Network 3 (1992) 177–186. Printed in the UK

Information capacity in recurrent McCulloch–Pitts networks with sparsely coded memory states

G Palm and F T Sommer

C u O Vogt Institut für Hirnforschung, Universitätsstrasse 1, D-4000 Düsseldorf, Federal Republic of Germany

Received 28 March 1991, in final form 17 October 1991

Abstract. A new access to the asymptotic analysis of autoassociation properties in recurrent McCulloch–Pitts networks in the range of low activity is proposed. Using information theory, this method examines the static structure of stable states imprinted by a Hebbian storing process. In addition to the definition of critical pattern capacity usually considered in the analysis of the Hopfield model, we introduce the definition of information capacity which guarantees content adressability and is a stricter upper bound of the information really accessible in an autoassociation process. We calculate these two types of capacities for two types of local learning rules which are very effective for sparsely coded patterns: the Hebb rule and the clipped Hebb rule. It turns out that for both rules the information capacity is exactly half the pattern capacity.

1. Introduction

How many patterns can be stored in a large associative memory? The answer is given by the critical pattern capacity α_c (in patterns per neuron). How much information can be stored with autoassociation in a large recurrent associative memory? The answer is given by the memory capacity C (in bits per synapse). While the first question is well treated in the literature, there are no answers at all yet to the second question. However, it is the second question which is relevant for applications of associative memory in information retrieval.

This paper provides answers to the two questions for networks of binary McCulloch–Pitts neurons where the stored patterns are sparse (low average activity p). The number of storable patterns per neuron α_c diverges for $p \to 0$, but their information content $I(p)$ tends to zero (p is the probability that a neuron is active in a pattern, $I(p)$ is the Shannon information). This divergence is removed by defining the critical pattern capacity as $P := \alpha_c I(p)$. But P is still different from the information capacity C as defined in section 2.

We propose a method of calculating information capacities considering the discrepancy between the set of memory states S and the set of fixed points $\mathcal{F}$ of the network dynamics, where the synaptic connectivity matrix is formed by a specified learning rule. Our method reveals the structure of these sets in the state space of the system by examining a retrieval procedure which checks the membership of each state to the set $\mathcal{F}$ and disregards the transient behaviour of states not belonging to $\mathcal{F}$. Although we concentrate on a model with parallel update the results are valid

0954-898X/92/020177+10$04.50

for sequential update as well, since the classification of fixed points is independent of whether the considered dynamics is sequential or parallel.

This treatment takes into account the effects of spurious states on the pattern completion property of the memory which yields new asymptotic estimates for the information capacity. These values may serve as stricter upper bounds to the information practically retrievable with more realistic iterative retrieval procedures than the critical pattern capacity values considered in the literature.

Our method also reproduces the critical pattern capacity results; in this case it is equivalent to the noise-to-signal treatment of Palm [1] and Nadal and Toulouse [2] for fixed activity of the memory states.

In section 2 we give the definitions of the two types of capacity in the framework of this method mentioned at the beginning. Section 3 outlines the calculation ansatz. Section 4 leads to the explicit results for the two learning rules considered. A detailed discussion situating our results in relation to other approaches using noise-to-signal calculations [1, 2, 3] or methods of statistical physics [4, 5, 6] is given in section 5.

1.1. The model

As storing process we consider two types of local learning rules. Let $\{0,1\}^n$ be the space of neural activity states. We choose a set of randomly generated memory states with fixed activity $S := \{\xi^\mu \in \mathcal{M}_k : \mu = 1, \ldots, M\}$ with $\mathcal{M}_k := \{x \in \{0,1\}^n : |x| = k\}$ to generate the memory matrix c_{ij} via the learning rule. As the learning rules we treat

The Hebb rule
$$c_{ij} = \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu \tag{1.1}$$

and the clipped Hebb rule
$$c_{ij} = H\left[\sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu\right]. \tag{1.2}$$

$H[x]$ is the Heaviside function.

