Information theoretical performance measure for associative memories and its application to neural networks

Mathias Schlüter, Oliver Kerschhaggl, and Friedrich Wagner
Institut für Theoretische Physik der Christian Albrechts Universität zu Kiel, Olshausenstraße 40, 24098 Kiel, Germany

(Received 15 September 1997; revised manuscript received 5 January 1999)

We present a general performance measure (information loss) for associative memories based on information theoretical concepts. This performance measure can be estimated, provided that mean values of observables have been determined for the associative memory. Then the estimation guarantees a minimal association quality. The formalism allows the application of the performance measure to complex systems where the relation between input and output of the associative memory is not explicitly known. Here we apply our formalism to the Hopfield model and estimate the storage capacity α_c from the numerically determined information loss. In contrast to other numerical methods the whole overlap distribution is taken into account. Our numerical value $\alpha_c = 0.1379(4)$ for the storage capacity in the Hopfield model is below numerical values obtained previously. This indicates that the consideration of small remnant overlaps lowers the storage capacity of the Hopfield model. [S1063-651X(99)02008-5]

PACS number(s): 87.10.+e, 07.05.Mh, 84.35.+i, 89.70.+c

I. INTRODUCTION

The advantage of associative memories where the data are stored in a spatially distributed way is the robustness against noisy input data on the one hand. On the other hand, due to the spatial distribution architecture these memories recollect the associated data with some retrieval error. So a strict person might conclude that the memory is useless. But in real situations the associative memory (AM) is often used as a preprocessor followed by some classification or decision process. For such processes a faulty output of the associative memory still could be valuable. In this paper we derive a general criterion for the quality of an associative memory considering also large retrieval errors.

Section II starts with a general description of an associative memory. Based on this description we formulate the real information loss. This quantity is by Shannon's theorem [1,2] closely related to the minimal expected number of yes-no questions needed to specify the recollection when the AM's output is given. In the literature it is also known as conditional information [1].

Unfortunately in most of the cases the associative memory is too complex and it is not possible to calculate the real information loss. Therefore we show, in Sec. III, how the real information loss can be estimated from above by measuring expectation values of observables.

In Sec. IV we choose the overlap between the associative memory's output and the recollection as a concrete observable and give an upper bound of the information loss both for the averaged overlap and the whole spectrum of the overlap. The results of Sec. IV are then applied (Sec. V) to a special neural associative memory model—the Hopfield model [3]. First we argue that the whole overlap spectrum is necessary to characterize the storage behavior. Then the information loss is estimated from the numerically determined overlap spectrum. The improved sensitivity of our estimate of the information loss allows for a finite size analysis leading to a precise numerical determination of the storage capacity α_c in the Hopfield model.

II. DESCRIPTION OF ASSOCIATIVE MEMORY

In this section we will present a general description of an associative memory and show how to measure its quality. We regard an AM as a stochastic system described by the probability $\mathcal{M}_{\underline{\xi},\underline{\eta}}(s^\infty|s^0)$ that the AM responds to an input s^0 out of a set K of key words with an output s^∞ out of a set R of recollections. The probability depends on p previously learned patterns $((\xi^1,\eta^1),\ldots,(\xi^p,\eta^p))$ each of them consisting of a key word ξ^μ and a recollection η^μ ($\underline{\xi}:=(\xi^1,\ldots,\xi^p)$, $\underline{\eta}:=(\eta^1,\ldots,\eta^p)$)

$$\sum_{s^\infty \in R} \mathcal{M}_{\underline{\xi},\underline{\eta}}(s^\infty|s^0)=1. \tag{1}$$

The mapping assigning every pair $(\underline{\xi}, \underline{\eta})$ its transition probabilities $\mathcal{M}_{\underline{\xi},\underline{\eta}}(s^\infty|s^0)$ is usually called the ''learning rule.'' Figure 1 sketches the typical situation an AM is used in.

