

## Enlarged basin of attraction in neural networks with persistent stimuli

A. Engel

*Sektion Physik der Humboldt Universität, Bereich 04, Invalidenstrasse 42,  
1040 Berlin, German Democratic Republic*

M. Bouten, A. Komoda, and R. Serneels

*Limburgs Universitair Centrum, Universitaire Campus, 3610 Diepenbeek, Belgium*

(Received 23 April 1990)

The basins of attraction of extremely diluted neural-network models in the presence of external neural stimuli parallel to the starting configuration are calculated analytically. For moderate values of the storage capacity  $\alpha$ , the basins of attraction can be enlarged significantly. For larger values of  $\alpha$ , the patterns are still locally stable but become dynamically blocked by the external stimuli so that the effective storage capacity decreases. The performance can be improved further by allowing for time-dependent stimuli.

### I. INTRODUCTION

Models of attractor neural networks as studied by statistical physicists are systems of  $N$  binary formal neurons  $S_i = \pm 1$  ( $i = 1, \dots, N$ ) interconnected via synaptic couplings  $J_{ij}$ ,  $i \neq j$  which show nontrivial emergent properties for information processing in the thermodynamic limit  $N \rightarrow \infty$ . In particular they can function as an associative memory by relaxing to one of  $p = \alpha N$  stored patterns  $\{\xi_i^\mu\}$  ( $i = 1, \dots, N$ ,  $\mu = 1, \dots, p$ ) if a noisy version  $\{S_i^0\}$  of this pattern has been presented to them as input. The central quantities of interest are the storage capacity  $\alpha_c$ , i.e., the threshold value for  $\alpha$  above which the cooperative behavior breaks down, and the basin of attraction quantifying the maximally tolerable amount of noise in the stimulus  $\{S_i^0\}$ .

Using specific learning rules  $J_{ij} = J_{ij}(\{\xi_i^\mu\})$  the storage capacity has been calculated for different statistics of the patterns  $\{\xi_i^\mu\}$  using concepts of the statistical mechanics of spin glasses.<sup>1-7</sup> The only approximation involved here is that of replica symmetry which can be shown to be rather accurate.<sup>8</sup> On the other hand the analytical determination of the basin of attraction involves usually less controllable approximations<sup>9-14</sup> calling for comparison with numerical simulations.<sup>15-16</sup> A notable exception is the extremely diluted model introduced by Derrida, Gardner, and Zippelius<sup>17</sup> which allows for an exact solution of the dynamics.

A major breakthrough in the field was achieved by Gardner,<sup>18</sup> who succeeded in calculating the maximal storage capacity  $\alpha_c$  as a function of the statistical properties of the patterns alone without reference to a particular learning rule. This allowed her to study the properties of networks which are optimized with respect to their storage capacity. In particular for the extremely diluted model introduced previously by Derrida, Gardner, and Zippelius one can calculate the basin of attraction analytically.<sup>19</sup> The result shows that although the maximal storage capacity  $\alpha_c$  is equal to 2, already for  $\alpha > 0.42$  the

basin of attraction becomes very small and therefore the ability of the network for correcting noisy inputs is already severely limited. It is hence necessary to look for possibilities to optimize neural networks with respect to their typical basin of attraction<sup>20</sup> or to find mechanisms that enlarge the basins of attraction for a given network.

In the present paper we show that a persistence of the stimuli enforcing the initial condition at reduced strength is a very simple but efficient mechanism to improve the typical basin of attraction of different neural network models. An associative memory always uses the information present in the stimulus (the initial condition resembling one of the stored patterns) as well as that contained in the network structure (the synaptic matrix  $J_{ij}$ ). In neural network models discussed so far both types of information are combined by enforcing the stimulus by a large external field and then switching of the field completely to allow for the relaxation process, then solely determined by the synaptic matrix. Here we investigate the dynamics of networks resulting from the interplay of the internal structure of the network and a small but persistent external field parallel to the input configuration  $\{S_i^0\}$ .

It has already been shown that external fields of this type improve the static properties of symmetric neural networks.<sup>21,22</sup> In particular they allow for high quality retrieval even if the network without fields is already in the state of confusion. Early numerical work of Kinzel<sup>23</sup> suggests that also the basin of attraction may improve.

In order to study the dynamic behavior analytically we have to confine our investigations to the extremely diluted model of Derrida, Gardner, and Zippelius. This model lacks the rather high connectivity of biological neural networks and is characterized by a lesser degree of internal feedback. Nevertheless it reproduces several properties of highly connected models rather accurately and might therefore serve as a first, interesting example for the usefulness of the proposed method.

The paper is organized as follows. In Sec. II we recall the determination of the basin of attraction for the ex-

tremely diluted model and show that the mathematical simplifications that allow a determination of the dynamics still hold when an external field is present. In Sec. III we show how the basin of attraction can be enlarged by an external field parallel to the initial configuration  $\{S_i^0\}$ . Section IV deals with a further improvement due to time-dependent external fields and reviews proposals to enlarge the basin of attraction by introducing diagonal couplings.<sup>10</sup> Finally, Sec. V contains our conclusions.

