Hebbian Learning Reconsidered:
Representation of Static and Dynamic Objects in Associative Neural Nets

A. Herz, B. Sulzer, R. Kühn, and J. L. van Hemmen

Sonderforschungsbereich 123 an der Universität Heidelberg, Im Neuenheimer Feld 294, D-6900 Heidelberg,
Federal Republic of Germany

Abstract. According to Hebb's postulate for learning, information presented to a neural net during a learning session is stored in the synaptic efficacies. Long-term potentiation occurs only if the postsynaptic neuron becomes active in a time window set up by the presynaptic one. We carefully interpret and mathematically implement the Hebb rule so as to handle both stationary and dynamic objects such as single patterns and cycles. Since the natural dynamics contains a rather broad distribution of delays, the key idea is to incorporate these delays in the learning session. As theory and numerical simulation show, the resulting procedure is surprisingly robust and faithful. It also turns out that pure Hebbian learning is by selection: the network produces synaptic representations that are selected according to their resonance with the input percepts.

1 Introduction

According to Hebb's neurophysiological postulate for learning (Hebb 1949), information presented to a neural net during a learning session is stored in the synapses. In a more modern formulation, postsynaptic spiking enables enhancement in just those synapses that are eligible to change by virtue of a time window set by presynaptic activity. Though a direct experimental demonstration that Hebbian synapses exist has been reported only recently (Kelso et al. 1986; Malinow and Miller 1986; Gustafsson et al. 1987), the Hebb rule has become a mainstay of many considerations concerning synaptic modifications through learning.

There have been early attempts to incorporate signal delays in the dynamical description of neural networks (Caianiello 1961; Grossberg 1968; Fukushima 1973). However, except for the recent prescription of Kleinfeld (1986) and Sompolinsky and Kanter (1986), who use a single transmission delay to construct a sequence generator, spin-glass models of neural networks have usually not taken delays into account (Little 1974; Little and Shaw 1978; Hopfield 1982; Amit et al. 1985, 1987; Amit 1987; Toulouse et al. 1986; van Hemmen and Kühn 1986; van Hemmen 1987).

Here we show how Hebb's principle, if interpreted in a careful way, can be implemented in a neural network with a *broad* distribution of delays so as to handle both static objects (single patterns) and temporal associations such as tunes and rhythms. That is, after the network has been taught both single, stationary, patterns and temporal sequences – of course, by the *same* principle! – it will reproduce them when triggered suitably.

2 Model

The key idea is that signal *delays*, omnipresent in the brain (Braitenberg 1967, 1986; Miller 1987) and well known for their stabilization of temporal sequences (Kleinfeld 1986; Sompolinsky and Kanter 1986; Amit 1987; Gutfreund and Mezard 1988; Riedel et al. 1988), have to be taken into account *during the learning phase* as well. This kind of Hebbian learning presupposes a broad distribution of delays (Braitenberg 1967, 1986; Lee et al. 1986; Miller 1987) and is therefore extremely robust. Neither does the performance of the network depend significantly on the characteristics of the distribution nor is it deteriorated by imprecise initial conditions such as noisy patterns or the theme BHCH instead of BACH.

In the present study we first concentrate on the underlying mechanism. Then we derive deterministic equations describing the macroscopic order parameters corresponding to the stored patterns, the so-called overlaps, and analyze the results of numerical simulations. Finally, we indicate an interesting appli-

cation to real-time expert systems, and present a discussion. Our analysis shows that pure Hebbian learning functions through *selection*, as advocated by Toulouse et al. (1986) and Dehaene et al. (1987).

Throughout what follows, the neurons are taken to be formal (McCulloch and Pitts 1943), i.e., Ising spins $S_i = \pm 1$ where i ranges between 1 and N, the size of the network. Patterns are specific Ising spin configurations $\{\xi_i^\mu; 1 \leq i \leq N\}$, labeled by $1 \leq \mu \leq q$. For the sake of convenience we work with unbiased random patterns where $\xi_i^\mu = \pm 1$ with equal probability. Though this choice is biologically somewhat implausible because of the high level of activity which it implies, we stick to it for now. There are two reasons for doing so. First, our choice simplifies the theoretical work and hence helps us in elucidating as clearly as possible the *mechanism* of learning with delays, the key issue of the present contribution. Second, it facilitates comparison with previous work on sequence generation through a delay mechanism (Kleinfeld 1986; Sompolinsky and Kanter 1986). We will indicate how low-activity patterns can be treated at the end of this paper, in Sect. 6.

For the moment we focus our attention on two specific neurons, say i and j, with J_{ij} as the synaptic efficacy for the information transport from j to i. If neuron j has fired, a soliton-like pulse propagates via the axon to the synapse at neuron i and it is here, at the synapse, where the information is stored, provided neuron i's activity is concurrent with or slightly after the arrival of the pulse coming from j (Kelso et al. 1986; Malinow and Miller 1986).

In a thick myelinated axon the transport is fast, taking at most a few milliseconds. However, a considerable portion of the axons – estimates range from 20 to 40% – are very fine myelinated fibers or unmyelinated ones. Though mostly neglected, these axons play quite an important role since they give rise to long delays, up to 100–200 ms (Lee et al. 1986; Miller 1987). In addition, there may be (post-)synaptic delays, so that long-term potentiation even occurs if the signal from j has arrived some time *before* i becomes active (Gustafsson et al. 1987). All this gives a rather broad time window for the time delays involved in long-term potentiation. Adding both types of delay we then find: At time t the signal $S_j(t - \tau)$ has to be paired with $S_i(t)$, the state neuron i is in; the delay τ may be of the order of hundreds of milliseconds.

A general delay mechanism can be imagined as a filter which transforms a family of signals, the "past" $\{S_j(s), s \leq t\}$, into

$$\bar{S}_j(t) = \int_0^\infty ds\, w(s) S_j(t - s).\tag{1}$$

Here $w(s) \geq 0$ is a memory kernel, which is normalized to one. A typical example is δ-function delay where

$w(s) = \delta(s - \tau)$. Since to excellent approximation a synapse itself is a low-pass filter (Scott 1977), the associated delay is exponential: $w(s) = \tau^{-1} \exp(-s/\tau)$ with $\tau < 20$ ms, a rather narrow time window. Combining axonal and synaptic delays we get a signal of the form $\bar{S}_j(t)$, which is concurrent with $S_i(t)$.

During a *learning session*, the system is offered a finite sequence of patterns $\{\xi_i^{v(t)}; 1 \leq i \leq N\}$ where $v(t)$ is a given function of t. According to Hebb's principle (Hebb 1949), at each instant of time the synaptic efficacy J_{ij} increases or decreases by an infinitesimal amount proportional to $S_i(t)\bar{S}_j(t)$. Let us suppose that the learning session lasts T milliseconds and, for the sake of simplicity, that we start with a *tabula rasa* ($J_{ij} = 0$). Then we obtain, adding the partial increments,

$$J_{ij}(\tau) = N^{-1}\varepsilon_{ij}(\tau) \cdot \frac{1}{T} \int_0^T dt\, S_i(t) S_j(t - \tau)\tag{2}$$

for simple δ-function delay and the very same expression with $S_j(t - \tau)$ replaced by $\bar{S}_j(t)$ for more general delays. In the integrand, $S_i(t) = \xi_i^{v(t)}$. The N^{-1} is a scaling factor which will turn out to be useful for the theoretical analysis of the fully connected systems considered in the present paper; see Sect. 4. The $\varepsilon_{ij}(\tau)$ is a weight of the synapse $J_{ij}(\tau)$ which takes morphological characteristics of the connection $j \to i$ with delay τ into account. In what follows, we shall restrict ourselves to the homogeneous case, $\varepsilon_{ij}(\tau) = \varepsilon(\tau)$. In passing we note that the increments in (2) are linear (new data are just added) and that by imposing upper and lower bounds (Nadal et al. 1986; Parisi 1986; van Hemmen et al. 1988b) or by allowing leakage one can induce forgetfulness. We will not pursue this issue here, however. It is also worthwhile to realize that the prescription (2) in conjunction with a *distribution* of delays τ strongly deviates from all presently known cycle generators (Willwacher 1976, 1982; Buhmann 1985; Kleinfeld 1986; Sompolinsky and Kanter 1986; Buhmann and Schulten 1987; Dehaene et al. 1987).

