Behavior of a neural network with three-spin interactions

R. P. Bajpai
Institute of Self Organising Systems and Biophysics, North-Eastern Hill University, Shillong 793 003, India

Prabodh Shukla
Department of Physics, North-Eastern Hill University, Shillong 793 003, India
(Received 6 July 1989; revised manuscript received 2 February 1990)

A neural network with three-spin interactions is studied numerically. The number of patterns stored in the memory of the network is varied from 1 to 1400 for a 100-spin system. The storage capacity of the network is found to be approximately 700 patterns. Various factors affecting the quality of retrieval are examined.

I. INTRODUCTION

Hopfield[1] has proposed a simple model to elucidate the workings of a neural network. His model is based on a system of N Ising spins. Each Ising spin represents a neuron in its firing or nonfiring state. The statistical mechanical content of the model is grasped more easily if we visualize the Ising spin as a pixel on a black and white television screen. The system of N Ising spins then corresponds to a screen having N pixels where each pixel can be either black or white (no shades). A system of N Ising spins can assume 2^N configurations each of which corresponds to a distinct N-bit picture on the screen. In the Hopfield model some of these pictures are "stored" in the form of pairwise interactions (synaptic efficiencies) between the spins (neurons). The stored pictures lose their individual indentity in the "interactions." However, when the system "sees" a new picture somewhat reminiscent of a stored picture, the dynamics of the system reconstructs the memorized picture. This is very similar to the phenomenon of human memory being jogged by familiar-looking objects. Therefore the Hopfield model is thought to provide a possible framework for understanding human associative memory.

The reconstructed pictures are the attractors of the network dynamics. In favorable circumstances, the attractors are either identical with the stored pictures or have a very high overlap with them. It can be shown analytically[2] that if the number of stored pictures is less than $N/(2\ln N)$ then the attractors are identical with the memories in the large-N limit. For a larger number of stored pictures up to $0.145N$, the attractors may not be identical to the memories but have a very high overlap with them (over 97%). The quantity $0.145N$ is generally taken as the storage capacity of the Hopfield model. If a larger number of patterns is stored, the attractors of the dynamics have drastically reduced (less than 45%) overlaps with the memories making meaningful retrieval of a memory difficult.[3,4]

Recently, generalizations of the Hofield model have been proposed where the storage capacity of the network scales as a power of N, rather than a fraction of it. These generalizations are based on higher-order[5] multispin interactions.[6] Gardner[6] has shown analytically that the storage capacity of a neural network based on n-spin interactions scales as N^{n-1} apart from logarithmic corrections. The Hopfield model corresponds to the $n=2$ case. In this paper we present results of the numerical simulation for a network having three-spin interactions. Our simulation confirms the analytical results due to Gardner, and provides additional results which are not available by analytical methods.

II. MODEL

Our model is characterized by the Hamiltonian

$$H=-\frac{1}{3!}\sum_{i,j,k}{}' J_{ijk}S_iS_jS_k \ , \tag{1}$$

with

$$J_{ijk}=\frac{1}{N^2}\sum_{\mu=1}^{p}\xi_i^\mu\xi_k^\mu\xi_k^\mu \ . \tag{2}$$

Here $\{\xi_1^\mu=\pm1;\ i=1,\ldots,N,\ \mu=1,\ldots,p\}$ are the Ising spins representing p N-bit patterns stored in the memory, and $\{S_i=\pm1;\ i=1,\ldots,N\}$ are the Ising spins representing an N-bit pattern which jogs the memory. The prime on the summation sign in Eq. (1) means that the sum is restricted to terms with indices i, j, and k all different from each other. This is a distinguishing feature between our simulations and that of Lee *et al.*[5] It improves the quality of retrieval from a partially corrupted image and is akin to the Hopfield model. It is convenient to write (1) in the form

$$H=-\sum_i h_iS_i \ , \tag{3}$$

where h_i, which can be thought of as the local field acting on the spin S_i, is given by

$$h_i=\tfrac{1}{6}\sum_{j,k}{}' J_{ijk}S_jS_k \ . \tag{4}$$

The input pattern is processed by the following set of discrete time equations of motion of the network:

$$S_i(t+1)=\mathrm{sgn}\,h_i(t) \ . \tag{5}$$

The above equations convert the input pattern $S_i(t=0)$ into a modified pattern $S_i(t=1)$ after a unit time interval. The pattern $S_i(t=1)$ is again inserted into Eq. (5) and a new modified pattern $S_i(t=2)$ is obtained. This process is continued. Under suitable conditions for the operation of the network as an associative memory, the initial pattern quickly flows to an attractor of the dynamics which is either one of the patterns stored in the memory or very close to it.

