Network Structures Arising from Spike-Timing Dependent Plasticity

Baktash Babadi

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

under the Executive Committee

of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Network Structures Arising from Spike-Timing Dependent Plasticity

Baktash Babadi

Spike-timing dependent plasticity (STDP), a widespread synaptic modification mechanism, is sensitive to correlations between presynaptic spike trains, and organizes neural circuits in functionally useful ways. In this dissertation, I study the structures arising from STDP in a population of synapses with an emphasis on the interplay between synaptic stability and Hebbian competition, explained in Chapter 1. Starting from the simplest description of STDP which relates synaptic modification to the intervals between pairs of pre- and postsynaptic spikes, I show in Chapter 2 that stability and Hebbian competition are incompatible in this class of "pair-based" STDP models, either when hard bounds or soft bounds are imposed to the synapses. In chapter 3, I propose an alternative biophysically inspired method for imposing bounds to synapses, i.e. introducing a small temporal shift in the STDP window. Shifted STDP overcomes the incompatibility of synaptic stability and competition and can implement both Hebbian and anti-Hebbian forms of competitive plasticity.

In light of experiments the explored a variety of spike patterns, STDP models have been augmented to account for interactions between multiple pre- and postsy-

naptic action potentials. In chapter 4, I study the stability/competition interplay in three different proposed multi-spike models of STDP. I show that the "triplet model" leads to a partially steady-state distribution of synaptic weights and induces Hebbian competition. The "suppression model" develops a stable distribution of weights when the average weight is high and shows predominantly anti-Hebbian competition. The NMDAR-based model can lead to either stable or partially stable synaptic weight distribution and exhibits both Hebbian and anti-Hebbian competition, depending on the parameters. I conclude that multi-spike STDP models can produce radically different effects at the population level depending on how they implement multi-spike interactions.

Finally in chapter 5, I focus on the types of global structures that arise from STDP in a recurrent network. By analyzing pairwise interactions of neurons through STDP and also numerical simulations of a large network, I show that conventional pair-based STDP functions as a loop-eliminating mechanism in a network of spiking neurons and organizes neurons into in- and out-hubs. Loop-elimination increases when depression dominates and decreases when potentiation dominates. STDP with dominant depression implements a buffering mechanism for network firing rates, and shifted STDP can generate recurrent connections in a network, and also functions as a homeostatic mechanism that maintains a roughly constant average value of the synaptic strengths. In conclusion, studying pairwise interactions of neurons through STDP provides a number of important insights about the structures that arise from this plasticity rule in large networks. This approach can be extended to networks with more complex STDP models and more structured external input.

Contents

List of Figures

To the green memory of the courageous women and men who lost their precious lives in their struggle for freedom in the streets of Tehran in 2009-2010.

Acknowledgments

I owe my deepest gratitude to my advisor Larry Abbott. This thesis would not have been possible without his encouragement, guidance and support from the initial to the final level. Apart from countless number of specific scientific details I learned from him, his greatest lesson for me was his eclectic approach toward scientific problems, his genuine appreciation of novel ideas, and the exceptional balance between flexibility and discipline is his research. I truly wish that his approach would have a reflection in my own future career. Science aside, I feel honored that I spent hours to discuss various social and political issues with Larry in the past four years, and found numerous common grounds with him.

I would like to thank Ken Miller for his invaluable guidance throughout my PhD research. Quite a few times, a problem that was perplexing me for weeks was solved at once by a keen comment from Ken. Whenever I had any question, I didn't hesitate to refer directly to Ken, whose office is always open to all students and PostDocs and his vast knowledge in virtually every field of neuroscience is an asset for everybody in the Center for Theoretical Neuroscience. Above all, Larry

and Ken are epitomes of socially responsible scientists to me, and it was truly a privilege to work in the center directed by them.

I should also thank Rafa Yuste and Liam Paninski for acquainting me with the great challenge of relating theory to experiments, for giving me the opportunity to work with them during my rotations, and for being on my thesis committee. I should also thank Robert Froemke, who generously accepted to be on my defense committee.

The theory center would not have been what it is without its current and previous brilliant members. My special thanks go to Max Nikitchenko, Greg Wayne, Xaq Pitkow, Michael Vidne, Yashar Ahmadian, and Dan Rubin for their friendship, support and camaraderie during my years in graduate school. I should also thank my friends and colleagues Evan Schaffer, Pablo Jercog, Alex Ramirez, Andrew Fink, late Daniel Ferreira, Mattia Rigotti, Saul Kato, David Sussillo, Qing Yuan, Sean Luo, Sean Escola, Kanaka Rajan, Vladimir Itskov, Anthony DeCostanzo, and Taro Toyoizomi.

I am grateful to my brothers Behtash and Mehrtash who patiently assisted me to fill up the gaps in my mathematical knowledge, in every step of my research. There are hardly any moments in my life more joyous than the hours of endless discussions that I had with them. Their presence was my sole solace in the most solemn moments of the past five years. I am afraid my words come short of conveying my gratitude toward my parents Shirin Mokhtari and Mohammad Babadi.

Their warm emotional and intellectual support is the most precious asset that I have in my entire life. Indubitably, I owe my passion for science and critical inquiry to their upbringing.

Finally, in the last two years of my PhD research, while my mind was engaged in research problems reported in this thesis, my heart was throbbing in solidarity with Iranian youth who were struggling for democracy and human rights in the streets of Tehran. Therefore, I humbly dedicate this dissertation to the memory of the dear ones who lost their lives in this historical battle.

Chapter 1

Introduction

Neither body is concealed from mind, nor mind is concealed from body,
however nobody is summoned to see the mind.
Mowlana Jalal ad-Din Rumi

1.1 Historical background

Questions about mind and its relation to body have bewildered humankind since the very dawn of ancient civilization. At least since the time of Hippocrates (400 BCE), who declared the brain to be the 'seat of intelligence' (Finger, 2001), accumulating evidence implied that mental phenomena are tightly correlated to the brain. However, as paradoxical as it may sound, it was not until the proposal of the notorious Cartesian dualism that systematic observations of the brain became widespread. By ascribing all mental processes to an intangible distant substance and advocating a mechanistic approach to studying the body, Descartes removed the halo of holiness that was spun around the body during the middle-ages, thereby catalyzed scientific observations of the body in general and the brain in particular (Chomsky, 1983), reportedly pioneered by Descartes himself.

Systematic observations of the brain culminated in the monumental work of Santiago Ramón y Cajal who established the 'neuron doctrine', the fundamental concept that the nervous system is made up of discrete individual cells (Finger, 2001). Ever since Cajal, an ultimate goal of brain sciences - possibly the most important one - is to relate the biological substrate of mental processes to their phenomenological aspects. In pursuit of this goal, studies of the brain have focused on three different levels of abstraction: the cellular-molecular level, the network-systems level, and the cognitive level. With the advent of molecular biology and modern genetics in recent decades, the study of cellular and sub-cellular mechanisms of the nervous system has gained unprecedented momentum. On the other hand, the phenomenological study of cognitive processes has a long history of its own, highly overlapping with the well-establish field of psychology, and spurred recently by functional neural imaging techniques. Nevertheless, the long-desired bridging between these two levels can only be accomplished by studying the mass actions emerging from functionally organized networks of neurons that give rise to cognitive phenomena.

1.2 The significance of theoretical neuroscience

The mid-level between cellular/molecular and psychological approaches is where theoretical and computational neuroscience can play its most significant role, borrowing techniques from statistical physics, nonlinear dynamics, theory of computation and related disciplines. Neural network studies are focussed on three different aspects: connectivity, activity and plasticity. Interestingly, these three aspects

make an indivisible triad within the realm of neural networks: connectivity determines the interaction of neurons in a population, neural activity arises from the pattern of connectivity, and plasticity in turn regulates connectivity according to activity. This triad and the interactions between its three components have provided my guideline in thinking about neural networks throughout my research.

1.3 Hebbian theory: Activity dependent plasticity

During my PhD research, I mainly concentrated on plasticity out of the above-mentioned triad. The conceptual basis for the role of plasticity in mediating network connectivity and activity is provided by the so-called 'Hebbian Theory'. Donald Hebb postulated that, through some biophysical mechanisms, neurons are organized into engrams, hypothetical neural network structures that can store memory traces. Hebbian postulate is best explained in the words of Hebb himself (Hebb, 1949):

> Let us assume that the persistence or repetition of a reverberatory activity (or 'trace') tends to induce lasting cellular changes that add to its stability. When anaxonof cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such thatA's efficiency, as one of the cells firing B, is increased.

This postulate gained strong experimental support with the discovery of long-term potentiation (LTP) and depression (LTD) of synaptic transmission, first observed in the hippocampus (Bliss and Gardner-Medwin, 1973; Bliss and Lomo, 1973) and

subsequently reported in numerous other neural structures, including neocortex (A Artola, 1987; Iriki et al., 1989; Hirsch et al., 1992), the amygdala (Chapman et al., 1990; Clugnet and LeDoux, 1990), and midbrain reward circuits (Liu et al., 2005; Pu et al., 2006).

1.4 Spike-timing dependent plasticity

Spike-timing dependent plasticity (STDP) is a specific instantiation of Hebb's postulate that relates the amount of synaptic modification to the temporal order of action potentials (spikes) arriving at a synapse. STDP was initially suggested by experimental observations Levy and Steward (1983) that led to theoretical studies (Abbott and Blum, 1996; Gerstner et al., 1996). It was revealed in its full form by Markram et al. (1997) in cortical networks and at the same time by Bell et al. (1997) in a cerebellum-like structure of electric fish.

STDP has now been observed at excitaory synapses in a wide variety of neural circuits including visual cortex (Sjöström et al., 2001; Froemke and Dan, 2002), barrel cortex (Egger et al., 1999; Feldman, 2000), hippocampus (Bi and Poo, 1998; Li et al., 2004), dorsal cochlear nucleus (Tzounopoulos et al., 2004), the magnocellular nucleus of songbirds (Boettiger and Doupe, 2001) and the olfactory system of locusts (Cassenaer and Laurent, 2007). Studies of STDP on the one hand relate this mechanism to the synaptic molecular machinery such as the kinetics of N-methyl-D-aspartate (NMDA) receptors (Malenka and Bear, 2004) and the dynamics of intracellular calcium (Shouval et al., 2002), and on the other hand extend it toward behavioral time scales (Drew and Abbott, 2006) and perceptual phenom-

ena (Fu et al., 2002), spanning the three aforementioned levels of abstraction (see Caporale and Dan (2008) for a review).

The most salient feature of STDP is that synaptic modification depends on the temporal order of pairs of pre- and postsynaptic spikes. With the exception of the cerebellum-like structure in electric fish (Bell et al., 1997), all the experimental results mentioned above show that when a presynaptic spike precedes a postsynaptic one, the synapse is potentiated, and when the temporal order is reversed, the synapse is depressed. Therefore, STDP augments causal relations among neurons. These observations motivated pair-based models for STDP that relate synaptic modification to the intervals between pairs of pre- and postsynaptic spikes. This class of models has been shown to organize neural circuits functionally and explain many neuro-computational phenomena such as formation of receptive fields (Meliza and Dan, 2006), development of orientation tuning (Yao et al., 2004) and direction selectivity (Senn and Buchs, 2003) in visual cortex, formation of place cells in hippocampus (Mehta et al., 2002), and learning of temporal sequences (Masquelier et al. (2008); D'Souza et al. (2010), see Dan and Poo (2006) for a review).

More recent experimental evidence (Froemke and Dan, 2002; Froemke et al., 2006; Wang et al., 2005) indicates that the dependence of synaptic plasticity on the timing of spikes is more complex than a simple pair-based interaction. These observations motivated the development of multi-spike models of STDP, which augment the pair-based model to include interactions of multiple pre- and postsynaptic spikes (see Morrison et al. (2008) for a review). In this dissertation both classes of STDP models (pair-based and multi-spike) will be studied systematically, particu-

larly with respect to the important challenge described bellow.

1.5 The stability/competition dilemma

Every form of Hebbian plasticity faces the challenge of maintaining the balance between forming new memories through modification of synaptic strengths, and preserving old synaptic configurations to maintain old memories (Abraham and Robins, 2005). A more specific aspect of this challenge is the dilemma between synaptic competition and stability. Hebbian plasticity in general and STDP in particular induces a competition among synapses in a network (Miller, 1996). This is a powerful mechanism for shaping and modifying neural activity, but it can also be problematic for the normal functioning of a neural network. Unless changes in synaptic strength are coordinated appropriately, the level of activity in a neural circuit can grow or shrink in an uncontrolled way (Abbott and Nelson, 2000). One simple way to resolve this problem is to impose bounds on the allowed range of synaptic strengths. The details of implementing these bounds have significantly different consequences on the structures that arise from STDP at the synaptic population level (Song et al., 2000; van Rossum et al., 2000; Rubin et al., 2001; Gütig et al., 2003; Babadi and Abbott, 2010).

In the chapters that follow, I examine different STDP models through the lens of the stability/competition dilemma. For each model, I explore cases where stability and competition coexist, and cases where they are not compatible. I propose mechanisms to resolve this incompatibility. I also explore structures that arise from STDP in large recurrent neural networks, which are in turn consequences of intro-

ducing physiological bounds into synaptic strengths.

1.6 Outline of the dissertation

In chapter 2, I review the stability/competition interplay in a simple pair-based STDP model and establish the methodological basis for the rest of the dissertation. In chapter 3, I propose a modification of the pair-based STDP model that can reconcile stability with competition. In chapter 4, I focus on stability and competition in more complex multi-spike STDP models and systematically study them in different parameter regimes. Finally, in chapter 5, I study the structures that arise from STDP in a large recurrent network of spiking neurons.

To avoid interrupting the text with details of mathematical derivations as much as possible, I explain the intuitions behind the calculations and use formulas and equations only for describing the models and core concepts in the main text. All the detailed mathematical calculations are provided in the appendices at the end.

During my PhD research, I also took part in projects that were not systematically related to the main theme of this dissertation. Reports of those projects are included as appendices at the end of the thesis.

Chapter 2

Stability and Competition in Pair-based STDP models

Plasticity, then, in the wide sense of the word, means the possession of a structure weak enough to yield to an influence, but strong enough not to yield all at once. Each relatively stable phase of equilibrium in such a structure is marked by what we may call a new set of habits.
William James

2.1 Introduction

The simplest description of STDP relates synaptic modification to the temporal order of pairs of pre- and postsynaptic spikes: a synapse is potentiated when a presynaptic action potential precedes a postsynaptic spike, and depressed otherwise (Markram et al., 1997). This 'pair-based' model of STDP has been shown to induce a competitive form of Hebbian plasticity by forcing synapses to compete for control of the timing of postsynaptic action potentials while being strengthened or weakened (Song et al., 2000). Here, we analyze this model to determine how it affects populations of synapses converging onto a postsynaptic neuron. We focus on two basic features: stability and competition.

Stability is a property of the distribution of synaptic weights arising from an STDP model, and we will distinguish three cases: unstable, partially stable, and stable. In the unstable case, synaptic weights perpetually increase under STDP, unless some upper limit is imposed. As we will see, this happens in pair-based STDP when its potentiation component is larger than the depression component. When hard limits are imposed on an unstable STDP model, the synaptic weights cluster tightly against the upper bound. Another more interesting case is partial stability, in which individual synaptic weights increase or decrease indefinitely, but the average of the weights across a synaptic population stays fixed. We will show that in pair-based STDP this happens when the depression component is larger than the potentiation component. When hard bounds are imposed on a partially stable STDP model to limit the increases and decreases of individual synapses, the synaptic weights tend to cluster at either end of their allowed range, forming a U-shaped distribution (Song et al., 2000). Finally, when an STDP model is stable, synaptic weights form a unimodel distribution. In pair-based STDP this happens when soft bounds are introduced (van Rossum et al., 2000; Rubin et al., 2001).

The competition induced by STDP to the weights of synapses onto a postsynaptic neuron depends on correlations between their presynaptic spike patterns. This can be studied by dividing the inputs to a neuron into two groups, one with correlated presynaptic activity and the other with uncorrelated presynaptic spiking. In cases that we call "Hebbian", the synapses with correlated input become stronger than those with uncorrelated input. In other "anti-Hebbian" cases (which will be encountered in following chapters), the reverse occurs and the correlated synapses

become weaker than the uncorrelated. Here we are interested in the effect of introducing bounds to pair-based STDP, on Hebbian competition through this model.

In this chapter we review stability and competition in pair-based STDP models in more details to establish our approach and introduce ways of characterizing the effects of plasticity, to be used for analyzing more complex STDP models in the following chapters. We will show that it is not possible to maintain both stability and Hebbian competition in pair-based STDP by imposing either hard bounds or soft bounds to this model.

2.2 Methods

2.2.1 Neuronal and synaptic models

We used a leaky integrate-and-fire (LIF) model neuron in our numerical simulations. The membrane potential of the LIF neuron obeys

$$\tau_{\mathrm{m}} \frac{dV}{dt} = (V_{\mathrm{r}} - V) + I_{\mathrm{ex}} - I_{\mathrm{in}} \,, \tag{2.1}$$

where τ_{m} is the membrane time constant, V_{r} is the resting potential, I_{ex} is the excitatory input and I_{in} the inhibitory input. Although these inputs appear as currents, they are actually measured in units of the membrane potential (mV) because a factor of the membrane resistance has been absorbed into their definition. When the membrane potential V reaches the firing threshold V_{th} , the neuron fires an action potential and the membrane potential resets to the resting value V_{r}. The numerical values of all parameters are given Table 2.1.

Parameter	Symbol	Default value
Membrane time constant	τ_{m}	20 ms
Spiking threshold	V_{th}	-40 mv
Resting membrane potential	V_{r}	-60 mv
Synaptic time constant	τ_{s}	5 ms
Number of excitatory synapses	N_{ex}	1000
Number of inhibitory synapses	N_{in}	250
Inhibitory synaptic strength	w_{in}	1 mv
Excitatory input rate	r_{pre}	10 Hz
Inhibitory input rate	r_{in}	10 Hz
Correlation coefficient *	c	0.2

Table 2.1: Neuronal, synaptic, and plasticity parameters. * Correlation is only introduced for simulations in which synaptic competition is examined.

Each presynaptic action potential arriving at an excitatory or inhibitory synapse induces an instantaneous jump in the corresponding synaptic input (I_{ex} or I_{in}), which decays exponentially between the input action potentials. The time course of the synaptic inputs can thus be expressed as

$$I_{\mathrm{ex}} = \sum_{i=1}^{N_{\mathrm{ex}}} w_i \sum_{t_i^k \le t} \exp\left(\frac{t_i^k - t}{\tau_{\mathrm{s}}}\right) \quad \text{and} \quad I_{\mathrm{in}} = w_{\mathrm{in}} \sum_{i=1}^{N_{\mathrm{in}}} \sum_{t_i^k \le t} \exp\left(\frac{t_i^k - t}{\tau_{\mathrm{s}}}\right), \tag{2.2}$$

where w_i is the weight for excitatory synapse i with $i = 1, 2, \ldots, N$, w_{in} is the common fixed weight for all N_{in} inhibitory synapses, and t_i^k is the time of the k^{th} action potential at synapse i. The sums over presynaptic spike times are limited to spikes that arrive prior to the time t. The synaptic time constant $\tau_{\mathrm{s}} = 5\,\mathrm{ms}$ is taken to be the same for excitatory and inhibitory synapses. The excitatory synaptic strengths, labeled collectively as w, are modified by STDP.

If the rate of the excitatory and inhibitory inputs is r_{pre} and r_{in} respectively,

the average firing rate of the LIF neuron can be approximated as (Brunel and Sergi, 1998)

$$\bar{r}_{\mathrm{post}} = \left(\tau_{\mathrm{m}} \sqrt{\pi} \int_{\frac{V_{\mathrm{r}}-\mu}{\sigma}+\alpha}^{\frac{V_{th}-\mu}{\sigma}+\alpha} dx \, \exp(x^2)\big(1+\mathrm{erf}(x)\big) \right)^{-1}, \tag{2.3}$$

where

$$\mu = (N_{\mathrm{ex}}\, r_{\mathrm{pre}}\, \langle w \rangle - N_{\mathrm{in}}\, r_{\mathrm{in}}\, w_{\mathrm{in}})\, \tau_{\mathrm{s}} \quad \text{and} \quad \sigma^2 = \frac{(N_{\mathrm{ex}}\, r_{\mathrm{pre}}\, \langle w \rangle^2 + N_{\mathrm{in}}\, r_{\mathrm{in}}\, w_{\mathrm{in}}^2)\, \tau_{\mathrm{s}}^2}{\tau_{\mathrm{m}}},$$

with $\langle w \rangle$ denoting the average value of the excitatory synaptic weights. The parameter $\alpha = |\zeta(1/2)| \sqrt{\tau_{\mathrm{s}}/2\tau_{\mathrm{m}}}$, where ζ is the Riemann zeta function, is a correction to account for the nonzero synaptic decay constant. The arrival of a presynaptic spike increases the firing rate of the postsynaptic neuron transiently. For an LIF neuron in the case where the average excitatory input dominates over the inhibitory input, the firing rate after the arrival of a presynaptic spike at time t_0 can be approximated as (Herrmann and Gerstner (2001); see Appendix A.1)

$$r_{\mathrm{post}}(t) \approx \bar{r}_{\mathrm{post}} + \frac{w \exp\big(-(t-t_0)/\tau_{\mathrm{s}}\big)}{(V_{th}-V_{\mathrm{r}})\tau_{\mathrm{m}}}\, \Theta(t-t_0) \tag{2.4}$$

where w is the strength of the synapse through which the presynaptic spike arrived, and Θ is the Heaviside step function.

2.2.2 Correlated spike trains

To study synaptic competition, we introduce correlations into half of the excitatory input spike trains. To generate Poisson spike trains with homogeneous pairwise (zero-lag) correlations, a "generating" spike train with rate r/c was first produced. The correlated spike trains were then obtained by trimming the generating spike

train, that is, by randomly deleting spikes with probability $1 - c$. The resulting spike trains all have rate r, and each pair is correlated with correlation coefficient c (Kuhn et al., 2003)

.

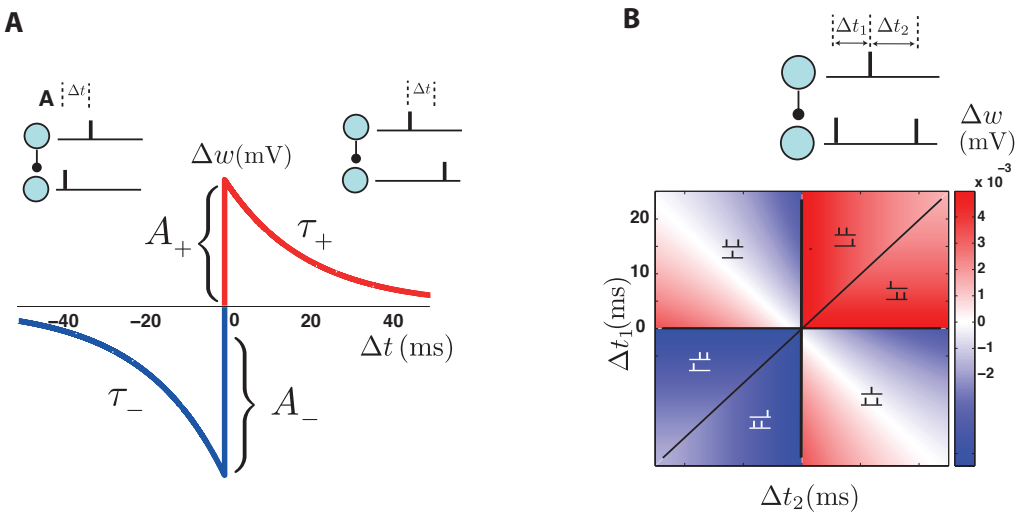

FIG. 2.1: Pair-based STDP. **A.** The parameters of pair-based STDP window. **B.** Top: a triplet of spikes composed of two pre-post pairs with intervals Δt_1 and Δt_2. Bottom: the amount of synaptic modification in response to triplets, which is symmetric in the pair-based model.

2.2.3 Pair-based STDP model

In pair-based STDP, a change of synaptic strength, Δw, is induced by a pair of pre- and postsynaptic action potentials with time difference (pairing interval) $\Delta t = t_{\text{post}} - t_{\text{pre}}$. The functional relation between the synaptic modification and the

pairing interval is

$$\Delta w = F(\Delta t) = \begin{cases} A_+ \exp(-\Delta t/\tau_+) & \text{if } \Delta t \geq 0 \\ \\ -A_- \exp(\Delta t/\tau_-) & \text{if } \Delta t < 0\,. \end{cases} \tag{2.5}$$

The positive parameters A_+ and A_- specify the maximum potentiation and depression, respectively. We express the synaptic strengths in units of the membrane potential (mV), so A_+ and A_- have mV units. The time constants τ_+ and τ_- determine the temporal spread of the STDP window for potentiation and depression (figure 2.1A). The synaptic strengths are maintained in their allowed range ($0 < w < w_{max}$) either by imposing hard bounds or soft bounds. Hard bounds are imposed by truncating any modification that would take a synaptic weight outside the allowed range. Soft bounds are imposed by making the maximum depression and potentiation weight-dependent so that when a synaptic strength approaches the bounds, its rate of change gradually decreases. This will be done by multiplying A_+ and A_- by $1 - (w/w_{max})$ and w/w_{max} respectively. In our analysis, we assume that the spike pairings are all-to-all, meaning that all possible pre-post pairs, not only the nearest neighbor pairs, contribute to plasticity. However, the following results apply qualitatively to a pair-based model with a nearest-neighbor restriction as well.

2.3 Results

In this chapter and also the next two chapters, to study the effects of different STDP models on synaptic weights, we simulated a single spiking neuron receiving

N_{ex} excitatory and N_{in} inhibitory presynaptic spike trains with Poisson statistics at rates r_{pre} and r_{in}, respectively. The strengths of the excitatory synapses, denoted by w, change according to STDP, while the strengths of inhibitory synapses remain constant. To examine the different forms of stability of each STDP model, we check whether the steady-state distribution of synaptic strengths is bounded without imposing external limits, or whether the increase or decrease of the weights is stopped only when they hit a boundary. We distinguish between the partially stable and unstable cases by computing the evolution of the average of the synaptic weights, which reaches a fixed point only in the partially stable case. Fully stable STDP is characterized by a fixed point for the average rate and bounded deviations for the strengths of individual synapses about this mean. As a probe of synaptic competition, we induce correlations in half of the excitatory inputs (Methods) and check whether STDP causes the synapses corresponding to correlated and uncorrelated subsets to compete for control of the postsynaptic firing. This also allows us to determine whether the effect of correlations is Hebbian or anti-Hebbian. The neuronal and input parameters used in our simulations are given in Table 2.1.

2.3.1 Stability and competition in pair-based STDP with hard bounds

To explain our method for analyzing synaptic stability and competition and also to provide a benchmark of comparison for the following chapters, we first examine a pair-based STDP model. In this model, synapses are modified only on the basis of the intervals between pairs of pre- and postsynaptic spikes in the pair-based model. When a synapse receives a larger ensembles of spikes, such as triplets or quadruplets,

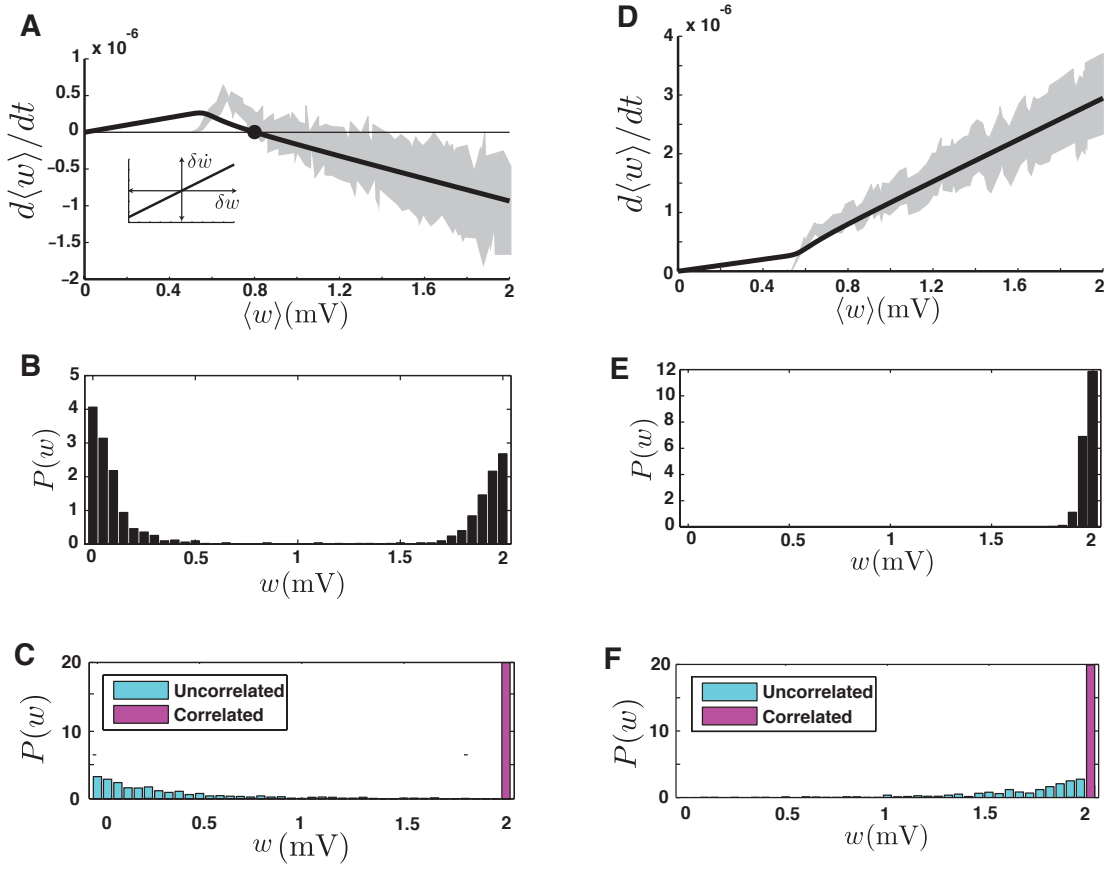

FIG. 2.2: Stability and competition in pair-based STDP with hard bounds. **A.** The average drift induced by the pair-based model on a population of excitatory synapses converging onto a single postsynaptic neuron, when $A_+ < A_-$. The black curve is a numerical evaluation of equation (2.7) and the gray area is the simulation results. The half-width of the gray area is the standard error. The filled circle is the stable fixed point. The inset shows the w-dependent drift (equation 2.8). **B.** The steady state distribution of synaptic weights obtained by simulation when $A_+ < A_-$. **C.** The steady state distribution of weights when half of the synapses receive correlated input (magenta) and the other half receive uncorrelated input (cyan). When $A_+ < A_-$ correlated synapses are strengthened. **D-F.** the same as A-C, but for $A_+ > A_-$. Note that there is no stable fixed point in **D**, and that all the synapses are pushed to the upper bound in **E** and **F**. For these simulations, the constants of the STDP model were $\tau_+ = \tau_- = 20$ ms, and $A_+ = 0.005$ mV and $A_- = 1.01A_+$ in A-C and $A_- = 0.005$ mV and $A_+ = 1.01A_-$ in D-F

plasticity is induced by the pre-post pairs within the ensemble independent of the higher-order structure of the ensemble. For instance, similarly spaced "pre-post-pre" and "post-pre-post" triplets induce the same amount of synaptic modification in this model (figure 2.1B). As we will see in Chapter 4, the response of multi-spike STDP models to large ensemble of spikes is quite different.

To quantify the average modification of a synapse under STDP, we first calculate the probability of a pairing interval Δt for spikes arriving at the synapse and then average equation (2.5) under that probability. We assume that the pre- and postsynaptic spike trains are both Poisson. The rate of the presynaptic spike train takes the constant value $r_{\rm pre}$. The baseline rate of postsynaptic firing is denoted by $\bar{r}_{\rm post}$ (equation 2.3). When a postsynaptic action potential is generated, presynaptic spikes are equally likely to arrive at any later time because the postsynaptic spike has no effect on presynaptic activity. However, when a presynaptic spike arrives at a particular synapse, it transiently increases the postsynaptic firing rate by an amount proportional to the strength of that synapse (equation 2.4). As a result, a postsynaptic action potential is more likely to be induced shortly after the arrival of a presynaptic spike. Including both the baseline rate and this brief enhancement, the average synaptic modification or the "drift" for a synaptic strength w is (see Methods)

$$\frac{dw}{dt} = (A_+\tau_+ - A_-\tau_-)\, r_{\rm pre}\, \bar{r}_{\rm post} + \frac{A_+\tau_+\tau_{\rm s}\, r_{\rm pre}\, w}{(\tau_{\rm s}+\tau_+)(V_{th}-V_{\rm r})\tau_{\rm m}}\,. \tag{2.6}$$

The first term in this equation relates the change in synaptic strength of a particular synapse, w, to the average strength of all the excitatory synapses, $\langle w\rangle$, through the

dependence of the baseline firing rate $\bar{r}_{\text{post}}$ on this average. This term is the same for all synapses, so we call it the "baseline drift". The second term depends on the synaptic strength of the particular synapse being considered, and it arises from the transient increase of postsynaptic firing rate following a presynaptic spike at this synapse. We call it the "w-dependent drift". The rate of change of the average of all the excitatory synaptic weights is given by the sum of the baseline drift and the average of the w-dependent drift,

$$\frac{d\langle w\rangle}{dt} = (A_{+}\tau_{+} - A_{-}\tau_{-})\, r_{\text{pre}}\, \bar{r}_{\text{post}} + \frac{A_{+}\tau_{+}\tau_{\text{s}}\, r_{\text{pre}}\, \langle w\rangle}{(\tau_{\text{s}} + \tau_{+})(V_{th} - V_{\text{r}})\tau_{\text{m}}}\,. \tag{2.7}$$

The average synaptic strength in the steady-state is the values of $\langle w\rangle$ that sets the right side of equation (2.7) to zero (i.e. a fixed point). One such a fixed point occurs when all the synapses are zero ($\langle w\rangle = 0$). This makes the postsynaptic neuron silent ($\bar{r}_{\text{post}} = 0$) and sets both the baseline and average w-dependent drifts to zero. This state is uninteresting and simply reflects the fact that no plasticity occurs when the postsynaptic neuron is silent. If the synaptic strengths are not zero, the average w-dependent drift is always positive because presynaptic spikes always enhance postsynaptic firing. As a result, a nontrivial fixed point for the average synaptic weight can occur only if the baseline drift is negative ($A_{-}\tau_{-} > A_{+}\tau_{+}$) so that it can cancel the w-dependent drift (figure 2.2A, closed circle). This fixed point is stable, because the positive w-dependent drift dominates if the average weight is smaller that the fixed-point value, and the negative baseline drift dominates if it is larger. Mathematically, stability requires the slope of the average drift to be negative at the fixed point (figure 2.2A), which always holds for the nontrivial fixed point of the pair-based model. In summary, the steady-state average synaptic

strength in pair-based STDP has a stable nontrivial mean only if the depression window is larger than the potentiation window ($A_{-}\tau_{-} > A_{+}\tau_{+}$). This fixed point is unique, so the mean of the steady-state distribution of synaptic weights converges to this value regardless its initial value.

The stability of the mean is not a sufficient condition for the steady-state distribution of synaptic strengths to be fully stable, each synapse must also have a stable deviation from the mean. The strength of a particular synapse can be expressed as $w = \langle w \rangle + \delta w$, where δw is the deviation of the synapse from the mean. If the deviation tends to grow over time, the synapses will drift away from the mean and the distribution will be partially stable and U-shaped (bimodal). If the deviation tends to decrease, the synapses will cluster around the mean and the distribution will be stable and unimodal. Assuming that the mean synaptic strength is at steady-state and that the deviation of an individual synapse (out of a few thousand) does not alter the mean significantly, the change of the deviation over time is governed solely by the w-dependent drift and can be derived from equation (2.6) as

$$\frac{d\delta w}{dt} = \frac{A_{+}\tau_{+}\tau_{\mathrm{s}}\, r_{\mathrm{pre}}\, \delta w}{(\tau_{\mathrm{s}} + \tau_{+})(V_{th} - V_{\mathrm{r}})\tau_{\mathrm{m}}}\,. \tag{2.8}$$

Because the coefficient of δw in this equation is positive (figure 2.2A, inset), the deviations tends to grow, and the final distribution of synaptic strengths for pair-based STDP is partially stable and U-shaped even though the mean is stable (figure 2.2B, Song et al. (2000)).

To check the accuracy of equations (2.7) and (2.8), we computed the synaptic

drift by averaging the amount of induced synaptic modification in simulations lasting 10^6 ms of simulated time, without implementing synaptic modification (figure 2.2A, gray shade). The discrepancy between the analytic and simulation results at low average synaptic strengths is due to the fact that our approximation for the transient postsynaptic firing rate (equation 2.4) is only accurate when the mean excitatory input is significantly larger than the mean inhibitory input. In the parameter regime where the potentiation window is larger than the depression window (figure 2.2D-F), the mean synaptic weight only has the trivial zero fixed point (figure 2.2D), so the plasticity is unstable and all of the synaptic strengths grow until they hit the upper bound, regardless of their initial values (figure 2.2E).

When the mean synaptic strength is stable and the w-dependent drift is positive, it is possible for STDP to discriminate between two groups of synapses based on the degree of correlation in their presynaptic spike trains. If the spike trains arriving at one group of synapses are correlated and those of the other synapses are not, the correlated group induces a larger transient increase in the postsynaptic firing rate and hence a larger w-dependent drift. Therefore the correlated group is more likely to become stronger than the mean, and the uncorrelated group tends to become weaker to maintain the balance around the mean (figure 2.2C). This results in a Hebbian competition among the synapses (Song et al., 2000). On the other hand, when there is no stable mean, all the synapses tend to grow regardless of their correlation and no competition takes place, although the correlated synapses still end up stronger than the uncorrelated group (figure 2.2F). Therefore, the condition for Hebbian competition through pair-based STDP is the existence of a stable mean, i.e. $A_{-}\tau_{-} > A_{+}\tau_{+}$, which is equivalent to partial stability.

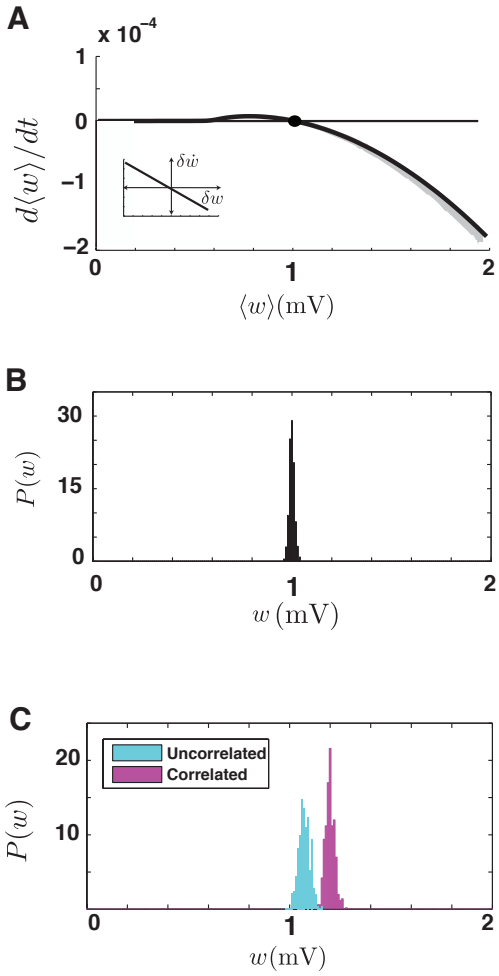

FIG. 2.3: Stability and competition in pair-based STDP with soft bounds. **A.** The average drift induced by the pair-based model on a population of excitatory synapses converging onto a single postsynaptic neuron, when soft bounds are imposed. The black curve is a numerical evaluation of equation (2.9) and the gray area is the simulation results. The half-width of the gray area is the standard error. The filled circle is the stable fixed point. The inset shows the growth rate of deviations from the mean (equation 2.10). **B.** The steady state distribution of synaptic weights obtained by simulation, which is stable. **C.** The steady state distribution of weights when half of the synapses receive correlated input (magenta) and the other half receive uncorrelated input (cyan). Competition is very weak. In these simulations, $A_+ = A_- = 0.005$.

2.3.2 Stability and competition in pair-based STDP with soft bounds

When soft bounds are imposed, A_+ and A_- will be multiplied by $1-(w/w_{max})$ and w/w_{max} respectively. This changes the baseline drift (equation (2.7)) to

$$\begin{aligned}\frac{d\langle w\rangle}{dt} &= A_+\tau_+\, r_{\mathrm{pre}}\,\bar{r}_{\mathrm{post}} \\ &\quad - \left[\left(A_+\tau_+ + A_-\tau_-\right) r_{\mathrm{pre}}\,\bar{r}_{\mathrm{post}} - \frac{A_+\tau_+\tau_{\mathrm{s}}\, r_{\mathrm{pre}}\, w_{max}}{(\tau_{\mathrm{s}}+\tau_+)(V_{th}-V_{\mathrm{r}})\tau_{\mathrm{m}}}\right]\frac{\langle w\rangle}{w_{max}} \\ &\quad - \frac{A_+\tau_+\tau_{\mathrm{s}}\, r_{\mathrm{pre}}}{(\tau_{\mathrm{s}}+\tau_+)(V_{th}-V_{\mathrm{r}})\tau_{\mathrm{m}}}\,\frac{\langle w\rangle^2}{w_{max}}\end{aligned} \tag{2.9}$$

and the change of the deviation (equation (2.8)) to

$$\frac{d\delta w}{dt} = \left[\left(A_+\tau_+ + A_-\tau_-\right) r_{\mathrm{pre}}\,\bar{r}_{\mathrm{post}} - \frac{A_+\tau_+\tau_{\mathrm{s}}\, r_{\mathrm{pre}}\left(2\langle w\rangle - w_{max}\right)}{(\tau_{\mathrm{s}}+\tau_+)(V_{th}-V_{\mathrm{r}})\tau_{\mathrm{m}}}\right]\frac{\delta w}{w_{max}}\,. \tag{2.10}$$

Equation (2.7) has always a single nontrivial stable fixed point, regardless of the values of STDP parameters (figure 2.3A). Calculating the fixed point from equation (2.7) and substituting in (2.10) shows that the coefficient of δw is always negative. Therefore, deviations tend to shrink and all synaptic weights tend to cluster around the mean and form a unimodal stable distribution (figure 2.3B).

The tendency of synaptic weights to cluster around the mean weakens synaptic competition considerably (figure 2.3C). Correlated synapses cannot become significantly stronger than uncorrelated ones because any modification toward higher

values of weights will be counteracted by a decrease of potentiation and an increase of depression. Likewise, any modification toward lower values will be counteracted by a decrease of depression and an increase of potentiation. Therefore, imposing soft bounds on STDP stabilizes the distribution of weights, but with the expense of losing synaptic competition.

2.4 Discussion

We reaffirmed the results of other studies which showed that if the depression window is larger than the potentiation window, pair-based STDP with hard bounds is partially stable and can induce Hebbian competition between correlated and uncorrelated synapses (Song et al., 2000; Song and Abbott, 2001; Cateau and Fukai, 2003). If the potentiation window is larger than the depression window, pair-based STDP with hard bounds is unstable and no competition takes place. Pair-based STDP with soft bounds is always stable regardless of the balance between potentiation and depression, but with the functionally important cost of losing synaptic competition, as shown by Rubin et al. (2001) and van Rossum et al. (2000) before. Soft bounds render STDP to a homeostatic plasticity mechanisms with negligible sensitivity to the correlation structure of the input.

As a conclusion, synaptic stability and competition cannot be simultaneously maintained in pair-based STDP model with either hard or soft bounds. Gütig et al. (2003) proposed interpolating between hard and soft bounds as a mechanism to obtain both synaptic competition and stability, but this approach is successful

only over a limited parameter range. In the next chapter, we propose our own alternative biophysically inspired way to introduce bounds, which can successfully maintain stability and competition over a wide range of parameters, in addition of having a number of other favorable computational properties.

Chapter 3

STDP with a Shifted Temporal Window

Research under a paradigm must be a particularly effective way of inducing paradigm shift.
Thomas S. Kuhn

3.1 Introductions

Synaptic competition and synaptic stability are desirable but often conflicting features of Hebbian synaptic plasticity. As we have seen in previous chapter, the instability of pair-based STDP can be eliminated by introducing strength-dependent modification (soft bounds; van Rossum et al. 2000; Rubin et al. 2001), but at the expense of eliminating synaptic competition. By interpolating between stable and unstable models of STDP, it is possible to obtain both synaptic competition and stability, but over a limited parameter range (Gütig et al., 2003). Here we propose an alternative solution inspired by the slow kinetics of NMDA receptors. We show that STDP can be stabilized if the boundary separating potentiation and depression does not occur for simultaneous pre- and postsynaptic spikes, but rather for

spikes separated by a small time interval. Through simulation as well as by solving the Fokker-Planck equation governing the distribution of synaptic strengths, we show that any positive shift of the STDP window can stabilize the distribution of synaptic strengths while preserving synaptic competition. Moreover, our simulations show that even a random symmetric jitter of a few milliseconds in the STDP window can also stabilize synaptic strengths while retaining these features. The results of this chapter have been published in Babadi and Abbott (2010).

3.2 Methods

The neural and synaptic models used in this chapter are similar to those introduced in the previous chapter. Most of the parameters also remain unchanged. The numerical values of the parameters are given in Table 3.1.

Parameter	**Symbol**	**Default value**
Membrane time constant	τ_{m}	$20\,ms$
Spiking threshold	V_{th}	$-40\,mV$
Resting membrane potential	V_r	$-60\,mV$
Maximum potentiation amplitude	A_+	$0.006\,mV$
Maximum depression amplitude	A_-	$0.005\,mV$
Potentiation time constant	τ_+	$20\,ms$
Depression time constant	τ_-	$20\,ms$
Window shift	d	$2\,ms$
Synaptic time constant	τ_{s}	$5\,ms$
Number of excitatory synapses	N_{ex}	1000
Number of inhibitory synapses	N_{in}	250
Inhibitory synaptic strength	w_{in}	$4\,mV$
Excitatory input rate	r_{ex}	$10\,Hz$
Inhibitory input rate	r_{in}	$10\,Hz$

Table 3.1: Neuronal, synaptic, and plasticity parameters

3.2.1 Shifted STDP model

In the shifted pair-based model the change in synaptic strength, Δw, induced by a pair of pre- and postsynaptic action potentials with time difference $\Delta t = t_{\text{post}} - t_{\text{pre}}$ is determined by

$$\Delta w = F(\Delta t) = \begin{cases} -A_{-}\, e^{(\Delta t - d)/\tau_{-}} & \text{if } \Delta t \leq d \\ \\ A_{+}\, e^{-(\Delta t - d)/\tau_{+}} & \text{if } \Delta t > d\,. \end{cases} \tag{3.1}$$

As before, the parameters A_{+} and A_{-}, both positive, determine the maximum amount of synaptic potentiation and depression, respectively. The time constants τ_{+} and τ_{-} determine the temporal extent of the STDP window for potentiation and depression. The parameter d, also positive, introduces a shift in the STDP window such that even in cases where a presynaptic action potential precedes the postsynaptic spike by a short interval ($0 < \Delta t < d$), the corresponding synapse gets depressed. Note that we recover conventional pair-based STDP by setting $d = 0$. Numerical values of the STDP parameters are given in Table 3.1. An important feature of the pair-based model we use in this chapter is that STDP arises solely from pairs of pre- and postsynaptic spikes that are nearest neighbors in time, in agreement with experimental results (Sjöström et al., 2001). Specifically, each postsynaptic action potential can only potentiate a synapses on the basis of the interval to the presynaptic spike immediately preceding it, and each presynaptic action potential can only depress a synapses on the basis of the timing interval to the immediately preceding postsynaptic spike. This assumption is important for the results we obtain using the pair-based STDP model, as discussed below.

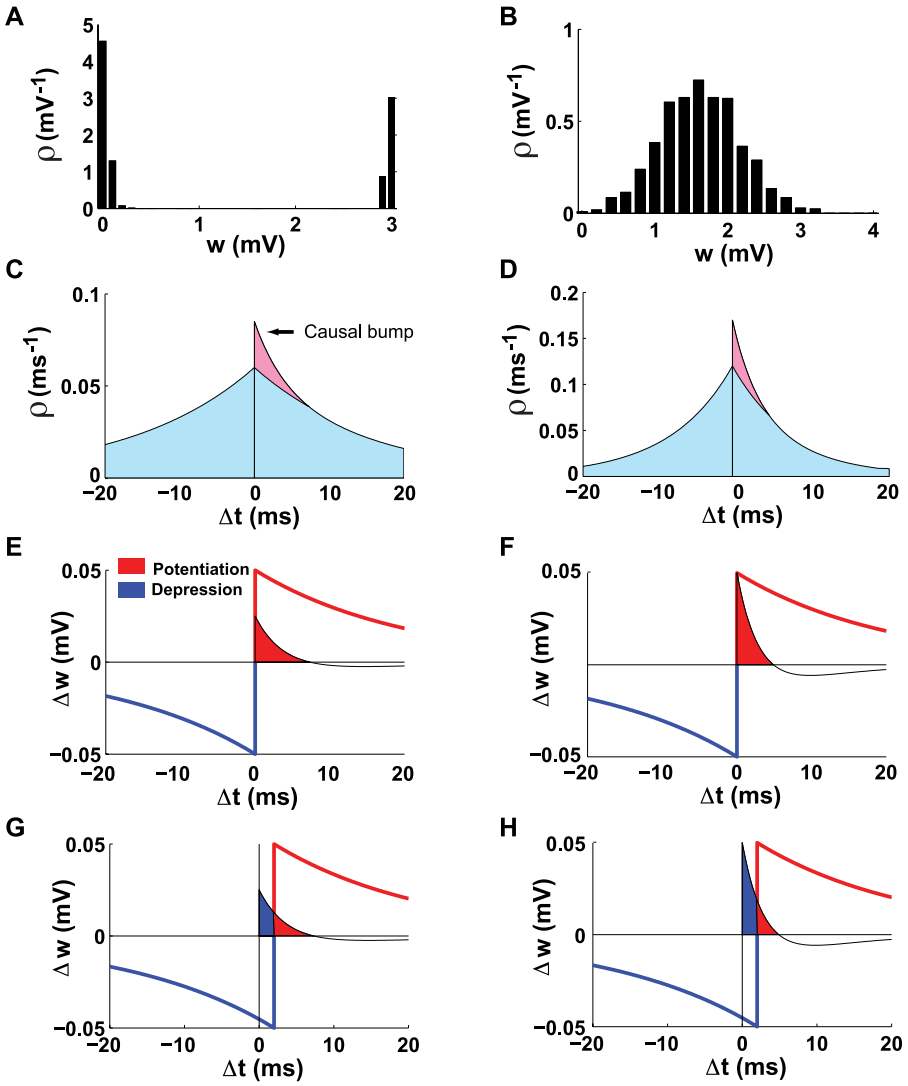

FIG. 3.1: Comparison of unshifted and shifted STDP. **A.** The U-shaped steady-state distribution of synaptic strengths for conventional unshifted STDP. **B.** The unimodal steady-state distribution of synaptic strengths for shifted STDP (d=2ms). **C.** The probability density of pairing intervals for presynaptic and post-synaptic spike trains. The blue area is the symmetric acausal contribution, and the pink area is the additional causal bump arising from postsynaptic spikes induced by the presynaptic input. **D.** Same as C, but for a stronger synapse. The causal bump is larger and closer to $\Delta t = 0$. **E.** The causal bump superimposed on the unshifted STDP window. The potentiation part of the STDP curve is red and the depression part is blue. The causal bump falls entirely within the potentiation domain (red shading). **F.** Same as E, but for a stronger synapse. The causal bump still falls within the potentiation region. **G.** Same as E, but for shifted STDP. Part of the causal bump falls into the depression region (blue shading). **H.** Same as G, but for a stronger synapse. More of the causal bump falls into the depression region.

3.3 Results

3.3.1 Stability of synaptic strengths

With conventional, unshifted STDP ($d = 0$), synaptic strengths grow or shrink indefinitely unless limits are imposed. These limits produce a U-shaped distribution of synaptic strengths (figure 3.1A, Song et al. (2000)). However, if we introduce a $d = 2ms$ shift into the STDP window, the steady-state distribution of synaptic strengths is unimodal and stable even when no limits are imposed (figure 3.1B). Why does this occur?

The total effect of a sequence of pre- and postsynaptic action potentials on the strength of a synapse can be computed by multiplying the STDP window function by the probability of a spike pair appearing with time difference Δt and then integrating over all values of Δt. If we assume Poisson spike trains and ignore the effects of the synapse, the probability distribution of nearest-neighbor pre-post pairs is an exponentially decaying function of the magnitude of the interval between them (figure 3.1C). The decay rate of this exponential is equal to the sum of the pre- and postsynaptic firing rates (Appendix A.2). The presence of a synapse induces an additional contribution to this distribution for small positive Δt arising from postsynaptic spikes induced by the synaptic input (figure 3.1C). The size of this "causal bump" is proportional to the probability of a presynaptic action potential evoking a postsynaptic response, and hence to the strength of the synapse. The stronger the synapse, the larger the bump. In addition, because the postsynaptic spike latency is shorter for stronger synapses, the bump moves closer to $\Delta t = 0$ as the synaptic strength increases (figure 3.1D). These features of the pre-post interval

distribution are crucial for our analyses.

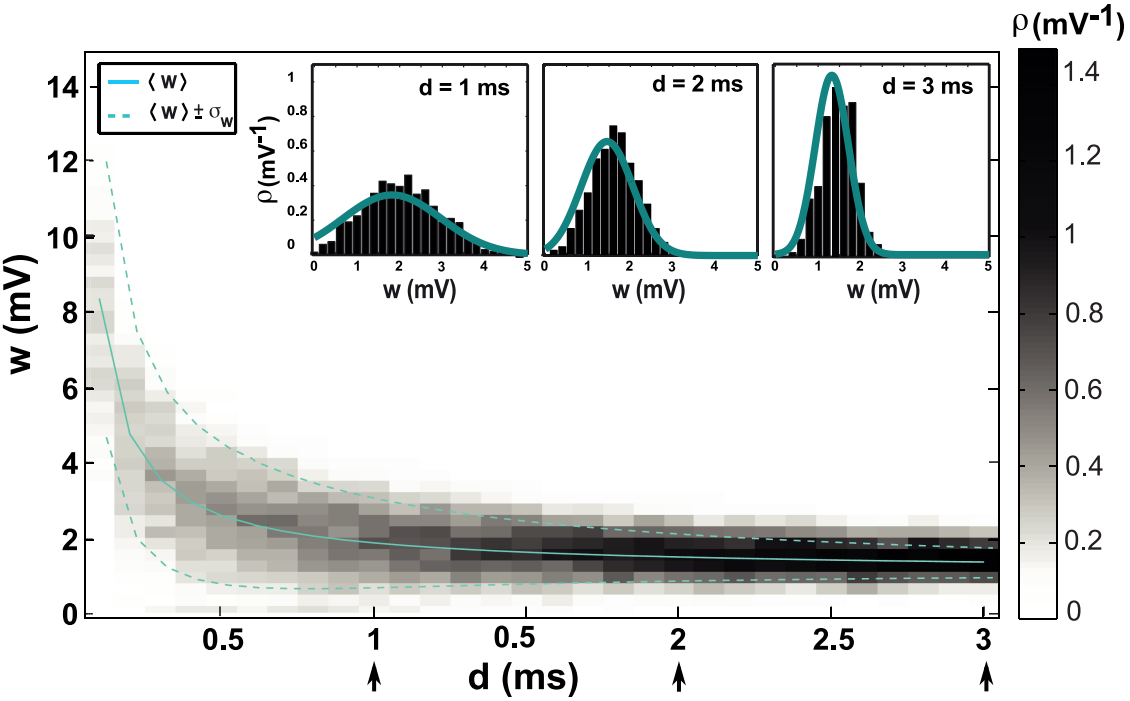

FIG. 3.2: Shifted STDP stabilizes the distribution of synaptic strengths. The horizontal axis is the value of the shift, the vertical axis is the synaptic strength and the gray level is the probability density of strengths, obtained by simulation. Solid line is the analytically calculated mean and dashed lines show the analytically calculated standard deviation around the mean. Insets show the distribution of synaptic strengths for different values of the shift. Solid curves are analytically calculated distributions. The arrows at the bottom of the horizontal axis of the main plot show the shift values corresponding to the insets.

