The Dynamics of Balanced Neural Networks under Spike–Timing Dependent Plasticity

by
Alan Eric Akil

A dissertation submitted to the Department of Mathematics,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Mathematics

Chair of Committee: Krešimir Josić

Committee Member: William Ott

Committee Member: Andrew Török

Committee Member: Robert Rosenbaum

University of Houston
May 2021

ACKNOWLEDGMENTS

The work presented in this dissertation would have not been possible without the emotional and professional support and help from a number of people. I owe them my deepest gratitude.

First and foremost, my advisor Dr. Krešimir Josić. I am eternally grateful to him for believing in me, supporting me through everything, and giving me invaluable opportunities to grow as a researcher. He was always closely involved in my work to make sure I was progressing towards my degree and growing as a scientist. I am forever thankful for having such a caring and outstanding mentor.

I had the great pleasure of working with Dr. Robert Rosenbaum, who was our closest collaborator and nearly served as an unofficial co–advisor to me. He was part of all the projects that are presented in this dissertation, and he was always available to meet, give feedback, and guide me. I am deeply thankful for the opportunity to collaborate with him.

The work in Chapter 4 is the result of a fruitful collaboration with Dr. Ariana Andrei and Dr. Valentin Dragoi from the University of Texas Health Science Center. I am grateful for the opportunity to work with them and for their valuable input throughout.

I would also like to thank Dr. Andrew Török and Dr. William Ott who have served in my committee since the first APR and always provided constructive feedback.

Special thanks to the staff at the Department of Mathematics at UH who helped me throughout all administrative matters: Emilio, Amy, Mariela, Linda, Neha, Gloria, Dave, Jason, Tai, and others. As well as to all the members at the Josić lab and graduate students at the Department of Mathematics whom I have interacted with in the last five years, especially Sarah, Jen, Kate, and Rita.

I am grateful for my closest friends who have supported me and offered advice throughout the last couple of years: Pruthil, Fer, Max, and Luis Diaz, to name a few.

Last but not least, I would like to thank Diana Anzures for her unconditional support, as well as my family. Particularly, my father Julio who not only encouraged my interest in science from a young age, but also gave me the opportunity to pursue my studies in the U.S. and supported me throughout.

ABSTRACT

The dynamics of local cortical networks are irregular, but correlated. Dynamic excitatory–inhibitory balance is a plausible mechanism that generates such irregular activity, but it remains unclear how balance is achieved and maintained in plastic neural networks. In particular, it is not fully understood how plasticity induced changes in the network affect balance, and in turn, how correlated, balanced activity impacts the structure of the network, and, ultimately, learning. How do the dynamics of balanced networks change under different plasticity rules? How does correlated spiking activity in recurrent networks change the evolution of weights, their eventual magnitude, and weight structure across the network? To address these questions, we develop a theory of spike–timing dependent plasticity in balanced networks. We show that balance can be attained and maintained under plasticity–induced weight changes. We find that correlations in the input mildly affect the evolution of synaptic weights. Under certain plasticity rules, we find an emergence of correlations between firing rates and synaptic weights. Under these rules, synaptic weights converge to a stable manifold in weight space with their final configuration dependent on the initial state of the network. Lastly, we show that our framework can describe the dynamics of plastic, balanced networks when subsets of neurons receive targeted optogenetic input.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Mammalian brains are comprised of complex large–scale networks of neurons that work jointly to process sensory information and guide behavior. They are also home to our conscious and subconscious thoughts, feelings, and emotions. The brain is a high dimensional, complex system, and even though we have some understanding of the function of certain brain areas and regions, we are a long way from fully understanding how it works. Although a desire to better understand the brain is often fueled by curiosity, there are immense potential benefits to humankind that would follow from such understanding. The most important benefits lie in the field of medicine: studying the brain eventually may lead to more effective treatments or even cures for incapacitating or deadly brain diseases such as epilepsy, Alzheimer's disease, dementia, depression, and others. Another field fueled by developments in Neuroscience, which spans an enormous number of applications, is Artificial Intelligence (AI) where models are frequently inspired by experimental observations from Neuroscience.

In this dissertation, we focus on the study of the dynamics of neural networks, which could eventually lead to better models of the brain that can be applied in other areas of Neuroscience. In mathematics, the dynamics of a system refers to a description of its evolution in time. The dynamics of a neural network can typically be described in terms of the activity of neurons, firing patterns, spike train correlations (measure of synchrony), and changes in synaptic weights (when

these are plastic).

1.1 The Dynamical Regime of the Cortex

The activity of neurons in the nervous system can be described in terms of spike patterns or firing rates. Spiking network models, which are the focus of this dissertation, simulate dynamics at a great level of detail since individual spikes take place over 1–3 ms. Firing rate models simulate average activity of individual neurons or of populations of neurons. In this case, rates represent the average number of spikes elicited in a time window.

The firing of neurons in a neuronal network may be regular or irregular [3]. Roughly, regular spiking is predictable and characterized by a very narrow distributions of inter–spike intervals (ISIs). The coefficient of variation of the ISIs, CV_{ISI}, is hence close to zero. Irregular spiking is harder to predict, and ISIs have a broader distribution. For instance, a spike train with the statistics of a Poisson process ($CV_{ISI} \approx 1$) would be irregular. Neurons in the mammalian cortex typically fire in an irregular fashion [26,50,121,127], although spiking in some areas of the neocortex and elsewhere may be more regular [89].

A natural question that arose in the 1990's was: What is the mechanism behind irregular firing of cortical neurons? It was first proposed theoretically [137,138], and later supported by experimental evidence, that irregular spiking was a consequence of balanced excitatory and inhibitory inputs [8, 12,29,41,56,61,103,104,147,151]. This question, and the proposed solutions, have been debated for many years, but it seems safe to assume, based on experimental evidence, that cortical networks operate in a balanced regime. However, this may not be the case for subcortical networks or cortical networks in pathological states.

Cortical dynamics can be characterized by the distribution of firing rates across neurons, and their tendency to synchronize. The firing rates of neurons in cortical networks range from less than one spike per second to nearly 100 spikes per second [53, 77]. More recent experiments have shown that firing rate distributions over neurons in several cortical areas are typically unimodal

and heavy–tailed [68, 101, 117, 122].

Activity of populations of neurons may be asynchronous or synchronous [3]. Neural synchrony can be measured by spike count correlations, which are a normalized measure of co–variability in the activity of pair of neurons. More precisely, spike count are obtained by partitioning spike trains into equal intervals of time of fixed size (time window) and counting spikes in each time window. The Pearson correlation coefficient between these spike counts defines the spike count correlation between the pair of neurons. In an asynchronous regime, mean spike count correlations are very weak. While in synchronous regimes, neurons may be strongly correlated and often tend to fire together. Real local cortical networks may spike asynchronously [25, 34, 35] or synchronously [103, 125, 126, 131], depending on several conditions (*e.g.*, awake vs. asleep, brain area, etc.).

Distributions of firing rates provide a large scale picture of the spiking in a network. When analyzing experimental or simulated data, frequently one reduces this distribution to its first moment: the mean firing rate. The behavior (or dynamics) of the mean rate has been the focus of many theoretical studies. In the late 1980's, it was proposed that the rates followed chaotic trajectories [128]. Others proposed multi–attractor models for rates, where random fluctuations drive firing rate from one basin of attraction to another [17, 22, 28, 86, 87]. Here we will assume that firing rates are fixed and stable for fixed input, and change in response to changes in the synaptic weights. This is a simplifying assumption that seems to hold approximately in sensory cortical areas [3, 94].

1.2 Dynamics of Balanced Networks

In the 1990's, it was proposed that experimentally observed irregular spiking activity is a consequence of a balance between excitation and inhibition [120, 121, 137, 138]. This led to the development of a class of models called balanced neuronal networks. Early theoretical work on such networks focused on irregular, asynchronous dynamics, with large networks exhibiting vanishing correlations [112, 137, 138]. However, more recent extensions have shown how correlated dynamics can be generated both endogenously and exogenously, while preserving irregular single cell

activity [10, 27, 69, 84, 93, 95, 116, 123, 145], showing the existence of both asynchronous and correlated states in balanced networks. These model networks also share relevant features with cortical networks including stable fixed firing rates for fixed input, and unimodal, long tailed firing rate distributions.

Balanced network theory provides a mean–field description of the macroscopic dynamics (see Chapter 2). However, we still lack a deeper understanding on how these networks behave under spike–timing dependent plasticity (STDP): Will the network remain stable and balanced as synaptic weights change? Can we extend the mean–field theory to describe networks as the network structure evolves? We will revisit these questions at the end of this chapter, after we review experiments and models of synaptic plasticity.

1.3 Brief Overview of Synaptic Plasticity

The term synaptic plasticity refers to the changes in the synaptic architecture of neural networks. In experiments, this is often quantified by measuring changes in the amplitude of the post–synaptic potential. In computational models, plasticity is most commonly measured by the change in synaptic weights (also known as synaptic strength, or coupling strength) which are determinant of the amplitude of the post–synaptic potential.

In 1949, Donald Hebb postulated in his book "The Organization of Behavior" a mechanism for the formation of memories and neural assemblies in the brain [60]:

> "When an axon of 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 that A's efficiency, as one of the cells firing B, is increased."

Hebb's postulate is often summarized as "neurons that fire together wire together", and it has had massive influence in the field of synaptic plasticity and learning. This general idea that co–activation of cells leads to plastic changes in synaptic weights is referred to as "Hebbian plasticity", and the resulting strongly connected groups of neurons are often called neural assemblies

or "Hebbian assemblies" [91].

There has been increasing evidence in favor of Hebb's postulate since it was originally proposed. Kandel and Tauc performed a classical behavior conditioning experiment in the abdominal ganglion of a sea slug *Aplysia depilans*, where they stimulated two different neural pathways [72, 73]. They demonstrated that repeated, temporal pairing of ineffective stimuli with effective stimuli can lead to long term potentiation (LTP) of the weaker neural pathway.

Complementary to Hebb's postulate, in the late 1970's, Lynch et al. reported evidence of long term depression (LTD) in a rat's hippocampus after repeated low frequency stimulation [32, 88].

More recently, in the 1990's, Markram et al. and later Bi and Poo showed in dual whole–cell voltage recordings of pyramidal cells the spike–time dependence of changes in synaptic weights by pairing different temporal patterns of pre– and post–synaptic spikes and analyzing the resulting changes in excitatory post–synaptic potentials (EPSP's) [14, 92].

STDP is not the only form of synaptic plasticity. Plastic changes in synaptic weights can depend on firing rates of pre– and post–synaptic cells, as well as on neuromodulators, or Calcium traces. In this dissertation, we only focus on plasticity rules dependent on spike timing, even though such rules can often be described in terms of firing rates.

1.4 Theory of Spike–Timing Dependent Plasticity

Experimental evidence outlined above suggests that the precise timing between pre-and-post synaptic spikes plays an important role in the process of synaptic plasticity [91]. Theories and models of synaptic plasticity have been developed over many years in line with these experimental findings. Before spike timing took the spotlight in synaptic plasticity, a number of rate–based models were developed and were actually shown to be biologically relevant. For example, in 1982, Oja's and the Bienenstock, Cooper, and Munro (BCM) rules were introduced [15, 102]. Inspired by Hebb's postulate, Oja proposed a model in which synaptic weights would change proportional to both pre–synaptic input and post–synaptic activity – so co–active pairs undergo LTP [102]. On the other

hand, Bienenstock, Cooper, and Munro proposed a model where the Hebb–like product of pre– and post–synaptic activities is controlled by a sliding threshold that controls the activity of the post–synaptic cell [15]. The threshold is typically some function of the inputs and hence also changes over time to induce LTD or LTP at different levels of activity, eventually leading to stabilization of synaptic weights. The BCM rule has been shown to model well selectivity in the visual cortex and hippocampus; and both BCM and Oja's rule can be used to perform principal component analysis (PCA) on a set of inputs.

In the 1990's the basic STDP model was introduced where synaptic weights change according to an STDP function that depends on the relative timing of pre– and post–synaptic spikes [42, 74]. Analytical treatments of such STDP rules have been proposed for a number of cases, starting with the description of mean synaptic dynamics on single post–synaptic neurons receiving Poisson feed–forward input [55, 74]. Kempter et al. developed a self–consistent theory for the case of a Poisson post–synaptic neuron that is the basis for the STDP theory that we will develop in this dissertation [74]. Later Gütig et al. studied mean synaptic dynamics when the post–synaptic neurons was integrate–and–fire [55].

The relationship between STDP and network dynamics is complex due to the large number of connections and the effect of recurrence, and so a large number of theoretical studies have constrained their analysis to numerical simulations only [9, 71, 87, 98].

However, eventually, analytical results were extended to the case of recurrent networks of Poisson neurons under no external input by Burkitt and his colleagues [23]. Right after, Gilson and others followed up with an analytical description of network dynamics and weight dynamics in recurrent networks along with a number of theoretical results in a series of five papers [44–48].

Other work provided analytical treatments of specific plasticity rules, such as homeostatic inhibitory plasticity [130, 141]. Using linear response and motif resumming techniques [135], Ocker et al. developed a theory describing the evolution of mean synaptic weights in recurrent neural networks of noisy integrate–and–fire neurons under STDP [100]. This approach relies on the assumption that the input to individual cells is dominated by white noise, local synaptic input is

weak, and integral of the STDP function is small.

Assuming that neurons fire as Poisson processes, Ravid Tannenbaum et al. showed that synfire chains and self connected assemblies can emerge autonomously in recurrent networks [111]. Montangie et al. showed that a more realistic form of STDP based on spike triplets also leads to autonomous emergence of assemblies [97].

1.5 Motivation of this Dissertation

Previous work presented in Sections 1.2 and 1.4 has a number of limitations that were briefly mentioned and we review here to motivate this dissertation.

On one hand, balanced networks are biologically relevant and sound models that can generate cortical–like activity. However, current extensions of the original theory still lack a description of changes in synaptic weights, at least at a mean–field level. This leads to a major issue in studies of numerical simulations: one cannot predict anything about the dynamics beforehand! For example, it would not be known under which STDP rules the network will remain stable or balanced, how dynamics can change as a consequence of weight changes, and how correlated activity can affect the weights themselves. Therefore, there is a clear need for an extension of balanced network's mean–field theory that accounts for changes in mean synaptic weights.

On the other hand, almost the entirety of previous work on STDP in recurrent networks assumes Poisson spiking. This presents as a limitation, because neurons may fire Poisson–like on average, but every single neurons is not exactly Poissonian [50, 89, 121, 127]. In addition, models of Poisson neurons generate irregular activity via randomness in spike generation, however internal noise is not the mechanism behind irregular firing in real neurons [143]. In fact, isolated neurons are spike quite reliably, and it is likely that irregularity emerges from network input noise or synaptic failure. Other more biologically–grounded neuron models such as the class of integrate–and–fire models, can generate irregular activity via near cancellation of excitatory and inhibitory inputs as described earlier in Sections 1.1 and 1.2. Previous theories of STDP that have used more realistic neuron

models, *e.g.*, see Ocker et al. [100]. However, their work has other limitations and assumptions. For example, they assume recurrent synaptic input is weak, and neurons are driven by strong Gaussian white noise; even though cortical neurons are known to receive temporally and spatially correlated inputs.

Here, we develop a complementary theory describing the evolution of synaptic weights and associated mean rates in tightly balanced networks in both correlated and asynchronous states. We combine the mean–field theory of firing rates and correlations in balanced networks [10, 83, 110, 115, 116, 137, 138] with an averaging approach assuming a separation of timescales between changes in spiking activity, and the evolution of synaptic weights [74]. We show how the weights and the network dynamics co–evolve under different classical rules, such as Hebbian plasticity, Kohonen's rule, and a form of inhibitory plasticity [14, 60, 78, 92, 141]. In general, the predictions of our theory agree well with empirical simulations. We also explain when the mean–field theory fails, leading to disagreements with simulations, and we develop a semi–analytic extension of the theory that explains these disagreements.

We find that spike train correlations, in general, have a mild effect on the synaptic weights and firing rates, in agreement with previous work [52, 100]. We also show that for some STDP rules, synaptic competition can introduce correlations between synaptic weights and firing rates, resulting in the formation of a stable manifold of fixed points in weight space, and hence asymptotic weight distributions that depend on the initial state. Finally, we apply this theory to show how inhibitory STDP [141] can lead to a reestablishment of an asynchronous, balanced state that is broken by optogenetic stimulation of a neuronal subpopulation [33]. We thus extend the classical theory of balanced networks to understand how synaptic plasticity shapes their dynamics.

In addition to our theoretical results on plastic, balanced networks, we applied our framework to model optogenetically–induced changes in functional connectivity in a local network of visual cortex V1 of macaque monkeys reported by Andrei et al. [6]. In particular, we show that inhibitory STDP in a balanced network can lead to a dynamic reversal of spike train correlations following repeated optogenetic activation of a subset of excitatory cells, similar to that reported in experiments [6].

Chapter 2

Mean Field Theory of Plastic Balanced Networks

In mammals, local cortical networks can be comprised of thousands of cells, with each neuron receiving thousands of inputs from cells within the local network, and other cortical layers, areas, and thalamus [16]. Predominantly excitatory, long–range inputs would lead to high, regular firing unless counteracted by local inhibition. To reproduce the sparse, irregular activity observed in cortex, model networks often exhibit a balance between excitatory and inhibitory inputs [4, 10, 30, 64, 112, 115, 137, 138]. This balance can be achieved robustly and without tuning, when synaptic weights are scaled like $\mathcal{O}(1/\sqrt{N})$, where N is the network size [137, 138]. In this balanced state mean excitatory and inhibitory inputs cancel one another, and the activity is asynchronous [112]. Inhibitory inputs can also track excitation at the level of cell pairs, cancelling each other in time, and produce a correlated state [10, 103].

We first review the mean–field description of asynchronous and correlated states in balanced networks, and provide expressions for firing rates and spike count covariances averaged over subpopulations that accurately describe networks of more than a few thousand neurons [10, 83, 110, 115, 116, 137, 138]: Let N be the total number of neurons in a recurrent network composed of N_e excitatory and N_i inhibitory neurons. Cells in this recurrent network also receive input from N_x external

Poisson neurons firing at rate r_{x}, and with pairwise correlation c_{x} (see Fig. 3.1.4 **A** in Chapter 3). We use subscripts e, i and x for the local excitatory, local inhibitory, and external populations, respectively. The spike train of neuron $j = 1, \ldots, N_a$ in population $a = \mathrm{e}, \mathrm{i}, \mathrm{x}$ is represented as a sum of Dirac delta functions,

$$S_j^a(t) = \sum_n \delta(t - t_n^{a,j}) \tag{2.1}$$

where $t_n^{a,j}$ is the n^{th} spike time of the neuron. The synaptic input current to neuron $j = 1, \ldots, N_a$ in population $a = \mathrm{e}, \mathrm{i}$ is given by

$$T_j^a(t) = R_j^a(t) + X_j^a(t) \tag{2.2}$$

where $R_j^a(t) = E_j^a(t) + I_j^a(t)$ is the recurrent input current and $X_j^a(t)$ is the external input, and

$$U_j^a(t) = \sum_{k=1}^{N_b} J_{jk}^{ab} \big(\alpha_b * S_k^b\big)(t) \tag{2.3}$$

sums inputs from populations $U = E, I, X$, and $b = \mathrm{e}, \mathrm{i}, \mathrm{x}$. Here, $*$ denotes the convolution operation, and J_{jk}^{ab} is the synaptic weight from pre–synaptic neuron k in population b to post–synaptic neuron j in population a, and $\alpha_b(t)$ is a post–synaptic current (PSC) waveform, and is defined by

$$\alpha_b(t) = \tau_b^{-1} e^{-t/\tau_b} H(t) \tag{2.4}$$

where $H(t)$ is the Heaviside step function, and τ_b is the synaptic timescale of neurons in population $b = \mathrm{e}, \mathrm{i}, \mathrm{x}$. In all simulations we take $\tau_{\mathrm{x}} = 10$ ms, $\tau_{\mathrm{e}} = 8$ ms, and $\tau_{\mathrm{i}} = 4$ ms [10]. Without loss of generality, we assume that $\int \alpha_b(t) = 1$.

We consider a random, blockwise Erdős–Rényi connectivity structure with

$$J_{jk}^{ab} = \frac{1}{\sqrt{N}} \begin{cases} j_{ab} & \text{with prob. } p_{ab} \\ 0 & \text{otherwise,} \end{cases} \tag{2.5}$$

where connections are statistically independent and $j_{ab}, p_{ab} \sim \mathcal{O}(1)$ for $b = \mathrm{e}, \mathrm{i}, \mathrm{x}$ and $a = \mathrm{e}, \mathrm{i}$. In all simulations, we set $p_{ab} = 0.1$ for all $b = \mathrm{e}, \mathrm{i}, \mathrm{x}$ and $a = \mathrm{e}, \mathrm{i}$.

2.1 Mean Field Theory of Firing Rates in Balanced Networks

We denote by r_a the mean firing rate of neurons in population $a = \mathrm{e}, \mathrm{i}, \mathrm{x}$, and let the components of the 2×1 vector,

$$\boldsymbol{r} = \begin{bmatrix} r_{\mathrm{e}} \\ r_{\mathrm{i}} \end{bmatrix}, \tag{2.6}$$

describe the mean activity of both subpopulations. The mean–field synaptic inputs to neurons in populations $a = \mathrm{e}, \mathrm{i}$ are denoted by $\overline{U}_a = \langle U_j^a(t) \rangle_j$ for $U = E, I, X, R, T$. In vector form,

$$\overline{\boldsymbol{U}} = \begin{bmatrix} \overline{U_{\mathrm{e}}} \\ \overline{U_{\mathrm{i}}} \end{bmatrix}. \tag{2.7}$$

We also define the recurrent and feed–forward mean–field connectivity matrices,

$$\overline{W} = \begin{bmatrix} \overline{w}_{\mathrm{ee}} & \overline{w}_{\mathrm{ei}} \\ \overline{w}_{\mathrm{ie}} & \overline{w}_{\mathrm{ii}} \end{bmatrix}, \quad \text{and} \quad \overline{W}_x = \begin{bmatrix} \overline{w}_{\mathrm{ex}} \\ \overline{w}_{\mathrm{ix}} \end{bmatrix}, \tag{2.8}$$

where $\overline{w}_{ab} = p_{ab} j_{ab} q_b \sim \mathcal{O}(1)$. Here, we have defined $q_b = N_b/N$ which are assumed to be $\mathcal{O}(1)$ for $b = \mathrm{e}, \mathrm{i}$. Throughout all examples, we set $q_{\mathrm{e}} = 0.8$ and $q_{\mathrm{i}} = 0.2$. We assume that both the probabilities, p_{ab}, and weights, j_{ab} are constant across pairs of subpopulations.

We review the mean–field analysis of firing rates in the balanced state [10, 83, 110, 115, 137, 138].

The mean external input to the E subpopulation is given by

$$\overline{X}_{\mathrm{e}} = \frac{1}{N_{\mathrm{e}}} \sum_{j=1}^{N_{\mathrm{e}}} \sum_{k=1}^{N_{\mathrm{x}}} J_{jk}^{\mathrm{ex}} \big(\alpha_{\mathrm{x}} * S_k^{\mathrm{x}}\big)(t) \tag{2.9}$$

$$= \Big[N_{\mathrm{x}} \Big[J_{jk}^{\mathrm{ex}} \big(\alpha_{\mathrm{x}} * S_k^{\mathrm{x}}\big)(t) \Big]_k \Big]_j \tag{2.10}$$

$$= N_{\mathrm{x}} \Big[\Big[J_{jk}^{\mathrm{ex}} \Big]_k \Big]_j \Big[\Big[\big(\alpha_{\mathrm{x}} * S_k^{\mathrm{x}}\big)(t) \Big]_k \Big]_j \tag{2.11}$$

$$= N_{\mathrm{x}} p_{\mathrm{ex}} \frac{j_{\mathrm{ex}}}{\sqrt{N}} r_{\mathrm{x}} \frac{N}{N} \tag{2.12}$$

$$= \sqrt{N} p_{\mathrm{ex}} j_{\mathrm{ex}} \frac{N_{\mathrm{x}}}{N} r_{\mathrm{x}} \tag{2.13}$$

$$= \sqrt{N} p_{\mathrm{ex}} j_{\mathrm{ex}} q_{\mathrm{x}} r_{\mathrm{x}} \tag{2.14}$$

$$= \sqrt{N} \overline{w}_{\mathrm{ex}} r_{\mathrm{x}}. \tag{2.15}$$

where we denoted averages over neurons by $[\cdot]$, we assumed correlations between synaptic weights J_{jk}^{ex} and activities $\big(\alpha_{\mathrm{x}} * S_k^{\mathrm{x}}\big)$ are negligible, and multiplied by $\frac{N}{N}$. In an identical derivation replacing e with i, we obtain the mean external input to the I population:

$$\overline{X}_{\mathrm{i}} = \sqrt{N} \overline{w}_{\mathrm{ix}} r_{\mathrm{x}}. \tag{2.16}$$

In matrix form, the mean external input is

$$\overline{\boldsymbol{X}} = \sqrt{N} \overline{W}_{\mathrm{x}} r_{\mathrm{x}}. \tag{2.17}$$

The mean recurrent input to the E subpopulation is

$$\overline{R}_{\mathrm{e}} = \frac{1}{N_{\mathrm{e}}} \sum_{j=1}^{N_{\mathrm{e}}} \sum_{k=1}^{N_{\mathrm{e}}} J_{jk}^{\mathrm{ee}} \big(\alpha_{\mathrm{e}} * S_k^{\mathrm{e}}\big)(t) + \frac{1}{N_{\mathrm{e}}} \sum_{j=1}^{N_{\mathrm{e}}} \sum_{m=1}^{N_{\mathrm{i}}} J_{jm}^{\mathrm{ei}} \big(\alpha_{\mathrm{i}} * S_m^{\mathrm{i}}\big)(t) \tag{2.18}$$

$$= \Big[N_{\mathrm{e}} \Big[J_{jk}^{\mathrm{ee}} \big(\alpha_{\mathrm{e}} * S_k^{\mathrm{e}}\big)(t) \Big]_k \Big]_j + \Big[N_{\mathrm{i}} \Big[J_{jm}^{\mathrm{ei}} \big(\alpha_{\mathrm{i}} * S_m^{\mathrm{i}}\big)(t) \Big]_m \Big]_j \tag{2.19}$$

$$= N_{\mathrm{e}} \Big[\Big[J_{jk}^{\mathrm{ee}} \Big]_k \Big]_j \Big[\Big[\big(\alpha_{\mathrm{e}} * S_k^{\mathrm{e}}\big)(t) \Big]_k \Big]_j + N_{\mathrm{i}} \Big[\Big[J_{jm}^{\mathrm{ei}} \Big]_m \Big]_j \Big[\Big[\big(\alpha_{\mathrm{i}} * S_m^{\mathrm{i}}\big)(t) \Big]_m \Big]_j \tag{2.20}$$

$$= N_{\mathrm{e}} p_{\mathrm{ee}} \frac{j_{\mathrm{ee}}}{\sqrt{N}} r_{\mathrm{e}} \frac{N}{N} + N_{\mathrm{i}} p_{\mathrm{ei}} \frac{j_{\mathrm{ei}}}{\sqrt{N}} r_{\mathrm{i}} \frac{N}{N} \tag{2.21}$$

$$= \sqrt{N} \big(\overline{w}_{\mathrm{ee}} r_{\mathrm{e}} + \overline{w}_{\mathrm{ei}} r_{\mathrm{i}} \big) \tag{2.22}$$

where we have separated the input into excitatory and inhibitory, and again ignored correlations between synaptic weights and rates. In particular, we have assumed $\Big[\Big[J_{jk}^{\mathrm{ee}} \big(\alpha_{\mathrm{e}} * S_k^{\mathrm{e}}\big)(t) \Big]_k \Big]_j \approx \Big[\Big[J_{jk}^{\mathrm{ee}} \Big]_k \Big]_j \Big[\Big[\big(\alpha_{\mathrm{e}} * S_k^{\mathrm{e}}\big)(t) \Big]_k \Big]_j$ and $\Big[\Big[J_{jm}^{\mathrm{ei}} \big(\alpha_{\mathrm{i}} * S_m^{\mathrm{i}}\big)(t) \Big]_m \Big]_j \approx \Big[\Big[J_{jm}^{\mathrm{ei}} \Big]_m \Big]_j \Big[\Big[\big(\alpha_{\mathrm{i}} * S_k^{\mathrm{i}}\big)(t) \Big]_m \Big]_j$. This is an assumption widely made across the non–plastic balanced network literature [10, 11, 83, 110, 115, 137, 138], however rarely presented explicitly [10]. We make use of this assumption when deriving mean–field equations for spike train covariances in Section 2.4, and we show in Section 3.2 that this assumption may be violated when synaptic weights become strongly correlated with firing rates under certain plasticity rules. The mean recurrent input to the I subpopulation is obtained in the exact same way, but replacing e post–synaptic cell with i

$$\overline{R}_{\mathrm{i}} = \sqrt{N} \big(\overline{w}_{\mathrm{ie}} r_{\mathrm{e}} + \overline{w}_{\mathrm{ii}} r_{\mathrm{i}} \big). \tag{2.23}$$

In matrix form, the mean recurrent input is given by

$$\overline{\boldsymbol{R}} = \sqrt{N} \overline{W} \boldsymbol{r}. \tag{2.24}$$

The mean total synaptic input is the sum of recurrent and external contributions

$$\overline{\boldsymbol{T}} = \sqrt{N} \left[\overline{W} \boldsymbol{r} + \overline{W}_{\mathrm{x}} r_{\mathrm{x}} \right]. \tag{2.25}$$

We next make the ansatz that in the balanced state the mean input and firing rates remain finite as the network grows, *i.e.*, $\overline{\boldsymbol{T}}, \boldsymbol{r} \sim \mathcal{O}(1)$ [10,83,110,115,137,138]. This can only be obtained by a precise cancellation between external and recurrent synaptic inputs. Such cancellation requires $\overline{W}\boldsymbol{r} + \overline{W}_{\mathrm{x}} r_{\mathrm{x}} \sim \mathcal{O}(1/\sqrt{N})$ so that

$$\lim_{N\to\infty} \boldsymbol{r} = -\overline{W}^{-1}\overline{W}_{\mathrm{x}} r_{\mathrm{x}} \tag{2.26}$$

in the balanced state, provided $\overline{X}_e/\overline{X}_i > \overline{w}_{ei}/\overline{w}_{ii} > \overline{w}_{ee}/\overline{w}_{ie}$ [137,138] to ensure positive rates are realized. The firing rates in Eq. (2.26) depend only on the recurrent structure and mean input, and are hence independent of the correlation structure in the network.

