Introduction

Synaptic plasticity is believed to be the biological substrate of experience-dependent changes to the brain (Lynch 2004; Martin and Morris 2002). Therefore it is appropriate to wonder how long synaptic memory traces last and how memory lifetime is regulated. Various ways have been suggested to create plastic yet stable memory systems, for instance, by combining slow (cortical) and fast (hippocampal) learning systems, or using neuro-modulators to adjust learning rates, while more recent studies have focused on receptor stability in the postsynaptic membrane. In this modeling study, we ask how the plasticity rules themselves affect the memory retention and how synapses should retain previously learned modifications despite subsequent ongoing activity.

We study this question using phenomenological models of spike-timing-dependent plasticity (STDP). STDP is the observation that synapses change their efficacy depending on the precise timing difference between presynaptic and postsynaptic spikes (Bi and Poo 1998; Levy and Steward 1983; Markram et al. 1997; Sjöström et al. 2001). STDP has been observed in many systems (Abbott and Nelson 2000) and is thought to play a key role in receptive field development (Mu and Poo 2006; Young et al. 2007) as well as adult visual plasticity (Dan and Poo 2006; Yao and Dan 2001). Memory persistence is a particularly prominent problem with STDP, as in its naïve form STDP implies that any pre/post spike pair can modify the synapse, potentially erasing memories.

STDP has attracted intense theoretical interest (Davison and Fregnac 2006; Gerstner et al. 1996; Kempter et al. 1999; Kistler 2002; Kistler and van Hemmen 2000; Levy 1996; Pfister and Gerstner 2006; Roberts 1999). It leads to receptive field development (Delorme et al. 2001; Masquelier and Thorpe 2007; Song and Abbott 2001) and maximizes mutual information (Toyozumi et al. 2007), while being consistent with the BCM rule (Izhikevich and Desai 2003; Pfister and Gerstner 2006; Shouval et al. 2002). An early and widely used STDP model modifies the synapses as a function of the time difference between pre- and postsynaptic spikes only, independently of the synaptic weight (Song et al. 2000). This non-weight-dependent STDP (nSTDP) requires imposing upper and lower bounds on the weights to prevent unlimited weight growth. nSTDP gives rise to strong competition between inputs to a neuron; this is reflected in a bimodal synaptic weight distribution, which selects certain inputs above others even in the absence of structured input.

In contrast, weight-dependent STDP (wSTDP) incorporates the observation that strong synapses are harder to potentiate than weak ones (Bi and Poo 1998; Debanne et al. 1996, 1999; Montgomery et al. 2001). Interestingly, this small modification eliminates the need for weight bounds and gives rise to a unimodal weight distribution (Rubin et al. 2001; van Rossum et al. 2000). This distribution closely matches the weight distributions observed experimentally (O’Brien et al. 1998; Song et al. 2005; Turrigiano et al. 1998) and thus wSTDP is perhaps more realistic. (An alternative explanation is that the weak weights in nSTDP are silent synapses or too weak to be measured.) However, in contrast to nSTDP, wSTDP has weaker competition. The dichotomy between nSTDP and wSTDP is not strict, and intermediate models have been proposed that combine stronger competition with stable learning (Gutig et al. 2003; Meffin et al. 2006; Morrison et al. 2007; Toyozumi et al. 2007); the nSTDP and wSTDP learning rules can be seen as limiting cases.

Recent studies of supervised learning rules have concentrated on erasure of old memories as a result of storing new ones (Barrett and van Rossum 2008; Fusi and Abbott 2007). In contrast here we investigate the persistence of synaptic weights subject to unsupervised wSTDP or nSTDP learning and how quickly changes in weights are erased by ongoing activity. We find that the precise learning rule has a very strong influence on the memory retention time. Second, we consider the formation and the stability of receptive fields in networks with STDP learning. We show that despite its lack of intrinsic competition, wSTDP can lead to the formation of receptive fields provided there is sufficient lateral inhibition in the network. Furthermore the stability of the receptive fields is modulated by the strength of lateral inhibition, suggesting a novel role for inhibition in network plasticity.

Part of these results were presented earlier in abstract form (Billings and van Rossum 2006).
METHODS

Single-neuron simulations

For single-neuron simulations, we use a leaky integrate and fire (LIF) neuron with membrane potential \( V(t) \) dynamics governed by:
\[
\tau_m \frac{dV(t)}{dt} = -V(t) + I(t) + R_n g(t),
\]
where \( I(t) \) is the input current to the neuron. The neuron fires when the membrane potential reaches a threshold value \( V_{th} \) and on firing resets to its resting value \( V_r \). The parameters are: membrane time constant \( \tau_m = 20 \) ms, threshold potential \( V_{th} = -54 \) mV, resting potential \( V_r = -74 \) mV, input resistance \( R_n = 100 \) M\( \Omega \) (Song and Abbott 2001). The neuron receives current inputs through 800 excitatory synapses. These excitatory AMPA-like synapses have an exponential time course with a time constant of 5 ms and a reversal potential \( V_0 = 0 \) mV. The input to the neuron at any time is the sum of the contributions from all inputs \( I(t) = \sum_i w_i g_i(t) [V_0 - V(t)] \), where \( g_i(t) \) is an exponential function representing the synaptic time constant and \( w_i \) is the synaptic weight. For the parameters detailed in the following about 30 inputs with average weight need to be simultaneously active to raise the membrane to spiking threshold from rest.

The input spike trains have Poisson statistics. Each input has a firing rate drawn from a Gaussian distribution of 10 ± 4 (SD) Hz. At the end of a random time interval, drawn from an exponential distribution with a mean of \( \tau_r = 20 \) ms, the rates are re-drawn from the Gaussian distribution. This ensures that the correlation between any two inputs \( i(t) \) and \( j(t) \) is proportional to \( \exp(-|t - t'|/\tau_r) \). This correlation was chosen in a previous study in rough analogy with input to the visual system (Song and Abbott 2001): to allow direct comparison, we use the same correlation structure here.

Implementation of the plasticity models

In STDP learning rules, the weight modification depends on the timing difference between pre- and postsynaptic spikes. In the nSTDP rule, the weight change is independent of the weight itself (Song et al. 2000). The weight change due to a presynaptic and postsynaptic spike pairing is given by
\[
\Delta w = \begin{cases} a_+ \exp(-s_{mn}/\tau_p) & \text{if } s_{mn} > 0 \\ -a_- \exp(s_{mn}/\tau_d) & \text{if } s_{mn} < 0 \end{cases}
\]
where \( s_{mn} = t_{pre}^{(m)} - t_{post}^{(n)} \) is the time difference between post- and presynaptic spikes with times labeled \( m \) and \( n \). The constants \( A_+ \) and \( A_\pm \) set the amount of potentiation and depression, respectively, while \( \tau_p \) and \( \tau_d \) set the duration of the potentiation and depression plasticity windows. The plasticity windows are exponential with \( \tau_p = 20 \) ms unless otherwise stated. Furthermore, we set \( A_- = 1 \) pS and a slightly larger \( A_+ = A_- + \varepsilon \). Like many plasticity rules, weights diverge unless hard upper and lower bounds are imposed. We impose a minimum value of 0 pS and a maximum value of \( w_{max} = 200 \) pS (Song and Abbott 2001).

In a number of studies it has been observed that the relative amount of synaptic potentiation (\( \Delta w_{pos} \)) is weaker for strong synapses, whereas the relative amount of depression shows no such dependence (Bi and Poo 1998; Debanne et al. 1999; Montgomery et al. 2001). This leads to wSTDP (van Rossum et al. 2000)
\[
\Delta w = \begin{cases} a_+ \exp(-s_{mn}/\tau_p) & \text{if } s_{mn} > 0 \\ -a_- w \exp(s_{mn}/\tau_d) & \text{if } s_{mn} < 0 \end{cases}
\]
Here the absolute amount of synaptic depression now depends on the current weight of the synapse, whereas the potentiation is independent of the weight as was the case for nSTDP. This rule gives rise to a unimodal and soft-bound weight distribution. We take the potentiation increment \( a_+ = 1 \) pS, the dimensionless depression constant \( a_- = 0.0114 \).

