PERGAMON

Neural Networks 14 (2001) 805–813

Neural Networks

www.elsevier.com/locate/neunet

2001 Special issue

Temporal receptive fields, spikes, and Hebbian delay selection

Christian Leibold, J. Leo van Hemmen*

Physik Department, Technische Universität München, D-85747 Garching bei München, Germany

Accepted 9 April 2001

Abstract

The intriguing concept of a receptive field evolving through Hebbian learning, mostly during ontogeny, has been discussed extensively in the context of the visual cortex receiving spatial input from the retina. Here, we analyze an extension of this idea to the temporal domain. In doing so, we indicate how a particular spike-based learning rule can be described by means of a mean-field learning equation and present a solution for a couple of illustrative examples. We argue that the success of the learning procedure strongly depends on an interplay of, in particular, the temporal parameters of neuron (model) and learning window, and show under what conditions the noisy synaptic dynamics can be regarded as a diffusion process. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Spike-based learning; Hebbian learning; Delay selection; Temporal receptive field

1. Introduction

A receptive field is defined classically as the set of stimuli that alters a nerve cell's response (Hartline, 1940). The receptive-field description of neuronal responses to an external stimulus has since been a very successful approach to understanding the brain's function from a neurophysiologist's point of view (e.g.Aertsen & Johannesma, 1980; DeAngelis, Ohzawa & Freeman, 1995; Eckhorn, Krause & Nelson, 1993; Kowalski, Depireux & Shamma, 1996; Theunissen, Sen & Doupe, 1999). From a modeler's perspective, however, the notion of a receptive field includes a severe problem. Almost none of the neurons in the brain are directly exposed to an external stimulus. Instead, they are part of a high-dimensional and strongly interacting dynamical system. A neuron's receptive field is, consequently, no inherent property to the cell under consideration. It is a projection of the network's topology and dynamical state upon a specific cell's activity. A model of the local topology surrounding a single cell is, therefore, unable to explain its classical receptive field. Since almost every neuronal model is a local one, the notion of a receptive field has undergone a reinterpretation (MacKay & Miller, 1990; Roberts, 1999; Wimbauer, Wenisch, Miller & van Hemmen, 1997). A modeler's receptive field is the set of spike trains, or the set of neurons behind them, that serves as input to a subsystem of the brain. Different classes of spike trains are then related to different stimuli, which leads to an effective model of the sensory periphery. This is what we mean by 'receptive field' in this paper.

One popular approach towards an explanation of how a local receptive field emerges is the analysis of Hebbian-type learning rules in a feed-forward synaptic topology as proposed by Linsker (1986) for the visual pathway. A synaptic weight J_{mn} between input cell n and output cell m is assumed to change according to the dynamic equation

$$\frac{\mathrm{d}}{\mathrm{d}t} J_{mn} = w_5 + w_4 v_m + w_3 v_n + w_2 C_{mn}, \tag{1}$$

where $v_{m/n}$ is interpreted as the respective cell's spike rate, C_{mn} is a measure of the correlation between the activities of cell m and n, and the w_i are constants. Eq. (1) is called Hebbian-type since the w_2 term on the right combines pre- and postsynaptic activity.

In spite of the considerable success of rate-based learning equations in explaining orientation selectivity (Erwin & Miller, 1998; Linsker, 1986; MacKay & Miller, 1990), ocular dominance (Miller, Keller & Stryker, 1989), and even directional selectivity (Wimbauer et al., 1997), these learning rules fail whenever a temporal aspect of a receptive field has to be learned at a neuronal timescale as in e.g. the auditory system (Carr, 1993). This failure has been remedied by spike-timing dependent learning equations (Bartsch & van Hemmen, 2001; Eurich, Pawelzik, Ernst, Cowan & Milton, 1999; Gerstner, Kempter, van Hemmen & Wagner, 1996; Kempter, Leibold, Wagner & van Hemmen, 2001; Kempter, Gerstner & van Hemmen, 1999; Kistler & van Hemmen, 2000; Roberts, 1999; Ruf & Schmitt, 1997;

* Corresponding author.

PII: S0893-6080(01)00081-8

Senn, Markram & Tsodyks, 2001; Song, Miller & Abbott, 2000; van Hemmen, 2000).

In analogy to Eq. (1), the change of synaptic weights under the influence of spikes can be derived as a function of pre- and postsynaptic rates and the time-averaged correlation function $C_{mn}(t, t+s)$,

$$\frac{\mathrm{d}}{\mathrm{d}t}J_{mn} = \eta\left[w^{\mathrm{in}}v_n + w^{\mathrm{out}}v_m + \int_{-\infty}^{\infty}\mathrm{d}s C_{mn}(t, t+s)W(s)\right] \tag{2}$$

where the integral kernel W is called the *learning window*. For a detailed explanation we refer to Section 2; Kempter et al. (1999); van Hemmen (2000). Experimental evidence for the existence of W is steadily growing; see e.g. Bell, Han, Sugawara and Grant (1997); Bi and Poo (1998, 1999); Debanne, Gähwiler and Thompson (1998); Feldman (2000); Levy and Stewart (1983); Markram, Lübke, Frotscher & Sakmann, 1997, and Zhang, Tao, Holt, Harris & Poo (1998), reviewed in Bi and Poo (2001); Linden (1999); and Paulsen and Sejnowski (2000).

The present paper gives a brief derivation of Eq. (2) from a learning algorithm based on spike timing. We then indicate a strategy of how such a learning rule can be treated analytically and apply our results to a number of standard paradigms.

2. Self-averaging learning equations

Since Eq. (2) comprises only temporally averaged quantities, it is insensitive to the exact temporal location in an averaging interval of length $\mathscr{T}$ but *not* to time differences of pre- and postsynaptic spikes that induce the actual modification of the synapse, provided they occur within the learning window. We therefore suppose that each pre- and postsynaptic spike gives rise to an immediate change of the synaptic weight J_{mn}, say a pre-synaptic spike at time $t_n^{(f)}$ alters J_{mn} by ηw^{in}, a postsynaptic spike at time $t_m^{(f)}$ by ηw^{out} and, additionally, for every *pair* of pre- and postsynaptic spikes the weight is changed by an amount $\eta W(t_n^{(f)} - t_m^{(f)})$ where η is assumed to be very 'small' in the sense that $0 < \eta w \tau \ll 1$ for all occurring rates w and all neuronal time constants τ. After a time $\mathscr{T}$, we then find the change of the synaptic weight between input cell n and output cell m to be

$$\Delta J_{mn}(t) = J_{mn}(t) - J_{mn}(t-\mathscr{T}) \tag{3}$$

$$= \eta\left[\sum_{t-\mathscr{T}\le t_n^{(f)}<t} w^{\mathrm{in}} + \sum_{t-\mathscr{T}\le t_m^{(f)}<t} w^{\mathrm{out}} + \sum_{t-\mathscr{T}\le t_n^{(f)},t_m^{(f)}<t} W(t_n^{(f)} - t_m^{(f)})\right].$$

