Reward-modulated spike-timing-dependent plasticity with a dynamic spike timing rule and inhibitory plasticity

Heidi Gonzalez *h6gonzal@ucsd.edu*

Jason Keller *jakeller@ucsd.edu*

Neurosciences Graduate Program
University of California, San Diego
La Jolla, CA 92093

Abstract

The viability of spike-timing-dependent plasticity (STDP) to explain learning processes is controversial, although recent developments of reward-modulated STDP (RM-STDP) models provide a plausible substrate. However, evidence has also emerged to show that rewards themselves can modify the STDP rule. In this modeling study, we use a dynamic STDP rule to show that such modification can lead to network instability, and furthermore that inhibitory STDP may be able to balance networks to restore asynchronous, stochastic firing. We conclude that further experimental and modeling work is necessary to arrive at a biologically plausible mechanism of learning.

1 Introduction

Although spike-timing-dependent plasticity (STDP) has emerged as a computationally powerful mechanism of cellular learning, its functional significance remains controversial. The canonical STDP rule indicates that causal presynaptic-before-postsynaptic firing patterns lead to long-term synaptic potentiation (LTP) at excitatory glutamatergic synapses, whereas acausal postsynaptic-before-presynaptic firing patterns lead to long-term synaptic depression (LTD) at those synapses [1]. Reward-modulated STDP (RM-STDP) supposes that neuromodulation can act as a third signal, in addition to presynaptic and postsynaptic spike times, to connect cellular and behavioral plasticity. This idea has spawned several models (see [2] and [3] for instructive examples) that attempt to explain the process of learning across several levels of analysis from molecules to behavior.

These models typically invoke an "eligibility trace" that is inspired by the synaptic tagging hypothesis [4] and represents the underlying biochemical processes that allow a synapse to be potentiated or depressed when change in behavioral state occurs, even if occurring several seconds after the presynaptic or postsynaptic neurons have fired. The mechanisms of these slow chemical processes are still emerging, but they likely involve a rise in intracellular calcium concentration, back-propagating action potentials (BPAPs), and relief of the Mg^{2+} block on NMDA receptors. The behavioral state change may be modeled as the dopamine (DA) concentration above or below some baseline level, since the role of DA in signaling reward and punishment has been well established from electrophysiology and microdialysis studies [5]. This DA reward signal is applied globally to all synapses in a noisy and stochastic network simulating *in vivo* conditions, and only eligible excitatory synaptic

weights are updated. Therefore, a network can learn to potentiate only those firing patterns which causally led to a reward in an unsupervised and thus biologically plausible manner.

RM-STDP models provide a realistic explanation of the neural correlates of learning, but many of their assumptions remain untested, and furthermore recent *in vitro* studies have proven them false. In particular, several studies have investigated the effect of neuromodulation on the STDP rule itself and found that neuromodulators such as DA, noradrenaline, and acetylcholine can alter STDP rules in complex ways that depend on the brain area and cell type under investigation [6], as well as the location of the synapse on the dendritic tree [7]. Additionally, inhibitory plasticity has received increasing attention recently [8] and has been hypothesized to balance network plasticity, even according to specific STDP rules [9]. Thus the present study addresses the following question: does the addition of both a dynamic STDP rule and inhibitory STDP into an existing RM-STDP model allow for a more biologically plausible model of learning?

2 Methods

Many RM-STDP models exist, both clock-driven and event-driven. Following the guidance of [10], we chose a clock-driven model since evaluating dynamics at each time step was shown to be critical for choosing the correct synapses that are causally related to a postsynaptic cell driven over threshold. Although Izhikevich's 2007 model [2] is over-simplified more than some other models, it is a freely available and compact model that allows for relatively straightforward modifications. This model includes 1000 neurons (800 excitatory, 200 inhibitory) in a cortical "minicolumn" which are randomly connected with 10% probability. All cells receive random noisy "thalamic" input, and one arbitrary excitatory synapse is conditioned by delivering a dopamine reward 1-3 seconds after the pre-post pair fires within a 20 millisecond window. All excitatory synapses are initialized at strength = 1 mV and have a maximum strength of 4 mV except for the conditioned synapse, which is zeroed at the start of the simulation. Inhibitory synapses are initialized to -1 mV and have a maximum strength of -4 mV. Over one hour of simulation time, the success of the RM-STDP model in unsupervised learning is qualitatively determined by two metrics: (1) the conditioned synapse is maximally potentiated, and (2) the network remains in an asynchronous, stable firing regime as indicated by an exponential distribution of synaptic strengths favoring low strengths.

