Journal of Statistical Physics, Vol. 58, Nos. 5/6, 1990

Effect of Damage in Neural Networks

Eva Koscielny-Bunde[1]

Received August 21, 1989

The effect of damage on the pattern recognition in the Hopfield-model of neural networks is studied. It is assumed that in a damaged network the synaptic efficacies $J_{i,j} = J_{j,i}$ between pairs of neurons S_i and S_j follow the Hebb rule with probability $(1-p)$ and are equal to zero with probability p. Numerical simulations are performed for a net consisting of 400 neurons. It is investigated in detail how the retrieval of noisy patterns and the storage capacity of the net depends, for varying initial noise, on the concentration p of the damaged synaptic efficacies.

KEY WORDS: Damaged neural networks; pattern recognition.

1. INTRODUCTION

In recent years, a large number of neural network models has been investigated (for recent reviews see, e.g., refs. 1–4). Perhaps the most prominent one is the Hopfield model,[5] which combines the early work of McCulloch and Pitts,[6] Hebb,[7] and Little[8] with modern ideas from spin-glass theory.

The Hopfield model has been studied extensively theoretically and has also found practical applications. The long-time retrieval behavior of the net, in the (thermodynamic) limit of an infinite number of neurons and for random nominated patterns, has been explored theoretically by Amit *et al.*[9] and the first attempts to make neural net chips were largely based on the Hopfield model.[10]

The Hopfield network consists of N neurons. Each neuron can be in two states $S_i = \pm 1$: In $S_i = 1$, the neuron is firing, in $S_i = -1$, the neuron is quiescent. Pairs of neurons are coupled by bonds $J_{i,j}$, which represent the synaptic efficacies between them.

[1] Fachbereich Informatik, Universität Hamburg, D-2000 Hamburg 50, West Germany.

0022-4715/90/0300-1257$06.00/0

The effect of learning is to modify the bonds so that the learnt patterns become dynamically stable configurations of the network and this way can be retrieved. In the Hopfield net, the synaptic efficacies $J_{i,j}$ and $J_{j,i}$ between two neurons are identical and the dynamics is purely relaxational: The state of the neuron i at time $t+1$ is obtained from the states of all other neurons at time t by the deterministic dynamic rule

$$S_i(t+1) = \operatorname{sign}\left[\sum_{j=1}^{N} J_{i,j} S_j(t)\right] \tag{1}$$

This way, starting from an initial configuration $\{S(0)\} = (S_1(0), S_2(0),..., S_N(0))$ the system evolves toward the local minima of the energy function (cost function)

$$E = -\frac{1}{2} \sum_{i,j=1}^{N} J_{i,j} S_i S_j \tag{2}$$

The local minima are neuron configurations where each neuron S_i is aligned in parallel with its local field $h_i = \sum_{j=1}^{N} J_{i,j} S_j$.

The couplings $J_{i,j}$ are determined by the M random patterns $\{\xi^\mu\} = (\xi_1^\mu,..., \xi_N^\mu)$ which are to be stored in the network. According to Hebb,[7] one chooses

$$J_{i,j} = \frac{1}{N} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \tag{3}$$

The overlap q^μ between a neuron configuration $\{S\}$ and a random pattern is defined as

$$q^\mu = \frac{1}{N} \sum_{i=1}^{N} S_i \xi_i^\mu \tag{4}$$

A particular pattern $\{\xi^\nu\}$ is called stored by the network if an initial configuration $\{S(0)\}$ near $\{\xi^\nu\}$ develops under the dynamic rule (1) to a state whose overlap with $\{\xi^\nu\}$ is large.

According to Amit *et al.*,[9] only a certain percentage $\alpha = M/N$ of patterns per neuron can be stored. For large nets, rigorous in the limit of an infinite number of neurons, the system approaches "retrieval states" which have at least 97% overlap with the stored patterns, as long as α is below a critical value $\alpha_c \cong 0.14$. As α decreases to zero, the mean overlap increases exponentially fast toward 1, $1-q \sim \exp[-1/(2\alpha)]$. Above α_c, the system flows into a locally stable state, where the remanent overlap is considerably smaller, $q \cong 0.35$.