If the pattern shown to the network is stationary, i.e., $S_i(t) = \xi_i^\mu$ for some fixed μ and $0 \leq t \leq T$ while T exceeds all the τ's of the system (a natural consequence of the pattern being stationary), then the J_{ij} become symmetric and, despite the delays, the performance of the network turns out to be that of the Hopfield model. Plainly, in general the J_{ij} are asymmetric.

The Hebb rule (2) gives rise to a *retrieval* mechanism which is largely (but not completely) independent of the distribution of the delays τ. To illustrate this statement we study three cases.

(A) For each pair (i, j) there is a *large* number of axons (possibly, interneurons) whose delays τ have a distribution independent of i and j. Summing over the incoming signals we obtain the postsynaptic potential

(PSP)

$$h_i(t) = \sum_{j(\neq i)} \sum_\tau J_{ij}(\tau) S_j(t-\tau). \tag{3a}$$

The weights $\varepsilon(\tau)$ are chosen according to a given distribution of delays τ. Of course, one can also imagine that only a *single* axon links each pair of neurons. This idea has led us to two further models.

(B) In addition to the axonal delay, which is assumed to depend on j only, we have a synaptic (exponential) delay at i. Then τ_{ij} is to be split into τ_i and τ_j so that

$$h_i(t) = \sum_{j(\neq i)} J_{ij}(\tau_i, \tau_j) \bar{S}_j(t), \tag{3b}$$

where $\bar{S}_j(t)$ incorporates the exponential and δ-function delay with times τ_i and τ_j, respectively. The τ_i and τ_j are drawn from given distributions and $\varepsilon(\tau) = 1$.

(C) For each pair (i, j) there exists a single axon with delay τ_{ij} which is sampled from a given distribution *independent* of i and j. Then we find

$$h_i(t) = \sum_{j(\neq i)} J_{ij}(\tau_{ij}) S_j(t-\tau_{ij}). \tag{3c}$$

Again, $\varepsilon(\tau) = 1$.

The models A and B are random-site problems which are analytically soluble in terms of so-called sublattice magnetizations (van Hemmen et al. 1986; van Hemmen and Kühn 1986); see Sect. 4. Model C, however, is a random-bond problem and no exact solution is known yet.

The simplest possible dynamics $S_i(t + \Delta t) = \text{sgn}[h_i(t)]$ is deterministic. Updating is either performed in parallel for *all* spins (Little 1974; Little and Shaw 1978), or sequentially in that only a *single* spin may be modified at each step Δt, which is adapted accordingly (Hopfield 1982). Noise can be taken care of by a stochastic dynamics of the Glauber type at a finite inverse temperature β. In that case, the probability of getting S_i at time $t + \Delta t$ equals $(1/2)\{1 + \tanh[\beta h_i(t) S_i]\}$.

3 Learning

As we already noted above, the problem of learning and recalling a temporal pattern sequence may serve as a paradigm to illustrate the most salient features of the Hebbian learning rule (2). For the sake of definiteness we consider a cycle of q unbiased random patterns $\{\xi_i^\mu; 1 \leq i \leq N\}$ with $1 \leq \mu \leq q$, each of duration Δ. During the learning session, which lasts T msec, we then impose upon the network a *time-dependent* stimulus $S_i(t)$, $0 \leq t \leq T$, of the form

$$S_i(t) = \xi_i^{\nu(t)}, \qquad \nu(t) = \mu \bmod q \quad \text{for} \quad (\mu-1)\Delta \leq t < \mu\Delta. \tag{4}$$

Since the cycle consists of q patterns and may be repeated $(T \geq q\Delta)$, μ in (4) is to be taken modulo q so that $1 \leq \nu(t) \leq q$.

Our first task is now to compute the synaptic efficacies $J_{ij}(\tau)$, assuming that they are generated according to the Hebb rule (2) in response to the external stimulus (4). The outcome will, of course, depend on the ratio τ/Δ and on the *delay mechanism*, i.e., on the form of the memory kernel w in (1). We shall consider three different cases, δ-function delay, exponential delay and a combination of the two.

To compute $J_{ij}(\tau)$ for δ-function delay, it is advantageous to split τ into integer and fractional parts of Δ,

$$\tau = (n_\tau + d_\tau)\Delta, \qquad n_\tau \in \mathbb{N} \cup \{0\}, \qquad 0 \leq d_\tau < 1. \tag{5}$$

If at time t the system is exposed to pattern μ so that $t = (\mu - 1 + s)\Delta$ for some $0 \leq s < 1$, then the pattern presented to the network at time $t - \tau$ was $(\mu - n_\tau - 1) \bmod q$ or $(\mu - n_\tau) \bmod q$ for $0 \leq s < d_\tau$ and $d_\tau \leq s < 1$, respectively. The contribution of pattern μ to $J_{ij}(\tau)$ therefore consists of two parts, viz., $\xi_i^\mu \xi_j^{(\mu - n_\tau - 1) \bmod q}$ and $\xi_i^\mu \xi_j^{(\mu - n_\tau) \bmod q}$, with relative weights d_τ and $1 - d_\tau$, respectively. After the learning session, which may take one or several complete sweeps through the cycle (4), the strength of the synapse connecting neuron j to neuron i is

$$J_{ij}(\tau) = \frac{\varepsilon(\tau)}{N} \left\{ (1 - d_\tau) \sum_{\mu=1}^q \xi_i^{(\mu + n_\tau) \bmod q} \xi_j^\mu \right.$$
$$\left. + d_\tau \sum_{\mu=1}^q \xi_i^{(\mu + n_\tau + 1) \bmod q} \xi_j^\mu \right\}. \tag{6}$$

This expression may be used for the models A and C. Before proceeding to the other delay mechanisms, however, we would like to point out the following.

First, apart from the a-priori weights $\varepsilon(\tau)$, (6) clearly exhibits a "resonance phenomenon": The $J_{ij}(\tau)$ with delays τ which are *integer* multiples of Δ and thus match the timing of the external stimulus are the ones that receive maximum strength (see Fig. 1). Note that they are also the ones that would support a stable cycle at exactly the same period (neglecting transition times) as that of the external stimulus (4). Thus, due to a subtle interplay between external stimulus and internal architecture (distribution of τ's), the Hebb rule (2), which *prima facie* appears to be instructive in character, has in fact also pronounced selective aspects.

Second, in a neural net with some given distribution of transmission delays τ, the external stimulus encoded in the $J_{ij}(\tau)$ will enjoy a rather multifacetted representation. For instance, the $J_{ij}(\tau)$ with $\tau/\Delta \ll 1$ or, equivalently, $n_\tau = 0$ and $d_\tau \ll 1$, will be almost symmetric and they will therefore encode the individual patterns of the cycle as *unrelated static objects*. On the other hand, synapses with transmission delays of the order of
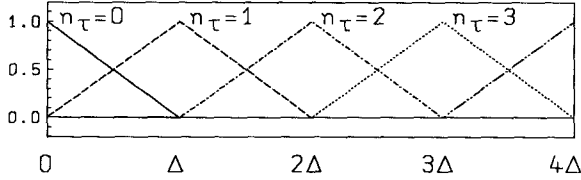
Fig. 1. Contributions to the synaptic efficacy $J_{ij}(\tau)$ as a function of the delay τ. See (6). The solid line indicates the contribution of terms of the form $\xi_i^\mu \xi_j^\mu$, corresponding to $n_\tau = 0$. The dashed line shows the contribution of the $\xi_i^{\mu+1} \xi_j^\mu$, where $n_\tau = 1$. The dashed-dotted line refers to terms of the form $\xi_i^{\mu+2} \xi_j^\mu$, where $n_\tau = 2$. And so on. Each term $\xi_i^{\mu+n_\tau} \xi_j^\mu$ aims at moving the system from state μ to state $\mu + n_\tau$. Precisely at $\tau = 0, \Delta, 2\Delta, \ldots$ we have the *maxima* of the contributions which want to shift the system $0, 1, 2, \ldots$ steps further (et cycl.) in the pattern sequence. This phenomenon exemplifies the synaptic selection principle: Synapses are strengthened according to their resonance with the input percepts

Δ will be able to detect the transitions between the patterns which occur in the expression (4). The corresponding synaptic efficacies will be asymmetric and establish various temporal relations between the patterns $\{\xi_i^\mu\}$ of the cycle, *thereby representing the complete cycle as a dynamic object*. As we will see later on, a network with this kind of synapse reproduces the external stimuli rather faithfully, for both a broad range of distributions of the transmission delays and many different initial conditions.

