Long Term Memory Storage Capacity
of Multiconnected Neural Networks

P. Peretto[1] and J. J. Niez[2]

[1] Centre d'Etudes Nucléaires de Grenoble DRF/PSC, BP 85 X, F-38041 Grenoble, France
[2] Centre d'Etudes Nucléaires de Grenoble LETI/MSC, BP 85 X, F-38041 Grenoble, France

Abstract. Quantitative expressions of long-term memory storage capacities of complex neural network are derived. The networks are made of neurons connected by synapses of any order, of the axono-axonal type considered by Kandel et al. for example. The effect of link deletion possibly related to aging, is also considered. The central result of this study is that, within the framework of Hebb's laws, the number of stored bits is proportional to the number of synapses. The proportionality factor however, decreases when the order of involved synaptic contact increases. This tends to favor neural architectures with low-order synaptic connectivities. It is finally shown that the memory storage capacities can be optimized by a partition of the network into neuron clusters with size comparable with that observed for cortical microcolumns.

1 Introduction

A number of authors have investigated the long-term memory storage capacity of neural networks. Whereas most derivations either appealed to qualitative considerations or have used computer simulations (cf. Kohonen 1978; Palm 1981), a simple analytical treatment of this problem has been proposed by Hopfield in 1982. In this study, the networks he considers are fully connected sets of binary neurons i.e., neurons with only two internal states, firing or silent. Patterns are stored using the Hebbian rule (cf. Hebb 1949) and the states of neurons are determined by a majority mechanism (see Sect. 2.1). Hopfield shows that the stored configurations are stable states as long as their number does not exceed a limit value M, the memory storage capacity, given by:

$$M = N/K \qquad (1)$$

where N is the number of neurons and K is a constant. K depends on the required degree of configuration stability. For example $K \simeq 20$ if 99% of patterns stored in a network comprising some tens of neurons, are to be stable.

The derivation of (1) rises questions, some relating to the dynamics of the system and others to the topology of the network. For example two-state neurons and abrupt threshold rules appear as crude oversimplifications of the real dynamics. Hopfield (1982) shows, however, that his approach is robust with respect to the smoothening of the response step function in a sigmoid function. We have also proven (Peretto 1984) that the storage capacity is weakly noise dependent as long as the noise parameter is below a critical value. Noise also transforms an all-or-nothing function into a continuous one of probabilistic character. To conclude, improvements in the modeling of the dynamics will probably not modify the Hopfield's conclusions.

Changing the structure of the networks certainly does. Electron micrographs reveal that the cortical neurons are packed in tightly intertwinned bundles of fibers (Roney et al. 1979; Shaw et al. 1982). The assumption that neurons are linked only by binary connections is too simple. Synapses can modify the membrane potentials of other synapses as well as those of dendrites. The role of axono-axonal processes has been emphasized by Kandel et al. (1965), Carew et al. (1971), Kandel (1976) in their studies of neural systems of invertebrates and it is likely that these junctions can also play a role in the functionning of central nervous systems of higher organisms. They have been shown to exist in the spinal chord of vertebrates in particular.

In Sect. 2.2 it is shown how the rules of cooperative neural networks models, described in Sect. 2.1 have to be modified to account for the existence of higher order synaptic junctions. Traces of stored patterns which were imprinted in axono-dendritic junctions are now dispatched in synaptic contacts of any order so increasing the memory capacity for a given number N

of neurons in the network. The optimal learning rules for synapses of any type are derived in Sect. 2.3. Section 3 is devoted to the calculation of memory storage capacities of ideal multi-connected systems i.e., fully connected networks with optimal synaptic efficacies determined by noiseless learned configurations. The results depend on a coupling parameter γ which settles the relative orders of magnitude between synapses of different complexities. The storage capacities are calculated in the low coupling and in the strong coupling limits in Sects. 3.2 and 3.3. Intermediate situations cannot treated exactly except for a mixture of axono-dendritic (binary) and axono-axono-dendritic (ternary) interactions (in Sect. 3.4). Some of the constraints are released in Sect. 4 which makes the model more realistic. In Sect. 4.1 the memory storage capacity of networks experiencing blurred configurations during the learning cession is derived. The effect of deleting links at random is studied in Sect. 4.2. This section ends with a discussion on the change induced by replacing the optimal efficacies by more genuine Hebbian expressions. Finally it is shown in Sect. 5 that a partition of the system strongly enhances its combinatorial capacities. The calculation of the best partition leads to size of the parts which are comparable with the sizes of cortical microcolumns.

2 A Description of Multiconnected Neural Networks

2.1 Neuronal Activity

The activity of a neural network manifests itself by the emission of standard signals, the action potentials. The state of a neuron i is set to $S_i = 1$ if it is firing and $S_i = 0$ if it is silent. S_i is determined by the relative value of the membrane potential V_i of the hillock zone of neuron i with respect to its threshold value θ_i:

$$S_i = \mathbf{1}\,(V_i - \theta_i). \tag{2}$$

$\mathbf{1}$ is the Heavyside step function: $\mathbf{1}\,(x > 0) = 1$, $\mathbf{1}\,(x < 0) = 0$.

The membrane potential V_i is a linear combination of the post-synaptic potentials triggered by the activities impinging onto the neuron i:

$$V_i = \sum_j C_{ij} S_j. \tag{3}$$

C_{ij} is the efficacy of a synapse (ij) linking the upstream neuron j to the downstream neuron i.

Equations (2) and (3), which have been proposed by McCulloch and Pitts (1943), greatly simplify the reality: all the dynamic aspects have been eliminated, noise has not been introduced, even the linearity involved in (3) is questionable. However as far as the memory storage capacities are the only quantities of interest, these complications can be forgotten because they only depend on the steady properties of the

system. For a more detailed discussion of the dynamics of noisy neural networks the reader is referred to Peretto and Niez (1986). A study of the dynamics of small stochastic systems can be found in Thomson and Gibson (1981).