When there is no shift in the STDP window, the causal bump falls entirely within the potentiation domain (figure 3.1E), which is why synaptic strengths grow until something else stops them (figure 3.1F). When the STDP window is shifted, part of the causal bump falls into the region where depression occurs (figure 3.1G). Furthermore as the synapse gets stronger, a larger portion of the causal bump falls into the depression domain, both because the causal bump gets bigger and because it moves closer to $\Delta t = 0$ (figure 3.1H). This prevents further growth of the synaptic

strength and explains why a shift stabilizes synaptic growth through STDP. Stabilization of synaptic weights occurs for any positive value of the delay (d), but larger delays result in lower mean values and sharper distributions for the weights (figure 3.2).

For a more quantitative evaluation of shifted STDP, we computed the steady-state solution of the Fokker-Planck equation governing the distribution of synaptic strengths (Risken (1996); Kempter et al. (2001); Cateau and Fukai (2003), Appendix A.2). With a few reasonable approximations and ignoring any limits or bounds, the steady-state distribution of synaptic strengths has the form of a gamma distribution,

$$\rho(w) = N_0(w+\mu)^{k-1}\exp\left(-\frac{w+\mu}{\theta}\right), \tag{3.2}$$

where N_0 is a normalization constant and μ, θ and k are computed parameters. If either k or θ is negative, this distribution cannot be normalized, implying unstable synaptic strengths. The calculations indicate that θ is positive for any positive shift ($d > 0$, Appendix A.2). Positivity of k requires that $A_+\tau_+ > A_-\tau_-$. Note that this is opposite to the condition required for partial stability of conventional, unshifted STDP (Chapter 2, Song et al. (2000)). Because it is easier to do the analytic calculations without imposing strict boundary conditions on the synaptic strengths, the analytic formula sometimes includes a small probability for negative strength synapses, which is not allowed in the simulations. Other than this small discrepancy, the agreement between the analytic distribution and the simulation results is good (figures 3.2 & 3.3). In what follows, $d = 2ms$, $A_+ = 0.006mV$, $A_- = 0.005mV$, and $\tau_+ = \tau_- = 20ms$, unless stated otherwise.

3.3.2 Steady-state firing rate

STDP has an interesting regulatory effect on the steady-state firing rate of a neuron (Song et al., 2000; Kempter et al., 2001). With unshifted STDP, this is a buffering effect making the steady-state postsynaptic firing rate relatively insensitive to the firing rates of excitatory and inhibitory inputs. Shifted STDP also buffers the postsynaptic firing rate, but the residual dependence on the presynaptic rates displays an interesting effect. Although the steady-state firing rate decreases when the inhibitory input rates are increased, it has a surprising non-monotonic dependence on the rates of excitatory inputs (figure 3.3).

The stabilization of synaptic strengths discussed in the previous section arises from the change of size and shape of the causal bump seen in figures 3.1C & D. Buffering of the steady-state postsynaptic firing rate is affected primarily by the shape of the symmetric, non-causal component of the spike-timing probability. As mentioned previously, this component falls off exponentially, for either positive or negative spike-timing differences, at a rate given by the sum of the presynaptic and postsynaptic firing rates (Appendix A.2). If this sum grows, the acausal part of the distribution gets more peaked near zero, bringing more spike pairs into the region of the STDP window where the shift leads to synaptic depression. The resulting reduction in synaptic strength then lowers the postsynaptic firing rate. This form of buffering would not be present if all spike pairs, rather than only nearest-neighbor pairs, were involved in STDP. If we allowed all spike pairs to induce synaptic plasticity the relevant symmetric, non-causal distribution would be flat, rather than exponentially decaying. In this case, there is no analogous stabilization and, in fact, postsynaptic rates slowly rise, making the plasticity unstable, even with shifted

STDP. This is why we require shifted STDP to be based only on nearest-neighbor spike pairs.

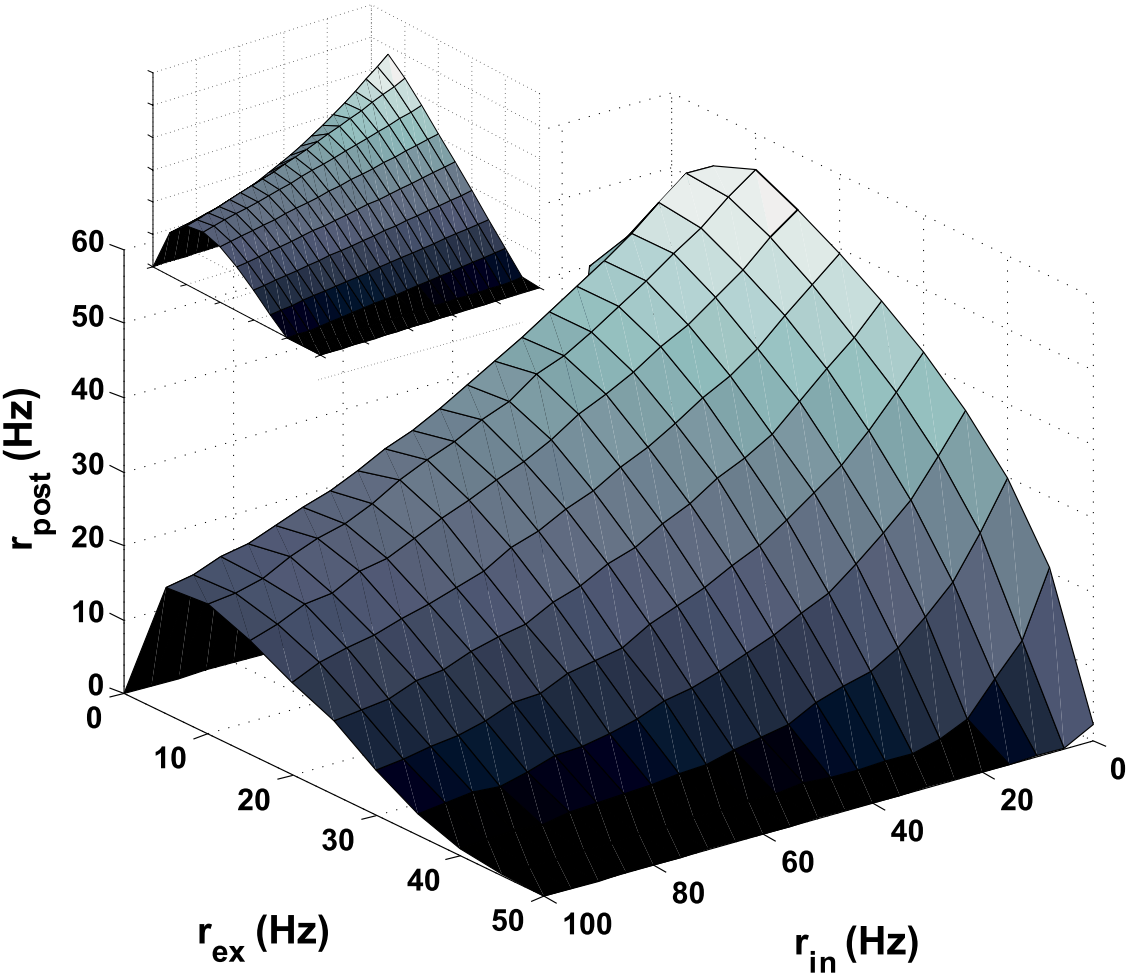

FIG. 3.3: The steady-state postsynaptic firing rate. The steady state firing rate is plotted as a function of the input rates for excitation and inhibition. The inset shows the corresponding analytic result.

In general, we expect the firing rate of a neuron to increase when its excitatory inputs fire more rapidly, and this is exactly what occurs for excitatory input rates below about 10 Hz in figure 3.3. However, for excitatory input rates higher than this, the steady-state (after STDP has equilibrated) postsynaptic firing rate de-

creases. This occurs for the reason outlined in the previous paragraph. Increasing the presynaptic rate causes the acausal distribution to sharpen and induces synaptic depression. This slows the postsynaptic rate, broadening the acausal distribution until the spike intervals in the delay region are sufficiently reduced in number. This is what causes the steady-state postsynaptic firing rate to drop when the excitatory presynaptic rates are raised to high levels.

Shifted STDP also has a buffering property on changes in the inhibitory input rate. In presence of strong inhibitory input, the postsynaptic firing rate falls. This broadens the acausal part of the spike-pair distribution, lowering the chance for pairs to fall into the depression domain caused by the shift and, thus, resulting in more potentiation. However, in this case, the effect is not strong enough to overcome the expected tendency of the postsynaptic rate to be suppressed by inhibition (figure 3.3).

3.3.3 Synaptic competition

Hebbian plasticity in general and STDP in particular allows neurons to become selective to correlated subsets of their inputs, but this requires synaptic competition (Miller, 1996). As in the previous chapter, we call synaptic plasticity "competitive" if correlating a subset of synaptic inputs causes both that set and the remaining synapses to change their strengths in an opposing manner, so than either the correlated or the uncorrelated set of synapses gains control of the postsynaptic firing (see for example Gütig et al. (2003)). In particular, if STDP is competitive, the strengths of either the correlated or uncorrelated subgroup of synapses should cluster near zero. To determine whether the necessary competition exists with shifted

STDP, we imposed pairwise correlations with a coefficient of 0.2 on one half of the incoming excitatory spike trains while leaving the other half uncorrelated. With unshifted STDP, this arrangement induces a competition that correlated synapses always win (Song et al., 2000). In other words, the synapses receiving correlated input become stronger and those receiving uncorrelated input get weaker.

Interestingly, with shifted STDP the outcome of the competition depends on the rate of inhibitory input to the neuron. When the rate of inhibitory input is 10 Hz for the parameters we use, the synapses receiving correlated spikes end up weaker than the synapses receiving uncorrelated spikes (figure 3.4A). This behavior is "anti-Hebbian" in that it is opposite to what is expected from normal Hebbian modification. However, when the rate of the inhibitory inputs is increased to 20 Hz, we obtain the usual Hebbian result in which correlated synapses win the competition and become stronger than uncorrelated synapses (figure 3.4B). Results obtained over a range of inhibitory input rates show a transition from anti-Hebbian to Hebbian modification (figure 3.4C). Choosing other values for the correlation coefficient within a range from 0.1 to 0.9 yielded qualitatively similar results. Competition also occurs between two correlated subgroups with different correlation coefficients, with the more correlated synapses dominating over the less correlated ones in the Hebbian (high inhibition) case and vice versa in the anti-Hebbian (low inhibition) mode. If the correlation coefficients for the two groups are the same, no competition takes place.

These results were obtained using spike trains with zero time-lag correlations, meaning that for any two correlated spike trains, a subset of spikes is perfectly synchronous. More realistic spike correlations can be generated by including a

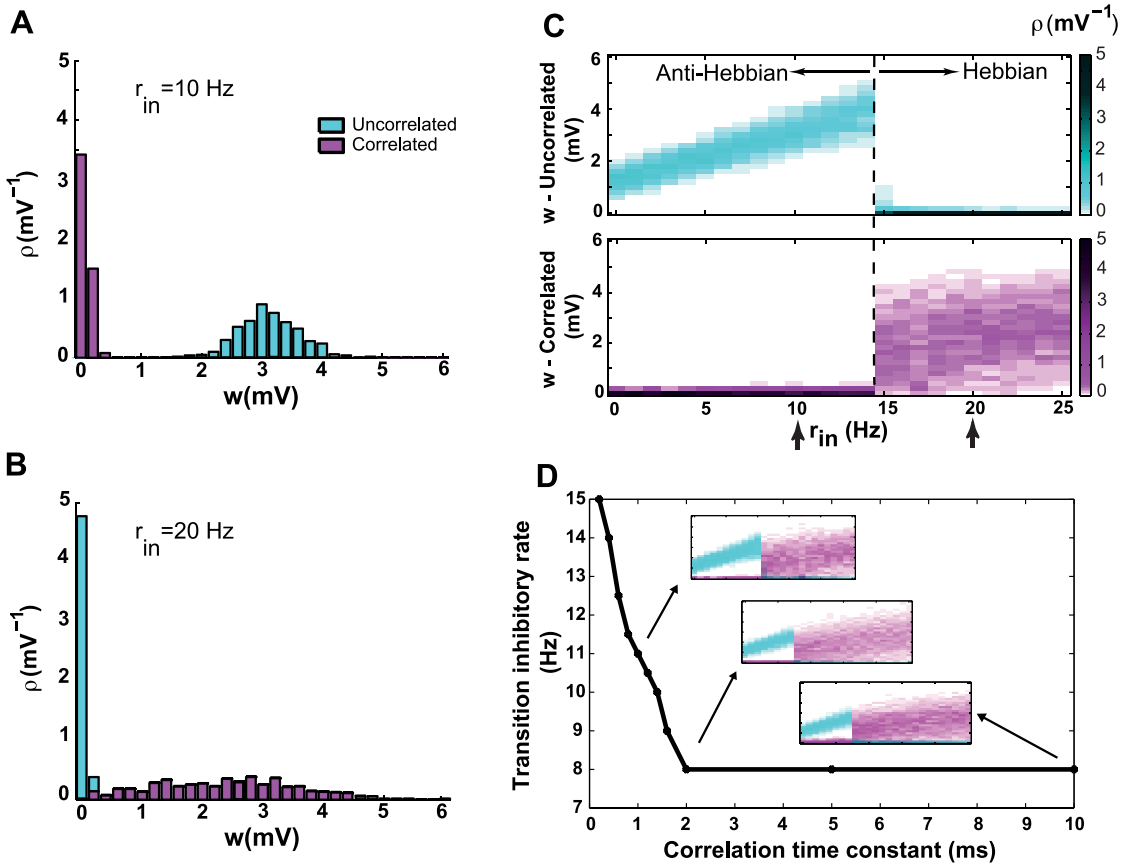

FIG. 3.4: Synaptic competition through shifted STDP. Cyan color indicates synapses with uncorrelated inputs, and magenta indicates correlated inputs. The rate of excitatory input is fixed at 10 Hz, and the correlation coefficient is 0.2 for correlated input spike trains. **A.** Steady-state distribution of synaptic strengths for an inhibitory rate of 10 Hz. Uncorrelated synapses become stronger than correlated. **B.** Steady-state distribution of synaptic strengths for an inhibitory rate of 20 Hz. Correlated synapses now become stronger than uncorrelated. **C.** Distributions of strengths for synapses receiving uncorrelated (top) and correlated (bottom) inputs as a function of the inhibitory input rate. The color level indicates the probability density of strengths. A transition from anti-Hebbian to Hebbian competition occurs at an inhibitory input rate of 14 Hz (dotted line). Arrows indicate the parameters for panels A and B. **D.** The transitional inhibitory rate as a function of correlation time constant. The transition takes place at lower inhibitory rates as the correlation time constant increases up to $2ms$, then remains constant at $7Hz$ for higher values. The insets show the full distribution of correlated and uncorrelated synaptic strengths as in C, for correlation time constants of 1, 2 and $10ms$.

small random jitter in the timing of the synchronous spikes. The mean of this jitter determines the correlation time constant. Breaking perfect synchrony does not change the above results qualitatively. However, the rate of inhibitory input needed to transition from anti-Hebbian to Hebbian competition is sensitive to the correlation time constant (figure 3.4D). When the correlation time constant increases, the inhibitory rate at the transition decreases, until the correlation time constant becomes greater than the shift of the STDP window ($2ms$). Further increase in the correlation time constant does not lead to any more lowering of the transitional inhibitory rate (figure 3.4D).

The dependence of the outcome of synaptic competition on the level of inhibitory input can be explained by evaluating the effect of inhibition on the firing regime of the postsynaptic neuron. When the inhibitory input to a neuron is low, it operates in a "mean-driven" regime, meaning that the time-averaged "free-running" membrane potential (that is, the membrane potential if the spike generation mechanism is turned off) is above the firing threshold (Gerstein and Mandelbrot, 2006). In the mean-driven regime, integrate-and-fire neurons spike regularly, so the coefficient of variation of the inter-spike-intervals (CV_{ISI}), which is a measure of the irregularity of firing, is small (Shadlen and Newsome, 1998). On the other hand, when the inhibitory input to the neuron is high, the mean membrane potential is below the firing threshold. In this case, large deviations in the membrane potential from its mean are required to make the neuron fire, and the neuron is said to be in the "fluctuation-driven" regime (Gerstein and Mandelbrot, 2006; Shadlen and Newsome, 1998; Troyer and Miller, 1997). This makes firing times irregular, resulting in a larger CV_{ISI}.

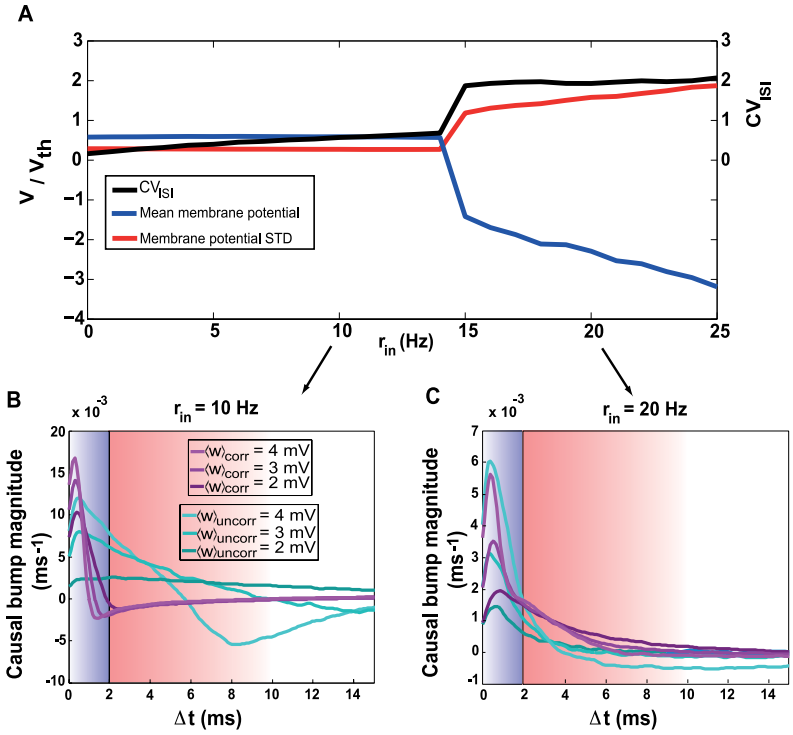

FIG. 3.5: The effect of the inhibitory input on synaptic competition. **A.** Transition from mean-driven to fluctuation-driven firing regimes when the rate of the inhibitory input is increased. The black curve is the coefficient of variation of postsynaptic interspike intervals (CV_{ISI}), the blue curve is the mean free-running membrane potential in units of the spiking threshold, and the red curve is the standard deviation of the membrane potential in the same units. For inhibitory input rates greater than 14 Hz, there is an abrupt switch from the mean-driven to the fluctuation-driven regime, corresponding to the transition from anti-Hebbian to Hebbian competition (figure 3.4). **B.** Postsynaptic causal bumps due to uncorrelated (cyan) and correlated (magenta) input spikes for different mean synaptic strengths (shading) when the inhibitory input rate is 10 Hz. The blue area shows the depression domain and the red area is the potentiation domain. Note that the correlated causal bumps (magenta) fall almost entirely into the depression domain (blue shading) in this case, so the correlated synapses lose the competition. **C.** Same as panel B, but for an inhibitory input rate of 20 Hz. Note the heavy tail of the correlated causal bumps (magenta), which extend into the potentiation domain of the STDP window. These curves were obtained by numerical simulations, changing the mean of the steady-state distribution of correlated or uncorrelated synapses to the desired value for each curve. Because the correlated synapses arrive in unison, their causal bump is the aggregate effect of all of their spikes. To show the contribution of individual correlated spikes, comparable to that of the uncorrelated ones, we therefore normalized the magnitude of the causal bump of the correlated synapses by their average cluster size ($\sim c\,N_{ex}/2$).

The model neuron we study traverses these regimes as the firing rate of its inhibitory inputs is varied (figure 3.5A). When the inhibitory input is small, the neuron operates in the mean-driven regime, with its mean free-running membrane potential above threshold and a small CV_{ISI}. When the inhibitory input rates increase beyond 14 Hz, the neuron suddenly switches to a fluctuation-driven regime in which the mean membrane potential is below threshold and CV_{ISI} is large. The transition between the mean-driven and fluctuation-driven regimes occurs exactly where synaptic competition switches from being anti-Hebbian to Hebbian (compare figure 3.5C with 3.5A). Thus, the key feature determining whether plasticity is anti-Hebbian or Hebbian is whether the postsynaptic neuron is in a mean-driven or fluctuation-driven state.

Recall that the causal bump is the excess probability of postsynaptic firing caused by an incoming input spike. As mentioned previously, the effect of shifted STDP on the distribution of synaptic strengths can be explained by considering the shape of the postsynaptic causal bump in relation to the STDP temporal window. When the postsynaptic neuron is in the mean-driven regime, the membrane potential rises rapidly to the threshold. As a result, presynaptic action potentials can only enhance postsynaptic firing if they occur during a relatively short time-interval prior to the postsynaptic spike. This means that the causal bump decays rapidly for longer intervals. The causal bump also has a higher amplitude and decays more rapidly for stronger synapses (figure 3.5B). Furthermore, the causal bump due to correlated inputs is even narrower and sharper (and more inside the depression region) than the bump due to uncorrelated inputs (figure 3.5B, magenta traces), because correlated spikes are more likely to induce a postsynaptic spike

rapidly when they occur in unison. As a result, the uncorrelated synapses win the synaptic competition when the level of inhibition is low.

When the postsynaptic neuron fires in the fluctuation-driven regime, the membrane potential spends a considerable time near but below the firing threshold before spiking. As a result, presynaptic input can affect postsynaptic firing over a longer time interval than in the mean-driven regime. This makes the causal bump broader than in the mean-driven case (figure 3.5C). Furthermore, the causal bump is even broader for correlated than for uncorrelated inputs because the simultaneous arrival of correlated spikes generates a stronger depolarization transient that makes it possible for subsequent weaker inputs to push the postsynaptic neuron above threshold over a longer time interval. This gives the causal bump for the correlated inputs a long tail that extends well into the potentiation domain of the STDP window (figure 3.5C, magenta traces), allowing them to win the competition in this case.

The transition from the mean-driven to the fluctuation-driven regime and correspondingly from anti-Hebbian to Hebbian competition is quite abrupt. This may be due to the interplay between the correlated inputs and the firing mode of the neuron. Correlated inputs increase membrane potential fluctuations and spiking irregularity (Salinas and Sejnowski, 2000). Therefore, within the context of shifted STDP, there is positive feedback between the fluctuation-driven regime and the dominance of correlated inputs. As the neuron transitions to the fluctuation-driven regime through increased inhibition, the correlated synapses start to strengthen more than the uncorrected ones which, in turn, increases the fluctuations of the membrane potential

and pushes the neuron further into the fluctuation-driven regime. This positive feedback continues until the correlated synapses dominate over the uncorrelated ones and the neuron falls completely into the fluctuation-driven mode.

3.3.4 Jittered STDP window

It is not necessary to introduce an explicit shift into the STDP window to assure stability. Any mechanism that causes depression to dominate over potentiation for short positive pairing intervals will have the same qualitative effect. One such mechanism is a symmetric random jitter introduced into an unshifted STDP window that has $A_- > A_+$. By jitter we mean that the time Δt used to determine the effect of STDP for any given pair of pre- and postsynaptic spikes, is not simply the difference between the times of their occurrence, but instead a random term is added. In other words, $\Delta t = t_{\text{post}} - t_{\text{pre}} + \eta$, where η is a random variable drawn from a distribution with zero mean and a certain variance (we use a Gaussian distribution). Although the STDP window has no explicit shift in this case (figure 3.3.6A, top), the effective window obtained by averaging over the symmetric random jitter (figure 3.6A, bottom), exhibits the required feature that depression occurs for small positive pairing intervals.

Simulations show that jittered STDP has all the qualitative properties of shifted STDP, although the maximum depression must be set to be greater rather than the maximum potentiation (we take $A_+ = 0.005mV$ and $A_- = 0.007mV$, although see Froemke et al. (2005)). To keep $A_+\tau_+ > A_-\tau_-$, as required for stability, the time constant of potentiation must be larger than that of depression (we take $\tau_+ = 20ms$ and $\tau_- = 10ms$). If the standard deviation of the jitter is less than $2ms$, the steady-state distribution of synaptic weights is not inherently stable and we

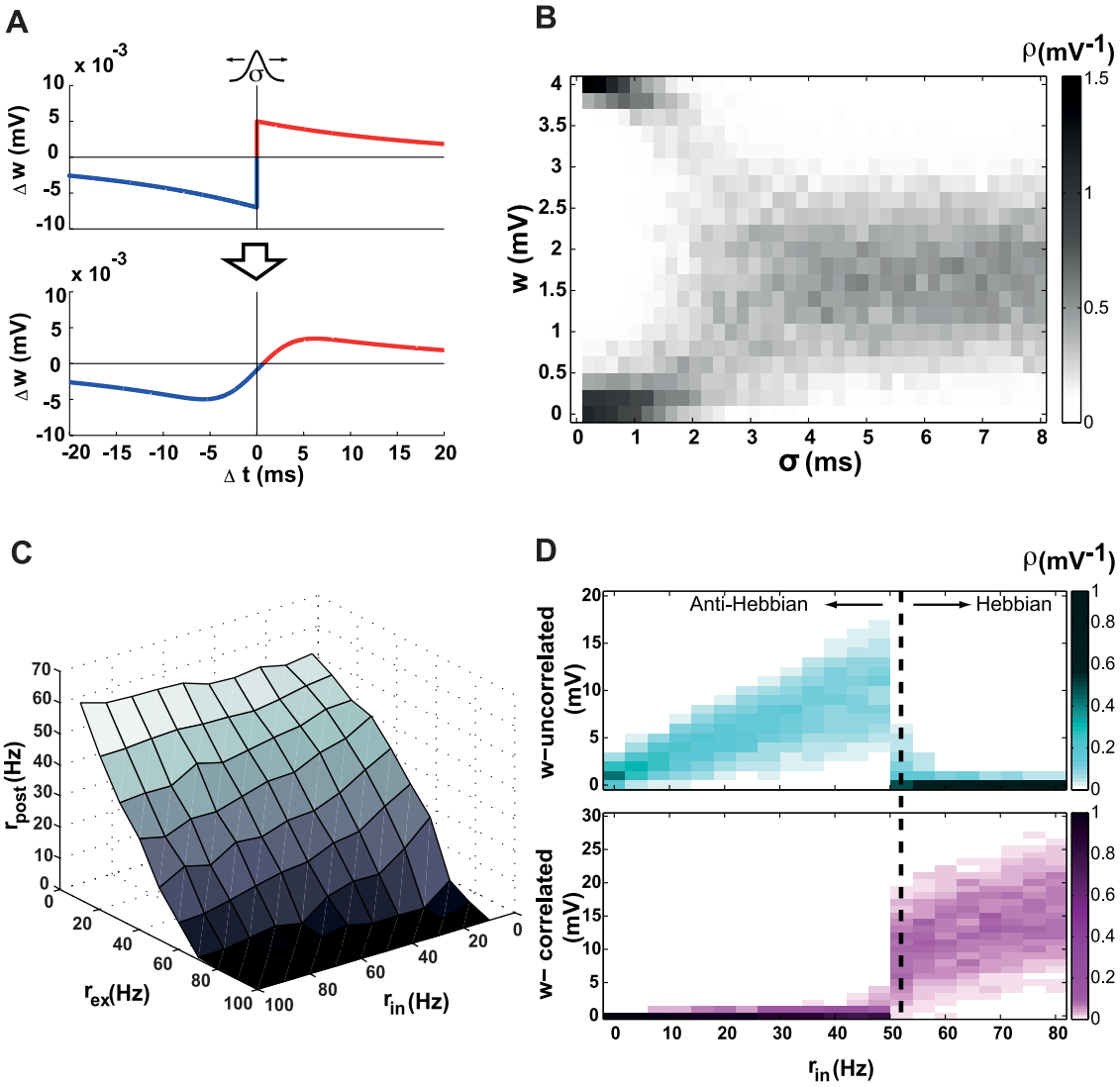

FIG. 3.6: Jittered STDP. **A.** A random symmetric jitter of the unshifted STDP window (top) results in an effective window function (bottom) in which depression is dominant for short positive pairing intervals (blue shading). **B.** Jittered STDP stabilizes the distribution of synaptic weights. The horizontal axis is the standard deviation of the jitter (σ), the vertical axis is synaptic strength and the gray level indicates the probability density of strengths. For jitters smaller than 2 ms the distribution is bimodal, but for larger jitters it is stable and unimodal. **C.** The steady-state firing rate of the postsynaptic neuron as a function of the excitatory and inhibitory input rates when the jitter is 3 ms. **D.** Jittered STDP ($\sigma = 3ms$) implements both Hebbian and anti-Hebbian competition. As in figure 3.4, the top panel shows the distribution of uncorrelated synapses (cyan) and the bottom panel shows the distribution of correlated synapses (magenta), both as functions of the inhibitory input rate. The transition from anti-Hebbian to Hebbian competition occurs when the inhibitory input rate is about 50 Hz in this case.

obtain a U-shaped distribution of synaptic strengths (figure 3.6B). However, for larger standard deviations of the jitter, the steady-state distribution is stable and unimodal as in the case of shifted STDP (figure 3.6B). Other features of shifted STDP are also reproduced. The steady-state firing rate of the postsynaptic neuron decreases when the rate of presynaptic input increases (figure 3.6C), and either anti-Hebbian or Hebbian competition occurs depending on the rate of inhibitory input to the neuron (figure 3.6D).

3.4 Discussion

We have shown that a slight shift in the effective STDP temporal window, such that postsynaptic spikes occurring shortly after presynaptic action potentials cause synaptic depression, can stabilize the distribution of synaptic strengths without loss of competition, both in pair-based and triplet-based models. The shift can be explicitly implemented in the STDP window or achieved by other means such as a symmetric spike-by-spike random jitter. In fact, any mechanism that causes synaptic depression for small but causal (positive by our convention) pre-post spike intervals should lead to the stabilization and other effects we report. What biophysical mechanisms could cause this to occur?

The sharp transition between depression and potentiation in STDP appears to be due to the abrupt onset of long-term potentiation (Sjöström et al., 2003; Bender et al., 2006). It is believed that the Ca^{2+} influx through NMDA receptors is responsible for this potentiation (Malenka and Bear, 2004) and that the abrupt

onset arises because the NMDA channel be in an open but blocked configuration before subsequent depolarization removes the Mg^{2+} block (Nowak et al., 1984). To assure a large Ca^{2+} influx and subsequent potentiation, it seems reasonable to assume that the depolarization that removes the Mg^{2+} block should occur near the peak of the NMDA activation. The Mg^{2+} removal by postsynaptic depolarization is extremely rapid (Jahr and Stevens, 1990) but the NMDA activation has a finite rise time, so the peak of NMDA activation occurs a few milliseconds after the arrival of the presynaptic spike (Destexhe et al., 1994). Therefore, it seems likely that the maximum potentiation should occur when the presynaptic spike precedes the postsynaptic action potential by several milliseconds, and that depression could result from timing differences shorter than this. Thus, the biophysics of the NMDA receptor appears to support the idea of a temporal shift in the STDP window. The shape of the STDP window has been inferred from models of NMDA receptor kinetics and back-propagating action potentials (Shouval et al., 2002; Karmarkar et al., 2002). However, the millisecond timing of the transition from depression to potentiation was not investigated systematically, because its significance was not evident at that time. Nevertheless, in some parameterizations of such models a small depression domain for short positive pairing intervals has been reported (Karmarkar et al., 2002).

Typically in electrophysiological recordings, action potentials are measured at the soma, but what matters for STDP is the timing of the events at the synapse. More precisely, the timing of the postsynaptic EPSP and that of the backpropagating action potential to the synapse control plasticity. Transmission delays may have their own interesting computational properties. For example, it has been shown that STDP in the presence of axonal transmission delays can have a desyn-

chronizing effect on population bursts and a synchronizing effect on random spiking in a recurrent network (Lubenov and Siapas, 2008). The transmission delay of the EPSP to the soma and that of the backpropagating action potential subtract from the delay we need for shifted STDP. For distal synapses where these delays are longer, there may be a higher probability that the causal bump falls out of the depression domain caused by the shift. This might be a mechanism for counterbalancing the attenuation of the EPSPs arising from distal dendrites (Magee and Cook, 2000; Andrasfalvy and Magee, 2001) along with other proposed mechanisms (Rumsey and Abbott, 2004; Gidon and Segev, 2009). It may explain the enhancement of LTD reported in studies of STDP at distal sites (Froemke et al., 2005; Sjöström and Häusser, 2006; Letzkus et al., 2006). If the delay becomes longer than the shift for very distal synapses, other mechanisms such as limits on synaptic strength must serve to stabilize STDP. Finally, if the speed of backpropagating action potential can be increased through modification of voltage-dependent conductances, the model predicts that synapses should be more readily depressed.

The most direct test of the shifted STDP hypothesis would be to observe the effect of almost synchronous pre- and postsynaptic spikes on synaptic strength. However, the results of such experiments could be difficult to interpret because of confounding factors such as the physiological delays mentioned above. For example, if the pre- and postsynaptic spikes are induced exactly at the same time, the timing of their arrival at the synapse is not necessarily synchronous. If a shift in the STDP window function acts as a stabilizing mechanism, synapses should get depressed when postsynaptic spikes are generated by presynaptic spikes with short latency. Therefore, as an alternative experiment we suggest inducing spikes only in

the presynaptic neuron and allowing the postsynaptic firing to be affected by this presynaptic activity. One possible way to perform such an experiment is to hold the voltage of the postsynaptic neuron close to its firing threshold, so that individual EPSPs can induce a postsynaptic spike. In this case, if there is a stabilizing shift in the STDP window, strong synapses that induce short-latency postsynaptic action potentials abruptly should get depressed.

Shifted STDP results in a unimodal distribution of synaptic strengths. This finding is in agreement with the measurements of quantal synaptic currents (Turrigiano et al., 1998; O'Brien et al., 1998) and from paired recordings (Song et al., 2005). However, the observed distribution of peak EPSP amplitudes has a heavier tail than the gamma distribution obtained from shifted STDP (see also Andrasfalvy and Magee (2001); Katz et al. (2009)). STDP is unlikely to be the only mechanisms involved in shaping the distribution of synaptic strengths. Nevertheless, figure 3.4 shows that in presence of correlated input, this distribution can be quite broad. Thus, in the context of shifted STDP, a heavy-tailed distribution may be a sign of multiple correlated subgroups of input spike trains.

The synapses in the model we considered were current-based, meaning that each excitatory or inhibitory input injects a current waveform to the neuron regardless of the value of its membrane potential. We have also studied an analogous model with conductance-based synapses, and this does not qualitatively change the reported results. These results show that the outcome of competition between correlated and uncorrelated spike trains with shifted STDP depends on the firing state of the postsynaptic neuron, which can be controlled by the rate of its inhibitory in-

puts. This allows for a dynamic switching between anti-Hebbian and Hebbian forms of plasticity, and it might be related to the role of local inhibitory interneurons in switching the activity-dependent development of visual cortical circuits during the critical period (Hensch, 2005).

In conclusion, a slightly shifted STDP window stabilizes synaptic strength, buffers firing rates, and can implement different modes of synaptic competition. The required shift may arise from properties of the NMDA receptor, or from random jitter. In light of their importance in determined the outcome of synaptic plasticity, we argue that the properties of STDP for short pairing intervals, which have not yet been clearly resolved, warrant a more detailed investigation.

Chapter 4

Stability and Competition in Multi-Spike STDP Models

Only the multitude through its practical experimentation will offer the models and determine when and how the possible becomes real.
Antonio Negri

4.1 Introduction

In the models of STDP that we studied in the previous chapters, pairs of pre- and postsynaptic action potentials potentiate a synapse when the presynaptic spike precedes the postsynaptic spike, and depress it for the reverse order (Markram et al., 1997). However, when multiple pre- and postsynaptic spikes occur across a synapses over a short interval of time, the resulting plasticity depends on their timing in a more complex manner. For example, pair-based STDP models predict that "pre-post-pre" and "post-pre-post" triplets of spikes with the same pairwise intervals should induce the same plasticity, but experiments indicate that these two triplet patterns have different effects (Froemke and Dan, 2002; Wang et al., 2005). This and similar contradictions motivated the development of multi-spike models

of STDP, which go beyond pairwise interactions of pre- and postsynaptic spikes (see Morrison et al. (2008) for a review). Here, we analyze three such models to determine how they affect populations of synapses converging onto a postsynaptic neuron. As in previous chapters, we focus here on two basic features: stability and competition.

The three multi-spike STDP models that we consider were proposed on the basis of different experimental results. In the "suppression model", inspired by experimental results in cortical slices, the plasticity-inducing effect of each pre- or postsynaptic spike is suppressed by preceding spikes (Froemke and Dan, 2002; Froemke et al., 2005). The "triplet model", inspired by experiments in hippocampal cultures (Wang et al., 2005), includes the effect of neighboring pre-post pairings as well as depression exerted by preceding presynaptic spikes and potentiation by preceding postsynaptic spikes (Pfister and Gerstner, 2006). Another model that incorporates multi-spike interactions, the "NMDAR-based model", is based on the kinetics of the NMDA receptor. It was proposed by Senn et al. (2001) before experimental results on multi-spike effects in STDP were available.

We apply the approach developed in Chapter 2 for characterizing the effects of plasticity through the multi-spike STDP models. For each case, we first consider the parameters originally proposed for the model, and then systematically explore a range of parameter values.

4.2 Methods

The neural and synaptic models are the same as in Chapter 2 (Table 2.1).

4.2.1 The triplet model

In the triplet model, synapses are modified on the basis of pre-post pairing events in a manner similar to the pair-based model (equation 2.5) but, in addition, when a synapse is potentiated by a pre-post pairing ($\Delta t > 0$), the postsynaptic potentiation variable $M_{\rm post}$ is added to the amount of the pair-based potentiation A_+. Similarly, when a synapse is depressed by a paring event ($\Delta t < 0$), the presynaptic depression variable $M_{\rm pre}$ is added to the pair-based depression A_-. Thus,

$$\Delta w = F_{trip}(\Delta t) = \begin{cases} \left[A_+ + M_{post}(t-\epsilon)\right] \exp(-\Delta t/\tau_+) & \text{if } \Delta t \geq 0 \\ \\ -\left[A_- + M_{pre}(t-\epsilon)\right] \exp(\Delta t/\tau_-) & \text{if } \Delta t < 0 \end{cases} . \qquad (4.1)$$

The small parameter ϵ ensures that the values of M_{pre} and M_{post} just before their update by the pre- or postsynaptic spikes are used. The postsynaptic potentiation and presynaptic depression variables are governed by the equations

$$\begin{aligned} \frac{dM_{\rm pre}}{dt} &= -\frac{M_{\rm pre}}{\tau_{\rm pre}} + A_{\rm pre} \sum_i \delta(t - t_{\rm pre}^{(i)}) \\ \frac{dM_{\rm post}}{dt} &= -\frac{M_{\rm post}}{\tau_{\rm post}} + A_{\rm post} \sum_i \delta(t - t_{\rm post}^{(i)}) \,, \end{aligned} \qquad (4.2)$$

where $\delta(t)$ is the Dirac delta function, and $t_{\rm pre}^{(i)}$ and $t_{\rm post}^{(i)}$ are the times of arrival of pre- and postsynaptic spikes respectively. This introduces 4 parameters into the model beyond those of the pair-based model: the time constants $\tau_{\rm pre}$ and $\tau_{\rm post}$ and

the increments A_{pre} and A_{post}.

4.2.2 The suppression model

In the suppression model with time constants τ_{pre} and τ_{post}, the change in a synaptic weight is determined by

$$\begin{aligned} \Delta w &= F_{supp}(\Delta t) = \Big[1 - \exp(-\Delta t_{\text{pre}}/\tau_{\text{pre}})\Big]\Big[1 - \exp(-\Delta t_{\text{post}}/\tau_{\text{post}})\Big] \\ &\times \begin{cases} A_+ \exp(-\Delta t/\tau_+) & \text{if } \Delta t \geq 0 \\ \\ A_- \exp(\Delta t/\tau_-) & \text{if } \Delta t < 0\,, \end{cases} \end{aligned} \tag{4.3}$$

where Δt_{pre} is the interval between the presynaptic spike in the pair and its preceding presynaptic spike, and Δt_{post} is the interval between the postsynaptic spike and its preceding spike. The suppression model introduces 2 new parameters beyond those of the pair-based model: the time constants τ_{pre} and τ_{post}.

4.2.3 The NMDAR-based model

The NMDAR-based model (Senn et al., 2001; Gerstner and Kistler, 2002) is based on the assumption that NMDARs can be in one of three different states, "rest", "up" and "down". The variables f^{rest}, f^{up} and f^{dn} denote the fraction of NMDARs in each state respectively ($f^{\text{up}} + f^{\text{dn}} + f^{\text{rest}} = 1$). In the absence of pre- and postsynaptic spikes, the receptors in up and down states return to the rest state with time constants τ_f^{up} and τ_f^{dn} respectively. Each presynaptic spike up-regulates the receptors immediately after its arrival by an amount proportional to a parameter

$A_f^{\rm up}$, and each postsynaptic spike down-regulates the receptors proportional to a parameter $A_f^{\rm dn}$. The dynamics of the NMDARs in the 'up' and 'down' states can be expressed as:

$$\begin{aligned}\frac{df^{\rm up}}{dt} &= -\frac{f^{\rm up}}{\tau_f^{\rm up}} + A_f^{\rm up}\, f^{\rm rest} \sum_i \delta(t - t_{\rm pre}^{(i)}) \\ \frac{df^{\rm dn}}{dt} &= -\frac{f^{\rm dn}}{\tau_f^{\rm dn}} + A_f^{\rm dn}\, f^{\rm rest} \sum_i \delta(t - t_{\rm post}^{(i)})\,, \end{aligned} \tag{4.4}$$

where the sums run over all pre- ($t_{\rm pre}^{(i)}$) or postsynaptic ($t_{\rm post}^{(i)}$) spike times, indexed by i. In this and subsequent equations, we assume the convention that a quantity multiplying a δ function is evaluated immediately before the time when the argument of the δ function is zero.

The fraction of active second messenger $M^{\rm up}$ is increased by postsynaptic spikes proportional to the amount of up-regulated NMDARs $f^{\rm up}$ and the available inactive messengers $1 - M^{\rm up}$. Likewise, the fraction of active second messenger $M^{\rm dn}$ is increased by presynaptic spikes proportional to the amount of down-regulated NMDARs $f^{\rm dn}$ and available inactive messenger $1 - M^{\rm dn}$. In the absence of spikes, these second messenger fractions decay with time constants $\tau_{M^{\rm up}}$ and $\tau_{M^{\rm dn}}$, respectively. Thus,

$$\begin{aligned}\frac{dM^{\rm up}}{dt} &= -\frac{M^{\rm up}}{\tau_M^{\rm up}} + A_M^{\rm up}\, f^{\rm up}(1 - M^{\rm up}) \sum_i \delta(t - t_{\rm post}^{(i)}) \\ \frac{dM^{\rm dn}}{dt} &= -\frac{M^{\rm dn}}{\tau_M^{\rm dn}} + A_M^{\rm dn}\, f^{\rm dn}(1 - M^{\rm dn}) \sum_i \delta(t - t_{\rm pre}^{(i)})\,, \end{aligned} \tag{4.5}$$

where the sums run over all pre- ($t_{\rm pre}^{(i)}$) or postsynaptic ($t_{\rm post}^{(i)}$) spike times. The

parameters $A_M^{\rm up}$ and $A_M^{\rm dn}$ governing the magnitude of the changes in the messengers on spiking events. Finally, synaptic potentiation occurs in response to postsynaptic spikes and depends on the amount of $M^{\rm up}$, and synaptic depression occurs in response to presynaptic spikes depending on the amount of $M^{\rm dn}$, so that

$$\frac{dw}{dt} = A_+ \left[M^{\rm up} - \theta^{\rm up}\right]^+ \sum_i \delta(t - t_{\rm post}^{(i)} - \epsilon) - A_- \left[M^{\rm dn} - \theta^{\rm dn}\right]^+ \sum_i \delta(t - t_{\rm pre}^{(i)} - \epsilon) \quad (4.6)$$

where $\theta^{\rm up}$ and $\theta^{\rm dn}$ are thresholds above which the corresponding messengers take part in plasticity, and $[x]^+$ denotes the piece-wise linear threshold function $[x]^+ = x$ for $x > 0$ and zero otherwise. The small parameter ϵ is included because, in this case, we evaluate the factors multiplying the δ functions *after* the time of a spike, as required by the model.

4.3 Results

To study the effect of different STDP models on synaptic weights, we simulated a single spiking neuron receiving $N_{\rm ex}$ excitatory and $N_{\rm in}$ inhibitory presynaptic spike trains with Poisson statistics at rates $r_{\rm pre}$ and $r_{\rm in}$, respectively. The strengths of the excitatory synapses, denoted by w, change according to STDP, while the strengths of inhibitory synapses remain constant. To examine the different forms of stability of each STDP model, we check whether the steady-state distribution of synaptic strengths is bounded without imposing external limits, or whether the increase or decrease of the weights is stopped only when they hit a boundary. We distinguish between the partially stable and unstable cases by computing the evolution of the average of the synaptic weights, which reaches a fixed point only in the partially

	Triplet	**Suppression**	**NMDAR-based**
A_+	5.3×10^{-3} mV	1.3×10^{-2} mV	10^{-3} mV
A_-	3.5×10^{-3} mV	5.1×10^{-3} mV	10^{-3} mV
A_{pre}	0	–	–
A_{post}	8×10^{-3} mV	–	–
A_f^{up}	–	–	1
A_f^{dn}	–	–	0.5
A_M^{up}	–	–	0.7
A_M^{dn}	–	–	0.7
τ_+	16.8 ms	13.3 ms	–
τ_-	33.7 ms	34.5 ms	–
τ_{pre}	–	28 ms	–
τ_{post}	40 ms	88 ms	–
τ_f^{up}	–	–	300 ms
τ_f^{dn}	–	–	300 ms
$\tau_{\text{m}}^{\text{up}}$	–	–	600 ms
$\tau_{\text{m}}^{\text{dn}}$	–	–	600 ms
θ^{up}	–	–	0.7
θ^{dn}	–	–	0.35

Table 4.1: Original parameters of the multi-spike STDP models used to generate figures 4.1B, 4.4B, 4.7B and 4.10

stable case. Fully stable STDP is characterized by a fixed point for the average rate and bounded deviations for the strengths of individual synapses about this mean. As a probe of synaptic competition, we induce correlations in half of the excitatory inputs (Chapter 2, Methods) and check whether STDP causes the synapses corresponding to correlated and uncorrelated subsets to compete for control of the postsynaptic firing. This also allows us to determine whether the effect of correlations is Hebbian or anti-Hebbian. The neuronal and input parameters used in our simulations are given in Chapter 2, Table 2.1.

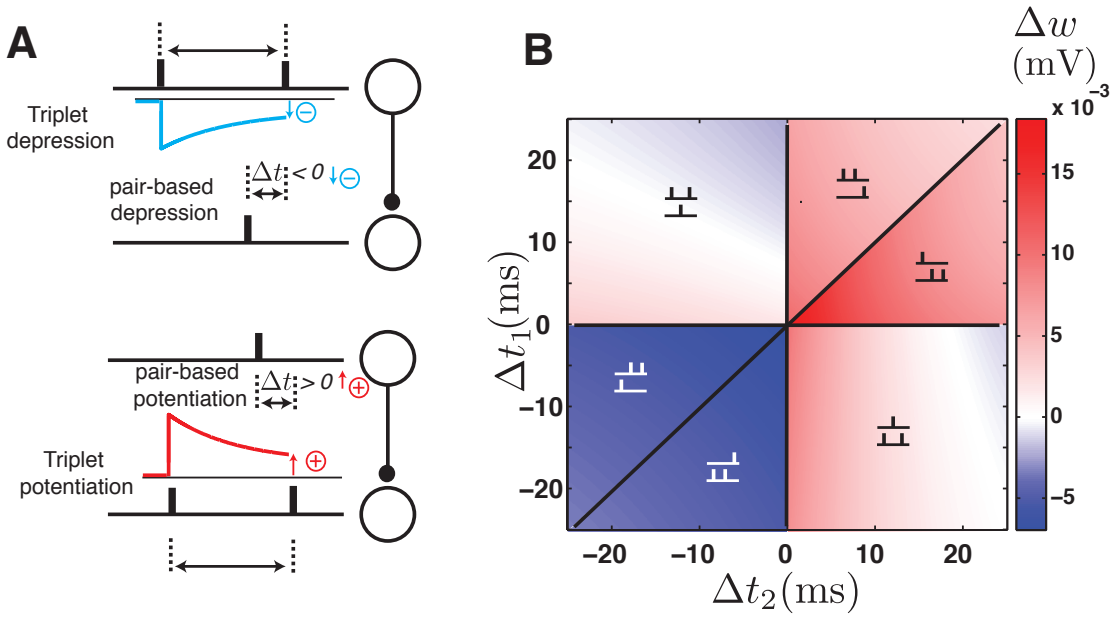

FIG. 4.1: **A.** Schematic illustration of spike interactions in the triplet model in which previous presynaptic spikes induce extra depression (top) and previous postsynaptic spikes induce extra potentiation (bottom). **B.** Plasticity due to triplets of spikes: pre-post-pre triplets induce depression or weak potentiation (top left), and post-pre-post ordering induces potentiation (bottom right). This figures is based on parameters fit to hippocampal data (Table 4.1).

4.3.1 Stability and competition in the triplet model

Experimental results on synapses in hippocampal cultures reveal a marked asymmetry in the plasticity induced by post-pre-post and pre-post-pre spike sequences, in contrast to the predictions of the pair-based model (Chapter 2, figure 2.1B). Post-pre-post sequences induce potentiation, and pre-post-pre have little or no effect (Wang et al., 2005). In addition, in experiments on cortical synapses, the balance between potentiation and depression shifts toward potentiation when the frequency of pre-post pairing events increases, another property not captured by pair-based

STDP (Sjöström et al., 2001). These results motivated Pfister and Gerstner (2006) to propose the triplet model, which takes into account interactions of spikes beyond pre-post pairings. In addition to the effect of pre-post pairings, the triplet model includes additional depression due to previous presynaptic spikes and additional potentiation from earlier postsynaptic spikes (figure 4.1A). This is accomplished through a presynaptic depression variable and a postsynaptic potentiation variable assigned to each synapse. In the absence of incoming presynaptic spikes, the presynaptic depression variable decays exponentially with time constant τ_{pre}. Likewise, the value of the postsynaptic potentiation variable decreases exponentially in the absence of postsynaptic spikes with time constant τ_{post}. When a presynaptic spike reaches the synapse, the presynaptic depression variable abruptly increases by the amount A_{pre}, and when a postsynaptic spike occurs, the postsynaptic variable increases by A_{post}. This is how the triplet model accounts for the asymmetry of synaptic modification in response to triplets. For a pre-post-pre triplet, the first presynaptic spike induces extra depression on the synapse, while for a post-pre-post triplet the first postsynaptic spike induces extra potentiation (figure 4.1B). The triplet model that we consider sums the contributions of all previous pre- and postsynaptic spikes as well as all pre-post pairings (all-to-all). Pfister and Gerstner (2006) also provided a version of the triplet model based only on nearest neighboring spikes, but the qualitative behavior of both versions is similar.

As we did for the pair-based STDP model (Chapter 2), we can derive equations governing the evolution of the mean synaptic strengths and deviations around the mean for individual synaptic weights. The average values of the presynaptic depression and postsynaptic potentiation variables, obtained from substituting rates

for spikes in equation (4.5), are $A_{\rm pre}\,\tau_{\rm pre}\,r_{\rm pre}$ and $A_{\rm post}\,\tau_{\rm post}\,\bar{r}_{\rm post}$. Using these values and averaging over the probability of pre-post pairings, the drift of the mean of the synaptic weights in the triplet model is

$$\begin{aligned}\frac{d\langle w\rangle}{dt} &= A_+\,\tau_+\,r_{\rm pre}\,\bar{r}_{\rm post}+A_{\rm post}\,\tau_{\rm post}\,\tau_+\,r_{\rm pre}\,\bar{r}^2_{\rm post}-A_-\,\tau_-\,r_{\rm pre}\,\bar{r}_{\rm post}-A_{\rm pre}\,\tau_{\rm pre}\,\tau_-\,r^2_{\rm pre}\,\bar{r}_{\rm post}\\ &\quad+\frac{(A_+ + A_{\rm post}\,\tau_{\rm post}\,\bar{r}_{\rm post})\tau_+\,\tau_{\rm s}\,r_{\rm pre}\,\langle w\rangle}{(\tau_{\rm s}+\tau_+)(V_{th}-V_{\rm r})\tau_{\rm m}}\,. \end{aligned} \tag{4.7}$$

As in the pair-based model, the last term in this equation is the w-dependent drift and the other terms make up the baseline drift. The dynamics of deviations of individual synapses from the mean is governed by the w-dependent drift, so

$$\frac{d\delta w}{dt}=\frac{(A_+ + A_{\rm post}\,\tau_{\rm post}\,\bar{r}_{\rm post})\tau_+\,\tau_{\rm s}\,r_{\rm pre}\,\delta w}{(\tau_{\rm s}+\tau_+)(V_{th}-V_{\rm r})\tau_{\rm m}}\,. \tag{4.8}$$

As in the pair-based model, the coefficient of δw is always positive, so individual weights will drift away from the mean for any choice of parameters, making individual synaptic weights unstable.

The parameters of the original model were optimized separately by Pfister and Gerstner (2006) to match experimental data from hippocampal cultures and cortical slices, resulting in two sets of parameters. Our simulation results indicate that, for both sets of parameters, the distribution of synaptic weights is unstable, so that all the synaptic weights cluster around the upper bound. In addition, no competition takes place between correlated and uncorrelated synapses with these parameter sets. This led us to consider properties of the triplet model for a range of parameter values. As in our discussion of pair-based STDP, we study the triplet model when pair-based potentiation is larger than pair-based depression ($A_+ = 0.005$ mV,

$A_- = 1.01\,A_+$) and when pair-based depression is larger than pair-based potentiation ($A_- = 0.005$ mV, $A_+ = 1.01\,A_-$). In each case, we varied the ratio between postsynaptic potentiation and presynaptic depression ($A_{\text{post}}/A_{\text{pre}}$) systematically, while keeping A_{post} constant at 0.001 mV.