2.2 Scaling of Synaptic Weights

A natural question that arises is why scale synaptic weights as $1/\sqrt{N}$? Why is this scaling used and why is preferred to other scaling such as $1/N$? To answer these questions, we can use a rate model that approximate rate dynamics of the spiking neural network:

$$\frac{d\mathbf{r}}{dt} = -\mathbf{r} + f\left(\overline{JK}[V\mathbf{r} + V_{\mathrm{x}} r_{\mathrm{x}}]\right) \tag{2.27}$$

where f is the input–response curve (also known as f-I curve), $\overline{JK} = \mathrm{mean}(|J_{ab}|K_{ab})$, K_{ab} is the mean number of connections from population b to a, and the components of V and V_{x} are given by $v_{ab} = J_{ab}K_{ab}/\overline{JK}$ for $a = \mathrm{e}, \mathrm{i}$ and $b = \mathrm{e}, \mathrm{i}, \mathrm{x}$.

Eq. (2.27) admits the following stable solution:

$$\mathbf{r} = f\left(\overline{JK}[V\mathbf{r} + V_{\mathrm{x}} r_{\mathrm{x}}]\right). \tag{2.28}$$

We rearrange Eq. (2.28) and arrive at

$$\frac{1}{\overline{JK}} f^{-1}(\mathbf{r}) = V\mathbf{r} + V_{\mathrm{x}} r_{\mathrm{x}}. \tag{2.29}$$

Scaling of $\overline{JK}$ as N^γ for $0 < \gamma \leq 1$ will lead to an expression for rates equivalent to Eq. (2.26):

$$\lim_{N\to\infty} \mathbf{r} = -V^{-1}V_{\mathrm{x}}r_{\mathrm{x}}. \tag{2.30}$$

In particular, using the scaling $J \sim \mathcal{O}(1/\sqrt{N})$, which implies $\overline{JK}$ grows as $N^{1/2}$, we can arrive at the balanced solution (Eqs. (2.26,2.30)). Other scalings, *i.e.*, $\gamma \neq 1/2$, lead to either frozen solutions ($\mathbf{r} = 0$), or narrow rate distributions [138]. Other scalings may lead to non–trivial solutions under different neuron models [118] (see Section 2.7 for more details.). Also, it follows that if $\overline{JK}$ scales as N^γ, the computation of firing rates does not require knowledge of the transfer function (f-I curve), which is an advantage of our theory compared to previous theories of STDP [44–48, 97, 100, 111].

We next show why $\gamma = 1/2$ is a preferable scaling. For the sake of the argument, let's assume we have a network of neurons where all excitatory synaptic weights are defined as J_E, and all inhibitory weights are J_I. Consider the total input to any cell:

$$Z(t) = J_E \sum_{j=1}^{N_{\mathrm{e}}} (\alpha_{\mathrm{e}} * S_j^{\mathrm{e}}(t)) + J_I \sum_{k=1}^{N_{\mathrm{i}}} (\alpha_{\mathrm{i}} * S_k^{\mathrm{i}}(t)). \tag{2.31}$$

The mean input is then

$$\begin{aligned}
\langle Z(t)\rangle_t &= \langle J_E \sum_{j=1}^{N_{\mathrm{e}}} (\alpha_{\mathrm{e}} * S_j^{\mathrm{e}}(t)) + J_I \sum_{k=1}^{N_{\mathrm{i}}} (\alpha_{\mathrm{i}} * S_k^{\mathrm{i}}(t))\rangle_t && (2.32)\\
&= J_E \sum_{j=1}^{N_{\mathrm{e}}} \langle \alpha_{\mathrm{e}} * S_j^{\mathrm{e}}(t)\rangle_t + J_I \sum_{k=1}^{N_{\mathrm{i}}} \langle \alpha_{\mathrm{i}} * S_k^{\mathrm{i}}(t)\rangle_t && (2.33)\\
&= J_E \sum_{j=1}^{N_{\mathrm{e}}} r_j^{\mathrm{e}} + J_I \sum_{k=1}^{N_{\mathrm{i}}} r_k^{\mathrm{i}} && (2.34)\\
&= J_E N_{\mathrm{e}} r_{\mathrm{e}} + J_I N_{\mathrm{i}} r_{\mathrm{i}} && (2.35)\\
&= J_E q_{\mathrm{e}} N r_{\mathrm{e}} + J_I q_{\mathrm{i}} N r_{\mathrm{i}} && (2.36)
\end{aligned}$$

here $\langle\cdot\rangle_t$ denotes time–average, and r_j^a is the time–averaged firing rate of neuron j in population a. The mean input, $\langle Z(t)\rangle_t$, will be order 1, so long as $J_E \approx -J_I$, for all values of N.

The variance of the input is

$$\mathrm{Var}[Z(t)]_t = \mathrm{Var}\left[J_E \sum_{j=1}^{N_\mathrm{e}} (\alpha_\mathrm{e} * S_j^\mathrm{e}(t)) + J_I \sum_{k=1}^{N_\mathrm{i}} (\alpha_\mathrm{i} * S_k^\mathrm{i}(t))\right]_t \tag{2.37}$$

$$= J_E^2 \sum_{j=1}^{N_\mathrm{e}} \mathrm{Var}\left[\alpha_\mathrm{e} * S_j^\mathrm{e}(t)\right]_t + J_I^2 \sum_{k=1}^{N_\mathrm{i}} \mathrm{Var}\left[\alpha_\mathrm{i} * S_k^\mathrm{i}(t)\right]_t \tag{2.38}$$

$$= J_E^2 \sum_{j=1}^{N_\mathrm{e}} \frac{1}{T_\mathrm{win}} r_j^\mathrm{e} + J_I^2 \sum_{k=1}^{N_\mathrm{i}} \frac{1}{T_\mathrm{win}} r_k^\mathrm{i} \tag{2.39}$$

$$= \frac{1}{T_\mathrm{win}} J_E^2 q_\mathrm{e} N r_\mathrm{e} + \frac{1}{T_\mathrm{win}} J_I^2 q_\mathrm{i} N r_\mathrm{i} \tag{2.40}$$

where we assumed neurons' firing is approximately independent and made the ansatz that spiking is Poissonian, so we can estimate the variance of rates using the mean rates over time (since the Fano factor, which is the ratio of variance to mean, of a Poisson process is 1). Particularly, we used

$$\mathrm{Fano[rate]} = \frac{\mathrm{Var[rate]}}{\mathrm{Mean[rate]}} = \frac{\mathrm{Var}[\frac{\text{spike count}}{T_\mathrm{win}}]}{\mathrm{Mean}[\frac{\text{spike count}}{T_\mathrm{win}}]} \tag{2.41}$$

$$= \frac{\frac{1}{{T_\mathrm{win}}^2}\mathrm{Var[spike\ count]}}{\frac{1}{T_\mathrm{win}}\mathrm{Mean[spike\ count]}} \tag{2.42}$$

$$= \frac{1}{T_\mathrm{win}}\mathrm{Fano[spike\ count]} \tag{2.43}$$

$$= \frac{1}{T_\mathrm{win}} \tag{2.44}$$

where T_win is the time window over which spikes are counted. In stationary dynamics, we can partition the time domain into sub–intervals of size T_win and count spikes in each sub–interval. In Eq. (2.41), the random variable 'spike count' refers to the number of spikes that is observed in each sub–interval, and the random variable 'rate' is the number of spikes normalized by the size of the sub–interval observed in each sub–interval.

Unlike the mean, the variance of the input (Eq. (2.37)) can now grow uncontrollably with N because weights are squared and hence cannot be used to cancel out large terms. However, if $J_E, J_I \sim \mathcal{O}(1/\sqrt{N})$, then the fluctuations in the input are order 1 and activity is irregular. In

general, J_E, J_I can be scaled as $N^{-\gamma}$ for $0 < \gamma \leq 1$. However, if $\gamma < 1/2$, then fluctuations grow very large with N and distributions of rates become narrow [138]. And, if $\gamma > 1/2$, then the variance of the input is zero, *i.e.*, there are no fluctuations in the input, and the network is in a frozen state where all neurons are quiescent [138].

2.3 Generating Correlated Spike Trains

As mentioned in Section 1.2, correlations in a recurrent network can be generated in a number of different ways [10, 27, 69, 84, 93, 95, 116, 123, 145]. In our framework, the source of correlations is an external population of correlated Poisson neurons called X (see Fig. 3.1.4 **A** in Section 3.1.2). Such external correlated input represents correlated synaptic input from other cortical networks or subcortical areas. We now explain how we simulated an external collection of correlated Poisson neurons.

In our simulations we generated spike trains, S_j^{x}, for the external population following [10, 81, 134]. Here, we review the Multiple Interaction Process (MIP) algorithm for generating correlated Poisson processes. To generate N_{x} processes, all having firing rate r_{x} and pairwise correlation c_{x}, we first generate a "mother" process with rate $r_m = r_{\mathrm{x}}/c_{\mathrm{x}}$. Then, to generate each of the N_{x} "daughter" processes, we first take all the spikes (events) in the mother process and then delete each spike independently with probability $1 - c_{\mathrm{x}}$. In other words, each event from the mother process is included in each daughter process independently with probability c_{x}. The rates of the daughter processes are therefore $r_m c_{\mathrm{x}} = r_{\mathrm{x}}$, as desired. Also, two daughter processes contain a proportion c_{x} of spike that occur at the same time (since the probability that a spike time in one process also appears in the other process is c_{x}). This algorithm produces spikes in several neurons that occur at precisely the same time. To make the spike trains less synchronous, but still correlated, we "jitter" all of the daughter spike times by adding i.i.d. random variables to each daughter spike times [135].

The cross spectral density (CSD) between two stationary processes X and Y is the Fourier

transform of their cross–covariance function:

$$\langle X, Y\rangle(f) := \int_{-\infty}^{\infty} c_{XY}(\tau) e^{-2\pi i f \tau} d\tau \tag{2.45}$$

where

$$c_{XY}(\tau) := \mathrm{cov}(X(t), Y(t+\tau)). \tag{2.46}$$

We also define the mean–field spectra,

$$\langle S_a, S_b\rangle = \mathrm{mean}_{j,k}[\langle S_j^a, S_k^b\rangle] \tag{2.47}$$

to be the average CSD between neurons in populations $a, b = \mathrm{e}, \mathrm{i}, \mathrm{x}$ in the network.

If the random variable (which is added to jitter spike times) has density $G(t)$ then the CSD is given by

$$\langle S_j^{\mathrm{x}}, S_k^{\mathrm{x}}\rangle(f) = \begin{cases} c r_{\mathrm{x}} |\widetilde{G}(f)|^2 & j \neq k \\ r_{\mathrm{x}} & j = k \end{cases} \tag{2.48}$$

where $\widetilde{G}(f)$ is the Fourier transform of $G(t)$.

In our simulations, we use Gaussian–distributed jitters with standard deviation $\tau_{\mathrm{jitter}} = 5$ ms, and we get

$$\langle S_j^{\mathrm{x}}, S_k^{\mathrm{x}}\rangle(f) = \begin{cases} c r_{\mathrm{x}} e^{-4f^2\pi^2\tau_{\mathrm{jitter}}^2} & j \neq k \\ r_{\mathrm{x}} & j = k \end{cases}. \tag{2.49}$$

2.4 Mean Field Theory of Covariances in Balanced Networks

In this section we closely follow the derivation from Baker et al. [10] for mean–field spike train covariances in asynchronous and correlated balanced networks. Similar expressions have been derived for similar models [54, 62, 112, 116, 132].

We define a 2×2 matrix of average cross–spectral densities between spike trains of each subpopulation:

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle = \begin{bmatrix} \langle S_{\mathrm{e}}, S_{\mathrm{e}} \rangle & \langle S_{\mathrm{e}}, S_{\mathrm{i}} \rangle \\ \langle S_{\mathrm{i}}, S_{\mathrm{e}} \rangle & \langle S_{\mathrm{i}}, S_{\mathrm{i}} \rangle \end{bmatrix}. \tag{2.50}$$

Since we will derive spike train covariances using CSDs, we define frequency–dependent connectivity matrices

$$W = \begin{bmatrix} w_{\mathrm{ee}} & w_{\mathrm{ei}} \\ w_{\mathrm{ie}} & w_{\mathrm{ii}} \end{bmatrix} \text{ and } W_x = \begin{bmatrix} w_{\mathrm{ex}} \\ w_{\mathrm{ix}} \end{bmatrix}, \tag{2.51}$$

where

$$w_{ab}(f) = p_{ab} j_{ab} q_b \widetilde{\alpha}_b(f) \sim \mathcal{O}(1) \tag{2.52}$$

and

$$\widetilde{\alpha}_b(f) = \frac{1}{1 + 2\pi i f \tau_b} \tag{2.53}$$

is the Fourier transform of $\alpha_b(t)$.

To better understand the relationship between CSD's and spike count covariances, note that for stationary processes, the covariance between integrals over large time windows is related to the zero–frequency CSD according to

$$\lim_{T_{\mathrm{win}} \to \infty} \frac{1}{T_{\mathrm{win}}} \mathrm{cov} \left(\int_0^{T_{\mathrm{win}}} X(t) dt, \int_0^{T_{\mathrm{win}}} Y(t) dt \right) = \langle X, Y \rangle(0). \tag{2.54}$$

so that for large T_{win},

$$\mathrm{cov} \left(\int_0^{T_{\mathrm{win}}} X(t) dt, \int_0^{T_{\mathrm{win}}} Y(t) dt \right) \approx T_{\mathrm{win}} \langle X, Y \rangle(0). \tag{2.55}$$

Note that a spike count over a window of size T_{win} is just an integral of spike trains over $[0, T_{\mathrm{win}}]$. We can write the zero–lag covariance between two stationary processes in terms of the CSD as

$$\mathrm{cov}(X(t), Y(t)) = \int_{-\infty}^{\infty} \langle X, Y \rangle(f) df. \tag{2.56}$$

This will be useful below in Section 2.6 when deriving an expression for the evolution of mean synaptic weights. More generally, the cross–covariance function, $c(\tau) = \mathrm{cov}(X(t), Y(t+\tau))$, is the inverse Fourier transform of $\langle X, Y \rangle(f)$.

Some other useful properties of the cross–spectral operator that we will use in derivations are:

$$\langle aX + bZ, Y \rangle = a\langle X, Y \rangle + b\langle Z, Y \rangle \tag{2.57}$$

for $a, b \in \mathbb{C}$

$$\langle X, Y \rangle = \langle Y, X \rangle^* \tag{2.58}$$

where z^* is the complex conjugate of z,

$$\langle K * X, Y \rangle = \widetilde{K} \langle X, Y \rangle \tag{2.59}$$

where $*$ denotes convolution, $K(t)$ is an L^2 kernel, and $\widetilde{K}(f)$ is the Fourier transform. The power spectral density of X is given by $\langle X, X \rangle \in \mathbb{R}$.

It is often useful to define the cross–spectral matrix between two multivariate processes. Specifically, suppose $\vec{X}(t) \in \mathbb{R}^m$ and $\vec{Y}(t) \in \mathbb{R}^n$ are multivariate, stationary processes then $\langle \vec{X}, \vec{Y} \rangle \in \mathbb{R}^{m \times n}$ with

$$[\langle \vec{X}, \vec{Y} \rangle]_{jk} = \langle \vec{X}_j, \vec{Y}_k \rangle. \tag{2.60}$$

This operator has essentially all the same properties elucidated above except that the complex conjugate turns into a conjugate–transpose.

Our goal in this section is to derive an expression for the mean–field matrix of spike train CSDs,

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle = \begin{bmatrix} \langle S_\mathrm{e}, S_\mathrm{e} \rangle & \langle S_\mathrm{e}, S_\mathrm{i} \rangle \\ \langle S_\mathrm{i}, S_\mathrm{e} \rangle & \langle S_\mathrm{i}, S_\mathrm{i} \rangle \end{bmatrix}. \tag{2.61}$$

We start noting that by Eq. (2.57), we can write the total input CSDs as

$$\langle \boldsymbol{T}, \boldsymbol{T} \rangle = \langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{R} \rangle + \langle \boldsymbol{R}, \boldsymbol{X} \rangle + \langle \boldsymbol{R}, \boldsymbol{R} \rangle \tag{2.62}$$

where $\langle \boldsymbol{X}, \boldsymbol{R} \rangle = \langle \boldsymbol{R}, \boldsymbol{X} \rangle^*$. We will derive expressions for each term on the right–hand side of Eq. (2.62). We start with $\langle \boldsymbol{X}, \boldsymbol{X} \rangle$. Consider the mean external input CSD between neurons in populations $a, b = \mathrm{e}, \mathrm{i}$

$$\begin{aligned}
\langle X_a, X_b \rangle &= \mathrm{mean}_{j,k} \langle X_j^a, X_k^b \rangle && (2.63)\\
&= \mathrm{mean}_{j,k} \left\langle \sum_{m=1}^{N_\mathrm{x}} J_{jm}^{a\mathrm{x}} \big(\alpha_\mathrm{x} * S_m^\mathrm{x}\big), \sum_{n=1}^{N_\mathrm{x}} J_{kn}^{b\mathrm{x}} \big(\alpha_\mathrm{x} * S_n^\mathrm{x}\big) \right\rangle && (2.64)\\
&= \mathrm{mean}_{j,k} \left\{ \sum_{m,n=1}^{N_\mathrm{x}} J_{jm}^{a\mathrm{x}} J_{kn}^{b\mathrm{x}} \tilde{\alpha}_\mathrm{x} \tilde{\alpha}_\mathrm{x}^* \langle S_m^\mathrm{x}, S_n^\mathrm{x} \rangle \right\} && (2.65)\\
&= \mathrm{mean}_{j,k} \left\{ |\tilde{\eta}_\mathrm{x}|^2 \langle S_\mathrm{x}, S_\mathrm{x} \rangle \sum_{m \neq n} J_{jm}^{a\mathrm{x}} J_{kn}^{b\mathrm{x}} + |\tilde{\eta}_\mathrm{x}|^2 r_\mathrm{x} \sum_{m=1}^{N_\mathrm{x}} J_{jm}^{a\mathrm{x}} J_{kn}^{b\mathrm{x}} \right\} && (2.66)\\
&= |\tilde{\eta}_\mathrm{x}|^2 j_{a\mathrm{x}} p_{a\mathrm{x}} j_{b\mathrm{x}} p_{b\mathrm{x}} q_\mathrm{x} \big[(N_\mathrm{x} - 1) \langle S_\mathrm{x}, S_\mathrm{x} \rangle + r_\mathrm{x} \big] && (2.67)
\end{aligned}$$

where we used bilinearity of the CSD operator (Eq. (2.57)), $\langle S_m^\mathrm{x}, S_n^\mathrm{x} \rangle = \langle S_\mathrm{x}, S_\mathrm{x} \rangle$ for $m \neq n$ and $\langle S_m^\mathrm{x}, S_n^\mathrm{x} \rangle = r_\mathrm{x}$ if $m = n$ (Eq. (2.49)), and Eq. (2.59). In matrix form, this yields

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle = N W_\mathrm{x} \langle S_\mathrm{x}, S_\mathrm{x} \rangle W_\mathrm{x}^* + q_\mathrm{x}^{-1} W_\mathrm{x} r_\mathrm{x} W_\mathrm{x}^* - q_\mathrm{x}^{-1} W_\mathrm{x} \langle S_\mathrm{x}, S_\mathrm{x} \rangle W_\mathrm{x}^*. \tag{2.68}$$

Note that $\langle S_\mathrm{x}, S_\mathrm{x} \rangle \sim \mathcal{O}(c_\mathrm{x})$, so $c_\mathrm{x} = 0$ in the asynchronous state and $c_\mathrm{x} \sim \mathcal{O}(1)$ in the correlated state. Thus, from Eq. (2.68) it follows that

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(c_\mathrm{x} N) + \mathcal{O}(1) \tag{2.69}$$

In particular, we see that in the asynchronous state, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(1)$. While in the correlated state, $\langle \boldsymbol{X}, \boldsymbol{X} \rangle \sim \mathcal{O}(N)$.

We continue with the $\langle \boldsymbol{R}, \boldsymbol{X} \rangle$ term. Recall that the recurrent input can be broken up into its excitatory and inhibitory contributions, so $\langle \boldsymbol{R}, \boldsymbol{X} \rangle = \langle \boldsymbol{E}, \boldsymbol{X} \rangle + \langle \boldsymbol{I}, \boldsymbol{X} \rangle$ (by Eq. (2.57)). Consider the external input CSD between excitatory and external inputs to neurons j and k in populations $a, b = \mathrm{e}, \mathrm{i}$, respectively,

$$\langle E_j^a, X_k^b \rangle = \left\langle \sum_{m=1}^{N_\mathrm{e}} J_{jm}^{a\mathrm{e}} \big(\alpha_\mathrm{e} * S_m^\mathrm{e}\big), X_k^b \right\rangle \tag{2.70}$$

$$= \tilde{\alpha}_\mathrm{e} \sum_{m=1}^{N_\mathrm{e}} J_{jm}^{a\mathrm{e}} \langle S_m^\mathrm{e}, X_k^b \rangle \tag{2.71}$$

$$\tag{2.72}$$

We take averages separately for $b = \mathrm{e}, \mathrm{i}$ to obtain

$$\langle E_a, X_\mathrm{e} \rangle = \frac{\tilde{\alpha}_\mathrm{e} j_{a\mathrm{e}} p_{a\mathrm{e}}}{\sqrt{N}} (N_\mathrm{e} - 1) \langle S_\mathrm{e}, X_\mathrm{e} \rangle + \{ S_\mathrm{e}, X_\mathrm{e} \} \tag{2.73}$$

$$= \sqrt{N} w_{a\mathrm{e}} \langle S_\mathrm{e}, X_\mathrm{e} \rangle + \frac{w_{a\mathrm{e}}}{q_\mathrm{e} \sqrt{N}} (\{ S_\mathrm{e}, X_\mathrm{e} \} - \langle S_\mathrm{e}, X_\mathrm{e} \rangle) \tag{2.74}$$

and

$$\langle E_a, X_\mathrm{i} \rangle = \frac{\tilde{\alpha}_\mathrm{e} j_{a\mathrm{e}} p_{a\mathrm{e}}}{\sqrt{N}} N_\mathrm{e} \langle S_\mathrm{e}, X_\mathrm{i} \rangle = \sqrt{N} w_{a\mathrm{e}} \langle S_\mathrm{e}, X_\mathrm{i} \rangle \tag{2.75}$$

At this step we have assumed that synaptic weights $J_{jm}^{a\mathrm{e}}$ and $\langle S_m^\mathrm{e}, X_k^b \rangle$ are uncorrelated. This assumption is used widely in balanced network theory, but it is sometimes not explicitly stated. We have made this assumption also in Section 2.1, Eqs. (2.9,2.16,2.18,2.23) and this eventually lead to the balance equation for firing rates (Eq. (2.26)). In Section 3.2, we will show a case in which this assumption does not hold.

Similar to Eq. (2.73), we obtain $\langle I_a, X_b \rangle$ for $b = \mathrm{e}, \mathrm{i}$

$$\langle I_a, X_{\rm e}\rangle = \sqrt{N} w_{a{\rm i}} \langle S_{\rm i}, X_{\rm e}\rangle \tag{2.76}$$

and

$$\langle I_a, X_{\rm i}\rangle = \sqrt{N} w_{a{\rm i}} \langle S_{\rm i}, X_{\rm i}\rangle + \frac{w_{a{\rm i}}}{q_{\rm i}\sqrt{N}} (\{S_{\rm i}, X_{\rm i}\} - \langle S_{\rm i}, X_{\rm i}\rangle) \tag{2.77}$$

Combining Eqs. (2.73–2.77) and writing in matrix form, we arrive at the following expression for the mean CSD between recurrent and external inputs

$$\langle \boldsymbol{R}, \boldsymbol{X}\rangle = \langle \boldsymbol{E}, \boldsymbol{X}\rangle + \langle \boldsymbol{I}, \boldsymbol{X}\rangle \tag{2.78}$$

$$= \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X}\rangle + \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{X}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{X}\rangle}{\sqrt{N}}\right). \tag{2.79}$$

where $Q = {\rm diag}(q_{\rm e}, q_{\rm i})$. The exact same derivation for $\langle \boldsymbol{R}, \boldsymbol{X}\rangle$ (replacing X by S) gives an expression for the mean–field CSD between spike trains and total external inputs

$$\langle \boldsymbol{T}, \boldsymbol{S}\rangle = \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{S}\rangle + \langle \boldsymbol{X}, \boldsymbol{S}\rangle + \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{S}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{S}\rangle}{\sqrt{N}}\right) \tag{2.80}$$

where

$$\{\boldsymbol{S}, \boldsymbol{S}\} = {\rm diag}\Big(\{S_{\rm e}, S_{\rm e}\}, \{S_{\rm i}, S_{\rm i}\}\Big) \tag{2.81}$$

and

$$\{S_a, S_a\} = \text{mean}_j \langle S_j^a, S_j^a \rangle. \tag{2.82}$$

Recall that we would like to arrive at an expression for the mean spike train CSDs, $\langle \boldsymbol{S}, \boldsymbol{S} \rangle$. To derive firing rates earlier in Section 2.1, we required $\overline{\boldsymbol{T}} \sim \boldsymbol{r}$, that is, $\mathcal{O}(1)$ transfer of mean input. Similarly, we now require that transfer of mean–field covariances is $\mathcal{O}(1)$. In particular, we assume

$$\lim_{N \to \infty} \frac{1}{\sqrt{N}} \frac{\langle T_a, U_b \rangle}{\langle S_a, U_b \rangle} = 0 \tag{2.83}$$

for $U = X, S$. Previous derivations of covariances throughout the literature require a linear transfer of covariances, a much stronger assumption, since we do not even need to know the actual value of $\frac{\langle T_a, U_b \rangle}{\langle S_a, U_b \rangle}$. Using Eq. (2.78) and the fact that $\langle \boldsymbol{T}, \boldsymbol{X} \rangle = \langle \boldsymbol{R}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle$, we have

$$\langle \boldsymbol{T}, \boldsymbol{X} \rangle = \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{X}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{X} \rangle}{\sqrt{N}} \right) + \langle \boldsymbol{X}, \boldsymbol{X} \rangle. \tag{2.84}$$

Assuming input and output CSD's are of the same order of magnitude (Eq. (2.83)),

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle = \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \frac{1}{\sqrt{N}} W Q^{-1} \{\boldsymbol{S}, \boldsymbol{X}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{X} \rangle}{\sqrt{N}} \right) + \langle \boldsymbol{X}, \boldsymbol{X} \rangle. \tag{2.85}$$

where we have divided both sides of Eq. (2.84) element–wise by $\sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle$ and applied Eq. (2.83). Note that Eq. (2.85) not only holds in the limit as N goes to infinity. In fact, for finite N, this result holds as long as

$$\left| \frac{\langle \boldsymbol{T}, \boldsymbol{X} \rangle}{W \langle \boldsymbol{S}, \boldsymbol{X} \rangle} \right| \ll \sqrt{N} \tag{2.86}$$

The order of the term $\{\boldsymbol{S}, \boldsymbol{X}\}$ with respect to N is not known. However, we use the Cauchy–Schwartz inequality to obtain

$$|\{\boldsymbol{S}, \boldsymbol{X}\}| \leq \sqrt{\{\boldsymbol{S}, \boldsymbol{S}\}\{\boldsymbol{X}, \boldsymbol{X}\}} \sim \mathcal{O}(\sqrt{\{\boldsymbol{X}, \boldsymbol{X}\}}) \tag{2.87}$$

since $\{\boldsymbol{S}, \boldsymbol{S}\} \sim \mathcal{O}(1)$. We compute an expression for $\{\boldsymbol{X}, \boldsymbol{X}\}$ by repeating the derivation that led to Eq. (2.68), but setting $j = k$,

$$\{\boldsymbol{X}, \boldsymbol{X}\} = N W_{\mathrm{x}} \langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle W_{\mathrm{x}}^{*} + q_{\mathrm{x}}^{-1} P_{\mathrm{x}}^{-1} W_{\mathrm{x}} r_{\mathrm{x}} W_{\mathrm{x}}^{*} - q_{\mathrm{x}}^{-1} W_{\mathrm{x}} \langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle W_{\mathrm{x}}^{*} \tag{2.88}$$

where

$$P_{\mathrm{x}} = \mathrm{diag}(p_{\mathrm{ex}}, p_{\mathrm{ix}}). \tag{2.89}$$

It is now clear that by Eq. (2.68), we must have

$$\{\boldsymbol{X}, \boldsymbol{X}\} \sim \langle \boldsymbol{X}, \boldsymbol{X} \rangle \tag{2.90}$$

be at least $\mathcal{O}(1)$. So by Eq. (2.87), $\{\boldsymbol{S}, \boldsymbol{X}\}$ is at most $\mathcal{O}(\langle \boldsymbol{X}, \boldsymbol{X} \rangle)$.
We then rewrite Eq. (2.85) as

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle \sim \sqrt{N} W \langle \boldsymbol{S}, \boldsymbol{X} \rangle + \langle \boldsymbol{X}, \boldsymbol{X} \rangle + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle + \langle \boldsymbol{S}, \boldsymbol{X} \rangle}{\sqrt{N}} \right). \tag{2.91}$$

In order to realize a self consistent solution, we require the right–hand side to cancel in the limit as $N \to \infty$, *i.e.*, we need $\langle \boldsymbol{S}, \boldsymbol{X} \rangle \sim \langle \boldsymbol{X}, \boldsymbol{X} \rangle / \sqrt{N}$. Thus, in the limit as $N \to \infty$, we have

$$\langle \boldsymbol{S}, \boldsymbol{X} \rangle = \frac{1}{\sqrt{N}} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}{N} \right). \tag{2.92}$$

We will use Eq. (2.92) later to arrive at our final expression for mean–field spike train CSDs, $\langle \boldsymbol{S}, \boldsymbol{S} \rangle$.

Starting with Eq. (2.80), we divide both sides element–wise by $\sqrt{N}W\langle \boldsymbol{S}, \boldsymbol{S}\rangle$ and apply Eq. (2.83),

$$\langle \boldsymbol{S}, \boldsymbol{S}\rangle \sim \sqrt{N}W\langle \boldsymbol{S}, \boldsymbol{S}\rangle + \langle \boldsymbol{X}, \boldsymbol{S}\rangle + \frac{1}{\sqrt{N}}WQ^{-1}\{\boldsymbol{S}, \boldsymbol{S}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{S}\rangle}{\sqrt{N}}\right). \tag{2.93}$$

We take the conjugate transpose of Eq. (2.92) and obtain

$$\langle \boldsymbol{X}, \boldsymbol{S}\rangle = \langle \boldsymbol{S}, \boldsymbol{X}\rangle^* = \frac{1}{\sqrt{N}}\langle \boldsymbol{X}, \boldsymbol{X}\rangle W^{-*} + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}{N}\right). \tag{2.94}$$

Substituting this expression into Eq. (2.93) gives

$$\langle \boldsymbol{S}, \boldsymbol{S}\rangle \sim \sqrt{N}W\langle \boldsymbol{S}, \boldsymbol{S}\rangle + \frac{1}{\sqrt{N}}\langle \boldsymbol{X}, \boldsymbol{X}\rangle W^{-*} + \frac{1}{\sqrt{N}}WQ^{-1}\{\boldsymbol{S}, \boldsymbol{S}\} \tag{2.95}$$

$$+ \mathcal{O}\left(\frac{\langle \boldsymbol{S}, \boldsymbol{S}\rangle}{\sqrt{N}}\right) + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}{N}\right). \tag{2.96}$$

Once again, to realize a self–consistent solution, we require a cancellation of the terms on the right–hand side in the limit as $N \to \infty$, so we need $\langle \boldsymbol{S}, \boldsymbol{S}\rangle \sim \mathcal{O}(\langle \boldsymbol{X}, \boldsymbol{X}\rangle/N + \{\boldsymbol{S}, \boldsymbol{S}\}/N)$. Hence, in the limit as $N \to \infty$,

$$\langle \boldsymbol{S}, \boldsymbol{S}\rangle = \frac{1}{N}W^{-1}\langle \boldsymbol{X}, \boldsymbol{X}\rangle W^{-*} - \frac{1}{N}Q^{-1}\{\boldsymbol{S}, \boldsymbol{S}\} + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X}\rangle}{N^{3/2}}\right). \tag{2.97}$$

The first term in Eq. (2.97) corresponds to externally generated covariability, and the second term arises from externally and internally generated covariability. In all comparisons of theory and simulations, we have ignored the second and third terms as they are negligible in the regimes studied here and analytically intractable. We do note that the second term can be estimated from numerical simulations and used in a semi–analytic approach to compute covariances, however this term only matters in extreme cases when $\langle \boldsymbol{X}, \boldsymbol{X}\rangle = 0$ [10].

