Coding Properties of Spiking Neurons: Reverse and Cross-Correlations

Wulfram Gerstner[*]

Swiss Federal Institute of Technology Lausanne, Laboratory of Computational Neuroscience, DI-LCN, CH 1015 Lausanne EPFL

To appear in: Neural Networks 14 (2001)

Abstract

What is the 'meaning' of a single spike? Spike-triggered averaging ('reverse correlations') yields the typical input just before a spike. Similarly, cross-correlations describe the probability of firing an output spike given (one additional) presynaptic input spike. In this paper we analytically calculate reverse and cross-correlations for a spiking neuron model with escape noise. The influence of neuronal parameters (such as the membrane time constant, the noise level, and the mean firing rate) on the form of the correlation function is illustrated. The calculation is done in the framework of a population theory that is reviewed. The relation of the population activity equations to population density methods is discussed. Finally we indicate the role of cross-correlations in spike-time dependent Hebbian plasticity.

Keywords: spiking neurons, integrate-and-fire model, reverse correlations, cross-correlations, population equations, population activity, rate model, Hebbian learning, spike-time dependent plasticity,noise.

1 Introduction

The problem of neural coding can be phrased as two questions (Optican and Richmond 1987; Bialek, Rieke, de Ruyter van Stevenick, and Warland 1991; Tovee, Rolls, Treves, and Belles 1993; Kjaer, Hertz, and Richmond 1994;

[*]Email address: wulfram.gerstner@epfl.ch

Softky 1995; König, Engel, and Singer 1996; Rieke, Warland, de Ruyter van Steveninck, and Bialek 1996; de Ruyter van Steveninck, Lowen, Strong, Koberle, and Bialek 1997; Berry and Meister 1998; Roddey, Girish, and Miller 2000).

(i) *Encoding*: What is the effect of a single presynaptic spike on the activity of a postsynaptic neuron, i.e., does the stimulation cause an output spike or not?

(ii) *Decoding*: What do we learn about the stimulus from the observation of a single output spike?

The first questions can be addressed by cross-correlation experiments. A neuron that is spontaneously active because of background activity receives at time t^{in} one additional presynaptic input spike. By averaging over many presynaptic spike arrivals at the same synapse, we can estimate the probability of generating an output spike $t^{(f)}$ given an input spike t^{in}. The cross-correlation $C^{\text{cross}}(t^{(f)}|t^{\text{in}})$ can therefore be evaluated via a Peri-Stimulus-Time Histogram (PSTH) triggered on the presynaptic spike. Systematic experiments along those lines have for example been performed by (Poliakov, Powers, and Binder 1997) in continuation of earlier models and experiments. (Moore, Segundo, Perkel, and Levitan 1970; Knox 1974; Kirkwood and Sears 1978; Fetz and Gustafsson 1983).

The second question can be addressed by information theoretic measurements (Optican and Richmond 1987; Tovee, Rolls, Treves, and Belles 1993; Bialek, Rieke, de Ruyter van Stevenick, and Warland 1991; de Ruyter van Steveninck, Lowen, Strong, Koberle, and Bialek 1997) or, in its simplest form by reverse-correlation measurements (de Boer and Kuyper 1968). In the reverse-correlation approach, a neuron is stimulated by a time-dependent stimulus $\mathcal{I}(t) = \mathcal{I}_0 + \Delta\mathcal{I}(t)$. Every time an output spike $t^{(f)}$ occurs, the time course $\Delta\mathcal{I}(t^{(f)} - s)$ just before the firing time $t^{(f)}$ is recorded. Averaging over many firing times $t^{(f)}$ yields the typical input

$$\langle \mathcal{I}(t^{(f)} - s) \rangle_{t^{(f)}} = C^{\text{rev}}(s) \,. \tag{1}$$

Thus the reverse correlation function $C^{\text{rev}}(s)$ is defined by a spike-triggered average.

For a *linear rate* model where output spikes $t^{(f)}$ are generated by an inhomogeneous Poisson process with input-dependent rate

$$\nu^{\text{out}}(t) = \nu_0^{\text{out}} + \int_0^\infty G(s)\, \Delta\mathcal{I}(t-s)\, \mathrm{d}s \tag{2}$$

it is well known (de Boer and Kuyper 1968) that the reverse correlation function is determined by the filter $G(s)$ and the auto-correlation $\langle \Delta\mathcal{I}(t)\, \Delta\mathcal{I}(t') \rangle$

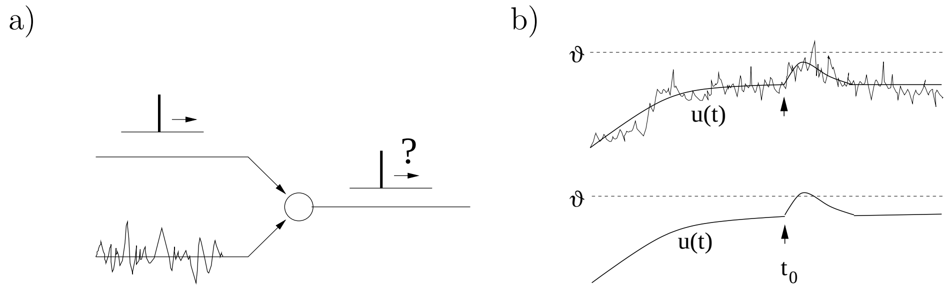

Figure 1: a) A neuron which is driven by a noisy background input receives at time t_0 one extra input spike. Does this extra input trigger an output spike? b) Two hypothetic scenarios: Top: With noisy input an output spike is the more likely the closer the mean membrane potential (thick solid line) is to the threshold ϑ (dashed horizontal line). The firing probability increases during the postsynaptic potential that is caused by the input pulse at t_0 (arrow). Bottom: Without noise, the membrane potential can reach threshold (for the first time) only during the rising phase of the postsynaptic potential.

of the input (see Appendix A for a review of the argument). In particular, for white-noise input with

$$\langle \Delta\mathcal{I}(t) \rangle = 0 \quad \text{and} \quad \langle \Delta\mathcal{I}(t)\, \Delta\mathcal{I}(t') \rangle = \sigma^2\, \delta(t-t') \tag{3}$$

we have

$$C^{\text{rev}}(s) = \frac{\sigma^2}{\nu_0^{\text{out}}}\, G(s)\,. \tag{4}$$

For *spiking neurons*, however, the form of the reverse correlation function and its relation to elementary model parameters like the membrane time constant has remained unclear. Apart from occasional simulation results [e.g., (Softky 1995)], no systematic study seems to have been undertaken. Oddly enough, it is a recent progress in the theory of population dynamics (Knight 1972a; Abbott and van Vreeswijk 1993; Gerstner and van Hemmen 1992; Gerstner 1995; Amit and Brunel 1997; Fusi and Matteo 1999; Brunel and Hakim 1999; Gerstner 2000; Nykamp and Tranchina 2000; Omurtag, Knight, and Sirovich 2000) that allows us now to calculate some of the coding properties of single neurons (Herrmann and Gerstner 1999; Herrmann and Gerstner 2001a; Brunel, Chance, Fourcaud, and Abbott 2001; Bethge, Pawelzik, Rothenstein, and Tsodyks 2001); see also (Knight 1972b). In fact, the typical behavior of a *single* neuron is identical to that of a population of independent neurons (i.e., without coupling).

The paper is organized as follows. After a brief discussion of the dynamics of a single spiking neuron in Section 2, we will review in Section 3 population

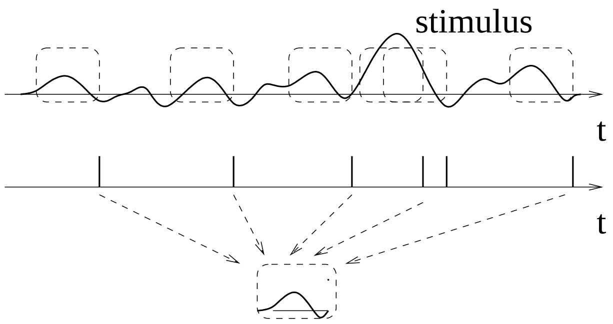

Figure 2: Reverse correlation technique (schematic). The stimulus $\mathcal{I}(t)$ in the top trace has caused the spike train shown immediately below. The time course of the stimulus just before the spikes (dashed boxes) has been averaged to yield the typical time course $\langle \mathcal{I}(t^{(f)} - s) \rangle_{t^{(f)}}$ (bottom).

equations that describe the dynamics of a pool of neurons. Linearization of the population equations allows us to analytically calculate the reverse correlation function for the case of input with low-amplitude fluctuations $\Delta\mathcal{I}(t)$. Reverse and cross-correlations are the topic of Section 4. In the discussion of Section 5 we indicate how the correlation function enters into the formalism of spike-time dependent Hebbian learning. Formal mathematical arguments have been moved to the appendix A and B.

2 Spike Response Model

In this section, we review the Spike Response Model (SRM) and specify the particular, somewhat simplified version that we will use throughout the paper.