The evaluation of the $J_{ij}(\tau)$ for exponential delay or for a combination of exponential and δ-function delays proceeds along the same lines as above. We first compute the integrand in (2), assuming that the system is exposed to pattern μ, so that $t = (\mu - 1 + s)\Delta$ for some $0 \leq s < 1$. For exponential delay we then obtain

$$S_i(t)\bar{S}_j^\tau(t) = (1 - e^{-s\Delta/\tau})\xi_i^\mu \xi_j^\mu$$

$$+ e^{-s\Delta/\tau} \frac{(e^{\Delta/\tau} - 1)}{1 - e^{-q\Delta/\tau}} \sum_{n=1}^{q} e^{-n\Delta/\tau} \xi_i^\mu \xi_j^{(\mu-n)\bmod q}. \tag{7}$$

To finish the analysis, we have to integrate contributions from a complete learning sweep through the cycle (4) and get

$$J_{ij}(\tau) = \frac{\varepsilon(\tau)}{N} \sum_\mu \left\{ \left[1 - \frac{\tau}{\Delta}(1 - e^{-\Delta/\tau}) \right] \xi_i^\mu \xi_j^\mu \right.$$

$$\left. + 4 \frac{\tau}{\Delta} \frac{\sinh^2\left(\frac{\Delta}{2\tau}\right)}{1 - e^{-q\Delta/\tau}} \sum_{n=1}^{q} e^{-n\Delta/\tau} \xi_i^\mu \xi_j^{(\mu-n)\bmod q} \right\}. \tag{8}$$

In a similar vein, one can compute $J_{ij}(\tau', \tau)$ for a combination of exponential and δ-function delays defined by

$$\bar{S}_j^{\tau', \tau}(t) = \frac{1}{\tau'} \int_0^\infty ds\, e^{-s/\tau'} S_j(t - \tau - s). \tag{9}$$

This is model B. The final expression for $J(\tau', \tau)$ is somewhat voluminous. It is given in the Appendix.

For all three models A–C, the J_{ij} are bilinear in the ξ's and therefore may be written

$$J_{ij}(\tau) = N^{-1} \sum_{\mu, \nu} \xi_i^\mu Q_{\mu\nu}^{(\tau)} \xi_j^\nu \equiv N^{-1}(\xi_i, Q^{(\tau)}\xi_j) \tag{10}$$

for some matrix $Q^{(\tau)}$. In the case of model B, τ stands for a pair (τ', τ).

4 Retrieval

The dynamics of the retrieval process is described most conveniently in terms of the *overlaps* $(N \to \infty)$

$$m_\mu(t) = N^{-1} \sum_{i=1}^{N} \xi_i^\mu S_i(t). \tag{11}$$

If $S_i(t) = \xi_i^\mu$, i.e., the network is in state μ, then $m_\mu(t) = 1$. In general, $m_\mu(t)$ is less than 1 and measures how near the system is to state μ.

In what follows, we present an analytic treatment of the time evolution of the overlaps $m_\mu(t)$. We concentrate on one specific finite cycle and refer to the literature for the treatment of extensively many patterns outside the cycle (Riedel et al. 1988). We will be rather brief on model A, only spelling out some details so as to be able to analytically treat model B. Model C is a random-bond problem whose analytic solution is not available. Readers who are not interested in these details can directly proceed to Sect. 5.

The key idea of our analysis is the following. Given q binary random patterns $\{\xi_i^\mu\}$, $1 \leq \mu \leq q$, there are only 2^q different positions available for the random vectors $\xi_i = (\xi_i^\mu, 1 \leq \mu \leq q)$ in (10), namely, the 2^q corners $\mathbf{x}$ of the hypercube $[-1, 1]^q$. Introducing two notions borrowed from spin-glass theory (van Hemmen et al. 1986, 1988a; Grensing and Kühn 1986), *viz.* the sublattices $I(\mathbf{x}) = \{i; \xi_i = \mathbf{x}\}$ and the corresponding sublattice magnetizations

$$m(\mathbf{x}; t) = |I(\mathbf{x})|^{-1} \sum_{i \in I(\mathbf{x})} S_i(t), \tag{12}$$

one easily verifies that the PSP (3) at i is

$$h_i(t) = \sum_\mathbf{y} \sum_\tau (\xi_i, Q^{(\tau)}\mathbf{y}) p_N(\mathbf{y})\bar{m}^\tau(\mathbf{y}; t). \tag{13}$$

Here $|I(\mathbf{x})|$ denotes the size of $I(\mathbf{x})$, $p_N(\mathbf{y}) = N^{-1}|I(\mathbf{y})|$, and $\bar{m}^\tau(\mathbf{y}, t)$ is a convolution of $m(\mathbf{y}; t)$ with a memory kernel; cf. (1). As $N \to \infty$, $p_N(\mathbf{y})$ converges to $p(\mathbf{y})$ with probability one; for unbiased patterns, $p(\mathbf{y}) = 2^{-q}$. Since $\xi_i = \mathbf{x}$ for *all* i in $I(\mathbf{x})$, the PSP depends on i only through the label $\mathbf{x}$ of the sublattice which i belongs to. In the bulk limit $h_i(t)$ therefore converges to

$$h(\mathbf{x}; t) = \sum_\mathbf{y} \sum_\tau (\mathbf{x}, Q^{(\tau)}\mathbf{y}) p(\mathbf{y})\bar{m}^\tau(\mathbf{y}; t), \tag{14}$$

whatever $i \in I(\mathbf{x})$.

If parallel updating is performed after each elementary time step Δt, the *thermal* average of $S_i(t + \Delta t)$ is $\tanh[\beta h(\mathbf{x}; t)]$ for all $i \in I(\mathbf{x})$. Since $|I(\mathbf{x})| \sim p(\mathbf{x})N$ as $N \to \infty$, we may apply the strong law of large numbers (Lamperti 1966) to (12) so as to get

$$m(\mathbf{x}; t + \Delta t) = \tanh[\beta h(\mathbf{x}; t)]$$ (15)

with $h(\mathbf{x}; t)$ given by (14).

In the case of sequential dynamics, also of the Glauber type, only a single spin is updated per elementary time step Δt, which must therefore scale with the system size N as $\Delta t \propto N^{-1}$. As $N \to \infty$ we then obtain a set of 2^q coupled nonlinear differential equations (Riedel et al. 1988)

$$\dot{m}(\mathbf{x}; t) = -\Gamma\{m(\mathbf{x}; t) - \tanh[\beta h(\mathbf{x}; t)]\},$$ (16)

where $\dot{m}$ denotes differentiation of m with respect to time and $\Gamma \geq 1$ is the mean attempt rate. Because of the delays in $h(\mathbf{x}; t)$, (16) is a so-called *functional* differential equation.

