Associative Memory in Damaged Neural Networks

Norman Hendrich

Universität Hamburg, Fachbereich Informatik, Troplowitzstraße 7, D-2000 Hamburg 54, Germany †

Short title: *Damaged Neural Networks*

PACS number(s): 75.10H — 87.30G — 89.70

Submitted to: *J. Phys. A: Math. Gen.*

Date: 20 Spetember 1990

† e-mail: hendrich@rz.informatik.uni-hamburg.dbp.de

Abstract. The content-addressability of random patterns stored in neural networks trained with iterative learning rules is studied numerically under the effects of synaptic damage. The model of damage is random dilution of couplings with some probability λ after learning.

A simple model based on the distribution of stabilities $\rho(\kappa)$ is proposed, which allows for the calculation of the storage and retrieval properties of the networks as a function of damage λ. Simulations of random sparse connected networks show that the basins of attraction are indeed optimal at storage ratios $\alpha \leq 0.42$.

1. Introduction

Spinglass neural networks have recently attracted considerable attention from physicists. The networks show collective computational properties which make them a paradigm for fault-tolerant massive-parallel computation (Hopfield 1982). The distributed representation of information in the networks—a pattern is not stored locally (with an address) but rather by tiny modifications to all couplings—makes it interesting to study the effects of damage on the networks. Fault-tolerance is a major topic in current research on large parallel computers.

Typically a network consists of a highly interconnected system of N Ising spins (neurons) $S_i = \pm 1$ with couplings J_{ij} from neuron j to neuron i. The state of the S_i at time step $t+1$ is usually computed by parallel dynamics

$$S_i(t+1) = \mathrm{sgn}\big(h_i(t)\big) = \mathrm{sgn}\Big(N^{-1/2} \sum_{j \neq i} J_{ij} S_j(t) \Big), \qquad i = 1, \ldots, N. \tag{1}$$

Other kinds of dynamics are possible and interesting. For example, an asynchronous (serial) dynamics with noise may be more plausible for the modelling of biological neural networks.

The problem is now to adjust the couplings J_{ij} so that the dynamics (1) will show interesting behaviour. Usually the synapses are chosen so that a given set of $P = \alpha N$ states ξ_i^μ $(i = 1, \ldots, N; \mu = 1, \ldots, P)$ of the system become fixed attractors of the dynamics, that is, $\xi_i^\mu = \mathrm{sgn}(h_i(\{\xi_j^\mu\}))$. These states are then memorized by the network in a content-addressable (auto-associative) way: starting from an initial state which partially resembles one of the patterns the system will rapidly evolve into that pattern.

Two questions which immediately arise are what is the storage capacity α of the system and how large are the basins of attraction of the stored patterns? Hitherto most research has concentrated on fully-connected networks. For example, the Hopfield-model (Hopfield 1982) uses symmetric couplings defined by the local learning rule $J_{ij} = (1/N) \sum_\mu \xi_i^\mu \xi_j^\mu$ and has an asymptotic storage capacity of $\alpha_c \approx 0.14$ (Amit, Gutfreund and Sompolinsky 1985, Amit, Gutfreund and Sompolinsky 1987). Previous studies of damage in neural networks (Koscielny-Bunde 1990) have also focused on the Hopfield-model. However, not only is the model unable to store biased patterns, but also the patterns are not stored perfectly and the basins of attraction are tiny above $\alpha \approx 0.06$.

A natural improvement on the Hopfield-model are neural networks with iterative learning rules. To guarantee finite basins of attraction around the stored patterns a positive constant κ is introduced and the couplings J_{ij} are chosen to fulfil the stronger constraints

$$\kappa_{i\mu} = \xi_i^\mu \cdot h_{i\mu} = \xi_i^\mu \Big(N^{-1/2} \sum_{j \neq i} J_{ij} \xi_j^\mu \Big) \geq \kappa > 0. \tag{2}$$

A spherical normalization $\sum_{j \neq i} J_{ij}^2 = N$ is used, because the dynamics (1) is invariant against rescaling of the synapses J_{ij}. A noisy version S_i^μ of pattern ξ_i^μ obtained by flipping a finite fraction of spins at random will be recovered by the network in one iteration if $\xi_i^\mu \cdot h_i(\{S_j^\mu\}) > 0$ for each site i. The content-addressability of the pattern ξ_i^μ will therefore depend on the stabilities $\kappa_{i\mu}$. If all couplings are of the same order (for example in the binary couplings network $J_{ij} = \pm 1$), patterns which differ in no more than $\kappa\sqrt{N}/2$ sites from a stored pattern will be restored in one time step. In fact, the basins of attraction are found to be much larger, of order $O(N)$ and increase with κ (Forrest 1988). A network with the optimal values of the stabilities $\kappa_{i\mu} \geq \kappa_{\text{opt}}$ is said to be 'saturated'.