In the learning process one pair $(\underline{\xi}, \underline{\eta})$ is chosen according to a probability distribution P_p and learned by the AM. After the learning process has finished the association process starts: A pattern index $\mu \in \{1,\ldots,p\}$ is chosen with probability $1/p$ and the key word ξ^μ is sent through a noisy channel $\mathcal{N}(s^0|\xi^\mu)$.

$$\sum_{k' \in K} \mathcal{N}(k'|k)=1 \quad \text{for every} \quad k \in K. \tag{2}$$

The AM receives a noisy version s^0 of the chosen key word ξ^μ with probability $\mathcal{N}(s^0|\xi^\mu)$. The patterns should have been

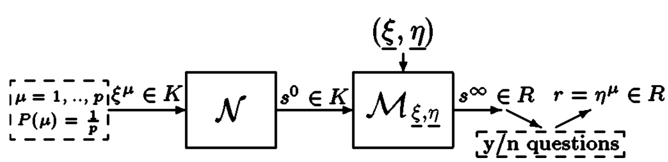

FIG. 1. Association of single patterns (ξ^μ, η^μ) previously learned by the AM.

learned by the AM in a way that the associated recollection s^∞ contains much information about $r=\eta^\mu$. In order to quantify this information we imagine a person at the output of the AM knowing s^∞ and having the task to specify the original recollection r. To achieve this goal the person is allowed to ask yes-no questions according to a question strategy S. This question strategy assigns every pair (r,s^∞) the number $\mathrm{len}_S(r|s^\infty)$ of yes-no questions needed to specify r for given s^∞.

To measure the association quality we imagine an experiment consisting of one learning process followed by a certain number of association processes. For this experiment, the average number of yes-no questions needed to specify r when s^∞ is known does not depend on the number of association processes. It is given by

$$q(S)=\sum_{r,s^\infty} P(r,s^\infty)\mathrm{len}_S(r|s^\infty), \tag{3}$$

where

$$P(r,s^\infty)=\frac{1}{p}\sum_{\mu,\underline{\xi},\underline{\eta}} P_p(\underline{\xi},\underline{\eta})T_{\underline{\xi},\underline{\eta}}(s^\infty|\xi^\mu)\,\delta_{\eta^\mu,r}, \tag{4}$$

with

$$T_{\underline{\xi},\underline{\eta}}(s^\infty|\xi^\mu):=\sum_{s^0} \mathcal{M}_{(\underline{\xi},\underline{\eta})}(s^\infty|s^0)\mathcal{N}(s^0|\xi^\mu).$$

If we define $(\log:=\log_2)$

$$\Phi(P):=-\sum_{r,s^\infty} P(r,s^\infty)\log P(r|s^\infty), \tag{5}$$

with the conditional probability

$$P(r|s^\infty)=P(r,s^\infty)\Big/\sum_r P(r,s^\infty),$$

Shannon's noiseless coding theorem (see Ref. [2], pp. 15 and 16) yields

$$\Phi(P)\leqslant\min q(S)\leqslant\Phi(P)+1, \tag{6}$$

where the minimum is taken over all possible question strategies S.

Thus $\Phi(P)$ has something to do with the minimal average number of yes-no questions needed to specify the original pattern r when the output s^∞ of the AM is known. We will therefore take $\Phi(P)$ to measure the quality of an AM and call it ''real information loss.'' Now the stored information I_s is defined as

$$I_s=p[I_R-\Phi(P)], \tag{7}$$

which is the information about the patterns available from the AM. Here

$$I_R=-\sum_r P(r)\log P(r), \tag{8}$$

with $P(r)=(1/p)\Sigma_{\underline{\eta},\mu}P_p(\underline{\eta})\delta_{r,\eta^\mu}$ is the Shannon entropy belonging to the experiment where the AM's output is unknown $[P_p(\underline{\eta})=\Sigma_{\underline{\xi}}P_p(\underline{\xi},\underline{\eta})]$.

Φ can also be used to measure the likelihood of two dynamical systems, where one of them is a ''black box'' and the other is established to model the black box. If $P(r,s^\infty)$ is the probability that the black box and the model are in the states r and s^∞, respectively, $\Phi(P)$ can directly be interpreted as a quality measure for the model.

In previous publications on neural associative memory [4,5] a quantity called ''missing information'' was suggested to regard small retrieval errors. This quantity is an estimation of $\Phi(P)$ with respect to special observables. This will become clear in Sec. IV. In the next section we will show how an estimation of the real information loss is connected to *every* possible choice of observables. Here the real information lost in the association process is transparently worked out and by Shannon's noiseless coding theorem related to the real information loss $\Phi(P)$ [see Eq. (6)].