Numerical simulations for a Hopfield net consisting of 400 neurons have been performed by Kinzel.[11] In the present work, I extend these calculations considerably to include also the effect of damage, a case which so far has not received a rigorous solution. Other types of damaged networks have already been discussed, with respect to different questions, by Kürten.[12]

2. THE HOPFIELD MODEL WITH DAMAGE

I assume that in a damaged network the synaptic efficacies $J_{i,j}=J_{j,i}$ between pairs of neurons follow the Hebb rule (3) only with probability $(1-p)$, with probability p a bond is broken and $J_{i,j}=J_{j,i}=0$. To investigate the Hopfield net under this damage, I have considered a net of 400 neurons.

In the first step, random patterns $\{\xi^\mu\}$, $\mu=1, 2,..., M$, have been generated, and the synaptic efficacies have been calculated according to the Hebb rule. In the second step, the network has been damaged by cutting the symmetric bonds with probability p. In the third step, one of the nominated patterns was chosen. To generate a noisy input pattern, the state of each neuron was changed with probability p_n. Accordingly, the overlap between noisy input pattern and nominated pattern is

$$q^\mu(0)=\frac{1}{N}\sum_{i=1}^{N} S_i(0)\,\xi_i^\mu \cong 1-2p_n \tag{5}$$

In the fourth step, then, the dynamic rule is applied to obtain the final overlap q^μ. The run stopped when the time-dependent overlap $q^\mu(t)$ stayed constant for about 20 time steps. For large M, also oscillations of period 2 with small amplitudes occurred. The run stopped when the amplitudes did not change for about 20 time steps. This procedure was repeated for typically 10^2 initial configurations, chosen from different nominated patterns. For large M, up to 500 initial configurations were considered. By averaging the results, I obtained, for a given number M of patterns, for fixed initial noise p_n, and for fixed damage parameter p, the mean retrieval overlap $\langle q \rangle$.

Figure 1 shows, for fixed initial noise $p_n=0.2$, the mean retrieval overlap $\langle q \rangle$ as a function of the concentration p of damaged synaptic efficacies,

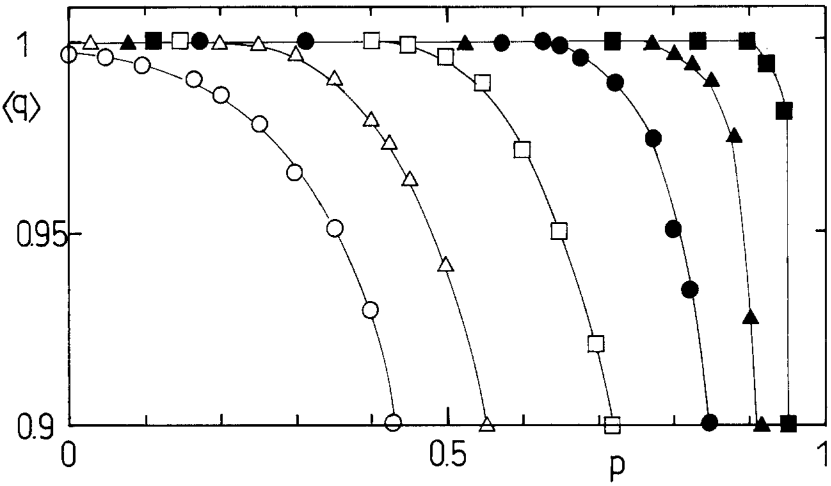

Fig. 1. Average retrieval overlap $\langle q \rangle$ as a function of the concentration p of damaged synaptic efficacies, for several fractions $\alpha = M/N$ of M random patterns, $p_n = 0.2$, and $N = 400$. Different symbols represent different values of α: $\alpha = 4/400$ (■), 10/400 (▲), 16/400 (●), 25/400 (□), 35/400 (△), 40/400 (○).

for various fractions α of nominated patterns. Similar pictures were obtained for other noise levels.