In the retrieval process we consider iteration of a parallel update step which is defined as the mapping $x \rightarrow x'$ with $x'_j = H\left[\sum_i c_{ij} x_i - \Theta\right] \quad \forall \quad j$, where Θ denotes a global threshold. If we restrict the retrieval to $x \in \mathcal{M}_k$ the global threshold can be chosen fixed during the iteration process in order to preserve the mean activity.

2. Capacities and retrieval quality criteria

We focus on the channel capacity of the information channel consisting of the local storing process and a certain retrieval procedure. In this retrieval procedure the subset $\mathcal{F} := \{x \in \mathcal{M}_k : x = x'\}$ of fixed points of c_{ij} is obtained by checking for every $x \in \mathcal{M}_k$ the fixed point condition $x = x'$.

Any definition of information capacity is combined with a quality criterion restricting the errors which are tolerated in the retrieval process. The error in our retrieval procedure can be expressed as the correction information necessary to obtain S from the retrieved $\mathcal{F}$ and is written as

$$I(S \mid \mathcal{F}) = I_{\mathrm{um}} + I_{\mathrm{ss}}. \tag{2.1}$$

$I_{\rm um}$ and $I_{\rm ss}$ are the contributions from the two types of errors which can be expressed in terms of the error probabilities:

$p_{\rm um} := p\,[x \notin \mathcal{F} \mid x \in \mathcal{S}]$	the probability of an unstable memory state
$p_{\rm ss} := p\,[x \in \mathcal{F} \mid x \notin \mathcal{S}]$	the probability of a spurious state.

The explicit expressions for $I_{\rm um}$ and $I_{\rm ss}$ are given in section 3.

With the retrieval quality criterion which requires

$$\frac{1}{n^2} I(\mathcal{S} \mid \mathcal{F}) \to 0 \qquad \text{as } n \to \infty \tag{2.2}$$

we define the *information capacity* as the information channel capacity measured in bits/synapse:

$$C := \frac{1}{n^2}\{I(\mathcal{S}) - I(\mathcal{S} \mid \mathcal{F})\} = I(\mathcal{S})/n^2 \tag{2.3}$$

where $I(\mathcal{S}) := \sharp\mathcal{M}_k I(p[x \in \mathcal{S}])$ is the information content in $\mathcal{S}$ and $I(p)$ the Shannon information (see section 3). With $p = p[x \in \mathcal{S}] = M/\sharp\mathcal{M}_k$ we obtain

$$C = MI(p)/n. \tag{2.4}$$

Inserting the maximal number M_1 of memory states for which the criterion (2.2) is fulfilled we obtain in (2.3) the information capacity.

The *critical pattern capacity* is usually defined in the physical literature [4, 5, 6] as $P := M_2 I(p)/n$ where now M_2 is the maximal number of memory states satisfying the so-called *embedding condition* (for $\kappa = 0$). This quality criterion is equivalent to the requirement that $\mathcal{S} \subseteq \mathcal{F}$ and can be expressed in our terms as

$$\frac{1}{n^2} I_{\rm um} \to 0 \qquad \text{as } n \to \infty. \tag{2.5}$$

Because (2.2) is more restrictive than (2.5) the information capacity should remain below the critical pattern capacity . The critical pattern capacity is no channel capacity for any storage and retrieval procedure. It is a measure of the information content of $\mathcal{S}$ and its quality criterion does not guarantee at all that this information is accessible with autoassociative retrieval.

3. Explicit quality criteria

For the prescribed retrieval procedure we derive explicit expressions for the two contributions in formula (2.1) describing the information loss due to the occurance of spurious states and to unstable memory states respectively.

Defining the Shannon information as usual as $I(p) = -p\,\mathrm{ld}[p] - (1-p)\,\mathrm{ld}[1-p]$ one can formulate explicitly the conditions on the error probabilities defined in section 2 which are necessary for the fulfillment of the quality criteria (2.2) and (2.5).

