

## Bistable gradient networks. II. Storage capacity and behavior near saturation

Patrick N. McGraw and Michael Menzinger

*Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6*

(Received 23 May 2002; published 30 January 2003)

We examine numerically the storage capacity and the behavior near saturation of an attractor neural network consisting of bistable elements with an adjustable coupling strength, the bistable gradient network. For strong coupling, we find evidence of a first-order “memory blackout” phase transition, as in the Hopfield network. For weak coupling, on the other hand, there is no evidence of such a transition and memorized patterns can be stable even at high levels of loading. The enhanced storage capacity comes, however, at the cost of imperfect retrieval of the patterns from corrupted versions.

DOI: 10.1103/PhysRevE.67.016119

PACS number(s): 84.35.+i, 07.05.Mh, 42.65.Pc, 42.79.Ta

### I. INTRODUCTION

In this paper we consider the behavior at high memory loading of a Hopfield-like attractor network of  $N$  bistable elements, the bistable gradient network or BGN [1,12]. This is a sequel to an earlier paper [2] that considered the BGN at low loading. We compare the BGN with the deterministic Hopfield network (HN) [3–9], examining the storage capacity and other key properties.

To begin, we review the BGN model and establish some notation. The BGN is described by the coupled differential equations

$$\frac{dx_i}{dt} = -\frac{\partial H}{\partial x_i}, \quad (1)$$

where  $x_i$  are  $N$  continuous-valued real variables (or components of an  $N$ -dimensional state vector  $\mathbf{x}$ ) representing the outputs of the  $N$  nodes of the network, and  $H$  is the Hamiltonian

$$H = H_0 + H_{int} = \sum_{i=1}^N \left( \frac{x_i^4}{4} - \frac{x_i^2}{2} \right) - \frac{1}{2} \gamma \sum_{i,j=1}^N w_{ij} x_i x_j. \quad (2)$$

The quantities  $w_{ij}$  are the elements of a symmetric matrix of connection strengths and  $\gamma$  is a control parameter determining the overall strength of the coupling among nodes. As in the Hopfield model [3,4], the connections or synaptic weights are determined by the Hebb learning rule [10]

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu - \frac{p}{N} \delta_{ij}, \quad (3)$$

where the  $N$ -dimensional vectors  $\xi^\mu$  represent a set of  $p$  distinct *memory patterns* to be recognized by the network. We take these patterns to consist of binary elements  $\pm 1$  only, and we assign them random values, thus introducing quenched disorder. The BGN’s key difference from the HN and from most of its continuous-valued relatives [11] lies in the presence of the local quartic potential term  $H_0$  in the Hamiltonian, which renders each node bistable.

The interaction term  $H_{int}$  furnishes an input to each node given by  $h_i = \gamma \sum_{\mu=1}^p w_{ij} x_j$  so that the dynamical equation for each node is given by

$$\frac{dx_i}{dt} = x_i - x_i^3 + h_i. \quad (4)$$

The input may also be referred to as a “magnetic field” by analogy with Ising spin systems. In the absence of input each node has two stable fixed points at  $\pm 1$ , but a nonzero field shifts the two fixed points. If the critical magnitude  $h_c = 2\sqrt{3}/9 \approx 0.38$  is exceeded then the field is strong enough to overcome the potential barrier in the quartic potential and there is then only one fixed point.

We define two sets of order parameters  $m_\mu$  and  $b_\mu$  by

$$m_\mu \equiv \xi^\mu \cdot \mathbf{x} = \frac{1}{N} \sum_{i=1}^N \xi_i^\mu x_i \quad (5)$$

and

$$b_\mu \equiv \frac{1}{N} \sum_{i=1}^N \xi_i^\mu \text{sgn}(x_i). \quad (6)$$

$m_\mu$  are inner products or overlaps of the state vector with the memorized patterns, while  $b_\mu$ , the “bit overlaps,” encode information about sign agreements between the state vector and the stored patterns. For the purposes of this paper, we will for the most part be more interested in  $b_\mu$  than in  $m_\mu$ , so where there is little risk of confusion we will drop the word “bit” and simply refer to  $b_\mu$  as an “overlap.”

The degree of loading of the network’s memory can be parametrized by the ratio  $\alpha \equiv p/N$ . In the companion paper [2] we examined the behavior of the BGN in the low-loading limit  $\alpha \ll 1$ . It was shown that the network can function as an associative memory and correct sign-flip errors in a stored pattern as long as  $\gamma > \frac{1}{3}$ . We found that the attractors of the BGN’s dynamics are readily classified into three categories that are separated from each other in energy. The lowest-energy states are the *memory* or *retrieval states*, each of which corresponds to one of the memorized patterns. In addition to these there are higher-energy spurious attractors of two types. The *mixture* or *spin glass states* have partial overlaps with several patterns and thus lie close to the subspace

spanned by the memory patterns. *Uncondensed states*, which have no counterpart in the HN, are states in which none of the fields acting on the nodes are strong enough to cause sign flips and the dynamics is therefore dominated by the local potential. They have energies per node close to  $-0.25$ . The spin glass states are intermediate in energy between the memory states and the uncondensed states. In the range  $\frac{1}{3} \leq \gamma \leq 1$ , pattern recognition by the network was found to be highly selective; the input must be close to the stored pattern in order for the pattern to be fully restored. The uncondensed states are numerous (of order  $2^N$ ) and fill most of the configuration space. The memorized patterns and their basins of attraction occupy isolated valleys among these states, but these valleys expand as  $\gamma$  increases. For  $\gamma \geq 1$ , on the other hand, the behavior resembles that of the HN: there are no uncondensed states, and the memory states have large basins of attraction.

In this paper we turn to the case where  $p/N$  is of order unity. We are interested in the maximum storage capacity, or the maximum number of patterns that can be successfully stored and retrieved, as well as changes in the network's performance as this limit is approached. Earlier work with small networks [12] suggested that at least under certain conditions the BGN could store many more patterns than the HN while possessing few spurious attractors. It is known that the Hopfield storage limit of  $p \approx 0.14N$  memory patterns can be exceeded if a more complicated learning algorithm is used [9], but in the BGN case improved capacity is achieved with the familiar Hebb rule. Since the previous results [1,12] were obtained with networks much too small to be of practical interest (e.g.,  $N=5$ ), we now examine larger networks, mainly through numerical simulations. (At the end of the paper we will return briefly to the small-network case.) We find that the high-loading behavior, like that at low loading, depends strongly on  $\gamma$ . For  $\gamma \geq 1$ , a Hopfield-like first-order phase transition results in the destabilization of all memory states at a critical value of  $\alpha$ . For  $\gamma=2$  this transition occurs at  $\alpha_c \approx 0.1$ , compared to  $\alpha_c \approx 0.14$  for the HN. For  $\gamma=0.5$ , on the other hand, we find that it is possible for the stored patterns to remain stable even at loading factors of  $\alpha \approx 0.3$  and higher. Furthermore, there is no sudden blackout; instead, the performance degrades gradually as  $\alpha$  increases. The price of this high capacity is that the retrieval of the patterns from corrupted versions may be imperfect.

The remainder of the paper is organized as follows. In Sec. II we discuss in general terms the effects of crosstalk, or interference between different stored patterns. It is crosstalk that is responsible for limitations on storage capacity. We compare the effects of crosstalk in the BGN and the HN. This discussion provides a framework for interpreting our numerical results. In Sec. III we examine numerically the stability of memory patterns as a function of  $\alpha$ . We find evidence of a first-order memory blackout phase transition in the BGN at high  $\gamma$ , but not at low  $\gamma$ . In Sec. IV, we examine the effects of high loading on the basins of attraction for the memory states and on their retrieval from corrupted input. We see that increasing the loading markedly alters the energy landscape. In Sec. V we comment briefly on the behavior of smaller networks and on the relation between the previous

small-network results [1,12] and those we have obtained for larger networks. We conclude with some discussion of the BGN results and possible directions for further study. We make some conjectures concerning the performance of the BGN in the presence of stochastic noise.