2.1 General Framework

During and immediately after an action potential, the membrane potential of a given neurons follows (at least in most cases) a rather stereotyped trajectory. In the Spike Response Model, this trajectory is described as a function $\eta(t - \hat{t})$ where $\hat{t}$ is the moment when the spike has been triggered. The form of the function η can be chosen so as to reproduce as closely as possible the time course of an action potential. If a short current pulse $\mathcal{I}(t)$ is applied during or immediately after an action potential, the membrane potential $u(t)$ will deviate slightly from its standard time course $\eta(t - \hat{t})$. The difference between the actual time course $u(t)$ and the reference time course $\eta(t - \hat{t})$ is described by a response kernel κ, i.e.,

$$u(t) = \eta(t-\hat{t}) + \int_0^\infty \kappa(t-\hat{t}, s)\, \mathcal{I}(t-s)\, \mathrm{d}s\,. \tag{5}$$

Note that the response kernel κ depends on the time $t-\hat{t}$ since the last spike. In order to emphasize this dependence we introduce a refractory variable $r = t - \hat{t}$ and rewrite Eq. (5) in the form

$$u_r(t) = \eta(r) + \int_0^\infty \kappa(r, s)\, \mathcal{I}(t-s)\, \mathrm{d}s\,. \tag{6}$$

If the last spike occurred a long time back in the past and if there has been no further stimulation, the neuron is at rest. In this case, the response to a small current pulse is described by $\kappa_0(s) = \lim_{r\to\infty}\kappa(r,s)$; typically the time course $\kappa_0(s)$ is characterized by the passive membrane time constant τ_m. During or a few milliseconds after an action potential, the response to the same current pulse can be quite different since sodium and potassium channels are open that are normally closed at rest. The dependence of κ upon the refractory variable r takes care of this effect (Kistler, Gerstner, and van Hemmen 1997).

In a noiseless model neuron, spikes are generated by a threshold process that resets the refractory variable r in Eq. (6)

$$u(t) = \vartheta \quad \Longrightarrow \quad r = 0 \quad (\hat{t} = t)\,. \tag{7}$$

Graphically speaking, we simply add (or 'paste in') an action potential of standard shape $\eta(t-\hat{t})$ if the membrane potential reaches the threshold at time $\hat{t}$. Noise can be induced by adding a stochastic component to the input current $\mathcal{I}(t) \longrightarrow \mathcal{I}(t) + \mathcal{I}_{\text{noise}}(t)$. The stochastic contribution $\mathcal{I}_{\text{noise}}$ can be motivated by stochastic spike arrival. Usually the noise is taken as white, i.e., $\langle \mathcal{I}_{\text{noise}} \rangle = 0$ and $\langle \mathcal{I}_{\text{noise}}(t)\mathcal{I}_{\text{noise}(t')} \rangle = \sigma_n^2\, \delta(t-t')$. We call this the diffusive noise model. The noisy input generates a distribution of membrane potentials with standard deviation σ_u that can be measured in experiments (Destexhe and Pare 1999). While such a noise model is not too far from reality, it suffers from the disadvantage that it renders the analytical calculation of interspike-interval distributions rather difficult.

It is therefore more convenient (albeit somewhat less realistic) to introduce noise by replacing the deterministic threshold by a 'noisy' threshold. More precisely, the neuron may fire in a time step Δt with some probability $\rho\,\Delta t$ even though the membrane potential u has not yet reached the threshold ϑ. The 'hazard' ρ of firing depends on the momentary distance between the membrane potential and the threshold and, possible, also on the slope $\dot{u} = \mathrm{d}u/\mathrm{d}t$, i.e.[1],

$$\rho(t) = f[u(t) - \vartheta, \dot{u}]\,. \tag{8}$$

We call f the 'escape rate'. A suitable function f should vanish for $u \ll \vartheta$ and increase rapidly for $u \gg \vartheta$. It turns out that, at least for standard integrate-and-fire neurons, the diffusive noise model introduced above can be replaced to a high degree of accuracy by an escape rate of the form (Plesser and Gerstner 2000; Herrmann and Gerstner 2001a)

$$f(u - \vartheta, \dot{u}) = a\left(\frac{1}{\tau} + 2\dot{u}\right)\frac{\mathcal{G}(u - \vartheta, \sigma_u)}{\mathrm{Erfc}(\frac{u-\vartheta}{\sigma_u})} \tag{9}$$

where $\mathcal{G}(x, \sigma_u)$ is a normalized Gaussian with standard deviation σ_u and and vanishing mean; $\mathrm{Erfc}(x)$ is the complementary error function; and a is a parameter. Since in an integrate-and-fire model, the variance σ_u^2 of the membrane potential is directly related to the (usually known) variance σ_n^2 of the noise current, a is in fact the only free parameter for a fit of the escape rate (9) to the diffusive noise model. Optimization of a over a large ensemble of periodic and aperiodic stimuli yields $a^{\mathrm{opt}} \approx 1.2$ (H. E. Plesser, private communication). Note that for $u \gg \vartheta$ and fixed $\dot{u}$, the escape rate (9) increases linearly with $u - \vartheta$.

A major advantage of escape rate models is the fact that statistical quantities such as the interval distribution can be calculated analytically. In fact, whatever the choice of the escape rate f, the probability that a neuron fires a spike at time t given that its last spike was at $\hat{t}$ is

$$P_{\mathcal{I}}(t|\hat{t}) = \rho_r(t)\, \exp\left\{-\int_{\hat{t}}^{t} \rho_r(t')\,\mathrm{d}t'\right\} \tag{10}$$

where $\rho_r(t') = f[u_r(t') - \vartheta, \dot{u}_r(t')]$ with $r = t' - \hat{t}$; cf. Eq. (6). For constant input $\mathcal{I}(t) = \mathcal{I}_0$, we get the stationary interval distribution $P_0(t - \hat{t}) \equiv P_{\mathcal{I}_0}(t|\hat{t})$. For arbitrary time-dependent input, we call $P_{\mathcal{I}}(t|\hat{t})$ the input-dependent interval distribution. Eq. (10) is a generalization of standard renewal theory (Cox 1962; Perkel, Gerstein, and Moore 1967) to the case of time-dependent input.

2.2 Specific Model

In this paper we work in the framework of the noisy spike response model defined by Eqs. (6) and (8). It has been shown that with an optimal choice

[1]In addition, we require that the rate vanishes during the first few milliseconds that correspond to the duration of the spike itself

a) b)

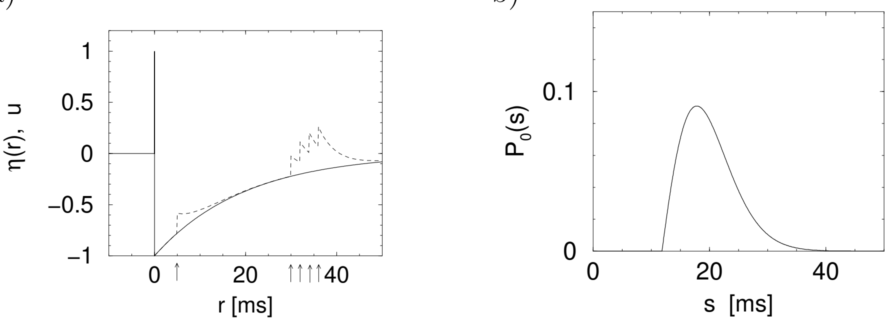

Figure 3: a) Spike Response Model. A neuron that has fired at $r = 0$ (vertical line) passes for $r > 0$ through a phase of refractoriness described by a kernel $\eta(r) = -\eta_0 \exp(-r/\tau_{\rm refr})$ with $\tau_{\rm refr} = 20\,$ms (solid line). Each presynaptic current pulse (arrow) causes a response $\kappa(r,s) \propto \exp(-s/\tau_m)$ with $\tau_m = 4\,$ms. The membrane potential u is the sum of the refractory kernel and the total effect of input current (dashed line). b) Interspike interval distribution $P_0(s)$ for the piecewise linear escape rate (13) with $\beta = 1$. The neuron is stimulated by a constant current $\mathcal{I}_0$ that has been adjusted so that the mean interval is $\langle s \rangle = \int_0^\infty s\, P_0(s)\, {\rm d}s = 20\,$ms.

of the kernels η and κ the spike response model can correctly predict about ninety percent of the firing times of the Hodgkin-Huxley model with time-dependent input (Kistler, Gerstner, and van Hemmen 1997). Thus the spike response model seems to offer a framework that is general enough to capture some essential aspects of neuronal dynamics.

To keep our arguments transparent, we focus in this paper on a much simpler model within the framework of Eq. (6) and take exponential kernels

$$\eta(r) = \delta(r) - \eta_0 \exp\left(-\frac{r}{\tau_{\rm refr}}\right) \mathcal{H}(r) \tag{11}$$

$$\kappa(r,s) = \exp\left(-\frac{s}{\tau_m}\right) \mathcal{H}(r-s)\, \mathcal{H}(s) \tag{12}$$

where $\mathcal{H}(.)$ is the Heaviside step function; cf. Fig. 3a. In Eq. (11) the form of an action potential has been reduced to a Dirac δ function which simply marks the firing time of the spike. The Heaviside function $\mathcal{H}(r-s)$ in Eq. (12) assures that the membrane potential trajectory restarts after each spike with an initial condition $\lim_{r\to 0} u_r(t) = -\eta_0$. Our specific choice of parameters is τ_m=4 ms, $\tau_{\rm refr} = 20\,$ms, and $\eta_0 = 1$; cf. Fig. 3a.

