A finite-size scaling investigation for Q-state Hopfield models: storage capacity and basins of attraction

1992 J. Phys. A: Math. Gen. 25 5919

(http://iopscience.iop.org/0305-4470/25/22/019)

J. Phys. A: Math. Gen. 25 (1992) 5919–5929. Printed in the UK

A finite-size scaling investigation for Q-state Hopfield models: storage capacity and basins of attraction

Thomas Stiefvater† and Klaus-Robert Müller‡

Institut für Logik, Komplexität und Deduktionssysteme, Universität Karlsruhe, Kaiserstrasse 12, D-7500 Karlsruhe, Federal Republic of Germany

Received 12 May 1992

Abstract. The storage capacity of a Q-state Hopfield network is determined via finite-size scaling for parallel dynamics and $Q \leqslant 8$. The results are in good agreement with theoretical predictions by Rieger. The basins of attraction and other associative memory properties are discussed for $Q = 4$, 6. A self-controlling Q-state model with improved basins of attraction is proposed.

1. Introduction

Recently binary Hopfield networks [1] have been generalized to models where the neuron behaviour is more diversified [2–7]. One side of this development draws us closer to biology where neurons possess a richer stucture than the formal binary neurons of Hopfield. On the other side the associative memory properties of more evolved models are important and should be investigated for possible applications. Especially, grey-level neurons could in principle be an interesting representation for the distributed processing of grey-tone data [4–6].

In this paper we show that for applications, grey-level neurons do not seem to be appropriate, because of the low storage capacity and slow dynamics of the network. Nevertheless they possess some interesting features. Our results are in good agreement with theoretical results by Rieger [4], Kohring [5], and Mertens *et al* [6] who claim that the storage capacity $\alpha_c \sim 1/Q^2$, where Q is the number of grey levels.

In section 2 we will define the Q-state Hopfield model. In section 3 the basins of attraction are investigated and the scaling results for the storage capacity are given in section 4. In section 5 we define a self-controlling Q-state model with better associative properties than the one defined in section 2. The basins of attraction are investigated in section 6 and some conclusions are given in section 7.

2. The Q-state Hopfield model

An interesting generalized type of associative memory is the Q-state Hopfield model. While binary neurons can take only values $s_i \in \{\pm 1\}$, a Q-state neuron S_i can take

† E-mail address: tstief@ira.uka.de
‡ E-mail address: klaus@ira.uka.de

0305-4470/92/225919+11$07.50

values in the discrete set $M = \{\sigma_1, \ldots, \sigma_Q\} \in \mathbb{Z}$. These values σ_i are selected equidistantly and symmetrically around 0, e.g. for $Q = 4$ the states are chosen to be $\{-3, -1, 1, 3\}$. The expectation value of the mean pattern-vector length is then given by

$$\beta = \left(\frac{N}{Q}\sum_{k=1}^{Q}(\sigma_k)^2\right)^{\frac{1}{2}} = \langle\|\xi^\mu\|\rangle = \left\langle\left(\sum_i \xi_i^{\mu\,2}\right)^{\frac{1}{2}}\right\rangle$$

for uniformly distributed patterns ξ_i^μ ($i = 1, \ldots, N$ and $\mu = 1, \ldots, p$). These patterns chosen from M are stored by a normalized Hebbian learning rule

$$J_{ij} = \frac{1}{\beta^2}\sum_\mu \xi_i^\mu \xi_j^\mu \tag{1}$$

as attractors of the network dynamics. A parallel deterministic updating

$$S_i(t + \Delta t) = \mathrm{dyn}\,(\gamma h_i(t)) \tag{2}$$

is used where the local field h_i is given by

$$h_i(t) = \sum_{j \neq i}^{N} J_{ij} S_j(t) \tag{3}$$

and the generalized step function dyn with steepness parameter $\gamma = 1$ is defined as

$$\mathrm{dyn}(z) = \begin{cases} \sigma_1 & \text{if } z \in [u(\sigma_1), o(\sigma_1)] \\ \vdots & \qquad\qquad (\sigma_i \in M) \\ \sigma_Q & \text{if } z \in [u(\sigma_Q), o(\sigma_Q)] \end{cases} \tag{4}$$

The lower and upper bound $u(\sigma_i)$ (respectively $o(\sigma_i)$) of the interval have to be chosen carefully in order to map h_i correctly onto σ_i. From a signal-to-noise analysis we can see that

$$|R| < \tfrac{1}{2}|o(\sigma) - u(\sigma)| \tag{5}$$

should hold for the modulus of the noise term $|R|$ in order to stabilize a given pattern ξ correctly, if the respective σ is taken to be in the middle of the interval $|o(\sigma) - u(\sigma)|$ (e.g. for $Q = 4$, $\sigma_3 = 1$, $o(\sigma_3) = 2$, $u(\sigma_3) = 0$). We will see that (5) can be violated depending on the initial state S_0. From theoretical results [4–6] we also know that $\alpha_c \sim 1/Q^2$, so it is clear that large network sizes have to be used during the simulations. Both facts make our investigations very time consuming.