The values of \( a_-, a_+ \), and \( A_- \) were informed by values already used in the literature but adjusted such that the mean weight is identical for wSTDP and nSTDP (−100 pS); this fixed the \( a_+/A_- \) and \( A_/A_- \) ratio. This is essential to ensure that the nSTDP and wSTDP learning leads to the same mean input to the neuron and thus the same postsynaptic firing rate, \( n_{posts} \). Similarly, for a fair comparison it is also necessary to ensure that the weight fluctuations are comparable. By tuning \( a_+ \) and \( A_- \), we ensured that the modification rate, \( \langle \Delta w \rangle \), of the synapses is the same for nSTDP and wSTDP.

In some wSTDP models, the weight change was randomized to incorporate the high variability seen in experiments (van Rossum et al. 2000). This leads to a broader weight distribution, more closely resembling the experimentally observed one. We ran a separate set of simulations with multiplicative Gaussian noise in the preceding update rule (\( \sigma = 0.015 \)). The parameters were then set to \( a_+ = 0.2 \) pS and \( a_- = 0.002 \). Note, that the weight changes need to be smaller in this case to maintain the same modification rate. After this readjustment, the single neuron wSTDP retention time increased to 120\( \tau_d \) for a bin width of 20 ms with 90% in the noiseless case, see results) but was still much shorter than the nSTDP retention time. Thus the qualitative difference between wSTDP and nSTDP retention time is independent of the wSTDP implementation details.

Depression and potentiation in nSTDP, Eq. 1, depend only on the pre- and postsynaptic spike timing difference. STDP learning rules of this form are sometimes termed additive. Rules of the kind in Eq. 2, where depression depends on the current synaptic weight while potentiation does not depend on the synaptic weight, can be termed mixed (Kepecs et al. 2002). Mixed rules tend to give a single fixed point of the weight dynamics and hence a unimodal weight distribution. Alternatively, one can use \( \Delta w_+ = (1 - w) \exp(-s_{mn}/\tau_p) \) in Eq. 2, which leads to comparable dynamics (Kistler and van Hemmen 2000; Rubin et al. 2001). Other authors have argued that experimental data are compatible with a power law dependence of the magnitude of potentiation on the synaptic weight (Morrison et al. 2007). Such a power law scheme also allows for interpolation between nSTDP and wSTDP (Gutig et al. 2003). The terms additive, mixed, and multiplicative STDP are somewhat confusing as the resulting behavior depends on whether the sum of depression and potentiation is weight independent. If that is the case, as in nSTDP, the dynamics are mainly driven by the spike correlations and are strongly competitive. In all other cases, the weight dynamics will be mainly driven by the net effect of potentiation and depression (Kepecs et al. 2002). A Fokker Planck approach can be used to calculate the weight distribution (Morrison et al. 2008).

Memory trace retention

To quantify the lifetime of a memory trace, we use the autocorrelation of the synaptic weights. Imagine that at time \( t_0 \), the synaptic weights have equilibrated; at this instant, we take a snapshot of the weights, \( w(t_0) \). Next we let the weights evolve for some further time \( t \). After this time, the weights will still reside in the equilibrium distribution, but individual weights will have moved due to random spike pairing events. We take again a snapshot of the weights; we now have two lists of weights. The autocorrelation of the weights is defined as
\[
A(t) = \frac{1}{t^2} \langle [w(t_0) - \langle w \rangle] [w(t_0 + t) - \langle w \rangle] \rangle
\]
where the average, indicated by the angular brackets, is over all synapses, and \( \sigma^2 \) is the variance in the weights. Note, that the equilibrium condition implies that \( \langle w \rangle = \langle w(t_0) \rangle = \langle w(t_0 + t) \rangle \). The value of \( A(t) \) measures how much of the original trace remains after the elapsed time \( t \) and can be considered as the memory strength that the system has of its state at
time \( t_p \). The autocorrelation function is typically a sum of exponentials with one exponential decaying the slowest. The time scale of this slowest decay corresponds to our measure of the retention time. However, note that deviations from single-exponential decay are sometimes visible on short time scales, e.g., Fig. 2C.

To measure the recognition performance of the neuron when explicit patterns are stored (Fig. 2), we used an SNR analysis. We measure the activation of the neuron when only inputs with enhanced weights were activated and compared that to the response to an unlearned, arbitrary pattern for which random inputs were activated. The neuron’s task is to distinguish between the two. The signal-to-noise ratio was defined as SNR = \( \mu_L - \mu_U \)/\( \sqrt{\text{var}(U)} \), where \( \mu_L(\text{var}(U)) \) is the mean response to a learned (unlearned) pattern, and \( \text{var}(U) \) is the associated variance. As is common in this type of analysis, the performance obtained is the best possible. Effects such as synaptic noise, stochastic release, or postsynaptic saturation could only reduce the performance.

Network model

The network consists of one layer of 60 integrate-and-fire neurons with parameters as in the preceding text. The network has periodic boundary conditions to eliminate edge effects and to ensure that all neurons operate under comparable conditions. In other words, the neurons and inputs are placed on a ring so that the neurons at both ends receive similar input. The neurons receive feed-forward input from a layer of 600 Poisson inputs through STDP synapses and receive all-to-all inhibition through lateral connections. The neurons do not self-inhibit. Now each neuron receives two current contributions: \( \tau_{\text{d}}(t) = I_{\text{fd}}(t) + I_{\text{inh}}(t) \). Feed-forward excitatory synapses are identical to the single-neuron case. Inhibitory synapses (conductance-based) are exponential with a time constant of 5 ms and have a reversal potential of −74 mV. The inhibitory synapses are not plastic and are uniform across the inhibitory population. All STDP networks are initialized with random input weights uniformly distributed between 0 and 200 pS.

In all networks, the mean excitatory input current to each unit is around 0.5 pA (averaged during 20 s over both selective and nonselective stimuli). As the inhibitory conductance was increased from 0 to 10 nS (Fig. 5D), the mean inhibitory current rose from 0 to 0.25 pA, ~50% of the average excitatory input. Thus inhibition does not have to be overly strong to stabilize the receptive fields; instead the inhibition is of same order of magnitude as the excitation as has been argued to be the case in visual cortex.

Inputs to the network are again Poisson trains, but the firing rate is spatially modulated as follows: input a has a rate \( \nu_a = \nu_0 + \nu_1 \left[ e^{-\lambda a} - e^{-\lambda b} \right] \), where the stimulus is centered at input \( s \), the background rate is \( \nu_0 = 10 \text{ Hz} \) and peak rate \( \nu_1 = 80 \text{ Hz} \), \( \lambda \) is the width of the network, and \( \sigma \) is the width of the stimulus set to be 1/10 of the number of inputs. The second and third term in the equation come from the periodic boundary conditions. The location of the center of the stimulus was randomly chosen at time intervals drawn from an exponential distribution with a mean of 20 ms. Again, this input structure is chosen to be comparable to previous work on nSTDP (Song and Abbott 2001).

Receptive field stability

Receptive fields of the neurons in the network are found as follows: at given times the synaptic weights are frozen and the same input stimulus as described in the preceding text is swept across the inputs. The tuning curve of each neuron is measured at \( m = 24 \) stimulus locations (25 stimuli around each location; response measured for 20 ms). The tuning curve is plotted in a polar plot and the vector average is calculated. Thus the receptive field of each neuron is characterized by the two-dimensional vector defined as, \( \vec{p} = \left\{ \frac{1}{m} \sum_k \nu_k \sin \left( \frac{2\pi k}{m} \right), \frac{1}{m} \sum_k \nu_k \cos \left( \frac{2\pi k}{m} \right) \right\} \), where \( k \) indicates the stimulus location, and \( \nu_k \) is the firing rate at that location.

