+b\sigma(t)}{T}\right]+1-m^2(t+1)\ . \tag{2.9}$$

The last equation follows from evaluating the expectation value of r^2 on both sides of (2.5) and using the previous equations for m and ρ. We will call this set of equations the $m-\rho-\sigma$ set.

Our system depends on three parameters: b, c, and T. The combination of b and c which is most relevant to our problem is

$$g=bc/(c-1)\ , \tag{2.10}$$

which represents the height which the threshold $\theta_i=bR_i$ can reach if the spin S_i stays constant in time. Varying these parameters we find different qualitative behavior in different regions. We observe the existence of three phases: a ferromagnetic phase characterized by finite-m values, a paramagnetic one in which $m\to 0$, and a periodic phase.

Let us start our discussion with the new feature of our model, the periodic phase. An example of the behavior of the variables m, ρ, and σ is shown in Fig. 1(a). This figure displays the solution to Eqs. (2.7)–(2.9) (the m-ρ-σ set) at the point $T=0.35$, $c=1.5$, $g=0.545$ and for the initial conditions $m=1$ and $\rho=\sigma=0$. To compare the oscillations of m with the results of simulations we display in Fig. 1(b) the behavior observed in a system of

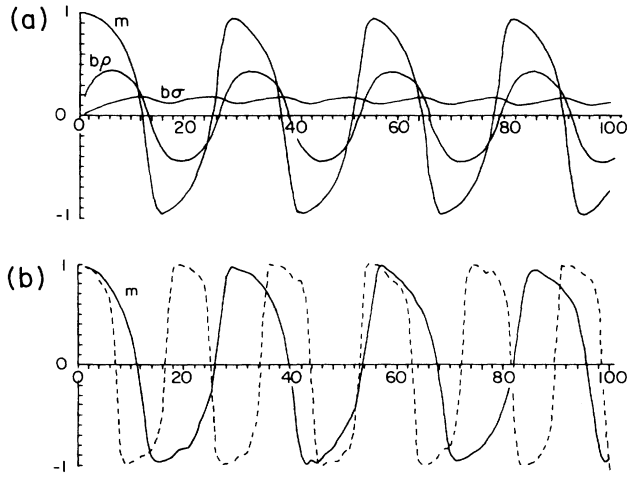

FIG. 1. (a) Graphical solutions of Eqs. (2.7)–(2.9) (the m-ρ-σ set) for $T=0.35$, $c=1.5$, $g=0.545$ and initial conditions $m=1$ and $\rho=\sigma=0$. The three curves represnt m, $b\rho$, and $b\sigma$ vs time (number of iterations). (b) Results for overlaps m at the same values of parameters from two simulations of a network of $N=400$ spins, one using sequential (dashed curve) and the other synchronous (solid curve) updating procedures.

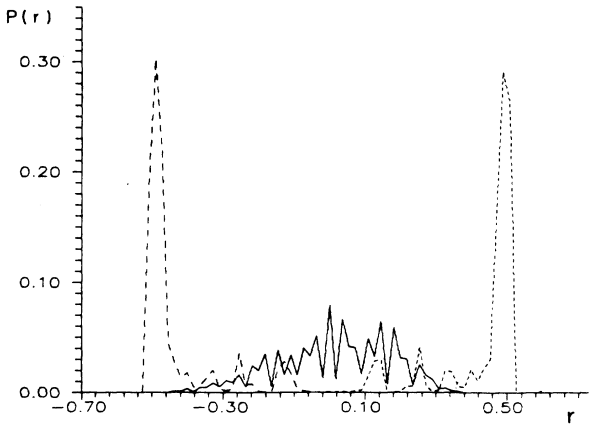

FIG. 2. The probability distribution $P(r)$ is plotted vs r for solutions to the master equations (2.3)–(2.5) at the same point in parameter space as in Fig. 1. The two dashed curves depict situations in which the average r is maximal or minimal. The solid curve corresponds to the case in which the average vanishes. These curves show that when the distribution is peaked at the external values of r it is also the narrowest, thus substantiating results of Fig. 1.

400 spins at the same value of the parameters. The solid curve describes the results of a synchronous updating mechanism which are practically identical with the solution of the m-ρ-σ set. The dashed curve is the outcome of a sequential (asynchronous) updating procedure. It is to be expected[9] that the synchronous and asynchronous systems show different behavior: The first corresponds indeed to an iterative set of equations such as the m-ρ-σ model, while the latter should be described by differential equations. We observe empirically that the main difference is the time scale of the process; the other features are quite similar.

While m oscillates between the values 1 and -1 $b\rho$ varies between ± 0.45. There is a small phase shift between m and ρ corresponding to the fact that ρ is being built up by the values of m. The distribution develops a width σ which oscillates together with m and ρ around some finite constant value which is significantly smaller than the extrema of ρ. The width grows when the state undergoes a transition (m crosses zero), and it decreases when m reaches its extrema. To test our approximation we solve numerically the master equation, i.e., (2.3)–(2.5), using the same parameters. We obtain again a periodic motion of m which drives—and is in turn driven by—a probability distribution which is shown in Fig. 2. Shown here are three distributions depicting the situation when the average r is maximal, crosses zero, and becomes minimal. We see again that at the extrema the width is narrow and it widens in between. This example shows that the simplified equations (2.7)–(2.9) describe correctly the general features of the master equation.

