# Towards a learning-theoretic analysis of spike-timing dependent plasticity


**David Balduzzi**
MPI for Intelligent Systems, Tübingen, Germany
ETH Zurich, Switzerland
david.balduzzi@inf.ethz.ch

**Michel Besserve**
MPI for Intelligent Systems and MPI for Biological Cybernetics
Tübingen, Germany
michel.besserve@tuebingen.mpg.de


## Abstract


This paper suggests a learning-theoretic perspective on how synaptic plasticity benefits global brain functioning. We introduce a model, the selectron, that (i) arises as the fast time constant limit of leaky integrate-and-fire neurons equipped with spiking timing dependent plasticity (STDP) and (ii) is amenable to theoretical analysis. We show that the selectron encodes reward estimates into spikes and that an error bound on spikes is controlled by a spiking margin and the sum of synaptic weights. Moreover, the efficacy of spikes (their usefulness to other reward maximizing selectrons) also depends on total synaptic strength. Finally, based on our analysis, we propose a regularized version of STDP, and show the regularization improves the robustness of neuronal learning when faced with multiple stimuli.


## 1  Introduction

Finding principles underlying learning in neural networks is an important problem for both artificial and biological networks. An elegant suggestion is that global objective functions may be optimized during learning [1]. For biological networks however, the currently known neural plasticity mechanisms use a very restricted set of data – largely consisting of spikes and diffuse neuromodulatory signals. How a *global* optimization procedure could be implemented at the *neuronal* (cellular) level is thus a difficult problem.

A successful approach to this question has been Rosenblatt's perceptron [2] and its extension to multilayer perceptrons via backpropagation [3]. Similarly, (restricted) Boltzmann machines, constructed from simple stochastic units, have provided a remarkably powerful approach to organizing distributed optimization across many layers [4]. By contrast, although there has been significant progress in developing and understanding more biologically realistic models of neuronal learning [5–10], these do not match the performance of simpler, more analytically and computationally tractable models in learning tasks.

**Overview.**  This paper constructs a bridge from biologically realistic to analytically tractable models. The *selectron* is a model derived from leaky integrate and fire neurons equipped with spike-timing dependent plasticity that is amenable to learning-theoretic analysis. Our aim is to extract some of the principles implicit in STDP by thoroughly investigating a limit case.

Section §2 introduces the selectron. We state a constrained reward maximization problem which implies that selectrons encode empirical reward estimates into spikes. Our first result, section §3,


1


is that the selectron arises as the fast time constant limit of well-established models of neuronal spiking and plasticity, suggesting that cortical neurons may also be encoding reward estimates into their spiketrains.

Two important questions immediately arise. First, what guarantees can be provided on spikes being reliable predictors of global (neuromodulatory) outcomes? Second, what guarantees can be provided on the usefulness of spikes to other neurons? Sections §4 and §5 answer these questions by providing an upper bound on a suitably defined $0/1$ loss and a lower bound on the *efficacy* of a selectron's spikes, measured in terms of its contribution to the expected reward of a downstream selectron. Both bounds are controlled by the sum of synaptic weights $\|\mathbf{w}\|_1$, thereby justifying the constraint introduced in §2. Finally, motivated by our analysis, §6 introduces a *regularized STDP* rule and shows that it learns more robustly than classical STDP. §7 concludes the paper. Proofs of theorems are provided in the supplementary material.

**Related work.** Spike-timing dependent plasticity and its implications for the neural code have been intensively studied in recent years. The work closest in spirit to our own is Seung's "hedonistic" synapses, which seek to increase average reward [6]. Our work provides guarantees on the finite sample behavior of a discrete-time analog of hedonistic neurons. Another related line of research derives from the information bottleneck method [9, 11] which provides an alternate constraint to the one considered here. An information-theoretic perspective on synaptic homeostasis and metabolic cost, complementing the results in this paper, can be found in [12, 13]. Simulations combining synaptic renormalization with burst-STDP can be found in [14].

Important aspects of plasticity that we have not considered here are properties specific to continuous-time models, such as STDP's behavior as a temporal filter [15], and also issues related to convergence [8, 10].

The learning-theoretic properties of neural networks have been intensively studied, mostly focusing on perceptrons, see for example [16]. A non-biologically motivated "large-margin" analog of the perceptron was proposed in [17].

## 2 The selectron

We introduce the selectron, which can be considered a biologically motivated adaptation of the perceptron, see §3. The mechanism governing whether or not the selectron spikes is a Heaviside function acting on a weighted sum of synaptic inputs; our contribution is to propose a new reward function and corresponding learning rule.

Let us establish some notation. Let $\mathbf{X}$ denote the set of $N$-dimensional $\{0, 1\}$-valued vectors forming synaptic inputs to a selectron, and $Y = \{0, 1\}$ the set of outputs. A selectron spikes according to

$$y = f_{\mathbf{w}}(\mathbf{x}) := H\left(\mathbf{w}^{\mathsf{T}}\mathbf{x} - \vartheta\right), \text{ where } H(z) := \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{else} \end{cases} \tag{1}$$

is the Heaviside function and $\mathbf{w}$ is a $[0, 1] \subset \mathbb{R}$ valued $N$-vector specifying the selectron's synaptic weights. Let $P(\mathbf{x})$ denote the probability of input $\mathbf{x}$ arising.

To model the neuromodulatory system we introduce random variable $\nu : \mathbf{X} \to \{-1, 0, +1\}$, where positive values correspond to desirable outcomes, negative to undesirable and zero to neutral. Let $P(\nu|\mathbf{x})$ denote the probability of the release of neuromodulatory signal subsequent to input $\mathbf{x}$.

**Definition 1.** *Define **reward function***

$$R(\mathbf{x}, f_{\mathbf{w}}, \nu) = \underbrace{\nu(\mathbf{x})}_{neuromodulators} \cdot \underbrace{(\mathbf{w}^{\mathsf{T}}\mathbf{x} - \vartheta)}_{margin} \cdot \underbrace{f_{\mathbf{w}}(\mathbf{x})}_{selectivity} = \begin{cases} \nu(\mathbf{x}) \cdot (\mathbf{w}^{\mathsf{T}}\mathbf{x} - \vartheta) & if\ y = 1 \\ 0 & else. \end{cases} \tag{2}$$

The reward consists in three components. The first term is the neuromodulatory signal, which acts as a supervisor. The second term is the total current $\mathbf{w}^{\mathsf{T}}\mathbf{x}$ minus the threshold $\vartheta$. It is analogous to the margin in support vector machines or boosting algorithms, see section §4 for a precise formulation.

2

The third term gates rewards according to whether or not the selectron spikes. The reward is thus *selected*[1]: neuromodulatory signals are ignored by the selectron's reward function when it does not spike, enabling specialization.

**Constrained reward maximization.**    The selectron solves the following optimization problem:

$$\underset{\mathbf{w}}{\text{maximize:}} \quad \widehat{R}_n := \sum_{i=1}^{n} \nu(\mathbf{x}^{(i)}) \cdot (\mathbf{w}^{\mathsf{T}}\mathbf{x}^{(i)} - \vartheta) \cdot f_{\mathbf{w}}(\mathbf{x}^{(i)}) \tag{3}$$

$$\text{subject to:} \quad \|\mathbf{w}\|_1 \leq \omega \ \text{ for some } \omega > 0.$$

**Remark 1** (spikes encode rewards)**.**
*Optimization problem* (3) *ensures that selectrons spike for inputs that, on the basis of their empirical sample, reliably lead to neuromodulatory rewards. Thus, spikes encode expectations about rewards.*

The constraint is motivated by the discussion after Theorem 1 and the analysis in §4 and §5. We postpone discussion of how to impose the constraint to §6, and focus on reward maximization here.

The reward maximization problem cannot be solved analytically in general. However, it is possible to use an iterative approach. Although $f_{\mathbf{w}}(\mathbf{x})$ is not continuous, the reward function is a continuous function of $\mathbf{w}$ and is differentiable everywhere except for the "corner" where $\mathbf{w}^{\mathsf{T}}\mathbf{x} - \vartheta = 0$. We therefore apply gradient ascent by computing the derivative of (3) with respect to synaptic weights to obtain online learning rule

$$\Delta\mathbf{w}_j = \alpha \cdot \nu(\mathbf{x}) \cdot \mathbf{x}_j \cdot f_{\mathbf{w}}(\mathbf{x}) = \begin{cases} \alpha \cdot \nu(\mathbf{x}) & \text{if } \mathbf{x}_j = 1 \text{ and } y = 1 \\ 0 & \text{else} \end{cases} \tag{4}$$

where update factor $\alpha$ controls the learning rate.

The learning rule is *selective*: regardless of the neuromodulatory signal, synapse $\mathbf{w}_{jk}$ is updated only if there is both an input $\mathbf{x}_j = 1$ and output spike $y = f_{\mathbf{w}}(\mathbf{x}) = 1$.

The selectron is not guaranteed to find a global optimum. It is prone to initial condition dependent local optima because rewards depend on output spikes in learning rule (4). Although this is an undesirable property for an isolated learner, it is less important, and perhaps even advantageous, in large populations where it encourages specialization.