The memory trace retention introduced in the preceding text measures how long weight correlations last. We define a similar measure to quantify retention of the receptive fields using the receptive field vectors. For a network with \( N \) neurons, we will have a 2N component vector \( \{ p_{1r}, p_{1l}, \ldots, p_{Nr}, p_{Nl} \} \) where \( r \in \{1, \ldots, N\} \). We calculate the autocorrelation of this vector in exactly the same way as we did for the weight vector. If the autocorrelation is one, the receptive fields have not changed from their initial state and have remained in their initial input locations. If, in contrast, the autocorrelation is zero, their receptive-field locations have become independent of the initial positions.

In the preceding text we described how we quantify the stability of the receptive fields in the case that they are subjected to ongoing presentation of the input stimulus used to train them. In the simulations of Fig. 6, we test the stability of the receptive fields when a blank stimulus is presented. In this case the input stimulus consists of unstructured, uncorrelated Poisson spike trains, i.e., \( \nu_a = \nu_{0a} \).

To further characterize the receptive field, we calculate the selectivity \( S \) of a neuron as \( S = 1 - \frac{1}{\nu_{\text{max}}} \sum_k \nu_k \), where \( \nu_{\text{max}} \) is the maximum firing rate of the neuron (occurring at the optimal stimulus position) (Bienenstock et al. 1982). In the case that the tuning curve is flat, we find a selectivity of 0. If the tuning curve is more peaked, the selectivity increases. In the limit that the peak is a delta function, we find a selectivity of \( S = 1 \).

Results

Synaptic weight persistence for single neurons

We first study the retention of synaptic weights in a single neuron equipped with different STDP rules. Before we address memory retention, we examine how the weights fluctuate under equilibrium conditions as this turns out to relate to how quickly weight modifications are erased.

We simulated a single integrate-and-fire neuron receiving stationary Poisson inputs. After an initial period, the synaptic weights reach an equilibrium distribution, Fig. 1, B and C, in which individual weights are still changing, Fig. 1, D and E, while the distribution remains stationary. As observed previously, the two plasticity rules give rise to very different equilibrium weight distributions. The wSTDP rule gives rise to a unimodal, soft-bound weight distribution, while nSTDP gives rise to a bimodal weight distribution requiring hard bounds on the minimal and maximal synaptic weight (Kistler and van Hemmen 2000; Rubin et al. 2001; Song et al. 2000; van Rossum et al. 2000). Both STDP rules reward synapses that cause postsynaptic spikes. In wSTDP, however, each weight experiences a strong force pulling it back to the mean value, leading to a central distribution. In nSTDP there is no such force, and hard bounds need to be imposed. Furthermore, in nSTDP depression is usually made somewhat stronger than potentiation so that about half the weights, too weak to induce enough postsynaptic spikes, are depressed to zero, whereas stronger weights grow to the upper bound. This leads to a bimodal distribution.
To quantify the retention time of random synaptic modifications in nSTDP and wSTDP, we calculate the temporal autocorrelation of the weights while they fluctuate within the equilibrium distributions. In Fig. 1, F and G, the autocorrelation of the weights of a single neuron with 800 Poisson inputs is plotted for nSTDP and wSTDP. For nSTDP, the autocorrelation decays exponentially at large time scales with a time constant of 18 h. Under comparable conditions, the wSTDP autocorrelation falls rapidly with a time constant of 29 s. For comparison, the nSTDP autocorrelation has been replotted on this time scale in Fig. 1G, emphasizing the difference; the nSTDP autocorrelation decay is approximately 2,200 times slower than the wSTDP decay. Thus a seemingly minor modification in the learning rule not only affects the weight distribution but also dramatically alters the synaptic retention time as was suggested earlier (Rubin et al. 2001).

Clearly not just the learning rule determines the speed with which the weights change. If the neuron is firing quickly or if the plasticity parameters are set such that weight modifications are large, previous synaptic modifications will decay rapidly. Conversely, small weight modifications and low firing rates will lead to long-lasting synaptic changes. To control for this, the parameters were set such that the postsynaptic firing rate $v_{\text{post}}$ was similar for both learning rules (~15 Hz). In addition, we matched the modification rates of the rules, defined as $\langle v_{\text{post}}|\Delta w\rangle$ where $|\Delta w|$ is the size of synaptic modification steps. Thus any differences between the retention times of the two rules are not simply down to differences in modification rate.

For wSTDP, the retention time scale can be calculated for general parameters (APPENDIX). It decays exactly exponentially with a time constant

$$\tau = \frac{1}{\tau_a v_{\text{pre}} v_{\text{post}}}$$

where $v_{\text{pre}}$ ($v_{\text{post}}$) is the pre (post)-synaptic firing rate, and $\tau_a$ and $a_+$ characterize the depression plasticity window and rate constants (METHODS). With our parameters, Eq. 3 gives a value of 27 s, which is in good agreement with the preceding simulation results, dashed curve in Fig. 1G. The fact that $\tau_a$ and $a_+$ do not occur in Eq. 3, might seem surprising; however, the equation is exact within certain approximations, see APPENDIX. Furthermore, long-term potentiation and depression (LTP and LTD) are not fully analogous in wSTDP as only LTD depends on the synaptic weight. Finally, the output firing rate $v_{\text{post}}$ is dependent on the LTP parameters, thus these parameters affect the retention time indirectly.

In contrast, the autocorrelation function for nSTDP weights is not a single exponential and is far more difficult to calculate exactly. However, two approximation schemes allow a calculation of the longest nSTDP retention time scale (see APPENDIX). First, we can estimate the nSTDP retention time scale by recasting nSTDP as a diffusion process in a double-well potential. This gives a retention time of 20 h for the parameters used in the simulation, Fig. 1F (curve labeled double well). Alternatively, one can interpret the stochastic weight evolution as a discrete Markov process, which also matches the simulation well (curve labeled one step).

FIG. 1. Weight distributions and weight persistence in single-cell nonweight-dependent and weight-dependent spike-timing-dependent plasticity (nSTDP and wSTDP) models. A: diagram of the single-neuron simulation. An integrate-and-fire neuron receives 800 Poisson inputs. B: the equilibrium weight distribution reached when the synapses are subject to the nSTDP rule. C: the equilibrium distribution in the case of the wSTDP rule. D: the evolution with time of 30 nSTDP weights randomly sampled from the distribution in B after equilibrium has been established. E: as D but for the wSTDP. F: the temporal autocorrelation of the nSTDP weights vs. time (---). The line labeled double well is an approximate calculation of the autocorrelation as found by considering nSTDP as diffusion in a double-well potential. The curve labeled one step is an approximate calculation of the autocorrelation (see APPENDIX). G: the temporal autocorrelation of wSTDP weights. The simulated autocorrelation (—) and the theoretical autocorrelation function (- - -). The nSTDP simulation data are replotted on this time scale for comparison (curve labeled nSTDP); the wSTDP autocorrelation decays >2,200 times more rapidly.
The numerical data together with the analysis show that synaptic modification due to random pre-post pairing is retained far longer by nSTDP than by wSTDP. The underlying reason for the slow decay in nSTDP is that it is bistable, which can be seen from its bimodal weight distribution. The nSTDP retention time is dominated by the time it takes for weights to wander from one maximum of the weight distribution to the other maximum across a region of low probability. As this is rather unlikely, the synaptic weights in nSTDP are much more persistent than those in wSTDP which has no such bistability.

**Relationship between forgetting and the autocorrelation time scale**

The preceding results raise the question how the autocorrelation time scale relates to how quickly explicitly stored patterns are erased by ongoing activity. Previously this has been studied using specialized measures (Toyoizumi et al. 2007). Here we address this question as follows: we wait until equilibrium is reached and then instantaneously embed a pattern in the weights. First, we set 10 of the weights to 200 pS, about twice the mean weight. (Although this is not the focus of this study, according to the STDP rules such patterns can be created by repeatedly pairing these inputs with timed postsynaptic spikes). After the intervention we continue stimulation with random inputs and track the mean values of these potentiated spikes. As is common in this type of analysis, this is a highly optimized input designed to find the maximal SNR achievable. Importantly, the SNR incorporates not only decay of the mean weights but also possible changes in their variance. A priori it is therefore not guaranteed that the SNR will decay as quickly as the mean weight.

