Activity-dependent selection of axonal and dendritic delays or, why synaptic transmission should be unreliable

Walter Senn*, Martin Schneider†,+ and Berthold Ruf†,⋆

* *Physiologisches Institut, Universität Bern, Bühlplatz 5, CH-3012 Bern, Switzerland*

† *Institute for Theoretical Computer Science, Technische Universität Graz, Klosterwiesgasse 32/2, A-8010-Graz, Austria*

*wsenn@cns.unibe.ch, +davidge@sbox.tu-graz.ac.at, ⋆bruf@igi.tu-graz.ac.at

July 6, 2001

Systematic temporal relations between single neuronal activities or population activities are ubiquitous in the brain. No experimental evidence, however, exists for a direct modification of neuronal delays during Hebbian type stimulation protocols. We show that, in fact, an explicit delay adaptation is not required if one assumes that the synaptic strengths are modified according to the recently observed temporally asymmetric learning rule with the downregulating branch dominating the upregulating branch. During development, slow unbiased fluctuations in the transmission time together with temporally correlated network activity may control neural growth and implicitly induce drifts in the axonal delays and dendritic latencies. These delays and latencies become optimally tuned in the sense that the synaptic response tends to peak in the soma of the postsynaptic cell if this is most likely to fire. The nature of the selection process, however, requires unreliable synapses in order to give 'successful' synapses an evolutionary advantage upon the others. Without unreliable transmission, the learning rule would equally modify all synapses with the same local time difference between the pre- and postsynaptic signal, irrespective whether the corresponding total axonal and dendritic delay supports the postsynaptic firing or not. Stochastic transmission may resolve this ambiguity by restricting the modification process to the active synapses only, giving those synapses a higher chance to be strengthened which contribute to the postsynaptic activity. The width of the learning window does also implicitely determine the preferred dendritic delay and the preferred width of the postsynaptic response. Hence, the learning rule may implicitly determine whether a synaptic connection provides precisely timed or rather 'contextual' information.

1 Introduction

The temporal structure in any type of cognitive or behavioral task such as memory recall, motor execution or navigation must be reflected in the temporal activity of neurons or neuronal populations. Time in this neuronal correlate is often considered to be encoded in the synaptic strengths which map to a phase advance of a postsynaptic neuron (see e.g. Roberts, 1999, cf. also Hopfield and Brody, 2000 and 2001). Time may also directly be encoded in axonal propagation delays and dendritic latencies. While in the case of a representation in synaptic weights the timing of a postsynaptic cell may be adapted by any synaptic modification, there is no experimental evidence for a direct adaptation of axonal and dendritic delays. Here we show that a temporally asymmetric spike-time dependent synaptic modification may substitute for such a direct adaptation of delays in favor of an implicit adaptation through selection. According to these rules the synaptic efficacy is upregulated if the presynaptic action potential (AP) is elicited before the postsynaptic AP, while it is downregulated otherwise (Markram et al., 1997; Bi and Poo, 1998; Feldman, 2000). The high temporal resolution occuring in these synaptic rules is puzzling, especially in view of the wide range of possible delays and latencies encountered in the brain. On the one hand, the temporal order of the pre- and postsynaptic signals can be distinguished with a remarkable sensitivity which may fall below 10 ms. On the other hand, substantial delays may be encountered between the generation of the presynaptic AP, the excitatory postsynaptic potential (EPSP) at the synaptic site, and the maximal response of the EPSP in the soma of the postsynaptic cell (Fig. 1a, see Discussion for references). We show that, despite this temporal discrepancy, the learning rule can select the correct pre- and postsynaptic delays (where 'correct' means that the somatic response peaks when the postsynaptic cell is most likely to fire), provided that the synaptic transmission is unreliable. We also show that in the presence of small unbiased delay fluctuations the learning rule makes the delays drift until the correct timing is achieved. In order to prevent an unbounded growth of the synaptic weights during this drift it is sufficient to require that downregulating branch in the learning rule dominates the upregulating branch in strength (cf. Abbott and Song, 1999, and Kempter et al., 1999). The paradigm of delay adaptation through selection may be of particular importance in the developing nervous system where it could explain the temporal fine tuning of specific synaptic pathways.

The idea that delays may be selected through an adaptation of synaptic weights was explored early in theoretical works which were focussed on learning spatio-temporal activity patterns (Sompolinsky and Kanter, 1986; Kühn and van Hemmen, 1991; Gerstner, 1993; Natschläger and Ruf, 1998). Delay selection is also a key mechanism for localizing sound signals. It has been suggested that during the ontogenetic development of the barn owl auditory system synaptic connections with matching delays are selected through an asymmetric learning rule (Gerstner et al., 1996, see also Carr et al., 1999). Other theoretical works discuss learning rules which explicitly adjust the synaptic delay as a function of the pre- and postsynaptic activities (Hüning et al., 1998; Eurich et al., 1999). Instead of postulating an explicit form of delay adaptation, however, we claim that delay learning is a natural consequence of a temporally asymmetric modification of the synaptic strength. Our analysis can be generalized to reinforcement learning which may work with a similar type of asymmetric learning rule, although on a larger time scale and on the level of neuronal populations.

An example of learning motor sequences within the basal ganglia−thalamocortical feedback loop outlined in the Discussion shows that the same delay structure between a presynaptic spike, the synaptic release, the postsynaptic spike and the backward reinforcement signal may also occur within cortical circuitries (compare the stages A, S and B in Figs. 1a and 8).

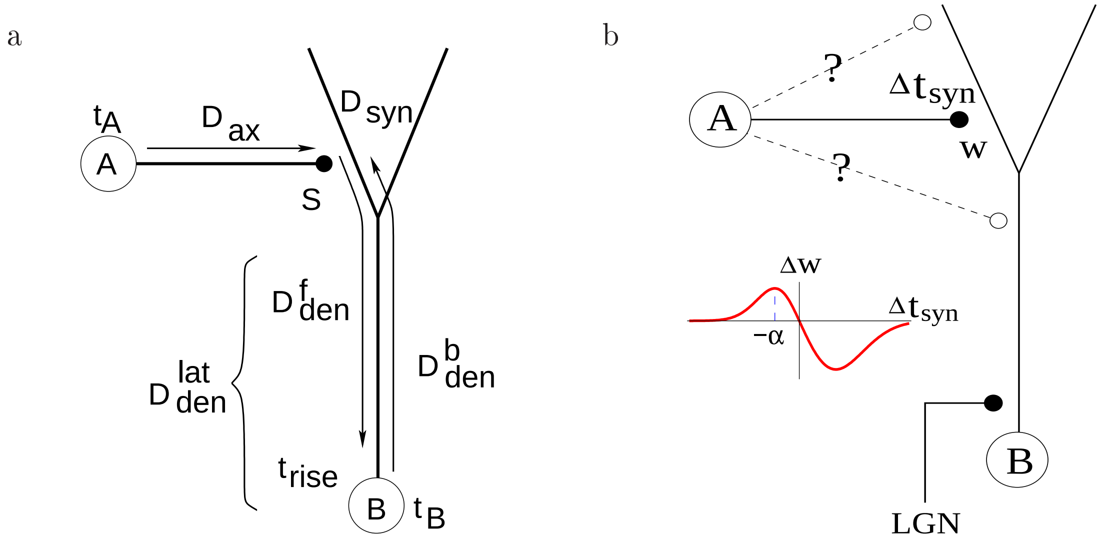

Figure 1: *The model.* **a:** *The different signal delays between a presynaptic neuron (or a neuronal population) A and a postsynaptic neuron B. The total forward delay is composed of an axonal delay D_{ax}, a (vanishing) synaptic delay D_{syn}, and the dendritic latency consisting of a forward dendritic (propagation) delay D^f_{den} and the EPSP rise time t_{rise} measured in the soma, $D^{lat}_{den} = D^f_{den} + t_{rise}$. At the peak of the EPSP there is an enhanced probability of neuron B to fire. An AP generated in the soma of neuron B arrives with a backward dendritic delay D^b_{den} at the synaptic site S.* **b:** *The synaptic weight w is modified depending on the local time difference $\triangle t_{syn}$ between the EPSP and the backward AP at site S (inset, $\triangle t_{syn} = (t_A + D_{ax}) - (t_B + D^b_{den})$). This rule, together with slow delay fluctuations and unreliable synaptic transmissions, induces a drift of the total forward delay from A to B towards the most probable spike time difference $t_B - t_A$ between the two cells. Depending on the time difference, α, that corresponds to the maximum of the learning function, the final synaptic position will be either close to the soma (α small) or far on the apical dendritic tree (α large). Hence, the shape of the learning function determines whether a synaptic connection tends to trigger a postsynaptic AP (through a proximal synapse) or whether it tends to support the postsynaptic activity by a slow 'contextual' EPSP (through a distal synapse), respectively.*

Integration field of a V1 cell: an example As a first example for delay adaptations we consider the integration field of a layer 5 pyramidal cell in the primary visual cortex. This cell is driven by feedforward visual input through the lateral geniculate nucleus (LGN) which synapses on the dendritic shaft in layer 4 (Nieuwenhuys, 1994). At the same time the cell receives lateral input onto the apical dendritic tree from neighboring cortical columns (cf. Fig. 1b). *In vivo* experiments show that the cell horizontally integrates neuronal activity over a cortical patch of roughly 1 cm in diameter (Bringuier et al., 1999). This cortical integration field corresponds to a visual receptive field on the retina of roughly 10°. However, only nearby cortical input, corresponding to visual stimulation within 2° of the receptive center, can activate the cell, while peripheral horizontal input provides only subthreshold input. The weak peripheral horizontal input can also be delayed up to 50 ms. It is assumed that these delays are mainly due to the slowly conducting horizontal pathways which show axonal conductance velocities ranging from 0.1 to 0.3 m/sec. The existence of the different latencies between the feedforward vertical input onto the proximal region and the peripheral intracortical input onto the distal portion of the apical tree suggests that the latencies are adjusted to optimize the somatic summation (Bringuier et al., 1999). Our simulations show that the temporally asymmetric learning rule is in fact a means to optimally tune these intracortical delays.

Outline of the basic mechanisms To explain the basic mechanisms we consider a neocortical cell (or population of cells) A and a layer 5 pyramidal cell B with different delay lines from A to B. In a first scenario we assume that the feedforward input triggers in sequence an AP in cell A and in cell B at times $t_A < t_B$, respectively. The presynaptic AP arrives at the synaptic site at time $t_A + D_{ax}$, while the backpropagating postsynaptic AP arrives at the same location at time $t_B + D^b_{den}$, with D_{ax} and D^b_{den} being the axonal delay and backward dendritic delay of the corresponding synaptic connection, respectively (see Fig. 1a). Assuming a competitive Hebbian modification among the synaptic connections (realized by a learning function with LTD dominating LTP, cf. Song et al., 2000), only these synapses experiencing the strongest upregulation survive, while all the others are suppressed. In other words, the rule selects the connection(s) for which the local time difference measured at the synaptic site, t_{syn}, corresponds to the maximum of the learning function, $t_{syn} = -\alpha$ (see Fig. 1b). Hence, the optimal delay line satisfies the condition $\triangle t_{syn} = (t_A + D_{ax}) - (t_B + D^b_{den}) = -\alpha$. Assuming that the maximum of the learning function covers the sum of the (forward) dendritic latency and the backward dendritic delay, $\alpha = D^{lat}_{den} + D^b_{den}$, this equation reduces to $D_{ax} + D^{lat}_{den} = t_B - t_A$. As dendritic latency D^{lat}_{den} we define the sum of the EPSP propagation delay and the rise time of the somatic EPSP. The last equation confirms our expectation that, if cell A should support the firing of cell B, the total forward delay should match the typical spike time difference between these cells.

In a second scenario we assume that third input provides only subthreshold input and that the spike in cell B is finally elicited by the synaptic responses from cell A. In this case the spike time t_B is roughly given by $t_B = t_A + D_{ax} + D^{lat}_{den}$. If, as above, we insert this expression in the condition $(t_A + D_{ax}) - (t_B + D^b_{den}) = -\alpha$ we obtain $D^{lat}_{den} + D^b_{den} = \alpha$. Hence, in the second scenario the rule selects the connection(s) for which the sum of the forward and backward dendritic delays corresponds to the peak of the learning function. For instance, if the maximum of the learning function is given by $\alpha = 9$ ms, that synaptic pathway is selected whose EPSP triggers 7 ms after the synaptic release a postsynaptic AP, which itself peaks after 2 ms back at the synaptic site. It is tempting to speculate that the match between the total dendritic delays and the peak in the learning function determines the position of a synapse on the dendritic tree, and that different learning functions (different α's) prefer different synaptic positions (see Fig. 1b).

To make our reasoning on the pre- and postsynaptic delay selection physiologically more plausible we are weakening the assumptions in two ways. First, we do not require that the full spectrum of possible delay lines must initially be present. Rather, we only assume a small number of synaptic connections (say 10) from A to B with similar pre- and postsynaptic delays. These delays, however, are assumed to exhibit slow stochastic fluctuations caused, for instance, by the dis- and re-appearance of synapses at different dendritic locations or by the modulation of active membrane conductances. The fluctuations, together with the asymmetric synaptic modifications, are enough to induce drifts in the average axonal and dendritic delays towards optimality. Second, we do not restrict ourselves to the above scenaria according to which the firing of cell B is determined always by third input or always by the firing of cell A. Rather, we assume some degree of temporal correlation in the Poisson activities of the two cells A and B which is implicitly induced by third input and which depends on the strength of the synaptic connections. In this generalized scenario, the axonal delay and the dendritic latency are both shifted towards optimality, i.e. they are adapted such that their sum becomes equal to the typical spike time difference, $D_{ax}+D_{den}^{lat}=t_B-t_A$. Note that there are different ways to distribute the total delay on the individual components D_{ax} and D_{den}^{lat}. In fact, after the adaptation process the additional relation $D_{den}^{lat}+D_{den}^{b}=\alpha$ is satisfied, and this implicitly determines D_{den}^{lat} and therefore the time course of the somatic EPSP. Hence, in the generalized scenario, the learning rule does not only adjust the correct total delay from A to B, but determines also whether the corresponding somatic EPSP is sharp (α small) or wide (α large).

