Neural Networks from Models to Applications
N'EURO '88, pp. 360-371 L. Personnaz and G.Dreyfus,
Eds. (EZIDET,Paris, 1989)

On Sparsely Coded Associative Memories*

J. BUHMANN†, R. DIVKO AND K. SCHULTEN‡

Physik-Department, Technische Universität München,
James-Franck-Straße, 8046 Garching, Fed. Rep. Germany

ABSTRACT

A class of neural networks with adaptive threshold and global inhibitory interactions is proposed. The networks are capable of nearly optimal storage of sparsely coded patterns, i.e. patterns with low level of activity. We present a replica symmetric solution of the mean field equations for the noise free case. The network shows the following, remarkable properties: 1) For low level of activity a the storage capacity increases as $-[a \ln a]^{-1}$; up to 0.38 bits/synapse can be stored. The network capacity approaches the theoretical upper bounds derived by Gardner (1988). 2) Spurious states representing superpositions of stored patterns can be suppressed by global inhibition; 3) The network is not opinionated, i.e. by assuming a state of low activity it categorizes respective inputs as not similar enough to any patterns stored.

*Part of this contribution will appear in Phys. Rev. A

†New address: Computer Science Department SAL, University of Southern California, University Park, Los Angeles, Ca 90089-0782

‡New address: Department of Physics, University of Illinois, 1110 W. Green Street, Urbana, Il 61801

1. Introduction

During the last few years a new wave of interest in simple models of neural systems swept physics, computer science and biology. The ability to build massively parallel computers, but also a breakthrough in understanding long-range spin glasses lead to various research activities on neural networks. One of the most intensely studied neural network models had been the one introduced by Hopfield. The model shows some remarkable computational capabilities, e.g. it functions as a fault tolerant associative memory. The computational capability emerges from dissipative network dynamics: relaxation to fixpoints representing stored patterns constitute the computational process of information retrieval. The mathematical tools of statistical mechanics of spin glasses allowed a systematic evaluation of the computational capabilities of the Hopfield model [1].

Starting from the Hopfield model search for neural networks with larger storage capacity and better recall behaviour began and followed two main strategies. The first and direct way to enhance storage capacity is to look for alternative synaptic connectivity rules which deviate from Hopfield's suggestion. We will follow this path and propose a sparse coding scheme with suitable connectivities. An alternative strategy to enhance storage capacity is based on iterative learning algorithms. These algorithms [2–4] derived from the well known perceptron algorithm, correct recall errors occuring during pattern association and, after a training period, produce a connectivity structure which guarantees optimum associative storage. Unfortunately, computing the optimum connectivity structure is very time consuming.

The question about the maximum amount of information which can be stored in neural network can be formulated as the following optimization problem first studied by Gardner [4]. Given a set of patterns to be stored in a network, how many connectivity structures do exist which garantee perfect recall of the whole pattern set? This optimization problem can be interpreted as an inverse problem of statistical mechanics, i.e. we define a set of pattern states of a neural network and look for suitable synaptic interactions to stabilize these equilibrium states. Gardner's approach yields upper bounds for the number p_c of correlated or uncorrelated patterns which can be stored and recalled in a neural network without errors. In the limit of strong correlations between patterns, i.e. for low activity in the network, p_c

increases as $p_c \sim -N/(a \ln a)$ where a denotes the percentage of active neurons in a given pattern.

2. Model for an Optimum Associative Memory

In this contribution we propose a neural network which stores biased patterns of small activity level a, i.e. patterns which correspond to only a small fraction of neurons firing*. The model can be interpreted as a generalization of the original Hopfield model [6] for arbitrary level of activity a. The network obeys the condition of Gardner for optimum storage of strongly correlated patterns, albeit tolerates a certain percentage of recall errors. If one whishes to develop a network with completely accurate recall our model yields a good initial connectivity to be improved by an iterative algorithm.

Storage of biased patterns in Hopfield networks was first proposed in [7]. However, the proposed model exhibits decreasing information content when the network activity, i.e. a, is decreased. An asymmetric version of our model had been proposed earlier in context with storing sequences of patterns in neural networks [8, 9].