2.2 Multiconnected Networks

Let us assume that a neuron k makes a contact through a synapse (ijk) with a binary synapse (ij) (see Fig. 1). The activity of k modifies the synaptic efficacy between the neurons j and i. The efficacy of (ij) is C_{ij} when $S_k = 0$ and $C_{ij} + C_{ijk}$ when $S_k = 1$. The efficacy of (ij) is therefore given by

$$C_{ij} + C_{ijk} S_k.$$

Summing up over all axono-axonal junctions k leads to a membrane potential V_i (3), given by:

$$V_i = \sum_j (C_{ij} + \sum_k C_{ijk} S_k) S_j$$

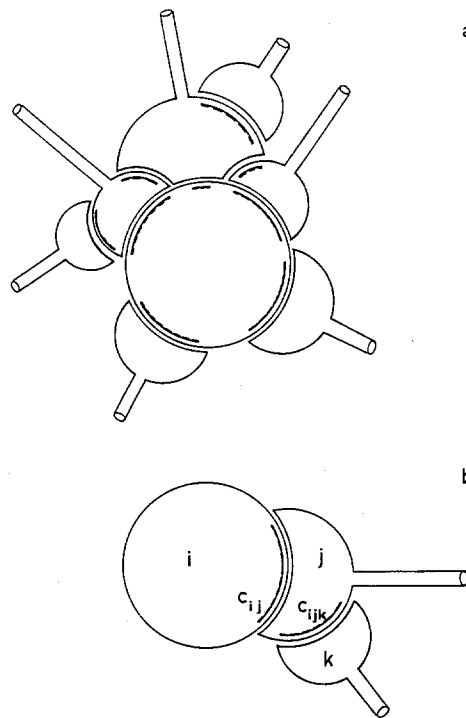$$= \sum_j C_{ij} S_j + \sum_{jk} C_{ijk} S_j S_k.$$

Fig. 1. a A schematic view of multi-synaptic contacts impinging on a dendrite as suggested by electron micrographs of central nervous tissues. Such an electron mirograph taken from a small region from the brain of a rat can be seen in B. Alberts et al. (1983), p. 1037 for example Fig. b. Labels of synaptic contacts: i, j and k are neuron labels. C_{ij} is an axono-dendritic synapse linking an axonal branch of neuron j to a dendrite of neuron i. C_{ijk} is an axono-axono-dendritic synapse which enables the activity of neuron k to modify the efficacy of synapse (ij)

The synapses (ijk) can in turn be modified by an other junction with a neuron l and so on. Finally

$$V_i = \sum_j C_{ij} S_j + \sum_{jk} C_{ijk} S_j S_k$$
$$+ \sum_{jkl} C_{ijkl} S_j S_k S_l + \dots \quad (4)$$

The memory storage capacities depend crucially on the relative magnitudes of junctions of different orders. As there is few quantitative results in this field, it will be assumed that the efficacies of p-ary synapses scale according to a parameter γ. For example

$$|C_{ijk}| \simeq \gamma |C_{ij}|,$$

γ only settles the overall strength of synapses of any order. The precise values of the synaptic efficacies c_{ij} are given by optimal learning rules which are discussed in the next section.

2.3 The Learning Rules

Memory is the ability for a system to store a number of patterns α which are sets

$$I^\alpha = \{S_i^\alpha\}; \quad \alpha = 1, \dots, M.$$

of N elementary states S_i^α. A pattern I^α is efficiently stored if it is stable; that is, according to (2):

$$V_i^\alpha > \theta_i; \; \forall i \quad \text{such as} \quad S_i = 1$$
$$V_i^\alpha < \theta_i; \; \forall i \quad \text{such as} \quad S_i = 0 \quad (5)$$

where (cf. (4))

$$V_i^\alpha = \sum_j C_{ij} S_j^\alpha + \sum_{jk} C_{ijk} S_j^\alpha S_k^\alpha + \dots \quad (6)$$

The conditions (5) can be lumped into one formula by introducing the variable $\sigma_i = 2S_i - 1$.

The equations (5) become:

$$E_i^\alpha = h_i^\alpha \sigma_i^\alpha > 0; \; \forall i, \alpha \quad (7)$$

with

$$h_i^\alpha = V_i^\alpha - \theta_i.$$

The "field" h_i^α is given by:

$$h_i^\alpha = J_i + \sum_j J_{ij} \sigma_j^\alpha + \sum_{jk} J_{ijk} \sigma_j^\alpha \sigma_k^\alpha$$
$$+ \sum_{jkl} J_{ijkl} \sigma_j^\alpha \sigma_k^\alpha \sigma_l^\alpha + \dots$$

where

$$J_i = -\theta_i + \sum_j \frac{C_{ij}}{2} + \sum_{jk} \frac{C_{ijk}}{4}$$
$$+ \sum_{jkl} \frac{C_{ijkl}}{8} + \dots$$

$$J_{ij} = \frac{C_{ij}}{2} + \sum_k \frac{C_{ijk} + C_{ikj}}{4}$$
$$+ \sum_{kl} \frac{C_{ijkl} + C_{ikjl} + C_{ilkj}}{8} + \dots$$
$$J_{ijk} = \frac{C_{ijk}}{4} + \sum_l \frac{C_{ijkl} + C_{ijlk} + C_{ilkj}}{8} + \dots \quad (8)$$

$$J_{ijkl} = \frac{C_{ijkl}}{8} + \dots$$

[for a general formula see (40)].

The best way of stabilizing I^α is achieved by aligning the local field h_i^α along the direction σ_i^α.

$$h_i^\alpha = \lambda \sigma_i^\alpha$$

where λ is some positive constant. From (7) and (8) one obtains:

$$J_i \sigma_i^\alpha + \sum_j J_{ij} \sigma_i^\alpha \sigma_j^\alpha$$
$$+ \sum_{jk} J_{ijk} \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha + \dots = \lambda.$$

Every term of this sum is maximum when the interactions J combine coherently with the patterns I^α.

$$J_{ij} = \sigma_i^\alpha \sigma_j^\alpha$$
$$J_{ijk} = \gamma \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha. \quad (9)$$
$$\vdots$$

It is often assumed (Little and Shaw 1978) that the neurons work close to their thresholds. In other words θ_i adjust itself so as $J_i = 0$: a non-zero J_i, would be equivalent to a local polarizing field which, if strong enough, would wipe out all memory properties. Indeed, assuming large J_i's, there would be an unique steady state $\{\sigma_i^0\}$ with $J_i \sigma_i^0 > 0$: the system would not be able to store more than one pattern. If the system experiences a dynamically noisy pattern $I^\alpha(t)$, the interactions J are replaced by time statistical averages:

$$J_{ij} = \langle \sigma_i^\alpha \sigma_j^\alpha \rangle$$
$$J_{ijk} = \gamma \langle \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha \rangle. \quad (10)$$
$$\vdots$$