We find that when few weights are modified, the SNR persists with a time scale that is identical to the equilibrium autocorrelation time scale, Fig. 2. A and B. The mean weight of the potentiated group decays exponentially back to the baseline. For both nSTDP and wSTDP, the time scale of relaxation of the means back to the equilibrium value matches the autocorrelation time scale (- - -). In contrast, when half the weights are set to 200 pS, the mean weight is not preserved and the output firing frequency increases. The evolution of the mean weight of the elevated group now no longer matches the equilibrium autocorrelation time scale, and the initial decay is very fast, Fig. 2, C and D. Likewise, if the pattern reduces the mean weight (and the output firing frequency), the time scale of evolution of the mean weight is longer than the autocorrelation time scale for large numbers of depressed weights, but matches it if the number of depressed weights is only small (not shown).

The correspondence between the time scale of decay of the mean weight of the modified weights and the autocorrelation time scale demonstrates a link between the time scale over which fluctuations remove correlations between synapses and the time scale over which small deviations from equilibrium persist in ensembles of STDP synapses. This is an instance of the fluctuation dissipation theorem of the first kind, linking the equilibrium autocorrelation to the time scale of survival of a small perturbation from equilibrium (Kubo et al. 1998). A consequence is that patterns stored in the weights decay with precisely the autocorrelation time scale provided that the system remains close to equilibrium.

To examine how well the stored pattern can be retrieved, we use a signal-to-noise ratio (SNR) analysis (METHODS). The neuron receives input corresponding to the stored pattern in which all inputs with high weights are spiking, and all other inputs are silent. The total input current is compared with the case where the neuron receives a scrambled version of the pattern. As is common in this type of analysis, this is a highly optimized input designed to find the maximal SNR achievable. The correspondence between the time scale of decay of the mean weight and the autocorrelation time scale is seen in Fig. 2. A and B (bottom). Thus the theory not only predicts decay of the mean weights, but also of memory performance as measured through an SNR analysis. The underlying reason is that the variance in the weight is virtually unaffected by the pattern. If the pattern is a large

![FIG. 2.](image)

**FIG. 2.** The relationship between the retention time scale for a stored pattern and the autocorrelation time scale for a single unit. A: storing a pattern in a nSTDP neuron. Inset: graph of the pattern stored in the weights: 10 weights were instantaneously set to 200 pS. The decay of the mean weight of the potentiated group (top graph, —), matches the autocorrelation time scale (- - -). The decay of the SNR (bottom graph) matches the autocorrelation time scale as well. B: same as A but for wSTDP. C and D: storing a pattern that causes a large deviation in the mean weight. Half the weights are instantaneously set to w_r and the remainder of the weights are unmodified (inset). The evolution of the mean value of this group of weights (—) no longer matches the autocorrelation time scale and is markedly nonexponential because the unbalanced pattern has disrupted the linearity of the weight dynamics. (Note the very sharp initial drop in the nSTDP case). Similarly, the SNR decay no longer matches the autocorrelation either.
perturbation to the equilibrium distribution, Fig. 2, C and D, the match among autocorrelation time, mean weight decay time, and SNR decay time breaks down. The SNR decay is slower than the mean decay as the pattern remains recognizable against other patterns, but the decay is still faster than the autocorrelation.

In summary, these results demonstrate that a characterization of the equilibrium autocorrelation function is a sufficient statistic for assessing the survival time of a stored pattern, provided that the pattern is stored with only a minor alteration to the equilibrium weight distribution. An SNR of one indicates that the pattern can be recalled with a 30% error. As both for the small and large perturbations the SNR is significantly larger than that, it can be argued that the synapses store usable memories. We therefore find that the long autocorrelation time scale of nSTDP allows more persistent storage of memory than in the wSTDP case.

Retention time and the parameters of the plasticity windows

So far we have examined cases for which the depression and potentiation plasticity windows are equal. However, experiments suggest that the STDP depression time window is approximately twice as long as the potentiation time window, e.g., $\tau_+ = 34 \pm 13$ ms and $\tau_- = 17 \pm 9$ ms (Bi and Poo 1998), raising the question of how robust our predictions are with respect to changes in the plasticity parameters.

First we change $\tau_-$ while $\tau_+$ is kept fixed. For wSTDP learning, the dependence of the autocorrelation time on the plasticity window is given by Eq. 3. However, as $\tau_-$ is increased, the mean weight decreases, as $\langle w \rangle = \tau_+ a_+ / \tau_- a_-$ (Burkitt et al. 2004), which lowers the output firing frequency that enters Eq. 3. With this taken into account, the theory matches the simulations well (Fig. 3B). Alternatively, the postsynaptic frequency can be held constant by scaling $a_+$ by the same factor as $\tau_-$. This effectively changes the average amount of weight modification per pairing event. Again the simulation results match the theory well, Fig. 3D. If $\tau_-$ is changed, but $\tau_+ a_+$ is kept fixed, the retention time does not alter (not shown).

In the nSTDP case, if $\tau_-$ is reduced, potentiation dominates, and the weights cluster at the upper bound and the output firing rate saturates, Fig. 3A (insets). In this case, the bimodality of the nSTDP weight distribution is completely lost and the autocorrelation time scale becomes short, Fig. 3A. Note, that the decay becomes much faster than can be explained by the change in the postsynaptic firing rate alone. The weight distribution has become unimodal, resulting in fast de-correlation also seen in the wSTDP case.

Conversely, as $\tau_-$ is increased, depression dominates and the synaptic weights congregate near the zero bound. However, at a certain time postsynaptic firing ceases, thus freezing the weights. The autocorrelation time scale is longer than when $\tau_+ = \tau_-$, Fig. 3A, but at the expense of a strongly decreased output firing rate. The strong dependence of the autocorrelation time scale on the output firing frequency can be compensated for in the nSTDP case, by reducing (increasing) $A_-$ by the same factor that increases (reduces) $\tau_-$. By compensating changes in $\tau_-$ by adjusting $A_-$ so that $\tau_+ A_-$ is constant, the synaptic weight distribution remains bimodal. In this case, the mean weight and output rate are fixed, and as a result the retention time varies much less as $\tau_-$ is changed (although the dependence is still substantial), Fig. 3C. These results show that nSTDP by itself is not sufficient for long retention times; its parameters need to be tuned to create a bimodal distribution.

![FIG. 3.](http://jn.physiology.org/)

The dependence of the memory retention on the size of the depression window $\tau_-$ (potentiation window kept at $\tau_+ = 20$ ms). A: the nSTDP weight autocorrelation time scale increases steeply as $\tau_-$ is increased. The balanced scenario is highlighted by the circle. The triangle is the autocorrelation time scale of the equivalent wSTDP weights. Insets: the concurrent changes in the weights and postsynaptic rate. B: the autocorrelation time scale in the wSTDP case when $\tau_-$ is altered. Mean weight and output firing frequency are reduced as $\tau_-$ is increased, leading to an overall increase in the autocorrelation time scale, in agreement with theory, Eq. 3 (- - -). C: in the nSTDP case, the output firing frequency can be kept fixed by changing $A_-$ when $\tau_-$ is altered (see text). In this case, we find that the autocorrelation time scale is much less sensitive to changes in $\tau_-$. D: in the wSTDP case, the output firing frequency can be kept fixed by changing $a_+$ (see text). The change in autocorrelation time scale as determined by the simulation (points) is again well matched by the change in autocorrelation time scale as predicted by Eq. 3 (- - -).
The size of the weight modifications also greatly influences the nSTDP retention time scale. For the case that the parameters are set such that the synaptic weight distribution is bimodal and $A_-$ and $A_+$ are scaled simultaneously so that the bimodality is preserved, the retention time depends on $A_-$ as (APPENDIX)

$$\tau \propto \frac{1}{A_-} \exp(1/A_-).$$

In other words, the retention time grows roughly exponentially as the potentiation/depression event size decreases; such exponential dependence on the step size is typical for processes that involve jumping across a barrier.