Explicit solutions to (15) and (16) are not known but numerical solutions can be obtained in a rather straightforward way. The outcomes are represented most conveniently in terms of the overlaps,

$$m_\mu(t) = \sum_{\mathbf{x}} p(\mathbf{x}) x_\mu m(\mathbf{x}; t) \equiv \langle x_\mu m(\mathbf{x}; t)\rangle.$$ (17)

Since $(\mathbf{x}, \mathbf{Q}^{(\tau)}\mathbf{y})$ in (14) is *bilinear* in $\mathbf{x}$ and $\mathbf{y}$, one can combine Eqs. (14), (15), and (17) to make a set of q equations for the overlaps themselves,

$$m_\mu(t + \Delta t) = \left\langle x_\mu \tanh\left[\beta \sum_\tau \sum_{\nu,\sigma} x_\nu Q_{\nu,\sigma}^{(\tau)} \bar{m}_\sigma^\tau(t)\right]\right\rangle.$$ (18)

Equation (18) describes parallel dynamics. Sequential dynamics can be treated in the same way. Taking advantage of (14), (16), and (17), one easily derives a set of q functional differential equations for the $m_\mu(t)$. This finishes the analysis of model A. We now turn to model B.

In the case of model B, the synaptic efficacies are of the form $J_{ij}(\tau_{ij})$ where τ_{ij} stands for the pair (τ_i^s, τ_j^a) of axonal (τ_j^a) and synaptic (τ_i^s) delays. To simplify the argument below, we will assume that the distributions $p_s(\tau_i^s)$ and $p_a(\tau_j^a)$ of τ_i^s and τ_j^a are discrete and finite. This assumption is immaterial and can be dispensed with (van Hemmen et al. 1986). By construction, the Hebbian synapses of model B are of the form $J_{ij}(\tau_{ij}) = N^{-1}(\xi_i, \mathbf{Q}^{(\tau_i^s, \tau_j^a)}\xi_j)$, which is bilinear in the ξ's but *not* in the τ's. We therefore introduce sublattices

$$I(\mathbf{x}, \vartheta_s, \vartheta_a) = \{i; \xi_i = \mathbf{x}, \tau_i^s = \vartheta_s, \tau_i^a = \vartheta_a\}.$$ (19)

As before, the PSP at i then only depends on the sublattice which i belongs to and

$$h_i(t) = h(\mathbf{x}, \vartheta_s, \vartheta_a; t)$$
$$= \sum_{\mathbf{y}, \vartheta_s', \vartheta_a'} (\mathbf{x}, \mathbf{Q}^{(\vartheta_s, \vartheta_a)}\mathbf{y}) p(\mathbf{y}) p(\vartheta_s') p(\vartheta_a') \bar{m}(\mathbf{y}, \vartheta_s', \vartheta_a'; t)$$ (20)

whatever $i \in I(\mathbf{x}, \vartheta_s, \vartheta_a)$. Here $\bar{m}(\mathbf{y}, \vartheta_s', \vartheta_a'; t)$ follows from (9) by putting $\tau = \vartheta_s'$ and $\tau' = \vartheta_a'$ and replacing S_j by $m(\mathbf{y}, \vartheta_s', \vartheta_a'; \cdot)$.

We can now proceed exactly as in the case of model A and derive deterministic equations for the sublattice magnetizations with either parallel or sequential dynamics. We only have to replace $m(\mathbf{x}; t)$ and $h(\mathbf{x}; t)$ in (15) and (16) by $m(\mathbf{x}, \vartheta_s, \vartheta_a; t)$ and $h(\mathbf{x}, \vartheta_s, \vartheta_a; t)$, respectively. In contrast to model A, a reduction to a set of q equations for the overlaps $m_\mu(t)$ is *not* possible since the delays vary from sublattice to sublattice; cf. (19). Of course, the $m_\mu(t)$ still can be computed through

$$m_\mu(t) = \sum_{\mathbf{x}} p(\mathbf{x}) x_\mu \left[\sum_{\vartheta_s, \vartheta_a} p(\vartheta_s) p(\vartheta_a) m(\mathbf{x}, \vartheta_s, \vartheta_a; t)\right].$$ (21)

As we will see shortly, the dynamical behaviour does not change appreciably when the sublattices do not have a macroscopic size yet and, strictly speaking, fluctuations cannot be ignored. This surprising effect will be analyzed elsewhere, however.

5 Results

There is ample evidence (Braitenberg 1967, 1986; Lee et al. 1986; Miller 1987) in favor of the basic postulate of the present paper, *viz.*, the existence of a *wide* distribution of delays in the brain. Though the duration of these delays seems to be in the range of hundreds of milliseconds, it has not been determined yet with absolute precision and one should be aware of this proviso. In spite of that, Hebbian learning *à la* (2) provides an extremely robust procedure to store and faithfully retrieve both static and dynamic objects, which in turn supplies further, indirect, evidence in favor of the existence of Hebbian synapses. We now consider this evidence in more detail.

As shown by Figs. 2–6, the Hebb rule (2) allows a faithful representation of a cycle which we have taught the network. It is satisfying, even remarkable, that the retrieval quality hardly depends on the specific model (Fig. 2). The Hebbian performance, though, exhibits a dichotomy: Either the job is done so well that the dependence upon the distribution of the delays, though existent, is extremely hard to discern, or the cycle does not run at all because, as a rule of thumb, the duration Δ of each pattern exceeds $\tau_{\max}$ so that the system cannot "measure" the pattern's lifetime (Fig. 3).

Thermal noise, which is measured by the inverse temperature β, gives rise to a similar dichotomic behaviour. Below a critical β_c, no stable cycle exists. On the other hand, if $\beta > \beta_c$, the performance of the system is only marginally temperature dependent. The critical β_c, however, is model dependent. For instance, with $N = 128$ and initial conditions as in Fig. 2, β_c turns out to be 8.0, 8.3, and 9.5 for models A, B, and C, respectively.
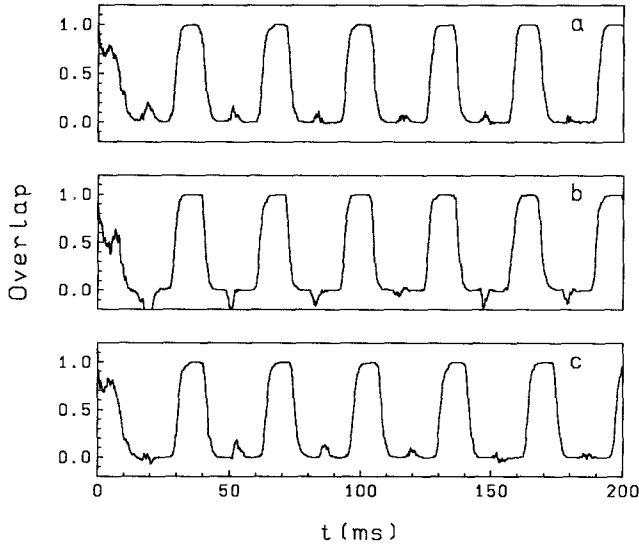
Fig. 2a–c. Model dependence. The overlap with the first pattern of a cycle consisting of three patterns (3-cycle) is shown as a function of time (t): **a** Model A, **b** Model B, **c** Model C. Though the transients are different, the long-term behaviour of the overlaps is strikingly similar. In all three cases, the number of neurons is $N = 256$. Dynamics is sequential and the inverse temperature β equals 10. The initial conditions are $m_1(t) = 1$ for $-1 \text{ ms} \leq t \leq 0$ and $m_\mu(t) = 0$ otherwise. We took a discrete distribution of axonal δ-function delays at $\tau = 0, 1, \ldots, 30$ ms. The a-priori weights are uniform and normalized so that $\sum_\tau \varepsilon(\tau) = 1$. During the learning session, each pattern lasted $\Delta = 10$ ms

Figure 4 shows that the performance of the network hardly depends on its size and that it agrees well with what is predicted by the analytic theory of the previous section.

If one teaches the system a set of patterns in, for example, a "trapezoidal" form and not as a block pulse of exact duration Δ, then the agreement between the patterns taught and those recalled is as good as in the previous figures; see Fig. 5.