The simultaneous selection of axonal and dendritic delays deserves a closer look. It starts with the observation that the learning rule cannot distinguish between delay lines with axonal and backward dendritic delays of the form $D_{ax}+\epsilon$ and $D_{den}^{b}+\epsilon$, simply because the argument $\triangle t_{syn}=(t_A+D_{ax})-(t_B+D_{den}^{b})$ of the learning function depends on the difference between these delays only. For instance, two delay lines, the second having a 3 ms longer axonal and a 3 ms longer backward dendritic delay than the first, have both the same signal time difference at the synaptic site, and therefore go through the same synaptic modifications. Assuming that the forward dendritic latency is equal to the backward dendritic delay, however, the somatic EPSP of the second delay line peaks systematically 6 ms ($=2\epsilon$) after the EPSP of the first, and therefore favors a different postsynaptic spike time. A way to nevertheless select between the two delay lines is to introduce unreliable synaptic transmissions, and to modify the synaptic strength only if neurotransmitter was released. A synapse which then actively supports a postsynaptic spike will have more chances to be upregulated than a synapse which does not, and this resolves the above ambiguity. From a general point of view, any selection process is built up on some source of randomness. In our case, the selection of the appropriate axonal delay requires slow delay fluctuations, while the selection of the appropriate dendritic latency requires unreliable synaptic transmissions.

Integration field of a V1 cell: some implications Coming back to the introductory example, A is an orientation selective layer 5 pyramidal cell in the primary visual cortex of a cat and B is a nearby neural population in the same cortical area. According to our reasoning, the axonal delays plus the dendritic latencies from B to A should reflect the time differences with which the corresponding retinal positions are most likely to be stimulated. Since an orientation selective cell is more sensitive to motions orthogonal to the preferred orientation the integration field should be more developed along this orthogonal orientation. For instance, we expect from our model that the subthreshold receptive field of such a cell is largest along the orthogonal orientation, and that along this same orientation the propagation velocity is highest. Both predictions seem to be confirmed to some degree by the *in vivo* recordings of Bringuier et al., 1999. In fact, the statistics of the orientation selective cells (Fig. 4 in the cited work) shows an average width of the subthreshold receptive field of 5° and 12° in the preferred and its orthogonal orientation, respectively, and a common latency of 50 ms across these differently sized half diameters. A selection of different axonal delays is further confirmed by the statistics of the 'apparent speed of horizontal propagation' which ranges from 0.05 to 0.6 m/sec. According to our model, the receptive field asymmetries concerning the axonal delays would be enhanced if during rearing the kittens were exposed to fast motions in exclusively one direction. A similar confirmation of the model concerning the adaptation of the dendritic latency is more difficult due to the lack of data. However, the intracellular recordings show that the relative width of the subthreshold depolarization is narrower for stimuli within 2° of the receptive field center than for stimuli in the periphery (Fig. 1 in Bringuier et al., 1999). This is in agreement with the anatomical finding that feedforward input projects proximal to the soma while horizontal cortical connections project onto the distal apical tree (Nieuwenhuys, 1994). According to our model, this synaptic distribution would be the result of a self-organizing process, provided that the modification of the proximal (feedforward) synaptic projections are governed by a narrower learning function than the distal (horizontal) ones.

Overview The paper is organized as follows: In Section 2 we formalize the model and state the assumptions on the synaptic modifications and the delay fluctuations. In Section 3 we introduce spike correlations between the inputs onto our neuron and derive the population equation for the dynamics of the synaptic strengths. In Section 4 we consider the change in the average delays induced by the weight modifications under the assumption of fixed spike correlations between the pre- and postsynaptic sites. In Section 5 the spike correlation is allowed to be influenced by the synaptic modifications and we study the induced shift in the dendritic latencies at fixed axonal delays. In Section 6 the general case is treated where correlated spike activity jointly entrains axonal and dendritic delays in the presence of stochastic delay fluctuations and unreliable synaptic transmissions. The Discussion provides physiological evidence for our assumptions and ends with an example of delay learning between different cortical regions.

2 The model

We consider different delay lines from a presynaptic neuron (or neuronal population) A to a postsynaptic neuron B. The delay lines may vary in the axonal pathway and the synaptic position on the dendritic tree (Fig. 1a). Each delay line is specified by an axonal delay, D_{ax}, a (negligible) synaptic delay, D_{syn}, a forward dendritic (propagation) delay, D^f_{den}, the EPSP rise time in the postsynaptic soma, t_{rise} (to which, together with D^f_{den}, we refer as the dendritic latency), and a backward dendritic delay, D^b_{den}, specifying the delay of the postsynaptically generated AP back to the subsynaptic site (i.e. to the site in the dendritic tree adjacent to the synapse). If pre- and postsynaptic APs are triggered at times t_A and t_B, respectively, the EPSP faces the postsynaptic AP at the subsynaptic site with a time difference

$$\triangle t_{syn} = (t_A + D_{ax} + D_{syn}) - (t_B + D^b_{den}) = \triangle t + \delta \ , \tag{1}$$

where $\triangle t = t_A - t_B$ and $\delta = D_{ax} + D_{syn} - D^b_{den}$ (cf. Fig. 2b). Note that δ is the relative delay encountered at the subsynaptic site for the case of equal spike times.

The synaptic strength w is assumed to be up- or downregulated depending on the sign of the local time difference $\triangle t_{syn}$. Specifically, we assume that w is modified proportionally to the bi-alpha-shaped function of the local time difference,

$$\triangle w \ \sim \ \psi_\circ(\triangle t_{syn}) = \begin{cases} \frac{\gamma}{\alpha}|\triangle t_{syn}| \, e^{-\frac{(\triangle t_{syn})^2}{2\alpha^2}} & , \quad \triangle t_{syn} < 0 \\ -\frac{\gamma}{\beta}|\triangle t_{syn}| \, e^{-\frac{(\triangle t_{syn})^2}{2\beta^2}} & , \quad \triangle t_{syn} \geq 0 \ , \end{cases} \tag{2}$$

with appropriate constants α, $\beta, \gamma > 0$ (see Fig. 2a). The form of $\psi_\circ$ is such that the weight is maximally upregulated if the presynaptic spike precedes the postsynaptic spike by α ms while it is maximally downregulated if the presynaptic spike follows the postsynaptic spike by β ms. The normalization is such that the absolute values of the peaks of the LTP and LTD branches are independent of their widths α and β, respectively. We assume that the proportionality factor in (2) is the actual weight itself, $\triangle w = w\,\psi_\circ(\triangle t)$. If w expresses the number of quantal release sites, for instance, the factor w states that each site separately undergoes long-term modification in that it either vanishes or forms another nearby release site. The same factor emerges if one deduces (2) from kinetic schemes governing the involved neurotransmitter receptors and secondary messengers (Senn et al., 2001). To avoid a blowing up of the synaptic strengths our analysis will use the intrinsic normalization property of the learning rule which is obtained when the downregulation dominates the upregulation ($\beta > \alpha$, see Eq. 8 below).

Next, we consider slow stochastic fluctuations in the axonal delays and dendritic latencies. We observe that, according to (1), it is reasonable to group those delay lines together with the same relative delay δ. This is possible since only the time difference $\triangle t_{syn} = \triangle t + \delta$ of the signals measured at the subsynaptic site enters in the synaptic dynamics (2). We may therefore parameterize the synaptic weights with the relative delay, $w(t) = w(\delta, t)$. Due to the linear scaling factor occuring in the rule these lumped synaptic weights are likewise adapted according to $\triangle w(\delta, t) = w(\delta, t)\,\psi_\circ(\triangle t)$. Similarly, stochastic changes in the axonal and dendritic delays are only relevant for the synaptic dynamics as far as they change the relative delay δ. For an individual pathway this relative delay may be shortened or extended by changing the axonal delay, D_{ax}, the backward dendritic delay, D^b_{den}, or both. To take account of these fluctuations we assume that for each delay line the relative delay δ may change during a time interval T by $\triangle\delta$ with a small probability $\epsilon \geq 0$. With probability $1 - 2\epsilon$ the delay of the synaptic pathway δ does not change. On the level of the lumped weights $w(\delta, t)$ this dynamics can be expressed by

$$w(\delta, t) = \epsilon w(\delta - \triangle\delta, t - T) + (1 - 2\epsilon) w(\delta, t - T) + \epsilon w(\delta + \triangle\delta, t - T) \ . \tag{3}$$

If the time scale T of these fluctuations is long compared to the time scale of the synaptic adaptation, the two processes do not interfere and the dynamics for the lumped weights is the same as that for the individual connections. On a time scale longer than T we are dealing with derivatives, and taking the limit $\triangle\delta,\, T \to 0$ we obtain after simple algebraic manipulations $\frac{\partial w(\delta,t)}{\partial t} = \epsilon \frac{\partial^2 w(\delta,t)}{\partial \delta^2}$, provided that $\lim_{\triangle\delta,T\to 0} \frac{(\triangle\delta)^2}{T} = 1$. Thus, the slow unbiased stochastic fluctuation introduces a diffusion of the synaptic weights along the relative delays.

The general scenario we consider is that neuron B fires, with some jitter, $|\overline{\triangle t}|$ after neuron A (note that in this case $\overline{\triangle t} < 0$). The firing time t_B depends on additional input from a third population onto B (as e.g. forward input from the LGN, see Fig. 1b), but may also be influenced by the firing of neuron A itself. The basic mechanism is that the rule (2) selects delay lines for which the pre- and postsynaptic signal time difference at the subsynaptic site equals the maximum of the learning function, $\triangle t_{syn} = \overline{\triangle t} + \delta = -\alpha$. The stochastic process (3) of building new delay lines (i.e. of assigning positive weights to delay lines with previously vanishing weights) ensures that after a while there is always a delay line with an appropriate relative delay δ and non-vanishing w. If the LTD branch of the learning function dominates the LTP branch, $\beta > \alpha$ in (2), delay lines which do not meet the above condition will be suppressed. If, in addition, we introduce synaptic transmission failures, those delay lines which do satisfy $\overline{\triangle t} + \delta = -\alpha$, but do not support the postsynaptic spike by virtue of their total forward delay, i.e. for which $D^f_{tot} = D_{ax} + D_{syn} + D^f_{den} + t_{rise} \neq \overline{\triangle t}$, will eventually be suppressed as well.

3 Weight modification induced by correlated activity

We assume that the pre- and postsynaptic neurons are firing with some instantaneous Poisson rates f_{pre} and f_{post}, respectively, which are themselves correlated to some degree. Let us consider a delay line with relative delay δ as defined in (1) and let us assume that the synaptic strength is modified according to (2). The change in w accumulated in the time interval T ($\gg \alpha, \beta$) preceding the current time t can then be approximated by

$$\triangle w(t) \approx w(\bar{t}) \int_{t-T}^{t} dt' \int_{t-t'-T}^{t-t'} d\tau\, \psi_\circ(\tau + \delta)\, f_{pre}(t' + \tau)\, f_{post}(t')\ , \tag{4}$$

with some appropriate value $\bar{t}$ between $t - T$ and t. Assuming that the individual weight changes in (2) are small we obtain, on a time scale which is large with respect to T, the differential equation

$$\frac{dw(t)}{dt} \approx \frac{\triangle w(t)}{T} = w(t) \int_{-\infty}^{\infty} d\tau\, \psi_\circ(\tau + \delta)\, C(\tau; t) \tag{5}$$

with correlation function

$$C(\tau; t) = \frac{1}{T} \int_{t-T}^{t} dt'\, f_{pre}(t' + \tau)\, f_{post}(t')\ . \tag{6}$$

Observe that the integral domain in (5) can be extended to infinity since T is larger than the width of $\psi_\circ$. Note further that in (4) and (5) we assume that the decay of the synaptic memory (i.e. of w) is on an even longer time scale, so that all the changes are accumulated without degradation. Next we require that the instantaneous pre- and postsynaptic firing

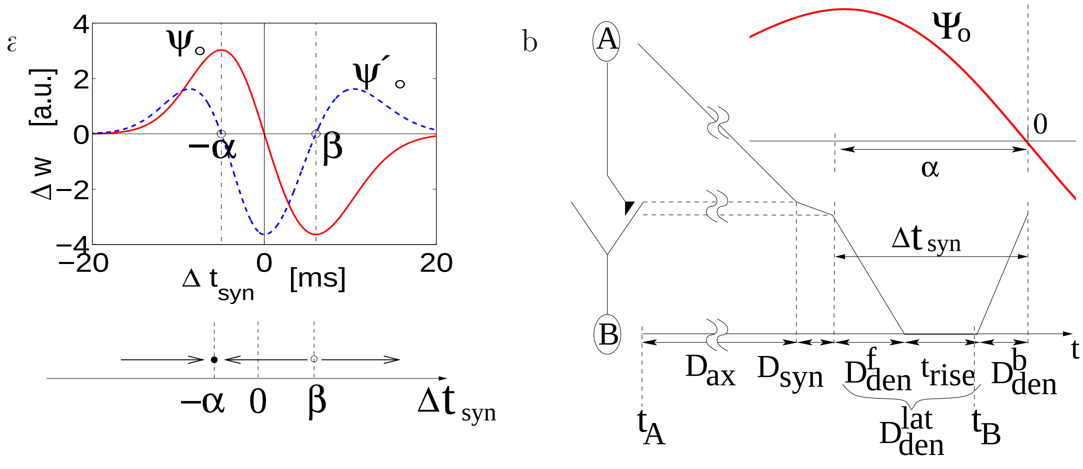

Figure 2: *The synaptic learning function and the different delays after learning.* **a:** *The learning function $\psi_\circ(\triangle t_{syn})$ for the synaptic weights as a function of the time difference between the pre- and postsynaptic signals at the subsynaptic site (full line, see Eq. 2 and 1). For repetitive regular stimulations ($\varsigma = 0$) the average axonal delay $\overline{D}_{ax}$ and the average dendritic latency $\overline{D}^{lat}_{den} = \overline{D}^{f}_{den} + t_{rise}$ are adapted according to the derivative $\psi'_\circ$ (dashed line, cf. Eqs. 16 and 22). The delay adaptation has an attracting fixed point $\triangle t_{syn} = -\alpha$, and a repulsive fixed point $\triangle t_{syn} = \beta$.* **b:** *The different delays involved in the signal transmission from neuron A to neuron B and back to the synaptic site. The bent curve represents the learning function $\psi_\circ$ centered at the time when the backpropagated AP reaches the subsynaptic site. The choice of the delays reflects the steady state behavior when $\triangle t_{syn} = -\alpha$.*

rates are correlated in the sense that there is a tendency for the postsynaptic neuron to fire around $\overline{\triangle t} \pm \varsigma$ ms after/before a presynaptic spike (after for $\overline{\triangle t} < 0$, before for $\overline{\triangle t} > 0$). Thus we assume a correlation function of the form

$$C(\tau; t) = c\,\mathcal{G}(\overline{\triangle t}, \varsigma, \tau) \;+\; \overline{f}_{pre}\,\overline{f}_{post}(t) \tag{7}$$