An efficient simulation of Eqs. (4) and (5) on a computer requires the storage of the $N\times N\times N$ interaction matrix J_{ijk} in the main memory of the computer. This is a strenuous requirement of the resources of the computer. Compromise can be achieved with less main memory of the computer if we make use of some algebraic manipulations. We rewrite the Hamiltonian and the local-field in terms of the overlaps of the starting pattern with the stored patterns. These overlaps R_μ are given by

$$R_\mu(t=0)=\sum_{i=1}^{N}\xi_i^\mu S_i \ . \tag{6}$$

It is easy to see that in terms of the overlaps the Hamiltonian (3) and the local field (4) take the form

$$H=-\frac{1}{6N^2}\sum_\mu [R_\mu^{(3)}-(3N-2)R_\mu] \ , \tag{7}$$

$$h_i=\frac{1}{6N^2}\sum_\mu [(R_\mu^2-N+2)\xi_i^\mu-2R_\mu S_i] \ . \tag{8}$$

The implementation of dynamics through Eq. (8) avoids the storage of the matrix J in the main memory of the computer. It requires only the knowledge of the individual stored pictures and the overlap of the starting picture with the individual stored pictures. Thus only a $p\times N$ matrix representing the memorized patterns has to be stored in the main memory of the computer. However, we must emphasize that this procedure is merely a calculational ploy, and it tends to conceal unintentionally an important feature of the neural networks. In the Hopfield model as well as in its generalizations, the stored patterns lose their individual identity. Reconstruction of a stored pattern is a nontrivial effect of the dynamics of the model. Our simulation algorithm retains the individual identity of the stored patterns only as a calculational convenience. It does not alter, nor does it seek to alter the basic physics and philosophy of the problem.

III. SIMULATION PROCEDURE AND RESULTS

The total number of spins in the network was fixed at $N=100$. Random and uncorrelated patterns $\{\xi_i^\mu\}$ were generated using a pseudorandom-number generator. The number of patterns so generated was varied from $p=1$ to 1400 in steps of 50 or less. The following procedure was repeated for each value of p.

A. Stability of stored patterns

One of the stored patterns was selected at random, and allowed to evolve under the equations of motion for up to ten iterations. The procedure was repeated for 100 different stored patterns if the number of stored patterns was much larger than 100. Otherwise, each stored pattern was tested individually. The results were classified into two broad categories: (i) when a fixed-point pattern was reached which did not differ from the corresponding stored pattern in more than nine pixels, and (ii) when either no fixed-point pattern was reached in ten iterations or the fixed-point pattern was off from the stored pattern in more than nine places. The number of spins in the fixed-point pattern which were off from the corresponding stored pattern were denoted by the letter e (for errors), and fixed points were binned into ten classes corresponding to $e=0,\ldots,9$. Experience showed that if an attractor was not reached in ten iterations then it was not reached in considerably more iterations if at all. The attractors which differed by more than 10% from the stored memories as well as the attractors which took considerably longer time to be reached were not very useful for the purpose of memory retrieval.

The above procedure tested the local stability of a stored pattern under the network dynamics. The stored patterns that remained unchanged under the dynamics were identified as the attractors of the dynamics. This corresponded to the case $e=0$ in the above classification. Our procedure also identified other attractors of the dynamics in the vicinity of the stored patterns. The behavior of the network with increasing p showed that all stored patterns remained unchanged under the dynamics for p up to about 300, and thereafter an increasing proportion of stored patterns become unstable under the

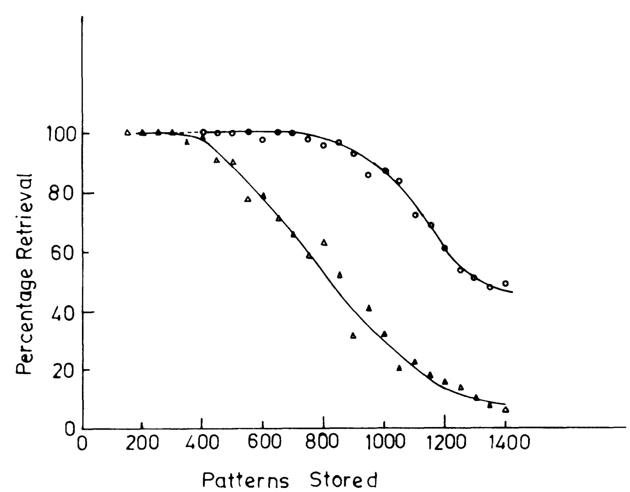

FIG. 1. Stability of stored patterns under the network dynamics. The lower curve shows the percentage of stable stored patterns as a function of the total number of stored patterns. The upper curve shows the percentage of stored patterns which lead to attractors having more than 97% bits in common with the corresponding memories.