For the escape rate we take the piecewise linear function

$$\rho_r(t) = \beta\, [u_r(t) - \vartheta]_+ = \beta\, [u_r(t) - \vartheta]\; \mathcal{H}\,[u_r(t) - \vartheta]\ . \tag{13}$$

A value of $\beta = 1$ corresponds to a medium noise level; cf. Fig. 3b.

We note that for $\tau_{\text{refr}} = \tau_m \equiv \tau$ we obtain the standard integrate-and-fire neuron as a special case. In fact, taking the derivative of Eq. (5) with exponential kernels η and κ with a common time constant τ yields between two spikes

$$\frac{\mathrm{d}}{\mathrm{d}t} u = -\frac{u}{\tau} + \mathcal{I}(t)\,. \tag{14}$$

After firing, the membrane potential is reset and integration restarts at a value $u_{\text{reset}} = -\eta_0$. For $\beta \to \infty$ firing occurs immediately when $u(t)$ crosses the threshold. In other words, in the limit of $\beta \to \infty$ and $\tau_{\text{refr}} = \tau_m$ we recover the noise-free integrate-and-fire neuron. On the other hand, a choice $\tau_{\text{refr}} \neq \tau_m$ allows us to separate the time scale of refractoriness from the time scale set by the membrane time constant which is not possible in the case of a standard integrate-and-fire model.

3 Population Equations

Instead of studying a single neuron and averaging over a long experiment (or several trials) to measure the reverse correlations, we will consider in this paper a population of N independent neurons and study its behavior in a single (short) trial. Let us denote the spike train of neuron j by $S_j(t) = \sum_f \delta(t - t_j^{(f)})$ where $t_j^{(f)}$ with $f = 1, 2, \ldots$ are the firing times of neuron j. The quantity we are interested in is the population activity

$$A(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t\, N} \sum_{j=1}^{N} \int_t^{t+\Delta t} S_j(t')\mathrm{d}t' = \frac{1}{N} \sum_{j=1}^{N} \delta(t - t_j^{(f)})\,, \tag{15}$$

i.e., the number of spikes that occur in a population of N neurons in a time window $[t, t + \Delta t]$ divided by the number of neurons and the length of the window. Since the population activity has units of a rate, it is also called the population rate (Nykamp and Tranchina 2000) or the space-rate (Maas and Nathschläger 2000).

In order to get a first intuitive understanding of the relation between the population activity and single-neuron dynamics, let us consider a population which is stimulated at $t = 0$ by a stimulating pulse $\mathcal{I}(t)$ of a duration of, say, 2 ms. Neurons in the population that happen to be close to threshold at the moment of input switching will be put across the threshold by the stimulation and will therefore emit an action potential. Other neurons that are in the refractory period or close to rest will stay quiescent. Thus, on a single-neuron level the response to a stimulus can be completely different depending on the internal state of the neuron. By calculating the population activity $A(t)$, we average over the internal state - in very much the same way as a repetition of the stimulus at random intervals would do.

In this section we review the population dynamics of spiking neurons and derive a population equation of the form (Gerstner and van Hemmen 1992; Gerstner 1995; Gerstner 2000)

$$A(t) = \int_{-\infty}^{t} P_{\mathcal{I}}(t|\hat{t})\, A(\hat{t})\, \mathrm{d}\hat{t} \tag{16}$$

where $P_{\mathcal{I}}(t|\hat{t})$ is the input-dependent interval distribution introduced in Eq. (10). In particular, we discuss the relation of this integral equation to the membrane potential density approach of (Abbott and van Vreeswijk 1993; Nykamp and Tranchina 2000; Omurtag, Knight, and Sirovich 2000; Brunel and Hakim 1999).

3.1 Review of Density Methods

Before we turn to the Spike Response Model, we make a short detour and start with the well-known integrate-and-fire model; cf. Eq. (14). Since the state of individual integrate-and-fire neurons is fully characterized by their membrane potential, the state of a large population of integrate-and-fire neurons can be described by a membrane potential density $p(u,t)$. The normalization is $\int_{-\infty}^{\vartheta} p(u,t)\, \mathrm{d}u = 1$, i.e., $\int_{u_0}^{u_1} p(u,t)\, \mathrm{d}u$ is the probability that a randomly chosen neuron has at time t a membrane potential between u_0 and u_1. The dynamics of the membrane potential density can be expressed by the continuity equation (Abbott and van Vreeswijk 1993; Nykamp and Tranchina 2000; Brunel and Hakim 1999; Omurtag, Knight, and Sirovich 2000)

$$\frac{\partial}{\partial t} p(u,t) = -\frac{\partial}{\partial u} J(u,t) + A(t)\, \delta(u - u_r) \tag{17}$$

where $J(u,t)$ is the flux along the voltage coordinate and $A(t)$ is the population activity. The last term in Eq. (17) describes the reset to $u = u_r$ after firing. For neurons with stochastic input $\mathcal{I}_{\rm noise}$ of power σ_n^2, the flux is

$$J(u,t) = p(u,t)\, \dot{u} + \frac{\sigma_n^2}{2\tau} - \frac{\partial}{\partial u} p(u,t) \tag{18}$$

where $\dot{u} = (-u/\tau) + \mathcal{I}(t)$ is the velocity of the flow. The second term in (18) arises due to the noise component $\mathcal{I}_{\rm noise}$ and is called the 'diffusion term'. The population activity $A(t)$ is given by the flux through the threshold; cf.

Fig. 4a. Since the threshold acts as an absorbing boundary (i.e. $p(u,t) = 0$ for $u \geq \vartheta$), we have

$$A(t) = J(\vartheta, t) = \frac{\sigma_n^2}{2\tau} - \frac{\partial}{\partial u} p(\vartheta, t) \tag{19}$$

For stationary input, Eqs. (17) – (19) can be solved analytically (Riccardi 1977); for arbitrary time-dependent input they must be integrated numerically (Nykamp and Tranchina 2000; Omurtag, Knight, and Sirovich 2000). This ends our short discussion of integrate-and-fire neurons with diffusive noise.

We now return to neurons of the Spike Response Model type. The internal state of a SRM-neuron can be characterized by the time $r = t - \hat{t}$ since its last output spike. We therefore introduce a 'refractory density' $q(r,t)$ with normalization

$$\int_0^\infty q(r,t)\, \mathrm{d}r = 1\,. \tag{20}$$

i.e., $\int_{r_0}^{r_1} q(r,t)\, \mathrm{d}r$ is the probability that a randomly chosen neuron has at time t a refractory variable between r_0 and r_1. In analogy to Eq. (17), the continuity equation for SRM neurons is given by (Gerstner and van Hemmen 1992)

$$\frac{\partial}{\partial t} q(r,t) = -\frac{\partial}{\partial r} J(r,t) - \rho_r(t)\, q(r,t) + A(t)\, \delta(r)\,. \tag{21}$$

The flux along the r-axis is

$$J(r,t) = q(r,t)\, \dot{r} = q(r,t) \tag{22}$$

since the velocity $\dot{r}$ of the refractory variable is constant and equal to one. The last term in Eq. (21) describes the 'reset' of the refractory variable to $r = 0$. Note that, in contrast to diffusive noise, escape noise does not enter into the flux equation (22) but directly as a 'sink' term $-\rho_r(t)\, q(r,t)$ on the right-hand side of Eq. (21). The magnitude of the 'sink' can be calculated from the escape rate

$$\rho_r(t) = f[u_r(t) - \vartheta, \dot{u}_r(t)] \tag{23}$$

with u_r given by (6). The population activity is

$$A(t) = \int_0^\infty \rho_r(t)\, q(r,t)\, \mathrm{d}r\,, \tag{24}$$

i.e., neurons with refractory variable r contribute with rate $\rho_r(t)$ to the activity $A(t)$. A schematic graph of the refractory dynamics is given in Fig. 4b.

We show in Appendix B that, for integrate-and-fire neurons with escape noise (and similarly for noiseless neurons), the continuity equation (17) for

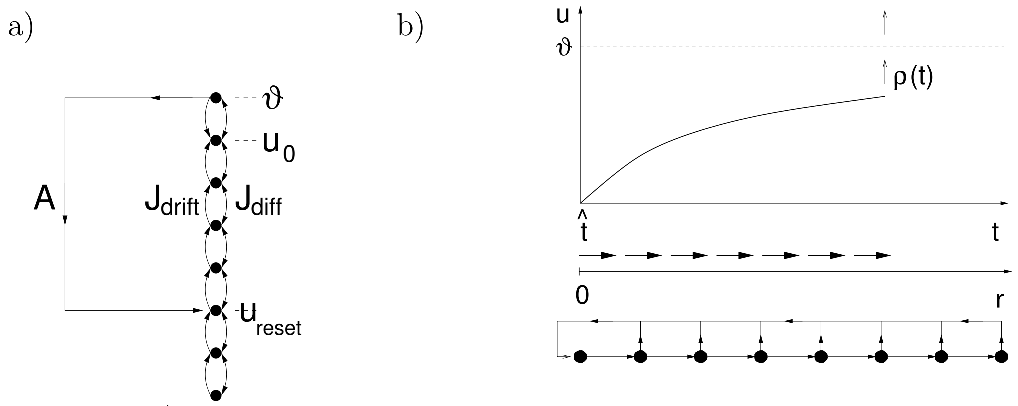

Figure 4: a) State transitions in a discretized implementation of the membrane potential density equation for stimulation with constant current $\mathcal{I}_0$ and diffusive noise. The flux $J_{\rm drift}$ drives the membrane potential towards u_0, while stochastic spike arrival (diffusive noise) leads to a flux $J_{\rm diff}$ upwards (excitation) or downwards (inhibition). The flux across threshold, i.e., the activity A, is reinjected at the reset potential u_r (schematic). b) Population dynamics of SRM neurons with escape noise. Top: Membrane potential of a neuron that has fired its last spike at $\hat{t}$. The membrane potential trajectory $u(t)$ ends when the neuron fires. The instantaneous firing rate is $\rho_r(t) = f[u_r(t) - \vartheta]$. Middle: The flow has constant speed along the r-axis. The trajectory of the refractory variable $r(t)$ ends when the neuron fires and restarts immediately at $r = 0$. Bottom: Discretized state diagram; the activity A consists of the firings of all neurons with different refractory variables r. The activity A is reinjected at $r = 0$.