Our network is composed of N neurons described by dynamic variables $\{S_i\}_{i=1}^N$. Neuron i is either firing ($S_i = 1$) or quiet ($S_i = 0$). The variables are updated asynchronously according to a probabilistic rule which represents the action of noise in the system. With probability $f_i = (1 + \exp[-(h_i - U)/T])^{-1}$ for a molecular field $h_i = \sum_k J_{ik} S_k$ the chosen neuron i fires at time $t + \Delta t$, otherwise it is quiet. The parameters U and T are the threshold potential and the network temperature. The patterns $\boldsymbol{\xi}^\nu = \{S_i\}_{i=1}^N$, which we intend to store in our network, are correlated and are chosen according to the distribution $P(\xi_i^\nu) = a\delta(\xi_i^\nu - 1) + (1 - a)\delta(\xi_i^\nu)$ with a small. The low average activity of neurons in the brain [10] suggests that the memory model considered, i.e. one for storage of patterns with small a, has a close correspondence to biological memories.

The synaptic connections between neurons are chosen according to the cell as-

*Independent of us Feigel'man and Tsydocs [5] have suggested the same neural network, albeit without global inhibition, and have discussed its properties in the limit $a \to 0$.

sembly hypothesis of Hebb by the rule

$$W_{ik} = \frac{1}{a(1-a)N} \sum_{\nu=1}^{p} (\xi_i^\nu - a)(\xi_k^\nu - a) - \frac{\gamma}{aN} \qquad i \neq k. \tag{1}$$

The first term in (1) describes the formation of cell assemblies, i.e. sets of excitatorily interacting neurons which represent patterns ν. The second term implies inhibition between all neurons of the network and enforces discrimination between different patterns. Synapses between two neurons are symmetric, i.e. $W_{ik} = W_{ki}$; self-interaction is forbidden. If one interpretes the W_{ik} given by (1) as interactions between the spin variable S_i, an energy (Hamiltonian) $\mathcal{H}$ can be associated with any configuration of spins $\{S_i\}_{i=1}^N$.

$$\mathcal{H} = -\frac{1}{2a(1-a)N} \sum_{i\neq k} \sum_{\nu} (\xi_i^\nu - a)(\xi_k^\nu - a) S_i S_k + \frac{\gamma}{2aN} \sum_{i \neq k} S_i S_k + U \sum_i S_i. \tag{2}$$

The existence of a Hamiltonian allows a statistical mechanical analysis of the network.

3. Finite Number of Patterns Stored

In case of a finite number p of stored patterns the network state can be completely characterized by two sets of parameters. The kind of parameter ($\nu = 1, 2, \ldots, p$)

$$m^\nu = \frac{1}{aN} \langle\!\langle \sum_i (\xi_i^\nu - a) \langle S_i \rangle \rangle\!\rangle \tag{3}$$

denotes the overlap of the network state with the pattern state ν diminished by the random correlation between different patterns. The brackets $\langle\!\langle \ldots \rangle\!\rangle$ denotes averaging over the random variables ξ^ν, whereas $\langle \ldots \rangle$ stands for thermal averaging. The second kind of parameter

$$x = \frac{1}{aN} \langle\!\langle \sum_i \langle S_i \rangle \rangle\!\rangle \, . \tag{4}$$

measures the total activity in the network. The sum $m^\nu + ax$ counts the number of neurons active in the actual network state $\mathbf{S}^\nu$ which are also active in pattern ν. All states with macroscopic overlap with one pattern ν are stable states of the Monte Carlo dynamics, i.e. $m^\mu = m\ \delta_{\nu\mu}$. Superpositions of several patterns, so-called spurious states, can be destabilized by global inhibition. To suppress superpositions of s patterns a minimal inhibition strength γ_c is necessary, i.e.

$$\gamma_c = \frac{1-(2s-1)a-U}{s(1-(s-1)a)} \qquad a \ll 1. \tag{5}$$

We find a hierarchy of spurious states with decreasing stability when more patterns are superimposed. There exist parameter values for U and γ for which no spurious states exist in the network. This feature is essential because it allows the network to store time sequences [8].

The network state with no activity ($m^\nu = 0 \ \ \forall \nu,\ x = 0$) is always an equilibrium state for positive threshold $U > 0$. This state plays the role of an indicator which signals that the network has no pattern associated to an initial state, i.e. the network responds with no activity if the initial state differs in too many bits from any stored pattern.

Studies of a network with two random patterns stored show that form and size of the bassins of attraction depend on the threshold U and on the inhibition term γ. For large values of U or γ the network tolerates only a small discrepancy between an initial state and a stored pattern for associative pattern reconstruction.