Error correction is also performed by a Hebbian network. If the system has been taught the theme BACH and afterwards it is presented BHCH, then it will readily reproduce the correct tune (Fig. 6). As in the static case the system functions as a pattern recognizer except that here "pattern" is a *spatio-temporal* object, where space refers to phase space and time to temporal order in the cycle. The "pattern" presented to the network (BHCH) will be recognized provided it is close enough in space *and* time to one of the stored prototypes (in casu, BACH).

Aspects of a real-time expert system can also be realized. In practical work one quite often wants a system to react *only if* first A and then B occurs, but not if either A or B or first B and then A appears. One can
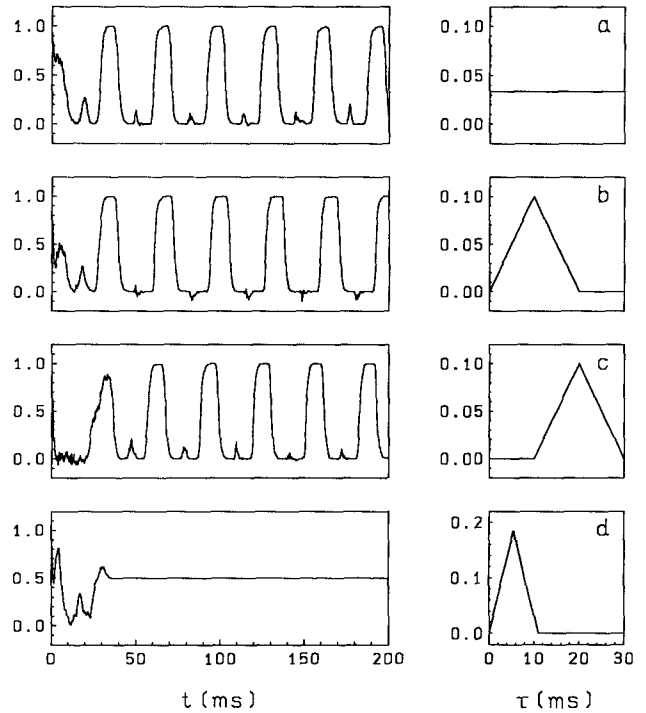
Fig. 3a–d. Performance of the network as a function of the distribution of delays for model B. Each track represents the time evolution of the overlap with the first pattern of a 3-cycle, $m_1(t)$, for a given distribution of axonal δ-function delays ($\tau_s = 0$). The delay distribution is displayed in the box to the right of the $m_1(t)$-plot to which it belongs. All the distributions are discrete, with a spacing of one ms. Note that a stable cycle is produced even in the *absence* of symmetric synapses without delay. The proviso is that, as in **a–c**, $\Delta < \tau_{max}$. If this condition is not satisfied, as in **d**, the cycle is not stable. In **a–c** variations of the τ distributions do affect the transients, but they hardly change the large-t behaviour. The system size is $N = 512$ and the dynamics is sequential with $\beta = 10$. During the learning session, each pattern lasted $\Delta = 10$ ms

teach the network the sequence $\{A, B, C\}$ and, if a threshold has been introduced (affects?), only the ordered pair (A, B) will induce the transition to C (see Fig. 7).

Once a cycle runs, a global estimate of its length is easy to obtain. During a learning session where the prescription (4) is used, with so-called block patterns, a cycle consisting of q patterns lasts $q\Delta$ ms. In the retrieval phase, the network is governed by a stochastic dynamics. At low temperatures it then takes $q(\Delta + 1)$ ms to complete a cycle since the patterns themselves last $q\Delta$ ms and each transition to a new pattern needs about one Monte Carlo step per spin (MCS). We have chosen our unit of time in such a way that $1 \text{ MCS} = 1$ ms. In passing we note that if the system has been taught a cycle in several runs and each run lasts $q\Delta_i$ ms; with Δ_i slightly fluctuating about a mean $\bar{\Delta}$, then the cycle is reproduced in $q(\bar{\Delta} + 1)$ ms. So
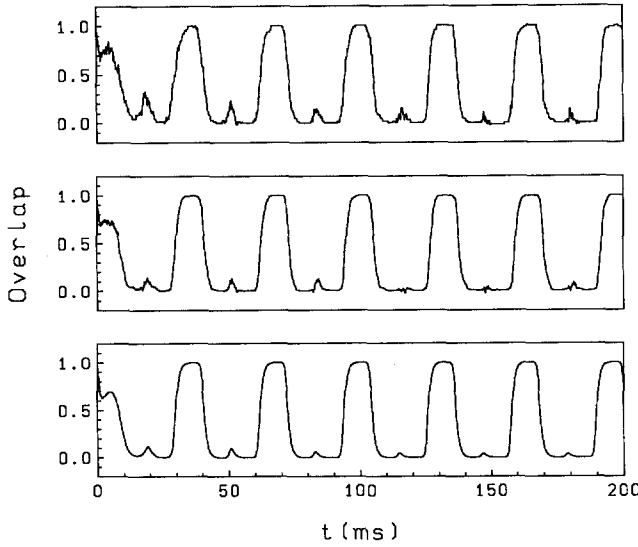
Fig. 4. System size dependence of simulations as compared to the analytic solution for the infinite system. From top to bottom, the system sizes are $N=128$, $N=512$, and $N=\infty$. The results shown are for model A. Initial conditions and other parameters are as in Fig. 2. Since 1 MCS = 1 ms (see text), $\Gamma=1$

absolute precision during the learning session is not required.

In view of Figs. 3 and 6, some comment on the usual cycle generating models (Kleinfeld 1986; Sompolinsky and Kanter 1986) seems to be in order. First, this type of model presupposes synaptic efficacies of the form $J_{ij}=J_{ij}^{(1)}+J_{ij}^{(2)}$ where $J_{ij}^{(1)}=J_{ji}^{(1)}=\sum_\mu \xi_i^\mu \xi_j^\mu$ is symmetric and $J_{ij}^{(2)}=\varepsilon \sum_\mu \xi_i^{\mu+1} \xi_j^\mu$ with $\varepsilon>1$ is to be paired with a delayed $\bar{S}_j(t)$. This interaction requires a very special tuning if it is to be obtained from the Hebb rule (2). We need model A with $\varepsilon(0)=1$, a single delay $\tau=\Delta$, and $\varepsilon(\Delta)=\varepsilon$; see Fig. 1. As to its interpretation, the following aspects may be noted.

The form of the J_{ij} suggests that *(i)* the second term "pushes" the system through an energy landscape created and stabilized by the first and *(ii)* the patterns neatly follow each other in the natural order $1,2,3,...$ et cycl. Figure 3 shows, however, that for δ-function delay the symmetric, "stabilizing" term may be *completely* absent and nevertheless a stable cycle may exist. We have also verified that, if exponential delay is dominating – which in the present situation is not very plausible – no stable cycle exists (Herz et al. 1988).