the membrane potential density $p(u,t)$ is *equivalent* to the continuity equation (21) for the refractory density $q(r,t)$. In fact, the transition from (17) to (21) is given by a simple coordinate transformation. The advantage of the formulation with refractory densities is that the partial differential equation (21) can be solved analytically. Since the flux has the simple form (22), we may switch to a moving coordinate frame. In the new frame, the partial differential equation has been reduced to an ordinary differential equation that can be integrated; cf. Appendix B. The result is (Gerstner and van Hemmen 1992; Gerstner and van Hemmen 1994)

$$A(t) = \int_{-\infty}^{t} \rho_{t-\hat{t}}(t)\, \exp\left[-\int_{\hat{t}}^{t} \rho_{t'-\hat{t}}(t')\, \mathrm{d}t'\right] A(\hat{t})\, \mathrm{d}\hat{t} \tag{25}$$

which has the form of the population equation (16) with an input-dependent interval distribution given by Eq. (10). The correct normalization of the population activity A follows from Eq. (20) and is given by

$$1 = \int_{-\infty}^{t} \exp\left[-\int_{\hat{t}}^{t} \rho_{t'-\hat{t}}(t')\, \mathrm{d}t'\right] A(\hat{t})\, \mathrm{d}\hat{t} \tag{26}$$

3.2 Linearization

In this subsection, we linearize the population equation (16) about a constant activity level A_0. The activity A_0 can be the result of stimulation by a constant current $\mathcal{I}_0 > 0$ or else due to spontaneous activity of the neurons ($\mathcal{I}_0 = 0$). Let us now assume that, in addition to the bias current $\mathcal{I}_0$, we also apply a time-dependent input $\Delta\mathcal{I}(t)$. The population activity will then respond by a change $\Delta A(t)$, i.e.,

$$A(t) = A_0 + \Delta A(t)\,. \tag{27}$$

Linearization of the population equation (16) yields (Gerstner 2000)

$$\Delta A(t) = \int_{-\infty}^{t} P_0(t-\hat{t})\, \Delta A(\hat{t})\, \mathrm{d}\hat{t} + A_0 \frac{\mathrm{d}}{\mathrm{d}t} \int_0^{\infty} \mathcal{L}(x) \int_0^{\infty} \kappa_0(s)\, \Delta\mathcal{I}(t-x-s)\, \mathrm{d}s\, \mathrm{d}x\,. \tag{28}$$

The first term on the right-hand side of Eq. (28) describes the after-effect of previous perturbations $\Delta A(\hat{t})$ in the past ($\hat{t} < t$). Perturbations that have occurred about one inter-spike interval earlier 'reappear' at time t, smoothed out by the interval distribution $P_0(t)$. If there have been no perturbations in the past, the first term vanishes. The second term in Eq. (28) is a linear filter that describes the immediate effect of recent input variations $\Delta\mathcal{I}(t')$.

Three aspects are important for an understanding of the filter properties generated by the second term in Eq. (28). Firstly, the integral $\Delta h(t) =$

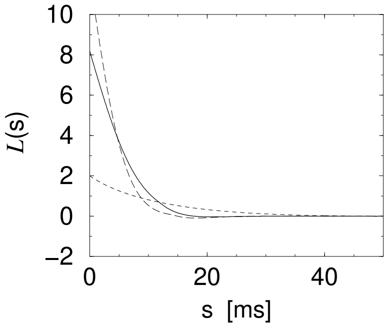

Figure 5: Filter $\mathcal{L}(s)$ for a SRM neuron with exponential kernels and a piecewise linear escape rate (13) at a noise level of $\beta = 1$ (solid line), $\beta = 0.1$ (dotted line), and $\beta = 2$ (long-dashed line). The bias current $\mathcal{I}_0$ has been readjusted so that the mean interval is always 20 ms (mean firing rate $\nu_0^{\text{out}} = 50\,\text{Hz}$). Parameters: membrane time constant $\tau_m = 4\,\text{ms}$ and refractory time constant $\tau_{\text{refr}} = 20\,\text{ms}$.

$\int_0^\infty \kappa_0(s)\, \Delta\mathcal{I}(t-s)\, \mathrm{d}s$ can be interpreted as the contribution to the membrane potential caused by the input $\Delta\mathcal{I}$. We call Δh the *input potential*.

Secondly, the input potential is convolved with a filter $\mathcal{L}(x)$. For escape rates $f(u)$ (i.e., escape rates that depend on the voltage u but not on the slope $\dot{u}$), the filter is (Gerstner 2000)

$$\mathcal{L}(x) = \int_x^\infty f'[u(\xi - x) - \vartheta]\, S_0(\xi)\, d\xi \;-\; S_0(x) \int_0^x e^{-\xi/\tau}\, f'[u(\xi) - \vartheta]\, d\xi\,, \quad (29)$$

with

$$S_0(s) = \exp\left\{ - \int_0^s f[u(\xi) - \vartheta]\, \mathrm{d}\xi \right\}\,. \quad (30)$$

Here $u(\xi)$ is the membrane potential trajectory of a neuron with constant stimulus $\mathcal{I}_0$. As we can see from Fig. 5, for high noise the filter $\mathcal{L}(x)$ exhibits low-pass characteristics (i.e., the filter extends over a broad time window of averaging), whereas for low noise it becomes sharper. In the limit of no noise ($\beta \to \infty$), it approaches a δ function.

Thirdly, Eq. (28) combines the integral over the filter $\mathcal{L}$ with a temporal derivative. For high noise, the derivative 'cancels' the convolution by the low-pass filter $\mathcal{L}$. The immediate response ΔA to a change in the input is therefore proportional to the input potential,

$$\text{high noise:} \quad \Delta A(t) \propto \quad \Delta h(t) = \int_0^\infty \kappa(s)\, \Delta\mathcal{I}(t-s)\, \mathrm{d}s\,. \quad (31)$$

For low noise, however, the immediate response becomes proportional to the derivative of the input potential

$$\text{low noise}: \quad \Delta A(t) \propto \frac{\mathrm{d}}{\mathrm{d}t}\Delta h(t) = \frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \kappa(s)\,\Delta\mathcal{I}(t-s)\,\mathrm{d}s\,. \tag{32}$$

An advantage of the analytical expressions (28) and (29) is that we can not only study the extreme cases but also intermediate noise levels. An illustration is given in the following section where we determine reverse and cross-correlation functions.

4 Results

In this section, we apply the linearized population equation to the problem of neuronal coding and calculate reverse and cross-correlation functions. The Fourier transform of Eq. (28) yields

$$\hat{A}(\omega) = i\omega\,\frac{A_0\,\hat{\mathcal{L}}(\omega)\,\hat{\kappa}(\omega)}{1-\hat{P}(\omega)}\,\hat{\mathcal{I}}(\omega) = \hat{G}(\omega)\,\hat{\mathcal{I}}(\omega)\,. \tag{33}$$

with a filter

$$\hat{G}(\omega) = \frac{A_0\,\hat{\mathcal{L}}(\omega)\,\hat{\kappa}(\omega)}{1-\hat{P}(\omega)\,.} \tag{34}$$

The inverse Fourier transform of Eq. (33) yields

$$A(t) = A_0 + \int_0^\infty G(s)\,\Delta\mathcal{I}(t-s)\,\mathrm{d}s \tag{35}$$

where A_0 is the population activity for stimulation with constant bias current $\mathcal{I}_0$. The filter G is

$$G(s) = \frac{1}{2\pi}\int_{-\infty}^\infty \hat{G}(\omega)\,e^{+i\omega s}\mathrm{d}\omega\,. \tag{36}$$

Equation (35) has the form of a linear rate model; cf. Eq.(2). We can therefore repeat the arguments of Appendix A to show that

$$C^{\mathrm{rev}}(s) = \frac{\sigma^2}{A_0}\,G(s)\,. \tag{37}$$

The central point of our approach is that the filter $G(s)$ can be predicted via Eq. (34) from the shape of the interval distribution $P_0(s)$, the response kernel $\kappa_0(s)$ and the filter $\mathcal{L}(s)$. Eqs. (37) and (34) establish therefore a relation between neuronal parameters and the reverse correlation function $C^{\mathrm{rev}}(s)$.