The values of the couplings J_{ij} are calculated iteratively using the Edinburgh-group algorithm (Gardner 1988a, Forrest 1988). In one learning step a coupling is updated, $J_{ij} = J_{ij} + \Delta J_{ij}$, if the stability of the pattern is less than the desired stability:

$$\Delta J_{ij} = N^{-1} \, \epsilon_{i\mu} \xi_i^\mu \xi_j^\mu \qquad \epsilon_{i\mu} = \Theta\Big(\kappa - \xi_i^\mu \cdot h_{i\mu}\Big). \tag{3}$$

The process is repeated until ΔJ_{ij} is zero for all patterns and neurons. The algorithm is proven to converge towards a solution of (2) in a finite number of learning steps below saturation. To guarantee optimal stabilities, the Minover learning algorithm (Krauth and Mézard 1987) may be used, which at each step learns the pattern with the lowest stability so far. The couplings constructed by these algorithms are not symmetric.

A major advance in the theory of neural networks was achieved by Gardner (Gardner 1987, Gardner 1988) who showed how to calculate the optimal storage capacity of a neuron with C connections in the mean-field replica-symmetric approximation. For unbiased random patterns the storage capacity $\alpha = P/C$ as a function of κ is given by

$$\alpha_c(\kappa)^{-1} = \frac{1}{\sqrt{2\pi}} \int_{-\kappa}^{\infty} dt \, \exp\left(-t^2/2\right) \, (t+\kappa)^2, \tag{4}$$

so that $\alpha_c(0) = 2$. This result for the optimal storage capacity also applies to diluted networks with connectivity $c = C/N \leq 1$ as well as to networks with high-order interactions with $c > 1$. The corresponding distribution of stabilities $\rho(\Lambda)$ can be calculated for a saturated network with parameters κ, $\alpha_c(\kappa)$ (Kepler and Abbott 1988, Gardner 1989),

$$\rho(\Lambda) = \frac{1}{\sqrt{2\pi}} \exp(-\Lambda^2/2) \, \Theta(\Lambda - \kappa) + \frac{1}{2}\Big(1 + \mathrm{erf}(\kappa/\sqrt{2})\Big) \, \delta(\Lambda - \kappa). \tag{5}$$

2. Storage properties of damaged networks

There are several models of network damage. The effects of damaged neurons which remain either permanently active or inactive are rather trivial. This is because the rest of the network decouples from these neurons during iterative learning: If neuron $S_k = 1$, then $\Delta J_{ik} = N^{-1} \xi_i^\mu S_k$ and the value of J_{ik} converges to $\langle \xi_i^\mu \rangle$.

Our model of damage is to destroy synapses randomly with a probability λ in the fully connected network, that is

$$J'_{ij} = C_{ij} J_{ij} \qquad C_{ij} = \begin{cases} 0 & \text{with probability } \lambda \\ 1 & \text{with probability } 1-\lambda. \end{cases}$$

Therefore, after damage each neuron has approximately $C \approx (1-\lambda)N$ couplings, that is, the network has a connectivity $c = C/N \approx 1 - \lambda$. The results (4) for the storage capacity remain valid in diluted networks, if the capacity is defined to be $\alpha' = P/C$. It is obvious that the iterative learning algorithm can be applied to diluted networks as well, updating the remaining couplings only, $\Delta J'_{ij} = C_{ij} \Delta J_{ij} = N^{-1} \, \epsilon_{i\mu} C_{ij} \xi_i^\mu \xi_j^\mu$.

Because learning is not local to each coupling, but rather local to each neuron (the error-mask $\epsilon_{i\mu}$ depends on the local field), the repetition of the learning algorithm in the damaged network will result in a new solution to (2), as long as $\alpha' < \alpha'_c$ (the number of patterns P to be stored has to be smaller than $\alpha'_c N$). The network is thereby able to adapt to the damage introduced by dilution and will recover completely (the value κ'_{opt} being smaller than κ_{opt}).

The natural measure of the damage is the (normalized) distribution of the stabilities $\rho(\kappa_{i\mu})$. As long as all stabilities are positive, all patterns are stored perfectly with finite basins of attraction.

The effect of damage is to deteriorate the distribution of the stabilities. Using a similar argument as with the basins of attraction, the network will store all patterns perfectly up to a level of damage of $\lambda \leq \kappa/\sqrt{N}$.

This is of no use in large networks, but it sets a strict lower limit on the level of damage up to which all patterns remain stored perfectly. However, analogous to the situation of the basins of attraction, one may expect that the networks are fault-tolerant up to a level of damage which is proportional to the number of connections per neuron.

In order to study the effects of damage on the storage of the patterns, random patterns ξ_i^μ were generated and the learning algorithm (3) was iterated until the networks were nearly saturated. Some of the networks were kept below saturation ($\kappa \approx 0.9 \cdot \kappa_c$) because learning becomes quite time-consuming near saturation. An example of the distribution of stabilities after learning may be seen from figure 1 for storage ratio $\alpha = 0.4$. The histogram records the fraction of all stabilities which lie in the intervals $[\kappa, \kappa + \Delta\kappa]$ with a resolution of $\Delta\kappa = 0.05$. The distribution does not show the true functional form expected in saturated networks (5), due to finite-size effects and limited learning time, but it is rather sharply peaked. The corresponding distributions of stabilities $\rho(\kappa)$ after damage with $\lambda = 0.05$, 0.10 and 0.15 are shown as well. Note that the mean value of the stabilities decreases only very slowly, while many $\kappa_{i\mu}$ with low values appear. Finally some stabilities are less than zero, thereby destabilizing the corresponding patterns.

