Journal of Statistical Physics, Vol. 59, Nos. 3/4, 1990

A High-Precision Study of the Hopfield Model in the Phase of Broken Replica Symmetry

G. A. Kohring[1]

Received December 6, 1989

Using a multi-spin coding algorithm, the Hopfield model is studied for network sizes up to $N = 33{,}968$. Thermodynamically stable states are found in a region where the replica-symmetric solution predicts none should exist, but where a one-step replica symmetry-breaking calculation predicts some should exist. Furthermore, the order parameter in this region is found to take on two distinct values, one of which is not predicted by any theory.

KEY WORDS: Hopfield model; multispin coding; neural networks; replica method; spin glasses.

1. INTRODUCTION

The Hopfield model[8] is an example of a physical system having random interconnections, the study of which is currently one of the most active subfields in physics research. This is partly due to the fact that knowledge gained in this area has applications not only to the widely studied case of spin glasses (see ref. 1 for a review), but also in areas outside the traditional sphere of physics, such as error-correcting codes,[3] evolution,[4] and, as exemplified by the Hopfield model, neural networks.[2] In studying such systems a very powerful calculational tool known as the "replica method" has emerged (see ref. 1 and references therein). This tool allows one to calculate the free energy averaged over a quenched distribution of random connections as the saddle-point of multidimensional integrals over so-called "replica variables." Normally this saddle point cannot be determined in general; instead, one makes a *replica-symmetric* approximation,[5] i.e., one assumes the replica variables are independent of the replica index and attempts to find the solution in that case. When this approximation is

[1] HLRZ, c/o KFA Jülich, D-5170 Jülich, West Germany.

0022-4715/90/0500-1077$06.00/0

stable, the replica-symmetric solution is in excellent agreement with Monte Carlo simulations.[7] However, as is the case for the Hopfield model, the replica-symmetric solution may not be stable and one must look for solutions which break replica symmetry.[1,2,7,9]

Normally, replica symmetry breaking is not directly observable; however, for the Hopfield model, Crisanti *et al.*[9] have shown that accounting for replica symmetry breaking leads to a new phase whose thermodynamic equilibrium is incompatible with the replica-symmetric solution. In fact, they searched for the existence of such a phase using standard Monte Carlo techniques.[2,9] Unfortunately, standard Monte Carlo techniques allow only systems of modest size to be simulated, which, due to the small size of the new phase, are insufficient to probe its properties. Indeed, the simulation results of refs. 2 and 10 were ambiguous as to the very existence of the proposed phase. In this paper a "multispin coding" version of the Hopfield model is used which allows very large system sizes and thus the possibility of probing for the existence of the predicted phase.

The next section describes the problem in more detail, including a brief discussion of the replica solution. Following that is a description of the numerical algorithm and the multispin coding procedure. Finally the results are presented and discussed.

2. MODEL

The Hopfield model[8] is a simplification of how processes like associative memory occur in higher-order cognitive systems such as the human brain. The model consist of N two-valued neurons $S_i = \{\pm 1\}$ connected by long-range interactions J_{ij}, which in turn are determined by the states to be "remembered." These interactions can, in their simplest approximation, be written as

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{P} \xi_i^\mu \xi_j^\mu \tag{2.1}$$

where ξ^μ represent the states to be remembered, or the "stored states," and P is the total number of these states. From these symmetric couplings the usual definition of energy can be employed: $E = -\sum_{i \neq j} S_i J_{ij} S_j$. By defining temperature $T = 1/\beta$ as a measure of the noise in the system,[2] i.e., the occasional spontaneous flipping of the neurons, one can consider the statistical mechanics of the model, so that associative memory is construed to occur when the system comes to thermodynamic equilibrium after having been initialized to a given state. If the equilibrium state is "close" to a stored state, then it is said that the stored state has been "remembered."

Apart from this unique choice of couplings, the model behaves in many respects similar to a spin glass, e.g., at low temperatures the system explores only a region of phase space severely limited by its initial state.