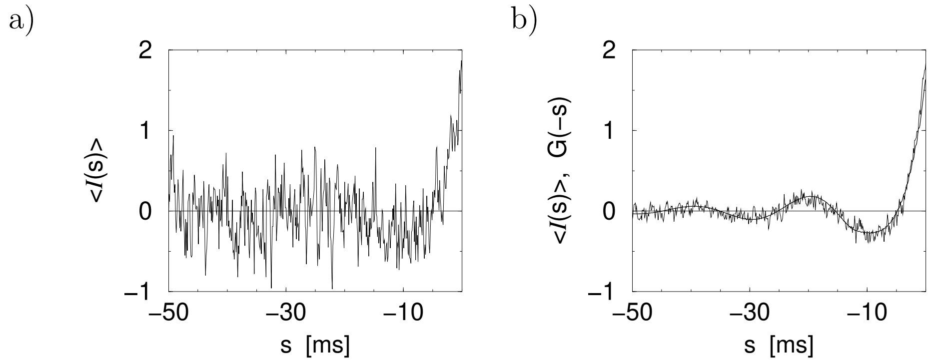

Figure 6: Reverse correlations. A SRM_0 neuron is stimulated by a constant bias current plus a stochastic input current. Each time an output spike is triggered, the time course of the input current is recorded. a) Average input $\langle \mathcal{I}(t - t^{(f)}) \rangle_{t^{(f)}}$ as a function of $s = t - t^{(f)}$ averaged over 1000 output spikes $f = 1, \ldots, 1000$. b) The same, but averaged over 25 000 spikes. The simulation result is compared with the time-reversed linear filter $G(-s)$ predicted from the theory (smooth line). Noise model: linear escape rate with $\beta = 1$.

4.1 Reverse Correlations

We consider the SRM neuron defined in Section 2.2. The response kernels are exponential with a time constant of $\tau_m = 4\,\mathrm{ms}$ for the kernel κ and $\tau_{\mathrm{refr}} = 20\,\mathrm{ms}$ for the refractory kernel η. The neuron is driven by a current $\mathcal{I}(t) = \mathcal{I}_0 + \Delta\mathcal{I}(t)$. The bias current $\mathcal{I}_0$ was adjusted so that the neuron fires at a mean rate of 50 Hz. The noise current was generated by the following procedure. In every time step of 0.1 ms we apply with a probability of 0.5 an input pulse of amplitude ± 1, i.e., the input has white-noise characteristics. To estimate the reverse correlation function, we build up a histogram of the average input $\langle \mathcal{I}(t - t^{(f)}) \rangle_{t^{(f)}}$ preceding a spike $t^{(f)}$. We see from Fig. 6a that the main characteristics of the reverse correlation function are already visible after 1000 spikes. After an average over 25 000 spikes, the time course is much cleaner and reproduces to a high degree of accuracy the time course of the time-reversed filter $G(-s)$ predicted by the theory. The oscillation with a period of about 20 ms reflects the intrinsic firing period of the neuron. This is shown more clearly in Fig. 7 where the bias current has been changed so that the mean interval is either shorter ($\langle s \rangle$ =10 ms) or longer ($\langle s \rangle$ =30 ms). The oscillations of the reverse correlation filter $G(-s)$ change accordingly. An increase in the noise level suppresses the oscillations; cf. Fig. 8.

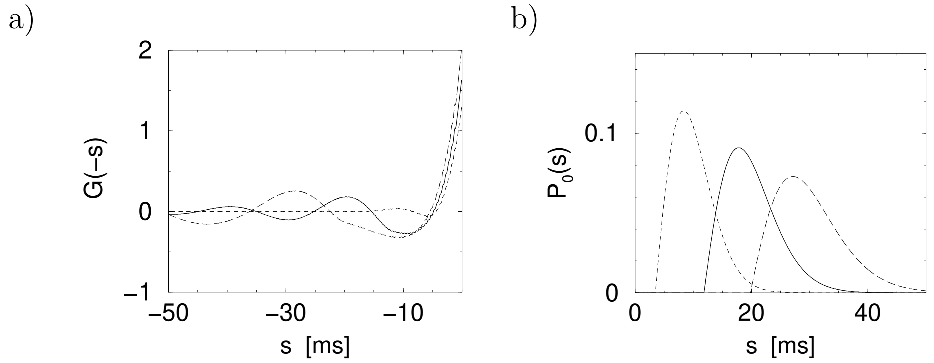

Figure 7: Reverse correlations (a) and interval distribution (b) for the same neuron as in Fig. 6, except that the bias current has been changed so that the mean interval is 10 (dotted line), 20 (solid line), or 30 milliseconds (long-dashed line). Linear escape rate with $\beta = 1$

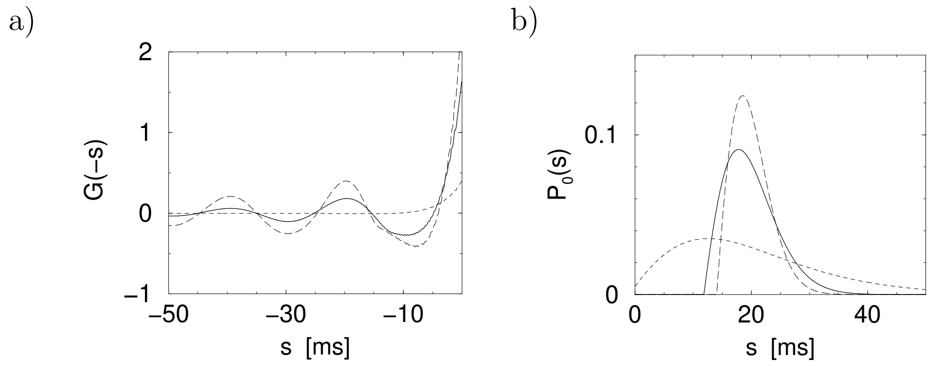

Figure 8: Reverse correlations (a) and interval distribution (b) for the same neuron as in Fig. 6, except that the noise level has been changed from $\beta = 1$ (solid line) to $\beta = 0.1$ (dotted line) or $\beta = 2$ (long-dashed line).

4.2 Cross-correlations

In the preceding subsection, the input $\Delta\mathcal{I}(t)$ was a random current with white-noise characteristics; cf. Eq. (3). In this subsection, we focus on a scenario that is somewhat closer to real synaptic input. To keep things simple,we model presynaptic spike arrival as *current* injection (rather than conductance change). More precisely, we assume that a presynaptic spike that arrives at time t^{in} generates an exponential current pulse

$$\mathcal{I}^{\text{syn}}(t) = \frac{q}{\tau_{\text{syn}}}\, \exp\left(-\frac{t-t^{\text{in}}}{\tau_{\text{syn}}}\right) \mathcal{H}(t-t^{\text{in}})\,, \tag{38}$$

where $\mathcal{H}(.)$ is the Heaviside step function. The postsynaptic potential (PSP) generated by the pulse is

$$\epsilon_0(s) = \int_0^\infty \kappa_0(s')\, \mathcal{I}^{\text{syn}}(s-s')\, \mathrm{d}s'\,. \tag{39}$$

Let us now suppose that we study a neuron that is spontaneously active at low rate. At $t = t^{\text{in}}$ a presynaptic spike arrives and evokes a PSP. Depending on the internal state of the neuron the extra PSP may or may not be sufficient to drive the membrane potential across threshold. Intuitively, the cross-correlation function $C^{\text{cross}}(t|t^{\text{in}})$ describes the probability of generating an output spike given the input. To make our intuition more precise, we take a population approach. Neurons in the population are firing spontaneously at a rate A_0. At time t^{in} all neurons receive a presynaptic input spike which generates a modulation $\Delta A(t)$ of the population activity. The change of the neuronal firing rate caused by one presynaptic input spike is $C^{\text{cross}}(t|t^{\text{in}}) = \Delta A(t)$. Using the linearized population equations, we can estimate the cross-correlation

$$\begin{aligned} C^{\text{cross}}(t|t^{\text{in}}) &= \Delta A(t) \\ &\approx \int_{-\infty}^{t} P_0(t-\hat{t})\, \Delta A(\hat{t})\, \mathrm{d}\hat{t} + A_0 \frac{\mathrm{d}}{\mathrm{d}t}\int_0^\infty \mathcal{L}(x)\, \epsilon_0(t-x)\, \mathrm{d}x \end{aligned} \tag{40}$$

where P_0 is the interval distribution during spontaneous activity and A_0 the spontaneous rate. From Eqs. (31) and (32) we conclude that the main peak of the cross-correlation is

$$C^{\text{cross}}(t^{\text{in}}+s|t^{\text{in}}) = \begin{cases} \epsilon_0(s) & \text{for high noise} \\ \frac{\mathrm{d}}{\mathrm{d}s}\epsilon_0(s) & \text{for low noise.} \end{cases} \tag{41}$$

Thus for high noise the cross-correlation is proportional to the PSP while for low noise it is proportional to its derivative (Moore, Segundo, Perkel, and