Since η is small, we can assume J_{mn} to be approximately constant in $[t-\mathscr{T}, t]$, and, taking single spike events to be independent, we apply the strong law of large numbers (Lamperti, 1996), i.e. we replace $\Delta J_{mn}(t)$ by its expectation value. Exploiting once more the fact that η is small, we introduce the temporal average $\bar{f}(t) := (1/\mathscr{T})\int_{t-\mathscr{T}}^{t}\mathrm{d}t' f(t')$ and assume $\mathscr{T}$ to greatly exceed the width of the learning window so as to obtain to fair approximation (Sanders & Verhulst, 1985)

$$\frac{\Delta J_{mn}}{\mathscr{T}} = \frac{\langle \Delta J_{mn}\rangle}{\mathscr{T}} = \eta\Bigg[w^{\mathrm{in}}\overline{\langle S_n(t)\rangle} + w^{\mathrm{out}}\overline{\langle S_m(t)\rangle} + \int_{-\infty}^{\infty}\mathrm{d}s W(s)\overline{\langle S_m(t)S_n(t+s-\Delta_{mn})\rangle}\Bigg], \tag{4}$$

where the spike trains $S(t)$ can be written as sequences of delta pulses $S(t) = \sum_{t^{(f)}}\delta(t - t^{(f)})$; cf. Gerstner and van Hemmen (1994). Since we are interested in weight changes on a time scale beyond $\mathscr{T}$, we place $\Delta J_{mn}/\mathscr{T} \to \mathrm{d}J_{mn}/\mathrm{d}t$ and end up with Eq. (2); cf. Kempter et al. (1999). As a result, we have found a *spike*-based interpretation of the quantities v and C in Eq. (2). We note that Eqs. (1) and (2) are equivalent as long as the time-averaged correlation function C is constant in time, which quite often does not hold. This is what we now focus on.

3. Poisson neurons

The learning equation (2) depends on statistical properties of pre- and postsynaptic spike trains. They can be calculated in explicit form for Poisson neurons (Kempter et al., 1999; van Hemmen, 2000), whose firing-probability density p_{F} is a function of the instantaneous membrane potential $v(t)$, so that we can write $p_{\mathrm{F}}[v(t)]$ for the density at time t. The probability that a neuron fires once during $[t, t+\Delta t)$ is $p_{\mathrm{F}}[v(t)]\Delta t$, firing twice or more has a probability $o(\Delta t)$ while the events in disjoint intervals are taken to be independent. For a straightforward introduction to inhomogeneous Poisson processes and mathematical details and derivations of results used below, the reader may consult Appendix A of Kempter, Gerstner, van Hemmen and Wagner (1998); the 'inhomogeneity' of the process stems from an explicit dependence of $p_F[v(t)]$ upon t.

Probability theory (Kempter et al., 1999, 1998; van Hemmen, 2000) now leads us to an explicit expression for the correlation function

$$C_{mn}(t+r, t) = \overline{\langle p_{\mathrm{F}}(v_m(t+r))\rangle|_{n,t}\, p_n^{\mathrm{in}}(t)}. \tag{5}$$

In order to calculate the time-averaged quantities v_m and C_{mn} in Eq. (2), we have to find an expression for the mean $\langle p_{\mathrm{F}}(v_m(t))\rangle$ and the average $\langle\cdots\rangle|_{n,t}$ under the condition of an input spike of cell n at time t. [If we are precise, a Poisson process assigns probability zero to an event at time t. What is meant is a spike occurring during the interval $[t, t+\Delta t)$ where the limit $\Delta t \to 0$ is taken afterwards.] Both expressions can be calculated for a Poissonian probability measure μ on the set of Poissonian input spike trains $\Omega = \cup_{F_1,\ldots,F_N}\Omega_{F_1\ldots F_N}$, where $\Omega_{F_1\ldots F_N}$ denotes the set of all spike trains with axon 1 conducting F_1 spikes, axon 2 conducting F_2 spikes, and so on;

all this is to happen during a period of duration $\mathscr{T}$. Since the subsets $\Omega_{F_1\ldots F_N}$ are disjoint, we can write

$$\langle p_{\mathrm{F}}(\upsilon_m)\rangle = \int_{\Omega} \mathrm{d}\mu\, p_{\mathrm{F}}(v_m) = \sum_{F_1,\ldots,F_N} \int_{\Omega_{F_1\ldots F_N}} \mathrm{d}\mu\, p_{\mathrm{F}}(\upsilon_m). \quad (6)$$

Though their occurrence is correlated through p_{F}, the individual spike events of a Poisson process are independent and, hence, the probability measure μ can be restricted to the subsets $\Omega_{F_1\ldots F_N}$ so as to give

$$\int_{\Omega_{F_1\ldots F_N}} \mathrm{d}\mu\, p_{\mathrm{F}}(\upsilon_m(t))$$

$$= \left(\prod_{n=1}^{N}\prod_{f=1}^{F_n}\int_{-\infty}^{t} \mathrm{d}t_j^{(f)}\right) P_{\{F_n\}}(t, \{t_j^{(f)}\}) p_{\mathrm{F}}(v_m(t)), \quad (7)$$

where, given the individual Poisson intensities p_n^{in},

$$P_{\{F_N\}}(t, \{t_j^{(f)}\}) = \mathrm{e}^{-\sum_n \int_{-\infty}^{t} \mathrm{d}s\, p_n^{\mathrm{in}}(s)} \left(\prod_{n=1}^{N} \frac{1}{F_n!} \prod_{f=1}^{F_n} p_n^{\mathrm{in}}(t_j^{(f)})\right) \quad (8)$$

is the probability density for N Poissonian input spike trains with $\{F_n\}$ spikes at times $\{t_n^{(f)}\}$. If we model the membrane potential v_m as a linear superposition of response kernels ϵ triggered by input spikes at time $t_n^{(f)}$ and delayed by Δ_{mn} (cf. Gerstner & van Hemmen, 1994), we get

$$v_m(t) = \sum_{nt_n^{(f)}} J_{mn}\epsilon(t - t_n^{(f)} - \Delta_{mn}). \quad (9)$$

The particular choice of the 'gain function'

$$p_{\mathrm{F}}(\upsilon) = v_0 \exp(\beta\upsilon) \quad (10)$$

then yields

$$\langle p_{\mathrm{F}}(\upsilon_m(t))\rangle = v_0 \exp\left\{\sum_{n=1}^{N} \int_0^{\infty} \mathrm{d}s\, p_n^{\mathrm{in}}(t - s - \Delta_{mn})[\mathrm{e}^{\beta J_{mn}\epsilon(s)} - 1]\right\}. \quad (11)$$

The calculation of the conditional mean can be performed analogously. We find

$$\langle p_{\mathrm{F}}(\upsilon_m(t + r))\rangle|_{n,t} = \langle p_{\mathrm{F}}(\upsilon_m(t + r))\rangle \mathrm{e}^{\beta J_{mn}\epsilon(r - \Delta_{mn})} \quad (12)$$

so that we end up with an explicit expression for the learning equation (2).