So the calculation of the local field was implemented as

$$h_i = \frac{1}{\beta^2}\sum_{\mu=1}^{p} \xi_i^\mu (\langle \xi^\mu, S\rangle - S_i \xi_i^\mu) \tag{6}$$

for efficiency reasons since $N \gg p$. In order to speed up the algorithm only the set of changing neurons I was taken into account:

$$\langle \xi^\mu, S(t)\rangle = \sum_{i=1}^{N} \xi_i^\mu S_i(t) = \langle \xi^\mu, S(t-1)\rangle + \sum_{i \in I} \xi_i^\mu (S_i(t) - S_i(t-1)) \tag{7}$$

where $I = \{i \,|\, S_i(t) \neq S_i(t-1)\}$. Note that for binary neurons (7) can be calculated using bit operations while for Q-state neurons this is no longer possible. This fact yields a rather slow performance.

3. Basins of attraction

It is not clear *a priori* which order parameter should be chosen to measure the extent to which we have retrieved a certain pattern. At present nothing is known about the geometry of the basins of attraction. They could be spherical or polyeders, so we define a number of order parameters, each of which could measure a different aspect of the attractor geometry. One can take the cosine between the normalized vectors ξ^μ and S

$$m^\mu = \frac{\langle \xi^\mu S \rangle}{\|\xi^\mu\| \, \|S\|} = \frac{\sum_i \xi_i^\mu S_i}{\|\xi^\mu\| \, \|S\|} \tag{8}$$

where $\|A\| = \sqrt{\sum_i (A_i)^2}$ and we define $m^\mu = 0$ for $S = 0$. Equivalently one can use an error rate

$$\mathrm{err}^\mu = \frac{1}{2NQ} \sum_i |\xi_i^\mu - S_i| \tag{9}$$

or the Tanimoto measure [8], which is a compromise between (8) and (9).

$$M^\mu = \frac{\langle \xi^\mu S \rangle}{\|\xi^\mu\|^2 + \|S\|^2 - \langle \xi^\mu S \rangle} = \frac{\langle \xi^\mu S \rangle}{(\xi^\mu - S)^2 + \langle \xi^\mu S \rangle} \,. \tag{10}$$

The Tanimoto measure was originally used to measure the similarity between two sets A and B. It is defined as the ratio between the cardinalities of the intersection $\mathrm{card}(A \cap B)$ and the union $\mathrm{card}(A \cup B)$ of the sets A and B. For real vectors $\mathrm{card}(A \cap B)$ is replaced by the scalar product $\langle \, , \, \rangle$ and $\mathrm{card}(A \cup B)$ by the difference of the sum of the moduli and the scalar product. One could also think of other order parameters as

$$M_n^\mu = \frac{\langle \xi^\mu S \rangle}{\left(\|\xi^\mu\|^2 + \|S\|^2 - \langle \xi^\mu S \rangle\right)^n} \,.$$

In the following we will use the cosine m^μ as a common measure for the initial state m_0. It is demonstrated in figures 1 and 2 that the order parameters (8)–(10) show comparable behaviour regarding the basins of attraction, i.e. regarding the measuring of the type of retrieved state in the associative memory.

For $Q = 4, 6$ and $N = 1024$, 1536, 2048 we average over 512 initial states at different memory loadings α. Clearly a breakdown of all target-state similarity measures can be seen at the interval borders where (5) is violated. The diagrams are meant to show a qualitative behaviour of the model.