The periodic motion observed in Figs. 1 and 2 is independent of the initial conditions of the system. Even if we start with $m=0$ and a random distribution $P(r)$, we observe a quick flow into the dynamical attractor which is the limit cycle which we described above.

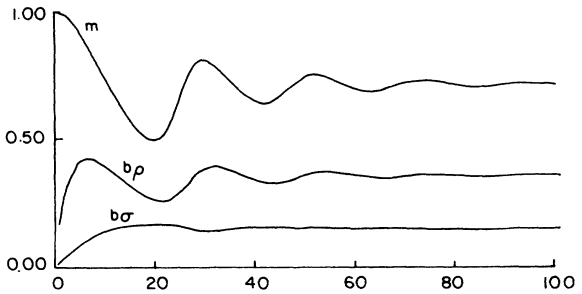

FIG. 3. Graphical solutions of Eqs. (2.7)–(2.9) (the m-ρ-σ) for $T=0.35$, $c=1.5$, $g=0.5$ and initial conditions $m=1$ and $\rho=\sigma=0$. The three curves of m, $b\rho$, and $b\sigma$ vs time display damped oscillations which settle into a fixed point characteristic of the ferromagnetic phase.

Reducing g slightly to $g=0.5$ we move into the ferromagnetic region, as seen in Fig. 3. Now we observe that the solutions of the m-ρ-σ set undergo damped oscillations after which they settle into a fixed point. The phase boundary at $T=0.35$ and $c=1.5$ has therefore to lie in between $g=0.545$ and $g=0.5$. On one side we find a limit cycle, i.e., a periodic phase, and on the other side a nontrivial fixed point, i.e., a ferromagnetic phase. To reach the third phase, characterized by $m=0$, one has to increase the temperature T. The general structure of the phase space is shown in Fig. 4. Using $c=1.5$ we have searched for the boundaries in three different ways: by solving the m-ρ-σ set, by solving the master equation (2.3)–(2.5), and by numerical simulations in a network of 2000 spins. All agree pretty well with one another.

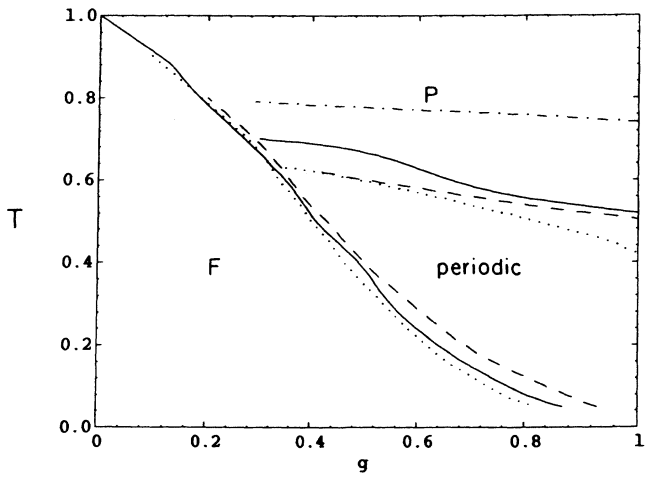

FIG. 4. The phase space of the single-pattern model has three different phases, ferromagnetic, paramagnetic, and periodic. The solid curve represents the results of simulations in a network of $N=2000$ spins. The dashed curves correspond to results of the master equation and the dotted curves are the results of the m-ρ-σ set. All these were calculated using $c=1.5$. For other c values we find that the border line between the ferromagnetic and periodic phases shows the same universal dependence on $g=bc/(c-1)$, whereas the transition between the periodic to paramagnetic phases changes with c. The dash-dot curve is the corresponding result of the master equation for $c=1.2$.

Varying c we find that the only qualitative change is that the phase boundary between the periodic phase and the paramagnetic one moves to higher T values. The dot-dash line represents the results of the master equation for $c=1.2$.

A further simplification of the model equations suggests itself by the fact that σ has values smaller than the extrema of ρ. This raises the possibility of using $\sigma=0$, i.e., assuming the probability distribution to be extremely peaked. We will mimic the effect of the distribution by introducing a small random fluctuation δ into the equation otherwise obtained from the first moment of (2.5)

$$m(t+1)=\tanh\frac{m(t)-b\rho(t)}{T}+\delta\ . \tag{2.11}$$

Together with

$$\rho(t+1)=\rho(t)/c+m(t+1) \tag{2.12}$$

it forms a simplified set of equations which we call the m-ρ set. Apart from the stochastic element δ it could be derived from (2.2) with the simplifying assumption $R_i=\rho_i^1$, i.e., complete dominance of the one pattern in the threshold factor.