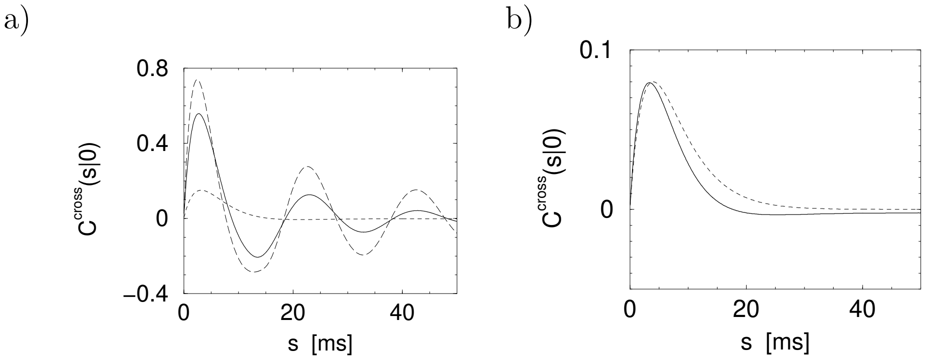

Figure 9: Cross-correlations. The input consists of a synaptic current pulse at $s = 0$ with time constant $\tau_{\rm syn} = 4\,$ms. a) Same neuron model as in Fig. 6, with noise levels $\beta = 1$ (solid line), $\beta = 0.1$ (dotted line), or $\beta = 2$ (long-dashed line). For medium and low noise ($\beta = 1$ or 2) the cross-correlations exhibits oscillations with the intrinsic firing period of 20 ms. b) In the case of high noise ($\beta = 0.05$, solid line) the cross-correlation is similar to the time course of the postsynaptic potential (dotted line). The postsynaptic potential has been scaled to approximately the same amplitude as $C^{\rm cross}$.

Levitan 1970; Kirkwood and Sears 1978; Fetz and Gustafsson 1983; Abeles 1991; Poliakov, Powers, and Binder 1997; Herrmann and Gerstner 2001a). For intermediate noise levels we calculate the cross-correlation from

$$C^{\rm cross}(t^{\rm in} + s|t^{\rm in}) = \int_0^\infty G(s')\,\frac{q}{\tau_{\rm syn}}\,\exp\left(-\frac{s-s'}{\tau_{\rm syn}}\right)\,{\rm d}s'\,. \tag{42}$$

The result is shown in Fig. 9. For high noise the time course of the cross-correlation in indeed comparable to that of the PSP; cf. Fig. 9b. For low noise, however, the sharp initial rise is followed by a trough and a second peak about one interspike-interval later. The overall pattern is dominated by oscillations; Fig. 9a.

5 Discussion

5.1 Correlations in Spike-Time Dependent Plasticity

Correlations between pre- and postsynaptic neurons are of eminent importance for synaptic plasticity. In standard Hebbian learning, the weight w_{ij} from a presynaptic neuron j to a postsynaptic neuron i changes, if both neurons are active 'at the same time'. Hebb's original formulation is slightly more precise and makes explicit reference to a causal relation between preand postsynaptic firing. It states that a change should occur if the 'presynaptic neuron contributes in firing' the postsynaptic one (Hebb 1949). Thus learning should occur if there are cross-correlations between pre- and postsynaptic activity.

More generally, we call a plasticity rule Hebbian if

(i) it only depends on *locally* availably information such as pre- and postsynaptic spike times and the current weight value w_{ij} and

(ii) it depends on the *correlations* between pre- and postsynaptic firings (i.e., plasticity is not due to pre- or postsynaptic activity alone).

An example of a Hebb-type learning rule is (Gerstner, Kempter, van Hemmen, and Wagner 1996; Kistler and van Hemmen 2000; Kempter, Gerstner, and van Hemmen 1999; Roberts and Bell 2000; Song, Miller, and Abbott 2000; Senn, Tsodyks, and Markram 2001)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) &= a_0(w_{ij}) + a_1^{\mathrm{in}}(w_{ij})\, S_j^{\mathrm{in}}(t) + a_1^{\mathrm{out}}(w_{ij})\, S_i^{\mathrm{out}}(t) \\ &+ a_2^{\mathrm{corr}}(w_{ij})\, S_j^{\mathrm{in}}(t) \int_{-\infty}^{t} \mathrm{d}t'\, W(t-t')\, S_i^{\mathrm{out}}(t') \\ &+ a_2^{\mathrm{corr}}(w_{ij})\, S_i^{\mathrm{out}}(t) \int_{-\infty}^{t} \mathrm{d}t'\, W(-t+t')\, S_j^{\mathrm{in}}(t')\ ,\end{aligned} \tag{43}$$

where $S_j^{\mathrm{in}}(t) = \sum_f \delta(t - t_j^{(f)})$ is the pre- and $S_i^{\mathrm{out}}(t) = \sum_f \delta(t - t_i^{(f)})$ the postsynaptic spike train. The contributions $a_0, a_1^{\mathrm{in}}, a_1^{\mathrm{out}}$ describe the change if there are no spikes at all, presynaptic spikes only, or postsynaptic spikes only. These are called the 'non-Hebbian' terms, since they describe adaptation processes; some or all of these terms may vanish - they are included in Eq. (43) for the sake of generality. All of the above parameters as well as a_2^{corr} may depend on the current weight value w_{ij}. The correlations between pre- and postsynaptic spikes are taken care off by the learning window $W(t_j^{(f)} - t_i^{(f)})$ (Gerstner, Ritz, and van Hemmen 1993; Gerstner, Kempter, van Hemmen, and Wagner 1996). The learning window can have two phases depending on the relative timing of pre- and postsynaptic neurons (Bell, Han, Sugawara, and Grant 1997; Markram, Lübke, Frotscher, and Sakmann 1997; Zhang, Tao, Holt, W.A.Harris, and Poo 1998; Debanne, Gähwiler, and Thompson 1998; Bi and Poo 1998). Since such two-phase learning windows are useful to detect temporal variations in neuronal firing patterns, they have been postulated on theoretical grounds (Gerstner, Kempter, van Hemmen, and Wagner 1996).

To see more clearly how correlations enter into the learning equation (43), we assume that learning is slow compared to typical inter-spike intervals. The expected evolution of weights on the slow time scale of learning is[2]

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = a_0 + a_1^{\mathrm{in}} \langle S_j^{\mathrm{in}}(t) \rangle + a_1^{\mathrm{out}} \langle {S^{\mathrm{out}}}_i(t) \rangle + \int_{-\infty}^{\infty} \mathrm{d}s \, a_2^{\mathrm{corr}} \, W(s) \, \langle S_j^{\mathrm{in}}(t) \, S_i^{\mathrm{out}}(t-s) \rangle \; . \tag{44}$$

The term $\langle S_j^{\mathrm{in}}(t) \rangle$ can be identified with the presynaptic rate $\nu_j^{\mathrm{in}}(t)$ and similarly $\langle {S^{\mathrm{out}}}_i(t) \rangle$ with the postsynaptic rate $\nu_i^{\mathrm{out}}(t)$. The integral over the learning window is driven by correlations between pre- and postsynaptic spikes. If the presynaptic firing is described by a Poisson process with constant rate, the correlations can be further separated into correlations between pre- and postsynaptic *rates* and additional spike-spike correlations. The final result is (Kempter, Gerstner, and van Hemmen 1999)

$$\frac{\mathrm{d}}{\mathrm{d}t} w_{ij}(t) = a_0 + a_1^{\mathrm{in}} \, \nu_j^{\mathrm{in}} + a_1^{\mathrm{out}} \, \nu_i^{\mathrm{out}} + a_2^{\mathrm{corr}} \, \nu_j^{\mathrm{in}} \, \nu^{\mathrm{out}} \int_{-\infty}^{\infty} W(s) \, \mathrm{d}s + a_2^{\mathrm{corr}} \, \nu_j^{\mathrm{in}} \int_{-\infty}^{\infty} W(-s) \, C_{ij}^{\mathrm{cross}}(t+s|t) \, \mathrm{d}s \tag{45}$$

Thus the cross-correlation $C_{ij}^{\mathrm{cross}}(t+s|t)$ enters as expected into the learning equation. Previously, the cross-correlation function $C^{\mathrm{cross}}(t+s|t)$ has only been evaluated for a particularly simple stochastically firing neuron model (Kempter, Gerstner, and van Hemmen 1999). For this simple model it was found that the cross-correlations follow the time course of the postsynaptic potential. The results in this paper allow to extend the analysis of Hebbian plasticity to more realistic spiking neuron models. As we have seen, the cross-correlation is a linearly filtered version of the postsynaptic potential; the time course of the cross-correlation function depends on the noise level and the mean firing rate of the neuron.

5.2 Extensions

In this paper we calculated reverse and cross-correlations for spiking neuron models. While the framework of the Spike Response Model introduced in Section 2 is rather general, we illustrated the results so far only for one specific, somewhat simplified neuron model with piecewise linear escape rate. Extensions could come in several directions. First, we could redo the analysis for more interesting realizations of the Spike Response Model (Kistler,

[2] for the sake of notational simplicity, we suppress the w_{ij} dependence of parameters

Gerstner, and van Hemmen 1997; Herrmann and Gerstner 2001b). Second, if the kernels η and κ are fitted to real neurons, we could try to explain experimental cross-correlation (Poliakov, Powers, and Binder 1997) or reverse correlation measurements. Third, so far the analysis is restricted to independent neurons. The results should be extended to cross-correlations in connected networks.

