Journal of Statistical Physics, Vol. 55, Nos. 5/6, 1989

Information Processing in Three-State Neural Networks

C. Meunier,[1] D. Hansel,[1,2] and A. Verga[1]

Received Novmeber 14, 1988

We introduce networks of three-state neurons, where the additional state embodies the absence of information. Their dynamical behavior is studied from the standpoint of information processing. These networks display strong pattern completion capabilities. Moreover, inference naturally occurs between coherent patterns.

KEY WORDS: Neural networks; three-state neurons; information processing; pattern completion; inferential properties.

1. INTRODUCTION

Connectionist models[1,2] embody the ideas that emergent properties can arise in complex systems from the interaction of many simple elementary units and that the distribution of computational power among the units entails a great robustness of the whole system. Their popularity in artificial intelligence[3] decreased at the end of the 1960s when Minsky and Pappert[4] brought to light the limitations of elementary models; however, a number of important papers[5-7] on connectionist models still appeared in the 1970s. The publication of Hopfield's seminal paper[8] in 1982 was a milestone and marked a revival of these models, which have achieved in recent years great favor among physicists under the name of neural networks. This interest may be explained in large part by the conjunction of such factors as the perception of the ever-growing importance of biological sciences, the recent studies of spin glasses[9] in statistical mechanics and the development of optimization methods,[10,11] and the growing interest in

[1] Centre de Physique Théorique, UPR A0014 CNRS, Ecole Polytechnique, 91128 Palaiseau Cedex, France.

[2] Present address: Racah Institute of Physics, Hebrew University, 91904 Jerusalem, Israel.

0022-4715/89/0600-0859$06.00/0

computer sciences and computationally intensive physics for massively parallel systems.[12-15]

Most of the models studied by physicists belong to two large classes:

1. Layered feedforward networks[16,17] whose connections are determined via the error backpropagation method,[18] which is a modified gradient descent, or more efficient algorithms.[19]

2. Hopfield-like models,[10,20,21] that is, in their simplest form, completely and symmetrically connected networks of binary units.

Other models, such as Aleksander's Boolean network,[22] nonetheless represent very interesting alternatives.

As the system we shall introduce here rather pertains to the second category, we shall briefly recall the properties of the Hopfield model which may serve as categorizer or associative memory. We shall assume that neurons are spinlike[23] with two states, -1 and $+1$ (Hopfield's neurons were bipolar with low state 0[10]). One wants to store a set $\{\xi^\mu, \mu = 1,..., p\}$ of patterns that may be retrieved from initial conditions corresponding to corrupted patterns (due to the basic symmetry $S \leftrightarrow -S$ the opposite patterns $-\xi^\mu$ will also be stored with equal stability). To this aim, one appropriately sets the connections between the N neurons of the network using generally the Hebb rule[24]:

$$C_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \qquad \text{for} \quad i \neq j, \quad C_{ii} = 0 \tag{1.1}$$

or variants of this rule[25] using, for instance, nonlinear synapses.[26,27] Starting from the initial condition, the system then evolves according to the following zero-temperature dynamics:

1. The state S_i of the ith neuron changes according to the value of the input coming from all the other neurons,

$$S_i(t+1) = \text{Sign}\left[\sum_j C_{ij} S_j(t)\right] \tag{1.2}$$

2. The neurons may update their state simultaneously (Little's dynamics[28,29]), one at a time (as the Hamming distance between two successive configurations is then equal to 1, this yields a simple one-flip stability criterion) or independently according to a Poissonian process with given mean rate (Hopfield's original hypothesis which aimed at crudely modelizing delays and low activity in biological neural networks).

The system finally converges to an asymptotic state (which in the useful regime of the network lies near the stored pattern the initial condi-

tion resembles most) as proved by the existence of a Liapunov function for the dynamics:

$$E = -\tfrac{1}{2}\sum_{i,j} C_{ij} S_i S_j \tag{1.3}$$

The very existence of the Liapunov function is related to the symmetry $S \leftrightarrow -S$ of the spins, the symmetry of the connections, and the absence of self-coupling terms C_{ii}. Here E is an energy function for the system, which is thus amenable to a thermodynamic study[20,21,26] when synaptic noise is taken into account; the updating of neurons is then probabilistic and controlled by a temperature parameter T. Due to the complete connectedness of the network, thermodynamic equilibrium properties are given by a mean-field theory. Such a thermodynamic approach is impossible in the case of asymmetric connections, which must be studied by purely dynamical methods.[30–33]

This model shows a great robustness with respect to dilution.[26,34] On the other hand, it suffers from two major drawbacks:

1. A low storage capacity[35]: The system cannot store more than approximately $0.14N$ patterns before saturation occurs.[20,21] The ability to retrieve any pattern then sharply decreases; for a large number of prototype patterns (that is, of patterns one wishes to store) the system is in a spin-glass phase (the connections have then typically a random Gaussian distribution) and the energy landscape shows many local minima, most of them not directly related to the prototype patterns. The complex structure of this landscape has some ultrametric properties which have received much attention in spin-glass physics.[36] The catastrophic degradation of the retrieval quality at saturation may be avoided by modifying the Hebb rule: Schemes such as marginalist learning or learning within bounds enable one to conserve good retrieval properties at the price of a still more reduced storage capacity (working memory models[37,38]). One may also increase the capacity by using higher order networks.[39,40]

2. Great difficulties to store strongly correlated patterns. This problem, too, originates in the existence of spurious states which are mixtures of the prototype patterns and play an important role in the dynamics. One may partly overcome this problem by storing only the mutually orthogonal parts of the patterns (projection rule[41,42]) or by adequately preprocessing the data to obtain weakly correlated patterns. One may also devise learning schemes that take into account the correlations between the prototype patterns (hierarchic memories[43,44]). Appropriate models have also been introduced for biased patterns which contain for instance, a small number of $+1$.[20,45] Finally, we recall that the effect of

these spurious states can be reduced by using some "unlearning" scheme[46] or by running the system at finite temperature[20] or adopting nonlinear synapses.[26,27]

In Hopfield-like models the use of binary neurons obviously entails that all the neurons and all the connections are involved in the storage of any pattern.

Therefore, the Hopfield model does not permit one to deal with situations involving partial rather than global information, that is, when only part of the neurons or connections play a role in the storage of a pattern or in the dynamics of retrieval. This entails serious limitations on the kind of patterns one may store or one may feed the network with as initial conditions:

1. A pattern cannot involve only part of the network; all neurons must contain relevant information with respect to any pattern.

2. Similarly, the initial condition is defined on the whole network; this rules out situations where the initial information is partial, that is, the state of part of the network is *a priori* unknown.

3. One cannot easily define partial patterns, that is, patterns defined on only part of the network. This rules out the possibility of defining a pattern as the union of several subpatterns in a natural way.

These remarks have stimulated the present study. Section 2 is devoted to a detailed presentation of our model, which is a Hopfield-like network of *three-state neurons.* To the high and low states -1 and $+1$ we add a *mid-state* 0 with given stability threshold. This additional neutral state embodies the absence of information and permits us to deal with situations where the state of part of the neurons is *unknown* or *irrelevant.* It also plays an important part in the dynamical behavior of the network, as populations of neurons may temporarily switch to 0; this point will be illustrated in Section 3 which is devoted to a general discussion of our system's behavior. We have already hinted at the fact that Hopfield's standard model, though largely motivated by biology, is of course a rather crude idealization of the biological situation: The neurons themselves can be either inactive (0 state) or active (1 state) like real neurons, but they do not display above threshold the frequency coding feature of real neurons[47]; moreover, the connectedness and symmetry of the network is not realistic from a biological standpoint. Our model departs still more from biological neurons and we do not aim at any biological relevance.

With such networks one easily retrieves a pattern starting from an *incomplete* initial condition where part of the information is unknown. This will be shown in Section 4. We shall consider there a situation close to

Hopfield's original model: We store global patterns which involve the whole network. We then study the retrieval properties of the system for two different classes of initial conditions: incomplete patterns, that is, with part of neurons set to the null state; and corrupted patterns, that is, with the same neurons flipped rather than set to 0. We are thus led to distinguish between pattern completion and noise correction.

These networks also allow for the storage of a pattern under the guise of overlapping prototype subpatterns of smaller size. The dynamics of small overlapping patterns will be studied in Section 5. We shall see that if the subpatterns display a sufficient coherence on their overlaps, the network shows strong *inferential* properties: Starting from an initial condition which corresponds to only one subpattern, other subpatterns will be finally excited. This behavior is controlled by the amount of overlap, the coherence on the overlaps, and the value of the stability threshold of the null state.

Our model can deal with *missing* or *partial* information and retrieve information not contained in the initial condition (inference and pattern completion properties). We think that refined models conserving these basic capabilities could be of interest in certain pattern recognition problems and for connectionist expert systems. This will be further discussed in the conclusion.

J. Yedidia is presently completing an analytical study[48] of this same model. He makes use of the method developped by B. Derrida *et al.*[30] for asymmetric neural networks, which is limited to high dilution rates (to avoid any indirect feedback in the network). This study nonetheless yields very interesting results which corroborate the results we obtain here for a nondiluted network.

2. THE THREE-STATE NEURAL NETWORK

We consider a network of N three-state (-1, 0, and $+1$) neurons and sets of prototype patterns $\{\xi^\mu,\ \mu = 1,..., p\}$ which are randomly generated according to the law

$$\text{Prob}(\xi_i^\mu = \xi) = P(-1)\,\delta(\xi + 1) + P(0)\,\delta(\xi) + P(1)\,\delta(\xi - 1) \qquad (2.1)$$

$P(-1)$, $P(0)$, and $P(1)$ are the probabilities to find any randomly chosen neuron in a pattern in the state -1, 0, or 1. We shall call (by analogy with functions) the subnetwork $A_\mu = \{i \text{ such that } \xi_i^\mu \neq 0\}$ the support of the pattern ξ^μ and shall denote by $N_\mu = \text{card}(A_\mu)$ the size of that pattern. To store the set of patterns $\{\xi^\mu\}$, we set up the connections using the Hebb learning rule (1.1). This choice has been made both for the sake of simplicity and

to obtain inference properties stemming from the correlations between small patterns (see Section 5). It implies that the storage of a new pattern ξ^μ does not affect the whole network but only the synapses which connect two neurons of A_μ. This entails the following:

1. When one stores a finite set of patterns, the dynamics takes place on a subnetwork of the original network of average size $[1-P(0)^p]N$.

2. Each synapse only carries the information relative to a *subset* {of average size $p[1-P(0)]^2$} of the complete set of patterns.

3. For large p, the distribution of coupling constants (i.e., synaptic strengths) is Gaussian with standard deviation $p^{1/2}[1-P(0)]/N$. Therefore the partial connectedness that may appear in our model in certain situations just reflect the properties of the set of patterns: small number of patterns, small size of the patterns or organization of the patterns in disjoint clusters. This is basically different from the partial connectedness which is voluntarily imposed in other models, namely:

4. There are diluted Hopfield models where one randomly destroys a fraction of synapses in order to test the robustness of the network.[26,34] For a large number p of stored patterns and for dilution rate c, the distribution of coupling constants is then a Gaussian distribution with standard deviation $p^{1/2}/N$ and total weight $1-c$ on which is superimposed a delta peak $c\delta(C_{ij})$. However, the whole network still takes part in the dynamics, even for finite p, except for a fraction of average size Nc^{N-1} which can be disregarded in the thermodynamic limit. Moreover, the unbroken connections still carry information of the whole set of patterns.

5. There are models with three-level synapses[26] where weak connections are suppressed while the other synapses take the values $+1$ or -1 according to their sign. In this case, too, one banks on the robustness of the standard Hopfield model to obtain a correct behavior of the network, and the distribution of synapses features a delta peak at 0 corresponding to the connections set to 0.

We now present the dynamics of our model. The whole network is updated at each time step according to the following rule (see Fig. 1). We define a threshold U which is the same for all the neurons. If the summed input on a neuron from all other neurons in the network strictly exceeds U, the new state is $+1$; similarly, if the input is smaller than $-U$, the new state is -1. For intermediate values of the input the neuron sets to the null state 0. We have chosen exactly opposite values for the low and high thresholds in order to preserve the symmetry $S \leftrightarrow -S$. Each pattern is then stored together with the reverse pattern and they have equal stability. The

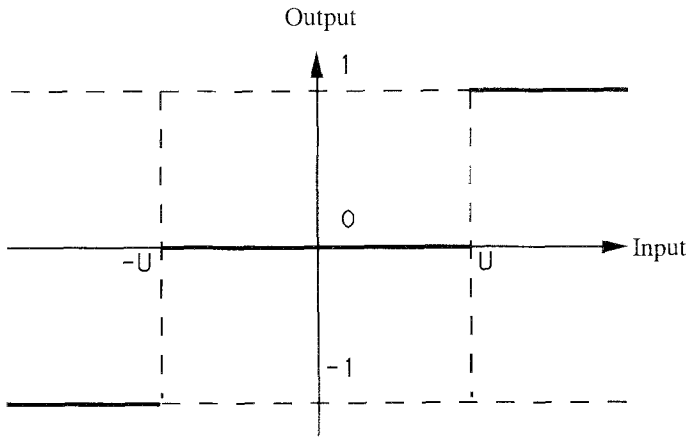


Fig. 1. Response function of individual neurons.

response of individual neurons is embodied in the zero-temperature dynamical equation

$$S_i(t+1)=\theta[h_i(t)-U]-\theta[-h_i(t)-U] \tag{2.2}$$

where $\theta(x)$ is the Heaviside function [with the convention $\theta(0)=0$] and where we have introduced the "local field" $h_i=\sum_{j=1}^{N} C_{ij}S_j$. Alternatively, it may be written in terms of the conditional probabilities:

$$\begin{aligned} P(S_i(t+1)=1;\, S(t)) &= \theta[h_i(t)-U] \\ P(S_i(t+1)=-1;\, S(t)) &= \theta[-h_i(t)-U] \\ P(S_i(t+1)=0;\, S(t)) &= 1-\theta[h_i(t)-U]-\theta[-h_i(t)-U] \end{aligned} \tag{2.3}$$

We found no conspicuous Liapunov function for this dynamics. One can nevertheless introduce a noisy dynamics characterized by the "temperature" T in the following way: in (2.3) one just replaces $\theta(x)$ by the smooth function

$$\frac{1}{1+\exp[-2x/T]}$$

and one performs the probabilistic update of neurons in agreement with the conditional probabilities then obtained. The average response of an individual neuron then reads

$$\frac{\mathrm{sh}[\beta h]}{\mathrm{ch}[\beta h]+\mathrm{ch}[\beta U]}, \qquad \beta=\frac{1}{T} \tag{2.4}$$

instead of (2.2) in the deterministic case (we recall that in Hopfield's model the average response is sigmoidal). For sheer convenience we shall restrict ourselves in what follows to the zero-temperature dynamics (the introduction of noise would not affect the basic properties of our model).