This set will be generalized in Sec. III for the purpose of investigating models with many patterns. Let us show here that it displays the correct characteristics of our model. We begin with $\delta=0$ and very low temperatures, where the tanh turns into a sgn function. Starting from $m=1$, $\rho=0$ the parameter ρ increases as

$$\rho(t)=[c/(c-1)](1-c^{-t})\ . \tag{2.13}$$

The threshold function is important if b is of order $(c-1)/c$ or more. This causes m to flip its value. It leads to a periodic motion of the type shown in Fig. 5. This simple example shows the relevance of the combination $g=bc/(c-1)$ in our problem.

Staying with the same values of b and c while increasing T we find that the oscillations get smoother until they reach an almost sinusoidal form, as shown in Fig. 6. At this point in parameter space we find that $|m-b\rho|<0.2$

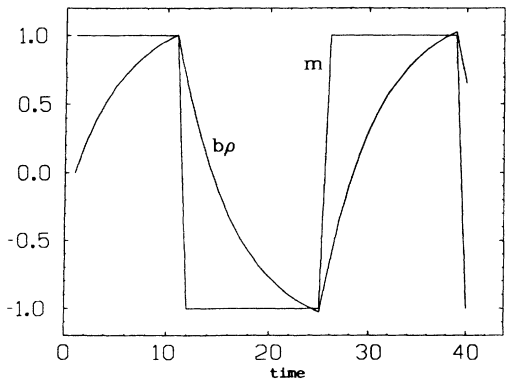

FIG. 5. Graphical solutions of Eq. (2.11) and (2.12) (the m-ρ set) for $T=0$, $c=1.2$, $b=0.2$. The amplitudes of the overlap m and the threshold function $b\rho$ are plotted vs the number of iterations (time). The ρ curve has the behavior of Eq. (2.13).

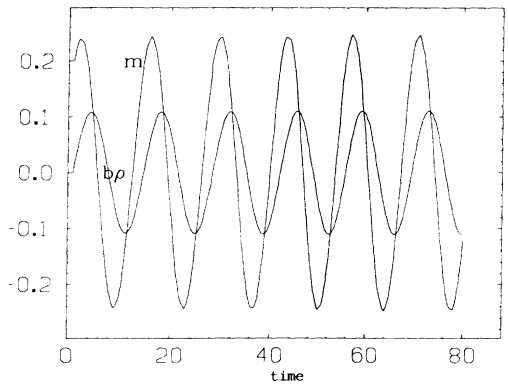

FIG. 6. The periodic behavior of Fig. 5 turns into almost pure sinusoidal functions for $T=0.82$. This plot uses the same values of b and c as in Fig. 5. The frequency is given by Eq. (2.14). This T value corresponds to the transition point between the periodic and paramagnetic phases for these values b and c in the m-ρ set of equations. For higher-T values one observes damped oscillations converging to zero.

hence we can replace the tanh of Eq. (2.11) by its linear approximation in order to estimate the period of oscillation. Solving the resulting set of linear recursion equations we obtain the following estimate for the frequency:

$$\tan^2\omega=\frac{4T}{c(1+T/c-b)^2}-1\ , \tag{2.14}$$

which fits very well the oscillations in Fig. 6.

Although Eq. (2.11) is quite a crude approximation to the correct physical behavior, it does represent a system with the same three phases, albeit in slightly shifted locations. The introduction of a small finite stochastic fluctuation δ can move the boundaries. Its main importance is in selecting the stable fixed points only. With $\delta=0$ the point $m=\rho=0$ is always a fixed point. In the ferromagnetic or periodic phases it is unstable. The fluctuations of δ will drive the system out of the unstable fixed point into the attractor for all initial conditions.

III. DYNAMICS OF SEVERAL PATTERNS

The dynamics of the system containing several patterns can be analyzed under the simplifying assumptions which we used at the end of Sec. II, the m-ρ set of equations. We are interested in solving the set of equations

$$m^\mu(t+1)=\frac{1}{N}\sum_i \xi_i^\mu \tanh\frac{h_i(t)-bR_i(t)}{T},\quad \mu=1,\ldots,p\ , \tag{3.1}$$

where the local field is given by

$$h_i=\sum_{\mu,\nu}(m^\mu+\lambda d_{\mu\nu}m^\nu)\xi_i^\mu \tag{3.2}$$

and the threshold function obeys the simplified relation

$$R_i=\sum_\mu \rho^\mu \xi_i^\mu\ . \tag{3.3}$$

To all this we will add stochastic noise as in (2.11). Note that in the local field we have included now the possible existence of pointers as defined in (1.5). We treat the case of a finite number of input patterns p in the thermodynamic limit $N\to\infty$ and neglect the overlaps between the patterns. Under this assumption the p patterns form an orthogonal set of vectors. This allows us to turn (3.1) into the set of equations

$$m^\mu=\frac{1}{2^{p-1}}\sum_{\eta^\mu=\pm1}{}'\tanh\frac{\sum_\nu(m^\nu-b\rho^\nu)\eta^\nu+\sum_{\sigma\nu}m^\sigma\lambda d_{\nu\sigma}\eta^\nu}{T}+\delta^\mu\ , \tag{3.4}$$

where the prime designates the fact that when $\nu=\mu$ only the factor $\eta^\nu=+1$ should be used. This set of equations together with

$$\rho^\mu(t+1)=\frac{\rho^\mu(t)}{c}+m^\mu(t+1) \tag{3.5}$$

forms the generalization of the m-ρ set which we are going to study in this section. Such equations can be regarded as mean-field approximations[5] which are valid in the limit of a dilute model.[9] These model equations serve then as a simplified system whose characteristics help us understand the behavior of the fully connected neural network. We will see that the approximation of this m-ρ set is good enough to describe correctly the qualitative features of the data obtained from numerical simulations.