**Remark 2** (unsupervised setting)**.**
*Define the* unsupervised setting *by $\nu(\mathbf{x}) = 1$ for all $\mathbf{x}$. The reward function reduces to $R(\mathbf{x}, f_{\mathbf{w}}) = (\mathbf{w}^{\mathsf{T}}\mathbf{x} - \vartheta) \cdot f_{\mathbf{w}}(\mathbf{x})$. Without the constraint synapses will saturate. Imposing the constraint yields a more interesting solution where the selectron finds a weight vector summing to $\omega$ which balances (i) frequent spikes and (ii) high margins.*

**Theorem 1** (Controlling the frequency of spikes)**.**
*Assuming synaptic inputs are i.i.d. Bernoulli variables with $P(\text{spike}) = p$, then*

$$P\Big(f_{\mathbf{w}}(\mathbf{x}) = 1\Big) \leq p \cdot \left( \frac{\|\mathbf{w}\|_1}{\vartheta} \right)^2 \leq p \cdot \left( \frac{\omega}{\vartheta} \right)^2.$$

The Bernoulli regime is the discrete-time analog of the homogeneous Poisson setting used to prove convergence of reward-modulated STDP in [8]. Interestingly, in this setting the constraint provides a lever for controlling (lower bounding) rewards per spike

$$\Big\{\text{reward per spike}\Big\} = \frac{\widehat{R}}{P(f_{\mathbf{w}}(\mathbf{x}) = 1)} \geq c_1 \cdot \frac{\widehat{R}}{\omega^2}.$$

If inputs are not Bernoulli i.i.d., then $P(y = 1)$ and $\omega$ still covary, although the precise relationship is more difficult to quantify. Although i.i.d. inputs are unrealistic, note that recent neurophysiological evidence suggests neuronal firing – even of nearby neurons – is uncorrelated [18].

---

[1]The name "selectron" was chosen to emphasize this selective aspect.

3

# 3 Relation to leaky integrate-and-fire neurons equipped with STDP

The literature contains an enormous variety of neuronal models, which vary dramatically in sophistication and the extent to which they incorporate the the details of the underlying biochemical processes. Similarly, there is a large menagerie of models of synaptic plasticity [19]. We consider two well-established models: Gerstner's Spike Response Model (SRM) which generalizes leaky integrate-and-fire neurons [20] and the original spike-timing dependent plasticity learning rule proposed by Song *et al* [5], and show that the selectron arises in the fast time constant limit of the two models.

First let us recall the SRM. Suppose neuron $n^k$ last outputted a spike at time $t_k$ and receives input spikes at times $t_j$ from neuron $n^j$. Neuron $n^k$ spikes or according to the Heaviside function applied to the membrane potential $M_{\mathbf{w}}$:

$$f_{\mathbf{w}}(t) = H\left(M_{\mathbf{w}}(t) - \vartheta\right) \text{ where } M_{\mathbf{w}}(t) = \eta(t - t_k) + \sum_{t_j \leq t} \mathbf{w}_{jk} \cdot \epsilon(t - t_j) \text{ at time } t \geq t_k.$$

Input and output spikes add

$$\epsilon(t - t_j) = K\left[e^{\left(\frac{t_j - t}{\tau_m}\right)} - e^{\left(\frac{t_j - t}{\tau_s}\right)}\right] \text{ and } \eta(t - t_k) = \vartheta\left[K_1 e^{\left(\frac{t_k - t}{\tau_m}\right)} - K_2\left(e^{\left(\frac{t_k - t}{\tau_m}\right)} - e^{\left(\frac{t_k - t}{\tau_s}\right)}\right)\right]$$

to the membrane potential for $t_j \leq t$ and $t_k \leq t$ respectively. Here $\tau_m$ and $\tau_s$ are the membrane and synapse time constants.

The original STDP update rule [5] is

$$\Delta\mathbf{w}_{jk} = \begin{cases} \alpha_+ \cdot e^{\left(\frac{t_j - t_k}{\tau_+}\right)} & \text{if } t_j \leq t_k \\ -\alpha_- \cdot e^{\left(\frac{t_k - t_j}{\tau_-}\right)} & \text{else} \end{cases} \tag{5}$$

where $\tau_+$ and $\tau_-$ are time constants. STDP potentiates input synapses that spike prior to output spikes and depotentiates input synapses that spike subsequent to output spikes.

**Theorem 2** (the selectron is the fast time constant limit of SRM + STDP)**.**
*In the fast time constant limit, $\lim_{\tau_\bullet} \to 0$, the SRM transforms into a selectron with*

$$f_{\mathbf{w}}(t) = H\left(M_{\mathbf{w}}(t) - \vartheta\right) \quad \text{where} \quad M_{\mathbf{w}} = \sum_{\{j \mid t_j \geq t_k\}} \mathbf{w}_{jk} \cdot \delta_{t_k}(t).$$

*Moreover, STDP transforms into learning rule* (4) *in the unsupervised setting with $\nu(\mathbf{x}) = 1$ for all* **x**. *Finally, STDP arises as gradient ascent on a reward function whose limit is the unsupervised setting of reward function* (2).

Theorem 2 shows that STDP implicitly maximizes a time-discounted analog of the reward function in (3). We expect many models of reward-modulated synaptic plasticity to be analytically tractable in the fast time constant limit. An important property shared by STDP and the selectron is that synaptic (de)potentiation is gated by output spikes, see §A.1 for a comparison with the perceptron which *does not* gate synaptic learning

# 4 An error bound

Maximizing reward function (3) implies that selectrons encode reward estimates into their spikes. Indeed, it recursively justifies incorporating spikes into the reward function via the margin ($\mathbf{w}^\mathsf{T}\mathbf{x} - \vartheta$), which only makes sense if upstream spikes predict reward. However, in a large system where estimates pile on top of each other there is a tendency to *overfit*, leading to poor generalizations [21]. It is therefore crucial to provide *guarantees* on the quality of spikes as estimators.

Boosting algorithms, where the outputs of many weak learners are aggregated into a classifier [22], are remarkably resistant to overfitting as the number of learners increases [23]. Cortical learning may be analogous to boosting: individual neurons have access to a tiny fraction of the total brain state, and so are weak learners; and in the fast time constant limit, neurons are essentially aggregators.

4

We sharpen the analogy using the selectron. As a first step towards understanding how the cortex combats overfitting, we adapt a theorem developed to explain the effectiveness of boosting [24]. The goal is to show how the margin and constraint on synaptic weights improve error bounds.

**Definition 2.** *A selectron incurs a $0/1$ **loss** if a spike is followed by negative neuromodulatory feedback*

$$l(\mathbf{x}, f_{\mathbf{w}}, \nu) = \mathbb{1}_{-f_{\mathbf{w}}(\mathbf{x}) \cdot \nu(\mathbf{x})} = \begin{cases} 1 & \text{if } y = 1 \text{ and } \nu(\mathbf{x}) = -1 \\ 0 & \text{else.} \end{cases} \tag{6}$$

*The $0/1$ loss fails to take the estimates (spikes) of other selectrons into account and is difficult to optimize, so we also introduce the **hinge loss**:*

$$h^{\kappa}(\mathbf{x}, f_{\mathbf{w}}, \nu) := \left( \kappa - (\mathbf{w}^{\mathsf{T}}\mathbf{x} - \vartheta) \cdot \nu(\mathbf{x}) \right)_{+} \cdot f_{\mathbf{w}}(\mathbf{x}), \text{ where } (x)_{+} := \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{else.} \end{cases} \tag{7}$$

*Note that $l \leq h^{\kappa}$ for all $\kappa \geq 1$. Parameter $\kappa$ controls the saturation point, beyond which the size of the margin makes no difference to $h^{\kappa}$.*

An alternate $0/1$ loss[2] penalizes a selectron if it (i) fires when it shouldn't, i.e. when $\nu(\mathbf{x}) = -1$ or (ii) does not fire when it should, i.e. when $\nu(\mathbf{x}) = 1$. However, since the cortex contains many neurons and spiking is metabolically expensive [25], we propose a conservative loss that only penalizes errors of commission ("first, do no harm") and does not penalize specialization.

**Theorem 3** (spike error bound)**.**
*Suppose each selectron has $\leq N$ synapses. For any selectron $n^k$, let $S^k = \{n^k\} \cup \{n^j : n^j \to n^k\}$ denote a 2-layer feedforward subnetwork. For all $\kappa \geq 1$, with probability at least $1 - \delta$,*

$$\mathbb{E}\underbrace{\left[ l(\mathbf{x}, f_{\mathbf{w}}, \nu) \right]}_{0/1 \text{ loss}} \leq \frac{1}{n} \sum_{i} \underbrace{h^{\kappa}\left( \mathbf{x}^{(i)}, f_{\mathbf{w}}, \nu(\mathbf{x}^{(i)}) \right)}_{\text{hinge loss}} + \omega \cdot 2B \cdot \underbrace{\frac{\sqrt{8(N+1)\log(n+1)} + 1}{\sqrt{n}}}_{\text{capacity term}}$$

$$+ 2B \cdot \underbrace{\sqrt{\frac{2 \log \frac{2}{\delta}}{n}}}_{\text{confidence term}} \qquad \text{where } B = \kappa + \omega - \vartheta.$$

**Remark 3** (theoretical justification for maximizing margin and constraining $\|\mathbf{w}\|_1$)**.**
*The theorem shows how subsets of distributed systems can avoid overfitting. First, it demonstrates the importance of maximizing the margin (i.e. the empirical reward). Second, it shows the capacity term depends on the number of synapses $N$ and the constraint $\omega$ on synaptic weights, rather than the capacity of $S^k$ – which can be very large.*

The hinge loss is difficult to optimize directly since gating with output spikes $f_{\mathbf{w}}(\mathbf{x})$ renders it discontinuous. However, in the Bernoulli regime, Theorem 1 implies the bound in Theorem 3 can be rewritten as

$$\mathbb{E}\left[ l(\mathbf{x}, f_{\mathbf{w}}, \nu) \right] \leq p\kappa \frac{\omega^2}{\vartheta^2} - \widehat{R}_n\left( \mathbf{x}^{(i)}, f_{\mathbf{w}}, \nu(\mathbf{x}^{(i)}) \right) + \omega \cdot \left\{ \text{capacity term} \right\} + \left\{ \text{confidence term} \right\} \tag{8}$$

and so $\omega$ again provides the lever required to control the $0/1$ loss. The constraint $\|\mathbf{w}\|_1 \leq \omega$ is best imposed offline, see §6.