with a constant presynaptic mean frequency $\overline{f}_{pre}$, a postsynaptic mean frequency $\overline{f}_{post}(t)$ which may depend on the synaptic weight $w(t)$ but is not correlated with the presynaptic spikes, a spike correlation factor $c \geq 0$, and $\mathcal{G}(\overline{\triangle t}, \varsigma, \tau) = (2\pi\varsigma^2)^{-\frac{1}{2}} \exp(-\frac{(\overline{\triangle t} - \tau)^2}{2\varsigma^2})$ being a Gaussian distribution around $\overline{\triangle t}$ with standard deviation ς. For the specific choices of the learning function (2) and the correlation function (7) the integral in (5) can be analytically calculated. One obtains

$$\frac{dw(t)}{dt} = w(t)\,(\,c\,\psi_\varsigma(\overline{\triangle t} + \delta) \;-\; \overline{\psi}_\circ\,\overline{f}_{pre}\,\overline{f}_{post}(t)) \tag{8}$$

where $\psi_\varsigma(\triangle t) = \int_{-\infty}^{\infty} \psi_\circ(\tau)\,\mathcal{G}(\triangle t, \varsigma, \tau)\,d\tau$ is approximately

$$\psi_\varsigma(\triangle t) \approx \begin{cases} \frac{\gamma\alpha^2}{(\alpha^2+\varsigma^2)^{\frac{3}{2}}} |\triangle t|\, e^{-\frac{(\Delta t)^2}{2(\alpha^2+\varsigma^2)}} & , \quad \triangle t < 0 \\ -\frac{\gamma\beta^2}{(\beta^2+\varsigma^2)^{\frac{3}{2}}} |\triangle t|\, e^{-\frac{(\Delta t)^2}{2(\beta^2+\varsigma^2)}} & , \quad \triangle t \geq 0\,, \end{cases} \tag{9}$$

and $-\overline{\psi}_\circ = \int_{-\infty}^{\infty} d\tau\, \psi_\circ(\tau) = -\gamma\,(\beta - \alpha)$. Note that for $\beta > \alpha$ one has $-\overline{\psi}_\circ < 0$. A negative integral of the learning function was postulated in (Abbott and Song, 1999; Kempter et al., 1999) to normalize the synaptic weights and stabilize the postsynaptic potential in a subthreshold regime. There is in fact later experimental evidence that, for synaptic connections onto pyramidal cells in layer II of the rat barrel cortex, the width of the LTD window, β, is larger than the one of the LTP window, α, and that the overall integral over the learning function is negative (Feldman, 2000). The exact form of ψ_ς is a smoothed version of (9) and the approximation is strict for $\alpha = \beta$. The shape of ψ_ς is thus again of the original form depicted in Fig. 2a, but now with maximum at $\vartriangle t = -\sqrt{\alpha^2 + \varsigma^2}$ and minimum at $\vartriangle t = \sqrt{\beta^2 + \varsigma^2}$. Note that for vanishing variance in the spike time differences, $\varsigma = 0$, one retrieves (2). The effective learning window for a synapse exposed to Gaussian distributed spike time differences appears therefore to be down-scaled and broadened.

To close the feedback loop we split the uncorrelated postsynaptic firing rate arising in (7) into a part originating from the specific presynaptic neuron and a constant background rate due to the remaining afferents,

$$\overline{f}_{post}(t) = \sum_\delta w(t)\,\overline{f}_{pre} + \overline{f}^{\,o}_{post}\,, \tag{10}$$

where the sum extends over the different synaptic delay lines from the specific presynaptic neuron (with weights $w = w_\delta$). In writing a linear sum we assume that the postsynaptic neuron is operating in a linear regime of its input-output function. We also assumed that the specific presynaptic neuron affects the correlation function (7) only by the overall synaptic input without considering the temporal structure of the interaction (but compare (23) in Sect. 6). Inserting (10) into (8) yields

$$\frac{dw(t)}{dt} = w(t)\,[\,c\,\psi_\varsigma(\overline{\vartriangle t} + \delta)\; - \;\overline{\psi}_\circ \overline{f}^{\,2}_{pre} \sum_\delta w(t) - \overline{\psi}_\circ \overline{f}_{pre}\,\overline{f}^{\,o}_{post}]\,. \tag{11}$$

Observe that in the presence of non-correlated spike activity, $c = 0$ in (7), the synaptic weight would asymptotically decay towards zero due to the negative overall integral over the learning function. Furthermore, the synaptic strength cannot grow to infinity due to the negative feedback loop represented by the second term in the brackets.

Finally, we consider a family of delay lines whose weights may stochastically fluctuate among nearest neighbors according to (3). In the limit of a continuum of delay lines the fluctuations result in a diffusion term which adds to (11),

$$\frac{dw(\delta,t)}{dt} = w(\delta,t)\,[\,c\,\psi_\varsigma(\overline{\vartriangle t} + \delta)\; - \;\overline{\psi}_\circ \overline{f}^{\,2}_{pre} \int_{-\infty}^{\infty} w(\delta,t)\, d\delta - \overline{\psi}_\circ \overline{f}_{pre}\,\overline{f}^{\,o}_{post}] \,+\, \epsilon \frac{\partial^2 w(\delta,t)}{\partial \delta^2}\,. \tag{12}$$

To reveal the structure of this population equation for the synaptic weights we rewrite it in the form

$$\frac{dw}{dt} = w\,[\psi_\varsigma(\overline{\vartriangle t} + \delta)\; - \; c_1\langle w\rangle \; - \; c_2] \,+\, \epsilon \frac{\partial^2 w}{\partial \delta^2}\,, \tag{13}$$

with $\langle w\rangle = \int_{-\infty}^{\infty} w(\delta,t) d\delta$ and constants $c_1 = \overline{\psi}_\circ \overline{f}^{\,2}_{pre}$ and $c_2 = \overline{\psi}_\circ \overline{f}_{pre}\,\overline{f}^{\,o}_{post}$. The correlation factor c was absorbed in the learning function ψ_ς.

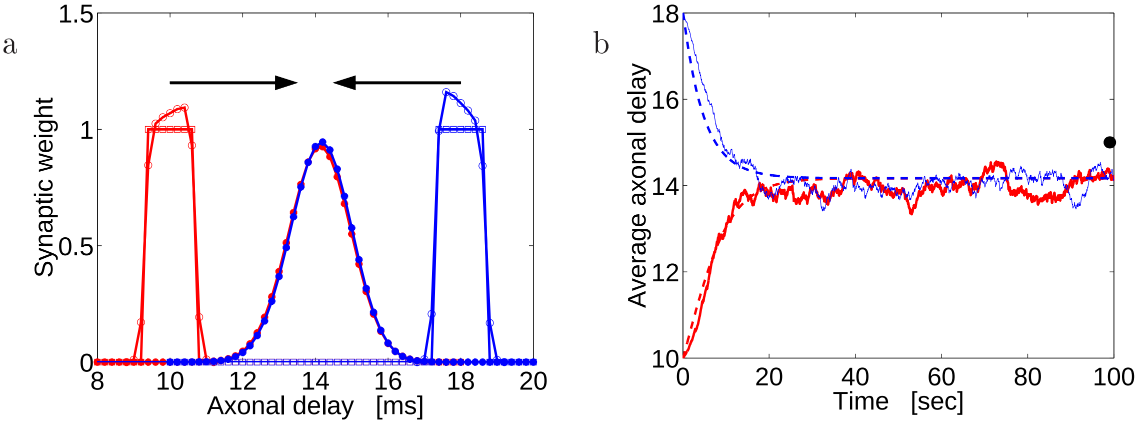

Figure 3: *Axonal delay adaptation with fixed synaptic and dendritic delays (= 1 ms) and with Gaussian spike time differences fluctuating around* $\overline{\triangle t} = -20$ *ms.* **a:** *Evolution of the axonal delays* D_{ax} *for two different initial distributions (left and right squared curves), shown after* 120 *ms (corresponding to 2 pairings, circled curves) and after* 20 *sec (corresponding to 400 pairings, starred curves) when converged to a Gaussian distribution centered at* $\overline{D}_{ax} = 14.2$ *ms.* **b:** *Time evolution of the average axonal delay* $\overline{D}_{ax}$ $(= \overline{\delta})$ *corresponding to the simulations in (a) (blurred lines) and according to the approximation (16) (dashed lines). The steady state for* $\overline{D}_{ax}$ *is well predicted by (16), according to which* $\overline{D}_{ax}$ *adapts such that eventually the argument of* ψ'_ς, $\overline{\triangle t} + \overline{D}_{ax}$, *is equal to the the zero* $-\sqrt{\alpha^2+\varsigma^2}$ *of* ψ'_ς *(which itself corresponds to the peak of* ψ_ς, *cf. Fig. 2a). In fact, in the steady state we have* $\overline{\triangle t} + \overline{D}_{ax} = -20 + 14.2 = -5.8$ *and* $-\sqrt{\alpha^2+\varsigma^2} = -5.83$. *Without jitter in the spike time differences* $(\varsigma = 0)$ *the average axonal delay* $\overline{D}_{ax}$ *would be roughly* 1 *ms longer (dot at* 100 *sec). Other parameters:* $\alpha = 5$ *ms and* $\varsigma = 3$ *ms.*

4 Induced delay shift for fixed spike correlations

Let us consider the 'supervised' learning scenario by imposing the stationary spike statistics (7) between the pre- and postsynaptic neuron. Thus, besides some uncorrelated component, the time differences of the signals at the synapse with delay δ are Gaussian distributed around $\overline{\triangle t} + \delta$ with standard deviation ς. We are interested in the dynamics of the relative delays δ which are implied by the weight modification (12).

To gain insight into this dynamics let us first neglect the stochastic renewal process and consider the synaptic modification in the form (13) with $\epsilon = 0$. Due to the normalization term $-c_1\langle w\rangle$ in (13), the delay adaptation is then purely governed by a selection process, selecting those connections that experience the strongest potentiation. [To prove this we first observe that due to the negative feedback induced by the uncorrelated spikes (the term $-c_1\langle w\rangle$) and the saturation towards zero (the factor w) the weights $w(\delta,t)$ do always converge to a steady state. In this limit either $w(\delta,t) = 0$ or $\psi_\varsigma(\overline{\triangle t}+\delta) - c_1\langle w\rangle - c_2 = 0$. Since $-c_1\langle w\rangle - c_2$ is independent of δ there are only discrete values of δ satisfying the latter equation. Now let us assume that the delay line δ_α with $\overline{\triangle t} + \delta_\alpha = -\alpha$ representing the maximum of ψ_ς initially has a positive weight, $w(\delta_\alpha, 0) > 0$. Since according to (13) weights with $\psi_\varsigma(\overline{\triangle t}+\delta) - c_1\langle w\rangle - c_2 > 0$ are potentiated while others are depressed, we conclude that in the steady state the delay line δ_α survived. Following the reasoning above we then must have $\psi_\varsigma(\overline{\triangle t}+\delta_\alpha)-c_1\langle w\rangle - c_2 = 0$ while for all other delay lines $w(\delta,t)=0$. In general, exactly those delay lines survive for which, among the initial weight distribution, the value $\psi_\varsigma(\overline{\triangle t}+\delta)$ is maximal (and positive).]

If we next consider stochastic fluctuations of the delays, $\epsilon > 0$ in (13), then the delays may shift beyond their initial regime and delays can finally be selected which were not present among the initial configuration. As can be guessed from the above analysis, the relative delays δ in general will shift such that the average time difference at the synaptic sites becomes $\overline{\triangle t}+\delta = -\alpha$. To formalize this statement we consider the average delay $\overline{\delta}$ defined by the center of gravity of the δ-distribution,

$$\overline{\delta}(t) = \frac{\int_{-\infty}^{\infty}\delta\, w(\delta,t)\, d\delta}{\int_{-\infty}^{\infty} w(\delta,t)\, d\delta} = \frac{\langle \delta w(\delta,t)\rangle}{\langle w(\delta,t)\rangle}\ , \tag{14}$$

where the brackets represent the integral over δ. Note that $\overline{\delta} = \overline{D}_{ax} + \overline{D}_{syn} - \overline{D}^{b}_{den}$. The dynamics of $\overline{\delta}$ is obtained by differentiating (14) with respect to time,

$$\dot{\overline{\delta}} = \frac{\langle \delta\dot{w}\rangle\langle w\rangle - \langle\delta w\rangle\langle\dot{w}\rangle}{\langle w\rangle^2} = \frac{1}{\langle w\rangle}(\langle\delta\dot{w}\rangle - \overline{\delta}\langle\dot{w}\rangle)\ , \tag{15}$$

and inserting for $\dot{w}$ the weight modification (13). As we show in Appendix A1 the equation can then be reduced to

$$\frac{d}{dt}\overline{\delta} \approx \sigma^2\psi'_\varsigma(\overline{\triangle t}+\overline{\delta})\ , \tag{16}$$

where ψ'_ς is the derivative of the learning function (9). The dynamics (16) has exactly two stationary solutions $\overline{\delta}_\alpha$ and $\overline{\delta}_\beta$ corresponding to the zeros $\overline{\triangle t}_{syn} = \overline{\triangle t}+\overline{\delta}_\alpha = -\sqrt{\alpha^2+\varsigma^2}$ and $\overline{\triangle t}_{syn} = \overline{\triangle t}+\overline{\delta}_\beta = \sqrt{\beta^2+\varsigma^2}$ of ψ'_ς (cf. Fig. 2a). To investigate the stability we consider the second derivative of ψ_ς at these points and find $\psi''_\varsigma(-\sqrt{\alpha^2+\varsigma^2}) < 0$ and $\psi''_\varsigma(\sqrt{\beta^2+\varsigma^2}) > 0$. We conclude that the delay $\overline{\delta}_\alpha$ is attracting with domain of attraction $(-\infty,\overline{\delta}_\beta)$ while the delay $\overline{\delta}_\beta$ is repulsive. The standard deviation of the delay distribution at the attracting steady state can qualitatively be approximated by

$$\sigma \approx \sqrt[4]{\frac{2\,\epsilon}{|\psi''_\varsigma(-\sqrt{\alpha^2+\varsigma^2})|}} = \left(\frac{\epsilon\sqrt{e}}{\gamma}\right)^{\frac{1}{4}} \frac{\alpha}{\sqrt{\alpha^2+\varsigma^2}}\ , \tag{17}$$

see Appendix A2. Since the σ steeply increases with small ϵ we conclude from (16) that small delay fluctuations are enough to make the delay drifting. Moreover, due to the term ψ''_ς, the steady state weight distribution is broad for learning functions with a flat peak.