III. ESTIMATION OF THE INFORMATION LOSS BY MEASURING OBSERVABLES

In most of the cases the relation $\mathcal{M}_{\underline{\xi},\underline{\eta}}(s^\infty|s^0)$ between the input s_0 and the output s^∞ of the AM is too complex to calculate the real information loss. Therefore we show in this section how to estimate the real information loss from above by the knowledge of mean values E_i of l observables $\mathcal{O}_i(r,s^\infty)$

$$E_i=\langle\mathcal{O}_i\rangle=\sum_{r,s^\infty} P(r,s^\infty)\mathcal{O}_i(r,s^\infty), \quad i=1,\ldots,l. \tag{9}$$

These equations together with the normalization of $P(r,s^\infty)$ build the constraints for the maximization of the real information loss.

We define W to be the set of mappings from $R\times R$ to $]0,1[$ and the functional $\Phi(P)$ with $P\in W$ is given by Eq. (5). The constraints form the restricted set of mappings:

$$\mathcal{A}_{\mathcal{O},E}=\{P\in W|E_i=\langle\mathcal{O}_i\rangle, \quad i=1,\ldots,l, \quad 1=\langle 1\rangle\}. \tag{10}$$

The maximization of the information loss with respect to the constraints means the maximization of Φ on the set $\mathcal{A}_{\mathcal{O},E}$.

We will use the following theorem to solve this maximization task.

Theorem III.1. For every distribution $P^*\in\mathcal{A}_{\mathcal{O},E}$ the following statement holds: P^* maximizes Φ on $\mathcal{A}_{\mathcal{O},E}$ if and only if there exist real numbers $\lambda_1,\ldots,\lambda_l$ and $Z>0$ such that the conditional distribution of P^* has the form

$$P^*(r|s^\infty)=\frac{1}{Z}\exp\left(-\sum_{i=1}^{l}\lambda_i\mathcal{O}_i(r,s^\infty)\right). \tag{11}$$

Since $P^*(r|s^\infty)$ depends exponentially on the observables and $P^*\in\mathcal{A}_{\mathcal{O},E},\Phi(P^*)$, we get for every $P\in\mathcal{A}_{\mathcal{O},E}$ the inequality

$$\Phi(P)\leqslant\Phi(P^*)=\left(\ln Z+\sum_{i=1}^{l}\lambda_iE_i\right)\Big/\ln 2. \tag{12}$$

This inequality allows the estimation of the real information loss from above.

The relations (11) and (12) are similar to the entropy maximization subject to observable constraints well known in the literature [6]. One direction of theorem III.1 can be proven with the theorem about the Lagrange multipliers and the other direction uses the inequality $\log x \leqslant (x-1)/\ln 2$. The Lagrange multipliers $\lambda_1,\ldots,\lambda_l$ and $\ln Z$ result from the constraints [Eq. (9) and the normalization of $P(r,s^\infty)$].

The real information loss differs from the conventional entropy because the logarithm in Eq. (5) has the conditional probability $P(r|s^\infty)$ instead of $P(r,s^\infty)$ as argument. As a consequence the maximal value of the real information loss $\Phi(P^*)$ depends only on the *conditional* probability distribution [see Eq. (12)] and a probability distribution satisfying the constraints and maximizing Φ is determined uniquely only up to its conditional distribution.

Since in practical applications the true distribution $P(r,s^\infty)$ will not be known, one uses the measured values E_i of the observables $\mathcal{O}_i(r,s^\infty)$ and applies the theorem III.1 to obtain via Eq. (12) an upper bound for $\Phi(P)$. In the next section we demonstrate this procedure by two (representative) examples of observables.

IV. THE OVERLAP

For the examples of observables we consider r, s^∞ to be Ising spin configurations, e.g., $R=\{+1,-1\}^N$. An obvious observable for estimating the information loss is the overlap

$$\mathcal{O}(r,s^\infty)=\frac{1}{N}\sum_{i=1}^{N} r_i s_i^\infty \tag{13}$$

between the AM's output s^∞ and the original recollection r. For given average

$$\overline{m}=\langle \mathcal{O}(r,s^\infty)\rangle \tag{14}$$

of $\mathcal{O}$, with $\overline{m}\neq\pm 1$ we choose for $P^*(r,s^\infty) = P(s^\infty)P^*(r|s^\infty)$ the conditional distribution to be

$$P^*(r|s^\infty)=\prod_{i=1}^{N}\frac{1}{2}(1+\overline{m}s_i^\infty r_i) \tag{15}$$

with arbitrary $P(s^\infty)>0$.