To understand the behaviour of the distribution of stabilities we consider a simple network with constant stabilities. The distribution of stabilities after learning is assumed to be

$$\rho^{(0)}(\kappa_{i\mu}) = \delta(\kappa_{i\mu} - \kappa_0), \tag{6}$$

that is, for every pattern μ at neuron i

$$\kappa_{i\mu} = \frac{1}{||J_i||}\left(\sum_{j\neq i}^{N} J_{ij}\xi_j^\mu \xi_i^\mu\right) = \kappa_0. \tag{7}$$

(with $||J_i|| = (\sum_{j\neq i} J_{ij}^2)^{1/2} = N^{1/2}$). After damage with probability λ we have

$$\kappa_{i\mu}^{(\lambda)} = \frac{1}{||J_i||'}\left(\sum_{j\neq i}^{N} C_{ij} J_{ij}\xi_j^\mu \xi_i^\mu\right) = \frac{1}{||J_i||'}\left(\sum_{j\neq i}^{N} J_{ij}\xi_j^\mu \xi_i^\mu - \sum_{o=1}^{\lambda N} J_{il_o}\xi_{l_o}^\mu \xi_i^\mu\right) \tag{8}$$

(the sum over o is over all $C_{il_o} = 0$). The norm after damage will be $||J_i||' \simeq \sqrt{1-\lambda}\ ||J_i||$.

Because of the condition (7) the terms in the sum over the $J_{ij}\xi_j^\mu\xi_i^\mu$ are correlated, $\langle J_{ij}\xi_j^\mu\xi_i^\mu\rangle \simeq \kappa_0/\sqrt{N}$. The mean value of the stabilities after damage is therefore

$$\kappa_\lambda = \sqrt{1-\lambda}\ \kappa_0. \tag{9}$$

The mean value of the stabilities found in the simulations of damaged networks is shown in figure 2. It agrees very well with the simple prediction (9).

We may assume that the distribution of stabilities will be approximately gaussian after damage. However, because of the correlation (7), the estimation of the variance of this distribution is not straightforward. For small values of λ, the terms in the second sum of (8) are nearly uncorrelated and we expect the variance to be $\sigma^2 \approx (\lambda/(1-\lambda))$. On the other hand, for $\lambda \approx 1$ we have $\sigma^2 \approx 1$.

The numerical data show that the variance of the sum can be approximated by $\sigma \simeq \sqrt{\lambda}$, see figure 3. (Note that the initial distribution of stabilities in the simulations is not a true δ-function, but is given by (5) and contains some stabilities $\kappa_{i\mu} < \kappa_c(\alpha)$, due to finite-size effects and limited learning time. Therefore one would expect a somewhat larger width of $\rho(\kappa)$, especially up to moderate concentration of damage.)

Therefore the distribution of stabilities in the damaged constant-stability network will be

$$\rho^{(\lambda)}(\kappa) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(\kappa-\kappa_\lambda)^2}{\sigma^2}\right] \tag{10}$$

with $\kappa_\lambda = \sqrt{1-\lambda}\,\kappa_0$ and $\sigma = \sqrt{\lambda}$.

This is consistent with the Hopfield-model, where the distribution of stabilities can be calculated analytically. With spherical normalization $\sum_{j\neq i} J_{ij}^2 = (1-\lambda)N$ the Hebb-rule reads

$$J_{ij} = C_{ij}\frac{1}{\sqrt{\alpha N}}\sum_{\mu=1}^{\alpha N} \xi_i^\mu \xi_j^\mu ,$$

and the distribution of stabilities in the fully connected Hopfield-model is given by (Kepler & Abbott 88)

$$\rho_H^{(0)}(\kappa) = \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{1}{2}\Big(\kappa - \frac{1}{\sqrt{\alpha}}\Big)^2\Big].$$

After damage (equivalent to dilution in the Hopfield-model) one has

$$\rho_H^{(\lambda)}(\kappa) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\kappa - \frac{\sqrt{1-\lambda}}{\sqrt{\alpha}}\right)^2\right]. \tag{11}$$

The initial distribution of stabilities in the Hopfield-model can be written as a superposition of δ-functions,

$$\rho_H^{(0)}(\kappa) = \int \mathrm{d}\kappa_0 \ \rho_H^{(0)}(\kappa_0) \ \delta(\kappa - \kappa_0).$$

Therefore after damage one should have

$$\rho_H^{(\lambda)}(\kappa) = \int \mathrm{d}\kappa_0 \ \rho_H^{(0)}(\kappa_0) \ \rho^{(\lambda)}(\kappa)$$

and this requires $\sigma = \sqrt{\lambda}$.