## 5 A bound on the efficacy of inter-neuronal communication

Even if a neuron's spikes perfectly predict positive neuromodulatory signals, the spikes only matter to the extent they affect other neurons in cortex. Spikes are produced for neurons by neurons. It is therefore crucial to provide guarantees on the usefulness of spikes.

In this section we quantify the effect of one selectron's spikes on another selectron's expected reward. We demonstrate a lower bound on efficacy and discuss its consequences.

---

[2]See §A.5 for an error bound.

5

**Definition 3.** *The **efficacy** of spikes from selectron $n^j$ on selectron $n^k$ is*

$$\frac{\delta R^k}{\delta \mathbf{x}_j} := \frac{\mathbb{E}[R^k|\mathbf{x}_j = 1] - \mathbb{E}[R^k|\mathbf{x}_j = 0]}{1 - 0},$$

*i.e. the expected contribution of spikes from selectron $n^j$ to selectron $n^k$'s expected reward, relative to not spiking. The notation is intended to suggest an analogy with differentiation – the infinitesimal difference made by spikes on a single synapse.*

Efficacy is zero if $\mathbb{E}[R^k|\mathbf{x}_j = 1] = \mathbb{E}[R^k|\mathbf{x}_j = 0]$. In other words, if spikes from $n^j$ make no difference to the expected reward of $n^k$.

The following theorem relies on the assumption that the average contribution of neuromodulators is higher after $n^j$ spikes than after it does not spike (i.e. upstream spikes predict reward), see §A.6 for precise statement. When the assumption is false the synapse $\mathbf{w}_{jk}$ should be pruned.

**Theorem 4** (spike efficacy bound).
*Let $p_j := \mathbb{E}[Y^j]$ denote the frequency of spikes from neuron $n^j$. The efficacy of $n^j$'s spikes on $n^k$ is lower bounded by*

$$c_2 \cdot \underbrace{\frac{\delta R^k}{\delta \mathbf{x}_j}}_{\text{efficacy}} \geq \underbrace{\frac{\mathbf{w}_j \cdot \mathbb{E}[Y^j Y^k]}{p_j}}_{\mathbf{w}_j\text{-weighted co-spike frequency}} + \underbrace{\frac{2\mathbb{E}\left[Y^j Y^k \cdot \left((\mathbf{w}^{\backslash j})^{\mathsf{T}}\mathbf{x} - \vartheta\right)\right]}{p_j(1 - p_j)}}_{\text{co-spike frequency}} - \underbrace{\frac{\mathbb{E}\left[Y^k \cdot \left((\mathbf{w}^{\backslash j})^{\mathsf{T}}\mathbf{x} - \vartheta\right)\right]}{1 - p_j}}_{n^k \text{ spike frequency}} \quad (9)$$

*where $c_2$ is described in §A.6 and $\mathbf{w}_i^{\backslash j} := \mathbf{w}_i$ if $i \neq j$ and $0$ if $i = j$.*

The efficacy guarantee is interpreted as follows. First, the guarantee improves as co-spiking by $n^j$ and $n^k$ increases. However, the denominators imply that increasing the frequency of $n^j$'s spikes *worsens* the guarantee, insofar as $n^j$ is not correlated with $n^k$. Similarly, from the third term, increasing $n^k$'s spikes *worsens* the guarantee if they do not correlate with $n^j$.

An immediate corollary of Theorem 4 is that Hebbian learning rules, such as STDP and the selectron learning rule (4), improve the efficacy of spikes. However, it also shows that naively increasing the frequency of spikes carries a cost. Neurons therefore face a tradeoff. In fact, in the Bernoulli regime, Theorem 1 implies (9) can be rewritten as

$$c_2 \cdot \frac{\delta R^k}{\delta \mathbf{x}_j} \geq \frac{\mathbf{w}_j}{p} \cdot \mathbb{E}[Y^j Y^k] + \frac{2}{p(1-p)}\mathbb{E}\left[Y^j Y^k \cdot \left((\mathbf{w}^{\backslash j})^{\mathsf{T}}\mathbf{x} - \vartheta\right)\right] - \frac{p \cdot \omega^2 \cdot (\omega - \vartheta)}{(1-p)\vartheta^2}, \quad (10)$$

so the constraint $\omega$ on synaptic strength can be used as a lever to improve guarantees on efficacy.

**Remark 4** (efficacy improved by pruning weak synapses).
*The $1^{st}$ term in (9) suggests that pruning weak synapses increases the efficacy of spikes, and so may aid learning in populations of selectrons or neurons.*

## 6  Experiments

Cortical neurons are constantly exposed to different input patterns as organisms engage in different activities. It is therefore important that what neurons learn is robust to changing inputs [26, 27]. In this section, as proof of principle, we investigate a simple tweak of classical STDP involving offline regularization. We show that it improves robustness when neurons are exposed to more than one pattern.

Observe that *regularizing* optimization problem (3) yields

$$\underset{\mathbf{w}}{\text{maximize:}} \quad \sum_{i=1}^{n} R\big(\mathbf{x}^{(i)}, f_{\mathbf{w}}, \nu(\mathbf{x}^{(i)})\big) - \frac{\gamma}{2}(\|\mathbf{w}\|_1 - \omega)^2 \quad (11)$$

$$\text{learning rule:} \quad \Delta \mathbf{w}_j = \alpha \cdot \nu(\mathbf{x}) \cdot \mathbf{x}_j \cdot f_{\mathbf{w}}(\mathbf{x}) - \gamma \cdot \big(\|\mathbf{w}\|_1 - \omega\big) \cdot \mathbf{w}_j \quad (12)$$

incorporates synaptic renormalization directly into the update. However, (12) requires continuously re-evaluating the sum of synaptic weights. We therefore *decouple* learning into an online reward maximization phase and an offline regularization phase which resets the synaptic weights.

6

A similar decoupling may occur in cortex. It has recently been proposed that a function of NREM sleep may be to regulate synaptic weights [28]. Indeed, neurophysiological evidence suggests that average cortical firing rates increase during wakefulness and decrease during sleep, possibly reflecting synaptic strengths [29, 30]. Experimental evidence also points to a net increase in dendritic spines (synapses) during waking and a net decrease during sleep [31].

**Setup.** We trained a neuron on a random input pattern for 10s to 87% accuracy with regularized STDP. See §A.7 for details on the structure of inputs. We then performed 700 trials (350 classical and 350 regularized) exposing the neuron to a new pattern for 20 seconds and observed performance under classical and regularized STDP.

**SRM neurons with classical STDP.** We used Gerstner's SRM model, recall §3, with parameters chosen to exactly coincide with [32]: $\tau_m = 10$, $\tau_s = 2.5$, $K = 2.2$, $K_1 = 2$, $K_2 = 4$ and $\vartheta = \frac{1}{4}\#$synapses. STDP was implemented via (5) with parameters $\alpha_+ = 0.03125$, $\tau_+ = 16.8$, $\alpha_- = 0.85\alpha_+$ and $\tau_- = 33.7$ also taken from [32]. Synaptic weights were clipped to fall in $[0, 1]$.

**Regularized STDP** consists of a small tweak of classical STDP in the online phase, and an additional offline regularization phase:

- *Online.* In the online phase, reduce the depotentiation bias from $0.85\alpha_+$ in the classical implementation to $\alpha_- = 0.75\alpha_+$.
- *Offline.* In the offline phase, modify synapses once per second according to

$$\Delta\mathbf{w}_j = \begin{cases} \gamma \cdot \left(\frac{3}{2} - \mathbf{w}_j\right) \cdot (\omega - s) & \text{if } \omega < s \\ \gamma \cdot (\omega - s) & \text{else,} \end{cases} \tag{13}$$

  where $s$ is output spikes per second, $\omega = 5Hz$ is the target rate and update factor $\gamma = 0.6$. The offline update rule is firing rate, and not spike, dependent.

Classical STDP has a depotentiation bias to prevent runaway potentiation feedback loops leading to seizures [5]. Since synapses are frequently renormalized offline we incorporate a weak exploratory (potentiation) bias during the online phase which helps avoid local minima.[3] This is in line with experimental evidence showing increased cortical activity during waking [30].