Eq. (2.97) holds both in asynchronous and correlated states in the limit of large networks (limit as $N \to \infty$), however $\langle \boldsymbol{X}, \boldsymbol{X}\rangle$ has different expressions when the network is in the asynchronous state ($c_{\mathrm{x}} = 0$) versus in the correlated state ($0 < c_{\mathrm{x}} \leq 1$). We evaluate these next.

2.4.1 Mean Field Theory of Covariances in the Asynchronous State

In the asynchronous state, the balanced network is driven by an external layer of uncorrelated Poisson neurons, so $c_{\mathrm{x}} = 0$ and $\langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle = 0$. Substituting this into Eq. (2.88), we obtain

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle = q_{\mathrm{x}}^{-1} W_{\mathrm{x}} r_{\mathrm{x}} W_{\mathrm{x}}^{*} \sim \mathcal{O}(1) \tag{2.98}$$

so the mean–field spike train CSDs in the asynchronous state are given by

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle = \frac{1}{N} W^{-1} q_{\mathrm{x}}^{-1} W_{\mathrm{x}} r_{\mathrm{x}} W_{\mathrm{x}}^{*} W^{-*} \sim \mathcal{O}(1/N) \tag{2.99}$$

ignoring negligible terms (see discussion after Eq. (2.97) at the end of the previous subsection for more details).

When comparing balanced network theory to empirical simulations, it is often convenient to bring our variables back to the time domain. Recall that spike train covariances are proportional to the zero–frequency CSD. Therefore we can define the mean–field spike count covariance matrix as:

$$C = \begin{bmatrix} C_{\mathrm{ee}} & C_{\mathrm{ei}} \\ C_{\mathrm{ie}} & C_{\mathrm{ii}} \end{bmatrix} \tag{2.100}$$

where C_{ab} is the mean spike count covariance between neurons in populations $a = \mathrm{e}, \mathrm{i}$ and $b = \mathrm{e}, \mathrm{i}$ respectively, counted over time windows of size T_{win}. In all simulations and evaluating theoretical predictions, we set $T_{\mathrm{win}} = 250$ ms, however the theory is flexible, and we can use a range of window sizes. Evaluating the CSDs at $f = 0$ and multiplying by the window size, T_{win} we arrive at the final expression for mean spike count covariances in the asynchronous state,

$$C \approx \frac{1}{N} \frac{r_{\mathrm{x}}}{q_{\mathrm{x}}} T_{\mathrm{win}} \overline{W}^{-1} \overline{W}_{\mathrm{x}} \overline{W}_{\mathrm{x}}^{T} \overline{W}^{-T}. \tag{2.101}$$

Note that $C \sim \mathcal{O}(1/N)$. So as N grows, mean covariances become extremely weak in the asynchronous state.

2.4.2 Mean Field Theory of Covariances in the Correlated State

In the correlated state, the balanced neural network is driven by a layer of correlated Poisson inputs, and $\langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle \neq 0$. From Eq. (2.68), we can then see that

$$\langle \boldsymbol{X}, \boldsymbol{X} \rangle = N W_{\mathrm{x}} \langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle W_{\mathrm{x}}^* + q_{\mathrm{x}}^{-1} W_{\mathrm{x}} r_{\mathrm{x}} W_{\mathrm{x}}^* - q_{\mathrm{x}}^{-1} W_{\mathrm{x}} \langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle W_{\mathrm{x}}^* \sim \mathcal{O}(N). \tag{2.102}$$

This is known as the "pooling" of covariances: If N random variables have $\mathcal{O}(1)$ covariances, then their sum will have covariance order $\mathcal{O}(N)$. Substituting Eq. (2.102) into Eq. (2.97), and ignoring terms of order $1/N$ or smaller, we obtain

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle = W^{-1} W_{\mathrm{x}} \langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle W_{\mathrm{x}}^* W^{-*} \sim \mathcal{O}(1). \tag{2.103}$$

Recall from Section 2.3 that spike trains in the external population are correlated Poisson processes generated with Gaussian–distributed jittering, and hence

$$\langle S_{\mathrm{x}}, S_{\mathrm{x}} \rangle = c_{\mathrm{x}} r_{\mathrm{x}} e^{-4 f^2 \pi^2 \tau_{\mathrm{jitter}}^2}. \tag{2.104}$$

Finally, to transform the CSDs into spike train covariances, we evaluate the CSDs at $f = 0$ and multiply the right–hand side of Eq. (2.103) by T_{win}. This yields the final expression for spike train covariances in the correlated state:

$$C \approx c_{\mathrm{x}} r_{\mathrm{x}} T_{\mathrm{win}} \overline{W}^{-1} \overline{W}_{\mathrm{x}} \overline{W}_{\mathrm{x}}^T \overline{W}^{-T}. \tag{2.105}$$

Note that now, $C \sim \mathcal{O}(1)$ and the network is in a correlated state.

2.5 Synaptic Plasticity Rules

To model activity–dependent changes in synaptic weights we used eligibility traces to define the propensity of a synapse to change [43, 59, 67, 70, 76]. The eligibility trace, $x_j^a(t)$, of neuron j in

population a evolves according to

$$\tau_{\mathrm{STDP}} \frac{dx_j^a(t)}{dt} = -x_j^a(t) + \tau_{\mathrm{STDP}} S_j^a(t), \tag{2.106}$$

for $a = \mathrm{e}, \mathrm{i}$, where $S_j^a(t) = \sum_n \delta(t - t_n^{a,j})$ is the sequence of spikes of neuron, j. The eligibility trace, and the time constant, τ_{STDP} define a period following a spike in the pre– or post–synaptic cell during which a synapse can be modified by a spike in its counterpart.

Table 2.1: **Examples of STDP rules.** A number of different plasticity rules can be obtained as special cases of the general form given in Eq. (2.107).

STDP Rule	Coefficients	Equation
Classical EE Hebbian [14, 60, 92]	$B_{(\mathrm{e},j),(\mathrm{e},k)} = -1$ $B_{(\mathrm{e},k),(\mathrm{e},j)} = 1$	$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(x_k^{\mathrm{e}} S_j^{\mathrm{e}} - x_j^{\mathrm{e}} S_k^{\mathrm{e}}\big)$
Classical EE Anti–Hebbian [14, 92]	$B_{(\mathrm{e},j),(\mathrm{e},k)} = 1$ $B_{(\mathrm{e},k),(\mathrm{e},j)} = -1$	$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(- x_k^{\mathrm{e}} S_j^{\mathrm{e}} + x_j^{\mathrm{e}} S_k^{\mathrm{e}}\big)$
Weight–dependent EE Hebbian [14, 60, 92]	$B_{(\mathrm{e},j),(\mathrm{e},k)} = -J_{jk}^{\mathrm{ee}}$ $B_{(\mathrm{e},k),(\mathrm{e},j)} = J_{\max}$	$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(J_{\max} x_k^{\mathrm{e}} S_j^{\mathrm{e}} - J_{jk}^{\mathrm{ee}} x_j^{\mathrm{e}} S_k^{\mathrm{e}}\big)$
Homeostatic Inhibitory [141]	$A_{\mathrm{i},k} = \alpha_{\mathrm{e}} \frac{J_{jk}^{\mathrm{ei}}}{J_{\mathrm{norm}}}$ $B_{(\mathrm{e},j),(\mathrm{i},k)} = -\frac{J_{jk}^{\mathrm{ei}}}{J_{\mathrm{norm}}}$ $B_{(\mathrm{i},k),(\mathrm{e},j)} = -\frac{J_{jk}^{\mathrm{ei}}}{J_{\mathrm{norm}}}$	$\frac{dJ_{jk}^{\mathrm{ei}}}{dt} = -\eta_{\mathrm{ei}} \frac{J_{jk}^{\mathrm{ei}}}{J_{\mathrm{norm}}} \Big[(x_j^{\mathrm{e}} - \alpha_{\mathrm{e}}) S_k^{\mathrm{i}} + x_k^{\mathrm{i}} S_j^{\mathrm{e}}\Big]$
Oja's Rule [102]	$B_{(\mathrm{e},j),(\mathrm{e},j)} = -J_{jk}^{\mathrm{ee}}$ $B_{(\mathrm{e},j),(\mathrm{e},k)} = \beta$	$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(\beta x_j^{\mathrm{e}} S_k^{\mathrm{e}} - J_{jk}^{\mathrm{ee}} x_j^{\mathrm{e}} S_j^{\mathrm{e}}\big)$
Kohonen's Rule [78]	$A_{\mathrm{e},j} = -J_{jk}^{\mathrm{ee}}$ $B_{(\mathrm{e},j),(\mathrm{e},k)} = \beta$	$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(\beta x_j^{\mathrm{e}} S_k^{\mathrm{e}} - J_{jk}^{\mathrm{ee}} S_j^{\mathrm{e}}\big)$

Our theory of synaptic plasticity allows any synaptic weight to be subject to constant drift, changes due to pre– or post–synaptic activity only, and/or due to pairwise interactions in activity between the pre– and post–synaptic cells (zero, first, and second order terms, respectively, in Eq. (2.107)). The theory can be extended to account for other types of interactions. Each synaptic

weight therefore evolves according to a generalized STDP rule:

$$\frac{dJ_{jk}^{ab}}{dt} = \eta_{ab}\left(A_0 + \sum_{\alpha=\{a,j\},\{b,k\}} A_\alpha S_\alpha + \sum_{\alpha,\beta=\{a,j\},\{b,k\}} B_{\alpha,\beta} x_\alpha S_\beta\right) \tag{2.107}$$

where η_{ab} is the learning rate that defines the timescale of synaptic weight changes, $A_0, A_\alpha, B_{\alpha\beta}$ are functions of the synaptic weight, J_{jk}^{ab}, and $a, b = \mathrm{e}, \mathrm{i}$. For instance, the term $B_{(\mathrm{e},k),(\mathrm{i},j)} x_k^{\mathrm{e}} S_j^{\mathrm{i}}$ represents the contribution due to a spike in post–synaptic cell j in the inhibitory subpopulation, at the value x_k^{e} of the eligibility trace in the pre–synaptic cell k in the excitatory subpopulation. Higher order interactions are at the heart of triplet rules [15,49,108], and other types of interactions may also be important, *e.g.*, for calcium–based update rules [51,124]. For simplicity we here focus on pairwise interactions between spikes and eligibility traces, and leave extensions to more complex rules for future work.

This general formulation captures a range of classical plasticity rules as special examples: Table 2.1 shows that different choices of parameters yield Hebbian [14,60,92], anti–Hebbian, as well as Oja's [102], and other rules (see Fig. 2.5.1 for illustrations of the STDP function of each rule in Table 2.1). We only considered interactions up to second order. Here, we will not consider the BCM rule [15], and other rules [49,108] that depend on interactions beyond second order.

2.6 Dynamics of Mean Synaptic Weights in Correlated Balanced Networks

2.6.1 Accounting for the Effects of Correlations in the General STDP Rule

We now provide a detailed derivation of the mean synaptic weights accounting for the effects of firing rates and spike count covariances (Eqs. (2.26,2.105)).

We assumed that changes in synaptic weights occur on longer timescales than the dynamics of

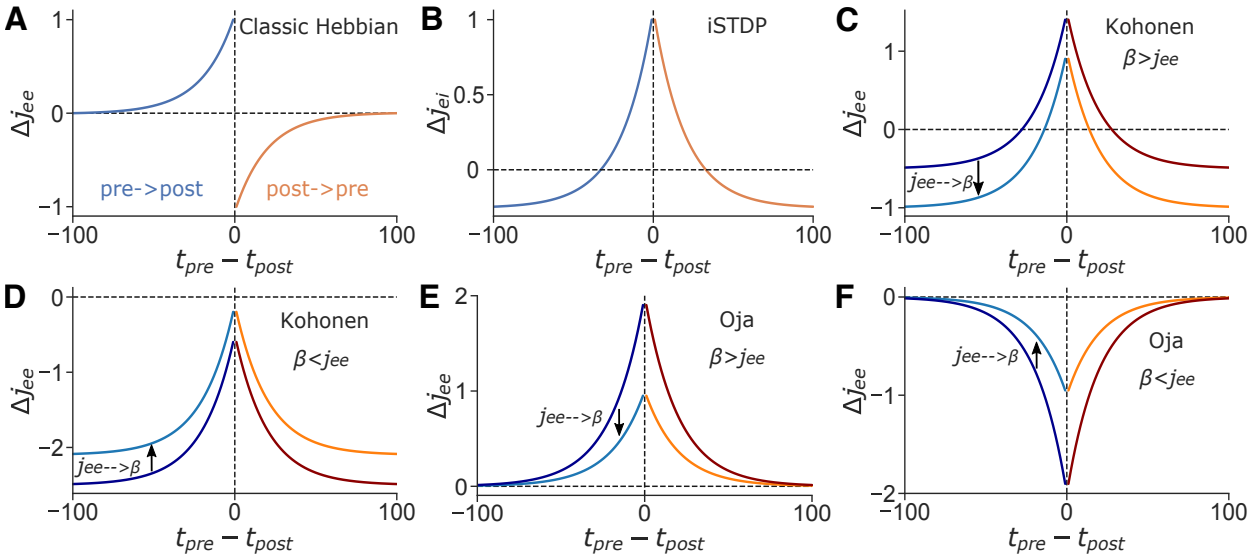

Figure 2.5.1: **STDP windows of different plasticity rules. A:** Change in synaptic weights as a function of the relative timing of pre– and post–synaptic spikes in Classical Hebbian STDP (same as weight–dependent Hebbian). **B:** Same as **A**, but for inhibitory STDP. **C:** Same as **A**, but for Kohonen's rule when weights are below parameter β. **D:** Same as **C**, but for the case when weights are above β. **E:** Same as **A**, but for Oja's rule when weights are below parameter β. **F:** Same as **E**, but for the case when weights are above β. Figure adapted from Akil et al. [5].

the eligibility trace and the correlation timescale [44–48, 74], *i.e.*,

$$1/\eta_{ab} \gg T_{\text{win}}, \tau_{\text{STDP}}. \tag{2.108}$$

This is a biologically sound assumption (see discussion in Chapter 5 for more details) [74]. Let ΔT be a time larger than the timescale of eligibility traces, τ_{STDP}, and T_{win}, but smaller than $1/\eta_{ab}$, so that the time differential of the weights satisfies [74]:

$$\frac{\Delta J_{jk}^{ab}}{\Delta T} = \frac{\eta_{ab}}{\Delta T} \int_0^{\Delta T} \Big[A_0 + \sum_{\alpha=\{a,j\},\{b,k\}} A_\alpha S_\alpha + \sum_{\alpha,\beta=\{a,j\},\{b,k\}} B_{\alpha,\beta} x_\alpha S_\beta \Big] dt. \tag{2.109}$$

The difference in timescales allows us to assume that the firing rates and covariances are in quasi–equilibrium. We used $1/\eta_{ab} = 10^5$ ms, and $\tau_{\text{STDP}} = 200$ ms, with correlation time window width $T_{\text{win}} = 250$ ms. Our derivations require $\tau_{\text{STDP}} \ll \Delta T \ll 1/\eta_{ab}$, however an exact numerical value

for ΔT is neither used nor needed. In Section 2.6.3, we address what happens when timescales are not separated.

We expand the sums and split the integral over the terms in the sums, to obtain

$$\begin{aligned}\frac{\Delta J_{jk}^{ab}}{\Delta T} = \frac{\eta_{ab}}{\Delta T}\Bigg(&\Delta T A_0 + A_{a,j}\int_0^{\Delta T} S_j^a dt + A_{b,k}\int_0^{\Delta T} S_k^b dt \\ &+ B_{\{a,j\},\{a,j\}}\int_0^{\Delta T} x_j^a S_j^a dt + B_{\{a,j\},\{b,k\}}\int_0^{\Delta T} x_j^a S_k^b dt \\ &+ B_{\{b,k\},\{a,j\}}\int_0^{\Delta T} x_k^b S_j^a dt + B_{\{b,k\},\{b,k\}}\int_0^{\Delta T} x_k^b S_k^b dt\Bigg).\end{aligned} \tag{2.110}$$

Each integral that is multiplied by $1/\Delta T$ in the previous equation is the expectation of the integrand. Thus we can write:

$$\begin{aligned}\frac{\Delta J_{jk}^{ab}}{\Delta T} = \eta_{ab}\Bigg(&A_0 + A_{a,j}\mathrm{E}[S_j^a] + A_{b,k}\mathrm{E}[S_k^b] + B_{\{a,j\},\{a,j\}}\mathrm{E}[x_j^a S_j^a] \\ &+ B_{\{a,j\},\{b,k\}}\mathrm{E}[x_j^a S_k^b] + B_{\{b,k\},\{a,j\}}\mathrm{E}[x_k^b S_j^a] + B_{\{b,k\},\{b,k\}}\mathrm{E}[x_k^b S_k^b]\Bigg)\end{aligned} \tag{2.111}$$

where $\mathrm{E}[\cdot]$ denotes expectation over time.

We provide a detailed derivation for the term $B_{\{a,j\},\{b,k\}}\mathrm{E}[x_j^a S_k^b]$. The other terms are derived in the same way. First note that

$$\mathrm{E}[x_j^a S_k^b] = \mathrm{E}[(K * S_j^a)(t) S_k^b(t)] = \mathrm{cov}(K * S_j^a, S_k^b) + \mathrm{E}[K * S_j^a]\mathrm{E}[S_k^b] \tag{2.112}$$

where $K(t) = e^{-t/\tau_{\mathrm{STDP}}} H(t)$ and $H(t)$ is the Heaviside function. The second term on the right hand side can be written in terms of the rates,

$$\mathrm{E}[K * S_j^a]\mathrm{E}[S_k^b] = \tau_{\mathrm{STDP}} r_j^a r_k^b \tag{2.113}$$

where r_j^a is the rate of neuron j in population a and we used the fact that $\int_{-\infty}^{\infty} K(t)dt = \tau_{\mathrm{STDP}}$.

We showed earlier in Section 2.4, Eq. (2.56) that we obtain,

$$\mathrm{cov}(K * S_j^a, S_k^b) = \int_{-\infty}^{\infty} \langle K * S_j^a, S_k^b \rangle(f) df. \tag{2.114}$$

This can be simplified by recalling that convolutions interact nicely with CSD's (Eq. (2.59)), and

$$\langle K * S_j^a, S_k^b \rangle = \widetilde{K} \langle S_j^a, S_k^b \rangle \tag{2.115}$$

where $\widetilde{K}(f)$ is the Fourier transform of the exponential kernel $K(t)$. This gives

$$\mathrm{cov}(K * S_j^a, S_k^b) = \int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_j^a, S_k^b \rangle(f) df. \tag{2.116}$$

Therefore,

$$\mathrm{E}[x_j^a S_k^b] = \int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_j^a, S_k^b \rangle(f) df + \tau_{\mathrm{STDP}} r_j^a r_k^b. \tag{2.117}$$

The first term on the right hand side of Eq. (2.117) depends on the spike count covariance between spike trains of populations a and b, while the second term depends on firing rates of populations a, b. Following the procedure demonstrated here and applying it to each term in Eq. (2.111), we obtain the following equation describing the evolution of weights,

$$\begin{aligned}
\frac{\Delta J_{jk}^{ab}}{\Delta T} = \eta_{ab} \Bigg(& A_0 + A_{a,j} r_j^a + A_{b,k} r_k^b \\
& + B_{\{a,j\},\{a,j\}} \left(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_j^a, S_j^a \rangle(f) df + \tau_{\mathrm{STDP}} r_j^a r_j^a \right) \\
& + B_{\{a,j\},\{b,k\}} \left(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_j^a, S_k^b \rangle(f) df + \tau_{\mathrm{STDP}} r_j^a r_k^b \right) \\
& + B_{\{b,k\},\{a,j\}} \left(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_k^b, S_j^a \rangle(f) df + \tau_{\mathrm{STDP}} r_k^b r_j^a \right) \\
& + B_{\{b,k\},\{b,k\}} \left(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_k^b, S_k^b \rangle(f) df + \tau_{\mathrm{STDP}} r_k^b r_k^b \right) \Bigg).
\end{aligned} \tag{2.118}$$

Averaging both sides of the equation over all neurons j and k in populations a and b, respectively, we obtained:

$$\begin{aligned}\frac{\Delta J_{ab}}{\Delta T} = \eta_{ab}\Bigg(& A_0 + A_{a,j} r_a + A_{b,k} r_b \\ & + B_{\{a,j\},\{a,j\}} \Bigg(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_a, S_a \rangle (f) df + \tau_{\mathrm{STDP}} r_a r_a \Bigg) \\ & + B_{\{a,j\},\{b,k\}} \Bigg(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_a, S_b \rangle (f) df + \tau_{\mathrm{STDP}} r_a r_b \Bigg) \\ & + B_{\{b,k\},\{a,j\}} \Bigg(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_b, S_a \rangle (f) df + \tau_{\mathrm{STDP}} r_b r_a \Bigg) \\ & + B_{\{b,k\},\{b,k\}} \Bigg(\int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_b, S_b \rangle (f) df + \tau_{\mathrm{STDP}} r_b r_b \Bigg) \Bigg).\end{aligned} \tag{2.119}$$

Lastly, we rearrange terms and arrive at a compressed expression dependent on rates and spike count covariances:

$$\frac{dJ_{ab}}{dt} = \eta_{ab}\Bigg(A_0 + \sum_{\alpha,\beta=\{a,b\}} \mathrm{Rate}_{\alpha,\beta} + \mathrm{Cov}_{\alpha,\beta} \Bigg), \tag{2.120}$$

where

$$\begin{aligned}\mathrm{Rate}_{\alpha,\beta} &= A_\alpha r_\alpha / 2 + B_{\alpha,\beta} \tau_{\mathrm{STDP}} r_\alpha r_\beta \\ \mathrm{Cov}_{\alpha,\beta} &= B_{\alpha,\beta} \int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_\alpha, S_\beta \rangle (f) df,\end{aligned} \tag{2.121}$$

where all the coefficients remained the same, but changed notation, *e.g.*, $B_{\{a,j\},\{b,k\}} = B_{a,b}$.

For example, weight–dependent Hebbian *EE* plasticity in Table 2.1 leads to the following mean–field equation,

$$\frac{dJ_{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}} \Bigg(J_{\max} - J_{\mathrm{ee}} \Bigg) \Bigg(\tau_{\mathrm{STDP}} r_{\mathrm{e}}^2 + \int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_{\mathrm{e}}, S_{\mathrm{e}} \rangle (f) df \Bigg). \tag{2.122}$$

2.6.2 Solving for the Dynamics of a Plastic Balanced Network

Eqs. (2.123,2.124,2.125) thus self–consistently describe the macroscopic dynamics of the balanced network.

$$\lim_{N\to\infty} \boldsymbol{r} = -\overline{W}^{-1}\overline{W}_{\mathrm{x}} r_{\mathrm{x}} \tag{2.123}$$

$$\langle \boldsymbol{S}, \boldsymbol{S} \rangle = \frac{1}{N} W^{-1} \langle \boldsymbol{X}, \boldsymbol{X} \rangle W^{-*} - \frac{1}{N} Q^{-1} \{ \boldsymbol{S}, \boldsymbol{S} \} + \mathcal{O}\left(\frac{\langle \boldsymbol{X}, \boldsymbol{X} \rangle}{N^{3/2}} \right). \tag{2.124}$$

$$\frac{dJ_{ab}}{dt} = \eta_{ab} \left(A_0 + \sum_{\alpha,\beta=\{a,b\}} \mathrm{Rate}_{\alpha,\beta} + \mathrm{Cov}_{\alpha,\beta} \right), \tag{2.125}$$

There are two approaches to analyzing this coupled system of ordinary differential equations: (1) solve directly for the steady–states of the system; or (2) apply numerical integration to obtain the evolution of the system in time. To obtain the equilibria, we first find the firing rates and covariances (both in terms of plastic weight J_{ab}) obtained using the mean–field description of the balanced network, Eqs. (2.123,2.124). We next substitute the results into Eq. (2.125), set $\frac{dJ_{ab}}{dt} = 0$, and find the roots. We denote the solution by J^*_{ab}. We then use the mean synaptic weight (root of Eq. (2.125), J^*_{ab}) to obtain the corresponding rates and covariances using Eqs. (2.123,2.124). Alternatively, we can solve the system iteratively over time and obtain the time evolution of rates, covariances, and weights. Starting at an arbitrary value of $J_{ab}(t)$, we proceed in the same way as in the first approach, but instead of setting $\frac{dJ_{ab}}{dt} = 0$, we use $J_{ab}(t)$ to compute the value of the derivative at time t, $\frac{dJ_{ab}}{dt}|_t$, and use it to update the mean weight at the next time step, $J_{ab}(t + \Delta t)$. We then use $J_{ab}(t + \Delta t)$ to update rates and covariances. We repeat this process until convergence (see Section 3.4 for sample trajectories under different rules, and for our criterion to determine stationarity of synaptic weights).

2.6.3 What Happens if Timescales Are Not Separated?

Our theory requires a separation of timescales between spiking dynamics and synaptic weight changes (Eq. (2.108)). We next asked how dynamics of balanced networks are affected when these timescales are not well separated. To address this question we considered three example STDP rules: weight–dependent Hebbian, Kohonen, and inhibitory plasticity (see Table 2.1 and Fig. 2.5.1). We simulated these networks with varying learning rate, and found that when synaptic weights updates occur on a fast timescale (compared to spiking dynamics), large fluctuations in weights and deviations from theory emerge (Fig. 2.6.1).

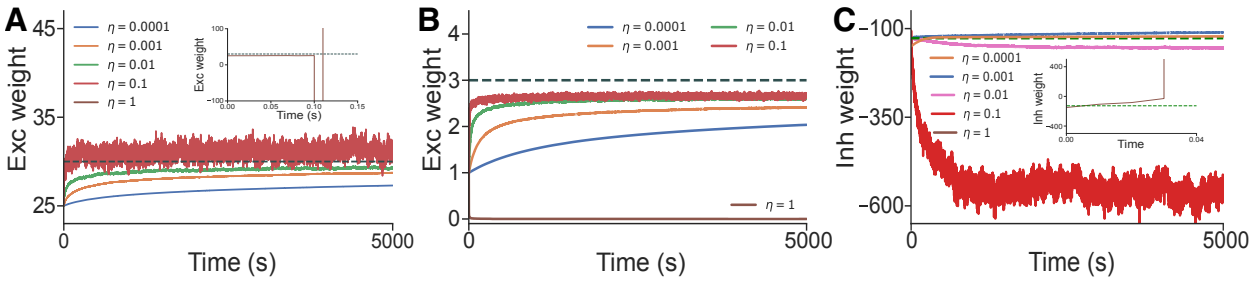

Figure 2.6.1: **Timescale of weight changes in plastic balanced networks.** **A:** Evolution of mean excitatory synaptic weight subject to weight–dependent Hebbian STDP in a balanced network. Fast weight timescales induce large fluctuations and deviations from the fixed point. Inset: When $\eta = 1$ weight updates are so large that the network becomes unstable. **B:** Same as **A:**, but for excitatory weights evolving according to Kohonen's rule. Convergence to the fixed point is very slow when $\eta < 0.01$, and the approximation breaks down for $\eta = 1$. **C:** Same as **A:**, but only inhibitory weights are plastic and follow inhibitory STDP updates. Deviations from the theory appear when timescales are not well separated ($\eta \geq 1/1000$). In all panels, dashed lines represent theory (Eqs. (2.26,2.101,3.4)), and solid lines represent numerical results. Relevant simulation parameters: $N = 5000$, $c_x = 0$. Other parameters as in Table 3.1. Figure adapted from Akil et al. [5].

2.7 Neuron Model

The novel mean–field theory of plastic balanced networks presented in Sections 2.1–2.6 does not depend on a particular model for dynamics of individual neurons. Complex models of neuron dynamics such as Hodgkin–Huxley [66] may be used, however they require the solution of four

differential equations with a large number of parameters that can result in many different spike trains patterns at a large computational cost. Such biological detail is not the focus of this work, so we instead use a reduced one–dimensional model for neuron's voltage dynamics called the current–based exponential integrate–and–fire (EIF) model [37], which can also recreate realistic spiking, with a much smaller number of parameters.

In the EIF model the membrane potential of neuron $j = 1, \ldots, N_a$ in population $a = \mathrm{e}, \mathrm{i}$ obeys exponential integrate–and–fire dynamics [37],

$$C_m \frac{dV_j^a}{dt} = -g_L(V_j^a - E_L) + g_L \Delta_T e^{(V_j^a - V_T)/\Delta_T} + I_j^a(t). \tag{2.126}$$

See Table 3.1 for a description and values of the parameters of the voltage dynamics. In our simulations, this equation was integrated using a Forward Euler method with step size $dt = 0.1$ ms.

Integrate–and–fire models can either current or conductance–based. In all of our work, we used the current–based formalism (Eq. (2.126)). However, in the latter, the membrane conductance is a function of time and the resulting neuron dynamics may be quite different to that of current–based models. In fact, there are separate mean–field theories for non–plastic networks of conductance–based synapses, and recently it has been shown that such networks require scaling synapses as $1/\log(N)$ [118], compared to the $1/\sqrt{N}$ scaling used here.