## II. PATTERN RETRIEVAL AND CROSSTALK

In a given state  $\mathbf{x}$ , the input to the  $i$ th node of the network can be expressed as

$$h_i = \gamma \sum_{j=1}^N w_{ij} x_j = \gamma \sum_{\mu=1}^p \xi_i^\mu m_\mu - \gamma \frac{p}{N} x_i, \quad (7)$$

using the Hebb rule (3) and the definition (5) of  $m_\mu$ . In Ref. [2], we showed that when  $\alpha \ll 1$  there are stable retrieval states in which  $b_\nu = 1$  for one particular  $\nu$ . For a state in which the  $\nu$ th overlap is large, we can decompose the field as follows:

$$h_i = \gamma \left( \xi_i^\nu m_\nu + \sum_{\mu \neq \nu} \xi_i^\mu m_\mu - \frac{p}{N} x_i \right) = \gamma \left( S_i + C_i - \frac{p}{N} x_i \right). \quad (8)$$

We refer to the first term,  $S_i$ , as the ‘‘signal’’ term and the second term,  $C_i$ , as the ‘‘crosstalk’’ term. In the limit  $p/N \rightarrow 0$ , the mutual overlaps between different patterns is small, so that  $m_\nu \ll 1$  ( $\nu \neq \mu$ ), and the last two terms in Eq. (8) can be neglected. The signal term is then dominant and there is a stable retrieval state given by  $\mathbf{x} = \sqrt{1 + \gamma} \boldsymbol{\xi}^\nu$  with  $m_\nu = \sqrt{1 + \gamma}$ . We expect this solution to be approximately valid for small but nonzero values of  $p/N$ . For this case, the overlaps  $m_\mu$  ( $\mu \neq \nu$ ) behave as Gaussian distributed random variables with zero mean and variance  $1/\sqrt{N}$ . Accordingly, the crosstalk term  $C_i \equiv \sum_{\mu \neq \nu} \xi_i^\mu m_\mu$  in Eq. (8), being a sum of  $p$  such quantities, is a random quantity with zero mean and variance  $\sqrt{p/N}$ . The third term, which arises from the subtraction of the diagonal elements, is of order  $p/N$  and thus is generally smaller than the crosstalk term. We will neglect it for the moment.

In the absence of crosstalk, a retrieval state is not only linearly stable (i.e., stable against small perturbations) but, for  $\gamma > \frac{1}{3}$ , it is also stable against individual sign flips. The latter means that if a retrieval state is corrupted by changing the sign of one or a small number  $\ll N$  of nodes, then the dynamics will reverse the flipped sign and restore the pattern. This happens because, in the absence of crosstalk, each node experiences a field  $\gamma S_i = \gamma \xi_i^\nu m_\nu$  that has the same sign as  $\xi_i^\nu$  and for  $\gamma > \frac{1}{3}$  that field is strong enough to overcome the potential barrier of the individual node. Now consider a given node (say, the  $i$ th node) in the presence of crosstalk. The crosstalk field acting on the  $i$ th node may be either aligned with or opposed to  $x_i$ . If it is aligned, then its effect on that node is to increase the magnitude of  $x_i$ , making it larger than  $\sqrt{1 + \gamma}$ . If it is opposed to  $x_i$ , then its effect is to decrease the equilibrium magnitude of  $x_i$ . If the crosstalk term is large enough, then it may be sufficient to overcome

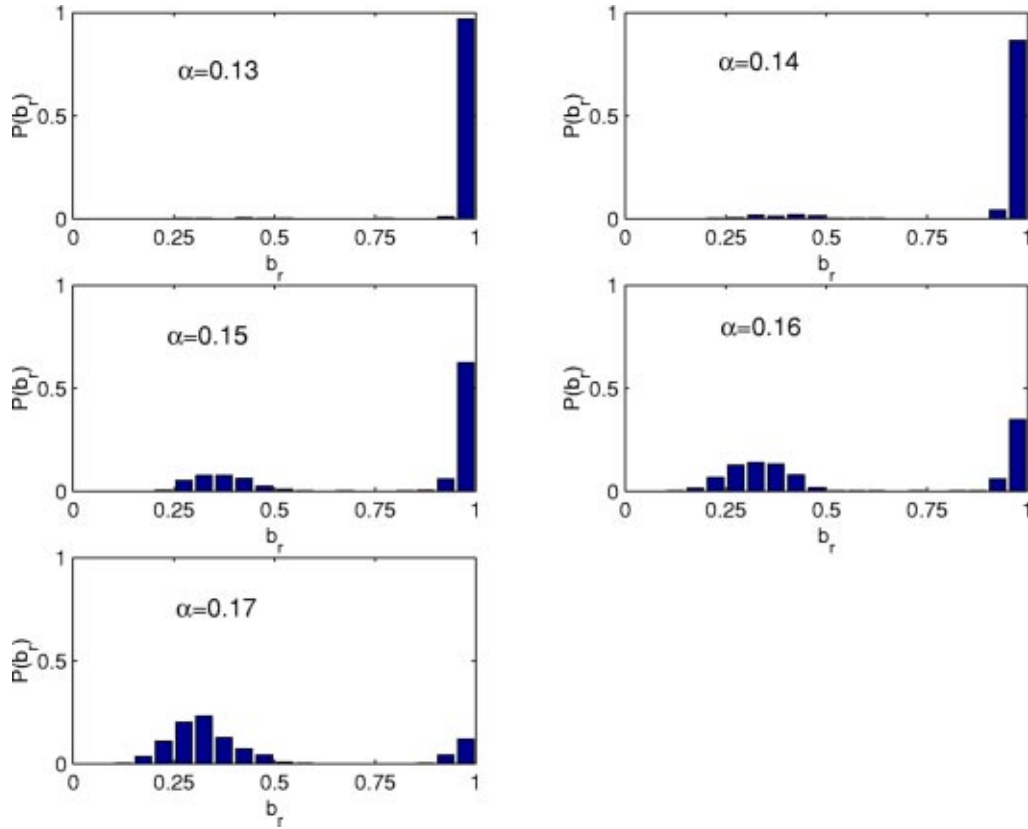


FIG. 1. A series of remanent overlap histograms for a HN with  $N=2000$  nodes, at different values of the loading fraction  $\alpha \equiv p/N$ . When the critical loading  $\alpha_c \approx 0.148$  is exceeded, the strong peak at  $b_r > 0.95$  decays and another peak appears at  $b_r \approx 0.3$ .

the local potential barrier and reverse the sign of  $x_i$ , thus introducing a sign error into the pattern. This will occur only if

$$C_i + S_i < -\frac{2\sqrt{3}}{9\gamma}.$$

By contrast, in the Hopfield model a sign error is introduced if

$$C_i + S_i < 0.$$

Thus the relative strength of crosstalk required to introduce sign errors is greater for the BGN than for the HN, and the discrepancy is greatest for small values of  $\gamma$ . One might expect that this would make the BGN less vulnerable to crosstalk (and the memory states more stable) than the HN, especially at low  $\gamma$ , but this is not a foregone conclusion as the BGN's dynamics include mechanisms that tend to amplify small initial overlaps [2] and could conceivably also amplify crosstalk. Our numerical results confirm that the BGN is in fact less prone to crosstalk-induced errors at low values of  $\gamma$ , but not at high values.

If

$$-\frac{2\sqrt{3}}{9\gamma} < C_i + S_i < \frac{2\sqrt{3}}{9\gamma},$$

the crosstalk will not be strong enough to reverse the sign of  $x_i$  if initially  $x_i$  is correctly aligned with  $\xi_i^\nu$ , but it will nonetheless destroy the stability against a sign flip. In other words, if  $x_i$  is initially misaligned, it will not be corrected. We may say that the node is “*bistabilized*” but not destabilized. (Such an effect cannot occur in the HN where the nodes are not individually bistable.) Since the crosstalk is random, it will in general bistabilize some nodes and not others, with the result that the memory state will be stable against sign flips of certain nodes but not of others. Thus, even though a memory state may be linearly stable, the dynamics may only be able to correct some sign errors, not all. This contrasts with the low-loading case where crosstalk is negligible and there is a single threshold coupling strength,  $\gamma = \frac{1}{3}$ , above which any single sign error can be corrected. In general, crosstalk results in nonuniform behavior among nodes, including different magnitudes of  $x_i$  for different  $i$ .