**Receptive-field development in STDP networks**

In the preceding we have seen that under continuous stimulation isolated single neurons with wSTDP synapses forget their weights very rapidly. As this could be disastrous for network function, we asked if rapid forgetting also occurs in networks. The framework we use is a single layer network with all to all lateral inhibitory connections and plastic feed-forward excitatory connections that receive input stimuli and are subject to unsupervised learning. This model can be interpreted as a simple model for orientation selectivity (Ben-Yishai et al. 1995; Shapley et al. 2003; Song and Abbot 2001; Yao et al. 2004). However, before the question of weight retention and stability of input selectivity can be addressed, we first examine receptive-field formation with STDP.

We trained wSTDP and nSTDP networks from random initial conditions on the input stimulus shown in Fig. 4A (see METHODS). One group of networks has lateral inhibitory connections, whereas the other has no lateral inhibition. The formation of receptive fields in these types of networks with nSTDP has been explored previously. It was found earlier that neurons develop receptive fields even in the absence of recurrent connections (Delorme et al. 2001; Song and Abbot 2001).

---

**FIG. 4.** Receptive-field development in nSTDP and wSTDP networks (with 60 units and 600 inputs). A, left: schematic of the network. Large circles represent integrate-and-fire neurons, whereas small circles represent inputs (Poisson point processes). Right: raster plot of a short sample of input to the network. B: tuning curves for 2 neurons in the nSTDP network, before (dashed line) and after (solid line) training. C: same as B but for the wSTDP network with lateral inhibition. D and E: histogram of the selectivities of the neurons after training with no inhibition (left) and with 7-nS lateral inhibitory connections (right). For the wSTDP network, inhibition is essential to develop selectivity. F: the leftmost nSTDP tuning curve of B is plotted along with the feed-forward weights. Inset: weight distribution for this neuron. G: same as F but for wSTDP. wSTDP forms receptive fields that are similar to nSTDP receptive fields when sufficient lateral inhibition is present, although the underlying weights are more centrally distributed.
This is a consequence of the strong competition of the nSTDP rule in the single unit, which selects one group of inputs above another. The winner is determined by the initial conditions. Because the initial weights are random, the map of the receptive fields is also random. When local recurrent excitatory connections are added, all neurons in the network become selective for the same area of the input range (like a single column). When, in addition, all-to-all inhibition is included, maps form in which the receptive fields of the neurons tile the input in a locally continuous manner. In nSTDP networks with lateral inhibition only, disordered maps develop.

As in these previous studies, receptive fields form readily in the absence of lateral inhibition in the nSTDP network, Fig. 4D. The average selectivity increased from around 0–0.65 during training. In the wSTDP case without inhibition, there is no receptive-field development and no increase in the mean selectivity of the neurons in the network, Fig. 4E. However, with inhibition present, input selectivity does also develop in the wSTDP network, Fig. 4E. The receptive fields sharpen, and in some cases, receptive fields develop where there was little initial structure. The mean selectivity increases from 0.65 to 0.8 during training, a change that is as pronounced as it is for the nSTDP network with inhibition, where the mean selectivity also increases from around 0.65 to 0.85 during training. These results demonstrate that the development of selectivity that is intrinsic to the nSTDP learning rule, but that is absent from the wSTDP learning rule, can occur in wSTDP networks with lateral inhibition.

Note that in both networks with inhibition, some selectivity already exists before training due to the random initial conditions of the feed-forward weights (METHODS). However, these initial receptive fields have peak rates that are typically <1/3 of the final rates for wSTDP learning (the difference is even larger for nSTDP learning). In addition, the initial tuning curves often have multiple peaks and are irregular. In contrast, after training, the receptive fields are smooth with only one peak and the receptive fields tend to evenly distribute across input space, Fig. 4, B and C. In parallel, the interspike interval coefficient of variation (CV) decreases from 1.54 to 0.32 after training (in the absence of inhibition the CV remains virtually the same, 0.67 and 0.73, respectively).

The underlying structure in the feed-forward weights is shown in Fig. 4, F and G. Associated with the receptive fields in nSTDP networks is a region of weights at the maximum weight value, while all other weights are zero, Fig. 4F. The characteristic bimodal distribution of nSTDP is still present, but the weights are spatially inhomogeneous. This is a result of the strong competitive behavior of nSTDP mentioned previously, driving elevation of the correlated input group at the stimulus location. In wSTDP networks however, the underlying feed-forward weight structure corresponding to the receptive fields remains unimodal, Fig. 4G.

The development of receptive fields in wSTDP networks is thus dependent on lateral inhibition, similar to the situation for rate-based competitive Hebbian learning (Hertz et al. 1991). Multiple processes contribute to the emergence of selectivity: First, rough receptive fields already emerge without plasticity as neurons compete for input and after a “race to spike,” dominant neurons suppress less-selective neurons. Second, STDP refines the receptive fields because the activity in the dominant neuron is positively correlated with the input, whereas the activity in the suppressed neuron is negatively correlated with the input (data not shown). On repeated presentation of the stimulus this effect will grow stronger as the firing of the losing unit becomes further anti-correlated with the inputs driving the dominant unit, leading to the final weight profiles, Fig. 4, F and G.

Receptive-field stability in STDP networks

Having established the development of receptive fields with wSTDP, we now examine the receptive-field stability. The network is presented again with the stimulus of Fig. 4A. As in the single-neuron case with random ongoing weight modification, there is no separate learning and testing phase, instead we measure the persistence under the continued stimulation with the same stimulus ensemble. We track the receptive fields of the neurons, by plotting the tuning curves on a polar plot and taking the vector sum of the responses (METHODS). The direction of this receptive-field vector gives the preferred direction, while its length combines the selectivity and the firing rate. The receptive fields form an unordered map, i.e., neighboring cells don’t necessarily have neighboring receptive fields (Song and Abbott 2001), Fig. 5A. In wSTDP, an unordered map forms as well provided that there is sufficient lateral inhibition, Fig. 5B, but the selectivity of the neurons is more variable.

To measure the stability of the receptive fields, we calculate the autocorrelation of the of the receptive-field vector (METHODS). If the autocorrelation is one, the receptive fields have not moved. If the receptive field autocorrelation falls to zero, the receptive fields have no relation to their previous position anymore. The nSTDP network with no lateral inhibition gives rise to a receptive field autocorrelation that decays with a time scale of 11 h, Fig. 5C. With inhibition, this increases to 93 h (an accurate figure is difficult to obtain in this case because the very slow decay means that an enormous simulation time would be needed to see substantial decay of the memory).

The wSTDP receptive fields decorrelate quickly in comparison to the nSTDP network. Importantly, however, the decorrelation time scale depends on the strength of the lateral inhibitory connections: When lateral connection strength is zero, no stable receptive-field vectors exist (because, as we have seen, no receptive fields form) and the correlation time is simply that of filtered noise. However, as the strength of the inhibitory connections is increased, and the receptive fields sharpen, the correlation time scale of the receptive field vectors increases.

The retention time depends smoothly on the inhibition, Fig. 5D. Thus the stability of the receptive fields in wSTDP networks can be varied by altering the level of lateral inhibition. When the inhibition is sufficiently large, the receptive fields remain correlated with their initial positions for more than one hour. Although this is shorter than for the nSTDP network, the persistence in the network is much longer than the wSTDP single neuron persistence which is only 29 s. Although in the nSTDP network inhibition also stabilizes the receptive fields, the improvement is much less dramatic than in the wSTDP case.

A simple explanation for the increased stability in the wSTDP network with inhibition could be that the inhibition reduces the firing rates and hence slows down the weight evolution. If this was the case, then one would expect that the
decorrelation rate would slow in proportion to the firing frequency of the neuron. However, this is not the case. We observe that as the inhibition was raised from 1 to 10 nS, the mean peak firing rates of the neurons actually rise slightly (41–44 Hz), while the receptive field retention time (as measured by the decay to 80%) rises from 120 s at 1-nS inhibition to ~1 h (3,280 s) at 10 nS, more than an order of magnitude difference.