If the system experiences several patterns a natural, but not optimal, generalization of (10) is:

$$J_{ij} = \sum_{\alpha=1}^{M} \langle \sigma_i^\alpha \sigma_j^\alpha \rangle$$
$$J_{ijk} = \gamma \sum_{\alpha=1}^{M} \langle \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha \rangle \quad (11)$$
$$\vdots$$
$$J_{ij\dots p} = \gamma^{p-2} \sum_{\alpha=1}^{M} \langle \sigma_i^\alpha \sigma_j^\alpha \dots \sigma_p^\alpha \rangle.$$

The optimal general binary synaptic efficacies have been found by Kohonen (1978, 1984) for associative

filters and the idea has been applied by Personnaz et al. (1985) to fully connected networks. The set of binary interactions $J = \{J_{ij}\}$ can be written as a $N \times N$ matrix J,

$$J = \xi \cdot \xi^T$$

where ξ is a $M \times N$ matrix whose columns are the stored patterns. Kohonen has shown that the optimal matrix is J':

$$J' = \xi \cdot \xi^+$$

where ξ^+ is the pseudo-inverse of ξ, rather than the matrix J. However the increase of memory capacity is not large (a factor of two or so) because the number of orthogonal patterns is limited to N (see Appendix A1). On the other hand the method is difficult to generalize to synaptic junctions of higher orders. Therefore the memory storage capacities of neural networks will be calculated using the simple expressions (11). These rules are reminiscent of the Hebbian rules although the later applies to real synaptic interactions $c_{ij}...$ rather than to the optimal ones $J_{ij}...$ This point is discussed in Sect. 4.3. The occurence of Hebbian mechanism in real neural network has been discussed by a number of authors, Stent (1973), Woody (1982), Lynch et al. (1984) in particular. A thorough account of the experimental situation is given by Changeux (1986). Theoretical considerations can be found in Palm (1982). Hebb's rules are also invoked in the explanation of neuropsychological experiments (Delacour 1981).

3 Memory Storage Capacities of Fully Connected Networks

3.1 General Derivation

A stored pattern I^α is stable if $E_i^\alpha > 0$ for all i's (cf. (7)). From (7), (8) and (11) E_i^α reads:

$$E_i^\alpha = \sum_j J_{ij}\sigma_i^\alpha\sigma_j^\alpha + \sum_{jk} J_{ijk}\sigma_i^\alpha\sigma_j^\alpha\sigma_k^\alpha + ...$$
$$= \sum_{\beta=1}^M \left(\sum_j \sigma_i^\alpha\sigma_j^\alpha\sigma_i^\beta\sigma_j^\beta + \gamma\sum_{jk}\sigma_i^\alpha\sigma_j^\alpha\sigma_k^\alpha\sigma_i^\beta\sigma_j^\beta\sigma_k^\beta + ... \right). \quad (12)$$

It is assumed that the input patterns are noiseless $\langle \sigma_i^\alpha\sigma_j^\alpha \rangle = \sigma_i^\alpha\sigma_j^\alpha$. The case of noisy inputs is treated in Sect. 4.1. Following the argument given by Hopfield (1982) the term $\beta = \alpha$ is singled out. This allows the quantity E_i^α to be written as a sum of two terms

$$E_i^\alpha = v_i^\alpha + \varrho_i^\alpha, \quad (13)$$

where v_i^α is a coherent contribution and ϱ_i^α an incoherent contribution to E_i^α. The coherent contribution is a positive number. The incoherent contribution is a stochastic variable with a zero mean value if the stored patterns are stochastically independent. The stability of an elementary state σ_i^α is limited by the fluctuations of the incoherent term. More precisely the probability Y for σ_i^α to be unstable is the probability that

$$\varrho_i^\alpha < -v_i^\alpha,$$

v_i^α is given by:

$$v_i^\alpha = v = \sum_j 1 + \gamma\sum_{jk} 1 + ... + \gamma^{p-2}\sum_{j...p} 1$$
$$= N\sum_{r=0}^{p-2} (\gamma N)^r = N \cdot \frac{1-(\gamma N)^{p-1}}{1-\gamma N}. \quad (14)$$

The incoherent part is a sum of random variables

$$\varrho_i^\alpha = \sum_{\beta \neq \alpha} Z_i^{\alpha\beta} \quad (15)$$

with

$$Z_i^{\alpha\beta} = \sigma_i^\alpha\sigma_i^\beta \left(\sum_j \sigma_j^\alpha\sigma_j^\beta + \gamma\sum_{jk}\sigma_j^\alpha\sigma_j^\beta\sigma_k^\alpha\sigma_k^\beta + ... \right).$$

If M is large enough for the central limit theorem to be applicable, ϱ_i^α is a Gaussian variable with a mean square deviation given by:

$$\langle (\varrho_i^\alpha)^2 \rangle = \langle \varrho^2 \rangle = (M-1) \langle (Z_i^{\alpha\beta})^2 \rangle$$
$$= (M-1) \langle Z^2 \rangle. \quad (16)$$

All indexes have been skipped because all patterns and all sites play the same role.

The probability distribution of ϱ is $P(\varrho)$:

$$P(\varrho) = \frac{1}{\sqrt{2\pi\langle\varrho^2\rangle}} \cdot \exp - \frac{\varrho^2}{2\langle\varrho^2\rangle} \quad (17)$$

and

$$Y = P(\varrho < -v) = \int_{-\infty}^{-v} P(\varrho)d\varrho$$
$$= \int_v^\infty P(\varrho)d\varrho = 1 - \phi(X) \quad (18)$$

where $\phi(X)$ is defined by

$$\phi(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^X \exp - \frac{t^2}{2} dt$$

and

$$X = \frac{v}{\sqrt{\langle\varrho^2\rangle}}. \quad (19)$$

Let P^* be the probability that the whole pattern is stable

$$P^* = (1-Y)^N = \phi(X)^N. \quad (20)$$

The Eq. (20) is inverted leading to

$$X = \phi^{-1}(P^{*1/N}). \quad (21)$$

Equations (16), (19) and (21) give the general expression of memory storage capacities.

$$M = 1 + \frac{v^2}{\langle Z^2 \rangle \cdot K(N, P^*)} \tag{22}$$

with

$$K(N, P^*) = (\phi^{-1}(P^{*\,1/N}))^2 .$$

In the limit $P^* \to 1$ the asymptotic expansion of $\phi(X)$

$$\lim_{X \to \infty} \phi(X) \simeq 1 - \frac{1}{X\sqrt{2\pi}} \exp - \frac{X^2}{2}$$

yields K as the solution of the implicit equation:

$$\frac{\exp - K/2}{\sqrt{K}} = - \frac{\sqrt{2\pi}}{N} \cdot \mathrm{Ln}\, P^* . \tag{23}$$

$K(N, P^*)$ is a very slowly varying function of N. For large N's it is nearly logarithmic. With $P^* = 0.5$, $K = 8.706$ for $N = 400$ and $K = 6.269$ for $N = 100$ i.e., values close to the ones obtained using numerical simulations.