### III. STABILITY OF THE MEMORY STATES AND REMANENT OVERLAP

To examine the stability of the memory states, we followed a procedure similar to that of Ref. [9]. Using an ensemble of realizations of the random patterns, we made a number of trials in which the network was initialized to the state  $\mathbf{x} = \xi^\nu$  for some pattern  $\nu$ . The initial bit overlap  $b_\nu$  was thus equal to 1. We then allowed the state to evolve until an attractor was reached. In each case we measured the final

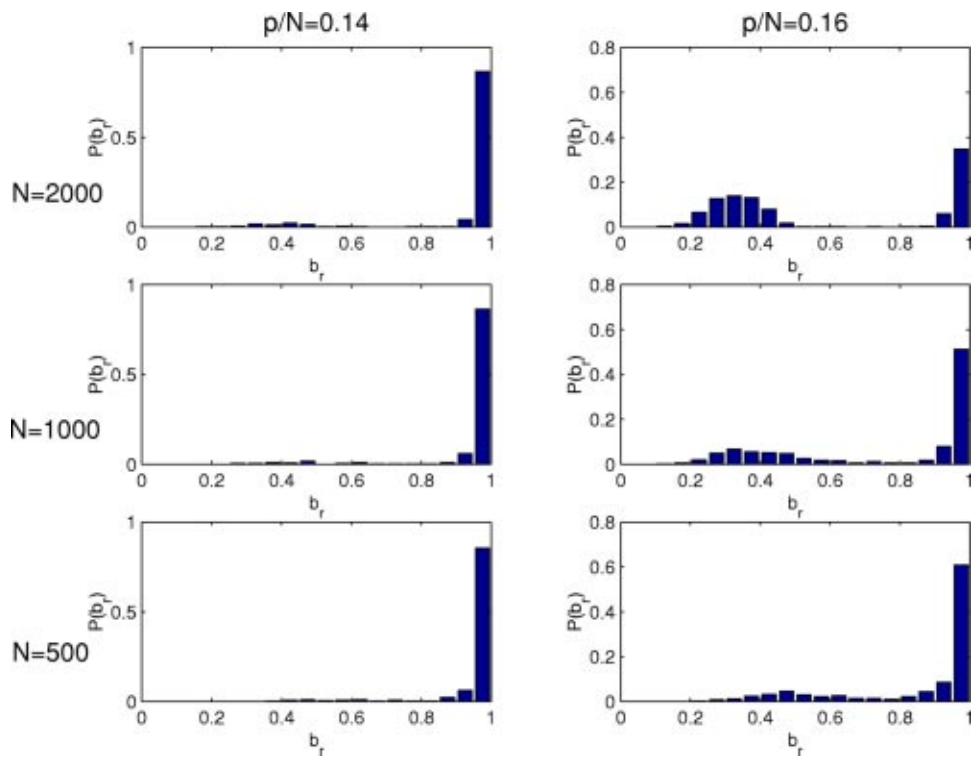


FIG. 2. Remanent overlap histograms for the Hopfield network below the critical loading (left column) and above (right column), showing finite-size effects. The transition becomes sharper as  $N$  increases.

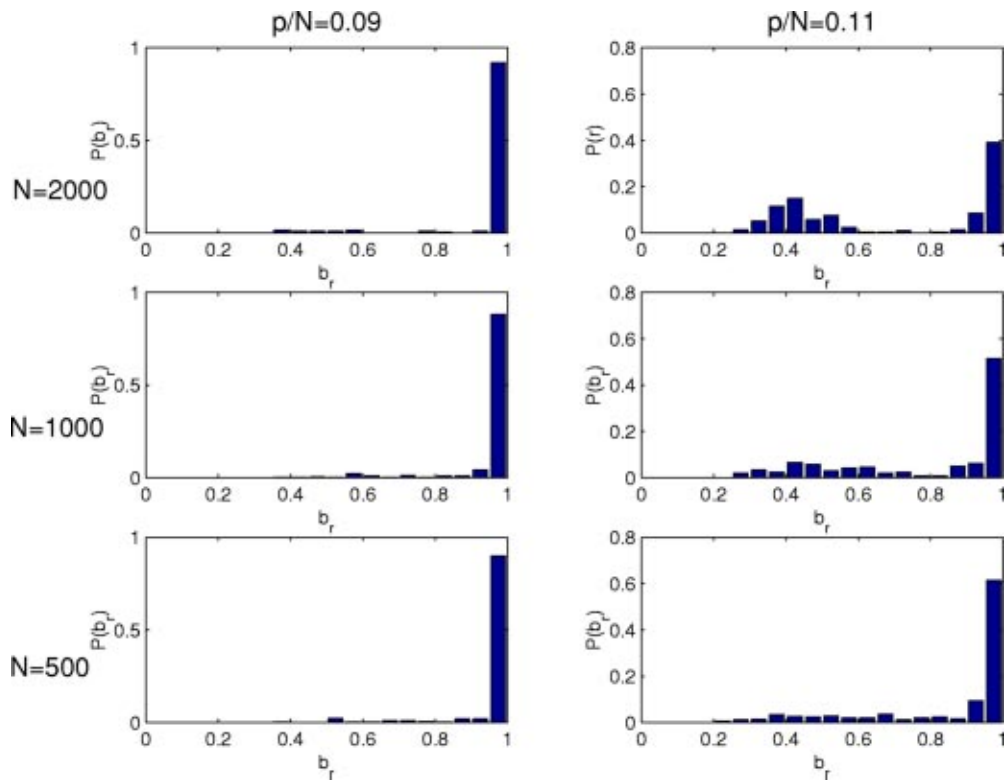


FIG. 3. Remanent overlap histograms for the BGN with  $\gamma=2$  show evidence of a first-order transition at  $0.9 < \alpha_c < 0.11$ . Note the similarity with Fig. 2.

energy as well as the bit overlap between the initial and final states, to which we refer as the *remanent* bit overlap  $b_r$ . We then constructed histograms showing the probabilities  $P(b_r)$  for  $b_r$  falling in intervals of width 0.05. Each of the histograms discussed in this section represents an ensemble of at least 200 initial conditions. If the memory state is entirely stable, as is the case at very low loading, then after the dynamics converges the final bit overlap will still be equal to 1. For higher loading, however, the crosstalk fields may introduce sign errors so that  $b_r < 1$ .

### A. HN and BGN with high $\gamma$

First consider the case of the HN. Figure 1 shows a series of  $P(b_r)$  histograms for the HN at different values of  $\alpha = p/N$ . (The data here are our own, but the results are comparable to those given in Ref. [9]. We include them for comparison with BGN results.) At lower levels of loading, crosstalk introduces few errors, so  $b_r > 0.95$  in the large majority of cases. However, as  $\alpha$  increases beyond the critical value  $\alpha_c \approx 0.148$ , the high- $b_r$  peak of the distribution begins to vanish and a second, broader peak begins to grow in the vicinity of  $b_r \approx 0.3$ . The states in the second peak are spin glass states. As shown in Fig. 2, this transition becomes sharper with increasing network size, and finite-size scaling analysis shows behavior characteristic of a first-order phase transition in the thermodynamic limit [9]. In the limit  $N \rightarrow \infty$  the associative memory fails suddenly as the critical loading is exceeded—the remanent overlap drops abruptly from near 1 to 0.3. This nonzero value of the remanent overlap above the critical loading was noted in Ref. [9] and attributed to replica symmetry breaking, as the replica symmetric theory predicts that  $b_r$  should drop to zero above the phase transition. This phenomenon is related to the nonzero remanent magnetization of a spin glass [13].

In the BGN with  $\gamma=2$ , a similar transition evidently occurs at  $\alpha_c \approx 0.1$ . As evidence, in Fig. 3 we show two series of histograms at increasing  $N$ , one below the suspected transition and one above. As in the HN case, the transition grows sharper with increasing network size. Below the critical loading, the high- $b$  peak remains robust as  $N$  increases, while above the critical loading the high- $b$  peak shrinks with increasing network size and the low- $b$  peak grows. Two quantitative differences are that the critical loading  $\alpha_c$  is lower in the BGN case,  $\alpha_c \approx 0.1$ , while the average remanent overlap above the critical loading is higher, near 0.45.

The first-order nature of the transition is confirmed by examining the energies of the final states. These energies and the overlaps are shown in a scatter plot in Fig. 4(a). The spin glass states associated with the low- $b_r$  peak are clustered at energies below those of the retrieval states. The gap in energy between these two clusters corresponds to the latent heat of the phase transition.