Since both $P^*(r|s^\infty)$ and $\mathcal{O}(r,s^\infty)$ depend only on the product $r_i s_i^\infty$, the constraint (14) is satisfied. The distribution (15) can be written in the form (11) with $\overline{m}=-\tanh\lambda$ and $Z=[2\cosh(\lambda/N)]^N$, which leads by theorem III.1 to the upper bound $\Phi(P^*)=i_{\overline{m}}N$ with

$$i_{\overline{m}}=1-\frac{1}{2}[(1-\overline{m})\log(1-\overline{m})]-\frac{1}{2}[(1+\overline{m})\log(1+\overline{m})] \tag{16}$$

for the real information loss $\Phi(P)$. The ''missing information'' for neural associative memory discussed in Refs. [4,5] would result from theorem III.1 if the average network and pattern magnetization are determined in addition to the average overlap (14). These additional observables lead to an upper bound of the real information loss slightly less than Eq. (16) for biased patterns.

If the overlap distribution has a complicated structure (for example, in the Hopfield model [7–9]) it is not sufficiently represented by its average $\overline{m}$. In these situations both Eq. (16) and the quantity discussed in Refs. [4,5] lead to a poor estimation of the real information loss. Therefore we choose as our second example the observables

$$\mathcal{O}_m(r,s^\infty)=\delta_{Nm,\Sigma r_i s_i^\infty}, \tag{17}$$

where $m=1-2k/N$, $k\in\{0,\ldots,N\}$. Averaging these observables leads to the overlap distribution

$$\rho(m)=\langle \mathcal{O}_m(r,s^\infty)\rangle, \tag{18}$$

where the $\rho(m)$'s have to be identified with the E_i's in Eq. (9).

In order to apply theorem III.1 we write the conditional distribution in the form of Eq. (11):

$$P^*(r|s^\infty)=\frac{1}{Z}\exp\left(-\sum_m \lambda(m)\mathcal{O}_m(r,s^\infty)\right), \tag{19}$$

where Z and $\lambda(m)$ have to be chosen such that $P^*(r,s^\infty)$ is normalized and satisfies the constraints (18). The normalization and the constraints (18) are satisfied if

$$Z>0$$

and

$$e^{-\lambda(m)}=Z\frac{\rho(m)}{g(m)}, \tag{20}$$

with

$$g(m)=\binom{N}{(N-mN)/2}.$$

Now theorem III.1 states that the distribution (19) maximizes the real information loss. The upper bound $\Phi(P^*)=i_\rho N$ with

$$i_\rho=\frac{1}{N}\sum_m \rho(m)\log\frac{g(m)}{\rho(m)} \tag{21}$$

for the real information loss can be obtained from Eq. (12).

To see the possibly substantial improvement of the bound (21) over Eq. (16), we expand $g(m)$ for large N according Stirling's formula and obtain

$$i_\rho\simeq\sum_m \rho(m)\left[1-\frac{1}{2}(1-m)\log(1-m)-\frac{1}{2}(1+m)\times\log(1+m)\right]-\frac{1}{N}\sum_m \rho(m)\log\rho(m). \tag{22}$$

The first term is a sum over m of terms as in Eq. (16) weighted with the probability $\rho(m)$. Since the right hand side of Eq. (16) is a concave function of $\overline{m}$ the first term of Eq. (22) is always less than or equal to $i_{\overline{m}}$ because of Jensens inequality. The second term in Eq. (22) is always bounded from above by $\log(N+1)/N$ which becomes negligible for $N\rightarrow\infty$. Thus the determination of the whole overlap distribution $\rho(m)$ may reduce the upper bound considerably, especially if the overlap has a bimodal distribution. The upper bound (21) for the real information loss cannot be derived from the considerations in previous publications [4,5].