To summarize, for fixed and non-flat pre- and postsynaptic spike correlations the synaptic pathways develop such that a presynaptic signal, on average, meets a postsynaptic signal at the synaptic site with a time difference of $\triangle t_{syn} = -\sqrt{\alpha^2+\varsigma^2}$ ms, where α is the peak of the LTP function and ς is the width of the Gaussian part in the spike correlation function (7). In particular, the average spike time difference $\overline{\triangle t}$ can be compensated by appropriate delays D_{ax} and D^b_{den} as long as, on average, the presynaptic signals arrive at the synaptic site not later than $\sqrt{\beta^2+\varsigma^2}$ ms after the backpropagated spike (cf. Fig. 2). Two assumptions are crucial for these delay adaptations: (1) the small stochastic fluctuations in the delays and (2) the negative integral over the learning function leading to a normalization of the synaptic weights through the uncorrelated spikes.

Simulation results (Fig. 3) To examine the quality of the approximation (16) we simulated a stochastic version of (13) with discrete weight updates and discrete random delay fluctuations after each pairing of pre- and postsynaptic spikes. The spike time differences $\triangle t = t_A - t_B$ were sampled from a Gaussian distribution with mean $\overline{\triangle t} = -20\,\mathrm{ms}$, standard deviation $\varsigma = 3\,\mathrm{ms}$, and a sampling rate of $\overline{f}_{pre} = 20\,\mathrm{Hz}$. We have chosen 7 delay lines with axonal delays D_{ax} evenly distributed between $9.4 - 11.6\,\mathrm{ms}$ and $17.4 - 18.6\,\mathrm{ms}$, respectively (left and right squares in Fig. 3a). The synaptic delay and the backward dendritic delay were fixed to one millisecond, $D_{syn} = D^b_{den} = 1$, so that $\delta = D_{ax}$. After each spike pair we evaluated the learning function (2) and set for the weight change $\triangle w(\delta,t) = w(\delta,t)\psi_\circ(\triangle t+\delta)$. The parameters for $\psi_\circ$ were $\alpha = 5\,\mathrm{ms}$, $\beta = 7\,\mathrm{ms}$, and $\gamma = 3.5$. The allowed range for the relative delays δ was between $\delta_{min} = 9$ and $\delta_{max} = 21\,\mathrm{ms}$ with mesh width $\triangle\delta = 0.2\,\mathrm{ms}$. The additional stochastic delay fluctuations were implemented by replacing a synaptic weight $w(\delta,t)$ after each presynaptic spike with probability $\epsilon = 0.1$ by one of its two neighbors $w(\delta \pm \triangle\delta, t)$, cf. Eq. 3. Finally, the different weights were reduced after each pairing by $-c_1 w(\delta,t)\,\langle w\rangle$ with $c_1 = 0.3$ and $\langle w\rangle = \sum_{i=0}^{60} w(\delta_{min} + i\,\triangle\delta,\, t)\triangle\delta$. The remaining parameter in (13) was set to $c_2 = 0$, thus assuming $\overline{f}^{o}_{post} = 0$.

The simulation was run for 100 sec biological time (which is determined by the learning rate γ). From both rectangular initial distributions the weights $w(\delta,t)$ converged to a Gaussian distribution centered at $\overline{\delta} = \overline{D}_{ax} = 14.2\,\mathrm{ms}$ with standard deviation $\sigma = 0.8$ (Fig. 3a). The time course of $\overline{D}_{ax}$, obtained by evaluating (14) with the corresponding weights from the simulation, can be matched with the dynamics (16) with $\sigma = 0.2$ (Fig. 3b). Although, due to the different approximation steps, the two σ's do not coincide (the one from 17, for comparison, would be 0.31), the steady state for the average axonal delay ($\approx 14\,\mathrm{ms}$) is well predicted by the dynamics (16). Note that the jitter in the spike times shortens the average axonal delay (dot in Fig. 3b).

5 Induced delay shift for variable spike correlations

We next consider an 'unsupervised' learning scenario according to which the presynaptic neuron may directly influence the postsynaptic spike time t_B, which in turn affects the synaptic strength from the considered presynaptic neuron. To investigate this feedback loop we first consider the simplified scenario where the postsynaptic spike is exclusively triggered by the presynaptic input. We further simplify matters by fixing the axonal and synaptic delays to some constant value, say $D_{ax} = D_{syn} = 0$, and focus on the adaptation of the dendritic delays and latencies.

To take account of the dependencies among the different dendritic delays we parameterize the synapses on the dendritic tree by their distance D from the soma and identify this distance by the forward dendritic delay, $D^f_{den} \equiv D$. Since distal synapses induce a longer (and smaller) somatic EPSP we assume for the rise time a monotonically increasing function of the distance, $t_{rise} = R(D)$. Likewise, for the backward dendritic delay we assume a monotonically increasing function, $D^b_{den} = B(D)$. The somatic EPSP induced by synapse D is assumed to have the form

$$E_{R(D)}(t) = \frac{t}{R^2(D)} \cdot e^{-\frac{t}{R(D)}}\,, \quad \text{for } t \geq 0\,, \tag{18}$$

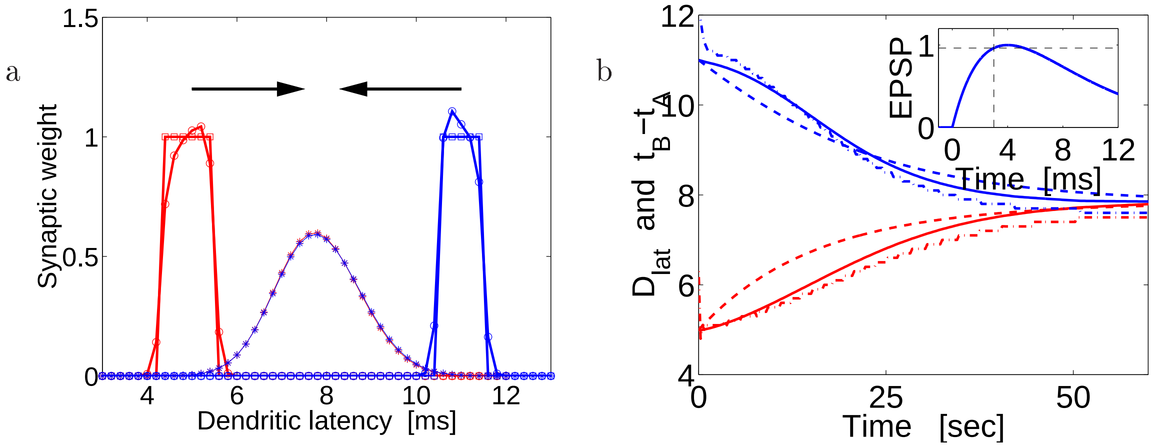

Figure 4: *Adaptation of the dendritic latency, with postsynaptic spikes triggered by the specific presynaptic neuron.* **a:** *Evolution of the weight distribution as a function of the dendritic latency* $D^{lat}_{den} = D^{f}_{den} + t_{rise}$*, with snapshots of the initial distributions (left and right squared curves), after* $120ms$ *(slightly deformed circled curves), and after* $60sec$ *(nearly identical Gaussians centered at* $\overline{D}^{lat}_{den} \approx 8\,ms$*).* **b:** *Time evolution of the average dendritic latency* $\overline{D}^{lat}_{den} = \overline{D}^{f}_{den} + t_{rise}$ *corresponding to the two simulations in (a) (full lines) and the approximation (22) with* $\sigma = 0.2$ *(dashed lines). Superimposed is the evolution of the spike time difference* $t_B - t_A$ *(dashed-dotted lines) for the two simulations in (a). The average dendritic latency is implicitly adapted such that in the steady state the argument of* $\psi'_\circ$ *is equal to the first zero,* $\overline{D}^{den}_{tot} \equiv \overline{D}^{lat}_{den} + \overline{D}^{b}_{den} = \alpha$ *(cf. Eq. 22 and Fig. 2a). In fact, from* $\overline{D}^{lat}_{den} \approx 8$*,* $t_{rise} = 4$ *and* $\overline{D}^{b}_{den} = \frac{1}{2}\overline{D}^{f}_{den}$ *one calculates* $\overline{D}^{b}_{den} \approx 2$*,* $\overline{D}^{den}_{tot} \approx 10$*, and this is in the range of* $\alpha = 10.5$ *(cf. also Fig. 2b). The inset shows the normalized EPSP according to (18) with the threshold crossing at* $\frac{3}{4}t_{rise}$ *(horizontal and vertical dashed lines).*

and $E_D(t) = 0$ for $t < 0$ (see inset of Fig. 4b). To obtain the time course of the somatic voltage induced by a single presynaptic spike at time t_A we integrate over all synaptic connections,

$$V(t) = \int_0^\infty w(D,t)\, E_{R(D)}(t - (t_A + D))\, dD\,. \tag{19}$$

where $w(D,t)$ denotes the synaptic weight of the connection from neuron A to neuron B with forward dendritic delay D (recall that $D_{ax} = 0$). We further assume that the postsynaptic neuron is sufficiently depolarized such that the input from neuron A alone may push the postsynaptic voltage across some fixed threshold θ. The postsynaptic spike time t_B is then given by the first crossing of θ. If θ is in the upper third of the maximal depolarization (19), say (cf. also inset in Fig. 4), and the delay distribution is not too wide, we may roughly estimate

$$t_B \approx t_A + \overline{D} + \overline{R} = t_A + \overline{D}^{lat}_{den}\,. \tag{20}$$

Here, $\overline{D}$ and $\overline{R}$ represent the average forward dendritic delay and the average rise time of the somatic EPSPs, respectively, and $\overline{D}^{lat}_{den} \equiv \overline{D} + \overline{R}$ is the average dendritic latency. The averages are taken with respect to the synaptic weight distribution. For instance, for the average dendritic latency we have

$$\overline{D}^{lat}_{den}(t) = \frac{\int_{-\infty}^{\infty} D^{lat}_{den}(D)\, w(D,t)\, dD}{\int_{-\infty}^{\infty} w(D,t)\, dD} \tag{21}$$

with $D^{lat}_{den}(D) = D + R(D)$. To estimate the shift in $\overline{D}^{lat}_{den}$ induced by the synaptic modifications we have to relate $\overline{D}^{lat}_{den}$ with the average synaptic time difference $\overline{\triangle t}_{syn}$. From (1) we get $\triangle t_{syn}(D) = t_A - (t_B + B(D))$. Setting $t_A = 0$ and inserting (20) yields $\overline{\triangle t}_{syn} = -\overline{D}^{lat}_{den} - \overline{B}$, where $\overline{B}$ is the average backward dendritic delay defined similarly to (21). If the synaptic weights $w(D,t)$ are now subject to the dynamics (12), with δ replaced by D, we obtain (see Appendix A3)

$$\frac{d}{dt}\overline{D}^{lat}_{den} \approx -\sigma^2\, \eta\, \psi'_\circ(-\overline{D}^{lat}_{den} - \overline{B}) \;, \quad \text{with } \eta = \frac{B'(\overline{D})\,(1 + R'(\overline{D}))}{1 + R'(\overline{D}) + B'(\overline{D})}\;. \tag{22}$$

Here, σ represents the standard deviation of the distribution of the forward dendritic delays D. It is given by (17) with $\varsigma = 0$.

The unique attracting steady state of (22) is again given by the zero of $\psi'_\circ$ with $\psi''_\circ < 0$, i.e. by $\overline{\triangle t}_{syn} = -\overline{D}^{lat}_{den} - \overline{B} = -\alpha$. Thus, assuming small unbiased fluctuations in the synaptic positions (encoded by D) together with repeatedly supra-threshold input from the considered presynaptic neuron, the synaptic population slowly shifts towards an average position $\overline{D}$ with total average dendritic delay $\overline{D}^{den}_{tot} \equiv \overline{D}^{lat}_{den} + \overline{B} \approx \alpha$. Observe that the derivative of $B(\overline{D})$ enters as a factor in (22) and that therefore the postsynaptic delay is only adapted if the backward delay is a non-constant function of the synaptic position. The reason is that we assume a simultaneous occurrence of the synaptic releases and, if the backward dendritic delay would be the same for all synapses, each synapse would see the same local time difference between the forward and backward signal. For $B(D) = \text{const}$, therefore, no shift in the average dendritic latency is expected. Note further that the sign in (16) and (22) is different. This reflects the fact that, in order to change from an initial $\overline{\triangle t}_{syn}$ with $-\alpha < \overline{\triangle t}_{syn} < 0$ to the final $\overline{\triangle t}_{syn} = -\alpha$, either the axonal delay has to decrease, as in (16) (at fixed spike times t_A, t_B), or the dendritic latency has to increase, as in (22) (and thereby increasing t_B).

Simulation results (Fig. 4) To qualitatively test the dynamics (22) for the average dendritic latency $\overline{D}^{lat}_{den}$ we simulated again the discretized version of (13). We have initially chosen 8 and 7 delay lines with dendritic latencies D^{lat}_{den} equally spaced within $4.2 - 5.6$ ms and $10.4 - 11.6$ ms, respectively (see Fig. 4). The dendritic latency was decomposed of a constant EPSP rise time $R(D) = t_{rise} = 4$ ms and a corresponding forward dendritic delay D with $D^{lat}_{den} = D + t_{rise}$. The backward dendritic delays were half the forward ones, $B(D) = \frac{1}{2}D$. We further set $\varsigma = 0$ and $D_{ax} = D_{syn} = 0$. The presynaptic cell was stimulated with a frequency of 20 Hz and whenever the sum of EPSPs crossed the threshold $\theta = 6.5$, cf. Eq. 19, a postsynaptic spike was elicited. The weight changes were calculated by evaluating $\psi_\circ$ at $\triangle t_{syn} = t_A - t_B - B(D)$, where t_A and t_B are the corresponding spike times of the pre- and postsynaptic neuron. The parameters in $\psi_\circ$ were $\alpha = 10.5$, $\beta = 14$, $\gamma = 0.7$. To

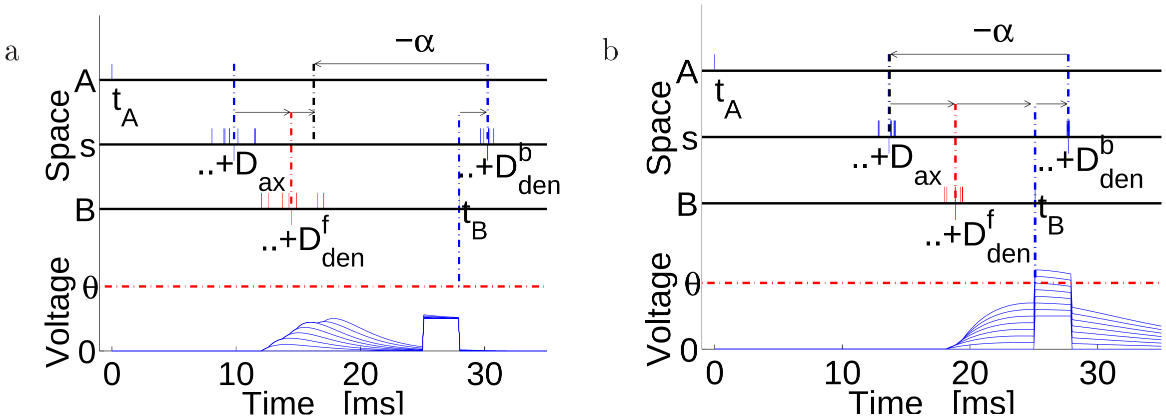

Figure 5: *Sketch of the joint adaptation of the axonal and dendritic delays in the postsynaptic priming protocol.* **a:** *Configuration of the different delays before learning.* **b:** *Repetitive stimulations of neuron A together with an increased background firing rate ('priming') of neuron B during the subsequent depolarization* $|\overline{\triangle t}|$ *later will lead to a drift of the axonal and dendritic delays such that eventually the average total dendritic delay matches the width of the learning function,* $\overline{D}^{den}_{tot} \equiv \overline{D}^{f}_{den} + t_{rise} + D^{b}_{den} \approx \alpha$*, and the average total forward delay matches the time delay of the depolarization,* $\overline{D}^{f}_{tot} \equiv \overline{D}_{ax} + \overline{D}^{f}_{den} + t_{rise} \approx |\overline{\triangle t}|$.

implement the other terms in (13) we reduced the weights by $-c_1 w(D,t)\langle w\rangle$ with $c_1 = 0.3$ and $c_2 = 0$, and implemented the stochastic delay fluctuations as described in the previous section.