## II. THE MODEL

We consider a set of  $N$  formal neurons  $S_i = \pm 1$  which are connected by synaptic couplings  $T_{ij}$  and dilute the couplings randomly. The remaining nonbroken couplings are denoted as  $J_{ij}$ . We have

$$J_{ij} = c_{ij} T_{ij}, \quad (2.1)$$

where the  $c_{ij}$  are independent random variables with distribution

$$P(c_{ij}) = \left[ 1 - \frac{C}{N} \right] \delta(c_{ij}) + \frac{C}{N} \delta(c_{ij} - 1). \quad (2.2)$$

Hence on the average each neuron is connected in an asymmetric way with  $C$  other neurons. We will be interested in the limits  $C \rightarrow \infty$ ,  $N \rightarrow \infty$  with  $C \ll N$ . The remaining synapses are assumed to be normalized such that for every neuron  $i = 1, \dots, N$

$$\sum_{j(\neq i)} J_{ij}^2 = C. \quad (2.3)$$

The dynamics of the network is defined by parallel updating according to

$$S_i(t+1) = \text{sgn} \left[ \frac{1}{\sqrt{C}} \sum_{j(\neq i)} J_{ij} S_j(t) + h_i(t) \right] \quad (2.4)$$

where  $h_i$  denotes the negative threshold of neuron  $i$  and using the analogy with spin systems will be referred to as external field. In Sec. III this field will be assumed to be parallel to the initial configuration  $\{S_i^0\} = \{S_i(t=0)\}$ ,

$$h_i = h S_i^0. \quad (2.5)$$

The remaining synaptic couplings  $J_{ij}$  are chosen such that a given set of patterns  $\{\xi_i^\mu\}$  ( $i = 1, \dots, N$ ,  $\mu = 1, \dots, p$ ) are stable fixed points of the dynamics (2.4) with  $h_i = 0$  for all  $i$ . We will consider only the case of unbiased patterns where the  $\xi_i^\mu = \pm 1$  with equal probability.

The generic initial condition  $\{S_i^0\}$  of an association problem is assumed to have a nonzero overlap  $m_0$  with one of the patterns, for example, the first, and to be uncorrelated with all the other ones. Hence we take

$$S_i^0 = \xi_i^1 \delta_i^0, \quad (2.6)$$

where the  $\delta_i^0$  are independent random variables with distribution

$$P(\delta_i^0) = \frac{1+m_0}{2} \delta(\delta_i^0 - 1) + \frac{1-m_0}{2} \delta(\delta_i^0 + 1). \quad (2.7)$$

Having described the model we now turn to the analysis of the network dynamics. It is easy to calculate the overlap  $m(t=1)$  of the configuration  $S_i(t=1)$  resulting after one time step using the statistical independence of the  $\delta_i^0$ .<sup>19,24</sup> To do so we determine the fraction of sites for which the quantity

$$\lambda_i = \xi_i^1 \left[ \frac{1}{\sqrt{C}} \sum_{j(\neq i)} J_{ij} S_j^0 + h S_i^0 \right] \quad (2.8)$$

is positive. Using the independence of the  $\delta_i^0$  we find for large  $C$

$$\begin{aligned} \text{Prob}(\lambda_i > 0) &= \int_0^\infty d\lambda_i [2\pi(1-m_0^2)]^{-1/2} \\ &\quad \times \exp \left[ -\frac{(\lambda_i - h\delta_i^0 - m_0\Delta_i)^2}{2(1-m_0^2)} \right], \end{aligned} \quad (2.9)$$

where we have introduced the stabilities

$$\Delta_i = \frac{1}{\sqrt{C}} \xi_i^1 \sum_j J_{ij} \xi_j^1. \quad (2.10)$$

For  $C \ll N$  the stabilities at different sites are only weakly correlated and we find

$$m(t=1) = \frac{1}{N} \sum_i \text{Erf} \left[ \frac{m_0 \Delta_i + h \delta_i^0}{\sqrt{2(1-m_0^2)}} \right] \quad (2.11)$$

$$= \left\langle \left\langle \text{Erf} \left[ \frac{m_0 \Delta + h \delta^0}{\sqrt{2(1-m_0^2)}} \right] \right\rangle \right\rangle_{\Delta, \delta^0} \quad (2.12)$$

$$= \int d\Delta P(\Delta) \left[ \frac{1+m_0}{2} \text{Erf} \left[ \frac{m_0 \Delta + h}{\sqrt{2(1-m_0^2)}} \right] + \frac{1-m_0}{2} \text{Erf} \left[ \frac{m_0 \Delta - h}{\sqrt{2(1-m_0^2)}} \right] \right], \quad (2.13)$$

where  $\text{Erf}(x)$  denotes the standard error function and the distribution  $P(\Delta)$  is to be determined via (2.10) using the statistics of the patterns  $\{\xi_i^\mu\}$ .