We first examine the fixed points of the mean synaptic weight (figure 4.2A & F). When $A_{\text{post}}/A_{\text{pre}}$ is small, the average synaptic weight has two nontrivial fixed points (figure 4.2B & G). The first is stable (figures 4.2B & G, filled circle) and the second is unstable (figure 4.2B & G, open circle). The appearance of the unstable fixed point in the triplet model is due to the dependence of the postsynaptic potentiation on the postsynaptic firing rate. This added potentiation increases when the mean synaptic weight increases, eventually overcoming the combined effect of presynaptic and pair-based depression. The existence of two fixed points makes the steady-state distribution of synaptic weights sensitive to the initial distribution. If the mean of the initial distribution is greater than the unstable fixed point, the distribution will be unstable and all of the weights will be pushed toward the upper bound (figure 4.2D & I, middle). If the mean of the initial distribution is lower than the unstable fixed point, the mean of the steady-state distribution converges to the stable fixed point and individual weights drift away from the mean toward the lower and upper bounds, resulting in partial stability and a U-shaped distribution similar to the pair-based model (figures 4.2D & I, left). When $A_{\text{post}}/A_{\text{pre}}$ reaches a critical value, the two fixed points coalesce and annihilate each other, and only the trivial unstable fixed point remains (figures 4.2C & H). In this case, regardless of the initial distribution, the final distribution is unstable and tightly clustered near the upper bound (figures 4.2D & I, right).

As we argued in the case of the pair-based model, synaptic competition can only take place when the steady-state distribution has a nontrivial stable mean and is partially stable. In the triplet model, when $A_{\text{post}}/A_{\text{pre}}$ is relatively small and the initial mean synaptic weight is lower than the unstable fixed point, this condition is fulfilled (figures 4.2A & F, dark gray areas; figure 4.2E, left). However, if the stable and unstable fixed points are too close together, there is no guarantee of synaptic competition (figure 4.2J, left). The reason is that when a subset of the synaptic inputs are correlated, presynaptic spikes tend to arrive in tandem and induce large transients in the postsynaptic firing rate, causing large fluctuations in the mean synaptic weight. This can cause the mean synaptic strength to fluctuate beyond the unstable fixed point, destabilizing the weight distribution. As a result, the parameter regime for synaptic competition in the triplet model is highly restricted to the region of small $A_{\text{post}}/A_{\text{pre}}$. This region is even smaller when the pair-based component of potentiation is larger than the pair-based depression (figure 4.2F, dark gray area). Thus, it is not surprising that the original parameters obtained by (Pfister and Gerstner, 2006) did not lead to competitive synaptic plasticity. In summary, what appear as novel properties of the triplet model compared to the pair-based model are the sensitivity to the initial distribution of weights and a tighter parameter range for Hebbian competition.

4.3.1.1 Stability and competition in the shifted triplet model

The distribution of synaptic weights can be stabilized by a shift in the triplet model as well, similar to the model described in Chapter 3. Simulations show that the final

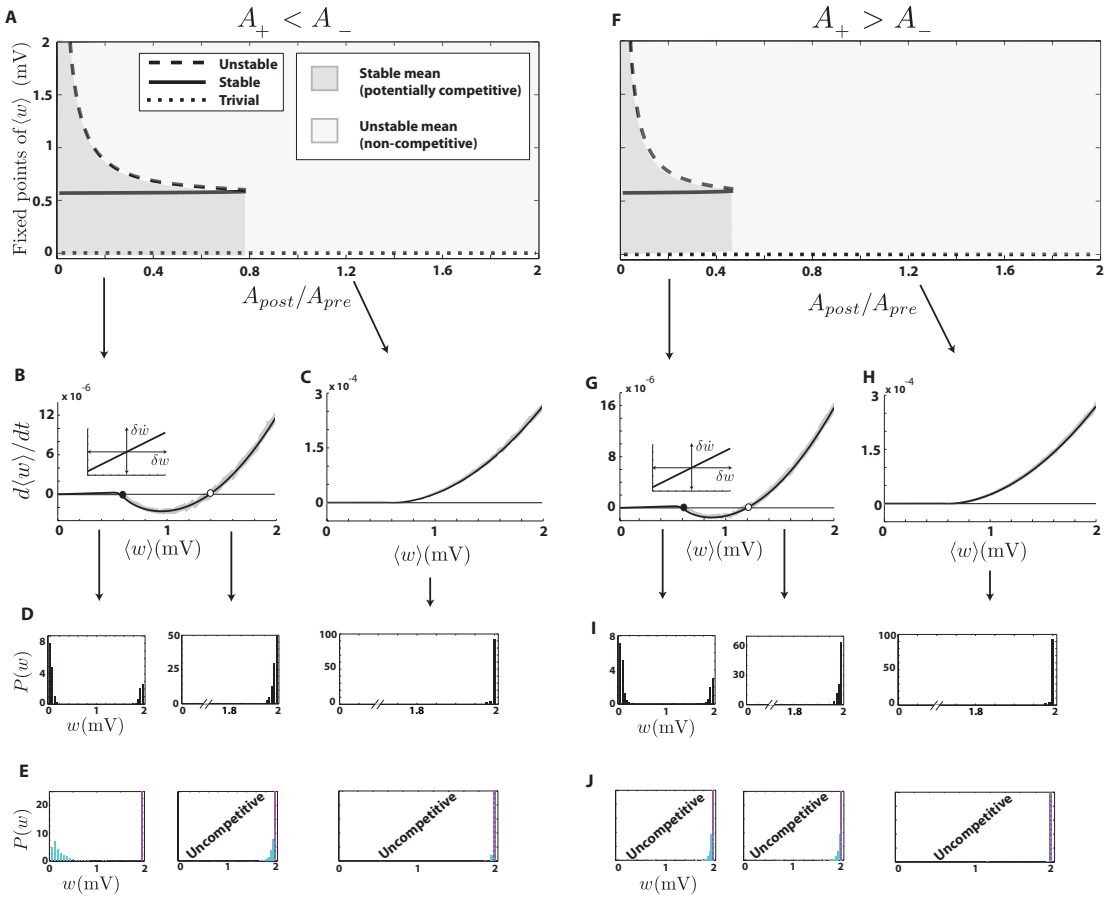

FIG. 4.2: Stability and competition in the triplet model. **A.** Fixed points of $\langle w \rangle$ as functions of the ratio $A_{\text{post}}/A_{\text{pre}}$. **B.** Average drift of the weights, when $A_{\text{post}}/A_{\text{pre}} = 0.2$. Gray area: simulation, solid curve: analytic, filled circle: stable fixed point, open circle: unstable fixed point, inset: w-dependent drift near the stable fixed point. **C.** Average drift of the weights when $A_{\text{post}}/A_{\text{pre}} = 1.2$. **D.** Distribution of synaptic weights obtained from simulation, with parameters as in B and an initial mean of 0.4 mV (left) or 1.6 (middle), or using parameters as in C. (right). **E.** Synaptic competition for the parameters and initial values used in D. **F-J.** Same as A-E, but when pair-based potentiation is larger than pair-based depression. For this figure, the time constants of presynaptic depression and postsynaptic potentiation were $\tau_{\text{pre}} = \tau_{\text{post}} = 40$ ms), and the pair-based time constants of the model equal were $\tau_+ = \tau_- = 20$ ms.

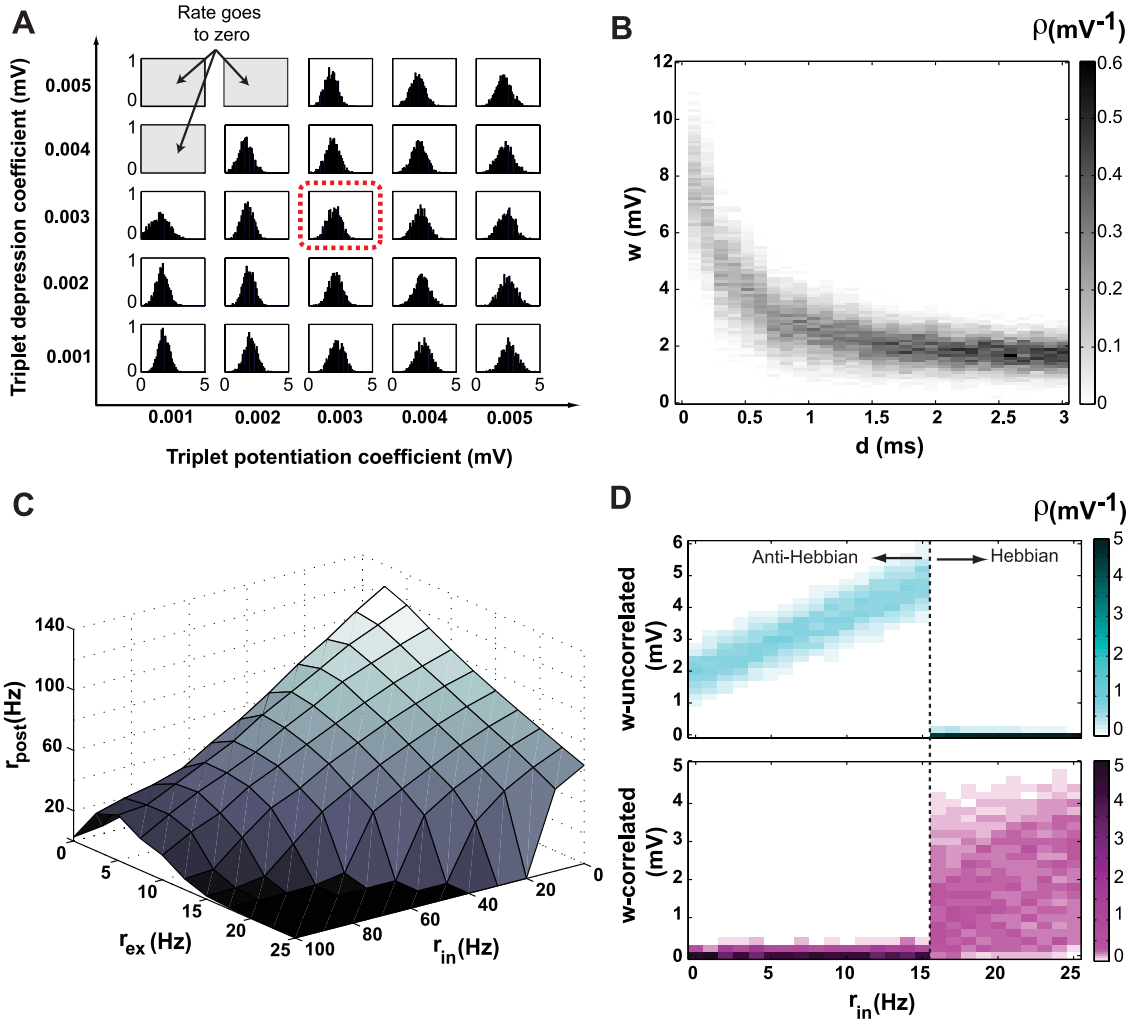

FIG. 4.3: The shifted triplet model. **A.** The final distribution of weights for different values of A_{pre} and A_{post}. Except for very high depression values, the distribution is unimodal and stable. We used the representative value of $0.003mV$ for both (A_{pre} and A_{post}, red dotted box) for the remaining results in this figure. **B.** The shift stabilizes the distribution of synaptic weights. The horizontal axis is the value of the shift, the vertical axis is the synaptic strength, and the gray level is the probability density of the strengths (as in figure 3.2), obtained by simulation. **C.** The steady-state firing rate of the postsynaptic neuron as a function of the excitatory and inhibitory input rates. **D.** The shift in the triplet model can implement both Hebbian and anti-Hebbian competition. As in figure 3.4, the top panel shows the distribution of the uncorrelated synapses (cyan) and the bottom panel shows the distribution of the correlated ones (magenta), as a function of the inhibitory input rate. The transition from anti-Hebbian to Hebbian competition occurs at an inhibitory input rate of 16 Hz.

distribution of weights is stable and unimodal (figure 4.3A) using a triplet model in which the window of the pre-post pairing is shifted by $2ms$, unless the presynaptic depression is extremely high, which causes the firing rate of the postsynaptic neuron to go to zero, terminating plasticity. After finding that the shifted STDP in the triplet model stabilizes weights for a wide range of parameters, we set A_{pre} and A_{post} to $0.003mV$ and examined other properties of shifted STDP in this model. Further simulations showed that shifted STDP within the framework of the triplet model has all the qualitative properties of the shifted pair-based STDP model (Chapter 3). A shift as low as 0.1 milliseconds is sufficient to stabilize the weights, with larger delays resulting in lower mean values and sharper distributions for the weights (figure 4.3B). The steady-state firing rate of the postsynaptic neuron decreases when the rate of the excitatory and/or inhibitory presynaptic input increases (figure 4.3C). Finally, either anti-Hebbian or Hebbian competition occurs depending on the rate of inhibitory input to the neuron (figure 4.3D).

4.3.2 Stability and competition in the suppression model

Plasticity experiments in cortical slices using triplets of spikes showed different effects than in cultured hippocampal neurons. In the synapses of the visual cortex of rats, pre-post-pre triplets induce potentiation whereas post-pre-post triplets induce depression (Froemke and Dan, 2002). These results led Froemke and Dan (2002) to propose the suppression model, in which plasticity is induced by nearest neighbor pre- and postsynaptic spikes. The plasticity is computed from the standard STDP curve, but the effect of the presynaptic spike in each pair is suppressed by previous presynaptic spikes and, similarly, the plasticity induced by the postsynaptic spike in each pair is suppressed by previous postsynaptic spikes (figure 4.4A). The

suppression is maximal immediately after each pre- or postsynaptic spikes, and it decreases exponentially as the interval between consecutive pre- or postsynaptic spike increases. The suppression model accounts for the asymmetry of synaptic modification in response to triplets. In the case of a pre-post-pre triplet, the first pair (pre-post) induces potentiation, but the amount of depression induced by the second pair (post-pre) is suppressed by the first presynaptic spike. For a post-pre-post triplet, the first pair (post-pre) induces depression, but the potentiation induced by the second pair (pre-post) is suppressed by the first postsynaptic spike (figure 4.4B).

The parameters of the model were originally set to match the synaptic modification seen in the experiments (Froemke and Dan (2002); Table 4.1). Our numerical simulations with these parameters show that the steady-state distribution is unstable and tightly clustered around the upper bound. When correlations are induced in half of the synaptic inputs, no competition takes place and all the weights are potentiated indiscriminately. To observe a range of behaviors of this model, we set the suppression time constants equal to the values given by Froemke and Dan (2002), namely $\tau_{\text{pre}} = 28$ ms, $\tau_{\text{post}} = 88$ ms. We also set the maximum potentiation and depression values equal ($A_+ = A_- = 0.005$ mV) and fixed the depression time constant ($\tau_- = 20$ ms). We then varied the potentiation time constant τ_+ to observe different behaviors of the model. Transitions to different behaviors can also be seen when changing other parameters (for example the ratio A_+/A_-), but our simulations showed that changing the ratio between the potentiation and depression time constants (τ_+/τ_-) reveals these transitions most clearly.

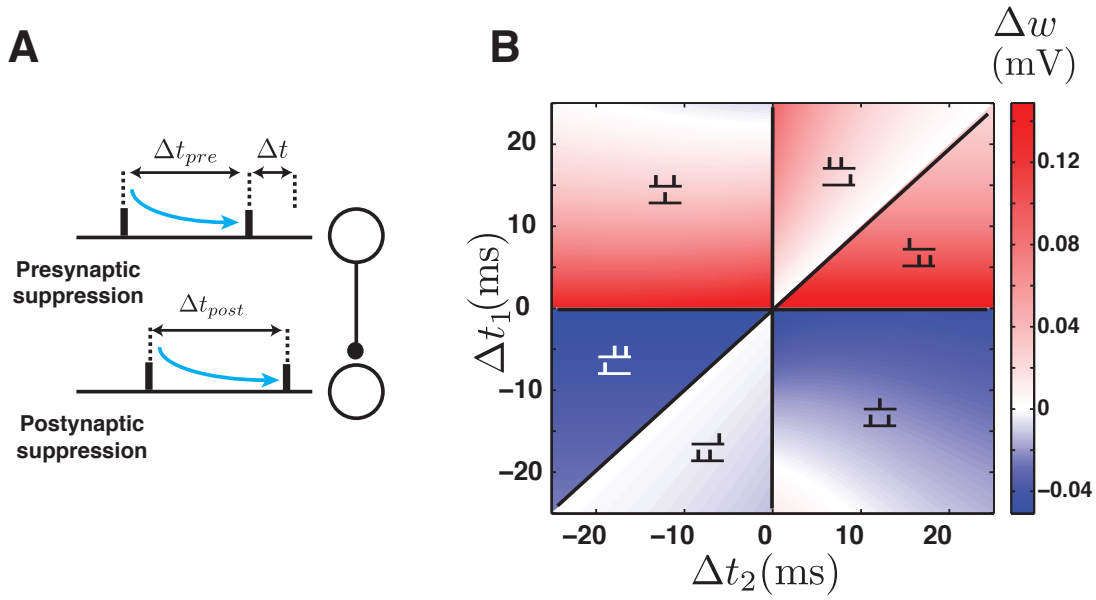

FIG. 4.4: The suppression model. **A.** Schematic of spike interactions in the suppression model, in which the effect of the presynaptic spike in a pair is suppressed by a previous presynatic spike (top), and the effect of the postsynaptic spike is suppressed by a previous postsynaptic spike (bottom). **B.** Plasticity in the suppression model induced by triplets of spikes: pre-post-pre triplets induce potentiation (top left), and post-pre-post induce depression (bottom right).

Calculating the drift of synapses in the suppression model is more complicated than in previous models. We leave the details to Appendix A.3 and report the results here. When $\tau_+/\tau_- < 1.2$, the average synaptic weight has a stable nontrivial fixed point (figure 4.5A-C). For higher values of τ_+/τ_-, the nontrivial fixed point disappears and the average synaptic weigh has only the trivial zero fixed point (figure 4.5A & D). For low τ_+/τ_- values, the steady-state distribution of weights is partially stable and U-shaped, as in the case of the pair-based model (figure 4.5E). However, for τ_+/τ_- between 1.05 and 1.2, the value of the average synaptic weight

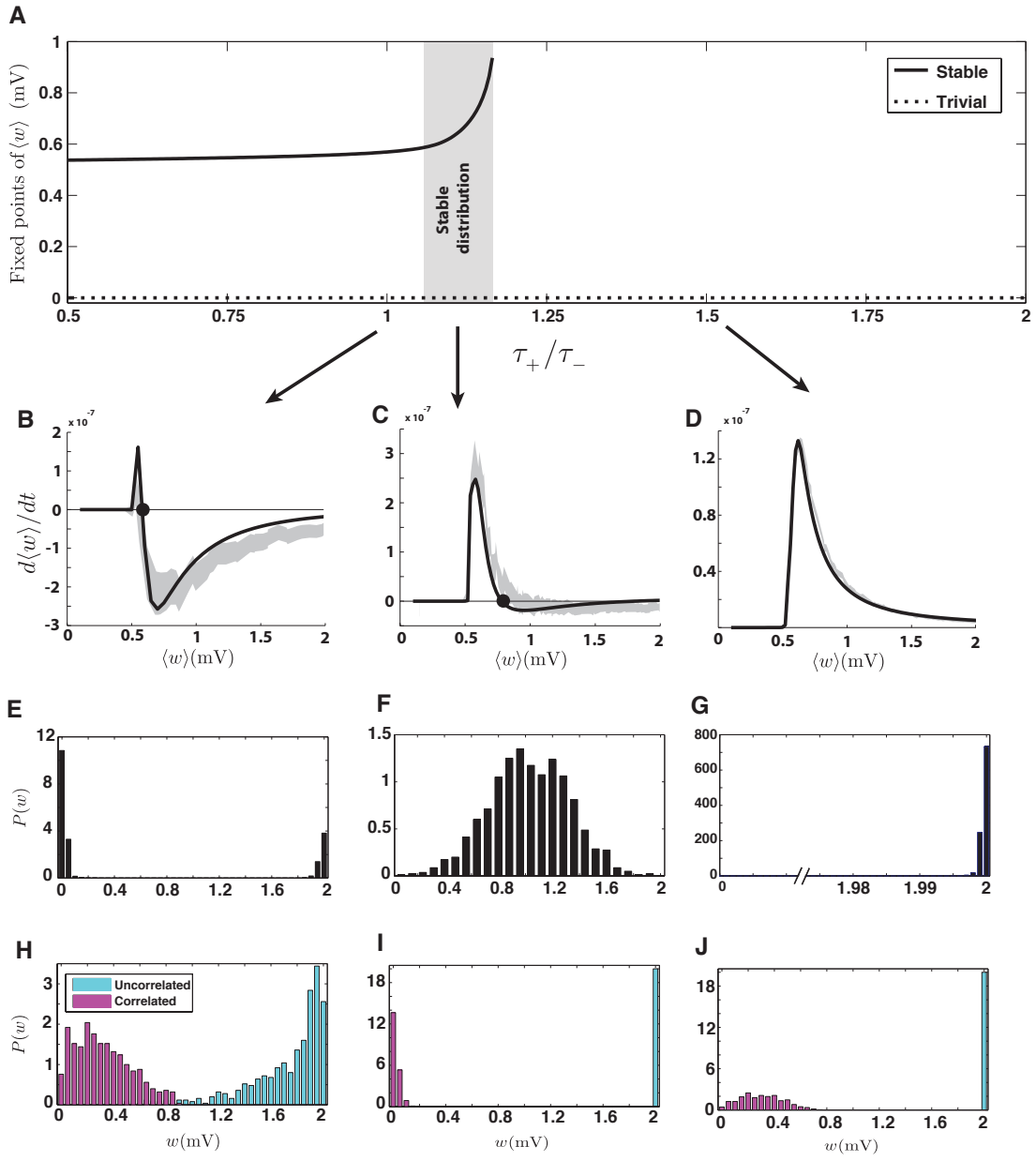

FIG. 4.5: **A.** Fixed points of $\langle w \rangle$ as functions of the ratio between the potentiation and depression time constants. **B.** The average drift when $\tau_+/\tau_- = 1$. Solid curve: analytical result, gray shading: simulation results, filled circle: stable fixed point. **C.** The average drift when $\tau_+/\tau_- = 1.1$. **D.** The average drift when $\tau_+/\tau_- = 1.5$. **E.** The partially stable bimodal steady-state distribution of weights corresponding to the parameters of B. **F.** The stable steady-state distribution of weights corresponding to the parameters of C. **G.** The unstable steady-state distribution of weights clustered around the upper bound corresponding to the parameters of D, when no stable fixed point exists. **H-J.** Competition between correlated and uncorrelated synapses with parameter corresponding to E-G. The competition is anti-Hebbian in all cases.

grows rapidly (figure 4.5A, gray area), and the steady-state distribution is stable and unimodal (figure 4.5F), implying that the w-dependent drift is negative in this range. Because of the complexity of spike interaction in the suppression model, a complete characterization of the w-dependent drift is beyond our analytical calculations (Appendix A.3). However, features of the response of an integrate-and-fire neuron to a pair of presynaptic spikes in the context of the suppression model explain why the w-dependent drift becomes negative when the average synaptic weight is large.

Suppose that two presynaptic spikes arrive at a neuron in quick succession, and we want to analyze the role of the second spike in inducing plasticity under the suppression model (figure 4.6). The second presynaptic spike participates in plasticity twice: once by pairing with the previous postsynaptic spike, and again by pairing with the next postsynaptic spike. When the strength of the synapse is low, the first presynaptic spike is not very likely to induce a postsynaptic action potential after its arrival, so the pairing interval between the second presynaptic spike and the preceding postsynaptic spike is typically long, which induces weak depression (figure 4.6A). However, if the synapse is strong, the first presynaptic spike is likely to induce a postsynaptic action potential, and its pairing interval with the second presynaptic spike is then short, inducing strong depression (figure 4.6B). In addition, because of the high probability of postsynaptic firing in response to both presynaptic spikes, the interval between the induced postsynaptic spikes is short, which strongly suppresses the potentiation caused by pairing the second presynaptic spike with its following postsynaptic spike. Therefore, depression dominates over potentiation in the suppression model when synapses are strong. When

this happens, deviations to even higher values lead to depression. This explains why w-dependent drift is negative when the average synaptic weight is large, which occurs when τ_+/τ_- approaches the critical value 1.2 (figure 4.5A, gray area).

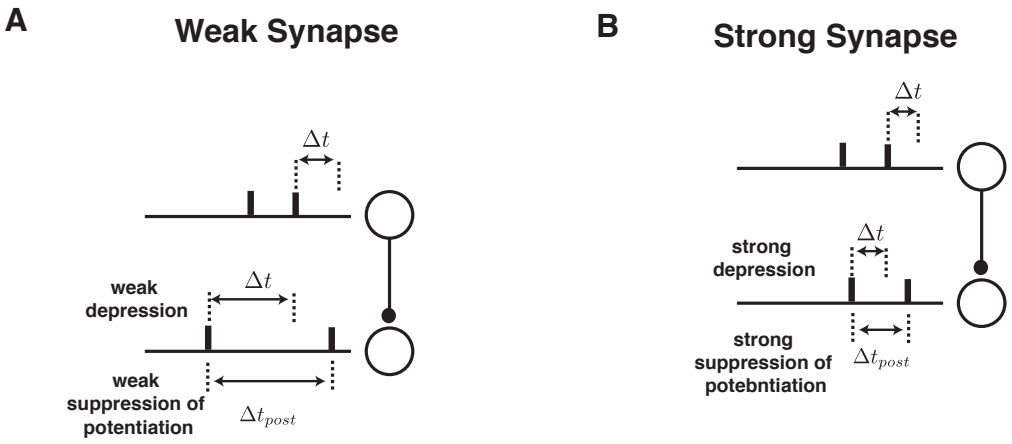

FIG. 4.6: Response of a neuron to a pair of presynaptic spikes and its consequences in the suppression model. **A.** When the synapse is weak, the probability of a postsynaptic spike does not increase significantly from the baseline. The interval between postsynaptic spikes and also the pairing interval between the second presynaptic and the first postsynaptic spike are likely to be long. The result is a weak depression and also a weak suppression of potentiation. **B.** When the synapse is strong, the neuron is likely to fire in response to both presynaptic spikes, which results in strong depression and also strong suppression of potentiation

When half of the synapses receive correlated spike trains and the other half uncorrelated inputs in the suppression model, a distinctive features is that anti-Hebbian competition takes place: the uncorrelated synapses become strong and the correlated ones weak (figures 5H-J). This is the result of postsynaptic suppression. When correlated presynaptic spikes arrive, they tend to induce a postsynaptic spike shortly after their arrival. This make the interval between the induced postsynaptic action potential and the previous spike shorter than for the postsynaptic

response to uncorrelated input. As a result, potentiation is suppressed for correlated synapses, and they eventually lose the competition with uncorrelated ones. In analogy with what was described in the previous paragraph, correlated inputs are similar to inputs with strong synapses and, in either case, the high probability of postsynaptic spiking makes the w-dependent drift negative. In summary, the characteristic properties of suppression model are anti-Hebbian competition and stability of the synaptic distribution when the mean synaptic strength is large.

4.3.3 Stability and competition in the NMDAR-based model

The NMDAR-based model (Senn et al., 2001) was proposed as an explanation for the original STDP experiments of Markram et al. (1997), and it predates both the triplet and suppression models and the data that inspired them. Nevertheless, as we will see below, it has features that resemble both of these models, and it is sensitive to spike interactions beyond pre-post pairings. The original version of the NMDAR-based model (Senn et al., 2001) includes the dynamics of the probability of presynaptic vesicle release. We focus on a simpler version that only models the modification of synaptic strengths by pre- and postsynaptic spikes (Gerstner and Kistler, 2002).

In the NMDAR-based model, the NMDAR is assumed to have three states, rest, up and down. Each incoming presynaptic spike moves a portion of the NMDARs in the rest state into the up state, and each postsynaptic spike transitions a portion of the rest state NMDARs into the down state (figure 4.7A). The NMDAR decays back to the rest state exponentially in the absence of spikes. In accord with the molecular kinetics of NMDARs (Shouval et al., 2002; Malenka and Bear, 2004),

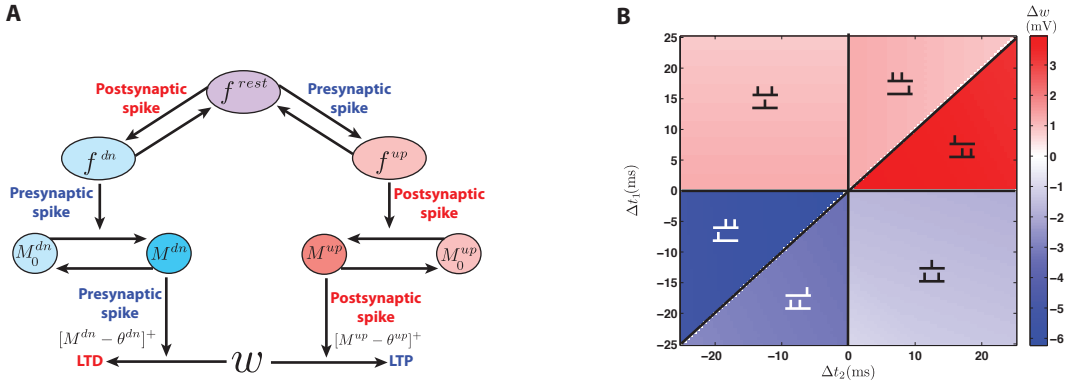

FIG. 4.7: The NMDAR-based model. **A.** Schematic of spike interactions in the NMDAR-based model. The presynaptic spike up-regulates f^{rest}, activates M^{dn} and potentiates the synapse. The postsynaptic spike down-regulates f^{rest}, activates M^{up} and potentiates the synapse. **B.** Plasticity in the NMDAR-based model due to triplets of spikes with parameters as in Table 4.1. The effect is asymmetric, with pre-post-pre triplets inducing potentiation (top left) and post-pre-post depression (bottom right).

the rest state can be interpreted as an NMDAR that is not bound to glutamate and is blocked by Mg^{2+}, the up state as an NMDAR that is bound to glutamate but blocked by Mg^{2+}, and the down state as an NMDAR that is not bound to glutamate but has had its Mg^{2+} block removed by a postsynaptic spike. The model also has two second messengers, called up and down messengers, that mediate potentiation and depression, respectively (figure 4.7A). These can be in either active or inactive states. When a presynaptic spike arrives, a fraction of the inactive down messengers transition to the active state, proportional to the amount of NMDAR in down state. Likewise, when a postsynaptic spike reaches the synapse, it moves a portion of the inactive up messengers into their active state, proportional to the amount of NMDAR in up state. The messengers decay back to their inactive states in the absence of spikes. Finally, upon arrival of a presynaptic spike, the synapse is depressed proportional to the amount of active down messenger, provided that

this is larger than a threshold θ^{dn}. Similarly, each postsynaptic spike causes the synapse to potentiate proportional to the amount of active up messenger provided that it is larger than a threshold θ^{up}. Thus, the presynaptic spike plays three roles in this model: it moves resting NMDARs into the up state, it activates the down messenger, and it induces depression. The postsynaptic spike also has three roles: it transitions resting NMDARs into the down state, activates the up messenger, and induces potentiation (figure 4.7A).

A key feature of the NMDAR based model is that preceding spikes decrease the amount of available resting NMDARs available to upcoming spikes. This implements a mechanism akin to the suppression model, in which previous spikes suppress the effect of subsequent spikes. If we assume that the second messengers are extremely fast such that we could disregard their dynamics, the NMDAR-based model reduces to the suppression model (figure 4.8, top-right). The roles of the second messengers are quite similar to those of the presynaptic depression and postsynaptic potentiation variables in the triplet model in that both integrate the effects of pre- and postsynaptic spiking to modify depression and potentiation. In fact, if we assume that the spikes have access to an unlimited pool of resting NMDARs and messengers, the NMDAR-based model is equivalent to the triplet model (figure 4.8, bottom-left). Finally, if we assume both unlimited messengers and resting NMDARs and also very fast messenger dynamics, the model will be reduced to simple pair-based STDP (figure 4.8, bottom-righ)

Given the multi-spike interactions in the NMDAR-based model, it is not surprising that it responds asymmetrically to triplets of spikes. Our numerical simula-

tions using the parameters provided by Senn et al. (2001) (Table 4.1) show that the synaptic modification in response to triplets in this model is qualitatively similar to that of the suppression model (figure 4.7B). They also show that, with the parameters provided by Senn et al. (2001), the steady-state distribution is unstable and tightly clustered around the upper bound. When correlations are induced in half of the synaptic inputs, no competition takes place and all the weights are potentiated indiscriminately.

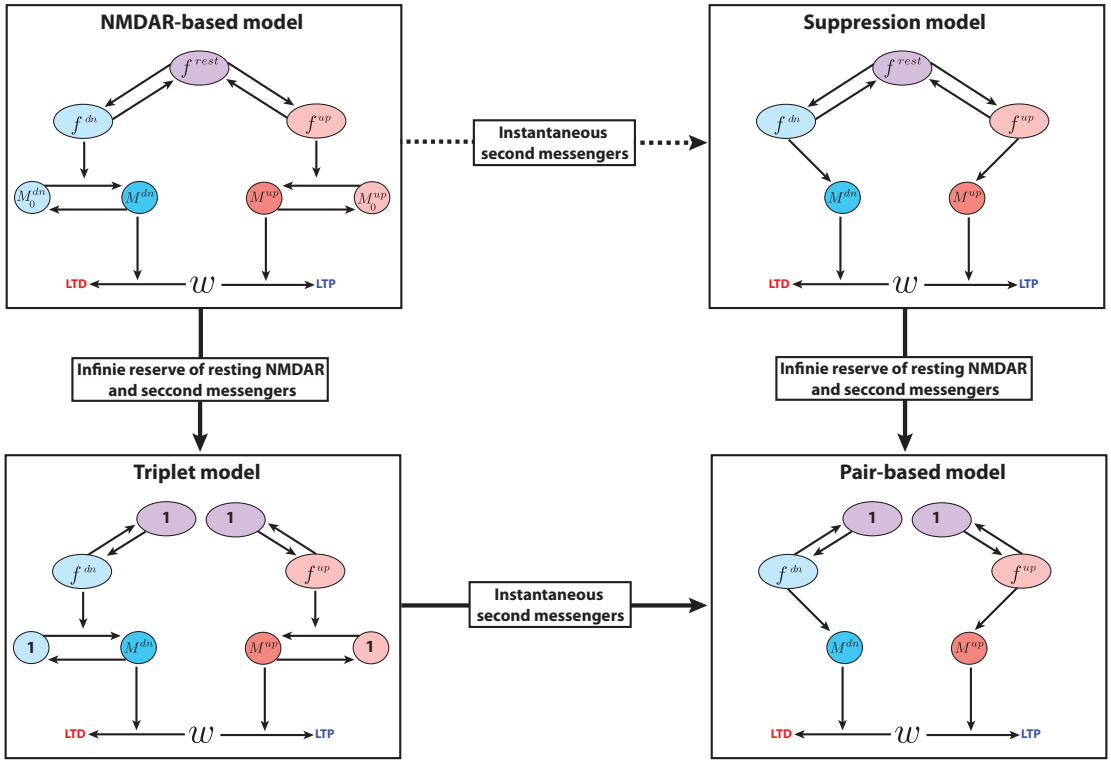

FIG. 4.8: Relationships between multi-spike STDP models. If the second messengers activate instantaneously by NMDRs, the NMDAR-based will be qualitatively equivalent to the suppression model (top, right). If there exists an infinite reserve of resting NMDARs and inactive second messengers, the NMDAR-base model will be reduced to the triplet model (bottom-left). If both assumptions are fulfilled, the NMDAR-based model will be reduced to the simple pair-based model (bottom-right)

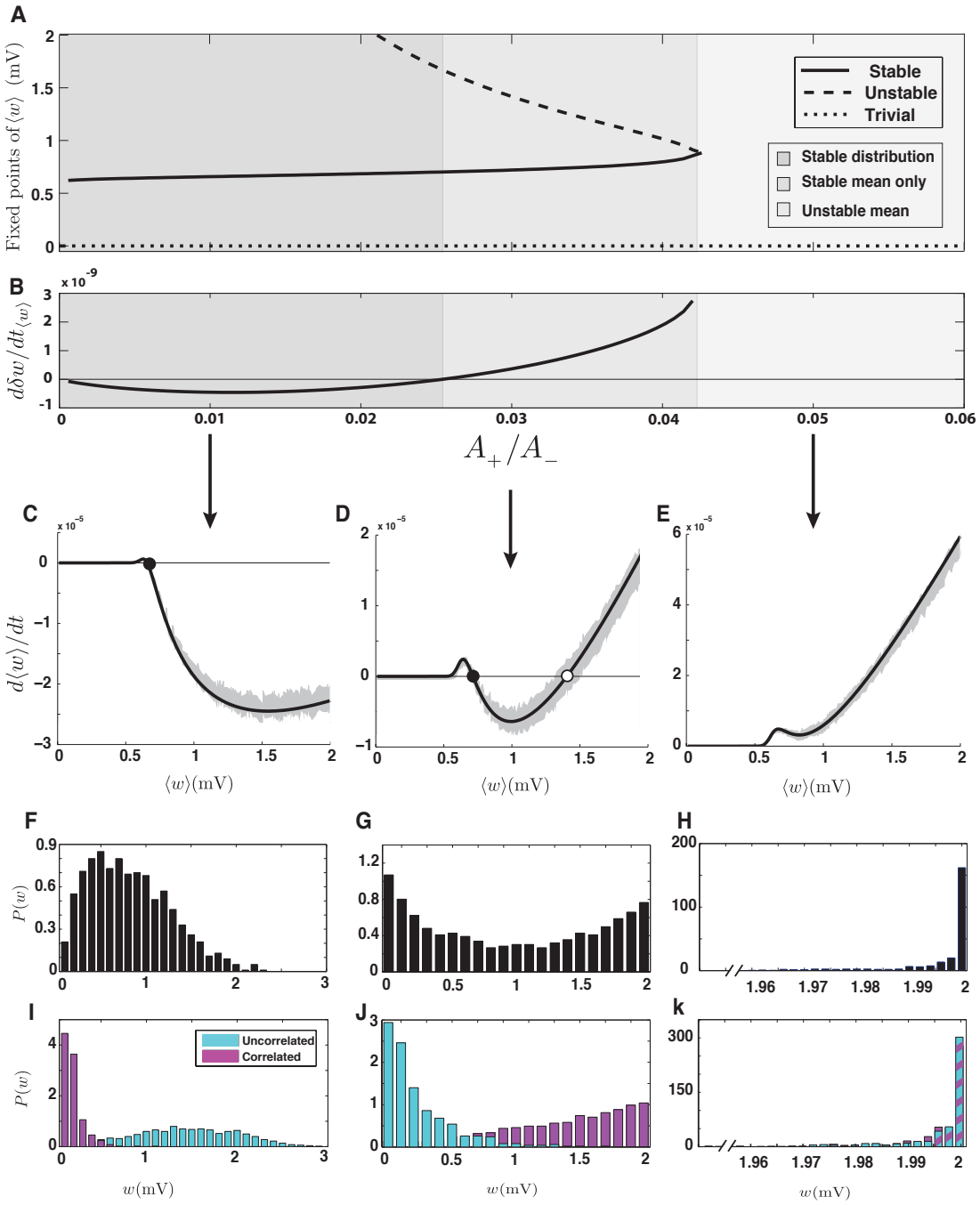

FIG. 4.9: Stability and competition in the NMDAR-based model. **A.** Fixed points of $\langle w \rangle$ as functions of the ratio between the maximum potentiation and depression parameters.**B.** The w-dependent drift at the stable fixed point, as a function of A_+/A_-, which changes sign at $A_+/A_- = 0.025$. **C-E.** The average drift when A_+/A_- is 0.01, 0.03 and 0.05, respectively. Solid curve: analytical result, gray shading: simulation results, filled circle: stable fixed point, open circle: unstable fixed point. **F-H.** The steady-state distributions corresponding to the parameters in C-E. **I-K.** Synaptic competition between correlated and uncorrelated synapses corresponding to the parameters in C-E. Parameters are $A_f^{\rm up} = A_f^{\rm dn} = A_M^{\rm up} = A_M^{\rm dn} = 0.1$, and the time constants are as in Table 4.1.

To examine the spectrum of behaviors in the NMDAR-based model, we calculated the synaptic drift (Appendix A.4). Interesting transitions into different regimes occur when the threshold of the down messenger is larger than that of the up messenger ($\theta^{\mathrm{dn}} = 0.2, \theta^{\mathrm{up}} = 0$), and the ratio between maximum potentiation and maximum depression (A_+/A_-) is varied (figure 4.9). All other parameters of the model are held constant at equal values for potentiation and depression components, and the time constants are set to the values provided by Senn et al. (2001)(Table 4.1). When A_+/A_- is smaller than a critical value (0.042), the average synaptic weight has both stable and unstable nontrivial fixed points. At the critical value, these two fixed points coalesce and disappear, and beyond the critical value the average synaptic weight has only the trivial fixed point at zero (figure 4.9A). The sign of w-dependent drift also changes as A_+/A_- varies. When A_+/A_- is smaller than 0.025, the w-dependent drift is negative, and for larger ratios it is positive (figure 4.9B). Taken together, three different behaviors are observed in the NMDAR-based model: 1) When a stable mean synaptic weight exists and w-dependent drift is negative ($0 < A_+/A_- < 0.025$, figures 4.9A & B, dark gray area), the steady-state distribution of synaptic weights is stable and unimodal (figures 4.9C & F). 2) When a stable mean synaptic weight exists and w-dependent drift is positive ($0.025 < A_+/A_- < 0.042$, figures 4.9A & B, light gray area), the steady-state distribution of synaptic weights is partially stable and U-shaped (figures 4.9D & G). 3) When the mean synaptic weight has no stable fixed point ($A_+/A_- > 0.042$), the steady state distribution is unstable, and it clusters near the upper bound (figures 4.9E & H).

Synaptic competition is different in these three regions of the parameter space.

When half of the input spike trains are correlated, the competition in the first region is anti-Hebbian because the w-dependent drift is negative and correlated synapses receive more depression (figure 4.9I). It is Hebbian in the second region because the w-dependent drift is positive (figure 4.9J). There is no competition in the third region because the mean is not stable (figure 4.9K). In short, the distinguishing features of the NMDAR-based model compared to the pair-based model are the possibility of a stable synaptic distribution and anti-Hebbian competition when the maximum depression is significantly larger than the maximum potentiation.

4.3.4 STDP with soft bounds

In previous sections, we imposed hard bounds on the synaptic strengths to confine them between zero and a maximum allowed value (w_{max}). As in the case of the pair-based model, it is also possible to confine the synapse by implementing soft bounds, that is, by making the maximum depression and potentiation weight-dependent so that when a synaptic strength approaches the bounds, its rate of change gradually decreases. This can be done by multiplying A_+ and A_- by $1-(w/w_{max})$ and w/w_{max} respectively. In the case of the triplet model, the presynaptic depression and postsynaptic potentiation variablea should be also multiplied by $1-(w/w_{max})$ and w/w_{max} respectively, bescause they appear as potentiation and depression factors as well.

The steady-state distribution of synaptic strengths is stable and unimodal for all three multi-spike STDP models with soft bounds (figure 4.10). This behavior is robust and holds for a wide range of parameters (we only show simulation results

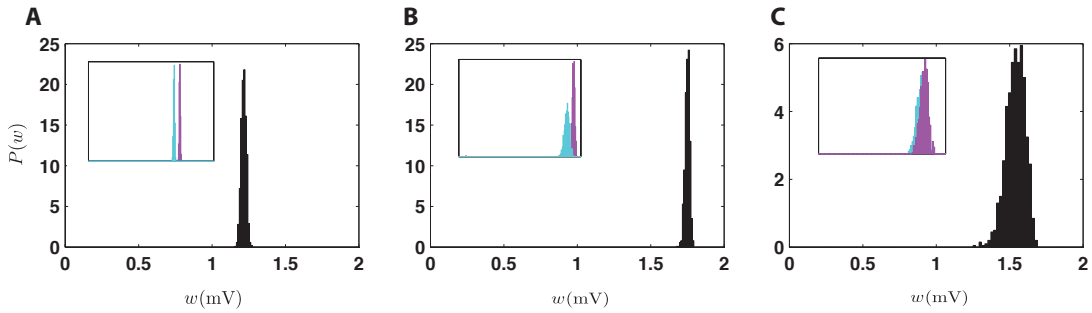

FIG. 4.10: Stability and competition in multi-spike STDP models with soft bounds. **A.** Steady-state distribution of synaptic strengths in the triplet model with soft bounds. **B.** Steady-state distribution of synaptic strengths in the suppression model with soft bounds. **C.** Steady-state distribution of synaptic strengths in the NMDAR-based model with soft bounds. Insets: steady-state distribution of weights when half of the synapses receive correlated input (magenta) and the other half receive uncorrelated input (cyan). In each case, the original parameters (Table 4.1) are used.

for the original parameters (Table 4.1) in each model). The soft bounds weaken synaptic competition drastically, so than the distributions of correlated and uncorrelated synapses are close to each other (figure 4.10, insets). As has been shown for pair-based STDP (Chapter 2, van Rossum et al. (2000); Rubin et al. (2001)), soft bounds turn STDP into a homeostatic plasticity mechanism with minimal sensitivity to the correlation structure of the external input.

4.4 Discussion

The main focus of this study was synaptic stability and competition, two desirable but often conflicting features of activity-dependent plasticity rules (Miller, 1996). Our analytical tool for assessing these properties was calculating the drift of a population of synapses under each multi-spike STDP model. This method has been applied to the pair-based STDP model in a number of previous studies (see Chapter

2). The pair-based model with hard bounds was shown to produce a partially stable U-shaped steady-state distribution of weights and Hebbian competition that favors correlated synapses over uncorrelated ones (Song et al., 2000; Song and Abbott, 2001; Cateau and Fukai, 2003). On the other hand, the pair-base model with soft bounds has been shown to have a stable steady-state distribution at the expense of losing synaptic competition and sensitivity to input correlations (van Rossum et al., 2000; Rubin et al., 2001). Our analysis in this chapter can be viewed as a reconfirmation and extension of these results into the domain of multi-spike STDP models.

Table 4.2 summarizes the results of this survey of stability and competition in multi-spike STDP models, together with the results for shifted STDP from the previous chapter. Like the pair-based model, the triplet model produces an unstable steady-state distribution of synaptic weights and Hebbian synaptic competition. The suppression model shows predominantly anti-Hebbian competition and a stable steady-state distribution of synaptic weights when the average weight is high. The NMDAR-based model displays both stable and unstable steady-state distributions depending on the parameters, with anti-Hebbian competition in the former case and Hebbian in the latter.

It is interesting that anti-Hebbian plasticity, which does not occur in the pair-based model unless a temporal shift is added (Babadi and Abbott, 2010), can arise in the suppression and NMDAR-based models. The dichotomy between full synaptic stability and Hebbian competition persists in multi-spike STDP models, and anti-Hebbian competition can coexist with full synaptic stability for at least at

	Stability	Partial stability	Hebbian competition	Anti-Hebbian competition
Pair-based	-	✓	✓	-
Shifted	✓	-	✓	✓
Triplet	-	✓	✓	-
Suppression	✓	✓	-	✓
NMDAR-based	✓*	✓	✓	✓

Table 4.2: Summary of stability/plasticity in STDP models. * Stability only coexists with anti-Hebbian competition is NMDAR-based model.

some parameter regimes in the suppression and NMDAR-based models. A number of solutions have been proposed to reconcile stability and competition in pair-based STDP. By interpolating between hard and soft bounds, it is possible to obtain both synaptic competition and stability, but over a limited parameter range (Gütig et al., 2003). Another solution, based on a small temporal shift in the STDP window as described in Chapter 3, can stabilize the distribution of synaptic weights while maintaining competitiveness. As we saw in this chapter, the shift has a similar effect in the triplet model (Babadi and Abbott, 2010).

To search the parameter space of the models for different stability/plasticity interplay, we systematically varied the balance between potentiation and depression parameters in each multi-spike STDP model. However, for each model, a fixed set of parameters was originally proposed to match experimental results. Our paremeter changes may cause the response profile of the model to deviate from its originally optimized form. This can be justified because both the temporal spread and the magnitude of potentiation and depression vary considerably along the dendritic tree (Froemke et al., 2005; Letzkus et al., 2006; Froemke et al., 2010). Therefore, each parameter set in our analyses and numerical simulations might coincide with the

characteristics of the STDP window at a particular location on the dendritic tree.

Although all of the models we considered were proposed on basis of observations of synaptic modification under similar conditions, their effect on a population of synapses onto a postsynaptic neuron can be quite different. One of the useful computational aspects of STDP is its ability to implement Hebbian learning and to functionally organize neural circuits. None of the three multi-spike models generated Hebbian competition among synapses when their original optimized parameters were used. Moreover, using these parameters, all three models produced an unstable distribution of weights tightly clustered near the upper bound of their allowed range. Given the observed broad distribution of synaptic weights that has been observed (Turrigiano et al., 1998; O'Brien et al., 1998; Song et al., 2005) this is implausible. As it is possible to construct several phenomenological models that explain a given experimental data set, it is reasonable to use the effects of plasticity at the population level (evaluated through simulations or analytical calculations) as a criteria for selecting an appropriate model.

It should be possible to unify different multi-spike STDP model into a single framework. The triplet and suppression models were motivated by different experimental data sets that showed opposite synaptic modification in response to triplets (Froemke and Dan (2002) vs. Wang et al. (2005)). However, the NMDAR-based model, which is phenomenologically closer to the molecular machinery involved in synaptic modification, can match the effects of either of these models, depending on the parameters used. This leads to the possibility that both the triplet and suppression models might arise from a single biophysical mechanism with different

parameters in different neural circuits (Shouval and Wittenberg, 2010). A simpler model that similarly encompasses both cases would be useful.

In conclusion, as long as the experimental methods are confided to observation of plasticity at a single synapse, we propose more stringent criteria in constructing phenomenological models, to take into account the behavior of the model at the level of synaptic populations along with the underlying molecular mechanisms.

Chapter 5

Structures arising from STDP in a network

The whole however, is merely the essential nature of parts reaching its completeness through the process of its own development.
Georg Wilhelm Friedrich Hegel

5.1 Introduction

STDP is inherently a local synaptic modification rule because its sole determinants of synaptic modification are the timings of pre- and postsynaptic spikes. Cortical neurons, on the other hand, are embedded in highly interconnected networks in which each neuron receives up to ten thousand synapses from other neurons (Braitenberg and Schuz, 1991; Stepanyants et al., 2008). Thus, although STDP operates locally, the synapses in a neural network do not change in isolation. In previous chapters, I studied the effect of STDP in shaping the distribution of weights for a population of synapses converging onto a single neuron. This is one step toward bridging the gap between the locality of STDP and the global structures that it generates. In this chapter, I take a further step beyond the level of a single neuron

and extend the focus to a recurrent network of spiking neurons by surveying the global structures that arise from this local synaptic modification mechanism within a large network.

The antisymmetric form of the STDP window led to the proposal that this synaptic modification rule should eliminate strong recurrent connections between two neurons (Abbott and Nelson, 2000). This principle has recently been expanded by Kozloski and Cecchi (2010) to larger polysynaptic loops in the case of a "balanced" STDP window in which the magnitudes of potentiation and depression are equal. These authors have also shown that balanced STDP organizes the neurons in a network into in- and out-hubs. Here, I show that, surprisingly, all of the above network properties can be explained through an understanding of the effect of STDP on pairwise interactions of neurons. This provides an analytically tractable way of relating the structures arising in a network to properties of the STDP model being used, even beyond the simple balanced case. The results show that the loop-elimination property of STDP without a shifted window lessens as potentiation dominates over depression. STDP with a shifted window (Chapter 3), on the other hand, can operate as a loop-generating mechanism and also has homeostatic properties that maintain the average value of the synaptic weights across a network.

5.2 Methods

5.2.1 Network model

A network of $N_{\rm ex} = 1000$ excitatory and $N_{\rm in} = 250$ inhibitory leaky integrate-and-fire (LIF) neurons was simulated. The dynamics and parameters of the LIF neurons are the same as in previous chapters (see Chapter 2, table 1). Each neuron receives excitatory and inhibitory inputs from all the other neurons in the network. The strengths of the excitatory-to-inhibitory, inhibitory-to-excitatory and inhibitory-to-inhibitory synapses are fixed. At the beginning of each simulation, their strengths are drawn from uniform distributions with means $\langle w_{\rm ex\to in}\rangle = 2$ mV, $\langle w_{\rm in\to ex}\rangle = 8$ mV, and $\langle w_{\rm in\to in}\rangle = 8$ mV respectively. The excitatory-to-excitatory connections are modified by pair-based STDP as described in Chapters 2 and 3. They are also initialized at the beginning of each run to random values from a uniform distribution with mean $\langle w_{\rm ex\to ex}\rangle = 2$ mV and range $w_{\rm min} = 0$ to $w_{\rm max} = 4$ mV. These values are chosen to establish an excitation/inhibition balance in the initial network. Self connections are prohibited for all neurons.

In addition to synaptic inputs originating from the neurons within the network, each neuron receives uncorrelated noise as an external input. Taken together, the input to the $i^{\rm th}$ excitatory or inhibitory neuron can be described by

$$\frac{dI_i}{dt} = -\frac{I_i}{\tau_{\rm s}} + \sum_{j=1, j\neq i}^{N_{\rm ex}, N_{\rm in}} w_{ij} \sum_k \delta(t - t_j^k) + \mu + \sigma\,\xi(t)\,. \tag{5.1}$$

Here, the first sum runs over all neurons ($N_{\rm ex}$ and $N_{\rm in}$ for excitatory and inhibitory populations, respectively). The second sum runs over all of the spike times t_j^k

produced by neuron j, indexed by k. In addition, w_{ij} is the synaptic strength from neuron j to neuron i, μ and σ are determine the mean and variability of the external input, and $\xi(t)$ is zero mean white noise. The parameter σ was set to 22 mV/$\sqrt{\text{ms}}$ to provide an average initial baseline firing rate of 10 Hz for the neurons in the network when μ is zero. In the simulations, the value of μ is changed systematically to modify the initial baseline firing rate, which connects the average initial firing rate of the network to the strength of the external input. The synaptic time constant $\tau_{\text{s}} = 5\,\text{mV}$ is taken to be the same for excitatory and inhibitory synapses. Each simulation is run until the excitatory-to-excitatory connections reach a steady-state in which the average firing rate, and the mean and variance of the weights remain constant.