With the distribution of stabilities (10) the fraction of stored patterns can be calculated as a function of initial stability κ_0 and damage λ. The fraction of patterns with stability $\kappa_{i\mu} < \kappa_{\min}$ can be written

$$\Gamma(\kappa < \kappa_{\min}, \lambda) = \int_{-\infty}^{\kappa_{\min}} \mathrm{d}\kappa \ \rho^{(\lambda)}(\kappa),$$

that is,

$$\Gamma(\kappa < \kappa_{\min}, \lambda) = \frac{1}{2}\Big[1 + \mathrm{erf}\Big(\frac{\kappa_{\min} - \kappa_\lambda}{\sigma}\Big)\Big]. \tag{12}$$

Therefore $1 - \Gamma(0, \lambda)$ gives the fraction of patterns that are stored at one neuron. The model is confirmed very well at low storage densities $\alpha \leq 0.4$, see figure 4. Error-bars are not shown, but are approximately the size of the symbols. Simulations with α up to 0.7 show that the constant-stability model agrees at least qualitatively with the simulations even at high levels of the storage density.

The deviations between the constant-stability model and the simulations were always less than 3%. They are due to the fact that the initial distribution of stabilities (5) is not a δ-function. (Naturally the model can be improved by integrating over the initial distribution of stabilities as shown above in the Hopfield-model.)

Any pattern that is stable with probability $1-\Gamma(0,\lambda)$ at one neuron will be stored perfectly in the network with probability $(1-\Gamma(0,\lambda))^N$. That is, at any finite value of damage the patterns will not be stored perfectly in the network. However, the overlap of the stored pattern with the initial pattern will remain very large. For example, at $\alpha = 0.4$ ($\kappa_c \approx 1.25$) the network will store the patterns with less than 1% errors up to a concentration of damage $\lambda_c = 0.224$. At $\alpha = 0.1$ the patterns are stored with less than 1% errors below $\lambda_c \leq 0.625$. An example of the fraction of perfectly stored patterns is shown in figure 5 ($N = 256$).

3. Basins of attraction after damage

To study the retrieval properties of the damaged networks numerically, random test-patterns $\xi_i^{\mu,r}$ having an initial overlap $m_0 = 1/N \sum_{i=0}^{N} \xi_i^\mu \xi_i^{\mu,r}$ with pattern ξ_i^μ were generated and iterated to stability, using the parallel dynamics (1). The measures of retrieval are the mean final overlap m_f of the iterated test-patterns with pattern ξ_i^μ and the fraction of perfectly recalled patterns f_p.

The values of m_c and f_p were recorded for various values of initial overlaps m_0. Typical simulations involved several hundred test-patterns per m_0 so as to obtain reliable statistics. Computer-time considerations limited system sizes to networks with less than $N = 512$ neurons. Most of the simulations were carried out with $N = 256$ and $N = 128$.

The fraction of perfectly recalled patterns at given storage ratio is well described by the scaling hypothesis

$$f_p(m_0)/(d(\lambda) - f_p(m_0)) = a_1 \exp(N a_2 (m_0 - m_c)) \qquad (13)$$

where m_c is the critical initial overlap a test pattern must have with a stored pattern to be recalled (the size of the basin of attraction). Because after damage some patterns may not be stored perfectly, the function $d(\lambda)$, depending on the damage λ, allows for a maximum fraction of perfectly recalled patterns of less than one. In the networks without damage $d(0) = 1$, confirming the finite-size scaling behaviour found by (Forrest 1988) for serial dynamics.

Figure 6 shows the values of m_c as determined by least-squares fits to equation (13) for various values of the storage ratio α. (Note that most of the error bars — typically the size of the symbols — were dropped in the interests of clarity. However, the errors of m_c near the critical damage at given storage ratio are much larger.)

With increasing level of damage the critical initial overlap m_c to ensure recall of patterns increases rather slowly, up to a certain level of damage, at which the networks lose all content-addressability within a small transition regime.

Recently two research groups have proposed to use the distribution of stabilities $\rho(\kappa)$ as a starting point for a calculation of the basins of attraction (Kepler and Abbott 1988, Gardner 1989). The mean final overlap $m_1(t+1)$ of a configuration with a pattern ξ_i^μ after one step of the parallel dynamics (1) starting from an initial-state having overlap m_0 can be calculated using Gardner's methods, giving

$$m_1 = F_\kappa(m_0) = \int_{-\infty}^{\infty} \mathrm{d}\Lambda \; \rho(\Lambda) \; \mathrm{erf}\Big(\frac{m_0 \Lambda}{\sqrt{2(1-m_0^2)}} \Big) \tag{14}$$

In extremely diluted networks ($C \leq \ln N$) the (unstable) fixed point $m_S = F(m_S)$ of equation (14) should give the basins of attraction, because the tree of ancestors (neurons at times $t-i$) influencing any neuron at time t contains no loops.