The mechanism responsible for the difference in retention time between the single neuron with wSTDP and the network is the same as the mechanism that leads to receptive-field formation. In the single neuron case, there is no process that prevents a weak weight becoming strong (or vice versa). But in a network with lateral inhibition there is competition. If a certain neuron is dominant for a given input location, then its weights will grow strong while weights to suppressed neurons do not. For the suppressed neuron to overcome this competition requires either a very large random weight jump or a smaller weight jump combined with a decreasing jump in the weight of the dominant neuron. As these processes are unlikely the retention time is long.

We also explored the consequences of adding local excitatory connections to the wSTDP networks and found that wSTDP networks that have local excitation can readily form ordered maps (not shown). The effect of local excitation is to partially cancel the inhibition, thus slightly increasing the fluctuations of the receptive fields and so reducing the auto-correlation time scale of the receptive fields. However, qualitatively, the results hold, namely wSTDP networks require inhibition to become input selective and that varying the level of inhibition leads to a variable degree of stability of the input selectivity.

In summary, lateral inhibition introduces competition in the wSTDP network. This inhibition can be varied, hence varying the competition and the readiness with which receptive fields form in the network. Conversely, for nSTDP strong competition is already present in the learning rule itself, and while inhibition sharpens the receptive fields, it is not necessary to form them.

**Forgetting of receptive fields**

So far, we have considered a situation in which there is no distinction between the learning and the test phase as an identical stimulus ensemble was presented throughout and learning was ongoing. We now examine how quickly the learned receptive fields are forgotten when subsequently different stimuli are presented. Of course, if stimulation is absent altogether, no pre- or postsynaptic spikes are generated, and the weights are maintained indefinitely. Therefore we measured forgetting using an unstructured Poisson stimulus with the same firing rates for all inputs (\(\nu_p = \nu_o\), METHODS). The forgetting is strongly dependent on the firing rate. When all inputs fire at the 10-Hz background rate, no significant postsynaptic firing results in either the nSTDP or wSTDP network (\(\nu_{post} < 0.5\) Hz), and the receptive fields are retained for very long.

For an input of 50 Hz, the postsynaptic firing rate in the nSTDP network with lateral inhibition is ~4 Hz, and the receptive fields do not decorrelate appreciably, Fig. 6A. The locations of the receptive fields remain fixed because the depressed weights are so weak that they cannot drive the target neuron and the competitive property of nSTDP ensures strong inputs remain strong. Eventually these weights can become strong by chance, but this takes place on a very long time scale.

In the case of the wSTDP network with lateral inhibition is ~50 s, the receptive fields are retained for very long time scale.

In the case of the wSTDP network with 7-nS lateral inhibitory connections, a 50-Hz stimulus leads to some forgetting as reflected in the quick initial decay of the correlation, on a time scale of ~50 s, Fig. 6A. Note however, that the correlation does not decay to zero. The reason is heterogeneity in the firing rates: some neurons fire at high rates (~10 Hz), and these neurons forget the receptive field quickly, e.g., Fig. 6C, while other neurons fall silent in response to the unstructured stimulus and hence retain their weights. The rapid and substantial
decorrelation of wSTDP receptive fields is due to the loss of selectivity in the fastest firing neurons. Note that this effect does not occur when the network is stimulated with structured input, Fig. 5. In that case, none of the neurons falls completely silent, Fig. 5C and the autocorrelation falls visibly to 0. In general, in cases with a high-input frequency the receptive fields remain stable when stimulated with unstructured stimuli provided that lateral inhibition is present.

Given the rapid forgetting of perturbed weights in single neurons using wSTDP, it might have been expected that removing the structured input that gave rise to learning of those receptive fields, would result in their rapid decay. Lateral inhibition prevents this. While some receptive fields are lost, many remain much longer than the autocorrelation time of a single unit would predict.

DISCUSSION

STDP has in recent years been observed in many systems, and there is evidence that it plays an important role in development and cortical reorganization (Dan and Poo 2006; Mu and Poo 2006; Yao and Dan 2001; Young et al. 2007). Previous investigations have examined the distribution of the weights and synaptic competition under STDP (Burkitt et al. 2004; Izhikevich and Desai 2003; Song et al. 2000; van Rossum et al. 2000) as well as its stability (Kempter et al. 1999, 2001). In this study, we examined memory stability in two STDP models. One STDP model updates the weights in a manner that is independent of the current synaptic weight (nSTDP) (Song et al. 2000), whereas the other model updates the weights in a way that does depend on the synaptic weight (wSTDP) (Rubin et al. 2001; van Rossum et al. 2000).

For single neurons the equilibrium autocorrelation, is several orders of magnitude shorter for wSTDP (29 s) than for nSTDP (18 h). The reason is that the bimodal synaptic weight distribution engendered by nSTDP can retain weights at its opposing stable points for long periods of time. These results generalize to the case where actual patterns are stored in the synaptic weights. So long as the patterns do not significantly distort the equilibrium weight distribution, the decay time scale of the SNR of their retrieval is similar to the time scale of decay of the autocorrelation. In comparing these rules, we made sure that the average weight changes were comparable. If the amount of weight change per pairing would be made different, the retention times would change correspondingly.

Next, we analyzed networks that develop receptive fields through STDP. For nSTDP networks, we find that receptive fields readily form and are very stable, taking hours to decorrelate with their initial positions. The wSTDP networks develop input selectivity with receptive fields similar to those in nSTDP networks provided that sufficient lateral inhibition is present. The resulting receptive fields with wSTDP are much more stable (about 1 h) than the single neurons (29 s). Likewise, nSTDP networks gain stability from lateral inhibition, but the increase is much less dramatic (from 11 to 93 h), as most stability in nSTDP is already intrinsic to the learning rule. Inhibition has been studied before in network studies (e.g., Tsodyks and Feigelman 1988), where it was necessary to maintain good storage capacity for sparse patterns. Here the inhibition plays a novel role: it stabilizes the plasticity. Importantly, using wSTDP the stability of the receptive fields is strongly modulated by the inhibition, suggesting how the nervous system could actively change learning rates as needed.

FIG. 6. Untraining receptive fields in nSTDP and wSTDP networks with lateral inhibition. After the completion of training and formation of receptive fields, the networks are presented with a blank stimulus where all inputs fire at a uniform rate of 50 Hz. A: the autocorrelation of the receptive-field vectors for both wSTDP and nSTDP, when the blank stimulus is presented. B and C: the weights and tuning curve for a neuron in the wSTDP and nSTDP network. Top: example receptive fields before and after untraining. B: the receptive field from a neuron in the nSTDP network before and after presentation of the blank stimulus. In the nSTDP case, the peak firing rate of the receptive fields is equally reduced for all neurons. C: a receptive field from the wSTDP network before and after presentation of the blank stimulus from a neuron that fires relatively rapidly at ~10 Hz during presentation of the blank stimulus. In the wSTDP case, receptive fields of rapidly firing neurons such as this one are wiped out, while the receptive fields of more slowly firing neurons take much longer to be forgotten (not shown). Bottom: the weights underlying the receptive fields plotted above before and after untraining.
The results presented here connect on a number of points with experimental observations. The relatively short retention times of networks with wSTDP might seem at odds with what is desirable for a receptive field. Yet interestingly, the effects of putative STDP in the visual cortex are short lasting (some 15 min) (Dan and Poo 2006; Yao and Dan 2001), and spontaneous activity can rapidly erase induced plasticity in sensory development (Zhou et al. 2003), consistent with the wSTDP results. The more general result that inhibition can modulate plasticity is consistent with a number of studies. First, it is known that the end of the ocular dominance critical period correlates strongly with increased inhibition (Fagiolini and Hensch 2000). Furthermore, a recent study observed reduced inhibition in auditory receptive field plasticity (Froemke et al. 2007). Our results suggest that transiently blocking inhibition combined with sensory stimulation can lead to rapid changes in receptive fields, whereas without such blocking, receptive fields should be much more stable.