5.2.2 Counting the loops

To count the number of closed loops (or recurrent connections) in the matrix of excitatory-to-excitatory synaptic weights (W), we first turn the network into a directed graph. This is done by comparing each synaptic weight to a threshold value h, and assigning the value 1 to the synapse if its weight is greater than or equal to h, and assigning a zero otherwise. This defines the adjacency matrix M of the resultant directed graph, which can be written formally as

$$M = \Theta(W - h)\,, \tag{5.2}$$

where $\Theta(x)$ is the Heaviside step function. The number of closed loops of length n in the adjacency matrix M is

$$L_n = \frac{\text{tr}(M^n)}{n}\,, \tag{5.3}$$

where tr denotes the matrix trace (the sum of the diagonal elements). To evaluate the degree of recurrence in a network, we compare the number of closed loops obtained from the above equation with the number in a randomly permuted (shuffled) version of the same matrix. This method distinguishes between recurrent connections formed purely by chance and those that arise from the structure of the connectivity matrix. In the following sections, whenever we mention the number of loops in a network, we are in fact referring to the number of loops in the adjacency matrix formed by turning the network into a directed graph as described above.

5.3 Results

Pair-based STDP is parameterized by the maximum values of potentiation and depression (A_+ and A_-, respectively) and by the time constants of the windows for potentiation and depression (τ_+ and τ_-, respectively). Except for the case of shifted STDP, we assume that spike interactions are all-to-all, i.e. all possible pre-post pairs, and not only the nearest neighbors, contribute to plasticity. Using an STDP model with nearest-neighbor interactions does not qualitatively alter the results obtained below (see Appendix A.7). To gain analytical insight into the structures that arise from STDP in a network, we first consider the simplest possible network element; a pair of connected neurons (figure 5.1). As will be shown in the following sections, many of the structures induced by STDP in a large network can be explained by properties of this simple two-neuron system.

Assume that two representative excitatory neurons from a network, labeled neuron 1 and neuron 2, are reciprocally connected (figure 5.1, middle). We denote

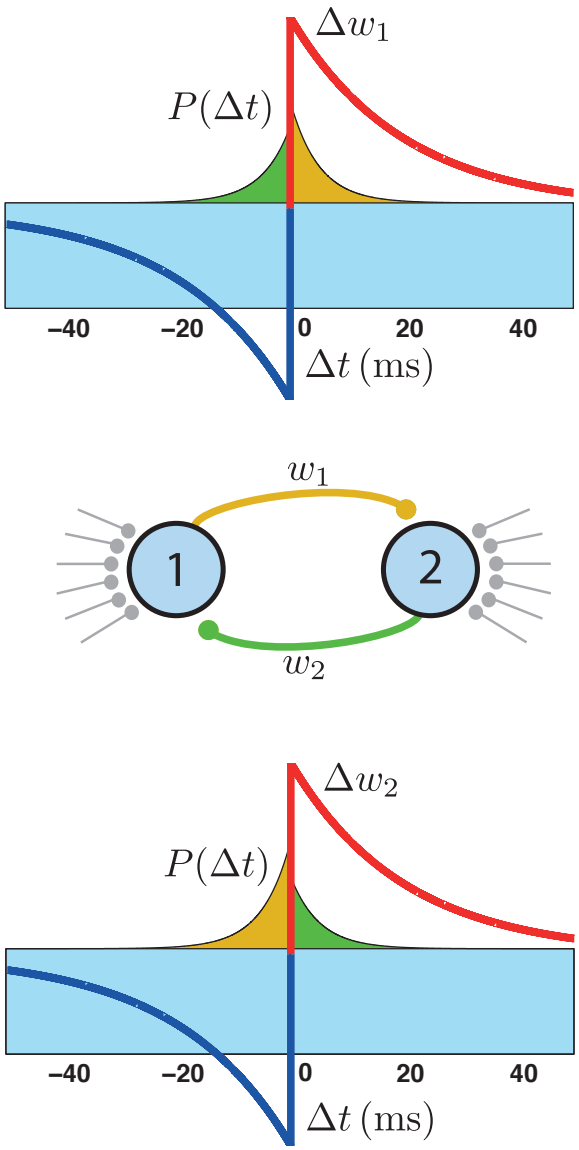

FIG. 5.1: Pairwise interactions of neurons through reciprocal synapses. Two representative neurons embedded in a network are shown with their reciprocal synapses (middle). Both synapses are modified by the pairings of the baseline spike trains of the neurons (blue area in top and bottom panels). Whenever neuron 1 fires, the synapse w_1 (yellow) induces a transient increase in the firing rate of neuron 2 (yellow areas in top and bottom panels). This transient firing rate increase potentiates w_1 because it falls into the potentiation domain of STDP (top), but depresses w_2 (green), as it falls into depression domain for this synapse (bottom). The transient increase in the firing rate of neuron 1 in response to spikes of neuron 2 (green areas in top and bottom panels) has the opposite effect.

the strength of the synapse from neuron 1 to neuron 2 as w_1 and the strength of the synapse from 2 to 1 as w_2. Because the two neurons receive uncorrelated external input and are embedded in a larger network, each fires at a constant baseline rate ($\bar{r}_1$ and $\bar{r}_2$, respectively). We assume that we can disregard correlations between the baseline spike trains of neurons 1 and 2 (Renart et al., 2010), but our results hold as long as any such correlations are symmetric functions of time. The two synapses are modified primarily by random pre-post pairings of their baseline spike trains. The average amount of modification is the same for both synapses. On top of the baseline firing, the reciprocal synaptic connections induce correlations into the spike trains of the two neurons. Each spike arriving from neuron 1 to neuron 2 transiently increases the firing rate of neuron 2 proportional to w_1 (figure 5.1, yellow areas). This transient increase (or causal bump, as we called it in previous chapters) on the one hand induces potentiation in w_1 (figure 5.1, top) and, on the other hand, depresses w_2 (figure 5.1, bottom). Likewise, the causal bump induced by neuron 2 into neuron 1 (figure 5.1, green areas) potentiates w_2 and depresses w_1. Taken together, the average drift of the synaptic pair can be expressed as

$$\begin{aligned} \frac{dw_1}{dt} &= A\,\bar{r}_1\,w_1 - B\,\bar{r}_2\,w_2 + C\,\bar{r}_1\,\bar{r}_2 \\ \\ \frac{dw_2}{dt} &= A\,\bar{r}_2\,w_2 - B\,\bar{r}_1\,w_1 + C\,\bar{r}_1\,\bar{r}_2\,, \end{aligned} \tag{5.4}$$

where the coefficients A, B and C can be calculated from the parameters of the neural and plasticity models (see Appendix A.5). The coefficient A represents the potentiation induced in a synaptic weight by its own causal effect on postsynaptic firing, B represents the depression induced by the causal effect of the other the

other synapse, and C characterizes the synaptic modification due to random pairings of the baseline spike trains of the two neurons. The drift of the synaptic pair due to the above linear system is restricted to a limited region in the state space, because the synaptic strengths are bounded between zero and w_{max}. As we will see in the following sections, this restriction results in a number of interesting effects beyond those expected for a simple linear system.

In what follows, we first examine the effect of different parameterizations of the STDP window on the synaptic pair. This leads to a number of predictions about the structures arising from STDP in networks. We then test each prediction with numerical simulations of a large network. The time constants of the STDP window are assumed to be equal ($\tau_+ = \tau_- = 20$ ms), but we vary the balance between potentiation and depression by changing the maximum values A_+ and A_-. The same qualitative results hold when the maximum values are set equal and the potentiation/depression balance is modified by changing the time constants.

5.3.1 Balanced STDP

The simplest form of STDP that we consider is balanced with equal potentiation and depression domains, i.e. with $A_+ = A_- = 0.005$ mV. In this case, the coefficient C vanishes because the baseline potentiation and depression cancel. In addition, the coefficients A and B are equal as is the amount of potentiation and depression induced by pre-post pairs and post-pre pairs. These conditions greatly simplify the system of equations (5.4). When the baseline firing rates are equal ($\bar{r}_1 = \bar{r}_2$), the values of the synaptic weights do not change when w_1 and w_2 are equal, i.e. the synaptic drift is zero on the line $w_1 = w_2$ (figure 5.2A, solid line).

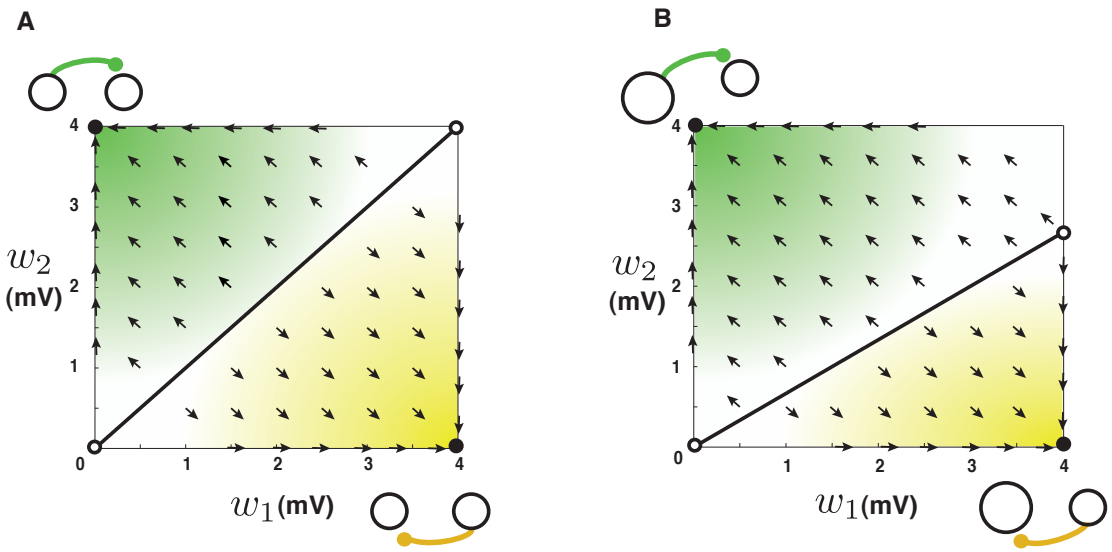

FIG. 5.2: Dynamics of reciprocal synapses when STDP is balanced. **A.** When the baseline firing rates of the two neurons are the same (10 Hz), synapses that have the same weight remain at equilibrium (solid diagonal line). When one of the synapses is initially larger than the other, it grows while the smaller one shrinks until they hit the boundaries and eventually settle into the attractors at the bottom right or top left. Attractors are depicted as filled circles and unstable fixed points as open circles. The attractors correspond to unidirectional connection as depicted schematically. The arrows show the flow of trajectories, obtained by numerical evaluation of equation (5.4) at each point. **B.** When the baseline firing rates are not the same (10 Hz vs. 15 Hz), the line of equilibria becomes tilted. The neuron with the higher rate (depicted larger) is more likely to send out a unidirectional synapse.

However, this equilibrium is unstable. If the two synapses have unequal strengths, the stronger synapse grows even stronger and weakens the other synapse until they reach the boundary of their allowed range (figure 5.2A, arrows). Then, the synapses continue their dynamics along the boundary edge until they reach the upper-left ($w_1 = 0, w_2 = w_{max}$) or lower-right ($w_1 = w_{max}, w_2 = 0$) corner of the state space (figure 5.2A, filled circles), depending on which synapse was stronger to begin with. These "attractors" of the synaptic dynamics indicate that, at steady-state, loops

between pairs of neurons are eliminated by this form of STDP. Note that a linear system of differential equation cannot have more than one attractor. The appearance of two attractors is a consequence of restricting the dynamics to a limited range. The existence of these attractors suggests that STDP favors unidirectional connections and eliminates loops in a network (figure 5.3A), as was shown by Kozloski and Cecchi (2010).

When the baseline rates of the two neurons are not equal, the line of equilibrium is tilted (figure 5.2B, see Appendix A.6). As a result, the size of the basins of the two attractors differ, and the neuron with the higher firing rate is more likely to strengthen its outgoing synapse and weaken its incoming synapse (figure 5.2B, top-left corner). Conversely, the neuron with the lower firing rate is more likely to weaken its outgoing synapse and strengthen its incoming synapse (figure 5.2B, bottom-right corner). If we generalize this behavior to the context of a network, an important prediction can be made: neurons with low firing rates attract strong excitatory synapses to themselves but project weaker synapse onto other neurons. Neurons with high firing rates experience the opposite trend; they lose incoming synaptic input through synaptic weakening. At the same time, the synapses they send to the other, typically slower firing neurons of the network tend to get stronger. Therefore, STDP organizes network neurons into in- and out-hubs, as has been shown by Kozloski and Cecchi (2010). Our simulation results also confirm this prediction (figure 5.3B). Also, since the weakening/strengthening effects act in opposite directions in neurons with low and high initial firing rates, it is expected to see an inverse relationship between the strength of incoming and outgoing synapses to each neuron in the steady-state connectivity matrix (figure 5.3B).

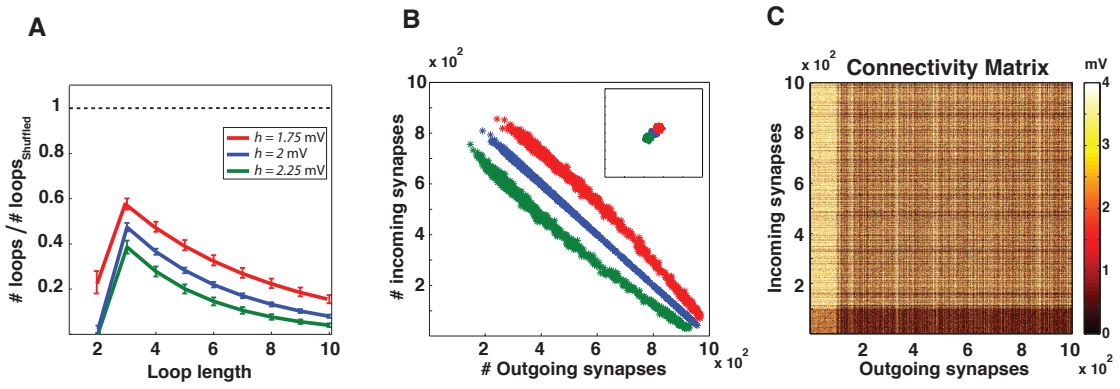

FIG. 5.3: Simulation results for a network with balanced STDP. **A.** The number of loops in the steady-state weight matrix, divided by the number of loops in a shuffled version of this matrix, as a function of the length of the loop. The initial firing rate of the network was 10 Hz. Error bars depict the standard deviations from using 10 different shuffled versions. Different colors correspond to different threshold values h for counting the loops (Methods). The ratios are less than one for all loop lengths, so STDP decrease the number of loops from the chance level. **B.** The number of outgoing synapses plotted against the number of incoming synapses in the steady-state connectivity matrix. Each marker depicts a neuron. Different colors correspond to different threshold values h for counting the synapses. Inset shows the same plot for a shuffled version of the steady-state weight matrix. **C.** The steady-state weight matrix when neurons 1-100 receive stronger external input and initially fire at 12 Hz, while the rest of the neurons fire initially at 10 Hz. Neurons 1-100 send out more synapses to the network (light vertical band) and receive less synapses (dark horizontal band), so they have turned into out-hubs.

Another related prediction is that, if the external input is biased to give a sub-population of excitatory neurons an initially higher firing rate than the rest of the network, these neurons will become out-hubs through STDP. We tested this by increasing the mean of the external input to the first 100 of 1000 excitatory neurons in a network, giving them a 20% higher initial rate (12 Hz in our example, compared to a baseline rate of 10 Hz for the rest of the network). The results show that this sub-population indeed turns into out-hubs once the synaptic weights reach steady

state (figure 5.3C). This also implies that these high-firing neurons are is insulated from the rest of the network, because their incoming connections are weakened.

5.3.2 STDP with dominant potentiation

When the potentiation/depression balance is tipped in favor of potentiation ($A_+ = 0.0055$ mV and $A_- = 0.0050$ mV in our examples), the coefficient A in equation (5.4) becomes larger than B (see Appendix A.6). In addition, the baseline parameter C is positive. By setting the right-hand-sides of equations (5.4) to zero, the fixed point for the values of the two synaptic weights can be determined to be at $\bar{w}_1 = Cr_2/(B - A)$ and $\bar{w}_2 = Cr_1/(B - A)$. Both of these values are negative, so the fixed point lies out of the allowed range of synaptic strengths. Furthermore, this fixed point is unstable, so the weights tend to drift away from it (figure 5.4; see also Appendix A.6).

We now examine the influence of the outlying fixed point on the dynamics within the allowed region of synaptic values when the baseline firing rates of the two neurons are equal. If the initial weights are fairly close to each other (figure 5.4A, red area), they eventually end up at the attractor in the upper-right corner of the phase space due to repulsion from the outlying fixed point. The attractor on the upper-right corner corresponds to strong recurrent connections. Trajectories of weights that hit the upper boundary ($w_2 = w_{max}$) perpendicularly, form another fixed point that is unstable (figure 5.4, open circles on top). Trajectories to the left of this critical line are eventually absorbed by the attractor at the top-left corner (corresponding to a unidirectional connection), while trajectories to its right are absorbed by the top-right attractor (corresponding to recurrent connections). A

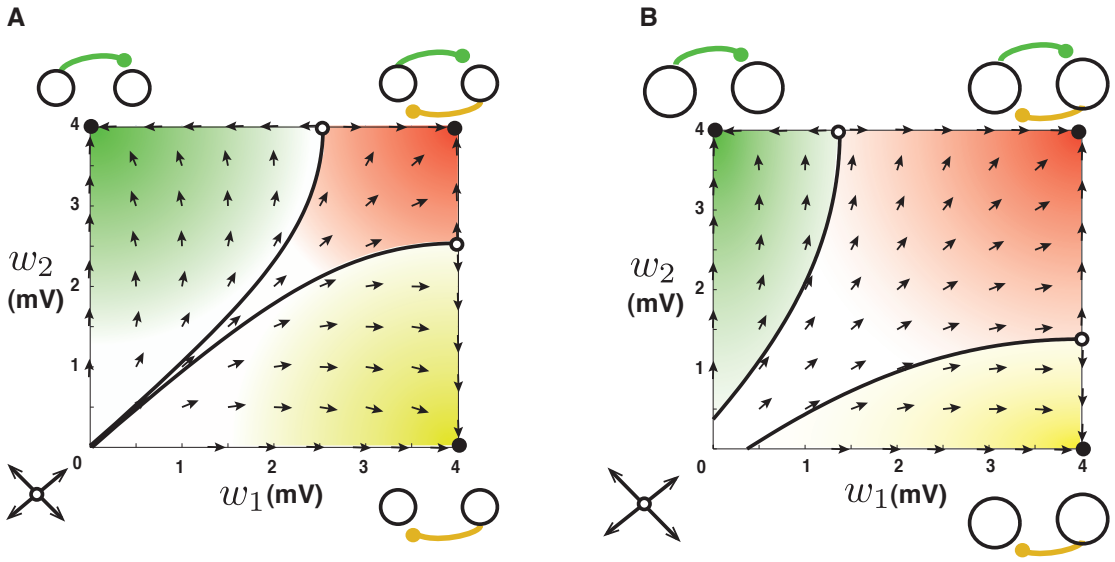

FIG. 5.4: Dynamics of reciprocal synapses when STDP is potentiation-dominated. **A.** When the baseline firing rates of the two neurons are both 10 Hz, an unstable fixed point exist out of the allowed range, schematically illustrated at the bottom left. Arrows show that the trajectories drift away from this outlying fixed point. Initial conditions starting within the red area end up in the attractor at the top-right corner, which corresponds to recurrent connections. Trajectories that hit the boundaries perpendicularly delineate the borders of the basins of attraction (solid curves). Initial conditions in the yellow area end up at the attractor at the bottom right, corresponding to a unidirectional connection from neuron 1 to neuron 2. Initial conditions within the green area go to the attractor at top left, corresponding to a unidirectional connection from neuron 2 to neuron 1. **B.** The same as A when the baseline firing rates are 15 Hz. The basin of attraction for recurrent connections (red area) becomes larger when the baseline firing rate increases.

similar unstable fixed point exists on the rightmost boundary ($w_1 = w_{max}$; figure 5.4, open circles on the right). As a result, the state-space of the weights is partitioned into three basins of attraction: one leading to the attractor corresponding to recurrent connections (figure 5.4, red shading) and the others to attractors that produce unidirectional connections (yellow and green shadings).

The appearance of the attractor corresponding to recurrent connections leads

to a prediction about networks: STDP with dominant potentiation is not as efficient as balanced STDP at eliminating loops in a network, because it maintains some of the recurrent connections. This prediction is confirmed by our numerical simulations showing that there are more loops in the steady-state weight matrix of a network in this case (compare figures 5.5A and 5.3A).

As the baseline firing rates of the neurons increase, the basin for the attractor with recurrent connections expands (figure 5.4B, red area). This leads to the prediction that when a network is driven by stronger external input and consequently has a higher initial average firing rate, it will have more loops in its steady-state weight matrix. Numerical simulation confirms this observation (figure 5.5B). To quantify the degree of recurrence in the network, we define a "recurrence index" as the sum of the number of loops with less than 10 synapses, divided by the sum of similar loops in a shuffled version of the network (see Methods). Simulation results show that the recurrence index increases rather abruptly when the initial rate of the network exceeds 70 Hz (figure 5.5B), even though the steady-state firing rate of the network increases smoothly as a function of the initial firing rate (figure 5.5C). Other factors, such as the regulatory role of inhibitory neurons, may contribute to defining the steady-state firing rate. Our conclusion is that the relationship between the degree of recurrence and the steady-state firing rate is nonlinear.

5.3.3 STDP with dominant depression

If depression dominates over potentiation in STDP ($A_+ = 0.0050$ mV and $A_- = 0.0055$ mV in our examples), the coefficient B in equations (5.4) is larger than A (see Appendix A.6), and the baseline parameter C is negative. For these con-

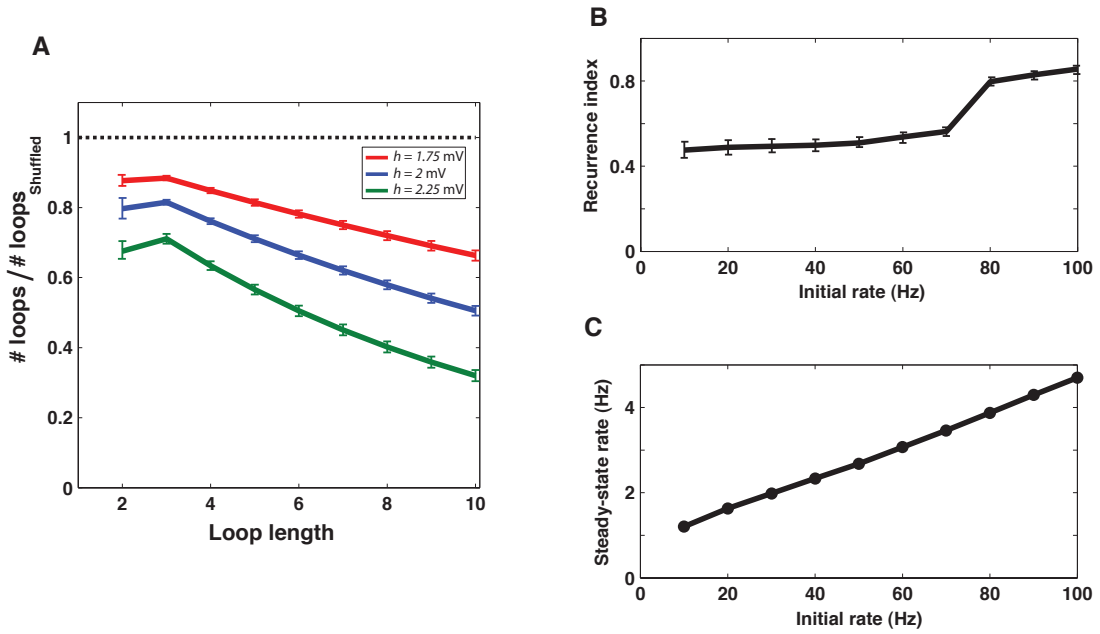

FIG. 5.5: Simulation results for a network with potentiation-dominated STDP. **A.** The number of loops in the steady-state weight matrix divided by the number of loops in a shuffled version of this matrix, as a function of the length of the loop. The initial firing rate of the network was 10 Hz. Error bars depict the standard deviations from 10 different shuffled versions. Different colors correspond to different threshold values h for counting the loops. The ratios are less than one for all loop lengths, so potentiation-dominated STDP decreases the number of loops, but less efficiently than balanced STDP (figure 5.3A). **B.** Recurrence index of the steady-state weight matrix as a function of the average initial firing rate. The recurrence index is defined as the total number of loops shorter than 10 divided by the same quantity for a shuffled network. The steady-state weight matrix becomes more recurrent when the initial rate (i.e. the external input) is higher. **C.** The average steady-state firing rate as a function of the average initial firing rate.

ditions, both elements of the fixed point of the weights, $\bar{w}_1 = Cr_2/(B - A)$ and $\bar{w}_2 = Cr_1/(B - A)$, are negative, which is once again outside of the allowed range of synaptic values. In this case, however, the fixed point is a saddle node, which attracts trajectories from one direction and repels them from the other (see Appendix A.6).

As before, we consider two neurons with equal baseline firing rates. The weight trajectories tend to approach the outlying fixed point in the direction that passes through the origin ($w_1 = w_2 = 0$; see figure 5.6A, arrows). This tendency makes the origin an attractor of the dynamics within the allowed range of synaptic weights. This attractor correspond to completely disconnected neurons. Because the outlying fixed point is a saddle node, the trajectories also tend to drift away from it in the direction perpendicular to the positive-slope diagonal. This tendency produces attractors corresponding to unidirectional connections (figure 5.6, top-left and bottom-right). Once again, trajectories that hit the borders perpendicularly partition the weight space into three basins of attractions corresponding to each attractor (figure 5.6).

The dynamics of the synaptic pair we have considered shows that some pairs of neurons tend to become disconnected when depression dominates over potentiation. This is a more potent mechanism for eliminating loops than the previous cases, so we expect that STDP with dominant depression eliminates more loops in a large network than the other forms we have considered. Numerical simulations confirm that, indeed, there are fewer loops in the steady-state of a network with depression-dominated STDP compared to the previous cases (figure 5.7A vs. 5.3A

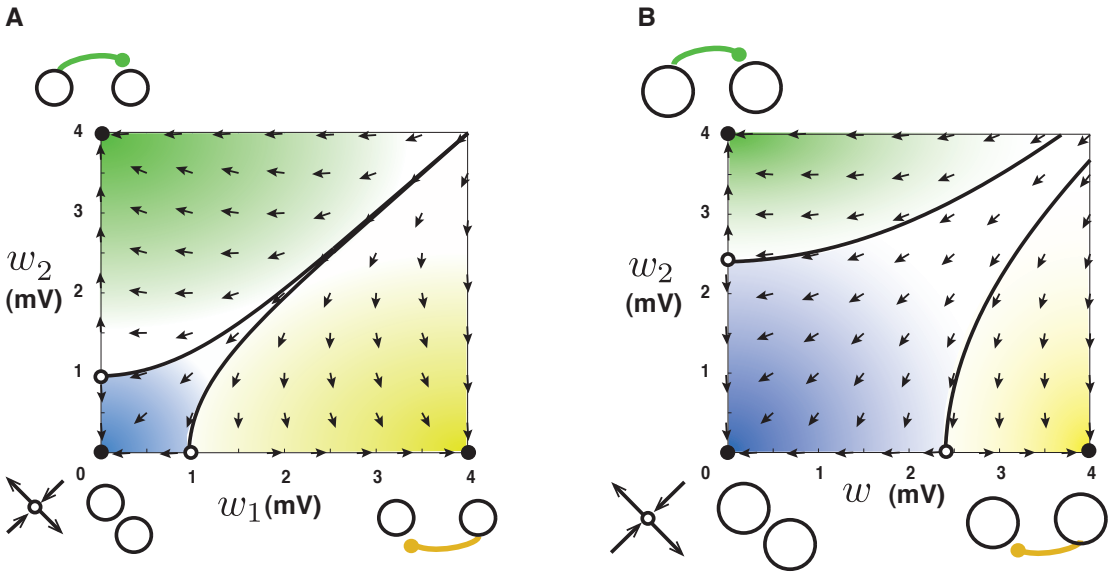

FIG. 5.6: Dynamics of reciprocal synapses when STDP is depression-dominated. **A.** When the baseline firing rates of the two neurons are 10 Hz, a saddle node exist out of the allowed range, schematically illustrated at the bottom left. Arrows show the movement of trajectories. Initial conditions starting within the blue area end up in the attractor at the origin, which corresponds to the loss of connectivity. Trajectories that hit the boundaries perpendicularly delineate the borders of the basins of attractions (solid curves). Initial conditions in the yellow area go to the attractor at the bottom right, corresponding to a unidirectional connection from neuron 1 to neuron 2. Initial conditions in the green area go to the attractor at top left, corresponding to a unidirectional connection from neuron 2 to neuron 1. **B.** The same as A but for a baseline firing rates of 15 Hz. The basin of attraction for the origin (connectivity loss) becomes larger when the baseline firing rate increases.

and 5.5A).

When the baseline rates of the two neurons increase, the basin of the attractor corresponding to disconnected pair becomes larger (figure 6B). This implements a buffering mechanism against high firing rates: when neurons become excessively active, the connections between them are eliminated, so the rate returns to a lower value. Thus, the steady-state firing rate of a network with depression-dominated

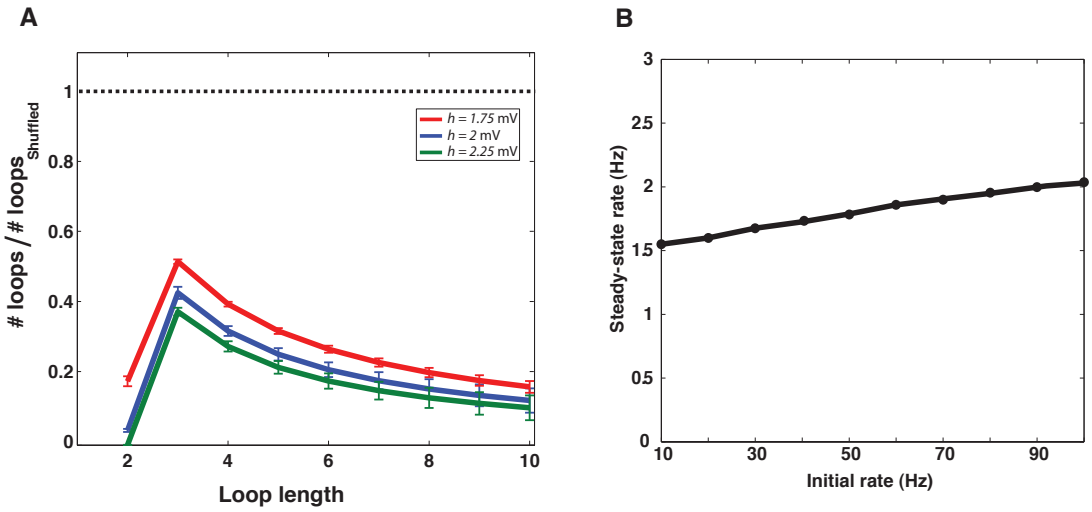

FIG. 5.7: Simulation results of a network with depression-dominated STDP. **A.** The number of loops in the steady-state weight matrix divided by the number of loops in a shuffled version of this matrix, as a function of the length of the loop. The initial firing rate of the network was 10 Hz. Error bars depict the standard deviation from 10 different shuffled versions. Different colors correspond to different threshold values h for counting the loops. The ratios are less than one for all loop lengths. Depression-dominated STDP decrease the loops more efficiently that balanced STDP (figure 5.3A). **B.** The average steady-state firing rate as a function of the average initial firing rate. Depression-dominated STDP buffers the effect of external input on the average steady-state firing rate

STDP should be quite insensitive to the initial firing rate. Simulation results corroborate this conjecture by showing that the steady-state firing rate of the network varies only slightly as a function of the initial firing rate (figure 5.7B).

5.3.4 STDP with a shifted window

The shifted STDP model, as defined in Chapter 3, only considers pairings between nearest neighbor pre- and postsynaptic spikes. This makes the dynamics of the pair of weights more complicated than in the previous cases, because the coefficients A, B and C in equations (5.4) depend on the baseline firing rates (see Appendix A.7).

Furthermore, the coefficient A can become negative at high firing rates, which makes the behavior of the system even more complicated. However, if we consider the system in three different rate regimes, we can elucidate the full range of its behaviors. As in Chapter 3, we assume that the potentiation domain is larger than the depression domain ($A_+ = 0.0085$mV and $A_- = 0.0050$mV). We set the amount of the shift to be $d = 2$ ms.

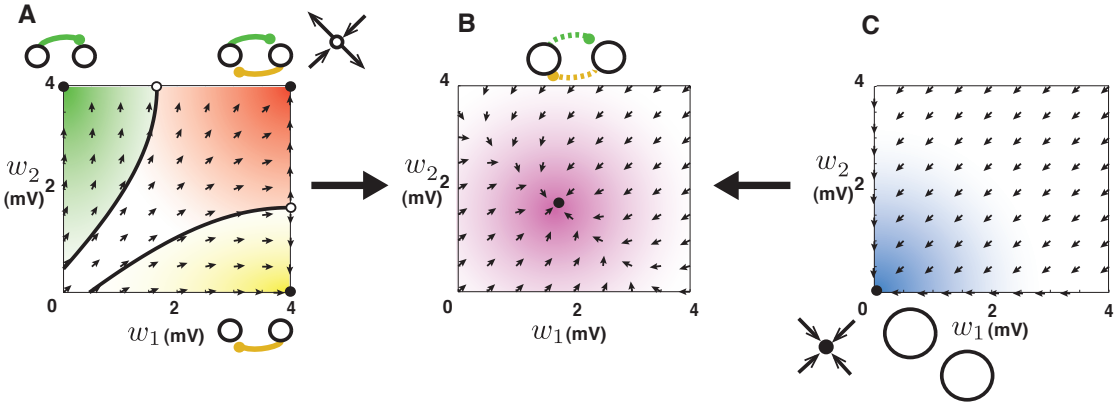

FIG. 5.8: Dynamics of reciprocal synapses with shifted STDP. **A.** When the baseline firing rates of the two neurons are 2 Hz, a saddle node exist out of the allowed range, schematically illustrated at the top right. Arrows show the movement of trajectories. Initial conditions starting within the red area end up in the attractor at the top right corner, which corresponds to strong recurrent connections. This increases the baseline firing rate of the embedding network and pushes the network into the regime shown in B. **B.** When the baseline firing rates of the two neurons are 30 Hz, a single stable fixed point exist within the allowed range of synaptic weights. All initial conditions end up at this fixed point, resulting in a recurrent reciprocal connection. **C.** When the baseline firing rates of the two neurons are 90 Hz, a stable fixed point exists out of the allowed range, schematically illustrated at the bottom left. Movement of trajectories toward the stable fixed point results in connectivity loss, regardless of the initial condition. This effect reduces the rate of the embedding network and pushes the system into the regime shown in B.

When the baseline firing rates of the two neurons are low, the coefficients A, B

and C are all positive. This is because the pairing intervals are not typically short enough to fall into the depression domain caused by the shift. In addition, the coefficient A is slightly smaller than B. This makes the fixed point for the weights positive and large, so once again it falls out of the range of allowed synaptic weights, but this time on the positive not the negative side (figure 5.8A). The fixed point is a saddle node (see Appendix A.7) and attracts the trajectories of weights along the direction toward the top-right corner of the state space (figure 5.8A, arrows), which corresponds to recurrent connections. This state is qualitatively similar to what we found for STDP with dominant potentiation (compare figures 5.8A and 5.4A & B).

As the strengths of recurrent connections in the embedding network increase, the baseline firing rates of the neurons go up, which eventually causes the coefficient A to become negative. This occurs because the pairing intervals between presynaptic spikes and their causally induced postsynaptic spikes become short enough to fall into the depression domain caused by the shift. This creates a single stable fixed point for the two weights located within the allowed range of the synaptic weights. Both weights are attracted to this fixed point, forming a recurrent connection (figure 5.8B, arrows; see also Appendix A.7), though not of maximal strength.

If the two neurons start with even higher baseline rates, the coefficients A and C are both negative. This follows because at very high firing rates, even the intervals between randomly paired spikes of the baseline activity are short enough to fall into the depression domain caused by the shift. This pushes the fixed point of the weights out of the allowed range (figure 5.8C) but, in this case, on the negative

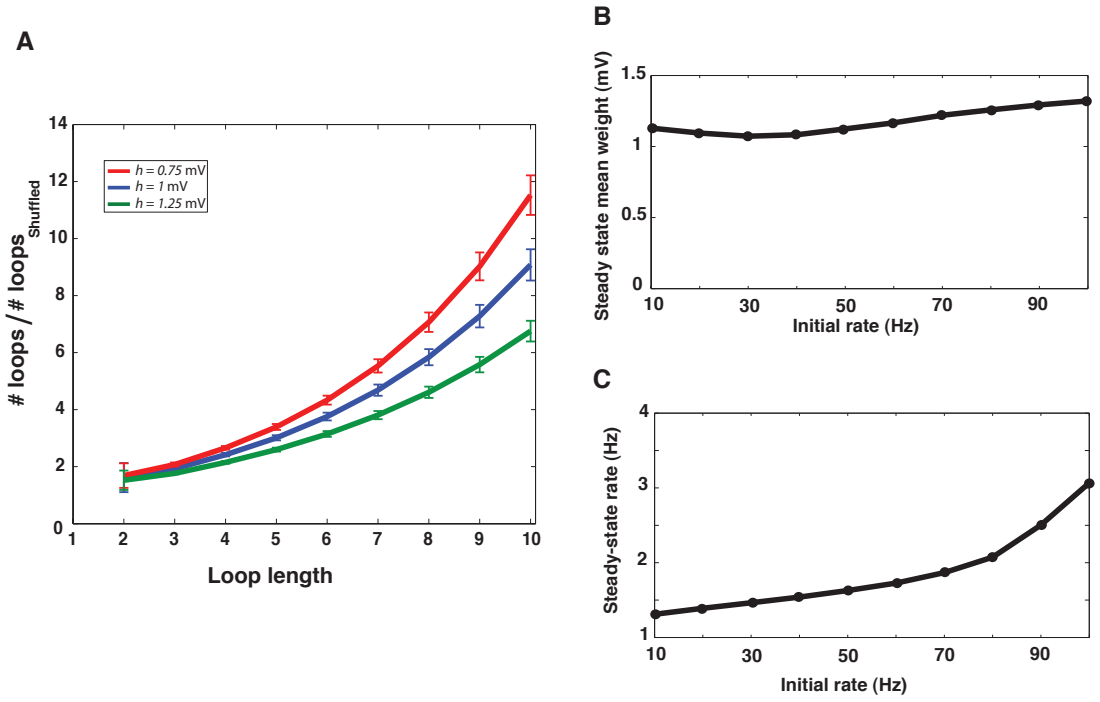

FIG. 5.9: Simulation results of a network with shifted STDP. **A** The number of loops in the steady-state weight matrix, divided by the number of loops in a shuffled version of this matrix, as a function of the length of the loop. The initial firing rate of the network was 10 Hz. Error bars illustrate the standard deviation from 10 different shuffled versions. Different colors correspond to different threshold values h for counting the loops. The ratios are all greater than one, showing that the network generates loops **B.** The average mean of the steady-state weights as a function of the average initial firing rate. Depression-dominated STDP buffers the effect of the external input on the average of the steady-state weights. **C.** The average steady-state firing rate as a function of the average initial firing rate.

side. Because this fixed point is stable, the weights tend to approach it, creating an attractor at the origin that eliminates both weights and disconnects the neurons. This mechanism prunes the weights in the embedding network until the baseline firing rate decreases enough to make the parameter C positive. Then, the regime with a stable fixed point within the allowed range (figure 5.8B) is restored.

Combining these effects, we find that, if the parameters of shifted STDP are set properly, a network will settle into a regime with a single stable fixed point within the allowed range of synaptic weights, regardless of the initial firing rate or the strength of the external input. By generalizing from the dynamics of a pair of synapses in this case, two predictions can be made. First, the steady-state matrix of synaptic weights will have many recurrent connections because there is no mechanism to eliminate loops, and reciprocal connections tend to be strengthened. This prediction is confirmed by numerical simulations that show a highly recurrent steady-state connectivity (figure 5.9A). Second, because the pairwise connections settle into a regime with a single stable fixed point regardless of the initial baseline rate, the mean excitatory-to-excitatory synaptic strength of the network in the steady-state should be resilient to changes in the external input or in the initial firing rate. Numerical simulations show that the steady-state mean of the synaptic weights changes only slightly as a function of the initial firing rate (figure 5.9B), even though the final firing rate of the network increases with the initial firing rate. Thus, shifted STDP implements a homeostatic mechanism that buffers the value of the mean synaptic strength from external influences.

5.4 Discussion

By analyzing pairwise interactions of neurons through STDP, we showed that conventional pair-based STDP functions as a loop-eliminating mechanism in a network of spiking neurons and organizes neurons into in- and out-hubs, as reported by Kozloski and Cecchi (2010). Loop-elimination increases when depression dominates and decreases when potentiation dominates. STDP with dominant depression implements a buffering mechanism for network firing rates, and shifted STDP can generate recurrent connections in a network, and it functions as a homeostatic mechanism that maintains a roughly constant average value of the synaptic strengths.

All the above analytical results were obtained by considering the effect of imposing weight constraints on a linear system describing pairwise interactions of neurons in the presence of STDP. The effect of constraints on Hebbian plasticity has been explored before by Miller and MacKay (1994) to explain the formation of visual receptive fields. Our work can be viewed as an extension of this approach to a specific form of Hebbian plasticity that involves the timing of spikes, namely STDP.

The network used in our numerical simulations was densely connected so that every neuron could potentially form a synaptic connection to every other one. This assumption is not compatible with the observed sparseness of cortical circuits, where each pyramidal neuron makes synaptic contacts with a small fraction of the other neurons in its vicinity (Markram and Tsodyks, 1996; Barthó et al., 2004). However, our analytical results does not rely on any particular assumption about density or sparsity of network connectivity. Instead, the results indicate that STDP can orga-

nize patterns of connectivity in particular ways within the framework provided by anatomical constraints, developmental hard-wiring and other physiological mechanisms, such as other forms of plasticity.

A prominent feature of STDP is its ability to organize neurons into in- and out-hubs. The dependence of hub-formation on baseline firing rate shows how heterogeneity of external inputs can influence the internal structure of a neural network. Moreover, this property of STDP can play an important protective role in pathological cases in which a sub-population of excitatory neurons fires at atypically high rates. Through STDP, most of the incoming synapses to this sub-population will be weakened (i.e. they become purely out-hubs), thereby preventing a further increase in their firing rate due to the intrinsic activity of the network. In fact, decoupling of a highly active sub-population from an embedding network through STDP has been observed previously by Morrison et al. (2007) in networks with an excitation-inhibition balance.

The relative magnitude of potentiation and depressions varies considerably along the dendritic tree (Froemke et al., 2005; Letzkus et al., 2006; Froemke et al., 2010). As hypothesized in Chapter 3, the effect of a shift in the STDP window might be more significant in proximal dendrites. Therefore, all the different parameter regimes of the STDP window considered in the present study could take place along the dendritic tree. A general prediction of our study is then that different regions of the dendritic tree might participate in different network structures as a result of variations of their STDP windows.

This study only addressed structures arising from pair-based STDP in a network. As in the case of synapses converging onto a single neuron (Chapter 4), multi-spike STDP models are expected to induce different structures into a network. A recent study by Clopath et al. (2010) showed that an STDP rule that is phenomenologically similar to the triplet model can induce feedback connections, feedforward structures or localized receptive fields in a small recurrent network, depending on the external input. Furthermore, correlations in the external input, which were not considered here, can have significant impacts on network structure through STDP. This issue has been explored recently in a series of articles by Gilson et al. (2009a;b;c;d; 2010a;b). Another simplifying assumption in our study was the similarity of the STDP windows for both of the synapses connecting the two neurons. Relaxing this assumption will lead to more complex synaptic weight dynamics under STDP.

In conclusion, studying pairwise interactions of neurons through STDP provides a number of important insights about the structures that arise from this plasticity rule in large networks. This approach can be extended to networks with more complex STDP models and more structured external input.

Chapter 6

Concluding Remarks

Experience without theory is blind, but theory without experience is mere intellectual play.
Immanuel Kant

I started this dissertation by characterizing simple pair-based STDP models with respect to the interplay of synaptic stability and competition, and showed that Hebbian competition is incompatible with full stability in these class of models, with either hard or soft bounds imposed to the range of synaptic strengths. Thereafter, I showed that a small temporal shift in the pair-based STDP window can stabilize synaptic strengths and implement both Hebbian and anti-Hebbian forms of competitive synaptic plasticity. This work points out that the detailed shape of the STDP window function near the transition from depression to potentiation is of the utmost importance in determining the consequences of STDP, suggesting that this region warrants further experimental study.

Next, I turned to the question of the interplay between stability and competition in multi-spike STDP models. Experimental results have inspired three models for STDP that take into account the contribution of ensembles of pre- and post-

synaptic spikes larger than mere pairs. By analyzing the steady-state distribution of synaptic weights on a neuron, I showed that multi-spike STDP rules can have radically different consequences at the population level, depending on the exact implementation of the multi-spike interactions. Nevertheless, the dichotomy between full stability and Hebbian competition persists in these models, although anti-Hebbian competition can coexist with synaptic stability in some parameter regimes.

As the next logical step, I focused on structures arising from STDP in recurrent networks of spiking neurons. As previously shown, STDP eliminates feedback loops and organizes neurons into in- and out-hubs (Kozloski and Cecchi, 2010). I found that all of these network properties could be explained by understanding the effect of STDP on pair-wise interactions. This provided an analytically tractable way of relating structures in a network to properties of the STDP model. Depending on the balance between potentiation and depression and also the temporal shift of the STDP window, feed-forward structures, feedback structures, buffering of the average firing rate and homeostasis of the average synaptic strength can emerge in a network.

From a larger perspective, the above results can be viewed as relating the phenomenological models of plasticity proposed on the basis of observations of single synapses, to the structures that arise in a population of synapses from these plasticity rules, thereby bridging the cellular/molecular and circuit/network levels described in the first chapter. From the viewpoint of the neural network triad proposed in the first chapter, the results can be interpreted as showing how neu-

ral activity gives rise to connectivity patterns through one of the most widespread synaptic plasticity mechanisms in the nervous system, namely STDP.

Our results are now awaiting experimental evaluation. Due to technical difficulties, to date the majority of experimental studies on plasticity have been confined to the level of a single synapse. Ideally, theoretical/computational studies can transcend these technical limitations and provide directions and insights for future experimental studies. Recent developments in methods for observing the activity of large populations of neurons such as calcium imaging (Cossart et al., 2005) and voltage sensitive dyes (Grinvald and Hildesheim, 2004), together with advanced statistical techniques for inferring neural connectivity from observed activity (Pillow et al., 2008), pave the way for experimental studies of synaptic structures in large neural networks. I hope this research suggests constructive questions that can be addressed by these new experimental techniques in the near future.

Appendix A

Mathematical appendices

A.1 Calculating the causal increase in postsynaptic firing

We first derive an approximate expression for the causal effect of presynaptic spikes on postsynaptic firing rate of the LIF neuron. Assuming that the excitatory input to the neuron is significantly larger than inhibitory input ($I_{\mathrm{ex}} >> I_{\mathrm{in}}$), the first term (leak) on the right-hand-side of equation (2.1) can be disregarded. The fluctuations of the input are also negligible in this regime, and the neuron fires almost regularly with inter-spike-intervals $1/\bar{r}_{\mathrm{post}}$. If we suppose that the neuron has fired the last postsynaptic spike at time $t = 0$ and, in addition to the baseline input, it receives a presynaptic spike at time t_{pre} via a synapse with strength w, the membrane potential at later times $t > t_{\mathrm{pre}}$ can be approximated as:

$$V(t) \approx V_{\mathrm{r}} + \frac{(I_{\mathrm{ex}} - I_{\mathrm{in}})\, t}{\tau_{\mathrm{m}}} + w\, \frac{\tau_{\mathrm{s}}}{\tau_{\mathrm{m}}} \left[1 - \exp\left(\frac{t_{\mathrm{pre}} - t}{\tau_{\mathrm{s}}}\right)\right] .$$

From the above equation, the relationship between the time of the incoming presynaptic spike $t_{\rm pre}$ and the time that the neuron fires its next postsynaptic spike after the presynaptic spike $\Delta t = t_{\rm post} - t_{\rm pre}$ can be expressed as

$$t_{\rm pre} = f(\Delta t) = \frac{\tau_{\rm m}\,(V_{th} - V_{\rm r})}{I_{\rm ex} - I_{\rm in}} - w\,\frac{\tau_{\rm s}}{I_{\rm ex} - I_{\rm in}}\left[1 - \exp\left(-\frac{\Delta t}{\tau_{\rm s}}\right)\right] - \Delta t\,. \qquad \text{(A.1)}$$

The probability of a presynaptic spike occurring somewhere within the typical inter-spike-interval of the postsynaptic neuron is the reciprocal of the duration of the interval, which is the baseline postsynaptic rate, so $P(t_{\rm pre}) = \bar{r}_{\rm post}$. The instantaneous firing rate of the neuron after the presynaptic spike is the probability density of Δt, which can be expressed as

$$\begin{aligned} r_{\rm post}(\Delta t) &= P(t_{\rm pre})\,|f'(\Delta t)| & \text{(A.2)}\\ &= \bar{r}_{\rm post}\left[1 + w\,\frac{\exp(-\Delta t/\tau_{\rm s})}{I_{\rm ex} - I_{\rm in}}\right] \\ &\approx \bar{r}_{\rm post} + w\,\frac{\exp(-\Delta t/\tau_{\rm s})}{(V_{th} - V_{\rm r})\tau_{\rm m}}\,. & \text{(A.3)} \end{aligned}$$

This is the equation (2.4) of the Methods section of Chapter 2.

A.2 Derivation of the steady-state distribution of weights for shifted STDP

The evolution of the distribution of synaptic strengths is described by the Fokker-Planck equation (Risken, 1996; Kempter et al., 2001; Cateau and Fukai, 2003).

$$\frac{\partial \rho(w,t)}{\partial t} = -\frac{\partial}{\partial w}\big(D_1(w)\rho(w,t)\big) + \frac{1}{2}\frac{\partial^2}{\partial w^2}\big(D_2(w)\rho(w,t)\big)\,, \qquad \text{(A.4)}$$

where D_1 and D_2 are drift and diffusion terms, respectively. To derive equilibrium distributions of synaptic strengths, we need the steady-state solution, obtained by setting the right side of equation A.4 to zero. Solving the resulting ordinary differential equation for $\rho(w)$, we obtain

$$\rho(w) = \frac{N_1}{D_2(w)} \exp\left(\int_0^{\omega} d\omega' \, \frac{2D_1(w')}{D_2(w')}\right), \tag{A.5}$$

where, N_1 is a normalization constant.

The terms D_1 and D_2 can be written as

$$D_1(\omega) = \int_{-\infty}^{\infty} d\epsilon \, \epsilon \, T(w, \epsilon) \quad \text{and} \quad D_2(\omega) = \int_{-\infty}^{\infty} d\epsilon \, \epsilon^2 \, T(w, \epsilon) \, . \tag{A.6}$$

Here, $T(w, \epsilon)$ is the probability density of a synaptic modification that changes the strength of a given synapse from w to $w + \epsilon$.

When the synaptic strengths are changing due to STDP, the only relevant stochastic variable is the interval between the pre- and postsynaptic spike pairs. If a pairing of pre- and postsynaptic spikes occurs with interval Δt, then $\epsilon = F(\Delta t)$, where F is the STDP window function (equation 3.1). To simplify the notation, we use t to denote Δt in the following equations. If the probability density of a pairing interval t is $P(t)$, then the transitional probability density can be written as

$$T(w, \epsilon) = \frac{P(t)}{|F'(t)|} \, . \tag{A.7}$$

With the transformations $d\epsilon \to |F'|dt$ and $\epsilon \to F$, the terms D_1 and D_2 can be

re-expressed as

$$D_1 = \int_{-\infty}^{\infty} dt\, P(t)F(t) \quad \text{and} \quad D_2 = \int_{-\infty}^{\infty} dt\, P(t)\, F^2(t)\,. \tag{A.8}$$

Thus, to determine D_1 and D_2 in terms of the parameters of the model, we only need to know the probability density of pairing intervals $P(t)$.

We approximate the spiking behavior of the integrate-and-fire neuron by that of a linear Poisson neuron firing at the same rate (Gütig et al., 2003; Kempter et al., 2001). We first consider the case that the presynaptic spike follows the postsynaptic spike ($t < 0$). In this case, the timing of the presynaptic spike has no causal effect on the postsynaptic spike time. If we assume that both the presynaptic and postsynaptic spike trains are Poisson, the probability density of nearest-neighbor pairing intervals is

$$P(t) = r \exp(rt)\,, \tag{A.9}$$

where $r = r_{\text{ex}} + \bar{r}_{\text{post}}$ is the sum of the excitatory presynaptic firing rate (r_{ex}) and the baseline postsynaptic firing rate ($\bar{r}_{\text{post}}$). If we assume that both the presynaptic and postsynaptic spike trains are Poisson, the probability density of pairing intervals is

$$P(t) = (r_{\text{ex}} + r_{\text{post}}(t)) \, \exp\left(- \int_0^t ds \, (r_{\text{ex}} + r_{\text{post}}(s)) \right) . \tag{A.10}$$

If we assume that $w \ll (V_{\text{th}} - V_r)$, we can Taylor expand equation A.10 to first order in w and, together with equation A.2, the probability density of pairing intervals

can be expressed as (see figure 3.1C & D)

$$P(t) \approx \begin{cases} r \exp(r\,t) & \text{if } t \leq 0 \\ \\ r \exp(-r\,t) + w\,\dfrac{\tau_{\mathrm{s}}}{\tau_{\mathrm{m}}(V_{\mathrm{th}} - V_r)}\,(\lambda\,\exp(-\lambda\,t) - r\,\exp(-r\,t)) & \text{if } t > 0\,, \end{cases} \tag{A.11}$$

with

$$\lambda = r + \frac{1}{\tau_{\mathrm{s}}}\,.$$

Note that the second term in equation A.11 for $t > 0$ corresponds to the causal bump in figures 3.1C & D. The shape of the causal bump resembles that calculated by Cateau & Fukai (Cateau and Fukai, 2003) from the equation for the first passage time of a noisy integrate-and-fire neuron.

If we substitute A.11 into equation A.6, we obtain D_1 and D_2 in terms of the parameters of the model. Because $P(t)$ is linear in w, D_1 and D_2 are also linear and can be written as

$$D_1 = \alpha w + \beta \quad \text{and} \quad D_2 = \gamma w + \delta\,. \tag{A.12}$$

Here, αw is the w-dependent drift and β is the baseline drift. Assuming that $d \ll \tau_{\pm}$,

these coefficients can be written as

$$\begin{aligned}
\alpha &= \frac{1}{\tau_{\mathrm{m}}(V_{\mathrm{th}}-V_r)}\left(\frac{A_{+}(\tau_{+}\tau_{\mathrm{s}}+d\tau_{\mathrm{s}})}{(1+r\tau_{+})(r\tau_{\mathrm{s}}\tau_{+}+\tau_{\mathrm{s}}+\tau_{+})}-(A_{+}+A_{-})d\right) \qquad \text{(A.13)}\\
\beta &= A_{+}r\tau_{+}\left(\frac{1-dr}{1+r\,\tau_{+}}\right)-A_{-}r\,\tau_{-}\left(\frac{1+d\,r}{1+r\,\tau_{-}}\right)\\
\gamma &= \frac{1}{\tau_{\mathrm{m}}(V_{\mathrm{th}}-V_r)}\left(\frac{A_{+}^{2}(2\tau_{+}\tau_{\mathrm{s}}+4d\tau_{\mathrm{s}})}{(2+r\tau_{+})(r\tau_{\mathrm{s}}\tau_{+}+2\tau_{\mathrm{s}}+\tau_{+})}-(A_{+}^{2}-A_{-}^{2})d\right)\\
\delta &= A_{+}^{2}r\tau_{+}\left(\frac{1-d\,r}{2+r\tau_{+}}\right)+A_{-}^{2}r\tau_{-}\left(\frac{1+dr}{2+r\tau_{-}}\right).
\end{aligned}$$

Finally, by inserting equations A.12 into equation A.5, we obtain the steady-state distribution

$$\rho(w)=N_{0}\,(w+\mu)^{k-1}\,\exp\left(-(w+\mu)/\theta\right) \qquad \text{(A.14)}$$

with

$$\mu=\frac{\delta}{\gamma}\,,\quad k=\frac{2(\beta\gamma-\alpha\delta)}{\gamma^{2}}\quad\text{and}\quad\theta=-\frac{\gamma}{2\alpha}\,. \qquad \text{(A.15)}$$

Equation A.14 is the same as equation 3.2 of the Results.