In fully connected networks, however, feedback-effects occur and the critical overlap m_c for perfect recall, as determined by the fit to the scaling hypothesis (13), is usually larger than the fixed point of equation (14). Using a numerical study (Kepler and Abbott 1988) proposed to use the fixpoint m_F of

$$F_\kappa(m_F) = (1 + m_F)/2 \tag{15}$$

to describe the basins of attraction in fully connected nearly saturated networks.

The values of the critical initial overlap m_c found numerically were always smaller than the corresponding results of the fixpoint m_F of equation (15), where the distribution of stabilities was taken from the simulations. As an example, compare the values of m_c found from the simulations with the predictions m_S and m_F at $\alpha = 0.3$ and 0.4, see figure 7. The results presented here would seem to be in agreement with previous simulations of iterative learning networks (Forrest 1988) and the Pseudoinverse-model (Kanter and Sompolinsky 1987). The estimate m_c found here is always bracketed by the values of m_F and m_S.

As our programs allowed for the efficient simulation of diluted networks some results on the basins of attraction m_c in random sparse connected networks, which compare favourably with the theoretical prediction m_S are presented: Figure 8 shows the values of the critical initial overlap m_c as determined by the fit to equation (13), as well as the values of m_S and m_F. The transition from the behaviour of the fully connected to the sparse connected network as given by (14) can clearly be seen. The size of the basin of attraction at storage ratios $\alpha < 0.42$ is not optimal (that is, $m_c > 0$) in the simulations, because the random overlaps of order $O(N^{-1/2})$ a test pattern has with other stored patterns can not be neglected ($m_{\rm random} \approx 0.07$ at $N = 256$). A test pattern $\xi_i^{\mu,r}$ having a very small overlap $m_i < N^{-1/2}$ will have a larger overlap with some other pattern and the network will recall this other pattern.

4. Conclusion

Neural networks trained with iterative learning rules prove to be fault tolerant against synaptic damage at any value of the storage ratio α. While the patterns are not stored perfectly at finite concentration λ of synaptic damage, the probability of errors is very small up to high values of λ. The distribution of stabilities $\rho(\kappa)$ found in damaged networks can be described well by a simple model based on a constant-stability network.

Any damage introduced by dilution can be compensated by repetition of iterative learning (relearning), as long as the storage ratio α' in the diluted network is kept below α'_c.

The current theoretical models to describe the basins of attraction in fully-connected networks, based on the distribution of stabilities, can be applied to the damaged networks as well. However, the prediction m_F seems a rather crude approximation of the basins of attraction at storage ratios $\alpha < 0.5$ in networks below saturation.

The calculation of the basins of attraction in random sparse connected networks (Gardner 1989) is confirmed by the simulations. In fact, even very small sparse connected networks have optimal basins of attraction at low storage ratios $\alpha < 0.42$.

This implies that associative memories using the neural network paradigm are useful even in the presence of faults. This will allow large-scale implementation of neural networks. However, the information content of the networks ($N^2\alpha$ bit) scales only as $O(1/\ln N)$ with the $N^2 \ln \sqrt{N}$ bit necessary to store the couplings. Therefore it might be interesting to study the binary-couplings neural network under the effects of damage, because this model has an information content of $O(1)$ and is very attractive for implementation.

Acknowledgments

I am grateful to the Fachbereich Informatik der Universität Hamburg for extensive computer time on their VAX 8550 and Apollo-workstations. I acknowledge financial support by the Studienstiftung des Deutschen Volkes.

References

Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev.A* **32** 1007–1018

——1987 *Ann. Phys., NY* **173** 30–67

Forrest B M 1988 *J. Phys. A: Math. Gen.* **21** 245–255

Gardner 1987 *Europhysics Letters* **4** 481–485

——1988a *J. Phys. A: Math. Gen.* **21** 257–270

——1989 *J. Phys. A: Math. Gen.* **21** 1969–1974

Hopfield J J 1982 *Proc. Nat. Acad. Sci.* **79** 2554–2558

Kanter I and Sompolinsky H 1987 *Phys. Rev.A* **35** 380–392

Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657–1662

Koscielny-Bunde E 1990 *J. Stat. Phys* **58** 1257–1260

Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745–L752

Krauth W, Nadal J P and Mézard M 1988 *J. Phys. A: Math. Gen.* **21** 2995–3011

Figure captions

Figure 1. Example of the distribution of stabilities $\rho(\kappa)$ in a nearly saturated network ($\alpha = 0.4$, $N = 256$) after learning and after damage with probabilities $\lambda = 0.05$, 0.10 and 0.15.

Figure 2. Mean value κ_λ of the distribution of stabilities $\rho^{(\lambda)}(\kappa)$ as a function of damage λ, divided by $\sqrt{1-\lambda}\,\kappa_0$. ($\alpha = 0.1$ $\bullet$, 0.2 $\circ$, $\ldots$, 0.7 $\otimes$).

Figure 3. Width σ of the distribution of stabilities $\rho^{(\lambda)}(\kappa)$ as determined from the simulations, divided by $\sqrt{\lambda}$. See text.