As we shall henceforth heavily rely on numerical simulations, it seems appropriate to define now the quantities that we shall use to characterize the system's behavior. To avoid large statistical errors in their computation, we shall average them over n_{stat} different set of prototype patterns (see Section 3).

To characterize the length of the dynamics, we start from one of the patterns ξ^μ and let the system evolve until a stationary state of the dynamics is reached after t_μ steps; the duration of the dynamics is then measured by the average over patterns $t_{\text{dyn}} = \langle t_\mu \rangle$.

To characterize the stability of patterns, we use two different sets of quantities. The first set is just an extension to the present case of quantities used for the standard Hopfield model:

1. m_μ is the overlap between the initial condition $S(0) = \xi^\mu$ and the final state $S(t_\mu)$:

$$m_\mu = \frac{1}{N} \sum_{i=1}^{N} \delta[\xi_i^\mu, S_i(t_\mu)]$$

where $\delta(x, y) = 1$ if $x = y$ and $\delta(x, y) = 0$ otherwise. If m_μ is barely different from 1, the pattern is well retrieved.

2. $m = (1/p) \sum_{\mu=1}^{p} m_\mu$ is the average of m_μ over the set of patterns $\{\xi^\mu\}$. It quantifies the global retrieval quality of the network.

3. The function $f(x)$, x ranging between 0 and 1, gives the fraction of patterns for which the quality of retrieval m_μ exceeds the given value x.

In order to characterize the amount of information *present in the initial condition* ξ^μ which can be actually retrieved, we introduce the overlap restricted to the subnetwork A_μ of the initial state $S(0) = \xi^\mu$ and the final state $S(t_\mu)$:

$$m_\mu^{\pm} = \frac{1}{N_\mu} \sum_{i \in A_\mu} \delta[\xi_i^\mu, S_i(t_\mu)]$$

This quantity does not take into account the evolution of the $N - N_\mu$ neurons that were initially in the null state. $m_\mu^{\pm}$ may be significantly greater than m_μ if the restriction of the initial pattern to its support A_μ remains unaltered but other patterns have been excited during the evolution of the system ("inferential regime"; see Section 5). We also define, as we did for m_μ, the two average quantities:

1. $m^{\pm}$, which is obtained from the $m_\mu^{\pm}$ by averaging over the set of patterns $\{\xi^\mu\}$.

2. $f^{\pm}(x)$, which is the mean fraction of patterns such that $m_\mu^{\pm} \geqslant x$.

Due to the influence of the other patterns, the network does not store, unless α is small, the prototype pattern ξ^μ, but rather a corrupted version of it. To quantify this alteration, we use the three following quantities, which, like the quantities we just discussed, are relative to the initial information:

1. $1 - m_\mu^{\pm}$ is the total alteration of the pattern on its support A_μ.

2. n_{flip}^μ is the proportion of neurons in the restriction of the final state to A_μ which are flipped with respect to the initial condition ξ^μ.

3. Similarly, n_{null}^μ is the proportion of neurons on the support A_μ of ξ^μ which have set to zero during the evolution ($n_{\text{null}}^\mu + m_\mu^{\pm} + n_{\text{flip}}^\mu = 1$).

The corresponding mean quantities $m^{\pm}$, n_{flip}, and n_{null} are obtained by averaging over the patterns $\{\xi^\mu\}$.

Finally, we measure as follows the excitation of the pattern ξ^μ during the evolution of the system initially fed with some prototype pattern ξ^ν. We first introduce $m_{\mu;\nu}^{\pm}$ and $n_{\text{flip}}^{\mu;\nu}$ (similar to $m_\mu^{\pm}$ and n_{flip}^μ) which quantify the discrepancy on A_μ between the exact pattern ξ^μ and the state resulting from the evolution of ξ^ν. We then define the excitation level $e_{\mu;\nu} = m_{\mu;\nu}^{\pm} - n_{\text{flip}}^{\mu;\nu}$. This quantity ranges between -1 and $+1$; the value $+1$ corresponds to the total excitation of the pattern ξ^μ on its support A_μ, whereas the value -1 corresponds to the total excitation of the opposite pattern. More generally $e_{\mu;\nu}$ quantifies the balance between he excitation levels of ξ^μ and the opposite pattern; it will prove useful in Section 5 to analyze the inferential properties of the network.

Before examining its influence on the dynamical behavior of our system, we first define an appropriate rescaling of the threshold. Consider a pattern ξ^μ of size N_μ which does not overlap any other pattern. As the summed input to any neuron of A_μ is obviously equal to $(N_\mu - 1)/N$, the pattern is perfectly stable provided that U does not exceed the critical threshold $U_\mu^* = (N_\mu - 1)/N$; otherwise, it immediately decays to the null pattern ξ^0 ($\xi_i^0 = 0$, $i = 1,..., N$) which has no information content. Thus, we see that for given threshold one cannot retrieve a pattern containing too little information (N_μ too small) unless it is possibly stabilized by the interaction with the other patterns; if the threshold is too high, the initial information disappears. This control of the amount of initial information required for information processing to take place is an interesting feature of our model.

As we shall always store random patterns with a mean number of

neutral neurons $NP(0)$, we define the mean critical threshold $U^* = \langle U_\mu^* \rangle = P(0) - 1/N$. It is convenient to choose as a parameter of our model the rescaled threshold U/U^*, which will from now on also be denoted by U.

3. THE DYNAMICAL BEHAVIOR; THE INTERMEDIATE CASE

We now study the main features of our model's behavior. To this end, we examine in some detail the typical intermediate case $P(0) = 1/2$, $P(-1) = P(1) = 1/4$; one then stores unbiased patterns with mean size $N/2$. Figure 2a displays the retrieval quality of the network for an intermediate value of threshold $U = 1/2$ (half the mean critical threshold). The dependence of both m and $m^{\pm}$ is qualitatively similar to what is observed for the Hopfield–Little standard model (see Fig. 2b):

1. For small values of α ($\alpha \leqslant 0.05$ approximately) the patterns interfere little and the average retrieval quality is optimal (better than 99%).

2. The retrieval quality is strongly decreased when α is increased past this value as a consequence of "collective" effects: The stability of most prototype patterns is affected by the "noise" due to the storage of the other patterns.

3. At $\alpha \cong 0.15$ one enters the saturation regime where the average retrieval qualities m and $m^{\pm}$ remain nearly constant.

The critical value of α where the retrieval quality drops off is about three times smaller in the present case than in the standard Hopfield–Little model. It seems therefore appropriate to discuss at this point the "signal to noise ratio"[20]; this will shed some light on the influence of $P(0)$ and U on the retrieval capacity of our model. Let us consider a specific prototype pattern ξ^ν and examine its stability. When the system's state is ξ^ν, any neuron ξ_i^ν of the support A_ν receives in the thermodynamic limit the input

$$\xi_i^\nu \frac{N_\nu}{N} + \xi_i^\nu \sum_{\mu \neq \nu} \xi^\mu \cdot \xi^\nu, \qquad \text{where} \quad \xi \cdot \eta = \frac{1}{N} \sum_{j=1}^{N} \xi_j \eta_j \tag{3.1}$$

If the pattern ξ^ν has no overlap with the other patterns ($\xi^\mu \cdot \xi^\nu = 0$ for $\mu \neq \nu$), the second term vanishes and the stability condition just reads (in the thermodynamic limit)

$$\frac{N_\nu}{N}(1 - U) > 0$$

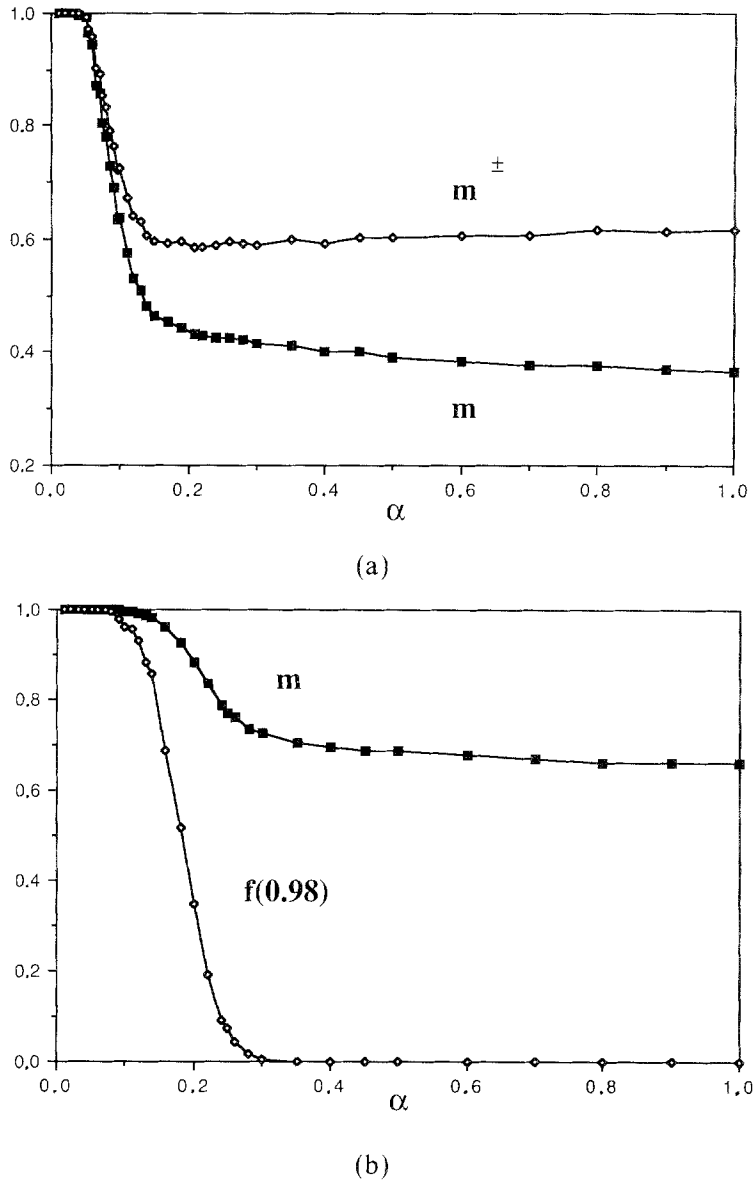


Fig. 2. Average retrieval quality as a function of the number of patterns stored ($\alpha = p/N$): (a) Intermediate case; the retrieval qualities on the whole network (m, solid squares) and on the support of the initial condition ($m^{\pm}$, open diamonds) are both displayed. (b) Hopfield–Little model; the average retrieval quality m (solid squares) plotted together with the proportion $f(0.98)$ of well-retrieved patterns (open diamonds).

Therefore, if no two patterns did overlap, the criterion for the stability of almost all patterns would read

$$[1-P(0)](1-U)>0 \tag{3.2}$$

However, the overlaps between patterns add to the expression (3.1) of the input a noiselike term with zero mean (for unbiased prototype patterns) and standard deviation

$$\alpha^{1/2}[1-P(0)]^{3/2} \tag{3.3}$$

A rough storage criterion for the network is then obtained by considering the ratio of the "noise term" (3.2) to the "signal term" (3.3):

$$\gamma\alpha^{1/2}\frac{[1-P(0)]^{1/2}}{1-U}<1$$

γ is some coefficient which slightly depends on the parameters ($P(0)$, U); in the intermediate case ($P(0)=1/2$, $U=1/2$), $\gamma\cong 3.2$, and for the Hopfield–Little model ($P(0)=0$, $U=0$), $\gamma\cong 2.6$. Thus, we see that the storage capacity is affected by both the threshold value U and the typical size of patterns: A higher threshold or a larger size of patterns reduces the storage capacity, as could be expected, as they diminish the stability of patterns and increase the interferences between patterns.