For the above distribution A.14 to be normalizable, θ and k must be positive. Equations A.15 indicate that these conditions are met if β, γ and δ, as given by equations A.13, are all positive and α is negative. Provided that r is less than of order $1/d$ (which it always is at steady-state), $\beta>0$ if $A_{+}\tau_{+}$ is sufficiently greater than $A_{-}\tau_{-}$, which is the condition stated in the text. Over the range we consider, $\gamma>0$ and $\delta>0$ without requiring any further conditions. When r is greater than of order $\sqrt{1/d\tau_{+}}$, which it is at steady state, $\alpha<0$, so stability is achieved.

If β is positive, the mean synaptic strength is approximately

$$\langle w \rangle = -\frac{\beta}{\alpha} \tag{A.16}$$

Solving the above equation simultaneously with equation 2.3, gives the steady-state firing rate of the neuron, as depicted in figure 3.3 (inset). Having solved for the steady-state postsynaptic firing rate and the mean synaptic strength, the parameters of the distribution (equation 3.2) are fully expressed in terms of the model parameters.

A.3 Calculating average drift of weights for the suppression model

Equation (A.2) enables us to calculate the probability density of the intervals between pre- and post synaptic spikes ($\Delta t = t_{\rm post} - t_{\rm pre}$). Assuming that the presynaptic spike train is Poisson and the postsynaptic spike train is regular (Appendix A.1), the probability density of Δt is

$$\begin{aligned} P(\Delta t) &= r_{\rm pre}\, r_{\rm post}(\Delta t) \exp(-\Delta t\, r_{\rm pre}) \qquad (A.17) \\ &= \begin{cases} r_{\rm pre}\left[\bar{r}_{\rm post} + w\, \dfrac{\exp(-\Delta t/\tau_{\rm s})}{(V_{th} - V_{\rm r})\tau_{\rm m}}\right] \exp(-\Delta t\, r_{\rm pre}) & \text{if } \Delta t \geq 0 \\ \\ r_{\rm pre}\, \bar{r}_{\rm post}\, \exp(-\Delta t\, r_{\rm pre}) & \text{if } \Delta t < 0\,. \end{cases} \end{aligned}$$

When the neuron fires almost regularly, the interval between two consecutive postsynaptic spikes can be considered to be $1/\bar{r}_{\text{post}}$. However, if a presynaptic spike arrives before a postsynaptic spike ($\Delta t > 0$), it shortens the interval between the upcoming postsynaptic spike and the preceding one. In this case, the interval between consecutive postsynaptic spikes (Δt_{post}) can expressed as a function of Δt from equation (A.1). Taken together,

$$\Delta t_{\text{post}} = \begin{cases} \dfrac{1}{\bar{r}_{\text{post}}}\left[1 - w\,\dfrac{\tau_{\text{s}}\left(1-\exp(-\Delta t/\tau_{\text{s}})\right)}{\tau_{\text{m}}\left(V_{th}-V_{\text{r}}\right)}\right] & \Delta t \geq 0 \\ \dfrac{1}{\bar{r}_{\text{post}}} & \Delta t < 0\,. \end{cases} \tag{A.18}$$

We now have all the components required to calculate the average drift of weights under the suppression model (see equation 4.3). In this model, each nearest neighboring pre-post pair of spikes with pairing interval Δt induces potentiation or depression depending on the ordering of the pair. In addition, the previous pre- and postsynaptic spikes participate in plasticity depending on their temporal distance from the spikes in the pair (Δt_{pre} and Δt_{post} respectively) . Of the three intervals participating in suppression model, Δt_{pre} and Δt are stochastic variables, while Δt_{post} is simply a function of Δt. Therefore, we should average the weight modification over all possibles values of Δt and Δt_{pre}. Given the assumption that the postsynaptic spike fires almost regularly, $|\Delta t|$ cannot be longer than the length of a typical postsynaptic ISI ($1/\bar{r}_{\text{post}}$). Also, in the case where the postsynaptic spike precedes the presynaptic one ($\Delta t < 0$), there is a lower limit on Δt_{pre}: it cannot be shorter than the pre-post interval $|\Delta t|$. By these considerations, the

average weight change can be calculated as

$$\frac{d\langle w\rangle}{dt} = \int_{-1/\bar{r}_{\rm post}}^{+1/\bar{r}_{\rm post}} d\Delta t\, P(\Delta t) \int_{\max(0,-\Delta t)}^{\infty} d\Delta t_{\rm pre}\, P(\Delta t_{\rm pre})\, F_{supp}(\Delta t, \Delta t_{\rm pre}, \Delta t_{\rm post})$$

where F_{supp} is the weight modification (equation 4.3). Because the presynaptic spike train is assumed to be Poisson, $P(\Delta t_{\rm pre})$ is the waiting time of the Poisson process, namely $r_{\rm pre}\,\exp(-\Delta t_{\rm pre}\, r_{\rm pre})$. Substituting equations (A.17) and (A.18) in the above equation and keeping only terms up to first order in $\langle w\rangle$ results in

$$\begin{aligned}\frac{d\langle w\rangle}{dt} &= \mathcal{E}(\tau_{\rm post})\,\bar{r}_{\rm post}\, r_{\rm pre}\left[\frac{A_+\,\tilde{\tau}_+\,\mathcal{E}(\tilde{\tau}_+)}{1+r_{\rm pre}\,\tau_{\rm pre}} - A_-\,\tilde{\tau}_-\mathcal{E}(\tilde{\tau}_-) + A_-\,\hat{\tau}_-\mathcal{E}(\hat{\tau}_-)\right] \\ &\quad + \langle w\rangle\,\frac{A_+\, r_{\rm pre}}{\tau_{\rm m}\,\tau_{\rm post}\,(V_{th}-V_{\rm r})}\left[\frac{\hat{\tau}_+\,\mathcal{E}(\hat{\tau}_+)\Big((\tau_{\rm post}-\tau_s)\mathcal{E}(\tau_{\rm post})+\tau_{\rm s}\Big)}{1+r_{\rm pre}\,\tau_{\rm pre}}\right. \\ &\quad \left. -\tau_{\rm s}\,\tilde{\tau}_+\,\mathcal{E}(\tilde{\tau}_+)\Big(1-\mathcal{E}(\tau_{\rm post})\Big)\right] \qquad \text{(A.19)}\end{aligned}$$

with $\mathcal{E}(\tau) = 1-\exp(-1/(r_{\rm post}\,\tau))$ and time constants defined as

$$\begin{aligned}\tilde{\tau}_+ &= \frac{\tau_+}{1+r_{\rm pre}\,\tau_+} \qquad & \hat{\tau}_+ &= \frac{\tau_{\rm s}\,\tau_+}{\tau_{\rm s}+\tau_+ +\tau_{\rm s}\,\tau_+} \\ \tilde{\tau}_- &= \frac{\tau_-}{1+r_{\rm pre}\,\tau_-} \qquad & \hat{\tau}_- &= \frac{\tau_{\rm pre}\,\tau_-}{\tau_{\rm pre}+\tau_- +\tau_{\rm pre}\,\tau_-}\,.\end{aligned}$$

Equation (A.19) is numerically evaluated in figure 4.5.

A.4 Calculating average drift of weights for the NMDAR-based model

To calculate the average synaptic change in the NMDAR-base model, we first assume that the pre- and postsynaptic spike trains are independent, i.e. we disregard the causal effect of presynaptic spikes on postsynaptic firing. Afterwards, we include the causal relation. Assuming that the pre- and postsynaptic spike trains are Poisson, we can replace the δ functions in equations (4.4) by their corresponding rates ($r_{\rm pre}$ and $\bar{r}_{\rm post}$) and, setting the right-hand-side equal to zero to obtain the mean values of the NMDAR state variables in up and down states, we find

$$\langle f^{\rm up}\rangle = \frac{A_f^{\rm up}\,\tau_f^{\rm up}\,r_{\rm pre}}{1+A_f^{\rm up}\,\tau_f^{\rm up}\,r_{\rm pre}+A_f^{\rm dn}\,\tau_f^{\rm dn}\,\bar{r}_{\rm post}}$$

$$\langle f^{\rm dn}\rangle = \frac{A_f^{\rm dn}\,\tau_f^{\rm dn}\,\bar{r}_{\rm post}}{1+A_f^{\rm up}\,\tau_f^{\rm up}\,r_{\rm pre}+A_f^{\rm dn}\,\tau_f^{\rm dn}\,\bar{r}_{\rm post}}\,. \tag{A.20}$$

The mean values of the second messengers $\langle M^{\rm up}\rangle$ and $\langle M^{\rm dn}\rangle$ could be calculated using the same technique. However we use a different method for the second messengers that is equivalent to the above technique for the mean and also enables us to calculate the variance around the mean. At any instant t that we observe the variable $M^{\rm up}(t)$, the last incoming postsynaptic spike has arrived Δt previously. The value of $M^{\rm up}(t-\Delta t)$ has increased instantly in response to the postsynaptic

spike and decayed afterwards so, according to equation (4.5),

$$M^{\mathrm{up}}(t) = \Big(M^{\mathrm{up}}(t-\Delta t) + A_M^{\mathrm{up}}\, f^{\mathrm{up}}(t-\Delta t)\,\big(1 - M^{\mathrm{up}}(t-\Delta t)\big)\Big)\, \exp(-\Delta t/\tau_M^{\mathrm{up}})\,. \tag{A.21}$$

Assuming that the system is in equilibrium, the mean value of M^{up} should be the same at the times $t-\Delta$ and t. Moreover, because the postsynaptic spike train is assumed to be Poisson, the probability of an incoming postsynaptic spike Δt before time t is $\bar{r}_{\mathrm{post}}\, \exp(-\Delta t\, \bar{r}_{\mathrm{post}})$. By setting the values of M^{up} equal to the mean and averaging over all possible Δt values we have

$$\langle M^{\mathrm{up}}\rangle = \Big(\langle M^{\mathrm{up}}\rangle + A_M^{\mathrm{up}}\, \langle f^{\mathrm{up}}\rangle\,\big(1 - \langle M^{\mathrm{up}}\rangle\big)\Big) \int_0^\infty dt\, \bar{r}_{\mathrm{post}}\, \exp(-t\, \bar{r}_{\mathrm{post}})\exp(-t/\tau_M^{\mathrm{up}})\,,$$

which results in

$$\langle M^{\mathrm{up}}\rangle = \frac{A_M^{\mathrm{up}}\, \tau_M^{\mathrm{up}}\, \langle f^{\mathrm{up}}\rangle\, \bar{r}_{\mathrm{post}}}{1 + A_M^{\mathrm{up}}\, \tau_M^{\mathrm{up}}\, \langle f^{\mathrm{up}}\rangle\, \bar{r}_{\mathrm{post}}}\,. \tag{A.22}$$

As mentioned before, the same result could be obtained by replacing the δ functions in equation (4.5) with their corresponding rates and solving for the steady-state. However, the same argument as for equation (A.21) can be used to calculate the variance of M^{up} around its mean. The square of $M^{\mathrm{up}}(t)$ can be obtained by squaring both sides of equation (A.21). Assuming the mean square of M^{up} to be the same at t and $t-\Delta t$ and averaging over all possible Δt values, we find

$$\begin{aligned}\langle M^{\mathrm{up}^2}\rangle \;=\; & \Big[(1 - A_M^{\mathrm{up}}\, \langle f^{\mathrm{up}}\rangle)^2 \langle M^{\mathrm{up}^2}\rangle + 2\,(1 - A_M^{\mathrm{up}}\, \langle f^{\mathrm{up}}\rangle) A_M^{\mathrm{up}}\, \langle f^{\mathrm{up}}\rangle\, \langle M^{\mathrm{up}}\rangle + (A_M^{\mathrm{up}})^2\, \langle f^{\mathrm{up}}\rangle^2 \Big] \\ & \times \int_0^\infty dt\, \bar{r}_{\mathrm{post}}\, \exp(-t\, \bar{r}_{\mathrm{post}})\exp(-2t/\tau_M^{\mathrm{up}})\,,\end{aligned}$$

which can be used to calculate the variance of $M^{\rm up}$ as

$$\sigma^2_{\rm up} = \langle {M^{\rm up}}^2 \rangle - \langle M^{\rm up} \rangle^2 \approx \frac{\langle M^{\rm up} \rangle^2}{2\,\bar{r}_{\rm post}\,\tau^{\rm up}_M (1 + A^{\rm up}_M\,\tau^{\rm up}_M\,\langle f^{\rm up} \rangle\,\bar{r}_{\rm post})}\,. \tag{A.23}$$

Similarly, the mean and the variance of $M^{\rm dn}$ are

$$\begin{aligned} \langle M^{\rm dn} \rangle &= \frac{A^{\rm dn}_M\,\tau^{\rm dn}_M\,\langle f^{\rm dn} \rangle\,r_{\rm pre}}{1 + A^{\rm dn}_M\,\tau^{\rm dn}_M\,\langle f^{\rm dn} \rangle\,r_{\rm pre}} \\ \sigma^2_{\rm dn} &\approx \frac{\langle M^{\rm dn} \rangle^2}{2\,r_{\rm pre}\,\tau^{\rm dn}_M (1 + A^{\rm dn}_M\,\tau^{\rm dn}_M\,\langle f^{\rm dn} \rangle\,r_{\rm pre})}\,. \end{aligned} \tag{A.24}$$

We have ignored the variance of $f^{\rm up}$ and $f^{\rm dn}$ in the above calculations because their contribution in $\sigma^2_{\rm up}$ and $\sigma^2_{\rm dn}$ is insignificant.

The second messengers $M^{\rm up}$ and $M^{\rm dn}$ participate in potentiation and depression immediately after their abrupt increase due to the most recent pre- or post-synaptic spikes. Based on equation (4.4), the sizes of these abrupt changes are $A^{\rm up}_M\,\langle f^{\rm up} \rangle (1 - \langle M^{\rm up} \rangle)$ and $A^{\rm dn}_M\,\langle f^{\rm dn} \rangle (1 - \langle M^{\rm dn} \rangle)$ respectively. Therefore, the mean value of the second messengers participating in potentiation and depression are

$$\mu^{\rm up} = \langle M^{\rm up} \rangle + A^{\rm up}_M\,\langle f^{\rm up} \rangle (1 - \langle M^{\rm up} \rangle)$$

$$\mu^{\rm dn} = \langle M^{\rm dn} \rangle + A^{\rm dn}_M\,\langle f^{\rm dn} \rangle (1 - \langle M^{\rm dn} \rangle) \tag{A.25}$$

Finally, assuming the steady-state values of $M^{\rm up}$ and $M^{\rm dn}$ to be Gaussian variables with the above mean and variance, the average synaptic drift according to equation

(4.6) can be calculated as

$$\begin{aligned}\frac{d\langle w\rangle_0}{dt} &= \frac{A_+\,\bar{r}_{\text{post}}}{\sqrt{2\pi}\,\sigma_{\text{up}}}\int_{-\infty}^{+\infty} dx\,[x-\theta^{\text{up}}]^+\,\exp\left(-\frac{(x-\mu^{\text{up}})^2}{2\,\sigma_{\text{up}}^2}\right) \\ &\quad -\frac{A_-\,r_{\text{pre}}}{\sqrt{2\pi}\,\sigma_{\text{dn}}}\int_{-\infty}^{+\infty} dx\,[x-\theta^{\text{dn}}]^+\,\exp\left(-\frac{(x-\mu^{\text{dn}})^2}{2\,\sigma_{\text{dn}}^2}\right) \\ &= A_+\,\bar{r}_{\text{post}}\left(\frac{\sigma_{\text{up}}}{\sqrt{2\pi}}\,\exp\left(-\frac{(\mu^{\text{up}}-\theta^{\text{up}})^2}{2\,\sigma_{\text{up}}^2}\right)+\frac{\mu^{\text{up}}-\theta^{\text{up}}}{2}\,(1+\text{erf}\,(\frac{\mu^{\text{up}}-\theta^{\text{up}}}{\sqrt{2}\,\sigma_{\text{up}}})\right) \\ &\quad -A_-\,r_{\text{pre}}\left(\frac{\sigma_{\text{dn}}}{\sqrt{2\pi}}\,\exp\left(-\frac{(\mu^{\text{dn}}-\theta^{\text{dn}})^2}{2\,\sigma_{\text{dn}}^2}\right)+\frac{\mu^{\text{dn}}-\theta^{\text{dn}}}{2}\,(1+\text{erf}\,(\frac{\mu^{\text{dn}}-\theta^{\text{dn}}}{\sqrt{2}\,\sigma_{\text{dn}}})\right).\end{aligned} \tag{A.26}$$

The subscript 0 on the right-hand-side denotes that we have not include the causal effect of presynaptic spikes on the postsynaptic spikes yet.

When a presynaptic spikes arrives, it transiently increases the postsynaptic firing rate proportional to the strength (w) of the synapse through which it arrived (Appendix A.1). Because the amount of NMDARs in down state (f^{dn}) increases through postsynaptic spikes, we should take into account the effect of this transient postsynaptic rate increase on f^{dn}. Apart from transiently increasing the postsynaptic firing rate, the presynaptic spike has another effect on the NMDARs: upon is arrival, it consumes part of the pool of NMDARs at rest step (f^{rest}) and turns them into f^{up}, leaving less f^{rest} available for the upcoming postsynaptic spikes. Because the NMDARs decay back to the rest state over time, the shorter the interval between pre- and postsynaptic spikes, the less time f^{rest} has the postsynaptic spike at its disposal to convert into f^{dn}. Considering this effect and averaging over the full range of intervals between the presynaptic spike and its induced postsynaptic

spikes in equation (4.4), we find

$$\begin{aligned}
\frac{d\langle f^{\mathrm{up}}\rangle}{dt} &= -\frac{\langle f^{\mathrm{up}}\rangle}{\tau_f^{\mathrm{up}}} + A_f^{\mathrm{up}}\,\langle f^{\mathrm{rest}}\rangle\, r_{\mathrm{pre}} \qquad \text{(A.27)}\\
\frac{d\langle f^{\mathrm{dn}}\rangle}{dt} &= -\frac{\langle f^{\mathrm{dn}}\rangle}{\tau_f^{\mathrm{dn}}} + A_f^{\mathrm{dn}}\,\langle f^{\mathrm{rest}}\rangle\,\bar{r}_{\mathrm{post}} + w\frac{r_{\mathrm{pre}}\,\tau_{\mathrm{s}}}{\tau_{\mathrm{m}}\,(V_{th}-V_{\mathrm{r}})}\,A_f^{\mathrm{dn}}\,\langle f^{\mathrm{rest}}\rangle\,\frac{\tau_{\mathrm{s}}+(1-A_f^{\mathrm{up}})\,\tau_f^{\mathrm{up}}}{\tau_{\mathrm{s}}+\tau_f^{\mathrm{up}}}\,.
\end{aligned}$$

As expected, if we disregard the w-dependent term (i.e. ignoring the causal effect of the presynaptic spikes on postsynaptic firing) and solve for the steady-state, we retrieve the result of equation (A.20). By keeping the w-dependent term in the above equations, the following w-dependent terms will be added to the steady-states calculated in equation (A.20),

$$\begin{aligned}
\langle f^{\mathrm{up}}\rangle_w &= -w\,\frac{\tau_{\mathrm{s}}\,A_f^{\mathrm{dn}}\,\tau_f^{\mathrm{dn}}\,\langle f^{\mathrm{up}}\rangle^2\left(\tau_{\mathrm{s}}+(1-A_f^{\mathrm{up}})\,\tau_f^{\mathrm{up}}\right)}{\tau_{\mathrm{m}}\,A_f^{\mathrm{up}}\,\tau_f^{\mathrm{up}}\,(V_{th}-V_{\mathrm{r}})(\tau_{\mathrm{s}}+\tau_f^{\mathrm{up}})}\\
\langle f^{\mathrm{dn}}\rangle_w &= w\,\frac{\tau_{\mathrm{s}}\,A_f^{\mathrm{dn}}\,\tau_f^{\mathrm{dn}}\,r_{\mathrm{pre}}\,(1-\langle f^{\mathrm{dn}}\rangle)^2\left(\tau_{\mathrm{s}}+(1-A_f^{\mathrm{up}})\,\tau_f^{\mathrm{up}}\right)}{\tau_{\mathrm{m}}\,(1+A_f^{\mathrm{up}}\,\tau_f^{\mathrm{up}}\,r_{\mathrm{pre}})\,(V_{th}-V_{\mathrm{r}})(\tau_{\mathrm{s}}+\tau_f^{\mathrm{up}})} \qquad \text{(A.28)}
\end{aligned}$$

The subscripts w denote the w-dependent contributions to the steady-state NMDAR ratios. Note that the contribution is negative for f^{up} and positive for f^{dn}. Intuitively, this implies that as the synapse gets stronger, the neuron fires more in response to the presynaptic spikes, hence a greater portion of NMDARs move to the down state and the share in the up state becomes smaller.

The second messenger M^{up} is also activated by postsynaptic spikes, therefore we should take into account the causal effect of presynaptic spikes on the postsynaptic rate in this case as well. Each presynaptic spike also increases the amount of f^{up} available for postsynaptic spikes according to equation (4.5). By taking these

effects into account, including the w-dependence from equation (A.28) and averaging over all possible pre-post intervals, the following w-dependent terms will be added to the steady-state of $M^{\rm up}$ and $M^{\rm dn}$,

$$\begin{aligned}\langle M^{\rm up}\rangle_w &= w\,\frac{A_M^{\rm up}\,\tau_M^{\rm up}\,\tau_{\rm s}\,r_{\rm pre}}{\tau_{\rm m}\,(V_{th}-V_{\rm r})\,(1+A_M^{\rm up}\,\tau_M^{\rm up}\,\langle f^{\rm up}\rangle\,\bar{r}_{\rm post})^2}\\ &\quad\times\left(\langle f^{\rm up}\rangle+\langle f^{\rm rest}\rangle\,\frac{A_f^{\rm up}\,\tau_f^{\rm up}}{\tau_{\rm s}+\tau_f^{\rm up}}\right)\end{aligned}$$

$$\langle M^{\rm dn}\rangle_w = \frac{\langle f^{\rm dn}\rangle_w\;A_M^{\rm dn}\,\tau_M^{\rm dn}\,r_{\rm pre}}{(1+A_M^{\rm dn}\,\tau_M^{\rm dn}\,\langle f^{\rm dn}\rangle\,r_{\rm pre})^2}\,. \tag{A.29}$$

The mean values in equation (A.25) inherit the w-dependence from equations (A.28) and (A.29) in the form

$$\begin{aligned}\mu_w^{\rm up} &= \left(1-A_M^{\rm up}\,\langle f^{\rm up}\rangle\right)\langle M^{\rm up}\rangle_w + A_M^{\rm up}\left(1-\langle M^{\rm up}\rangle\right)\langle f^{\rm up}\rangle_w\\ \mu_w^{\rm dn} &= \left(1-A_M^{\rm dn}\,\langle f^{\rm dn}\rangle\right)\langle M^{\rm dn}\rangle_w + A_M^{\rm dn}\left(1-\langle M^{\rm dn}\rangle\right)\langle f^{\rm dn}\rangle_w\,.\end{aligned} \tag{A.30}$$

Finally, by inserting the above results into equation (A.26) and keeping only the terms linear in w, the w-dependent drift in the NMDAR-based model is obtained,

$$\begin{aligned}\frac{d\langle w\rangle_w}{dt} &= \frac{A_+}{2}\,\bar{r}_{\rm post}\,\mu_w^{\rm up}\left(1+{\rm erf}\,(\frac{\mu^{\rm up}-\theta^{\rm up}}{\sqrt{2}\,\sigma_{\rm up}})\right)\\ &\quad-\frac{A_-}{2}\,r_{\rm pre}\,\mu_w^{\rm dn}\left(1+{\rm erf}\,(\frac{\mu^{\rm dn}-\theta^{\rm dn}}{\sqrt{2}\,\sigma_{\rm dn}})\right)\,.\end{aligned} \tag{A.31}$$

.

Equation (A.26) together with equation (A.31) describe mean drift of weights through NMDAR-base model,

$$\frac{d\langle w\rangle}{dt} = \frac{d\langle w\rangle_0}{dt} + \frac{d\langle w\rangle_{\langle w\rangle}}{dt}\,,$$

which is used in analytical results in figure 4.9. For small deviations of individual synaptic weights around the mean, equation (A.31) can be used giving

$$\frac{d\,\delta w}{dt} = \frac{d\langle w\rangle_{\delta w}}{dt}\,,$$

which is used in figure 4.9B.

A.5 Deriving the equations for pairwise interactions of weights

Assuming an all-to-all interaction and Poisson statistics for the spike trains, the probability distribution of pairing intervals between the pre- and postsynaptic spike trains is simply the product of their rates. In the case of pairwise interactions of neurons, the spike train of neuron 1 is presynaptic with respect to the synapse w_1 and postsynaptic with respect to w_2. Conversely, the spike train of neuron 2 is presynaptic to the synapse w_2 and postsynaptic to w_1. The rate of the each neuron has a baseline value ($\bar{r}_1$ and $\bar{r}_2$, respectively) and transiently increases at a times t after the arrival of a spike from the other neuron such that

$$r_1(t) \approx \bar{r}_1 + \frac{w_2 \exp\big(-t/\tau_{\rm s}\big)}{(V_{\rm th}-V_{\rm r})\tau_{\rm m}}, \quad \text{and} \quad r_2(t) \approx \bar{r}_2 + \frac{w_1 \exp\big(-t/\tau_{\rm s}\big)}{(V_{\rm th}-V_{\rm r})\tau_{\rm m}}\,.$$

For w_1 the transient increase of the rate of neuron 2 falls into the potentiation domain of STDP, and the transient increase of the rate of neuron 1 falls into the depression domain. Therefore, the average drift of w_1 is

$$\begin{aligned}\frac{dw_1}{dt} &= A_+ \int_0^{+\infty} dt\, \bar{r}_1\, r_2(t)\, \exp(-t/\tau_+) - A_- \int_{-\infty}^{0} dt\, \bar{r}_2\, r_1(t)\, \exp(t/\tau_-) \\ &= A\, \bar{r}_1\, w_1 - B\, \bar{r}_2\, w_2 + C\, \bar{r}_1\, \bar{r}_2 \end{aligned} \tag{A.32}$$

with coefficients

$$\begin{aligned} A &= \frac{A_+\, \tau_+\, \tau_{\mathrm{s}}}{\tau_{\mathrm{m}}\, (V_{\mathrm{th}} - V_{\mathrm{r}})\, (\tau_+ + \tau_{\mathrm{s}})} \\ B &= \frac{A_-\, \tau_-\, \tau_{\mathrm{s}}}{\tau_{\mathrm{m}}\, (V_{\mathrm{th}} - V_{\mathrm{r}})\, (\tau_- + \tau_{\mathrm{s}})} \\ C &= A_+\, \tau_+ - A_-\, \tau_- \,. \end{aligned} \tag{A.33}$$

Together with a similar calculation for w_2 the system of equations (5.4) are obtained.

Setting the right sides of equations (5.4) to zero, yields the fixed point of the weights,

$$\bar{w}_1 = \frac{C\, \bar{r}_2}{B - A} \quad \text{and} \quad \bar{w}_2 = \frac{C\, \bar{r}_1}{B - A}\,. \tag{A.34}$$

Note that when the numerator and denominator are zero (the balanced STDP) the fixed point is not defined. In that case, all values satisfying $w_1 = \bar{r}_2 w_2 / \bar{r}_1$ are equilibrium points of the system.

A.6 Dynamics of the pair of weights

Equations (5.4) can be expressed in vectorial form as

$$\begin{bmatrix} \dot{w}_1 \\ \dot{w}_2 \end{bmatrix} = \begin{bmatrix} A\,\bar{r}_1 & -B\,\bar{r}_2 \\ -B\,\bar{r}_1 & A\,\bar{r}_2 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} + C\,\bar{r}_1\,\bar{r}_2\,. \tag{A.35}$$

The eigenvalues of the above matrix of coefficients that govern the dynamics of the system, are

$$\lambda_{\pm} = \frac{A\,(\bar{r}_1 + \bar{r}_2) \pm \sqrt{A^2\,(\bar{r}_1 - \bar{r}_2)^2 + 4B^2\,\bar{r}_1\,\bar{r}_2}}{2} \tag{A.36}$$

with the associated eigenvectors $[1 \quad \frac{A\,r_1 - \lambda_{\pm}}{B\,r_2}]^{\mathrm{T}}$.

In the case of balanced STDP ($A = B$), one eigenvalue is zero and the other is positive. Therefore, the weights drift away from the equilibrium line. In the case of potentiation-dominated STDP ($A > B$), both eigenvalues are positive, so the fixed point is unstable. In the case of depression-dominated STDP ($A < B$), one eigenvalue is positive and the other negative, so the fixed point is a saddle node.

Assuming that the baseline rates are equal ($\bar{r}_1 = \bar{r}_2 = \bar{r}$), for the case of potentiation-dominated STDP, the point where w_1 hits its maximum and $dw_2/dt = 0$ (i.e. the end point of the trajectory perpendicular to the rightmost boundary, figure 5.4) is

$$w_1 = w_{max}\,,\; w_2 = \frac{B\,w_{max} - C\,\bar{r}}{A}\,. \tag{A.37}$$

This point defines the boundary between the basin of attractions (figure 5.4). As equation A.37 shows, w_2 becomes smaller when the baseline rate increases, and the

boundary moves downward. Consequently, the basin of attraction for the top-right corner (figure 5.4) grows.

In the case of potentiation-dominated STDP, the point at which w_1 hits zero and $dw_2/dt = 0$ (i.e. the trajectory perpendicular to the rightmost boundary) is

$$w_1 = 0\,,\; w_2 = \frac{-C\,\bar{r}}{A}\,. \tag{A.38}$$

This point defines the boundary between the basin of attractions (figure 5.6). C is negative in this case, thus as the baseline rate increases, w_2 becomes larger, the boundary moves upward, and the basin of the attractor at origin (figure 5.6) grows.

A.7 The drift of the pair of synapses under shifted STDP

In the shifted STDP model, only nearest-neighboring pre- and postsynaptic spike pairs participate in plasticity. We assume that the spike trains of neurons 1 and 2 are Poisson. In this case, the probability density that neuron 2 fires an interval t after a spike from neuron 1, and no other spike occurs between these two (i.e. nearest-neighbor condition), is

$$P(t)_{1\to 2} = \bar{r}_1\, r_2(t)\, \exp\Big[-\int_0^t ds\, \big(\bar{r}_1 + r_2(s)\big)\Big]\,. \tag{A.39}$$

Because the spike of neuron 1 precedes that of neuron 2, the firing rate of neuron 2 increases transiently during the pairing interval (equation A.32). Assuming that

$w_1 >> V_{\rm th} - V_r$, the above probability can be approximated as

$$\begin{aligned} P(t)_{1\to 2} &\approx \bar{r}_1\,\bar{r}_2\,\exp\Big[-(\bar{r}_1+\bar{r}_2)\,t\Big] \\ &\quad + w_1\,\frac{\bar{r}_1\,\tau_{\rm s}}{\tau_{\rm m}\,(V_{\rm th}-V_{\rm r})}\,\exp(-\bar{r}_1\,t)\,\Big[\lambda_2\,\exp(-\lambda_2 t) - \bar{r}_2\,\exp(-\bar{r}_2 t)\Big] \end{aligned} \tag{A.40}$$

where $\lambda_2 = \bar{r}_2 + 1/\tau_{\rm s}$. The drift of w_1 can now be calculated as

$$\begin{aligned} \frac{dw_1}{dt} &= A_+ \int_d^{+\infty} dt\, P(t)_{1\to 2}\,\exp\big[-(t-d)/\tau_+\big] \\ &\quad -A_- \int_d^{0} dt\; P(t)_{1\to 2}\,\exp\big[(t-d)/\tau_-\big] \\ &\quad -A_- \int_0^{+\infty} dt\; P(t)_{2\to 1}\,\exp\big[-(t-d)/\tau_-\big] \\ &= A(\bar{r}_1,\bar{r}_2)\,\bar{r}_1\,w_1 - B(\bar{r}_1,\bar{r}_2)\,\bar{r}_1\,w_2 + C(\bar{r}_1,\bar{r}_2)\,\bar{r}_1\,\bar{r}_2 \end{aligned} \tag{A.41}$$

with coefficients

$$\begin{aligned} A(\bar{r}_1,\bar{r}_2) &= \frac{1}{\tau_{\rm m}\,(V_{\rm th}-V_{\rm r})}\left(\frac{A_+\,\tau_{\rm s}\,(\tau_+ + d)\,(1+\bar{r}_1\tau_+)}{(1+\bar{r}\,\tau_+)\,(\tau_{\rm s}+\tau_+ + \bar{r}\,\tau_{\rm s}\,\tau_+)} - (A_+ + A_-)\,d\right) \\ B(\bar{r}_1,\bar{r}_2) &= \frac{A_-\,\tau_{\rm s}\,(\tau_- - d)\,(1+\bar{r}_2\tau_-)}{\tau_{\rm m}\,(V_{\rm th}-V_{\rm r})\,(1+\bar{r}\,\tau_-)\,(\tau_{\rm s}+\tau_- + \bar{r}\,\tau_{\rm s}\,\tau_-)} \\ C(\bar{r}_1,\bar{r}_2) &= \frac{A_+\,\tau_+\,(1-\bar{r}\,d)}{1+\bar{r}\,\tau_+} - \frac{A_-\,\tau_+\,(1+\bar{r}\,d)}{1+\bar{r}\,\tau_-}\,, \end{aligned} \tag{A.42}$$

where $\bar{r} = \bar{r}_1 + \bar{r}_2$. Note that, in this case, the coefficients are functions of the baseline rates.

The case $d = 0$ is equivalent to unshifted STDP with nearest-neighbor interactions. The above coefficients show that in this case the model has the same qualitative behavior as STDP with all-to-all spike interactions, because the rate-dependence does not change the fixed points and eigenvalues qualitatively. For $d > 0$ (shifted window), the coefficient A decreases and eventually becomes negative as the baseline rate increases. This is because the first term in the parenthesis, which is positive, decreases with rate, while the second term, which subtracts a positive amount, remains constant. At higher firing rates, the coefficient C also becomes negative, because its first term (positive) decreases with rate while the second term increases.

At low rates, when $0 < A < B$ and $C > 0$, both components of the fixed point (equations A.34) are positive. Also, one of the eigenvalues (equation A.36) is positive and the other is negative, so the fixed point is a saddle node. At high rates, when $A < 0 < B$ and $C > 0$, both components of the fixed point remain positive, and both eigenvalues are negative, so the fixed point is stable. At even higher firing rates, C becomes negative as well, therefore the components of the fixed point become negative, while the eigenvalues remain negative. As a result, the system has a stable fixed point out of its allowed range.

Appendix B

Miscellaneous Research Projects

B.1 A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus

Relay neurons in the lateral geniculate nucleus (LGN) receive direct visual input predominantly from a single retinal ganglion cell (RGC), in addition to indirect input from other sources including interneurons, thalamic reticular nucleus (TRN), and the visual cortex. To address the extent of inuence of these indirect sources on the response properties of the LGN neurons, we fit a Generalized Linear Model (GLM) to the spike responses of cat LGN neurons. We conclude that the indirect source of response modulation of the LGNrelay neurons arises from inhibitory sources, compatible with thalamic interneurons or TRN. The results are published in Journal of Vision (2010) 10(10):22, 1-14. A reprint is appended to the end of this dissertation.

B.2 Fast nonnegative deconvolution for spike train inference from population calcium imaging.

This work presents a fast nonnegative deconvolution filter to infer the approximately most likely spike train of each neuron, given the calcium fluorescence observations. This algorithm outperforms optimal linear deconvolution (Wiener filtering) on both simulated and biological data. My contribution was limited to implementing the "barrier method" in the optimization algorithm. The results are published in Journal of Neurophysiology (2010) 104(6): 3691-704. A reprint is appended to the end of this dissertation.

Bibliography

A Artola, W. S. (1987). Long-term potentiation and NMDA receptors in rat visual cortex. *Nature*, 330(6149):649–52.

Abbott, L. & Blum, K. (1996). Functional significance of long-term potentiation for sequence learning and prediction. *Cereb Cortex*, 6(3):406–16.

Abbott, L. & Nelson, S. (2000). Synaptic plasticity: taming the beast. *Nature Neuroscience*, 3:1178–1183.

Abraham, W. & Robins, A. (2005). Memory retention–the synaptic stability versus plasticity dilemma. *Trends in Neuroscience*, 28(2):73–8.

Andrasfalvy, B. & Magee, J. (2001). Distance-dependent increase in AMPA receptor number in the dendrites of adult hippocampal CA1 pyramidal neurons. *Journal of Neuroscience*, 21(23):9151–9159.

Babadi, B. & Abbott, L. (2010). Intrinsic stability of temporally shifted spike-timing dependent plasticity. *PLoS Computational Biology*, 6(11).

Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K., & Buzsáki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology*, 92(1):600–8.

Bell, C., Han, V., Sugawara, Y., & Grant, K. (1997). Synaptic plasticity in a cerebellum-like structure depends on temporal order. *Nature*, 387(6630):278–81.

Bender, V., Bender, K., Brasier, D., & Feldman, D. (2006). Two coincidence detectors for spike timing-dependent plasticity in somatosensory cortex. *Journal of Neuroscience*, 26(16):4166–77.

Bi, G. & Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–72.

Bliss, T. & Gardner-Medwin, A. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *Journal of Neurophysiology*, 232(2):357–74.

Bliss, T. & Lomo, T. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Neurophysiology*, 232(2):331–56.

Boettiger, C. & Doupe, A. (2001). Developmentally restricted synaptic plasticity in a songbird nucleus required for song learning. *Neuron*, 31(5):809–18.

Braitenberg, V. & Schuz, A. (1991). *Anatomy of the cortex—Statistics and geometry*. Springer-Verlag.

Brunel, N. & Sergi, S. (1998). Firing frequency of leaky integrate-and-" re neurons with synaptic current dynamics. *Journal of Theoretical Biology*.

Caporale, N. & Dan, Y. (2008). Spike timing-dependent plasticity: a hebbian learning rule. *Annual Review of Neuroscience*, 31:25–46.

Cassenaer, S. & Laurent, G. (2007). Hebbian stdp in mushroom bodies facilitates the synchronous flow of olfactory information in locusts. *Nature*, 448(7154):709–13.

Cateau, H. & Fukai, T. (2003). A stochastic method to predict the consequence of arbitrary forms of spike-timing-dependent plasticity. *Neural Computation*, 15(3):597–620.

Chapman, P., Kairiss, E., Keenan, C., & Brown, T. (1990). Long-term synaptic potentiation in the amygdala. *Synapse*, 6(3):271–8.

Chomsky, N. (1983). *Cartesian Linguistics: A Chapter in the History of Rationalist Thought*. University Press of America.

Clopath, C., Büsing, L., Vasilaki, E., & Gerstner, W. (2010). Connectivity reflects coding: a model of voltage-based stdp with homeostasis. *Nature Neuroscience*, 13(3):344–352.

Clugnet, M. & LeDoux, J. (1990). Synaptic plasticity in fear conditioning circuits: induction of ltp in the lateral nucleus of the amygdala by stimulation of the medial geniculate body. *Journal of Neuroscience*, 10(8):2818–24.

Cossart, R., Ikegaya, Y., & Yuste, R. (2005). Calcium imaging of cortical networks dynamics. *Cell Calcium*, 37(5):451–7.

Dan, Y. & Poo, M. (2006). Spike timing-dependent plasticity: from synapse to perception. *Physiological Reviews*, 86(3):1033.

Destexhe, A., Mainen, Z., & Sejnowski, T. (1994). Synthesis of models for excitable membranes, synaptic transmission and neuromodulation using a common kinetic formalism. *Journal of Computational Neuroscience*, 1(3):195–230.

Drew, P. & Abbott, L. (2006). Extending the effects of spike-timing-dependent plasticity to behavioral timescales. *Proceedings of the National Academy of Sciences of the United States of America*, 103(23):8876–81.

D'Souza, P., Liu, S., & Hahnloser, R. (2010). Perceptron learning rule derived from spike-frequency adaptation and spike-time-dependent plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 107(10):4722–7.

Egger, V., Feldmeyer, D., & Sakmann, B. (1999). Coincidence detection and changes of synaptic efficacy in spiny stellate neurons in rat barrel cortex. *Nature Neuroscience*, 2(12):1098–105.

Feldman, D. (2000). Timing-based ltp and ltd at vertical inputs to layer ii/iii pyramidal cells in rat barrel cortex. *Neuron*, 27(1):45–56.

Finger, S. (2001). *Origins of Neuroscience: A History of Explorations into Brain Function*. Oxford University Press, USA.

Froemke, R., Letzkus, J., Kampa, B., Hang, G., & Stuart, G. (2010). Dendritic synapse location and neocortical spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience*, 2.

Froemke, R., Poo, M., & Dan, Y. (2005). Spike-timing-dependent synaptic plasticity depends on dendritic location. *Nature*, 434(7030):221–225.

Froemke, R., Tsay, I., Raad, M., Long, J., & Dan, Y. (2006). Contribution of individual spikes in burst-induced long-term synaptic modification. *Journal of neurophysiology*, 95(3):1620–1629.

Froemke, R. C. & Dan, Y. (2002). Spike-timing-dependent synaptic modification induced by natural spike trains. *Nature*, 416(6879):433–238.

Fu, Y., Djupsund, K., Gao, H., Hayden, B., Shen, K., & Dan, Y. (2002). Temporal specificity in the cortical plasticity of visual space representation. *Science*, 296(5575):1999–2003.

Gerstein, G. L. & Mandelbrot, B. (2006). Random walk models for the spike activity of a single neuron. *Biophysical Journal*, 4(1):41–68.

Gerstner, W., Kempter, R., van Hemmen, J., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76–81.

Gerstner, W. & Kistler, W. (2002). *Spiking Neuron Models*. Cambridge University Press.

Gidon, A. & Segev, I. (2009). Spike-timing-dependent synaptic plasticity and synaptic democracy in dendrites. *Journal of Neurophysiology*, 101(6):3226–3234.

Gilson, M., Burkitt, A., Grayden, D., Thomas, D., & van Hemmen, J. (2009a). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. i. input selectivity–strengthening correlated input pathways. *Biological Cybernetics*, 101(2):81–102.

Gilson, M., Burkitt, A., Grayden, D., Thomas, D., & van Hemmen, J. (2009b). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. ii. input selectivity—symmetry breaking. *Biological Cybernetics*, 101(2):103–114.

Gilson, M., Burkitt, A., Grayden, D., Thomas, D., & van Hemmen, J. (2009c). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks iii: Partially connected neurons driven by spontaneous activity. *Biological Cybernetics*, 101(5):411–426.

Gilson, M., Burkitt, A., Grayden, D., Thomas, D., & van Hemmen, J. (2009d). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks iv. *Biological Cybernetics*, 101(5):427–444.

Gilson, M., Burkitt, A., Grayden, D., Thomas, D., & van Hemmen, J. (2010a). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks v: self-organization schemes and weight dependence. *Biological Cybernetics*, pages 1–22.

Gilson, M., Burkitt, A., & van Hemmen, L. (2010b). Stdp in recurrent neuronal networks. *Frontiers in Computational Neuroscience*, 4.

Grinvald, A. & Hildesheim, R. (2004). VSDI: a new era in functional imaging of cortical dynamics. *Nature Reviews Neuroscience*, 5(11):874–85.

Gütig, R., Aharonov, R., Rotter, S., & Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric hebbian plasticity. *Journal of Neuroscience*, 23:3697– 3714.

Hebb, D. O. (1949). *The Organization of Behavior*. John Wiley & Sons.

Hensch, T. (2005). Critical period plasticity in local cortical circuits. *Nature Reviews Neuroscience*, 6(11):877–888.

Herrmann, A. & Gerstner, W. (2001). Noise and the psth response to current transients: I. general theory and application to the integrate-and-fire neuron. *Journal of computational neuroscience*, 11(2):135–151.

Hirsch, J., Barrionuevo, G., & Crepel, F. (1992). Homo- and heterosynaptic changes in efficacy are expressed in prefrontal neurons: an in vitro study in the rat. *Synapse*, 12(1):82–5.

Iriki, A., Pavlides, C., Keller, A., & Asanuma, H. (1989). Long-term potentiation in the motor cortex. *Science*, 245(4924):1385–7.

Jahr, C. & Stevens, C. (1990). A quantitative description of nmda receptor-channel kinetic behavior. *Journal of Neuroscience*, 10(6):1830–1837.

Karmarkar, U. R., Najarian, M. T., & Buonomano, D. V. (2002). Mechanisms and significance of spike-timing dependent plasticity. *Biological Cybernetics*, 87(5-6):373–382.

Katz, Y., Menon, V., Nicholson, D. A., Geinisman, Y., Kath, W. L., & Spruston, N. (2009). Synapse distribution suggests a two-stage model of dendritic integration in ca1 pyramidal neurons. *Neuron*, 63(2):171–177.

Kempter, R., Gerstner, W., & van Hemmen, J. (2001). Intrinsic stabilization of output rates by spike-based hebbian learning. *Neural Computation*, 13:2709–2741.

Kozloski, J. & Cecchi, G. (2010). A theory of loop formation and elimination by spike timing-dependent plasticity. *Front. Neural Circuits*, 4.

Kuhn, A., Aertsen, A., & Rotter, S. (2003). Higher-order statistics of input ensembles and the response of simple model neurons. *Neural Computation*, 15(1):67–101.

Letzkus, J. J., Kampa, B. M., & Stuart, G. J. (2006). Learning rules for spike timing-dependent plasticity depend on dendritic synapse location. *Journal of Neuroscience*, 26(41):10420–10429.

Levy, W. & Steward, O. (1983). Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791–7.

Li, C., Lu, J., Wu, C., Duan, S., & Poo, M. (2004). Bidirectional modification of presynaptic neuronal excitability accompanying spike timing-dependent synaptic plasticity. *Neuron*, 41(2):257–68.

Liu, Q., Pu, L., & Poo, M. (2005). Repeated cocaine exposure in vivo facilitates ltp induction in midbrain dopamine neurons. *Nature*, 437(7061):1027–31.

Lubenov, E. & Siapas, A. (2008). Decoupling through synchrony in neuronal circuits with propagation delays. *Neuron*, 58(1):118–131.

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

Malenka, R. & Bear, M. (2004). Ltp and ltd:: An embarrassment of riches. *Neuron*, 44(1):5–21.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213.

Markram, H. & Tsodyks, M. (1996). Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594):807–810.

Masquelier, T., Guyonneau, R., & Thorpe, S. (2008). Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. *PLoS One*, 3(1):e1377.

Mehta, M., Lee, A., & Wilson, M. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, 417(6890):741–6.

Meliza, C. & Dan, Y. (2006). Receptive-field modification in rat visual cortex induced by paired visual stimulation and single-cell spiking. *Neuron*, 49(2):183–9.

Miller, K. (1996). Synaptic economics: Competition and cooperation in correlation-based synaptic plasticity. *Neuron*, 17:371–374.

Miller, K. & MacKay, D. (1994). The role of constraints in hebbian learning. *Neural Computation*, 6(1):100–126.

Morrison, A., Aertsen, A., & Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Computation*, 19(6):1437–1467.

Morrison, A., Diesmann, M., & Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike timing. *Biological Cybernetics*, 98(6):459–478.

Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., & Prochiantz, A. (1984). Magnesium gates glutamate-activated channels in mouse central neurones. *Nature*, 307:462–465.

O'Brien, R., Kamboj, S., Ehlers, M., Rosen, K., Fischbach, G., & Huganir, R. (1998). Activity-dependent modulation of synaptic AMPA receptor accumulation. *Neuron*, 21(5):1067–1078.

Pfister, J. & Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *The Journal of neuroscience*, 26(38):9673.

Pillow, J., Shlens, J., Paninski, L., Sher, A., Litke, A., Chichilnisky, E., & Simoncelli, E. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–9.

Pu, L., Liu, Q., & Poo, M. (2006). BDNF-dependent synaptic sensitization in midbrain dopamine neurons after cocaine withdrawal. *Nature Neuroscience*, 9(5):605–7.

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590.

Risken, H. (1996). *The Fokker-Planck equation.* Springer-Verlag. New York.

Rubin, J., Lee, D., & Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical Review Letters*, 86(2):364–367.

Rumsey, C. & Abbott, L. (2004). Equalization of synaptic efficacy by activity-and timing-dependent synaptic plasticity. *Journal of neurophysiology*, 91(5):2273–2280.

Salinas, E. & Sejnowski, T. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *Journal of Neuroscience*, 20(16):6193–6209.

Senn, W. & Buchs, N. (2003). Spike-based synaptic plasticity and the emergence of direction selective simple cells: mathematical analysis. *Journal of Computational Neuroscience*, 14(2):119–138.

Senn, W., Markram, H., & Tsodyks, M. (2001). An algorithm for modifying neurotransmitter release probability based on pre- and postsynaptic spike timing. *Neural Computation*, 13(1):35–67.

Shadlen, M. & Newsome, W. (1998). The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *Journal of Neuroscience*, 18(10):3870–3896.

Shouval, H., Bear, M., & Cooper, L. (2002). A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10831–6.

Shouval, H. & Wittenberg, G. (2010). Spike timing dependent plasticity: a consequence of more fundamental learning rules. *Frontiers in Computational Neuroscience*, 4.

Sjöström, P. & Häusser, M. (2006). A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons. *Neuron*, 51(2):227–238.

Sjöström, P., Turrigiano, G., & Nelson, S. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32(6):1149–1164.

Sjöström, P., Turrigiano, G., & Nelson, S. (2003). Neocortical LTD via coincident activation of presynaptic nmda and cannabinoid receptors. *Neuron*, 39(4):641–654.

Song, S. & Abbott, L. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, 32(2):339–350.

Song, S., Miller, K., & Abbott, L. (2000). Competitive hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, 3(9):919–26.

Song, S., Sjostrom, P., Reigl, M., Nelson, S., & Chklovskii, D. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol*, 3(3):e68.

Stepanyants, A., Hirsch, J., Martinez, L., Kisvárday, Z., Ferecskó, A., & Chklovskii, D. (2008). Local potential connectivity in cat primary visual cortex. *Cerebral Cortex*, 18(1):13–28.

Troyer, T. & Miller, K. (1997). Physiological gain leads to high isi variability in a simple model of a cortical regular spiking cell. *Neural Computation*, 9(5):971–983.

Turrigiano, G., Leslie, K., Desai, N., Rutherford, L., & Nelson, S. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892–895.

Tzounopoulos, T., Kim, Y., Oertel, D., & Trussell, L. (2004). Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. *Nature Neuroscience*, 7(7):719–25.

van Rossum, M., Bi, G., & Turrigiano, G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *Journal of neuroscience*, 20(23):8812–21.

Wang, H.-X., Gerkin, R. C., Nauen, D. W., & Bi, G.-Q. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience*, 8(2):187–193.

Yao, H., Shen, Y., & Dan, Y. (2004). Intracortical mechanism of stimulus-timing-dependent plasticity in visual cortical orientation tuning. *Proceedings of the National Academy of Sciences of the United States of America*, 101(14):5081–6.

Journal of Vision (2010) 10(10):22, 1–14 http://www.journalofvision.org/content/10/10/22 1

A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus

Baktash Babadi — Center for Theoretical Neuroscience, Columbia University, New York, NY, USA

Alexander Casti — Department of Neuroscience, The Mount Sinai School of Medicine, New York, NY, USA, & Department of Mathematics, The Cooper Union School of Engineering, New York, NY, USA

Youping Xiao — Department of Neuroscience, The Mount Sinai School of Medicine, New York, NY, USA

Ehud Kaplan — Department of Neuroscience, The Mount Sinai School of Medicine, New York, NY, USA

Liam Paninski — Center for Theoretical Neuroscience, Columbia University, New York, NY, USA, & Department of Statistics, Columbia University, New York, NY, USA

Relay neurons in the lateral geniculate nucleus (LGN) receive direct visual input predominantly from a single retinal ganglion cell (RGC), in addition to indirect input from other sources including interneurons, thalamic reticular nucleus (TRN), and the visual cortex. To address the extent of influence of these indirect sources on the response properties of the LGN neurons, we fit a Generalized Linear Model (GLM) to the spike responses of cat LGN neurons driven by spatially homogeneous spots that were rapidly modulated by a pseudorandom luminance sequence. Several spot sizes were used to probe the spatial extent of the indirect visual effects. Our extracellular recordings captured both the LGN spikes and the incoming RGC input (S potentials), allowing us to divide the inputs to the GLM into two categories: the direct RGC input and the indirect input to which we have access through the luminance of the visual stimulus. For spots no larger than the receptive field center, the effect of the indirect input is negligible, while for larger spots its effect can, on average, account for 5% of the variance of the data and for as much as 25% in some cells. The polarity of the indirect visual influence is opposite to that of the linear receptive field of the neurons. We conclude that the indirect source of response modulation of the LGN relay neurons arises from inhibitory sources, compatible with thalamic interneurons or TRN.

Keywords: lateral geniculate nucleus (LGN), S potentials, generalized linear model (GLM), retinogeniculate transmission

Citation: Babadi, B., Casti, A., Xiao, Y., Kaplan, E., & Paninski, L. (2010). A generalized linear model of the impact of direct and indirect inputs to the lateral geniculate nucleus. *Journal of Vision, 10*(10):22, 1–14, http://www.journalofvision.org/content/10/10/22, doi:10.1167/10.10.22.