V. APPLICATION TO NEURAL ASSOCIATIVE MEMORIES

We want to apply our formalism to neural associative memories. These memories usually recall the stored data with some retrieval error, which is related to the quality of the associative memory. In our formalism the retrieval error is quantified by the real information loss. As an example of neural associative memory we choose the Hopfield model [3] and estimate its storage capacity with the help of the numerically determined overlap spectrum (18). In the Hopfield model the pattern statistic is given by $P_p(\underline{\xi},\underline{\eta})=2^{-pN}\delta_{\underline{\xi},\underline{\eta}}$, which means autoassociation of equally distributed patterns. The network is a dynamical system evolving according to

$$s_i^{t+1}=\operatorname{sgn} h_i^t, \tag{23}$$

with

$$h_i^t=\sum_{j\neq i} w_{ij}s_j^t, \tag{24}$$

where only one neuron is updated per time step (asynchronous dynamics). The patterns are stored in the synaptic matrix w by the Hebb rule [3,10]:

$$w_{ij}=\frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu. \tag{25}$$

This learning prescription is motivated by Hebb's postulate about learning [11]. The convergence of the network state s^t to a fixpoint s^∞ is guaranteed because of the symmetry of the synaptic matrix. The stable fix point s^∞ is then regarded as the network's proposal for the recollection $r=\xi^\mu$, since for autoassociation the set of key words $\underline{\xi}$ and the set of recollections $\underline{\eta}$ coincide.

The Hopfield network is a complex dynamical system and it is not possible to calculate the $\Phi(P)$. Therefore we will estimate the real information loss with the help of the numerically determined overlap spectrum (18) via Eq. (21). Afterwards a finite size scaling analysis of the estimated real information loss is performed which leads to an accurate estimation of the storage capacity α_c. We use the whole spectrum of the overlap because in finite systems the overlap distribution has a double peak structure due to small remnant overlaps [7–9] indicating that the average of the overlap is not sufficient.

The numerical estimation of the real information loss is performed as follows: At first $p=\alpha N$ equally distributed patterns ξ^μ, $\mu=1,\dots,p$ are generated and stored in the synaptic matrix by Eq. (25). Then we start from the noisy version s^0 of the pattern ξ^μ and iterate dynamics (23) until a stable state s^∞ is reached. After the iteration the overlap between the network state and the original pattern is measured. By doing this for every pattern from several pattern sets we can determine the overlap spectrum (18) and from Eq. (21) an upper bound for the real information loss. The error of i_ρ has been calculated from the fluctuation of the mean values i_ρ calculated for one set according to Eq. (21). In Fig. 4 the upper bound of the real information loss per spin is presented for different system sizes as a function of α.

TABLE I. Different results for the storage capacity α_c, the average overlap $\overline{m}$, and upper bounds for the information loss per spin i_c at α_c. In the upper part of the table the analytical results of the replica symmetry (RS), the one (1RSB), and the two step (2RSB) replica symmetry breaking calculation are presented. In the lower part numerical estimates are given.

	RS [7]	1RSB [13]	1RSB [14]	2RSB [14]
α_c	0.137905	0.144	0.138186	0.138187
$\overline{m}(\alpha_c)$	0.967417	0.982	0.966777	0.966776
i_c	0.120078	0.074	0.121967	0.121970
	[8]	[9]		present paper
α_c	0.143(1)	0.1420(15)		0.1379(4)
i_c				0.121(7)

We have no theoretical founded finite size scaling law as used in Refs. [7–9]. However, based on the system sizes investigated here we can show the following finite size law: If for a given system size N the information loss can be described by straight lines $i_N(\alpha)=a_N+b_N\alpha$ for $\alpha\in[0.1375,0.1452]$, then there exists a loading parameter α_c where the information loss for all system sizes $N\geqslant 2000$ is equal to some value i_c.

First we determine a reference straight line by fitting independently offset a_N and slope b_N of the information loss for the simulated system sizes and extrapolate them to infinite system sizes. The fitted slopes and offsets are plotted in Fig. 2 vs $1/N$. From Fig. 2 we see that all system sizes except the smallest one ($N=1000$) satisfy the scaling law

$$a_N=a_\infty(1+a_{\mathrm{corr}}/N), \quad b_N=b_\infty(1-b_{\mathrm{corr}}/N).$$

The reference straight line is defined by the offset $a_\infty=-3.77(18)$ and the slope $b_\infty=28.2(1)$, which are obtained by fitting the above scaling law. Now we determine independently for each system size $N\geqslant 2000$ the point of intersection (α_c,i_c) with the reference straight line. Figure 3 demonstrates that the points of intersection are equal within the statistical errors for all system sizes, which proves the finite size law we mentioned above. The law justifies the fitting of a *common* point of intersection defining the storage capacity α_c and critical information loss per spin i_c.