2.8 Perturbative Analysis

We next show how rates and spike count covariances are impacted by perturbations in synaptic weights. At steady state the average firing rates in a balanced network with mean–field connectivity matrix $\overline{W}_0$ are given by

$$\boldsymbol{r}_0 = -\overline{W}_0^{-1} \overline{W}_{\mathrm{x}} r_{\mathrm{x}} \tag{2.127}$$

We assume that the mean–field connectivity matrix is perturbed to $\overline{W}_{\mathrm{perturb}} = \overline{W}_0 + \Delta\overline{W}$. Using Neumann's approximation [75], $(I + H)^{-1} \approx (I - H)$, which holds for any square matrix H with

$\|H\| < 1$, and ignoring terms of power 2 and larger, we obtain,

$$\overline{W}^{-1}_{\text{perturb}} = (\overline{W}_0 + \Delta\overline{W})^{-1} = \big(\overline{W}_0(I + \overline{W}_0^{-1}\Delta\overline{W})\big)^{-1} \tag{2.128}$$

$$\approx \big(I - \overline{W}_0^{-1}\Delta\overline{W}\big)\overline{W}_0^{-1}, \tag{2.129}$$

where I is the identity matrix of appropriate size. We use this approximation of the perturbed weights to estimate the rates and spike count covariances using Eqs. (2.26,2.105). The 2×2 mean–field connectivity matrix, $\overline{W}_0$, must be non–singular for the balanced state to exist and for Neumann's approximation to hold [138]. While the non–singularity of $\overline{W}_0$ is a non-restrictive condition for two neural populations, $\overline{W}_0$ can become singular in some models with several neural sub-populations [33, 83].

2.9 Comparison of Theory with Numerical Experiments

We define spike trains of individual neurons in the population as sums of Dirac delta functions, $S_i(t) = \sum_j \delta(t - t_{ij})$, where the t_{ij} is the time of the j^{th} spike of neuron i. Assuming the system has reached equilibrium, we partition the interval over which activity has been measured into K equal subintervals, and define the spike count covariance between two cells as,

$$\text{cov}(n_{1k}, n_{2k}) = \sum_k (n_1^k - \overline{n}_1)(n_2^k - \overline{n}_2), \tag{2.130}$$

where n_{ik} is the spike count of neuron i in subinterval, or time window, k, and $\overline{n}_i = \frac{1}{K}\sum_k n_{ik}$ is the average spike count over all subintervals. In simulations we used subintervals of size $T_{\text{win}} = 250$ ms, although the theory applies to sufficiently long subintervals, and can be extended to shorter intervals as well. The spike count covariance thus captures shared fluctuations in firing rates between the two neurons [31].

2.10 Remarks on the General STDP Rule

We would like to know what the impact of rates and covariances is in the synaptic weights. We estimated the magnitude of the rate and covariance terms in Eq. (2.120) in Figs. 2.11.1 and 3.1.6. The remaining question is: Given the magnitude of each term, what is the qualitative effect that increasing rates or covariances will have on synaptic weights? Here, we show that changes in rates or covariances have two distinct effects. Such changes can shift the location of the fixed point, or they can modulate the speed of convergence to that equilibrium. We prove this in *Claims* 1 and 2 next.

Claim 1. Consider a general STDP rule that satisfies Eq. (2.107) and involves synapses between neurons of different populations ($a \neq b$). In this case, both the rates and covariances determine the location of the fixed point of the synaptic weights. In particular, increasing firing rates or spike count covariances can shift the location of the fixed point of the mean synaptic weight.

Proof. For simplicity, assume the STDP only involves second order terms (only these give rise to covariance terms in the mean–field equation of the weights),

$$\frac{dJ_{jk}^{ab}}{dt} = \eta_{ab}\Big(B_{\{a,j\},\{a,j\}} x_j^a S_j^a + B_{\{a,j\},\{b,k\}} x_j^a S_k^b + B_{\{b,k\},\{a,j\}} x_k^b S_j^a + B_{\{b,k\},\{b,k\}} x_k^b S_k^b \Big). \tag{2.131}$$

As shown before, the mean synaptic weight evolves according to:

$$\begin{aligned} \frac{dJ_{ab}}{dt} = \eta_{ab}\Big[& B_{\{a,j\},\{a,j\}}\Big(\int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_a, S_a\rangle(f)df + \tau_{\text{STDP}} r_a r_a \Big) \\ & + B_{\{a,j\},\{b,k\}}\Big(\int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_a, S_b\rangle(f)df + \tau_{\text{STDP}} r_a r_b \Big) \\ & + B_{\{b,k\},\{a,j\}}\Big(\int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_b, S_a\rangle(f)df + \tau_{\text{STDP}} r_b r_a \Big) \\ & + B_{\{b,k\},\{b,k\}}\Big(\int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_b, S_b\rangle(f)df + \tau_{\text{STDP}} r_b r_b \Big)\Big] \end{aligned} \tag{2.132}$$

By assumption, $a \neq b$, and re–grouping terms, we get:

$$\begin{aligned}\frac{dJ_{ab}}{dt} = \eta_{ab}\Big(& B_{\{a,j\},\{a,j\}}\tau_{\text{STDP}} r_a r_a + (B_{\{a,j\},\{b,k\}} + B_{\{b,k\},\{a,j\}})\tau_{\text{STDP}} r_a r_b \\ & + B_{\{b,k\},\{b,k\}}\tau_{\text{STDP}} r_b r_b + B_{\{a,j\},\{a,j\}} \int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_a, S_a\rangle(f) df \\ & + (B_{\{a,j\},\{b,k\}} + B_{\{b,k\},\{a,j\}}) \int_{-\infty}^{\infty} \widetilde{K}(f)\text{Re}[\langle S_a, S_b\rangle(f)] df \\ & + B_{\{b,k\},\{b,k\}} \int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_b, S_b\rangle(f) df \Big).\end{aligned} \tag{2.133}$$

Grouping the first three terms and the last three terms we obtained:

$$\frac{dJ_{ab}}{dt} = \eta_{ab}\big(\text{Rate}_{a,b} + \text{Cov}_{a,b}\big). \tag{2.134}$$

Thus rates and spike count covariances have an additive effect in the equation for dJ_{ab}/dt. This implies that both rates and covariances determine the location of the fixed point of J_{ab}. Thus, increases in any of the two terms can lead to a shift in the location of the fixed point. In our framework, rates and covariances can be controlled independently through the r_{x} and c_{x}, respectively. Therefore, increasing c_{x} can put the network in regimes where $\text{Rate}_{a,b}$ and $\text{Cov}_{a,b}$ are of the same magnitude and hence increasing covariances can moderately shift the location of the fixed point J_{ab}. (This is the case for Kohonen's rule – see Fig. 3.1.6 in Section 3.1.) □

Claim 2. Let us now consider a general STDP rule that satisfies Eq. (2.107) and changes the weights of synapses between neurons of the same population ($a = b$), but these changes are only due to order 2 interactions. In this case, the rates and covariances do not determine the fixed point of the synaptic weights. However, rates and covariances can modulate the speed of convergence of the synaptic weights to steady state.

Proof. By assumption, the STDP only involves second order terms. Then:

$$\frac{dJ^{ab}_{jk}}{dt} = \eta_{ab}\Big(B_{\{a,j\},\{a,j\}} x^a_j S^a_j + B_{\{a,j\},\{b,k\}} x^a_j S^b_k + B_{\{b,k\},\{a,j\}} x^b_k S^a_j + B_{\{b,k\},\{b,k\}} x^b_k S^b_k \Big). \tag{2.135}$$

As shown before, the mean synaptic weight evolves according to Eq. (2.132). Since, $a = b$, and re–grouping terms, we get:

$$\frac{dJ_{aa}}{dt} = \eta_{ab}\Big(B_{\{a,j\},\{a,j\}} + B_{\{a,j\},\{b,k\}} + B_{\{b,k\},\{a,j\}} + B_{\{b,k\},\{b,k\}}\Big)\times \Big(\tau_{\text{STDP}} r_a^2 + \int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_a, S_a\rangle(f) df\Big). \tag{2.136}$$

Note that

$$\tau_{\text{STDP}} r_a^2 + \int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_a, S_a\rangle(f) df > 0 \tag{2.137}$$

therefore, the only fixed point of J_{aa} is determined by the roots of

$$B_{\{a,j\},\{a,j\}} + B_{\{a,j\},\{b,k\}} + B_{\{b,k\},\{a,j\}} + B_{\{b,k\},\{b,k\}} = 0. \tag{2.138}$$

Thus the location of the fixed point of J_{ab} is independent of rates and spike count covariances. However, since rates and covariances come in multiplicatively in the derivative of J_{ab} (*i.e.*, the derivative is proportional to the sum of rate and covariance contributions), rates and covariances can modulate the speed of convergence of the synaptic weights to the fixed point. This can happen through changes in r_{x} or c_{x}. This is the case for weight–dependent Hebbian STDP – see Fig. 3.1.3 in Section 3.1 for more details. □

2.11 Which Cause More Weight Changes: Rates or Covariances?

In Chapter 3, we use our theory to explore the interaction between weights, rates, and covariances for networks with plastic synapses that followed Kohonen's rule and weight–dependent EE Hebbian STDP. We found that increasing correlations mildly impact the dynamics of the synaptic weights. We now compare the magnitude of the covariance term to that of the rate term in Eq. (2.120) in order to determine which term will, generally, drive larger changes in synaptic weights.

To do this, we estimate the covariance terms in our equation for the mean weights (Eq. (2.120)) and compare them to the rate terms in that same equation. We assume there are interactions of

order 2 only, and that these have constant coefficients. In other words, we assume that the synaptic weights evolve according to,

$$\frac{dJ_{jk}^{ab}}{dt} = \eta_{ab}\big(x_j^a S_j^a + x_j^a S_k^b + x_k^b S_j^a + x_k^b S_k^b\big), \tag{2.139}$$

where, without loss of generality, we set all nonzero coefficients to one. We do this in order to give the same coefficient/weight to each term so that we can compare rates and covariance terms without giving more weight to one or the other. We then obtain an equation for the mean synaptic weights depending on the rates and covariances:

$$\frac{dJ_{ab}}{dt} = \eta_{ab} \sum_{\alpha,\beta=\{a,b\}} \mathrm{Rate}_{\alpha,\beta} + \mathrm{Cov}_{\alpha,\beta}, \tag{2.140}$$

where $\mathrm{Rate}_{\alpha,\beta} = \tau_{STDP} r_\alpha r_\beta$, $\mathrm{Cov}_{\alpha,\beta} = \int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_\alpha, S_\beta\rangle(f)df$. The fact that all nonzero coefficients are set to be equal allows us to estimate the raw values of the covariances and rates for each synapses (for $a, b = \mathrm{e}, \mathrm{i}$), and compare them.

We found that for a plasticity rule (Eq. (2.140)), which is an example of our general STDP rule Eq. (2.107), acting on *any* synapse, the impact of correlations in the synaptic weights is at least one order of magnitude smaller than the contribution of the rates to the dynamics of the weights (Fig. 2.11.1 **A**,**B**,**C**). This implies that changes in the firing rates will usually have a stronger effect on the weights than changes in spike count covariances. However, increasing spike count covariances can still have a mild impact in synaptic weight dynamics when rates remain fixed. This last point is illustrated in Chapter 3 in example networks undergoing excitatory plasticity (Kohonen and weight–dependent Hebbian).

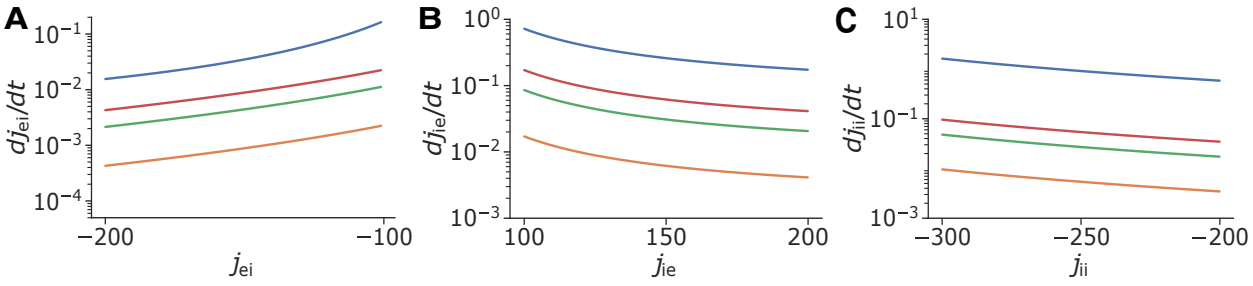

Figure 2.11.1: **General impact of correlations in synaptic weights. A:** Plot of the time derivative of the mean weight as a function of j_{ee} assuming interactions of order 2 only. **B** , **C**, **D:** Same as **A:** but for EI, IE, and II synapses, respectively. The contribution of rates to the evolution of synaptic weights is always larger than that of the covariances. Solid lines represent the theory (Eqs. (2.26,2.105,2.140)). Figure adapted from Akil et al. [5].

2.12 Implementation of Inhibitory Plasticity in Numerical Simulations

We next describe how we implemented the plasticity rules using the example of homeostatic inhibitory–to–excitatory STDP as in [141], which is a special case of our general STDP rule (see Eq. (2.107) and Table 2.1). Note that for the implementation of this rule, we will use the convention that $J^{\text{ei}}_{jk} < 0$ or inhibitory weights. After each presynaptic spike, $t^{\text{i}}_{k,n}$, of inhibitory neuron k, we make the update:

$$J^{\text{ei}}_{jk}(t+\Delta t) = J^{\text{ei}}_{jk}(t) - \eta_{\text{ei}} \frac{J^{\text{ei}}_{jk}(t)}{J_{\text{norm}}} (x^{\text{e}}_j(t) - \alpha_{\text{e}}) \tag{2.141}$$

and after each postsynaptic spike, $t^{\text{e}}_{j,m}$, of excitatory neuron j, we make the update:

$$J^{\text{ei}}_{jk}(t+\Delta t) = J^{\text{ei}}_{jk}(t) - \eta_{\text{ei}} \frac{J^{\text{ei}}_{jk}(t)}{J_{\text{norm}}} x^{\text{i}}_k(t) \tag{2.142}$$

where $J_{\text{norm}} \sim \mathcal{O}(1/\sqrt{N})$ is a normalization constant. Note that $J_{\text{norm}} < 0$, in order to yield a positive fraction in the previous two equations. Other STDP rules were implemented similarly.

We slightly modified the rule described in [141] to prevent I to E weights from becoming positive. In particular, we multiplied the right hand side of the updates by $\frac{J^{\text{ei}}_{jk}(t)}{J_{\text{norm}}}$, in order to create an unstable zero fixed point and ensure that weights remain negative. See Section 3.2 for an

example where inhibitory weights change signs if the zero fixed point is not present.

This modification guarantees that *EI* weights will remain negative in continuous–time rate–dynamics. However, *EI* weights could still change sign, since in network simulations, changes in synaptic weights occur in spike–based discrete–time updates. We thus proceed to derive a condition under which this modification ensures that synaptic weights will not change signs. Consider updates due to post–synaptic spikes, since pre–synaptic spikes always strengthen the connection. We would like for the weights to always be negative:

$$J_{jk}^{ei} = J_{jk}^{ei} - \frac{J_{jk}^{ei}}{J_{\text{norm}}} \eta_{\text{ei}} (x_j^e - \alpha_{\text{e}}) < 0. \tag{2.143}$$

Dividing through by J_{jk}^{ei} and rearranging,

$$-\frac{1}{J_{\text{norm}}} \eta_{\text{ei}} x_j^e + \alpha_{\text{e}} \eta_{\text{ei}} \frac{1}{J_{\text{norm}}} > -1. \tag{2.144}$$

Since $J_{\text{norm}} < 0$, it suffices to have:

$$\alpha_{\text{e}} \eta_{\text{ei}} \frac{1}{|J_{\text{norm}}|} < 1 \tag{2.145}$$

Therefore, as long as this condition, which depends on network and plasticity parameters, is satisfied, I to E weights will not turn positive.

Chapter 3

Numerical Simulations of Balanced Networks

In the following sections, we use the theory described in the Chapter 2 to show how synaptic weights and correlated activity co–evolve in tightly balanced networks. We review the issue of unstable weights under Classical Hebbian STDP [14, 60, 92] (Table 2.1), and show how it can be resolved in a biologically realistic way arriving at the weight–dependent Hebbian STDP rule. We then apply our theory to a balanced network with plastic excitatory synapses that evolve according to such weight–dependent Hebbian STDP, and investigate how correlations, rates, and synaptic weights interact.

We next examine balanced networks under Kohonen's STDP rule [78]. The theory predicts that the network is in a stable, balanced state, and provides asymptotic values for rates, spike count covariances, and synaptic weights. We also show that correlated activity can have a moderate impact on the dynamics of synaptic weights.

We also show that the theory does not accurately predict weight dynamics for certain STDP rules for which heterogeneity in rates can lead to competition between synaptic weights. The resulting correlations between rates and weights are inconsistent with classical balanced network theory. We show how our framework can be extended to describe such networks. We then demonstrate

that our theory can describe the robust response of iSTDP to targeted optogenetic input to subsets of neurons.

Lastly, we show that the statistics and dynamics of plastic, balanced networks simulated here tend to be similar to those in real cortical circuits. We also demonstrate that our theory can be used to describe transient dynamics as long as timescales are well separated.

3.1 Balanced Networks under Excitatory Plasticity

3.1.1 Hebbian Plasticity

Excitatory plasticity plays a central role in theories of learning, but can lead to instabilities [9, 14, 92, 100]. Our theory can play a fundamental role in stabilizing network dynamics under Hebbian STDP.

Stabilizing Excitatory Hebbian STDP. Consider a balanced network in an asynchronous state as described in Chapter 2. Assume that the EE weights change according to:

$$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(x_k^{\mathrm{e}} S_j^{\mathrm{e}} - x_j^{\mathrm{e}} S_k^{\mathrm{e}}\big). \tag{3.1}$$

We have shown that the mean EE synaptic weight then evolves as:

$$\frac{dJ_{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\bigg(1-1\bigg)\bigg(\tau_{\mathrm{STDP}} r_{\mathrm{e}}^2 + \int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_{\mathrm{e}}, S_{\mathrm{e}}\rangle(f) df\bigg) = 0. \tag{3.2}$$

Mean–field theory predicts stable dynamics when the system has a stable fixed point determined by $dJ_{ee}/dt = 0$. However, simulations show that network dynamics are unstable due to the divergence of individual weights to $\pm\infty$ (Fig. 3.1.1 **A–D**), so that the mean stays finite while the variance across the population diverges. This is a well known fact in the literature [9, 100]. Therefore, it is possible that the mean–field theory predicts stable dynamics, despite instabilities in higher moments that can cause individual weights to blow up. On the other hand, network dynamics will be unstable if the mean-field theory predicts that the first moment diverges.

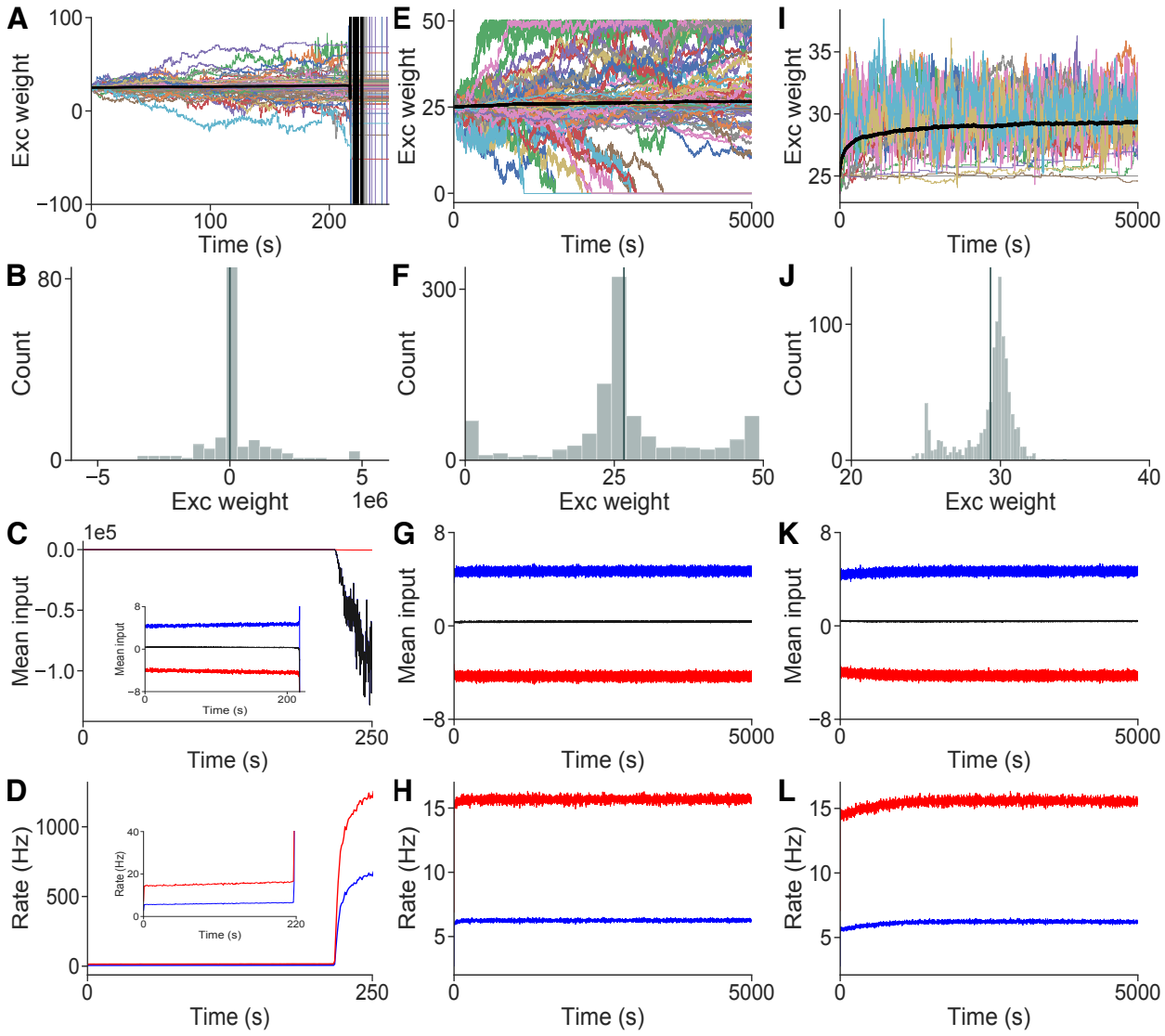

Figure 3.1.1: **Stabilization of Hebbian STDP. A:** Excitatory synaptic weights subject to classical Hebbian STDP (see Table 2.1). The mean synaptic weight is shown in black. Individual weights diverge to $\pm\infty$, destabilizing network dynamics. **B:** Distribution of synaptic weights right before the simulation is terminated. **C:** The mean synaptic input diverges as individual weights diverge. **D:** Positive feedback loops due to unconstrained Hebbian STDP lead to uncontrolled activity in the network. The simulation is terminated when input and rates grow with no bounds. **E:** Same as **A–D**, but for Classical Hebbian plasticity with hard constraints at 0 and 50. Mean synaptic weight (in black) remains unchanged. Network remains balanced and stable throughout the simulation. **F:** Same as **A–D**, but for weight dependent Hebbian STDP (see Table 2.1). The network maintains a stable, balanced state. Relevant simulation parameters: $N = 5000$, $c_x = 0$. Others as in Table 3.1. Figure adapted from Akil et al. [5].

In simulations, Hebbian STDP is frequently stabilized by imposing hard lower and upper bounds on synaptic weights (Fig. 3.1.1 **E,F**). With appropriately chosen bounds, the network is stabilized and balance is achieved (Fig. 3.1.1 **G,H**). We can use mean–field theory to find alternative ways to modify learning rules in order to stabilize network dynamics: One such modification is to multiply the LTP term in Eq. (3.1) by a constant ($J_{\max} \sim \mathcal{O}(J^{ee}_{jk})$) and the LTD term by the current value of the synaptic weight, J^{ee}_{jk}, in what we call the "Weight–dependent Hebbian STDP rule" (see Table 2.1). This can be interpreted as a constraint on synaptic weights due to limitations on the size and strength of a synapse. Excitatory synaptic weights then converge to a fixed point and the network remains in the balanced regime (Fig. 3.1.1 **I–L**). This approach can be generalized to other plasticity rules while maintaining the underlying dependence on spike timing.

Asymptotic behavior in weight–dependent Hebbian STDP. Assume that EE synaptic weights evolve according to a weight–dependent Hebbian STDP rule [14, 60, 92] (see Table 2.1),

$$\frac{dJ^{\mathrm{ee}}_{jk}}{dt} = \eta_{\mathrm{ee}}\left(J_{\max} x^{\mathrm{e}}_k S^{\mathrm{e}}_j - J^{\mathrm{ee}}_{jk} x^{\mathrm{e}}_j S^{\mathrm{e}}_k \right) \tag{3.3}$$

This rule yields the classical behavior originally proposed by Hebb: a pre–synaptic spike followed by a post–synaptic spike potentiates the synapse. This equation also enforces that a post–synaptic spike followed by a pre–synaptic spike depresses the synapse.

On average, the EE synaptic weights evolve as:

$$\frac{dJ_{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\left(J_{\max} - J_{\mathrm{ee}} \right)\left(\tau_{\mathrm{STDP}} r^2_{\mathrm{e}} + \int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_{\mathrm{e}}, S_{\mathrm{e}} \rangle (f) df \right). \tag{3.4}$$

The fixed point of Eq. (3.4) is simply $J^*_{\mathrm{ee}} = J_{\max}$ because rate and covariances equations (Eqs. (2.26,2.105)) show that the second factor on the right–hand side is strictly positive. We can still solve for the firing rates and spike count covariances through their mean–field equations using the fixed point $J^*_{\mathrm{ee}} = J_{\max}$.

Our theoretical framework predicts that the network attains a stable balanced state, and gives the location of the fixed point in terms of weights, rates, and covariances. We confirm our predictions with numerical simulations of the network in a correlated state (Fig. 3.1.2).

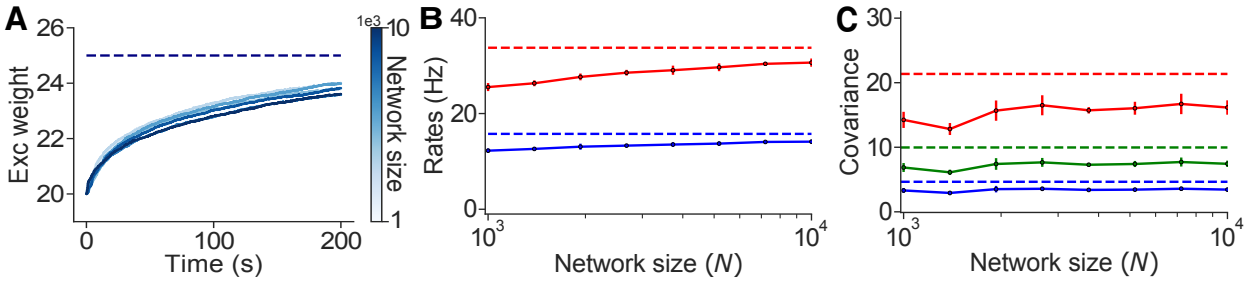

Figure 3.1.2: **Weight–dependent Hebbian STDP yields a stable balanced state. A:** Mean excitatory synaptic weight, j_{ee}, evolving over time for increasing network sizes. **B:** Mean E and I firing rates for increasing network sizes. **C:** Mean spike count covariance between E–E, E–I, and I–I spike trains. Dashed lines represent the theory (Eqs. (2.26,2.105,3.4)). Solid lines and dots are obtained from numerical simulations. Figure adapted from Akil et al. [5].

Our theory (Eqs. (2.26,2.105,3.4)) predicted that spike count covariances are of order 1, and this is confirmed by numerical simulations (Fig. 3.1.2 **C**). As N grows, firing rates and synaptic weights converge to the predicted values (Fig. 3.1.2 **A**,**B**).

We have thus shown, that for N large enough, the dynamics of a plastic balanced network can be well approximated by our theoretical results for the introduced weight–dependent Hebbian STDP rule, as long as the balance condition is satisfied (Eq. (2.26)).

Impact of correlations on weights under weight–dependent Hebbian STDP. In Section 2.10, we showed mathematically that in STDP rules of only order 2 interactions between neurons in the same subpopulation, correlations and rates modulate the speed of convergence of synaptic weights to equilibrium. We now show this is the case for the weight–dependent Hebbian STDP rule.

In particular, we used our theoretical framework to show that in a balanced network with EE weights that change according to Eq. (3.3), increasing correlations does not affect the location of the fixed point, but increases the speed of convergence to that equilibrium.

Simulations confirmed that the fixed point of synaptic weights does not change with increasing

correlations (Eq. (3.4)) (Fig. 3.1.3 **A**). Since synaptic weights are not affected by these changes in correlations, rates also remain unchanged (Fig. 3.1.3 **B**). This is also confirmed by empirical simulations. To measure how the speed of convergence is increased by stronger correlations, we compute t_{50} which is defined as the time it takes for the mean synaptic weight to reach 50% convergence to the fixed point for increasing values of input correlations, and found that t_{50} is dramatically reduced as correlations become larger, hence the mean weights converge faster to their equilibrium value (Fig. 3.1.3 **C**).

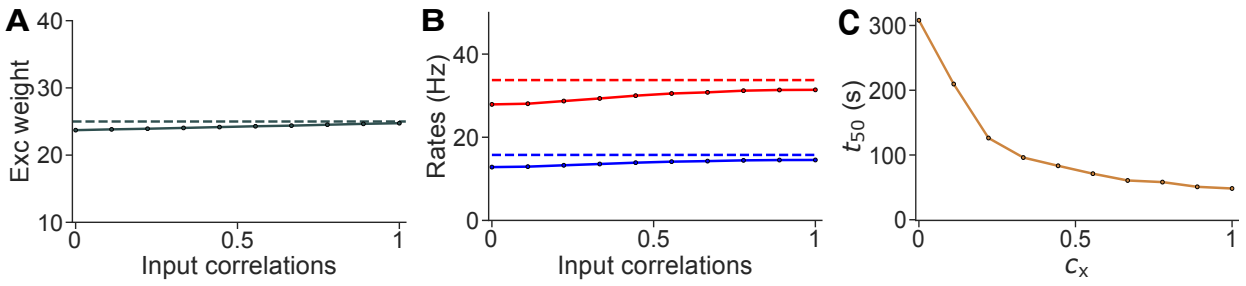

Figure 3.1.3: **Increasing input correlations speeds up convergence to equilibrium under Hebbian STDP. A:** Mean excitatory synaptic weight, j_{ee}, obtained at different values of input correlations. **B:** Mean E and I firing rates for increasing input correlations. **C:** Half time between initial condition and equilibrium for networks with increasing input correlations. Dashed lines represent the theory (Eq. (2.26,2.105,3.4)). Solid lines and dots are obtained from numerical simulations. Figure adapted from Akil et al. [5].

We found that, as predicted in Section 2.10, the effect of correlations in the weight dynamics is different to that seen with Kohonen's rule (shown in Section 3.1.2). In this case, our theory predicts that increasing covariances (through our parameter c_x), increases the rate of convergence to equilibrium, while maintaining the location of the fixed point unchanged.

3.1.2 Kohonen's Plasticity Rule

We have illustrated the power of our theory through examples of Hebbian STDP. In particular, our theory allows us to modify STDP rules in biologically plausible ways to achieve stability, to predict the network and weight dynamics, and to better understand the role of spike train correlations in

the evolution of synaptic weights.