We could iterate (2.13) if the spin variables  $S_i(t=1)$  at

different sites were again independent random variables. This is however, usually not the case since different  $S_i(t=1)$  could depend on the same  $\{S_i^0\}$ . In fact these correlations are the reason why the dynamics of fully

connected networks could not be calculated analytically beyond the first few time steps.<sup>25</sup>

On the other hand the  $S_i(t)$  become independent random variables if the network is extremely diluted, as observed by Derrida, Gardner, and Zippelius.<sup>17</sup> If the synaptic couplings are chosen according to (2.1) and (2.2) the state  $S_i(t)$  of neuron  $i$  at time  $t$  depends on a random tree of ancestors with  $t+1$  levels and average branching ratio  $C$ . The independence of the  $S_i(t)$  for all  $t$  is ensured in the absence of a field if the sites at any *given level* of this tree are different. In the presence of the fields *all sites* of the tree have to be different. Only in this case all  $S_j(t-1)$  determining  $S_i(t)$  depend on a different subset

$\{S_k^0\}$  not containing  $S_i^0$  and are therefore independent of each other and independent of  $S_i^0$ . The independence of  $S_i^0$  is essential since  $S_i^0$  determines the external field at site  $i$  for all time steps  $t$ . For large  $C$  the conditions with and without field give rise to the same bound on  $C$ , namely,<sup>26</sup>

$$\frac{\ln N}{\ln C} \rightarrow \infty \quad \text{for } N \rightarrow \infty. \quad (2.14)$$

As a simple example one might think of  $C \sim \ln N$  satisfying on the one hand (2.14) and allowing on the other hand  $C \rightarrow \infty$  for  $N \rightarrow \infty$ . For such an extremely diluted network we can therefore repeat the deviation (2.6) to (2.13) for any time step and get as basic equation

$$m(t+1) = \int d\Delta P(\Delta) \left[ \frac{1+m_0}{2} \text{Erf} \left[ \frac{m(t)\Delta+h}{\sqrt{2[1-m(t)^2]}} \right] + \frac{1-m_0}{2} \text{Erf} \left[ \frac{m(t)\Delta-h}{\sqrt{2[1-m(t)^2]}} \right] \right]. \quad (2.15)$$

The basin of attraction can be determined from the fixed point structure of this dynamical law.

We will consider two different types of networks characterized by two different distributions  $P(\Delta)$  in detail. The first is a network with constant stabilities for which  $P(\Delta)$  is a  $\delta$  function. This is motivated by the fact that for fully connected networks with constant stabilities the basin of attraction near saturation can be determined approximately<sup>13</sup> and the result is the same as for the extremely diluted system (without external field). Hence our results on the improvement of the attraction basin by an external field for an extremely diluted system with constant stabilities may serve as a first hint on the efficiency of the method for a fully connected model designed, e.g., by the projection rule. The second case is the optimized network of Gardner which allows for a maximal value of the storage capacity  $\alpha_c$ .<sup>18,19</sup>

The probability distributions for these two cases can be determined using the methods developed by Gardner<sup>18</sup> yielding the well known results<sup>11,19,27</sup>

$$P(\Delta) = \delta \left[ \Delta - \left[ \frac{1-\alpha}{\alpha} \right]^{1/2} \right] \quad (2.16)$$

and

$$P(\Delta) = \frac{1}{2} [1 + \text{Erf}(\kappa/\sqrt{2})] \delta(\Delta - \kappa) + \frac{1}{\sqrt{2\pi}} \theta(\Delta - \kappa) \exp \left[ -\frac{\Delta^2}{2} \right], \quad (2.17)$$

respectively, where the parameter  $\kappa$  in (2.17) is related to the storage capacity  $\alpha$  by the Gardner formula

$$\frac{1}{\alpha} = \int_{-\kappa}^{+\infty} Dt (t+\kappa)^2. \quad (2.18)$$

Here we have used the shorthand notation

$$Dt = \frac{\exp \left[ \frac{-t^2}{2} \right]}{\sqrt{2\pi}} dt. \quad (2.19)$$

In (2.16) and (2.17)  $\alpha$  gives the ratio of the number of stored patterns  $p$  to the average number  $C$  of synapses per neuron.

We note finally that the behavior of a fully connected network designed by some learning rule and then diluted according to (2.2) and (2.14) is rather different from what we are interested in here. In this case  $P(\Delta)$  is always Gaussian<sup>28</sup> as in the Hopfield model and the full retrieval fixed point  $m^* = 1$  at  $\alpha = 0$  decreases continuously when  $\alpha$  increases until it merges with the  $m^* = 0$  solution at  $\alpha_c$  in a second-order transition. The effect of persistent stimuli in this case is just a smoothing of the second-order transition giving rise to a nonzero (but rather small) value of the fixed point  $m^*$  also for  $\alpha > \alpha_c$  (Ref. 29) as suggested by linear response theory. In particular no unstable fixed point of the recursion relation and hence no nontrivial behavior of the basin of attraction occurs.

### III. BASIN OF ATTRACTION FOR A FIELD PARALLEL TO THE INPUT

We first consider a network with constant stabilities. From (2.15) and (2.16) we get

$$m(t+1) = \frac{1+m_0}{2} \text{Erf} \left[ \frac{m(t)\kappa+h}{\sqrt{2[1-m(t)^2]}} \right] + \frac{1-m_0}{2} \text{Erf} \left[ \frac{m(t)\kappa-h}{\sqrt{2[1-m(t)^2]}} \right] \quad (3.1)$$

with

$$\kappa = \left[ \frac{1-\alpha}{\alpha} \right]^{1/2}. \quad (3.2)$$

The fixed point structure of the dynamical law (3.1) is to be determined numerically and is shown in Fig. 1 for  $h=0$  where stable (unstable) fixed points are shown in solid (dashed) line. The full retrieval fixed point  $m^* = 1$  is stable for  $0 < \alpha < 1$  and disappears at  $\alpha = 1$ . The  $m^* = 0$  fixed point is unstable for  $0 < \alpha < 0.39$  and becomes stable for  $\alpha > 0.39$ . For  $0.39 < \alpha < 1$  the basin of attraction is