On Fig. 3, we examine in more detail the corruption of the prototype

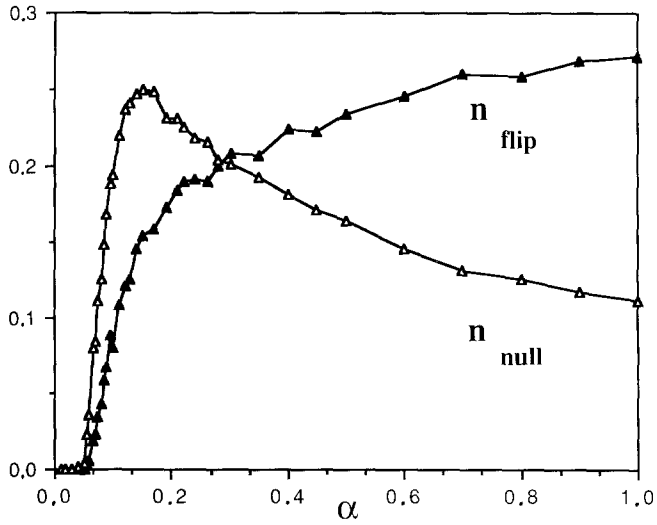


Fig. 3. Corruption of the prototype patterns during storage. The average proportions of neurons in the support that flip (n_{flip}, solid triangles) or set to zero (n_{null}, open triangles) are plotted versus α.

patterns (which are perfectly stored for $\alpha \leqslant 0.05$) due to their interference: We plot the average proportion of neurons in the support flipped (n_{flip}) or set to zero (n_{null}) with respect to the prototype pattern. For $0.05 \leqslant \alpha \leqslant 0.32$ the perturbation is mainly due to neurons setting to zero (at the peak $\alpha \cong 0.15$, we have 25% of neurons which have set to zero versus 15% of flipped neurons), whereas for higher values of α ($\alpha \geqslant 0.32$), flips become predominant; in this region the average level of corruption rate $1 - m^{\pm}$ remains almost constant (of the order of 60%). One also observes (see Fig. 4) a very good correlation between n_{flip} and the average duration t_{dyn} of the dynamics ($t_{\text{dyn}} \cong 350 n_{\text{flip}}$ beyond the retrieval regime) which both keep increasing for high α, in contrast with the saturation of $m^{\pm}$: The processing time is thus directly proportional to the "distance" n_{flip} between the prototype pattern and the attractor of the dynamics it evolves to. Note that the three different regimes of the network clearly stand out on all these curves: individual regime (perfect retrieval, $t_{\text{dyn}} \cong 1$), transition regime (sharp drop of the retrieval quality, strong dependence on α of t_{dyn} and n_{flip}), and saturation regime (constant retrieval quality, weak dependence of t_{dyn} and n_{flip} on α).

The dynamical properties of the system strongly depend on the threshold U. A higher threshold decreases the stability of patterns and limits inference, as will be shown in Section 5. However, more subtle dynamical effects also take place (see Fig. 5). Let us consider, for example, the storage of a small set ($\alpha = p/N = 0.05$) of patterns. The best retrieval

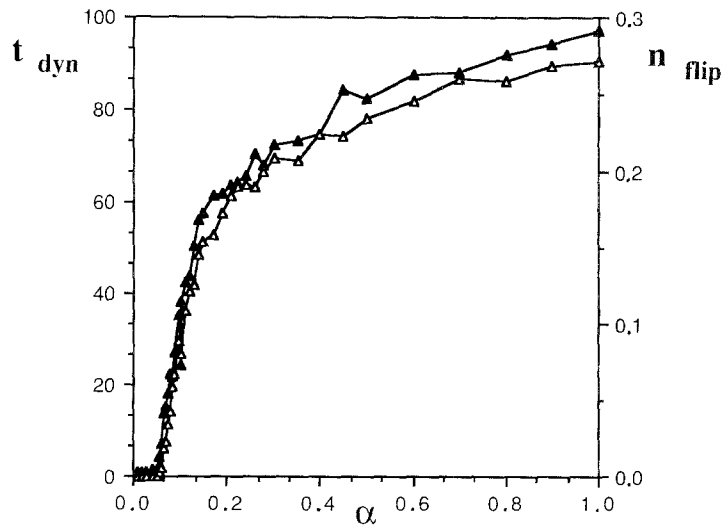


Fig. 4. Average length of the dynamics t_{dyn} (solid triangles) versus α in the intermediate case. Observe the strong correlation with n_{flip} (open triangles) in the saturated regime.

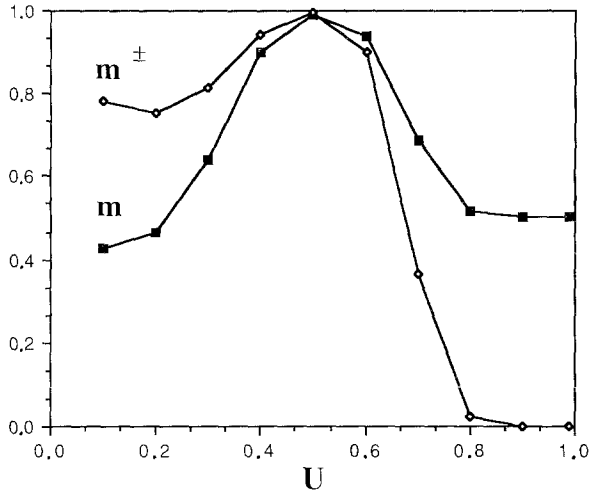
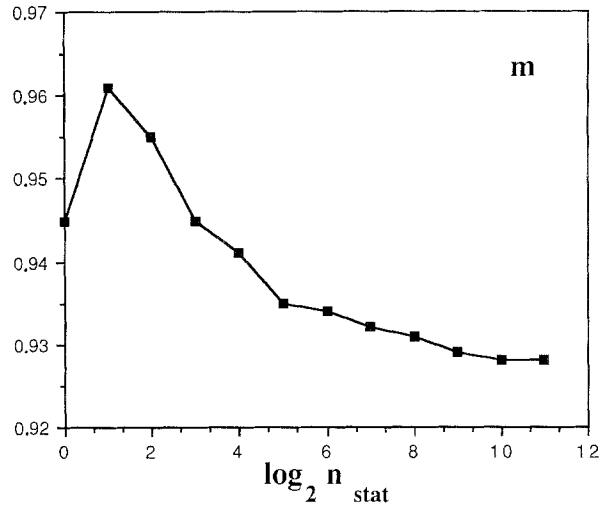


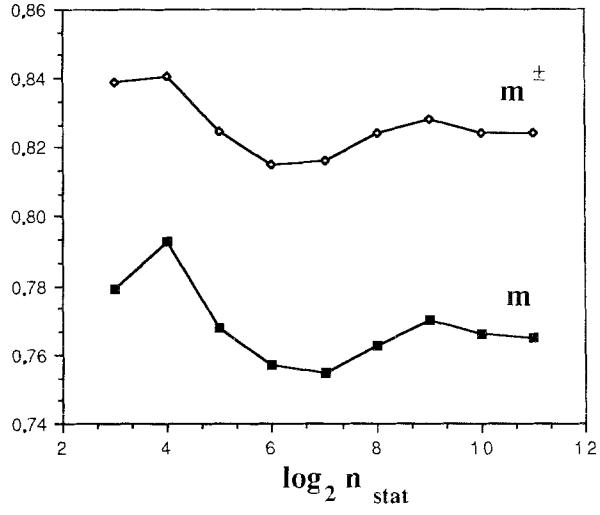
Fig. 5. Dependence of the average retrieval qualities on the whole network (m, solid triangles) and on the support of the initial condition ($m^{\pm}$, open diamonds) on the threshold U in the intermediate case for $\alpha = 0.05$.

quality is achieved for $U \cong 1/2$. This may be explained as follows. During the evolution of the network, some neurons are submitted to conflicting influences, which leads to a small input value. If this input is below threshold, the neurons set to zero, which cancels most of the frustration present in the current state of the network. This population then plays temporarily no role into the dynamics and enters again into play only later after the remaining "coherent" part of the network has further evolved. This subtle dynamical effect enhances the quality of retrieval of any prototype pattern by reducing the "noise" due to the other patterns. It is observed for intermediate values of the threshold ($U \approx 0.5$) and is limited to situations of low storage where no inference takes place: For the chosen values of the parameters, the best retrieval is obtained for $U = 0$ as soon as $\alpha \approx 0.08$ and no local enhancement of the retrieval quality at $U = 0.5$ is observed for $\alpha > 0.11$. We think nevertheless that this effect deserves attention as a striking illustration of the positive effects that the introduction of a neutral state may entail.

The present work is essentially numerical. Therefore we have paid attention to both finite-size effects and statistical effects, as will now be discussed. We have also compared the influence of such effects on our system to the influence they have on the Hopfield–Little model. Figure 6 clearly shows that the retrieval quality depends very little on the number n_{stat} of sets of prototype patterns in our statistical sample (here $\alpha = 0.2$; for other

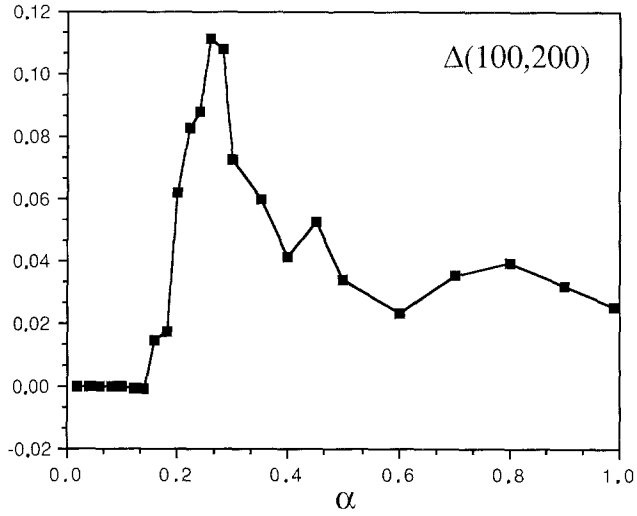


(a)

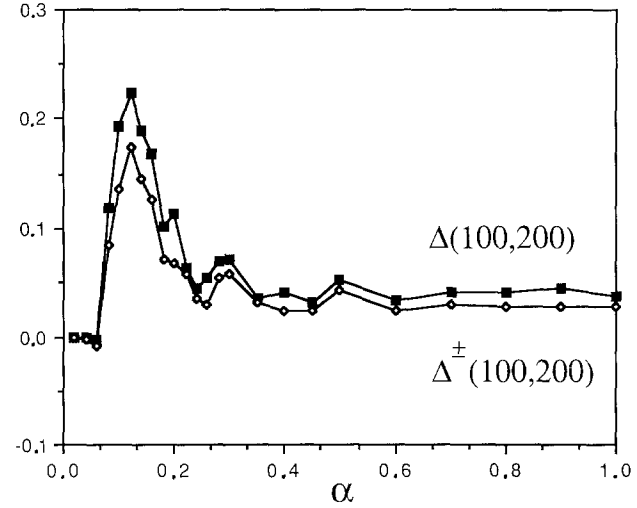


(b)

Fig. 6. Average retrieval quality m (solid squares) as a function of the size n_{stat} of the statistical sample: (a) Hopfield–Little model, $\alpha = 0.2$. (b) Intermediate case, $\alpha = 0.1$. The average retrieval quality of the initial information $m^{\pm}$ (open diamonds) is also plotted.



(a)



(b)

Fig. 7. Dependence of the average retrieval qualities on the size of the network. Networks with size $N=100$ and $N=200$ are compared. The relative variation $\Delta(100, 200)=[m(100)-m(200)]/m(200)$ (solid squares) is plotted versus α: (a) Hopfield–Little model. (b) Intermediate case. The relative change $\Delta^{\pm}(100, 200)=[m^{\pm}(100)-m^{\pm}(200)]/m^{\pm}(200)$ in $m^{\pm}$ is also plotted (open diamonds).

values of α, statistical effects were still lower); therefore, we have always chosen $n_{\text{stat}} = 16$ throughout our study. Figure 7 displays the dependence of the retrieval quality on the size of the network. In the case of the Hopfield standard model the number of neurons has no significant influence as long as saturation ($\alpha = 0.14$) is not reached. In the transition region where the retrieval quality drops off, we observe much stronger size effects, which nonetheless never exceeds 11%. They diminish for higher values of α, but remain at a level of about 3% in the saturation regime (spin-glass phase). Our model displays the same general behavior: Size effects are negligible in the individual regime and peak in the transition region, where they reach 20%. In the present study we have always considered networks of 200 neurons for the following reasons:

1. This allows for quick numerical computations on the SCS40 and VP200 vector computers we used.

2. Much larger networks (more than 1000 neurons) are not amenable to study, as the processing time of the programs we used grows as N^3 or N^4.

3. Finite-size effects are important only for intermediate values of α.

4. We are more interested in understanding the qualitative behavior of our system than obtaining precise quantitative results.

4. PATTERN COMPLETION

We now study the pattern completion properties of our network in the "Hopfield-like" limit where $P(0) = 0$ (the size of almost every pattern is N) and the threshold U is weak; $U = 0$ corresponds to the standard Hopfield model. The question of how large the threshold can be chosen in that limit is answered by examining the dependence on U of the retrieval quality in the retrieval regime (typically $\alpha \leqslant 0.14$). The results are displayed on Fig. 8. For low values of the threshold (typically $U \leqslant 0.2$) the retrieval remains almost perfect ($m \cong 1$). As U is further increased, the average retrieval quality steadily decreases and it finally vanishes for $U \cong 0.9$. The behavior of the average proportion of flips n_{flip} between the initial and final states is quite different, as it presents a maximum ($n_{\text{flip}} \cong 0.15$) for $U = 0.5$. These features stem from the following fact: For high values of U a large proportion of the initial states decay to the null state ξ^0 and therefore the average alteration of the prototype patterns during the dynamics mainly comes from neurons setting to the state 0. Actually, for $U \geqslant 0.9$ all patterns evolve to the null state and both m and n_{flip} vanish. In view of these results we chose for the scaled threshold throughout the present analysis the value

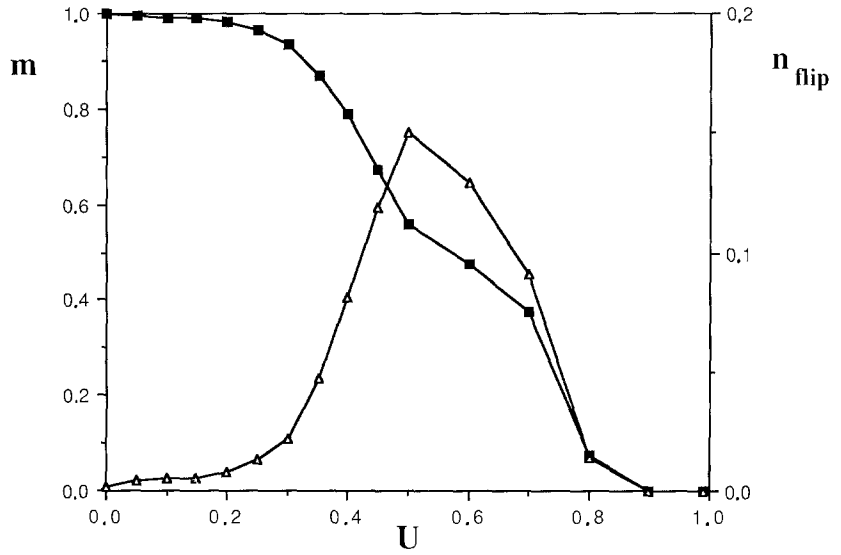


Fig. 8. Dependence on the threshold U of the retrieval quality in the Hopfield-like limit. The average retrieval quality (m, solid squares) and the average proportion of flips with respect to the prototype patterns (n_{flip}, open triangles) are plotted versus U. Here $\alpha = 0.1$.