The simulation was run for 60 sec biological time. From both initial distributions the weights $w(D,t)$ converged to a Gaussian distribution centered at $\overline{D} = 4.8$ ms (Fig. 4a). The time course of the spike time differences $\triangle t = t_A - t_B$ and of the average dendritic latency $\overline{D}^{lat}_{den}$ extracted from the simulation described above are shown in Fig. 4b. Superposed is the evolution of $\overline{D}^{lat}_{den}(t)$ obtained from the approximated dynamics (22). According to the steady state of (22), the average total dendritic delay, $D^{den}_{tot} \equiv \overline{D}^{lat}_{den} + \overline{B(D)} \approx \overline{D} + t_{rise} + B(\overline{D}) \approx$ 11.2 ms, is related to the first zero of $\psi'_\circ$, i.e. to $\alpha = 10.5$. The difference arises from the fact that the postsynaptic spike is typically triggered before the EPSP culmination, and the effective dendritic latency is therefore smaller than the one considered (see inset of Fig. 4b). Note that the time course of $\overline{D}^{lat}_{den}$ is roughly 3 times slower than that of $\overline{D}_{ax}$ (Fig. 3b) as predicted by the factor $\eta = \frac{1}{3}$ obtained from the formula in (22).

6 Co-evolution of axonal delays and dendritic latencies

In the previous simulations we fixed either the dendritic or the axonal delay while only fluctuations in the other delay were considered. Under this restriction we showed that each of the delays will separately evolve until the optimal timing imposed by the learning function is reached. For fixed dendritic delays and fixed spike correlations the axonal delays adapt such that the induced EPSPs peak at the time the postsynaptic neuron is expected to fire. For fixed axonal delays, in turn, the dendritic delays are adapted such that they finally match

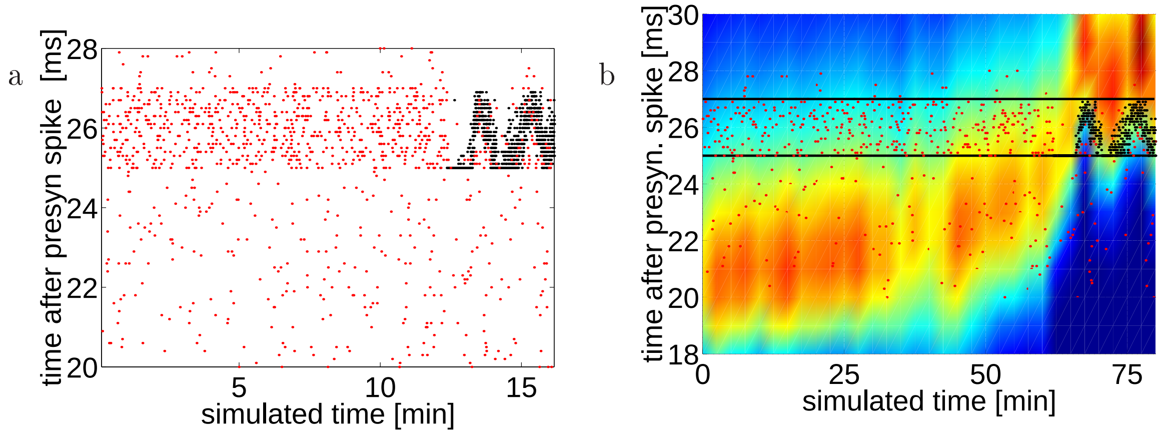

Figure 6: *Postsynaptic activity during the priming protocol (cf. Fig. 5)* **a:** *Raster plot of the postsynaptic spikes. Only each tenth trial is shown. During the time interval of* $25-27$ *ms after a presynaptic spike a postsynaptic depolarization increased the probability of a spontaneous postsynaptic spike. After* 28 *minutes of simulated time (corresponding to* $\sim 34'000$ *pairings) the presynaptic neuron learned to trigger the postsynaptic spike during the period of the postsynaptic depolarization (darker points top right).* **b:** *Same as in (a) but superposed with the sum of EPSPs induced by the presynaptic neuron. The horizontal lines indicate the period of the additional postsynaptic depolarization. For parameter values see text.*

the width of the learning window. Such a self-organization towards unique delays, however, is not possible anymore if the fluctuations would affect the axonal and the dendritic delays both at the same time. This is because the synaptic strength is changed only based on the local time difference measured at the synaptic site, and this may be the same for different pairs of axonal and dendritic delays. In fact, according to (1), the local time difference is the same for all axonal and backward dendritic delays with $\delta = D_{ax} - D^b_{den} = \text{const}$ and the stability condition $\triangle t_{syn} = \triangle t + \delta = -\alpha$ (see the discussion after (16) and (22)) is formally met for all pairs of the form $(D_{ax} + \epsilon, D^b_{den} + \epsilon)$ with $\epsilon > 0$. The same degeneracy is also present if D_{ax} confluently changes with the width α of the learning function. In either case synapses are going to be strengthened which are not causally contributing to the postsynaptic spike, although they see the pre- and postsynaptic signal in the optimal timing (i.e with α ms delay). It turns out that the unreliability of the synaptic transmission is the property which prevents an acausal development by disadvantaging synapses which are only "blind passengers" and do not contribute to the postsynaptic spike.

To reveal this point we first observe that, if the synaptic modification is induced e.g. via activation of the postsynaptic NMDA-receptors, a synaptic release and not just a presynaptic spike must occur to cause the synaptic modification. Let us now consider two unreliable delay lines with the same release probability $P_{rel} < 1$ and the same relative delay δ, but different axonal and dendritic delays and thus different total forward delays. Let us assume that the EPSP of the first delay line falls preferentially together with a subthreshold depolarization caused by a third input onto the same postsynaptic neuron. In this case the EPSP from the first delay line is expected to trigger more often a spike than the EPSP from the second.

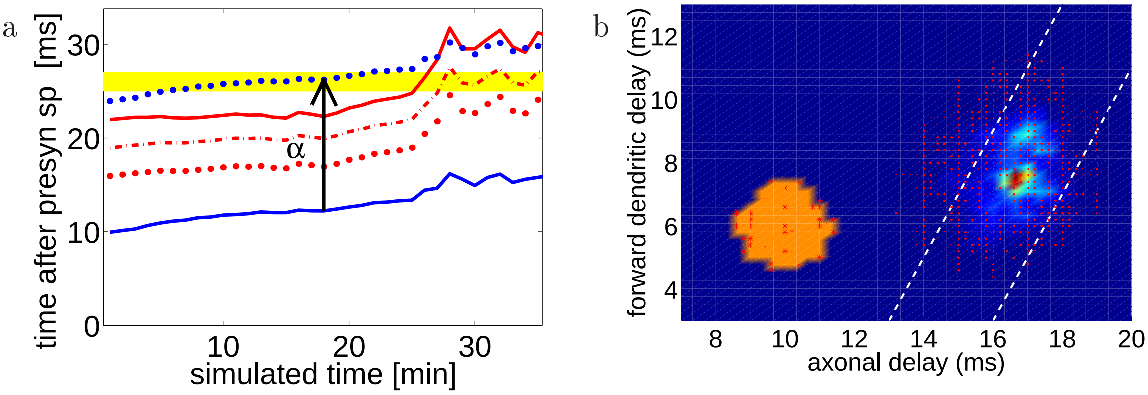

Figure 7: *Implicit adaptation of the different delays during the postsynaptic priming protocol (cf. Fig. 5).* **a:** *The evolution of the mean axonal delay and mean dendritic latency during the simulation shows that the synaptic modification implicitly adjusts the different delays according to (16) and (22), i.e. such that eventually the average total forward delay matches the average spike time difference,* $\overline{D}^f_{tot} = \overline{D}_{ax} + \overline{D}^{lat}_{den} = |\overline{\triangle t}| = 26\,\mathrm{ms}$*, and such that the total average dendritic delay matches the width of the learning function,* $\overline{D}^{den}_{tot} = \overline{D}^{lat}_{den} + \overline{D}^{b}_{den} \approx \alpha$ *(with* $\overline{D}^{lat}_{den} = \overline{D}^{f}_{den} + t_{rise}$*). (Lower fat line:* $\overline{D}_{ax}$*, dots:* $\overline{D}_{ax} + \overline{D}^{f}_{den}$*, dashed-dotted line:* $\overline{D}^{f}_{tot} = \overline{D}_{ax} + \overline{D}^{f}_{den} + t_{rise}$*, upper fat line:* $\overline{D}^{f}_{tot} + \overline{D}^{b}_{den}$*, upper dots:* $\overline{D}_{ax} + \alpha$*, shaded band: time of subthreshold input.)* **b:** *Distribution of the axonal and forward dendritic delays at the initial configuration (left blur) and after* 35 *minutes of the simulated time (right blur). The gray level is coding the synaptic weight corresponding to the different delays. Without transmission failures the final distribution of the delays would be smeared out within the two dashed lines.*

This implies that, compared to the second delay line, the synaptic release from the first line is stronger correlated with a postsynaptic spike and, since the occurrence of a release is a prerequisite to induce a synaptic change, the first delay line is more often upregulated. The second delay line which shows the same number of releases with the same local time difference is less upregulated since these releases do not exhibit the strong correlation with an immediately following spike. Additional uncorrelated postsynaptic spikes may downregulate both weights such that only the first delay line survives.

To formalize this reasoning in the general case we reparameterize the synaptic weights according to $w(D_{ax}, D)$, where $D \equiv D^f_{den}$ again abbreviates the forward dendritic delay. We consider a postsynaptic priming scenario with additional synaptic input from other neurons onto B roughly $\overline{\triangle t}$ after the activity of the presynaptic neuron A. In addition, we now assume that the individual releases may also influence the timing of the postsynaptic spikes in a statistical sense. To simplify matters we again assume a linear transfer function with threshold at 0 and, to reduce the number of constants, with gain 1. Instead of the correlation between the (instantaneous) pre- and postsynaptic firing rate it is now the correlation between the presynaptic release rate and the postsynaptic firing rate given a release, which determines the synaptic weight change. To formally describe the weight change of the synapse (D_{ax}, D) we introduce the postsynaptic firing rate $f_{post}(\tau, D_{ax}, D; t)$ conditional to a presynaptic spike at time $t+\tau$ *and* a subsequent release at synapse (D_{ax}, D). In extension of (10), this conditional postsynaptic firing rate is

$$\begin{aligned}\overline{f}_{post}(\tau, D_{ax}, D; t) &= \overline{f}_{pre} w(D_{ax}, D; t) E_{R(D)}(\tau + D_{ax} + D) + \ldots \\ \ldots + \; &P_{rel}\,\overline{f}_{pre} \textstyle\sum_{D'_{ax}\neq D_{ax}, D'\neq D} w(D'_{ax}, D'; t)\, E_{R(D')}(\tau + D'_{ax} + D') \; + \; \overline{f}^{o}_{post} \,.\end{aligned} \tag{23}$$

Note that, since we assume a release at synapse (D_{ax}, D), no factor P_{rel} occurs in the first term on the right hand side. In favor of the notational load we set $D_{syn} = 0$. The last term f^o_{post} represents the contribution of uncorrelated afferents to the postsynaptic firing rate. By considering the effect of the individual spikes onto the instantaneous postsynaptic rate we now obtain an additional dependency of the spike correlation function (6) on the individual synapses and the relative presynaptic spike times τ. Substituting $\overline{f}_{pre}(t)$ with $\overline{f}_{rel}(t)$ and $\overline{f}_{post}(t)$ with $f_{post}(\tau, D_{ax}, D; t)$ in (7) we obtain a conditional release-spike correlation function for synapse (D_{ax}, D) of the form

$$C_{(D_{ax},D)}(\tau; t) = c\,\mathcal{G}(\overline{\triangle t}, \varsigma, \tau) \; + \; \overline{f}_{rel}\,\overline{f}_{post}(\tau, D_{ax}, D; t)\,. \tag{24}$$

Recall that the first term represents the contribution to the correlation function caused by a postsynaptic spike which is induced by *other* input. Instead of (5) and (11) we now get

$$\begin{aligned}\tfrac{dw(D_{ax},D;t)}{dt} &= w(D_{ax}, D; t) \textstyle\int_{-\infty}^{\infty} d\tau\, \psi_\circ(\tau + \delta)\, C_{(D_{ax},D)}(\tau; t) \\ &\approx w(D_{ax}, D; t) \, \big[\, c\,\psi_\varsigma(\overline{\triangle t} + \delta) \; + \; (1 - P_{rel})\overline{f}_{pre}\, w(D_{ax}, D; t)\, \psi_{R(D)}(-D^{den}_{tot}) \\ &\ldots + \; \overline{f}_{rel} \textstyle\sum_{D'_{ax}, D'} w(D'_{ax}, D'; t) \psi_{R(D')}(-D'^{f}_{tot} + \delta) \; - \; \overline{\psi}_\circ\, \overline{f}_{rel}\, \overline{f}^{o}_{post} \Big] \,,\end{aligned} \tag{25}$$

where $D^{den}_{tot} = D + R(D) + B(D)$ is the sum of forward dendritic delay plus the EPSP rise time plus backward dendritic delay, $\delta = D_{ax} - B(D)$, and $D'^{f}_{tot} = D'_{ax} + D' + R(D')$ is the total forward delay of another delay line. The '$\approx$' in (25) comes from the approximation of the EPSPs by Gaussian functions of normalized area, with center $\tau + D'^{f}_{tot}$ and standard deviation $R(D')$, so that the integrals reduce to the evaluation of $\psi_{R(D')}$ given by (9).