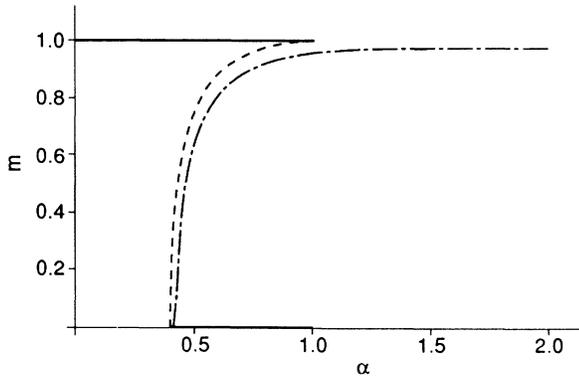


FIG. 1. Zero-field point structure for two extremely diluted models. For the network with constant stabilities the stable (unstable) fixed points are shown by the solid (dashed) line. For the Gardner diluted model only the unstable fixed points appearing for  $\alpha > 0.42$  are shown (dash-dotted line).

determined by an unstable fixed point which rapidly rises with  $\alpha$ .

The influence of the external field is demonstrated in Fig. 2(a) where the fixed points are plotted versus the field strength for  $\kappa=1$  corresponding to  $\alpha=0.5$  and an initial overlap of  $m_0=0.4$ . Two features clearly emerge. First the unstable fixed point determining the basin of attraction decreases from its value at zero field and the stable fixed point originating from the  $m^*=0$  fixed point at zero field increases, both fixed points merge at  $h=h_{\min}(\alpha, m_0)$ . Second the full retrieval fixed point  $m^*=1$  disappears at  $k=\kappa$  via a first order transition to another stable fixed point tending to  $m_0$  for  $h \rightarrow \infty$ . This fixed point corresponds to the stabilization of the initial conditions by too strong an external field. The first-order transition necessitates the existence of another unstable fixed point which hides the  $m^*=1$  attractor for  $h > h_{\max}(\alpha, m_0)$ . Therefore though locally stable the full retrieval fixed point cannot be reached but is dynamically blocked by the external field. Hence for  $h_{\min}(\alpha, m_0) < h < h_{\max}(\alpha, m_0)$  the external field enables us to reach the  $m^*=1$  fixed point from the initial overlap  $m_0=0.4$ , i.e., from far outside the attraction region at  $h=0$  hereby improving the content addressability considerably. This increase in the basin of attraction occurs however at the expense of the storage capacity since the  $m^*=1$  fixed point is already destabilized by the external field for  $h=\kappa$ , which by (3.2) corresponds to  $\alpha=1/(1+h^2) < 1$ . Increasing  $\alpha$ , thus decreasing  $\kappa$  results in larger values for  $h_{\min}$  and smaller values for  $h_{\max}$  [see Fig. 2(b)] until they merge at some value  $\alpha_{\max}(m_0)$ . For  $\alpha > \alpha_{\max}(m_0)$  the fixed point topology is qualitatively

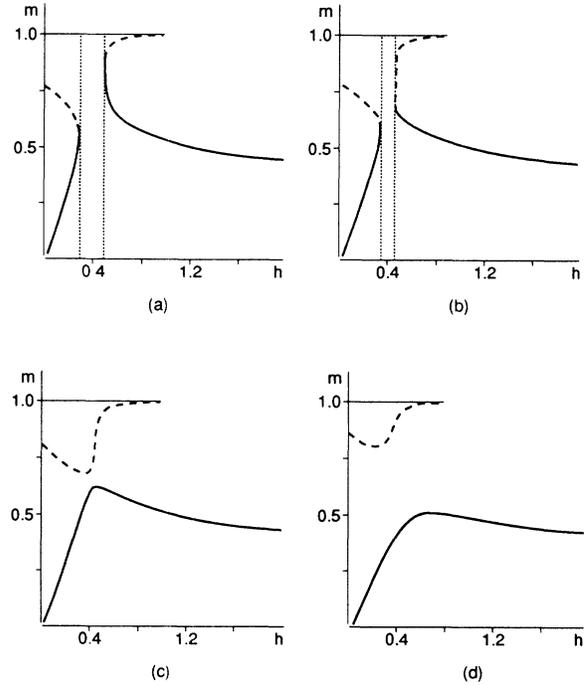


FIG. 2. Fixed point structure vs field strength for an extremely diluted model with constant stabilities for  $m_0=0.4$  for different  $\kappa$  values (a)  $\alpha=0.5$ , (b)  $\alpha=0.51$ , (c)  $\alpha=0.515$ , (d)  $\alpha=0.61$ . The solid (dashed) lines show the stable (unstable) fixed points. The field region where the dynamics yields full retrieval is bounded by dotted lines in (a) and (b).

different as shown in Figs. 2(c) and 2(d). Still the unstable fixed point decreases initially a little for small field values  $h$  but then rises with  $h$  and merges with the  $m^*=1$  fixed point at  $h=\kappa$ . The  $m^*=0$  fixed point rises with increasing field  $h$  and tends asymptotically to the value  $m^*=m_0$  corresponding to a stabilization of the starting pattern. Hence for  $\alpha > \alpha_{\max}(m_0)$  a field of the proposed type is unable to notably improve retrieval.