Since computing the sum of synaptic weights is non-physiological, we draw on Theorem 1 and use the neuron's firing rate when responding to uncorrelated inputs as a proxy for $\|\mathbf{w}\|_1$. Thus, in the offline phase, synapses receive inputs generated as in the online phase but without repeated patterns. Note that (12) has a larger pruning effect on stronger synapses, discouraging specialization. Motivated by Remark 4, we introduce bias $\left(\frac{3}{2} - \mathbf{w}_j\right)$ in the offline phase to ensure weaker synapses are downscaled more than strong synapses. For example, a synapse with $\mathbf{w}_i = 0.5$ is downscaled by *twice* as much as a synapse with weight $\mathbf{w}_j = 1.0$.

Regularized STDP alternates between 2 seconds online and 4 seconds offline, which suffices to renormalize synaptic strengths. The frequency of the offline phase could be reduced by decreasing the update factors $\alpha_\pm$, presenting stimuli less frequently (than 7 times per second), or adding inhibitory neurons to the system.

**Results.** A summary of results is presented in the table below: accuracy quantifies the fraction of spikes that co-occur with each pattern. Regularized STDP outperforms classical STDP on both patterns on average. It should be noted that regularized neurons were not only online for 20 seconds but also offline – and exposed to Poisson noise – for $40$ seconds. Interestingly, exposure to Poisson noise improves performance.

| Algorithm | Accuracy | |
| --- | --- | --- |
| | Pattern 1 | Pattern 2 |
| Classical | 54% | 39% |
| Regularized | 59% | 48% |

---

[3]The input stream contains a repeated pattern, so there is a potentiation bias in practice even though the net integral of STDP in the online phase is negative.

7

(a) Classical STDP &nbsp;&nbsp;&nbsp;&nbsp;&nbsp;&nbsp; (b) Regularized STDP
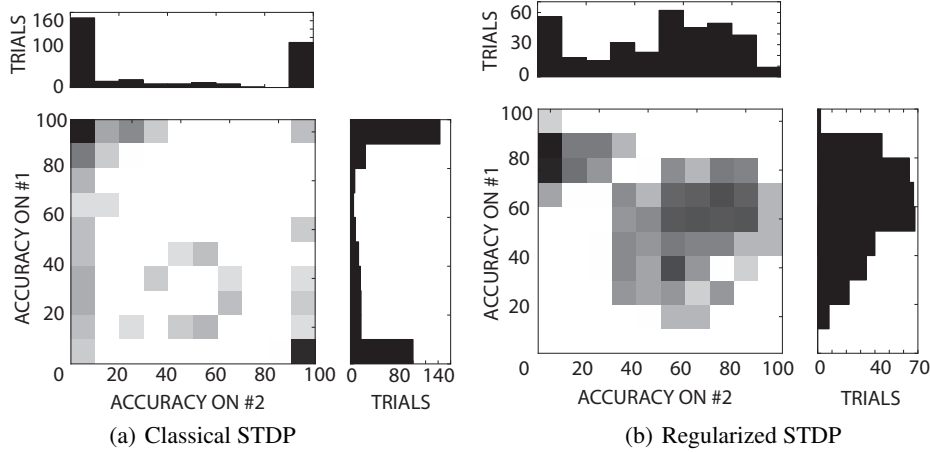
Figure 1: Accuracy after 20 seconds of exposure to a novel pattern.

Fig. 1 provides a more detailed analysis. Each panel shows a 2D-histogram (darker shades of gray correspond to more trials) plotting accuracies on both patterns simultaneously, and two 1D histograms plotting accuracies on the two patterns separately. The 1D histogram for regularized STDP shows a unimodal distribution for pattern #2, with most of the mass over accuracies of 50-90%. For pattern #1, which has been "unlearned" for twice as long as the training period, most of the mass is over accuracies of 50% to 90%, with a significant fraction "unlearnt". By contrast, classical STDP exhibits extremely brittle behavior. It completely unlearns the original pattern in about half the trials, and also fails to learn the new pattern in most of the trials.

Thus, as suggested by our analysis, introducing a regularization both improves the robustness of STDP and enables an exploratory bias by preventing runaway feedback leading to epileptic seizures.

## 7 &nbsp; Discussion

The selectron provides a bridge between a particular model of spiking neurons – the Spike Response Model [20] with the original spike-timing dependent plasticity rule [5] – and models that are amenable to learning-theoretic analysis. Our hope is that the selectron and related models lead to an improved understanding of the principles underlying learning in cortex. It remains to be seen whether other STDP-based models also have tractable discrete-time analogs.

The selectron is an interesting model in its own right: it embeds reward estimates into spikes and maximizes a margin that improves error bounds. It imposes a constraint on synaptic weights that: concentrates rewards/spike, tightens error bounds and improves guarantees on spiking efficacy. Although the analysis does not apply directly to continuous-time models, experiments show that a tweak inspired by our analysis improves the performance of a more realistic model. An important avenue for future research is investigating the role of feedback in cortex, specifically NMDA synapses, which may have interesting learning-theoretic implications.

**Acknowledgements.** We thank Timothée Masquelier for generously sharing his source code [32] and Samory Kpotufe for useful discussions.

## References

[1] Friston K, Kilner J, Harrison L: **A free energy principle for the brain**. *J. Phys. Paris* 2006, **100**:70–87.

[2] Rosenblatt F: **The perceptron: a probabilistic model for information storage and organization in the brain**. *Psychol Rev* 1958, **65**(6):386–408.

[3] Rumelhart DE, Hinton GE, Williams RJ: **Learning representations by back-propagating errors**. *Nature* 1986, **323**:533–536.

[4] Hinton G, Osindero S, Teh YW: **A Fast Learning Algorithm for Deep Belief Nets**. *Neural Computation* 2006, **18**:1527–1554.

8


[5] Song S, Miller KD, Abbott LF: **Competitive Hebbian learning through spike-timing-dependent synaptic plasticity**. *Nature Neuroscience* 2000, **3**(9).

[6] Seung HS: **Learning in Spiking Neural Networks by Reinforcement of Stochastic Synaptic Transmission**. *Neuron* 2003, **40**(1063-1073).

[7] Bohte SM, Mozer MC: **Reducing spike train variability: A computational theory of spike-timing dependent plasticity**. In *Advances in Neural Information Processing Systems (NIPS)* 2005.

[8] Legenstein R, Maass W: **A criterion for the convergence of learning with spike timing dependent plasticity**. In *Advances in Neural Information Processing Systems (NIPS)* 2006.

[9] Buesing L, Maass W: **Simplified rules and theoretical analysis for information bottleneck optimization and PCA with spiking neurons**. In *Adv in Neural Information Processing Systems (NIPS)* 2007.

[10] Legenstein R, Pecevski D, Maass W: **Theoretical analysis of learning with reward-modulated spike-timing-dependent plasticity**. In *Advances in Neural Information Processing Systems (NIPS)* 2008.

[11] Tishby N, Pereira F, Bialek W: **The information bottleneck method**. In *Proc. of the 37-th Annual Allerton Conference on Communication, Control and Computing*. Edited by Hajek B, Sreenivas R 1999.

[12] Balduzzi D, Tononi G: **What can neurons do for their brain? Communicate selectivity with spikes**. *To appear in Theory in Biosciences* 2012.

[13] Balduzzi D, Ortega PA, Besserve M: **Metabolic cost as an organizing principle for cooperative learning**. *Under review,* 2012.

[14] Nere A, Olcese U, Balduzzi D, Tononi G: **A neuromorphic architecture for object recognition and motion anticipation using burst-STDP**. *PLoS One* 2012, **7**(5):e36958.

[15] Schmiedt J, Albers C, Pawelzik K: **Spike timing-dependent plasticity as dynamic filter**. In *Advances in Neural Information Processing Systems (NIPS)* 2010.

[16] Anthony M, Bartlett PL: *Neural Network Learning: Theoretical Foundations*. Cambridge Univ Press 1999.

[17] Freund Y, Schapire RE: **Large Margin Classification Using the Perceptron Algorithm**. *Machine Learning* 1999, **37**(3):277–296.

[18] Ecker AS, Berens P, Keliris GA, Bethge M, Logothetis NK, Tolias AS: **Decorrelated neuronal firing in cortical microcircuits**. *Science* 2010, **327**(5965):584–7.

[19] Dan Y, Poo MM: **Spike timing-dependent plasticity of neural circuits.** *Neuron* 2004, **44**:23–30.

[20] Gerstner W: **Time structure of the activity in neural network models**. *Phys. Rev. E* 1995, **51**:738–758.

[21] Geman S, Bienenstock E, Doursat R: **Neural Networks and the Bias/Variance Dilemma**. *Neural Comp* 1992, **4**:1–58.

[22] Freund Y, Schapire RE: **Experiments with a New Boosting Algorithm**. In *Machine Learning: Proceedings of the Thirteenth International Conference* 1996.

[23] Schapire RE, Freund Y, Bartlett P, Lee WS: **Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods**. *The Annals of Statistics* 1998, **26**(5).

[24] Boucheron S, Bousquet O, Lugosi G: **Theory of classification: A survey of some recent advances**. *ESAIM: PS* 2005, **9**:323–375.

[25] Hasenstaub A, Otte S, Callaway E, Sejnowski TJ: **Metabolic cost as a unifying principle governing neuronal biophysics**. *Proc Natl Acad Sci U S A* 2010, **107**(27):12329–34.