For $P_{rel} = 1$ we recover in (25) our problem that the learning rule cannot distinguish between delay lines (D_{ax}, D) with common relative delay δ (since then the bracket in (25) is independent of D_{ax} and D). If the synapses are unreliable, however, the second term in the brackets becomes positive and the synaptic weights evolve differently for each delay line. Due to this positive feedback term, delay lines for which the value $\psi_{R(D)}(-D^{den}_{tot})$ is large have an evolutionary advantage upon the other delay lines with the same δ. Provided that $\sqrt{\alpha^2 + R(D)^2}$ does not vary too much the value of $\psi_{R(D)}$ is largest for that delay line satisfying $D^{den}_{tot} \equiv D + R(D) + B(D) \approx \sqrt{\alpha^2 + R(D)^2}$. Note that the analogous condition is satisfied for the attracting steady state of the dynamics (22) which we deduced under the assumption of fixed axonal delays. On top, the first term in the brackets favors delay lines for which $\psi_\varsigma(\overline{\triangle t}+\delta)$ is maximal and thus delays satisfying $\overline{\triangle t}_{syn} \equiv \overline{\triangle t}+\delta = \overline{\triangle t}+D_{ax}-B(D) \approx -\sqrt{\alpha^2 + \varsigma^2}$. Together with the previous condition this fixes the optimal axonal delay D_{ax} as well as the optimal forward dendritic delay D characterizing the delay line. Recall that the latter condition is also satisfied in the attracting steady state of the dynamics (16). Note further that combining the two conditions while assuming $R(D) \approx \varsigma$ yields $D^{f}_{tot} \equiv D_{ax} + D + R(D) \approx -\overline{\triangle t}$, and that this is also compatible with the third term within the bracket in (25). Thus, the learning rule prefers a total dendritic delay corresponding to the width of the effective learning function and a total forward delay corresponding to the the average spike time differences between the pre- and postsynaptic neuron (Fig. 5).

By comparing (11) with (25) we see that the negative feedback loop preventing the synaptic weights from growing to infinity is lost when we consider the spike correlation induced by the presynaptic neuron. The stability property is regained if we consider synaptic projections from other neurons being subject to the same type of synaptic modifications of the weights. In fact, if the spike activity among the presynaptic neurons has an uncorrelated component, an overall increase in the synaptic weights leads to an enhanced component $\overline{f}^{o}_{post}$ and this establishes a negative feedback loop through the last term in (25). Recall that a positive $\overline{\psi}_{\circ}$ is equivalent to a negative integral over the learning function $\psi_{\circ}$. Alternatively, we may postulate that with each postsynaptic spike the synaptic weight slightly degrades, independently of any spike timing (Kempter et al., 1999). In this case, we obtain an additional term in the brackets of (25) which is proportional to $-(\overline{f}^{o}_{post} + \overline{f}_{rel} \sum_{D'_{ax}, D'} w(D'_{ax}, D'; t))$. Since this term is equal for all delay lines it does only act as a normalization without distorting the weight distribution and therefore does neither change the average axonal nor dendritic delays. Finally, small unbiased stochastic fluctuations in the axonal and dendritic delay will cause the average delays to move towards the steady state given by the above conditions $D^{den}_{tot} \approx \sqrt{\alpha^2 + R(D)^2}$ and $D^{f}_{tot} \approx -\overline{\triangle t}$. Similarly to the last term in (12), these stochastic fluctuations will appear in (25) as diffusion term of the form $\epsilon \left(\frac{\partial^2 w}{\partial D_{ax}^2} + \frac{\partial^2 w}{\partial D^2} \right)$.

Simulation results (Fig. 6, 7) To test whether the unreliability of the synaptic transmission indeed helps to select a specific configuration among the 2-dimensional parameterization of delay lines we combined the 'supervised' and 'unsupervised' learning scenario from the previous sections. We initially distributed 20 delay lines with axonal delays D_{ax} between $8.5 - 11.5$ ms and forward dendritic delays $D \equiv D^{f}_{den}$ between $4.5 - 7.5$ ms and set for the corresponding synaptic weights $w(D_{ax}, D) = 0.4$ while the remaining weights were zero. The backward dendritic delay of each delay line was half the forward dendritic delay, $B(D) \equiv D^{b}_{den}(D) = \frac{1}{2}D$, and the EPSP rise time was set to $R(D) \equiv t_{rise} = 3$ ms. The common release probability was $P_{rel} = 0.5$ and the threshold of the postsynaptic neuron was set to $\theta = 20$ mV. We stimulated the presynaptic neuron with a periodic train of $\overline{f}_{pre} = 20$ Hz and primed the postsynaptic cell with a 10 mV subthreshold depolarization between $25 - 27$ ms after each presynaptic spike (cf. Fig. 5). This induces the additional spike correlation described in (24) with $\overline{\triangle t} = 26$ ms and $\varsigma \approx 1$ ms. Whenever the postsynaptic potential, composed of this depolarization and the sum of the evoked EPSPs (cf. Eq. 18 and 19), crossed the threshold, a postsynaptic spike was elicited, unless a spontaneous spike was already triggered within 5 ms before. The postsynaptic cell showed a spontaneous background firing rate $\overline{f}^{o}_{post} = 3$ Hz which was increased to an instantaneous Poisson rate of 30 Hz during the 2 ms period of the additional depolarization. For each pair of synaptic release and postsynaptic spike we modified the corresponding synaptic weight according to $\triangle w = w\, \psi_{\circ}(\triangle t_{syn})$ with learning function $\psi_{\circ}$ given by (2) and local time difference $\triangle t_{syn}$ given by (1). The parameters of $\psi_{\circ}$ were $\alpha = 14$ ms, $\beta = 20$ ms, and $\gamma = 0.8$. To prevent an infinite synaptic growth we reduced the synaptic weights after each postsynaptic spike by 0.01 times w times the sum of all synaptic weights as discussed at the end of the last paragraph. To simulate the stochastic fluctuations we assumed a 2-dimensional mesh-grid of axonal and dendritic delays (D_{ax}, D) with a granularity of 0.2 ms. After each presynaptic spike we redistributed each synaptic weight with probability 0.1 by 80% of its own weight and 5% of the weights of the 4 neighbors.

The simulation was run for 35 minutes biological time. Although initially the presynaptic neuron could not contribute to the firing of the postsynaptic cell, it learned to do so after 28 minutes of correlated spike activity (Fig. 6a). After this period the postsynaptic spikes were always evoked by the EPSPs from the presynaptic cell which were strengthened by the rule. The connections adapted their total forward delay such that $\overline{D}^f_{tot} \equiv \overline{D}_{ax} + \overline{D}^f_{den} + t_{rise} = 27.0 \approx -\overline{\triangle t} = 26$ ms, and this lead to a peak of the EPSPs during the period of the additional postsynaptic depolarization (Fig. 6b and dashed-dotted line in Fig. 7a).

The simulation shows that it is possible to implicitly learn the peak EPSP with a precision of almost 1 ms although the width α of the learning function itself is more than 10 times broader. To explain this fact we inspect the different dendritic delays after reaching the steady state (Fig. 7a). Towards 30 min simulation time, the total dendritic delay eventually fully covers the learning window, $\overline{D}^{den}_{tot} \equiv \overline{D}^f_{den} + t_{rise} + \overline{D}^b_{den} = 13.5 \approx \sqrt{\alpha^2 + (t_{rise})^2} = 14.3$ ms, as predicted by the theoretical consideration above (Fig. 7a, cf. sketch in Fig. 5). Independently of the width of the LTP branch, the total forward delay $\overline{D}^f_{tot}$ adapts towards the typical spike time difference $\overline{\triangle t}$, and thus eventually supports existing temporal structures in the spike activity, as long as the width α of the learning function can be absorbed by the total dendritic delay $\overline{D}^{den}_{tot}$.

Finally, the stochastic synaptic transmission ($P_{rel} < 1$) led to a unique average axonal delay $\overline{D}_{ax} \approx 17$ ms and a unique average forward dendritic delay $\overline{D}^f_{den} \approx 7$ ms after the application of the stimulation protocol (Fig. 7b). Without transmission failures ($P_{rel} = 1$) an elongated ridge would evolve in the (D_{ax}, D^f_{den})-space corresponding to those delays with fixed synaptic time difference $\triangle t_{syn}$ between the pre- and postsynaptic signals. This shows that unreliable synapses are in fact necessary for a unique development of the axonal and dendritic delays.

7 Discussion

Summary The present work extends the idea of delay selection by showing that slow stochastic fluctuations in the axonal and dendritic delays leads to delay drifts beyond the original delay distributions (cf. Figs. 3 and 4). We showed that during correlated pre- and postsynaptic activity the temporally asymmetric synaptic modification implicitly (1) adjusts the total forward delay composed of the axonal, synaptic, forward dendritic delay and EPSP rise time until it matches the statistically dominant time difference between the pre- and postsynaptic spikes, and (2) tunes the total dendritic delays composed of the forward dendritic delay, the EPSP rise time and the backward dendritic delay until it fits the width of the effective learning function (cf. Figs. 5, 6 and 7). Although the evolution of these delays is perturbed by uncorrelated spontaneous activity, the precision of the delay adaptation may be of an order higher than the width of the learning function itself. Interestingly, the tight selection of the optimal delays is only possible if the synaptic transmission is unreliable. This unreliability gives the "successful" delay lines an evolutionary advantage upon the others.

To stabilize the self-organization of the different delays one must assume that the integral over the learning function is negative. Thus, three properties are important for an activity-dependent development of axonal delays and dendritic latencies through asymmetric weight modifications: slow unbiased delay fluctuations, unreliable synapses, and LTD dominating LTP.

Physiological evidence for the hypothesized mechanisms The shape of the temporally asymmetric learning function (see Eq. 2 and Fig. 2) is motivated by different recent experiments (Markram et al., 1997; Zhang et al., 1998; Bi and Poo, 1998; Feldman, 2000). These experiments do not allow to discriminate between axonal, synaptic and dendritic delays as defined in (1). Instead, they provide the change in the synaptic strength as a function of the spike time differences $t_A - t_B$ between the two neurons recorded from. In these experiments the neurons lay nearby and, in case of cortical recordings, they mostly synapse on each others proximal dendritic tree. As a consequence, the axonal and dendritic delays appear to be small, and the time difference at which maximal upregulation is observed corresponds roughly to our time difference α defining the peak in the learning function (2). Depending on the experimental study this peak lays between -25 and 0 ms.

Different sources for the pre- and postsynaptic delays exist. Axonal propagation velocities for horizontal cortico-cortical connections are in the order of 0.2 m/s, corresponding to a propagation time of 50 ms for cells which are 1 cm apart, while the propagation velocities of vertical fibers are roughly 10 times faster (Bringuier et al., 1999). Considerable delays may also be present postsynaptically. In passive dendritic structures delays up to 30 ms and more were calculated for the time difference between the centroids of an injected current, the locally induced excitatory postsynaptic potentials (EPSP) and the EPSP propagated into the soma (Agmon-Snir and Segev, 1993). The peak-to-peak time from a distal EPSP to the forward propagated somatic EPSP, and from the somatic AP to the backpropagated dendritic AP can both reach up to 8 ms (Segev et al., 1992; Stuart and Sakmann, 1994). In addition to these peak to peak delays the signal is delayed by the rise time of the somatic EPSP which is in the order of $3 - 10$ ms and more. Moreover, in active dendritic structures transient potassium currents can delay the generation of a postsynaptic AP by tens up to hundreds of milliseconds (McCormick, 1991).

The small delay fluctuations we assume may come along with the appearance of new dendritic spines and synapses observed during induction of long-term modifications or their unspecific stochastic disappearing (Engert and Bonhoeffer, 1999; Toni et al., 1999). Other sources of delay fluctuations may be new axonal collaterals bifurcating from an existing synaptic connection and targeting onto a different dendritic subtree. The axonal and dendritic delays of such new connections may have slightly changed due to different channel densities or morphological properties of the corresponding axonal collateral or the dendritic subtrees (Engert and Bonhoeffer, 1999; Maletic-Savatic et al., 1999). Variable delays are particularly present at the developmental stage of the CNS during which neurites can grow with speeds up to 50 μm/h (Gomez and Spitzer, 1999). This speed is indirectly proportional to the frequency of the endogenous Ca^{2+}-transients and it is tempting to relate the Ca^{2+}-signals decelerating neurite growth with the Ca^{2+}-signals accelerating synaptic long-term changes once the connections are formed (see e.g. Bliss et al, 1993). Evidence for formation of new connections during the development is given by the functional differences between the axonal arborization of cortical V1 cells in young and adult cats. In the juvenile animal horizontal axon collaterals often form connections only in a restricted proximal region, while they are found on distal axon terminals in the adult animal (Katz and Shatz, 1996). In pyramidal cells dendritic latencies may change during the gated growth of the apical tree towards the pial surface where it integrates input from the superficial layers (Polleux et al., 2000). The hypothesis that the emergence and selection of interneuronal delays might be guided by a temporally asymmetric learning rule is further supported by the fact that this type of synaptic modification is particularly prominent in embryonic cultures (Bi and Poo, 1998; Bi and Poo, 1999) and in cortical slices of animals younger than 2-3 weeks (Markram et al., 1997; Feldman, 2000), while long-term plasticity cannot be found at synapses from the thalamic input after this stage (Crair and Malenka, 1995; Feldman et al., 1998).