### B. BGN at low $\gamma$

For  $\gamma=0.5$ , in contrast to  $\gamma=2$ , the BGN's behavior differs markedly from that of the HN. A series of  $P(b_r)$  histograms for different values of  $p/N$  is shown in Fig. 5. Two features are evident. First, the stored patterns remain stable

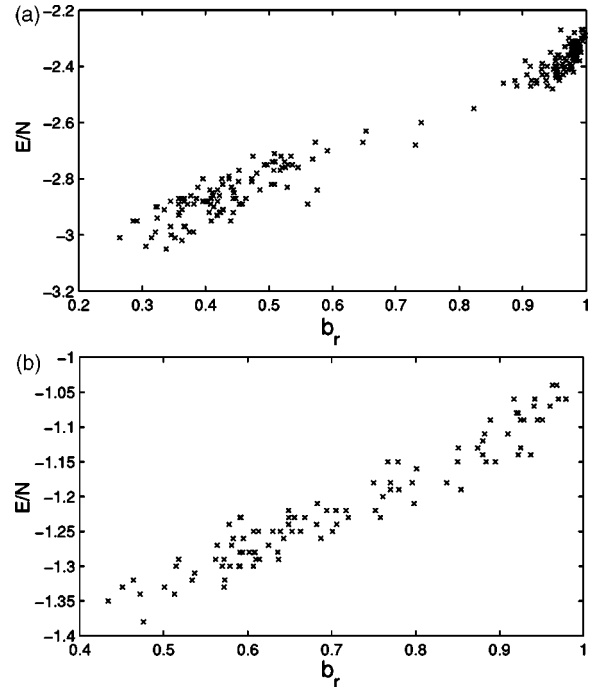


FIG. 4. Scatter plots of final energies and remanent overlaps for trajectories starting from memory patterns, for BGN above critical loading. (a) BGN with  $N=2000$ ,  $\gamma=2$ , and  $\alpha=p/N=0.11$ . Note that the overlap is strongly correlated with the energy and there is a gap in energy between the high- and low- $b_r$  states. This gap represents the latent heat of the phase transition. (b) BGN with  $N=2000$ ,  $\gamma=1$ , and  $\alpha=0.2$ . As in the  $\gamma=2$  case, the energy is strongly correlated with  $b_r$ , but the high and low groups overlap and the energy difference is smaller.

with few errors even up to high levels of storage,  $\alpha=0.3$ . Second, there is no evidence of a discontinuous memory failure; rather, the retrieval quality as measured by  $b_r$  appears to degrade continuously as  $\alpha$  increases. No second peak appears in the histograms; instead, the high- $b$  peak first spreads and then begins to drift downward as errors accumulate.

At an intermediate value  $\gamma=1$  the  $P(b_r)$  histograms (Fig. 6) suggest a first-order transition, although the evidence is less pronounced than in the  $\gamma=2$  case. A second peak appears above the transition, and the high- $b_r$  peak shows clear signs of shrinking as  $N$  increases, but the tails of the two peaks overlap substantially. The greater overlap between the peaks comes about for two reasons. First, the high- $b_r$  peak above the critical loading is broader than in the  $\gamma=2$  case. Second, the remanent magnetization is much higher (i.e., the drop in  $b_r$  at the critical point is much smaller). The latent heat is also much smaller, as can be seen from the scatter plot of the energy [Fig. 4(b)]. The critical loading, or storage capacity, is approximately 0.17, higher than for  $\gamma=2$  and higher than for the HN.

## IV. ATTRACTORS, BASINS, AND THE ENERGY LANDSCAPE AT HIGH LOADING

In the preceding section, we examined the trajectories of initial conditions corresponding to memorized patterns and

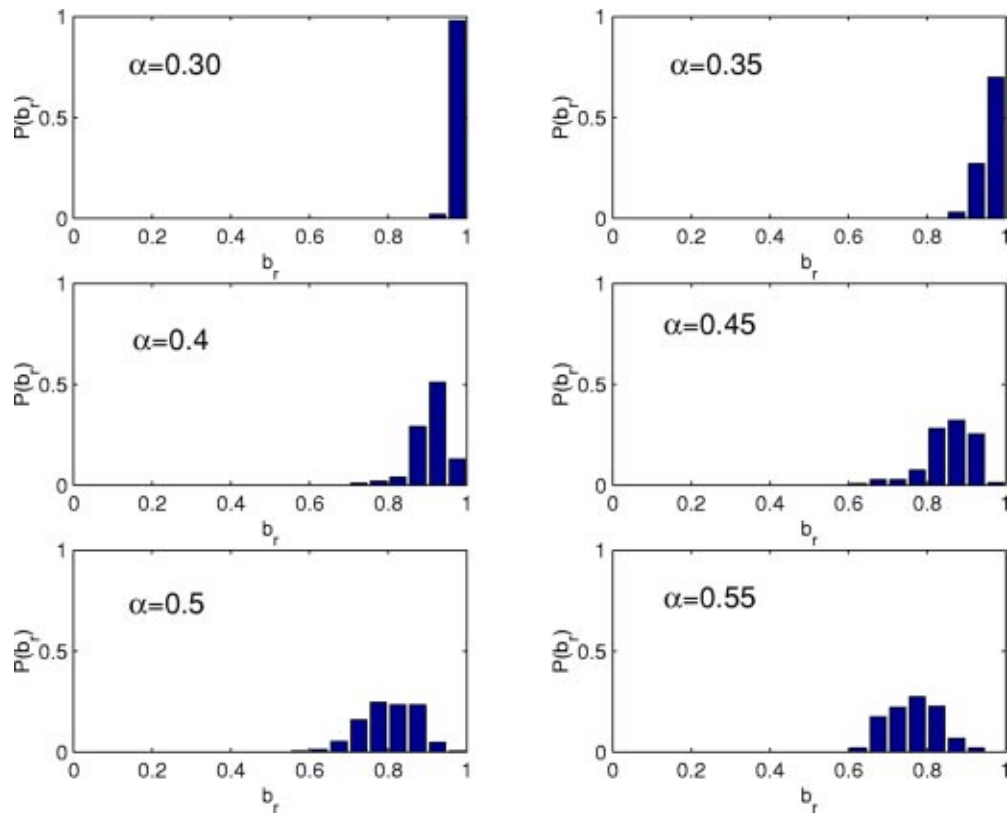


FIG. 5. Remanent bit overlap histograms for the BGN with  $N=1000$  and  $\gamma=0.5$ , at a series of increasing values of the loading factor  $\alpha$ . Few errors occur even at  $\alpha=0.3$ , and the memory degrades gradually rather than abruptly with increasing  $\alpha$ .

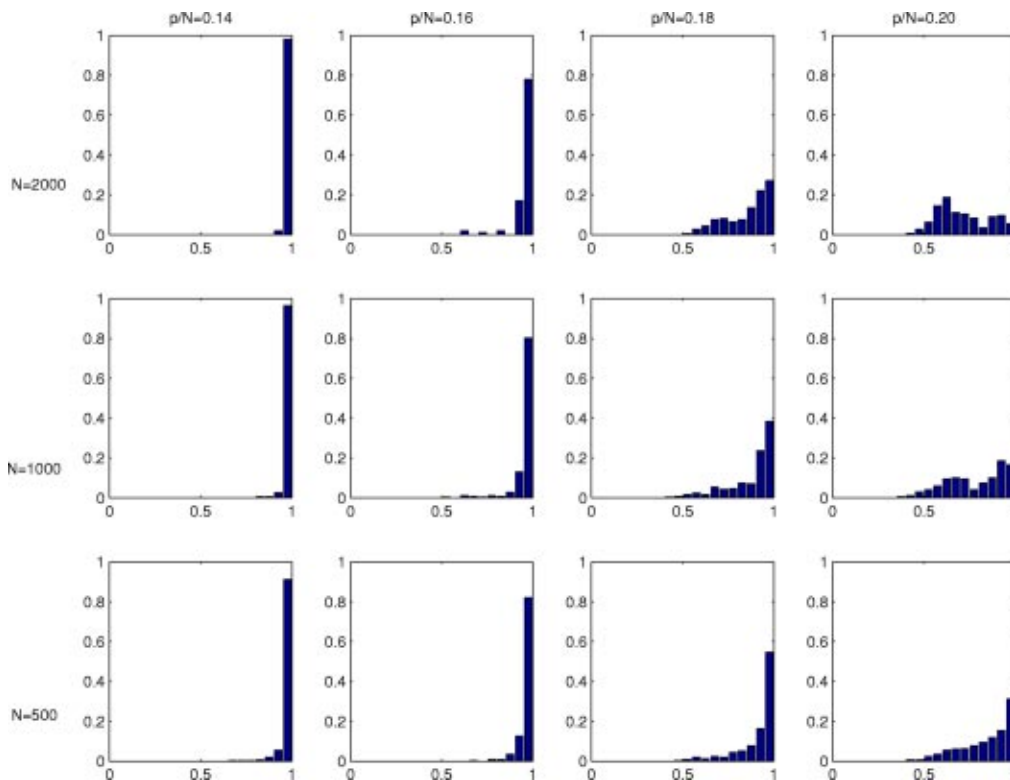


FIG. 6. Remanent overlap histograms for BGN with  $\gamma=1$  are consistent with a first-order transition at  $0.16 < \alpha_c < 0.18$ .