Figure 4. Fraction of stored patterns (with stability $\kappa_{i\mu} > 0$) per neuron as a function of damage λ in nearly saturated networks. ($N = 256$, $\alpha = 0.1$ $\circ$, 0.2 $\diamond$, 0.3 $\star$, 0.4 $*$, 0.6 $\oplus$). The curves show the theoretical model with the corresponding initial stabilities (from above) $\kappa_0 \approx \kappa_c(\alpha) = 2.99$, 2.04, 1.58, 1.30, 0.91, 0.5 and 0.2.

Figure 5. Fraction of perfectly stored patterns f_p in nearly saturated networks as a function of damage λ for different values of the storage ratio α. ($\alpha = 0.1$ $\bullet$, 0.2 $\circ$, 0.3 $\diamond$, 0.4 $\star$, 0.5 $*$, 0.6 $\oplus$. $N = 256$.)

Figure 6. Critical initial overlap m_c determined numerically as a function of damage λ for various values of the storage ratio α.

Figure 7. Critical initial overlap m_c ($\bullet$) as determined numerically compared with the theoretical predictions m_S ($\circ$) and m_F ($\diamond$), $\alpha = 0.3$, 0.4.

Figure 8. Critical initial overlap m_c in random sparse connected networks. The lower curve gives the fixpoint of $m_S = F_K(m_S)$ and the upper $2F_K(m_F) = 1 + m_F$. Connectivity: $c = 1.0$ $\star$, 0.2 $\bullet$, 0.1 $\circ$, 0.05 $\diamond$. $N = 256$.

Figure 1.

Example of the distribution of stabilities $\rho(\kappa_{i\mu})$ after learning ($\alpha = 0.4$, $N = 256$) and after damaging with probabilities $\lambda = 0.05$, $\lambda = 0.10$ and $\lambda = 0.15$.

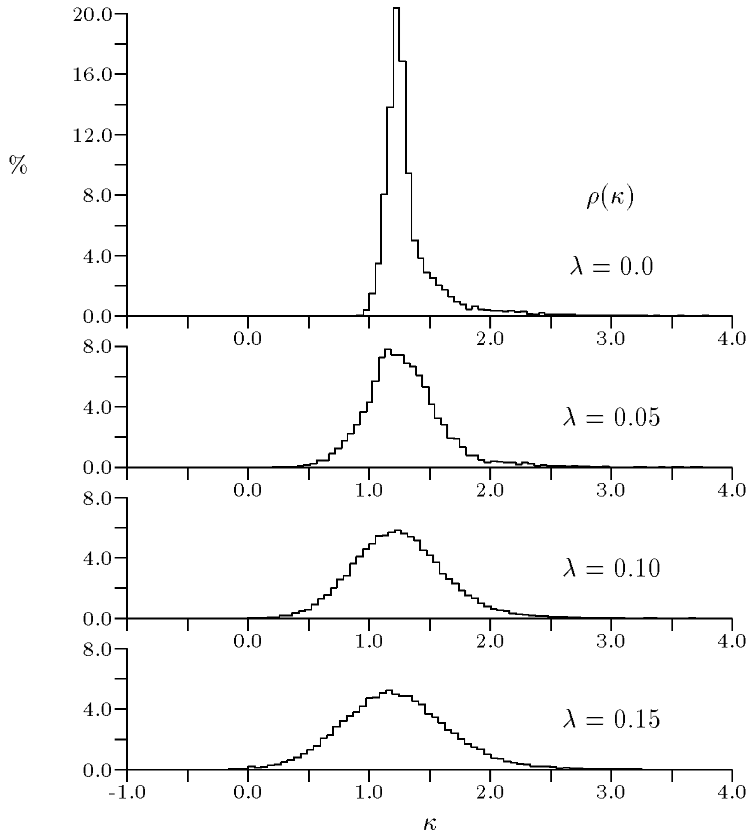

Figure 2.

Mean value κ_λ of the distribution of stabilities $\rho(\kappa_{i\mu})$ as a function of damage λ, divided by $\sqrt{1-\lambda}\kappa_0$. ($\alpha = 0.1$ $\bullet$, 0.2 $\circ$, ..., 0.7 $\otimes$).

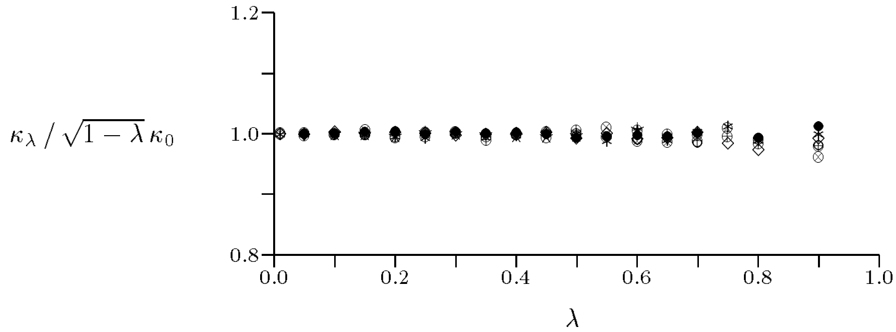

Figure 3.