There are two possible concerns related to recent experimental findings which are worth discussing. First, we would like to emphasize that short-term synaptic plasticity, as found between cortico-cortical pyramidal cells (Markram et al., 1998), does not fundamentally change the present picture of delay adaptation. Short-term depression, for instance, is believed to reduce the vesicle release probability as a function of the presynaptic activity, and hence could be easily incorporated in our analysis where the release probability explicitly enters. On a phenomenological level, synaptic depression pronounces changes in the presynaptic firing rates which may even support the temporal structure between the pre- and postsynaptic cells (compare also Tsodyks and Markram, 1996, Abbott et al., 1997, and Senn et al. 1998). In this context, the long-term modification of synaptic depression (Markram and Tsodyks, 1996) has a similar effect on delay selection as the long-term modification of the absolute synaptic strength. Second, recent findings show that for CA1 pyramidal neurons in the hippocampus the shape of the somatic EPSP is independent on the synaptic location on the dendritic tree (Magee and Cook, 2000). In our framework this implies that there is no one-to-one correspondence between the synaptic location and the peak of the learning function. It should be emphasized, however, that experimental support for the location independence of the synaptic response in mammals is only present in hippocampal neurons and, to some degree, in motoneurons (see Magee and Cook, 2000, for references). In contrast, neocortical layer 5 pyramidal cells, for instance, do actively attenuate synaptic inputs from the apical tree to the soma and show a wide range of somatic EPSP rise times (Berger et al., 2001). These neurons can well be integrator and coincidence detector in one (Larkum et al., 1999), and therefore require a mechanism which determines whether a synaptic input should supply contextual information (by a broad, distally generated EPSP) or timing information (by a sharp, proximally generated EPSP).

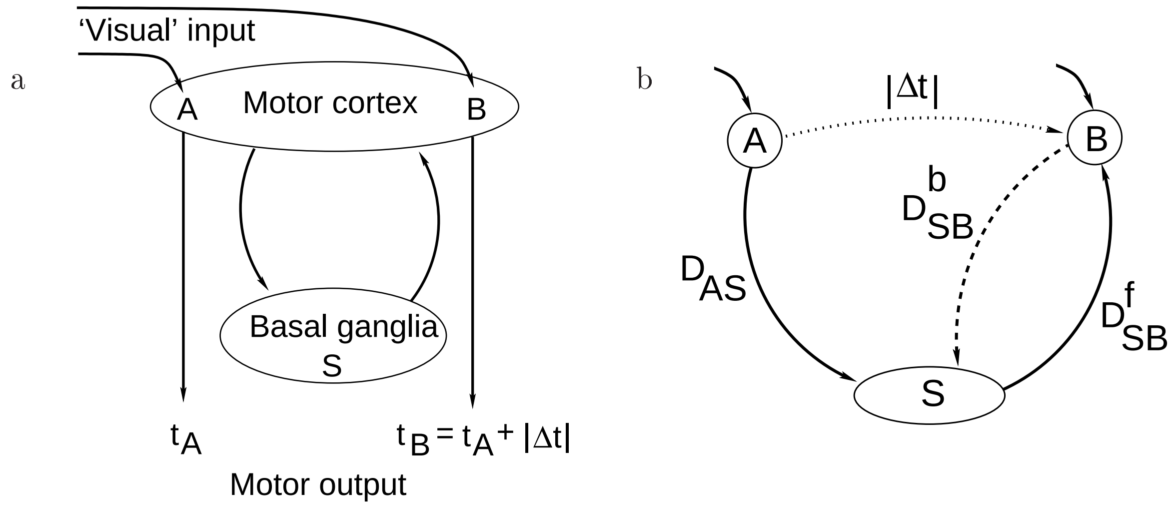

Figure 8: *Learning time delays between neuronal populations.* **a:** *Sequences of finger movements, e.g. for playing a music instrument, require the temporal control of activity in the motor cortex. For instance, when looking at the notes $a - b$, the visual stimulus activates through cortico-cortical pathways in sequence the populations A and B in the motor cortex at times t_A and $t_B = t_A + |\triangle t|$. When automatizing such finger sequences the corresponding activity pattern is believed to be stored in the basal ganglia (S) from where it can be retrieved through the basal ganglia – thalamocortical feedback loop.* **b:** *The temporal structure between the activities at the sites A, B and S in the motor cortex and the basal ganglia, respectively, is similar to that considered for a single synaptic connection (see Fig. 1a), although it may extend over a larger time scale. A reinforcement signal from B to S (mediated e.g. by dopamine projections) modifying the strength of the 'relay' S according to a temporally asymmetric rule may similarly select pathways through S which support a sequence of activities in the upper area. Assuming a variety of delay lines with stochastic transmission, our analysis predicts that after repeated presentation of the visual stimulus sequence, the total forward delay through S will eventually match the time difference between the cortical populations A and B, $D_{AS} + D^f_{SB} = |\triangle t|$. This is even possible in the presence of a delay D^b_{SB} of the reinforcement signal. In this case, the forward delays are selected such that in addition $D^f_{SB} + D^b_{SB} = \alpha$ holds, where α is the peak in the learning function of the 'relay' S (cf. also Fig. 5).*

Some predictions Synaptic delay lines and their modification have recently been the focus of an experimental study in cultured networks of embryonic rat hippocampus (Bi and Poo, 1999). This work shows that the different mono- or poly-synaptic delay lines between two neurons may indeed play a crucial role in the induction of either LTP or LTD. If paired pulse stimulation of a presynaptic neuron triggers an AP in a postsynaptic neuron through a specific pathway, then the strength of that and putative shorter pathways are upregulated while pathways with putative longer delays are downregulated. Our present investigation suggests that simultaneous single pulse stimulations at two different locations with interpulse intervals of 10, 30, and 50 ms, would shift, after minutes of repetitions, the (axonal) delays of the pathways towards 10, 30, and 50 ms, respectively. Beside studying network effects it would be interesting to investigate *in vitro* synaptic modifications at neurons with delayed responses, and to test whether the learning windows are in fact broadly tuned to match the dendritic latencies. In the context of subthreshold receptive fields (Bringuier et al., 1999), our work predicts an asymmetric distribution of horizontal propagation velocities in the primary visual cortex of cats reared under unidirectional background motions (cf. Introduction). Another prediction is that, due to the self-organization of the synaptic locations, distal synapses with slow somatic EPSPs should have broader learning functions than proximal ones with fast somatic EPSPs. In fact, for certain connections to layer IV spiny stellate cells the peak of the spike-based temporal modification function is measured to be at $t_{post} - t_{pre} = 60\,\mathrm{ms}$ (Cowan and Stricker, 1999).

A further example: learning delays between cortical areas Our analysis also applies to the problem of delay adaptation between neuronal populations with correlated activities. As an example we may consider the learning of finger sequences, say for playing a music instrument. The storage and recall of such sequences requires a temporal control of the activity in the motor cortex from which the appropriate spinal motor patterns are recruited. According to a current hypothesis, the automation of motor sequences is a form of 'habit' learning which involves the basal ganglia - thalamocortical feedback loop (Petri and Mishkin, 1994; Rolls and Treves, 1998). Such learning is driven by an external input and probably internally involves a delayed reinforcement signal. Let us assume that looking in sequence to two notes, say $a - b$, the visual input eventually evokes through the parietal pathway a well timed activity at the sites A and B, respectively, in the motor cortex (see Fig. 8a, cf. also Ungerleider, 1995). During the many repetitions of the finger sequence, the basal ganglia are taught by the motor cortex to store a mirror image of the cortical spatio-temporal activity pattern. The input from the basal ganglia to the motor cortex which correctly predicts the ongoing visual input may evoke a reinforcement signal back to the basal ganglia, for instance through dopamine projections which strengthens the appropriate pathways through a temporally asymmetric rule. After learning, the motor sequence can be retrieved in the 'habit' mode by a sequence of look-ups from the basal ganglia, thereby freeing motor cortex for other, 'higher level' tasks.

The temporal structure between the two cortical populations and the corresponding memory trace in the basal ganglia is similar to that between a pre- and postsynaptic neuron and the connecting synapses (Fig. 8b). Since the reinforcement signal intrinsically has a delay which is beyond the temporal precision of the motion, an additional gating mechanism is required which takes this delay into account. Our analysis shows that a temporally asymmetric learning rule is enough to solve this timing problem. Put into a general framework, the problem is to adjust the delay from A to B, given that the tuning mechanism is located at some intermediate site S, and hence receives itself only delayed signals from A and B. It is not evident how such a delay adaptation can work at all. Summarized, the constraints are: (1) *Implicitness:* no explicit mechanism exists for a directed adaptation of the individual delays. Rather, the average delay of different nearby delay lines can only implicitly be adapted by strengthening or weakening the connections in the presence of slow unbiased delay fluctuations. (2) *Locality:* there is no bird's view which tells the time elapsed between the events at the two sites. Rather, only the time differences between a forward and a backward signal can be locally measured at the site S *in between* the two signal sources.

(3) *Directionality:* in adapting the delay line based on the local time differences one should take into account the backward delay from site B to site S, and the fact that this backward delay may be different from the corresponding forward delay. (4) *Recurrence:* the change in the connection strengths may itself influence the activity at location B and thus change the statistics of the temporal relationship between the signals at the two locations. As we showed, a temporally asymmetric learning rule may simultaneously cope with all these constraints.

Comparison with recent theoretical work The apparent importance of temporal relationships among cortical activities motivated different studies on delay adaptations. Hüning et al. (1998) suggested that synaptic delays should adapt proportionally to the negative temporal derivative of the EPSP with peak centered at the postsynaptic spike time (Hüning et al., 1998). Such an explicit delay adaptation which assumes the temporal modifications of intracellular messenger cascades, however, is difficult to motivate, apart from the fact that the effective synaptic delays measured in central synapses are in the range of 0.4 ms (Lloyd, 1960), a number which is down-corrected in a later study to 0.17 ms (Munson and Sypert, 1979). Other works suggested that delay changes should have qualitatively the same form as the learning function for the synaptic weights (Eurich et al., 1997). If deduced from the modification of the synaptic strength, however, the rule for the delay adaptation is neither given by the derivative of the EPSP, nor is it identical to the original weight modification. In contrast, as our analysis shows, the best motivated form for an explicit delay adaptation is given by the derivative of the original weight modification function (cf. Fig. 2a). In a series of other works, delay modifications are suggested to result from two independent processes, delay selection through weight modification and delay shifts through explicit delay adaptation (Eurich et al., 1998; Eurich et al., 1999). However, no physiological evidence for an explicit delay adaptation exist so far and, as we argue, no such rule is in fact necessary. In the presence of weak delay fluctuations and stochastic synaptic transmissions, a temporally asymmetric synaptic weight modification is enough to explain axonal and dendritic delay shifts. While the adaptation of the total forward delay occurs independently of the peak in the learning function, this peak implicitly determines the width of the somatic EPSP. Hence, during development, the rule controls axonal delays and dendritic latencies to support stimulus driven temporal structures and to generate different types of causal relationships between neuronal activities.

A Appendix

A1 *Proof of (16):* Defining the normalized density $p(\delta) = p(\delta, t) = \frac{w(\delta,t)}{\langle w \rangle}$ of connections from neuron B to neuron A with relative delay δ we may write

$$\overline{\delta} = \int_{-\infty}^{+\infty} \delta p(\delta) d\delta \equiv \langle \delta p(\delta) \rangle \ ,$$

where by definition $\langle . \rangle$ denotes the integral over δ. Inserting the expansion

$$\psi(\overline{\triangle t} + \delta) \approx \psi(\overline{\triangle t} + \overline{\delta}) + (\delta - \overline{\delta})\psi'(\overline{\triangle t} + \overline{\delta}) \qquad (26)$$

into the rule (13) and averaging with respect to δ we get

$$\langle\dot{w}\rangle \approx \langle w\rangle\psi(\overline{\triangle t}+\overline{\delta}) + \langle w(\delta-\overline{\delta})\rangle\psi'(\overline{\triangle t}+\overline{\delta}) - c_1\langle w\rangle^2 - c_2\langle w\rangle + \epsilon\langle\frac{\partial^2 w}{\partial\delta^2}\rangle \ . \tag{27}$$

This approximation is good if we assume that the width σ of the delay distribution is narrow compared to the width $2(\alpha+\varsigma)$ of ψ. Evaluating the integral in the last term of the right hand side we find $\langle\frac{\partial^2 w}{\partial\delta^2}\rangle = 0$, since we may assume that the derivative $\frac{\partial w}{\partial\delta}$ at the boundaries $\delta = \pm\infty$ vanishes. Since

$$\langle w(\delta-\overline{\delta})\rangle = \langle w\delta\rangle - \langle w\rangle\overline{\delta} = \langle w\rangle\frac{\langle w\delta\rangle}{\langle w\rangle} - \langle w\rangle\overline{\delta} = 0$$

the second term in (27) cancels as well. Multiplying the remainder of Equation(27) with $\overline{\delta}$ we get

$$\overline{\delta}\langle\dot{w}\rangle \approx \overline{\delta}\langle w\rangle\psi(\overline{\triangle t}+\overline{\delta}) - \overline{\delta}\langle w\rangle(c_1\langle w\rangle + c_2) \ . \tag{28}$$

By (13) and the Taylor expansion (26) we also approximate after averaging

$$\langle\delta\,\dot{w}\rangle \approx \langle\delta w\rangle\psi(\overline{\triangle t}) + (\langle\delta^2 w\rangle - \langle\delta w\rangle\overline{\delta})\psi'(\overline{\triangle t}) - \langle\delta w\rangle(c_1\langle w\rangle + c_2) + \epsilon\langle\frac{\partial^2(\delta w)}{\partial\delta^2}\rangle \ . \tag{29}$$

Again, the last term on the right hand side vanishes by partial integration with boundary conditions $w^{\pm\infty} = \frac{\partial w^{\pm\infty}}{\partial\delta} = 0$. The factor in the second term reduces to

$$\langle\delta^2 w\rangle - \langle\delta w\rangle\overline{\delta} = \frac{\langle\delta^2 w\rangle}{\langle w\rangle}\langle w\rangle - \frac{\langle\delta w\rangle}{\langle w\rangle}\overline{\delta}\langle w\rangle = (\overline{\delta^2} - \overline{\delta}^2)\langle w\rangle = \sigma^2\langle w\rangle \ ,$$

where σ^2 is the variation of the delay distribution given by the density $p(\delta)$. Using $\langle\delta w\rangle = \overline{\delta}\langle w\rangle$ Equation (29) simplifies to

$$\langle\delta\,\dot{w}\rangle \approx \overline{\delta}\langle w\rangle\psi(\overline{\triangle t}+\overline{\delta}) + \sigma^2\langle w\rangle\psi'(\overline{\triangle t}+\overline{\delta}) - \overline{\delta}\langle w\rangle(c_1\langle w\rangle + c_2) \ . \tag{30}$$

Subtracting (28) from (30) we obtain

$$\langle\delta\,\dot{w}\rangle - \overline{\delta}\langle\dot{w}\rangle \approx \langle w\rangle\sigma^2\psi'(\overline{\triangle t}+\overline{\delta})$$

and combining this with (15) we get the desired rule (16) governing the adaptation of the average relative delay.