For figure 1(*a*) we see a breakdown in the cosine measure (8) for $m_0 = \frac{2}{3}$, which can be explained by signal-to-noise analysis. For $Q = 4$ the σ_i are defined as above, and at the interval border 2 (between $\sigma_4 = 3$ and $\sigma_3 = 1$), choosing $m_0 = \frac{2}{3}$ and $\xi_i = 3$, it is only the sign of the noise term R that decides whether the local field h_i tends towards 3 or 1. So the dip seen in the curves is an artefact caused by the special form of the dyn function. If the dyn function is modified, e.g. the σ_i are not chosen equidistantly any more, or the steepness parameter γ is altered, then the characteristic m_0 also changes. The dip at $m_0 = \frac{2}{3}$ is increased in depth with growing α.

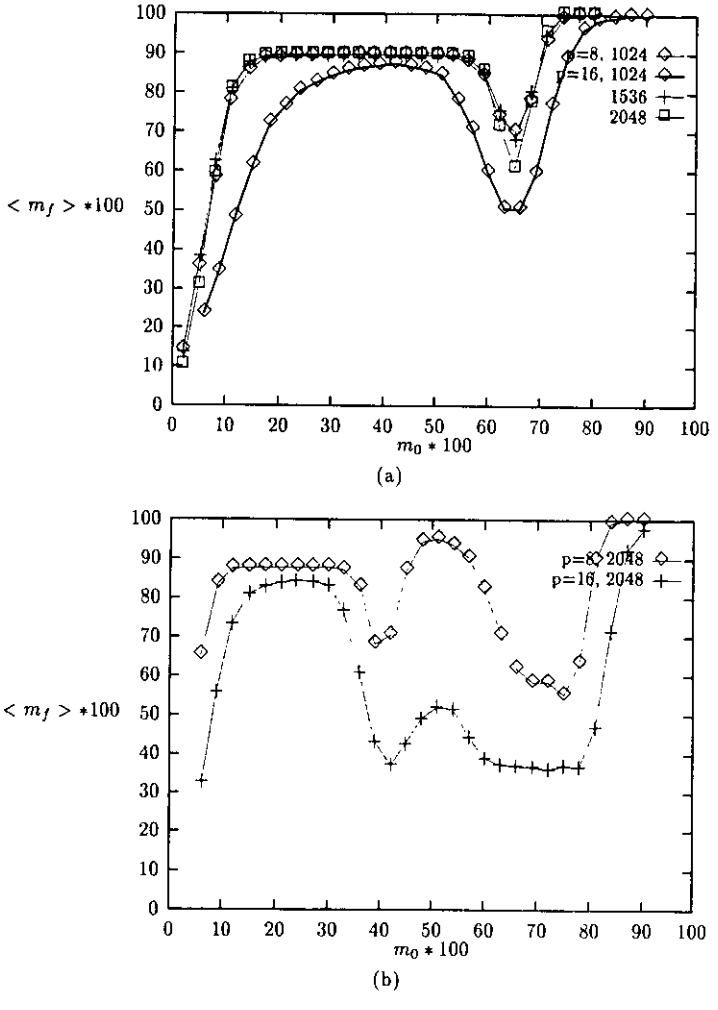

Figure 1. Cosine measure $\langle m_f \rangle$ for (*a*) $Q = 4$ and $N = 1024$, 1536, 2048, where $\alpha = 0.0156$ (0.008) for the lower (upper) curve and for (*b*) $Q = 6$ and $N = 2048$, where $\alpha = 0.008$ (0.004) for the lower (upper) curve.

We see that the upper curve in figure 1(*a*) with $\alpha = 0.008$ is rather sharp, the plateau at $\langle m_f \rangle \sim 0.89$ corresponds to a state where all $\{\pm 3\}$ are mapped onto $\{\pm 1\}$. Therefore we expect $\langle m_f \rangle$ to be $\langle m_f \rangle \sim 2/\sqrt{5} \sim 0.89$. For this state $\langle M_f \rangle$ is expected to be 0.5, since 50% of the values in both vectors coincide. The error rate is $\langle \mathrm{err}_f \rangle \sim 0.13$. The lower curve is drawn for $\alpha = 0.015$ which is closer to the critical $\alpha_c \sim 0.018$.

For higher Q values we find a higher number of dips. In figure 1(*b*) we see $Q = 6$ with $\alpha = 0.004$ (upper curve) and $\alpha = 0.008$ (lower curve) both of which are below $\alpha_c \sim 0.0083$. The system size for the simulation is $N = 2048$. We find the following mixture states: the states with $\langle m_f \rangle = 0.88$ arise because nodes which should be in the $\{\pm 5\}$ or $\{\pm 3\}$ state actually take a $\{\pm 1\}$ state if the initial overlap m_0 is in the interval [0.1, 0.35]. In this configuration the other measures take the values $\langle M_f \rangle \sim 0.3$ and $\langle \mathrm{err}_f \rangle \sim 0.16$.