4. Linearization of the learning equation

By means of an expansion of the exponentials in Eq. (11) it is easy to see that a small β accounts for a learning equation that is approximately linear in J_{mn}, whereas a steep p_{F}, i.e. large β, yields a nonlinear dynamics. In other words, there is a nonlinear $\mathscr{N}$ so that Eq. (2) reads

$$\frac{\mathrm{d}}{\mathrm{d}t}|\mathbf{J}\rangle = \mathscr{N}|\mathbf{J}\rangle. \quad (13)$$

Our strategy is as follows. First, we will show that under rather general conditions, Eq. (13) has the fixed point $|\mathbf{J}^{\mathrm{fix}}\rangle = J^{\mathrm{fix}}|\mathbf{1}\rangle$, where $|\mathbf{1}\rangle$ is the constant unit vector with $\langle\mathbf{1}|\mathbf{J}\rangle = \sum_{\mathbf{n}} \mathrm{J}_{\mathbf{mn}}$. Second, we will give the linear differential equation for the deviations $|\iota\rangle = |\mathbf{J}\rangle = |\mathbf{J}^{\mathrm{fix}}\rangle$ from the fixed point,

$$\frac{\mathrm{d}}{\mathrm{d}t}|\iota\rangle = D\mathscr{N}|\iota\rangle. \quad (14)$$

Linear equations can be solved exactly. We therefore only need to determine the eigenspaces of $D\mathscr{N}$,

$$D\mathscr{N}|\mathbf{\Phi}_\lambda\rangle = \lambda|\mathbf{\Phi}_\lambda\rangle, \quad (15)$$

and, assuming a diagonable $D\mathscr{N}$, we can stick to expanding the solution to Eq. (14) in terms of eigenvectors,

$$|\iota\rangle = \sum_{\lambda} a_\lambda|\mathbf{\Phi}_\lambda\rangle. \quad (16)$$

Here, the time dependence of $a_\lambda(t)$ is given by Eq. (19). Since every basis $|\mathbf{\Phi}_\lambda\rangle$ has a reciprocal basis $\langle\mathbf{\Phi}^\lambda|$, here of eigenvectors of the hermitian adjoint operator $(D\mathscr{N})^*$ with eigenvalues λ^*, satisfying

$$\langle\mathbf{\Phi}^\lambda|\mathbf{\Phi}_{\lambda'}\rangle = \delta^\lambda_{\lambda'}, \quad (17)$$

we find that the coefficients a_λ evolve independently of each other according to

$$\left\langle\mathbf{\Phi}^\lambda\right|\frac{\mathrm{d}}{\mathrm{d}t}\iota\rangle = \frac{\mathrm{d}}{\mathrm{d}t}a_\lambda = \lambda a_\lambda, \quad (18)$$

and thus

$$a_\lambda(t) = \exp(t\lambda)\, a_\lambda(0). \quad (19)$$

In conclusion, the emerging synaptic structure reflects the eigenvectors with largest positive real part of λ, at least as long as the growth of the coefficients $a_\lambda(t)$ is unlimited. Since in the real synapse resources are restricted, synaptic growth saturates. In other words, if the initial conditions are such that one a_λ has too much of a head start it dominates structure formation, even if there is another eigenvalue with larger real part. That is to say, both the spectral properties of $D\mathscr{N}$ and the initial conditions have to be checked in order to make predictions concerning the resulting structure; cf. MacKay and Miller (1990); van Hemmen (2000).

4.1. Fixed point of the nonlinear dynamics

In order that the ensuing theory unfolds, we introduce a few biophysically motivated and highly plausible postulates.

Postulate 1. The distribution of latencies is broad in the sense that, to excellent approximation, the average membrane potential for $|\mathbf{J}\rangle := |\mathbf{J}^{\mathrm{fix}}\rangle = J^{\mathrm{fix}}|\mathbf{1}\rangle$ is constant. This holds, for example, in the barn owl (Kempter et al., 2001). We can, therefore, take

$$N^{-1}\sum_n p_n^{\mathrm{in}}(t - \Delta_{mn}) = v^{\mathrm{in}}, \quad (20)$$

meaning that the delay distribution totally smears out the temporal structure of the input process. Since we have announced a discussion of paradigms of delay selection, we refrain here from a more general analysis and, additionally, restrict the input processes to those having identical input rate $\nu_n^{\rm in} = \nu^{\rm in}$.

Application of Postulate 1 to Eq. (11) yields the temporally constant output rate

$$\nu_m^{\rm out} = \langle p_F(\upsilon_m(t))\rangle = \nu_0 \exp\left\{N\nu^{\rm in}\int_0^\infty {\rm d}s[{\rm e}^{\beta J^{\rm fix}\epsilon(s)} - 1]\right\}, \quad (21)$$

with ν_0 stemming from Eq. (10), and correlation function

$$C_{mn}(t+r,t) = \overline{p_n^{{\rm in}(t)}\langle p_F(\upsilon_m(t+r))\rangle|_{n,t}}$$
$$= \nu_m^{\rm out}\nu^{\rm in}\exp[\beta J^{\rm fix}\epsilon(r - \Delta_{mn})]. \quad (22)$$

If we insert Eqs. (21) and (22) into Eq. (2) the fixed point condition $\mathcal{N}|\mathbf{J}^{\rm fix}\rangle = 0$ is an implicit equation for $\beta J^{\rm fix}$,

$$\gamma := \frac{-w^{\rm in}}{\nu^{\rm in}\int {\rm d}s W(s)\exp[\epsilon(-s)\beta J^{\rm fix}] + w^{\rm out}} = \frac{\nu_m^{\rm out}}{\nu^{\rm in}}. \quad (23)$$

Postulate 2. The learning parameters $w^{\rm in}$, $w^{\rm out}$ as well as the learning window W and the response kernel ϵ are such that the fraction γ given by Eq. (23) is positive.

Postulate 3. We have $\gamma\nu^{\rm in}/\nu_0 = \nu_m^{\rm out}/\nu_0 > 1$.

Postulates 2 and 3 are more a restriction for reasonable model parameters than real biological constraints. They enable us to combine Eqs. (21) and (23) so as to get

$$\ln(\gamma\nu^{\rm in}/\nu_0)/(N\nu^{\rm in}) = \int_0^\infty {\rm d}s[{\rm e}^{\beta J^{\rm fix}\epsilon(s)} - 1] =: \psi(\beta J^{\rm fix}), \quad (24)$$

where the left-hand-side is a positive number and ψ is a monotonically increasing positive function of $\beta J^{\rm fix} > 0$ with $\psi(0) = 0$. Consequently, ψ is invertible and $\beta J^{\rm fix}$ is unique.