For the remainder of Section 3.1.2, we consider a network in a correlated state with excitatory–to–excitatory (EE) weights that evolve according to Kohonen's rule [78]. This rule was first introduced in artificial neural networks [78, 79], and was later shown to lead to the formation of self–organizing maps in model biological networks [80]. We used our theory to show that Kohonen's rule leads to stable asynchronous or correlated balanced states, and verified these predictions in simulations. We analyzed the particular dynamics of weights under Kohonen's rule that can result in a saddle–node bifurcation, and lastly showed that correlations have a small impact on weight dynamics, mildly shifting the fixed point of the system.

Asymptotic behavior of balanced networks under Kohonen's rule. Kohonen's Rule can be implemented by letting EE synaptic weights evolve according to [78] (see Table 2.1),

$$\frac{dJ_{jk}^{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\big(\beta x_j^{\mathrm{e}} S_k^{\mathrm{e}} - J_{jk}^{\mathrm{ee}} S_j^{\mathrm{e}}\big), \tag{3.5}$$

where $\beta > 0$ is a parameter that can change the fixed point of the system (see Fig. 3.1.5). This STDP rule is competitive as weight updates only occur when the pre–synaptic neuron is active, so that the most active pre–synaptic neurons change their synapses more than less active pre–synaptic cells.

The mean–field approximation describing the evolution of synaptic weights given in Eq. (2.120) has the form:

$$\frac{dJ_{\mathrm{ee}}}{dt} = \eta_{\mathrm{ee}}\left(\beta\tau_{STDP} r_{\mathrm{e}}^2 - J_{\mathrm{ee}} r_{\mathrm{e}} + \beta\int_{-\infty}^{\infty} \widetilde{K}(f)\langle S_{\mathrm{e}}, S_{\mathrm{e}}\rangle df\right). \tag{3.6}$$

The fixed point of Eq. (3.6) can be obtained using the expressions for the rates and covariances obtained in the balanced state (Eqs. (2.26,2.105)). The rates and covariances at steady–state can then be obtained from the resulting weights (see Section 2.6 for details on how to solve these equations).

Our theory predicts that the network attains a stable balanced state, and the average rates,

weights, and covariances at this equilibrium (Fig. 3.1.4) (see Section 3.4 for empirical distributions under Kohonen's and other rules). These predictions agree with numerical simulations in both the asynchronous and correlated states (Fig. 3.1.4 **B**,**C**). As expected, predictions improve with network size, N, and spike count covariances scale as $1/N$ in the asynchronous state (Fig. 3.1.4 **D**–**F**). Similar agreement holds in the correlated state, including the impact of the correction introduced in Eq. (3.6) (Fig. 3.1.4 **G**–**I**).

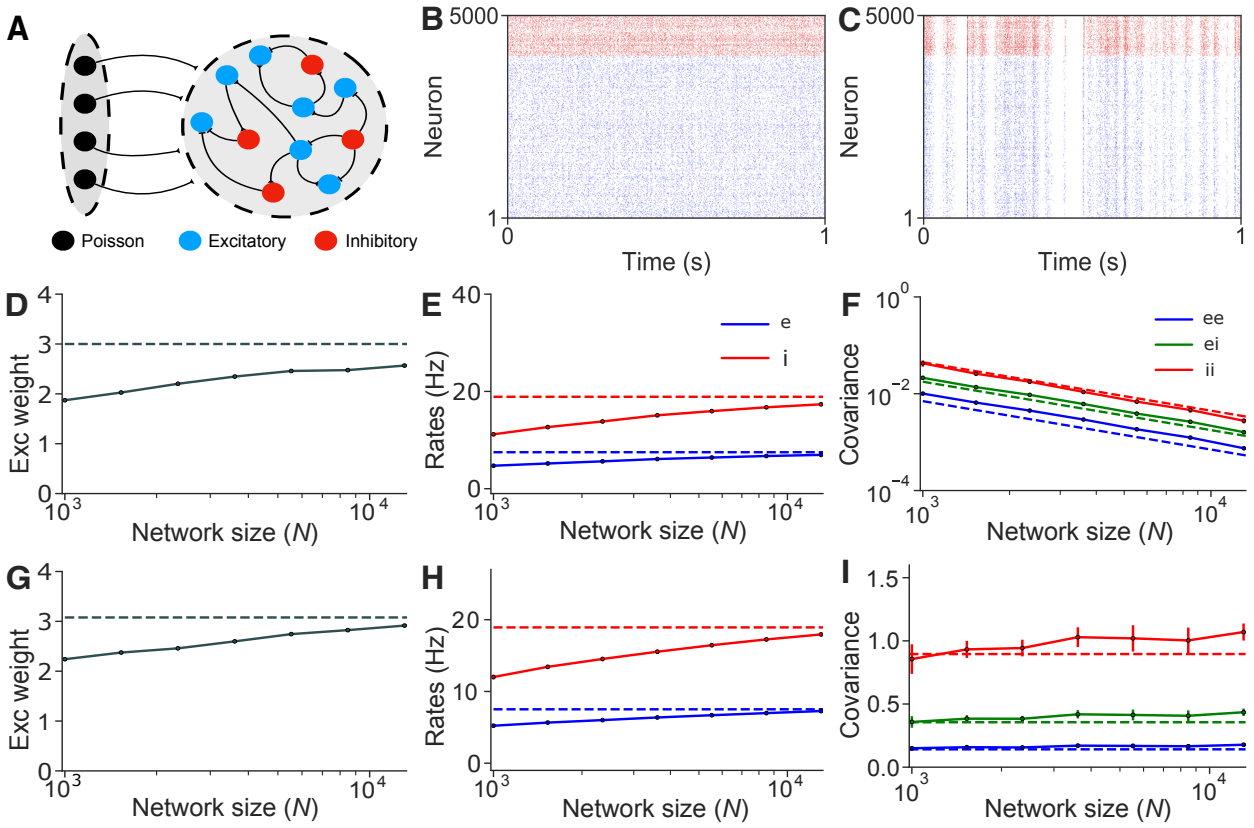

Figure 3.1.4: **A plastic, balanced network in asynchronous and correlated regimes. A:** A recurrent network of excitatory, E, and inhibitory, I, neurons is driven by an external feedforward layer, X, of correlated Poisson neurons. **B:** Raster plot of all neurons in a network of $N = 5000$ neurons in an asynchronous state. E cells in blue, I neurons in red . **C:** Same as (**B**), but in a correlated state. **D:** Mean steady state EE synaptic weight, j_{ee}, in an asynchronous state. **E:** Mean E and I firing rates for different network sizes, N, in an asynchronous state. **F:** Mean EE, II and EI spike count covariances in an asynchronous state. **G–I:** Same as (**D–F**) but for a network in a correlated state. Solid lines represent simulations, and dashed lines are values obtained using Eqs. (2.26,2.105,3.6). All empirical results were averaged over 10 realizations. In the asynchronous state $c_{\mathrm{x}} = 0$, and in the correlated state $c_{\mathrm{x}} = 0.1$. Unless otherwise stated, colors carry the same meaning in all figures. Figure adapted from Akil et al. [5].

Saddle–node bifurcation of excitatory weights in Kohonen's STDP rule. Recall from Eq. (3.5) that β is a parameter that represents the amplitude of the LTP term. Therefore it is natural to ask: How does the stability of the synaptic weights depend on β? Our mean-field theory shows that the stable network state disappears in a saddle–node bifurcation (Fig. 3.1.5 **A**). For the chosen set of parameters (see Table 3.1), the theory predicts a saddle–node bifurcation occurs at $\beta \approx 8$ (Fig. 3.1.5 **A**). We confirmed the presence of a saddle–node bifurcation in numerical simulations (Fig. 3.1.5 **B**). However, the bifurcation occurs at a lower value of β due to divergence of individual synapses (Fig. 3.1.5 **C–F**). This is not unexpected, as fluctuations can destabilize a system close to a saddle–node bifurcation by pushing it across the unstable fixed point.

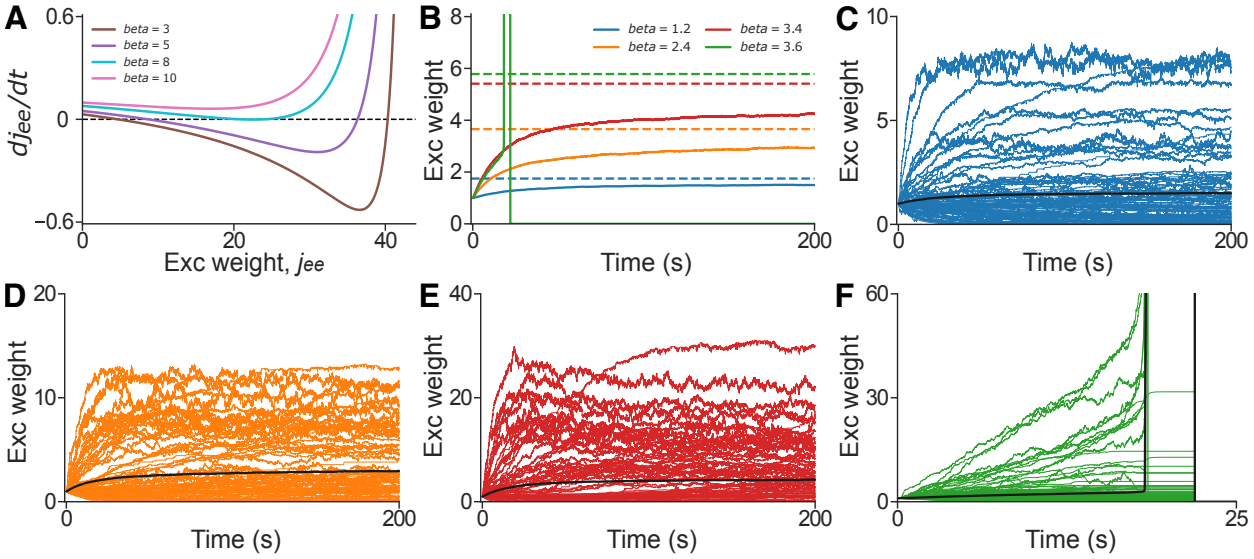

Figure 3.1.5: **Stability of weights in Kohonen's rule. A:** The derivative of the mean excitatory weight for different values of excitatory weights subject to Kohonen's STDP rule obtained from Eq. (3.6). The excitatory weight dynamics undergo a saddle–node bifurcation near $\beta \approx 8$. **B:** Evolution of mean synaptic weights over a range of values of β. A bifurcation occurs in simulations at $\beta \approx 3.6$, as noise pushes the system past the critical point earlier than the mean–field theory predicts. Dashed lines represent theoretical values, solid lines were obtained from simulations. **C:** Evolution of individual synaptic weights for $\beta = 1.2$. Mean synaptic weight in black. **D:** Same as **C:** for $\beta = 2.4$. **E:** Same as **C:** for $\beta = 3.4$. Variance of individual synaptic weight increases with β. **F:** Same as **C:** for $\beta = 3.6$. The mean, and some individual excitatory weights diverge. In panels **C–F**, solid lines were obtained from numerical simulations. (Eqs. (2.26,2.105,3.6)), and solid lines represent numerical results. Relevant simulation parameters: $N = 5000$, $c_x = 0$. Other parameters are given in Table 3.1. Figure adapted from Akil et al. [5].

Dynamics of correlated balanced networks under Kohonen's STDP. We next asked

whether and how the equilibrium and its stability are affected by correlated inputs to a plastic, balanced network. In particular, we used our theory to determine whether changes in synaptic weights are driven predominantly by the firing rates of the pre– and post–synaptic cells, or correlations in their activity. We also asked whether correlations in neural activity can change the equilibrium, the speed of convergence to the equilibrium, or both? We have seen that for Hebbian STDP increases in correlations can increase the rate of convergence to the fixed point. We now address this question for networks under Kohonen STDP.

We first address the role of correlations. Our theory predicts that a plastic, balanced network remains stable under Kohonen's rule, and that an increase in the mean *EE* weights by $10-20\%$ as correlations in the input are increased (Eqs. (2.26,2.105,3.6)). Both predictions were confirmed by simulations (Fig. 3.1.6 **A**,**B**). We also used the theory to predict that this increase in synaptic weights results in negligible changes in firing rates, which simulations again confirmed (Fig. 3.1.6 **C**).

How large is the impact of correlations in plastic balanced networks more generally? To address this question, we assumed that only pairwise interactions affect *EE* synapses, as first order interactions depend only on rates after averaging. We thus set $B_{\alpha,\beta} \equiv 1$, and all other coefficients to zero in Eq. (2.107). While the network does not reach a stable state under this arbitrary plasticity rule, it allows us to estimate the contribution of rates and covariances to the evolution of synaptic weights. Here $B_{\alpha,\beta}$ can have any nonzero value, since it scales both the rate and covariance terms. Under these conditions, our theory predicts that the rate term is at least an order of magnitude larger than the correlation term (even when rates themselves are small, *i.e.*, when j_{ee} is small), and so correlations only have a low impact on the dynamics of synaptic weights (Fig. 3.1.6 **D**). Therefore, our theory predicts that, in general, changes in synaptic weights will largely be driven by changes in firing rate patterns, rather than modulations in pairwise correlations.

We next ask the opposite question: How do changes in synaptic weights impact firing rates, and covariances? We predicted, using the full theory (see Eqs. (2.26,2.105), and Section 2.8), that potentiation of *EE* weights leads to large increases in rates and spike count covariances. This

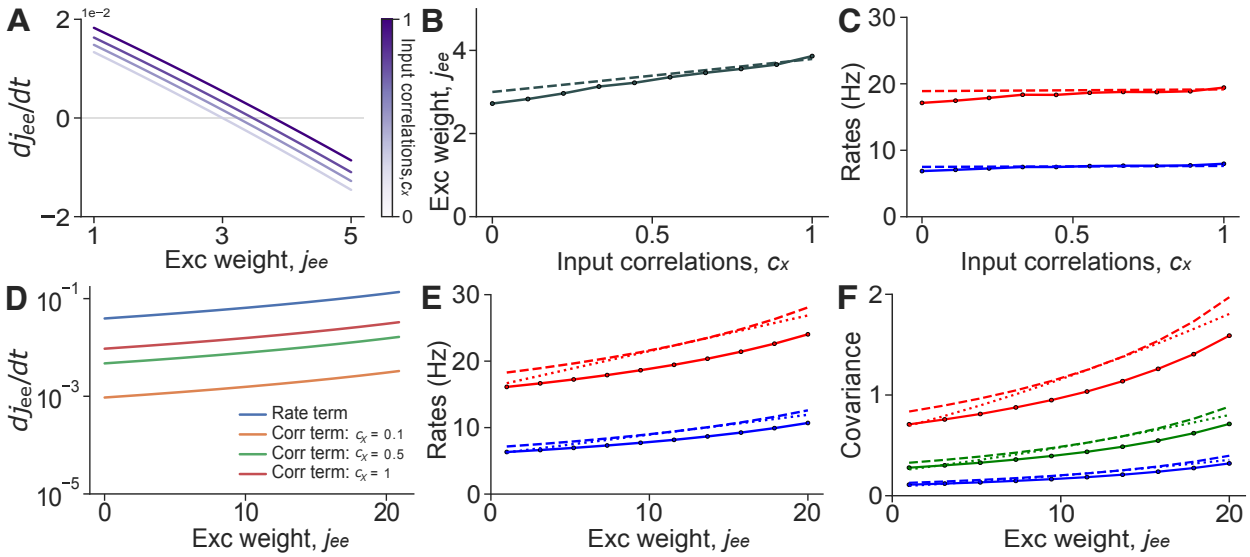

Figure 3.1.6: **Spike count covariances mildly impact the fixed point of synaptic weights and firing rates. A:** The rate of change of EE weights as function of the weight, j_{ee}, at different levels of input correlations, c_{x}. **B:** Mean steady–state EE synaptic weight for a range of input correlations, c_{x}. **C:** Mean E and I firing rates as a function of input correlations. **D:** Same as (**A**) but for an EE STDP rule with all coefficients involving order 2 interactions set equal to 1, and all other coefficients set equal to zero. **E:** Mean E and I firing rates as a function of mean EE synaptic weights. **F:** Mean spike count covariances between E spike trains, I spike trains, and between E–I spike trains as a function of EE synaptic weight, j_{ee}. Solid lines represent simulations (except in **A**,**D**), dashed lines are values obtained from theory (Eqs. (2.26,2.105,3.6)), and dotted lines were obtained from the perturbative analysis. Note that in all panels, 'Exc weight' refers to j_{ee} rather than J_{ee}, as the former does not depend on N. Figure adapted from Akil et al. [5].

prediction was again confirmed by numerical simulations (Fig. 3.1.6 **E**,**F**). This observation holds generally, and STDP rules that results in large changes in synaptic weights will produce large changes in rates and covariances.

Our theory thus shows that *in general* weight dynamics can be moderately affected by correlations when these are large enough (see Fig. 3.1.3 for a similar analysis on Hebbian STDP). In turn, changes in synaptic weights will generally change the structure of correlated activity in a balanced network.

The predictions of the theory hold in all cases we tested (see Sections 3.1.1 and 3.2). Understanding when plasticity will support a stable balanced state allows one to implement Kohonen's rule and other rules in complex contexts and tasks, without the emergence of pathological states

(such as those described in Section 3.1.1 in networks under classical Hebbian STDP).

3.2 Balanced Networks under Inhibitory Plasticity

Next, we show that in its basic form our theory can fail in networks subject to inhibitory STDP, and how the theory can be extended to capture such dynamics. The failure is due to correlations between weights and pre–synaptic rates which are typically ignored [10, 83, 110, 115, 137, 138] (we used this assumption in our derivation of Eqs. (2.9,2.16,2.18,2.23,2.73, 2.75,2.76,2.77)), but can cause the mean–field description of network dynamics to become inaccurate. This is similar to the breakdown of balanced state theory in the presence of correlations between in– and out–degrees discussed by Vegué and Roxin, 2019 [139].

To illustrate this, we consider a balanced network subject to homeostatic plasticity [141]. This type of plasticity has been shown to stabilize the asynchronous balanced state and conjectured to play a role in the maintenance of memories [63, 87, 141]. Following [141] we assume that *EI* weights evolve according to:

$$\frac{dJ^{\mathrm{ei}}_{jk}}{dt} = -\eta_{\mathrm{ei}} \frac{J^{\mathrm{ei}}_{jk}}{J_{\mathrm{norm}}} \left[(x^{\mathrm{e}}_j - \alpha_{\mathrm{e}}) S^{\mathrm{i}}_k + x^{\mathrm{i}}_k S^{\mathrm{e}}_j \right] \tag{3.7}$$

where α_{e} is a constant that determines the target firing rates of E cells and $J_{\mathrm{norm}} \sim \mathcal{O}(1/\sqrt{N})$ is a normalization constant. Note that $J_{\mathrm{norm}} < 0$ so the fraction in Eq. (3.7) is positive assuming $J^{\mathrm{ei}}_{jk} < 0$. In a departure from the rule originally proposed by Vogels et al. [141], we chose to multiply the time derivative by the current weight value. This modification creates an unstable fixed point at zero, prevents *EI* weights from changing signs, and keeps the analysis mathematically tractable (see Section 2.12 for more details). An alternative way to prevent weights from changing sign would be to place a hard bound at zero, but this would create a discontinuity in the vector field of J_{ei}, complicating the analysis.

On average, the inhibitory weights change as,

$$\frac{dJ_{\rm ei}}{dt} = -\eta_{\rm ei}\frac{J_{\rm ei}}{J_{\rm norm}}\Big((2\tau_{STDP}r_{\rm e}-\alpha_{\rm e})r_{\rm i}+2\int_{-\infty}^{\infty}\widetilde{K}(f){\rm Re}[\langle S_{\rm e},S_{\rm i}\rangle]df\Big). \tag{3.8}$$

It is easy to see that this modification imposes an unstable fixed point at $J_{\rm ei}=0$.

Weights in original iSTDP rule may switch signs. We mentioned that the form of inhibitory STDP in [141] was slightly modified to include a zero unstable fixed point in the synaptic weights. This was done because under certain relevant conditions, the synaptic weights can change sign in an effort to maintain E rates at the target.

One simple case in which this happens, is when all I neurons are stimulated (Fig. 3.2.1). The network is put in a semi–balanced regime [11] where neurons receive excess inhibition on average (Fig. 3.2.1 **A** inset). The inhibitory rates increase, which would effectively decrease E rates. However, I to E synaptic weights decrease so that the E rates are maintained near the target (Fig. 3.2.1 **A** and **B**). Since EI weights are not controlled to stay negative, excess inhibition causes some synapses to switch signs in order to maintain E rates at the target (Fig. 3.2.1 **C**). This violates Dale's law, so in all other simulations of inhibitory STDP, the learning rule is modified to incorporate a zero unstable fixed point, which prevents synaptic weights from switching signs (as proven in Section 2.12).

Dynamics of correlated balanced networks under inhibitory STDP. Under the rule described by Eq. (3.7) a lone pre–synaptic spike depresses the synapse, while near–coincident pre– and post–synaptic spikes potentiate the synapse. Changes in EI weights steer the rates of individual excitatory cells to the target value $\rho_{\rm e}:=\frac{\alpha_{\rm e}}{2\tau_{STDP}}$. Indeed, individual EI weights are potentiated if post–synaptic firing rates are higher than $\rho_{\rm e}$, and depressed if the rate is below $\rho_{\rm e}$. Our theory predicts that the network converges to a stable balanced state (Fig. 3.2.2 **A**). Correlations again have only a mild impact on the evolution of synaptic weights (Fig. 3.2.2 **A**).

Although our theory predicts a single stable fixed point for the average EI weight, simulations

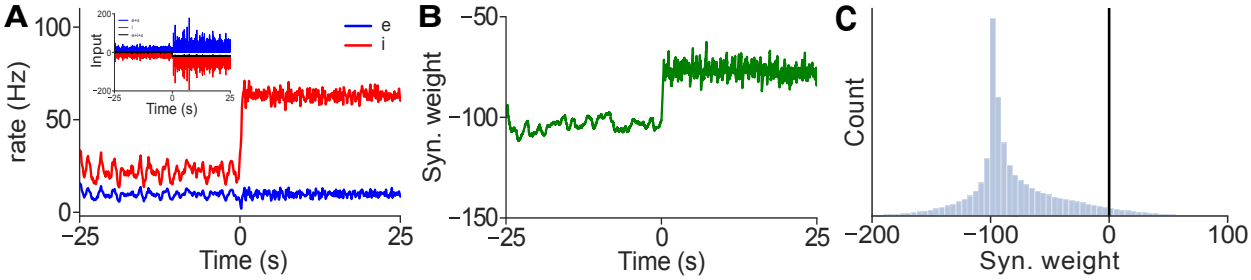

Figure 3.2.1: **Homeostatic nature of inhibitory STDP can break Dale's Law. A:** Excitatory and inhibitory firing rates over time. External stimulation starting at time 0 sec increases the firing rate of I cells. Inset: Average excitatory, inhibitory, and total inputs. Total input is negative when I cells are stimulated. **B:** Mean EI synaptic weight over time. After stimulation onset, EI weights decrease in magnitude. **C:** Distribution of EI synaptic weights. A number of inhibitory synapses have turned positive. All plots obtained from a single simulation, where added external input to all inhibitory neurons starts at time 0 sec. Figure adapted from Akil et al. [5].

show that weights converge to a different average depending on the initial EI weights (Fig. 3.2.2 **B–E** solid lines). A manifold of stable fixed points emerges due to synaptic competition, which is a consequence of heterogeneity in inhibitory firing rates in the network: Weights of highly active pre–synaptic inhibitory cells are potentiated more strongly compared to those of lower firing cells (Fig. 3.2.2 **E**). Thus, while inhibitory rates and EI weights are initially uncorrelated, correlations emerge as the excitatory rates approach their target. Networks with different initial EI synaptic weights, converge to different final distributions, and the emergent correlations between weights and rates drive the system to different fixed points (Fig. 3.2.2 **C,D**).

We used a semi–analytical approach to confirm that correlations between weights and rates explain the discrepancy between predictions of the mean field theory, and simulations. To do so we introduced a correlation–dependent correction term into the expression for the rates:

$$\lim_{N\to\infty} \vec{r} = -\overline{W}^{-1}\big(\overline{W}_{\mathrm{x}} r_{\mathrm{x}} + \mathrm{cov}(J_{\mathrm{ei}}, r_{\mathrm{i}})\big), \tag{3.9}$$

where $\mathrm{cov}(J_{\mathrm{ei}}, r_{\mathrm{i}}) := [\langle\langle J^{\mathrm{ei}}_{jk} r^{\mathrm{i}}_k\rangle_k - \langle J^{\mathrm{ei}}_{jk}\rangle_k \langle r^{\mathrm{i}}_k\rangle_k\rangle_j, 0]^T$. Recall that in our derivation of mean firing rates we ignored this term (see Eq. (2.18)). That was a valid assumption in non–plastic networks,

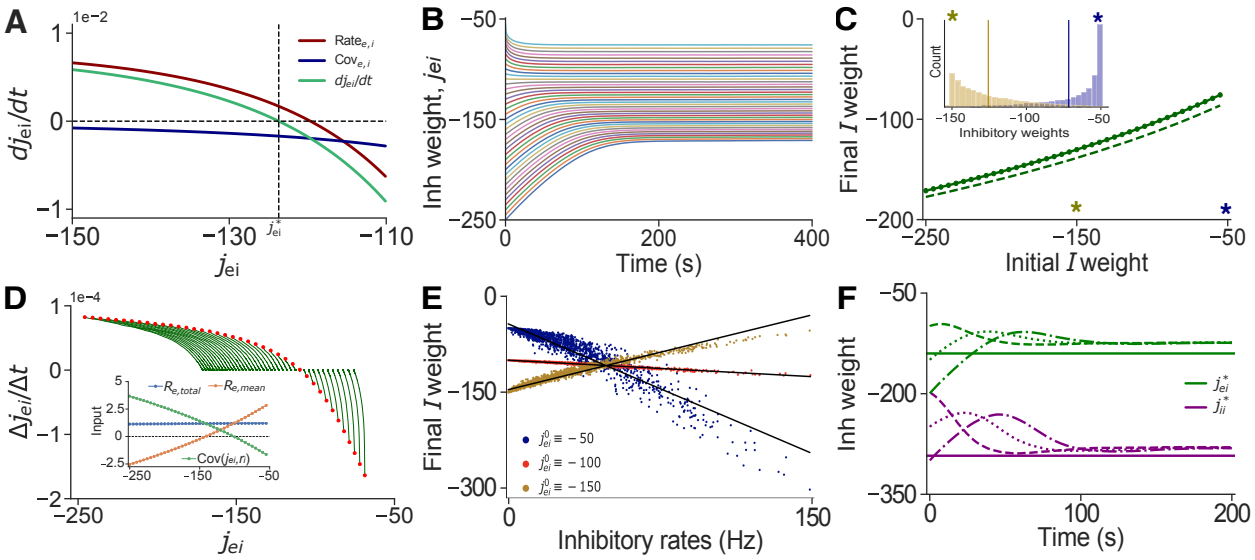

Figure 3.2.2: **Correlations between synaptic weights and inhibitory rates lead to the formation of a manifold in weight space. A:** The rate of change of EI weights as a function of the weights themselves. The contributions of the covariance (blue) is considerably smaller than the contribution of the rate (red), and the theory predicts a stable fixed point. **B:** Evolution of inhibitory weights showing that different initial weights converge to different fixed points. Inset: weights starting at different initial conditions converge to equilibrium at different times for fixed $1/\eta_{\rm ei} = 10^{-5}$). **C:** A manifold of fixed points in $j^*_{\rm ei} - j^0_{\rm ei}$ space emerges due to correlations between weights and rates. Solid line represents simulations, dashed line are values obtained from the modified theory (Eqs. (3.9,2.105,3.8)). Inset: Final distribution of EI weights for a network with initial weights $j^0_{\rm ei} = 150$ (yellow), and $j^0_{\rm ei} = 50$ (blue). Modified theory predicts the manifold of fixed points. **D:** Same as **A**, but obtained from simulations. Lines represents trajectories from different initial weights (red dots). Inset: Total recurrent input to E neurons, $R_{\rm e,total} = \langle\langle J^{\rm ee}_{jk} r^{\rm e}_j\rangle_j + \langle J^{\rm ei}_{kl} r^{\rm i}_l\rangle_l\rangle_k$ for a range of initial weights. Mean recurrent input to E cells, $R_{\rm e,mean} = w_{\rm ee} r_{\rm e} + w_{\rm ei} r_{\rm i}$. The mean input deviates from the total input due to emergent correlations between weights and rates, ${\rm Cov}(J_{\rm ei}, r_{\rm i}) = R_{\rm e,total} R_{\rm e,mean}$. **E:** The weights of individual EI synapses corresponding to the same post–synaptic E cell as a function of the equilibrium firing rates of pre-–synaptic I neurons. Each color represents a different simulation of the network with different initial EI weight. Equilibrium inhibitory weights and pre–synaptic rates are correlated (Blue: $R^2 = 0.952$, Red: $R^2 = 0.9865$, Yellow: $R^2 = 0.979$). **F:** Sample trajectories of the $j_{\rm ei} - j_{\rm ii}$ system for a network of $N = 10^4$ neurons in an asynchronous state. Simulations with different initial weights (dashed lines), converge to a fixed point close to the one predicted by the theory (solid line). Figure adapted from Akil et al. [5].

however for this rule the presence of strong correlations between pre–synaptic rates and weights is evident and needs to be accounted for (Fig. 3.2.2 **E**). The average covariances between weights and rates obtained numerically explain the departure from the mean–field predictions (Fig. 3.2.2 **C**). Using the corrected equation (Eq. (3.9)), we predicted mean equilibrium weights that agree well

with simulations (Fig. 3.2.2 **C** dashed line).

iSTDP on *II* synapses decorrelates weights and rates. We next asked whether the mean–field theory provides a good description of network dynamics in the absence of correlations between weights and rates. Such correlations disappear in a network with homogeneous inhibitory firing rates. Finding an initial distribution of weights that result in a balanced state with uniform inhibitory firing rates is non–trivial, and may not be possible outside of unstable regimes exhibiting rate–chaos where mean–field theory ceases to be valid [105]. However, allowing *II* synapses to evolve under the same plasticity rule we used for *EI* synapses can homogenize inhibitory firing rates: If we let

$$\frac{dJ^{\mathrm{ii}}_{jk}}{dt} = -\eta_{\mathrm{ii}} \frac{J^{\mathrm{ii}}_{jk}}{J_{\mathrm{norm}}} \left[(x^{\mathrm{i}}_j - \alpha_{\mathrm{i}}) S^{\mathrm{i}}_k + x^{\mathrm{i}}_k S^{\mathrm{i}}_j \right], \tag{3.10}$$

all inhibitory responses approach a target rate $\rho_{\mathrm{i}} = \frac{\alpha_{\mathrm{i}}}{2\tau_{STDP}}$, effectively removing the variability in I rates. The evolution of the mean II and EI synaptic weights is now given by

$$\begin{aligned} \frac{dJ_{\mathrm{ei}}}{dt} &= -\eta_{\mathrm{ei}} \frac{J_{\mathrm{ei}}}{J_{\mathrm{norm}}} \Big((2\tau_{STDP} r_{\mathrm{e}} - \alpha_{\mathrm{e}}) r_{\mathrm{i}} + 2 \int_{-\infty}^{\infty} \widetilde{K}(f) \mathrm{Re}[\langle S_{\mathrm{e}}, S_{\mathrm{i}} \rangle] df \Big), \\ \frac{dJ_{\mathrm{ii}}}{dt} &= -\eta_{\mathrm{ii}} \frac{J_{\mathrm{ii}}}{J_{\mathrm{norm}}} \Big((2\tau_{STDP} r_{\mathrm{i}} - \alpha_{\mathrm{i}}) r_{\mathrm{i}} + 2 \int_{-\infty}^{\infty} \widetilde{K}(f) \langle S_{\mathrm{i}}, S_{\mathrm{i}} \rangle df \Big). \end{aligned} \tag{3.11}$$

We conjectured that if inhibitory rates converge to a common target, synaptic competition would be eliminated, and no correlations between weights and rates would emerge. This, in turn, would remove the main obstacle to the validity of a mean–field description. The fixed point of these equations can again be obtained using Eqs. (2.26,2.105) which predict that the network remains in a stable balanced state (asynchronous or correlated). We also require $\eta_{\mathrm{ei}} \geq \eta_{\mathrm{ii}}$, since when η_{ei} is much slower than η_{ii}, the network becomes unstable as homogeneous inhibitory weights and rates are not able to stabilize the heterogeneous distribution of E activity (see below "Stability of iSTDP in *EI* and *II* connections depends on weights timescales." and Fig. 3.2.3). We chose the same STDP timescale for both *EI* and *II* synapses, and our predictions agree with the results of

simulations (Fig. 3.2.2 **F**). The stable manifold of fixed points is replaced by a single stable fixed point, and the average weights and rates approach a state that is independent of the initial weight distribution.