For the determination of the common point of intersection we minimize χ^2. Due to the requirement of a common point of intersection this consists in a nonlinear problem, for which we used the program described in Ref. [12]. The fitted straight lines are plotted as dotted lines in Fig. 4. The values of α_c and i_c obtained from the fit are given in Table I. From the error matrix of the fit parameters we calculated the one standard deviation error ellipse around (α_c,i_c) shown in the inset of Fig. 4.

To compare our values for i_ρ with replica theory we need a calculation of i_ρ from replica theory. This can be achieved by the observation, that in this theory the overlap distribution

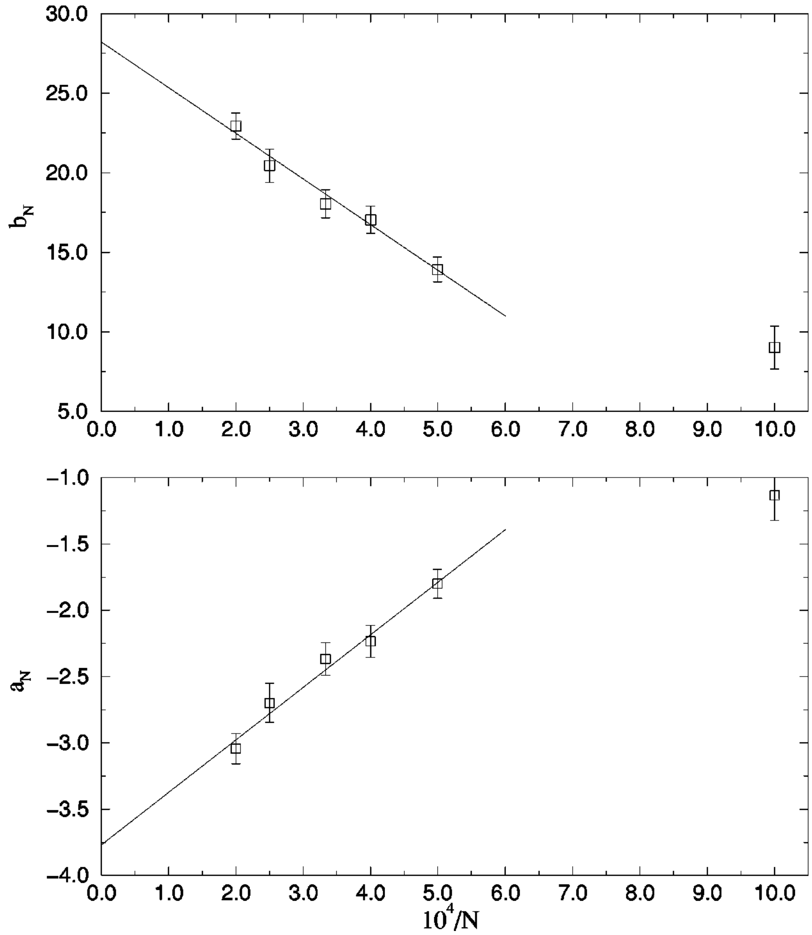

FIG. 2. Offset a_N and slope b_N of the information loss fitted in the region $0.1375 \leqslant \alpha \leqslant 0.1452$ as a function of $1/N$. The solid lines are linear fits of the data for $N \geqslant 2000$. The reference straight line of the information loss is defined by offset $\alpha_\infty = -3.77$ and slope $b_\infty = 28.2$ at $1/N = 0$.

is sharply peaked around its mean value $\overline{m}(\alpha)$ either little less than one in the retrieval phase or zero in the spin glass phase. In this case both upper bounds according to Eqs. (16) and (22) become equal for $N \rightarrow \infty$, and we can calculate i_ρ by inserting $\overline{m}(\alpha)$ known from the replica theory into Eq. (16). The resulting $i_{\overline{m}} = i_\rho$ from the replica symmetric theory [7] is shown in Fig. 4 as a dashed curve. Since $\overline{m}(\alpha)$ jumps discontinuously at $\alpha_c \simeq 0.137905$ from $\overline{m}(\alpha_c) \simeq 0.967$ to $\overline{m}(\alpha) = 0$ for $\alpha > \alpha_c$ due to a first order phase transition, $i_{\overline{m}}$ changes from $i_{\overline{m}} \simeq 0.12$ at α_c to $i_{\overline{m}} = 1$ for $\alpha > \alpha_c$. The analytical results for α_c and i_c of the replica symmetry (RS), the one (1RSB), and the two step (2RSB) replica symmetry breaking calculation are presented in Table I.