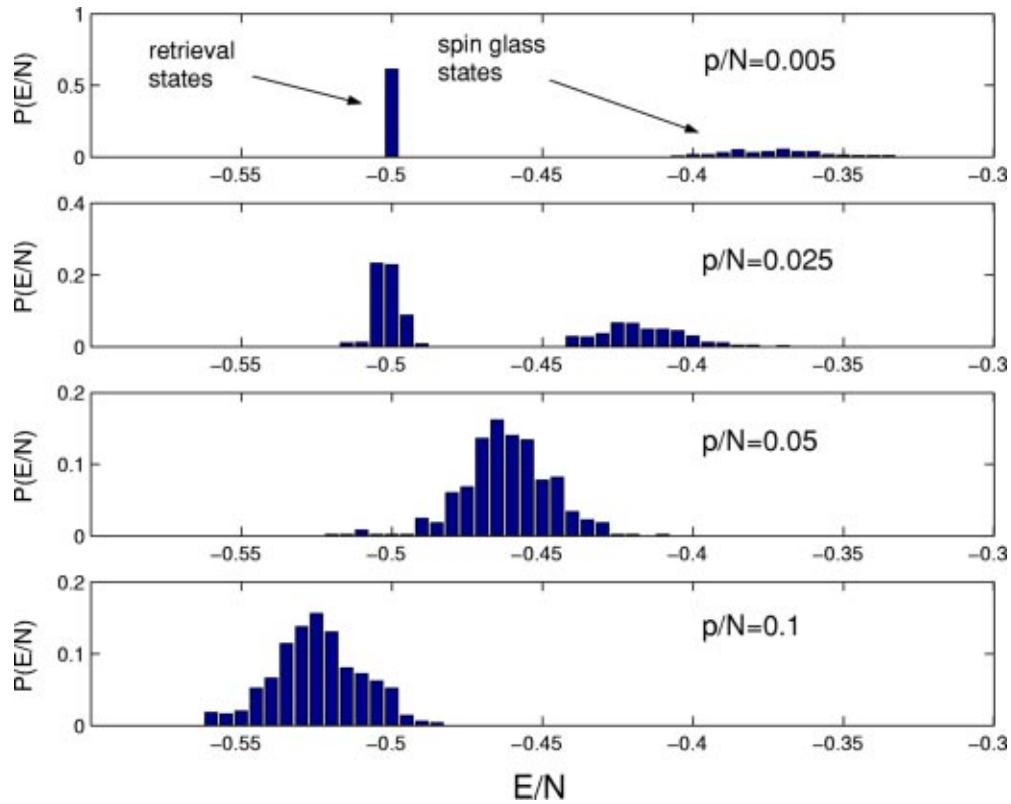


FIG. 7. Energy histograms of  $P(E/N)$  for attractors reached from random initial states of the HN with  $N=1000$  nodes. As the number  $p$  of stored patterns increases, the peak at  $E/N = -0.5$  corresponding to the retrieval states shrinks (and also spreads slightly) while the spin glass states drift downward in energy until they are lower than the retrieval state energies.

determined whether these trajectories remain close to the pattern or move away from it. Such experiments, however, probe only one aspect of network performance. We are interested not only in the stability of memory states but also in the sizes of their basins of attraction. The function of associative memory depends on the ability of the dynamics to correct partially corrupted patterns. More generally, we are interested in the evolution of the energy landscape with increasing  $\alpha$ . To address these issues, we performed two additional sets of numerical experiments. First, we examined the attractors reached from a large number of *random* initial conditions to obtain a uniform sampling of the phase space and a broad picture of the energy landscape. Second, we examined the fate of initial conditions at specified Hamming distances from memory patterns. The latter set of experiments probes the landscape in the vicinity of the memory states. Results of similar experiments were given in Ref. [2] for the case of low loading.

#### A. Evolution of the energy landscape: attractors reached from random initial states

In the case of the HN, it is known that in the thermodynamic limit the memory states are the lowest-energy states for  $\alpha < 0.05$ , while for higher values of  $\alpha$  spin glass states arise which have lower energies. Up to  $\alpha_c \approx 0.148$  the memory states remain local minima of the energy even though they are not the global minima. Above  $\alpha_c$  they cease

even to be local minima and therefore become unstable [9].<sup>1</sup> The drop in energy from the memory states to the spin glass states at  $\alpha_c$  is the latent heat. One way to observe the evolution of the energy landscape is to examine the attractors reached from an ensemble of random initial conditions which effectively samples the configuration space. Figures 7–9 show histograms for the energies of attractors sampled in this manner. In each case, we sampled a total of at least 200 random initial conditions with several realizations of the random patterns  $\xi^\mu$ . In Fig. 7, for the HN, we can see that at low loading the attractors are separated into two groups, the retrieval states at  $E/N = -0.5$  and spurious states at a range of higher energies. The probability of retrieving a memory state from a random initial condition is high. With increasing  $\alpha$ , the spurious states move to lower energies until they are below the memory states. At the same time, their basins of attraction take up a larger portion of the configuration space, as is apparent from the growing size of the spin glass peak in the histogram and the shrinking size of the retrieval state peak. Figure 8 shows that for the BGN with  $\gamma=2$  the evolution is qualitatively similar. In Fig. 9, for  $\gamma=0.75$ , we observe that an additional effect of high loading is to destabilize the uncondensed states. At low loading, the uncondensed states dominate the configuration space—almost all random

<sup>1</sup>For a schematic illustration of the evolution of the energy landscape, see Fig. 2.18 of Ref. [7].

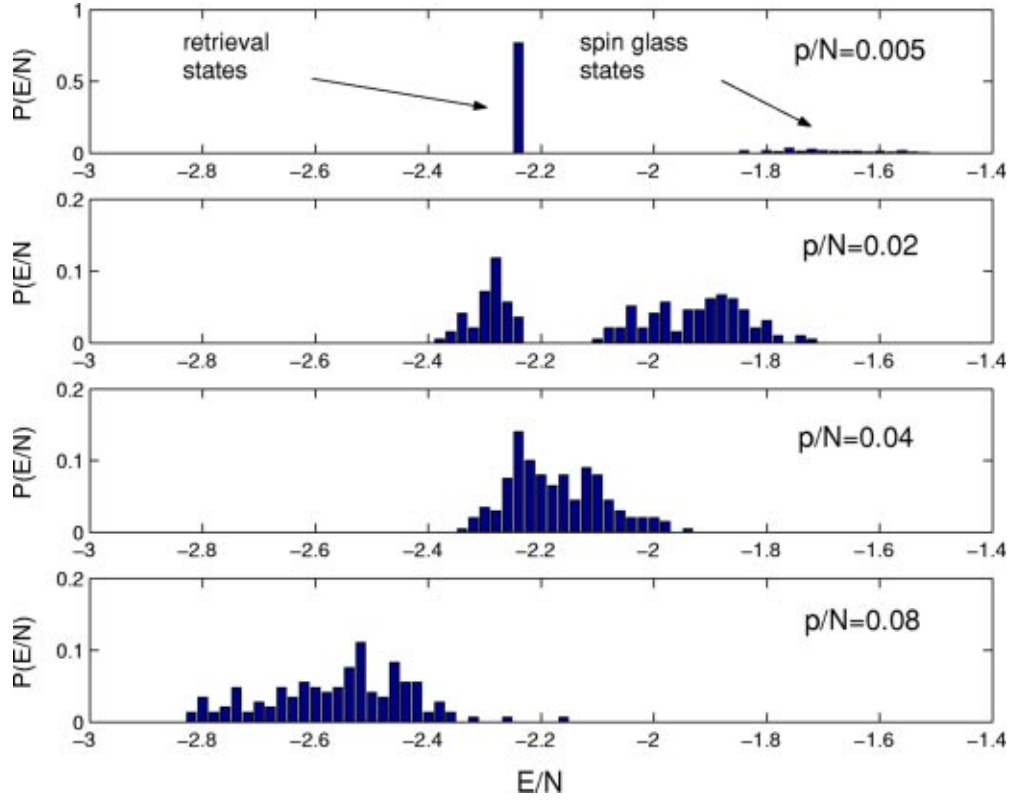


FIG. 8. Attractor energies from random initial conditions for BGN with  $\gamma=2$ . Qualitative behavior resembles that of the HN as shown in Fig. 7. One difference is that the spreading of the retrieval state energies with increasing loading is more pronounced.

initial conditions land on an uncondensed state, as was noted in Ref. [2]. The other two types of state begin to show their presence as the loading is increased, while the uncondensed states eventually disappear.