equations of motion. However, the majority of the unstable patterns flowed into nearby attractors which differed from the corresponding memories in only a few places.

The lower curve in Fig. 1 shows the percentage of starting patterns which remained unchanged under the dynamics as a function of the total number of stored patterns p. The stored patterns begin to yield around $p=300$, and the percentage of stable patterns drops steadily with increasing p. The upper curve in the same figure shows the cumulative percentage of starting patterns which yield attractors with $e=0$, 1, 2, or 3. The attractors in these cases have sufficiently high overlaps with the stored patterns to be considered useful for memory retention. This curve shows a downward trend around $p=700$, and suggests that $p=700$ may be the limiting storage capacity of the 100-spin network with three-spin interactions. The data pints in Fig. 1 are shown at intervals of 50 pictures in order not to crowd the figure. Data points at much smaller intervals were obtained and the two solid curves are freehand drawings through all the data points, respectively. As already mentioned, data points for the cases $e=0,\ldots,9$ were obtained separately. The figure shows only a cumulative curve with $e=0$, 1, 2, or 3 as a representative example.

It is remarkable that even for $p=1400$ the attractors possess considerable remanence of the stored memories. The histogram in Fig. 2 shows the frequency with which attractors occur which differ from the corresponding memories in less than ten places. The data is based on 100 trials starting from a different stored pattern each time. It shows a peak at $e=3$ which could not have been anticipated beforehand. Lee *et al.*[5] classified all attractors with $e=0$ undesirable traps. A considerable number of these traps are still suitable for memory purposes as they have only a slight error.

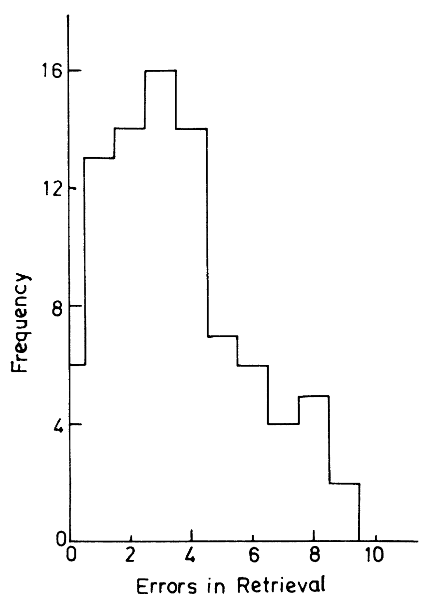

FIG. 2. Remanence of memories in a network with $N=100$, and $p=1400$. The histogram shows the frequency with which stored patterns yield attractors which differ from them in $e=0,1,\ldots,9$ places. The results are based on 100 trials.

B. Basins of attraction

In the preceding section we selected a stored pattern randomly and studied its evolution under the dynamics. In the present section we will modify the selected pattern in a controlled and systematic manner, and study the evolution of the modified pattern under the equations of motion of the network. The modification is made by flipping a number of spins in the stored pattern. Subsequently the overlap of the modified pattern with all the stored patterns are evaluated and that stored pattern is identified which has the highest overlap R with the modified pattern. The overlap measures the degree of similarity between the two patterns. $R=N$ means the two patterns are identical, $R=0$ means the two patterns are completely uncorrelated, and $R=-N$ means the two patterns are mirror images of each other.

We selected a stored pattern randomly and modified it to generate progressively decreasing values of R. Each modified pattern was allowed to evolve under the dynamics for up to ten iterations. The result of the dynamical processing was classified into various categories similar to those discussed in the earlier section. If the modified pattern reached a fixed-point pattern in ten iterations, the fixed-point pattern was compared with the corresponding stored pattern, and the number of spins which differed in the two patterns was counted. This was denoted by the letter e (number of errors). Fixed-point patterns were classified into ten classes corresponding to $e=0,1,\ldots,9$. If e was greater than 9, or the modified pattern did not reach a fixed point in ten iterations, the modified pattern was considered outside a useful domain of attraction of the corresponding memory.