4.2. Linearized dynamics

The total differential $D\mathcal{N}$ of Eq. (14) contains partial derivatives of rate and correlation function. Defining an effective learning window and response kernel,

$$W^{\rm eff}(s;x) := W(s)\exp[x\epsilon(-s)], \quad \epsilon^{\rm eff}(s;x) := \epsilon(s)\exp[x\epsilon(s)] \quad (25)$$

we find

$$\frac{\rm d}{{\rm d}t}\iota_{mn} = \sum_n [k_2 + \delta_{nn'}k_3 + Q_{nn'}]\iota_{mn'}, \quad (26)$$

where the constants k_2 and k_3 are given by

$$k_2 = \eta\beta\nu_m^{\rm out}\nu^{\rm in}w^{\rm out},$$
$$k_3 = \eta\beta\nu_m^{\rm out}\nu^{\rm in}\int {\rm d}s\epsilon(-s)W^{\rm eff}(s;\beta J^{\rm fix}). \quad (27)$$

The Hebbian matrix reads

$$Q_{nn'} = \eta\beta\nu_m^{\rm out}\int {\rm d}s W^{\rm eff}(s;\beta J^{\rm fix})$$
$$\times \int {\rm d}s'\epsilon^{\rm eff}(s;\beta J^{\rm fix})\overline{p_n^{\rm in}(t - \Delta_{mn} + s)p_{n'}^{\rm in}(t - \Delta_{mn'} - s')}. \quad (28)$$

Defining a basis $|\mathbf{e}_{mn}\rangle$ by the property that it singles out components, viz., $\langle\mathbf{e}^{mn}|\iota\rangle = \iota_{mn}$ we can write Eq. (26) in a more sophisticated and, thus, mathematically more meaningful way,

$$\left(\frac{\rm d}{{\rm d}t} - k_3\right)|\iota\rangle = [|\mathbf{1}\rangle k_2\langle\mathbf{1}| + \sum_{nn'}|\mathbf{e}_{mn}\rangle Q_{nn'}\langle\mathbf{e}^{mn'}|]|\iota\rangle. \quad (29)$$

5. Examples 1. No delay section

In this section we specify the Hebbian matrix Q for the case of two simple Poissonian input processes in order to discuss the nature of Eq. (26) and $|\mathbf{J}^{\rm fix}\rangle$. Both subsections can also be discussed in rate description (1), i.e. they do not provide examples of delay selection as present in, say, the barn owl (Carr, 1993) and discussed in Section 6.

5.1. Homogeneous Poissonian input

The simplest Poisson process is a homogeneous process with intensity $\nu^{\rm in}$. From Eq. (28) we see that $Q_{nn'} = Q_0$ is merely a constant so that the dynamics is given by

$$\left(\frac{d}{dt} - k_3\right)|\iota\rangle = |\mathbf{1}\rangle(k_2 + Q_0)\langle\mathbf{1}|\iota\rangle. \quad (30)$$

In this case, the operator is highly degenerate. It has eigenvectors

$$|\mathbf{\Phi}_\mu\rangle = N^{-1/2}\sum_n \exp(2\pi i\mu n/N)|\mathbf{e}_{mn}\rangle \quad (31)$$

with eigenvalues $\lambda_\mu = N\delta_{\mu 0}(k_2 + Q_0) + k_3$, for $0 \le \mu < N$. That is to say, there are $N - 1$ eigenvectors ($\mu \neq 0$) with the same eigenvalue k_3 and there is only one eigenvector $|\mathbf{\Phi}_0\rangle = |\mathbf{1}\rangle$ with a different eigenvalue $\lambda_0 = N(Q_0 + k_2) + k_3$.

Two points ought to be stressed. First, if the system should be capable of selecting delays, N has to be quite large; in reality it nearly always is. Then, the absolute value $|k_3|$ is negligibly small compared to $N|(Q_0 + k_2)|$. Second, we find that the coefficient $\langle\mathbf{\Phi}^0|\iota\rangle = N^{-1/2}\sum_n \iota_{mn}$ represents the dynamics of the mean synaptic weight.

Because of

$$\langle \mathbf{\Phi}^{\lambda} | \iota \rangle(t) = \exp(t\lambda) \langle \mathbf{\Phi}^{\lambda} | \iota \rangle(0), \tag{32}$$

a positive eigenvalue $\lambda_0 > 0$ means that the learning rule gives rise to an alteration of the mean synaptic weight on the time scale $[N(Q + k_2)]^{-1}$.

If, in contrast, $\lambda_0 = N(Q_0 + k_2) < 0$ and $k_3 > 0$, Eq. (32) shows that the average synaptic weight $N^{-1/2}\langle \mathbf{\Phi}^0 | \mathbf{J} \rangle$ stays constant asymptotically, and coefficients $a_\mu = \langle \mathbf{\Phi}^{\mu} | \iota \rangle$ with $\mu \neq 0$, change at the same rate k_3^{-1}. Then Eq. (31) tells us we end up with plane waves and random phases giving an uncoordinated development of the individual synaptic weights: some grow, others decrease.

We conclude that in the present situation the fixed point $|\mathbf{J}^{\text{fix}}\rangle$ is in general a saddle point having Jacobian eigenvalues with negative and positive real parts. Eigenvectors with positive real parts are degrees of freedom for structure formation, though without any delay selection, negative real parts give $\lim_{t\to\infty} \langle \mathbf{\Phi}^{\lambda} | \iota \rangle(t) = 0$.

5.2. Correlated subpopulations

A more exciting situation originates if we divide the set of input cells $\mathcal{I}$ into two subpopulations $\mathcal{I}_+$ and $\mathcal{I}_-$ with $N_+ \approx N_- \approx N/2$ neurons, respectively. We assume that the former group behaves in a more coherent way than the latter. In terms of the matrix Q this situation is described by

$$Q_{nn'} = \begin{cases} Q_0 + Q_1 & \text{for } n, n' \in \mathcal{I}_+ \\ Q_0 & \text{else} \end{cases} \tag{33}$$

with $Q_1 > 0$. There are four subclasses of eigenvectors,

$$|\mathbf{\Phi}_{\pm,0}\rangle = \left(\sqrt{N/2(\alpha_\pm^2 + 1)}\right)^{-1} \left(\alpha_\pm \sum_{n\in\mathcal{I}_+} |\mathbf{e}_{mn}\rangle + \sum_{n\in\mathcal{I}_-} |\mathbf{e}_{mn}\rangle \right) \tag{34}$$

$$|\mathbf{\Phi}_{\pm,\mu}\rangle = N_\pm^{-1/2} \sum_{n\in\mathcal{I}_\pm} \exp[2\pi i \mu n / N_\pm] |\mathbf{e}_{mn}\rangle, \qquad \mu \neq 0 \tag{35}$$

with eigenvalues

$$\lambda_{\pm,0} = \delta_{\mu 0}[N/2(k_2 + Q_0)(1 + \alpha_\pm)] + k_3, \tag{36}$$

where

$$\alpha_\pm = \frac{Q_1}{2(k_2 + Q_0)} \pm \sqrt{\left(\frac{Q_1}{2(k_2 + Q_0)}\right)^2 + 1}. \tag{37}$$

Again, k_3 can be neglected if $N_- \approx N_+ = N/2$ is large. The two eigenvectors $|\mathbf{\Phi}_{\pm,0}\rangle$ with non-zero eigenvalue represent mixtures of the average synaptic weights in either subpopulation, $\langle \mathbf{\Phi}^{\pm,0} | \iota \rangle \propto (\alpha_\pm \sum_{n\in\mathcal{I}_+} \iota_{mn} + \sum_{n\in\mathcal{I}_-} \iota_{mn})$. Since Q_1 was assumed to be positive, Eq. (37) yields $\alpha_\pm \gtrless 0$ for any choice of $Q_0 + k_2$. In terms of the eigenvectors, $|\mathbf{\Phi}_{-,0}\rangle$ represents changes of the mean synaptic weights of both subpopulations into different directions, whereas $|\mathbf{\Phi}_{+,0}\rangle$ represents changes into the same direction. In the case of negative $Q_0 + k_2$, the corresponding eigenvalues $\lambda_{\pm,0} \gtrless 0$ have different signs, so that the importance of the eigenstate $|\mathbf{\Phi}_{-,0}\rangle$ of structure formation increases and that of $|\mathbf{\Phi}_{+,0}\rangle$ decreases. In the case of positive $Q_0 + k_2$, both eigenvalues are positive and a growing coefficient $\langle \mathbf{\Phi}^{+,0} | \iota \rangle$ leads to a trivial saturation.