Furthermore, $J_{ij}^{(2)}$ alone does not induce the natural order. To wit, suppose that $\bar{S}_j(t)=S_j(t-\tau)$ while $\tau=4\Delta$ and that we start with the *initial* condition $v(t)=1,3,4,2$ (or BCHA) for $-\tau\leq t\leq 0$, each pattern lasting Δ ms. For $t\geq 0$ a conventional cycle generator
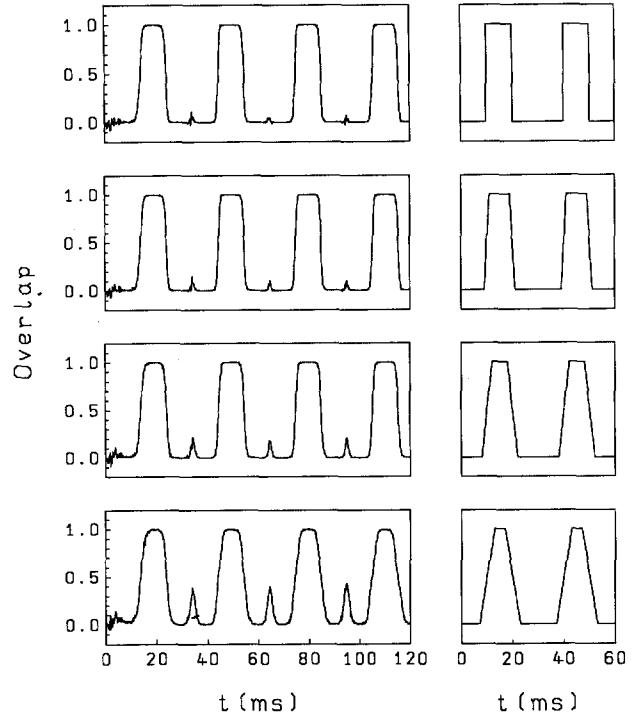

Fig. 5. Overlap with the first pattern of a 3-cycle as a function of time, for four different retrieval sessions. The time evolution of the overlap during the learning session is displayed in a box to the right of the corresponding retrieval session. The external stimulus encoded in the $J_{ij}(\tau)$ is of trapezoidal shape and each pattern of the stimulus evolves continuously out of its predecessor in such a way that $\sum_\mu m_\mu(t)\equiv 1$. The results shown are for model A. Here $N=512$ and we have *parallel* dynamics with $\Delta t=1/5$ ms at $\beta=10$. The initial conditions are $m_1(t)=1$ for -5 ms $\leq t\leq 0$ and $m_\mu(t)=0$ otherwise. We took a discrete distribution of axonal δ-function delays at $\tau=0$, $1/5$, $2/5$, ..., 30 ms. The a-priori weights are uniform and normalized so that $\sum_\tau \varepsilon(\tau)=1$

then will play 2, 4, 1, 3 and shifted variations thereof. Simply note that in this case

$$h_i(t)=\varepsilon\sum_\mu \xi_i^{\mu+1} m_\mu(t-\tau),\tag{22}$$

so that the natural order 1, 2, 3, 4 (or BACH) never appears[1]. We could also start with an initial condition where each pattern is given a duration $\Delta'\neq\Delta$. A mere δ-function delay, which gives rise to (22), would repeat Δ' forever. However, if one uses the Hebbian learning rule (2) with a broad distribution of τ's, then both the right order and the right duration Δ soon will appear provided (rule of thumb) $\Delta<\tau_{max}$. This inequality guarantees that the patterns, as seen on the time scale

---

[1] This picture does not change if in (22) $\xi_i^{\mu+1}$ is replaced by ξ_i^μ, or if the symmetric term is added and $\varepsilon>1$, as in (Kleinfeld 1986; Sompolinsky and Kanter 1986; Riedel et al. 1988)
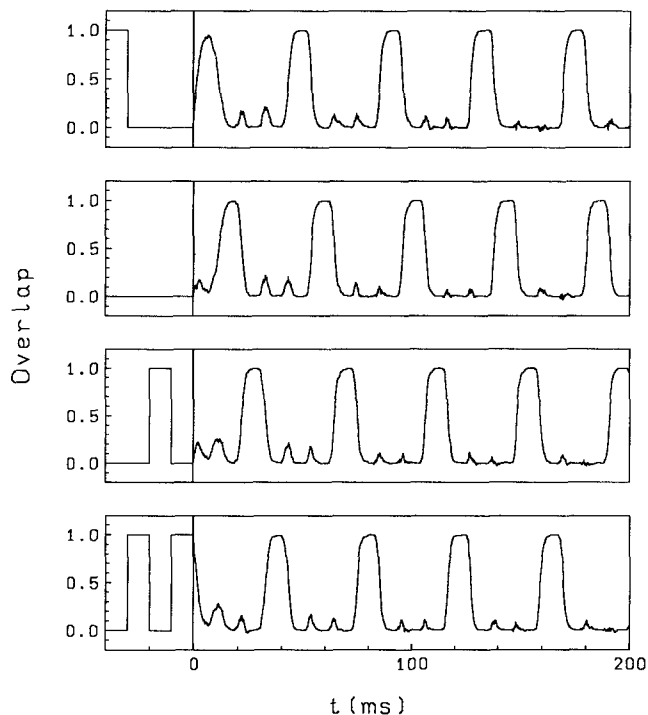
Fig. 6. Dependence upon initial conditions. The overlaps with the patterns 1, 2, 3, and 4 (from top to bottom) have been plotted as a function of time. The network has learnt the cycle 1, 2, 3, 4 (or BACH). After it has been given the faulty pattern sequence 1, 4, 3, 4 (or BHCH) as initial condition for $-\tau_{max} \leqq t \leqq 0$, the correct order BACH is spontaneously retrieved. Thus BHCH is in the *entrance domain* of BACH. The simulation result is shown for model A, with $N = 512$ and a uniform discrete distribution of δ-function delays ($\tau = 0, 1, ..., 40$ ms). Here we have sequential dynamics with $\beta = 10$ and $\Delta = 10$ ms
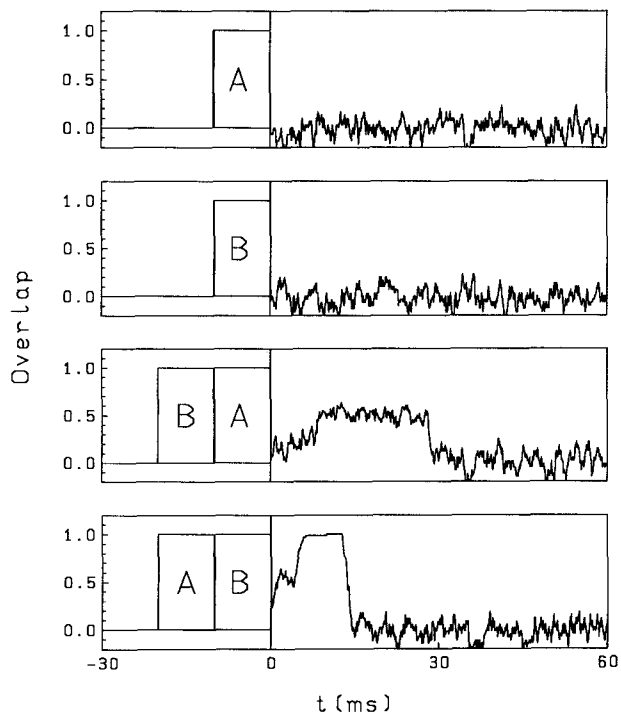
Fig. 7. The network providing a typical function of a real-time expert system. If the network has been taught the sequence $\{A, B, C\}$ (*not* a closed 3-cycle) and we have a deterministic threshold dynamics $S_i(t + \Delta t) = \text{sgn}[h_i(t)S_i(t) - \theta]$ with $\theta \approx 0.8$, then the net has learnt the rule that *only* the ordered pair (A, B) gives C. The overlap with pattern C is displayed as a function of time for $t > 0$, given different initial conditions for $t \leqq 0$. The size of the system is $N = 120$ and the model is A

of the network, have a finite duration. An illustration of the above considerations is provided by Fig. 6.

6 Discussion

In a theoretical analysis of brain functions such as learning one is bound to make approximations or assumptions. The investigations of the present study were inspired by experimental evidence (Lee et al. 1986; Miller 1987) in favour of long transmission delays, hundreds of milliseconds. We would, however, like to stress that it is the *broad* distribution of the delays, and not their exact duration, which enables a robust and faithful performance of Hebbian learning. The network will accommodate all processes consisting of patterns which do not last too long ($\Delta < \tau_{max}$) so that it can notice their duration and which do not change too often so that the delays can stabilize a new pattern ($\Gamma^{-1} < \Delta$ where Γ is the mean attempt rate).

At the moment, the very idea that the brain will use the delay mechanism to represent temporal sequences

is still somewhat speculative. For sequences which are slow in terms of neurophysiological time scales it may even be implausible. But one has to realize that *if* Hebbian syapses exist — and there is quite a bit of evidence that they do (Kelso et al. 1986; Malinow and Miller 1986; Gustafsson et al. 1987) — then they are bound to record various temporal aspects of external stimuli in a way which has been indicated in the present work.