4. Infinite Number of Patterns Stored

In case of infinitely many patterns stored ($\alpha \equiv p/N$ finite) the analysis of the signal to noise ratio provides much insight into the properties of the network. Assuming that the network stays in pattern state ξ^1 the part of field h_i^p which stabilizes neuron i in pattern state ξ_i^1 takes the two values

$$h_i^p = \begin{cases} 1-a-U & \text{for } \xi_i^1 = 1 \\ -a-U & \text{for } \xi_i^1 = 0 \end{cases} . \tag{6}$$

The overlap of pattern 1 with the infinitely many patterns $\nu > 1$ effects a Gaussian noise with variance $\langle\!\langle\, (h_i^f)^2 \,\rangle\!\rangle = \alpha a$. The resulting signal to noise ratios assume the values $\rho_1 = (1-a-\gamma-U)/\sqrt{\alpha a}$ and $\rho_0 = (a+\gamma+U)/\sqrt{\alpha a}$ for an active and a quiet neuron, respectively. The optimal threshold is given by the condition that $\rho_1 = \rho_0$, i.e. $U_{opt} = 1/2 - a - \gamma$ with the resulting optimal signal to noise ratio $\rho_{opt} = 1/\sqrt{4\alpha a}$. This expression shows that the structural noise with strength $\sqrt{\alpha a}$, generated by overlap with infinitely many patterns, limits the storage capacity in our network. The requirement that ρ_{opt} should not exceed a critical value yields a rough estimate of the storage capacity $\alpha_c = 1/a$.

5. Analysis of Storage Capacity

We want to investigate the storage capacities of the network in case of an infinite network storing infinite patterns. For this purpose we have to investigate the stability of pattern states, i.e. investigate in how far structural noise with strength $\sqrt{a\alpha}$ can destabilize the pattern states ξ_i^ν. For calculating the free energy density of the network we assume that a network state has macroscopic overlap with a finite number s of patterns. Averaging over the $p - s$ patterns with microscopic overlap with the network state considered yields an additional noise source. Following Amit et al [1] the partition function of the network can be evaluated by means of the replica method [11]. We find five different sets of order parameters which describe the network. The first two orderparameters are identical to that which characterize the network when only a finite number of patterns are beeing stored. Three further mean field parameters describe the randomness due to the storage of infinitly many patterns. The third order parameter is

$$q \;=\; \frac{1}{aN} \langle\!\langle \sum_i \langle\, S_i \,\rangle^2 \rangle\!\rangle \tag{7}$$

and corresponds to the Edwards-Anderson parameter in spin glass theories. The fourth and the fifth parameters are

$$r \;=\; \frac{1}{\alpha(1-a)} \sum_{\mu>s} \left(\frac{1}{aN} \sum_i (\xi_i^\nu - a) \langle\, S_i \,\rangle \right)^2 \tag{8}$$

$$y \;=\; \frac{1}{\alpha(1-a)} \sum_{\mu>s} \langle \left(\frac{1}{aN} \sum_i (\xi_i^\nu - a)\, S_i \right)^2 \rangle - \frac{2}{\alpha\bar{\beta}} \left(\gamma\, x + U + \frac{\alpha}{2} \right) \tag{9}$$

and characterize the mean and thermal fluctuations of the overlap between the thermodynamic state and the patterns which are not condensed.

The order parameters derive their importance from the fact that the free energy density f for the network can be expressed in terms of them. In fact, one can derive for the free energy density the expression

$$f = \frac{a}{2(1-a)} \sum_\nu m^{\nu 2} + \frac{a^2 \gamma}{2} x^2 + a \left(U + \frac{\alpha}{2} \right) x - \frac{a \alpha \overline{\beta}}{2} (rq - xy) \\ + \frac{\alpha}{2\beta} \left(\ln(1-C) - \frac{\overline{\beta} q}{1-C} \right) - \frac{1}{\beta} \int \mathcal{D}z \langle\!\langle \ln[1 + \exp \beta\Phi] \rangle\!\rangle$$

where we have defined $\int \mathcal{D}z f(z) = \int\limits_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) f(z)$, $C = \overline{\beta}(x-q)$ and $\Phi = \sum_\nu \frac{\xi^\nu - a}{1-a} m^\nu + \frac{\alpha \overline{\beta}}{2}(y-r) + \sqrt{a \alpha r} z$. One seeks to determine now for which values of the order parameters the free energy density assumes minimum values. After a lengthy calculation one can show that the corresponding order parameters are solutions of the equations