The case $p=2$ is particularly simple. Using symmetric pointers, $d_{12}=d_{21}=1$, we find that Eqs. (3.4) turn into independent equations for the combinations $m^\pm=m^1\pm m^2$:

$$m^\pm(t+1)=\tanh\frac{m^\pm(t)(1\pm\lambda)-b\rho^\pm(t)}{T}+\delta^\pm\ . \tag{3.6}$$

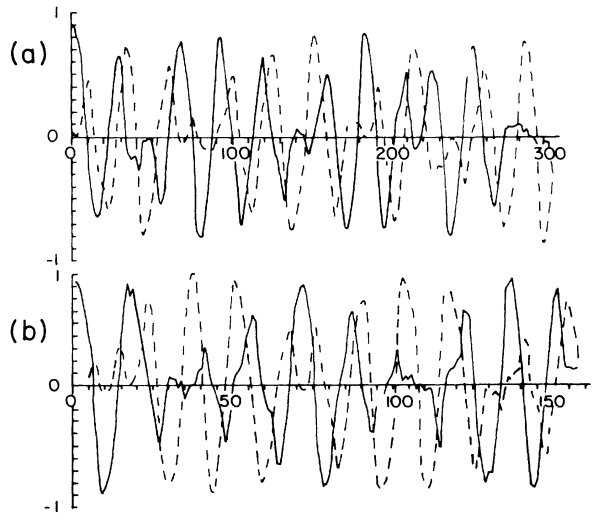

FIG. 7. (a) m^1 and m^2 vs time as derived from solutions to the set of equations (3.5) and (3.6) at the point $T=0.6$, $g=0.4$, $c=1.5$, $\lambda=0.05$. The stochastic noise used in these equations was $\delta^{1,2}=0.04$ multiplied by a random number between $+1$ and -1. (b) Results of a sequential simulation calculation performed on a system of $N=400$ at the same point in parameter space. We have used two orthogonal patterns in the construction of this model.

A solution to these equations is presented in Fig. 7 together with the results of a simulation with $N=400$ spins. We chose parameters lying inside the periodic phase: $T=0.6$, $g=0.4$, $c=1.5$, and $\lambda=0.05$. For the m-ρ set we have used stochastic noise of $\delta^{1,2}=0.04\times$(a random number between $+1$ and -1). Both figures show the existence of beats due to the two different frequencies caused by the $(1\pm\lambda)$ factor in Eq. (3.6). We observe clear alternations of strong and weak oscillations of the two patterns as in a system of coupled oscillators. The simulation displayed in the lower half of this figure was performed with sequential (asynchronous) updating, using two orthogonal patterns. It displays the same characteristics as the m-ρ set but different frequencies. This difference is due to the asynchronous updating procedure, as explained in the discussion of Fig. 1. Replacing it with a synchronous one we obtain similar frequencies to the ones seen in the m-ρ set of equations.

Next we move to the neighborhood of the border line between the periodic and ferromagnetic phases in the one-pattern phase space. Figure 8 shows the results for the values $T=0.6$, $g=0.28$, $c=1.5$ and $\lambda=0.1$. Now we observe completely different characteristics: continuous interchange between the two patterns with the overlap keeping the same sign. This can be easily explained by the m-ρ model equations. We have already observed that the $+$ and $-$ combinations evolve independently. At the border line they will have different characteristics. The equation for the $+$ combination will be in the ferromagnetic phase, while the $-$ combination will find itself in the periodic phase because of the different weights of $1\pm\lambda$ appearing in them. The result is the observed $m^1\leftrightarrow m^2$ cycle. By proper tuning of g we can obtain such results even for very small λ.

We observed such cyclic behavior even for $\lambda=0$. In general, however, the $\lambda=0$ case shows random motion between the two patterns due to random phase shifts between $m^{\pm}$.