To summarize, in a regime of Hebbian plasticity depending on spike timing, temporal correlations in a subpopulation of input cells give rise to a coherent synaptic development among synapses of this subset, if $Q_0 + k_2 < 0$; cf. Kempter et al. (1999); van Hemmen (2000).

6. Examples II. Delay selection

We now turn to input regimes that allow selection of synapses according to their transmission delays Δ_{mn}. From now on, as a consequence of Section 5.1, we put $k_3 = 0$ and $N(Q_0 + k_2) < 0$.

6.1. Temporally correlated input

A Hebbian matrix Q that depends on transmission delays Δ_{mn} requires that the temporal average of the correlation function does so too. For instance, we could take a kind of 'low-pass filtered white noise'. This is an inhomogeneous Poisson process; cf. Appendix A of Kempter et al. (1998). It is constructed as follows. We start with a *homogeneous* Poisson process $\mathcal{P}$ that has rate constant $\nu^{\text{in}} > 0$ and, thus, generates a series of random event times t_n. We then take such a series, fix it once and for all, and define the rate function

$$p_{\{t_n\}}(t) = \sum_{t_n < t} \tau_c^{-1} \exp[-(t - t_n)/\tau_c]. \tag{38}$$

The above sum is over all n with $t_n < t$. The resulting rate function $p_{\{t_n\}}$ of our new process depends on a *specific* realization $\{t_n\}$ of the underlying stationary Poisson process $\mathcal{P}$. We call our new process 'time-correlated Poisson' (TCP). It is very convenient for doing explicit calculations and grasps part of the stochasticity of the biophysics involved. It is easy to see why. Each time a new t_n appears, $p_{\{t_n\}}(t_n)$ makes a jump $+1/\tau_c$, inducing more 'spikes' than before in dependence upon the correlation time τ_c. Because the rate constant ν^{in} of the underlying Poisson process $\mathcal{P}$ is de facto a constant rate, we get little groups of TCP spikes each time a t_n appears; it is doing so at a rate ν^{in}.

For a TCP process with rate constant $\nu^{\text{in}} > 0$ and correlation time τ_c we obtain

$$\overline{p_n^{\text{in}}(t - \Delta_{mn} + s) p_{n'}^{\text{in}}(t - \Delta_{mn'} - s')} = (\nu^{\text{in}})^2 + \nu^{\text{in}}/(2\tau_c) \exp[-|s + s' + \Delta_{mn'} - \Delta_{mn}|/\tau_c]. \tag{39}$$

Due to Eq. (28), this leads to

$$Q_{nn'} = Q_0 + Q_1(\Delta_{mn'} - \Delta_{mn}) \tag{40}$$

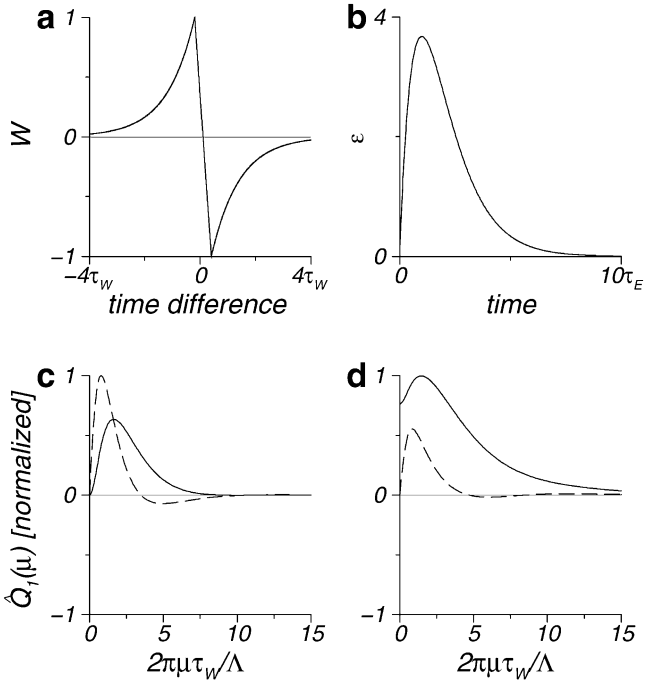

Fig. 1. A learning window (a) with exponential decay τ_W and a response kernel ϵ (b) with rise and decay time $\tau_E = \tau_W/5$ (see the Appendix for details) yield eigenvalues $N\hat{Q}_1$, which have been plotted for $\beta J^{\text{fix}} \approx 0$ (c) and $\beta J^{\text{fix}} = 0.5$ (d); real parts are shown as solid lines, imaginary parts as dashed lines.

where

$$Q_0 = \eta \nu_m^{\text{out}} (\nu^{\text{in}})^2 \int \mathrm{d}s W^{\text{eff}}(s) \int \mathrm{d}s' \epsilon^{\text{eff}}(s'),$$

$$Q_1(\Delta) = \eta \nu_m^{\text{out}} \nu^{\text{in}} / (2\tau_c) \int \mathrm{d}s W^{\text{eff}}(s) \int \mathrm{d}s' \epsilon^{\text{eff}}(s') \tag{41}$$

$$\exp(-|s + s' + \Delta|/\tau_c).$$

If τ_c is far beyond the maximal delay Δ and the width of W^{eff}, say τ_W, Q_1 is nearly constant and we have arrived at the case of Section 5. If, on the other hand, τ_c is small enough, Q_1 varies distinctly with Δ and so does $D\mathcal{N}$ in Eq. (14), i.e. the learning dynamics depends on the delay coordinate Δ. Consequently, we are in a regime that is capable of delay selection in that synaptic weights are strengthened or weakened depending on their delay Δ. The most accurate temporal resolution of delay selection is restricted to the width of W so that, for $\tau_c \to 0$, the resolution of delay selection is $\mathcal{O}(\tau_W)$.

To discuss structure formation on the set of delays Δ_{mn}, we introduce an ordered discrete topology. The simplest and, nevertheless, in many cases realistic (Carr, 1993) manner to do this is

$$\Delta_{mn} = n(\Lambda/N) \tag{42}$$

for positive integers $0 \le n < N$. Here, Λ denotes the maximal latency present. Because of Eq. (42), the matrix Q is a circulant (Bellman, 1970) and, hence, its eigenspaces are given by

$$|\mathbf{\Phi}_\mu\rangle = N^{-1/2} \sum_n \exp[2\pi i \mu n/N] |\mathbf{e}_{mn}\rangle,$$
$$\lambda_\mu = N[\delta_{\mu 0}(k_2 + Q_0) + \hat{Q}_1(\mu)], \tag{43}$$

where $\hat{Q}_1(\mu) = \int \mathrm{d}s Q_1(s) \exp[-2\pi i \mu s/\Lambda]$ denotes an ordinary Fourier transform.