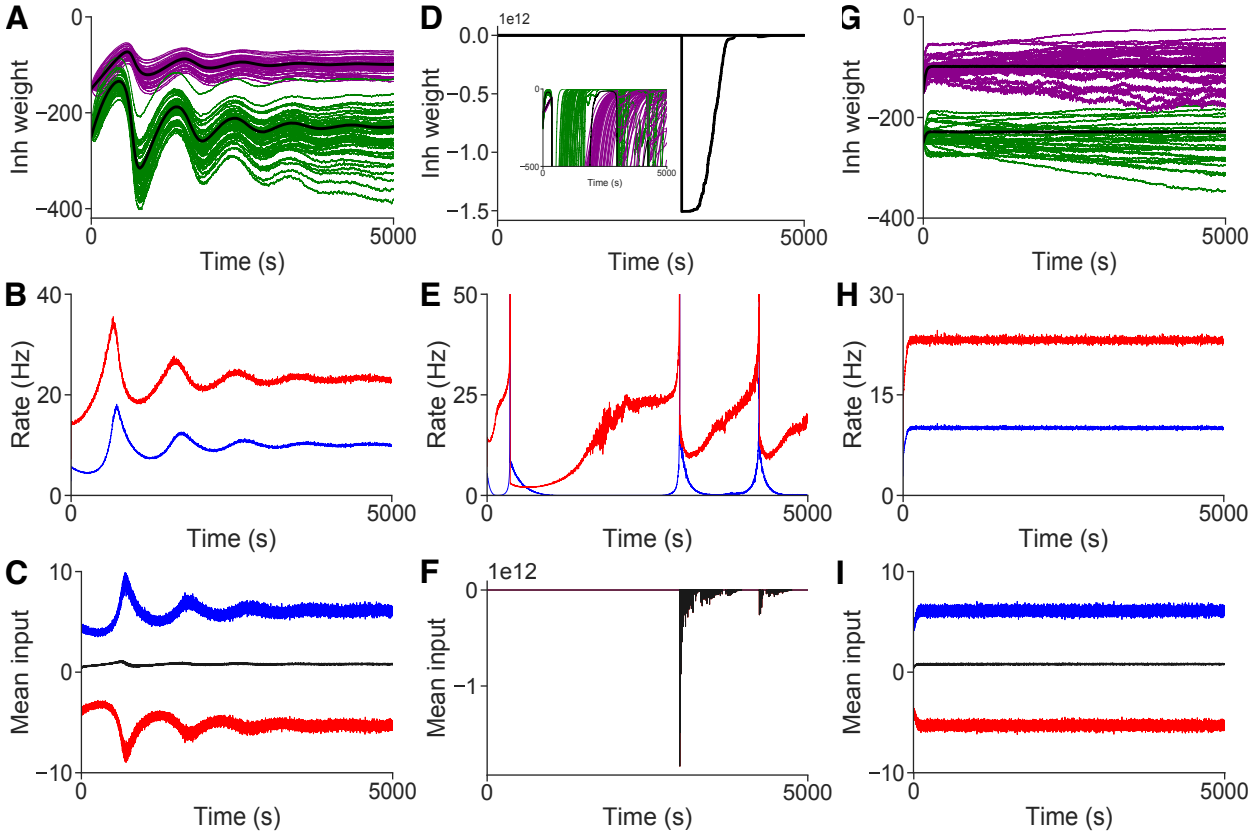

Figure 3.2.3: **Inhibitory STDP: different timescales on *EI,II* synapses lead to different behavior.** **A:** Individual *EI* (green) and *II* (dark magenta) synaptic weights. Average of each type in black. **B:** Mean firing rates of E and I populations. **C:** Mean excitatory (blue), inhibitory (red) and total (black) input. Network is in balanced state. In **A–C**, $\eta_{\text{ei}} = \eta_{\text{ii}} = 10^{-4}$. **D–F:** Same as **A–C**, but here $\eta_{\text{ei}} = 10^{-4}$ and $\eta_{\text{ii}} = 10^{-3}$. Weights and consequently the mean input grow uncontrollably. Rates oscillate and undergo periods of quiescence of E neurons. **G–I:** Same as **A–C**, but now $\eta_{\text{ei}} = 10^{-3}$ and $\eta_{\text{ii}} = 10^{-4}$. Activity in the network is now balanced and stable. Relevant parameters: $N = 5000$, $c_{\text{x}} = 0$. Other parameters as in Supporting Table 3.1. Figure adapted from Akil et al. [5].

This model of inhibitory plasticity is likely a large oversimplification. Synapses of different interneuron subtypes are likely subject to different plasticity rules operating on different timescales [24, 146], and would therefore not lead to uniform inhibitory firing rates. The mean-field theory we presented here can be extended to account for multiple inhibitory subtypes with

different plasticity rules.

Stability of iSTDP in *EI* and *II* connections depends on weights timescales. Consider again a network with inhibitory STDP in *EI* and *II* connections. As we showed above, the theory predicts one stable fixed point for the weights. However, convergence to that stable fixed point depends on individual timescales of synaptic weights. In particular, if $\eta_{\text{ei}} \geq \eta_{\text{ii}}$, then the network realizes a stable balanced state (Fig. 3.2.3 **A–C,G–I**). However, if *EI* weights change at least an order of magnitude slower than *II* weights, then the network is destabilized in attempt to control excitatory rates (Fig. 3.2.3 **D–F**): Early in the simulation, *II* weights change fast in order to push *I* cells to the target rate (Fig. 3.2.3 **D**). As a consequence, inhibitory feedback increases and *E* neurons are silenced (Fig. 3.2.3 **E**), since *EI* STDP works on a slower timescale and is not able to catch up with *II* STDP to push *E* rates to the target. At this point, *EI* STDP makes relatively large updates since the *E* rates are far from target, and it strongly overshoots which causes rates to rise very fast past the target. This also increases *I* rates away from target, and so *II* weights change very fast to pull *I* rates towards their target. As a result both *E* and *I* rates drop below target (Fig. 3.2.3 **E** – first oscillation). The cycle then repeats. These oscillations cause weights and inputs to grow unrealistically large, and push the network out of balance. (Fig. 3.2.3 **F**).

We next show that balanced networks subject to iSTDP are robust to perturbatory, optogenetic inputs. Our theory predicts, and simulations confirm, that this learning rule maintains balance when non–plastic networks do not, and it can return the network to its original state after stimulation.

3.3 Inhibitory Plasticity Adapts Network's Response to Stimuli

Thus far, we analyzed the dynamics of plastic networks in isolation. However, cortical networks are constantly driven by sensory input, as well as feedback from other cortical and subcortical areas. We next ask whether and how balance is restored if a subset of pyramidal (excitatory) neurons are stimulated [33].

Optogenetic techniques are frequently used to target and perturb the activity of neural subpopulations in order to better understand cortical dynamics and neural computation. In such experiments, firing rates of targeted neurons increase when exposed to light due to the expression of microbial rhodopsins such as channelrhodopsin 2 (ChR2).

However, not all target neurons express the ChR2 protein [1, 20, 107, 109]. Thus stimulation separates the target, *e.g.*, pyramidal cell population into stimulated and unstimulated subpopulations. Although classical mean–field theory produced singular solutions, Ebsch, et al. showed that the theory can be extended, and that a non–classical balanced state is realized as long as $\overline{W}_{\mathrm{x}}r_{\mathrm{x}}$ is in the column space of $\overline{W}$: Balance at the level of population averages (E and I) is maintained, while balance at the level of the three subpopulations is broken [33]. Since local connectivity is not tuned to account for the extra stimulation (optogenetics), local synaptic input cannot cancel external input to the individual subpopulations. However, the input averaged over the stimulated and unstimulated excitatory population is cancelled.

We show that inhibitory STDP, as described by Eq. (3.7), can restore balance in the inputs to the stimulated and unstimulated subpopulations. Similarly, Vogels, et al. showed numerically that such plasticity restores balance in memory networks (*i.e.*, networks containing Hebbian assemblies) [141]. Here, we present an accompanying theory that describes the evolution of rates, covariances, and weights before, during, and after stimulation, and confirm the prediction of the theory numerically.

We assume that a subpopulation of pyramidal neurons in a correlated balanced network receives a transient excitatory input. This could be a longer transient input from another subnetwork, or an experimentally applied stimulus. To model this drive, we assume that the network receives input from two populations Poisson neurons, X_1 and X_2. The first population drives all neurons in the recurrent network, and was denoted by X above. The second population, X_2, provides an additional input to a subset of excitatory cells in the network, for instance ChR2–expressing pyramidal neurons (E_{expr} in Fig. 3.3.1 **A**). The resulting connectivity matrix between the stimulated (e_1), unstimulated (e_2) and inhibitory (i) subpopulations, and the feed–forward input weight matrix have the form:

$$\overline{W} = \begin{bmatrix} \overline{w}_{\mathrm{e_1e_1}} & \overline{w}_{\mathrm{e_1e_2}} & \overline{w}_{\mathrm{e_1i}} \\ \overline{w}_{\mathrm{e_2e_1}} & \overline{w}_{\mathrm{e_2e_2}} & \overline{w}_{\mathrm{e_2i}} \\ \overline{w}_{\mathrm{ie_1}} & \overline{w}_{\mathrm{ie_2}} & \overline{w}_{\mathrm{ii}} \end{bmatrix}, \text{ and } \overline{W}_x = \begin{bmatrix} \overline{w}_{\mathrm{e_1x_1}} & \overline{w}_{\mathrm{e_1x_2}} \\ \overline{w}_{\mathrm{e_2x_1}} & 0 \\ \overline{w}_{\mathrm{ix_1}} & 0 \end{bmatrix}, \tag{3.12}$$

where $\overline{w}_{ab} = p_{ab} j_{ab} q_b \sim \mathcal{O}(1)$, as before.

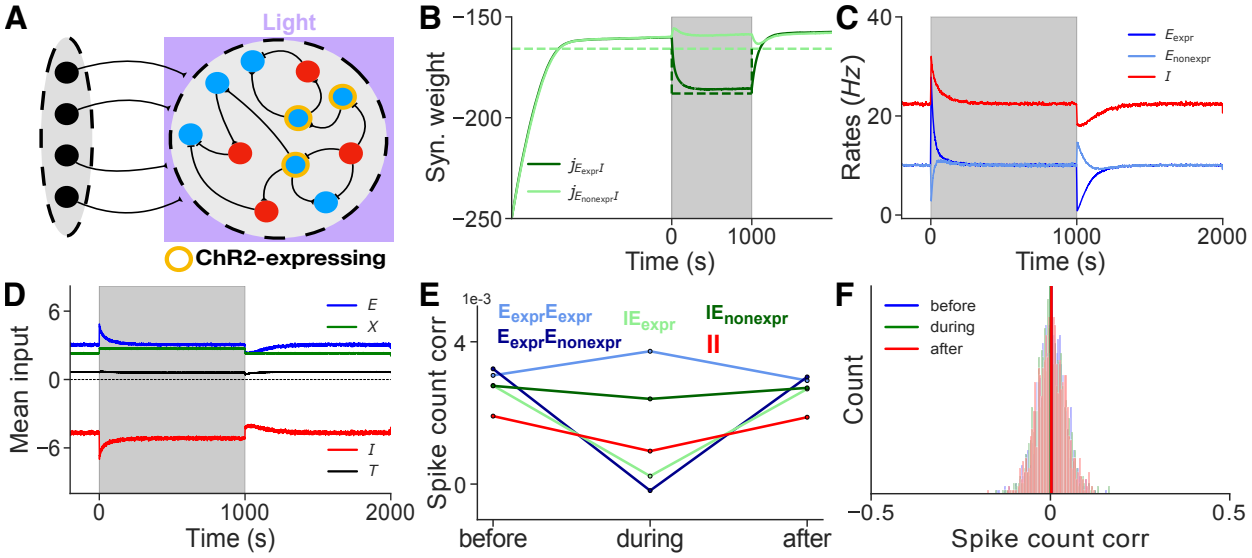

Figure 3.3.1: **Framework for STDP in balanced networks describes the dynamics of networks receiving optogenetic input. A:** A recurrent network of excitatory, E, and inhibitory, I, neurons is driven by an external feedforward layer X_1 of uncorrelated Poisson neurons. Neurons that express ChR2 are driven by optogenetic input, which is modeled as an extra layer of Poisson neurons denoted by X_2. **B:** Evolution of mean synaptic weights over the course of the experiment. **C:** Evolution of mean firing rates. Inhibitory STDP maintains E rates near the target, $\frac{\alpha_e}{2\tau_{\mathrm{STDP}}}$. **D:** Evolution of mean excitatory, external, inhibitory, and total currents. Balance is transiently disrupted at stimulus onset and offset, but it is quickly restored by iSTDP. **E:** Mean spike count correlations before, during, and after stimulation remain very weak for all pairs. **F:** The distribution of spike count correlations also remains nearly unchanged with weak mean correlations before, during, and after stimulation. Solid lines represent simulations, dashed lines are values obtained from theory (Eqs. (2.26,2.105,3.13,3.14)). Figure adapted from Akil et al. [5].

The mean–field equation relating firing rates to average weights and input (Eq. (2.26)) holds, with the vector of rates $r = [r_{\mathrm{e_1}}, r_{\mathrm{e_2}}, r_{\mathrm{i}}]^T$, and input vector $r_{\mathrm{x}} = [r_{\mathrm{x_1}}, r_{\mathrm{x_2}}]^T$. Similarly, mean spike count covariances are now represented by a 3×3 matrix that satisfies Eq. (2.105). The mean $E_1 I$

and E_2I weights evolves according to

$$\frac{dJ_{\mathrm{e_1i}}}{dt} = -\eta_{\mathrm{e_1i}}\frac{J_{\mathrm{e_1i}}}{J^0_{\mathrm{e_1i}}}\Big((2\tau_{\mathrm{STDP}}r_{\mathrm{e_1}} - \alpha_{\mathrm{e}})r_{\mathrm{i}} + 2\int_{-\infty}^{\infty}\widetilde{K}(f)\mathrm{Re}[\langle S_{\mathrm{e_1}}, S_{\mathrm{i}}\rangle]df\Big) \tag{3.13}$$

$$\frac{dJ_{\mathrm{e_2i}}}{dt} = -\eta_{\mathrm{e_2i}}\frac{J_{\mathrm{e_2i}}}{J^0_{\mathrm{e_2i}}}\Big((2\tau_{\mathrm{STDP}}r_{\mathrm{e_2}} - \alpha_{\mathrm{e}})r_{\mathrm{i}} + 2\int_{-\infty}^{\infty}\widetilde{K}(f)\mathrm{Re}[\langle S_{\mathrm{e_2}}, S_{\mathrm{i}}\rangle]df\Big). \tag{3.14}$$

We simulated a network of $N = 10^4$ neurons in an asynchronous state with $c_{\mathrm{x_1}} = c_{\mathrm{x_2}} = 0$. A subpopulation of 4000 E cells receives transient input. Solving Eqs. (3.13,3.14) predicts that inhibitory plasticity will alter EI synaptic weights so that the firing rates of both the E_{expr} and the $E_{\mathrm{non-expr}}$ approach the target firing rate before, during, and after stimulation. Once the network reaches steady state the mean inputs to each subpopulation cancel. Thus changes in EI weights restore balance at the level of individual subpopulations or "detailed balance," consistent with previous studies [63, 141]. Simulations confirm these predictions (Fig. 3.3.1 **B–D**).

When the input is removed, the inhibitory weights onto cells in the E_{expr} subpopulation converge to their pre–stimulus values, returning E_{expr} rates to the target value, and reestablishing balance (Fig. 3.3.1 **B–D**). Correlations remain low ($\mathcal{O} \sim 10^{-4}$) before, during, and after stimulation (Fig. 3.3.1 **E,F**), suggesting that at equilibrium the network is in the asynchronous state. Our theory thus describes how homeostatic inhibitory STDP increases the stability and robustness of balanced networks to perturbations by balancing inputs at a level of individual cells, maintaining balance in regimes in which non–plastic networks cannot maintain balance. We presented an example in which only one subpopulation is stimulated. However, the theory can be extended to any number of subpopulations in asynchronous or correlated balanced networks receiving a variety of transient stimulus.

We have thus shown that our theory captures the dynamics of stimulated plastic balanced networks. Homeostatic inhibitory STDP increases the stability and robustness of balanced networks to perturbations by balancing inputs at a level of individual cells, maintaining balance in regimes in which non–plastic networks are unstable. We presented an example in which only one subpopulation is stimulated. However, the theory can be extended to any number of subpopulations in

asynchronous or correlated balanced networks receiving a variety of transient stimulus.

3.4 Statistics and Dynamics of Balanced Networks under Pairwise STDP Rules

Our theoretical results describe the evolution of averages (see Chapter 2, Eqs. (2.26,2.105,2.120)). However, it is also important to understand changes in higher order statistics of individual, and collective activity under plasticity. Although we do not go beyond a mean–field theory, we provide the results of corresponding numerical simulations in three example balanced networks where: (1) excitatory weights were subject to weight–dependent Hebbian STDP; (2) *EE* weights changed according to Kohonen's rule; and (3) inhibitory weights changed according to an iSTDP rule (see Table 2.1).

Under weight–dependent Hebbian STDP individual synaptic weights converge to a unimodal distribution. This is in agreement with experimental results [129] (Fig. 3.4.1 **A,B**). Balance is preserved throughout the simulation (Fig. 3.4.1 **C**). Distributions of rates are unimodal and skewed to the right (Fig. 3.4.1 **D** and inset), also agreeing with experimental findings [129]. Spike count covariances and correlations are distributed around zero in the asynchronous state (Fig. 3.4.1 **E,F**).

EE weights under Kohonen's rule also converge to a unimodal distribution (Fig. 3.4.2 **A,B**). Again, the network is stable and balance is preserved (Fig. 3.4.2 **C**). The distributions of rates is again unimodal and skewed to the right (Fig. 3.4.2 **D**), and covariances and correlations are close to zero (Fig. 3.4.2 **E,F**). Lastly, *EI* weights subject to an inhibitory plasticity rule also follow a unimodal distribution (Fig. 3.4.3 **A,B**), distributions of I rates are unimodal and skewed to the right, and covariances and correlations are centered at zero in the balanced, asynchronous state (Fig. 3.4.3 **C–F**). However, the E rates are concentrated at the target rate (Fig. 3.4.3 **D** – inset), in disagreement with experimental findings [129]. In short, all three STDP rules show realistic distributions of weights, rates, and covariances, except for the distribution of E rates under iSTDP.

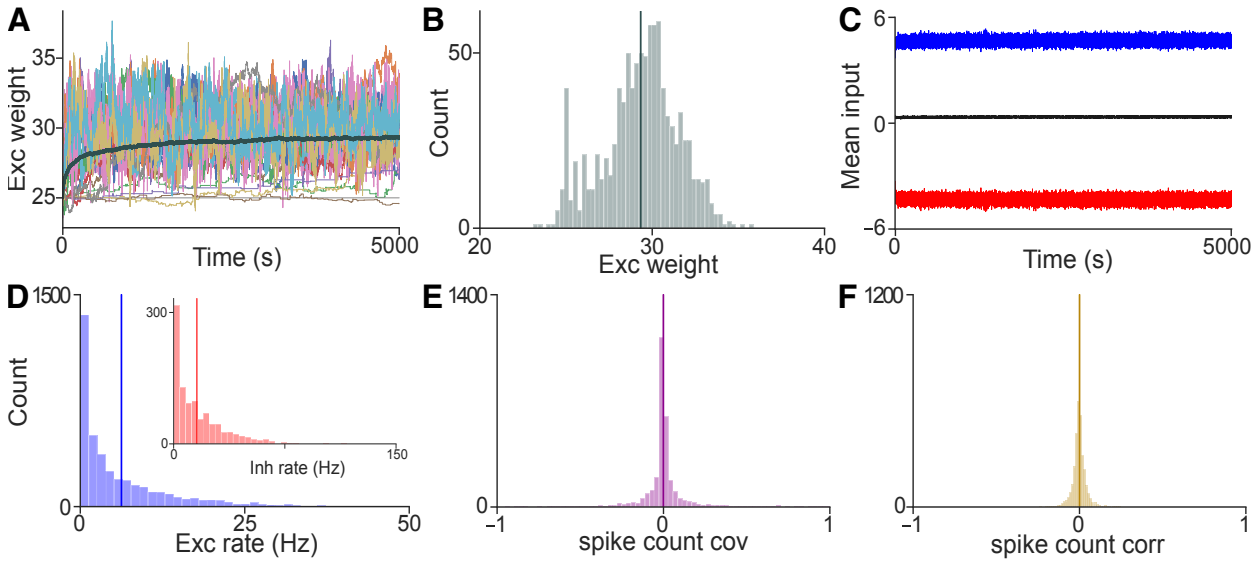

Figure 3.4.1: **Statistics of weights and activity under weight–dependent Hebbian STDP.** **A:** Evolution of individual synaptic weights and their mean (black). **B:** *EE* synaptic weights follow a unimodal distribution at steady state. **C:** Evolution of synaptic currents showing that balance is achieved and maintained. Color scheme as in Fig. 3.1.4: blue denotes mean excitatory input, and red the mean inhibitory input. The total mean input is shown in black. **D:** Steady state distribution of excitatory and inhibitory (inset) firing rates. In equilibrium most neurons fire at low rates, and a few neurons fire at high rates. **E:** Distribution of spike count covariances. The mean is close to zero indicating an asynchronous state. **F:** Same as **E:**, but for spike count correlations, again showing a mean that is close to 0. Relevant simulation parameters: $N = 5000$, $c_{\mathrm{x}} = 0$. Others as in Table 3.1. Figure adapted from Akil et al. [5].

3.4.1 Transient Dynamics of Synaptic Weights

In much of this work we focused on steady state behavior. However, the system of equations (Eqs. 2.26,2.105,2.120) relating rates, covariances, and weights can also be solved iteratively over time, provided the timescales are well separated. Here, we show three examples: weight–dependent Hebbian STDP, Kohonen's rule, and iSTDP. In all three cases, the observed trajectory of synaptic weights closely follows the theoretical predictions (Figs. 3.4.4 **A–C**).

To determine if synaptic weights had reached steady state we sampled 1000 synaptic weights from the whole population (of plastic weights) at times $t = 4000$ sec and $t = 5000$ sec (end of simulation). We then compared the two distributions using the Kolmogorov–Smirnov 2–sample test. We found that in the three networks simulated in Figs. 3.4.1, 3.4.2, and 3.4.3, the distributions were

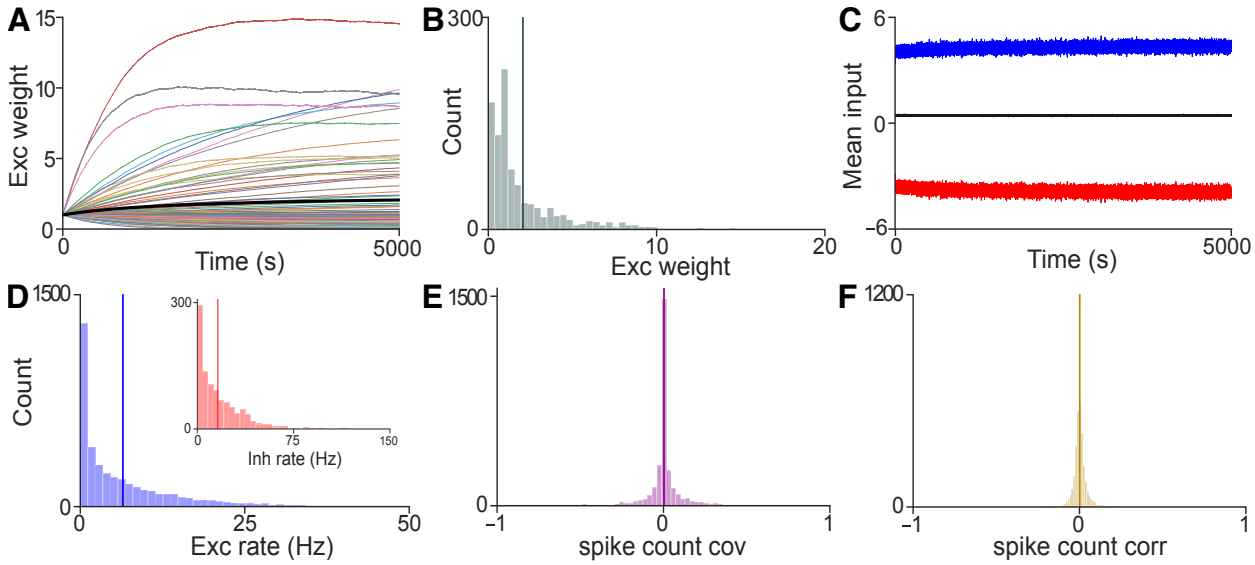

Figure 3.4.2: **Statistics and dynamics of balanced networks undergoing Kohonen's STDP rule. A:** Evolution of mean synaptic weight subject to Kohonen's STDP rule with mean synaptic weight in black. **B:** The distribution of *EE* synaptic weights at steady state is unimodal. **C:** Evolution of synaptic currents in the network. Balance is achieved and maintained. **D:** Equilibrium distribution of excitatory and inhibitory (inset) firing rates. Most neurons fire at low rates, and a few neurons fire at high rates. **E:** Distribution of spike count covariances shows that the mean is close to zero (asynchronous state). **F:** Same as **E:**, but for spike count correlations. The mean is close to zero, suggesting that the network is in an asynchronous state. Color scheme and simulation parameters as in Fig. 3.4.1. Figure adapted from Akil et al. [5].

not distinguishable, and hence had reached equilibrium by the end of the simulation (Fig. 3.4.5 **A–C**).

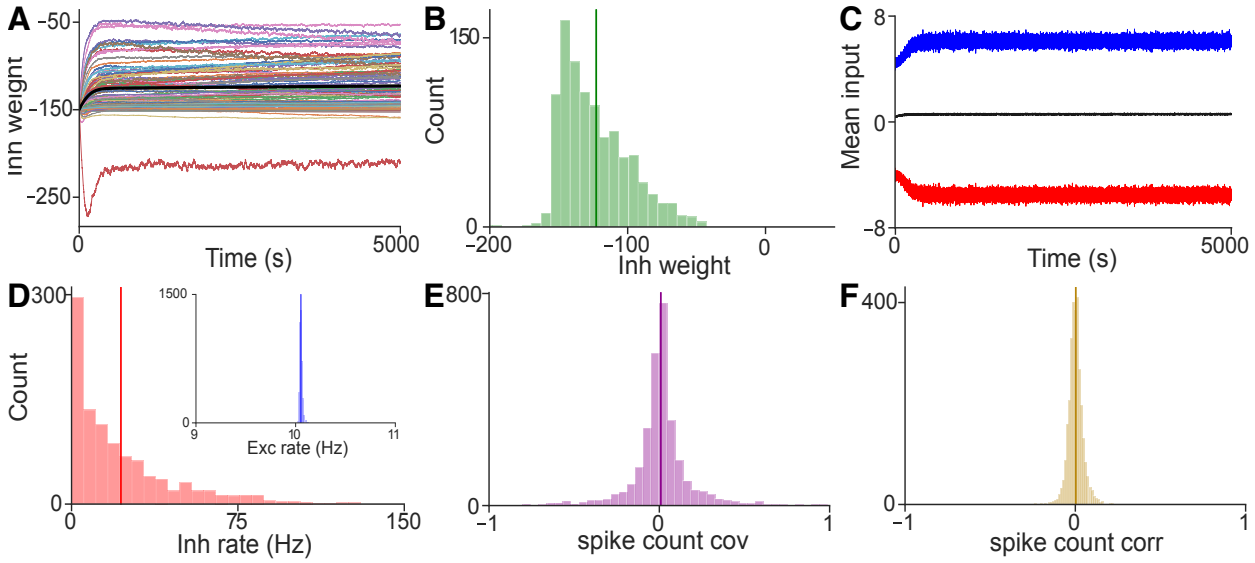

Figure 3.4.3: **Statistics and dynamics of balanced networks experiencing inhibitory STDP. A:** Evolution of *EI* synaptic weights subject to iSTDP. Mean synaptic weight in black. **B:** Distribution of *EI* synaptic weights at steady state. **C:** Evolution of synaptic currents in the network. Balance is achieved and maintained. **D:** Distribution of excitatory (inset) and inhibitory firing rates at equilibrium. Most inhibitory neurons fire at low rates, and a few neurons fire at high rates. All excitatory neurons fire at rates close to the 10 Hz target rate. **E:** Distribution of spike count covariances with mean near zero (asynchronous state). **F:** Same as **E:**, but for spike count correlations. The mean is close to zero, suggesting that the network is in an asynchronous state. Color scheme and simulation parameters as in Fig. 3.4.1. Figure adapted from Akil et al. [5].