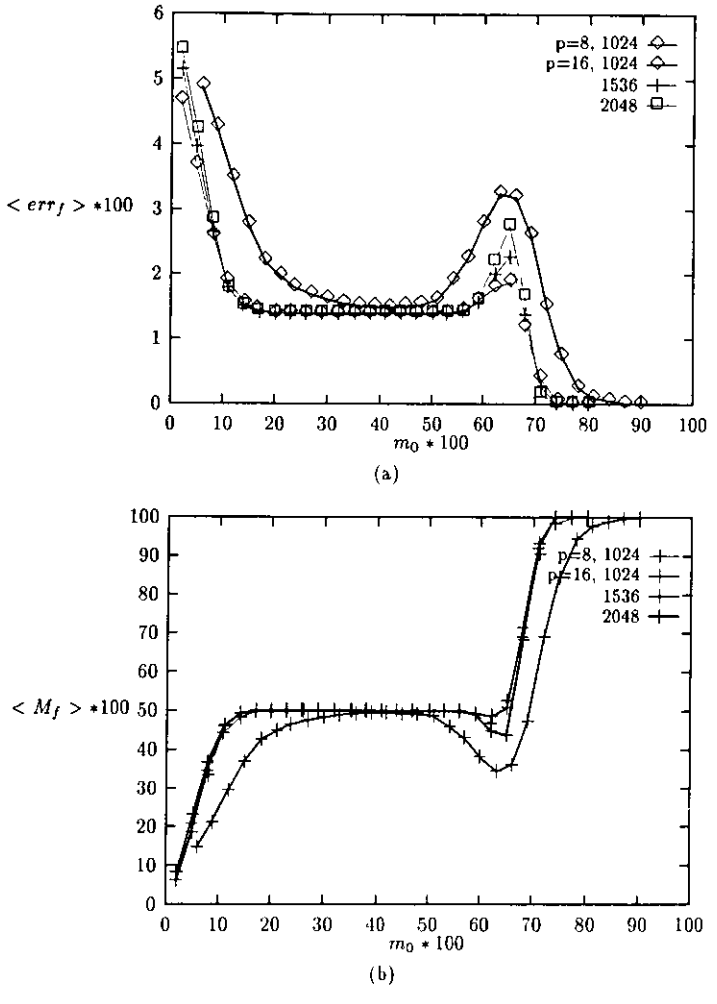

Figure 2. Mean final-state characteristics of the Q-state network dynamics (2) as a function of initial overlap $m^\mu(0) = m_0$ for $Q = 4$ and $N = 1024, 1536, 2048$, where $\alpha = 0.0156$ (0.008) for the lower (upper) curve is measured (a) by the error rate $\langle \mathrm{err}_f \rangle$ and (b) by the Tanimoto measure $\langle M_f \rangle$.

For an initial state $m_0 \in [0.45,\ 0.6]$, we see from the Tanimoto measure that $\frac{2}{3}$ of the states stay unchanged. A superposition of the transitions $[\frac{1}{2}(5 \to 3) + \frac{1}{2}(5 \to 1)]$ or $[\frac{1}{2}(5 \to 3) + \frac{1}{2}(3 \to 1)]$ would keep $\frac{1}{3}$ of the states unchanged. Both transitions would yield the correct error rate of $\langle \mathrm{err}_f \rangle \sim 0.075$, but the value $\langle m_f \rangle \sim 0.95$ holds only for the transition $[\frac{1}{2}(5 \to 3) + \frac{1}{2}(3 \to 1)]$.

We see that the three measures all provide a good way to describe the multitude of spurious states in figures 1 and 2, i.e. the structure of the basins of attraction.

It is also clear that a Q-state network has a rather small robustness against distortion, dependent on the special form of the dyn function. The transition from correct recognition to non-recognition proceeds in several steps, in the sense that the system undergoes a number of transitions through various phases, where we get spurious states with correct signs but wrong vector length (cf figures 1 and 2), e.g.

for $Q = 4$ a state $\xi_i^\mu = -3$ is mapped onto $S_i = -1$. The phase borders depend on the number of Q-states, the steepness parameter γ of the dyn function and α as we have seen in the figures. In the following we will use the cosine m^μ, to study more of the model's interesting features.