### B. Retrieval of patterns from corrupted versions

To obtain information about the landscape in the vicinity of the memory states and about the shapes of their basins of attraction, we examined the fates of initial conditions that were not random but rather at specified initial overlaps with particular stored patterns. As in Ref. [2], these initial configurations were generated by starting with a particular “target” pattern and flipping the signs of a specified number of randomly chosen nodes. For each value of the initial overlap  $b_{init}$ , we generated an ensemble of initial conditions for several different realizations of the random set of memory patterns. We then evolved these states until the dynamics converged, and evaluated  $b_{final}$ , the final overlap with the target pattern, for each trajectory. If  $b_{final}=1$ , this signifies that all signs that were initially flipped have been corrected and the target pattern has been retrieved perfectly. If  $b_{init} < b_{final} < 1$ , then the pattern has been retrieved imperfectly. The final state is closer to the stored pattern, but not all errors have been corrected. If  $b_{init} > b_{final}$ , then the trajectory has moved farther away from the stored pattern. As discussed in Sec. II, the ability of the network to correct sign errors depends on the competition among the signal, the crosstalk, and the local potential. The initial states currently discussed have at least a moderately large overlap with the target pat-

tern, resulting in a signal, and random overlaps with the other memory patterns, resulting in crosstalk.

Consider now the HN. At low loading, there is very little crosstalk. The crosstalk term is typically of order  $\sqrt{p/N}$ . If the HN is set in an initial condition whose overlap with the target pattern  $\xi$  is  $b_{init} > \sqrt{p/N}$ , then most nodes experience a net field that is aligned with the target pattern. Nodes that are initially misaligned with the target state ( $x_i = -\xi_i$ ) will tend to change their signs and align with  $\xi_i$ . Each node that realigns in this way increases the value of  $b$  and thus increases the strength of the signal acting on the remaining nodes. As a result, even if in the initial state some fraction of the nodes experience a net field opposed to  $\xi_i$ , eventually the growing signal may overcome the crosstalk and correct those nodes as well. Therefore for low loading, as long as the initial state has  $b_{init} > \sqrt{p/N}$ , the probability of completely retrieving the target pattern is close to unity. As the loading ratio  $\alpha$  increases, however, the typical crosstalk becomes stronger, and a higher signal is required to overcome the crosstalk noise. Therefore sign errors are not likely to be corrected unless the initial overlap is above a threshold, which grows higher with increasing  $\alpha$ . If the crosstalk is too large, then some signs that are initially aligned with the pattern may be flipped, and the state may move away from the target pattern instead of toward it. Each node that flips out of alignment with the target pattern reduces the size of the overlap and hence of the signal, which makes other nodes more susceptible to crosstalk-induced errors, and a cascade of errors can occur. The critical loading  $\alpha_c$  is the level at which even a state with



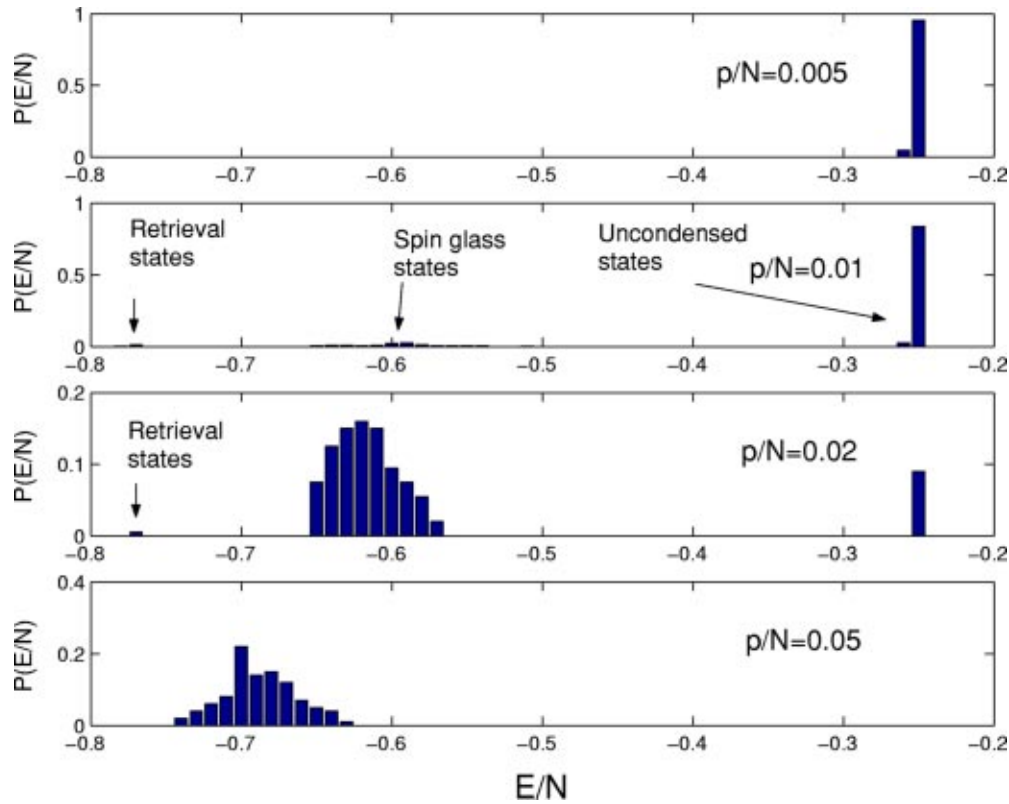


FIG. 9. Attractor energies from random initial conditions for BGN with  $N=1000$  and  $\gamma=0.75$ . For low loading, the uncondensed states (with energy  $E/N \approx -0.25$ ) fill most of the configuration space. The other two peaks in the histogram are very small. As the loading increases, however, the uncondensed states disappear. As in all cases, the “spin glass” peak grows larger with increasing  $\alpha$  and shifts downward in energy.

$b_{init} = 1$  becomes unstable against such a cascade of errors.

Figure 10(a) shows a scatter plot of  $b_{final}$  vs  $b_{init}$  for a HN slightly below critical loading. There is a threshold overlap for retrieval. If  $b_{init} > 0.5$  the signal is strong enough to correct errors and the majority of trajectories flow towards the target pattern. For  $b_{init} < 0.5$  the majority of trajectories move instead away from the target pattern. As  $\alpha$  increases, the threshold value of  $b_{init}$  for retrieval increases until at  $\alpha_c$  it reaches 1. The plot in Fig. 10(a) is for a network with  $N=2000$ ; experiments with networks of different sizes reveal that the retrieval threshold becomes sharper as  $N$  increases.

In the BGN, on the other hand, the dynamics of error correction is more complicated due to the local potential. At strong coupling  $\gamma \geq 1$ , the potential barriers against sign flips are less important than at weak coupling. As a result the BGN in this regime behaves in many respects like the HN. It is not surprising, then, that the scatter plot of  $b_{final}$  vs  $b_{init}$  for a BGN with  $\gamma=2$  slightly below its critical loading [Fig. 10(b)] appears qualitatively similar to the corresponding Fig. 10(a) for the HN. There is a threshold (approximately  $b_{init} = 0.6$ ) below which the probability of fully retrieving the target pattern drops sharply. Above this threshold, the signal is evidently strong enough to correct most sign errors. A key difference, however, is that even below this threshold the average  $b_{final}$  is larger than  $b_{init}$ . This means that the majority of trajectories move toward the target pattern rather

than away from it, but only some of the sign errors are corrected, not all.