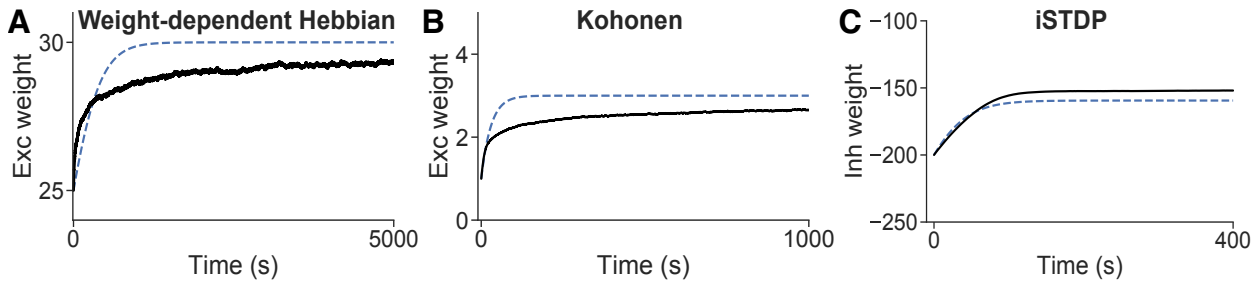

Figure 3.4.4: **Transient dynamics of synaptic weights. A:** Mean synaptic weight in a network subject to Weight–dependent Hebbian STDP. **B:** Same as **A**, but for Kohonen STDP. **C:** Same as **A**, but for inhibitory STDP. In all panels, dashed lines represent the theory (Eqs. (2.26,2.105,2.120)); solid lines represent numerical simulations. Simulation parameters as in Figs. 3.4.1,3.4.2,3.4.3. Figure adapted from Akil et al. [5].

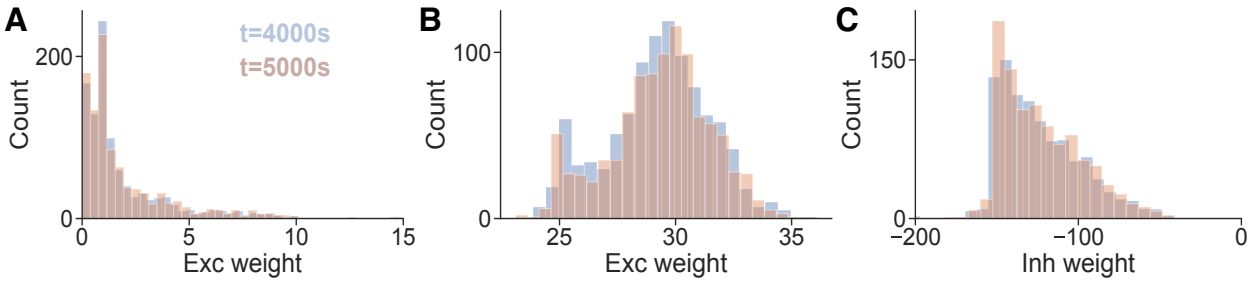

Figure 3.4.5: **Distributions of synaptic weights are stationary.** **A:** Distributions of *EE* synaptic weights for a balanced network with excitatory weights evolving according to Kohonen's rule at times $t = 4000$ s and $t = 5000$ s. The two distributions are not different according to the Kolmogorov–Smirnov 2–sample test ($D_n = 0.029$, p–value= 0.8). Hence, *EE* weights are at steady state. **B:** Same as **A**, but for a network with weights that follow weight–dependent Hebbian STDP showing excitatory weights are in equilibrium ($D_n = 0.045$, p–value= 0.26). **C:** Same as **A**, but for a network where *EI* weights are subject to iSTDP. Inhibitory weights are in equilibrium ($D_n = 0.024$, p–value= 0.94). All simulation parameters as in Figs. 3.4.1,3.4.2,3.4.3. Figure adapted from Akil et al. [5].

3.5 Technical Details of Simulations

Neural networks described in the Section 2 were simulated numerically using the set of parameters shown in Table 3.1. Code is available at https://github.com/alanakil/PlasticBalancedNetsPackage.

Table 3.1: **Summary of simulation parameters.** List of simulation parameters used for initial connectivity, neuron dynamics, and plasticity. Unless otherwise stated, these parameters were used in all numerical simulations presented in this dissertation.

Connectivity		
Parameter	**Value**	**Description**
p_{ab}	0.1	Probability of connection for all $a, b = \mathrm{e}, \mathrm{i}$
j_{ee}/C_m	25 mV	Weight of E to E synapses
j_{ei}/C_m	-100 mV	Weight of E to I synapses
j_{ie}/C_m	112.5 mV	Weight of I to E synapses
j_{ii}/C_m	-250 mV	Weight of I to I synapses

Neuron Model		
Parameter	**Value**	**Description**
C_m	1	Membrane capacitance
g_L	$C_m/15$	Leak conductance
E_L	-72 mV	Resting potential
V_{th}	-50 mV	Spiking threshold
V_{re}	-75 mV	Reset potential
Δ_T	1 mV	'Sharpness' parameter
V_T	-55 mV	Threshold

Plasticity Model		
Parameter	**Value**	**Description**
η_{ab}	10^{-4}	Learning rate of synaptic weights
τ_{STDP}	200 ms	Decay constant of eligibility traces
ρ_{e}	10 Hz	Target rate of E cells
ρ_{i}	20 Hz	Target rate of I cells

Chapter 4

Modelling Spike–Timing Dependent Plasticity in Real Cortical Networks

In Chapter 2, we have presented a computational model of a balanced neural networks, described by a theory that predicts average firing rates, spike train covariances, and synaptic weights. In Chapter 3, we have demonstrated how this theory can be used to extract general principles about the evolution of synaptic weights in balanced networks, and applied the theory to specific, biologically relevant forms of pairwise STDP.

In this chapter, we will apply our framework to neural data collected by Ariana Andrei [6] working in Valentin Dragoi's lab at the University of Texas Health Science Center. The Dragoi lab focuses on information processing in the primary visual cortex and its influence on behavioral performance. One powerful technique used in this lab is optogenetics, which was introduced about 10 years ago. As mentioned before, a great advantage of optogenetic techniques is that specific neuron subtypes can be targeted to be excited or inhibited via the injection of a lentiviral vector and light delivery to ChR2–expressing cells. Such activity modulation of targeted populations of cortical neurons is a powerful way to uncover hidden mechanisms that help better understand the operating dynamical regime of the cortex.

Here, we used our framework to model optogenetic perturbation of pyramidal cells of macaque

V1 cortex, in a similar, but not identical context as in our numerical experiment of Section 3.3. We first described the experimental findings from neural recordings obtained by Andrei et al. [6]. Next, we showed how our framework can qualitatively model experimental findings using a correlated balanced network undergoing iSTDP. We explained why other types of STDP cannot capture the same effects, and provide an intuition as to why iSTDP can.

4.1 Rapid, State–dependent, Compensatory Plasticity Revealed by Functional Connectivity Dynamics in Vivo

4.1.1 Materials and Methods

In the experiments performed by Andrei et al. [6] glutamatergic (excitatory) neurons in the primary visual cortex V1 in three hemispheres of two macaque monkeys were targeted using a lentiviral vector to express channelrhodopsin–2 (ChR2) under the control of an alpha–CaMKII promoter (Fig. 4.1.1 **A**). Activity of cortical populations of neurons was recorded using multichannel laminar electrodes coupled to a fiber–optic for light delivery (Fig. 4.1.1 **A**). After transfection, targeted excitatory cells were excited (increased their firing rate) when exposed to light.

Andrei et al. [6] employed a stimulation protocol similar to previous studies [36, 90] (Fig. 4.1.1 **B**), where upon subject's fixation at the center of a screen, light was delivered in 10 pulses of 10 ms at 35 Hz (total stimulation time was 300 ms). We refer to trials where light is delivered as "laser trials", and to trials where no light is shined as "control trials". Laser and control trials were randomly interleaved throughout the course of the experiment.

When neural activity is recorded over a number of trials, firing rates and spike train correlations can be computed in several different ways. Here, firing rates and correlations were computed in a dynamic way described next [18,125]. As mentioned in Section 2, spike trains of individual neurons can be viewed as sums of Dirac delta functions, $S_i(t) = \sum_j \delta(t - t_{ij})$, where the t_{ij} is the time of the j^{th} spike of neuron i. The data over which neuronal activity is recorded was partitioned into subintervals of $T_{\text{win}} = 200$ ms within a trial and into blocks of block-size = 20 trials, sliding

every 50 ms over time and every trial. To obtain the correlations we fix a partition k, and compute the spike count of each neuron i at each trial t, denoted by n_i^t. We then estimate the spike count correlations between neurons 1 and 2 as:

$$\mathrm{Corr}(n_{1k}, n_{2k}) = \frac{\sum_t (n_1^t - \overline{n}_1)(n_2^t - \overline{n}_2)}{\sqrt{\sum_t (n_1^t - \overline{n}_1)}\sqrt{\sum_t (n_2^t - \overline{n}_2)}} \tag{4.1}$$

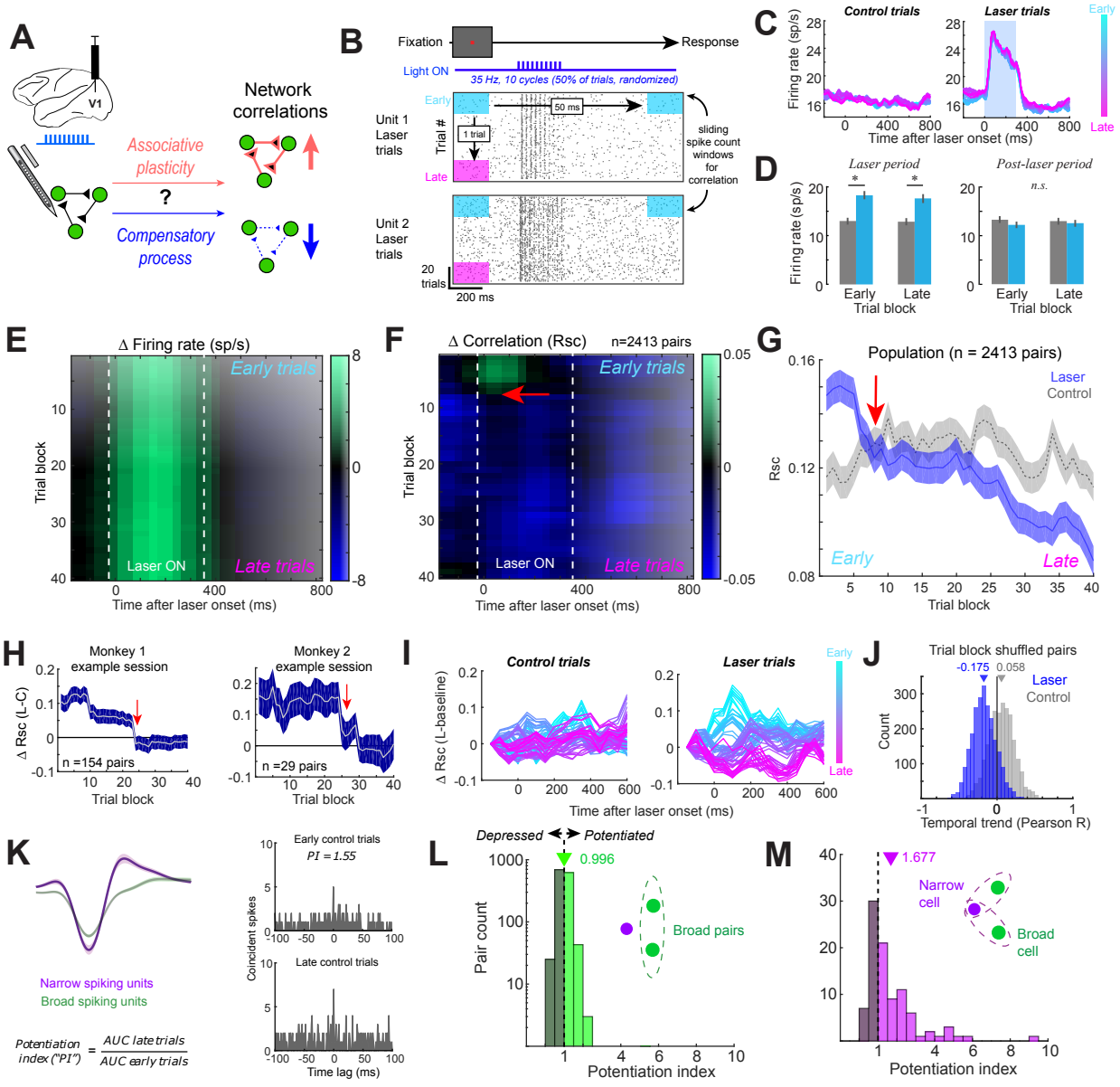

Figure 4.1.1: *See next page.*

Figure 4.1.1 *(previous page)*: **Functional connectivity progressively decreases during repeated optogenetic stimulation trials.** **A:** Repeated optogenetic stimulation of excitatory neurons (green circles) could lead to increases or decreases in functional connectivity between cells, via associative potentiation (red connections), or compensatory processes (blue connections), respectively. **B:** Monkeys performed a contrast detection task, maintaining fixation throughout the trial and reported the presence/absence of a visual stimulus to receive a juice reward. Optogenetic stimulation (35 Hz, 10 pulses at 10 ms width) were randomly interleaved on 50% of trials. Spike count correlations were computed sliding a window of size 20 trials and 200 ms over trials (every trial) and time (every 50 ms), as described in Section 4.1.1. **C:** Population firing rates on laser (right) and control (left) trials across early to late trial blocks. ($n = 310$ units). **D:** Changes in firing rate were not significantly different between early and late trial blocks (first, last 20 trials, respectively). **E:** Firing rate difference between laser and mean control across all blocks of trials for all pairs of laser-responsive neurons. Blocks consist of 20 trial windows, sliding every trial. **F:** Population mean correlation difference (laser minus mean control trials) across all trial blocks. Red arrow shows trial block when correlations reverse from positive to negative. **G:** Correlations on laser (blue) and control (gray) trials across early to late trial blocks, 150 ms after laser onset. Traces and envelope show mean ± s.e.m. **H:** Difference in correlations (laser minus control trials) for example sessions from each monkey, for the time bin 150 ms after laser onset. Line shows mean, envelope shows s.e.m. **I:** Example session showing baseline (200 ms pre–laser) subtracted correlations differences on control (left panel) and laser trials (trials). Same session as panel **H**, left. **J:** Histogram showing temporal trends of shuffled pairs within each trial block for each session on laser (blue) and control (gray) conditions. Arrowheads show median of distributions (both are significantly different from zero, $P < 0.001$ Wilcoxon signed rank test). **K:** Left: A potentiation index (PI) was calculated for all pairs of broad (green) and narrow (purple) single unit pairs on control trials. Right: Example raw cross–correlogram for one narrow-broad waveform pair on early (upper) and late (lower) trials. **L:** Distribution of all PIs for all broad waveform pairs. Arrowhead shows mean. **M:** Same as **L**, but for all broad-narrow waveform pairs. This figure has been adapted from Andrei et al. [6].

where $\overline{n}_i$ is the average spike count of neuron i over all trials in subinterval k. The firing rate of each neuron was obtained by normalizing the spike count n_i^t by the time window and the trial block size: $\text{rate}_i^t = \frac{n_i^t}{T_{\text{win}} \times \text{block-size}}$ (where block-size = 20). See Fig. 4.1.1 **B** for a graphical description.

In both neural experimental data and simulated data, firing rates and spike count correlations were computed as described above and separately for control trials (spontaneous), and for laser trials (evoked).

For each pair of individual units, the cross correlogram (CCG's with lag=200 ms) was computed using all the spikes from control (no laser), no visual stimulus trials. CCG's were calculated for early and late trials, taking the first 20 blocks and last 20 blocks, respectively. Only sessions with

a minimum of 59 trials in this condition were considered, as in the noise correlation calculations. The potentiation index (PI) for each pair was defined as the area under the CCG curve (over 150 ms in the positive lag and 150 ms in the negative lag) on early trials divided by the area under the CCG curve on late trials. Hence, if PI = 1, then there was no difference in the number of joint spikes between early and late trials. If PI > 1, then there were more coincident spikes in later trials (putative potentiation); and lastly, if PI < 1, then there were less coincident spikes in later trials (putative depression). All sessions had consistent waveform characteristics throughout the recording, and contained mixtures of both depressed and potentiated pairs, indicating that differences in early versus late CCG's are not attributable to unstable recordings.

4.1.2 Results

Dynamic reversal of correlations. On one hand, optogenetic activation of ChR2–expressing cells lead to a uniform increase in the activity that persisted over blocks of trials (Fig. 4.1.1 **C,D,E**), with no difference between the amplitude of modulation between early and late trials.

On the other hand, spike count correlations initially increased compared to control trials, but reversed on later blocks of trials (Fig. 4.1.1 **F,G**). The drop in correlations occured due to a change in correlations over laser trials rather than in correlations over control trials (Fig. 4.1.1 **G**). This dynamic reversal of correlations was also seen in individual sessions from each monkey (Fig. 4.1.1 **H**), though the transition point where correlations over laser trials dropped below control happened at different times. Correlation modulation was also evident compared to baseline correlations before laser onset within a trial (Fig. 4.1.1 **I**). Reversal of correlations with time was not due a decrease in overall excitability, as evoked stimulus responses (visual stimulus > 5% contrast) on early versus late trials were not different (P=0.197, Wilcoxon signed rank test control trials, n=42 units).

The decreasing temporal trend persisted even when spike count correlations from individual pairs within each trial block were shuffled prior to calculating the Pearson temporal trend (Fig. 4.1.1 **J**).

In order to understand the mechanism behind changes in functional connectivity, Andrei et

al. divided the experimental, single–unit data into broad (putative excitatory) and narrow (putative inhibitory) units based on waveform width [7, 133] (Fig. 4.1.1 **K**), and quantified the connection strength of simultaneously recorded pairs of neurons by measuring the area under the cross–correlogram on early versus late control trials (Potentiation index, "PI", see Section 4.1.1). Control trials were used to assess stable connectivity changes, without the synchronous activation due to optogenetic stimulation. Putative *EI* cell pairs were significantly potentiated (PI mean 1.68 ± 0.15 s.e.m.), while putative *EE* pairs were overall unchanged (PI mean 0.996 ± 0.007 s.e.m.), (Fig. 4.1.1 **L,M**, $P = 2.94 \times 10^{-4}$, Wilcoxon rank sum test, *EI* versus *EE*, $n = 1476$ pairs). This strongly suggests that rapid compensatory processes are mediated by dynamic changes to *EI* synapses [141, 150]. We will come back to this hypothesis and validate it using our modeling framework in Section 4.2.

Repeated stimulation reverses correlations, independent of laser frequency. One natural question that arises is: how do temporally decreasing correlations depend on laser stimulation frequency? To answer this question, an additional set of experiments were performed with varying laser frequency (10, 20, 35 Hz; Fig. 4.1.2 **A,B**). Each frequency block comprised at least 96 trials (Fig. 4.1.2 **A,B**). Again, firing rates increased uniformly, with no temporal changes over blocks of trials for all frequencies (Fig. 4.1.2 **C**). Similar to the original experiments (Fig. 4.1.1), a significant, dynamic reversal of correlations was observed (Fig. 4.1.2 **D,E,F**); implying that temporal changes in correlations do not depend on stimulation frequency, but are instead caused by stimulation cycles.

Rest experiments: no temporal trend in correlations. Synaptic plasticity and measures of spike count correlations are highly sensitive to cortical states [34, 142]. Therefore, similar experiments to those presented in Fig. 4.1.1 were carried out, but with the animal in a resting state, in order to understand under which conditions a dynamic reversal of correlations may take place. For this set of experiments, daily recording sessions were split based on the two conditions: (1) the detection task, (Fig. 4.1.3 **A**, awake condition); and (2) a rest period (Fig. 4.1.3 **E**), during which the lights were turned off and the animal was allowed to rest while laser and control trials took place. Light was delivered under the same protocol as in Fig. 4.1.1 **A,B**.

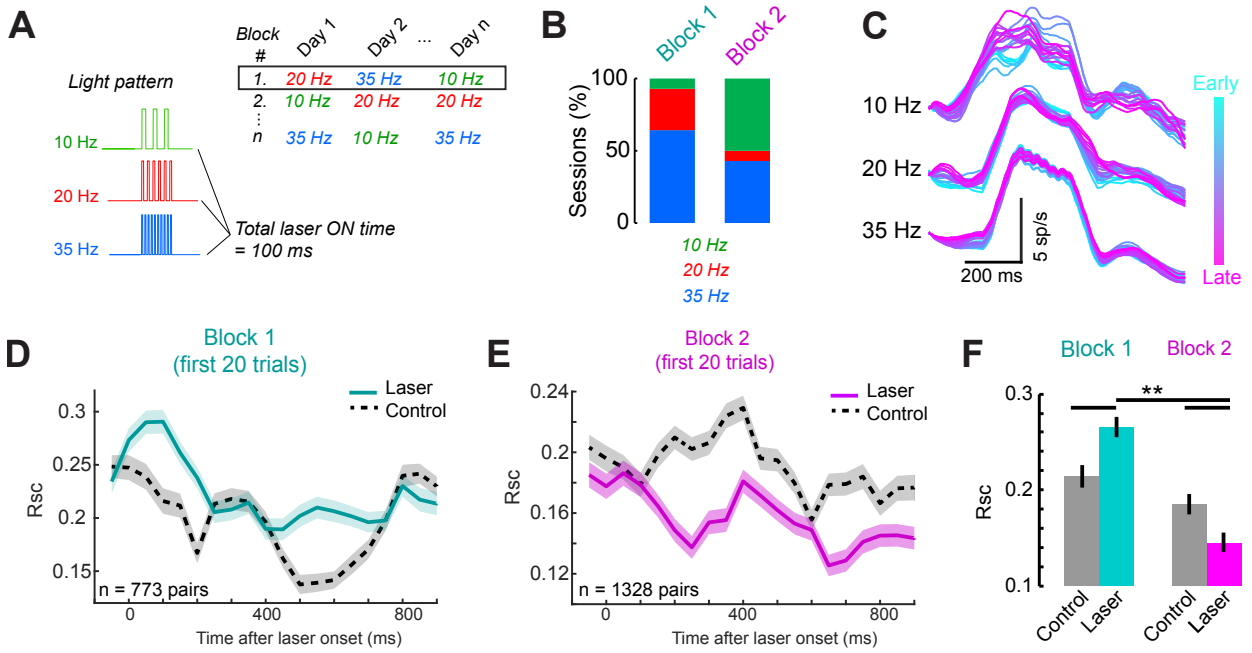

Figure 4.1.2: **Stimulation cycles, not frequency, leads to temporally decreasing correlations. A:** Follow up experiments using different stimulation frequencies. On individual recording days (upper right, columns) monkeys performed multiple iterations of the contrast detection task, with laser stimulation delivered at different frequencies (upper right, rows; left panels). **B:** Distributions of frequencies included in Block 1 (left, $n = 11$ sessions) and Block 2 (right, $n = 11$ sessions). **C:** All laser stimulation frequencies lead to significantly increases in firing rates of neural populations consistently across early to late trial blocks (20 trials per block). **D:** Correlated variability of neuron pairs during the first 20 trials of the first trial block of each recording day on laser (blue) and control (black) trials. **E:** Same conventions as **D**, but for the first 20 trials of the second frequency block across all days. **F:** Correlations are significantly increased in the first block of trials on laser versus control trials. On the second block of trials, laser stimulation induces a significant decrease in correlations ($** P = 2.010^{-12}$, Kruskal–Wallis test, d.f.=3, post hoc signed rank test for within block comparisons, rank sum test for across block comparisons). This figure is adapted from Andrei et al. [6].

During the rest period, there were no behavioral requirements and thus inter–trial intervals were minimal. A common way of measuring a resting state is by analyzing the time series of the local field potential (LFP, average membrane potential of neurons near an electrode). In particular, measurements of the low–to–high frequency power ratio of the LFP [13] confirmed the subject was in a resting state.

Recall that in the awake condition, firing rates do not change temporally over blocks (Fig. 4.1.1 **C,D,E**), but correlations dynamically reverse (Fig. 4.1.1 **F,G,H,I**). This was also true in this set of

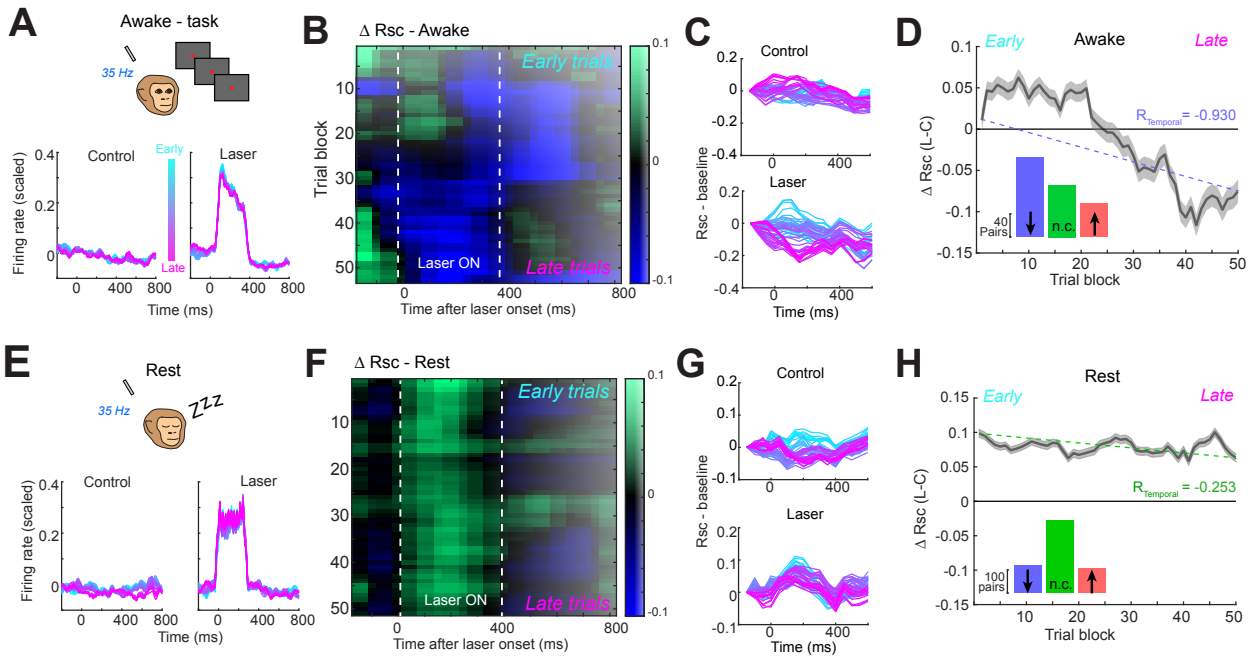

Figure 4.1.3: **Dynamic correlation reversal depends on overall behavioral state.** Follow up experiments using 35 Hz optogenetic stimulation during recordings of neural populations during two states; awake detection task **A–D** and rest **E–H**. **A, E:** Optical stimulation strongly increased firing rates of neural populations during the contrast detection task component **(A)** and during the rest **(E)** component. **B, F:** Correlation difference between laser and mean control trials during the awake component (**B**, $n = 269$ pairs) and the rest component (**F**, $n = 600$ pairs). Red arrow shows the transition from increased to decreased correlational states only during the awake condition. **C, G:** Since there is no meaningful trial structure during the rest condition, we also present the change in correlated variability during laser stimulation compared to baseline state (200 ms window, prior to laser stimulation) across stimulation blocks during the awake **(C)** and rest **(G)** conditions. **D, H:** Difference in correlations between laser and control trials across trial blocks (200 ms after laser onset) during awake **(D)** and rest **(H)** components. Dashed line connects first and last trial blocks. Insets show counts of pairs with decreasing (blue), increasing (red), or no change in temporal trends (green). This figure has been taken and modified from Andrei et al. [6].

experiments in the awake state (Fig. 4.1.3 **A,B,C,D**). However, in the rest condition, correlations in laser trials remained high compared to control in all blocks, with no temporal change over blocks (Fig. 4.1.3 **F,G,H**). This effect persisted when comparing to baseline pre–laser correlations (Fig. 4.1.3 **G**). The time interval between the end of the awake and start of the rest components was 8 minutes 33 seconds ± 1 minute 29 seconds (median ± s.e.m.), hence there was enough time for optogenetically–induced dynamic correlation changes. Sleep is known to be associated with a net reduction in excitatory synaptic weights [142]. Hence any changes in excitatory synapses that

may have occurred during the awake condition are likely undone in the resting state. Interestingly, neurons seemed to be more excitable in the resting state, even though the electrode position was not changed compared to awake sessions. Lastly, the inhibitory tone in the awake state is typically higher than in resting states [21, 57]. This suggests that some form of optogenetically–evoked inhibitory plasticity may lead to reported temporal changes in correlations.

4.2 Dynamic Reversal of Correlations in Correlated Balanced Networks under iSTDP

4.2.1 Materials and Methods

Network model. To try to understand the mechanism behind temporal changes in functional connectivity observed in these experiments, we considered a balanced, recurrent computational network model (Fig. 4.2.1 **A**) composed of excitatory (E) and inhibitory (I) neurons ($N = 10,000$) with plastic synapses, as defined in Section 2. Optogenetic stimulation was modeled as an increase in firing rates in 50% of E neurons, and was patterned to mimic the experimental trial structure.

Experiment structure. Following the experimental protocol as closely as possible, in the numerical experiment, each trial was 1.4 seconds long. Half of the excitatory neurons were stimulated by providing input in pulses of width 10 ms, at 35 Hz, and 10 cycles. The laser onset was at 500 ms and lasted for 300 ms, only in half of the trials (laser trials). Each numerical experiment was composed of 60 laser trials randomly interleaved with 60 control trials.

Comparison of experiments with simulations. Firing rates and spike count correlations were computed in the same way described in Section 4.1.1.

Simulation of a resting state. Cortical networks experience a change in dynamics when a subject falls asleep [142]. Even though balance is maintained, synapses undergo net potentiation in wake and net depression during rest/sleep [142]. We modeled the dynamics of a cortical network during periods in which the subject is resting using the same balanced network model, but with weaker synaptic connectivity compared to that used to model the awake state [142]. Thus, to enter

the resting state, all synapses were downscaled by a factor $0 < f \leq 1$:

$$J_{\text{rest}} = f \cdot J_{\text{awake}}, \tag{4.2}$$

where J_{rest} denotes the full connectivity matrix of the network in the resting state, and similarly J_{awake}, the full connectivity in the awake state.