$U = 0.1$, which does not significantly degrade the retrieval quality with respect to the standard Hopfield–Little model while enabling the specific properties of our model to play an important role.

As already mentioned in the introduction, we consider two very different situations:

1. Pattern completion. The initial conditions are then derived from the prototype patterns ξ^μ by setting to the null state 0 a fraction x_{mod} of the neurons in the support.

2. Retrieval of prototype patterns from noisy initial conditions, that is, correction of noisy patterns. These initial conditions are obtained by flipping part of the neurons.

These two cases are compared on Figs. 9a and 9b, which respectively correspond to the retrieval regime ($\alpha = 0.1$) and the transition regime ($\alpha = 0.2$) of the Hopfield–Little model. To obtain these results we randomly chose $x_{\text{mod}} N$ neurons for each prototype pattern ξ^μ, either flipped or set to 0 these neurons depending on the case, and computed the overlap m_μ on the whole network between the final state resulting from the evolution of the initial condition thus obtained and ξ^μ. We then averaged m_μ over the n_{stat} sets of patterns $\{\xi^\mu\}$ (but not on the symmetric counterparts $-\xi^\mu$) to

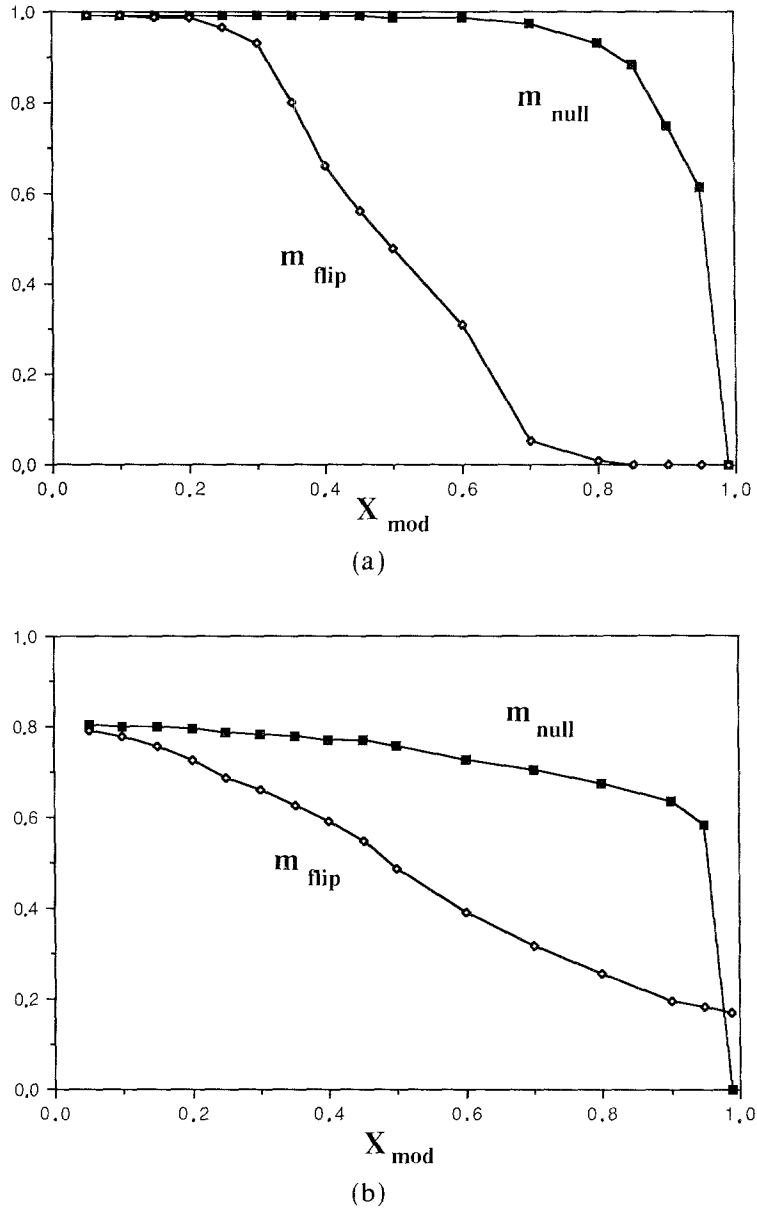


Fig. 9. Pattern completion capabilities for two different kinds of initial conditions: Prototype patterns with a proportion x_{mod} of neurons flipped and patterns with the same neurons set to zero. The respective values of m in these two cases, denoted by m_{flip} (open diamonds) and m_{null} (solid squares), are plotted versus x_{mod}: (a) Retrieval regime $\alpha = 0.1$. (b) Transition regime $\alpha = 0.2$.

obtain m. Here the meaning of the quantities m_μ and m is slightly different from what it is in the other sections, as the initial condition is an altered version of a stored pattern; however, no confusion should arise.

In the retrieval regime (see Fig. 9a), pattern completion is excellent even if the initial condition is only a small part of the prototype pattern; more than 90% of the pattern is recovered as long as the initial condition contains more than 15% of the information relative to that pattern. This strikingly illustrates the redundancy of storage in neural networks. For smaller initial information content the retrieval quality sharply drops off to 0 as most initial conditions evolve to the null state ξ^0. In contrast, the retrieval quality is much lower if one starts from a noisy pattern: It becomes lower than 90% when 35% of the prototype pattern is flipped and vanishes when 85% of the neurons are flipped. This difference can be explained as follows: For a noisy initial condition one is faced with a mixture of the pattern with the opposite pattern and one cannot expect a good retrieval quality when a 50% mixture is approached. On the other hand, a partial initial condition contains only information relative to the pattern itself, which enhances the retrieval quality. Note that the symmetry of the curve corresponding to noisy initial conditions with respect to the point (0.5, 0.5) also stems from the basic symmetry $S \leftrightarrow -S$ of the system. The proportion of perfectly retrieved patterns (retrieval quality better than 98%) is plotted on Fig. 10 in both situations. As expected, no pattern is perfectly retrieved from a noisy initial condition when the alteration level x_{mod} exceeds 50%; on the other hand, 90% of the patterns are well retrieved as long as the initial condition contains at least 50% of the total information on the prototype pattern.

In the transition regime, the behavior of the network is globally similar to what it was in the previous case. However, the decay of the retrieval quality when x_{mod} is increased is faster. Note also that for very noisy initial conditions the retrieval quality no longer vanishes; this, together with the imperfect storage of the prototype patterns ($m \cong 0.8$ for $x_{\text{mod}} = 0$), is obviously due to the presence of spurious attractors.

Thus we see that the use of three-state neurons enables us to achieve true pattern completion, that is, retrieval of prototype patterns starting from only part of the information. The capability of our network to perform such a task is clear especially in the storage regime of the network and is substantially higher than the ability to eliminate noise superimposed on a prototype pattern. It degrades abruptly when the initial information becomes too small, as most initial conditions then decay to the null pattern. The quality of pattern completion of course depends on the storage level α in the network and the exact value of the threshold U.

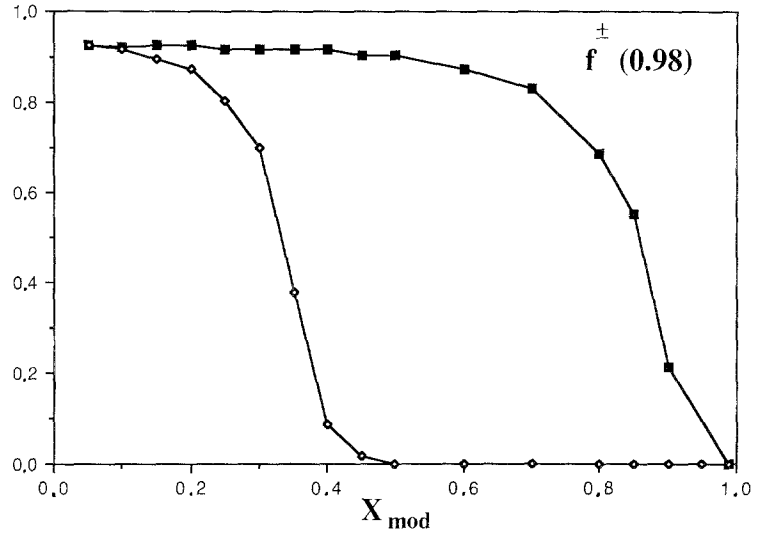


Fig. 10. Average proportion $f^{\pm}(0.98)$ of prototype patterns well retrieved as a function of the discrepancy x_{mod} between the initial condition and the corresponding prototype pattern. Conventions are the same as in Fig. 9.

5. THE INFERENTIAL PROPERTIES

The present section is devoted to the inferential properties of our network, that is, to the excitation of prototype patterns (or their symmetric counterparts) during the dynamics, starting from one of the stored patterns as initial condition. As we shall see, such a behavior occurs when small patterns are stored, whereas for large patterns the behavior is rather similar to the behavior of the Hopfield model. This originates in the nature of the spurious states, which in the former case mainly consists in "mergings" of small patterns, as will be further explained below. Such spurious attractors will play here a quite positive role, as they will enable us to store implicitly in the network a pattern as the union of several prototype patterns.

Though it is not a sufficient condition (as exemplified by the Hopfield model, where the retrieval of the prototype patterns degrades only for large enough α), a collective dynamics of patterns requires the supports of the patterns to overlap. This requirement is easily met. Indeed, when $P(0) > 0$, two randomly chosen patterns overlap with probability 1 in the thermodynamic limit: For two patterns with respective size $(1-c)N$ and $(1-c')N$ the probability $P(c, c')$ that they do not overlap behaves for large N as

$$P(c, c') = \frac{(1-c)^{N(1-c)+1/2}\,(1-c')^{N(1-c')+1/2}}{(1-c-c')^{N(1-c-c')+1/2}} \qquad \text{for} \quad c + c' \leqslant 1$$

$$P(c, c') = 0 \qquad \text{for} \quad c + c' > 1$$

It tends to 0 when N goes to infinity except when $c = c' = 0$, a situation that almost never occurs. Even if the prototype patterns have finite support with probability 1, they do overlap in the thermodynamic limit if their number grows fast enough, as shown by the following percolationlike argument. Indeed, consider a network of N neurons and p prototype patterns with same finite size n (for simplicity). The probability P for two randomly chosen patterns to overlap then reads

$$P = 1 - \frac{C_{N-n}^{n}}{C_{N}^{n}}$$

and behaves in the thermodynamic limit as n^2/N. Consider the p prototype patterns as p sites, any two of them being connected with probability P (a link corresponds to the existence of an overlap). The mean number of chains (without loops) of overlapping patterns connecting i and i' is then

$$n_{\text{chains}} = \sum_{k=2}^{n} A_{p-2}^{k-2} P^{k-1}$$

As $P \cong n^2/N$, the quantity n_{chains} vanishes in the thermodynamic limit if p is finite, as could be expected. On the other hand, if p scales like N, we have

$$n_{\text{chains}} = \exp(1/P)\, n!\, P^{n-1}$$

Using Stirling's formula, we see that n_{chains} grows without bound when $N \to \infty$ provided that $p > N/n^2$ and vanishes if this condition is not fulfilled. This means that with probability 1, randomly chosen patterns do not overlap if $p < N/n^2$ and that, on the contrary, an infinite cluster of overlapping patterns exists in the thermodynamic limit when $p > N/n^2$. Actually, one could expect the transition to occur for $Pp = 1$ in such a mean-field theory, as Pp is the average number of patterns overlapping with any given pattern.