Except for $M = 40$, the curves in the figure approach unity as the damage parameter approaches zero. The data seem to suggest that the approach to unity is exponentially fast, according to

$$1 - \langle q \rangle \sim \exp[-a(M)(1-p)] \tag{6}$$

with $a(M) \sim 1/M$ increasing when M decreases. An accurate analysis of the behavior near unity is beyond the scope of the present work, since it requires much larger systems and a large number of configurations. For a network consisting of N neurons and for a fixed number K of configurations, the minimum nonzero difference between $\langle q \rangle$ and unity equals $2/(NK)$. Accordingly, the exponential decay can only be studied in a p and M range where

$$1 - \langle q \rangle \cong \exp[-a(M)(1-p)] \gg 2/(NK)$$

This may be the case for some intermediate values of M considered in Fig. 1, $M = 35$, and this curve indeed is described well by (6); for an actual test of (6), however, also smaller M must be studied, which requires much larger systems.

The mean retrieval overlap increases monotonically when the noise level decreases. The maximum overlap is obtained for initial configurations consisting of the pure patterns, $p_n = 0$. The dependence of this maximum retrieval overlap on the fraction $\alpha = M/N$ of nominated patterns is shown in Fig. 2, for several values of the damage parameter p.

The upper curve is for the undamaged case, $p = 0$. Below $\alpha = 0.1$, $\langle q \rangle$ cannot be distinguished from unity. As α increases, $\langle q \rangle$ first decreases slightly until $\alpha \cong 0.15$, where $\langle q \rangle \cong 0.96$. When α is increased further, $\langle q \rangle$ decays rapidly, reaching a value of 0.76 at $\alpha = 0.175$. Below 90% retrieval, large fluctuations in the mean retrieval overlap occurred. In this region, the data are based on averages over up to 500 configurations each.

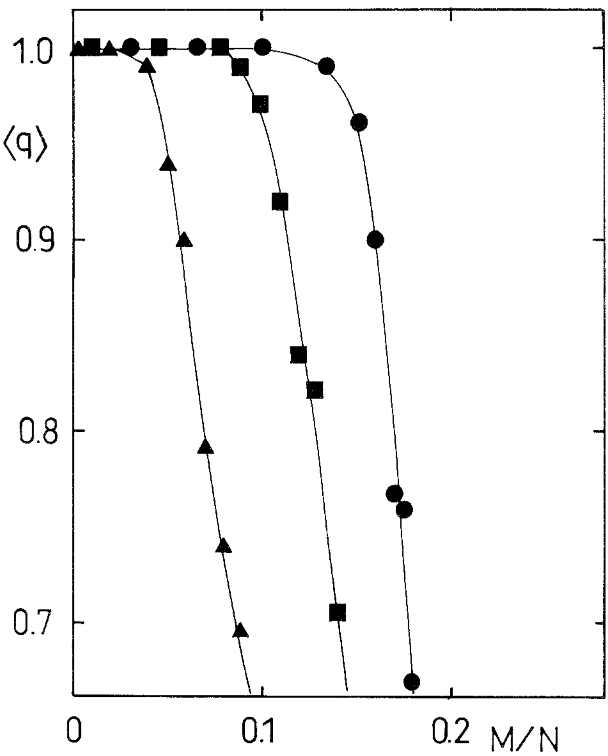

Fig. 2. Average retrieval overlap $\langle q \rangle$ of the pure random patterns versus fraction M/N of the nominated patterns, for $p = 0$ (●), 0.4 (■), and 0.8 (▲).

It has to be noted, however, that Fig. 2 is in contrast to the findings of Kinzel,[11] who also considered a network of 400 neurons. While for small α both results seem to coincide, the curves differ strongly at larger values of α. For example, Kinzel obtained $\langle q \rangle \cong 0.9$ for $\alpha = 0.25$ and $\langle q \rangle \cong 0.7$ for $\alpha = 0.5$. It seems, however, that the results of Amit *et al.*[9] are in favor of Fig. 2. Amit *et al.* found, for very large networks, a sharp transition at $\alpha_c \cong 0.14$. In smaller networks such as the one considered here, one expects qualitatively similar behavior, but with a broadened transition regime, in analogy to the situation in critical phenomena (see, e.g., ref. 13); and this is what is observed in Fig. 2.