The assumption that the ξ_i^μ are ± 1 with equal probability can be dispensed with. Let the ξ_i^μ have a mean a. Then $(\xi_i^\mu - a) (1 - a^2)^{-1/2}$ has mean zero and variance one. In the Hebb rule (2) we now replace $S_i(t)S_j(t - \tau)$ by $(S_i(t) - a) (S_j(t - \tau) - a)/(1 - a^2)$ and add a ferromagnetic background. In the case $a = 0$ we regain the situation which we have studied in this work. Moreover, for stationary patterns the modified interaction of Feigel'man and Ioffe (1987) and Bös (1988) is recovered. Finally, the above generalization has the advantage of reducing to the original Hebb rule (1949) in the low-activity limit where $a \approx -1$. Then the

increments $\Delta J_{ij}(\tau)$ are nonvanishing only if *both j and i* are active at the corresponding times t and $t-\tau$ (otherwise at least one of the factors of the integrand is about zero). Phrased differently, the representation which we have chosen in (2) is a correct version of the Hebb rule if $a=0$. The performance of the above and related schemes for nonzero a will be treated in a forthcoming contribution.

As to alternatives, one might, of course, object that receptors and afferent neurons have not been included, despite the fact that they play an essential role in transmitting the data presented to the system. Our reply is that this simplification was intentional since we do not expect, by good reason (Herz 1989), that the inclusion of an afferent system will drastically modify the network's performance. Furthermore, our aim was to reduce Hebbian learning to its bare essentials, storing data in a highly connected network with a broad distribution of delays. In Braitenberg's terms (Braitenberg 1974), we have concentrated more on the syntax, the rules which govern the succession of terms, than the dictionary, the various entries which are delivered by the afferents.

Though the degree of connectivity in different parts of the brain is in general extremely high, the relevant processing units are themselves not fully connected. Here we have studied a fully connected network but it is to be noted that dilution, though technically more complicated, can be included in a rather straightforward way (van Hemmen 1987). We will return to dilution elsewhere. It is to be emphasized, however, that the Hebb rule as embodied in (6) does allow *dynamic* disconnection of the network. That is, if during the learning session different parts of the net are exposed to uncorrelated external stimuli, then the efficacies of the synapses connecting these parts will be small by virtue of (6). The reason is simply that the $J_{ij}(\tau)$ are nothing but correlation functions.

The idea to introduce a delay synapse mechanism for the storage and recall of regular temporal sequences was evoked earlier (Kleinfeld 1986; Sompolinsky and Kanter 1986). The starting point of our work was the observation that the very same delays which appear in the retrieval dynamics (cf. (3)) should also be taken into account *during the learning phase*. Remarkably, this matching – being a natural aspect of Hebbian learning – produces synapses which in general *differ* from those introduced by Kleinfeld and by Sompolinsky and Kanter. Only in the very special case where there is just a single delay with $\tau=\Delta$ do the prescriptions agree (see e.g. Fig. 1).

Another key point of our approach is that the system learns *by selection*: If the network is taught a cycle of patterns, each with duration Δ, mainly those synapses will be strengthened whose delays match Δ, a

resonance phenomenon which is clearly exemplified by (6) and Fig. 1. On the other hand, a single stationary pattern will not discriminate between synapses with different delays τ. The above selective aspect was advocated in a general context by Toulouse et al. (1986) and Dehaene et al. (1987).

During the learning phase, one usually teaches the system one cycle or a single pattern and before going to a new object one inserts a blank. If desired, however, one may teach the network a succession of cycles and single, stationary, patterns. The idea is now that a sequence of separate entities is to be accommodated during the very same session $(0 \leq t \leq T)$. Notwithstanding this, the retrieval procedure will recognize each of these entities *as such*.

Also in other aspects, Hebbian learning is robust. This is nicely illustrated by Figs 2–7. The retrieval is faithful. Imprecise or partial initial conditions are corrected or completed and, when prepared suitably, the network can function as a real-time expert system.

Another interesting possibility of the Hebbian learning rule is the following. One might wish to generate a cycle of patterns that do not all take the same amount of time. For instance, A should appear twice as long as B and C. Here the role of preprocessing becomes apparent. As compared to B and C, pattern A saturates. Saturation effects are well-known. They are taken care of by preprocessing. We therefore suggest to use A, $\varnothing$, B, C, where $\varnothing$ is a blank during the learning session, simulating saturation. During retrieval, a set of hidden units then detects the blank and forces the system to stay in the state preceding the blank until the next state, here B, appears.

Finally, one aspect of Hebbian performance, which has been overlooked in the literature, deserves special attention. If the dynamics of a network operates with, say, δ-function delays, one has to specify not only the initial condition at time $t=0$ but also over a *whole* interval $[-\tau_{max}, 0]$. That is, the initial conditions span a Banach space $\Re$ of bounded (measurable) functions on the interval $[-\tau_{max}, 0]$. Let us denote by $\mathbf{D}_c$ the set of all functions in $\Re$ which eventually, i.e., as $t \to \infty$, give rise to the cycle c. The set $\mathbf{D}_c$ is something like a basin with c as exit. We therefore call $\mathbf{D}_c$ an *entrance domain* where the word "entrance" refers to the fact that $\mathbf{D}_c$ is in the space of initial conditions.

In a conventional context (Kleinfeld 1986; Sompolinsky and Kanter 1986; Riedel et al. 1988), only a single delay τ is present and we have seen that if a faulty melody is played during $[-\tau, 0]$, shifted variations thereof will spoil the air forever. (On the other hand, if one starts with a single tone, say 1, at $t=0$ and zero for $-\tau \leq t < 0$, an initial condition which is in $\Re$, then the system will perform $1, 2, \ldots$ et cycl. The explanation is that $J_{ij} = J_{ij}^{(1)} + J_{ij}^{(2)}$, $J_{ij}^{(1)}$ stabilizes 1 and is not per-

turbed by the past, which is zero, until $t = \tau$. Then $J_{ij}^{(2)}$ will throw the system into 2 and everything proceeds as usual.) Hebbian learning, however, also guarantees that faulty themes (say, BDCH or BAACH instead of BACH) are transformed asymptotically into their correct counterparts. So Hebbian learning allows much larger entrance domains than most conventional techniques (Kleinfeld 1986; Sompolinsky and Kanter 1986; Amit 1987; Dehaene et al. 1987; Buhmann and Schulten 1987; Gutfreund and Mézard 1988). This in turn is, in our opinion, further evidence in favor of the remarkably powerful Hebbian learning procedure.

In summary, a prerequisite for the working of any learning scheme is that the structure of the learning task is compatible with the network architecture and the learning algorithm. In the present context, the task is to store spatio-temporal objects, such as stationary patterns, sequences and cycles. The internal representability of these objects is guaranteed by a broad distribution of delays τ in conjunction with a high connectivity. The representation itself is accomplished by the Hebb rule, (6), in that the Hebbian synapses $J_{ij}(\tau)$ measure and store the correlations of the external stimulus in space (ij) *and* time (τ). The dynamics of the neural network, operating with the very same delays, is able to extract the spatio-temporal information encoded in the $J_{ij}(\tau)$. Retrieval is therefore extremely robust.

Note Added

After submittal of this paper, we became aware of related work by Kerszberg and Zippelius (1988) about synchronization of neural networks with delays, and by Coolen and Gielen (1988), who address problems of sequence learning in neural nets, albeit through a different approach.