We now discuss synaptic structure formation by means of $\hat{Q}_1(\mu)$; see Fig. 1c and d. Because

$$\langle \mathbf{e}^{mn} | \mathbf{\Phi}_\mu \rangle \langle \mathbf{\Phi}^\mu | \iota \rangle(t) = \exp\{t\Re(\lambda(\mu)) + i[t\Im(\lambda(\mu)) + 2\pi\mu n/N]\} \langle \mathbf{\Phi}^\mu | \iota \rangle(0), \tag{44}$$

real parts quantify the growth rate of the spectral component $2\pi\mu/\Lambda$ whereas imaginary parts measure its drifting velocity. A positive $\mathcal{I}(\lambda(\mu))$ for $\mu > 0$ yields a drift toward shorter delays (smaller n), a negative $\mathcal{I}(\lambda(\mu))$ results in a phase drift toward longer delays (larger n).

The real part $\mathcal{R}(\lambda(\mu))$ shows band-pass character (Fig. 1c and d) as a function of $2\pi\mu/N$. As a consequence, it gives rise to a coordinated development of synaptic weights on the domain of transmission delays Δ_{mn} and it does so in such a way that the eigenvector with largest real part dominates structure formation; cf. Song et al. (2000).

6.2. Periodic input processes

We now turn to an input regime that is biologically relevant especially in early auditory brainstem processing. Since the auditory system is tonotopically organized, i.e. different

locations process different frequency components, each cell is exposed to spike input with a predominant periodicity. The idea behind this is—simply stated—that the cochlea is a kind of 'inverse piano'. Input correlation functions have temporal correlation lengths of a couple of periods. If the memory time scale τ_W of the Hebbian learning rule (2) does not exceed the input correlation width, the input spikes can be modeled by a strictly periodic Poisson process. If we then focus on a period T_p, we obtain

$$\overline{p_n^{\rm in}(t-\Delta_{mn}+s)p_{n'}^{\rm in}(t-\Delta_{mn'}-s')} = (\nu^{\rm in})^2\sum_p |\hat{g}_\mu|^2\exp[2\pi i\mu(\Delta_{mn}-\Delta_{mn'})/\tau_p], \quad (45)$$

and, hence,

$$Q_{nn'} = \sum_\mu \exp[2\pi i\mu(\Delta_{mn}-\Delta_{mn'})/T_p]\hat{Q}(\mu) \quad (46)$$

where

$$\hat{Q}(\mu) = \eta\nu_m^{\rm out}(\nu^{\rm in})^2|\hat{g}_\mu|^2\hat{W}^{\rm eff}(\mu)\hat{\epsilon}^{\rm eff}(\mu). \quad (47)$$

For the discrete delay topology $\Delta_{mn} = n/NT_p$, we obtain eigenspaces and eigenvalues of the form

$$|\mathbf{\Phi}_\mu\rangle = N^{-1/2}\sum_n \exp[2\pi i\mu/N]|\mathbf{e}_{mn}\rangle, \quad \lambda_\mu = [\delta_{\mu 0}(k_2+Q_0)+\hat{Q}_1(\mu)]. \quad (48)$$

Here, $Q_0 = \sum_{\mu'\in N\mathbb{Z}}\hat{Q}(\mu)$ and $\hat{Q}_1(\mu) = (1-\delta_{\mu 0})\times\sum_{\mu'\in\mu+N\mathbb{Z}}\hat{Q}(\mu)$. A glance at Fig. 1c and d suffices to reveal that, for a sufficiently small period T_p, the largest real part belongs to $\mu = 1$ and, thus, structure formation reflects the periodicity of the input.

In view of the above periodic-delay selection, it is handy to introduce a quality measure for quantifying a particular parameter's capability to generate the required delay selectivity. Using the Heaviside step function $\Theta(x)$ ($=1$ for $x>0$ and $=0$ for $x\le 0$), we call

$$\mathcal{M}(\omega_0,\omega_1) := (\omega_1-\omega_0)^{-1}\int_{\omega_0}^{\omega_1} {\rm d}\omega\Theta(\Re[\lambda(\omega)])\Re[\lambda(\omega)]/|\lambda(\omega)| \quad (49)$$

the *average periodic momentum* (APM) between ω_0 and ω_1. It is the average ratio of positive real part and total length of the leading eigenvalue for input signals with periods between $2\pi/\omega_1$ and $2\pi/\omega_0$. For each period $2\pi/\omega$, the ratio $\Theta(\Re[\lambda(\omega)])\Re[\lambda(\omega)]/|\lambda(\omega)|$ quantifies the fraction of synaptic modification that is responsible for an increase of the leading eigenvector—and not for its oscillation.

In Fig. 2, we show the dependence of $\mathcal{M}(\pi/\tau_W, 4\pi/\tau_W)$ upon the shift $s_{\rm Peak}$ of the positive peak of the learning window (see Fig. 2a), the time constant of the response kernel ϵ (Fig. 1b) and steepness β of the gain function p_F. Two features are evident. First, the APM increases with β; see Fig. 2e. Second, the time constant τ_E of the membrane response kernels has to be adapted to the shift $s_{\rm Peak}$ of the positive peak of the learning window in such a way that short response kernels come along with short $s_{\rm Peak}$; cf. Fig. 2f. In conclusion, periodic-structure formation is a result of an intricate cooperation of the temporal features of a neuron (model) and the synaptic learning window. Nonlinearities generally improve structure formation.

7. Noise

Eq. (2) is an ensemble-averaged mean-field equation. Its solution describes the mean trajectory of a noisy system. The mean-field description is valid, if the variance is small enough so as to avoid disturbing the mean trajectory 'too much'. It has been shown by Kempter et al. (1999) that this is indeed the case, if η is small enough. The reason is that the variance scales with η^2, whereas structure formation, i.e. the eigenvalues, scale with η.