[26] Fusi S, Drew P, Abbott L: **Cascade Models of Synaptically Stored Memories**. *Neuron* 2005, **45**:599–611.

[27] Fusi S, Abbott L: **Limits on the memory storage capacity of bounded synapses**. *Nature Neuroscience* 2007, **10**(4):485–493.

[28] Tononi G, Cirelli C: **Sleep function and synaptic homeostasis.** *Sleep Med Rev* 2006, **10**:49–62.

[29] Vyazovskiy VV, Cirelli C, Pfister-Genskow M, Faraguna U, Tononi G: **Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep**. *Nat Neurosci* 2008, **11**(2):200–8.

[30] Vyazovskiy VV, Olcese U, Lazimy Y, Faraguna U, Esser SK, Williams JC, Cirelli C, Tononi G: **Cortical firing and sleep homeostasis**. *Neuron* 2009, **63**(6):865–78.

[31] Maret S, Faraguna U, Nelson AB, Cirelli C, Tononi G: **Sleep and waking modulate spine turnover in the adolescent mouse cortex.** *Nat Neurosci* 2011, **14**(11):1418–1420.

[32] Masquelier T, Guyonneau, R and Thorpe SJ: **Competitive STDP-Based Spike Pattern Learning**. *Neural Computation* 2009, **21**(5):1259–1276.

[33] Roelfsema PR, van Ooyen A: **Attention-gated reinforcement learning of internal representations for classification.** *Neural Comput* 2005, **17**(10):2176–2214.



9


# Appendices

## A.1   The perceptron

We describe the perceptron to facilitate comparison with the selectron.

The perceptron's loss function and learning rule are most naturally expressed when inputs and outputs take values in $\{\pm 1\}$. We therefore present the perceptron in both $\pm 1$ and $0/1$ "coordinate systems".

Let us first relabel inputs and outputs from $0/1$ to $\pm 1$: $\mathbf{A} = 2\mathbf{X} - 1$ and $B = 2Y - 1$. Given input $\mathbf{a}$, the perceptron's output is determined according to

$$b = \rho_{\mathbf{w}}(\mathbf{a}) := \text{sign}\left(\mathbf{w}^{\mathsf{T}}\mathbf{a}\right),$$

where $\mathbf{w}$ is a real-valued $N$-vector specifying the perceptron's synaptic weights.

Given supervisor $\sigma : \mathbf{A} \to \{\pm 1\}$ that labels inputs as belonging to one of two classes, define $0/1$-valued loss function for the perceptron as

$$l_p(\mathbf{a}, \rho_{\mathbf{w}}, \sigma) := \mathbb{1}_{-\rho_{\mathbf{w}} \cdot \sigma(\mathbf{a})} = \begin{cases} 0 & \text{if } \rho_{\mathbf{w}}(\mathbf{a}) = \sigma(\mathbf{a}) \\ 1 & \text{else.} \end{cases}$$

The following learning rule

$$\Delta \mathbf{w}_j = \alpha \cdot \mathbf{a}_j \cdot \Big(\sigma(\mathbf{a}) - \rho_{\mathbf{w}}(\mathbf{a})\Big) = \alpha \cdot \begin{cases} \mathbf{a}_j & \text{if } \sigma(\mathbf{a}) = 1, \rho_{\mathbf{w}}(\mathbf{a}) = -1 \\ -\mathbf{a}_j & \text{if } \sigma(\mathbf{a}) = -1, \rho_{\mathbf{w}}(\mathbf{a}) = 1 \\ 0 & \text{else} \end{cases}$$

converges onto an optimal solution if the classes are linearly separable.

**The perceptron in "$0/1$ coordinates".**   For the sake of comparison, we reformulate the perceptron in "$0/1$ coordinates". If $\mathbf{w}_j \geq 0$ for all $j$, the mechanism of the perceptron is

$$y = f_{\mathbf{w}}(\mathbf{x}) := H\left(\mathbf{w}^{\mathsf{T}}\mathbf{x} - \frac{\|\mathbf{w}\|_1}{2}\right). \tag{A.14}$$

Similarly, we obtain loss function

$$l_p(\mathbf{x}, f_{\mathbf{w}}, \sigma) := \mathbb{1}_{-(2f_{\mathbf{w}}(\mathbf{x}) - 1) \cdot \sigma(\mathbf{x})} = \begin{cases} 0 & \text{if } \mathbf{x} = 1, \sigma(\mathbf{x}) = 1 \text{ or } \mathbf{x} = 0, \sigma(\mathbf{x}) = -1 \\ 1 & \text{else} \end{cases}$$

and learning rule

$$\Delta \mathbf{w}_j = \alpha \cdot (2\mathbf{x}_j - 1) \cdot \Big(\sigma(\mathbf{x}) - 2f_{\mathbf{w}}(\mathbf{x}) + 1\Big) = \alpha \cdot \begin{cases} \mathbf{a}_j & \text{if } \sigma(\mathbf{a}) = 1, \rho_{\mathbf{w}}(\mathbf{a}) = -1 \\ -\mathbf{a}_j & \text{if } \sigma(\mathbf{a}) = -1, \rho_{\mathbf{w}}(\mathbf{a}) = 1 \\ 0 & \text{else.} \end{cases} \tag{A.15}$$

**Non-biological features of the perceptron.**   We highlight two features of the perceptron. First, learning rule (A.15) is not selective. If the perceptron classifies an input incorrectly, it updates its synaptic weights regardless of whether it outputted a 0 or a 1. It is thus not an output spike-dependent learning rule. The main consequence of this is that the perceptron is forced to classify every input. The selectron, by contrast, actively ignores neuromodulatory signals when it does not spike.

Second, the perceptron requires a local error signal. Multilayer perceptrons are constructed by replacing the sign($\bullet$) in (A.14) with a differentiable function, such as the sigmoid, and backpropagating errors. However, backpropagation requires two pathways: one for feedforward spikes and another for feedback errors. In other words, the perceptron requires local error signals. A dedicated error pathway is biologically implausible [33], suggesting that the cortex relies on alternate mechanisms.

10

## A.2  Proof of Theorem 1

The theorem requires computing the second moment of a selectron's total current, given i.i.d. Bernoulli inputs. We also compute the expectation and the variance since these are of intrinsic interest.

**Lemma 5** (moments for i.i.d. Bernoulli inputs).
*For Bernoulli i.i.d. inputs on synapses, i.e. $P(\mathbf{x}_j = 0) = p$ for all $j$, we have*

$$\mathbb{E}[\langle \mathbf{w}, \mathbf{x}\rangle - \vartheta] = p \cdot \|\mathbf{w}\|_1 - \vartheta$$
$$\mathbb{V}[\langle \mathbf{w}, \mathbf{x}\rangle - \vartheta] = p(1-p) \cdot \|\mathbf{w}\|_2^2$$
$$\mathbb{E}[\langle \mathbf{w}, \mathbf{x}\rangle^2] = p(1-p) \cdot \|\mathbf{w}\|_2^2 + p^2 \cdot \|\mathbf{w}\|_1^2.$$

*Proof.* For the mean,

$$\mathbb{E}\Big[\langle \mathbf{w}, \mathbf{x}\rangle\Big] = \sum_{\mathbf{x}\in\mathbf{X}}\left[ P(\mathbf{x}_1)\cdots P(\mathbf{x}_n)\cdot\left(\sum_{j=1}^n \mathbf{w}_j \cdot \mathbf{x}_j\right)\right]$$
$$= \sum_{j=1}^n P(\mathbf{x}_j = 1)\cdot \mathbf{w}_j = \sum_{j=1}^n p\cdot\mathbf{w}_j = p\cdot\|\mathbf{w}\|_1,$$

since $\mathbf{w}_j \geq 0$ for all $j$.

For the variance,

$$\mathbb{V}\Big[\langle \mathbf{w}, \mathbf{x}\rangle\Big] = \sum_{\mathbf{x}\in\mathbf{X}} P(\mathbf{x})\langle \mathbf{w}, \mathbf{x}\rangle^2 - p^2\cdot\|\mathbf{w}\|_1^2$$
$$= \left[\sum_{i\neq j} p^2\cdot\mathbf{w}_i\mathbf{w}_j + \sum_j p\cdot\mathbf{w}_j^2\right] - \left[\sum_j p^2\cdot\mathbf{w}_j^2 + \sum_{i\neq j} p^2\cdot\mathbf{w}_i\mathbf{w}_j\right]$$
$$= p(1-p)\cdot\|\mathbf{w}\|_2^2.$$

The expression for $\mathbb{E}[\langle \mathbf{w}, \mathbf{x}\rangle^2]$ follows immediately. $\square$

**Theorem 1.** *Assuming the inputs on each synapse are i.i.d. Bernoulli variables with $P(spike) = p$, we have*

$$P\Big(f_{\mathbf{w}}(\mathbf{x}) = 1\Big) \leq p\cdot\left(\frac{\|\mathbf{w}\|_1}{\vartheta}\right)^2.$$

*Proof.* By Lemma 5, $\mathbb{E}\langle \mathbf{w}, \mathbf{x}\rangle^2 = p(1-p)\cdot\|\mathbf{w}\|_2^2 + p^2\cdot\|\mathbf{w}\|_1^2$. Applying Chebyshev's equality, $P(|X| > \epsilon) \leq \frac{\mathbb{E}X^2}{\epsilon^2}$, obtains

$$P\Big(f_{\mathbf{w}}(\mathbf{x}) = 1\Big) = P\Big(\langle \mathbf{w}, \mathbf{x}\rangle > \vartheta\Big) \leq \frac{p(1-p)\cdot\|\mathbf{w}\|_2^2 + p^2\cdot\|\mathbf{w}\|_1^2}{\vartheta^2}.$$

The result follows since $\|\mathbf{w}\|_2 \leq \|\mathbf{w}\|_1$. $\square$

## A.3  Proof of Theorem 2

We first compute the limit for STDP, and then move on to leaky integrate-and-fire neurons.