Figure 3 shows  $h_{\min}$  and  $h_{\max}$  versus  $\alpha$  for different values of the initial overlap  $m_0$ . Clearly the method works best for large values of  $m_0$  since then the external field acts almost parallel to the desired pattern. Nevertheless for  $m_0$  as small as 0.2 one can accomplish retrieval until  $\alpha_{\max}(0.2)=0.44$ . From Fig. 1 one infers that for this value of  $\alpha$  retrieval without field is possible only for  $m_0 > 0.6$ .

For a network designed by the optimal learning rule of Gardner the behavior is rather similar. From (2.15) and (2.17) we find instead of (3.1) and (3.2) now<sup>19</sup>

$$m(t+1) = \frac{1+m_0}{2} \left[ [1 + \text{Erf}(\kappa/\sqrt{2})] \text{Erf} \left[ \frac{m(t)\kappa + h}{\sqrt{2[1-m(t)^2]}} \right] + \int_{\kappa}^{\infty} D\Delta \text{Erf} \left[ \frac{m(t)\Delta + h}{\sqrt{2[1-m(t)^2]}} \right] \right] \\ + \frac{1-m_0}{2} \left[ [1 + \text{Erf}(\kappa/\sqrt{2})] \text{Erf} \left[ \frac{m(t)\kappa - h}{\sqrt{2[1-m(t)^2]}} \right] + \int_{\kappa}^{\infty} D\Delta \text{Erf} \left[ \frac{m(t)\Delta - h}{\sqrt{2[1-m(t)^2]}} \right] \right] \quad (3.3)$$

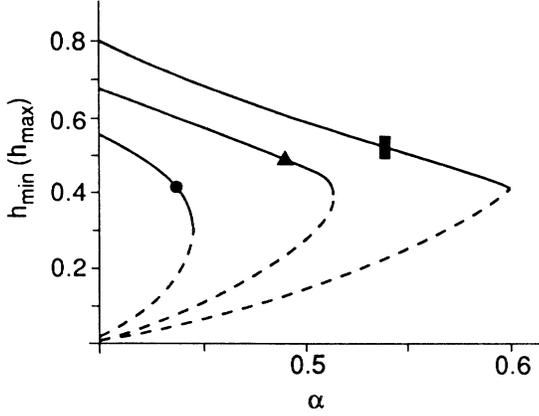


FIG. 3. Critical field bounds  $h_{\min}(m_0, \alpha)$  (---) and  $h_{\max}(m_0, \alpha)$  (—) which limit the domain of field values where the external field allows full retrieval (network with constant stabilities) for  $m_0 = 0.2$  (●),  $m_0 = 0.4$  (▲),  $m_0 = 0.6$  (■).

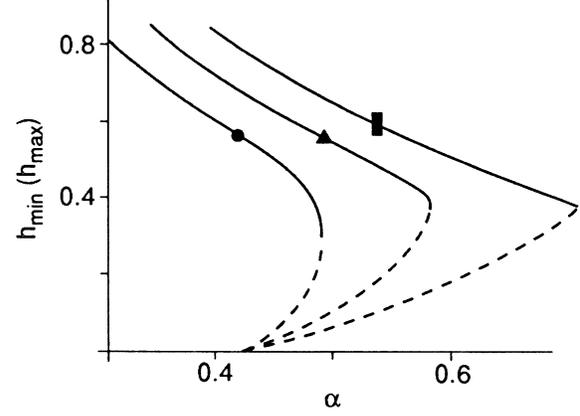


FIG. 4. Critical field bounds  $h_{\min}(m_0, \alpha)$  (---) and  $h_{\max}(m_0, \alpha)$  (—) which limit the domain of field values where the external field allows full retrieval (optimized Gardner model) for  $m_0 = 0.2$  (●),  $m_0 = 0.4$  (▲),  $m_0 = 0.6$  (■).

with  $\alpha$  related to  $\kappa$  following (2.18).

Here for  $h=0$  the full retrieval fixed point  $m^*=1$  is stable for  $\alpha < 2$ , the  $m^*=0$  fixed point is unstable for  $\alpha < 0.42$  and stable for  $\alpha > 0.42$ . Just as in the case with constant stabilities an unstable fixed point rises rapidly from  $\alpha=0.42$ . This unstable fixed point line is shown in Fig. 1 (dash-dotted line) for  $h=0$ .<sup>19</sup> Note that although the maximum storage capacity is twice that of the network with constant stabilities the basin of attraction is extremely small for  $\alpha > 0.42$ . In particular the value of  $\alpha=0.42$  where the unstable fixed point rises from  $m^*=0$  is only slightly larger than in the previous case.

The plots of the fixed points versus field strength for given values of  $\kappa$  and  $m_0$  are very similar to Fig. 2 and so are not reproduced here. Figure 4 gives a plot of  $h_{\min}$  and  $h_{\max}$  versus  $\alpha$  to be compared with Fig. 3. Even quantitatively both solutions are rather similar which suggests that the proposed procedure is rather general. For all values of  $m_0$  one can in the second case penetrate a little bit further into the region of small basins of attraction by including the external field. The optimal field strength is in all cases about 0.4.