We modeled the dynamics of the STDP rule as shown in Figure 1 below and motivated by the *in vitro* results of [11]. This study showed that acausal firing patterns that normally mediate LTD could mediate LTP if DA was applied to the slice preparation. Although this experiment was done on hippocampal slices, no data is currently available in cortical areas. Furthermore, we do not know the time course of the change in the STDP rule, but we make the assumption that the rise in second messenger molecules such as cyclic AMP (cAMP) that underlie the change can be modeled with a Gaussian time course (see Equation 1) that is slower than the extracellular DA concentration, which rises instantaneously after reward and decays exponentially.

$$\Delta negSpikeTiming = gaussianMax * e^{\frac{(currentTime - lastRewardTime)^2}{2 * gaussianStdDev^2}} \quad (1)$$

Following the inhibitory STDP model from [9], we modeled changes in inhibition with a static STDP curve in which inhibitory synapses are strengthened for near-coincident pre- and post-synaptic spikes, but weakened otherwise. Additionally, inhibitory plasticity, in contrast to excitatory plasticity, is not gated by the DA reward signal. This scheme has been demonstrated experimentally by [12], although the exact time constants for decay and values for maximum and minimum inhibitory synaptic strength were chosen based on the excitatory model parameters.

3 Results

3.1 Addition of a dynamic STDP rule

To verify that our dynamic STDP rule worked as expected, we examined the synaptic eligibility trace, c(t), of the conditioned synapse using the dynamic rule along with the original c(t) of that synapse using a static rule. A positive value for c(t) means that LTP will be induced if the current extracellular DA concentration is above baseline, and a negative value means that LTD will be induced accordingly. As shown in Figure 2 below, c(t) using the dynamic STDP rule can indeed switch LTD to LTP for acausal spike timings.

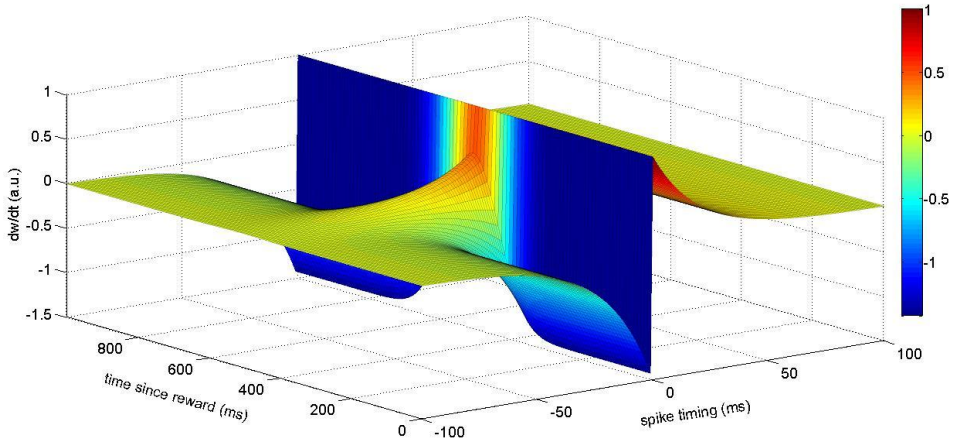

Figure 1: The dynamic (three dimensional) STDP rule used in our simulations. A reward (rise in extracellular DA concentration) is applied at "time since reward" = 0. The "spike timing" axis denotes $t_{POST} - t_{PRE}$ such that negative numbers on this axis represent the acausal, LTD side of the standard STDP curve. The dw/dt axis shows that the absolute value of LTD before reward is larger than that of LTP. At negative spike timings, LTD slowly changes to LTP and then back to LTD at a rate corresponding to the standard deviation of the Gaussian in Equation 1, whilst LTP is constant.