The quality criterion (2.5) demanded by the definition of the critical pattern capacity considers only:

$$I_{um} = \sharp\{\mathcal{M}_k \setminus \mathcal{F}\}\, I\left(\frac{\sharp\{\mathcal{S}\setminus\mathcal{F}\}}{\sharp\{\mathcal{M}_k\setminus\mathcal{F}\}}\right)$$
$$= \binom{n}{k} p[x\notin\mathcal{F}\cap x\in\mathcal{S}]\ \mathrm{ld}\left[\frac{p[x\notin\mathcal{F}]}{p[x\notin\mathcal{F}\cap x\in\mathcal{S}]}\right].$$

Using the fact that $\sharp\mathcal{S}, \sharp\mathcal{F} \ll \sharp\mathcal{M}_k$ we arrive at

$$\frac{I_{um}}{n^2} = \frac{Mp_{um}}{n^2}\mathrm{ld}\left[\frac{\binom{n}{k}}{Mp_{um}}\right].$$

Thus criterion (2.5) holds if the error probability p_{um} fulfils

$$\frac{k\ \mathrm{ld}\,[n]\, M}{n^2} p_{um} \to 0 \qquad \text{as } n \to \infty. \tag{3.1}$$

Furthermore, the stronger quality criterion (2.2) requires that the term:

$$\frac{I_{ss}}{n^2} = \frac{1}{n^2}\sharp\mathcal{F}\ I\left(\frac{\sharp\{\mathcal{F}\setminus\mathcal{S}\}}{\sharp\mathcal{F}}\right)$$
$$= \frac{1}{n^2}\binom{n}{k}\Bigg\{p[x\in\mathcal{F}\cap x\notin\mathcal{S}]\ \mathrm{ld}\left[\frac{p[x\in\mathcal{F}]}{p[x\in\mathcal{F}\cap x\notin\mathcal{S}]}\right]$$
$$+ (p[x\in\mathcal{F}] - p[x\in\mathcal{F}\cap x\notin\mathcal{S}])\ \mathrm{ld}\left[\frac{p[x\in\mathcal{F}]}{p[x\in\mathcal{F}] - p[x\in\mathcal{F}\cap x\notin\mathcal{S}]}\right]\Bigg\}$$

vanishes. Using $\sharp\mathcal{S} \ll \sharp\mathcal{M}_k$ and $p[x\notin\mathcal{S}] \simeq 1$ and $p[x\in\mathcal{F}] \simeq p[x\in\mathcal{F}\cap x\notin\mathcal{S}] + p[x \in \mathcal{S}]$ we obtain

$$\frac{I_{ss}}{n^2} = \binom{n}{k}\frac{p_{ss}}{n^2}\ \mathrm{ld}\left[1 + \frac{M}{\binom{n}{k}p_{ss}}\right] + \frac{M}{n^2}\ \mathrm{ld}\left[1 + \frac{\binom{n}{k}p_{ss}}{M}\right].$$

Since for both rules the order of M is less than n^2, with $u := \binom{n}{k}p_{ss}/n^2$ we can estimate

$$\frac{I_{ss}}{n^2} \leqslant u\ \mathrm{ld}\left[1 + \frac{1}{u}\right] + \mathrm{ld}[1 + u]$$

which vanishes only for $u \to 0$ as $n \to \infty$. Thus the quality criterion (2.2) holds if p_{um} satisfies (3.1) and

$$u \simeq \binom{n}{k}p_{ss} \to 0 \qquad \text{as } n \to \infty. \tag{3.2}$$

4. Calculation of pattern and information capacities

We use the conditions on the error probabilities (3.1) and (3.2), which are necessary for satisfaction of the quality criteria, to obtain the upper capacity bounds for the two examined learning rules in the following subsections .