Necessarily this study makes a number of assumptions that are well worth remembering. First, the STDP rules used an all-to-all spike implementation, i.e., all spikes are included in the synaptic modifications and the contributions from each spike pairing sum linearly. However, there is evidence that non-linear correlations exist (Froemke and Dan 2002; Sjöström et al. 2001; Wang et al. 2005). Although the spike triplet data has been modeled heuristically (Pfister and Gerstner 2006), a unified model of these effects is still lacking. Such effects are outside the scope of this study. Nevertheless our results can be generalized to wSTDP rules with different pairing interactions (Burkitt et al. 2004). The second approximation is that the temporal and correlation structure of actual input and output spike trains is likely much more complicated than assumed here. Third, in classical LTP different stimulus strengths and paradigms activate different biochemical plasticity pathways, resulting in varying LTP longevity, see e.g., (Abraham 2003; Barrett et al. 2009), and similar effects have been observed in sensory plasticity (Zhou et al. 2003). Similar modulation might be present in STDP as well. Finally, neuromodulation might be able to gate plasticity, allowing for modulation of memory retention. Nevertheless, the differences between the two learning rules are so dramatic that they likely generalize to more complex models, including rate-based plasticity rules.

The result that wSTDP weights de-correlate rapidly as compared with nSTDP is strongly related to the bimodal weight distribution of nSTDP (Toyoizumi et al. 2007). Dynamics that support bistability are more stable than dynamics with similar fluctuations but no such bistability. The dichotomy between a learning rule with quick forgetting and an unimodal weight distribution versus a rule that yields long memory retention and learning rule with slow forgetting and a bimodal weight distribution is therefore quite general, although STDP learning rules can be devised that display both (Gutig et al. 2003; Meffin et al. 2006; Toyoizumi et al. 2007); this can be used to prevent development of strong selectivity in response to even the weakest input correlation. The fact that the bistability of nSTDP yields very robust memory is reminiscent of the suggestion that a biophysical bistability, such as CaMKII autophosphorylation, can stabilize single synapse memory (Crick 1984; Lisman 1994). However, as this study shows, bistability does not need to occur at the biophysical level. It can be achieved on the level of single-neuron activity by rewarding pre- to post-spike correlations as happens in nSTDP. Or ultimately, the stability can also be achieved on the network level by lateral inhibition. Whether the faster forgetting of wSTDP is a bug or a feature is hard to determine, as this appears strongly task dependent: some learning or processing might require very stable, little changing synaptic weights. On the other hand there is no doubt that adaptability is important in certain behaviors in which case having highly persistent receptive fields might be a drawback.

**APPENDIX**

**Retention time for wSTDP**

The wSTDP retention time can be calculated exactly under Poisson stimulation. The changes in the weights of synapses subject to STDP can be regarded as a stochastic process (van Rossum et al. 2000). The evolution of a weight subject to wSTDP with Poisson inputs and an all-to-all spike implementation can be described in the Langevin formalism (van Kampen 1992). This is a first-order time evolution equation with a noise term

\[ w(t + dt) = w(t) + A(w)dt + N(0, c) \sqrt{dt} \]  (A1)

where \( A(w) \) is the drift term. Under the assumption of independent Poisson firing, \( A(w) = \nu_{\text{pre}}\nu_{\text{post}}(\tau_+ a_+ - \tau_- a_-) \) (Burkitt et al. 2004). This can be understood as follows: the term \( \nu_{\text{pre}}\nu_{\text{post}} \) gives the rate of pre-post and post-pre pairs, while \( \tau_+ a_+ \) and \( -\tau_- a_- \) give the average amount of LTP and LTD, respectively, incurred per pairing. We can rewrite the drift as

\[ A(w) = \alpha [w_0 - w(t)] \]

where the mean weight is given by \( w_0 = (\tau_+ a_+) / (\tau_+ a_+) = \tau_- a_- \nu_{\text{pre}}\nu_{\text{post}} \). This expression shows that the drift will always pull the weights toward the mean value: when the weight goes above (below) the mean, depression (potentiation) starts to dominate, moving it back to the mean value.

The second term in Eq. A1 is the noise term, where \( N(0, c) \) denotes a Gaussian distribution with zero mean and variance \( c \). Although in general the noise is weight dependent, for the choice of parameters in our simulations, it varies only negligibly with the weight as compared with the drift, so it is assumed constant.

To obtain the autocorrelation time scale we multiply Eq. A1 by the weight at time 0, \( w(0) \), and take the ensemble average

\[ \langle w(0)w(t + dt) - w(0)w(t) \rangle = \alpha [w_0 - w(0)] \delta t \]  (A2)

where we use that \( \langle w(0) \rangle = w_0 \) because the system is at equilibrium. Thus \( \langle d/dt w(0) \rangle = \alpha [w_0^2 - w(0)] \) with the solution \( \langle w(0) \rangle = e^{-\alpha t} w_0^2 \). Hence the autocorrelation is

\[ C(t) = \frac{1}{\sigma^2} \left[ \langle (w(0)w(t) - \langle w(0) \rangle^2) \right] \]  (A3,A4)

\[ = \exp(\tau_- a_- \nu_{\text{post}}) \]

The autocorrelation decays exponentially with a time constant \( \tau = 1/(\tau_- a_- \nu_{\text{post}}) \), which is the reciprocal of the gradient of the drift with respect to the weight. Note that the result is independent of the variance of the noise term \( c \).

Although this is not immediately obvious, the autocorrelation time also depends on \( \tau_+ \) and \( a_+ \). If \( \tau_- \) or \( a_- \) is modified, then the mean synaptic weight \( w_0 = (\tau_+ a_+)/(\tau_+ a_+) \), and consequently the output firing frequency \( \nu_{\text{post}} \) change, modifying the correlation time.

If the neuron’s firing rate is fully linear in its input (which is only approximately true in the simulations), the postsynaptic rate is \( \nu_{\text{post}} = kw_0 \nu_{\text{pre}} \), where \( k \) is some constant. This allows the inverse decrece-
The diffusion term is input that the neuron receives, with the steady-state solution matches the simulation well. Although the eigen system cannot easily be solved manually, it describes the correlation of the process. Because we are investigating centroid of the probability distribution, identical to the drift in the Turrigiano (2000). The quantity is analogous to a stochastic escape problem (van Kampen 1992).

The retention time of the nSTDP rule is far harder to analyze. In nSTDP, the weights congregate near 0 and \( w_{\text{max}} \), where the weight distribution will have maxima. The wandering of the synaptic weights is analogous to a stochastic escape problem (van Kampen 1992). While at short time scales small fluctuations around the maxima dominate, at long time scales, the autocorrelation will depend on how quickly the weights randomly move from one maximum to the other. Here we describe two methods that allow us to approximate the autocorrelation time scale of nSTDP.

The first method exploits the idea that weight evolution under nSTDP can be described by the Fokker-Planck formalism. The Fokker-Planck equation expresses the evolution of a probability distribution in terms of a drift process that determines the movement of the centroid of the probability distribution, identical to the drift in the Langevin equation [provided that the fluctuations are independent of the synaptic weight (Risken 1996)], and 2) fluctuations that give rise to a diffusive process

\[
\dot{\phi}(w, t) = -\dot{\phi}[A(w)P(w, t)] + \frac{1}{2} \dot{\phi}^{2}[B(w)P(w, t)]
\]  

(A5)

Assuming identical potentiation and depression time constants (\( \tau_+ = \tau_- \)), the nSTDP drift is \( A(w) = \nu_{\text{pre}}\nu_{\text{post}}[1 + \text{sign}(w + \delta w)] - A_0 \) and the diffusion term is \( B(w) = \nu_{\text{pre}}\nu_{\text{post}}A_2 \) (van Rossum, Bii, and Turrigiano 2000). The quantity \( W_{\text{tot}} = N\nu_{\text{pre}}\nu_{\text{post}}w \) describes the total input that the neuron receives, with \( N \) the number of inputs.