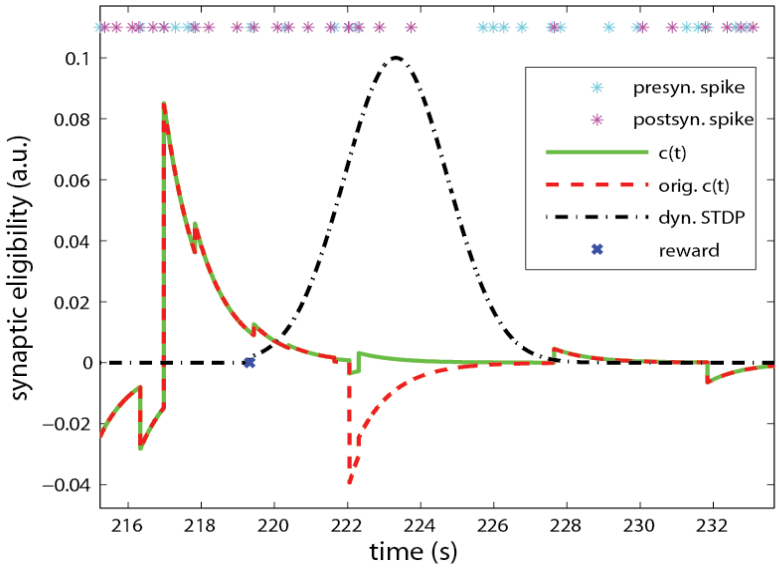

Figure 2: Example synaptic eligibility traces for the conditioned synapse using a dynamic STDP rule (c(t), green solid line) and a static STDP rule (orig. c(t), red dashed line). The black dash-dot line shows the time course of the STDP rule modification after a reward indicated by a blue 'x'. Asterisks indicate presynaptic (cyan) and postsynaptic (magenta) spike times for the conditioned synapse.

Next we evaluated the success of the RM-STDP model as we varied the *gaussianStdDev* parameter from Equation 1, using the metrics described above in the Methods section. In all simulations the conditioned synapse reached the maximum synaptic strength (capped at 4 mV). However, as anticipated the network became more unstable for wider Gaussian STDP modulations since synapses had more time to be eligible for LTP instead of LTD. The results are presented in Figure 3, which shows the excitatory synapse strength distributions for different values of *gaussianStdDev,* after the full hour of simulation time. Although this the distributions shown are only snapshots of rapidly evolving networks, the general trend shows that the synaptic strengths tend to more positive values (as a qualitative indicator of instability) for slower STDP rule dynamics, subject to random variation in the network connectivity.