First we calculate the error probabilities for the described retrieval process in general. We have to assume asymptotic statistical independence of different synaptical values c_{ij}. In the case of the clipped Hebb rule the asymptotic independence is shown in appendix 1. Since in one retrieval step it is decided whether any input state $x \in \mathcal{M}_k$ is a fixed point or not we obtain the defined error probabilities from the treatment of the one-step retrieval process. In the one-step case, if a memory state is entered as input, e_{01} and e_{10} are defined as the error probabilities for a single off-output neuron and on-output neuron respectively. Without loss of generality we may assume that $x = (1, 1, ..., 0, 0, ...)$. Let us define the probabilities that an input x activates all of its on output neurons

for spurious states $$Q_{\rm ss} := p\left[\sum_{i=1}^{k} c_{ij} \geqslant \Theta \ \text{ for } \ j = 1, ..., k \mid x \notin S\right] \tag{4.1}$$

for memory states $$Q_{\rm ms} := p\left[\sum_{i=1}^{k} c_{ij} \geqslant \Theta \ \text{ for } \ j = 1, ..., k \mid x \in S\right] . \tag{4.2}$$

Since c_{ij} is symmetric for local learning rules, there are statistical dependencies of the threshold problems of the first k columns and the single-neuron error probabilities only yield lower bounds $Q_{\rm ss} \geqslant e_{01}^k$ and $Q_{\rm ms} \geqslant (1 - e_{10})^k$.

For the error probabilities we obtain

$$\begin{aligned} p_{\rm ss} &= Q_{\rm ss}\, p\left[\sum_{i=1}^{k} c_{ij} < \Theta \ \text{ for } \ j = k+1, ..., n \mid x \notin S\right] \\ &\simeq Q_{\rm ss}\,(1 - e_{01})^{n-k} \end{aligned} \tag{4.3}$$

because the threshold problems for the columns $j = k + 1, ..., n$ can be regarded as independent and

$$p\left[\sum_{i=1}^{k} c_{ij} x_j \geqslant \Theta \mid x \notin S\right] \simeq p\left[\sum_{i=1}^{k} c_{ij} \xi_j^l \geqslant \Theta \mid j \in \{j : \xi_j^l = 0\}\right] .$$

Similarly

$$\begin{aligned} p_{\rm um} &= 1 - Q_{\rm ms} p\left[\sum_{i=1}^{k} c_{ij} < \Theta \ \text{ for } j = k+1, ..., n \mid x \in S\right] \\ &\simeq 1 - Q_{\rm ms}\,(1 - e_{01})^{n-k} . \end{aligned} \tag{4.4}$$

4.1. The clipped Hebb rule

The memory matrix is generated by (1.2). For this rule which is treated in the one-step retrieval case [7, 8] we know: $p := k/n$, $c_{ij} \in \{0,1\}$, $p(c_{ij} = 1) = 1 - (1-p^2)^M =: q$. In [7] it is shown that for $q = 1/2$ the channel capacity of the learning process reaches its optimum. If the threshold is set equal to $\Theta = k$ the error probabilities are $e_{01} = 2^{-k}$, $e_{10} = 0$ and the number of memory states is

$$M = \frac{n^2}{k^2} \ln[2]. \tag{4.5}$$

It is straightforward to see that $Q_{ss} = 2^{-k^2/2}$ and $Q_{ms} = 1$. To achieve high capacity one has to use sparse-coded patterns, i.e.

$$k = a \ln[n] \tag{4.6}$$

with a a positive constant.