The effect of damage is to shift the abrupt decay to smaller values of

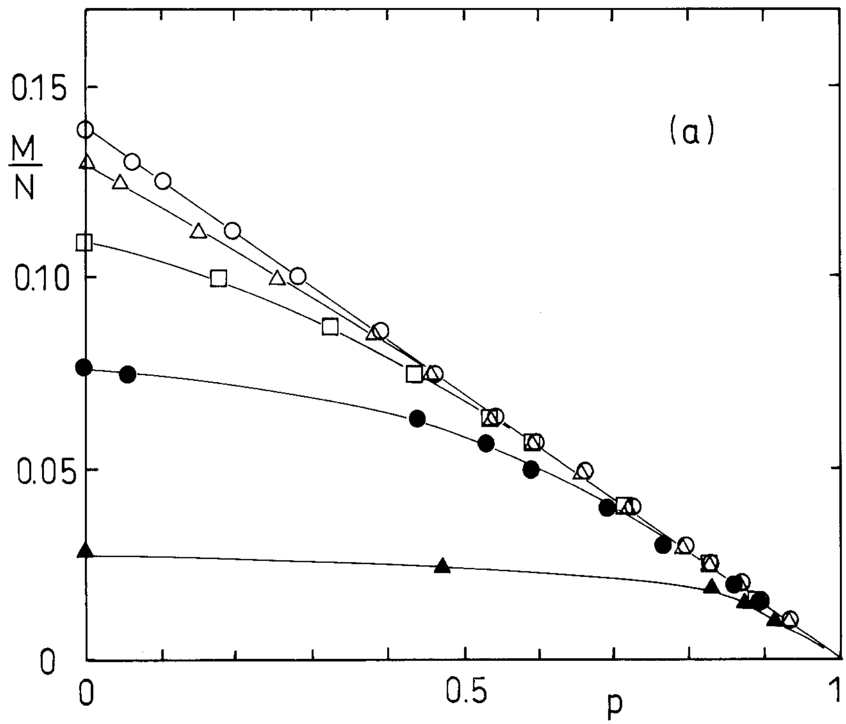

Fig. 3. Phase boundary of the fraction M/N of learnt random patterns as a function of the damage parameter p, for the initial noise $p_n = 0$ (○), 0.1 (△), 0.2 (□), 0.3 (●), and 0.4 (▲). In the left region of the curves in (a), noisy patterns are recognized with less than 0.5% error ($\langle q \rangle \leqslant 0.99$). In the left region of the curves in (b), noisy patterns are recognized with less than 5% error ($\langle q \rangle \leqslant 0.90$).

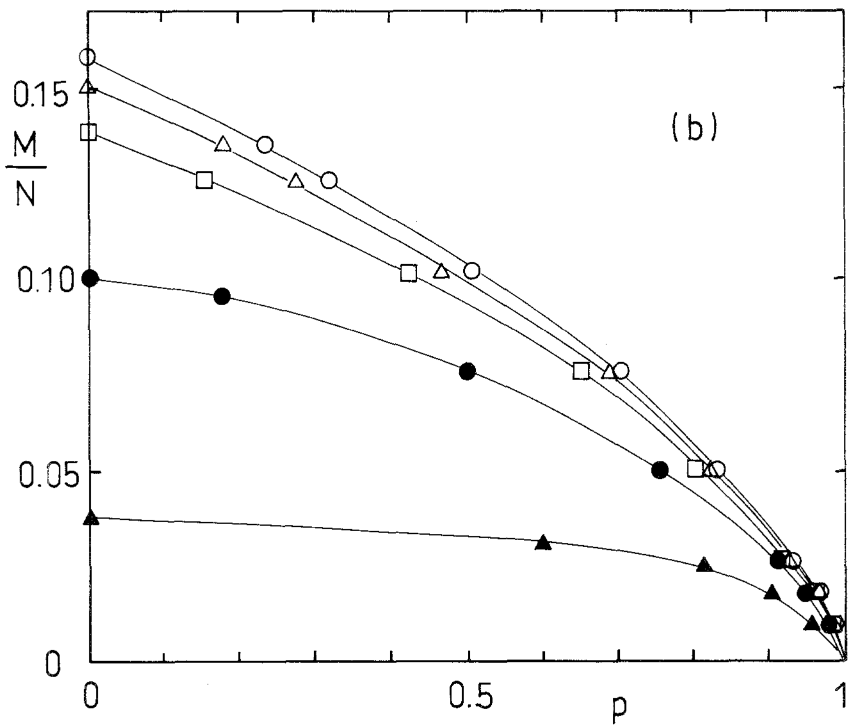

Fig. 3. (*Continued*)