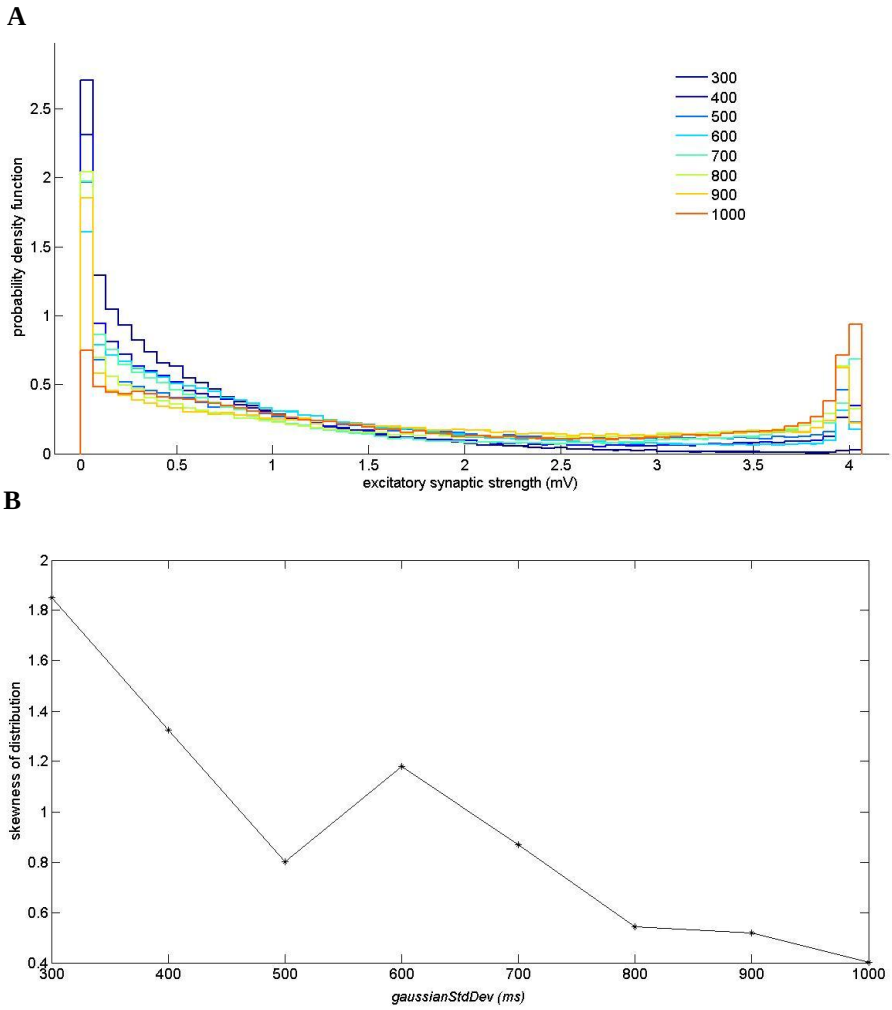

Figure 3: Excitatory synapse strength distributions and their skewness after full simulations for different values of *gaussianStdDev* from Equation 1. (A) Synapse strength distributions; the legend gives the Gaussian standard deviations in milliseconds. (B) Skewness of the distributions in (A). More positive values indicate more unimodal, positively-skewed distributions with values clustered at low synaptic strength.

3.2 Addition of inhibitory STDP

The rules governing STDP have been extensively studied for excitatory connections, however little attention has been given to the rules of STDP for inhibitory synapses. However, it has

recently become clear that inhibitory connections must be plastic in order to balance the strengthening of excitatory synapses. We reasoned that any increase in inhibitory synaptic weights might be able to balance the superfluous excitatory plasticity induced by a dynamic STDP rule that allowed LTD to become LTP at excitatory synapses. When we applied the static inhibitory STDP described in the Methods section to our model with a dynamic STDP rule at excitatory synapses, we observed a tendency for inhibitory synaptic weights to scale up and attempt to normalize the excitatory network as it grew unstable, as shown in Figure 4 below.

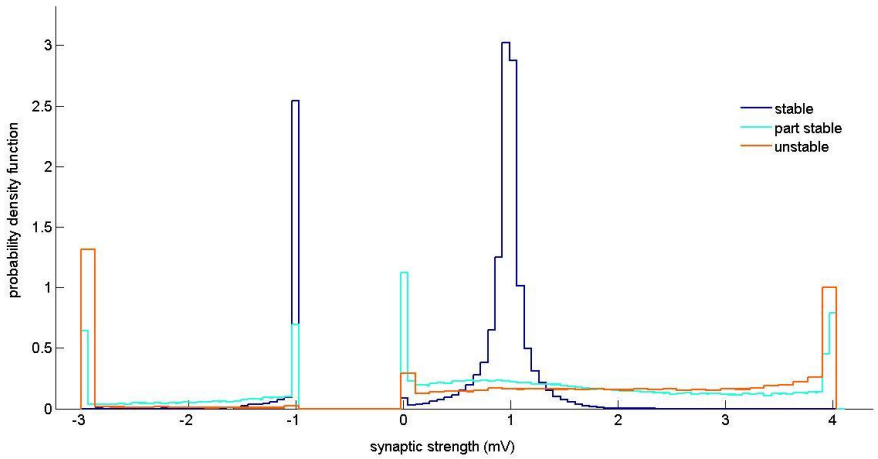

Figure 4: Both inhibitory (< 0 mV) and excitatory (> 0 mV) synapse strength distributions at different levels of instability with a model using both a dynamic excitatory STDP rule and a static inhibitory STDP rule. In each simulation, *gaussianStdDev* from Equation 1 was 1000 ms. In the "stable" case near the beginning of the simulation, synaptic strengths are near their initial values. In the "part stable" case many excitatory strengths have increased to their maximum value, and inhibitory strengths quickly follow to attempt to normalize the network. In the "unstable" case, both excitatory and inhibitory distributions converge on their maximum strengths.