$$m^\nu = \frac{1}{a} \int \mathcal{D}z \langle\!\langle (\xi^\nu - a) f(\beta\Phi) \rangle\!\rangle \tag{10a}$$

$$x = \frac{1}{a} \int \mathcal{D}z \langle\!\langle f(\beta\Phi) \rangle\!\rangle \tag{10b}$$

$$q = x - \frac{1}{a} \int \mathcal{D}z \langle\!\langle \frac{1}{4 \cosh^2\left(\frac{\beta}{2}\Phi\right)} \rangle\!\rangle \tag{10c}$$

$$\frac{\alpha}{2} \overline{\beta}(y-r) = -U - \gamma x + \frac{\alpha}{2} \frac{C}{1-C} \tag{10d}$$

$$r = \frac{q}{(1-C)^2} \tag{10e}$$

where $f(x) = 1/(1 + \exp(-x))$. For a discussion of the solutions of these equations we study the special case that the network has macroscopic overlap with only one pattern, i.e. $m^\nu = m\,\delta_{\nu,1}$, and that the temperature is zero. For $m^\nu = m\,\delta_{\nu,1}$ equations (10a–c) are

$$m = \frac{1-a}{2} \Big(\mathrm{erfc}(-\Phi_1) - \mathrm{erfc}(-\Phi_0) \Big) \tag{11a}$$

$$x = \frac{1}{2} \mathrm{erfc}(-\Phi_1) + \frac{1-a}{2a} \mathrm{erfc}(-\Phi_0) \tag{11b}$$

$$C = \frac{1-C}{\sqrt{2\pi \alpha a x}} \left(a \exp(-{\Phi_1}^2) + (1-a) \exp(-{\Phi_0}^2) \right) \tag{11c}$$

with

$$\Phi_1 = \frac{1-C}{\sqrt{2 \alpha a x}} \left(m - U - \gamma x + \frac{\alpha}{2} \frac{C}{1-C} \right) \quad \text{and}$$

$$\Phi_0 = \frac{1-C}{\sqrt{2 \alpha a x}} \left(-\frac{a}{1-a} m - U - \gamma x + \frac{\alpha}{2} \frac{C}{1-C} \right).$$

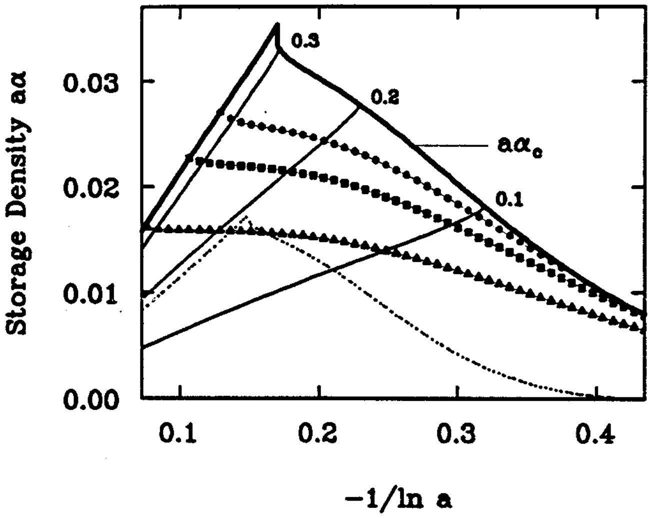

Fig. 1: Storage capacity α_c multiplied by a (bold line) as a function of $-[\ln a]^{-1}$ ($U = 0.7$, $\gamma = 0.0$). The thin lines denote (α, a) values with constant information (0.1,0.2,0.3 bits) stored per synapse. The symbols ○, □, △ indicate contour lines of (α, a) values for which networks recall with 0.95 (○), 0.97 (□) and 0.99 (△) accuracy.

The parameters r and y can be expressed in terms of the other parameters and have been inserted into ($11a$–c).

Equations ($11a$–c) have to be solved numerically. As long as solutions with $m + ax \leq 1$ exist, the network can be assumed to work properly. One investigates then for which values α_c of α the solutions of ($11a$–c) do not satisfy the condition $m + ax \leq 1$. This value is called the critical storage capacity and corresponds to the maximum fraction of patterns stored. This analysis shows a remarkable dependence of the critical storage capacity α_c on the activity parameter a. Storage in the network is limited by the variance of the structural noise, i.e. by the value of $(1 - C)/\sqrt{2\alpha a x}$. The dependence of the critical storage capacity α_c is shown in Fig. 1. In the range $-1/\ln a < 0.17$ α_c scales as

$$a\alpha_c \approx -\frac{0.1991}{\ln a} + 1.394 \cdot 10^{-3}. \tag{12}$$

In the limit of extremely small activation ($a < 10^{-50}$) the storage capacity is given by $\alpha_c = C(U)/(a \ln a)$ where $C(U)$ depends on the threshold U ($C(0.7) = 0.22$).