Figures 7 and 8 present characteristic attractors of our system. In general, we see interchanges between the two

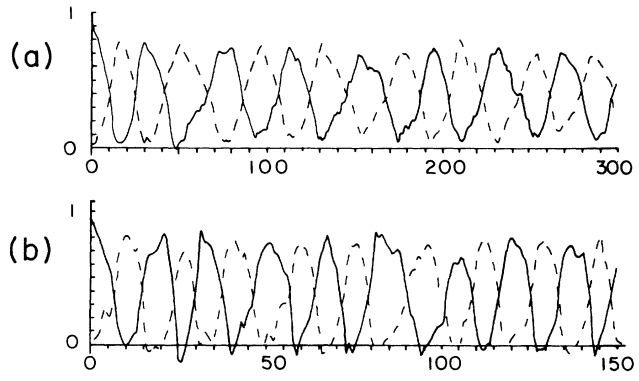

FIG. 8. (a) Same as Fig. 7 but at the point $T=0.6$, $g=0.28$, $c=1.5$, $\lambda=0.1$. The T, g, c point lies on the border line between the ferromagnetic and periodic phases of the one-m-ρ set. This is the reason for the observed behavior. (b) Same characteristics are obtained from the simulation calculation. The difference in time scale was already noted in Fig. 1.

patterns. Even if we start with the initial condition $m^1=m^2=\frac{1}{2}$ the system moves quickly to a situation where one pattern dominates the other. In the m-ρ set this comes about by the effect of the noise which destabilizes the $m^+=1$, $m^-=0$ solution. This is a reflection of the fact that in the Hopfield model this is an unstable solution.

For $p>2$ the system (3.4) does not decompose into independent equations. Still we can find solutions in which two patterns oscillate as in the $p=2$ case while the other patterns remain inactive. As an illustrative example let us look at the case $p=3$ and discuss the stability of a solution with $m^3=\rho^3=0$. We choose the set of equations (3.4) with no pointers, i.e., $\lambda=0$. The equation for m^3 is

$$m^3=\tfrac{1}{4}\left[\tanh\frac{m^3-b\rho^3+a^+}{T}+\tanh\frac{m^3-b\rho^3-a^+}{T}+\tanh\frac{m^3-b\rho^3+a^-}{T}+\tanh\frac{m^3-b\rho^3-a^-}{T}\right]+\delta^3\,, \tag{3.7}$$

where

$$a^{\pm}=m^1\pm m^2-(b\rho^1\pm b\rho^2)\,.$$

To find if $m^3=\rho^3=0$ is a stable fixed point we take the derivative with respect to m^3 and require that the slope of the right-hand side be smaller than 1. This leads to the condition

$$T^{-1}<\tfrac{1}{2}[\cosh^2(a^+/T)+\cosh^2(a^1/T)]\,. \tag{3.8}$$

As long as either a^+ or a^- have an absolute value larger than 0.59 this equation is satisfied for all T. If both are equal it suffices for them to reach 0.45 to have a stable solution for all T. This means that there exist domains in parameter space where such stability will be found.

In practice we indeed observe such behavior. It is quite common for a pair of patterns to dominate the scene, interchanging roles as in Fig. 7 if the parameters are in the periodic phase. After a while, one of them may pick a different pattern to serve as its partner while all others remain relatively inactive. We find this type of behavior to be quite common in simulations which we have run with 3, 4, and more patterns.

IV. MOTION IN PATTERN SPACE AND LOCAL FATIGUE

In this section we limit ourselves to motion in pattern space in which only positive overlaps with the patterns are obtained. When a large overlap with an input pattern occurs we will refer to it as being activated. We are interested in the phenomenon of transitions between the activated patterns. We wish to cause the motion in pattern space by a threshold function which may be interpreted as a local-fatigue effect

$$\theta_i=b(R_i+|R_i|)/2 \tag{4.1}$$

as explained in the Introduction. The physiological term

which we attribute to this function reflects the fact that this type of threshold becomes active only for spins which stay positive too long, i.e., neurons which keep firing for a long time. The function (4.2) guarantees that such a spin variable will flip its sign. It will continue to affect it after the flip until R_i vanishes. This inertial effect causes sign reversal of pattern overlaps. To avoid the latter one can utilize a model in which the inverse patterns are not attractors. We will make use of a model[10] which accommodates memories with negative activity, i.e., negative average magnetization $a=\langle \xi^\mu \rangle$. This model is defined by the Hamiltonian

$$H=-\tfrac{1}{2}\sum_{i,j} J_{ij}S_iS_j+\frac{G}{2N}\left[\sum_i S_i-Na\right]^2, \tag{4.2}$$

where the couplings are

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}(\xi_i^\mu-a)(\xi_j^\mu-a) \tag{4.3}$$

and G is a coefficient enforcing the constraint. Adding to the dynamics of this Hamiltonian our threshold function (4.1) we obtain transitions without sign reversals. The sign reversal is avoided because the inertial effect mentioned above operates on a small number of spins and because the inverse of a pattern is no longer an attractor. Staying within the periodic phase we get the wanted feature of self-driven temporal sequence of patterns.

Our results are displayed in Figs. 9 and 10. Figure 9 is the result of simulation calculations for $N=500$ spins and four random memories with average activity $a=-0.6$. We used the parameters $c=1.2$, $T=0.05$, $g=1.2$, and $G=2$. The overlap of the spin configuration of our system with four different memories is shown in four frames, with the fifth displaying the average fatigue factor of all spins $[(1/N)\sum_i \theta_i]$ on the same time scale. The panorama of overlaps includes plateaus at 1 as well as intermediary peaks of smaller magnitudes.