Introduction

As the prototypical sensory nucleus of the visual thalamus, the dorsal lateral geniculate nucleus (LGN) has been the subject of numerous studies for decades. However, the precise role of the LGN in processing visual information remains a matter of debate (Carandini et al., 2005; Mante, Bonin, & Carandini, 2008; Mayo, 2009; Rees, 2009; Sherman, 2005; Sherman & Guillery, 2002). One reason for this is the complicated feedforward and feedback circuitry that drives and modulates relay cell responses (Sherman & Guillery, 1998), which makes isolation of the separate inputs difficult. Despite this, much progress has been made in demonstrating that nonretinal inputs to the relay cells significantly alter LGN activity, and enough is known to make a plausible case for the functional role of some of these effects (Andolina, Jones, Wang, & Sillito, 2007; McAlonan, Cavanaugh, & Wurtz, 2008; Rees, 2009; Wang et al., 2007). To say that the LGN acts as more than a "mere relay" is now a cliche of neuroscience literature, replacing the older one to the contrary. The aim of this work is to probe the influence of nonretinal inputs with a quantitative statistical model and to assess their importance in shaping fine details of LGN response.

doi: 10.1167/10.10.22 Received April 6, 2010; published August 23, 2010 ISSN 1534-7362

Thalamic relay neurons receive strong feedforward excitation predominantly from one or two retinal ganglion cells (RGCs) with overlapping receptive fields (Cleland, Dubin, & Levick, 1971; Usrey, Reppas, & Reid, 1999), from which the LGN neuron inherits the dominant features of its concentric, center–surround receptive field. A fundamental difference between the RGC and LGN receptive fields is a stronger, spatially extended inhibitory surround in the LGN, observed long ago by Hubel and Wiesel (1961). This enhanced inhibition likely originates in either local interneurons (Blitz & Regehr, 2005; Dubin & Cleland, 1977) or processes emanating from the thalamic reticular nucleus (TRN; Wang et al., 2001), with which the LGN is known to be reciprocally connected. The LGN and TRN also receive excitatory feedback from the cortex; together, the LGN, TRN, and cortex form a closed circuit in which each area has a reciprocal connection to the others. Thus, a key issue is the extent to which these extraretinal sources influence the visual information processing of the LGN neurons.

Any model of the retinogeniculate pathway (Carandini, Horton, & Sincich, 2007; Casti, Hayot, Xiao, & Kaplan, 2008) must account for the fact that the LGN transmits only a fraction of the retinal inputs it receives (see Figure 1). It is clear from recordings that a single spike from the retina is sufficient (perhaps in conjunction with unseen inputs) to drive a relay cell past threshold, yet the fact is that in anesthetized preparations, on average, less than half of the primary driver ganglion cell inputs trigger an LGN spike (Kaplan, Purpura, & Shapley, 1987). The LGN compensates for its lower firing rate by making each spike more informative (Sincich, Horton, & Sharpee, 2009; Uglesich, Casti, Hayot, & Kaplan, 2009). This suggests that the mechanisms responsible for repressing LGN activity do so with a purpose and are not random.

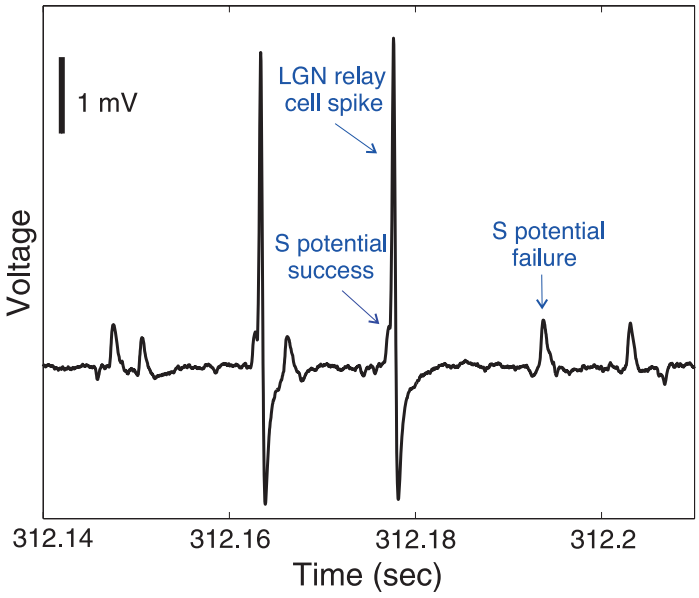

Figure 1. A sample extracellular voltage recording of an LGN neuron, corresponding to the cell in Figure 5B. This record shows seven S potentials, which are large EPSPs driven by RGC input spikes. A "failed" S potential does not have a concomitant LGN spike, while a "successful" S potential is typically embedded within an LGN spike.

The trend in recent studies of retinogeniculate (RG) transmission has been to find the simplest model or mechanism capable of explaining the variance in LGN relay cell recordings (Carandini et al., 2007; Casti et al., 2008; Sincich et al., 2009). Here, we take a different approach: we try to quantify the indirect visual influences on the RG transmission, no matter how modest, and examine its properties and physiological origin. By "indirect" visual influences we mean any visual input to the LGN relay cell beyond monosynaptic RG transmission. In our model, we take advantage of single-cell extracellular recordings of LGN neurons that capture the timing of incoming retinal spikes in the form of S potentials (Bishop, 1953; Kaplan & Shapley, 1984). We thus have access to both the main input and the output of single neurons in the thalamus. Since the visual stimulus is under our experimental control, it is possible, by fitting a realistic neural model to these data, to examine whether the response of the LGN neuron to the visual stimulus is entirely governed by the direct monosynaptic retinogeniculate transmission, or whether some aspects of the LGN response are dictated by indirect sources of visual input. This is the underlying rationale of the method used in the present study. By fitting the parameters of a generalized linear model (GLM; Paninski, Pillow, & Lewi, 2007; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005) to the extracellular data, we show that the visual stimulus influences the response of the LGN neurons by paths other than the monosynaptic RG transmission. The time scales and spatial extent of the indirect contributions we found are consistent with feedforward inhibition from thalamic interneurons or feedback inhibition from the TRN.

Methods

Surgery

The experimental methods were similar to those described in Casti et al. (2008) in accordance with the National Institutes of Health Guidelines and the Mount Sinai School of Medicine Institutional Animal Care and Use Committee. In brief, adult cats were anesthetized initially with an intramuscular (IM) injection of xylazine (Rompun, 2 mg/kg) followed by ketamine hydrochloride (Ketaset, 10 mg/kg). Anesthesia was maintained with a mixture of propofol (4 mg/kg h) and sufentanil (0.05 μg/kg h), which was given intravenously (IV) during the experiment. Propofol anesthesia has been shown to cause no changes in blood flow in the occipital cortex (Fiset et al.,

1999) and appears to be optimal for brain studies. Phenylephrine hydrochloride (10%) and atropine sulfate (1%) were applied to the eyes. The animal was mounted in a stereotaxic apparatus. The corneas were protected with plastic gas-permeable contact lenses, and a 3-mm diameter artificial pupil was placed in front of each eye. Blood pressure, ECG, and body temperature were measured and kept within the physiological range. Paralysis was produced by an infusion of pancuronium bromide (0.25 mg/kg h), and the animal was artificially respired. Such preparations are usually stable in our setup for more than 96 h.

Visual stimuli

Visual stimuli were presented monocularly on a CRT (mean luminance of 25 cd/m^2; frame rate of 160 Hz) driven by a VSG 2/5 stimulator (Cambridge Research Systems, Cambridge, UK). Stimuli consisted of spatially homogeneous circular spots of various diameters, ranging from 0.5° to full field, modulated temporally according to a pseudorandom sequence (Reinagel & Reid, 2000; van Hateren, 1997). For each spot size, we presented a sequence of 256 stimulus segments of random luminance modulation, each 8 s long, in which 128 repeated segments (repeats) were interleaved with 128 nonrepeating segments (uniques). The entire stimulus run thus lasted 8 × 256 = 2048 s, during which the spot size was fixed. A filtered version of the repeated segment is shown in the top panel of Figure 4A.

Recording of LGN spikes and S potentials

Extracellular recordings were taken from layers A and A1 of the LGN of 6 adult anesthetized cats. Amplified electrical signals were sampled at 40 kHz by a data acquisition interface (Spike 2, CED) for subsequent spike recording and sorting. To map the receptive field (RF) of LGN relay neurons, we first moved a mouse-controlled light or dark bar on the CRT to find the approximate position of the RF, and then reverse correlated the spike train with a 16 × 16 checkerboard, in which each check (spanning 0.25° at 57 cm, in each linear dimension) was modulated by an independent *m*-sequence (Reid, Victor, & Shapley, 1997; Shutter, 1987). Neurons were classified as X or Y based on the responses to contrast reversal of fine gratings (Hochstein & Shapley, 1976). None of our Y cell recordings were sufficiently stable to be used in this work, so all the model results presented are for X cells. All cells were within 15° of the *area centralis*. The RF center size was estimated by fitting a Difference of Gaussians (DOG) model to the spatial response map that resulted from the reverse correlation procedure. The center radius was taken to be twice the standard deviation of the Gaussian fit. In the figures, spot sizes are reported as a multiple of the estimated size of the RF center and are referred to as relative spot sizes.

When the electrode tip is sufficiently close to the LGN cell body, the retinal input to the relay cell can be captured in the form of slow synaptic potentials (S potentials), as shown in Figure 1 (Bishop, 1953; Kaplan & Shapley, 1984). Because the electrode tip needs to be so close to the cell, we used finely tapered, high-resistance (10 MΩ) tungsten electrodes with platinum black tip conditioning (FHC; epoxylite insulation). Despite the danger of damaging the cell in this type of recording, we were often able to hold the recordings for 12 h or more with no apparent degradation in visual responsiveness. Owing to the length of each recording (>4 h) and the difficulty in obtaining S potentials that are easily discriminable over baseline noise, we typically were able to get just 1 or 2 successful recordings per animal that were suitable for the analysis presented here.

Spike times were determined offline by setting two threshold voltages: a low threshold of approximately 0.5 mV for the S potentials and a higher threshold for the LGN spikes. An event time was defined by the threshold crossing at half-height. As illustrated in Figure 1, a single retinal S potential has variable success in eliciting an LGN relay cell response. A retinal input "failure" appears as an isolated, small-amplitude event, and a "success" manifests itself as a shoulder on the rising phase of the large-amplitude relay cell spike it triggers. The S potential is fused with the relay cell spike to various degrees, sometimes separated entirely from the action potential by ~1 ms (Sincich, Adams, Economides, & Horton, 2007; Weyand, 2007). In cases where a retinal input induced an LGN spike, our method of detecting S potentials created a slight delay between the S potential and the LGN spike time, ranging from 0 to 1 ms, depending on the extent to which the S potential was embedded within the LGN spike. For a small fraction (<5%) of LGN events, the presence of a spike-triggering S potential is not detectable in the extracellular trace either within the spike or nearby. In these cases, it is possible that the most recent S potential served to "prime" the relay cell, allowing another, perhaps smaller, input to drive it past threshold. Whatever the case, whether an S potential is detectable or not, we assume in the model that each LGN spike has an S potential accompanying it. The validity of this assumption has received experimental support in recordings from both anesthetized (Sincich et al., 2007) and awake (Weyand, 2007) preparations. In all the cells analyzed, successive S potentials arrived at intervals >2 ms, making it unlikely that there were two or more dominant drivers from separate ganglion cells.

Modeling and data analysis

To estimate distinct contributions to the response of LGN neurons, we fit a Generalized Linear Model (GLM;

Paninski, 2004; Paninski et al., 2007) to the stimulus and spike train data. The GLM in this case has the following form:

$$\lambda_t = f\left(b + \sum_{i=1}^{\dim D} D_i x_{t-i+1} + \sum_{j=1}^{\dim H} H_j n_{t-j} + \sum_{m=1}^{\dim K} K_m l_{t-m+1}\right). \quad (1)$$

The inputs to the model are x_t, n_t, and l_t, which represent discrete time series for the RGC spikes (S potentials), LGN spikes, and the luminance of the visual stimulus at time t, respectively. A distinct linear temporal filter is convolved with each source of input to the model: The filter $\vec{D}$ acts on the RGC spikes (x_t), the filter $\vec{H}$ acts on the past LGN spikes (n_t) and models the spike-history effects on the present activity of the neuron, and the filter $\vec{K}$ acts on the luminance of the visual stimulus (l_t). The parameter b is a constant offset that defines the background firing rate of the LGN neuron. Because of the presence of the constant b, which represents the sum of all constant inputs to the function f, the filters are mainly responsive to deviations of the corresponding inputs around their means. In line with this inherent separation of mean and variance in the model and in order to make the interpretation of the filter $\vec{K}$ more clear, we subtracted the mean luminance (25 cd/m^2) from the values l_t at each time. After convolving the inputs with the filters, the result is fed into a nonlinear, monotonically increasing function f to calculate the instantaneous firing rate λ_t of the LGN neuron at time t. The main role of f is to capture nonlinear thresholding effects. Finally, the number of spikes in each time bin of duration d_t is drawn from a Poisson distribution such that $n_t \sim \text{Poiss}(\lambda_t, dt)$. Figure 2 depicts a schematic structure of the GLM and its components. An important feature of the GLM used here is that the visual input enters through two distinct routes: first, from the RGC spikes (x_t) with its corresponding temporal filter $\vec{D}$, and second, through the luminance of the visual stimuli (l_t), with its corresponding filter $\vec{K}$ (which we call the "luminance" or "indirect" filter). The rationale for incorporating the latter in the model is to account for possible information about the visual stimulus that affects the response of the LGN neuron but is not directly mediated by the RGC spike train to the LGN neuron. Thus, the filter $\vec{D}$ represents the monosynaptic RG transmission, whereas $\vec{K}$ filters stimulus information beyond that directly transmitted to the LGN by the retina (such as cortical feedback and intrageniculate inhibition).

In order to simplify the notation, the inputs to the model can be lumped into a single vector as

$$\vec{X}_t = [1 \quad x_t \ \ldots \ x_{t-\dim(D)+1} \quad n_{t-1} \ \ldots \ n_{t-\dim(H)} \quad l_t \ \ldots \ l_{t-\dim(K)+1}]. \quad (2)$$

The parameters of the model, i.e., the constant offset and the linear temporal filters, can also be lumped into the single vector

$$\vec{\theta} = [b \quad \vec{D} \quad \vec{H} \quad \vec{K}]. \quad (3)$$

Now Equation 1 can be rewritten as

$$\lambda_t = f(\vec{\theta} \cdot \vec{X}_t). \quad (4)$$

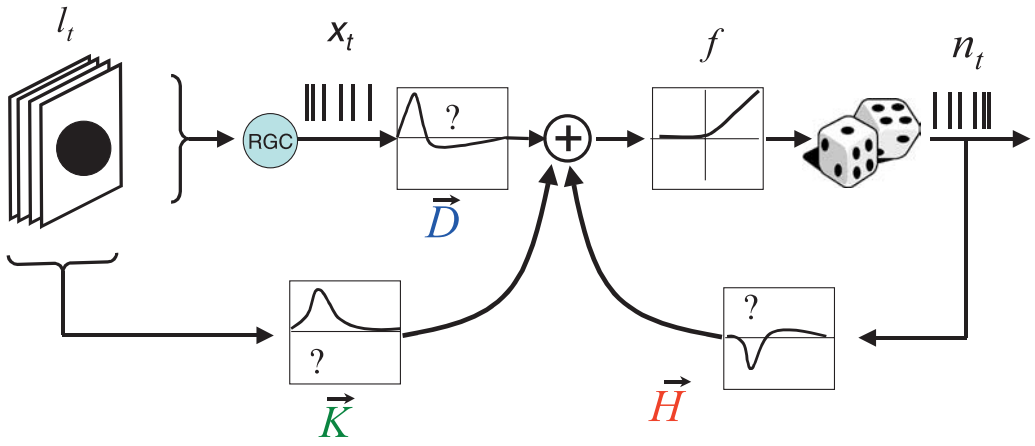

Figure 2. A schematic of the GLM that is fit to the recorded LGN data. The inputs to the model are the RGC input events x_t (S potential times; Figure 1), the LGN cell spike time history n_t, and the luminance of the visual stimulus l_t. The linear filters acting on the inputs are $\vec{D}$ (blue), $\vec{H}$ (red), and $\vec{K}$ (green). A static nonlinearity f (Equation 6) transforms the sum of filtered inputs to obtain the instantaneous firing rate of the neuron. The output spikes are generated as an inhomogeneous Poisson process with the rate parameter λ_t given by Equation 1. The model inputs derive from the experimental data, while the linear filters are given by maximizing the likelihood of the model that reproduces the experimental LGN spike train. The temporal filter $\vec{D}$ represents the monosynaptic retinogeniculate transmission, and the filter $\vec{K}$ captures the effects of indirect visual inputs.

The log-likelihood L of the GLM producing the observed LGN spike train (O) can be written as in Snyder and Miller (1991):

$$L = \log p\left(O|\vec{X},\vec{\theta}\right) = c + \sum_t \left(n_t \log f\left(\vec{\theta}\cdot\vec{X}_t\right) - f\left(\vec{\theta}\cdot\vec{X}_t\right)dt\right), \quad (5)$$

where c is a constant unrelated to the model parameters. If $f(u)$ is a convex function of its scalar argument u, and $\log f(u)$ is concave in u, then the above log-likelihood is guaranteed to be a concave function of the parameter $\vec{\theta}$, since in this case the log-likelihood is just a sum of concave functions of $\vec{\theta}$ (Paninski, 2004). This ensures that the likelihood has a unique maximum for some parameter vector $\vec{\theta}_{ML}$, which can be found easily by numerical ascent techniques. We use the standard Hessian-based estimate for the standard error of the optimization: diag $[(\nabla^2_{\vec{\theta}_{ML}} L)^{-1}]^{\frac{1}{2}}$ provides error bars for each corresponding entry in the parameter vector $\vec{\theta}$, where $\nabla^2_{\vec{\theta}_{ML}} L$ denotes the Hessian of the log-likelihood evaluated at the maximum likelihood estimate θ_{ML} (Paninski, 2004; Paninski et al., 2007; Truccolo et al., 2005; see Figures 3 and 5). The function $f(u)$ used here that satisfies the above conditions is

$$f(u) = \log(1 + e^u). \quad (6)$$

Other choices of convex and log-concave functions (including exponential and quadratic) did not alter the results of this study.

For each stimulus size, the recorded data consist of LGN responses to 128 unique and 128 repeated trials. The parameters of the model ($\vec{\theta}$) are optimized by maximizing the log-likelihood (Equation 5) of the model in reproducing the spike train of the unique trials. Thereafter, the repeated trials (which are not part of the GLM training data) are used to cross-validate the model. To assess the quality of the model fit to the observed data, we calculated the percentage of the variance in the poststimulus time histogram (PSTH) of the recorded LGN data accounted for by the GLM for the repeated trials (Pillow, Paninski, Uzzell, Simoncelli, & Chichilnisky, 2005):

$$r = 100 \times \left(1 - \frac{\left\langle (\mathrm{PSTH}_{\mathrm{data}} - \mathrm{PSTH}_{\mathrm{model}})^2 \right\rangle}{\left\langle (\mathrm{PSTH}_{\mathrm{data}} - \langle \mathrm{PSTH}_{\mathrm{data}} \rangle)^2 \right\rangle}\right), \quad (7)$$

where $\mathrm{PSTH}_{\mathrm{data}}$ and $\mathrm{PSTH}_{\mathrm{model}}$ refer to the PSTH of the laboratory data and the GLM, respectively, and $\langle\cdot\rangle$ denotes a time average across trials. We computed the PSTH using a bin width of 6.25 ms, equivalent to the frame interval of the 160 Hz stimulus presentation on the CRT. For each experimental condition, two GLMs are fit to the observed LGN spike trains. First, a full model that contains all the temporal filters ($\vec{D}$, $\vec{H}$, and $\vec{K}$); and second, a GLM that

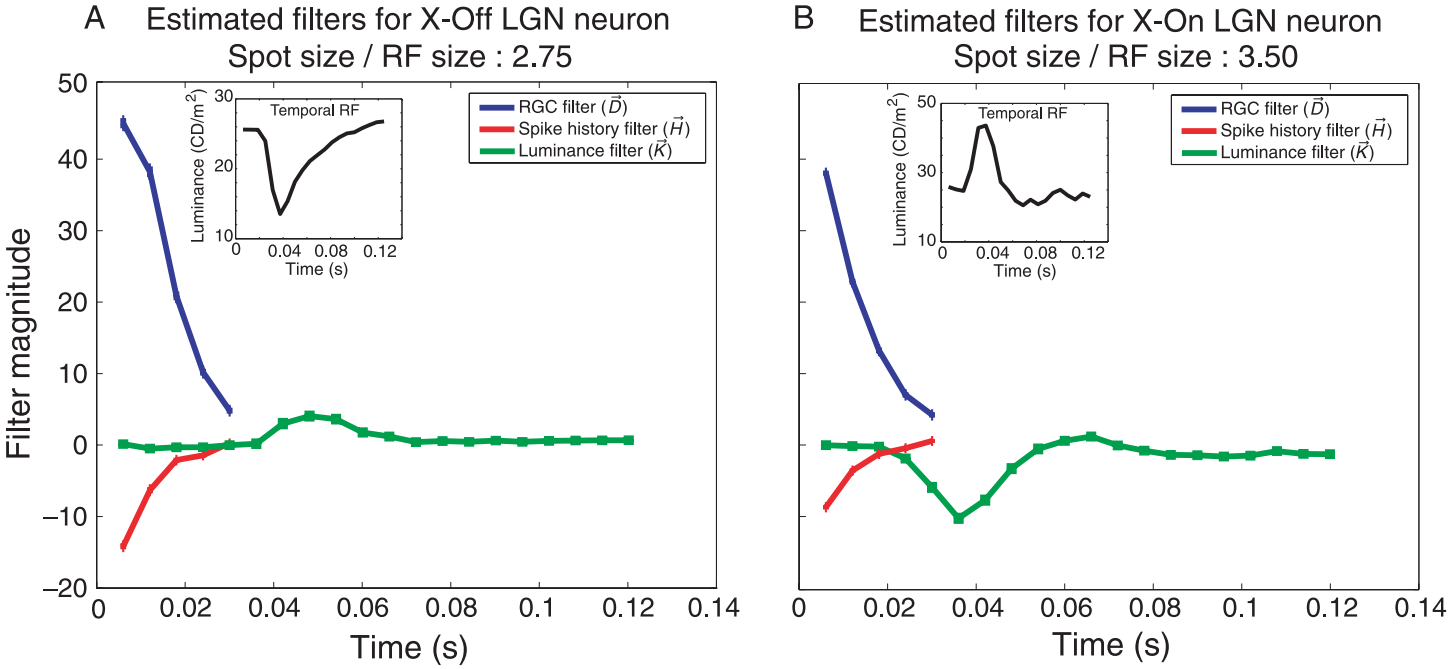

Figure 3. An example of the optimized linear temporal filters for (A) an X-Off LGN neuron and (B) an X-On LGN neuron. The error bars indicate the standard error of the optimization for each point (see Methods section). The time course of the response of each cell, obtained by reverse correlation with checkerboard *m*-sequence stimuli (Recording of LGN spikes and S potentials section), is shown in the insets. Each filter is normalized by the standard deviation of its corresponding input, so that the magnitude of the three filters may be compared directly.

lacks the indirect filter $\vec{K}$. Comparison of these two models provides the contribution of the indirect or extraretinal routes of visual input to the LGN response.

Note that the time courses of the three filters $\vec{D}$, $\vec{H}$, and $\vec{K}$ indicate LGN spike probability contributions of different events that happen at different times. Relative to the "zero time" (the time at which the filters are convolved with the input and spike histories), a peak in the retinal filter $\vec{D}$ roughly corresponds to the most recent postsynaptic S potential event at the LGN cell body. A trough in the $\vec{H}$ filter is associated with the suppression that generally follows the most recent LGN action potential, and a peak or trough in the indirect kernel $\vec{K}$ is associated with the stimulus history as it appeared on the laboratory monitor. To get an estimate of the relative time delay between relay cell responses and any indirect inputs, we compare the time evolution of the $\vec{K}$ filter with the time evolution of the LGN linear receptive field at the RF center, as given by reverse correlation with a spatiotemporal *m*-sequence (see Recording of LGN spikes and S potentials section).

Results

We optimized the parameters of the GLM for 10 recorded LGN neurons (6 X-Off and 4 X-On cells). Figure 3 shows the optimized temporal filters for two representative LGN neurons. Before discussing the contributions of each model component to the LGN response, we evaluate the overall performance of the model in capturing the statistical properties of the recorded LGN spike trains.

Statistics of the LGN discharge

Figure 4A juxtaposes the response of a recorded X-Off LGN neuron and the GLM fit to the repeated trials. As can be seen, the full GLM can accurately reproduce the instantaneous firing rate and the variance of the firing rate of the real neuron across different trials. The inter-spike-interval (ISI) distribution of the spikes generated by the GLM also matches that of the real neuron (Figure 4B); similar results were observed for the other analyzed cells (data not shown).

Apart from the statistical parameters of the LGN response, the GLM can capture other important features of LGN response as well. For instance, in both the real data and the model, the LGN neuron produces fewer spikes than the RGC neuron. Thus, each S potential is not necessarily accompanied by an LGN spike (Figure 4A, rasters). Some S potentials succeed in being transmitted by LGN to the visual cortex and some fail (Carandini et al., 2007; Casti et al., 2008; Uglesich et al., 2009); in this data

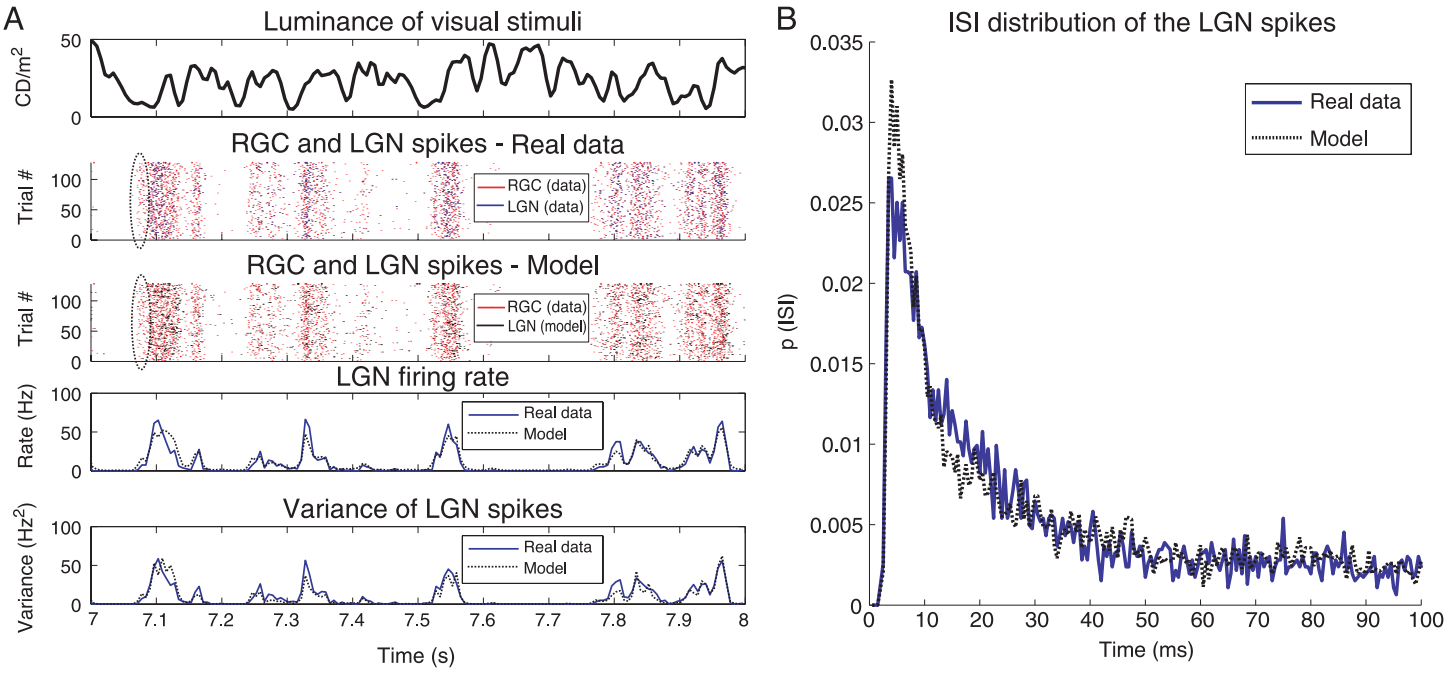

Figure 4. The spike train of the GLM compared to the recorded (real) spike train of the LGN X-Off neuron from Figure 3A, for the 128 repeated trials. (A) The luminance of the visual stimulus is shown in the top panel; the RGC spikes (red) and the real LGN spikes (blue) are shown in the second panel. In the third panel, the same RGC spikes (red) and the spike trains of the model LGN neuron (black) are illustrated. The ovals highlight the tendency of the RGC spikes to cluster before the LGN spikes both in the real data and the model. The fourth panel shows the instantaneous firing rate of the real (blue) and model (dotted black) LGN neurons averaged across trails. The bottom panel shows the variance of the firing rate for real (blue) and model (dotted black) LGN neuron across trials. (B) The inter-spike-interval (ISI) distribution of the spike trains of the real (blue) and model (dotted black) LGN neurons.

set, the "transfer ratio" (the number of LGN spikes divided by the number of retinal S potentials) was 0.24 (i.e., about one quarter of S potentials triggered an LGN spike). The GLM predicted a similar transfer ratio of ~0.24. Both in the real data and in the model, the failed S potentials tend to cluster prior to successful S potentials accompanied by an LGN spike (Figure 4A, ovals); this phenomenon is well explained by temporal summation of RGC inputs, as has been shown previously (Carandini et al., 2007; Casti et al., 2008; see the Discussion section).

The GLM filters

For all spot sizes and for both RF center polarities (On and Off), the time course of the estimated RGC filter $\vec{D}$ is approximately an exponential decay with a time constant of ~20 ms (Figure 3, blue curves). The spike history filter $\vec{H}$ is also approximately exponential but is negative (suppressive) and has a shorter time constant (~5 ms) for all spot sizes (Figure 3, red curves). The filters $\vec{D}$ and $\vec{H}$ were traced back up to 30 ms prior to the stimulus time t. For times longer than that, these filters were effectively zero, and taking them into account only slowed down the optimization process. Moreover, using the repeated trials to cross-validate the model with filters having more time bins showed that adding redundant time bins to the model filters did not improve the fit, and eventually made the fit worse. This can be due to over-fitting of the model to the details of the noise in the data. With these considerations in mind, we found a 30-ms history to be appropriate for the GLM optimization.

The luminance filter $\vec{K}$ is most significant in time intervals beyond the range of the $\vec{D}$ and $\vec{H}$ filters: It has a peak around 40–45 ms after stimulus onset for the Off cells (Figure 3A) and a trough around the same time for the On cells (Figure 3B). To give this result some physiological context, we compared the time evolution of $\vec{K}$ with the first-order temporal linear kernel (i.e., the temporal receptive field; Figure 3, insets). The linear receptive field has a peak for the On cells and a trough for the Off cells, both appearing about 35 ms after stimulus onset. Thus, the $\vec{K}$ filter is about 5–10 ms delayed compared to the peak response of the LGN neurons. The $\vec{K}$ filter represents the combined effect of the visual response of a presynaptic neuron plus the synaptic transmission to the LGN cell. Therefore, any positive or negative fluctuations in $\vec{K}$ can have a variety of interpretations. A positive excursion (a peak) in the $\vec{K}$ filter can be due to an excitatory input with an On origin, or an inhibitory input with an Off origin. Similarly, a negative excursion (a trough) in the $\vec{K}$ filter can be attributed to an excitatory Off or an inhibitory On input. Alternatively, since the $\vec{K}$ filter is tuned to variations of the luminance around its mean, a peak can be interpreted as the removal of an excitatory Off or inhibitory On input.

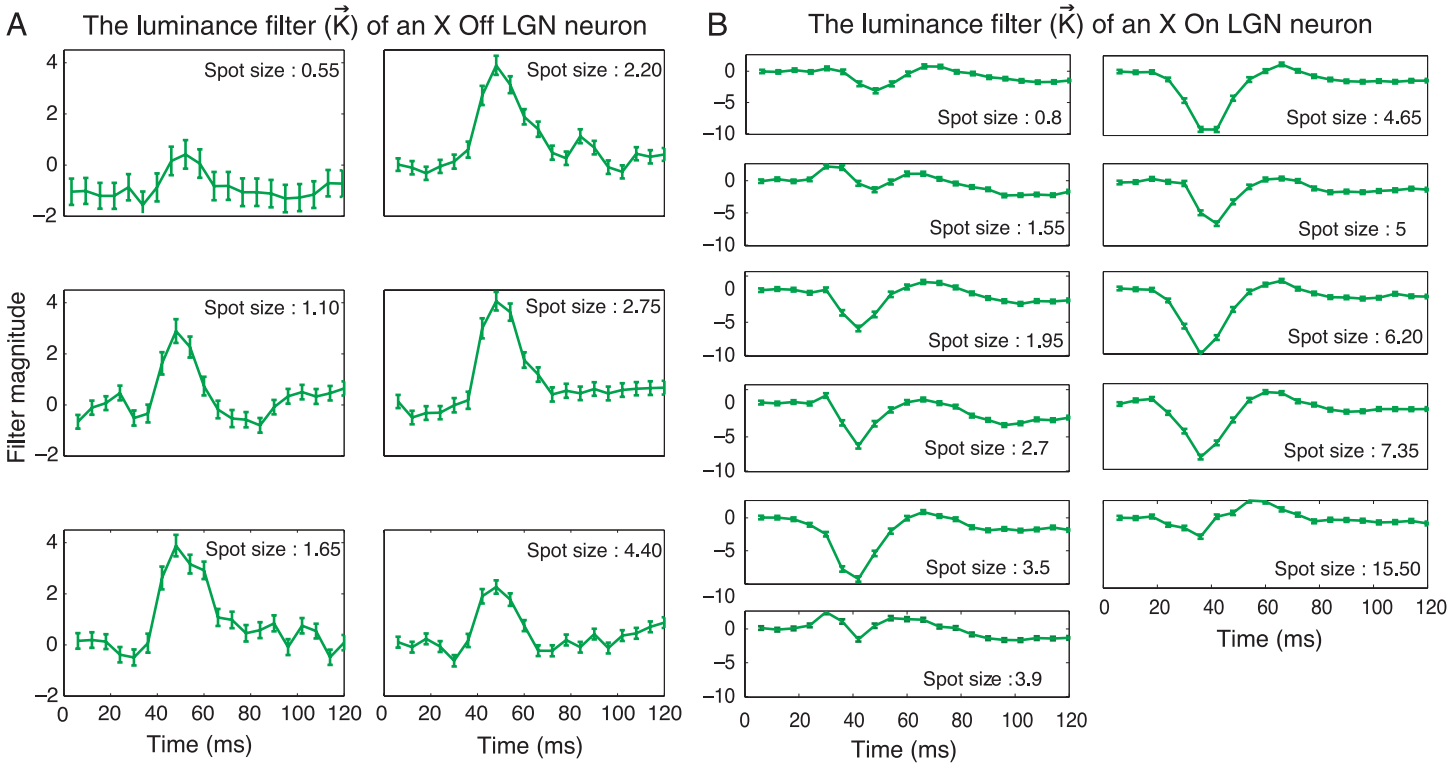

Figure 5. The luminance filter $\vec{K}$ for several spot sizes (relative to the RF center size), (A) for the representative Off cell corresponding to Figure 3A and (B) for a representative On cell corresponding to Figure 3B. The error bars show the standard error of the optimization for each point (see Methods section).

Finally, a trough can be interpreted as the removal of an excitatory On or inhibitory Off input. Although a linear model (up to spike generation) like ours cannot in principle distinguish between increased inhibition and decreased excitation, in what follows we will discuss all the above possibilities in light of the available literature about thalamic circuitry and argue that the indirect visual influence corresponds to an inhibitory source, possibly with an opposite polarity to the RF center of the LGN relay cell (see Discussion section).

While the RGC and spike history filters do not vary with the size of the visual stimulus, the luminance filter $\vec{K}$ changes as a function of spot size. Figure 5 shows the progression of the $\vec{K}$ filter for different visual spot sizes, corresponding to the representative Off and On cells in Figure 3. The magnitude of the luminance filter $\vec{K}$ is almost zero for small spot sizes and becomes more significant as spot size increases. The magnitude of the $\vec{K}$ filter tends to shrink again for even larger spot sizes. Among other things, this suggests the absence of indirect excitation or inhibition for small spot sizes and enhanced inhibition at larger spot sizes. If this indirect effect were due to decreased excitation, then we would have expected to see evidence of indirect excitation at small spot sizes. It is worth noting that in some cells the $\vec{K}$ filter may show a polarity reversal for later times (e.g., Figure 3B). This is expected as the luminance filter itself indirectly originates from the activity of an RGC whose temporal RF is typically biphasic.

Contribution of direct visual input

To quantify the indirect contribution of luminance information through the $\vec{K}$ filter on the RG transmission—as opposed to the contribution from the RGC input and its associated $\vec{D}$ filter—we fit a GLM without the $\vec{K}$ filter to have a baseline for comparison. We then calculated the performance of the full model and the model without the $\vec{K}$ filter according to Equation 7. Figure 6A shows the performance of the GLMs for a representative Off cell, and Figure 6B illustrates the same results for a representative On cell (corresponding to Figures 3 and 5, respectively). Incorporating the luminance filter $\vec{K}$ into the model generally improves the fit, particularly for mid-range stimulus sizes. This finding is more evident in Figures 6C and 6D, where the contribution of $\vec{K}$ filter is demonstrated as a function of the stimulus size. For the mid-ranged spot sizes, the $\vec{K}$ filter increases the performance of the GLM up to 7% for both cells.

The bar charts in Figure 7 show the average performance of the GLMs with and without the luminance filter

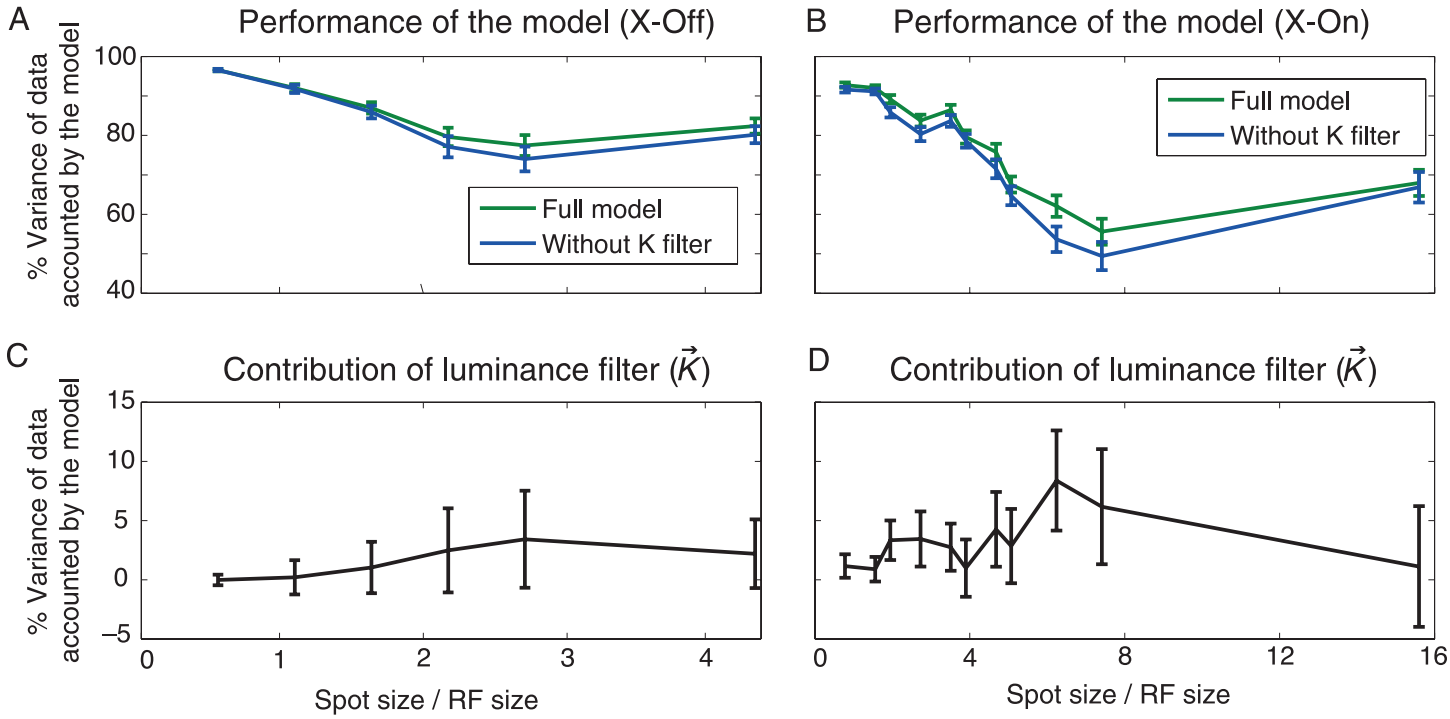

Figure 6. The performance of the GLMs and the contribution of the luminance filter ($\vec{K}$). (A) The performance of the GLM in reproducing the real LGN spike trains in the repeated trials, calculated according to Equation 7. The graph shows the performance of the full GLM (green) and the GLM without $\vec{K}$ filter (blue) as a function of the relative size of the stimulus spot, for the same Off cell as in Figures 3A and 5A. (B) Similar results for the representative On cell of Figures 3B and 5B. (C) The contribution of the luminance filter $\vec{K}$ in improving the performance of the GLM for the representative Off cell. (D) The same results for the representative On cell. Error bars indicate the standard deviation over the 128 repeated trials in each case. Note that the contribution of the luminance filter is fairly small.

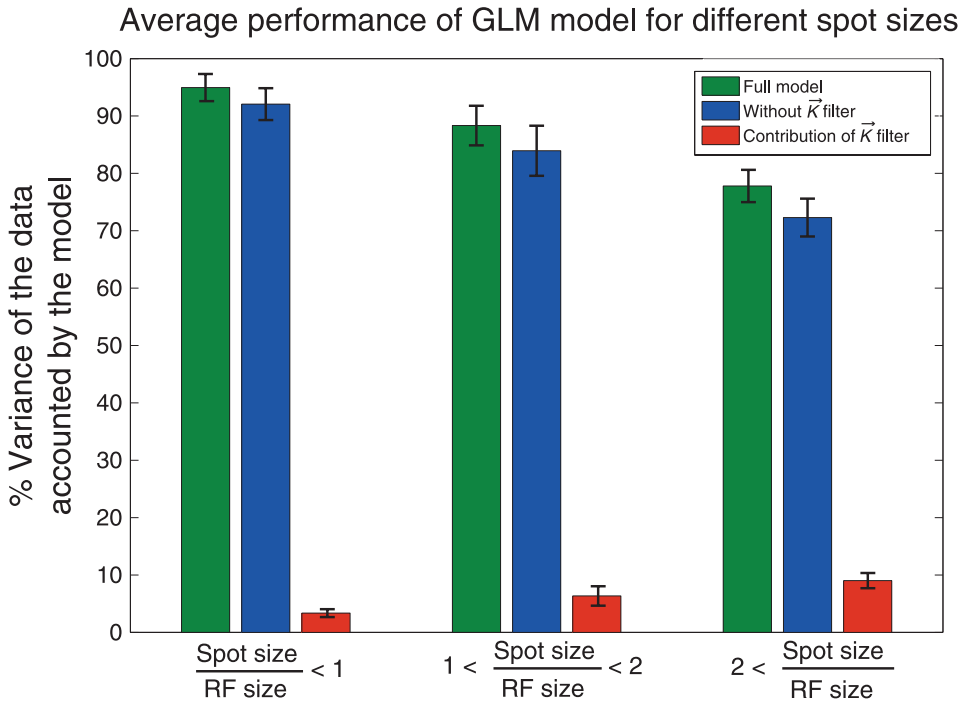

Figure 7. The average performance of the GLMs and the contribution of the luminance filter ($\vec{K}$) for different relative spot sizes. The performance of the model is calculated according to Equation 7. The GLM predicts the LGN response better for small stimuli. The contribution of the luminance filter ($\vec{K}$) increases for larger stimuli. The error bars depict the standard error.

$\vec{K}$, together with the contribution of $\vec{K}$ for three ranges of the relative stimulus size for all the recorded cells. In general, the model performs better for smaller spots and the contribution of $\vec{K}$ filter increases for larger spots. Figure 8 shows the contribution of the luminance filter $\vec{K}$ as a function of the stimulus size for all 10 recorded LGN neurons. Here also a clear trend is visible: the $\vec{K}$ filter contributes to the LGN response predominantly for stimuli that are larger than the receptive field center. The maximum contribution of the $\vec{K}$ filter can reach up to 25% for some cells, though in most cases the contribution is smaller. The data from some of the cells analyzed here were also used in previous studies from our group (Casti et al., 2008; Uglesich et al., 2009), as indicated in the inset of Figure 8.

Discussion

In order to quantify the effect of indirect visual inputs to the LGN relay neurons, we fit a generalized linear model to spike trains recorded from these neurons, along with their RGC inputs (as reflected by the recorded S potentials), while the size of the visual stimulus was varied systematically. We found that the indirect input sources are most significant for stimulus sizes that were more than two times larger than the diameter of the RF center of LGN neurons.

The GLM class of models is a natural mathematical extension of the basic physiological concept of a "receptive field" that also includes spike-history effects such as refractoriness, burstiness, and adaptation. Because of the separation of linear and nonlinear components and the incorporation of spike-history effects into the linear component, the GLM is well suited for modeling neural spike trains and gives an intuitive explanation of the underlying neuronal processes. In fact, the GLM has proved useful in a wide variety of experimental preparations (Okatan, Wilson, & Brown, 2005; Pillow et al., 2005, 2008; Truccolo et al., 2005) and has been shown in many cases to be of comparable accuracy to the more widely used integrate-and-fire (IF) model (Paninski et al., 2007). In the particular case of LGN response, IF model neurons with a biologically realistic circuitry have been used to predict experimental data similar to those of the present study (Casti et al., 2008). The accuracy of the GLM is quite comparable to that of the IF model in this case (i.e., more than 90% variance captured for small spots and around 80% for large spot stimuli). The IF model has the advantage of matching more closely the physiological intuition about the neural circuits, at the expense of losing the speed and guarantee of finding the optimal parameters, which are inherent in the GLM. The GLM retains most of the appealing biophysical interpretations of the individual model components of the IF type models and very often offers significant computational savings in the optimization. Besides the general properties of the GLM, what made this model an

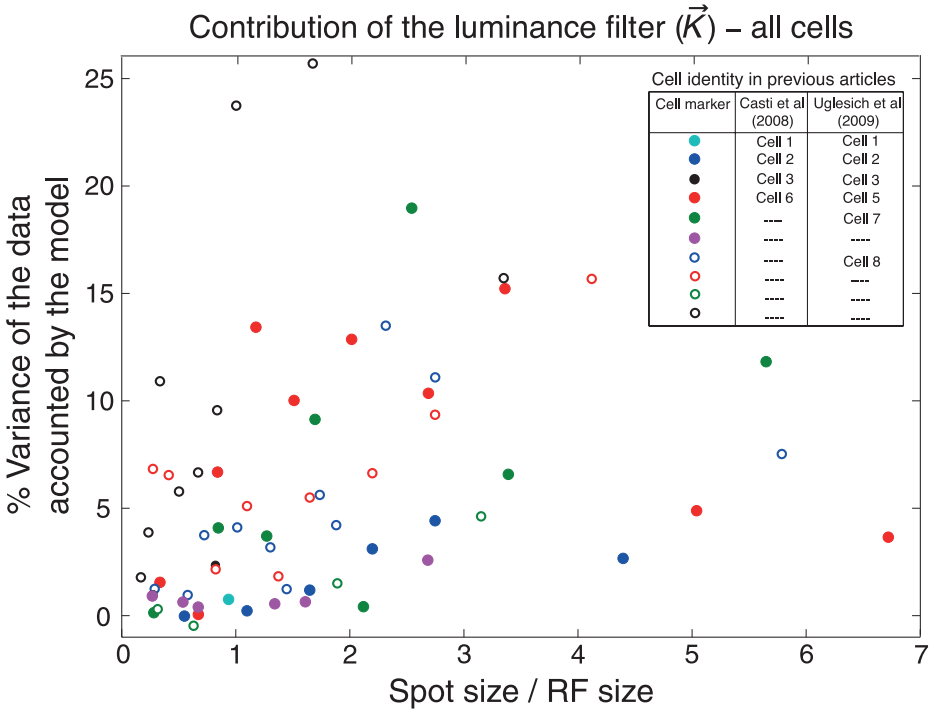

Figure 8. The contribution of the luminance filter $\vec{K}$ to the performance of the GLM for all 10 analyzed LGN neurons. The horizontal axis indicates the relative spot size. The vertical axis indicates the contribution of the $\vec{K}$ filter (in the percentage units illustrated in Figure 6). Each marker shows the results for a single neuron and a specific stimulus size. The filled circles represent Off cells and open circles represent On cells. Each color corresponds to a distinct cell. The inset shows the identity of the cells that were also analyzed in two previous articles of our group.

appropriate choice in the present study was the availability of the retinal S potentials, which allowed us to separate in the model the direct retinal input from the other, less easily detected inputs. In the present study, we detected a statistically significant, though modest, contribution of inhibitory inputs to the performance of a generalized linear model. That we were able to capture these effects highlights an important distinction between GLM and IF models: while the performance of the IF model is restricted by the specifics of the implemented circuitry (such as fixed synaptic delay, polarity of the inhibitory input, etc.), the components of the GLM are more phenomenological and do not require an *a priori* implementation of the specific biophysical properties. It is, therefore, more flexible in revealing the factors influencing the neural response.

Both the exponentially decaying shape and the time course of the estimated RGC filter $\vec{D}$ here are consistent with retinogeniculate synaptic transmission. The results show that on average the RGC input component of the model accounts for more than 90% of the variance of the observed LGN activity in all stimulus conditions. Although the contribution of indirect inputs in the LGN visual response may reach up to 25% for some neurons, it is generally small compared to the monosynaptic retinogeniculate transmission. This result is in line with previous studies showing that the LGN response is predominantly governed by the RGC driving input and the postsynaptic temporal summation of S potentials (Carandini et al., 2007; Casti et al., 2008; Cudeiro & Sillito, 1996; Sincich et al., 2007). In particular, although the S potentials that fail to elicit an LGN spike depolarize the membrane potential of the LGN neuron, they are not sufficient to push it over the spiking threshold. However, these "failures" make it easier for subsequent S potentials to evoke an LGN spike. Therefore, the tendency of the RGC spikes to cluster before the LGN spikes, as shown in Figure 4, could be explained solely by temporal summation of RGC inputs (Carandini et al., 2007; Casti et al., 2008). The same short-term temporal summation has been shown (Sincich et al., 2009; Uglesich et al., 2009) to account for the majority of the formal information (bits) transmitted by LGN relay neurons. These recent studies have shown that LGN relay neurons reencoded features of the visual stimulus so that each output spike carried more information about the stimulus than each RGC spike. Further, Sincich et al. (2009) found that the most informative dimension in the vector space of RGC inputs corresponds to a monotonically increasing function that starts to deviate from zero around 35 ms before the LGN spike. This analysis is compatible with the temporal summation of RGC spikes within a 35-ms window before the LGN spike.

In our GLM, all potential indirect inputs are lumped into the single luminance filter $\vec{K}$. Although the contribu-

tion of the indirect inputs in shaping the LGN response is modest compared to the direct retinogeniculate transmission, our results enable us to put forward a few conjectures about its origin. As mentioned before, the GLM per se cannot distinguish between excitatory or inhibitory origin of the $\vec{K}$ filter. Although there is evidence that LGN relay neurons may receive weak excitatory input from a few RGC cells other than their main excitatory drive (Cleland et al., 1971; Hamos, van Horn, Raczkowski, & Sherman, 1987; Usrey et al., 1999; Wang et al., 2007), the receptive fields of those RGCs are tightly overlapping and have the same polarity as the main RGC/LGN neuron (Usrey et al., 1999). The luminance filter $\vec{K}$ has an excitatory effect on Off cells and an inhibitory effect on On cells, which was observed in all 10 neurons we analyzed. Its polarity is therefore opposite to that of the temporal linear receptive field of the LGN neuron. An excitatory origin with the same polarity as the LGN neuron cannot explain the reversal of the polarity of the observed $\vec{K}$ filter. The maximum effect of the $\vec{K}$ filter has ~40–45 ms delay with respect to the stimulus onset (Figure 3), while the peak response of the linear temporal receptive fields of LGN neurons occurs ~35 ms after the stimulus onset, giving an approximate relative delay of ~5–10 ms between the retinal excitation at the LGN cell body and the indirect input. Again, the 5–10 ms delay of the peak (or the trough in case of On cells) of the $\vec{K}$ filter is not compatible with a feedforward excitatory input from other RGCs. When all spot sizes are considered, our results are consistent with the interpretation of these indirect effects as inhibition, as we found no evidence for indirect excitation at small spot sizes (~0.5–1 times the RF center size). The greater impact of the contribution of the $\vec{K}$ filter for larger stimuli implies that the receptive field of the inhibitory source is either larger than the receptive field of the LGN neuron itself or displaced toward its receptive field periphery.

Sources for inhibition

The above-mentioned properties of the observed indirect visual input seem consistent with a local feedforward interneuronal source or feedback inhibition from the TRN (or both). Interneuronal inhibition was found to be prominent in the intracellular recordings from LGN relay neurons by Wang et al. (2007), who had access to both the excitatory (EPSCs) and inhibitory (IPSCs) inputs to visually driven relay neurons. They mapped the RF of both inhibitory and excitatory inputs and showed that the inhibitory input most likely arises from thalamic interneurons with an overlapping center–surround receptive field structure similar to the LGN relay neurons (Dubin & Cleland, 1977; Humphrey & Weller, 1988; Sherman & Friedlander, 1988). It has been shown that the polarity of the receptive field of the inhibitory input is opposite to that of the LGN target, resulting in a pull inhibitory effect (Wang et al., 2007). Our results deviate from this strict pull inhibitory input. For instance, in the case of pull inhibition to an Off LGN neuron, the inhibitory interneuron responds to On transitions in the stimulus (dark to bright). Therefore, such an interneuron would inhibit the LGN cell for bright spot stimuli, and the $\vec{K}$ filter would have to be negative even for small spot stimuli. However, we observed no significant $\vec{K}$ filter influence for small spots, and for larger spots, it is positive, contrary to what one would expect of pull inhibition. One possible explanation is that the $\vec{K}$ filter represents a partial removal of a pull inhibitory input for larger spots that hit the surround of the RF of the inhibitory interneuron. Neurons in the TRN, on the other hand, have large receptive fields that are often less well organized compared to LGN relay neurons (Uhlrich, Cucchiaro, Humphrey, & Sherman, 1991). The TRN activity is also a plausible source for the inhibition we observe since it tends to be anti-correlated with LGN responses. This has been observed for retinotopically matched LGN and TRN cells in anesthetized cat (Funke & Eysel, 1998) and awake monkey (McAlonan et al., 2008) preparations. The 5–10 ms delay between the observed inhibition and the relay cell peak response could accommodate a round trip to V1 and back to the LGN through the TRN (Briggs & Usrey, 2007).