The rest of the paper is mainly devoted to the calculation of the average v and to the calculation of the mean square deviation $\langle Z^2 \rangle$ to find out the memory storage capacities, in a number of specific situations.

3.2 The Weak Coupling Limit

According to (15) $Z_i^{\alpha\beta}$ can be written as

$$Z_i^{\alpha\beta} = \sigma_i^\alpha \sigma_i^\beta \, (I^{\alpha\beta} + \gamma \, (I^{\alpha\beta})^2 + \gamma^2 (I^{\alpha\beta})^3 + \ldots) \tag{24}$$

where

$$I^{\alpha\beta} = \sum_j \sigma_j^\alpha \sigma_j^\beta = I^\alpha \cdot I^\beta$$

is the scalar product between the stored patterns I^α and I^β. In the weak coupling limit it is assumed that

$$\gamma \lesssim 1/N$$

(and a fortiori $\gamma \ll 1/\sqrt{N}$).
Equation (24) becomes

$$Z_i^{\alpha\beta} \simeq \sigma_i^\alpha \sigma_i^\beta \cdot \frac{I^{\alpha\beta}}{1 - \gamma I^{\alpha\beta}}, \tag{25}$$

$I^{\alpha\beta}$ is a Gaussian variable with a mean square deviation of N

$$P(I^{\alpha\beta}) = \frac{1}{\sqrt{2\pi N}} \cdot \exp - \frac{(I^{\alpha\beta})^2}{2N} . \tag{26}$$

This implies

$$\gamma \cdot \sqrt{\langle (I^{\alpha\beta})^2 \rangle} \simeq \gamma \cdot \sqrt{N} \lesssim \frac{1}{\sqrt{N}} \ll 1$$

and

$$Z_i^{\alpha\beta} \simeq \sigma_i^\alpha \sigma_i^\beta I^{\alpha\beta} .$$

One deduces $\langle Z^2 \rangle \simeq \langle (I^{\alpha\beta})^2 \rangle = N$. On the other hand $v \simeq N/(1 - \gamma N)$ in the weak coupling limit. The memory storage capacity is therefore given by:

$$M \simeq 1 + \frac{N}{(1 - \gamma N)^2 \cdot K(N, P^*)} . \tag{27}$$

When the neurons interact through only binary synapses $\gamma = 0$ and the memory capacity is:

$$M_2 \simeq \frac{N}{K(N, P^*)} . \tag{28}$$

A similar result with $K(N, P^*)$ replaced by a logarithmic function on N has been derived by Weisbuch et al. (1985).

3.3 The Strong Coupling Limit

Here it is assumed that $\gamma\sqrt{N} > 1$. To avoid a divergence of the series (15) the order of connection is limited to p. The behavior of the network is dominated by the last term of the series

$$Z_i^{\alpha\beta} = \sigma_i^\alpha \sigma_i^\beta \gamma^{p-2} (I^{\alpha\beta})^{p-1} .$$

Now, if two stochastic variables X and Z are related by

$$Z = f(X) \quad \text{i.e.} \quad X = f^{-1}(Z)$$

the probability distribution $Q(Z)$ of Z is given by:

$$Q(Z) = P(f^{-1}(Z)) \cdot \frac{d}{dZ} (f^{-1}(Z)),$$

where $P(X)$ is the probability distribution of X.

Here $Z = \gamma^{p-2} X^{p-1}$. Dropping the $\sigma_i^\alpha \sigma_i^\beta$ factor which is unessential in the calculation of $\langle Z^2 \rangle$, one obtains:

$$Q(Z) = \frac{1}{(p-1)\gamma^{\frac{p-2}{p-1}} \sqrt{2\pi N}} \cdot Z^{-\frac{p-2}{p-1}}$$

$$\cdot \exp - \frac{Z^{\frac{2}{p-1}}}{2\gamma^{2\left(\frac{p-2}{p-1}\right)} N} .$$

The mean square deviation of z

$$\langle Z^2 \rangle = 2 \int\limits_0^\infty Z^2 \, Q(Z) dZ$$

is calculated using the transformation

$$z = \frac{Z^{\frac{2}{p-1}}}{2\gamma^2\left(\frac{p-2}{p-1}\right) N}$$

leading to:

$$\langle Z^2 \rangle = \frac{1}{\sqrt{\pi}} \, 2^{p-1} \gamma^{2(p-2)} \Gamma(p-\tfrac{1}{2}) N^{p-1}$$

where Γ is the gamma-function. Finally the memory storage capacity of a p-connected network is given by

$$M_p \sim 1 + \frac{\sqrt{\pi} N^{p-1}}{2^{p-1} \Gamma(p-\tfrac{1}{2})} \cdot \frac{1}{K(N, P^*)} . \tag{29}$$

For binary network $p = 2$, the formula (28) is recovered. The memory storage capacity of ternary networks $p = 3$ reads, using $\Gamma(5/2) = 3\sqrt{\pi}/4$,

$$M_3 \sim 1 + \frac{N^2}{3K(N, P^*)}$$

and therefore $M_3/M_2 = N/3$. The number of stored bits per neuron increases as p increases but the number of stored bits per synapse decreases. Indeed this number is

$$b_p = \frac{M_p N}{N^p} \simeq \frac{\sqrt{\pi}}{2^{p-1} \Gamma(p-\tfrac{1}{2}) K(N, P^*)} \tag{30}$$

in particular $b_3/b_2 = 1/3$.

3.4 Networks Connected Through Binary and Ternary Synapses

The best synaptic organization is probably a mixture of synapses of low orders. It is therefore interesting to have an analytical expression for the memory storage capacities of neural networks involving only binary and ternary connections. The coherent part of such a network is given by:

$$v = N + \gamma N^2 .$$

The mean square deviation of the stochastic variables $Z_i^{\alpha\beta}$

$$Z_i^{\alpha\beta} = \sigma_i^\alpha \sigma_i^\beta (I^{\alpha\beta} + \gamma(I^{\alpha\beta})^2)$$

is derived in Appendix A2

$$\langle Z^2 \rangle = N(1 + 2^{5/2} \pi^{-1/2} \gamma N^{1/2} + 3\gamma^2 N) . \tag{31}$$

The memory storage capacity reads

$$M \sim 1 + \frac{N(1 + \gamma N)^2}{1 + 2^{5/2} \pi^{-1/2} \gamma N^{1/2} + 3\gamma^2 N}$$
$$\cdot \frac{1}{K(N, P^*)} . \tag{32}$$

It is worth-noting that this formula is valid in the whole range of values of γ.