**Lemma 6** (fast time constant limit of STDP).
*The fast time constant limit of STDP, recall (5), is*

$$\lim_{\tau_\bullet \to 0} \Delta\mathbf{w}_{jk} = (\alpha_+ - \alpha_-)\cdot\delta_0(\Delta t)$$

*where $\delta_0$ is the Dirac delta.*

11

*Proof.* We start from the STDP curve:

$$\Delta \mathbf{w}_{jk} := \frac{\alpha_+}{\tau_+} \exp\left(\frac{t_j - t_k}{\tau_+}\right) H(t_k - t_j) - \frac{\alpha_-}{\tau_-} \exp\left(\frac{t_k - t_j}{\tau_-}\right) H(t_j - t_k)$$

$$= \frac{\alpha_+}{\tau_+} \exp\left(\frac{-\Delta t}{\tau_+}\right) H(\Delta t) - \frac{\alpha_-}{\tau_-} \exp\left(\frac{\Delta t}{\tau_-}\right) H(-\Delta t)$$

where $H$ is the Heaviside function. Note that we divide by $\tau_{\pm}$ to ensure that the area is constant as $\tau_{\pm} \to 0$. An equivalent approach is to rescale $\alpha$ and $\tau$ proportionately, so that both tend to zero.

Let us now compute the limit, in the sense of distributions, as $\tau_{\pm} \to 0$. Let $f$ be a test function in the Schwartz space. The linear form associated to

$$S_+(\Delta t) = \frac{\alpha_+}{\tau_+} \exp\left(\frac{-\Delta t}{\tau_+}\right) H(\Delta t)$$

is

$$\mathcal{S}_+ : f(x) \mapsto \int S_+(x) f(x) dx.$$

Thus,

$$\mathcal{S}_+(f) = \int \frac{\alpha_+}{\tau_+} \exp\left(\frac{-x}{\tau_+}\right) H(x) f(x) dx.$$

Integrating by parts obtains

$$\mathcal{S}_+(f) = \left[ -\alpha_+ \exp\left(\frac{-x}{\tau_+}\right) f(x) \right]_0^\infty - \int_0^\infty -\alpha_+ \exp\left(\frac{-x}{\tau_+}\right) f'(x) dx$$

so that

$$\mathcal{S}_+(f) = \alpha_+ \cdot f(0) + \tau_+ \cdot \mathcal{S}_+(f')$$

and $\lim_{\tau_+ \to 0} [\mathcal{S}_+(f)] = \alpha_+ \cdot f(0) = \alpha_+ \cdot \delta_0(f)$. A similar argument holds for the negative part of the STDP curve, so the fast time constant limit is

$$\Delta \mathbf{w}_{jk} = (\alpha_+ - \alpha_-) \cdot \delta_0(\Delta t).$$

$\square$

**Remark 5.** *If STDP incorporates a potentiation bias (in the sense that the net integral is positive [5]), then the fast time constant limit acts exclusively by potentiation.*

**Theorem 2.** *In the fast time constant limit, $\lim_{\tau_\bullet} \to 0$, the SRM transforms into a selectron with*

$$f_{\mathbf{w}}(t) = H\left(M_{\mathbf{w}}(t) - \vartheta\right) \quad \text{where} \quad M_{\mathbf{w}} = \sum_{\{j | t_j \geq t_k\}} \mathbf{w}_{jk} \cdot \delta_{t_k}(t).$$

*Moreover, STDP transforms into learning rule (4) in the unsupervised setting with $\nu(\mathbf{x}) = 1$ for all $\mathbf{x}$. Finally, STDP arises as gradient ascent on a reward function whose limit is the unsupervised setting of reward function (2).*

*Proof.* Setting $K_1 = K_2 = 0$, $\tau_s = \frac{1}{2}\tau_m$, $\frac{1}{K} = e^{-1} - e^{-2}$, and taking the limit $\tau_m \to 0$ yields

$$M_{\mathbf{w}}(t) = \sum_{\{j | t_j \geq t_k\}} \mathbf{w}_{jk} \cdot \delta_{t_k}(t).$$

By Lemma 6, taking limits $\tau_{\pm} \to 0$ transforms STDP into

$$\Delta \mathbf{w}_{jk} = \begin{cases} (\alpha_+ - \alpha_-) & \text{if } t_k = t_j \\ 0 & \text{else,} \end{cases}$$

which is a special case of the online learning rule for the selectron derived in §2, where the neuromodulatory response is $\nu(\mathbf{x}) = 1$ for all $\mathbf{x}$.

STDP is easily seen to arise as gradient ascent on

$$\arg\max_{\mathbf{w}} \sum_{t_k} \left[ \sum_{t_{k-1} < t_j \leq t_k} \mathbf{w}_j \cdot e^{\left(\frac{t_j - t_k}{\tau_+}\right)} - \sum_{t_k < t_j < t_{k+1}} \mathbf{w}_j \cdot e^{\left(\frac{t_k - t_j}{\tau_-}\right)} \right]. \qquad (A.16)$$

Taking the limit of (A.16) yields the special case of (2) where $\nu(\mathbf{x}) = 1$ for all $\mathbf{x}$. $\square$

12

Eq. (A.16) can be expressed in the shorthand

$$\arg\max_{\mathbf{w}} \left[ \sum_{t_k} \left( \mathbf{w}^{\mathsf{T}} \cdot \mathbf{d}(t_k) \right) \cdot f_{\mathbf{w}}(t_k) \right], \quad \text{where } \mathbf{d}_j(t) = \begin{cases} e^{\left( \frac{t_j - t}{\tau_+} \right)} & \text{if } t_j \le t \\ -e^{\left( \frac{t - t_j}{\tau_-} \right)} & \text{else} \end{cases}.$$

## A.4 Proof of Theorem 3

To prove the theorem, we first recall an error bound from [24]. To state their result, we need the following notation. Let

$$\mathcal{C}_\omega = \left\{ f(\mathbf{x}) = \text{sign} \left( \sum_{j=1}^{N} \mathbf{a}_j \cdot g^j(\mathbf{x}) \right) \ \middle| \ \mathbf{a}_j \in \mathbb{R}, \|\mathbf{a}\|_1 \le \omega \text{ and } g^j \in \mathcal{C} \right\}$$

where $\mathcal{C}$ is a class of base classifiers taking values in $\pm 1$.

Let $\phi : \mathbb{R} \to \mathbb{R}_+$ be a nonnegative cost function such that $\mathbb{1}_{x>0} \le \phi(x)$, for example $\phi(x) = (1 - x)_+$. Define

$$L(f) = \mathbb{E}\mathbb{1}_{-f(X)Y<0} \quad \text{and} \quad \widehat{L}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{-f(X_i)Y_i<0} \tag{A.17}$$

and

$$A(f) = \mathbb{E}\phi(-f(X)Y) \quad \text{and} \quad \widehat{A}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \phi(-f(X_i)Y_i). \tag{A.18}$$

**Theorem 7.** *Let $f$ be a function chosen from $\mathcal{C}_\omega$ based on data $(X_i, Y_i)_{i=1}^n$. With probability at least $1 - \delta$,*

$$L(f) \le \widehat{A}(f_n) + 2L_\phi \cdot \mathbb{E}\mathcal{R}ad_n(\mathcal{C}(X_1^n | f = 1)) + B\sqrt{\frac{2 \log \frac{1}{\delta}}{n}}, \tag{A.19}$$

*where $L_\phi$ is the Lipschitz constant of $\phi$, $B$ is a uniform upper bound on $\phi(-f(x)y)$, and $\mathcal{R}ad\big(\mathcal{C}(X_1^n)\big)$ is the Rademacher complexity of $\mathcal{C}$ on data $X_1^n$.*

*Proof.* See §4.1 of [24]. □

We are interested in errors of commission, not omission, so we consider modified versions of (A.17),

$$\mathcal{L}(f) = \mathbb{E}\left[ \mathbb{1}_{-f(X)Y<0} \cdot \mathbb{1}_{f>0} \right] \quad \text{and} \quad \widehat{\mathcal{L}}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \left[ \mathbb{1}_{-f(X)Y<0} \cdot \mathbb{1}_{f>0} \right], \tag{A.20}$$

and (A.18),

$$\mathcal{A}(f) = \mathbb{E}\left[ \phi(-f(X)Y) \cdot \mathbb{1}_{f>0} \right] \quad \text{and} \quad \widehat{\mathcal{A}}_n(f) = \frac{1}{n} \sum_{i=1}^{n} \left[ \phi(-f(X_i)Y_i) \cdot \mathbb{1}_{f>0} \right]. \tag{A.21}$$

where we multiply by the indicator function $\mathbb{1}_{f>0}$. This results in a discontinuous hinge function. We therefore have to modify the above theorem.