Appendix A: Reverse correlations as a linear filter

In this Appendix we want to show that the reverse correlation function $C^{\rm rev}(s)$ is proportional to the filter $G(s)$ of the linear rate model

$$\nu^{\rm out}(t) = \nu_0^{\rm out} + \int_0^\infty G(s)\, \Delta\mathcal{I}(t-s)\, {\rm d}s\,. \tag{46}$$

Here, $\nu_0^{\rm out}$ is the mean output rate and $\Delta\mathcal{I}(t)$ is the fluctuating part of the input. Eq. (46) describes the relation between a known (deterministic) input $\Delta\mathcal{I}(t)$ and the rate. We adopt a statistical point of view and assume that the input $\Delta\mathcal{I}(t)$ is drawn from a statistical ensemble of stimuli with mean $\langle\Delta\mathcal{I}(t)\rangle = 0\,$. Angular brackets denote averaging over the input ensemble or, equivalently, over an infinite input sequence. We are interested in the correlation

$$C_{\nu I}(s) = {\rm lim}_{T\to\infty} \frac{1}{T}\int_0^T \nu^{\rm out}(t+s)\, \Delta\mathcal{I}(t)\, {\rm d}t = \langle \nu^{\rm out}(t+s)\, \Delta\mathcal{I}(t)\rangle \tag{47}$$

between input $\Delta\mathcal{I}$ and activity ΔA. With the linear rate equation (46) we find

$$C_{\nu I}(s) = \int_0^\infty G(s) \langle \Delta\mathcal{I}(t+s-s')\, \Delta\mathcal{I}(t)\rangle {\rm d}s'\,. \tag{48}$$

where we have used $\nu_0^{\rm out}\,\langle\Delta\mathcal{I}(t)\rangle = 0$. Thus the correlation function depends on the filter $G(s)$ as well as on the autocorrelation $\langle\Delta\mathcal{I}(t')\,\Delta\mathcal{I}(t)\rangle$ of the input ensemble.

For the sake of simplicity, we assume that the input consists of white noise i.e., the input has an autocorrelation

$$\langle\Delta\mathcal{I}(t')\,\Delta\mathcal{I}(t)\rangle = \sigma^2\, \delta(t'-t)\,. \tag{49}$$

In this case Eq. (48) reduces to

$$C_{\nu I}(s) = \sigma^2\, G(s)\,, \tag{50}$$

i.e., the correlation function $C_{\nu I}$ is proportional to the filter $G(s)$.

In order to relate the correlation function $C_{\nu I}$ to the reverse correlation $C^{\rm rev}$, we think of the rate $\nu^{\rm out}$ as the intensity of an inhomogeneous Poisson process that generates spike events $t^{(f)}$. The rate can be estimated if we average over a large number $1 \le i \le N$ of repetitions of the experiment (or else over a large number of independent neurons)

$$\nu^{\rm out}(t) = \frac{1}{N} \sum_{i=1}^{N} \sum_f \delta(t - t_i^{(f)}) \,. \tag{51}$$

The correlation function (47) is therefore

$$\begin{aligned} C_{\nu I}(s) &= \lim_{T\to\infty} \frac{1}{T} \left[\frac{1}{N} \sum_{i=1}^{N} \sum_f \Delta\mathcal{I}(t_i^{(f)} - s) \right] \\ &= \nu_0^{\rm out} \, \langle \mathcal{I}(t^{(f)} - s) \rangle_{t^{(f)}} = \nu_0^{\rm out} \, C^{\rm rev}(s) \,. \end{aligned} \tag{52}$$

For the second equality sign we have used that the expected number of spike events in an interval of length T is $\nu_0^{\rm out}\,T$ where $\nu_0^{\rm out}$ is the mean rate. The last equality follows from the definition of the reverse correlation function; cf. Eq. (1). The combination of Eqs. (50) and (52) yields

$$C^{\rm rev}(s) = \frac{\sigma^2}{\nu_0^{\rm out}} \, G(s) \,. \tag{53}$$

Thus the reverse correlation function is proportional to the linear filter G. In Section 4 we calculate the filter $G(s)$ for *spiking* neurons.

Appendix B1: Integrating the Partial Differential Equation

In this appendix we integrate Eq. (21) that describes the dynamics of refractory densities $q(r,t)$. All neurons that have fired together at time $\hat{t}$ form a group that moves along the r-axis at constant speed; cf. Fig. 4b. To solve Eq. (21) we turn to a reference frame that moves along with the group. Formally we replace the variable r by $x = t - r$ and define a new density

$$Q(x,t) = q(t-x,t) \qquad \text{for } x \le t \tag{54}$$

and $Q(x,t) = 0$ for $x > t$. The temporal derivative of Q for $x < t$ is

$$\frac{\partial}{\partial t} Q(x,t) = \frac{\partial}{\partial r} q(r,t)|_{r=t-x} \, \frac{{\rm d}r}{{\rm d}t} + \frac{\partial}{\partial t} q(r,t)|_{r=t-x} \tag{55}$$

with $\mathrm{d}r/\mathrm{d}t = 1$. We insert Eq. (21) on the right-hand side of (55). For $x \leq t$, the result is

$$\frac{\partial}{\partial t} Q(x,t) = -\rho_{t-x}(t)\, Q(x,t) \; + A(t)\, \delta(t-x) \quad \text{for } x \leq t \tag{56}$$

and $Q(x,t) \equiv 0$ for $x > t$. Thus the partial differential equation (21) has been transformed into an ordinary differential equation. The solution of Eq. (56) is (starting the integration at $t_0 > x$)

$$Q(x,t) = Q(x,t_0)\, \exp\left[-\int_{t_0}^{t} \rho_{t'-x}(t')\, \mathrm{d}t'\right] \tag{57}$$

where $Q(x,t_0)$ is the initial condition of the integration. How can we interpret Eq. (57)? The variable $x = t - r$ is in fact the last firing time $\hat{t}$. Evaluating (57) in the limit of $t_0 \to x$ yields (with $x = \hat{t}$)

$$Q(\hat{t},t) = A(\hat{t})\, \exp\left[-\int_{\hat{t}}^{t} \rho_{t'-\hat{t}}(t')\, \mathrm{d}t'\right] . \tag{58}$$

Here we have used $\lim_{t_0 \to \hat{t}} Q(\hat{t}, t_0) = A(\hat{t})$ which follows from the δ term in Eq. (56). On the other hand, from (24) we have

$$A(t) = \int_0^{\infty} \rho_{t-\hat{t}}(t)\, Q(\hat{t},t)\, \mathrm{d}\hat{t} . \tag{59}$$

If we insert (58) into (59), we find

$$A(t) = \int_{-\infty}^{t} \rho_{t-\hat{t}}(t)\, \exp\left[-\int_{\hat{t}}^{t} \rho_{t'-\hat{t}}(t')\, \mathrm{d}t'\right] A(\hat{t})\, \mathrm{d}\hat{t} \tag{60}$$

which is the population equation (16) for escape noise. If we insert Eq. (58) into the normalization condition for the Q-variable, $1 = \int_{-\infty}^{t} Q(\hat{t},t)\, \mathrm{d}\hat{t}$, we arrive at a normalization equation for the activity:

$$1 = \int_{-\infty}^{t} \exp\left[-\int_{\hat{t}}^{t} \rho_{t'-\hat{t}}(t')\, \mathrm{d}t'\right] A(\hat{t})\, \mathrm{d}\hat{t} . \tag{61}$$

Equation (61) can be used as a starting point for a theory of population dynamics (Gerstner 2000). In fact, Eq. (60) can be derived from (61) by taking the temporal derivative.

Appendix B2: From membrane potential densities to refractory densities

In this appendix we study a population of integrate-and-fire neurons with escape noise. We want to show that we can use the membrane potential trajectory of the integrate-and-fire model

$$u(t) = \eta_0 \exp\left(-\frac{t-\hat{t}}{\tau}\right) + \int_{\hat{t}}^{t} \exp\left(-\frac{t-t'}{\tau}\right) \mathcal{I}(t')\,\mathrm{d}t' \tag{62}$$

to define a transformation from voltage to refractory variables: $u \longrightarrow r$ with $r = t - \hat{t}$. It turns out that the final equations are even simpler if we take $\hat{t}$ instead of r as our new variable — and we therefore consider the transformation $u \longrightarrow \hat{t}$.

Before we start, we calculate the derivatives of Eq. (62). The derivative with respect to t yields $\partial u/\partial t = [-u/\tau] + \mathcal{I}(t)$ as expected for integrate-and-fire neurons. The derivative with respect to $\hat{t}$ is

$$\frac{\partial u}{\partial \hat{t}} = \left[\frac{\eta_0}{\tau} - \mathcal{I}(\hat{t})\right] \exp\left(-\frac{t-\hat{t}}{\tau}\right) = F(t,\hat{t}) \tag{63}$$

where the function F is defined by Eq. (63).