Even with overlapping patterns, the appearance of an inferential behavior depends on the size of the overlaps, the threshold value, and the "coherence" of the patterns on their overlaps. Before analyzing the behavior of our model, let us examine the influence of these three factors in the very simple case of two overlapping patterns ξ^A and ξ^B with respective supports A and B and same size n. To quantify the size of the overlap,

we introduce the overlap ratio r, which is the ratio of the number of neurons in the overlap to the common size n of the patterns. The connection matrix then reads

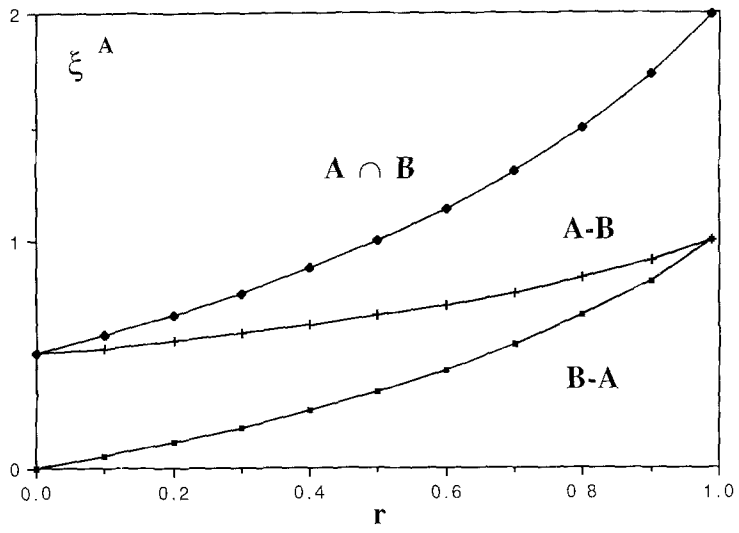
$$C_{ij} = \frac{1}{n(2-r)} (\xi_i^A \xi_j^A + \xi_i^B \xi_j^B)$$

Table I and Fig. 11 display in the thermodynamic limit $n \to \infty$ the input level as a function of the overlap ratio r for the three different groups of neurons $A \cap B$, $A - B$, and $B - A$ and for the three different initial conditions $\xi^{A \cap B}$, ξ^{A-B}, and ξ^A, assuming a perfect coherence of the two patterns A and B on their overlap $A \cap B$; $\xi^{A \cap B}$ is the restriction of ξ^A to $A \cap B$ and ξ^{A-B} its restriction to the set $A - B$. By complete coherence we mean that the two prototype patterns ξ^A and ξ^B coincide perfectly on $A \cap B$; $\xi_i^A = \xi_i^B$ for i in $A \cap B$. We see that for $U < 1/2$ the pattern ξ^A is retrieved for all overlap ratios (see Fig. 11a). On the contrary, inference may take place, that is, the whole pattern ξ^B be excited, only if r is high enough. Notice also the effect of coherent overlap (all input levels grow with r). When the initial condition is ξ^{A-B} (see Fig. 11b), inference never occurs; this could be expected, as the initial condition ξ^{A-B} contains no information on the pattern ξ^B. In addition, the larger r is, the lower must U be chosen in order to ensure the stability of ξ^A as the size of the initial condition, that is, the amount of initial information on the pattern, decreases with growing r. Finally, if the initial condition is $\xi^{A \cap B}$ (see Fig. 11c) we remark that the input level for the neurons of $A \cap B$ is twice the level for the neurons of $A \triangle B$ as a consequence, loosely speaking, of the double storage of the subpattern $\xi^{A \cap B}$.

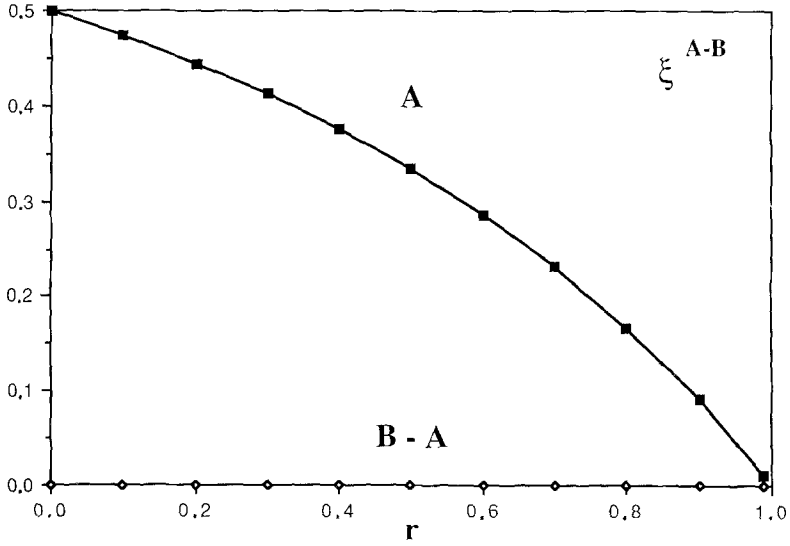
Table I. Excitation Levels of the Neurons in the Three Groups $A - B$, $A \cap B$, and $B - A$ for the Three Different Initial Conditions ξ^A, $\xi^{A \cap B}$, and ξ^{A-B} in a Situation Where Two Patterns A and B Overlap[a]

Initial condition	Neurons $A - B$	Neurons $A \cap B$	Neurons $B - A$
ξ^A	$1/(2-r)$	$(1+r)/(2-r)$	$r/(2-r)$
	$1/(2-r)$	$1/(2-r)$	0
ξ^{A-B}	$(1-r)/(2-r)$	$(1-r)/(2-r)$	0
	$(1-r)/(2-r)$	$(1-r)/(2-r)$	0
$\xi^{A \cap B}$	$r/(2-r)$	$2r/(2-r)$	$r/(2-r)$
	0	$r/(2-r)$	0

[a] The excitation levels are given as functions of the overlap ratio r; the upper expression corresponds to perfect coherence, the lower expression to total incoherence.

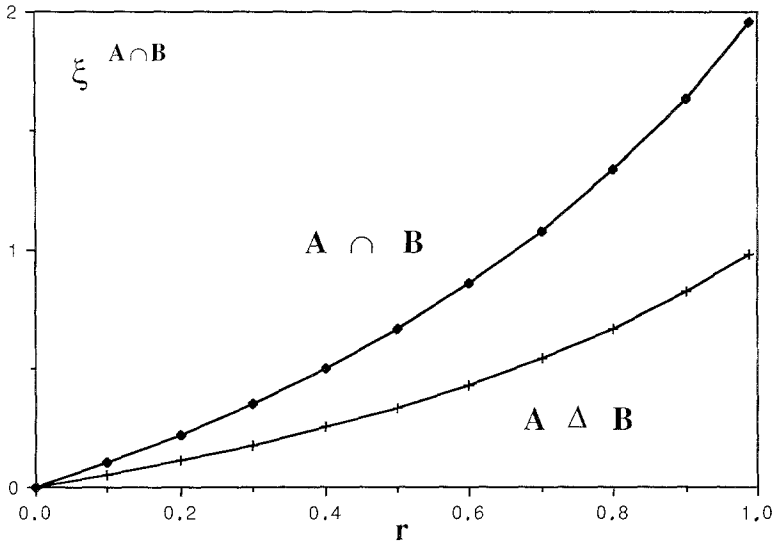


(a)



(b)

Fig. 11. Input level for the neurons of the three groups $A \cap B$, $A - B$, and $B - A$ as a function of the overlap ratio of the patterns ξ^A and ξ^B for three different initial conditions. Perfect coherence is assumed: (a) Initial condition ξ^A. (b) Initial condition ξ^{A-B}. (c) Initial condition $\xi^{A \cap B}$.



(c)

Fig. 11 (*continued*)

In the opposite case where the patterns ξ^A and ξ^B display no coherence on their overlap, that is, when there is no correlation between the states of the neurons of the prototype patterns ξ^A and ξ^B in their overlapping part $A \cap B$, the results are quite different (see Table I): Inference never takes place and the neurons of A all receive the same input, whether they belong to $A \cap B$ or $A - B$. Thus, this simple example clearly illustrates some of the basic prerequisites for inference to occur: low threshold, large overlap, and high coherence.

We shall now study the inferential properties of our model. Before investigating the case where small patterns are stored, we first discuss briefly the intermediate case presented in Section 3. In the saturation regime, $m \cong 0.4$ and $m^{\pm} \cong 0.6$ (see Fig. 2a). This means that about 80% of the neurons outside the support of the initial condition leave the null state during the system's evolution, while about 40% of the initial information is altered in the same time. The system thus displays two unwelcome features from the standpoint of inference: (1) Excitaton of neurons outside the support of the initial condition is not hindered by the lack of coherence between prototype patterns. (2) Most initial conditions are badly preserved ($m^{\pm}$ significantly differs from 1).

To further emphasize this point, we have plotted on Fig. 12 the proportion of well-retrieved patterns (retrieval quality better than 98%)

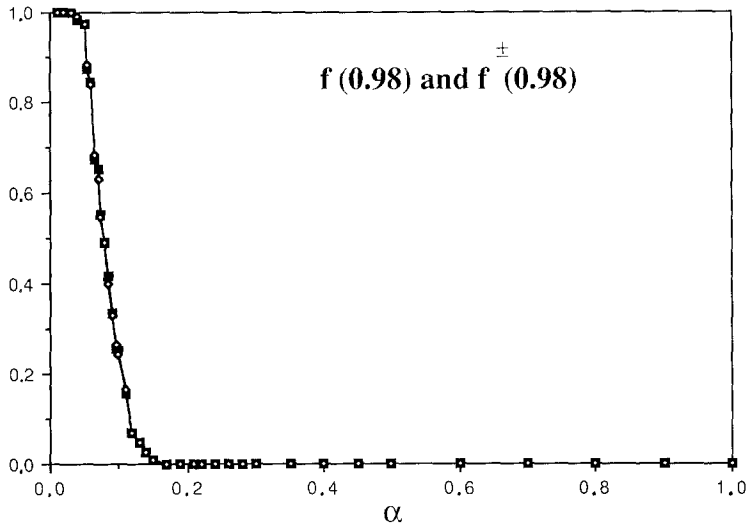


Fig. 12. Average proportion of well-retrieved patterns versus α. The graphs of $f(0.98)$ (retrieval quality on the whole network; solid squares) and $f^{\pm}(0.98)$ (retrieval limited to the support; open diamonds) are perfectly superimposed.

taking into account either the whole network (f) or just the support of the initial condition ($f^{\pm}$). The two curves are perfectly superimposed. This proves that the initial information is well retained only for those prototype patterns which do not excite any other pattern during the system's evolution: Preservation of the initial information cannot be achieved if inference takes place.

On the contrary, when small patterns are stored, the initial information is well preserved while the patterns which have a coherent overlap are excited. This is illustrated by the following analysis [where $P(0)=0.9$, that is, the average size of patterns is $N/10$].

Consider first small patterns with a perfectly coherent overlap (a situation which should also be amenable to an analytic study). To achieve such a coherence of the prototype patterns we randomly choose them according to (2.1) with $P(-1)=0$, $P(0)=0.9$, and $P(1)=0.1$; then $\xi_i^\mu=1$ on the support A_μ of the pattern ξ^μ. The behavior of the system is displayed on Fig. 13: When α is increased, $m^{\pm}$ remains close to 1, whereas m sharply drops off. This means that the initial condition is almost wholly retained while many neurons not belonging to its support are excited. Notice also that the retrieval quality on the support $m^{\pm}$ is slightly improved when α is increased as a consequence of the coherent overlap of patterns which increases their effective storage weight.

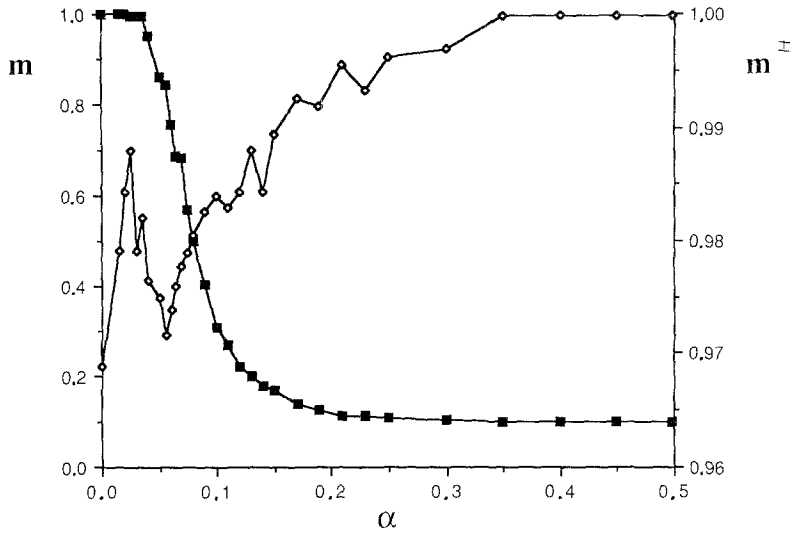
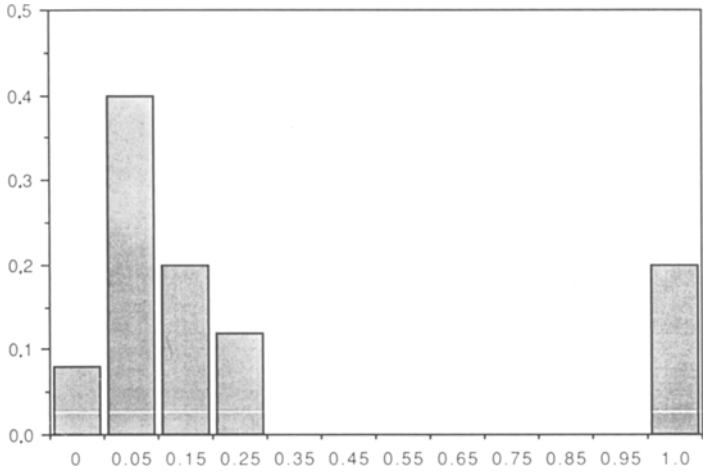
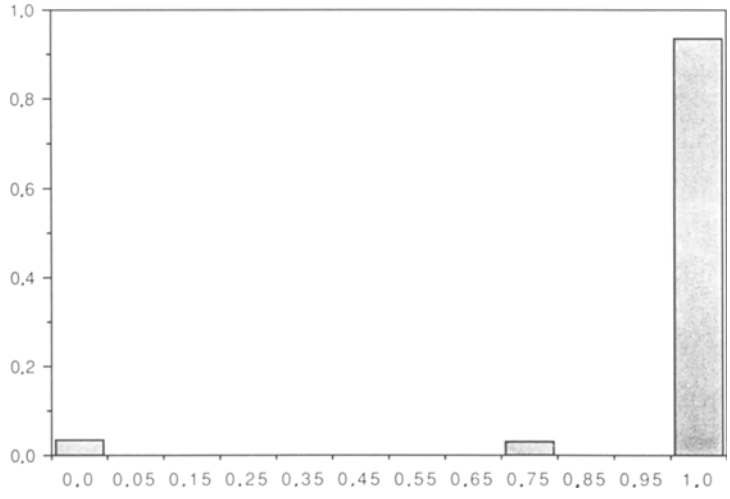


Fig. 13. Dependence on α of the average retrieval qualities m (solid squares) and $m^{\pm}$ (open diamonds) in the case where small patterns $[P(0)=0.9]$ are stored.