#### IV. TIME-DEPENDENT EXTERNAL FIELDS

It is tempting to improve the performance of the network further by allowing for a time dependence of the guiding external field, where both the strength and the direction of the field could vary. An obvious choice for a changing direction is a field parallel to the configuration at time  $t$ :

$$h_i(t) = hS_i(t). \quad (4.1)$$

Then instead of (2.4) we have

$$S_i(t+1) = \text{sgn} \left[ \frac{1}{\sqrt{C}} \sum_{j(\neq i)} J_{ij} S_j(t) + hS_i(t) \right]. \quad (4.2)$$

A field of this type is equivalent to a diagonal coupling  $J_{ii} = h$  ( $i=1, \dots, N$ ). Still (2.14) is sufficient to ensure the absence of correlations between  $S_i(t)$  and  $S_j(t)$  and we find similarly to (2.15):

$$m(t+1) = \int d\Delta P(\Delta) \left[ \frac{1+m(t)}{2} \text{Erf} \left[ \frac{m(t)\Delta + h}{\sqrt{2[1-m(t)^2]}} \right] + \frac{1-m(t)}{2} \text{Erf} \left[ \frac{m(t)\Delta - h}{\sqrt{2[1-m(t)^2]}} \right] \right]. \quad (4.3)$$

Unlike (2.15) the dynamical law (4.3) possesses the fixed points  $m^*=0$  and  $m^*=1$  for all values of  $h$ . Figure 5 gives a plot of the fixed point structure versus field strength for a network with constant stabilities and different values of  $\alpha$ . The unstable fixed point determining the basin of attraction always decreases and merges at some value  $h_c(\alpha)$  with the  $m^*=0$  fixed point. This field value is given by

$$\text{Erf}(h_c/\sqrt{2}) + \sqrt{2(1-\alpha)}/\pi\alpha \exp(-h_c^2/2) = 1 \quad (4.4)$$

and for  $\alpha \rightarrow 1$  tends to infinity as  $h_c \sim [\alpha/(1-\alpha)]^{1/2}$ . Figure 6 shows  $h_c$  versus  $\alpha$ , i.e., to the left of this curve the  $m=0$  fixed point is unstable and the basin of attraction is 100%. Again the behavior of a network optimized with the Gardner prescription is very similar.

From these results it seems that diagonal couplings are very efficient for enlarging the basin of attraction even without reducing the storage capacity. This is however limited to the extremely diluted model. For a fully connected model with constant stabilities it is known from

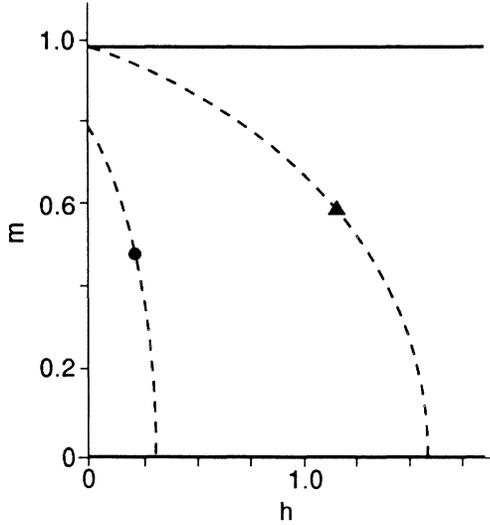


FIG. 5. Fixed point structure vs field strength for the model with constant stabilities when the field is parallel to the configuration at time  $t$ . The lines  $m=1$  and  $m=0$  are stable fixed points. The dashed lines (marked by  $\bullet$  for  $\alpha=0.5$  and by  $\blacktriangle$  for  $\alpha=0.8$ ) are unstable fixed points.

numerical simulations<sup>10</sup> that there is an optimum value for the diagonal couplings and the basin of attraction shrinks again for larger values. This is probably due to the stabilization of spurious states by the field. Hence for fully connected models the usefulness of the method is limited although a notable increase for the basin of attraction for a suitable value of  $h$  is possible.<sup>10</sup>

$$m(t+1) = \int d\Delta P(\Delta) \left[ \frac{1+m_0}{2} \operatorname{Erf} \left[ \frac{m(t)\Delta + h(t)}{\sqrt{2[1-m(t)^2]}} \right] + \frac{1-m_0}{2} \operatorname{Erf} \left[ \frac{m(t)\Delta - h(t)}{\sqrt{2[1-m(t)^2]}} \right] \right]. \quad (4.6)$$

It is impossible now to determine the basin of attraction by inspection of the fixed point structure because of the explicit time dependence contained in  $h(t)$ . Concentrating however on the situation where the field decreases to zero very slowly we can assume that  $m(t)$  is always near to the fixed point for the given value of  $h$  and may use the results of Sec. III for a qualitative analysis of the retrieval dynamics.

One always starts with  $m=m_0$  corresponding to the fixed point for  $h \rightarrow \infty$ . For  $\alpha < \alpha_{\max}(m_0)$  [cf. Fig. 2(a)]  $m(t)$  slowly increases with decreasing  $h$  and for  $h \leq h_{\max}$  relaxes to the full retrieval fixed point  $m^*=1$  where it remains when  $h$  tends to zero. Hence a slowly decreasing field allows in this case to reach the full retrieval fixed point  $m^*=1$  without knowing the values of  $h_{\min}(m_0, \alpha)$  and  $h_{\max}(m_0, \alpha)$  and is therefore advantageous as compared with the constant field studied in Sec. III. For  $\alpha > \alpha_{\max}(m_0)$  on the other hand one infers from Fig. 2(c) that after a small increase of  $m$  with decreasing  $h$  it soon starts to decrease again tending finally to  $m^*=0$  for  $h=0$ .