The densities in the variable $\hat{t}$ are denoted as $Q(\hat{t},t)$. Since the density Q is normalized to one, the coordinate transformation from $p(u,t)$ to $Q(\hat{t},t)$ must respect $Q(\hat{t},t)\,\mathrm{d}\hat{t} = p(u,t)\,\mathrm{d}u$. This yields

$$Q(\hat{t},t) = p(u,t)\,F(t,\hat{t}) \tag{64}$$

where u is taken as a function of t and $\hat{t}$; cf. Eq. (62). We now want to show that the differential equation for the density $Q(\hat{t},t)$ that we derived in (56)

$$\frac{\partial}{\partial t} Q(\hat{t},t) = -\rho_{t-\hat{t}}(t)\,Q(\hat{t},t) \quad \text{for } \hat{t} < t\,. \tag{65}$$

is equivalent to the continuity equation (17) for the membrane potential densities. If we insert Eq. (64) into Eq. (65) we find

$$\frac{\partial p}{\partial u}\frac{\partial u}{\partial t} F + \frac{\partial p}{\partial t} F + p\,\frac{\partial F}{\partial t} = -\rho\,p\,F\,. \tag{66}$$

For the integrate-and-fire neuron we have $\partial F/\partial t = -F/\tau$. Furthermore for $R\mathcal{I}(t) > \eta_0$ we have $F \neq 0$. Thus we can divide (66) by F and rewrite Eq. (66) in the form

$$\frac{\partial p(u,t)}{\partial t} = -\frac{\partial}{\partial u}\left[\dot{u}\,p(u,t)\right] - f(u-\vartheta)\,p(u,t) \quad \text{for } u > u_r \tag{67}$$

where we have used the definition of the hazard via the escape function $\rho(t) = f[u(t)-\vartheta]$; the definition of the reset potential $u_r = \eta_0$; and the differential equation $\dot{u} = (-u + R\mathcal{I})/\tau$ of the integrate-and-fire neuron. If we compare Eq. (67) with the membrane potential density equations (17) and (18) of the noisy integrate-and-fire neuron we see that the main difference is in the treatment of the noise, i.e., diffusive noise in (18) is replaced by escape noise in (67). For noise-free integrate-and-fire neurons we set in Eq. (67) $f(u - \vartheta) = 0$ for all values $u < \vartheta$. The (noise-free) continuity equation for the membrane potential densities $p(u, t)$ is therefore equivalent to the noise-free equation of refractory densities, i.e., $\partial Q(\hat{t}, t)/\partial t = 0$ that we retrieve from (65) in the limit of vanishing ρ.

References

Abbott, L. F. and C. van Vreeswijk (1993). Asynchronous states in a network of pulse-coupled oscillators. *Phys. Rev. E 48*, 1483–1490.

Abeles, M. (1991). *Corticonics*. Cambridge: Cambridge University Press.

Amit, D. J. and N. Brunel (1997). A model of spontaneous activity and local delay activity during delay periods in the cerebral cortex. *Cerebral Cortex 7*, 237–252.

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

Berry, M. and M. Meister (1998). Refractoriness and neural precision. *J. of Neuroscience 18*, 2200–2211.

Bethge, M., K. Pawelzik, R. Rothenstein, and M. Tsodyks (2001). Noise as a signal for neuronal populations. *Phys. Rev. Letters xx*, xx.

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

Bialek, W., F. Rieke, R. R. de Ruyter van Stevenick, and D. Warland (1991). Reading a neural code. *Science 252*, 1854–1857.

Brunel, N., F. Chance, N. Fourcaud, and L. Abbott (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Physical Review Letters*.

Brunel, N. and V. Hakim (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Computation 11*, 1621–1671.

Cox, D. R. (1962). *Renewal theory*. London: Methuen.

de Boer, E. and P. Kuyper (1968). Triggered correlation. *IEEE Trans. Biomedical Engineering 15*, 169–179.

de Ruyter van Steveninck, R. R., G. D. Lowen, S. P. Strong, R. Koberle, and W. Bialek (1997). Reproducibility and variability in neural spike trains. *Science 275*, 1805.

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

Destexhe, A. and D. Pare (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *Journal of Neurophysiology 81 81*, 1531–1547.

Fetz, E. E. and B. Gustafsson (1983). Relation between shapes of post-synaptic potentials and changes in firing probability of cat motoneurones. *J. Physiol. 341*, 387–410.

Fusi, S. and Matteo, M. Collective behavior of networks with linear (VLSI) Integrate and Fire Neurons. *Neural Computation 11*, 633-652

Gerstner, W. (1995). Time structure of the activity in neural network models. *Phys. Rev. E 51*(1), 738–758.

Gerstner, W. (2000). Population dynamics of spiking neurons: fast transients, asynchronous states and locking. *Neural Computation 12*, 43–89.

Gerstner, W., R. Kempter, J. L. van Hemmen, and H. Wagner (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature 386*, 76–78.

Gerstner, W., R. Ritz, and J. L. van Hemmen (1993). Why spikes? Hebbian learning and retrieval of time–resolved excitation patterns. *Biol. Cybern. 69*, 503–515.

Gerstner, W. and J. L. van Hemmen (1992). Associative memory in a network of 'spiking' neurons. *Network 3*, 139–164.

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

Hebb, D. O. (1949). *The organization of behavior.* New York: Wiley.

Herrmann, A. and W. Gerstner (1999). Understanding the psth response to synaptic input. In *ICANN'99 Artificial Neural Networks*, Volume 470, pp. 1012–1017. IEE Conference Publication.

Herrmann, A. and W. Gerstner (2001a). Noise and the psth response to current transients. *J. Computational Neuroscience submitted*, xx–xx.

Herrmann, A. and W. Gerstner (2001b). Noise and the psth response to current transients. *J. Computational Neuroscience submitted*, xx–xx.

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

Kirkwood, P. and P. Sears (1978). The synaptic connexions to intercostal motoneurones as revealed by the average common excitation potential. *J. Physiology 275*, 103–134.

Kistler, W. M., W. Gerstner, and J. L. van Hemmen (1997). Reduction of Hodgkin-Huxley equations to a single-variable threshold model. *Neural Comput. 9*, 1015–1045.

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

Kjaer, T. W., J. A. Hertz, and B. J. Richmond (1994). Decoding cortical neuronal signals: network models, information estimation and spatial tuning. *J. Comput. Neuroscience 1*, 109–139.

Knight, B. W. (1972a). Dynamics of encoding in a population of neurons. *J. Gen. Physiology 59*, 734–766.

Knight, B. W. (1972b). The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. *J. Gen. Physiology 59*, 767–778.

Knox, C. K. (1974). Cross-corrlation functions for a neuronal model. *Biophysical J. 14*, 567–582.

König, P., A. K. Engel, and W. Singer (1996). Integrator or coincidence detector? The role of the cortical neuron revisited. *TINS 19*(4), 130–137.

Maas, W. and T. Nathschläger (2000). A model for fast analog computation based on unreliable synapses. *Neural Computation 12*, 1679–1704.

Markram, H., J. Lübke, M. Frotscher, and B. Sakmann (1997). Regulation of synaptic efficacy by coincidence of postysnaptic ap and epsp. *Science 27*, 213–215.

Moore, G., J. Segundo, D. Perkel, and H. Levitan (1970). Statistical signs of synaptic interaction in neurons. *Biophysical Journal 10*, 876–900.

Nykamp, D. and D. Tranchina (2000). A population density approach that facilitates large-scale modeling of neural networks: Analysis and application to orientation tuning. *J. Computational Neuroscience 8*, 19–50.

Omurtag, A., B. Knight, and L. Sirovich (2000). On the simulation of a large population of neurons. *J. Computational Neuroscience 8*, 51–63.

Optican, L. M. and B. J. Richmond (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. 3. Information theoretic analysis. *J. Neurophysiol. 57*, 162–178.

Perkel, D. H., G. L. Gerstein, and G. P. Moore (1967). Neuronal spike trains and stochastic point processes I. the single spike train. *Biophys. J. 7*, 391–418.

Plesser, H. E. and W. Gerstner (2000). Noise in integrate-and-fire models: from stochastic input to escape rates. *Neural Computation 12*, 367–384.

Poliakov, A. V., R. K. Powers, and M. C. Binder (1997). Functional identification of input-output transforms of motoneurons in cat. *J. Physiology 504*, 401–424.

Riccardi, L.M. *Diffusion Processes and related topics in biology.* Springer Verlag, Berlin.

Rieke, F., D. Warland, R. de Ruyter van Steveninck, and W. Bialek (1996). *Spikes - Exploring the neural code.* MIT Press, Cambridge, MA.

Roberts, P. and C. Bell (2000). Computational consequences of temporally asymmetric learning rules: sensory image cancellation. *Computational Neuroscience 9*, 67–83.

Roddey, J., B. Girish, and J. Miller (2000). Assessing the performance of neural encoding models in the presence of noise. *J. Compuational Neuroscience 8*, 95–112.

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

Softky, W. R. (1995). Simple codes versus efficient codes. *Current Opinion in Neurobiology 5*, 239–247.

Song, S., K. Miller, and L. Abbott (2000). Competitive hebbian learning through spike-time-dependent synaptic plasticity. *Nature Neuroscience 3*, 919–926.

Tovee, M. J., E. T. Rolls, A. Treves, and R. P. Belles (1993). Information encoding and the responses of single neurons in the primate visual cortex. *J. Neurophysiol. 70*, 640–654.

Zhang, L., H. Tao, C. Holt, W.A.Harris, and M.-M. Poo (1998). A critical window for cooperation and competition among developing retinotectal synapses. *Nature 395*, 37–40.