As in the preceding section the number of stored patterns was increased in steps of 50 or less, and for each set of p stored patterns, the overlap of the modified pattern with one of the stored patterns was varied from $R=98$ down to about $R=0$. For each value of R, 100 different modified patterns were examined. The results of a representative case for $p=501$ are shown in Fig. 3. The x axis shows the overlap of the modified pattern with the corresponding stored pattern. The y axis shows the percentage of cases which yielded a fixed-point pattern identical to the appropriate memory (triangles), fixed-point patterns which differed from the memory in three places or less (open circles), and cases which either did not yield a fixed-point pattern or yielded one which was off from the memory in more than nine places (solid circles). The data can be divided broadly into three regions. If R is less than 30, the memory is not retrieved essentially. There is a transition region from around $R=30$ to 60 where the chance of retrieving a memory increases steadily with increasing R. If R is greater than 60, the chance of retrieving a memory is quite high and nearly independent of the values of R. This is the most useful region for the purpose of memory retrieval. We may take $R=60$ as

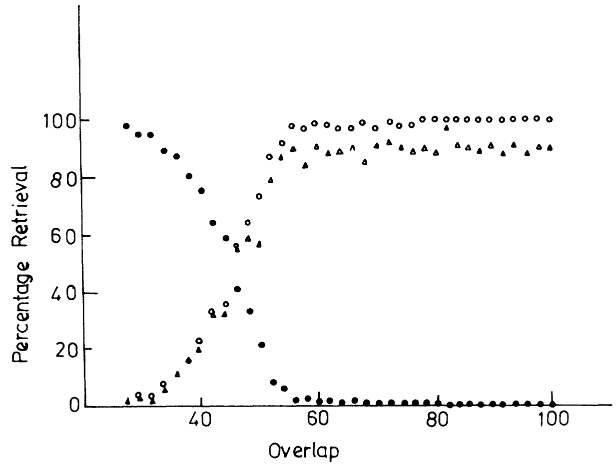

FIG. 3. Retrieval of a memory starting from a corrupted pattern. The x axis shows the overlap of the corrupted pattern with the memory. The triangles show the cases when the exact memory was retrieved. Open circles show the cases when the retrieved pattern differed from the memory in three places or less. The solid circles denote the cases when the fixed-point pattern was either not obtained or differed from the memory in more than nine places.

the minimum overlap which a "sight" must have with a memory in order to jog that memory.

The variation of the minimum overlap with the number of patterns stored in the memory is shown in Fig. 4. As in Sec. III A, the total number of spins in the network was fixed at $N=100$, and the number of stored patterns was varied from $p=1$ to 1400 steps of 50 or less. The number of stored patterns is shown on a logarithmic scale along the x axis. The open circles show the minimum overlap required to retrieve a memory in a neural network based on three-spin interactions. The solid circles show the results for the Hopfield model (two-spin interaction) by way of a comparison. As we saw in the representative case of $p=501$ shown in Fig. 3, the minimum overlap is not a very sharply defined quantity. It rises steadily over a rather broad transition region. The minimum gap shown in Fig. 4 corresponds to the value in the middle of the transition region of the retrieval curve for the case of three errors or less in the retrieved pattern. It corresponds approximately to the chance of 75% or more that a memory will be retrieved.

It is obvious from Fig. 4 that the network based on three-spin interactions can store many more patterns than the one based on two-spin interactions. Our results for a 100-spin network show that the two-spin network can store about 14 patterns, which agrees with the known results for the Hopfield model. In comparison, the three-spin network can store about 700 patterns. However, our numerically found storage capacity for the three-spin network falls short of the value predicted by Gardner. Gardner's prediction for the storage capacity of the three-spin network is approximately $0.13N^2$ patterns. We are not able to say definitely whether the discrepancy comes from the assumptions made in Gardner's calculations or the inadequacy of the size of the network used in our numerical simulations. The 100-spin network is the largest network that we can study conveniently on our machine.