Width σ of the distribution of stabilities $\rho(\kappa_{i\mu})$ as determined numerically, divided by $\sqrt{\lambda}$. See text.

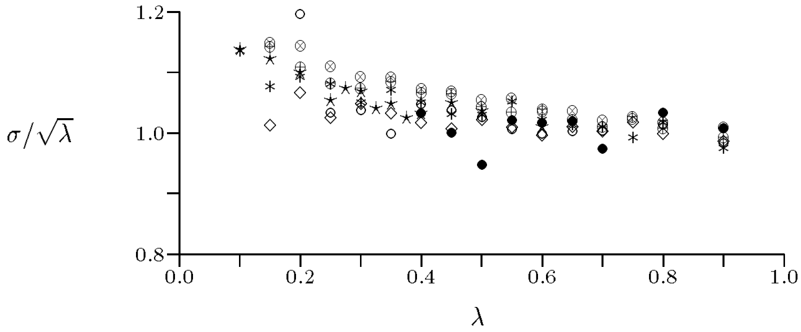

Figure 4.

Fraction of stored patterns (with stability $\kappa_{i\mu} > 0$) per neuron as a function of damage λ in nearly saturated networks. ($N = 256$, $\alpha = 0.1$ $\diamond$, 0.2 $\diamond$, 0.3 $\star$, 0.4 $*$, 0.6 $\oplus$). The curves show the theoretical model with (from above) $\kappa_0 = 2.99$, 2.04, 1.58, 1.30, 0.91, 0.5 and 0.2.

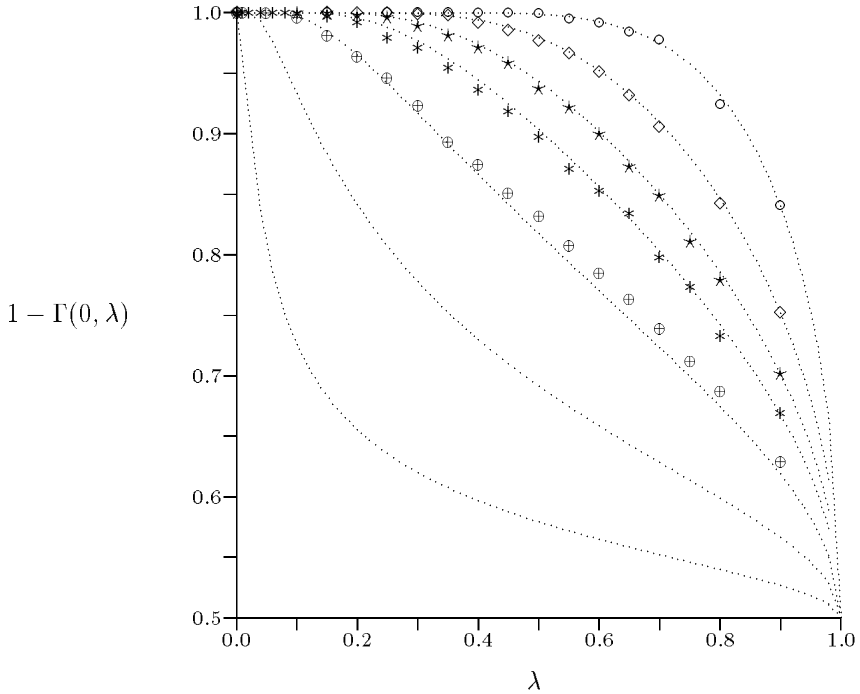

Figure 5.

Fraction of perfectly stored patterns f_p (all stabilities $\kappa_{i\mu} > 0$) averaged over a large number of coupling matrices as a function of damage λ for storage ratios $\alpha = 0.1\ \bullet$, $0.2\ \circ$, $0.3\ \diamond$, $0.4\ \star$, $0.5\ *$, $0.6\ \oplus$. $N = 256$.

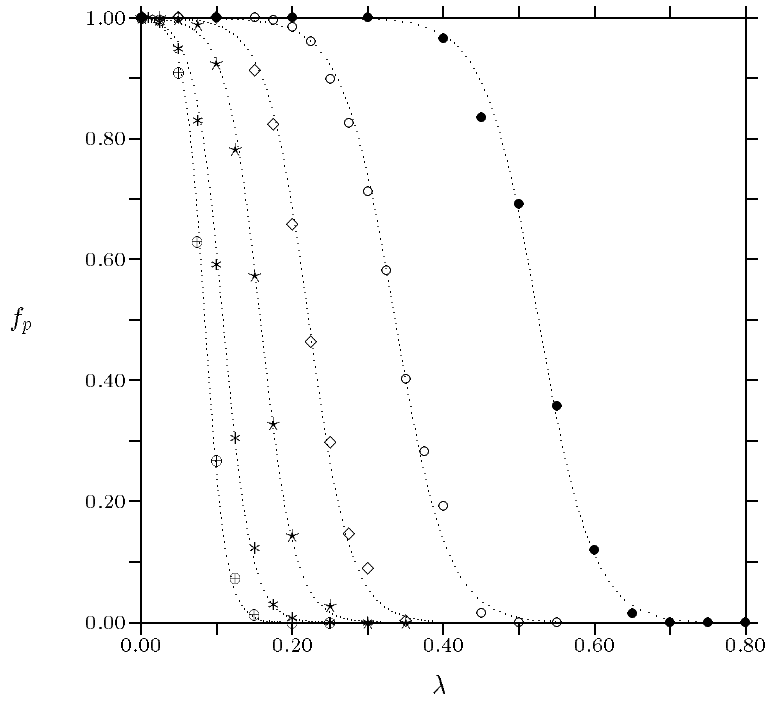

Figure 6.