Figure 14 displays the histograms of excitation levels $e_{\mu;\nu}$ for a typical (randomly chosen) set of patterns $\{\xi^{\mu}\}$. When α is small (Fig. 14a) the prototype patterns are stable: The initial information is perfectly preserved during the dynamics ($e_{\mu;\mu}=1$), but no significant inference takes place (the excitation levels $e_{\mu;\nu}$, $\nu \neq \mu$, are small and just correspond to the overlaps of patterns). The situation is quite different for higher α (Fig. 14b): Excitation levels are high and the category $e_{\mu;\nu}=1$ regroups most of them (few intermediate values are observed). This proves that when the average overlap is large enough, most prototype patterns are no longer stable and other patterns are wholly excited during the system's evolution. We have also noticed that if the information initially present in the pattern ξ^{μ} is not conserved ($e_{\mu;\mu}=0$), then this pattern decays to the null state ξ^{0} and does not excite any other prototype patterns during its evolution; however, it can be excited during the evolution of another pattern ξ^{ν} ($e_{\mu;\nu}\neq 0$). More generally, the fact that a pattern ξ^{ν} is excited during the evolution of the initial condition ξ^{μ} does not entail that the pattern ξ^{μ} is excited during the evolution of ξ^{ν} ($e_{\mu;\nu}\neq e_{\nu;\mu}$); this stems from the basic asymmetry of the dynamics of patterns. It seems now appropriate to discuss the nature of the spurious states. In the Hopfield–Little model the retrieval quality degrades for high α as spurious states originating in the similarity between prototype patterns become attractors of the dynamics. These unwanted spurious states, which limit the system's storage capacity, are mixtures of prototype

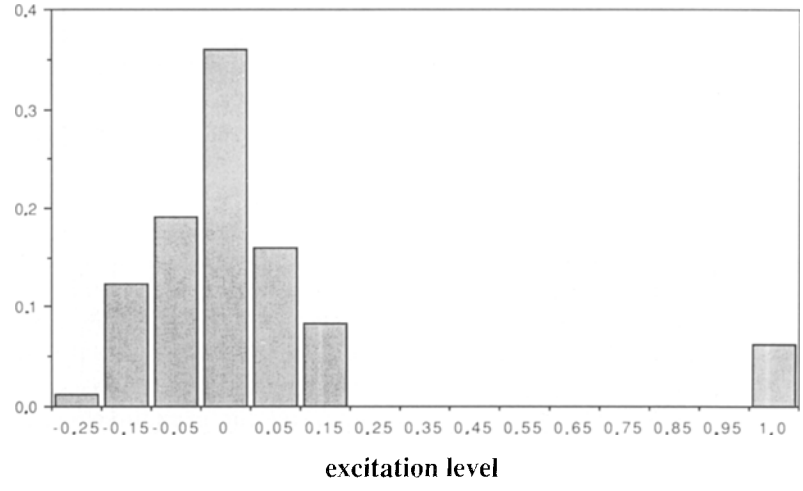


(a)

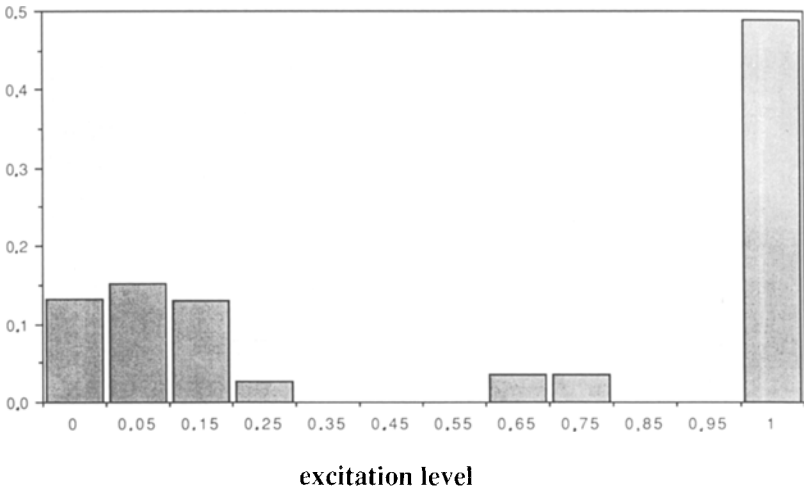


(b)

Fig. 14. Histograms of the excitation levels $e_{\mu;\nu}$ in a situation of perfect coherence. For each initial condition ξ^{ν}, we computed the excitation level $e_{\mu;\nu}$ of the pattern ξ^{μ} at the end of the system's evolution. We then classified the $(\alpha N)^2$ excitation levels thus obtained in several categories and plotted for each category the proportion of $e_{\mu;\nu}$ in that category. Each category corresponds to an interval of width 0.1 centered on the value indicated, except the categories $e_{\mu;\nu}=0$ and $e_{\mu;\nu}=1$: (a) $\alpha=0.025$ (retrieval regime). (b) $\alpha=0.135$ (inferential regime).



(a)



(b)

Fig. 15. Histograms of the excitation levels $e_{\mu;\nu}$ for $\alpha = 0.075$ and two different levels of coherence: (a) Total incoherence, $P(-1) = P(1)$. (b) Total coherence, $P(-1) = 0$.

patterns.[20,21,49] The same phenomenon is observed in the intermediate case we discussed above. On the other hand, when one stores small coherent patterns, the "spurious" attractors consist in mergings of overlapping patterns: the "typical" spurious state ξ, which originates in the coherent overlap of the prototype patterns $\xi^{\mu_1},..., \xi^{\mu_r}$, has support $A = \bigcup A_{\mu_j}$, $j = 1,..., r$, and is defined by $\xi_i = \xi_i^{\mu_j}$ for $j = 1,..., r$ and i in A_{μ_j}. Such spurious attractors, which actually exist as long as the coherence between patterns is strong, as will be shown below, thus play a *positive* role in our system's behavior, since they are responsible for the inferential properties.

We now examine the dependence of the inferential properties on the coherence between the prototype patterns. Figure 15 displays the histograms of excitation levels in a typical (randomly chosen) situation. Apart from the category $e_{\mu;\mu} = 1$, which corresponds to the conservation of the initial information during the dynamics, all the excitation levels are roughly symmetrically distributed around 0 in the incoherent case (see Fig. 15a); no inference takes place and one just observes overlap effects. The negative values of $e_{\mu;\nu}$ corresponds to slight excitations by overlap effects of the symmetric counterparts $\{-\xi^\mu\}$ of the prototype patterns $\{\xi^\mu\}$. On the other hand, when coherence is perfect, about 50% of the excitation levels fall into the category $e_{\mu;\nu} = 1$, which is the sign of strong inference; note also that quite naturally only the prototype patterns are

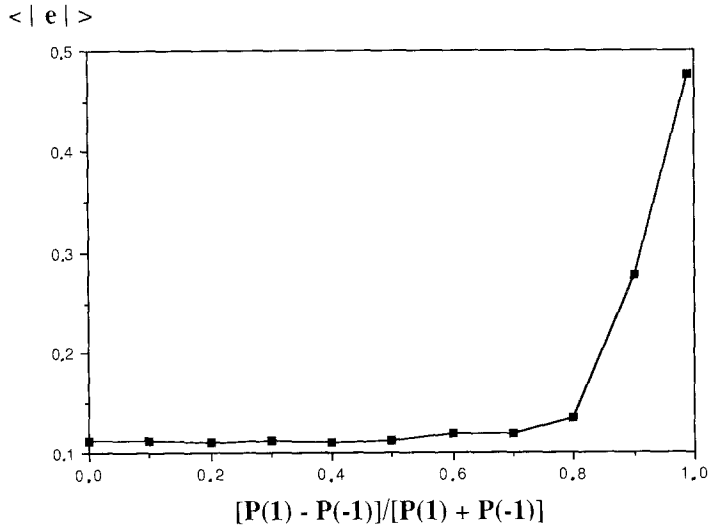
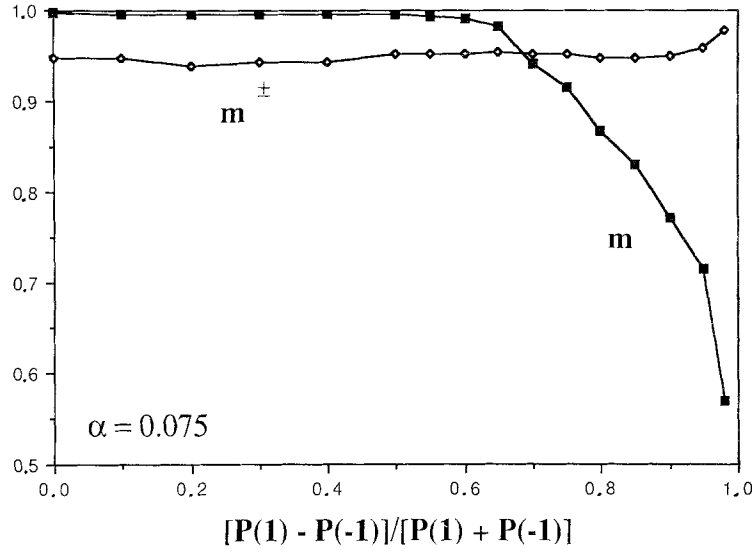
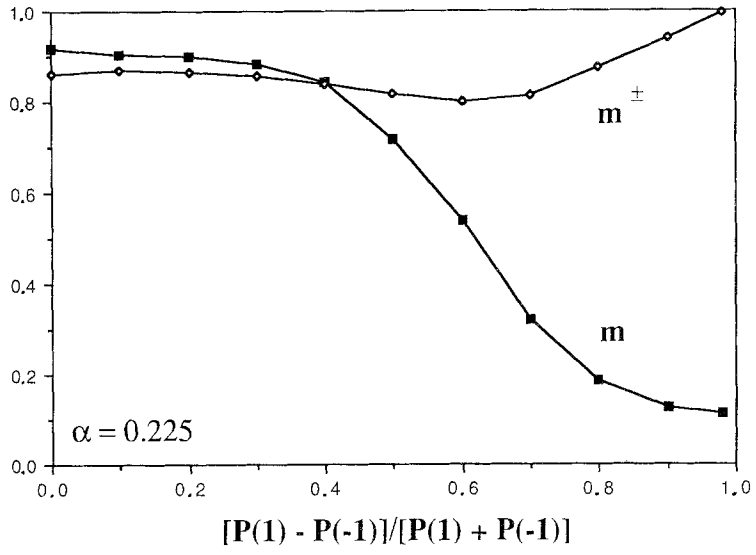


Fig. 16. Mean value $\langle |e| \rangle = 1/(\alpha N)^2 \sum |e_{\mu,\nu}|$ of the absolute magnitudes of the excitation levels versus the coherence level of the patterns $[P(1) - P(-1)]/[P(1) + P(-1)]$; here $\alpha = 0.075$.

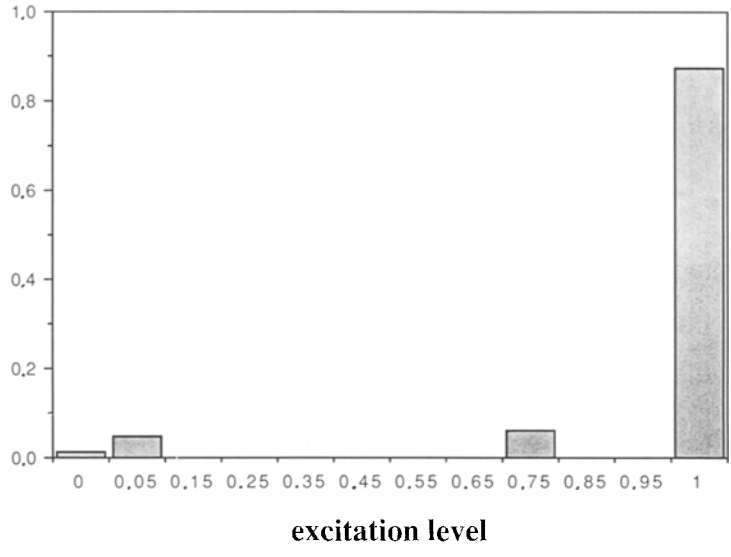


(a)

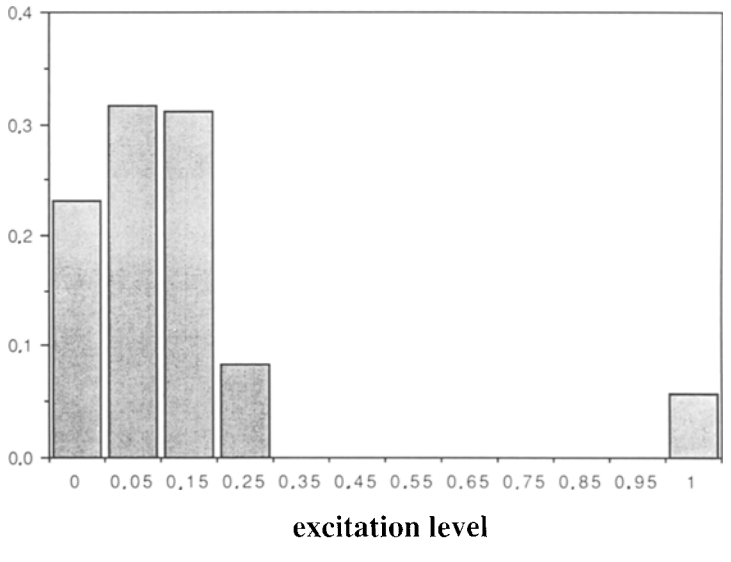


(b)

Fig. 17. Dependence of the average retrieval qualities m (solid squares) and $m^{\pm}$ (open diamonds) on the level of coherence $[P(1)-P(-1)]/[P(1)+P(-1)]$: (a) $\alpha=0.075$, (b) $\alpha=0.225$.