A2 *Proof of (17):* We first show that the weight distribution is non-singular. To this end we derive the differential equation for σ^2 and show that for small σ^2 we always have $\frac{d\sigma^2}{dt} > 0$. Substituting (13) for $\frac{d\,w}{dt}$ one calculates

$$\begin{aligned}\frac{d}{dt}\sigma^2 &= \frac{d}{dt}(\overline{\delta^2} - \overline{\delta}^2) = \frac{d}{dt}\left(\frac{\langle\delta^2 w\rangle}{\langle w\rangle} - \frac{\langle\delta w\rangle^2}{\langle w\rangle^2}\right) = \ldots = \\ &= 2\,\epsilon + (\overline{\delta^2\psi} - \overline{\delta^2}\;\overline{\psi}) + 2\,\overline{\delta}\,(\overline{\delta}\,\overline{\psi} - \overline{\delta\psi}) \ ,\end{aligned}$$

where the overlining denotes the average with respect to the normalized density $p(\delta)$. At the course of this calculation we used that by partial integration with boundary conditions

$w^{\pm\infty} = \frac{\partial w^{\pm\infty}}{\partial \delta} = 0$ one finds $\langle \frac{\partial^2}{\partial \delta^2} w \rangle = \langle \delta \, \frac{\partial^2}{\partial \delta^2} w \rangle = 0$ and $\langle \delta^2 \, \frac{\partial^2}{\partial \delta^2} w \rangle = 2 \, \langle w \rangle$. Assuming that $\sigma^2 \to 0$ the density p would become a delta function and the average would commute with the product. Hence we would also have $(\overline{\delta^2 \psi} - \overline{\delta^2} \; \overline{\psi}) \to 0$ and $(\overline{\delta} \, \overline{\psi} - \overline{\delta \psi}) \to 0$ and therefore $\frac{d}{dt} \sigma^2 \to 2\epsilon$. But this contradicts the above assumption and we conclude that, in fact, σ^2 cannot shrink to zero.

We can now fit the non-singular steady state distribution with the Gaussian $\tilde{w}(\delta) = w(\overline{\delta}) \, \exp(-\frac{(\delta - \overline{\delta})^2}{2\sigma^2})$. Inserting $\tilde{w}$ for w in (13) and evaluating at $\overline{\delta}$ and $\overline{\delta} \pm \sigma$ yields $0 = \psi(-\sqrt{\alpha^2 + \varsigma^2}) - c_1 \langle \tilde{w} \rangle - c_2 - \epsilon/\sigma^2$ and $0 = \psi(-\sqrt{\alpha^2 + \varsigma} \pm \sigma) - c_1 \langle \tilde{w} \rangle - c_2$. Eliminating $c_1 \langle \tilde{w} \rangle + c_2$ and using the approximation $\psi(-\sqrt{\alpha + \varsigma}) - \psi(-\sqrt{\alpha + \varsigma} \pm \sigma) \approx -\psi''(-\sqrt{\alpha^2 + \varsigma^2}) \, \sigma^2/2$ gives $\psi''(-\sqrt{\alpha^2 + \varsigma^2}) \, \sigma^2/2 + \epsilon/\sigma^2 = 0$ from which (17) follows.

Finally, the total synaptic weight at the steady state can be qualitatively approximated by $\langle \tilde{w} \rangle \approx (\gamma \, e^{\frac{1}{2}} - \epsilon/\sigma^2)/c_1$ with σ given by (17). Since $\langle \tilde{w} \rangle = \sigma \sqrt{2\pi} \, w(\overline{\delta})$, the maximal weight at steady state is roughly $w(\overline{\delta}) \approx \frac{\psi(-\sqrt{\alpha^2 + \varsigma^2}) - \epsilon/\sigma^2}{c_1 \, \sigma \sqrt{2\pi}}$. •

A3 *Proof of (22):* For fixed presynaptic delays we have according to (20)

$$\frac{d \, \overline{D}_{post}}{d t} = \frac{d \, \overline{t_B}}{d t} = \frac{d \, \overline{t_B}}{d \, \overline{D}} \frac{d \, \overline{D}}{d t} \approx (1 + R'(\overline{D})) \frac{d \, \overline{D}}{d t} \,, \tag{31}$$

where $\overline{D}$ is the average forward dendritic delay and $R(\overline{D}) \approx \overline{R(D)}$ the average rise time. Since we are interested in replacing the factor $d\overline{D}/dt$ in (31) by a term containing the learning function $\psi_\circ$ we first replace the change in $\overline{D}$ by a change in $\overline{B}$ from where we bridge to $\psi_\circ$. Thus we first write

$$\frac{d \, \overline{D}}{d t} = \left(\frac{d \, B}{d \, \overline{D}} \right)^{-1} \frac{d \, B(\overline{D})}{d t} \,. \tag{32}$$

In the same way as we deduced (16) in Appendix A1 (note that there we have $\frac{d \, \overline{\triangle t}_{syn}}{d t} = \frac{d \, (\overline{\triangle t} + \overline{\delta})}{d t} = \frac{d \, \overline{\delta}}{d t}$) we may approximate the change of the average time difference $\overline{\triangle t}_{syn} = t_A + \overline{D}_{pre} - (\overline{t_B} + \overline{B})$ at the synaptic site by

$$\frac{d \, \overline{\triangle t}_{syn}}{d t} = - \frac{d \, \overline{t_B}}{d t} - \frac{d \, \overline{B}}{d t} \approx \sigma_B^2 \, \psi_\circ'(\overline{\triangle t}) \,, \tag{33}$$

where σ_B is the standard deviation of the distribution of $B(D)$. Solving (33) for $\frac{d \overline{B}}{d t}$, identifying $\frac{d \overline{B}}{d t} \approx \frac{d \, B(\overline{D})}{d t}$, and inserting into (32) yields

$$\frac{d \, \overline{D}}{d t} \approx \left(\frac{d \, \overline{t_B}}{dt} - \sigma_B^2 \psi_\circ' \right) \left(\frac{d \, B(\overline{D})}{d\overline{D}} \right)^{-1} \,. \tag{34}$$

Inserting (34) into (31) and solving for $\overline{t_B}$ we obtain

$$\frac{d \, \overline{t_B}}{d t} \approx - \sigma_B^2 \psi_\circ' \frac{B'^{-1} (1 + R')}{1 + B' + R'} \,.$$

Estimating the standard deviation of the backward delay, σ_B, by the one for the forward dendritic delay, $\sigma_B \approx B'(\overline{D}) \, \sigma$, yields the desired formula (22). •

Acknowledgments This study was supported by the Swiss National Science Foundation (grant 21-57076.99) and the Silva Casa foundation (WS). WS would like to thank Stefano Fusi for helpful discussions.

References

Abbott, L. & Song, S. (1999). Temporally asymmetric hebbian learning, spike timing and neuronal response variability. In: *Advances in Neural Information Processing Systems*, M. Kearns, S. Solla, & D. Cohn, ed., volume 11. MIT Press, Cambridge MA.

Abbott, L., Varela, J., Sen, K., & Nelson, S. (1997). Synaptic depression and cortical gain control. *Science*, 275:220–224.

Agmon-Snir, H. & Segev, I. (1993). Signal delay and input synchronization in passive dendritic structures. *J. Neurophysiol.*, 70(5):2066–2085.

Berger, T., Larkum, M., & Lüscher, H.-R. (2001). A high I_h channel density in the distal apical dendrite of layer V neocortical pyramidal cells increases bidirectional attenuation of EPSPs. *J. Neurophysiol.*, 85:855–868.

Bi, G. & Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neuroscience*, 18(24):10464–10472.

Bi, G. & Poo, M. (1999). Distributed synaptic modifications in neural networks induced by patterned stimulation. *Nature*, 401:792–796.

Bliss, T. & Collingridge, G. (1993). A synaptic model of memory: long term potentiation in the Hippocampus. *Nature*, 361:31–39.

Bringuier, V., Chavane, F., Glaeser, L., & Frégnac, Y. (1999). Horizontal propagation of visual activity in the synaptic integration filed of area 17 neurons. *Science*, 283:695–699.

Carr, C. & Friedman, M. (1999). Evolution of time coding systems. *Neural Computation*, 11:1–20.

Cowan, A. & Stricker, C. (1999). Long-term synaptic plasticity in the granular layer of rat barrel cortex. In: *Society For Neuroscience, Abstracts, No. 689.7*.

Crair, M. & Malenka, R. (1995). A critical period for long-term potentiation at thalamocortical synapses. *Nature*, 375:325–328.

Engert, F. & Bonhoeffer, T. (1999). Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature*, 399:66–70.

Eurich, C., Cowan, J., & Milton, J. (1997). Hebbian delay adaptation in a network of integrate-and-fire neurons. In: *Artificial Neural Networks - ICANN'97*, W. Gerstner, A. Germond, M. Hasler, & J.-D. Nicoud, ed., volume 1327 of *Lecture Notes in Computer Science*, pages 157–162. Springer, Berlin.

Eurich, C., Ernst, U., & Pawelzik, K. (1998). Continuous dynamics of neuronal delay adaptation. In: *Artificial Neural Networks - ICANN'98*, M. B. L. Niklasson & T. Ziemke, ed., pages 355–360. Springer, Berlin.

Eurich, C., Pawelzik, K., Ernst, U., Cowan, J., & Milton, J. (1999). Dynamics of self-organized delay adaptation. *Phys. Rev. Lett.*, 82(7):1594–1597.

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

Feldman, D., Nicoll, R., Malenka, R., & Isaac, J. (1998). Long-term depression at thalamocortical synapses in developing rat somatosensory cortex. *Neuron*, 21:347–357.

Gerstner, W. (1993). Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biol. Cybernetics*, 69:503–515.

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

Gomez, T. & Spitzer, N. (1999). In vivo regualation of axon extension and pathfinding by growth-cone calcium transients. *Nature*, 397:350–355.

Hopfield, J. & Brody, C. (2000). What is a moment? cortical sensory integration over brief intervals. *PNAS*, 97:13919–13924.

Hopfield, J. & Brody, C. (2001). What is a moment? transient synchrony as a collective mechanism for spatiotemporal integration. *PNAS*, 98:1282–1287.

Hüning, H., Glünder, H., & Palm, G. (1998). Synaptic delay learning in pulse-coupled neurons. *Neural Computation*, 10:555–565.

Katz, L. & Shatz, C. (1996). Synaptic activity and the construction of cortical circuits. *Science*, 274:1133–1138.

Kempter, R., Gerstner, W., & van Hemmen, J. (1999). Spike-based compared to rate-based Hebbian learning. In: *Advances in Neural Information Processing Systems 11*, S. S. et al., ed., page in press.

Kühn, R. & van Hemmen, J. (1991). Temporal Association. In: *Model of neural networks*, E. Domany, J. van Hemmen, & K. Schulten, ed., chapter 7, pages 213–280. Springer.

Larkum, M., Zhu, J., & Sakmann, B. (1999). A novel cellular mechanism for coupling inputs arriving at different cortical layers. *Nature*, 398:338–341.

Lloyd, D. (1960). Section I: Neurophysiology. In: *Handbook of Physiology*, H. Magoun, ed., volume II, pages 929–949. American Physiology Society, Washington, DC.

Magee, J. & Cook, E. (2000). Somatic EPSP amplitude is independent of synapse location in hippocampal pyramidal neurons. *Nature Neuroscience*, 3(9):895–903.

Maletic-Savatic, M., Malinow, R., & Svoboda, K. (1999). Rapid Dendritic Morphogenesis in CA1 Hippocampal Dendrites Induced by Synaptic Activity. *Science*, 283:1923–1927.

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

Markram, H., Pikus, D., Gupta, A., & Tsodyks, M. (1998). Potential for multiple mechanisms, phenomena and algorithms for synaptic plasticity at single synapses. *Neuropharmacology*, 37:489–500.

Markram, H. & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382:807–810.

McCormick, D. (1991). Functional properties of slowly inactivating potassium current in guinea pig dorsal lateral geniculate relay neurons. *J. Neuroscience*, 66:1176–1189.

Munson, J. & Sypert, G. (1979). Properties of single fibre excitatory post-synaptic potentials in triceps surae motoneurones. *J. Physiol. (London)*, 296:329–342.

Natschläger, T. & Ruf, B. (1998). Spatial and temporal pattern analysis via spiking neurons. *Network: Computation in Neural Systems*, 9:319–332.

Nieuwenhuys, R. (1994). The neocortex. An overview of its evolutionary development, structural organization and synaptolgy. *Anat Embryol*, 190:307–337.

Petri, H. & Mishkin, M. (1994). Behaviorism, cognitivism, and the neurophysiology of memory. *American Scientist*, 82:30–37.

Polleux, F., Morrow, T., & Glosh, A. (2000). Semaphorin 3A is a chemoattractant for cortical apical dendrites. *Nature*, 404:567–573.

Roberts, P. (1999). Computational consequences of temporally asymmetric learning rules: I. differential hebbian learning. *J. Computational Neuroscience*, 7:235–246.

Rolls, E. & Treves, A. (1998). *Neural Networks and Brain Functions*. Oxford University Press.

Segev, I., Rapp, M., Manor, Y., & Yarom, Y. (1992). Analog and digital processing in single nerve cells: dendritic integration and axonal propagation. In: *Single Neuron Computation*, T. McKenna, J. Davis, & S. Zornetzer, ed., pages 173–198. Academic Press Inc.

Senn, W., Segev, I., & Tsodyks, M. (1998). Reading neural synchrony with depressing synapses. *Neural Computation*, 10:815–819.

Senn, W., Tsodyks, M., & Markram, H. (2001). An algorithm for modifying neurotransmitter release probability based on pre-and post-synaptic spike timing. *Neural Computation*, 13(1):35–68.

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

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

Stuart, G. J. & Sakmann, B. (1994). Active propagation of somatic action potentials in cerebellar Purkinje cells. *Nature*, 367:69–72.

Toni, N., Buchs, P.-A., Nikonenko, I., Bron, C., & Muller, D. (1999). LTP promotes formation of multile spine synapses between a single axon terminal and a dendrite. *Nature*, 402:421–425.

Tsodyks, M. & Markram, H. (1996). Plasticity of Neocortical Synapses Enables Transitions Between Rate & Temporal Coding. In: *Proceedings of the ICANN'96*, C. von der Malsburg, ed., volume 112 of *Lecture Notes in Comp. Sci.*, pages 445–450. Springer.

Ungerleider, L. (1995). Functional brain image studies of cortical mechanisms for memory. *Science*, 270:769–775.

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