The statistical mechanics of the model has been solved in some detail by Amit *et al.*[2] In doing so, they were not interested in the behavior of the system for any particular set of patterns, but in the behavior for an "average" set of stored patterns; hence, the free energy must be averaged over the stored patterns. This is done by the replica method:

$$f = \lim_{N\to\infty} \frac{-1}{\beta N} \langle\langle \ln Z \rangle\rangle$$

$$= \lim_{n\to 0} \lim_{N\to\infty} \frac{-1}{\beta n N} (\langle\langle Z^n \rangle\rangle - 1) \tag{2.2}$$

where Z is the usual partition function and the double brackets indicate an average over all the stored states. For a complete discussion of this solution, the reader is referred to the original paper of Amit *et al.* Here it suffices to say that for the case of replica symmetry, they showed that there indeed existed solutions of the saddle-point equations having large remembrance, i.e., a value of $m^\mu \equiv (1/N)\sum_j \xi_j^\mu S_j$ close to one, provided $\alpha < 0.1378$ ($\alpha \equiv P/N$) and $T_R < T < T_M$. For low temperatures, $T < T_R$, this replica-symmetric solution was shown to be unstable; however, when α is near 0.1378, the value of T_R for which the solution becomes unstable is on the order of 0.01. On present-day computers, it would be difficult to distinguish such temperatures from $T = 0$; hence, for practical purposes the replica-symmetric solution can be considered stable for these values of α.

For the case of replica symmetry breaking, Crisanti *et al.*[9] have shown, using a generalization of Parisi's method, that there exists at $T = 0$ a region $0.1378 \leqslant \alpha \leqslant 0.144$ in which there are thermodynamic equilibrium states having large remembrance. If the existence of such a region, along with its properties, were confirmed by Monte Carlo simulations, it would constitute a direct test showing that replica symmetry breaking has physical consequences. Previous simulations were restricted in the size of the networks which could be probed and, as mentioned above, could not unambiguously confirm these predictions. It was with the hope of directly seeing these effects that the following simulations were carried out.

3. NUMERICAL METHOD

The numerical method used for this research was first published by Penna and Oliveira.[10] In the form discussed here, the program is $O(N)$

times faster than their version. The storage requirements of this algorithm grow like $O(\alpha N^2)$ bits; hence, as will be discussed below, cpu time is the chief limiting factor in implementing this program.

For a zero-temperature Monte Carlo simulation, the updating rule at each time step is simply

$$S_i(t+1) = \operatorname{sgn}(h_i(t)) \tag{3.1}$$

where $h_i(t)$ is the local field at time t. In the Hopfield model, the local field can be written as

$$\begin{aligned} h_i(t) &= \sum_{j \neq i} J_{ij} S_j(t) \\ &= \frac{1}{N} \sum_{j \neq i} \sum_{\mu}^{P} \xi_i^\mu \xi_j^\mu S_j(t) \\ &= \sum_{\mu} m^\mu(t)\, \xi_i^\mu - \frac{P}{N} S_i(t) \\ &= \frac{1}{N} H_i(t) - \frac{P}{N} S_i(t) \end{aligned} \tag{3.2}$$

where $m^\mu(t) \equiv (1/N) \sum_j \xi_j^\mu S_j(t)$ and $H_i(t) \equiv \sum_{j,\mu} \xi_j^\mu \xi_i^\mu S_j(t)$. Obviously, for numerical simulations one need store only the ξ_i^μ and not the complete coupling matrix J_{ij}. Since ξ_i^μ and S_i are two state variables only, it is more efficient to store them as single bits rather than whole words, i.e., one should use the so-called "multispin coding" approach.[10] On a computer with B bits per word, all the stable patterns can be stored in $\alpha N^2/B$ integer words. Compared to algorithms which store one spin in one word, this multispin coding scheme improves by a factor B the total size of the systems which can be simulated.