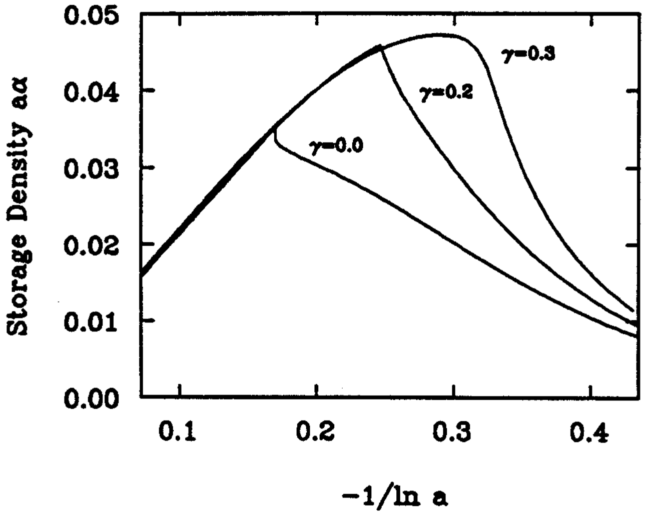

Fig. 2: Critical storage capacity for networks with different inhibition strength ($U + \gamma = 0.7$).

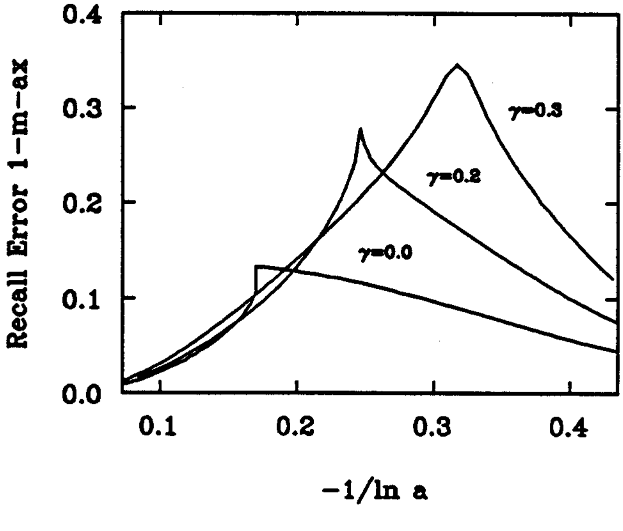

Fig. 3: Recall error $1 - m - ax$ at α_c for networks with different global inhibition ($U + \gamma = 0.7$).

Feigel'man and Tsodycs have shown $C(1) = 0.5$ [5]. The functional dependence of α_c on a originates from the growth of $\frac{1-a}{2a}\mathrm{erfc}(-\Phi_0)$ in (11b) which destabilizes the solution $m^\nu = m\,\delta_{\nu,1}$. The scaling behaviour (12) coincides with the upper bound determined by Gardner [12] who derived for the optimum storage capacity a dependence $\alpha_c \sim -[a \ln a]^{-1}$ as well. This coincidence suggests that storage in our network for small a is nearly optimal.

One can analyze the properties of the solutions of equation ($11a$–c) for values $\alpha < \alpha_c$ and determine the recall error, defined as $R = 1 - (m + ax)$ as well as the information stored in the network in units of bits/synapse [see below]. One can then seek solutions of the equations (11) with $m^1 \neq 0 \wedge m^2 \neq 0 \wedge m^\nu \equiv 0 \; \forall \nu > 2$. Such solutions correspond to situations when the memory errouneously retrieves patterns which are superpositions of stored patterns. Such network states are referred to as spurious states. Such states do not exist near α_c since in this region they are destabilized by structural noise. The region in the (α, a)–space where spurious states exist is located below the dashed line in Fig. 2.

The dependence of the critical storage density α_c on a for networks with global inhibition is shown in Fig. 2. The threshold U is adjusted to different inhibition values according to the formula $U + \gamma = 0.7$. α_c is seen to increase for intermediate activation levels a. The range of the increase depends on the parameter γ, i.e. on the strength of global inhibition. The critical storage capacity for small activity levels a depends only very slightly on γ.