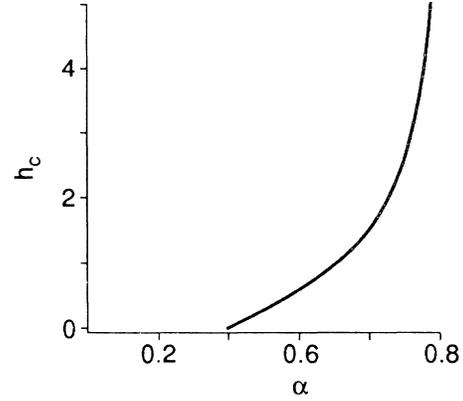


FIG. 6. Critical field value  $h_c$  for the model with constant stabilities limiting the full retrieval field values ( $h > h_c$ ) when the field is parallel to the configuration at time  $t$ .

Another interesting possibility for a time-dependent field is one parallel to the input and steadily decreasing in strength:

$$h_i(t) = h(t) S_i^0. \quad (4.5)$$

This means that the input  $S_i^0$  is enforced by a strong [ $h(0) \gg 1$ ] stimulus which then disappears slowly still guiding the relaxation process for some time. Hence the dynamics becomes only gradually more and more determined by the network structure which is reminiscent of the simulated annealing approach<sup>30</sup> to complex optimization. Similar to (2.15) we find the dynamical law

## V. CONCLUSIONS

In the present paper we have shown how the typical basin of attraction of neural network models with given synaptic matrix can be enlarged considerably by persistent external fields parallel to the input configuration  $\{S_i^0\}$ . In order to make analytical progress we had to confine our investigations to an extremely diluted network which was shown to remain exactly solvable with respect to the dynamics also in the presence of external fields of the discussed type. Nevertheless the procedure is sufficiently general and should be of value for fully connected networks too. This is of interest since most models of neural networks studied so far are optimized with respect to their storage capacity  $\alpha_c$  which gives rise to nonoptimal basins of attraction. For the optimal learning of Gardner, e.g., one has  $\alpha_c=2$  but only for  $\alpha < 0.42$  the attraction basins are large.<sup>19</sup>

Persistent stimuli are convenient tools to combine the information contained in the initial configuration  $\{S_i^0\}$  and that in the network structure  $\{J_{ij}\}$  in an efficient

way. They are easily implemented in simulation programs as well as in hardware realizations. Of course the improvement is largest if the initial overlap  $m_0$  between  $\{S_i^0\}$  and the desired pattern is large, nevertheless our investigations show that even for values of  $m_0$  as small as 0.2 a notable enlargement of the basin of attraction occurs. For an efficient use of persistent stimuli the proper choice of their intensity, i.e. the strength  $h$  of the field parallel to  $\{S_i^0\}$  is essential. As we have shown it has to lie within a definite interval  $(h_{\min}, h_{\max})$  depending both on  $\alpha$  and  $m_0$ . For  $h < h_{\min}(m_0, \alpha)$  the unstable fixed point of the dynamical recursion relation determining the basin of attraction is still larger than  $m_0$ , for  $h > h_{\max}(m_0, \alpha)$  the full retrieval fixed point though still stable cannot be reached due to a dynamical blocking by the external field. Since  $m_0$  and perhaps also  $\alpha$  are for a generic association problem not known in advance the value of  $h$  has to be fixed somewhat tentatively, a drawback which can be avoided using a field strength  $h(t)$  slowly decreasing with time (see below). Moreover our results suggest that  $h \approx 0.4$  is a suitable value for  $h$  for various values of  $m_0$  and  $\alpha$ . (cf. Figs. 3 and 4).

It should be emphasized that the described increase in the basin of attraction occurs at the expense of the storage capacity  $\alpha_c$ . Not only is the full retrieval fixed point  $m^* = 1$  already destabilized by the external field for  $\alpha < \alpha_c$ , but even being still locally stable it is blocked by the external field for too large a value of  $\alpha$ . Both effects result in a decrease of  $h_{\max}(m_0, \alpha)$  with  $\alpha$ . Since  $h_{\min}(m_0, \alpha)$  increases with  $\alpha$  there is for all values of  $m_0$  a maximal value  $\alpha_{\max}(m_0)$  of  $\alpha$  where  $h_{\min}$  and  $h_{\max}$  merge. Beyond this value of  $\alpha$  persistent stimuli are of no use to improve the network performance. It is an advantage of our dynamical approach to yield a critical storage capacity  $\alpha_{\max}(m_0)$  depending on the initial overlap  $m_0$ . A static approach only gives  $\alpha_c = \alpha_{\max}(m_0 = 1)$  a quantity with limited relevance to practical association problems only.

A simple generalization of the proposed method involves time-dependent external fields where both the direction and the strength of the field could vary. If the strength remains constant and the direction is chosen to

be always parallel to the current configuration  $\{S_i(t)\}$  the dynamics is the same as for a network with constant diagonal couplings  $J_{ii} = h$ . This choice of field seems to be very efficient for the extremely diluted model for not too large  $\alpha$  values since the full retrieval fixed point  $m^* = 1$  remains stable for all values of  $h$  and no dynamical blocking occurs. Numerical investigations of *fully* connected networks<sup>10</sup> suggest, however, that in this case diagonal couplings are of limited use. This is probably due to the creation of spurious states by the selfcouplings which act like an external field parallel to *any* configuration thereby stabilizing configurations which have not been stored.