It is also evident that the m^μ need not be recalculated in full at every site of the lattice. Instead they can be calculated at the time the network is initialized to a given state and then updated only when in asynchronous updating a spin is flipped, or in parallel updating after each parallel update. This saves a factor $O(N)$ in cpu time compared to the approach used in ref. 10 of recalculating these quantities at each lattice site. Within the multispin coding framework, the $m^\mu(t)$ are calculated by

$$Nm^\mu(t) = N - 2 \sum_{j=1}^{N/B} POPCNT(\zeta_j^\mu \otimes \sigma_j(t))$$

where ζ_j^μ is an integer word containing B spins from the state ξ^μ, $\sigma_j(t)$ is an integer word containing B spins from the state $S(t)$, $\otimes$ stands for the "exclusive-or" logical operator, and $POPCNT$ is a function which counts the number of bits set to one in the argument.

In calculating the $H_i(t)$, one sums over the μ in Eq. (3.2) by multiplying Nm^μ by $+1$ (-1) if the bit representing ξ_i^μ is 1 (0). The final step is carried out by using integer division, $S_i(t+1)=[H_i(t)+NP]/(PN+P)$ if the bit representing $S_i(t)=1$, or $S_i(t+1)=[H_i(t)+NP]/(PN-P)$ if the bit representing $S_i(t)=0$.

The cpu bottleneck is in calculating the local fields $h_i(t)$, and in actual running, the program is about as fast, for small networks, as an equivalently coded version of a program which does not use multispin coding. For large networks there is a clear advantage in not using too large a fraction of the computer's main memory. For example, on the HLRZ's Cray-YMP/832, one can simulate system sizes up to $PN=\alpha N^2\approx 2.0\times 10^7$ using the standard, single-spin per word approach. On the other hand, a multispin coding version of the program with the same values of α and N requires less than 350,000 words. Although the Cray-YMP/832 has handled networks up to $\alpha N^2\approx 1.41\times 10^9$ in test runs of the present multispin coding program, cpu time limitations make it impractical to use systems much larger than $\alpha N^2\approx 2\times 10^8$ for actual studies. In calculating the local fields $h_i(t)$, the algorithm described above achieved speeds of 580 conditional adds per microsecond while running in asynchronous updating mode.

Effects attributable to replica symmetry breaking were tested for by first choosing at random $P=\alpha N$ patterns to be stored, then initializing the system to one of the stored patterns and applying the above zero-temperature Monte Carlo algorithm to each neuron in sequence until a stable state or cycle of length 2 was reached. When such a state was reached, the final overlap with the initial state was recorded. At $N=1088$, a total of 100 initial starting states, in each of 400 sets of patterns, for a total of 40,000 initial states were used in order to determine the distribution of the final overlap, while at $N=33{,}968$, averages over only 20 initial starts in each of 20 sets of patterns, for a total of 400 initial states, could be simulated within a reasonable time. (These statistics should be compared to refs. 2 and 9, where at their largest system size of $N=3000$, only 200 initial states were used.) These simulations were carried out for values of α in the range $\{0.110, 0.148\}$, with most of the data being taken in the region of $0.142<\alpha<0.146$, where the results of replica symmetry breaking should have been the strongest.

4. RESULTS AND DISCUSSION

The first quantity measured was the distribution of the final overlaps m_f, given that the initial starting state was one of the stored patterns. Figure 1 shows two typical histograms of m_f for $\alpha=0.142$ and $\alpha=0.144$