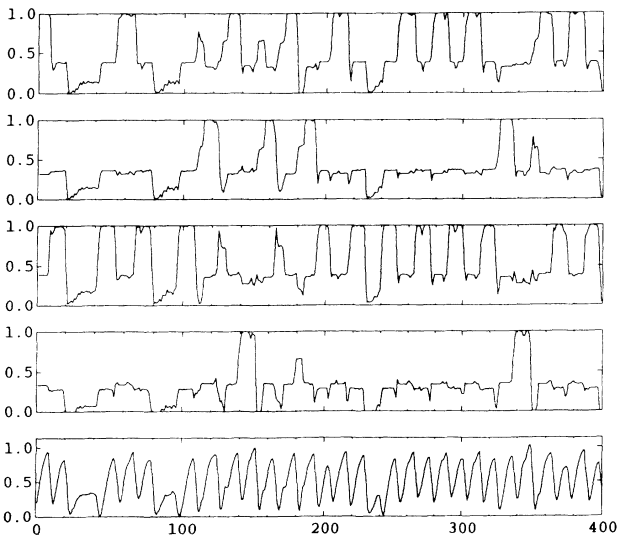

FIG. 9. Results of a simulation calculation of a system of $N=500$ spins and four random memories with average activity $a=-0.6$. We used dynamics based on the Hamiltonian (4.2), (4.3) and the fatigue factor (4.1) with sequential updating. The parameters are $c=1.2$, $T=0.05$, $g=1.2$, $G=2$. The first four frames display the overlap of the spin configuration of our system with the four different memories as a function of time (number of iterations). The fifth frame displays the average fatigue factor of all spins $[(1/N)\sum_i \theta_i]$ on the same time scale.

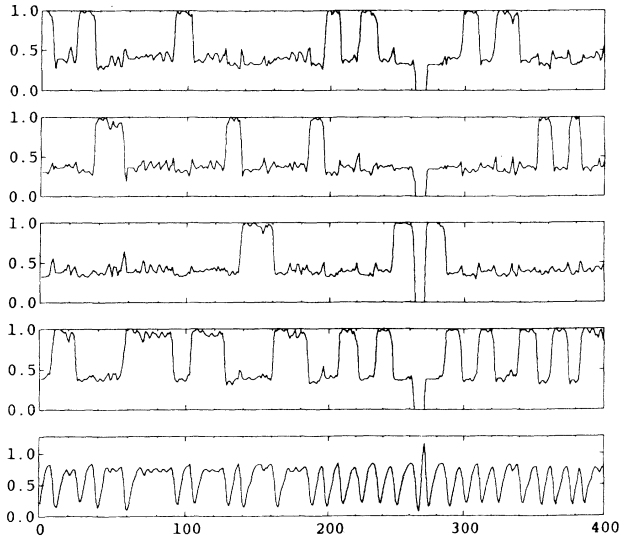

FIG. 10. Same display as Fig. 9 for a system based on the asymmetric dynamics of (4.4). The parameters are the same as those of Fig. 9 but for $g=0.9$.

A slightly different picture is shown in Fig. 10. This is based on dynamics of the same type in which we use for the couplings an asymmetric structure

$$J_{ij}=\frac{1}{N}\sum_{\mu=1}^{p}\xi_i^\mu(\xi_j^\mu-a). \tag{4.4}$$

Since we solve dynamical equations of the type (1.3) rather than a Hamiltonian problem we can handle the asymmetric choice as well. The results shown in Fig. 10 were obtained for the parameters $c=1.2$, $T=0.05$, $g=0.9$, and $G=2$. The activity was also chosen as $a=-0.6$. The main difference between the two figures is that the latter does not have intermediary peaks. The spin configuration of the system tends to stay for some time in one pattern and after being destabilized it moves to another pattern in a random fashion. We see here the chain $1\to4\to1\to2\to4\to1\to4\to2\to3$ and so on, continuing indefinitely. Within the time period of 400 updatings of the system we observe only one occurrence of negative overlaps, in which a major fraction of all spins turned positive. The fatigue factor rose quickly to revert this trend.

In Fig. 10 we see that the system spends most of its time in one of the original patterns. We will refer to this fact as an active memory. For quantitative purposes let us call the memory active when one of the overlaps is larger than 0.9. The fraction of the time in which the memory is active, which will be called the active memory duration, is a dynamical order parameter of our system. Clearly it varies with all the parameters we have specified before. For example, increasing T decreases the active

memory duration. An interesting question is how does the number of input patterns affect it. It is well known that memory models have a critical value of p after which the patterns lose their basins of attraction. In the Hopfield model it is[2] $p_{cr}=0.14N$. Clearly our model cannot have temporary fixed points beyond the critical point of the version without dynamic thresholds. It seems logical to expect the active memory duration to decrease as p increases, vanishing as the basins of attraction shrink to zero.