**Theorem 8.** *Let $f$ be a function chosen from $\mathcal{C}_\omega$ based on data $(X_i, Y_i)_{i=1}^n$. With probability at least $1 - \delta$,*

$$\mathcal{L}(f) \le \widehat{\mathcal{A}}_n(f) + 2(\mathbb{E}[\mathbb{1}_f] \cdot L_\phi + B) \cdot \mathbb{E}\mathcal{R}ad_n(\mathcal{C}(X_1^n | f = 1)) + 2B\sqrt{\frac{2 \log \frac{2}{\delta}}{n}},$$

*where $L_\phi$ is the Lipschitz constant of $\phi$, $B$ is a uniform upper bound on $\phi(-f(x)y)$, and $\mathcal{R}ad\big(\mathcal{C}(X_1^n)\big)$ is the Rademacher complexity of $\mathcal{C}$ on data $X_1^n$.*

13

*Proof.* We adapt the proof in [24]. By definition it follows that

$$\mathcal{L}(f) \leq \mathcal{A}(f)$$
$$= \widehat{\mathcal{A}}_n(f) + \mathcal{A}(f) - \widehat{\mathcal{A}}_n(f)$$

Now observe that $\sum_x [p(x)f(x)\mathbb{1}_S] = \sum_{x \in S} p(x)f(x) = p(S)\sum_x q(x)f(x)$, where $q(x) := p(x|x \in A)$. Thus, changing distributions from $P(x)$ to $Q(x) = P(x|f = 1)$, we can write

$$\mathcal{L}(f) \leq \widehat{\mathcal{A}}_n(f) + \left[ P(f = 1) \cdot \mathbb{E}_Q[\phi(-f(X)Y)] - \widehat{P}_n(f = 1) \cdot \sum_{\{i: f(x_i=1\}} \frac{\phi(-f(x_i)y_i)}{|\{i: f(x_i = 1\}|} \right]$$

$$= \widehat{\mathcal{A}}_n(f) + P(f = 1) \cdot \left[ A_Q(f) - \widehat{A}_{\widehat{Q}_n}(f) \right] + \widehat{A}_{\widehat{Q}_n}(f) \cdot \left[ \mathbb{E}_P f - \mathbb{E}_{\widehat{P}_n} f \right]$$

where $\widehat{P}_n(x_i) = \frac{1}{n}$ is the empirical distribution, $\widehat{P}_n(f = 1) = \frac{|\{i: f(x_i=1\}|}{n}$ and $\widehat{Q}_n(x_i) = \frac{1}{|\{i: f(x_i=1\}|}$.

Continuing,

$$\mathcal{L}(f) \leq \widehat{\mathcal{A}}_n(f) + \mathbb{E}_P[\mathbb{1}_f] \cdot \sup_{g \in \mathcal{C}} \left[ A_Q(g) - \widehat{A}_{\widehat{Q}_n}(g) \right] + \widehat{A}_{\widehat{Q}_n}(f) \cdot \sup_{g \in \mathcal{C}} \left[ \mathbb{E}_P g - \mathbb{E}_{\widehat{P}_n} g \right]$$

$$\leq \widehat{\mathcal{A}}_n(f) + 2\mathbb{E}[\mathbb{1}_f] \cdot L_\phi \cdot \mathbb{E}\mathcal{R}ad_n(\mathcal{C}(X_1^n|f = 1))$$

$$+ 2\widehat{A}_{\widehat{Q}_n} \cdot \mathbb{E}\mathcal{R}ad_n(\mathcal{C}(X_1^n|f = 1)) + 2B\sqrt{\frac{2\log\frac{2}{\delta}}{n}}$$

$$\leq \widehat{\mathcal{A}}_n(f) + 2(\mathbb{E}[\mathbb{1}_f] \cdot L_\phi + B) \cdot \mathbb{E}\mathcal{R}ad_n(\mathcal{C}(X_1^n|f = 1)) + 2B\sqrt{\frac{2\log\frac{2}{\delta}}{n}}$$

where we bound the two supremum's with high probability separately using Rademacher complexity and apply the union bound. The last inequality follows since $\phi \leq B$. $\square$

**Remark 6.** *The change of distribution is not important, since the Rademacher complexity is bounded by the distribution free VC-dimension in Lemma 9.*

To recover the setting of Theorem 3, we specialize to $\phi^\kappa(f) = (\kappa - f(X)Y)_+$, for $\kappa \geq 1$, so that the hinge loss can be written as $h^\kappa = \phi^\kappa(\mathbf{w}^\intercal \mathbf{x} - \vartheta) \cdot \mathbb{1}_{f(X)>0}$. The Lipschitz constant is $L_\phi = 1$.

Now let

$$\mathcal{F}_\omega = \left\{ f(\mathbf{x}) = H\left( \sum_{j=1}^N \mathbf{w}^j \cdot g^j(\mathbf{x}) - \vartheta \right) \; \middle| \; \|\mathbf{w}\|_1 \leq \omega \text{ and } g^j \in \mathcal{F} \right\},$$

where functions in $\mathcal{F}$ take values in $0/1$.

Function class $\mathcal{F}_\omega$ denotes functions implementable by a two-layer network of selectrons $S^k = \{n^k\} \cup \{n^j : n^j \to n^k\}$. The outputs of $g^j$ are aggregated, so that the function class $\mathcal{F}_\omega$ of subnetwork $S^k$ is larger than that of selectrons considered individually, i.e. $\mathcal{F}$.

**Lemma 9.** *The Rademacher complexity of $\mathcal{F}_\omega$ is upper bounded by*

$$\mathcal{R}ad_n(\mathcal{F}_\omega) \leq \omega \cdot \frac{\sqrt{2(N+1)\log(n+1)} + \frac{1}{2}}{\sqrt{n}}.$$

*Proof.* Given selectron $f_\mathbf{w}(\mathbf{x}) = \sum_j \mathbf{w}_j \cdot g_\mathbf{w}^j(\mathbf{x})$, let $\bar{g}^j := 2g^j - 1$ be the corresponding $\{\pm 1\}$-valued function. Then

$$\sum_{j=1}^N \mathbf{w}_j g^j(\mathbf{x}) - \vartheta = \frac{1}{2}\sum_{j=1}^N \mathbf{w}_j(\bar{g}^j(\mathbf{x}) + 1) - \vartheta = \frac{1}{2}\|\mathbf{w}\|_1 - \vartheta + \frac{1}{2}\sum_{j=1}^N \mathbf{w}_j \bar{g}^j(\mathbf{x}),$$

since $\mathbf{w}_i \geq 0$ for all $i$. Thus,

$$\mathcal{R}ad_n(\mathcal{F}_\omega) \leq \frac{\left|\frac{1}{2}\omega - \vartheta\right|}{\sqrt{n}} + \omega \cdot \mathcal{R}ad_n(\mathcal{F}).$$

14

The Rademacher complexity of $\mathcal{F}$ is upper bounded by the VC-dimension,

$$\mathcal{R}ad(\mathcal{F}(X_1^n)) \leq \sqrt{\frac{2V_\mathcal{F} \cdot \log(n+1)}{n}}$$

which for a selectron with $N$ synapses is at most $N + 1$.

Finally, note that if $\vartheta < 0$ then the selectron always spikes, and if $\vartheta > \|\mathbf{w}\|_1$ then it never spikes. We therefore assume that $0 \leq \vartheta \leq \omega$, which implies $|2\vartheta - \omega| \leq \omega$ and so

$$\frac{\left|\frac{1}{2}\omega - \vartheta\right|}{\sqrt{n}} \leq \frac{\omega}{2\sqrt{n}}.$$

$\square$

**Theorem 3.** *Suppose each selectron has $\leq N$ synapses. For any selectron $n^k$, let $S^k = \{n^k\} \cup \{n^j : n^j \to n^k\}$ denote a 2-layer feedforward subnetwork. For all $\kappa \geq 1$, with probability at least $1 - \delta$,*

$$\mathbb{E}\underbrace{[l(\mathbf{x}, f_\mathbf{w}, \nu)]}_{0/1 \text{ loss}} \leq \frac{1}{n}\sum_i \underbrace{h^\kappa\left(\mathbf{x}^{(i)}, f_\mathbf{w}, \nu(\mathbf{x}^{(i)})\right)}_{\text{hinge loss}} + \underbrace{\omega \cdot 2B \cdot \frac{\sqrt{8(N+1)\log(n+1)} + 1}{\sqrt{n}}}_{\text{capacity term}}$$

$$+ B \cdot \underbrace{\sqrt{\frac{2\log\frac{2}{\delta}}{n}}}_{\text{confidence term}} \qquad \text{where } B = \kappa + \omega - \vartheta.$$

*Proof.* Applying Theorem 8, Lemma 9, and noting that $\mathbb{E}[\mathbb{1}_f] \leq 1 \leq B = \kappa + \omega - \vartheta$, where $B$ is an upper bound on the hinge loss, obtains the result. $\square$

## A.5 An alternate error bound with hard margins

For the sake of completeness, we prove Corollary 11, an alternate to Theorem 3 that bounds a symmetric $0/1$ loss:

$$\mathbb{1}_{(2f_\mathbf{w}(\mathbf{x})-1)\cdot\nu(\mathbf{x})<0} = \begin{cases} 1 & \text{if } f_\mathbf{w}(\mathbf{x}) = 1, \nu(\mathbf{x}) = -1 \\ 1 & \text{if } f_\mathbf{w}(\mathbf{x}) = 0, \nu(\mathbf{x}) = 1 \\ 0 & \text{else.} \end{cases}$$

The loss penalizes the selectron when either (i) it fires when it shouldn't or (ii) it doesn't fire when it should.