(a)



(b)

Fig. 18. Histograms of excitation levels for $\alpha = 0.075$, perfect coherence, and two different values of the threshold: (a) $U = 0.25$, (b) $U = 0.75$.

excited and not their symmetric counterparts (the excitation levels are never negative). In fact, inference requires a strong coherence of the patterns. We see on Fig. 16, where the mean $\langle |e| \rangle = 1/(\alpha N)^2 \sum |e_{\mu;\nu}|$ of the $|e_{\mu;\nu}|$ has been plotted as a function of the coherence level {which is measured by the quantity $[P(1)-P(-1)]/[P(1)+P(-1)]$} that inference remains weak as long as the coherence level does not exceed 80% and sharply increases beyond this value. The retrieval qualities clearly display the same behavior (see Fig. 17). For values of α corresponding to the retrieval regime of the incoherent case (see Fig. 17a) the average retrieval quality remains quite constant (no inference) for coherence levels lower than 50% and then sharply drops off (inference); on the other hand, the initial information is always well retained and its stability even improves at high coherence levels. When α is increased, that is, for values which correspond to the beginning of the saturation regime in the incoherent case (see Fig. 17b), the behavior remains qualitatively similar; however, the decrease of m is much smoother and inference appears more gradually than before. Notice here the conspicuous improvement in the stability of the initial information at high levels of coherence.

We now consider the last factor that affects the inferential properties: the value of the threshold. As expected, the higher the threshold, the more inference is inhibited, as illustrated by Figs. 18a and 18b, which display two

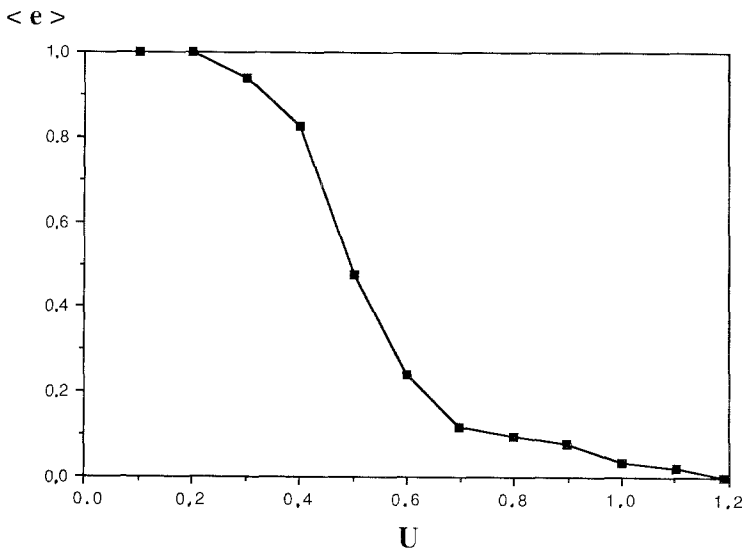


Fig. 19. Average value $\langle e \rangle = 1/(\alpha N)^2 \sum e_{\mu;\nu}$ of the excitation levels $e_{\mu;\nu}$ as a function of the threshold U in a situation of perfect coherence; here $\alpha = 0.075$.

cases of perfect coherence differing just by the value of the threshold. As shown by Figs. 19 and 20, the effect of the threshold is more gradual than was the effect of coherence. Moreover, if a high threshold inhibits inference as did a low coherence, it also drastically affects the stability of the initial information (see Fig. 20). Finally, we display on Fig. 21 the average retrieval qualities as a function of α: Notice that for low threshold (Fig. 21a) inference occurs even for small average overlap of the prototype patterns.

Thus, we see that our three-state neuron network displays interesting inferential properties when small patterns are stored: The information contained in the initial condition (one of the prototype patterns in our study) is preserved during the system's evolution, but new information, not present in the initial condition, may be brought to light if it has a high level of coherence with the information initially present. Moreover, the inference, which originates in the positive role played here by the spurious states, may be controlled to some extent by the value of the threshold.

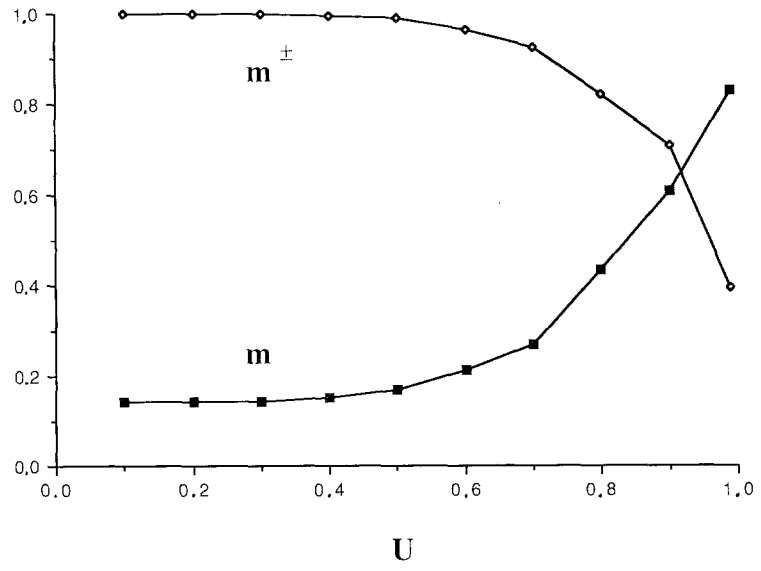


Fig. 20. Dependence on the threshold U of the average retrieval qualities m (solid squares) and $m^{\pm}$ (open diamonds) for $\alpha = 0.075$ and a perfect coherence between patterns.

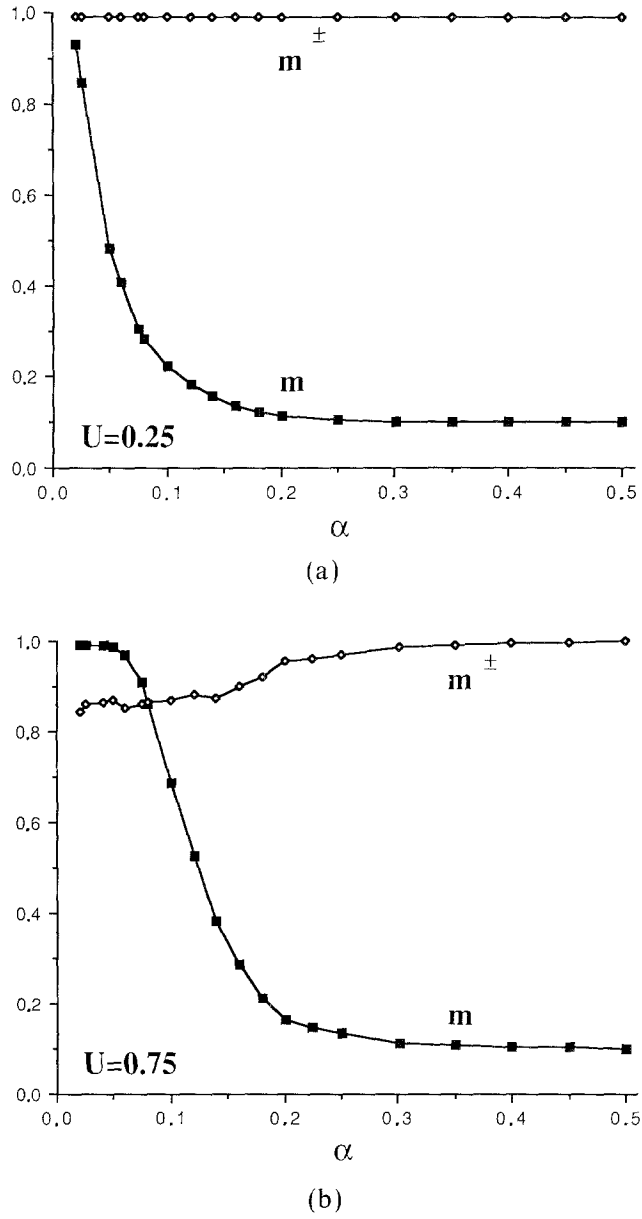


Fig. 21. Variations with α of the average retrieval qualities m (solid squares) and $m^{\pm}$ (open diamonds) for perfect coherence $[P(-1)=0]$ for two different values of the threshold: (a) Low threshold $U=0.25$. (b) High threshold $U=0.75$.

6. CONCLUSION: THE INFORMATION PROCESSING CAPABILITIES

In this section we summarize the information processing capabilities of our network and discuss their possible interest in certain AI problems. Let us recall that connectionist architectures have already been used in the representation of knowledge (knowledge networks), in natural language understanding (word disambiguation), vision research,[50] and expert systems.[51] From the standpoint of information processing the basic features of our model may be described as follows:

1. Each neuron corresponds to an elementary fact which can be true $(+1)$, false (-1), or undefined (0). The third state, 0, embodies the irrelevance of the elementary fact in a given context as well as the absence of significant evidence on its truth or falsehood (trivalued logic). The whole network may be thought of as the universe of all the elementary facts pertaining to the description of a given problem.

2. A prototype pattern is a group of mutually correlated facts. Generally it contains information relative to only part of the network. One may think of the patterns as factual statements, categories, rules, or fragments of a picture. The term rule should be used here with caution, as causality is implicit rather than explicit; however, such is often the case when one derives empirical rules from the observations of past associations of facts (in many situations of medical diagnostics, for instance).

3. The interrelationship among neurons is specified by the connection matrix C, which indicates the correlation between any two elementary facts: if $C_{ij}=0$, the facts i and j are not directly related; if $C_{ij}>0$, they are correlated (the correlation being proportional to the magnitude of C_{ij}), whereas they are anticorrelated for $C_{ij}<0$. The connection matrix is constructed from the prototype patterns, that is, from the patterns or rules initially prescribed. Once the connections are specified, the network becomes the analog of a base of knowledge. Notice that due to the symmetry $S \leftrightarrow -S$, which gives naturally symmetric roles to the logical values "true" and "false," the knowledge base contains with each category ξ^{μ} its opposite counterpart $-\xi^{\mu}$.

4. Information processing occurs during the system's evolution: Starting from some (partial) initial information, the system evolves during successive cycles, extracting new information (transition of neurons from 0 to ± 1 between the initial and final states). During this process some initial facts (neurons initially in the $+1$ or -1 states) may set to the undefined state 0 if they are not supported by enough additional evidence or even flip

if they prove in contradiction with the context [i.e., the current state $S(t)$ of the network].

Let us discuss more fully this dynamical behavior, which is somehow analogous to the processing of data by an inference engine in a classical expert system. The prescriptions for the dynamics are extremely simple in the present case. No explicit selection of rules, activation of rules, or conflict resolution strategy is required; the complex dynamical behavior just appears as an emergent property of the system grounded on the response function of elementary neurons. This response function may be rephrased as follows: any neuron changes its state so as to fit a best in the context and adopt the undefined state 0 in case of ambiguity. This is done by updating the state of the elementary neuron depending on the information it receives from the neurons to which it is directly connected. The neuron sets to the null state 0 if the input is low (with respect to the threshold U), that is, if it receives information from too few neurons or it receives contradictory information. If it receives a large and meaningful information from the other neurons, it sets to a nonzero state. The threshold U thus appears as a certainty criterion: It is the minimal amount of external evidence which must be supplied to assign a definite logical value to the basic fact one considers.

Before turning to the possible practical interest of such networks, we would also like to emphasize the two following points:

1. The processing of information is simple and meaningful, as was illustrated above by the response function of individual neurons. Let us give another example to support this assertion: the evolution of an initial condition corresponding to part of a prototype pattern. If the initial condition contains little information or the threshold is too high, the system evolves to the null state ξ^0. This stable null state ξ^0 corresponds to a total lack of information. Such a situation is highly desirable: the system does not validate the initial information if it is too uncertain to lead to a significant conclusion. If the initial information content is higher, the whole prototype pattern will be recovered. Finally, if it overlaps other patterns, inference will take place, for small prototype patterns, provided that the coherence of the prototype patterns is sufficient. This is reasonable, as no inference effects should ever occur between uncorrelated bodies of data. That inferential behavior appears when the number of prototype patterns and the coherence are increased via what we could perhaps call with great caution a percolative transition (see Section 5).

2. The system's dynamics is highly parallel: All neurons are updated synchronously (Little dynamics), that is, all facts are examined at the same time; many patterns may become activated during the same dynamical

cycle. This feature is one of the strong points of connectionist systems, as it allows for a quick processing of large bodies of data, and can be opposed, for instance, to the sequential activation of rules in classical expert systems.