4. Critical storage capacity

In order to calculate the critical storage capacity α_c we use a finite-size scaling ansatz for the weight of the retrieval peak f. Therefore 2000–3000 pure patterns ξ_i (from different random pattern sets) are used as intitial states and the final overlaps m_f are determined. We assume that f, the weight of the retrieval peak (cf figure 4) or the percentage of final overlaps with $m_f > 0.9$ exhibits the following form for $\alpha_0 \sim \alpha_c$ and large N:

$$\frac{f(\alpha_0, N)}{1 - f(\alpha_0, N)} = A \exp\left[BN(\alpha_c - \alpha_0)\right] \tag{11}$$

which is expected for a first-order phase transition at α_c. $\alpha_0 = p/N$ and A, B are constants. For fixed Q and α_0 the values $\ln(f(\alpha_0, N)/1 - f(\alpha_0, N))$ should be straight lines intersecting for different values of N (figure 5(*a*)):

$$g_N(\alpha_0) = \ln(A) + BN(\alpha_c - \alpha_0)\ . \tag{12}$$

For two appropriately chosen fixed values of α_0, $g(\alpha_0)$ from (12) is plotted as a function $f(1/N)$ for different values of N (figure 5(*b*)):

$$\alpha_0 = \frac{1}{N}\left(\frac{\ln(A) + g(\alpha)}{B}\right) + \alpha_c\,. \tag{13}$$

α_c can then be taken from the intersection of both curves. In table 1 a summary for different values of Q is given and the corresponding values of α_c are shown. In figure 3 we compare the data of our simulation to results of an analytic calculation of Rieger [4]. Both results are in good agreement. The prediction by Rieger for high values of Q

$$\alpha_c \sim \frac{0.3}{Q^2} \tag{14}$$

is given in figure 3 together with our estimates of α_c. In the large-Q regime ($Q \geqslant 6$) the agreement is very good, while for low values of Q we find the same kind of deviation as Rieger did in his analytic study. Figure 4 shows the weights of the retrieval peak below and above α_c. In figure 5 we show the results of interpolation (13) for $Q = 4$ and different values of N.

Table 1. Critical storage capacities α_c for different values of Q.

Q	α_c
4	0.021 821 (±0.000 221)
5	0.012 724 (±0.000 268)
6	0.008 028 (±0.000 219)
8	0.004 387 (±0.000 189)

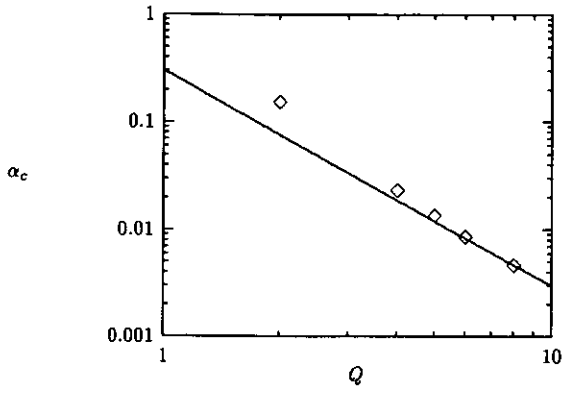

Figure 3. Comparison of the critical storage capacities α_c for different values of Q with the theoretical result [4].