The active memory duration must be closely related to the question of coverage capacity[11] in the Hopfield model: We expect that when the basins of attraction of the memories cover the configuration space in a model without threshold parameters then the introduction of dynamical thresholds will lead to motion from one attractor to another without being lost for long periods in regions which are not associated with any memory. The active memory duration must be proportional to the region of configuration space covered by the basins of attraction. One should then expect the active memory duration to be much higher in multiconnected models[11] which are characterized by coverage capacities larger than the bilinear models we use in this paper. Moreover, it is worthwhile noting that odd interactions avoid also the problem of sign reversal of the overlaps.

The behavior of the active memory duration in a model governed by the fatigue factor (4.1) and the asymmetric couplings (4.4) is displayed in Fig. 11. We work here at the same point in parameter space as in Fig. 10. We use systems of 500 and 1000 spins and measure the fraction of the active memeory duration on 300 random trial runs for each p. In spite of the statistical fluctuations one can observe a clear trend. We use linear fits to the curves to determine the critical points where the memory becomes inactive. In this range of N this point may be fitted by $p_{cr}=33+0.054N$.

The models discussed so far did not contain any pointers. The transitions from one pattern to another occurred in a stochastic fashion. By introducing an appropriate set of pointers into the couplings one can obtain predetermined temporal sequences. The pointers do not cause the exit from an attractor; this is due to the threshold functions. The pointers help, however, to direct the movement in pattern space, i.e., they assist the entry into certain attractors. By allowing more than one pointer $d_{\mu\nu}$ for the pattern ν we reach a situation in which there exists competition between the various possible transitions. The pointers may be thought of as defining associations between patterns, which by themselves can be built in a learning process which develops on a long time scale. Thus the pointers define families of patterns which are related dynamically. This dynamical relation does not depend at all on the Hamming distance between the related patterns. Families are naturally defined as groups of attractors with high connectivity between them. When one runs a system like this, one observes that the spin configuration travels in pattern space between memories which belong to the same family and moves from one family to the next according to the various pathways which are given by the net of pointers or in some stochastic fashion.

As an example let us look at a system with eight patterns which we divide into two disconnected families with the following pointers:

$$\begin{aligned}&(1,2,8):\ 8\leftrightarrow 2,\ 8\rightarrow 1,\ 1\leftrightarrow 2\ ,\\ &(3,4,5,6,7):\ 3\leftrightarrow 5,\ 7\rightarrow 3,\ 7\rightarrow 5,\ 3\leftrightarrow 4\ ,\\ &4\rightarrow 5,\ 4\rightarrow 6,\ 6\rightarrow 3,\ 6\rightarrow 5\ .\end{aligned}\tag{4.5}$$

All pointers were given equal strength $\lambda=0.09$. Other-

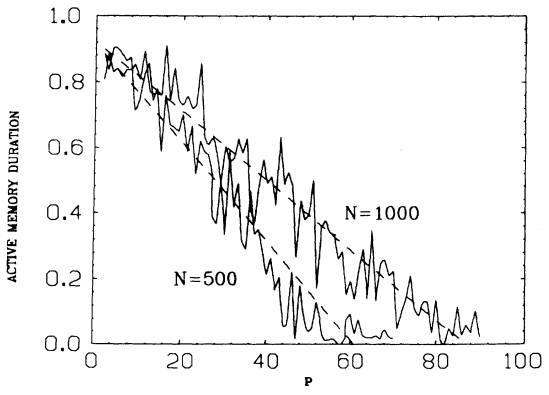

FIG. 11. The active memory duration (i.e, the relative amount of time the system has an overlap greater than 0.9 with one of the input patterns) is measured as a function of the number of memories p. We have used the same point of phase space as in Fig. 9 and performed 300 random trial runs for every value of p. We show results for $N=500$ and 1000 spins and linear fits which we use to estimate the critical point.

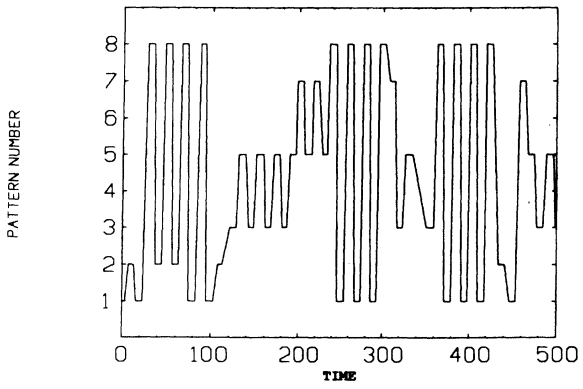

FIG. 12. A system of $N=2000$ spins and eight memories with the pointers defined in Eq. (4.5) was run at the point $c=g=1.2$, $T=0.2$, $G=2$ using the dynamics of Eqs. (4.1), (4.2), and (4.4). The eight input patterns were generated randomly with the constraint that the average activity should be $a=-0.5$. The actual activities were -0.485, -0.472, -0.494, -0.486, -0.502, -0.465, -0.506, -0.533 for the patterns 1–8. The figure shows the activation of the patterns (numbered along the y axis) as a function of the number of iterations. When the spin configuration does not have an overlap greater than 0.9 with one of the memories, it is regarded as forming a transition between two patterns. It spends most of the time staying inside the families of (4.5).