Although we could not fully stabilize the network using our excitatory plasticity rule, several parameters could potentially affect the results. For instance, allowing for greater inhibitory synaptic strengths, greater inhibitory connectivity, or simply more inhibitory neurons may allow network stability given the initial results above. Clearly more work and a quantitative description of stability, as discussed below, are needed to draw solid conclusions.

4 Conclusions

In light of new evidence regarding the alterations of the STDP curve at excitatory synapses during neuromodulation, the RM-STDP model of learning becomes less plausible. However, our experiments show that the addition of a dynamic STDP rule can allow for a more biologically plausible RM-STDP model, albeit dependent on underlying assumptions of the timing of the intracellular chemical processes produced by neuromodulation. We conclude that reasonable estimates of this timing on the order seconds can cause a network to go unstable, but that inhibitory STDP has the potential to balance this instability.

Our results are mainly qualitative due to limited simulation time, but future work could make more quantitative predictions using repeated simulations and a probabilistic description of network instability. Also, numerous other improvements could be made to the modified Izhikevich model used here. First, with greater computing power the network could obviously be expanded to contain more neurons, more realistic anatomy, more cell types, and compartmental neuron models. This will likely have profound implications since the location of synapses on a tree greatly affects their STDP rule, and neuromodulation can alter the propagation of BPAPs [7]. Second, although data is still scarce regarding the effect of different neuromodulators on the STDP curve and their combinatorial effects, future

dynamic STDP rules could account for multiple sources of neuromodulation. Third, different inhibitory STDP rules, or even a dynamic inhibitory STDP rule, could be used, especially for different inhibitory cell types in a more detailed network. Again, data is scarce here, so future *in vitro* experiments are needed to provide initial guidance on model parameters.

Even from this short and incomplete laundry list it is clear that many important questions remain outstanding in establishing the significance of STDP in neural computation. For example, how do these STDP rules vary across different areas of the brain and different cell types with various neuromodulator receptor subtypes and intracellular signaling molecules? We hope that future efforts in applying dynamical systems approaches to STDP and specifically RM-STDP can yield a greater understanding of learning processes from molecules to behavior.

References

[1] Markram H, Gerstner W & Sjöström PJ (2011). A history of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 3.

[2] Izhikevich EM (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cerebral Cortex*, 17(10), pp.2443–2452.

[3] Legenstein R, Pecevski D & Maass W (2008). A Learning Theory for Reward-Modulated Spike-Timing-Dependent Plasticity with Application to Biofeedback L. J. Graham, ed. *PLoS Computational Biology*, 4(10), p.e1000180.

[4] Redondo RL & Morris RGM (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, 12(1), pp.17-30.

[5] Schultz W (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80, pp.1-27.

[6] Pawlak V, Wickens JR, Kirkwood A & Kerr JND (2010). Timing is not everything: neuromodulation opens the STDP gate. *Frontiers in Synaptic Neuroscience*, 2.

[7] Froemke RC, Letzkus JJ, Kampa BM, Hang GB & Stuart GJ (2010). Dendritic synapse location and neocortical spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 2.

[8] Kullmann DM & Lamsa KP (2011). LTP and LTD in cortical GABAergic interneurons: emerging rules and roles. *Neuropharmacology*, 60, pp.712-719.

[9] Vogels TP, Sprekeler H, Zenke F, Clopath C & Gerstner W (2011). Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. *Science*. Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1211095 [Accessed December 6, 2011].

[10] Rudolph M & Destexhe A (2007). How much can we trust neural simulation strategies? *Neurocomputing*, 70, pp.1966 - 1969.

[11] Zhang JC, Lau PM & Bi GQ (2009). Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. *Proceedings of the National Academy of Sciences*, 106(31), pp.13028–13033.

[12] Woodin MA, Ganguly K & Poo M (2003). Coincident pre-and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl-transporter activity. *Neuron*, 39(5), pp.807–820.