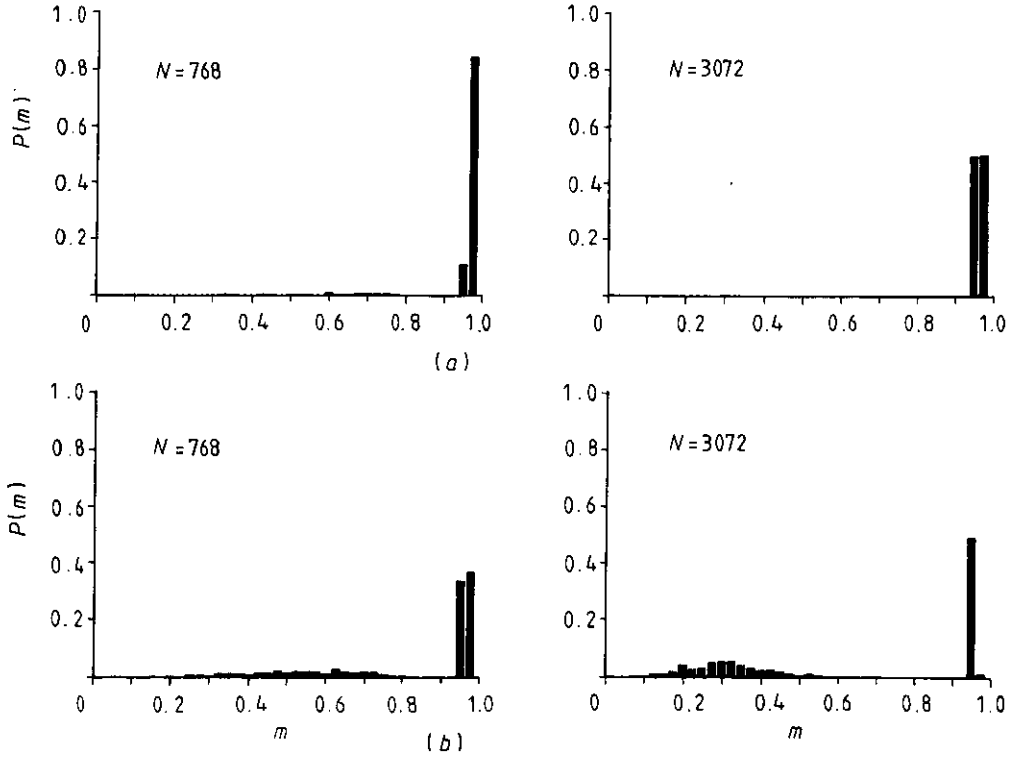

Figure 4. Weight of the retrieval peak below and above $\alpha_c = 0.018$ for (*a*) $\alpha = 0.016$ and (*b*) $\alpha = 0.039$ for $Q = 4$ and different network sizes N.

5. A self-controlling model

We now propose a network or rather a dynamical system, which suppresses the spurious states encountered in section 3 by self-adjusting γ in the dyn function, thereby increasing the basin size drastically. In the following we study the dynamics

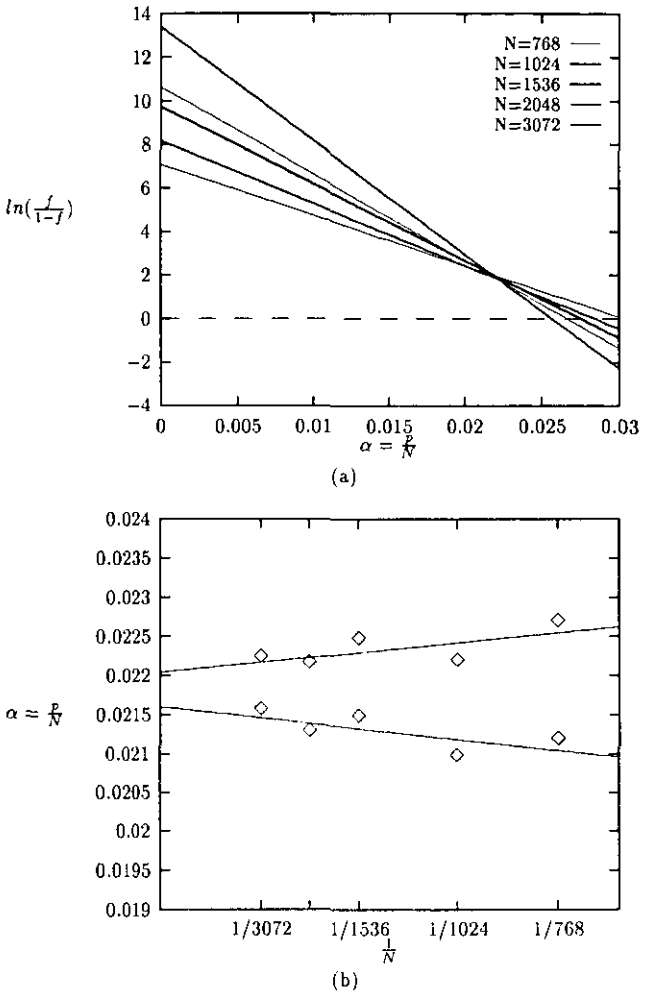

Figure 5. Finite-size scaling interpolation for $Q = 4$.

of this model,

$$S_i(t + \Delta t) = \mathrm{dyn}(\gamma h_i(t)) = \mathrm{dyn}\left(\frac{\beta}{\|S\|}\sum_{j\neq i}^{N} J_{ij}S_j(t)\right). \tag{15}$$

The key idea is that the state vector is normalized, i.e. $\gamma = \beta/\|S\|$. Of course a normalization as in (15) is a non-local process so we actually leave the field of biologically plausible neural networks.