Appendix

In the case of model B, which is a combination of an exponential delay with characteristic time τ' and a δ-function delay with characteristic time τ, the synaptic efficacy is

$$J_{ij}(\tau', \tau) = \frac{\varepsilon(\tau', \tau)}{N} \left\{ \left[d_\tau - \frac{\tau'}{\varDelta}(1 - e^{d_\tau \varDelta/\tau'}) e^{-(1 - d_\tau)\varDelta/\tau'} \right] \right.$$

$$\times \sum_\mu \zeta_i^\mu \zeta_j^{(\mu - n_\tau - 1) \bmod q}$$

$$+ \frac{\tau'}{\varDelta} \cdot \frac{(1 - e^{-d_\tau \varDelta/\tau'}) e^{-(1 - d_\tau)\varDelta/\tau'} (e^{\varDelta/\tau'} - 1)}{1 - e^{-q\varDelta/\tau'}}$$

$$\times \sum_\mu \sum_{n=1}^q e^{-n\varDelta/\tau'} \zeta_i^\mu \zeta_j^{(\mu - n_\tau - n - 1) \bmod q}$$

$$+ \left[1 - d_\tau - \frac{\tau'}{\varDelta}(e^{-d_\tau \varDelta/\tau'} - e^{-\varDelta/\tau'}) e^{d_\tau \varDelta/\tau'} \right]$$

$$\times \sum_\mu \zeta_i^\mu \zeta_j^{(\mu - n_\tau) \bmod q}$$

$$+ \frac{\tau'}{\varDelta} \cdot \frac{(e^{d_\tau \varDelta/\tau'} - e^{-\varDelta/\tau'}) e^{d_\tau \varDelta/\tau'} (e^{\varDelta/\tau'} - 1)}{1 - e^{-q\varDelta/\tau'}}$$

$$\left. \times \sum_\mu \sum_{n=1}^q e^{-n\varDelta/\tau'} \zeta_i^\mu \zeta_j^{(\mu - n_\tau - n) \bmod q} \right\}. \tag{A.1}$$

Here $\tau = (n_\tau + d_\tau)\varDelta$, as in (5). Equation (A.1) contains (6) and (8) as limiting cases for $\tau' \to 0$ and $\tau \to 0$, respectively.

References

Amit DJ (1987) Neural networks counting chimes. RIP Preprint RI/87/49

Amit DJ, Gutfreund H, Sompolinsky H (1985) Spin-glass models of neural networks. Phys Rev A 32:1007–1018

Amit DJ, Gutfreund H, Sompolinsky H (1987) Statistical mechanics of neural networks near saturation. Ann Phys (NY) 173:30–67

Bös S (1988) Neuronales Netzwerk mit hierarchisch strukturierter Information. Diplomarbeit, Universität Heidelberg

Braitenberg V (1967) Is the cerebellar cortex a biological clock in the millisecond range? In: Fox CA, Snider RS (eds) The cerebellum. Progress in Brain Research, vol 25. Elsevier, Amsterdam, pp 334–346

Braitenberg V (1974) On the representation of objects and their relation in the brain. In: Conrad M, Güttinger W, Dal Cin M (eds) Lecture Notes in Biomathematics, vol 4. Springer, Berlin Heidelberg New York, pp 290–298

Braitenberg V (1986) Two views of the cerebral cortex. In: Palm G, Aertsen A (eds) Brain theory. Springer, Berlin Heidelberg New York, pp 81–96

Buhmann J (1985) Mustererkennung in selbstorganisierenden, neuronalen Netzwerken. Diplomarbeit. Technische Universität München

Buhmann J, Schulten K (1987) Noise-driven temporal association in neural networks. Europhys Lett 4:1205–1209

Caianiello ER (1961) Outline of a theory of thought processes and thinking machines. J Theor Biol 1:204–235

Coolen ACC, Gielen CCAM (1988) Delays in neural networks. Europhys Lett 7:281–285

Dehaene S, Changeux J-P, Nadal J-P (1987) Neural networks that learn temporal sequences by selection. Proc Natl Acad Sci USA 84:2727–2731

Feigel'man MV, Ioffe LB (1987) The augmented models of associative memory: asymmetric interaction and hierarchy of patterns. Int J Mod Phys 1:51–68

Fukushima K (1973) A model of associative memory in the brain. Kybernetik 12:58–63

Grensing D, Kühn R (1986) Random-site spin glass models. J Phys A 19:L1153–1157

Grossberg S (1968) Prediction theory for some nonlinear functional differential equations I. Learning of lists. J Math Anal Appl 21:643–694

Gustafsson B, Wigström H, Abraham WS, Huang Y-Y (1987) Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. J Neurosci 7:774–780

Gutfreund H, Mézard M (1988) Processing of temporal sequences in neural networks. Phys Rev Lett 61:235–238

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

Hemmen JL van (1987) Nonlinear neural networks near saturation. Phys Rev A 36:1959–1962

Hemmen JL van, Kühn R (1986) Nonlinear neural networks. Phys Rev Lett 57:913–916

Hemmen JL van, Grensing D, Huber A, Kühn R (1986) Elementary solution of classical spin-glass models. Z Phys B-Condensed Matter 65:53–63

Hemmen JL van, Grensing D, Huber A, Kühn R (1988a) Nonlinear neural networks I. General theory; II. Information processing. J Stat Phys 50:231–257; 259–293

Hemmen JL van, Keller G, Kühn R (1988b) Forgetful memories. Europhys Lett 5:663–668

Herz A (1989) Representation and recognition of spatiotemporal objects within a generalized Hopfield scheme. Proc Connectionism in Perspective (Zürich): (to be published)

Herz A, Sulzer B, Kühn R, Hemmen JL van (1988) Hebbian learning – a canonical way of representing static and dynamic objects in an associative network. Proc nEuro '88 (Paris): (to be published)

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

Kelso SR, Ganong AH, Brown TH (1986) Hebbian synapses in hippocampus. Proc Natl Acad Sci USA 83:5326–5330

Kerszberg M, Zippelius A (1988) Synchronization in neural assemblies. Preprint

Kleinfeld D (1986) Sequential state generation by model neural networks. Proc Natl Acad Sci USA 83:9469–9473

Lamperti J (1966) Probability. Benjamin, New York. For non-random patterns one simply requires that $p_N(\mathbf{y})$ converge to a limit $p(\mathbf{y})$

Lee KH, Chung K, Chung JM, Coggeshall RE (1986) Correlation of cell body size, axon size, and signal conduction velocity for individually labelled dorsal root ganglion cells in the cat. J Comp Neurol 243:335–346

Little WA (1974) The existence of persistent states in the brain. Math Biosci 19:101–120

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

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

Malinow R, Miller JP (1986) Postsynaptic hyperpolarization during conditioning reversibly blocks induction of long-term potentiation. Nature 320:529–530

Miller R (1987) Representation of brief temporal patterns, Hebbian synapses, and the left-hemisphere dominance for phoneme recognition. Psychobiology 15:241–247

Nadal J-P, Toulouse G, Changeux J-P, Dehaene S (1986) Networks of formal neurons and memory palimpsests. Europhys Lett 1:535–542

Parisi G (1986) A memory which forgets. J Phys A: Math Gen 19:L617–620

Peretto P, Niez JJ (1986) Collective properties of neural networks. In: Bienenstock E, Fogelman-Soulié F, Weisbuch G (eds) Disordered systems and biological organization. Springer, Berlin Heidelberg New York, pp 171–185

Riedel U, Kühn R, Hemmen JL van (1988) Temporal sequences and chaos in neural nets. Phys Rev A 38:1105–1108

Scott AC (1977) Neurophysics. Wiley, New York

Sompolinsky H, Kanter L (1986) Temporal association in asymmetric neural networks. Phys Rev Lett 57:2861–2864

Toulouse G, Dehaene S, Changeux J-P (1986) A spin-glass model of learning by selection. Proc Natl Acad Sci USA 83:1695–1698

Willwacher G (1976) Fähigkeiten eines assoziativen Speichersystems im Vergleich zu Gehirnfunktionen. Biol Cybern 24:181–198

Willwacher G (1982) Storage of a temporal pattern sequence in a network. Biol Cybern 43:115–126

Priv.-Doz. Dr. J. L. van Hemmen
Sonderforschungsbereich 123 an der Universität Heidelberg
Im Neuenheimer Feld 294
D-6900 Heidelberg 1
Federal Republic of Germany