4 A More Detailed Analysis of Memory Storage Capacities of Neural Networks

4.1 Noisy Input Pattern

A number of assumptions have been made in the above derivation. They are more closely considered in the present section.

One of the hypotheses was the absence of noise in stored patterns. Introduction of noise transforms the coherent term v into a stochastic variable. Let m be the probability that a bit is erroneous during the learning session.

$$\langle \sigma_i \rangle = \sigma_i(1 - 2m)$$

Consider the contribution, due to binary interactions, to the coherent term

$$\sum_j \langle \sigma_i^\alpha \sigma_j^\alpha \rangle \sigma_i^\alpha \sigma_j^\alpha$$
$$= N((1-m)^2 - 2m(1-m) + m^2) = N(1-2m)^2$$

where the various factors in the bracket relate to the probability for σ_i^α and σ_j^α to be both correct, for one of these to be uncorrect and for both to be uncorrect. More generally

$$\sum_{j \ldots l} \langle \sigma_i^\alpha \sigma_j^\alpha \ldots \sigma_l^\alpha \rangle \, \sigma_i^\alpha \sigma_j^\alpha \ldots \sigma_l^\alpha = N^{p-1}(1-2m)^p .$$

The mean value of the coherent term is given by:

$$\bar{v} = \sum_{r=2}^{p} \gamma^{r-2} N^{r-1} (1 - 2m)^r . \tag{33}$$

Assuming that the generation of errors is a Poissonian process, the mean square deviation of v is approximately

$$\langle \Delta v^2 \rangle = m \sum_{r=2}^{p} r \gamma^{r-2} N^{r-1} .$$

The probability Y is now given by an integral over a correlation of the Gaussian probability distributions of v and ϱ which is also a Gaussian distribution: therefore all results found in Sect. 3-1 hold with v replaced by $\bar{v}$ and $\langle \varrho^2 \rangle$ by $\langle \varrho^2 \rangle + \langle \Delta v^2 \rangle$. It appears that $\langle \Delta v^2 \rangle$ can be neglected in any cases. In the weak coupling limit the series (33) is summed out and the memory storage capacity is:

$$M \simeq 1 + \frac{(1-2m)^4 N}{(1 - \gamma N(1-2m))^2} \cdot \frac{1}{K(N, P^*)} \tag{34}$$

whereas in the limit of strong coupling

$$M \simeq 1 + \frac{\pi^{1/2}(1-2m)^{2p} N^{p-1}}{2^{p-1} \Gamma(p-\tfrac{1}{2})} \cdot \frac{1}{K(N, P^*)} . \tag{35}$$

These results show that the blurring of input patterns sensitively reduces the memory storage capacities of neural networks.

4.2 Study of Randomly Connected Networks

It has been assumed so far that the graphs of connections are complete, at least up to order p. If the network is uncompletely connected the synaptic efficacies can be written as

$$C_{ij...l} = U_{ij...l} \cdot J_{ij...l} \qquad (36)$$

where $U_{ij...l} = 1$ if the whole chain of synaptic junctions from neuron 1 to neuron i is complete and $U_{ij...l} = 0$ otherwise. The parameters $J_{ij...l}$ depend on learning as in (11).

One first consider the binary connected networks. The coherent part is given by:

$$\bar{v}_i = \sum_j \bar{U}_{ij} = N\bar{\omega}$$

where the bar is an ensemble average over the realizations of synaptic connections. $\bar{\omega}$ is the probability for a binary synapse to exist. The distribution of the incoherent term

$$Z_i^{\alpha\beta} = \sigma_i^\alpha \sigma_i^\beta \sum_j U_{ij} \sigma_j^\alpha \sigma_j^\beta \qquad (37)$$

is still Gaussian. Indeed $Z_i^{\alpha\beta}$ is a sum of independent variable v_j

$$Z_i^{\alpha\beta} = \sum_j v_j.$$

The probability distribution of v_j is:

$$P(v_j) = (1 - \bar{\omega})\delta(v_j) + \frac{\bar{\omega}}{2}\delta(v_j - 1) + \frac{\bar{\omega}}{2}\delta(v_j + 1).$$

Therefore

$$\langle \Delta v_j^2 \rangle = \int v_j^2 P(v_j) dv_j = \bar{\omega}$$

and $\langle Z^2 \rangle = N\bar{\omega}$.

The memory strage capacity of binary connected networks therefore depends linearly on the amount $\bar{\omega}N$ of existing binary synapses:

$$M_2 \simeq \frac{\bar{\omega}N}{K(N, P^*)}. \qquad (38)$$

In general networks the coherent term reads:

$$\bar{v}_i = \sum_j \bar{U}_{ij} + \gamma \sum_{jk} \bar{U}_{ijk} + \cdots$$

The coefficients $U_{ij...l}$ factorize

$$U_{ij...l} = u_2(ij) \cdot u_3(jk)...u_p(ml)$$

the subscript refers to the order of the synapse in the chain. Assuming that the synapses are deleted at random one finds

$$\overline{U_{ij...l}} = \prod_r \bar{\omega}_r.$$

If, moreover, all the probabilities $\bar{\omega}_r$ are the same, $\bar{\omega}_r = \bar{\omega}$,

$$\bar{v}_i = \sum_{r=2}^{p} \gamma^{r-2}(\bar{\omega}N)^{r-1}.$$

In binary connected network one observes that the probability distribution of

$$Z_i^{\alpha\beta} = \sigma_i^\alpha \sigma_i^\beta \sum_j U_{ij} \sigma_j^\alpha \sigma_j^\beta$$

once the various links realizations have been averaged out, is the same as that of

$$Z_i^{\alpha\beta\prime} = (\overline{U_{ij}^2})^{1/2} \sum_j \sigma_j^\alpha \sigma_j^\beta = \bar{\omega}^{1/2} \sum_j \sigma_j^\alpha \sigma_j^\beta.$$

In multiconnected networks

$$Z_i^{\alpha\beta} = \sigma_i^\alpha \sigma_i^\beta \left(\sum_j U_{ij} \sigma_j^\alpha \sigma_j^\beta + \gamma \sum_{jk} U_{ijk} \sigma_j^\alpha \sigma_j^\beta \sigma_k^\alpha \sigma_k^\beta + ... \right)$$

it will be assumed that a corresponding decoupling can be carried out i.e., the mean square value of $Z_i^{\alpha\beta}$ is identical to that of

$$Z_i^{\alpha\beta\prime} = \bar{\omega}^{1/2} \sum_j \sigma_j^\alpha \sigma_j^\beta + \gamma\bar{\omega} \sum_{jk} \sigma_j^\alpha \sigma_j^\beta \sigma_k^\alpha \sigma_k^\beta + ...$$

Then all the subsequent analysis simply amounts for replacing N by $\bar{\omega}N$ (except in the LnN term) in all formulae.