We now present an alternative noise model. If we knew the eigensystems $|\mathbf{\Phi}_\lambda\rangle$ of the linear operator $D\mathcal{N}$ in Eq. (14), we could write down an effective stochastic dynamics for the coefficients $a_\lambda = \langle\mathbf{\Phi}^\lambda|\iota\rangle$,

$$\frac{\rm d}{{\rm d}t}a_\lambda = \lambda a_\lambda + \xi_\lambda, \quad (50)$$

where ξ_λ denotes the fluctuations of the coefficients, and, as a result, a_λ are now stochastic variables. Using the identity $\mathbb{1} = \sum_n |\mathbf{e}_{mn}\rangle\langle\mathbf{e}^{mn}|$, we then calculate the average variance of the synaptic weights,

$$\begin{aligned}{\rm Var}(J) &:= N^{-1}\sum_n(\langle J_{mn}^2\rangle - \langle J_{mn}\rangle^2)\\ &= \sum_{n,\lambda\lambda'}(\langle a_\lambda a_{\lambda'}^*\rangle - \langle a_\lambda\rangle\langle a_{\lambda'}^*\rangle)\langle\mathbf{\Phi}^\lambda|\mathbf{e}_{mn}\rangle\langle\mathbf{e}^{mn}|\mathbf{\Phi}_{\lambda'}\rangle\\ &= \sum_\lambda(\langle|a\lambda|^2\rangle - |\langle a_\lambda\rangle|^2) = \sum_\lambda\exp(2t\lambda)\int_0^t{\rm d}s\\ &\quad\times\int_0^t{\rm d}s'\exp[-\lambda(s+s')]\langle\xi_\lambda(s)\xi_\lambda(s')\rangle. \end{aligned} \quad (51)$$

Synaptic growth is often thought of as random walk, an approximation of a diffusion process (Lamperti, 1996). We therefore postulate ξ_λ to be white noise with $\langle\xi_\lambda(s)\xi_\lambda(s')\rangle = \sigma_\lambda^2\delta(s-s')$ and obtain for the average variance

$${\rm Var}(J) = \sum_\lambda\frac{\sigma_\lambda^2}{2\lambda}({\rm e}^{2t\lambda}-1). \quad (52)$$

For small λt, Eq. (52) yields ${\rm Var}(J) = t\sum_\lambda\sigma_\lambda^2[1+\mathcal{O}(\lambda t)]$ and, thus, on the time scale of $\min_{\Re(\lambda>0)}[1/\Re(\lambda)]$ spike-based learning is, indeed, a diffusion process.

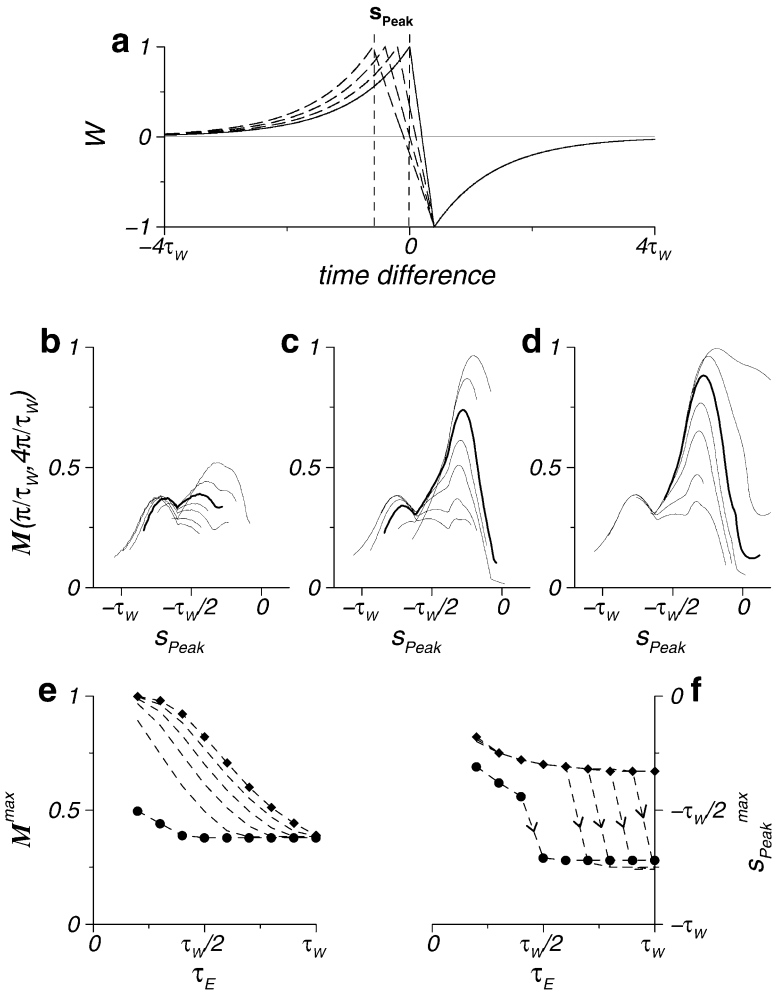

Fig. 2. Average periodic momentum (AMP) is a measure of the quality of periodic-structure formation and quantified by $\mathscr{M}$. It depends on the interplay between learning window W, membrane response kernel ϵ, and the gain function p_F. We define s_{Peak} to be the shift of the positive peak of W (a) and plot $\mathscr{M}(\pi/\tau_W, 4\pi/\tau_W)$ for different values of τ_E as a function of s_{Peak} for $\beta J^{fix} \approx 0$ (b), $\beta J^{fix} = 0.2$ (c), and $\beta J^{fix} = 0.4$ (d): black lines from large to small peaks denote r $\tau_E = 0.2\tau_W, 0.3\tau_W, 0.4\tau_W$ [thick line], $0.5\tau_W, 0.6\tau_W, 0.8\tau_W, \tau_W$. All graphs have a local maximal value of approximately $-0.75\tau_W$ and another one for $s_{Peak} > -\tau_W/2$. The former is more significant for small β and large τ_E, whereas the latter increases with smaller τ_E and larger β. (e) Shows the dependence of the global maximum $\mathscr{M}^{max}$ of the APM on the membrane time constant τ_E. Diamonds denote $\beta J^{fix} = 0.5$, circles $\beta J^{fix} = 0$, while the other lines are obtained with $\beta J^{fix} = 0.1, 0.2, 0.3, 0.4$ and show a monotonic increase of $\mathscr{M}^{max}$. In the same way, (f) shows a monotonically decreasing dependence of the shift s^{max}_{Peak} at $\mathscr{M}^{max}$ upon τ_E. A sharp decrease reflects a transition (arrows) of $\mathscr{M}^{max}$ from the short-τ_E (right) to the long-τ_E (left) peak in (b)–(d).

8. Discussion

Delay selection by means of Hebbian learning (Gerstner et al., 1996; Roberts, 1999; Song et al., 2000) can account for the emergence of temporal receptive fields on a time scale τ_W and, hence, for temporal maps (Kempter et al., 2001) on the very same time scale set by the width τ_W of the learning window W. We have shown that synaptic structure formation is a result of an interplay between neuron (model) and learning window. So, the following, natural, question comes up: how, then, are learning window and neuron properties adjusted to each other in the real brain? As far as we know, there are as yet mere speculations only. Amongst those are ideas of meta-plasticity, i.e. the learning parameters somehow tune themselves by means of second-order 'meta' learning rules. Disregarding the lack of experimental evidence, meta-plasticity simply postpones the key question: how is a meta learning rule adapted to fit physiological reality?

A second way of looking at the above problem may be more promising. Learning parameters as well as spike generation are both results of the electrophysiological properties of one and the same nerve cell. Fast cells should be fast in spike generation as well as in learning. It would, therefore, be fruitful to find out whether and to what extent both timescales reflect a common physiological origin.