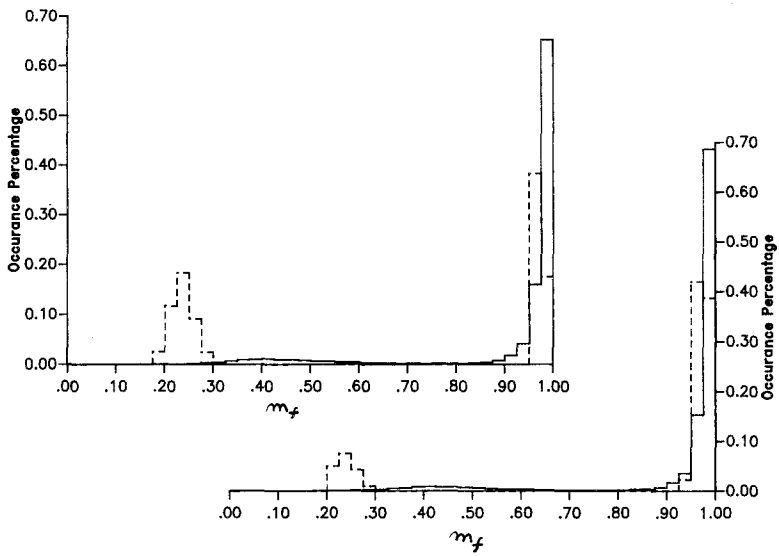

Fig. 1. Histograms for relative occurrence of the final overlap at $\alpha = 0.142$ and $\alpha = 0.144$. The solid line is at $N = 1088$, and the dashed line is at $N = 33{,}968$ for $\alpha = 0.142$ and $N = 32{,}000$ for $\alpha = 0.144$.

each at two values of N. It is obvious from both of these plots that the order parameter m_f has a two-state structure characteristic of first-order phase transitions. This would seem to be in agreement with the theoretical predictions of Amit *et al.* In order to determine to which phase these particular values of α belong, one must examine the finite-size behavior

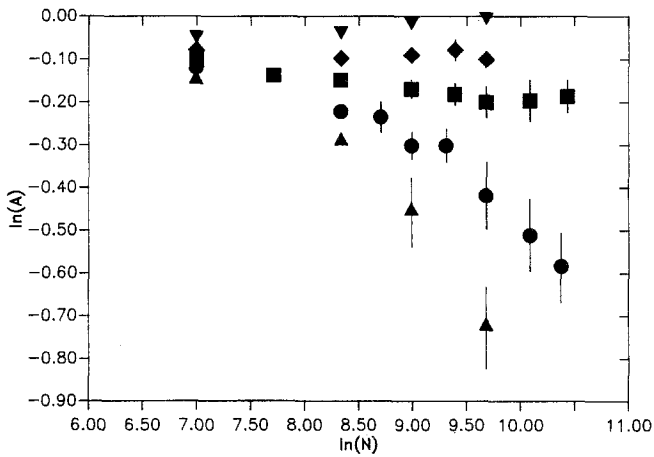

Fig. 2. Plot of ln A versus ln N for $\alpha = 0.136$ (▼), $\alpha = 0.140$ (◆), $\alpha = 0.142$ (■), $\alpha = 0.144$ (●), and $\alpha = 0.146$ (▲).

of the total area under one of the peaks to determine how fast that area is growing or shrinking.

Figure 2 shows such a plot for the area under the large m_f peak, A, at $\alpha = 0.136$, 0.140, 0.142, 0.144, and 0.146. From this plot it is clearly seen that the area under the large-m peak in the histogram is continually decreasing for $\alpha \geqslant 0.144$; hence, these values of α lie in the phase of low recallability. For $\alpha < 0.136$ the area appears to be increasing up to the maximum $A = 1$, indicating a phase of high remembrance. But for $\alpha = 0.140$ and $\alpha = 0.142$, A decreases slightly as N increases before reaching a nonzero plateau, which, within the error bars, is neither decreasing nor increasing as N becomes large, i.e., the two-state structure for m is apparently stable. These values of α correspond to the region in which there are no replica-symmetric, thermodynamically stable states and can be interpreted as a confirmation of replica symmetry breaking. Figure 3 shows the values of A extrapolated to $N = \infty$. At $\alpha \approx 0.138$, where the replica-symmetric approximation predicts a first-order phase transition, A first shows deviations from $A = 1$. At $\alpha = 0.143 \pm 0.001$ the system undergoes a phase transition, which could be either first or second order. From Figs. 1 and 2 it is evident that the large-N behavior becomes manifest only for network sizes considerably larger than those previously simulated. Simulations up to $N = 3000$ as in refs. 2 and 9 do not give a clear indication about the limit $N \to \infty$ and, as a consequence, failed to see the narrow region in which the extrapolated value of A is neither zero nor unity.