If we put (4.5) and (4.6) into (3.1) the following condition results

$$\frac{\ln[2]}{k} \operatorname{ld}[n]\, p_{\mathrm{um}} = p_{\mathrm{um}}/a \to 0 \qquad \text{as } n \to \infty.$$

With (4.4) we obtain as necessary condition for the quality criterion (2.5)

$$n\, e_{01} \to 0 \qquad \text{as } n \to \infty \tag{4.7}$$

which is equivalent to a low noise-to-signal criterion in the one-step retrieval process called in [10] the *hifi condition*. If we put the error behaviour for the clipped rule in (4.7) the requirement: $a > 1/\ln[2]$ on the constant a in (4.6) is demanded. Using this constraint on a in the equation (2.4) with (4.5) we obtain the *critical pattern capacity*

$$P = \ln[2]. \tag{4.8}$$

In this range of sparseness from (4.3) the error probability can be written as

$$p_{ss} \simeq \exp\left[-\frac{\ln[2]}{2} (a \ln[n])^2\right].$$

This expression is inserted into (3.2) for the second quality criterion to be satisfied. In this condition the leading exponent is $a(\ln[n])^2 [1 - a \ln[2]/2]$. Negative exponent requires the stronger condition on the constant in (4.6): $a > 2/\ln[2]$. With this constraint as in (4.8) we find for the *information capacity*

$$C = \tfrac{1}{2} \ln[2]. \tag{4.9}$$

4.2. The unclipped Hebb rule

In this case the matrix c_{ij} is generated in the storing process according to (1.1). For the unclipped Hebb rule the error probabilities for one-step retrieval are approximately [2, 9, 11]

$$e_{01} \simeq G\left(-\frac{\vartheta n}{\sqrt{kM}}\right) \qquad e_{10} \simeq G\left(-\frac{(1-\vartheta)n}{\sqrt{kM}}\right)$$

($G(x)$ being the Gauss integral, $0 \leqslant \vartheta \leqslant 1$ the normalized threshold, see [9]). Low errors can only be expected in the range of high signal-to-noise ratio: $r := n/\sqrt{kM} \to \infty$. Then the large-$n$ behaviour of the error probabilities is given by

$$e_{01} \propto \exp\left[-\frac{\vartheta^2}{2kM}n^2\right] \qquad e_{10} \propto \exp\left[-\frac{(1-\vartheta)^2}{2kM}n^2\right]$$

(see proposition 1 in appendix 1).

If we put

$$M = b\ n^2/(k \ln[n]) \tag{4.10}$$

with b a positive constant the condition (3.1) reduces to $p_{\text{um}} \to 0$ as in the clipped case. Again the condition (4.7) has to be satisfied. If we insert the error behaviour of the unclipped Hebb rule we obtain the condition on the constant in (4.10) $b < \vartheta^2/2$. Using this in (2.4) it leads with $\vartheta \to 1$ to the *critical pattern capacity*

$$P = \frac{1}{2\ln[2]}. \tag{4.11}$$

In proposition 2 of appendix 1 we show that the error probability p_{ss} behaves like

$$p_{\text{ss}} \propto \exp\left[-\frac{\vartheta^2}{4M}n^2\right] \qquad \text{as } n \to \infty.$$

Inserting this into (3.2) we obtain for the leading exponent:

$$k\ln[n] - \frac{\vartheta^2}{4M}n^2 \to 0$$

which becomes negative for $b < \vartheta^2/4$. This yields the *information capacity*

$$C = \frac{1}{4\ln[2]}. \tag{4.12}$$

5. Discussion

Our iterative retrieval procedure just extracts the set $\mathcal{F}$ of fixed points of the system. The capacity of the information channel consisting of the storing process and our retrieval process (the information capacity) can be treated using noise-to-signal calculations for one-step retrieval. The quality criterion of the information capacity is that the learned patterns are recognized as known and all other patterns are classified as unknown.

In our framework it is also possible to define a quality criterion associated with the critical pattern capacity (see section 2) that has been investigated in the literature before by several methods. In this quality criterion one only requires that the learned patterns are recognized; the requirement that unlearned patterns should be classified as unknown is dropped. The two quality criteria fix a range of the memory load M in which the mean number of spurious states varies between the maximum ($p_{\text{ss}} \to 1$) for the critical pattern capacity and zero ($p_{\text{ss}} \to 0$) for the information capacity.