Gardner's analysis also predicts a sharply defined storage capacity which is approximately equal to $0.13N^2$ random patterns for an N-spin network with three-spin interactions. A first-order phase transition takes place at this critical value of storage. The order parameter of the transition is simply related to the number of errors in the retrieved pattern. The number of errors below the criti-

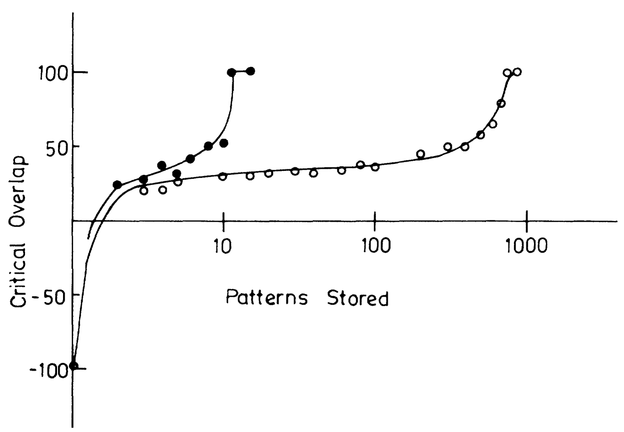

FIG. 4. The threshold of recognition: the y axis shows the minimum overlap that an input pattern must have with a memory in order to have at least a 75% chance of jogging that memory, the x axis shows on a logarithmic scale the total number of patterns stored in the network. The solid circles show the data for the case of two-spin interaction, and the open circles for three-spin interaction. The rise in the two curves around $p=14$ and 700 signal the storage limits of the two networks respectively.

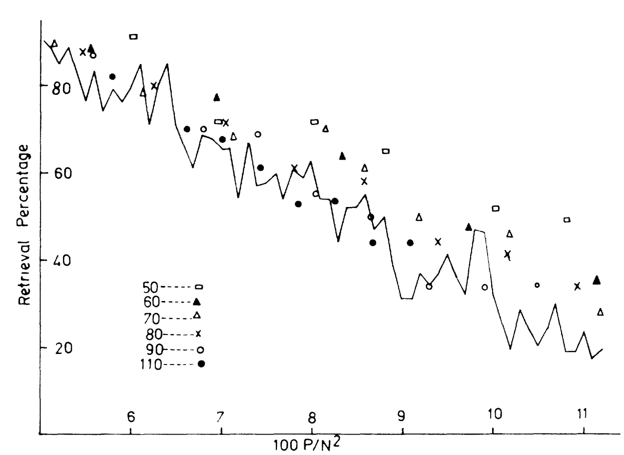

FIG. 5. Scaling behavior of networks. The percentage of stable stored random patterns as a function of P/N^2 for networks with $N=100$.

cal storage is small (less than $0.08N$). It jumps to a value of the order of N as the critical storage is exceeded. Our numerical study shows that the number of errors in the retrieved pattern is indeed small up to the storage limit but we find a gradual transition rather than a sharp transition.

Another important prediction of Gardner's analysis is the scaling behavior as a function of P/N^2. The stability analysis of the stored patterns was repeated for networks of sizes varying between $N=50$ and 110 in steps of 10. Bak and Little[7] had earlier reported the results for $N=10$ to 40. Our results for the case of error-free retrieval are plotted in Fig. 5 in the transition region. The detailed curve corresponds to the $N=100$ network. Considering the expected large variation in the transition region, the observed behavior is consistent with scaling, though no decrease of the transition region is noticeable. Perhaps error-free retrieval is not a good order parameter. It is quite possible that this may be due to the finite size of the network we have studied. The width of the transition region may decrease with increasing size of the network. However, it is not possible to settle this question within the scope of the simulations presented in this paper, and the resources of our computing machine.

[1]J. J. Hopfield, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[2]C. Weisbuch and F. Fogelman Soulie, J. Phys. (Paris) Lett. **46**, L623 (1985).

[3]D. J. Amit, H. Gutfreund, and H. Sopolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).

[4]D. J. Amit, in *The Physics of Structure Formation,* edited by W. Guttinger and G. Dangelmayr (Springer-Verlag, Berlin, 1987), pp. 2–21.

[5]Y. C. Lee *et al.*, Physica D **22**, 276 (1986).

[6]E. Gardner, J. Phys. A **20**, 3453 (1987).

[7]C. S. Bak and M. J. Little, in *IEEE International Conference on Neural Networks, San Diego, CA, 1988* (IEEE, New York, 1988) (IEEE Catalog No. 88 CH 2632-8), Vol. 1, pp. 207–216.