Figure 4 shows the extrapolated position of the center of both the large-m and the small-m peaks as a function of α. For comparison, the

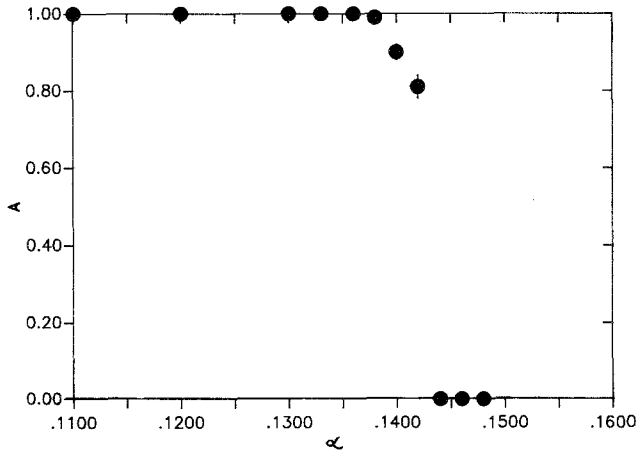

Fig. 3. The $N \to \infty$ extrapolated values of A versus α.

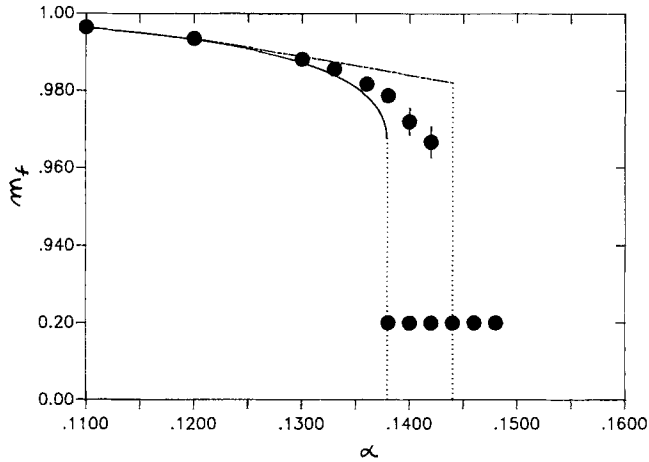

Fig. 4. Plot of m_f versus α. The solid line is the replica-symmetric prediction and the dashed line is the one-step replica-symmetry-breaking prediction of ref. 9.

predictions of the replica-symmetric approximation are shown along with the one-step replica-symmetry-breaking calculation of Crisanti *et al.*[9] In the region in which replica symmetry is broken there exist two stable values for the order parameter m. From Fig. 3 this is interpreted as meaning that in the neighborhood of a finite fraction of the stored states, there