Critical initial overlap m_c as a function of damage λ in fully connected networks ($N = 256$) with storage ratios $\alpha = 0.1$ $\bullet$, 0.2 $\circ$, 0.3 $\diamond$, 0.4 $\star$, 0.5 $*$.

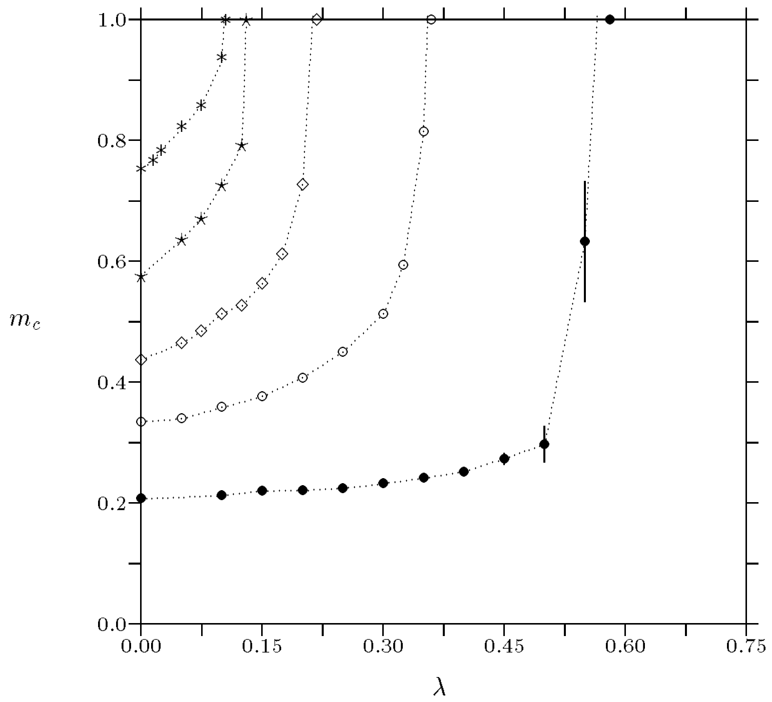

Figure 7.

Critical initial overlap m_c ($\bullet$) as determined numerically compared with the theoretical predictions m_S ($\circ$) and m_F ($\diamond$) as a function of damage λ.

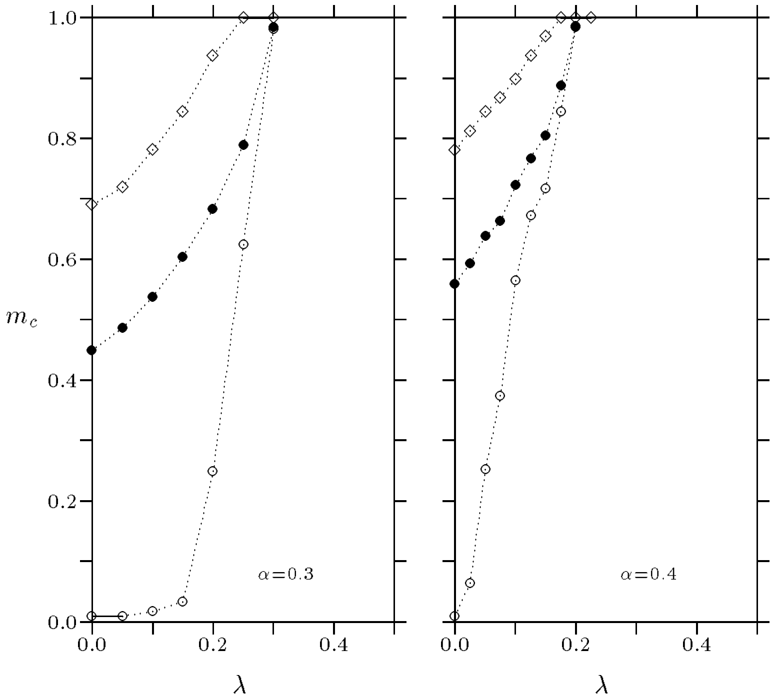

Figure 8.

Critical initial overlap m_c in diluted networks. The lower curve gives the fixpoint of $m_c = F_K(m_c)$ and the upper $2F_K(m_c) = 1 + m_c$. Simulations: N=256. Connectivity: c=1.0 $*$, c=0.2 $\bullet$, c=0.1 $\circ$, c=0.05 $\diamond$.