wise we have used the same model discussed above with $N=2000$ and $c=g=1.2$, $T=0.2$, $G=2$, and average activity $a=-0.5$. We display the results of 500 simulation steps in Fig. 12. In this example we observe 25 transitions within the first family, 18 within the second, and 5 transitions between the families. Clearly the latter do not stem from predetermined input, but also some of the former were not initiated by the pointers, e.g., the $1\rightarrow 8$ transitions which did not exist in the input. We suspect that in this particular example this is due to the fact that pattern 8 happens to have the lowest activity. Using the fatigue factor we find that the spin system is mostly negative when it leaves the attractor, hence patterns with lower activity have a higher chance to become the next attractor. In any case, even if the system moves from pattern 1 to 8 because of the big number of negative spins in the latter, it moves back because of the original pointer. The net result is the observed motion within families. This type of motion in pattern space has similarity to an associative thinking process.

V. SUMMARY AND DISCUSSION

The Hopfield model and its generalized dynamical systems are based on simplifications of the biological systems. They serve to demonstrate that in complexity of the networks lies the secret of their capacity to store large amounts of information. One of the simplifying assumptions made in these models is to neglect threshold dynamics. We have demonstrated that by including simple dynamic threshold behavior one can create nontrivial motion in the space of memory patterns.

Using a threshold function which depends on the history of the local spin variable one introduces a non-Markovian element of feedback into the dynamics of the system. This leads to the possibility of obtaining periodic behavior which is an interesting effect by itself. Clearly one needs neural networks which have such possibilities to serve as frequency filters for sensory functions which depend on temporal sequences of signals. Using the local threshold variables one could envisage an adaptive system which develops the necessary ability through the regulation of the relevant parameters in the threshold function as well as in coupling space. We have seen in Sec. II examples of periodic motion which this system can create. In Sec. III we saw that a system of two patterns can behave like coupled oscillators. All these elements can be useful for the purpose of analyzing signals or providing natural clocks.

In our models we encounter different parameters which could have an adaptive character i.e., change dynamically in a fashion which is connected to the development of the system under the influence of external interactions. Thus we could envisage a corrective procedure for obtaining the best choice of couplings to increase the capacity of the model, including some selective procedure for pointers by demanding dynamical correlations between patterns. We have concentrated on understanding the effect of the threshold functions only. For this reason we have employed in our simulation calculations the simplest algorithm for the couplings, Eq. (1.2), and have not opted for maximal capacity or efficiency.

In Sec. II we have solved the one-pattern model and shown the existence of the periodic phase. Moreover, using simplified model equations, we were able to give correct characterizations for some of the important features of the model such as estimating the frequency in Eq. (2.14). The simplified one-pattern approximation served as a guide in our discussion of multiple pattern structures in Sec. III. Thus we were able to explain the phenomenon of interchange among patterns, the appearance of beats, and the existence of regions dominated by positive overlaps only.

The threshold functions which we chose to deal with depended on the accumulated spin variable R in a specific linear fashion. Clearly the choice of this particular variable as well as the functional dependence on it are quite arbitrary. We have chosen them because they seem to be the simplest and most natural ones. In spite of being so simple they led to interesting and complicated dynamics.

We have seen in Sec. IV that using the threshold function (4.1), which we associated with fatigue of the single neuron, one obtains self-driven temporal sequences of patterns under appropriate conditions. One is then tempted to make the statement that fatigue drives the thinking process. One should, however, keep in mind that we have discussed a simplified model whose biological or neurological relevance has still to be demonstrated. In the meantime one should accept it for what it is, a mathematical neural-network structure with new degrees of freedom. Its importance, in our opinion, lies in the fact that these local degrees of freedom drive global changes of the system; their destabilizing effect causes the complex system to move from one attractor to another thus creating a nontrivial motion in pattern space.

ACKNOWLEDGMENTS

We are grateful to C. Lustig for suggesting that changes in thresholds can lead to temporal sequences of patterns. We wish to thank Y. Dothan for helpful discussions and G. Toulouse for pointing out the book by Braitenberg and its relevance to our work.

[1]J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2254 (1982); **81**, 3088 (1984).

[2]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); Phys. Rev. Lett. **55**, 1530 (1985).

[3]H. Sompolinsky and I. Kanter, Phys. Rev. Lett. **57**, 2861 (1986). D. Kleinfeld, Proc. Natl. Acad. Sci. U.S.A. **83**, 9469 (1986).

[4]D. J. Amit, Proc. Natl. Acad. Sci. U.S.A. **85**, 2141 (1988).

[5]H. Gutfreund and M. Mezard, Phys. Rev. Lett. **61**, 235 (1988).

[6]D. Lehman (unpublished).

[7]J. Buhmann and K. Schulten, Europhys. Lett. **4**, 1205 (1987).

[8]V. Braitenberg, *Vehicles* (MIT Press, Cambridge, Massachusetts, 1984).

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

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

[11]D. Horn and M. Usher, J. Phys. (Paris) **49**, 389 (1988).