A more interesting possibility is given by an external field always parallel to the input but of decreasing strength  $h(t)$ . If this decrease is sufficiently slow one can assume that  $m(t)$  follows adiabatically the fixed point corresponding to the current value of  $h(t)$ . From Fig. 2(a) one infers that starting from  $m(t=0) = m_0$  for  $h(t=0) \gg 1$   $m(t)$  slightly increases with decreasing  $h(t)$  as long as  $h(t) > h_{\max}(m_0, \alpha)$ . As  $h(t) = h_{\max}(m_0, \alpha)$  there is a discontinuous transition to the full retrieval fixed point  $m^* = 1$  and the external field becomes irrelevant. No *a priori* knowledge of  $h_{\max}(m_0, \alpha)$  is necessary. This is a rather attractive scenario. The relaxation process is at the beginning firmly guided by the stimulus. If this guidance is relaxed step by step and the amount of noise in the stimulus was not too high ( $m_0$  not too low) at some point the system "realizes" the similarity with a stored pattern and completes the information by its intrinsic dynamics. If, however,  $m_0$  was too small, i.e.,  $\alpha_{\max}(m_0) < \alpha$ , the overlap  $m(t)$  is always mainly determined by the field  $h(t)$  and despite a possible slight increase at the beginning, tends to zero for  $h(t) \rightarrow 0$  [cf. Fig. 2(b)].

#### ACKNOWLEDGMENTS

One of us (A.K.) is indebted to the Belgian Inter-University Institute for Nuclear Sciences (IIKW) for financial support. This research has been funded by the Belgian program on Inter-University Attraction Poles (IUAP).

<sup>1</sup>D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys. (N.Y.)* **173**, 30 (1987).

<sup>2</sup>H. Sompolinsky, *Phys. Rev. A* **34**, 2571 (1986).

<sup>3</sup>I. Kanter and H. Sompolinsky, *Phys. Rev. A* **35**, 380 (1987).

<sup>4</sup>J. L. Van Hemmen, *Phys. Rev. A* **36**, 1959 (1987).

<sup>5</sup>D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **35**, 2293 (1987).

<sup>6</sup>M. V. Feigelman and L. B. Ioffe, *Int. J. Mod. Phys. B* **1**, 51 (1987).

<sup>7</sup>N. Parga and M. A. Virasoro, *J. Phys.* **47**, 1857 (1986).

<sup>8</sup>A. Crisanti, D. J. Amit, and H. Gutfreund, *Europhys. Lett.* **2**, 337 (1986).

<sup>9</sup>W. Krauth, J. P. Nadal, and M. Mezard, *J. Phys. A* **21**, 2995 (1988).

<sup>10</sup>W. Krauth, M. Mezard, and J. P. Nadal, *Complex Systems* **2**, 387 (1988).

<sup>11</sup>T. B. Kepler and L. T. Abbott, *J. Phys. (Paris)* **49**, 1657 (1988).

<sup>12</sup>G. Kohring, *Europhys. Lett.* **8**, 697 (1988).

<sup>13</sup>M. Opper, J. Klein, H. Kohler, and W. Kinzel, *J. Phys. A* **22**, L407 (1989).

<sup>14</sup>R. D. Henkel and M. Opper, *Europhys. Lett.* **11**, 403 (1990).

<sup>15</sup>W. Kinzel, *Z. Phys. B* **60**, 205 (1985).

<sup>16</sup>B. M. Forest, *J. Phys. A* **21**, 245 (1988).

<sup>17</sup>B. Derrida, E. Gardner, and A. Zippelius, *Europhys. Lett.* **4**, 167 (1987).

<sup>18</sup>E. Gardner, *J. Phys. A* **21**, 257 (1988).

<sup>19</sup>E. Gardner, *J. Phys. A* **22**, 1969 (1989).

<sup>20</sup>E. Gardner, N. Stroud, and D. J. Wallace, *J. Phys. A* **22**, 2019 (1989).

<sup>21</sup>A. Engel, H. Englisch, and A. Schutte, *Europhys. Lett.* **8**, 393 (1989).

<sup>22</sup>D. J. Amit, G. Parisi, and S. Nicosi, *Networks* **1**, 75 (1990).

<sup>23</sup>W. Kinzel, *Z. Phys. B* **62**, 267 (1986).

<sup>24</sup>T. B. Kepler and L. T. Abbott, *J. Phys. (Paris)* **49**, 1657 (1988).

<sup>25</sup>E. Gardner, B. Derrida, and P. Mottishaw, *J. Phys. (Paris)* **48**, 741 (1987).

<sup>26</sup>R. Kree and A. Zippelius (unpublished).

<sup>27</sup>L. T. Abbott and T. B. Kepler, *J. Phys. A* **22**, 2031 (1988).

<sup>28</sup>B. Derrida and J. P. Nadal, *J. Stat. Phys.* **43**, 993 (1987).

<sup>29</sup>A. Rau and D. Sherrington, *Europhys. Lett.* **11**, 499 (1990).

<sup>30</sup>S. Kirkpatrick, C. D. Gelatt, and H. P. Vecchi, *Science* **220**, 671 (1983).