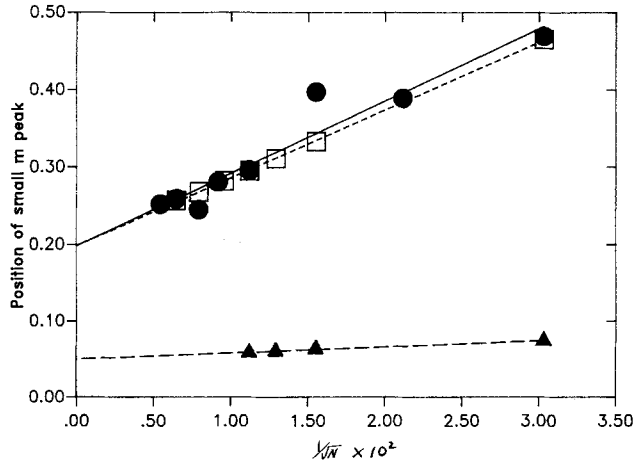

Fig. 5. Position of the dmall-m peak versus $1/\sqrt{N}$. The open squares are at $\alpha = 0.144$ and the closed circles at $\alpha = 0.142$. The lower data show the value of the remanent overlap starting from an arbitrary initial state.

are energy minima having large overlap with the stored state, while for other stored states the nearest energy minima have small overlaps. Below $\alpha \approx 0.138$ almost all stored states have high overlap with an energy minimum, while above $\alpha \approx 0.144$ almost all stored states have only a small overlap with the nearest energy minima.

These energy minima having small overlap do not simply correspond to the remanent overlap seen in spin glasses.[11] Figure 5 shows the position of the center of the small-m peak as a function of $1/\sqrt{N}$, and the average overlap of the final state with the initial state when the initial state is chosen at random. The extrapolation to $N = \infty$ clearly yields different values. Note that the remanent overlap when starting from an arbitrary state is much smaller than that for the case of a spin glass. Hence, the system does distinguish between the stored states and an arbitrary state even in the overloading phase.

It is also of interest to note that simulation results for m_f fall between the values predicted by replica-symmetric calculations and one-step replica symmetry breaking. From this it can be concluded that the limit of many-step replica symmetry breaking is not approached monotonically in this model. If such behavior were true in general, it would provide quite a powerful method for placing bounds on physically meaningful quantities.

One outstanding puzzle is why the theoretical calculations fail to see the states having small-m overlap with the stored states? Both replica-symmetric and one-step replica symmetry breaking predict a transition to $m = 0$, although such a transition is ruled out by the present simulations.

In summary, the numerical simulations presented here confirm the existence of a new phase in the Hopfield model which is predicted only in a replica-symmetry-breaking approximation. The new phase is characterized by a two-state order parameter, although the lower value of m_f is not predicted by theoretical calculations. Furthermore, the numerical algorithm shows that multispin coding can give significant advantages even when it is not faster than a normally coded program.

NOTE ADDED IN PROOF

The qualitative features of the phase diagram in fig. 4 were anticipated in early work by W. Kinzel in *Lecture Notes in Physics, Number 275*, J. L. van Hemmen and I. Morgenstern, eds. (Springer-Verlag, Berlin 1986).

ACKNOWLEDGMENTS

I would like to thank D. Stauffer and R. W. Gerling for their many helpful conversations in relation to this work.

REFERENCES

1. K. Binder and P. Young, *Rev. Mod. Phys.* **58**:801 (1986).
2. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. Lett.* **55**:1530 (1985); *Ann. Phys.* **173**:30 (1987).
3. N. Sourlas, *Nature* **339**:693 (1989).
4. S. Diederich and M. Opper, *Phys. Rev. A* **39**:4333 (1989).
5. D. Sherrington and S. Kirkpatric, *Phys. Rev. Lett.* **35**:1972 (1975).
6. E. Gardner, *J. Phys. A* **21**:257 (1988); W. Krauth and M. Mézard, *J. Phys. A* **20**:L745 (1987); G. A. Kohring, *J. Phys.* (Paris) **51** (1990).
7. J. R. L. de Almeida and D. J. Thouless, *J. Phys. A* **11**:983 (1987).
8. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
9. A. Crisanti, D. J. Amit, and H. Gutfreund, *Europhys. Lett.* **2**:337 (1986).
10. T. J. P. Penna and P. M. C. Oliveira, *J. Phys. A* **22**:L719 (1989).
11. W. Kinzel, *Phys. Rev. B* **33**:5086 (1986).

Communicated by D. Stauffer