We see that, although the same dips as in figure 1(*a*) occur in figure 6, a normalized dynamics as in (15) is more robust against distortion, because the spurious states are destabilized by the self-adjusting γ. For example if the sign of the pattern state is correct but its length is incorrect, e.g. $|S_i| < |\xi_i^\mu|$, then this state is not stable since $\beta/\|S\|$ is increased above 1. Therefore in the next step of (15) the network will be able to reach a higher value of $|S_i|$ and step by step the original pattern length is

reconstructed adaptively. A corresponding correction mechanism holds also for the other direction, where $|S_i| > |\xi_i^\mu|$. At the fixed point we find $\beta \sim \|S\|$ and $\gamma \sim 1$ as in section 2.

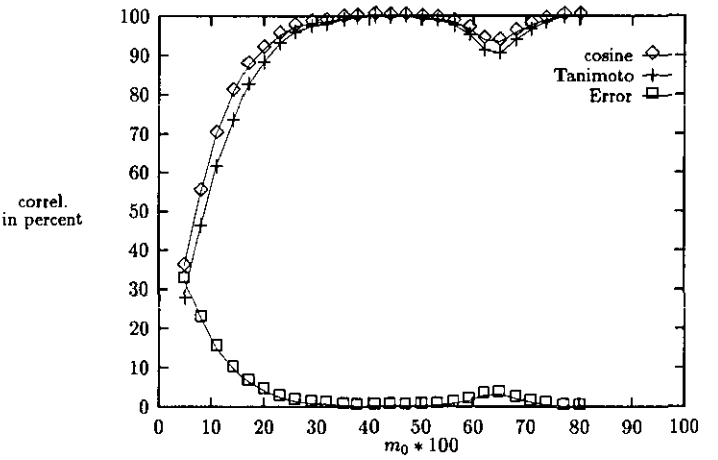

Figure 6. Comparison between the different measures for the self-controlling model, where $\alpha = 0.008$, $N = 1024$, $Q = 4$.

So the normalization does not affect the storage capacity, since the fixed points of the mean-field equations of (15) are not changed by renormalization, but (15) is crucial for an improvement of the fault-tolerance, respectively the basin size of the model.

In principle one could also define a neural network using in (6) M^μ (respectively M_n^μ) instead of m^μ. This increases the storage capacity and the basin size, because the noise term is reduced nonlinearly since $M^\mu \leqslant m^\mu$. The respective synaptic couplings then have to be changed from (1) to either non-local or higher-order weights.

6. Basins of attraction for the self-controlling model

In order to calculate the critical overlap m_c for the self-controlling model again we use a finite-size scaling ansatz. For this a set of 500 distorted initial states (from different random pattern sets) with overlap m_0 is chosen and the average final overlap m_f is determined. The distorted states are chosen equally distributed over M.

The steep decrease of the weight of the retrieval peak f close to m_c is interpolated for different network sizes N. f is assumed to show the following form for $m_0 \sim m_c$:

$$\frac{f(m_0, N)}{1 - f(m_0, N)} = A \exp\left[BN(m_c - m_0)\right] \tag{16}$$

similar to the finite-size scaling ansatz for the critical storage capacity in section 4. A and B are constants. m_c can then be taken from the intersection of both curves in figure 7 and a summary for $Q = 4$ can be determined from table 2.

For the dynamics (2) we obtain $m_c \geqslant \frac{2}{3}$ showing less fault tolerance than the self-controling model.