5.1. Critical pattern capacity results

Our evaluation is essentially equivalent to the calculation in the work of Palm [9] and Nadal and Toulouse [2] (in their so-called 'low-error-regime') albeit both works consider the case of hetero association. The new aspect of our results is the possible extension of the treatment to memory states that are not restricted to $\mathcal{M}_k$ but have a low average density $p = k/n$ of ones, see [12], which keeps the results unchanged. Therefore we can compare our results to works treating memory states with low average activity with methods of statistical physics.

Our value for the Hebb rule (4.11) is a confirmation of the result of Tsodyks and Feigelman [4] calculated with mean field theory *à la* Amit *et al* [13]. They use a learning rule which approaches the Hebb rule for vanishing p. The result also coincides with the Gardner bound [5] also obtained applying techniques of statistical physics. It is an upper bound on the pattern capacity for any storing processes.

The fact that the critical pattern capacity for the clipped Hebb rule exceeds the Gardner bound of 0.29 calculated by Gutfreund and Stein [6] led to a discussion started in [6] and [3]. Nadal [3] explained this inconsistency with the fixed activity level in the Willshaw calculation. Because of our results for fluctuating activity of the memory states [12], we believe that the Gutfreund and Stein value is so low, because the the error constraint in the Gardner calculation requires exactly zero error instead of asymptotically vanishing average error.

The values (4.11) and (4.8) can be reproduced in computer simulations with reasonable accuracy: see [14] and [15].

5.2. Information capacity results

For the Hebb rule our estimate of the information capacity (4.12) turns out to reach the local learning bound defined in [16]. Also for the clipped Hebb rule our result (4.9) coincides with the channel capacity of an optimal but non-constructive iterative retrieval process as calculated in [8]. Thus the retrieval procedure considered here is very idealistic and our treatment could not describe the more realistic iterative retrieval process, where one starts with an initial adress pattern and gradually updates it to find the closest fixed point. It does not regard the transient behaviour of states which are not fixed points; in real iterative retrieval there will occur effects like confusion due to irregular shapes of the basins of individual memory states and the existence of initial states whose dynamics ends outside $\mathcal{M}_k$ or in cycles. Up to now, however, no other measure can be found in the literature that considers these effects. Our information capacity value is in fact a better upper bound on the information which can effectively be retrieved with autoassociation using any realistic iterative retrieval process than the critical pattern capacity.

Since fixed point retrieval turns out to yield twice the values of one-step autoassociation (see [9]), the information capacity for practical iterative retrieval procedures should be between $C/2$ and C.

Appendix 1

In the following we present some calculations, which are necessary for the treatment of the Hebb rule.

Proposition 1.

$$(2\pi t^2)^{-1/2}e^{-t^2/2}(1-t^2) \leqslant G(-t) = 1 - G(t) \leqslant (2\pi t^2)^{-1/2}e^{-t^2/2}$$

Proof. Since $x^2 = t^2 + (x-t)^2 + 2t(x-t)$, we have

$$\int_t^\infty e^{-x^2/2}\,dx = e^{-t^2/2}\int_0^\infty e^{-x^2/2}e^{-xt}\,dx.$$

From this and with $e^{-x^2/2} \leqslant 1$ we obtain the second inequality directly (since $\int_0^\infty e^{-xt}\,dx = 1/t$) and the first one after partial integration (since $\int_0^\infty xe^{-xt}\,dx = 1/t$). □

Proposition 2. With the definitions of section 4 for the Hebb rule, it holds that

$$Q_{ss} \propto \exp\left[\frac{-\theta^2 n^2}{4M}\right] \qquad \text{as } n \to \infty. \tag{5.1}$$

Proof. Again $x = (1,1,\dots,0,0,\dots)$, so in the following we consider only the left upper k^2 block of c_{ij}.