We replace the modified hinge loss in Theorem 3 with a hard-margin loss. Following [24], let

$$\widehat{L}_n^\gamma(f) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{f(X_i)Y_i<\gamma} \quad \text{and} \quad \phi^\gamma(x) = \begin{cases} 0 & \text{if } x \leq -\gamma \\ 1 & \text{if } x \geq 0 \\ 1 + x/\gamma & \text{else.} \end{cases}$$

The following corollary of Theorem 7 is shown in [24].

**Corollary 10.** *Let $f_n$ be a function chosen from $\mathcal{C}_\omega$. For any $\gamma > 0$, with probability at least $1 - \delta$,*

$$L(f_n) \leq \widehat{L}_n^\gamma(f_n) + \frac{\omega}{\gamma}\sqrt{\frac{2V_\mathcal{C}\log(n+1)}{n}} + \sqrt{\frac{2\log\frac{1}{\delta}}{n}}.$$

*where $V_\mathcal{C}$ is the VC-dimension of $\mathcal{C}$.*

We use Corollary 10 to derive an alternate error bound for the selectron. Introduce hard-margin loss

$$\mathbb{1}_{\left(\mathbf{w}^\mathsf{T}\mathbf{x}-\vartheta\right)\cdot\nu(\mathbf{x})<\gamma} = \begin{cases} 1 & \text{if } \text{sign}(\mathbf{w}^\mathsf{T}\mathbf{x} - \vartheta) = \text{sign}(\nu(\mathbf{x})) \text{ and } |\mathbf{w}^\mathsf{T}\mathbf{x} - \vartheta| > \gamma \\ 0 & \text{else} \end{cases}$$

The following error bound exhibits a trade-off on the $0/1$ loss of a selectron that is controlled by $\gamma$.

15

**Corollary 11** (error bound with hard margins).
*For any $\gamma > 0$, with probability at least $1 - \delta$,*

$$\mathbb{E}\mathbb{1}_{(2f_{\mathbf{w}}(\mathbf{x})-1)\cdot\nu(\mathbf{x})<0} \leq \frac{1}{n}\sum_i \mathbb{1}_{(\mathbf{w}^{\mathsf{T}}\mathbf{x}^{(i)}-\vartheta)\cdot\nu(\mathbf{x}^{(i)})<\gamma} + \frac{\omega}{\gamma}\cdot\frac{\sqrt{8(N+1)\log(n+1)}+1}{\sqrt{n}} + \sqrt{\frac{2\log\frac{1}{\delta}}{n}}.$$

*Proof.* Follows from Lemma 9 and Corollary 10. □

As $\gamma$ increases, the size of the margin required to avoid counting towards a loss increases. However, the capacity term is multiplied by $\frac{1}{\gamma}$, and so reduces as the size of $\gamma$ increases.

Thus, the larger the margin, on average, the higher the value we can choose for $\gamma$ without incurring a penalty, and the better the bound in Corollary 11.

## A.6  Proof of Theorem 4

It is helpful to introduce some notation. Given $\mathbf{x} \in \mathbf{X}$, let $S_a^j = \{\mathbf{x}|\mathbf{x}_j = a\}$. Let $\nu_{jk=11}$, defined by equation

$$\nu_{jk=11} \cdot \left[\sum_{\mathbf{x}\in S_1^j} P(\mathbf{x}|\mathbf{x}_j=1)(\mathbf{w}^{\mathsf{T}}\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x})\right] = \sum_{\mathbf{x}\in S_1^j} P(\mathbf{x}|\mathbf{x}_j=1)\nu(\mathbf{x})(\mathbf{w}^{\mathsf{T}}\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x}),$$

quantify the average contribution of neuromodulators when selectrons $n^j$ and $n^k$ both spike. Similarly, let $\nu_{jk=01}$ quantify the average contribution of neuromodulators when $n^k$ spikes and $n^j$ does not, i.e. the sum over $\mathbf{x} \in S_0^j$.

If upstream neurons are reward maximizers then spikes by $n^j$ should predict higher neuromodulatory rewards, on average, than not firing. We therefore assume

$$\nu_{jk=11} \geq \nu_{j=10}. \tag{*}$$

**Theorem 4**  *Let $p_j := \mathbb{E}[Y^j]$ denote the frequency of spikes from neuron $n^j$. Under assumption (*), the efficacy of $n^j$'s spikes on $n^k$ is lower bounded by*

$$\frac{1}{\nu_{jk=11}}\cdot\frac{\delta R^k}{\delta \mathbf{x}_j} \geq \frac{\mathbf{w}_j \cdot \mathbb{E}[Y^jY^k]}{p_j} + \frac{2\mathbb{E}\left[Y^jY^k\cdot\left((\mathbf{w}^{\backslash j})^{\mathsf{T}}\mathbf{x}-\vartheta\right)\right]}{p_j(1-p_j)} - \frac{\mathbb{E}\left[Y^k\cdot\left((\mathbf{w}^{\backslash j})^{\mathsf{T}}\mathbf{x}-\vartheta\right)\right]}{1-p_j}$$

*where*

$$\mathbf{w}_i^{\backslash j} := \begin{cases} \mathbf{w}_i & \text{if } i \neq j \\ 0 & \text{else.} \end{cases}$$

*Proof.*

$$\frac{\delta R^k}{\delta \mathbf{x}_j} = \mathbb{E}[R^k|\mathbf{x}_j=1] - \mathbb{E}[R^k|\mathbf{x}_j=0]$$

$$= \sum_{\mathbf{x}\in S_1^j} P(\mathbf{x}|\mathbf{x}_j=1)\nu(\mathbf{x})(\mathbf{w}^{\mathsf{T}}\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x}) - \sum_{\mathbf{x}\in S_0^j} P(\mathbf{x}|\mathbf{x}_j=0)\nu(\mathbf{x})(\mathbf{w}^{\mathsf{T}}\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x})$$

$$= \nu_{jk=11}\cdot\sum_{\mathbf{x}\in S_1^j} P(\mathbf{x}|\mathbf{x}_j=1)(\mathbf{w}^{\mathsf{T}}\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x}) - \nu_{jk=01}\cdot\sum_{\mathbf{x}\in S_0^j} P(\mathbf{x}|\mathbf{x}_j=0)(\mathbf{w}^{\mathsf{T}}\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x}).$$

16

Assumption (*) implies

$$\frac{1}{\nu_{jk=11}} \cdot \frac{\delta R^k}{\delta \mathbf{x}_j} \geq \sum_{\mathbf{x}\in S_1^j} P(\mathbf{x}|\mathbf{x}_j=1)(\mathbf{w}^\intercal\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x}) - \sum_{\mathbf{x}\in S_0^j} P(\mathbf{x}|\mathbf{x}_j=0)(\mathbf{w}^\intercal\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x})$$

$$= \sum_{\mathbf{x}\in S_1^j} \frac{P(\mathbf{x})}{P(\mathbf{x}_j=1)}(\mathbf{w}^\intercal\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x}) - \sum_{\mathbf{x}\in S_0^j} \frac{P(\mathbf{x})}{P(\mathbf{x}_j=0)}(\mathbf{w}^\intercal\mathbf{x}-\vartheta)f_{\mathbf{w}}(\mathbf{x})$$

$$= \frac{1}{p_j}\mathbb{E}\left[Y^j Y^k \cdot (\mathbf{w}^\intercal\mathbf{x}-\vartheta)\right] - \frac{1}{1-p_j}\mathbb{E}\left[(1-Y^j)Y^k \cdot (\mathbf{w}^\intercal\mathbf{x}-\vartheta)\right]$$

The result follows by direct computation. $\qquad\square$

## A.7 Generation of Poisson spike trains

Spike trains are generated following [32]. Neurons have 200 synaptic inputs. Time is discretized into $1ms$ bins. At each time step spikes are generated according to a Poisson process where the rate varies as follows:

1. a synapse has probability $r \cdot dt$ of emitting a spike where $r$ is clipped in $[0, 90]Hz$.
2. $dr = s \cdot dt$ where $s$ is clipped in $[-1800, 1800]Hz$.
3. the rate of change $ds$ of $s$ is uniformly picked from $[-360, 360]Hz$.

The resulting spike train has an average firing rate of about $44Hz$. Masquelier *et al* also add a mechanism to ensure each synapse transmits a spike after at most $50ms$ which we do not implement.

Repeated patterns are sampled from the process above for $50ms$. The pattern is then cut-and-paste over the original Poisson spike train for $\frac{1}{3}$ of the total number of $50ms$ blocks. The cut-and-pasting is imposed on a randomly chosen (but fixed) subset containing $\frac{1}{2}$ of the neuron's synapses.

17