The recall error R at the critical storage density α_c for various values of γ is shown in Fig. 3. In the limit $a \to 0$ all recall errors vanish. Such behaviour has also been found for matrix memory models [13]. Figure 3 shows that the error rate of networks with strong global inhibition at $\alpha = \alpha_c$ can grow to values larger than 30 percent. This result implies, that in case strong inhibition is required to prevent spurious states, acceptable recall behaviour can be expected only at low mean activity levels.

The effectivity of the network as an associative memory can be evaluated by calculating the information content per synapse, i.e. the number of bits stored by one synapse. We obtain the information content of a network by calculating the entropy of the pattern states diminished by the loss of information due to errors in the equilibrium state. The entropy of the network state assumes the value

$$\begin{aligned} S_N &= -a\overline{m}\ln(a\overline{m}) - a(1-\overline{m})\ln(a - a\overline{m}) - a(x-\overline{m})\ln(ax - a\overline{m}) \\ &\quad (1 - a(1 + x - \overline{m}))\ln(1 - a(1 + x - \overline{m})) \end{aligned}$$

$(\overline{m} = m^\nu + ax)$ whereas the entropy of an arbitrary pattern state is $S_P = -a\ln a - (1-a)\ln(1-a)$. The total information stored in the network in units of bits is $I(\alpha, a) = \alpha N^2(S_P - \Delta S)/\ln 2$ with the loss of information accounted for by $\Delta S = S_N - S_P$. In Fig. 1 the contour lines in the (α, a) space with $I/N^2 = 0.1, 0.2, 0.3$ are shown. The phase diagram shows clearly that more information can be stored in a network with a sparse coding scheme, i.e. small a. A quantitative analysis reveals that for $U = 0.7$ and $\gamma = 0$ the information content per synapse I/N^2 is as large as 0.38 bits/synapse. The case of clipped synapses can be treated as in the Hopfield model [14] and yields a slightly reduced information content ($I = 0.28$ for $a = 10^{-6}$, $U = 0.75$).

6. Conclusion

In this letter we have proposed a neural network which can store and recall patterns with low activity, i.e. sparce patterns. The system has been analyzed by means of methods developed for statistical mechanics of spin glasses. The network proposed assumes optimum storage capacity α_c. Since many interesting information processing tasks involve sparse coding, the proposed network appears to be a most promising candidate for practical applications of model neural networks.

This work has been supported by the German Ministry for Science and Technology (ITR-8800-G9). JB is supported by a NATO grant (DAAD 300/402/513/9).

REFERENCES

[1] D.J. Amit, H. Gutfreund and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); Phys. Rev. Lett. **55**, 1530 (1985).

[2] W. Krauth, M. Mezard, J. Physics A **20**, L745 (1987)

[3] G. Pöppel, U. Krey, Europhys. Lett. **4**, 979 (1987)

[4] E. J. Gardner, J. Phys. A **21**, 257 (1988).

[5] M. V. Tsodycs and M. V. Feigel'man, Europhys. Lett. **6**, 101 (1988)

[6] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982), and bf 81, 3088 (1984).

[7] D.J. Amit, H. Gutfreund and H. Sompolinsky, Phys. Rev. A **35**, 2293 (1987).

[8] J. Buhmann, K. Schulten, Europhys. Lett. **4**, 1205 (1987); pp. 231 in *Neural Computers*, R. Eckmiller, C. v.d. Malsburg, Eds. (Springer, Berlin, 1987).

[9] J. Buhmann, Thesis, Technische Universität München, (1988).

[10] M. Abeles *Local Cortical Circuits.* (Springer, Berlin, 1982).

[11] A. D. Bruce, E. J. Gardner, D. J. Wallace, J. Phys. A **20**, 2909 (1987); after completion of our calculations we took notice of this paper where the case $a = 0.5$, $U = 0$, $\gamma = 0$ is solved.

[12] Gardner [4] yields an information content per synapse $I = 1/(2\ln 2) = 0.72$ (0.36) for asymmetric (symmetric) networks.

[13] D. J. Willshaw, O. P. Buneman and H. C. Longuet–Higgins, Nature 222, 960 (1969); G. Palm, Biol. Cybern. **36**, 19 (1980).

[14] H. Sompolinsky, Phys. Rev. A **34**, 2571 (1986)