In the case of the BGN with  $\gamma=0.5$ , the local potential barriers are important, and interesting dynamics results from the competition among the signal from the target pattern, the crosstalk from other patterns, and the local potential. In the  $\alpha \ll 1$  case [2], sign errors can be corrected only if the signal is strong enough to overcome the local potential barriers, and therefore there is a threshold value of the initial overlap which is much larger than  $\sqrt{p/N}$ . For the case  $N=1000$ ,  $p=5$ , for example, we find that this threshold is approximately  $b_{init} = 0.5$ . For  $b_{init} > 0.5$  there is a large probability that the target pattern is retrieved perfectly, while for  $b_{init} < 0.5$  there is a large probability that the network will be stuck in an uncondensed state with  $b_{final} = b_{init}$ . In the intermediate range  $0.4 < b_{init} < 0.6$ , there is also a significant probability that the trajectory is attracted to an asymmetric spurious state in which  $b_{final}$  is large but not unity and there are larger than random overlaps with one or more other memory patterns. This behavior is illustrated by the scatter plot of Fig. 11(a). As the loading increases [Figs. 11(a)–11(d)], something surprising happens: at first, the basins of attraction of the memory states *expand* slightly, contrary to what one would expect from the HN. The frontier of the uncondensed states is pushed back to lower values of  $b_{init}$ . For  $p=50$ , or  $\alpha=0.05$  [Fig. 11(d)], almost all states with  $b_{init} > 0.1$  undergo some motion toward the target pattern,

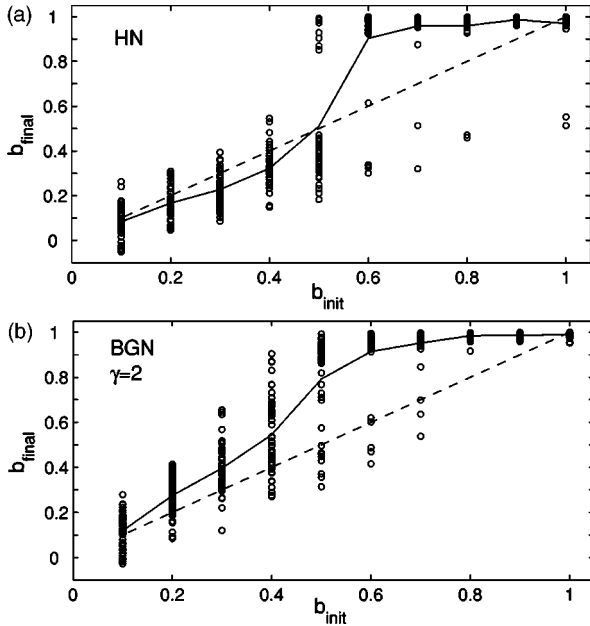


FIG. 10. Scatter plots of  $b_{final}$  vs  $b_{init}$  for networks slightly below critical loading. Points show  $b_{final}$  for an ensemble of initial conditions with specified  $b_{init}$ . The average of  $b_{final}$  is shown by the solid curve. The dotted diagonal line  $b_{final}=b_{init}$  is drawn for comparison: points above the line have  $b_{final}>b_{init}$ . (a) HN with  $N=2000$ ,  $p=260$ . (b) BGN with  $N=2000$ ,  $\gamma=2$ , and  $p=170$ . In both cases, the retrieval quality as measured by  $b_{final}$  drops sharply for  $b_{init}<0.6$ . For the BGN, however, the average  $b_{final}$  is always greater than  $b_{init}$ .

even if the pattern is not retrieved perfectly. As  $p$  increases further, the  $b_{final}$  vs  $b_{init}$  plot preserves the approximate shape of Fig. 11(d).

The dynamics of retrieval and error correction for the BGN with weak coupling and high loading is evidently quite different from that of the HN. In the  $\alpha=0.05$  case of Fig. 11(d),  $b_{final}$  is always 1 if  $b_{init}=1$ , which means that the memory state is stable. However, it is retrieved only imperfectly when sign errors are introduced: if  $b_{init}<1$  then  $b_{init}<b_{final}<1$ . This indicates that some nodes remain bistable and cannot be corrected. If the initial state is close to the target pattern, then the majority of errors are corrected, but that fraction decreases with decreasing  $b_{init}$ . Such a partial correction of errors does not often occur in the HN case. In the latter case, an initial condition either flows all the way to the retrieval state or moves away from it toward another attractor. The retrieval state may itself have a small number of errors due to crosstalk, but the presence of these errors does not depend on the initial state. The energy landscape of the HN in the neighborhood of a memory state apparently has the shape of a smooth basin—once the basin is entered, the trajectory usually runs without obstruction to the attractor at the bottom. For the weakly coupled BGN, on the other hand, the landscape appears to have the structure of a “funnel,” [16–18], i.e., a sequence of local minima at decreasing energies, with low potential barriers separating each state from the next. There is a region of configuration space that has an overall tilt toward the retrieval state, but in which

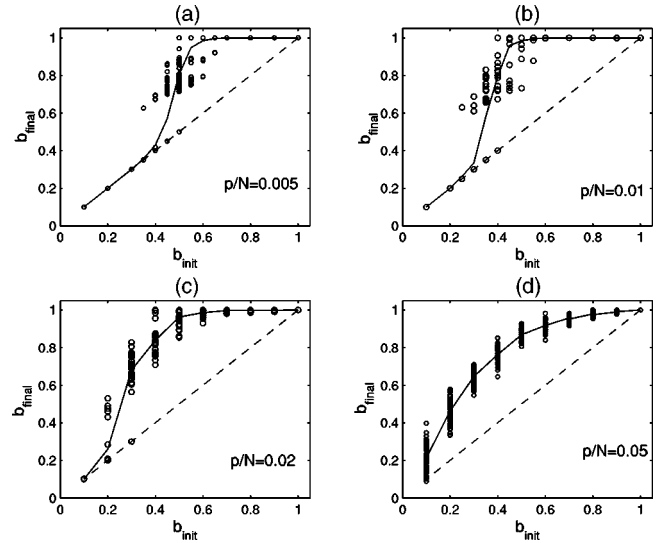


FIG. 11. Scatter plots showing pattern retrieval by the BGN with  $N=1000$ ,  $\gamma=0.5$ . As in Fig. 10, the solid curve shows the average value of  $b_{final}$ . Uncondensed states lie on the line  $b_{final}=b_{init}$ : no sign flips occur and  $b$  maintains exactly its initial value. (a)  $p=5$ , (b)  $p=10$ , (c)  $p=20$ , and (d)  $p=50$ .

there are many local minima which may trap the trajectory before it reaches the retrieval state. This is illustrated schematically in Fig. 12. Funnel-shaped energy landscapes were first suggested in the context of protein folding dynamics.

V. A COMMENT ON SMALL NETWORKS

So far, this paper and the companion paper [2] have focused mainly on large networks of  $N=1000$  or more. However, some applications of neural network algorithms to robotics and other areas make use of networks of only 20 or 100 nodes. Experimental studies of BGN-like chemical reactor networks [14,15] used fewer than ten nodes. The current work on the BGN was inspired in part by results suggesting that the BGN could store many more patterns than the HN,

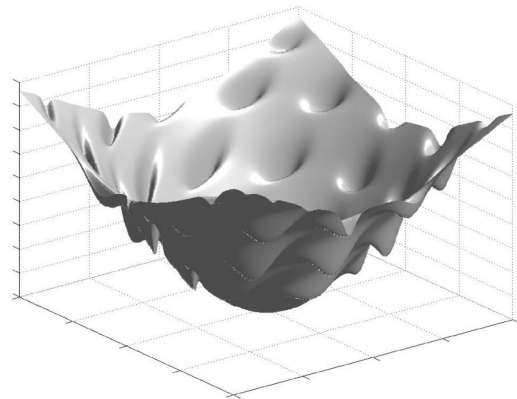


FIG. 12. Schematic picture of a funnel, or bumpy basin. Most trajectories travel some distance toward the global minimum (representing, for example, a memory state) but become trapped in a local minimum (a state with some errors) before reaching the bottom.

TABLE I. Properties of small networks.