Appendix

Learning window W and membrane response kernel ϵ

that have been used in Figs. 1 and 2 are given by

$$W(s) = \begin{cases} \exp[-(s - s_{\text{Trough}})/\tau_W] & s > s_{\text{Trough}} \\ \alpha\{\exp[-(s - s_{\text{Trough}})/\tau_W] - \exp[-(s - s_{\text{Trough}})/\tau_W]\} & s_{\text{Peak}} < s \le s_{\text{Trough}} \\ \exp[-(s - s_{\text{Peak}})/\tau_W] & s \le s_{\text{Peak}} \end{cases}$$

and

$$\epsilon(s) = \begin{cases} s/(\tau_E)^2 \exp(-s/\tau_E) & s > 0 \\ 0 & \text{else} \end{cases},$$

for $s_{\text{Trough}} = \tau_W/5$ and $s_{\text{Peak}} = 0$ unless mentioned otherwise. The factor $\alpha = 1/\{1 - \exp[-(s_{\text{Trough}} - s_{\text{Peak}})/\tau_W]\}$ is chosen so that W is a continuous function.

References

Aertsen, A. M. H. J., & Johannesma, P. I. M. (1980). Spectro-temporal receptive fields of auditory neurons in the grass frog. I. Characterization of tonal and natural stimuli. *Biological Cybernetics*, *38*, 223–234.

Bartsch, A. P., & van Hemmen, J. L. (2001). Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex. *Biological Cybernetics*, *84* (1), 41–55.

Bell, C. C., Han, V. Z., Sugawara, Y., & Grant, K. (1997). Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, *387*, 278–281.

Bellman, R. (1970). *Introduction to matrix analysis*, (2nd ed.). New York: McGraw-Hill.

Bi, G. -Q., & Poo, M.-m. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.*, *18* (9), 10464–10472.

Bi, G. -Q., & Poo, M.-m. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, *401* (6755), 792–796.

Bi, G. -Q., & Poo, M.-m. (2001). Synaptic modification by correlated activity: Hebb's postulate revisited. *Annual Reviews of Neuroscience*, *24*, 139–166.

Carr, C. E. (1993). Processing of temporal information in the brain. *Annual Reviews of Neuroscience*, *16*, 223–243.

DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1995). Receptive-field dynamics in the central visual pathway. *Trends in Neuroscience*, *18* (10), 451–458.

Debanne, D., Gähwiler, B. H., & Thompson, S. M. (1998). Long-term synaptic plasticity between pairs of individual CA3 pyramidal cells in rat hippocampal slice cultures. *J. Physiol.*, *507* (1), 237–247.

Eckhorn, R., Krause, F., & Nelson, J. I. (1993). The RF-cinematogram. A cross-correlation technique for mapping several visual receptive fields at once. *Biological Cybernetics*, *69* (1), 37–55.

Erwin, E., & Miller, K. D. (1998). Correlation-based development of ocularly matched orientation and ocular dominance maps: determination of required input activities. *J. Neurosci.*, *19* (16), 9870–9895.

Eurich, C. W., Pawelzik, K., Ernst, U., Cowan, J. D., & Milton, J. G. (1999). Dynamics of self-organized delay adaption. *Physical Review Letters*, *82* (7), 1594–1597.

Feldman, D. E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. *Neuron*, *27* (1), 45–56.

Gerstner, W., & van Hemmen, J. L. (1994). Coding and information processing in neural networks. In E. Domany, J. L. van Hemmen & K. Schulten, *Models of neural networks II* (pp. 1–93). New York: Springer-Verlag.

Gerstner, W., Kempter, R., van Hemmen, J. L., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, *383* (6595), 76–78.

Hartline, H. K. (1940). The receptive fields of optic nerve fibers. *American Journal of Physiology*, *130*, 690–699.

Kempter, R., Leibold, C., Wagner, H., & van Hemmen, J. L. (2001). Formation of temporal-feature maps by axonal propagation of synaptic learning. *Proceedings of the National Academy of Sciences of the USA*, *98*, 4166–4171.

Kempter, R., Gerstner, W., & van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E*, *59* (4), 4498–4514.

Kempter, R., Gerstner, W., van Hemmen, J. L., & Wagner, H. (1998). Extracting oscillations: neuronal coincidence detection with noisy periodic spike input. *Neural Computation*, *10* (8), 1987–2017.

Kistler, W. M., & van Hemmen, J. L. (2000). Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic action potentials. *Neural Computation*, *12* (2), 385–405.

Kowalski, N., Deprieux, D. A., & Shamma, S. A. (1996). Analysis of dynamic spectra in ferret primary auditory cortex. I & II. *J. Neurophysiol.*, *76*(5), 3503–3523, 3524–3534.

Lamperti, J. W. (1996). *Probability: a survey of the mathematical theory*, (2nd ed.). New York: Wiley.

Levy, W. B., & Stewart, D. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, *8* (4), 791–797.

Linden, D. J. (1999). The return of the spike: postsynaptic action potentials and the induction of LTP and LTD. *Neuron*, *22*, 661–666.

Linsker, R. (1986). From basic network principles to neural architecture (series). *Proceedings of the National Academy of Sciences of the USA*, *83*(19), 7508–7512, (20), 8390–8394, (21), 8779–8783.

MacKay, D. J. C., & Miller, K. D. (1990). Analysis of Linsker's application of Hebbian rules to linear networks. *Network*, *1*, 257–297.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, *275* (5297), 213–215.

Miller, K. D., Keller, J. B., & Stryker, M. P. (1989). Ocular dominance column development: analysis and simulation. *Science*, *245* (4918), 605–615.

Paulsen, O., & Sejnowski, T. J. (2000). Natural patterns of activity and long-term synaptic plasticity. *Current Opinion in Neurobiology*, *10* (2), 172–179.

Roberts, P. D. (1999). Computational consequences of temporally asymmetric learning rules: I. Differential Hebbian learning. *Journal of Computational Neuroscience*, *7* (3), 235–246.

Ruf, B., & Schmitt, M. (1997). Unsupervised learning in networks of spiking neuron using temporal coding. In W. Gerstner, A. Germond, M. Hasler & J.-D. Nicoud, *Proceedings of the Seventh International Conference on Artificial Neural Networks (ICANN '97)* (pp. 361–366). Heidelberg: Springer.

Sanders, J. A., & Verhulst, F. (1985). *Averaging methods in nonlinear dynamical systems*. New York: Springer.

Senn, W., Markram, H., & Tsodyks, M. (2001). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Computation*, *13* (1), 35–67.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, *3* (9), 919–926.

Theunissen, F. E., Sen, K., & Doupe, A. J. (2000). Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J. Neurosci.*, *20* (6), 2315–2331.

van Hemmen, J. L. (2000). Theory of synaptic plasticity. In F. Moss & S. Gielen, *Handbook of biophysics* (pp. 749–801), vol. 4. Elsevier: Amsterdam.

Wimbauer, S., Wenisch, O. G., Miller, K. D., & van Hemmen, J. L. (1997). Development of spatiotemporal receptive fields of simple cells I & II. *Biological Cybernetics*, *77*(6), 453 –461, 463–477.

Zhang, L. A., Tao, H. W., Holt, C. E., Harris, W. A., & Poo, M.-m. (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, *395* (6697), 37–44.