The feedforward inhibitory input arising from intrathalamic interneurons is classified into two distinct types with presumably different physiological roles (Blitz & Regehr, 2005). The "locked" feedforward inhibition, which is observed in ~33% of relay cells, is tightly correlated with the excitatory input and is believed to arise from the same RGC as the excitatory input. In contrast, the "nonlocked" feedforward inhibition seen in ~67% of relay cells is believed to originate from RGCs other than the main excitatory drive of the LGN cells. The locked variety of inhibition has been suggested as a mechanism for sharpening the precision of LGN responses, while the unlocked type provides surround inhibition to LGN receptive fields (Blitz & Regehr, 2005). Since the maximum effect of the inhibitory input in our data is observed when the size of the spot stimulus is larger than the receptive field center, this suggests that the indirect visual influence we observe tallies better with the nonlocked feedforward inhibition that arises from RGC neurons in the vicinity of the predominant excitatory driver of the LGN neuron. This does not imply the complete absence of a form of locked inhibition driven by the primary RGC input, but in the context of our GLM and stimulus paradigm, it means that such influences contribute little to the model performance or cannot be disambiguated from the spike history filter $\vec{H}$. If pull inhibition played a prominent role in the generation of LGN burst events, as implied by the results of Wang et al. (2007), who used natural stimuli, then perhaps this form of inhibition would have been captured in our model if we had used a different stimulus. Further, a time scale of

5–10 ms before the onset of the inhibition seems too long to implicate a locked form of inhibition and would be more consistent with disynaptic inhibition originating in other ganglion cells that might be spread across multiple interneurons with variable delays. Richer stimuli may reveal further details about the shape, location, and origin of the indirect visual inputs to the LGN relay neurons. Note also that we did not include Y cells in the present study as they did not yield stable recordings of S potentials. Whether the indirect inputs play a similar role in those cells is an open question.

Due to the limitations of extracellular recording and the phenomenological nature of the GLM, it is difficult to draw more conclusive interpretations about the exact origin of the indirect visual input from the present results. Nevertheless, our results show that the indirect input contributes to the output of the LGN neuron and is more significant for larger stimuli. To resolve the exact nature of the indirect visual influence, intracellular recordings with membrane potential of LGN neuron clamped at a high depolarizing value will be required, to reveal the IPSCs more distinctly than was possible in previous studies (Wang et al., 2007).

Conclusions

The indirect visual inputs to LGN relay neurons have a modest influence on the visual response. Their contribution is larger for large stimuli, with a polarity opposite to that of the receptive field of the LGN neuron itself. The properties of the indirect inputs are compatible with several possibilities: feedforward inhibition from thalamic interneurons innervated by the same RGC that provides the main drive to the relay cell (locked inhibition), feedforward inhibition from interneurons driven by RGCs other than the main driver (nonlocked inhibition), or feedback projections from the thalamic reticular nucleus. Further experiments with prolonged extracellular recording with spatiotemporal noise stimuli or natural images, or intracellular recording from LGN neurons, or a combination of electrophysiological and anatomical techniques will be needed to determine the exact source and the spatial extent of these inputs.

Acknowledgments

The authors gratefully acknowledge the following grant support. AC: K25 MH67225; EK: NEI EY016371, EY16224, NIGM71558, and Core Grant EY12867; LP: Sloan Research Fellowship and NSF CAREER Award.

Commercial relationships: none.
Corresponding author: Baktash Babadi.
Email: bb2280@columbia.edu.
Address: Center for Theoretical Neuroscience, Columbia University, 1051 Riverside Dr., New York, NY 10032, USA.

References

Andolina, I. M., Jones, H. E., Wang, W., & Sillito, A. M. (2007). Corticothalamic feedback enhances stimulus response precision in the visual system. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 1685–1690.

Bishop, P. O. (1953). Synaptic transmission; an analysis of the electrical activity of the lateral geniculate nucleus in the cat after optic nerve stimulation. *Proceedings of the Royal Society of London B: Biological Sciences, 141,* 362–392.

Blitz, D. M., & Regehr, W. G. (2005). Timing and specificity of feed-forward inhibition within the LGN. *Neuron, 45,* 917–928.

Briggs, F., & Usrey, W. M. (2007). A fast, reciprocal pathway between the lateral geniculate nucleus and visual cortex in the macaque monkey. *Journal of Neuroscience, 27,* 5431–5436.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., et al. (2005). Do we know what the early visual system does? *Journal of Neuroscience, 25,* 10577–10597.

Carandini, M., Horton, J. C., & Sincich, L. C. (2007). Thalamic filtering of retinal spike trains by postsynaptic summation. *Journal of Vision, 7*(14):20, 1–11, http://www.journalofvision.org/content/7/14/20, doi:10.1167/7.14.20. [PubMed] [Article]

Casti, A., Hayot, F., Xiao, Y., & Kaplan, E. (2008). A simple model of retina–LGN transmission. *Journal of Computerized Neuroscience, 24,* 235–252.

Cleland, B. G., Dubin, M. W., & Levick, W. R. (1971). Simultaneous recording of input and output of lateral geniculate neurones. *Nature—New Biology, 231,* 191–192.

Cudeiro, J., & Sillito, A. M. (1996). Spatial frequency tuning of orientation-discontinuity-sensitive corticofugal feedback to the cat lateral geniculate nucleus. *The Journal of Physiology, 490,* 481–492.

Dubin, M. W., & Cleland, B. G. (1977). Organization of visual inputs to interneurons of lateral geniculate nucleus of the cat. *Journal of Neurophysiology, 40,* 410–427.

Fiset, P., Paus, T., Daloze, T., Plourde, G., Meuret, P., Bonhomme, V., et al. (1999). Brain mechanisms of propofol-induced loss of consciousness in humans:

A positron emission tomographic study. *Journal of Neuroscience, 19,* 5506–5513.

Funke, K., & Eysel, U. T. (1998). Inverse correlation of firing patterns of single topographically matched perigeniculate neurons and cat dorsal lateral geniculate relay cells. *Visual Neuroscience, 15,* 711–729.

Hamos, J. E., van Horn, S. C., Raczkowski, D., & Sherman, S. M. (1987). Synaptic circuits involving an individual retinogeniculate axon in the cat. *Journal of Computerized Neurology, 259,* 165–192.

Hochstein, S., & Shapley, R. M. (1976). Quantitative analysis of retinal ganglion cell classifications. *The Journal of Physiology, 262,* 237–264.

Hubel, D. H., & Wiesel, T. N. (1961). Integrative action in the cat's lateral geniculate body. *The Journal of Physiology, 155,* 385–398.

Humphrey, A. L., & Weller, R. E. (1988). Structural correlates of functionally distinct x-cells in the lateral geniculate nucleus of the cat. *Journal of Computerized Neurology, 268,* 448–468.

Kaplan, E., Purpura, K., & Shapley, R. M. (1987). Contrast affects the transmission of visual information through the mammalian lateral geniculate nucleus. *The Journal of Physiology, 391,* 267–288.

Kaplan, E., & Shapley, R. (1984). The origin of the S (slow) potential in the mammalian lateral geniculate nucleus. *Experimental Brain Research, 55,* 111–116.

Mante, V., Bonin, V., & Carandini, M. (2008). Functional mechanisms shaping lateral geniculate responses to artificial and natural stimuli. *Neuron, 58,* 625–638.

Mayo, J. P. (2009). Intrathalamic mechanisms of visual attention. *Journal of Neurophysiology, 101,* 1123–1125.

McAlonan, K., Cavanaugh, J., & Wurtz, R. H. (2008). Guarding the gateway to cortex with attention in visual thalamus. *Nature, 456,* 391–394.

Okatan, M., Wilson, M., & Brown, E. (2005). Analyzing functional connectivity using a network likelihood model of ensemble neural spiking activity. *Neural Computation, 17,* 1927–1961.

Paninski, L. (2004). Maximum likelihood estimation of cascade point-process neural encoding models. *Network, 15,* 243–262.

Paninski, L., Pillow, J., & Lewi, J. (2007). Statistical models for neural encoding, decoding, and optimal stimulus design. *Progressive Brain Research, 165,* 493–507.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *Journal of Neuroscience, 25,* 11003–11013.

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., et al. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature, 454,* 995–999.

Rees, G. (2009). Visual attention: The thalamus at the centre? *Current Biology, 19,* R213–R214.

Reid, R., Victor, J., & Shapley, R. (1997). The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. *Visual Neuroscience, 14,* 1015–1027.

Reinagel, P., & Reid, R. C. (2000). Temporal coding of visual information in the thalamus. *Journal of Neuroscience, 20,* 5392–5400.

Sherman, S. M. (2005). Thalamic relays and cortical functioning. *Progressive Brain Research, 149,* 107–126.

Sherman, S. M., & Friedlander, M. J. (1988). Identification of x versus y properties for interneurons in the a-laminae of the cat's lateral geniculate nucleus. *Experimental Brain Research, 73,* 384–392.

Sherman, S. M., & Guillery, R. W. (1998). On the actions that one nerve cell can have on another: Distinguishing "drivers" from "modulators". *Proceedings of National Academy of Sciences of the United States of America, 95,* 7121–7126.

Sherman, S. M., & Guillery, R. W. (2002). The role of the thalamus in the flow of information to the cortex. *Philosophical Transactions Royal Society of London B: Biological Sciences, 357,* 1695–1708.

Shutter, E. E. (1987). A practical nonstochastic approach to nonlinear time-domain analysis. In V. Z. Marmarelis (Ed.), *Advanced methods of physiological systems modeling* (vol. 1, pp. 303–315). Los Angeles: University of Southern California.

Sincich, L. C., Adams, D. L., Economides, J. R., & Horton, J. C. (2007). Transmission of spike trains at the retinogeniculate synapse. *Journal of Neuroscience, 27,* 2683–2692.

Sincich, L. C., Horton, J. C., & Sharpee, T. O. (2009). Preserving information in neural transmission. *Journal of Neuroscience, 29,* 6207–6216.

Snyder, D., & Miller, M. (1991). *Random point processes in time and space*. New York: Springer-Verlag.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of Neurophysiology, 93,* 1074–1089.

Uglesich, R., Casti, A., Hayot, F., & Kaplan, E. (2009). Stimulus size dependence of information transfer from retina to thalamus. *Frontiers in Systems Neuroscience, 3,* 10.

Uhlrich, D. J., Cucchiaro, J. B., Humphrey, A. L., & Sherman, S. M. (1991). Morphology and axonal projection patterns of individual neurons in the cat perigeniculate nucleus. *Journal of Neurophysiology, 65,* 1528–1541.

Usrey, W. M., Reppas, J. B., & Reid, C. (1999). Specificity and strength of retinogeniculate connections. *Journal of Neurophysiology, 82,* 3527–3540.

van Hateren, J. H. (1997). Processing of natural time series of intensities by the visual system of the blowfly. *Vision Research, 37,* 3407–3416.

Wang, S., Bickford, M. E., Horn, S. C. V., Erisir, A., Godwin, D. W., & Sherman, S. M. (2001). Synaptic targets of thalamic reticular nucleus terminals in the visual thalamus of the cat. *Journal of Computerized Neurology, 440,* 321–341.

Wang, X., Wei, Y., Vaingankar, V., Wang, Q., Koepsell, K., Sommer, F. T., et al. (2007). Feedforward excitation and inhibition evoke dual modes of firing in the cat's visual thalamus during naturalistic viewing. *Neuron, 55,* 465–478.

Weyand, T. G. (2007). Retinogeniculate transmission in wakefulness. *Journal of Neurophysiology, 98,* 769–785.

Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging

Joshua T. Vogelstein, Adam M. Packer, Timothy A. Machado, Tanya Sippy, Baktash Babadi, Rafael Yuste and Liam Paninski

J Neurophysiol 104:3691-3704, 2010. First published 16 June 2010; doi:10.1152/jn.01073.2009

Journal of Neurophysiology publishes original articles on the function of the nervous system. It is published 12 times a year (monthly) by the American Physiological Society, 9650 Rockville Pike, Bethesda MD 20814-3991. ISSN: 0022-3077, ESSN: 1522-1598. Visit our website at http://www.the-aps.org/.

J Neurophysiol 104: 3691–3704, 2010.
First published June 16, 2010; doi:10.1152/jn.01073.2009.

Fast Nonnegative Deconvolution for Spike Train Inference From Population Calcium Imaging

Joshua T. Vogelstein,[1] Adam M. Packer,[2,3] Timothy A. Machado,[2,3] Tanya Sippy,[2,3] Baktash Babadi,[4] Rafael Yuste,[2,3] and Liam Paninski[4,5]

[1]*Department of Neuroscience, Johns Hopkins University, Baltimore, Maryland;* [2]*Howard Hughes Medical Institute, Chevy Chase, Maryland;* [3]*Department of Biological Sciences,* [4]*Center for Theoretical Neuroscience, and* [5]*Department of Statistics, Columbia University, New York, New York*

Submitted 9 December 2009; accepted in final form 3 June 2010

Vogelstein JT, Packer AM, Machado TA, Sippy T, Babadi B, Yuste R, Paninski L. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J Neurophysiol* 104: 3691–3704, 2010. First published June 16, 2010; doi:10.1152/jn.01073.2009. Fluorescent calcium indicators are becoming increasingly popular as a means for observing the spiking activity of large neuronal populations. Unfortunately, extracting the spike train of each neuron from a raw fluorescence movie is a nontrivial problem. This work presents a fast nonnegative deconvolution filter to infer the approximately most likely spike train of each neuron, given the fluorescence observations. This algorithm outperforms optimal linear deconvolution (Wiener filtering) on both simulated and biological data. The performance gains come from restricting the inferred spike trains to be positive (using an interior-point method), unlike the Wiener filter. The algorithm runs in linear time, and is fast enough that even when simultaneously imaging 100 neurons, inference can be performed on the set of all observed traces faster than real time. Performing optimal spatial filtering on the images further refines the inferred spike train estimates. Importantly, all the parameters required to perform the inference can be estimated using only the fluorescence data, obviating the need to perform joint electrophysiological and imaging calibration experiments.

INTRODUCTION

Simultaneously imaging large populations of neurons using calcium sensors is becoming increasingly popular (Yuste and Katz 1991; Yuste and Konnerth 2005), both in vitro (Ikegaya et al. 2004; Smetters et al. 1999) and in vivo (Göbel and Helmchen 2007; Luo et al. 2008; Nagayama et al. 2007), and will likely continue to improve as the signal-to-noise ratio (SNR) of genetic sensors continues to improve (Garaschuk et al. 2007; Mank et al. 2008; Wallace et al. 2008). Whereas the data from these experiments are movies of time-varying fluorescence intensities, the desired signal consists of spike trains of the observable neurons. Unfortunately, finding the most likely spike train is a challenging computational task, due to limitations on the SNR and temporal resolution, unknown parameters, and analytical intractability.

A number of groups have therefore proposed algorithms to infer spike trains from calcium fluorescence data using very different approaches. Early approaches simply thresholded dF/F [typically defined as $(F \quad F_b)/F_b$, where F_b is baseline fluorescence; e.g., Mao et al. 2001; Schwartz et al. 1998] to obtain "event onset times." More recently, Greenberg et al. (2008) developed a dynamic programming algorithm to identify individual spikes. Holekamp et al. (2008) then applied an optimal linear deconvolution (i.e., the Wiener filter) to the fluorescence data. This approach is natural from a signal processing standpoint, but does not realize the knowledge that spikes are always positive. Sasaki et al. (2008) proposed using machine learning techniques to build a nonlinear supervised classifier, requiring many hundreds of examples of joint electrophysiological and imaging data to "train" the algorithm to learn what effect spikes have on fluorescence. Vogelstein and colleagues (2009) proposed a biophysical model-based sequential Monte Carlo (SMC) method to efficiently estimate the probability of a spike in each image frame, given the entire fluorescence time series. Although effective, that approach is not suitable for on-line analyses of populations of neurons because the computations run in about real time per neuron (i.e., analyzing 1 min of data requires about 1 min of computational time on a standard laptop computer).

In the present work, a simple model is proposed relating spiking activity to fluorescence traces. Unfortunately, inferring the most likely spike train, given this model, is computationally intractable. Making some reasonable approximations leads to an algorithm that infers the approximately most likely spike train, given the fluorescence data. This algorithm has a few particularly noteworthy features, relative to other approaches. First, spikes are assumed to be positive. This assumption often improves filtering results when the underlying signal has this property (Cunningham et al. 2008; Huys et al. 2006; Lee and Seung 1999; Lin et al. 2004; Markham and Conchello 1999; O'Grady and Pearlmutter 2006; Paninski et al. 2009; Portugal et al. 1994). Second, the algorithm is fast: it can process a calcium trace from 50,000 images in about 1 s on a standard laptop computer. In fact, filtering the signals for an entire population of 100 neurons runs faster than real time. This speed facilitates using this filter on-line, as observations are being collected. In addition to these two features, the model may be generalized in a number of ways, including incorporating spatial filtering of the raw movie, which can improve effective SNR. The utility of the proposed filter is demonstrated on several biological data sets, suggesting that this algorithm is a powerful and robust tool for on-line spike train inference. The code (which is a simple Matlab script) is available for free download from http://www.optophysiology.org.

METHODS

Data-driven generative model

Figure 1 shows data from a typical in vitro epifluorescence experiment (for data collection details see *Experimental methods*

Address for reprint requests and other correspondence: J. T. Vogelstein, Johns Hopkins University, Department of Neuroscience, 3400 N. Charles St., Baltimore, MD 21205 (E-mail: joshuav@jhu.edu).

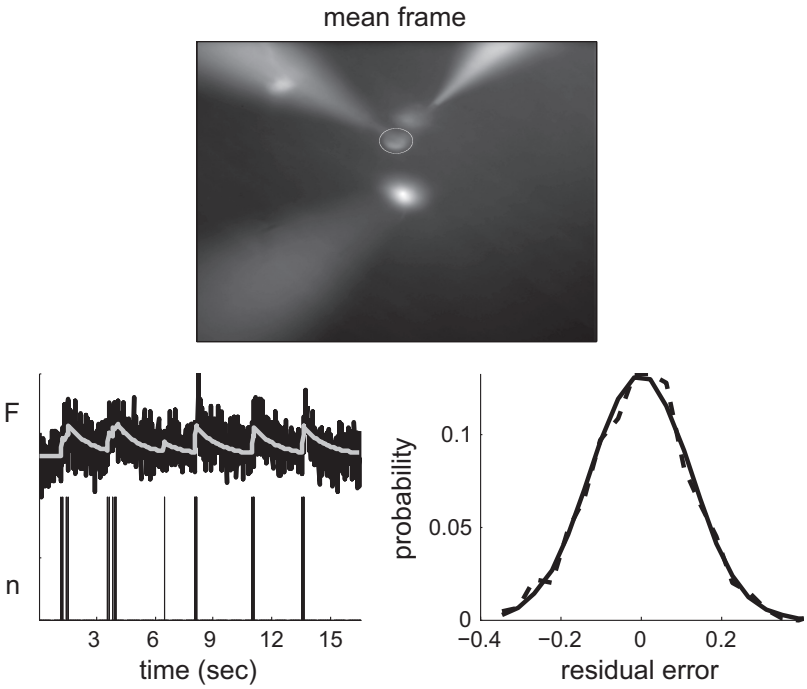

FIG. 1. Typical in vitro data suggest that a reasonable first-order model may be constructed by convolving the spike train with an exponential and adding Gaussian noise. *Top panel*: the average (over frames) of a field of view. *Bottom left*: true spike train recorded via a patch electrode (black bars), convolved with an exponential (gray line), superimposed on the Oregon Green BAPTA 1 (OGB-1) fluorescence trace (black line). Whereas the spike train and fluorescence trace are measured data, the calcium is not directly measured, but rather, inferred. *Bottom right*: a histogram of the residual error between the gray and black lines from the *bottom left panel* (dashed line) and the best-fit Gaussian (solid line). Note that the Gaussian model provides a good fit for the residuals here.

later in this section). The *top panel* shows the mean frame of this movie, including four neurons, three of which are patched. To build the model, the pixels within a region of interest (ROI) are selected (white circle). Given the ROI, all the pixel intensities of each frame can be averaged, to get a one-dimensional fluorescence time series, as shown in the *bottom left panel* (black line). By patching onto this neuron, the spike train can also be directly observed (black bars; *bottom left*). Previous work suggests that this fluorescence signal might be well characterized by convolving the spike train with an exponential and adding noise (Yuste and Konnerth 2005). This model is confirmed by convolving the true spike train with an exponential (gray line; *bottom left*) and then looking at the distribution of the residuals. The *bottom right panel* shows a histogram of the residuals (dashed line) and the best-fit Gaussian distribution (solid line).

The preceding observations may be formalized as follows. Assume there is a one-dimensional fluorescence trace $\boldsymbol{F}$ from a neuron [throughout this text $\boldsymbol{X}$ indicates the vector $(X_1, \ldots, X_T)$, where T is the index of the final frame]. At time t, the fluorescence measurement F_t is a linear-Gaussian function of the intracellular calcium concentration at that time $[\mathrm{Ca}^2\ ]_t$:

$$F_t \quad [\mathrm{Ca}^2\ ]_t \quad _t, \quad _t \overset{iid}{} \quad (0, \ ^2). \tag{1}$$

The parameter absorbs all experimental variables influencing the scale of the signal, including the number of sensors within the cell, photons per calcium ion, amplification of the imaging system, and so on. Similarly, the offset absorbs, for example, the baseline calcium concentration of the cell, background fluorescence of the fluorophore, and imaging system offset. The noise at each time $_t$ is independently and identically distributed according to a normal distribution with zero mean and ^{2} variance, as indicated by the notation $\overset{iid}{}$ (0, 1). This noise results from calcium fluctuations independent of spiking activity, fluorescence fluctuations independent of calcium, and other sources of imaging noise.

Then, assuming that the intracellular calcium concentration $[\mathrm{Ca}^2\ ]_t$ jumps by A M after each spike and subsequently decays back down to baseline C_b M, with time constant s, one can write:

$$[\mathrm{Ca}^2\ ]_t \quad _1 \quad (1 \quad / \ )[\mathrm{Ca}^2\ ]_t \quad (\ / \ )C_b \quad An_t \tag{2}$$

where is the time step size—which is the frame duration, or 1/(frame rate)—and n_t indicates the number of times the neuron spiked in frame t. Note that because $[\mathrm{Ca}^2\ ]_t$ and F_t are linearly related to one another, the fluorescence scale and calcium scale A are not identifiable. In other words, either can be set to unity without loss of generality because the other can absorb the scale entirely. Similarly, the fluorescence offset and calcium baseline C_b are not identifiable, so either can be set to zero without loss of generality. Finally, letting (1 /), *Eq. 2* can be rewritten by replacing $[\mathrm{Ca}^2\ ]_t$ with its nondimensionalized counterpart C_t:

$$C_t \quad C_t \quad _1 \quad n_t . \tag{3}$$

Note that C_t does not refer to absolute intracellular concentration of calcium, but rather, a relative measure (for a more general model see Vogelstein et al. 2009). The gray line in the *bottom left panel* of Fig. 1 corresponds to the putative $\boldsymbol{C}$ of the observed neuron.

To complete the "generative model" (i.e., a model from which simulations can be generated), the distribution from which spikes are sampled must be defined. Perhaps the simplest first-order description of spike trains is that at each time, spikes are sampled according to a Poisson distribution with some rate:

$$n_t \overset{iid}{} \text{Poisson} \ (\quad) \tag{4}$$

where is the expected firing rate per bin and is included to ensure that the expected firing rate is independent of the frame rate. Thus *Eqs. 1*, *3*, and *4* complete the generative model.

Goal

Given the above model, the goal is to find the maximum a posteriori (MAP) spike train, i.e., the most likely spike train $\hat{\boldsymbol{n}}$, given the fluorescence measurements, $\boldsymbol{F}$:

$$\hat{\boldsymbol{n}} \quad \underset{n_t \in \mathbb{N}_0 \forall t}{\operatorname{argmax}} \ P[\boldsymbol{n} \ \boldsymbol{F}], \tag{5}$$

where $P[\boldsymbol{n}|\boldsymbol{F}]$ is the posterior probability of a spike train $\boldsymbol{n}$, given the fluorescent trace $\boldsymbol{F}$, and n_t is constrained to be an integer $\mathbb{N}_0$ {0, 1,

2, . . . } because of the above assumed Poisson distribution. From Bayes' rule, the posterior can be rewritten:

$$P[\boldsymbol{n}\ \boldsymbol{F}] \quad \frac{P[\boldsymbol{n},\boldsymbol{F}]}{P[\boldsymbol{F}]} \quad \frac{1}{P[\boldsymbol{F}]}P[\boldsymbol{F}\ \boldsymbol{n}]P[\boldsymbol{n}], \tag{6}$$

where $P[\boldsymbol{F}]$ is the evidence of the data, $P[\boldsymbol{F}|\boldsymbol{n}]$ is the likelihood of observing a particular fluorescence trace $\boldsymbol{F}$, given the spike train $\boldsymbol{n}$, and $P[\boldsymbol{n}]$ is the prior probability of a spike train. Plugging the far right-hand side of *Eq. 6* into *Eq. 5*, yields:

$$\hat{\boldsymbol{n}} \quad \underset{n_t \in \mathbb{N}_0 \forall t}{\operatorname{argmax}} \frac{1}{P[\boldsymbol{F}]}P[\boldsymbol{F}\ \boldsymbol{n}]P[\boldsymbol{n}] \quad \underset{n_t \in \mathbb{N}_0 \forall t}{\operatorname{argmax}} P[\boldsymbol{F}\ \boldsymbol{n}]P[\boldsymbol{n}], \tag{7}$$

where the second equality follows because $P[\boldsymbol{F}]$ merely scales the results, but does not change the relative quality of any particular spike train. Note that the prior $P[\boldsymbol{n}]$ acts as a regularizing term, potentially imposing sparseness or smoothness, depending on the assumed distribution (Seeger 2008; Wu et al. 2006). Both $P[\boldsymbol{F}|\boldsymbol{n}]$ and $P[\boldsymbol{n}]$ are available from the preceding model:

$$P[\boldsymbol{F}\ \boldsymbol{n}] \quad P[\boldsymbol{F}\ \boldsymbol{C}] \quad \prod_{t\ 1}^{T} P[F_t\ C_t], \tag{8a}$$

$$P[\boldsymbol{n}] \quad \prod_{t\ 1}^{T} P[n_t], \tag{8b}$$

where the first equality in *Eq. 8*a follows because $\boldsymbol{C}$ is deterministic given $\boldsymbol{n}$, and the second equality follows from *Eq. 1*. Further, *Eq. 8*b follows from the Poisson process assumption, *Eq. 4*. Both $P[F_t|C_t]$ and $P[n_t]$ can be written explicitly:

$$P[F_t\ C_t] \quad (\ C_t \quad ,\ ^2), \tag{9a}$$

$$P[n_t] \quad \text{Poisson}(\quad), \tag{9b}$$

where both equations follow from the preceding model and the Poisson distribution acts as a sparse prior. Now, plugging *Eq. 9* back into *Eq. 8*, and plugging that result into *Eq. 7*, yields:

$$\hat{\boldsymbol{n}} \quad \underset{n_t \in \mathbb{N}_0 \forall t\ t\ 1}{\operatorname{argmax}}^{T} \frac{1}{\sqrt{2\ ^2}} \exp\left[\frac{1}{2}\frac{(F_t \quad C_t \quad)^2}{^2}\right] \frac{\exp[\quad(\quad)^{n_t}]}{n_t!} \tag{10a}$$

$$\underset{n_t \in \mathbb{N}_0 \forall t\ t\ 1}{\operatorname{argmax}}^{T} \frac{1}{2\ ^2}(F_t \quad C_t \quad)^2 \quad n_t \ln \quad \ln n_t!\ , \tag{10b}$$

where the second equality follows from taking the logarithm of the right-hand side and dropping terms that do not depend on $\boldsymbol{n}$. Unfortunately, solving *Eq. 10*b exactly is analytically intractable because it requires a nonlinear search over an infinite number of possible spike trains. The search space could be restricted by imposing an upper bound k on the number of spikes within a frame. However, in that case, the computational complexity scales *exponentially* with the number of image frames—i.e., the number of computations required would scale with k^T—which for pragmatic reasons is intractable.

Inferring the approximately most likely spike train, given a fluorescence trace

The goal here is to develop an algorithm to efficiently approximate $\hat{\boldsymbol{n}}$, the most likely spike train given the fluorescence trace. Because of the intractability described earlier, one can approximate *Eq. 4* by replacing the Poisson distribution with an exponential distribution of the same mean (note that potentially more accurate approximations are possible, as described in the DISCUSSION). Modifying *Eq. 10* to incorporate this approximation yields:

$$\hat{\boldsymbol{n}} \quad \underset{n_t \quad 0 \forall t\ t\ 1}{\operatorname{argmax}}^{T} \frac{1}{\sqrt{2\ ^2}} \exp\left[\frac{1}{2}\frac{(F_t \quad C_t \quad)^2}{^2}\right] (\quad)\exp[\quad n_t] \tag{11a}$$

$$\underset{n_t \quad 0 \forall t\ t\ 1}{\operatorname{argmax}}^{T} \frac{1}{2\ ^2}(F_t \quad C_t \quad)^2 \quad n_t \tag{11b}$$

where the second equality follows from taking the log of the right-hand side (logarithm is a monotone function and therefore does not change the relative likelihood of particular spike trains) and dropping terms constant in n_t. Note that the constraint on n_t has been relaxed from $n_t \in \mathbb{N}_0$ to $n_t \quad 0$ (since the exponential distribution can yield any nonnegative number). The exponential prior, much like the Poisson prior, imposes a sparsening effect, by penalizing the objective function for large values of n_t. Further, the exponential approximation makes the optimization problem concave in $\boldsymbol{C}$, meaning that any gradient ascent method guarantees achieving the global maximum (because there are no local maxima, other than the single global maximum). To see that *Eq. 11*b is concave in $\boldsymbol{C}$, rearrange *Eq. 3* to obtain $n_t \quad C_t \quad C_{t\ 1}$, so *Eq. 11*b can be rewritten:

$$\hat{\boldsymbol{C}} \quad \underset{C_t \quad C_{t\ 1} \quad 0 \forall t\ t\ 1}{\operatorname{argmax}}^{T} \frac{1}{2\ ^2}(F_t \quad C_t \quad)^2 \quad (C_t \quad C_{t\ 1}) \tag{12}$$

which is a sum of terms that are concave in $\boldsymbol{C}$, so the whole right-hand side is concave in $\boldsymbol{C}$. Unfortunately, the integer constraint has been lost, i.e., the answer could include "partial" spikes. This disadvantage can be remedied by thresholding (i.e., setting $n_t \quad 1$ for all n_t greater than some threshold and the rest setting to zero) or by considering the magnitude of a partial spike at time t as a rough indication of the probability of a spike occurring during frame t. Note the relaxation of a difficult discrete optimization problem into an easier continuous problem is a common approximation technique in the machine learning literature (Boyd and Vandenberghe 2004; Paninski et al. 2009). In particular, the exponential distribution is a convenient nonnegative log-concave approximation of the Poisson (see the DISCUSSION for more details).

Although this convex relaxation makes the problem tractable, the "sharp" threshold imposed by the nonnegativity constraint prohibits the use of standard gradient ascent techniques. This may be rectified by using an "interior-point" method (Boyd and Vandenberghe 2004). Interior-point methods solve nondifferentiable problems indirectly by instead solving a series of differentiable subproblems that converge to the solution of the original nondifferentiable problem. In particular, each subproblem within the series drops the sharp threshold and adds a weighted barrier term that approaches as n_t approaches zero. Iteratively reducing the weight of the barrier term guarantees convergence to the correct solution. Thus the goal is to efficiently solve:

$$\hat{\boldsymbol{C}}_z \quad \underset{\boldsymbol{C}\quad t\ 1}{\operatorname{argmax}}^{T} \frac{1}{2\ ^2}(F_t \quad C_t \quad)^2 \quad (C_t \quad C_{t\ 1}) \quad z \ln(C_t \quad C_{t\ 1}), \tag{13}$$

where $\ln(\cdot)$ is the "barrier term" and z is the weight of the barrier term (note that the constraint has been dropped). Iteratively solving for $\hat{\boldsymbol{C}}_z$ for z going down to nearly zero guarantees convergence to $\hat{\boldsymbol{C}}$ (Boyd and Vandenberghe 2004). The concavity of *Eq. 13* facilitates using any number of techniques guaranteed to find the global maximum. Because the argument of *Eq. 13* is twice analytically differentiable, one can use the Newton–Raphson technique (Press et al. 1992). The special tridiagonal structure of the Hessian enables each Newton–Raphson step to be very efficient (as described below). To proceed, *Eq. 13* is first rewritten in more compact matrix notation. Note that:

$$\boldsymbol{MC} \begin{bmatrix} & 1 & 0 & 0 & \cdots & 0 \\ 0 & & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & & 1 & 0 \\ 0 & \cdots & 0 & 0 & & 1 \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_{T\ 1} \\ C_T \end{bmatrix} \quad \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_{T\ 1} \end{bmatrix}, \tag{14}$$

where $\boldsymbol{M} \in \mathbb{R}^{(T\ 1)\ T}$ is a bidiagonal matrix. Then, letting $\mathbf{1}$ be a (T 1)-dimensional column vector, $\boldsymbol{\beta}$ a T-dimensional column vector of values, and $\boldsymbol{\lambda}$ $\mathbf{1}$ yields the objective function (*Eq. 13*) in more compact matrix notation (note that throughout we will use the subscript $\odot$ to indicate element-wise operations):

$$\hat{\boldsymbol{C}}_z \quad \underset{\boldsymbol{MC}\ \odot\mathbf{0}}{\operatorname{argmax}} \quad \frac{1}{2^{\ 2}} \| \boldsymbol{F} \quad \boldsymbol{\alpha C} \quad \boldsymbol{\beta} \|_2^2 \quad (\boldsymbol{MC})^{\mathrm{T}} \boldsymbol{\lambda} \quad z \ln_{\odot} (\boldsymbol{MC})^{\mathrm{T}} \mathbf{1}, \tag{15}$$

where $\boldsymbol{MC}\ _{\odot}\mathbf{0}$ indicates an element-wise greater than or equal to zero, $\ln_{\odot}(\cdot)$ indicates an element-wise logarithm, and $\|x\|_2$ is the standard L_2 norm, i.e., $\|x\|_2^2$ $_i x_i^2$. When using Newton–Raphson to ascend a surface, one iteratively computes both the gradient $\boldsymbol{g}$ (first derivative) and Hessian $\boldsymbol{H}$ (second derivative) of the argument to be maximized, with respect to the variables of interest ($\boldsymbol{C}$ here). Then, the estimate is updated using $\boldsymbol{C}_z \leftarrow \boldsymbol{C}_z$ $s\boldsymbol{d}$, where s is the step size and $\boldsymbol{d}$ is the step direction obtained by solving $\boldsymbol{Hd}$ $\boldsymbol{g}$. The gradient and Hessian for this model, with respect to $\boldsymbol{C}$, are given by:

$$\boldsymbol{g} \quad -_2(\boldsymbol{F} \quad \boldsymbol{C} \quad \boldsymbol{\beta}) \quad \boldsymbol{M}^T \boldsymbol{\lambda} \quad z\boldsymbol{M}^T (\boldsymbol{MC})_{\odot}^{\ 1} \tag{16a}$$

$$\boldsymbol{H} \quad -_2^{\ 2}\boldsymbol{I} \quad z\boldsymbol{M}^{\mathrm{T}} (\boldsymbol{MC})_{\odot}^{\ 2} \boldsymbol{M} \tag{16b}$$

where the exponents on the vector $\boldsymbol{MC}$ indicate element-wise operations. The step size s is found using "backtracking linesearches," which finds the maximal s that increases the posterior and is between 0 and 1 (Press et al. 1992).

Standard implementations of the Newton–Raphson algorithm require inverting the Hessian, i.e., solving $\boldsymbol{d}$ $\boldsymbol{H}^{\ 1}\boldsymbol{g}$, a computation that scales *cubically* with T (requires on the order of T^3 operations). Already, this would be a drastic improvement over the most efficient algorithm assuming Poisson spikes, which would require k^T operations (where k is the maximum number of spikes per frame). Here, because $\boldsymbol{M}$ is bidiagonal, the Hessian is tridiagonal, so the solution may be found in about T operations, via standard banded Gaussian elimination techniques (which can be implemented efficiently in Matlab using $\boldsymbol{H}\backslash\boldsymbol{g}$, assuming $\boldsymbol{H}$ is represented as a sparse matrix) (Paninski et al. 2009). In other words, the above approximation and inference algorithm reduces computations from *exponential* to *linear* time. APPENDIX A contains pseudocode for this algorithm, including learning the parameters, as described in the next section. Note that once $\hat{\boldsymbol{C}}$ is obtained, it is a simple linear transformation to obtain $\hat{\boldsymbol{n}}$, the approximate MAP spike train.

Learning the parameters

In practice, the model parameters $\boldsymbol{\theta}$ { , , , , } tend to be unknown. An algorithm to estimate the most likely parameters $\hat{\boldsymbol{\theta}}$ could proceed as follows: *1*) initialize some estimate of the parameters $\hat{\boldsymbol{\theta}}$, then *2*) recursively compute $\hat{\boldsymbol{n}}$ using those parameters and update $\hat{\boldsymbol{\theta}}$ given the new $\hat{\boldsymbol{n}}$ until some convergence criterion is met. This approach may be thought of as a pseudoexpectation-maximization algorithm (Dempster et al. 1977; Vogelstein et al. 2009). In the following text, details are provided for each step.

INITIALIZING THE PARAMETERS. Because the model introduced earlier is linear, the scale of $\boldsymbol{F}$ relative to $\boldsymbol{n}$ is arbitrary. Therefore before filtering, $\boldsymbol{F}$ is linearly mapped between zero and one, i.e., $\boldsymbol{F} \leftarrow (\boldsymbol{F}$ $F_{\min})/(F_{\max}$ $F_{\min})$, where $F_{\min}$ and $F_{\max}$ are the observed minimum and maximum of $\boldsymbol{F}$, respectively. Given this normalization, is set to one. Because spiking is sparse in many experimental settings, $\boldsymbol{F}$ tends to be around baseline, so is initialized to be the median of $\boldsymbol{F}$ and is initialized as the median absolute deviation of $\boldsymbol{F}$, i.e., $\text{median}_t(|F_t$ $\text{median}_s(F_s)|)/K$, where $\text{median}_i(X_i)$ indicates the median of X with respect to index i and K 1.4785 is the correction factor when using the median absolute deviation as a robust estimator of the SD of a normal distribution. Because in these data the posterior tends to be relatively flat along the dimension (i.e., large changes in result in relatively small changes in the posterior), estimating is difficult. Further, previous work has shown that results are somewhat robust to minor variations in the time constant (Yaksi and Friedrich 2006); therefore is initialized at 1 /(1 s), which is fairly standard (Pologruto et al. 2004). Finally, is initialized at 1 Hz, which is between average baseline and evoked spike rate for data of interest.

ESTIMATING THE PARAMETERS GIVEN $\hat{\boldsymbol{n}}$. Ideally, one could integrate out the hidden variables, to find the most likely parameters:

$$\hat{\boldsymbol{\theta}} \quad \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \int P[\boldsymbol{F}, \boldsymbol{C}\ \boldsymbol{\theta}] d\boldsymbol{C} \quad \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \int P[\boldsymbol{F}\ \boldsymbol{C}; \boldsymbol{\theta}] P[\boldsymbol{C}\ \boldsymbol{\theta}] d\boldsymbol{C}. \tag{17}$$

However, evaluating those integrals is not currently tractable. Therefore *Eq. 17* is approximated by simply maximizing the parameters given the MAP estimate of the hidden variables:

$$\hat{\boldsymbol{\theta}} \approx \underset{\boldsymbol{\theta}}{\operatorname{argmax}} P[\boldsymbol{F}, \hat{\boldsymbol{C}}\ \boldsymbol{\theta}] \quad \underset{\boldsymbol{\theta}}{\operatorname{argmax}} P[\boldsymbol{F}\ \hat{\boldsymbol{C}}; \boldsymbol{\theta}] P[\hat{\boldsymbol{n}}\ \boldsymbol{\theta}] \quad \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \ln P[\boldsymbol{F}\ \hat{\boldsymbol{C}}; \boldsymbol{\theta}] \quad \ln P[\hat{\boldsymbol{n}}\ \boldsymbol{\theta}], \tag{18}$$

where $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{n}}$ are determined using the above-described inference algorithm. The approximation in *Eq. 18* is good whenever most of the mass in the integral in *Eq. 18* is around the MAP sequence $\hat{\boldsymbol{C}}$.[1] The argument from the right-hand side of *Eq. 18* may be expanded:

$$\ln P[\boldsymbol{F}\ \hat{\boldsymbol{C}}; \boldsymbol{\theta}] \quad \ln P[\hat{\boldsymbol{n}}\ \boldsymbol{\theta}] \quad \sum_{t\ 1}^{T} \ln P[F_t\ \hat{C}_t; \quad , \quad , \quad] \quad \sum_{t\ 1}^{T} \ln P[\hat{n}_t \quad]. \tag{19}$$

Note that the right-hand side of *Eq. 19* decouples from the other parameters. The maximum likelihood estimate (MLE) for the observation parameters { , , } is therefore given by:

$$\left\{ \hat{\ }, \hat{\ }, \hat{\ } \right\} \quad \underset{, \ , \quad 0}{\operatorname{argmax}} \sum_{t\ 1}^{T} \ln P[F_t\ \hat{C}_t; \quad , \quad]$$

$$\underset{, \ , \quad 0}{\operatorname{argmax}} \quad \frac{1}{2}(2 \quad ^2) \quad \frac{1}{2} \left(\frac{F_t \quad \hat{C}_t}{\quad} \right)^2. \tag{20}$$

Note that a rescaling of may be offset by a complementary rescaling of $\hat{\boldsymbol{C}}$. Therefore because the scale of $\hat{\boldsymbol{C}}$ is arbitrary (see *Eqs. 2* and *3*), can be set to one without loss of generality. Plugging 1 into *Eq. 20* and maximizing with respect to yields:

$$\hat{\ } \quad \underset{0}{\operatorname{argmax}} \sum_{t\ 1}^{T} \quad (F_t \quad \hat{C}_t \quad)^2. \tag{21}$$

Computing the gradient with respect to , setting the answer to zero, and solving for $\hat{\ }$ yields $\hat{\ }$ $(1/T) \sum_t (F_t \quad \hat{C}_t)$. Similarly, computing the gradient of *Eq. 20* with respect to , setting it to zero, and solving for $\hat{\ }$ yields:

[1] *Equation 18* may be considered a crude Laplace approximation (Kass and Raftery 1995).

$$\hat{\sigma} = \sqrt{\frac{1}{T}\sum_t (F_t - \hat{C}_t - \hat{\beta})^2}, \quad (22)$$

which is simply the root-mean-square of the residual error. Finally, the MLE of λ is given by solving:

$$\hat{\lambda} = \underset{\lambda > 0}{\operatorname{argmax}} \sum_t [\ln(\lambda\Delta) - \hat{n}_t \lambda \Delta], \quad (23)$$

which, again, computing the gradient with respect to λ, setting it to zero, and solving for $\hat{\lambda}$, yields $\hat{\lambda}$ $T/(\sum_t \hat{n}_t)$, which is the inverse of the inferred average firing rate.

Iterations stop whenever *1*) the iteration number exceeds some upper bound or *2*) the relative change in likelihood does not exceed some lower bound. In practice, parameter estimates tend to converge after several iterations, given the above initializations.

Spatial filtering

In the preceding text, we assumed that the raw movie of fluorescence measurements collected by the experimenter had undergone two stages of preprocessing before filtering. First, the movie was segmented, to determine ROIs, yielding a vector $\vec{F}_t$ $(F_{1,t}, \ldots, F_{N_p,t})$, which corresponded to the fluorescence intensity at time t for each of the N_p pixels in the ROI (note that we use the $\vec{X}$ throughout to indicate row vectors in space vs. $\boldsymbol{X}$ to indicate column vectors in time). Second, at each time t, that vector was projected into a scalar, yielding F_t, the assumed input to the filter. In this section, the optimal projection is determined by considering a more general model:

$$F_{x,t} = \alpha_x C_t + \beta_x + \sigma\varepsilon_{x,t}, \quad \varepsilon_{x,t} \overset{iid}{\sim} (0, 1), \quad (24)$$

where α_x corresponds to the number of photons that are contributed due to calcium fluctuations C_t, and β_x corresponds to the static photon emission at pixel x. Further, the noise is assumed to be both spatially and temporally white, with standard deviation (SD) σ, in each pixel (this assumption can always be approximately accurate by prewhitening; alternately, one could relax the spatial independence by representing joint noise over all pixels with a covariance matrix $\quad_t$, with arbitrary structure). Performing inference in this more general model proceeds in a nearly identical manner as before. In particular, the maximization, gradient, and Hessian become:

$$\hat{\boldsymbol{C}}_z = \underset{\boldsymbol{MC} \geq_{\odot} \boldsymbol{0}}{\operatorname{argmax}} -\frac{1}{2\sigma^2}\|\vec{\boldsymbol{F}} - \boldsymbol{C}\vec{\alpha} - \boldsymbol{1}_T\vec{\beta}\|_F^2 - (\boldsymbol{MC})^{\mathrm{T}}\boldsymbol{\lambda} + z\ \ln_{\odot}(\boldsymbol{MC})^{\mathrm{T}}\boldsymbol{1} \quad (25)$$

$$\boldsymbol{g} = (\vec{\boldsymbol{F}} - \boldsymbol{C}\vec{\alpha} - \boldsymbol{1}_T\vec{\beta})^{\mathrm{T}}\frac{\vec{\alpha}^{\mathrm{T}}}{\sigma^2} - \boldsymbol{M}^{\mathrm{T}}\lambda + z\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{MC})_{\odot}^{-1} \quad (26)$$

$$\boldsymbol{H} = -\frac{\vec{\alpha}\vec{\alpha}^{\mathrm{T}}}{\sigma^2}\boldsymbol{I} - z\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{MC})_{\odot}^{-2}\boldsymbol{M}, \quad (27)$$

where $\vec{\boldsymbol{F}}$ is an N_p T element matrix, $\boldsymbol{1}_T$ is a column vector of ones with length T, $\boldsymbol{I}$ is an N_p N_p identity matrix, and $\|x\|_F$ indicates the Frobenius norm, i.e., $\|x\|_F^2$ $\sum_{i,j} x_{i,j}^2$, and the exponents and log operator on the vector $\boldsymbol{MC}$ again indicate element-wise operations. Note that to speed up computation, one can first project the background subtracted (N_c T)-dimensional movie onto the spatial filter $\vec{\alpha}$, yielding a one-dimensional time series $\boldsymbol{F}$, reducing the problem to evaluating a T 1 vector norm, as in *Eq. 15*.

The parameters $\vec{\alpha}$ and $\vec{\beta}$ tend to be unknown and thus must be estimated from the data. Following the strategy developed in the previous section, we first initialize the parameters. Because each voxel contains some number of fluorophores, which sets both the baseline fluorescence and the fluorescence due to calcium fluctuations, let both the initial spatial filter and initial background be the median image frame, i.e., $\hat{\alpha}_x$ $\hat{\beta}_x$ median$_t$ $(F_{x,t})$. Given these robust initializations, the maximum likelihood estimator for each α_x and β_x is given by:

$$\{\hat{\alpha}_x, \hat{\beta}_x\} = \underset{\alpha_x,\beta_x}{\operatorname{argmax}}\ P[\boldsymbol{F}_x|\hat{C}] \quad (28a)$$

$$= \underset{\alpha_x,\beta_x}{\operatorname{argmax}} \sum_t \ln P[F_{x,t}|\hat{C}_t] \quad (28b)$$

$$= \underset{\alpha_x,\beta_x}{\operatorname{argmax}} \sum_t \left\{ -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(F_{x,t} - \alpha_x\hat{C}_t - \beta_x)^2 \right\} \quad (28c)$$

$$= \underset{\alpha_x,\beta_x}{\operatorname{argmax}} -\sum_t (F_{x,t} - \alpha_x\hat{C}_t - \beta_x)^2, \quad (28d)$$

where the first equalities follow from *Eq. 1* and the last equality follows from dropping irrelevant constants. Because this is a standard linear regression problem, let $\boldsymbol{A}$ $[\hat{\boldsymbol{C}}, \boldsymbol{1}_T]^{\mathrm{T}}$ be a 2 T element matrix and $\boldsymbol{Y}_x$ $[\alpha_x, \beta_x]^{\mathrm{T}}$ be a 2 1 element column vector. Substituting $\boldsymbol{A}$ and $\boldsymbol{Y}_x$ into *Eq. 28*d yields:

$$\hat{\boldsymbol{Y}}_x = \underset{\boldsymbol{Y}_x}{\operatorname{argmax}} -\|\boldsymbol{F}_x - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{Y}_x\|_2^2, \quad (29)$$

which can be solved by computing the derivative of *Eq. 29* with respect to $\boldsymbol{Y}_x$ and setting to zero, or using Matlab notation: $\hat{\boldsymbol{Y}}_x$ $\boldsymbol{A}\backslash\boldsymbol{F}_x$. Note that solving N_p two-dimensional quadratic problems is more efficient than solving a single (2 N_p)-dimensional quadratic problem. Also note that this approach does not regularize the parameters at all, by smoothing or sparsening, for instance. In the DISCUSSION we propose several avenues for further development, including the elastic net (Zou and Hastie 2005) and simple parametric models of the neuron. As in the scalar F_t case, we iterate estimating the parameters of this model $\boldsymbol{\theta}$ $\{\tilde{\alpha}, \tilde{\beta}, \sigma, \gamma, \lambda\}$ and the spike train $\boldsymbol{n}$. Because of the free scale term discussed earlier, the absolute magnitude of $\tilde{\alpha}$ is not identifiable. Thus convergence is defined here by the "shape" of the spike train converging, i.e., the norm of the difference between the inferred spike trains from subsequent iterations, both normalized such that max$(\hat{n}_t)$ 1. In practice, this procedure converged after several iterations.

Overlapping spatial filters

It is not always possible to segment the movie into pixels containing only fluorescence from a single neuron. Therefore the above-cited model can be generalized to incorporate multiple neurons within an ROI. Specifically, letting the superscript i index the N_c neurons in this ROI yields:

$$\vec{F}_t = \sum_{i=1}^{N_c} \vec{\alpha}^i C_t^i + \vec{\beta} + \vec{\varepsilon}_t, \quad \vec{\varepsilon}_t \overset{iid}{\sim} (\boldsymbol{0}, \sigma^2 I) \quad (30)$$

$$C_t^i = \gamma^i C_{t-1}^i + n_t^i, \quad n_t^i \overset{iid}{\sim} \text{Poisson}(n_t^i; \lambda_i\Delta) \quad (31)$$

where each neuron is implicitly assumed to be independent and each pixel is conditionally independent and identically distributed with variance σ^2, given the underlying calcium signals. To perform inference in this more general model, let $\boldsymbol{n}_t$ $[n_t^1, \ldots, n_t^{N_c}]$ and $\boldsymbol{C}_t$ $[C_t^1, \ldots, C_t^{N_c}]$ be N_c-dimensional column vectors. Then, let diag $(\gamma^1, \ldots, \gamma^{N_c})$ be an N_c N_c diagonal matrix and let $\boldsymbol{I}$ and $\boldsymbol{0}$ be an identity and zero matrix of the same size, respectively, yielding:

$$\boldsymbol{MC}\begin{bmatrix} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & & \mathbf{I} \end{bmatrix}\begin{bmatrix} \boldsymbol{C}_1 \\ \boldsymbol{C}_2 \\ \vdots \\ \boldsymbol{C}_{T-1} \\ \boldsymbol{C}_T \end{bmatrix} \quad \begin{bmatrix} \boldsymbol{n}_1 \\ \boldsymbol{n}_2 \\ \vdots \\ \boldsymbol{n}_{T-1} \end{bmatrix} \tag{32}$$

and proceed as before. Note that *Eq. 32* is very similar to *Eq. 14*, except that $\boldsymbol{M}$ is no longer bidiagonal, but rather, block bidiagonal (and $\boldsymbol{C}_t$ and $\boldsymbol{n}_t$ are vectors instead of scalars), making the Hessian block-tridiagonal. Importantly, the Thomas algorithm, which is a simplified form of Gaussian elimination, finds the solution to linear equations with block-tridiagonal matrices in linear time, so the efficiency gained from using the tridiagonal structure is maintained for this block-tridiagonal structure (Press et al. 1992). Performing inference in this more general model proceeds similarly as before, letting $\vec{\boldsymbol{\alpha}} \quad [\vec{}_1, \ldots, \vec{}_{N_c}]$:

$$\hat{\boldsymbol{C}}_z \quad \underset{\boldsymbol{MC} \ \odot \mathbf{0}}{\operatorname{argmax}} \quad \frac{1}{2^{\ 2}} \left\| \vec{\boldsymbol{F}} \quad \boldsymbol{C}\vec{\boldsymbol{\alpha}} \quad \mathbf{1}_T \vec{} \right\|_F^2 \quad (\boldsymbol{MC})^{\mathrm{T}}\boldsymbol{\lambda} \quad z \ln_{\odot} (\boldsymbol{MC})^{\mathrm{T}}\mathbf{1}, \tag{33}$$

$$\boldsymbol{g} \quad (\vec{\boldsymbol{F}} \quad \boldsymbol{C}\vec{\boldsymbol{\alpha}} \quad \mathbf{1}_T \vec{})^{\mathrm{T}} \frac{\vec{\boldsymbol{\alpha}}^{\mathrm{T}}}{2} \quad \boldsymbol{M}^{\mathrm{T}} \quad z\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{MC})_{\odot}^{\ \mathbf{1}} \tag{34}$$

$$\boldsymbol{H} \quad \frac{\vec{\boldsymbol{\alpha}}\vec{\boldsymbol{\alpha}}^{\mathrm{T}}}{2}\mathrm{I} \quad z\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{MC})_{\odot}^{\ 2}\boldsymbol{M}. \tag{35}$$

If the parameters are unknown, they must be estimated. Initialize $\vec{}$ as above. Then, define $\boldsymbol{\alpha}_x \quad [{}^{1}_{x}, \ldots, {}^{N_c}_{x}]^{\mathrm{T}}$ and initialize manually by assigning some pixels to each neuron (of course, more sophisticated algorithms could be used, as described in the DISCUSSION). Given this initialization, iterations and stopping criteria proceed as before, with the minor modification of incorporating multiple spatial filters, yielding:

$$\left\{ \hat{}_x, \hat{}_x \right\} \quad \underset{x, \ x}{\operatorname{argmax}} \quad \frac{1}{2}\sum_t (F_{x,t} \quad \sum_{i \ 1}^{N_c} {}^{i}_{x}\hat{C}_t^i \quad {}_x)^2, \tag{36}$$

Now, generalizing the above single spatial filter case, let $\boldsymbol{A} \quad [\hat{\boldsymbol{C}}, \mathbf{1}_T]^{\mathrm{T}}$ be an $(N_c \quad 1) \quad T$ element matrix and $\boldsymbol{Y}_x \quad [\boldsymbol{\alpha}_x, \quad _x]^{\mathrm{T}}$ be an $(N_c \quad 1)$-dimensional column vector. Then, one can again use *Eq. 29* to solve to for $\hat{\boldsymbol{\alpha}}_{\mathbf{x}}$ and $\hat{}_x$ for all x.