The same calculations can also be performed for a network characterized by a complete binary graph of connections and an uncomplete ternary graph of connections. These constraints intend to mimic the situation encountered in microcolumns in which, on the one hand, clusters of ternary synapses would be associated to every binary synapse and, on the other hand, the neurons are so tightly linked that the graph of binary interactions can be supposed to be complete. One finds

$$M \simeq 1 + \frac{(1 + \gamma\bar{\omega}N)^2 N}{(1 + 2^{5/2}\pi^{-1/2}\gamma(\bar{\omega}N)^{1/2} + 3\gamma^2\bar{\omega}N)} \cdot \frac{1}{K(N, P^*)}.$$

If, moreover, one assumes that a fixed number z of ternary junctions is attached to every binary synapse, the formula becomes ($\bar{\omega} = z/N$):

$$M \sim 1 + \frac{(1 + \gamma z)^2}{1 + 2^{5/2}\pi^{-1/2}\gamma z^{1/2} + 3\gamma^2 z} \cdot \frac{1}{K(N, P^*)}.$$

In the limit $\gamma \gg 1/z$

$$M \sim \frac{zN}{3K(N, P^*)}. \qquad (39)$$

4.3 Comment on the Hebbian Rules

All calculations in the preceeding sections have assumed that the effective interactions $J_{ij...l}$ were given

the Hebbian form of (11). But the Hebbian rules actually determine the real connections $c_{ij...l}$ rather than the effectives ones $J_{ij...l}$. It is therefore more natural to set

$$C_{ij...l} = \gamma^{q-2} \sum_{\alpha=1}^{M} \sigma_i^\alpha \sigma_j^\alpha ... \sigma_l^\alpha$$

for a q-ary synaptic chain.

Then, according to (8),

$$J_{\underset{p\,\text{terms}}{ij...k}} = \sum_{q=p}^{q_{max}} \frac{1}{2^{q-1}} \binom{q-1}{p-1} \sum_{\{q\}} C_{ij...l} \qquad (40)$$

where $\{q\}$ describes all sets of q sites. The study of memory capacities using (40) is not easily tractable. It will be limited to networks with a maximum synaptic order $q_{max} = 3$.

The synaptic efficacies are given by

$$J_{ij} = \frac{1}{2} \sum_\alpha \sigma_i^\alpha \sigma_j^\alpha \left(1 + \gamma \sum_k \sigma_k^\alpha\right)$$

and

$$J_{ijk} = \frac{\gamma}{4} \sum_\alpha \sigma_i^\alpha \sigma_j^\alpha \sigma_k^\alpha .$$

The coherent term is a stochastic variable with an average value

$$\bar{v}_i = \bar{v} = \frac{N}{2}\left(1 + \frac{N\gamma}{2}\right)$$

and a mean square deviation

$$\langle \Delta v^2 \rangle = \frac{\gamma^2}{4} N^3 .$$

The mean square deviation of the incoherent term ϱ_i^α

$$\varrho_i^\alpha = \frac{1}{2} \sum_{\beta \neq \alpha} \left(\sum_j \sigma_i^\alpha \sigma_j^\alpha \sigma_i^\beta \sigma_j^\beta \right.$$
$$\left. + \gamma \sum_{jk} \sigma_i^\alpha \sigma_j^\alpha \sigma_i^\beta \sigma_j^\beta \sigma_k^\beta \left(1 + \frac{\sigma_k^\alpha}{2}\right)\right)$$

is

$$\langle \Delta \varrho^2 \rangle = \frac{M}{4}\left(N + \frac{5}{4}\gamma^2 N^2\right) .$$

The approximate memory storage capacity, given by the solution of

$$\bar{v}^2 = \langle \Delta v^2 \rangle + \langle \Delta \varrho^2 \rangle$$

is

$$M \sim N \frac{\left(1 + \dfrac{\gamma N}{2}\right)^2 - \gamma^2 N}{1 + \dfrac{5}{4}\gamma^2 N} .$$

For large γ's $M \sim N^2/5$ whereas the formula obtained with parameters C instead of J would have been $M \sim N^2$.

The conclusion is that the use of Hebbian interactions instead of optimal interactions does not change the power laws of memory storage capacities but it reduces significantly the number of stored bits per synapse, the higher the maximal synaptic order the larger the reduction factor.

5 Partitions of Neural Networks

The memory storage capacity of neural networks can be considered as low. For example $K(N \sim 10^2, p = 0.99) = 20$. But it can be improved by dividing the system into smaller subsets or moduli. Let k be the number of moduli. Each modulus comprises

$$S = N/k$$

neurons. A stored pattern is now an association of k memorized sub-patterns, one for each of the moduli. Let M^* be the maximum number of these patterns. M^* is given by:

$$M^* = (M(S))^k$$

where $M(S)$ is the memory storage capacity of a S neurons network.

The optimal partition is given by

$$\frac{dM^*}{dS} = 0 = \frac{d(M(S))^{N/S}}{dS}$$
$$= M^* \frac{N}{S}\left(-\frac{1}{S} \ln M + \frac{M'(S)}{M(S)}\right)$$

or

$$M(S) \ln M(S) = S M'(S) . \qquad (41)$$

It is worth noting that the optimal size S of moduli does not depend on the size N of the network but only on the scaling law $M(S)$ of the memory storage capacity. The capacities are approximately given by

$$M(S) \simeq \frac{S^{p-1}}{K_p} . \qquad (42)$$

With (42) in (41) the moduli size is given by:

$$S = e K_p^{1/p-1}$$

where $e = 2.718...$
For binary neurons $S \simeq 50$ if the patterns have to retrieved with 1% errors at most. The moduli therefore comprise about 50 (pyramidal) neurons. The overall capacity of the network is

$$M^* = 2^{\mu N}$$

with

$$\mu = \frac{p-1}{eK_p^{1/p-1}} \, \text{Log}_2 e \, .$$

In the above example $\mu \sim 3 \times 10^{-2}$ which is to be compared with $\mu = 1$ corresponding to the total number of states (2^N).