(i)
$$Q_{ss} \leqslant p\left[\underbrace{\sum_{i=1,j\leqslant i}^{k,k} c_{ij}}_{=:s} \geqslant \frac{k\Theta}{2}\right] = G\left(-\frac{n\theta}{\sqrt{2m}}\right) \leqslant \frac{\sqrt{2m}}{n\sqrt{2\pi}}\exp\left[-\frac{n^2\theta^2}{4M}\right]$$

since $E(s) = k^2Mp^2/2$, $\sigma^2(s) = E(s)$ and $\Theta = kMp^2 + k\theta$ (see proposition 1).

(ii) Let $\{N_{ij} : i = 1,\dots,k/l\}$ be a partition of each column j with $|N_{ij}| = l$. T denotes the smallest set of N_{ij} which covers the left lower triangle completely. We have

$$Q_{ss} \geqslant p\left[\underbrace{\sum_{h\in N_{ij}} c_{hj}}_{=:s'} \geqslant \frac{\Theta l}{k}\ \forall\ N_{ij}\in T\right] = p\left[s' \geqslant \frac{\Theta l}{k}\right]^{(1+2+\dots+k/l)l}$$

$$= G\left(-\frac{n\theta\sqrt{l}}{k\sqrt{M}}\right)^{k^2/2l} \simeq \left(\frac{k\sqrt{M}}{\sqrt{2\pi l}n\theta}\right)^{k^2/2l}\exp\left[-\frac{n^2\theta^2}{4M}\right].$$

The last estimate follows from proposition 1. □

References

[1] Palm G 1988 On the asymptotic storage capacity on neural networks *Neural Computers* ed R Eckmiller and Chr v d Malsburg (Berlin: Springer)

[2] Nadal J-P and Toulouse G 1990 Information storage in sparsely coded memory nets *Network* **1** 61–74

[3] Nadal J-P 1991 Associative memory: on the (puzzling) sparse coding limit *J. Phys. A: Math. Gen.* **24** 1093–101

[4] Tsodyks M V and Feigelman M V 1988 The enhanced storage capacity in neural networks with low activity level *Europhys. Lett.* 6 101–5
[5] Gardner E 1988 The space of interactions in neural network models *J. Phys. A: Math. Gen.* 21 257–70
[6] Gutfreund H and Stein Y 1990 Capacity of neural networks with discrete synaptic couplings *J. Phys. A: Math. Gen.* 23 2613–30
[7] Willshaw D J, Buneman O P and Longuet-Higgins H C 1969 Nonholographic associative memory *Nature* 222 960–2
[8] Palm G 1980 On associative memory *Biol. Cybern.* 36 19–31
[9] Palm G 1990 Local learning rules and sparse coding in neural networks *Advanced Neural Computers* ed R Eckmiller (Amsterdam: North Holland)
[10] Amari S I 1989 Statistical neurodynamics of associative memory *Neural Networks* 1 63–73
[11] Palm G 1991 Memory capacities of local learning rules for synaptic modification *Concepts in Neuroscience* 2 97–128
[12] Palm G and Sommer F T 1991 Information and pattern capacities in neural associative memories with feedback for sparse memory patterns *Conf. Proc. Workshop on Complex Dynamics of Neural Networks (Vietri, Italy, June 1991)* (Berlin: Springer)
[13] Amit D J and Gutfreund H 1987 Statistical mechanics of neural networks near saturation *Ann. Phys.* 173 30–67
[14] Horner H 1989 Neural networks with low levels of activity: Ising vs. McCulloch–Pitts neurons *Z. Phys.* B 75 133–6
[15] Schwenker F, Sommer F T and Palm G 1992 Simulations of recurrent networks of threshold neurons with sparse activity, in preparation
[16] Palm G 1992 On the information storage capacity of local learning rules *Neural Computation* in press