α. The transition region broadens slightly when the damage parameter is enhanced.

Figures 3a and 3b show the maximum number of patterns which can be stored with less than 0.5% and 5% error, as a function of the damage parameter p, for various values of initial noise p_n. For $p_n = 0$, the number of patterns that can be stored is, to a very good approximation, a simple straight line,

$$M(p, p_n = 0) = M(0, p_n = 0)(1 - p) \tag{7}$$

As the initial noise increases, the curve bends down for small p values. There exists a crossover value $p^x(p_n)$, which increases with increasing noise level. Above p^x, $\alpha \equiv M/N$ follows (7); below p^x, α bends down. For large noise, the curve becomes rather flat, showing only a weak dependence on the concentration p of damaged synaptic efficacies.

Qualitatively, Fig. 3b is similar to Fig. 3a, but as the accuracy of retrieval is decreased in Fig. 3b, the curve for the pure patterns no longer follows a straight line, and the maximum value of M/N is enhanced. It is anticipated that the relatively smooth transition between retrieval and non-retrieval is an effect of the finite size of the network. In the idealized case of an infinite number of neurons, the transition should be abrupt, and Figs. 3a and 3b should become identical.

The maximum number of patterns which can be stored with less than 0.5% and 5% error, as a function of the initial noise p_n, is shown in Figs. 4a and 4b for various values of the concentration p of damaged synaptic efficacies. For $p = 0$, the data agree well with the result of Kinzel.[11] The effect of the damage is twofold: The number of random patterns which can be stored is reduced and, with an increasing amount of damage, becomes less sensitive to initial noise.

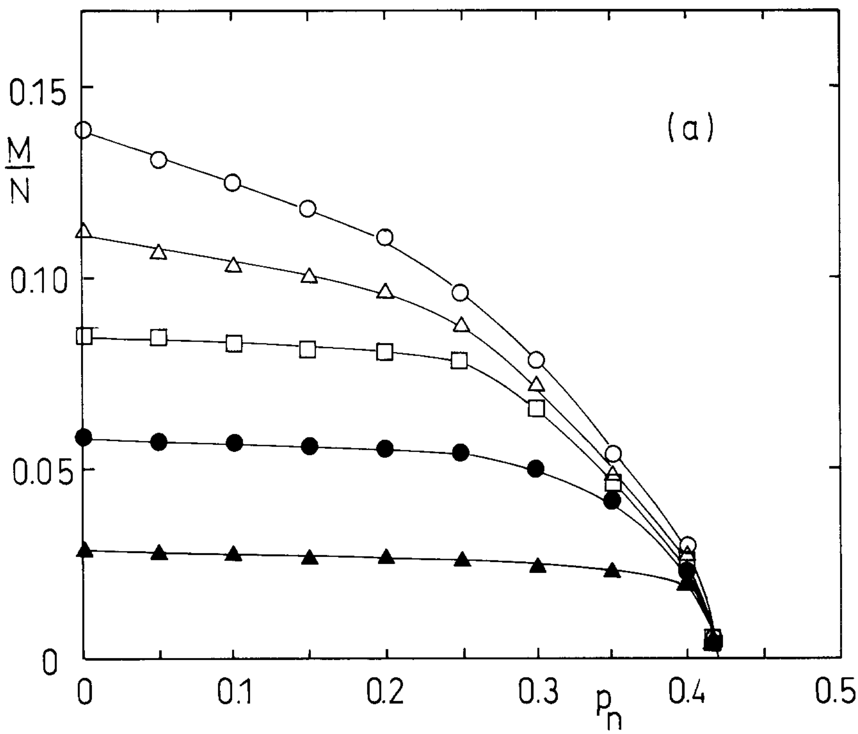

Fig. 4. Maximum fraction M/N of learnt random patterns that are recognized (a) with less than 0.5% error and (b) with less than 5% error, as a function of initial noise p_n, for several values of the damage parameter p: $p = 0$ (○), 0.2 (△), 0.4 (□), 0.6 (●), 0.8 (▲).