Experimental methods

SLICE PREPARATION AND IMAGING. All animal handling and experimentation were done according to the National Institutes of Health and local Institutional Animal Care and Use Committee guidelines. Somatosensory thalamocortical or coronal slices 350–400 m thick were prepared from C57BL/6 mice at age P14 as described (MacLean et al. 2005). Pyramidal neurons from layer V somatosensory cortex were filled with 50 M Oregon Green BAPTA 1 hexapotassium salt (OGB-1; Invitrogen, Carlsbad, CA) through the recording pipette or bulk loaded with an acetoxymethyl ester of Fura-2 (Fura-2 AM; Invitrogen). The pipette solution contained 130 mM K-methylsulfate, 2 mM $MgCl_2$, 0.6 mM EGTA, 10 mM HEPES, 4 mM ATP-Mg, and 0.3 mM GTP-Tris (pH 7.2, 295 mOsm). After cells were fully loaded with dye, imaging was performed in one of two ways. First, when using Fura-2, images were collected using a modified BX50-WI upright microscope (Olympus, Melville, NY) with a confocal spinning disk (Solamere Technology Group, Salt Lake City, UT) and an Orca charge-coupled device (CCD) camera from Hamamatsu Photonics (Shizuoka, Japan), at 33 Hz. Second, when using Oregon Green, images were collected using epifluorescence with the C9100-12 CCD camera from Hamamatsu Photonics, with arc-lamp illumination with excitation and emission band-pass filters at 480–500 and 510–550 nm, respectively (Chroma, Rockingham, VT). Images were saved and analyzed using custom software written in Matlab (The MathWorks, Natick, MA).

ELECTROPHYSIOLOGY. All recordings were made using the Multiclamp 700B amplifier (Molecular Devices, Sunnyvale, CA), digitized with National Instruments 6259 multichannel cards and recorded using custom software written using the LabVIEW platform (National Instruments, Austin, TX). Square pulses of sufficient amplitude to yield the desired number of action potentials were given as current commands to the amplifier using the LabVIEW and National Instruments system.

FLUORESCENCE PREPROCESSING. Traces were extracted using custom Matlab scripts to segment the mean image into ROIs. The Fura-2 fluorescence traces were inverted. Because some slow drift was sometimes present in the traces, each trace was Fourier transformed, and all frequencies 0.5 Hz were set to zero (0.5 Hz was chosen by eye); the resulting fluorescence trace was then normalized to be between zero and one.

RESULTS

Main result

The main result of this study is that the fast filter can find the approximately most likely spike train $\hat{\boldsymbol{n}}$, very efficiently, and that this approach yields more accurate spike train estimates than optimal linear deconvolution. Figure 2 depicts a simulation showing this result. Clearly, the fast filter's inferred "spike train" (*third panel*) more closely resembles the true spike train (*second panel*) than the optimal linear deconvolution's inferred spike train (*bottom panel*; Wiener filter). Note that neither filter results in an integer sequence, but rather, each infers a real number at each time.

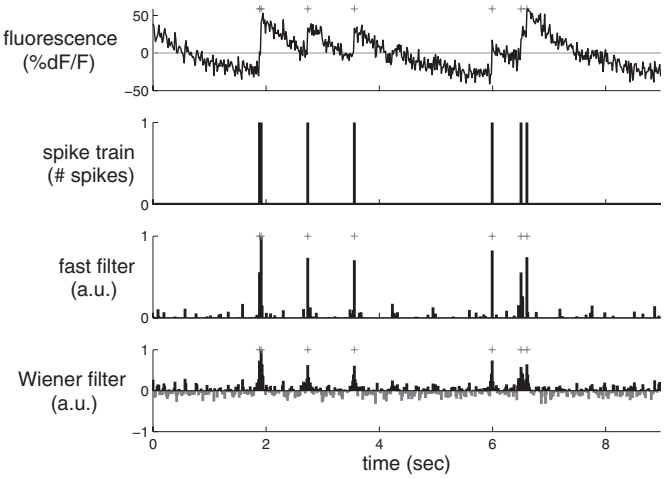

FIG. 2. A simulation showing that the fast filter's inferred spike train is significantly more accurate than the output of the optimal linear deconvolution (Wiener filter). Note that neither filter constrains the inference to be a sequence of integers; rather, the fast filter relaxes the constraint to allow all nonnegative numbers and the Wiener filter allows for all real numbers. The restriction of the fast filter to exclude negative numbers eliminates the ringing effect seen in the Wiener filter output, resulting in a much cleaner inference. Note that the magnitude of the inferred spikes in the fast filter output is proportional to the inferred calcium jump size. *Top panel*: fluorescence trace. *Second panel*: spike train. *Third panel*: fast filter inference. *Bottom panel*: Wiener filter inference. Note that the gray bars in the *bottom panel* indicate negative spikes. Gray symbols indicate true spike times. Simulation details: T 400 time steps, 33.3 ms, 1, 0, 0.2, 1 s, 1 Hz. Parameters and conventions are consistent across figures, unless indicated otherwise.

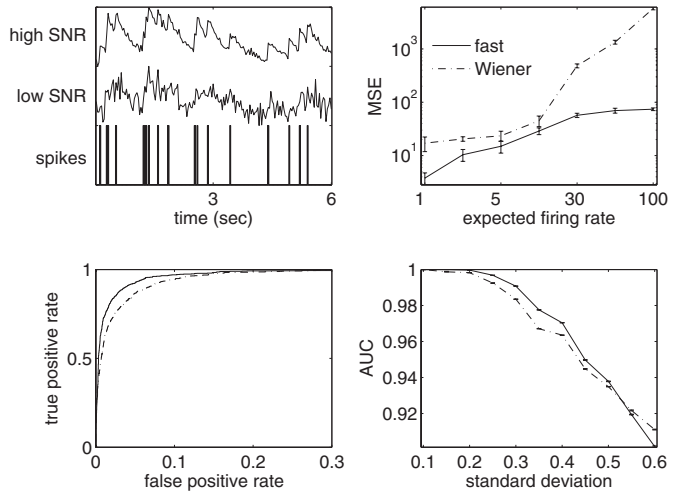

FIG. 3. In simulations, the fast filter quantitatively and significantly achieves higher accuracy than that of the Wiener filter. *Top left*: a spike train (*bottom*) and 2 simulated fluorescence traces, using the same spike train, one with low signal-to-noise ratio (SNR) (*middle*) and one with high SNR (*top*). Simulation parameters: τ 0.5 s, λ 3 Hz, 1/30 s, σ 0.6 (low SNR) and 0.1 (high SNR). Simulation parameters in other panels are the same, except where explicitly noted. *Top right*: mean-squared-error (MSE) for the fast (solid line) and Wiener (dashed-dotted line) filter, for varying the expected firing rate λ. Note that both axes are on a log-scale. Further note that the fast filter has a better (lower) MSE for all expected firing rates. Error bars show SD over 10 repeats. Simulation parameters: σ 0.2, T 1,000 time steps. *Bottom left*: receiver-operator-characteristic (ROC) curve comparing the fast (solid line) and Wiener (dashed-dotted line) filter. Note that for any given threshold, the Wiener filter has a better (higher) ratio of true positive rate to false positive rate. Simulation parameters as in *top right panel*, except σ 0.35 and T 10,000 time steps. *Bottom right*: area under the curve (AUC) for fast (solid line) and Wiener (dashed-dotted line) filter as a function of SD (σ). Note that the fast filter has a better (higher) AUC for all σ values until noise gets very high. The 2 simulated fluorescence traces in the *top left panel* show the bounds for SD here. Error bars show SD over 10 repeats.

The Wiener filter implicitly approximates the Poisson spike rate with a Gaussian spike rate (see APPENDIX B for details). A Poisson spike rate indicates that in each frame, the number of possible spikes is an integer, e.g., 0, 1, 2,.... The Gaussian approximation, however, allows any real number of spikes in each frame, including both partial spikes (e.g., 1.4) and *negative* spikes (e.g., 0.8). Although a Gaussian well approximates a Poisson distribution when rates are about 10 spikes per frame, this example is very far from that regime, so the Gaussian approximation performs relatively poorly. Further, the Wiener filter exhibits a "ringing" effect. Whenever fluorescence drops rapidly, the most likely underlying spiking signal is a proportional drop. Because the Wiener filter does not impose a nonnegative constraint on the underlying spiking signal, it infers such a drop, even when it causes n_t to go below zero. After such a drop has been inferred, since no corresponding drop occurred in the true underlying signal here, a complementary jump is often then inferred, to realign the inferred signal with the observations. This oscillatory behavior results in poor inference quality. The nonnegative constraint imposed by the fast filter prevents this because the underlying signal never drops below zero, so the complementary jump never occurs either.

The inferred "spikes," however, are still not binary events when using the fast filter. This is a by-product of approximating the Poisson distribution on spikes with an exponential (cf. *Eq. 11*a) because the exponential is a continuous distribution, versus the Poisson, which is discrete. The height of each spike is therefore proportional to the inferred calcium jump size and can be thought of as a proxy for the confidence with which the algorithm believes a spike occurred. Importantly, by using the Gaussian elimination and interior-point methods, as described in METHODS, the computational complexity of the fast filter is the same as an efficient implementation of the Wiener filter. Note that whereas the Gaussian approximation imposes a shrinkage prior on the inferred spike trains (Wu et al. 2006), the exponential approximation imposes a sparse prior on the inferred spike trains (Seeger 2008).

Figure 3 quantifies the relative performance of the fast and Wiener filters. The *top left panel* shows a typical simulated spike train (*bottom*), a corresponding relatively low SNR fluorescence trace (*middle*), and a relatively high SNR fluorescence trace (*top*), as examples. The *top right panel* compares the mean-squared-error (MSE) of the inferred spike trains using the fast (solid) and Wiener (dashed) filters, as a function of expected firing rate. Clearly, the fast filter has a better (lower) MSE for all rates. The *bottom left panel* shows a receiver-operator-characteristic (ROC) curve (Green and Swets 1966) for another simulation. Again, the fast filter dominates the Wiener filter, having a higher true positive rate for every false negative rate. Finally, the *bottom right panel* shows that the area under the curve (AUC) of the fast filter is better (higher) than that of the Wiener filter until the noise is very large. Collectively, these analyses suggest that for a wide range of firing rates and signal quality, the fast filter outperforms the Wiener filter.

Although in Fig. 2 the model parameters were provided, in the general case, the parameters are unknown and must therefore be estimated from the observations (as described in *Learning the parameters* in METHODS). Importantly, this algorithm does not require labeled training data, i.e., there is no need for joint imaging and electrophysiological experiments to estimate the parameters governing the relationship between the two. Figure 4 shows another simulated example; in this example, however, the parameters are estimated from the observed

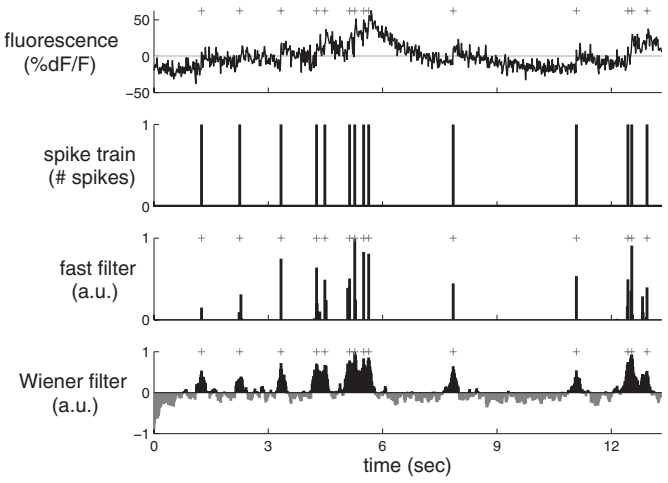

FIG. 4. A simulation showing that the fast filter achieves significantly more accurate inference than that of the Wiener filter, even when the parameters are unknown. For both filters, the appropriate parameters were estimated using only the data shown above, unlike Fig. 2, in which the true parameters were provided to the filters. Simulation details different from those in Fig. 2: T 1,000 time steps, 16.7 ms, σ 0.4.

fluorescence trace. Again, it is clear that the fast filter far outperforms the Wiener filter.

Given the preceding two results, the fast filter was applied to biological data. More specifically, by jointly recording electrophysiologically and imaging, the true spike times are known and the accuracy of the two filters can be compared. Figure 5 shows a result typical of the 12 joint electrophysiological and imaging experiments conducted (see METHODS for details). As in the simulated data, the fast filter output is much "cleaner" than the Wiener filter: spikes are more well defined, and not spread out, due to the sparse prior imposed by the exponential approximation. Note that this trace is typical of epifluorescence techniques, which makes resolving individual spikes quite difficult, as evidenced by a few false positives in the fast filter. Regardless, the fast filter output is still more accurate than the Wiener filter, both as determined qualitatively by eye and as quantified (described in the following text). Furthermore, although it is difficult to see in this figure, the first four events are actually pairs of spikes, reflected by the width and height of the corresponding inferred spikes when using the fast filter. This suggests that although the scale of $\boldsymbol{n}$ is arbitrary, the fast filter can correctly ascertain the number of spikes within spike events.

Figure 6 further evaluates this claim. While recording and imaging, the cell was forced to spike once, twice, or thrice for each spiking event. The fast filter infers the correct number of spikes in each event. On the contrary, there is no obvious way to count the number of spikes within each event when using the Wiener filter. We confirm this impression by computing the correlation coefficient, r^2, between the sum of each filter's output and the true number of spikes, for all 12 joint electrophysiological and imaging traces. Indeed, whereas the fast filter's r^2 was 0.47, the Wiener filter's r^2 was 0.01 (after thresholding all negative spikes), confirming that the Wiener filter output cannot reliably convey the number of spikes in a fluorescence trace, whereas the fast filter can. Furthermore, varying the magnitude of the threshold for the Wiener filter to discard more "low-amplitude noise" could increase the magnitude of r^2 ($\leq$0.24), still significantly lower than the fast filter's r^2 value. On the other hand, no amount of thresholding

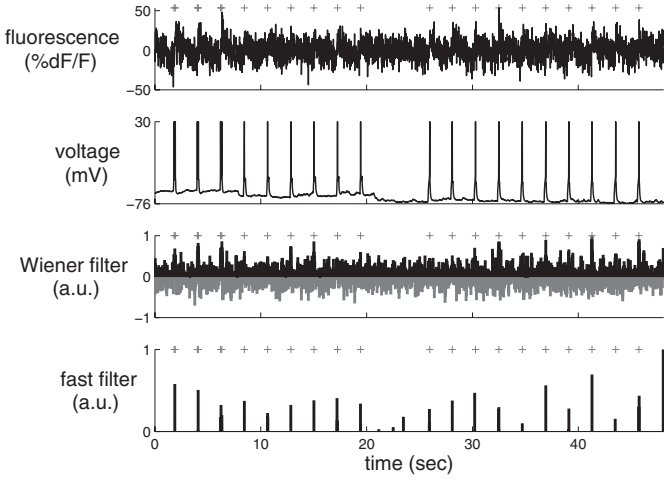

FIG. 5. In vitro data showing that the fast filter significantly outperforms the Wiener filter, using OGB-1. Note that all the parameters for both filters were estimated only from the fluorescence data in the *top panel* (i.e., not considering the voltage data at all). symbols denote true spike times extracted from the patch data, not inferred spike times from $\boldsymbol{F}$.

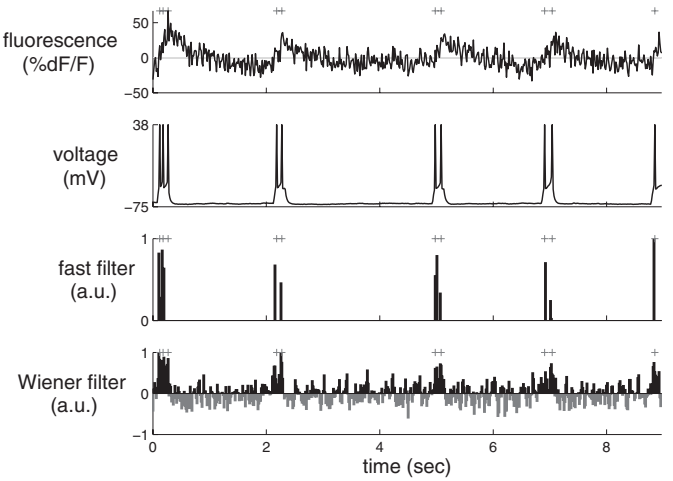

FIG. 6. In vitro data with multispike events, showing that the fast filter can often resolve the correct number of spikes within each spiking event, while imaging using OGB-1, given sufficiently high SNR. It is difficult, if not impossible, to count the number of spikes given the Wiener filter output. Recording and fitting parameters as in Fig. 5. Note that the parameters were estimated using a 60-s-long recording, of which only a fraction is shown here, to more clearly depict the number of spikes per event.

the fast filter yielded an improved r^2, indicating that thresholding the output of the fast filter is unlikely to improve spike inference quality.

On-line analysis of spike trains using the fast filter

A central aim for this work was the development of an algorithm that infers spikes fast enough to use on-line while imaging a large population of neurons (e.g., 100). Figure 7 shows a segment of the results of running the fast filter on 136 neurons, recorded simultaneously, as described earlier in *Experimental methods*. Note that the filtered fluorescence signals show fluctuations in spiking much more clearly than the unfiltered fluorescence trace. These spike trains were inferred in less than imaging time, meaning that one could infer spike trains for the past experiment while conducting the subsequent experiment. More specifically, a movie with 5,000 frames of 100 neurons can be analyzed in about 10 s on a standard desktop computer. Thus if that movie was recorded at 50 Hz, whereas collecting the data would require 100 s, inferring spikes would require only 10 s, a 10-fold improvement over real time.

Extensions

Earlier in METHODS, *Data-driven generative model* describes a simple principled first-order model relating the spike trains to the fluorescence trace. A number of the simplifying assumptions can be straightforwardly relaxed, as described next.

Replacing Gaussian observations with poisson. In the preceding text, observations were assumed to have a Gaussian distribution. The statistics of photon emission and counting, however, suggest that a Poisson distribution would be more natural in some conditions, especially for two-photon data (Sjulson and Miesenböck 2007), yielding:

$$F_t \overset{iid}{\sim} \text{Poisson} \quad (\alpha C_t + \beta), \tag{37}$$

where αC_t $\beta \geq 0$. One additional advantage to this model over the Gaussian model is that the variance parameter σ^2 no

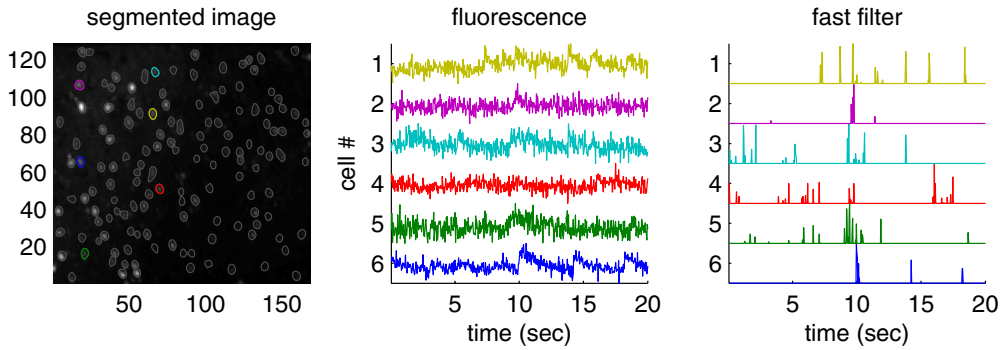

FIG. 7. The fast filter infers spike trains from a large population of neurons imaged simultaneously in vitro, faster than real time. Specifically, inferring the spike trains from this 400-s-long movie including 136 neurons required only about 40 s on a standard laptop computer. The inferred spike trains much more clearly convey neural activity than the raw fluorescence traces. Although no intracellular "ground truth" is available from these population data, the noise seems to be reduced, consistent with the other examples with ground truth. *Left*: mean image field, automatically segmented into regions of interest (ROIs), each containing a single neuron using custom software. *Middle*: example fluorescence traces. *Right*: fast filter output corresponding to each associated trace. Note that neuron identity is indicated by color across the 3 panels. Data were collected using a confocal microscope and Fura-2, as described in METHODS.

longer exists, which might make learning the parameters simpler. Importantly, the log-posterior is still concave in $\boldsymbol{C}$, as the prior remains unchanged, and the new log-likelihood term is a sum of terms concave in $\boldsymbol{C}$:

$$\ln P[\boldsymbol{F}\ \boldsymbol{C}] \quad \sum_{t\ 1}^{T} \ln P[F_t\ C_t] \quad \sum_{t\ 1}^{T} F_t \ln(\ C_t\ \ ) \quad (\ C_t\ \ ) \quad \ln(F_t!)\ . \quad (38)$$

The gradient and Hessian of the log-posterior can therefore be computed analytically by substituting the above likelihood terms for those implied by *Eq. 1*. In practice, however, modifying the filter for this model extension did not seem to significantly improve inference results in any simulations or data available at this time (not shown).

Allowing for a time-varying prior. In *Eq. 4*, the rate of spiking is a constant. Often, additional knowledge about the experiment, including external stimuli or other neurons spiking, can provide strong time-varying prior information (Vogelstein et al. 2009). A simple model modification can incorporate that feature:

$$n_t \overset{iid}{\sim} \text{Poisson}(\ _t\ ), \quad (39)$$

where $_t$ is now a function of time. Approximating this time-varying Poisson with a time-varying exponential with the same time-varying mean (similar to *Eq. 11*a) and letting $\boldsymbol{\lambda}$ $[\ _1, \ldots, \ _T]^{\mathrm{T}}$, yields an objective function very similar to *Eq. 15*, so log-concavity is maintained and the same techniques may be applied. However, as before, this model extension did not yield any significantly improved filtering results (not shown).

Saturating fluorescence. Although all the abovementioned models assumed a *linear* relationship between F_t and C_t, the relationship between fluorescence and calcium is often better approximated by the nonlinear Hill equation (Pologruto et al. 2004). Modifying *Eq. 1* to reflect this change yields:

$$F_t \quad \frac{C_t}{C_t \quad k_d} \quad _t \quad _t \overset{iid}{\sim} (0, \ ^2). \quad (40)$$

Importantly, log-concavity of the posterior is no longer guaranteed in this nonlinear model, meaning that converging to the global maximum is no longer guaranteed. Assuming a good initialization can be found, however, and *Eq. 40* is more accurate than *Eq. 1*, then ascending the gradient for this model is likely to yield improved inference results. In practice, initializing with the inference from the fast filter assuming a linear model (e.g., *Eq. 30*) often resulted in nearly equally accurate inference, but inference assuming the above nonlinearity was far less robust than the inference assuming the linear model (not shown).

Using the fast filter to initialize the SMC filter. A sequential Monte Carlo (SMC) method to infer spike trains can incorporate this saturating nonlinearity, as well as other model extensions discussed earlier (Vogelstein et al. 2009). However, this SMC filter is not nearly as computationally efficient as the fast filter proposed here. Like the fast filter, the SMC filter estimates the model parameters in a completely unsupervised fashion, i.e., from the fluorescence observations, using an expectation-maximization algorithm (which requires iterating between computing the expected value of the hidden variables—$\boldsymbol{C}$ and $\boldsymbol{n}$—and updating the parameters). In Vogelstein and colleagues (2009), parameters for the SMC filter were initialized based on other data. Although effective, this initialization was often far from the final estimates and thus required a relatively large number of iterations (e.g., 20–25) before converging. Thus it seemed that the fast filter could be used to obtain an improvement to the initial parameter estimates, given an appropriate rescaling to account for the nonlinearity, thereby reducing the required number of iterations to convergence. Indeed, Fig. 8 shows how the SMC filter outperforms the fast filter on biological data and required only three to five iterations to converge on these data, given the initialization from the fast filter (which was typical). Note that the first few events of the spike train are individual spikes, resulting in relatively small fluorescence fluctuations, whereas the next events are actually spike doublets or triplets, causing a much larger fluorescence fluctuation. Only the SMC filter correctly infers the presence of isolated spikes in this trace, a frequently occurring result when the SNR is poor. Thus these two inference algorithms are complementary: the fast filter can be used for rapid, on-line inference, and for initializing the SMC filter, which can then be used to further refine the spike train

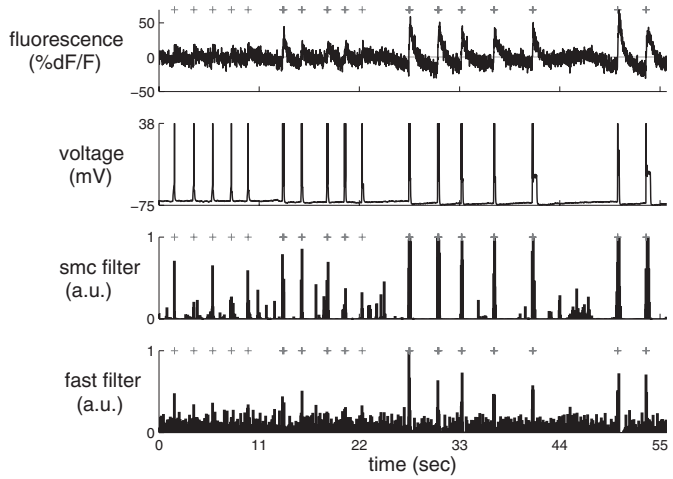

FIG. 8. In vitro data with SNR of only about 3 (estimated by dividing the fluorescent jump size by the SD of the baseline fluorescence) for single action potentials depicting the fast filter, effectively initializing the parameters for the sequential Monte Carlo (SMC) filter, significantly reducing the number of expectation-maximization iterations to convergence, using OGB-1. Note that whereas the fast filter clearly infers the spiking events in the end of the trace, those in the beginning of the trace are less clear. On the other hand, the SMC filter more clearly separates nonspiking activity from true spikes. Also note that the ordinate on the third panel corresponds to the inferred probability of a spike having occurred in each frame.

estimate. Importantly, although the SMC filter often outperforms the fast filter, the fast filter is more robust, meaning that it more often works "out of the box." This follows because the SMC filter operates on a highly nonlinear model that is not log-concave. Thus although the expectation-maximization algorithm used often converges to reasonable local maxima, it is not guaranteed to converge to global maxima and its performance in general will depend on the quality of the initial parameter estimates.

Spatial filter

In the preceding text, the filters operated on one-dimensional fluorescence traces. The raw data are in fact a time series of images that are first segmented into regions of interest (ROIs) and then (usually) spatially averaged to obtain a one-dimensional time series $\boldsymbol{F}$. In theory, one could improve the effective SNR of the fluorescence trace by scaling each pixel according to its SNR. In particular, pixels not containing any information about calcium fluctuations can be ignored and pixels that are partially anticorrelated with one another could have weights with opposing signs.

Figure 9 demonstrates the potential utility of this approach. The *top row* shows different depictions of an ROI containing a single neuron. On the *far left panel* is the true spatial filter for this neuron. This particular spatial filter was chosen based on experience analyzing both in vitro and in vivo movies; often, it seems that the pixels immediately around the soma are anticorrelated with those in the soma (MacLean et al. 2005; Watson et al. 2008). This effect is possibly due to the influx of calcium from the extracellular space immediately around the soma. The standard approach, given such a noisy movie, would be to first segment the movie to find an ROI corresponding to the soma of this cell and then spatially average all the pixels found to be within this ROI. The *second panel* shows this standard "boxcar spatial filter." The *third panel* shows the mean frame. The *fourth panel* shows the learned filter, using *Eq. 29* to estimate the spatial filter and background. Clearly, the learned filter is very similar to the mean filter and the true filter.

The *middle panels* of Figure 9 show the fluorescence traces obtained by background subtracting and then projecting each frame onto the corresponding spatial filter (black line) and true spike train (gray symbols). The *bottom panels* show the inferred spike trains (black bars) using these various spatial filters and, again, the true spike train (gray symbols). Although the

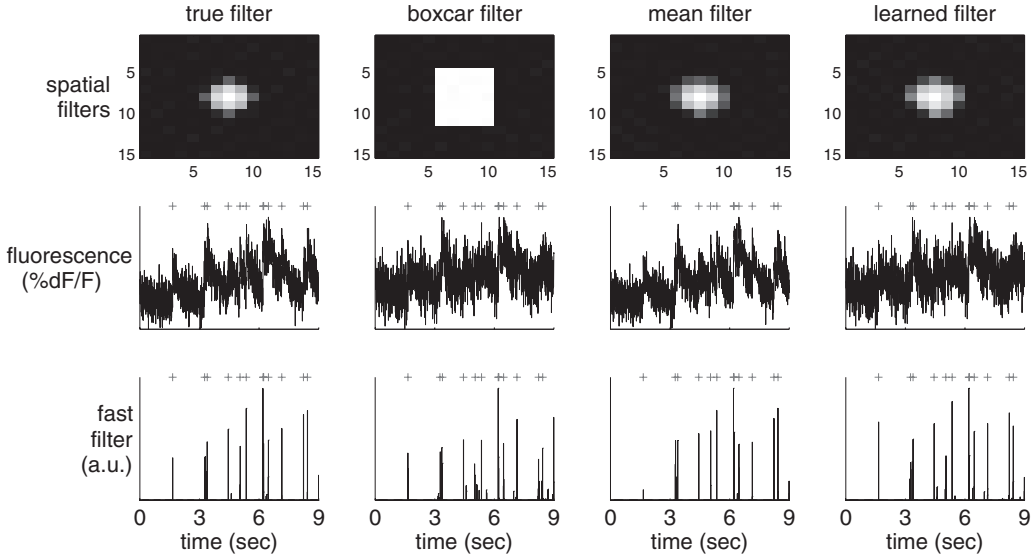

FIG. 9. A simulation demonstrating that using a better spatial filter can significantly enhance the effective SNR. The true spatial filter was a difference of Gaussians: a positively weighted Gaussian of small width and a negatively weighted Gaussian with larger width (both with the same center). Each column shows the spatial filter (*top*), one-dimensional fluorescence projection using that spatial filter (*middle*), and inferred spike train (*bottom*). From *left* to *right*, columns use the true, boxcar, mean, and learned spatial filter obtained using *Eq. 29*. Note that the learned filter's inferred spike train has fewer false positives and negatives than the boxcar and mean filters. Simulation parameters: $\vec{\alpha}$ $(\mathbf{0}, 2\boldsymbol{I})$ 0.5 $(\mathbf{0}, 2.5\boldsymbol{I})$, where $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates a 2-dimensional Gaussian with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, $\vec{\beta}$ $\mathbf{0}$, σ 0.2, τ 0.85 s, λ 5 Hz, 5 ms, T 1,200 time steps.

performance is very similar for all of them, the boxcar filter's inferred spike train is not as clean.

Overlapping spatial filters

The preceding text shows that if the ROI contains only a single neuron, the effective SNR can be enhanced by spatially filtering. However, this analysis assumes that only a single neuron is in the ROI. Often, ROIs are overlapping, or nearly overlapping, making the segmentation problem more difficult. Therefore it is desirable to have an ability to crudely segment, yielding only a few neurons in each ROI, and then spatially filter within each ROI to pick out the spike trains of each neuron. This may be achieved in a principled manner by generalizing the model as described in *Overlapping spatial filters* in METHODS. The true spatial filters of the neurons in the ROI are often unknown and thus must be estimated from the data. This problem may be considered a special case of blind source separation (Bell and Sejnowski 1995; Mukamel et al. 2009). Figure 10 shows that given reasonable assumptions of spiking correlations and SNR, multiple signals can be separated. Note that separation occurs even though the signal is significantly overlapping (*top panels*). To estimate the spatial filters, they are initialized using the boxcar filters (*middle panels*). After a few iterations, the spatial filters converge to very close approximation to the true spatial filters [compare true (*left*) and learned (*right*) spatial filters for the two neurons]. Note that both the true and learned spatial filters yield much improved spike inference relative to the boxcar filter. This suggests that even when spatial filters of multiple neurons are significantly overlapping, each spike train is potentially independently recoverable.

DISCUSSION

Summary

This work describes an algorithm that finds the approximate maximum a posteriori (MAP) spike train, given a calcium fluorescence movie. The approximation is required because finding the actual MAP estimate is not currently computationally tractable. Replacing the assumed Poisson distribution on spikes with an exponential distribution yields a log-concave optimization problem, which can be solved using standard gradient ascent techniques (such as Newton–Raphson). This exponential distribution has an advantage over a Gaussian distribution by restricting spikes to be positive, which improves inference quality (cf. Fig. 2), is a better approximation to a Poisson distribution with low rate, and imposes a sparse constraint on spiking. Furthermore, all the parameters can be estimated from only the fluorescence observations, obviating the need for joint electrophysiology and imaging (cf. Fig. 4). This approach is robust, in that it works "out of the box" on all the in vitro data analyzed (cf. Figs. 5 and 6). By using the special banded structure of the Hessian matrix of the log-posterior, this approximate MAP spike train can be inferred fast enough on standard computers to use it for on-line analyses of over 100 neurons simultaneously (cf. Fig. 7).

Finally, the fast filter is based on a biophysical model capturing key features of the data and may therefore be straightforwardly generalized in several ways to improve ac-

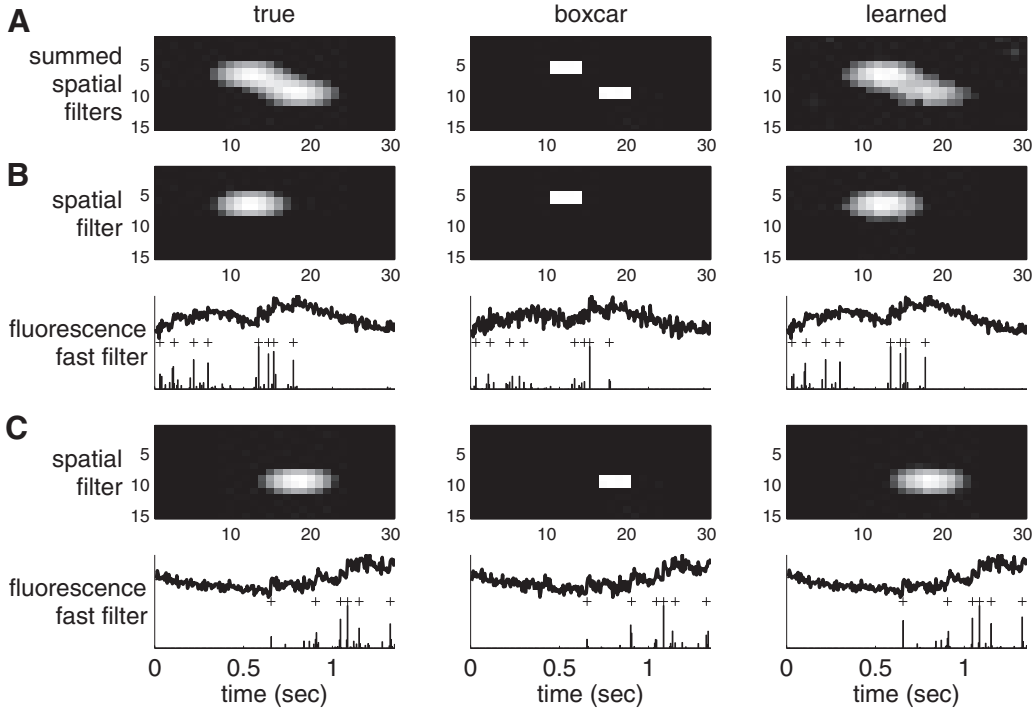

FIG. 10. Simulation showing that when 2 neurons' spatial filters are largely overlapping, learning the optimal spatial filters using *Eq. 36* can yield improved inference of the standard boxcar type filters. The 3 columns show the effect of the true (*left*), boxcar (*center*), and learned (*right*) spatial filters. *A*: the sum of the 2 spatial filters for each approach, clearly depicting overlap. *B*: the spatial filters (*top row*), one-dimensional fluorescence projection, and inferred spike train (*bottom row*) for one of the neurons. *C*: same as *B* for the other neuron. Note that the inferred spike trains when using the learned filter are close to optimal, unlike the boxcar filter. Simulation parameters: $\vec{\alpha}^1$ ([1, 0], $2\boldsymbol{I}$) 0.5 ([1, 0], $2.5\boldsymbol{I}$), $\vec{\alpha}^2$ ([1, 0], $2\boldsymbol{I}$) 0.5 ([1, 0], $2.5\boldsymbol{I}$), $\vec{\beta}$ $\mathbf{0}$, σ 0.02, τ 0.5 s, λ 5 Hz, 5 ms, T 1,200 time steps (not all time steps are shown).

curacy. Unfortunately, some of these generalizations do not improve inference accuracy, perhaps because of the exponential approximation. Instead, the fast filter output can be used to initialize the more general SMC filter (Vogelstein et al. 2009), to further improve inference quality (cf. Fig. 8). Another model generalization allows incorporation of spatial filtering of the raw movie into this approach (cf. Fig. 9). Even when multiple neurons are overlapping, spatial filters may be estimated to obtain improved spike inference results (cf. Fig. 10).

Alternate algorithms

This work describes but one specific approach to solving a problem that does not admit an exact solution that is computationally feasible. Several other approaches warrant consideration, including *1*) a Bayesian approach, *2*) a greedy approach, and *3*) different analytical approximations.

First, a Bayesian approach could use Markov Chain Monte Carlo methods to recursively sample spikes to estimate the full joint posterior distribution of the entire spike train, conditioned on the fluorescence data (Andrieu et al. 2001; Joucla et al. 2010; Mishchenko et al. 2010). Although enjoying several desirable statistical properties that are lacking in the current approach (such as consistency), the computational complexity of such an approach renders it inappropriate for the aims of this work.

Second, a common relatively expedient approximation to Bayesian sampling is a so-called greedy approach. Greedy algorithms are iterative, with each iteration adding another spike to the putative spike train. Each spike that is added is the most likely spike (thus the greedy term) or the one that most increases the likelihood of the fluorescence trace. Template matching, projection pursuit regression (Friedman and Stuetzle 1981), and matching pursuit (Mallat and Zhang 1993) are examples of such a greedy approach (the algorithm proposed by Grewe et al. (2010) could also be considered a special case of such a greedy approach).

Third, approximations other than the exponential distribution are possible. For instance, the Gaussian approximation is more appropriate for high firing rates, although in simulations, this more accurate approximation did not improve the Wiener filter output relative to the fast filter output (cf. Fig. 3). Perhaps the best approximation would use the closest log-concave relaxation to the Poisson model (Koenker and Mizera 2010). More formally, let $P(i)$ represent the Poisson mass at i and let $\ln Q$ be some concave density. Then, one could find the log-density Q such that Q maximizes $\sum_i P(i)Q(i) - \lambda \int \exp\{Q(x)\}dx$ over the space of all concave Q. The first term corresponds to the log-likelihood, equivalent to the Kullback–Leibler divergence (Cover and Thomas 1991), and the second is a Lagrange multiplier to ensure that the density $\exp\{Q(x)\}$ integrates to unity. This is a convex problem because the space of all concave Q is convex and the objective function is concave in Q. In addition, it is easy to show that the optimal Q has to be piecewise linear; this means that one need not search over all possible densities, but rather, simply vary $Q(i)$ at the integers. Note that $\int \exp\{Q(x)\}dx$ can be computed explicitly for any piecewise linear Q. This optimization problem can be solved using simple interior point methods and, in fact, the Hessian of the inner loop of the interior point method will be banded (because enforcing concavity of Q is a local constraint). This approximation could potentially be more accurate than our exponential approximation. Further, this approximation encourages integer solutions for n_t and is therefore of interest for future work.

The abovementioned three approaches may be thought of as complementary because each has unique advantages relative to the others. Both the greedy methods and the analytic approximations could potentially be used to initialize a Bayesian approach, possibly limiting the burn-in period, which can be computationally prohibitive in certain contexts. A greedy approach has the advantage of providing actual spike trains (i.e., binary sequences), unlike the analytic approximations. However, the actual spike trains could be quite far from the MAP spike train because greedy approaches, in general, have no guarantee of consistency. The analytic approximations, on the other hand, are guaranteed to converge to solutions close to the MAP spike train, where closeness is determined by the accuracy of the above approximation. Thus developing these distinct approaches and combining them is a potential avenue for further research.

Spatial filtering

Spatial filtering could be improved in a number of ways. For instance, pairing this approach with a crude but automatic segmentation tool to obtain ROIs would create a completely automatic algorithm that converts raw movies of populations of neurons into populations of spike trains. Furthermore, this filter could be coupled with more sophisticated algorithms to initialize the spatial filters when they are overlapping [for instance, principal component analysis (Horn and Johnson 1990) or independent component analysis (Mukamel et al. 2009)]. One could also use a more sophisticated model to estimate the spatial filters. One option would be to assume a simple parametric form of the spatial filter for each neuron (e.g., a basis set) and then merely estimate the parameters of that model. Alternately, one could regularize the spatial filters, using an elastic net type approach (Grosenick et al. 2009; Zou and Hastie 2005), to enforce both sparseness and smoothness.

Model generalizations

In this work, we made two simplifying assumptions that can easily be relaxed: *1*) instantaneous rise time of the fluorescence transient after a spike and *2*) constant background. In practice, often either or both of these assumptions are inaccurate. Specifically, genetic sensors tend to have a much slower rise time than that of organic dyes (Reiff et al. 2005). Further, the background often exhibits slow baseline drift due to movement, temperature fluctuations, laser power, and so forth, not to mention bleaching, which is ubiquitous for long imaging experiments. Both slow rise and baseline drift can be incorporated into our forward model using a straightforward generalization.

Consider the following illustrative example: the fluorescence rise time in a particular data set is quite slow, much slower than that of a single image frame. Thus fluorescence might be well modeled as the difference of two different calcium extrusion mechanisms, with different time constants. To model this scenario, one might proceed as follows: posit the existence of a two-dimensional time-varying signal, each like the calcium

APPENDIX A: PSEUDOCODE

Algorithm 1 Pseudocode for inferring the approximately most likely spike train, given fluorescence data. Note that the algorithm is robust to small variations $_z$, $_n$. The equations listed below refer to the most general equations in the text (simpler equations could be substituted when appropriate). Curly brackets, {·}, indicate comments.

```
initialize parameters, θ (see Initializing the parameters in METHODS)
while convergence criteria not met do
  for z  1,0.1,0.01, . . . ,  z do {interior point method to find Ĉ}
    Initialize n_t  _n for all t  1, . . . , T, C_1  0 and C_t  γC_t 1  n_t for all t  2, . . . , T
    let C_z be the initialized calcium, and P̂_z, be the posterior given this initialization
    while P̂_z'  P̂_z do {Newton–Raphson with backtracking line searches}
      compute g using Eq. 34
      compute H using Eq. 35
      compute d using H\g {block-tridiagonal Gaussian elimination}
      let C_z'  C_z  sd, where s is between 0 and 1, and P̂_z'  P̂_z {backtracking line search}
    end while
  end for
  check convergence criteria
  update α̃ and → using Eq. 36 {only if spatial filtering}
  let  be the root-mean square of the residual
  let  T/( ,n̂_t)
end while
```

signal assumed in the simpler models described earlier. Therefore each signal has a time constant and each signal is dependent on spiking. Finally, the fluorescence could be a weighted difference of the two signals. To formalize this model and to generalize it, let *1*) $\boldsymbol{X}$ $(X_1, \ldots, X_d)$ be a d-dimensional time-varying signal; *2*) $\boldsymbol{\Gamma}$ be a d d dynamics matrix, where diagonal elements correspond to time constants of individual variables, and off-diagonal elements correspond to dependencies across variables; *3*) $\boldsymbol{A}$ be a d-dimensional binary column vector encoding whether each variable depends on spiking; and *4*) $\boldsymbol{\alpha}$ be a d-dimensional column vector of weights, determining the relative impact of each dimension on the total fluorescence signal. Given these conventions, we have the following generalized model:

$$F_t \quad {}^{\mathrm{T}}\boldsymbol{X}_t \quad , \quad {}^{\mathrm{iid}} \quad (0, \ {}^2) \tag{41}$$

$$\boldsymbol{X}_t \quad \boldsymbol{X}_{t\ 1} \quad \boldsymbol{A}n_t, \quad n_t \ {}^{\mathrm{iid}}\mathrm{Poisson}(\ ) \tag{42}$$

Note that this model simplifies to the model proposed earlier when d 1. Because $\boldsymbol{X}$ is still Markov, all the theory developed above still applies directly for this model. There are, however, additional complexities with regard to identifiability. Specifically, the parameters and $\boldsymbol{A}$ are closely related. Thus we enforce that $\boldsymbol{A}$ is a known binary vector, simply encoding whether a particular element responds to spiking. The matrix $\boldsymbol{\Gamma}$ will not be uniquely identifiable, for the same reason that was not identifiable, as described in *Learning the parameters* in METHODS. Thus we would assume $\boldsymbol{\Gamma}$ was known, a priori. Note that other approaches to dealing with baseline drift are also possible, such as letting be a time-varying state: $_t \ _2 \ _{t\ 1} \ _t$, where $_t$ is a normal random variable with variance ^{2} that sets the effective drift rate. Both these models are the subjects of further development.

Concluding thoughts

In summary, the model and algorithm proposed in this work potentially provide a useful tool to aid in the analysis of calcium-dependent fluorescence imaging and establish the groundwork for significant further development.

APPENDIX B: WIENER FILTER

The Poisson distribution in *Eq. 4* can be replaced with a Gaussian instead of an exponential distribution, i.e., $n_t \ {}^{\mathrm{iid}}$, that, when plugged into *Eq. 7*, yields:

$$\hat{\boldsymbol{n}} \quad \underset{n_t \quad t\ 1}{\mathrm{argmax}} \ {}^{T} \quad \frac{1}{2\ ^2}(F_t \quad C_t \quad)^2 \quad \frac{1}{2}(n_t \quad)^2 \ . \tag{B1}$$

Note that since fluorescence integrates over , it makes sense that the mean scales with . Further, since the Gaussian here is approximating a Poisson with high rate (Sjulson and Miesenböck 2007), the variance should scale with the mean. Using the same tridiagonal trick as before, *Eq. 11*b can be solved using Newton–Raphson once (because this expression is quadratic in $\boldsymbol{n}$). Writing the above in matrix notation, substituting C_t $\gamma C_{t\ 1}$ for n_t, and letting 1 yields:

$$\hat{\boldsymbol{C}} \quad \underset{\boldsymbol{C}}{\mathrm{argmax}} \quad \frac{1}{2\ ^2} \ \boldsymbol{F} \quad \boldsymbol{C} \quad \mathbf{1}_T \ ^2 \quad \frac{1}{2} \quad \boldsymbol{MC} \quad 1 \ ^2, \tag{B2}$$

which is quadratic in $\boldsymbol{C}$. The gradient and Hessian are given by:

$$\boldsymbol{g} \quad \frac{1}{\ _2}(\boldsymbol{C} \quad \boldsymbol{F} \quad \mathbf{1}_T) \quad \frac{1}{\ }[(\boldsymbol{M}\hat{\boldsymbol{C}})^{\mathrm{T}}\boldsymbol{M} \quad \boldsymbol{M}^{\mathrm{T}}\mathbf{1}] \tag{B3}$$

$$\boldsymbol{H} \quad \frac{1}{\ _2}\boldsymbol{I} \quad \frac{1}{\ }\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M}. \tag{B4}$$

Note that this solution is the optimal linear solution, under the assumption that spikes follow a Gaussian distribution, and is often referred to as the Wiener filter, regression with a smoothing prior, or ridge regression (Boyd and Vandenberghe 2004). Estimating the parameters for this model follows a pattern similar to that described in *Learning the parameters* in METHODS.

ACKNOWLEDGMENTS

We thank V. Bonin for helpful discussions.

GRANTS

This work was supported by National Institute on Deafness and Other Communication Disorders Grant DC-00109 to J. T. Vogelstein; National Science Foundation (NSF) Faculty Early Career Development award, an Alfred P. Sloan Research Fellowship, and a McKnight Scholar Award to L.

Paninski; National Eye Institute Grant EY-11787 and the Kavli Institute for Brain Studies award to R. Yuste and the Yuste laboratory; and an NSF Collaborative Research in Computational Neuroscience award IIS-0904353, awarded jointly to L. Paninski and R. Yuste.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES

Andrieu C, Barat É, Doucet A. Bayesian deconvolution of noisy filtered point processes. *IEEE Trans Signal Process* 49: 134–146, 2001.

Bell AJ, Sejnowski TJ. An information-maximisation approach to blind separation and blind deconvolution. *Neural Comput* 7: 1129–1159, 1995.

Boyd S, Vandenberghe L. *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.

Cover TM, Thomas JA. *Elements of Information Theory*: New York: Wiley–Interscience, 1991.

Cunningham JP, Shenoy KV, Sahani M. Fast Gaussian process methods for point process intensity estimation. In: *Proceedings of the 25th International Conference on Machine Learning* (ICML 2008). New York: IEEE Press, 2008, p. 192–199.

Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B Methodol* 39: 1–38, 1977.

Friedman JH, Stuetzle W. Projection pursuit regression. *J Am Stat Assoc* 76: 817–823, 1981.

Garaschuk O, Griesbeck O, Konnerth A. Troponin c-based biosensors: a new family of genetically encoded indicators for in vivo calcium imaging in the nervous system. *Cell Calcium* 42: 351–361, 2007.

Göbel W, Helmchen F. In vivo calcium imaging of neural network function. *Physiology* (Bethesda) 22: 358–365, 2007.

Green DM, Swets JA. *Signal Detection Theory and Psychophysics*. New York: Wiley, 1966.

Greenberg DS, Houweling AR, Kerr JND. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat Neurosci* 11: 749–751, 2008.

Grewe BF, Langer D, Kasper H, Kampa BM, Helmchen F. High-speed in vivo calcium imaging reveals neuronal network activity with near-millisecond precision. *Nat Methods* 7: 399–405, 2010.

Grosenick L, Anderson T, Smith SJ. Elastic source selection for in vivo imaging of neuronal ensembles. In: *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging: From Nano to Macro* (ISBI '09). New York: IEEE Press, 2009, p. 1263–1266.

Holekamp TF, Turaga D, Holy TE. Fast three-dimensional fluorescence imaging of activity in neural populations by objective-coupled planar illumination microscopy. *Neuron* 57: 661–672, 2008.

Horn R, Johnson C. *Matrix Analysis.* Cambridge, UK: Cambridge Univ. Press, 1990.

Huys QJM, Ahrens MB, Paninski L. Efficient estimation of detailed single-neuron models. *J Neurophysiol* 96: 872–890, 2006.

Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, Yuste R. Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 304: 559–564, 2004.

Joucla S, Pippow A, Kloppenburg P, Pouzat C. Quantitative estimation of calcium dynamics from radiometric measurements: a direct, nonratioing method. *J Neurophysiol* 103: 1130–1144, 2010.

Kass R, Raftery A. Bayes factors. *J Am Stat Assoc* 90: 773–795, 1995.

Koenker R, Mizera I. Quasi-concave density estimation. *Ann Stat* 38: 2998–3027, 2010.

Lee DD, Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 401: 788–791, 1999.

Lin Y, Lee DD, Saul LK. Nonnegative deconvolution for time of arrival estimation. In: *Proceedings of the 2004 International Conference on Acoustics, Speech, and Signal Processing* (ICASSP 2004). New York: IEEE Press, 2004, p. 377–380.

Luo L, Callaway EM, Svoboda K. Genetic dissection of neural circuits. *Neuron* 57: 634–660, 2008.

MacLean J, Watson B, Aaron G, Yuste R. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron* 48: 811–823, 2005.

Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Trans Signal Process* 41: 3397–3415, 1993.

Mank M, Santos AF, Direnberger S, Mrsic-Flogel TD, Hofer SB, Stein V, Hendel T, Reiff DF, Levelt C, Borst A, Bonhoeffer T, Hbener M, Griesbeck O. A genetically encoded calcium indicator for chronic in vivo two-photon imaging. *Nat Methods* 5: 805–811, 2008.

Mao B, Hamzei-Sichani F, Aronov D, Froemke R, Yuste R. Dynamics of spontaneous activity in neocortical slices. *Neuron* 32: 883–898, 2001.

Markham J, Conchello J-A. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *J Opt Soc Am A Opt Image Sci Vis* 16: 2377–2391, 1999.

Mishchenko Y, Vogelstein J, Paninski L. A Bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. *Ann Appl Stat* http://www.imstat.org/aoas/nextissue.html.

Mukamel EA, Nimmerjahn A, Schnitzer MJ. Automated analysis of cellular signals from large-scale calcium imaging data. *Neuron* 63: 747–760, 2009.

Nagayama S, Zeng S, Xiong W, Fletcher ML, Masurkar AV, Davis DJ, Pieribone VA, Chen WR. In vivo simultaneous tracing and Ca^2 imaging of local neuronal circuits. *Neuron* 53: 789–803, 2007.

O'Grady PD, Pearlmutter BA. Convolutive non-negative matrix factorisation with a sparseness constraint. In: *Proceedings of the International Workshop on Machine Learning for Signal Processing, 2006.* New York: IEEE Press, 2006, p. 427–432.

Paninski L, Ahmadian Y, Ferreira D, Koyama S, Rad KR, Vidne M, Vogenstein J, Wu W. A new look at state-space models for neural data. *J Comput Neurosci*. doi: 10.1007/s10827-009-0179-x, 1–20, 2009.

Pologruto TA, Yasuda R, Svoboda K. Monitoring neural activity and [Ca^2] with genetically encoded Ca^2 indicators. *J Neurosci* 24: 9572–9579, 2004.

Portugal LF, Judice JJ, Vicente LN. A comparison of block pivoting and interior-point algorithms for linear least squares problems with nonnegative variables. *Math Comput* 63: 625–643, 1994.

Press W, Teukolsky S, Vetterling W, Flannery B. *Numerical Recipes in C*. Cambridge, UK: Cambridge Univ. Press, 1992.

Reiff DF, Ihring A, Guerrero G, Isacoff EY, Joesch M, Nakai J, Borst A. In vivo performance of genetically encoded indictors of neural activity in flies. *J Neurosci* 25: 4766–4778, 2005.

Sasaki T, Takahashi N, Matsuki N, Ikegaya Y. Fast and accurate detection of action potentials from somatic calcium fluctuations. *J Neurophysiol* 100: 1668–1676, 2008.

Schwartz T, Rabinowitz D, Unni VK, Kumar VS, Smetters DK, Tsiola A, Yuste R. Networks of coactive neurons in developing layer 1. *Neuron* 20: 1271–1283, 1998.

Seeger M. Bayesian inference and optimal design for the sparse linear model. *J Machine Learn Res* 9: 759–813, 2008.

Sjulson L, Miesenböck G. Optical recording of action potentials and other discrete physiological events: a perspective from signal detection theory. *Physiology* (Bethesda) 22: 47–55, 2007.

Smetters D, Majewska A, Yuste R. Detecting action potentials in neuronal populations with calcium imaging. *Methods* 18: 215–221, 1999.

Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninski L. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophys J* 97: 636–655, 2009.

Wallace DJ, zum Alten Borgloh SM, Astori S, Yang Y, Bausen M, Kgler S, Palmer AE, Tsien RY, Sprengel R, Kerr JND, Denk W, Hasan MT. Single-spike detection in vitro and in vivo with a genetic Ca^2 sensor. *Nat Methods* 5: 797–804, 2008.

Watson BO, MacLean JN, Yuste R. Up states protect ongoing cortical activity from thalamic inputs. *PLoS ONE* 3: e3971, 2008.

Wu MC-K, David SV, Gallant JL. Complete functional characterization of sensory neurons by system identification. *Annu Rev Neurosci* 29: 477–505, 2006.

Yaksi E, Friedrich RW. Reconstruction of firing rate changes across neuronal populations by temporally deconvolved Ca^2 imaging. *Nat Methods* 3: 377–383, 2006.

Yuste R, Konnerth A. Editors. *Imaging in Neuroscience and Development: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 2005.

Yuste R, Katz LC. Control of postsynaptic Ca^2 influx in developing neocortex by excitatory and inhibitory neurotransmitters. *Neuron* 6: 333–344, 1991.

Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Statist Soc B Stat Methodol* 67: 301–320, 2005.