This simple calculation however is not very satisfactory because it assumes that all combinations of memorized sub-patterns are allowed and are given a significance. This means, from the point of view of information theory, that every modulus is able to produce a maximum amount of informations of its own, independently on the environment. This is certainly not so. The moduli are under the control of higher structures which restrict their combinatorial possibilities. Only those patterns which satisfy the constraints prescribed by the higher structures are given a meaning. This certainly lowers the storage capacity of the system but a compromise between the combinatorial possibilities and the constraints leaves room for an enormous memory capacity, much larger and versatile than that associated with a non-partitionned system. These qualitative considerations naturally lead to a hierarchical structure for neural networks. It is interesting to observe that the first three levels of organization seem indeed to exist in cortical structure. The lower level would be that of micro-columns in which about 50 to 100 neurons are packed (Hubel et al. 1978; Eccles 1981; Szentagothai 1975, 1983). Bunches of about 50 to 100 microcolumns form columns and a column is linked to approximately 100, not necessarily close by, other columns.

6 Discussion and Conclusions

Two sorts of questions have been adressed in this article. The first deals with the calculation of memory storage capacities of well characterized networks i.e., networks with a given topology of connections, known synaptic efficacies and well defined dynamics. The other, which relies upon the answers to the former, is the determination of the best neuronal organization, i.e., organization which ensures a maximum memory storage capacity for a given set of biological constraints.

The central result of the article is that the maximum number of stored bits is approximately proportionnal to the number of synapses whatever the topology of the network. In particular this conclusion holds for networks comprising synapses of any order such as the axono-axonal contacts of the type studied by Kandel. Also a deletion of a percent of synapses reduces the capacity by the same amount.

There are however important limitations:

The proportionality between the number of stored bits and the overall number of synapses is valid only for large enough relative efficacy γ, a parameter which settles the magnitude ratios between synapses of different orders. γ must satisfy

$$\gamma > 1/\sqrt{N}$$

where N is the number of neurons. This condition is applicable to fully connected networks. It must be replaced by $\gamma > 1/\sqrt{\bar{\omega}N}$ in non-fully connected systems. $\bar{\omega}$ is the average probability for a synapse to exist.

In the limit $\gamma \lesssim \frac{1}{N}$, $\left(\text{resp.} \ \frac{1}{\bar{\omega}N} \right)$ the memory capacity is determined chiefly by the binary connections.

Between these two limits no analytical results have been derived except for the interesting case where binary and ternary connections, and only these connections, coexist (see (32)). The found formula could account for the memory storage capacity of cortical microcolumns.

It has been shown that the memory storage capacity is severely spoiled by the possible blurring of input patterns: In binary connected network the capacity varies as the fourth power of the probability for the input bits to be correct.

It has also been found that the proportionality factor between the number of stored bits and the number of synapses is a decreasing function of the synaptic order. On the other hand using genuine Hebbian rules, i.e. synaptic efficacies proportional to the correlated activities of system being trained in place of optimal efficacies (see Sect. 2.3), acts exactly in the same way. This rises questions related to the second topic, that of the best organization of neural networks. The answer is not easy because it relies upon the ill-defined notion of biological cost (or selective advantage). If the biological cost of a neuron is very high the system must strive to maximize the number of stored bits per neuron and therefore to develop connections of the highest possible orders. But if the cost of synapses is large this is the number of stored bit per synapse which must be as large as possible. As this number decreases for increasing synaptic orders the best neural network will be binary connected. The reality must lie somewhere inbetween these two extreme cases. A predominance of synapses of low orders means that the cost of synapses is relatively high as compared to the cost of neurons.

A related question is that of topological organization. It has been argued that the combinatorial possibilities of neural network are greatly enhanced

by dividing the system in smaller moduli. It has been shown that the optimal size of moduli does not depend on the overall size of the network. This size must therefore be a universal constant for all species. It depends only on the scaling behavior of the memory storage capacities. For binary networks and for a degree of confidence for stored patterns to be retrieved of 99%, it has been found that this size is close to 50 neurons which can be compared with the size of microcolumns.

Finally it has been stressed that the combinatorial possibilities explode exponentially with N, the size of the network. The set of these possibilities can be strictly adapted to the responses to environmental constraints in primitive organisms. But if, by chance, evolution induces a significant increase of the number N of neuron say in the neocortex, the number of possibilites becomes extremely large, much larger in particular than that would be strictly needed by survival conditions.

Appendix A1

Number of Orthogonal Binary Patterns

According to (24)

$$Z_i^{\alpha\beta} \alpha I^{\alpha\beta}$$

and therefore if $I^{\alpha\beta} = 0$, $Z_i^{\alpha\beta} = 0$.

If all stored patterns are orthogonal $I^{\alpha\beta} = 0$; $\forall \alpha, \beta$ and the incoherent term ϱ_i^α is zero. The memory storage capacity in this case is only limited by the maximum number M_m of patterns of N bits which can be made orthogonal. This number is N. Proof: Let consider M patterns I^α i.e.

$$I^\alpha = \{\sigma_i^\alpha\}; \ \sigma_i^\alpha = \pm 1.$$

The componants σ_i^α of I^α can be considered as special values of componants of vectors I^α, taken in R, in a N-dimensional vector space V. One knows that the maximum number of orthogonal vectors one can build in V is N. Therefore

$$M_m \leq N.$$

On the other hand it is possible to build N orthogonal sets of N binary states. As the construction eventually misses other possibilities one has

$$M_m \geq N$$

and therefore $M_m = N$.

The construction is as follows:

To one state $+$ one associates two states $+ +$ and $+ -$ of length 2. These two states are orthogonal.

$$+ \rightarrow \begin{cases} + + \\ + - \end{cases}$$

Each of the two states is considered as one new state $\pm$ on which the same production rule is applied. For example

$$+ + (\equiv \pm) \rightarrow \begin{cases} \pm \pm \ (\equiv + + + +) \\ \pm \approx \ (\equiv + + - -) \end{cases}$$

and

$$+ - (\equiv \pm) \rightarrow \begin{cases} \pm \pm \ (\equiv + - + -) \\ \pm \approx \ (\equiv + - - +) \end{cases}.$$

The orthogonality holds for the state $\pm$ as it holds for state $+$: the four states are orthogonal. The process can be pursued

$$+ \rightarrow + + (\equiv \pm) \rightarrow \pm \pm (\equiv \pm) \rightarrow \pm \pm (\equiv \pm) \rightarrow \dots$$

It generates after τ steps 2^τ orthogonal states of lengths $N = 2^\tau$.