Our determination of the storage capacity α_c considers remnant overlaps (the information loss depends on the whole overlap distribution) while previous numerical investigations [7–9] use the retrieval rate for the determination of α_c, which does not include remnant overlaps. In contrast to that numerical work our value for the storage capacity α_c is slightly below the analytical value obtained from 2RSB calculations. This indicates that the consideration of small remnant overlaps lowers the storage capacity of the Hopfield model.

If the assumption that the information loss can be described by straight lines for $\alpha \in [0.1375, 0.1452]$ is true and the system sizes investigated here are large enough to extrapolate to infinite system sizes correctly, then we predict the information loss to behave as

$$i_\infty(\alpha) = i_c + b_\infty(\alpha - \alpha_c) + O[(\alpha - \alpha_c)^2] \qquad (26)$$

for $\alpha \geqslant \alpha_c$. The values

$$b_\infty = 28.2(1), \quad \alpha_c = 0.1379(4), \quad i_c = 0.121(7)$$

are obtained from the slope of the reference straight line and the common point of intersection. Since the information loss has a finite slope at α_c, there is no discontinuity in the information loss as predicted by replica theory.

These two observations indicate that the remnant overlaps influence the storage behavior of the Hopfield model considerably, which should be further investigated by numerical and analytical calculations. If an analytical theory which takes remnant overlaps into account could be worked out it should be possible to calculate an information loss based on the whole overlap distribution. The behavior of that information loss could be compared to Eq. (26).

VI. CONCLUDING REMARKS

In the present paper we have introduced a universal performance measure for associative memories called informa-

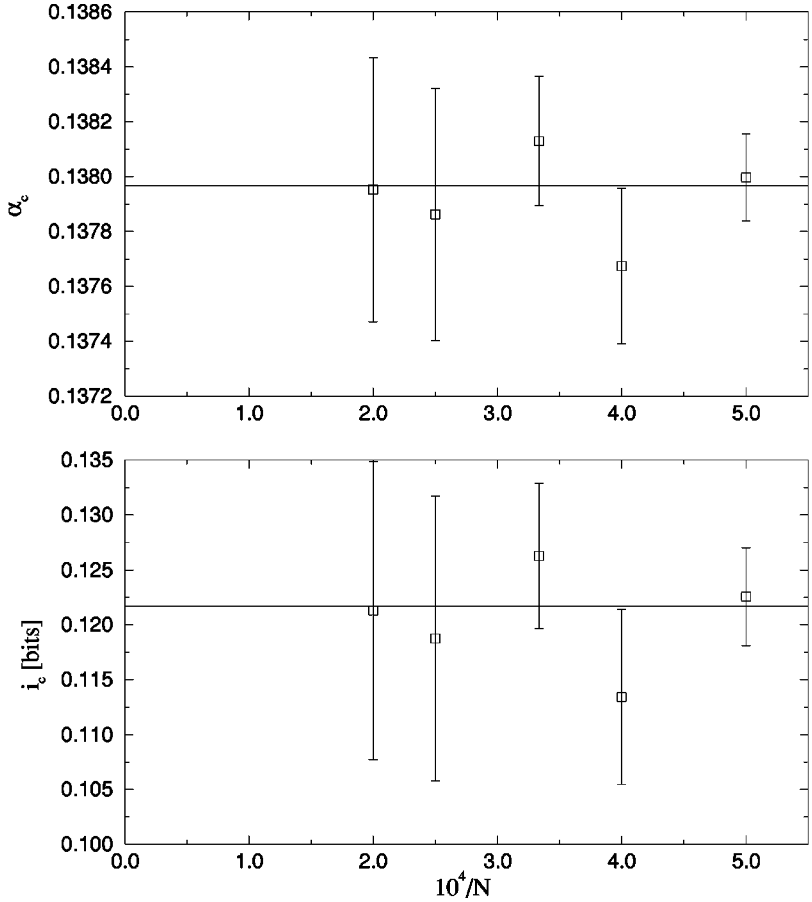

FIG. 3. Points of intersection (α_c, i_c) of the fitted information losses for $N \geqslant 2000$ with the reference straight line as a function of $1/N$. The constant solid lines $\alpha_c = 0.13797$ and $i_c = 0.1217$ are obtained by fitting a constant to the data.