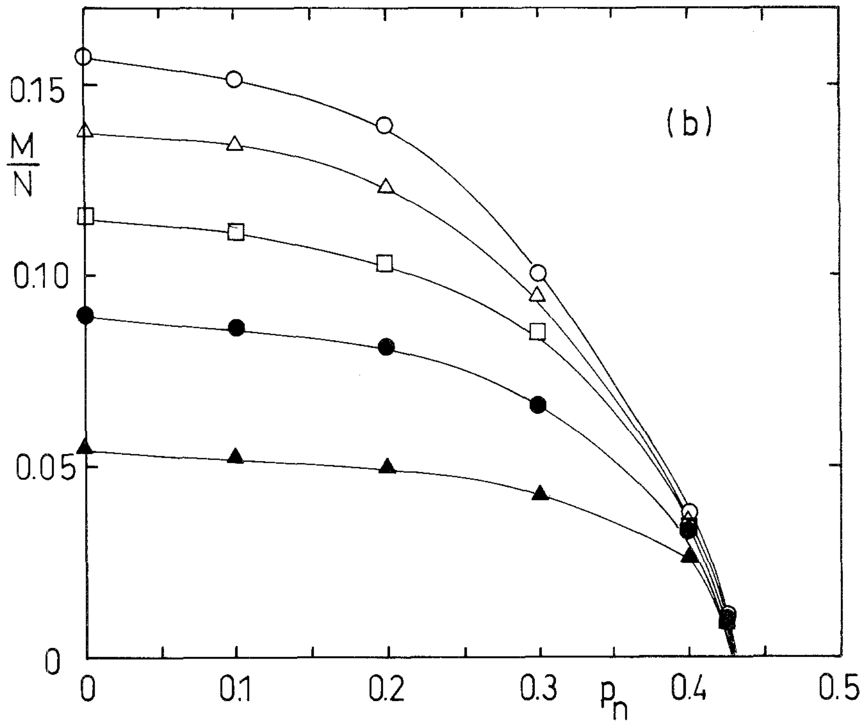

Fig. 4. (*Continued*)

ACKNOWLEDGMENTS

I thank Profs. F. Schwenkel and S. Havlin for valuable discussions on associative memory and neural networks, and Profs. D. Stauffer and A. Bunde for critical reading of the manuscript.

REFERENCES

1. J. A. Anderson and E. Rosenfeld, eds., *Neurocomputing* (MIT Press, Cambridge, Massachusetts, 1988).
2. R. Eckmiller and Ch. v. Malsburg, eds., *Neural Computers* (Springer-Verlag, Berlin, 1988).
3. S. I. Amari, S. Grossberg, and T. Kohonen, eds., *Neural Networks I* (Pergamon Press, New York, 1988).
4. H. Sompolinsky, *Phys. Today* **1988** (December):70 (1988).
5. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
6. W. S. McCulloch and W. Pitts, *Bull. Math. Biophys.* **5**:127 (1943).
7. D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).

8. W. A. Little, *Math. Biosci.* **19**:101 (1974); W. A. Little and G. L. Shaw, *Math. Biosci.* **39**:281 (1978).
9. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys.* **173**:30 (1987).
10. M. A. Sivilotti, M. A. Mahowald, and C. A. Mead, in *Proceedings of the 1987 Stanford Conference*, (MIT Press, Cambridge, Massachusetts, 1988), p. 295.
11. W. Kinzel, *Z. Phys. B* **60**:205 (1985).
12. K. E. Kürten, *Phys. Lett.* **129**:157 (1988).
13. D. Stauffer, *Introduction to Percolation Theory* (Taylor and Francis, London, 1985).

Communicated by D. Stauffer