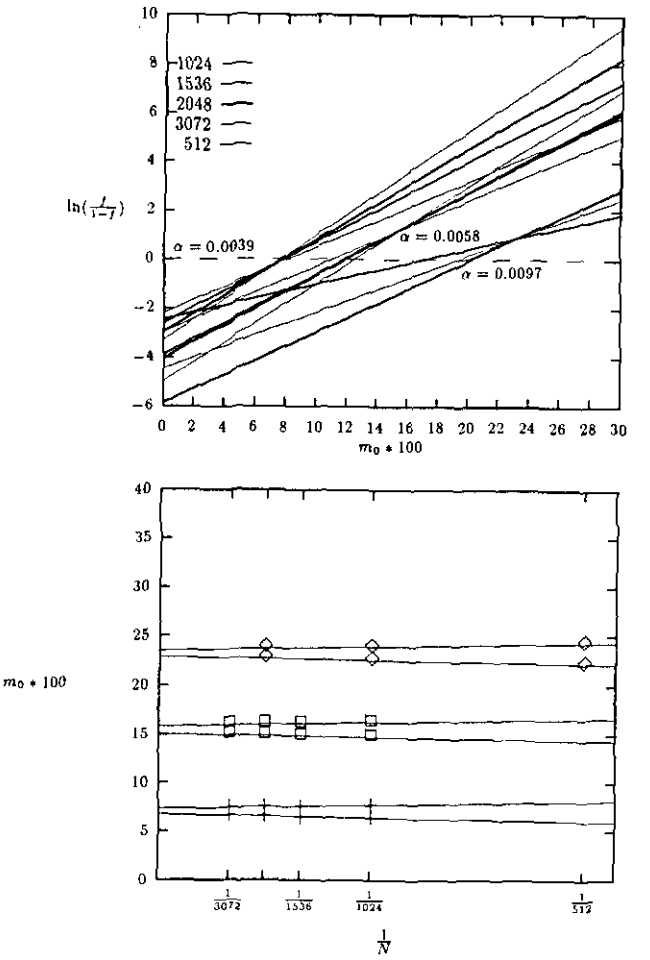

Figure 7. Scaling results for the critical overlap m_c for $Q = 4$, different N and α.

Table 2. Critical values m_c for $Q = 4$ and different storage capacities α for the self-controlling model.

α	m_c in % $(Q = 4)$
0.0039	7.053 801 (±0.285 765)
0.0049	15.434 90 (±0.392 263)
0.0097	23.199 53 (±0.317 813)

7. Conclusions

We have shown in a number of very time-consuming simulations how the important parameters of the associative memory α_c, m_c scale with system size. The results are in good agreement with theoretical predictions.

Concerning robustness of the associative memory, we could see that the Q-state model undergoes several transitions from correct recognition to spin-glass phase

(cf [9]). The phase borders depend on Q, γ and α. The type of spurious state can be explained in terms of different measures, each covering its own aspect of the basin shape. In order to get a network with high robustness being able to reconstruct all grey tones correctly below m_c we propose a self-controlling network model. This model improves the retrieval qualities of the Q-state Hopfield network drastically.

From our investigations we conclude that a grey-tone network defined as above cannot be used for real practical purposes, because of its low storage capacity and its slow performance. In pattern recognition, it seems to be unreasonable to use a Q-state Hopfield network as proposed in [4–6]. One should rather spend more time for the preprocessing of the images, and then employ a more efficient model, possibly involving sparse coding [10] or a grey-level perceptron [11].

Acknowledgments

One of us (K-R M) gratefully acknowledges financial support by Landesgraduiertenförderung Baden-Württemberg and A Glenz. It is a pleasure to thank R Kühn for valuable discussions. The computing time was partly granted by the Informatik Rechner Abteilung Karlsruhe. K-R M is supported by Landesgraduiertenförderung Baden-Württemberg.

References

[1] Hopfield J J 1982 *Proc. Nat. Acad. Sci. USA* **79** 2554
Little W A 1974 *Math. Biosci.* **19** 101
[2] Hopfield J J 1984 *Proc. Nat. Acad. Sci. USA* **81**
[3] Yedida J S 1989 *J. Phys. A: Math. Gen.* **23** 2265
[4] Rieger H 1990 *J. Phys. A: Math. Gen.* **23** L1273
[5] Kohring G A 1991 *J. Stat. Phys.* **62** 563
[6] Mertens S, Köhler H M and Bös S 1991 Learning Grey toned Patterns in Neural Networks *Preprint* Göttingen
[7] Kühn R 1990 *Proc. 11th Sitges conf. (Springer Lecture Notes in Physics vol 368)* ed L Garrido (Berlin: Springer) p 19
Kühn R, Bös S and van Hemmen J L 1991 *Phys. Rev.* A **43** 2084
[8] Tanimoto T 1958 *An elementary mathematical theory of classifications and prediction* IBM Corp. internal report
[9] Kühn R in preparation
[10] Stiefvater T, Müller K-R and Janßen H in preparation
Müller K-R 1992 *Proc. Int. Conf. on Artificial Neural Networks* ed I Aleksander and R Taylor (New York: Elsevier)
[11] Gerl S, Bauer K and Krey U 1992 *Learning with Q-state clock neurons—optimal storage capacity and adatron algorithm* submitted