$\alpha$	$N=5, p=5$		$N=10, p=5$		$N=10, p=10$		$N=15, p=10$	
	$V_{spur}$	$f_{stable}$	$V_{spur}$	$f_{stable}$	$V_{spur}$	$f_{stable}$	$V_{spur}$	$f_{stable}$
0.25	0.38	0.87	0.96	1.00	0.85	0.92	0.97	0.97
0.5	0.10	0.68	0.64	0.91	0.61	0.51	0.86	0.65
0.75	0.06	0.49	0.51	0.75	0.56	0.32	0.83	0.37
1.0	0.05	0.37	0.49	0.62	0.53	0.23	0.82	0.23
1.5	0.08	0.37	0.51	0.47	0.53	0.16	0.80	0.15
2	0.6	0.34	0.43	0.40	0.51	0.15	0.85	0.11
(HN)	0.03	0.34	0.38	0.48	0.50	0.16	0.75	0.15
$\alpha$	$N=20, p=5$		$N=20, p=10$		$N=50, p=5$		$N=100, p=5$	
	$V_{spur}$	$f_{stable}$	$V_{spur}$	$f_{stable}$	$V_{spur}$	$f_{stable}$	$V_{spur}$	$f_{stable}$
0.25	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
0.5	0.92	0.99	0.95	0.74	0.99	1.00	1.00	1.00
0.75	0.78	0.93	0.92	0.44	0.78	1.00	0.84	1.00
1.0	0.71	0.86	0.92	0.26	0.61	1.00	0.55	1.00
1.5	0.71	0.71	0.91	0.16	0.57	0.98	0.46	1.00
2	0.74	0.62	0.93	0.11	0.57	0.97	0.46	1.00
(HN)	0.47	0.78	0.91	0.20	0.34	1.00	0.34	1.00

with fewer spurious attractors [12]. The latter results were inferred from a few selected cases using very small networks, and so we attempted to test the genericness of these results for a variety of small networks as well as for larger networks.

It should be noted that with very small networks, there are great fluctuations in properties depending on the particular set of memory patterns chosen, as it is impossible to ignore the mutual correlations among patterns. Results for the maximum storage capacity of a network are well defined only in the thermodynamic limit  $N \rightarrow \infty$ , and for small  $N$  even an HN may in particular cases be able to store more than  $0.14N$  stable patterns. For these reasons one cannot draw strong general conclusions about storage capacity based on small networks alone, but it is nonetheless instructive to make some comparative studies of pattern stability in small networks. For several small values of  $N$  and  $p$ , we generated random sets of stored patterns and tested their stability, using the HN and the BGN at several different values of  $\gamma$ . We counted the average fraction of memory patterns that were stable,  $f_{stable}$ .<sup>2</sup> In addition, we estimated the fraction of configuration space  $V_{spur}$  occupied by spurious attractors by following the trajectories of random initial conditions inside the hypercube  $|x_i| < 2$ . These results are collected in Table I. In all cases the results were averaged over at least 100 sets of randomly generated patterns. In generating random sets of patterns, we did not eliminate cases in which two or more patterns are identical. The results show that, contrary to the selected examples discussed in Refs. [1] and [12], the spuri-

ous attractors do *not* generically occupy much less configuration space in the BGN than in the HN (in fact, they normally occupy more, especially at low values of  $\gamma$ ). However, for the most part the percentage of memory patterns that are stable is larger in the BGN than in the HN as long as  $\gamma \lesssim 1$ . As with larger networks, the BGN becomes most similar to the HN when  $\gamma > 1$ , while lower  $\gamma$  leads to increased pattern stability. The increased pattern stability at low  $\gamma$  is associated, however, with smaller basins of attraction for the memory states and therefore with a greater volume of phase space occupied by spurious states.

## VI. DISCUSSION

We have studied the properties of the BGN at high memory loading  $\alpha$ . Our results can be summarized as follows: For high values of  $\gamma$ , such as 2, there is a first-order transition similar to that of a HN. For  $\gamma = 2$ , the transition occurs at a critical loading  $\alpha_c \approx 0.1$ , which is lower than the critical loading of  $\alpha \approx 0.148$  for the HN. As  $\gamma$  decreases, the critical loading increases and the phase transition evidently becomes weaker and eventually disappears. For  $\gamma = 1$  the critical loading is apparently  $\alpha_c \approx 0.17$  (higher than for the HN) and the phase transition is much less pronounced for the finite-size networks we have studied. For  $\gamma = 0.5$  there is no evidence of a phase transition at all and patterns are stable with very few errors up to  $\alpha \approx 0.3$ .

A phenomenon that occurs in the BGN much more than in the HN is the partial retrieval of a pattern, whereby the dynamics corrects some sign errors in a pattern without correcting all of them. This is especially noticeable in the case of low  $\gamma$  and high  $\alpha$ . This phenomenon suggests that in this case the energy landscape in the vicinity of a stored pattern has the shape of a funnel rather than a smooth basin of attraction. By a smooth basin of attraction, we mean a connected neighborhood that is sloped toward an attractor and in

<sup>2</sup>In the thermodynamic limit there is an approximate permutation symmetry among the memory patterns, so that, in general, either all will be stable or none will be. By contrast, in small networks it is not unusual for some memorized patterns to remain stable while others are not.

which there are few local minima that might obstruct a trajectory once it has begun to flow toward the attractor. By a funnel we mean a region with an overall average slope toward an attractor, which, however, contains many other local minima in which the trajectory might become stuck before reaching the bottom. The presence of these local minima is due to the bistability of the individual nodes and the local potential barriers against spin flips. These same potential barriers are also responsible for stabilizing the patterns. They reduce the likelihood of crosstalk noise inducing an error in a pattern, but they can also inhibit the correction of an error that is present initially. Funnel-shaped landscapes were first examined in the context of protein folding dynamics [16,17]. It was suggested that at a finite temperature such a landscape would allow the protein dynamics efficiently to find the global minimum of energy in spite of the presence of numerous local minima separated by potential barriers. If indeed the landscape of the BGN at low  $\gamma$  forms a funnel, then it is possible that the introduction of some stochastic noise (i.e.,

finite temperature) could improve pattern retrieval by allowing trajectories to jump over the comparatively small potential barriers into lower-energy minima, just as in the protein case. The effect of stochastic noise could be a fruitful subject for further study. An interesting question is whether there is an optimum level of noise which would improve the pattern retrieval ability of the BGN with low  $\gamma$  while at the same time maintaining the larger storage capacity. Intriguing questions remain concerning the dynamics of the BGN at low  $\gamma$ . The apparent initial expansion of the basins of attraction of the memory states with increasing loading is counterintuitive, and the patterns visible in Fig. 11 hint at some dynamics that is not yet fully understood.

#### ACKNOWLEDGMENTS

This work was supported by Materials and Manufacturing Ontario (MMO), a provincial center of excellence. Discussions with V. Chinarov are also acknowledged.

- 
- [1] V. Chinarov and M. Menzinger, *Biosystems* **55**, 137 (2000).  
 [2] P. McGraw and M. Menzinger, *Phys. Rev. E* **67**, 016118 (2003).  
 [3] J.J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2554 (1982).  
 [4] W.A. Little, *Math. Biosci.* **19**, 101 (1974); W.A. Little and G.L. Shaw, *ibid.* **39**, 281 (1978).  
 [5] S. Haykin, *Neural Networks: A Comprehensive Foundation* (Prentice-Hall, Upper Saddle River, NJ, 1999).  
 [6] Daniel J. Amit, *Modeling Brain Function: The World Of Attractor Neural Networks* (Cambridge University Press, Cambridge, 1989).  
 [7] J. Hertz, A. Krogh, and R.G. Palmer, *Introduction to the Theory of Neural Computation* (Perseus Books, Cambridge, MA, 1991).  
 [8] D.J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **32**, 1007 (1985); *Phys. Rev. Lett.* **55**, 1530 (1985).  
 [9] D.J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys. (Paris)* **173**, 30 (1987); A. Chrisanti, D.J. Amit, and H. Gutfreund, *Europhys. Lett.* **2**, 337 (1986).  
 [10] D.O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).  
 [11] J.J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3088 (1984).  
 [12] V. Chinarov and M. Menzinger, *Connectionist Models of Neurons, Learning Processes, and Artificial Intelligence*, edited by J. Mira and A. Prieto (Springer-Verlag, Berlin, 2001), p. 333.  
 [13] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).  
 [14] W. Hohmann, M. Krauss, and F.W. Schneider, *J. Phys. Chem. A* **103**, 7606 (1999); **102**, 3103 (1998); **101**, 7364 (1997); G. Dechert, K.-P. Zeyer, D. Lebender, and F.W. Schneider, *ibid.* **100**, 19043 (1996).  
 [15] J.-P. Laplante, M. Pemberton, A. Hjelmfelt, and J. Ross, *J. Phys. Chem.* **99**, 10 063 (1995); V. Booth, T. Erneux, and J.-P. Laplante, *ibid.* **98**, 6537 (1994); A. Hjelmfelt and J. Ross, *ibid.* **97**, 7998 (1993).  
 [16] P.E. Leopold, M. Montal, and J.N. Onuchic, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8721 (1992).  
 [17] P. Wolynes *et al.*, *Science* **267**, 1618 (1995).  
 [18] O.M. Becker and M. Karplus, *J. Chem. Phys.* **106**, 1495 (1997).