One concludes that N orthogonal patterns are faithfully retrieved. One extra pattern cannot be made orthogonal to the N first. The best which can be achieved is to choose it such as $I^{N+1} \cdot I^\alpha \sim \sqrt{N}$. Then $\langle (\varrho_i^{N+1})^2 \rangle^{1/2} \simeq N$ which is of the same order as the coherent part $v_i^{N+1} : N$ is a strict upper limit of memory storage capacities.

Appendix A2

Derivation of (31)

The quantity to be calculated is

$$\langle Z^2 \rangle = 2 \int_0^\infty Z^2 Q(Z) dZ$$

given $Z = X + \gamma X^2$
and

$$P(X) = \frac{1}{\sqrt{2\pi N}} \exp - \frac{X^2}{2N}.$$

The relation between Z and X is inverted

$$X = \frac{-1 + \sqrt{1 + 4\gamma Z}}{2\gamma}$$

and the probability distribution $Q(Z)$ is deduced along the lines given in Sect. 3.3

$$Q(Z) = \frac{1}{\sqrt{2\pi N}} \frac{1}{\sqrt{1 + 4\gamma Z}} \exp - \frac{(-1 + \sqrt{1 + 4\gamma Z})^2}{8\gamma^2 N}.$$

A first change of variable

$$u = \sqrt{1 + 4\gamma Z}$$

leads to

$$\langle Z^2 \rangle = \frac{1}{16\gamma^3\sqrt{2\pi N}} \int_1^\infty (u^2-1)^2 \exp-\frac{(u-1)^2}{8\gamma^2 N}\, du.$$

A second one

$$v = u - 1$$

and a last one

$$t = \frac{v^2}{8\gamma^2 N}$$

to

$$\langle Z^2 \rangle = \frac{2N}{\sqrt{\pi}} \int_0^\infty t^{1/2}(1+2^{1/2}\gamma N^{1/2}t^{1/2})^2$$

$$\cdot \exp-t\, dt. \tag{A2.1}$$

The square of the integrand of (A2.1) is expanded. The resulting expression is a sum of Euler gamma functions

$$\langle Z^2 \rangle = \frac{2N}{\sqrt{\pi}} (\Gamma(3/2)+2^{3/2}\gamma N^{1/2}\Gamma(2)+2\gamma^2 N\Gamma(5/2)).$$

Using

$$\Gamma(3/2) = \frac{\sqrt{\pi}}{2};\ \Gamma(2)=1;\ \Gamma(5/2)=\frac{3\sqrt{\pi}}{4}$$

the Eq. (31) recovered.

References

Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD (1983) Molecular biology of the cell. Garland Publishing Co, New York

Carew TJ, Castellucci VF, Kandel ER (1971) An analysis of dishabituation and sensitization of the gill-withdrawal reflex in Aplysia. Int J Neurosci 2:79–98

Changeux JP, Heidmann T (1986) Allosteric receptors and molecular models of learning. In: Edelman G, Gall WE, Cowman WSM (eds.) New insights into synaptic function. Wiley, New York (in press)

Delacour J (1981) Conditionnement et biologie. Masson, Paris

Eccles JC (1981) The modular operation of the cerebral neocortex considered as the material basis of neural events. Neurosciences 6:1839–1856

Hebb DO (1949) The organization of behaviour. Wiley, New York

Hopfield JJ (1982) Neural networks and physical system with emergent collective computational abilities. Proc Natl Acad Sci USA 79:2554–2558

Hopfield JJ, Computation in simple collective "neural" systems. To appear in proceedings of Les Houches Winter session (26 February–8 March 1985)

Hubel D, Wiesel TN, Stryker MP (1978) Anatomical demonstration of orientation columns in macaque monkey. J Comparative Neurol 177:361–379

Kandel ER (1976) Cellular basis of behavior. Freeman, San Francisco

Kandel ER, Tauc L (1965) Heterosynaptic facilitation in neuron of the abdominal ganglion of Aplysia depilans. J Physiol 181:1–27

Kohonen T (1978) Associative memory. A system theoretical approach. Springer, Berlin Heidelberg New York

Kohonen T (1984) Self-organization and associative memory. Springer, Berlin Heidelberg New York Tokyo

Little WA, Shaw GL (1978) Analytic study of the memory storage capacity of neural networks. Math Biosci 39:281–290

Lynch G, Baudry M (1984) The biochemistry of memory: a new and specific hypothesis. Science 224:1057–1063

McCulloch WW, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 5:115–133

Palm G (1981) On the storage capacity of an associative memory with randomly distributed elements. Biol Cybern 39:125–135

Palm G (1982) Neural assemblies: an alternative approach of artificial intelligence: Springer, Berlin Heidelberg New York

Peretto P (1984) Collective properties of neural networks. A statistical physics approach. Biol Cybern 50:51–62

Peretto P, Niez JJ (1986) Stochastic dynamics of neural networks. IEEE Trans Syst Man Cabern 16:1

Personnaz L, Guyon I, Dreyfus G (1985) Information storage and retrieval in spinglass like neural networks. J Physique Lett 46:359–365

Roney KJ, Scheibel AB, Shaw GL (1979) Dendritic bundles: survey of anatomical experiments and physiological theories. Brain Res Rev 1:225–271

Shaw GL, Harth E, Scheibel AB (1982) Cooperativity in brain function: assemblies of approximately 30 neurons. Exp Neurol 77:324–3587

Stent GS (1973) A physiological mechanism for Hebb's postulate of learning. Proc Natl Acad Sci USA 70:997–1001

Szentagothai J (1975) The "module concept" in cerebral corterx architecture. Brain Res 95:475–496

Szentagothai J (1983) The modular architectonic principle of neural centers. Rev Physiol Biochem Pharmacol 98:11–61

Thomson RS, Gibson WG (1981) Neural model with probabilistic behavior II. One and two-neuron networks. Math Biosci 56:255–285

Weisbuch G, Fogelman-Soulié F (1985) Scaling laws for the attractors of Hopfield networks. J Phys Lett 46:623–630

Woody CD (1982) Memory learning and higher function: a cellular view. Springer, Berlin Heidelberg New York

Dr. P. Peretto
Centre d'Etudes Nucléaires de Grenoble
DRF/PSC
BP 85 X
F-38041 Grenoble
France