We now discuss some possible applications of such three-state networks, beginning with the image processing capabilities. The analysis of Section 4 shows that an entire prototype pattern can be retrieved even if the initial condition contains no more than 15% of the information relative to that pattern. This strong pattern completion capability does not impair the ability to retrieve a prototype pattern from a noisy version, which is similar to that of the Hopfield network. In addition, the analysis of the system's dynamics when small overlapping patterns are initially stored suggests an interesting possibility: the implicit storage of a pattern stemming from the coherent overlap of prototype subpatterns. Suppose one wants to retrieve a given collection of large patterns. One can consider each pattern as the merging of small subpatterns with coherent overlap, the same subpattern being possibly part of several large patterns. If the set of subpatterns which are stored is adequately chosen, one can hope to retrieve entirely any given pattern by feeding the network with any meaningful initial condition. Such an initial condition could, for instance, be made of one subpattern belonging to that pattern and possibly to some other patterns together with some additional information on subpatterns specific to the pattern to be retrieved. Of course, the correct implementation of this idea would require more sophisticated schemes than the basic network we present here.

That inferential behavior could also be of interest for connectionist expert systems. In the framework of zero-order logic, each prototype pattern corresponds to a rule. Starting from some initial condition (some initial body of data), new information is brought to light as successive rules are activated during the system dynamics until the system reaches a stationary state. A few comments on the network's behavior in such a framework are now necessary:

1. We have seen in Section 4 that generally whole patterns are excited during the inference process; this allows us to speak of rules being activated.

2. Each dynamical cycle corresponds to the processing of a given body of data and the activation of the relevant rules. These rules are activated at the same time and not sequentially (parallelism of the dynamics).

3. The infomation processing displays many welcome features: for

instance, the system does not conclude if the initial body of data is too small or contradictory (see above).

4. The logical value of an elementary information may sometimes change several times during the processing, as the current context [the state $S(t)$ of the network] evolves, before finally stabilizing; this, together with the existence of the validation threshold U, avoids the appearance of inner contradictions in the body of data during the processing.

5. The system is not limited to one peculiar type of inference: backward as well as forward or mixed inference can take place within the network.

6. Our system naturally involves two different hierarchical levels in the inferential regime: the basic set of subpatterns which are known *a priori* and the implicitly stored patterns (spurious attractors) which are naturally obtained via the inference process.

7. It displays strong robustness properties as is usual with neural networks. The information processing capabilities gracefully degrade when one adds to the base of knowledge some rules which contradict the other rules and thus decrease the coherence of the system. In addition, some inner contradictions can be present in the initial condition; the system will then quickly set the corresponding part of the data to the undefined state 0 and process the remainder of the initial data. These are strong points when compared to the brittleness problems faced by many classical expert systems.

8. The high speed of processing due to the parallelism might make neural expert systems better suited than classical expert systems to the real-time processing of information. Moreover, the number of neurons and the number of rules with which one can realistically hope to deal in hardware implementation of neural networks is no limitation: the biggest classical expert systems make use of a few thousands rules and most expert systems are limited to a few hundred rules; in comparison, 10^8–10^{10} connections can be *a priori* stored in a 1 cm^3 volume hologram.[15]

9. The model we have presented here is of course very far from performing realistic information processing on a given problem; nevertheless, we think that such three-level schemes could be of interest in the devising of neural expert systems. Actually, many refinements should be added to improve the information processing, such as the use of different values of the threshold for the different neurons (which would correspond to more or less stringent validation criteria), dynamical changes of the threshold depending on the context, clamping of the neurons which correspond to perfectly known elementary facts, dialog with the user during a consulation

session (input of new information at the system's request when the system cannot conclude or when new information is available in some real-time tasks), and introduction of a finite temperature (this would smooth the average response function of neurons and give a fuzzier and more realistic behavior for input values close to the threshold). Moreover, much attention should also be paid to the correct representation of knowledge in the network (choice of coherent rules with sufficient overlap in the construction of the base of knowledge, choice of appropriate storage weights for the different rules, introduction of asymmetry in the connections to take explicit causality into account at the level of rules).

As a final word, we shall briefly comment on some basic differences with the Hopfield–Little model. Our system displays a complex dynamics of patterns with interesting properties in the collective regime (inference), whereas the Hopfield model's useful regime is the individual regime where prototype patterns can be perfectly retrieved. This originates in the positive role that "spurious" attractors play in our model as opposed to the Hopfield network.[32] The dynamics we have studied is basically asymmetric. This stems from the fact that two given patterns have *a priori* different sizes; therefore, if two patterns do overlap, the smaller one will be more easily excited by inference than the larger one. Note also that the patterns have an implicit stability threshold which depends on their size: small patterns cannot be stable unless the threshold U is chosen low enough or the pattern is stabilized by an effect of coherent overlap with other patterns. In addition, the introduction of the additional state 0 results in the suppression of most of the frustration in the current state of the network (that is, eliminating inner contradictions in the body of data); this was impossible in the Hopfield–Little model and is one of the keys to the strong information processing capability of our network.

APPENDIX

One can easily derive a formal expression for the dynamics of patterns in the thermodynamic limit; we assume that the ratio $\alpha = p/N$ of the number of patterns to the number of neurons vanishes in this limit. For each pattern ξ^μ we introduce the "magnetization"

$$M_\mu[S] = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu S_i$$

which ranges between $-N_\mu/N$ and N_μ/N. The evolution of these quantities is given in view of (2.2) by p equations of the form

$$M_\mu[S(t+1)] = \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \left\{ \theta\left[\sum_{j\neq i} \frac{1}{N} \sum_\upsilon \xi_i^\upsilon \xi_j^\upsilon S_j(t) - U \right] - \theta\left[-\sum_{j\neq i} \frac{1}{N} \sum_\upsilon \xi_i^\upsilon \xi_j^\upsilon S_j(t) - U \right] \right\}$$

where U is the unscaled threshold introduced in Eq. (2.2). Taking into account the definition of $M_\nu(t) = M_\nu[S(t)]$, we obtain in the thermodynamic limit

$$M_\mu(t+1) = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \xi_i^\mu \left\{ \theta\left[\sum_\upsilon \xi_i^\upsilon M_\upsilon(t) - U \right] - \theta\left[-\sum_\upsilon \xi_i^\upsilon M_\upsilon(t) - U \right] \right\}$$

The self-averaging property allows us to replace in this expression the average on sites by an average on the probability distribution (2.4) of the variables ξ_i^μ, thus obtaining

$$M_\mu(t+1) = \left\langle \xi_i^\mu \left\{ \theta\left[\sum_\upsilon \xi_i^\upsilon M_\upsilon(t) - U \right] - \theta\left[-\sum_\upsilon \xi_i^\upsilon M_\upsilon(t) - U \right] \right\} \right\rangle$$

Adopting the integral representation of the Heaviside function

$$\theta(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dy \int_0^\infty dz\, e^{iy(z-x)}$$

and performing the average over the distribution of ξ, we finally obtain

$$M_\mu(t+1) = \frac{1}{i\pi} \int_{-\infty}^{\infty} dy \int_0^\infty dz\, e^{iy(z+U)} [1 - P(0)] \sin[yM_\mu(t)] \times \prod_{\nu\neq\mu} \{P(0) + [1-P(0)] \cos[yM_\nu(t)]\} \tag{A1}$$

Specializing the dynamical equation (A1) to the case of a single pattern yields

$$M_\mu(t+1) = \frac{N_\mu}{N} [\theta(M_\mu(t) - U) - \theta(-M_\mu(t) - U)]$$

and one thus recovers the stability criterion for isolated patterns of Section 2 (except for the term $-1/N$, which vanishes in the thermodynamic limit). Equation (A1) can also be used to study simple situations, such as the dynamics of two or three overlapping patterns.

REFERENCES

1. S. Grossberg, *The Adaptive Brain*, Vols. 1 and 2 (North-Holland, Amsterdam, 1987).
2. T. Kohonen, *Self-Organization and Associative Memory* (Springer-Verlag, New York, 1984).
3. F. Rosenblatt, *Principles of Neurodynamics* (Spartan, New York, 1962).
4. M. Minsky and S. Papert, *Perceptrons: An Introduction to Computational Geometry* (MIT Press, Cambridge, Massachusetts, 1969).
5. S. Amari, *Proc. IEEE* **59**:35–47 (1972).
6. S. Amari, *Kybernetic* **14**:201–215 (1974).
7. T. Kohonen, *IEEE Trans.* **C-21**:353–359 (1972).
8. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554–2558 (1982).
9. M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1986).
10. J. J. Hopfield and D. W. Tank, *Biol. Cybernet.* **52**:147–152 (1985).
11. I. Morgenstern, Spin-glasses, optimization and neural networks, in *Proceedings of the Heidelberg Colloquium on Glassy Dynamics and Optimization; Heidelberg 1986*, J. L. Van Hemmen and I. Morgenstern, eds. (Springer-Verlag, Berlin, 1987).
12. W. D. Hillis, *The Connection Machine* (MIT Press, Cambridge, Massachusetts, 1986).
13. U. Frisch, B. Hasslacher, and Y. Pomeau, *Phys. Rev. Lett.* **56**(14):1506–1508 (1986).
14. B. Kosko and C. Guest, Optical bidirectional associative memories, *Proc. SPIE: Image Understanding* **1987**:758.
15. D. Psaltis, D. Brady, X. Gu, and K. Hsu, Optical implementation of neural computers, in *Optical Processing and Computing*, H. Arsenault, ed. (Academic Press, New York, 1988).
16. E. Domany, R. Meir, and W. Kinzel, *Europhys. Lett.* **2**(3):175–185 (1986).
17. T. J. Sejnowski and C. R. Rosenberg, *Complex Syst.* **1**:145–168 (1987).
18. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning internal representation by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructures of Cognition* (MIT Press, Cambridge, Massachusetts, 1986).
19. P. Rùjan and M. Marchand, Learning by activating neurons: A new approach to learning in neural networks, Preprint Institut für Festkörperforschungsanlage der Kernforschungsanlage Jülich (1988).
20. D. J. Amit, The properties of models of simple neural networks, in *Proceedings of the Heidelberg Colloquium on Glassy Dynamics and Optimization; Heidelberg 1986*, J. L. Van Hemmen and I. Morgenstern, eds. (Springer-Verlag, Berlin, 1987).
21. D. J. Amit, H. Gutfreund, and H. Sompolinski, *Ann. Phys. (N.Y.)* **173**:30–67 (1987).
22. K. Y. M. Wong and D. Sherrington, *Europhys. Lett.* **7**(3):197–202 (1988).
23. W. S. MacCulloch and W. Pitss, *Bull. Math. Biophys.* **5**:115–133 (1943).
24. D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
25. J. S. Denker, *Physica D* **22**:216–232 (1986).
26. H. Sompolinski, The theory of neural networks: The Hebb rule and beyond, in *Proceedings of the Heidelberg Colloquium on Glassy Dynamics and Optimization; Heidelberg 1986*, J. L. Van Hemmen and I. Morgenstern, eds. (Springer-Verlag, Berlin, 1987).
27. J. L. Van Hemmen, D. Grensing, A. Huber, and R. Kühn, *J. Stat. Phys.* **50**(1/2):231–293 (1988).
28. W. A. Little, *Math. Biosci.* **19**:101 (1974).
29. J. F. Fontanari and R. Köberle, *J. Phys. France* **49**:13–23 (1988).
30. B. Derrida, E. Gardner, and A. Zippelius, *Europhys. Lett.* **4**(2):167–173 (1987).

31. W. Kinzel, Neural networks with asymmetric bonds, in *Proceedings of the Heidelberg Colloquium on Glassy Dynamics and Optimization; Heidelberg 1986*, J. L. Van Hemmen and I. Morgenstern, eds. (Springer-Verlag, Berlin, 1987).
32. G. Parisi, *J. Phys. A: Math. Gen.* **19**:675–680 (1986).
33. P. Peretto, *J. Phys. France* **49**:711–726 (1988).
34. A. Canning and E. Gardner, Partially connected models of neural networks, *J. Phys. A*, submitted.
35. E. Gardner, *J. Phys. A: Math. Gen.* **21**:257–270 (1988).
36. M. Mézard and M. A. Virosoro, *J. Phys. France* **46**:1293–1307 (1985).
37. M. Mézard, J. P. Nadal, and G. Toulouse, *J. Phys. France* **47**:1457–1462 (1986).
38. J. P. Nadal, G. Toulouse, J. P. Changeux, and S. Dehaene, *Europhys. Lett.* **1**(10):535–542 (1986).
39. M. V. Feigelman and L. B. Ioffe, *Int. J. Mod. Phys. B* **1**:51–68 (1987).
40. L. Personnaz, I. Guyon, and G. Dreyfus, *J. Phys. Lett.* **46**:359–365 (1985).
41. T. Kohonen and M. Ruohonen, *IEEE Trans. Comput.* **C22**:701 (1973).
42. L. Personnaz, I. Guyon, and G. Dreyfus, *Europhys. Lett.* **4**(8):863–867 (1987).
43. H. Gutfreund, *Phys. Rev. A* **37**(2):570–577 (1988).
44. N. Parga and M. A. Virasoro, *J. Phys. France* **47**:1857–1864 (1986).
45. D. J. Amit, H. Gutfreund, and H. Sompolinski, *Phys. Rev. A* **35**(5):2293–2303 (1987).
46. J. J. Hopfield, D. I. Feinstein, and R. G. Palmer, *Nature* **304**:158–159 (1983).
47. S. W. Kuffler and J. Nicholls, *From Neuron to Brain* (Sinauer Associates, Sunderland, Massachusetts, 1967).
48. J. Yedidia, private communication.
49. E. Gardner, *J. Phys. A: Math. Gen.* **19**:1047 (1986).
50. D. H. Ballard, *Artificial Intelligence* **22**:235–267 (1984).
51. Y. Lecun, in *Proceedings Cognitiva 85* (CESTA-AFCET, 1985), p. 599.