We discretize the continuous weight \( w \) into \( M \) bins (states) of width \( \delta w \) and approximate \( \dot{\phi}(w) = (1/\delta w)[f(w + \delta w) - f(w)] \) and similarly \( \dot{\phi}^{2}(w) = [1/(\delta w)^2][f(w - \delta w) + f(w + \delta w) - 2f(w)] \). Equation A5 becomes a matrix equation, describing a Markov process on the states \( w_i \), \( i = 1, 2, \ldots, M \), with a transition matrix \( M \). The evolution of the weights is given by \( \dot{\phi}(t, i, \tau_+ = \tau_-) \) depending on how the weight is given by \( \phi(t) = \sum_i \phi_i(t) \). The correlation function \( \langle w(0)w(t) \rangle \) can be calculated by tracking the evolution of all bins

\[
\langle w(0)w(t) \rangle = \sum_i w_i(0)w_i(t)P(i, t = 0)P(j, t = 0)
\]  

(A6)

We decompose the probability \( P(i, t = 0) \) into a linear combination of the eigenvectors of the transition matrix of \( M \). Hereto define the matrix \( C_{ik} \), such that \( \sum_j C_{ik} e^{s_j^{(k)}} = \delta_{ik} \) where \( s_j^{(k)} \) is the \( k \)th eigenvector of \( M \). Now each of the eigenvectors evolves independently according to \( \dot{\phi}_i(t, t, 0, \cdots, 0) = \sum_j e^{\lambda_j^{(k)}} C_{ik} \phi_j^{(k)} \).

\[
\langle w(0)w(t) \rangle = \sum_k e^{\lambda_k^{(k)}} \langle \sum_i P(i, 0)C_{ik}w_i \rangle \langle \sum_j s_j^{(k)}w_j \rangle
\]  

(A7)

describes the correlation of the process. Because we are investigating the equilibrium case, one can insert \( P(i, 0) = s_i^{(1)} \sum_j s_j^{(1)}, \) so that \( s_j^{(1)} \) is the eigenvalue with zero eigenvalue. The dominant time scale of Eq. A7 is the inverse of the smallest, nonzero eigenvalue of \( M \).

Although the eigen system cannot easily be solved manually, it is easily solved numerically, yielding the autocorrelation time scale. It matches the simulation well.

The second technique for calculating the nSTDP autocorrelation approximates the weight evolution as diffusion in a double-well potential. If a Fokker-Planck equation for a stochastic process has a steady-state solution \( P_{ss}(w) \) (as is the case here), the potential for that process is (Miguel and Toral 1997)

\[
V(w) = -\frac{\sigma^2}{Z} \ln P_{ss}(w)
\]  

(A8)

where \( \sigma \) is the amplitude of fluctuations and \( Z \) normalizes \( P_{ss}(w) \). Equation A8 expresses that if some particle diffuses in a potential, then the particle will most likely be found near the minima of the potential. For nSTDP, the equilibrium weight distribution is \( P_{ss}(w) = Z \exp \left(-K \cdot \sum w_i^2/2(\text{max} w_i) \right) \). \( \text{max} \) is the total input to the neuron (van Rossum et al. 2000).

\[
V(w) = \frac{\sigma^2}{2A_{\text{tot}}} \left(2w_{\text{tot}} - w^2 \right) \quad (0 < w < w_{\text{max}})
\]  

(A9)

where the magnitude of the fluctuations is given by \( \sigma^2 = \nu_{\text{pre}}\nu_{\text{post}} \cdot A_2 \). Outside the limits the potential is infinitely high due to the imposition of the hard bounds.

These hard bounds make this a difficult problem; however, we can approximate the potential with a quartic “double-well” potential \( V_{\text{approx}} \), the minima of which coincide with 0 and \( w_{\text{max}} \), and that provide a potential barrier between the two stable points with the same height as the original potential. Thus we fit a quartic requiring that \( V_{\text{approx}}(0) = V(0), V_{\text{approx}}(w_{\text{max}}) = V(w_{\text{max}}), V_{\text{approx}}(w_{\text{approx}}) = 0 \), and \( V_{\text{approx}}(0) = V(0) \), where \( w_{\text{approx}} \) is the point at which the drift vanishes. This potential is given by

\[
V_{\text{approx}}(w) = \frac{\sigma^2}{2w_{\text{tot}}^{2}} \left[ w^2(4w_{\text{tot}}^2 - 6w_{\text{tot}}w + 2w_{\text{tot}}^2) - 4w_{\text{tot}}^2w_{\text{approx}} \right]
\]  

(A10)

Next we assume that the central maximum of \( V_{\text{approx}} \) is sufficiently high so as to separate the time scale of diffusion within the well from the time scale of diffusion between wells. We can then approximate the mean first passage time of a weight to cross the center of the double-well potential as (van Kampen 1992)

\[
\tau_+ = \frac{2\pi}{\sqrt{V_{\text{approx}}''(w_{\text{approx}})(w_{\text{approx}})}} \exp \left[V_{\text{approx}}(w_{\text{approx}}) - V(0) \right] \frac{\sigma}{\sigma}
\]  

(A11)

where \( w_{\text{approx}} \) is the location of the central maximum of the potential. Equation A11 describes how long a weight near 0 on average takes to switch to the other well at \( w_{\text{approx}} \). The time to jump the other way, \( \tau_- \), is given by replacing 0 by \( w_{\text{approx}} \) in Eq. A11. The autocorrelation function for this two-state system is \( A(t) = \exp(-t/\tau_+ - t/\tau_-) \) with an autocorrelation time

\[
\tau_c = \frac{\tau_+ \tau_-}{\tau_+ + \tau_-}
\]  

(A12)

On long time scales, the dynamics is dominated by this switching of the weights between the stable fixed points. The autocorrelation will be exponential and dominated by this switching process as long as the distribution is strongly bimodal. However, at shorter time scales, the autocorrelation function will be dominated by phenomena occurring on shorter time scales - such as weights fluctuating around the maxima of the distribution.

\section*{Influence of the depression constant}

Here we examine how the autocorrelation for nSTDP depends on the size of the weight change \( A_+ \) and \( A_- \). Unlike changing the plasticity window, changing \( A_+ \) and \( A_- \) simultaneously does not change the balance of the nSTDP distribution. Thus its effects are somewhat simpler than the manipulations in the text that can only be analyzed through simulation. We assume the nSTDP potential is balanced so that \( \tau_+ = \tau_- = \tau \) and \( \tau = \pi/2 \) where \( \pi \) is the time associated with crossing from one well to the other. We express the double-well potential as \( V_{\text{approx}}(w) = (\sigma^2/2w_{\text{tot}}^2)A_{\text{tot}}(w) \).

Because we have asserted that all other variables are constant, the
function \( f(s) \) is also constant w.r.t. \( A \). Defining the constants \( \xi = \sqrt{\phi(w_{\text{in}})\phi(w_{\text{out}})} \) and \( \kappa = \phi(w_{\text{in}}) - \phi(0) = \phi(w_{\text{in}}) - \phi(w_{\text{out}}) \), the autocorrelation time scale is

\[
\tau_c = \frac{\pi \sigma_{\text{in}}^2 \sigma_{\text{out}}}{\nu_{\text{pre}} \nu_{\text{post}} A \cdot \kappa} \exp\left(\frac{\kappa}{2A \cdot w_{\text{in}}^2 \cdot w_{\text{out}}^2}\right)
\]  

(A13)

which is in good agreement with simulations, Fig. 3C.

ACKNOWLEDGMENTS

We thank R. Morris, S. Martin, J. Dwek, A. Lewis, S. Fusi, T. Sejnowski, and A. Barrett for discussion.

GRANTS

G. Billings was funded by the EPSRC through the Neuroinformatics Doctoral Training Centre. M.C.W. van Rossum was supported by Human Frontier Science Program (HFSP) and the Engineering and Physical Sciences Research Council (EPSRC).

REFERENCES

... (Continued)


Levy WB. A sequence predicting CA3 is a flexible associator that learns and uses context to solve hippocampal-like tasks. Hippocampus 6: 579–590, 1996.


J Neurophysiol • VOL 101 • JUNE 2009 • www.jn.org