tion loss. It is derived from information theoretical concepts. Our method for getting upper bounds of the real information loss by measuring observables allows the application to more complex systems (e.g., biological neural nets) than the ones discussed here. Provided appropriate observables have been chosen, these bounds can also be used for the numerical determination of the storage capacity with high precision as we have demonstrated for the Hopfield model.

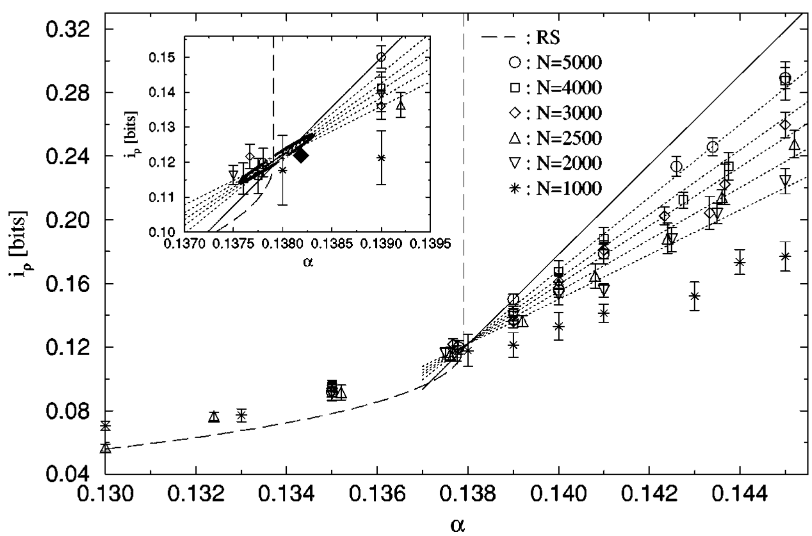

FIG. 4. The data points show the numerical estimate of the real information loss per spin i_ρ for the Hopfield model as a function of the load parameter α for different system sizes N. The dotted lines correspond to straight line fits for $N \geqslant 2000$ with a common point of intersection at (α_c, i_c). The solid line is the reference straight line obtained from Fig. 2. The dashed curve represents i_ρ predicted by the replica symmetric theory for infinite system sizes. The insert gives a magnified view around α_c including the error ellipse. For comparison the 2RSB values for the storage capacity α_c and the information loss per spin at α_c from Ref. [14] are indicated by a filled diamond.

If correlations are introduced to the stored patterns the information contained in the patterns decreases. This leads directly to the important question whether there are existing learning mechanisms such that the *stored information* specified in this paper remains constant. In our research on correlated patterns we have found explicit learning mechanisms which in deed leave the stored information constant. The results will be published in a forthcoming paper [15].

ACKNOWLEDGMENT

We would like to thank Gerhard Scheffler for his interest and valuable discussions.

[1] R. B. Ash, *Information Theory* (Dover, New York, 1990).
[2] C. M. Goldie and R. G. E. Pinch, *Communication Theory* (Cambridge University Press, Cambridge, 1991).
[3] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[4] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).
[5] H. Horner, Z. Phys. B **75**, 133 (1989).
[6] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[7] D. Amit, H. Gutfreund, and H. Sompolinsky, Ann. Phys. (N.Y.) **173**, 30 (1987).
[8] G. A. Kohring, J. Stat. Phys. **59**, 1077 (1990).
[9] T. Stiefvater, K. R. Müller, and R. Kühn, Physica A **232**, 61 (1996).
[10] G. Palm, Biol. Cybern. **39**, 19 (1980).
[11] D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).
[12] C. Lovelace and F. Wagner, Nucl. Phys. B **28**, 141 (1971).
[13] A. Crisanti, D. J. Amit, and H. Gutfreund, Europhys. Lett. **2**, 337 (1986).
[14] H. Steffan and R. Kühn, Z. Phys. B **95**, 249 (1994).
[15] M. Schlüter and F. Wagner (unpublished).