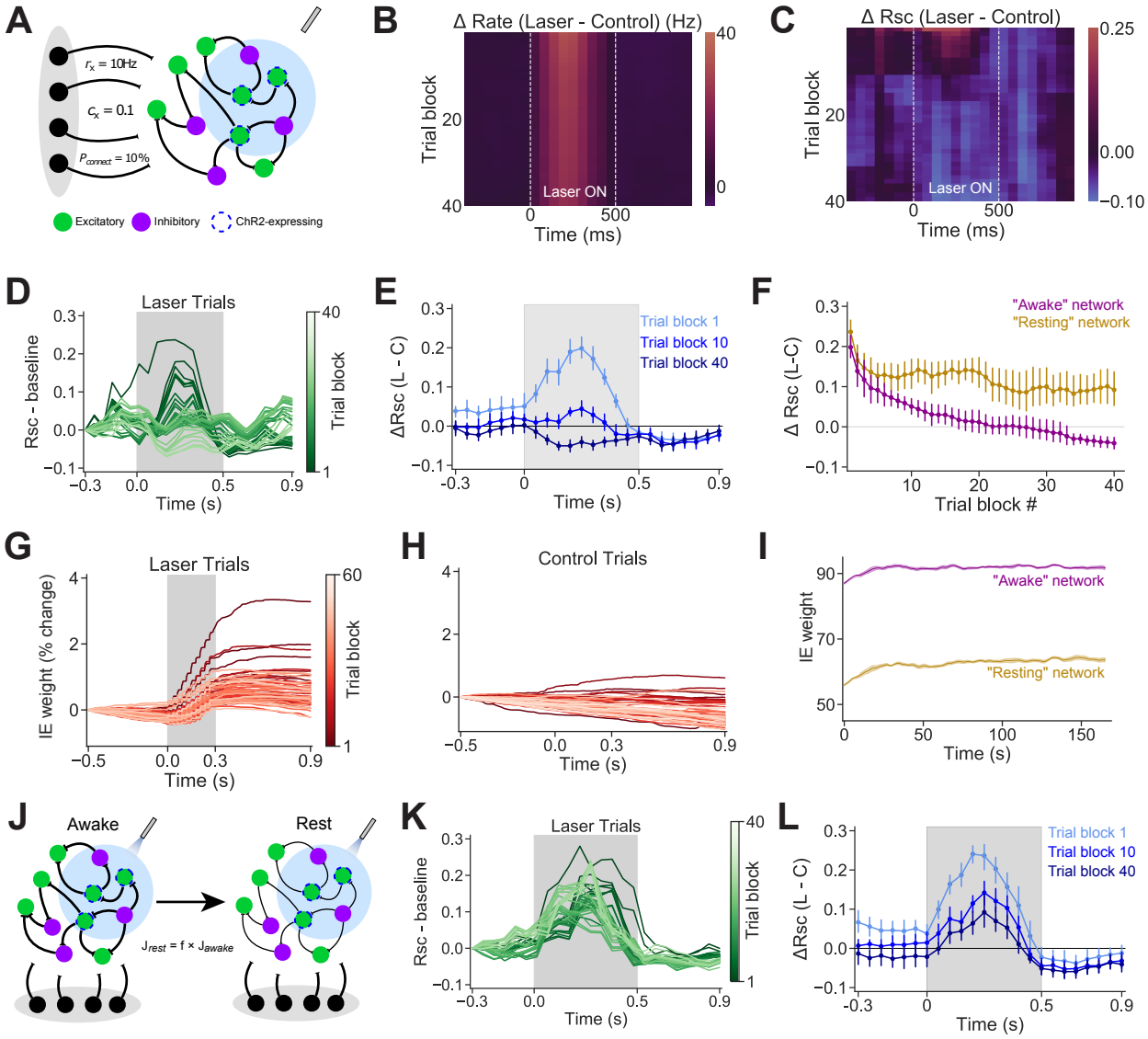

Figure 4.2.1: *See next page.*

Figure 4.2.1 *(previous page)*: **Dynamic reversal of functional connectivity modeled by inhibitory plasticity. A:** A recurrent network of 8000 E and 2000 I cells is driven by an external feedforward layer of Poisson neurons that connects with 10% of all recurrent neurons with rate 10 Hz, and pairwise correlations 0.1. Only inhibitory-to-excitatory synapses were plastic, and we assumed that 4000 E neurons expressed ChR2 and were driven by laser stimulation during laser trials. **B:** Change in mean firing rate (laser minus control) over time and trials. **C:** Change in mean pairwise noise correlations (laser minus control) over trials. **D:** Change in mean pairwise noise correlation during laser trials compared to the first time bin over the time course of a trial. **E:** Change in mean pairwise noise correlations averaged over ChR2-expressing pairs and over realizations of the connectivity matrix. Error bars indicate 95% confidence interval. Light blue indicates block #1, blue is block 10, and dark blue represents block #40. **F:** Change in mean noise pairwise correlations at the center time bin of the laser stimulation period ($t = 250$ ms) across trials averaged over 9 realizations of the connectivity matrix. Error bars indicate 95% confidence interval. Golden line corresponds to the network in the rest state (f=0.4). Magenta line corresponds to awake state (f=1). Dynamic reversal of noise correlations is robust to randomness in the network's connectivity. **G:** Percent change of mean synaptic weight compared to baseline (first time bin) during laser trials. **H:** Same as **G**, but for control trials. **I:** Evolution of mean EI synaptic weight over the course of the entire experiment under the awake and rest conditions. Shaded area indicates standard error. Inhibitory weights were potentiated due to repeated activation of E cells. Inhibitory weights in the rest state are weaker than in the awake state due to synaptic downscaling. Panels **B**–**I** correspond to the awake condition. **J:** Same as **A**, but during rest. All synapses were downscaled by a factor of $f = 0.4$. **K:** Same as **D**, but during rest. Laser trials exhibited larger mean pairwise noise correlations than control trials during the period of stimulation and across blocks of trials. **L:** Same as **E**, but during rest. Dynamic changes in correlations in the rest state were robust to random perturbations in network connectivity. All mean firing rates and mean noise correlations were computed over cells that expressed ChR2. This has been apapted from Andrei et al. [6].

Forms of STDP tested. We tested two types of plasticity rules: "Weight–dependent" Hebbian STDP [14, 92] modulating the strength of excitatory synapses, and inhibitory STDP [141] modulating the strength inhibitory synapses.

Excitatory synaptic weights, EE and IE, followed classical Hebbian STDP and hence evolved according to,

$$\frac{dJ_{jk}^{ae}}{dt} = \eta_{ae}\big(J_{\max} x_k^e S_j^a - J_{jk}^{ae} x_j^a S_k^e\big) \tag{4.3}$$

where η_{ae} is the learning rate that defined the timescale of changes in excitatory synapses, $J_{\max,\mathrm{ae}}$ is a soft upper bound for the excitatory synaptic weights that prevents divergence of weights to infinity, and $a = e, i$. Pre–synaptic spikes followed by post–synaptic spikes potentiate the synapse; and spikes occurring in the reverse order depress the synapse.

Inhibitory synaptic strenghts, *EI* and *II*, evolved as [141]:

$$\frac{dJ_{jk}^{ai}}{dt} = -\eta_{ai}\frac{J_{jk}^{ai}}{J_0^{ai}}\left[(x_j^a - \alpha_a)S_k^i + x_k^i S_j^a\right] \tag{4.4}$$

where α_a determines a target firing rate for a neurons, and $a = e, i$. Note that, as in Chapters 2 and 3, we used a slightly modified version of the original inhibitory plasticity rule to prevent weights from changing signs (see Section 2.12). As mentioned before, this rule acts as a homeostatic mechanism since it controls excitation by pushing the activity of the cells in population $a = e, i$ to parameter–dependent firing rates, $\rho_a = \frac{\alpha_a}{2\tau_{STDP}}$.

4.2.2 Computational Results

Dynamic reversal of correlations under iSTDP. We initially assumed that inhibitory–to–excitatory synapses were potentiated in response to stimulation (Fig. 4.2.1 **G,H,I**) in order to maintain E firing rates near a target rate (see inhibitory plasticity in Table 2.1). Optogenetic stimulation lead to an increase in firing rates uniformly across trials (Fig. 4.2.1 **B**). Moreover, we found a dynamic reversal of correlations following repeated activation across early to late trials (Fig. 4.2.1 **C,D,E,F** – magenta line), consistent with experimental findings (Fig. 4.1.1 **F,G,H,I**). Repeated stimulation increased the firing rates of targeted cells. As a consequence, inhibitory neurons received more excitation and inhibitory–to–excitatory weights were potentiated (Fig. 4.2.1 **G,I**), leading to an increase in inhibitory feedback to ChR2–expressing cells (Fig. 4.2.2 **F**) which decorrelated their responses over the timecourse of the experiment [58, 119, 125].

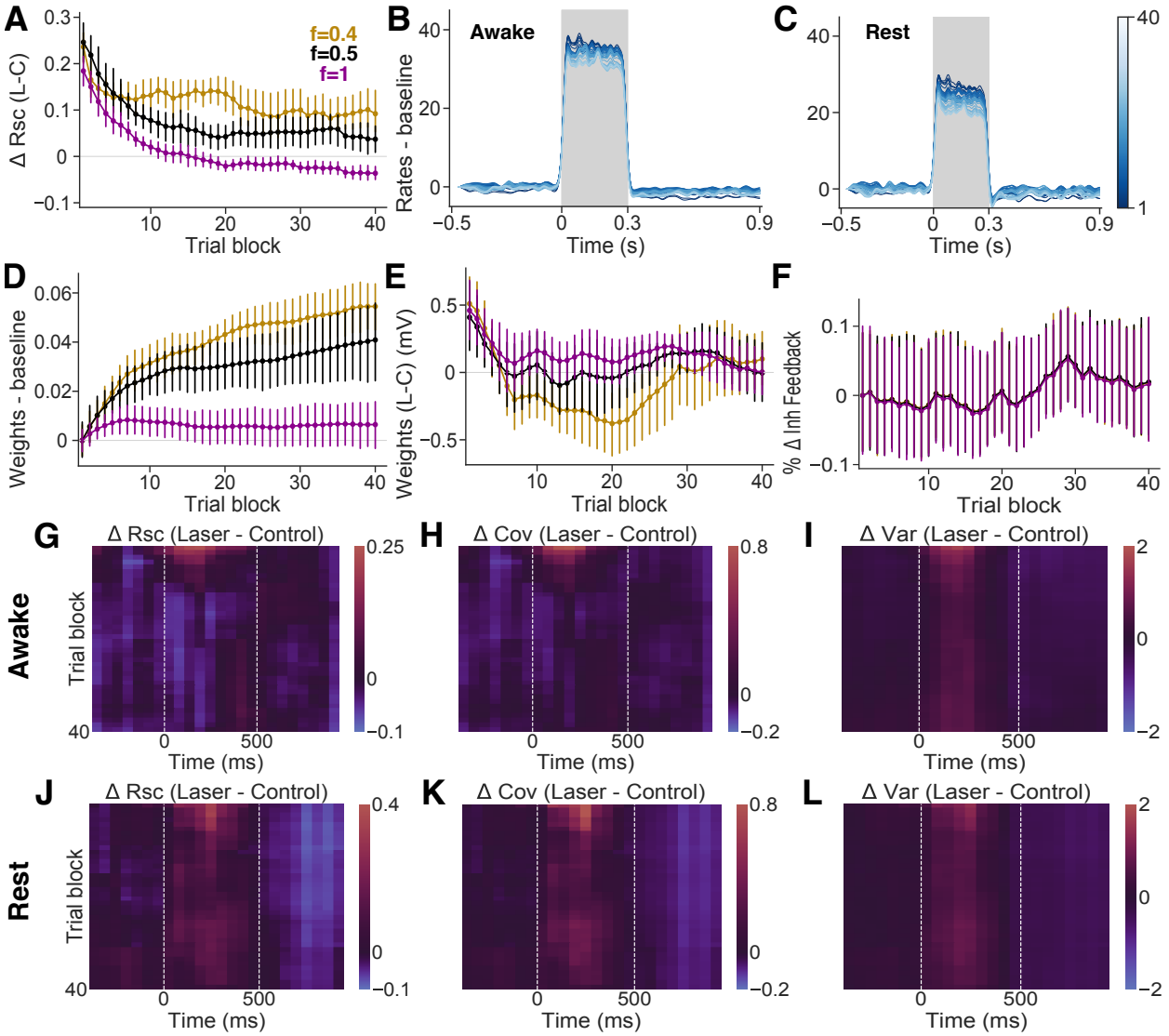

Figure 4.2.2: **Transition of network model from awake to rest condition.** **A:** Drop in mean correlations at center time bin of laser period for networks with varying synaptic weight scaling factors. **B:** Peri–stimulus time histogram (PSTH) of stimulated cells in awake condition. **C:** PSTH of stimulated cells in rest condition. The response is stronger than in the awake condition, consistent with experimental findings. **D:** Change in synaptic weights compared to baseline (first block) and computed over a window 200 ms in length, and over 20 trials. **E:** Evolution of the difference in synaptic weights in laser vs control trials for different values of the scaling factor f. **F:** Mean inhibitory input to ChR2–expressing neurons increased due to potentiation of inhibitory weights. The inhibitory input was also computed over 200 ms window, and over 20 trials. **G:** Change in mean spike count correlations in a sample simulation of the numerical experiment in the awake condition. Correlations over the laser period drop below control due to changes in synaptic weights. **H–I:** Same as **G**, but for spike count covariances and variances. Changes in correlations are due to changes in covariability, not in variance. **K–L:** Same as **G–I** but for the rest condition ($f = 0.4$). Correlations, covariance, and variance remain high throughout the laser period. This figure has been taken and modified from Andrei et al. [6].

Correlations have no temporal trend under rest. To model experimental outcomes under the rest condition, we downscaled all synapses in the recurrent network model by a factor of $f = 0.4$ [142] (see Fig. 4.2.1 **J** and Section 4.2.1 for more details). After this change, potentiation of *EI* synapses (Fig. 4.2.1 **I** – gold line) subtly decorrelated responses across trial blocks (Fig. 4.2.1 **K,L**, 4.2.1 **F** – gold line), but correlations on laser trials remained greater than control trials throughout all blocks. Both observations are consistent with experimental findings (Fig. 4.1.3 **F,G,H**), and are due to decreased inhibitory feedback from weaker *EI* synapses during rest compared to the awake state (Fig. 4.2.1 **I** – compare magenta and gold lines).

Transition from awake to rest and role of inhibitory feedback in decorrelation. We simulated a balanced network, as described above, undergoing the same protocol used in experiments, and using scaling factors $f = 0.4, 0.5, 1$ (where $f = 1$ is the awake state). As f was decreased, the drop in correlations becomes smaller. Keeping with experiments in the literature that suggest a downscale of synaptic weights during sleep [142], we modeled the resting state using a scaling factor of $f = 0.4$. Correlations then remained larger in laser trials compared to control trials (Fig. 4.2.2 **A**). This reversal was due to stimulation of ChR2–expressing cells that respond (Fig. 4.2.2 **B,C**) and induce potentiation of *EI* synapses (Fig. 4.2.2 **D,E**), effectively amplifying inhibitory feedback to stimulated cells (Fig. 4.2.2 **F**), decorrelating their response (Fig. 4.2.2 **G**). Both in the awake and rest state, the changes in correlations were a consequence of changes in covariability rather than variability (Fig. 4.2.2 **G–L**).

Correlation reversal is independent of stimulation frequency. We also simulated the same balanced network with $f = 1$ in parallel, applying different frequencies of laser stimulation (Fig. 4.2.3). We found that the dynamic reversal of correlations was present for laser frequencies between 10–100 Hz (Fig. 4.2.3 **A,B**), even though cells responded more weakly to lower frequencies (Fig. 4.2.3 **C**). The underlying mechanism remains the same: stimulation induces plasticity (Fig. 4.2.3 **D,E**), and an increase in inhibitory feedback decorrelated the responses of excitatory cells in the network (Fig. 4.2.3 **F**).

Other forms of STDP do not exhibit correlation reversal. Other forms of STDP in every

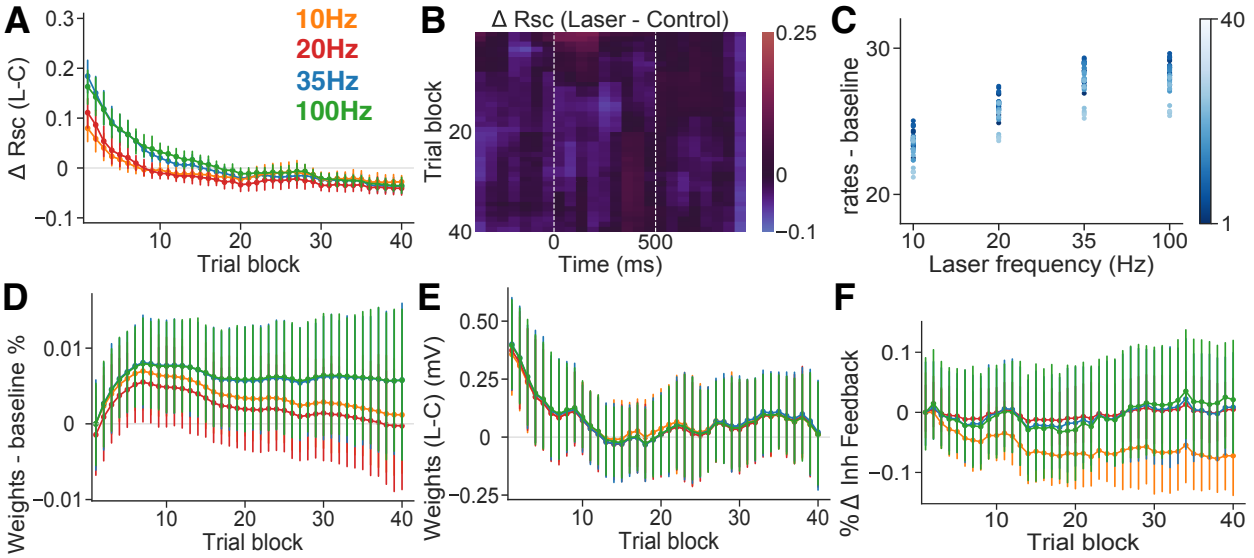

Figure 4.2.3: **Dynamic reversal of correlations is independent of laser stimulation frequency in network model. A:** Drop in mean correlations at center time bin of laser period for networks with varying laser stimulation frequencies. **B:** Change in mean spike count correlations in a sample simulation of the numerical experiment in the awake condition with laser frequency 10 Hz. Correlations over the laser period drop below control due to changes in synaptic weights. **C:** Firing rates of ChR2–expressing neurons averaged over the laser period. Higher laser frequencies evoke stronger responses. **D:** Change in synaptic weights compared to baseline (first block) and computed over a window of 200 ms and 20 trials. **E:** Evolution of difference in synaptic weights in laser vs control trials for different values of scaling factor f. **F:** Mean inhibitory input to ChR2–expressing neurons increases due to potentiation of inhibitory weights. Inhibitory input, defined as the inhibitory current scaled by the corresponding synaptic weight to ChR2–expressing neurons, is also computed over a time window of 200 ms and 20 trials. This figure has been taken and modified from Andrei et al. [6].

synapse were tested. We found that the dynamic reversal of correlations (as reported in our experiments) only occurred when *EI* synapses were plastic (Fig. 4.2.4 **A**), even though all types of plasticity modulated synaptic weights and correlations in different ways (Fig. 4.2.4 **A,B,C**).

Optogenetic techniques are powerful tools that can uncover hidden mechanisms not otherwise evident in neural networks [58]. In this case, repeated optogenetic stimulation leads to changes in functional connectivity seen via noise correlations in a local cortical network in the primary visual cortex V1. We have shown that using a balanced network, which can generate cortical–like activity in certain dynamical regimes, we can unveil the mechanism behind changes in functional

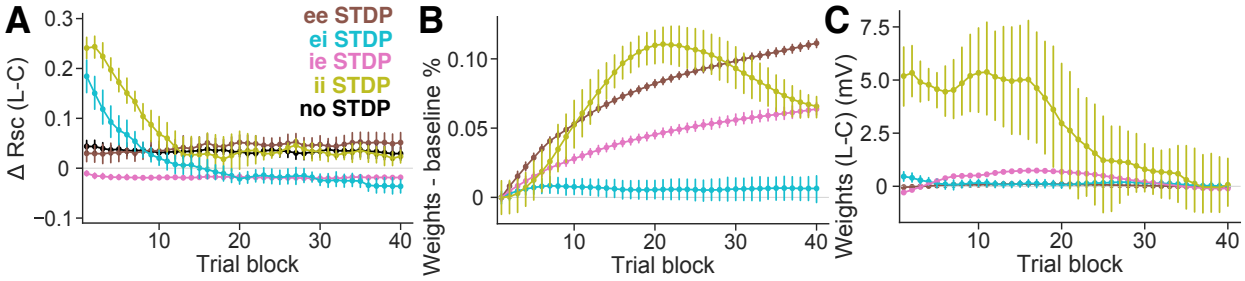

Figure 4.2.4: **Dynamic reversal of correlations only occurs when inhibitory connections are plastic in network model. A:** Drop in mean correlations at center time bin of laser period for networks with varying laser stimulation frequencies. **B:** Change in synaptic weights compared to baseline (first block) and computed over a window of 200 ms and 20 trials. **C:** Evolution of difference in synaptic weights in laser vs control trials for different values of scaling factor f. This figure has been taken and modified from Andrei et al. [6].

connectivity: homeostatic inhibitory plasticity. In particular, we show that inhibitory plasticity is a plausible mechanism, that its effect is largest in networks in an awake state, and that the timescale of such iSTDP is in line with that of RCP's. Hence, our framework proves not only useful for deeper theoretical understanding of dynamics in plastic, balanced networks (as seen in Chapters 2 and 3), but also for making sensible experimental predictions that were confirmed by real neural recordings and help further understand the complex dynamics of the mammalian cortex.

Chapter 5

Discussion

We developed an analytical framework to predict how a general class of second order STDP mechanisms affects the evolution of synaptic weights between neurons, and the dynamics of balanced networks. We found that the balanced state is generally maintained under synaptic weight changes, as long as the rates remain bounded. Additionally, we showed that correlations in spiking activity can introduce a small shift in the steady state, and can change how quickly the fixed point is reached.

One of the most important issues in understanding neural dynamics is establishing conditions under which the network remains active, yet stable as synaptic weights change. The theory we developed can help us address these questions, but it does have limitations. Since we used a mean–field approach, we can only capture first moments. While mean weight stability may not imply stable network dynamics (consider the case when weight variance diverges in Classical Hebbian STDP in Section 3.1.1), instability in the mean weights does imply that the network is also unstable.

As mentioned above, our theory can be used to show that small modifications to weight updates can stabilize different STDP rules. The question remains whether classical, unconstrained Hebbian EE plasticity can be stabilized through an interaction with STDP rules at different synapses. For instance, Litwin–Kumar and Doiron used a triplet voltage STDP rule that was stabilized by hard constraints and weight normalization to produce neural assemblies [87]. This rule by itself lead to

stable but pathological behavior, and they introduced iSTDP to restore a balanced, asynchronous network state. While such voltage–based triplet rules are outside the scope of the present study, we could use extensions of the mean–field theory to describe the impact of second and higher order moments on the evolution of weights, and network dynamics [99]. Our theory suggests that classical pairwise Hebbian STDP cannot be stabilized by other STDP rules such as iSTDP.

In the tight balance regime, large excitatory and inhibitory inputs cancel on average [56], resulting in a fluctuation–driven state exhibiting irregular spiking. This cancellation is achieved when synaptic weights are scaled by $1/\sqrt{N}$ and external input is strong [112,136–138]. Our main assumption was that synaptic weights change slowly compared to firing rates. As this assumption holds generally, we believe that our approach can be extended to other dynamical regimes. For instance supralinear stabilized networks (SSNs) operate in a loosely balanced regime where the net input is comparable in size to the excitatory and inhibitory inputs, and firing rates depend nonlinearly on inputs. Balanced networks and SSNs can behave differently, as they operate in different regimes. However, as shown in [4], SSNs and balanced networks may be derived from the same model under appropriate choices of parameters. In other words, the tight balanced solution can be realized in an SSN, and SSN–like solutions can be attained in a balanced network. This suggests that an extension of our theory of plasticity rules to SSNs should be possible.

We obtained a mean–field description of the balanced network by averaging over the entire inhibitory and excitatory subpopulation, and a single external population providing correlated inputs. As shown in Section 3.3, the theory can naturally be extended to describe networks consisting of multiple physiologically or functionally distinct subpopulations, as well as multiple input sources.

As just mentioned, the theoretical framework we presented is flexible, and can describe more intricate dynamics in circuits containing multiple inhibitory subtypes, and multiple plasticity rules, as well as networks in different dynamical regimes. Moreover, the theory can be extended to plasticity rules that depend on third order interactions [49, 108], such as the BCM rule [15]. This may produce richer dynamics, and change the impact of correlations.

The present theory relies on a separation of timescales between spiking dynamics and weight

changes. Such timescale separation is supported by a number of experiments [14,40,74,92,150]. We show in Section 2.6.3 that reducing this timescale separation, and hence increasing weight updates leads to deviations from the theory, and can result in network instability.

In mammalian brains, timescales of weight changes may not always be separated from rate and correlation timescales. The size and timescale of weight updates is likely to depend on many factors that can modulate STDP, such as spiking patterns, synapses type, brain area, network state, neuromodulation, and others. Separation of timescales may not be pronounced in certain non–cortical areas, such as the hippocampus, which can be rapidly modified [150]. For example, Petersen et al., 1998 and Froemke et al., 2006 found significant changes in putative synaptic weights over short timescales in hippocampal CA1/CA3 slices [106] and in visual cortical slices subject to multispike pre– and post–synaptic bursts [39], respectively. However, it is possible that the rate of change of synaptic weights may be overestimated *in vitro* [150].

How is our separation of timescales assumption affected when rapid compensatory processes are needed for homeostasis, given that experiments show that homeostasis is a process that is even slower than the timescale of STDP? Experimental evidence suggests that homeostatic processes can take hours or days [38,40,63,82,130,140,148–150]. On the other hand, theoretical models show that synaptic plasticity can be unstable in the presence of such slow homeostasis, and needs to be coupled with rapid compensatory processes such as inhibitory STDP [149,150]. The separation of timescales in our theory still puts synaptic dynamics on the "fast" side of the spectrum, as it separates network dynamics that occur over milliseconds from weight dynamics that take place over seconds or minutes. Hence, the assumption of timescale separation is still valid in our implementation of homeostatic inhibitory plasticity.

We showed in Section 3.2, that the mean–field description cannot capture the effect of some second order STDP rules as synaptic competition can correlate synaptic weights and pre–synaptic rates. We have shown that this can lead to different initial weight distributions converging to different equilibria. This can be interpreted as the maintenance of a previous state of the network over time.

However, a natural question that arises is why do correlations between weights and pre–synaptic rates only seem to play a role in iSTDP? In the examples of excitatory STDP we analyzed (Kohonen's rule and weight–dependent Hebbian rule), weights at equilibrium are determined by other parameters (Weight–dependent Hebbian rule) or rates (Kohonen's rule). Therefore weights are updated until those steady state values are achieved, yielding values independent of initial conditions. On the other hand, in the case of the inhibitory plasticity rule, inhibitory weights at equilibrium are determined by the firing rates alone. Since the firing rate vectors are lower–dimensional than the weight matrices, the equilibrium solution does not fully determine the weight matrices. This is shown in Fig. 3.2.2 **C** inset, where different distributions of weights can result in the same equilibrium firing rate when weights and pre–synaptic rates become correlated.

Generally, in plastic networks, correlations between weights and other features such as in–degrees, or out–degrees can emerge [139]. We have shown how the theory can capture the case in which synaptic weights and pre–synaptic rates are correlated. While we were not able to find analytical expressions for these correlations, we showed that a second–order correction is sufficient to explain the observed dynamics. Eventually, the mean–field theory would need to be extended to account for higher order network motifs and their potential correlations with synaptic weights and firing rates. This might be possible by extending our approach, but we leave these extensions for future work.

We have assumed that connection probabilities are homogeneous which translates to a narrow distribution of in–degrees. Cortical networks are heterogeneous, and a broad distributions of in–degrees can break the classical balanced state [83, 110]. Balance can be restored with the introduction of homeostatic plasticity [83], or by including heterogeneous out–degrees correlated with in–degrees [110]. As we mentioned previously, in such cases our theory would need to be extended to account for possible emerging correlations between weights and in-degrees or out–degrees. We relegate such extensions to future work.

We have shown that different plasticity rules can result in distinct firing rate distributions in different subpopulations. As shown by Mongillo et al. this can result in an increase or decrease in

sensitivity of activity patterns and memories to perturbation of different synapse classes [96].

A number of mathematical theories have been proposed to describe the coevolution of weights, rates, and the structure of correlations under STDP in recurrent neural networks [23, 44–47, 97, 100, 111]. All of these approaches require knowledge of neurons' transfer functions (f-I curves and/or correlation susceptibility functions). Often neurons are assumed to be Poissonian, and their responses to inputs (f-I curves) are prescribed [23, 44–47, 97, 111]. Other work [100] uses Fokker–Planck techniques to compute transfer functions. These approaches rely on an assumption that the input to each neuron is relatively weak or dominated by Gaussian white noise [85, 113] and efficient, direct Fokker–Planck approaches are not available for two–dimensional integrate–and–fire models such as those with adaptation currents, though one–dimensional approximations have been derived [65, 114]. Some previous work [100] also assumes that STDP curves are approximately anti–symmetric, *i.e.*, there is a cancellation between the positive and negative parts of the curves (as in Fig. 2.5.1 **A**).

Our approach uses balanced network theory to avoid the computation of neurons' transfer functions. As such, the resulting theory does not require an assumption of weak synaptic interactions or dominant Gaussian white noise input, but can be applied to networks with highly non–Gaussian, temporally correlated input (such as the networks in the correlated state considered here). Moreover, the balanced network theory we used is applicable to a variety of neuron models, including those with adaptation currents [11, 33]. Our extended balanced network theory also applies for non–anti–symmetric STDP curves (as in Fig. 2.5.1 **B–F**), which were not captured in previous studies [100]. However, the balanced network theory used in our approach relies on large N asymptotics, which yielded accurate approximations for $N \sim 10,000$ in our case (Fig. 3.1.4), but become less accurate for smaller networks. Moreover, our approach would not be accurate for modeling neural circuits that do not exhibit excitatory–inhibitory balance, as observed in some disease states, developmental stages, and sub–cortical neuronal networks. Finally, we used a mean–field approach that only yields approximations to population–averaged firing rates, synaptic weights, and covariances, while other approaches [23, 44–47, 97, 100, 111] give approximations to these quantities at

the level of individual neurons. Despite these limitations, our analytical approach was sufficient for answering the questions related to the interaction between excitatory–inhibitory balance, correlated neuronal activity, and plasticity that we considered.

We found that even in the correlated state, when the network receives temporally correlated input, changes in synaptic weights are dominated by firing rates, with correlations playing a secondary role (see Fig. 3.1.6 **A,B**). These findings are in agreement with previous work on STDP already mentioned before [52,100]. Results by Ocker et al. [100] were obtained in recurrent neural networks in different dynamical regimes and under different assumptions (see above for more details), while Graupner et al. [52] used networks of two neurons with varying natural firing patterns.

Partial stimulation of a population of E neurons has been shown to break balance due to the inability of the network in cancelling inputs when weights are static [33]. Ebsch et al. showed how classical balanced network theory can be modified to account for effects of input perturbations that break the classical balanced state [33]. Vogels et al. [141] (in addition to subsequent studies [2, 19, 63, 87, 130, 144]) showed empirically using simulations that inhibitory iSTDP can restore balance. We here provide a theoretical framework that describes the evolution of rates and weights before, during, and after a perturbation that breaks balance.

In addition to our theoretical analysis of an optogenetic perturbation of a subset of excitatory cells in Section 3.3, we used our framework to assess the origin of changes in functional connectivity in a local cortical network. In particular, in Chapter 4, we showed that repeated optogenetic stimulation can induce iSTDP, which in turn, causes changes in functional connectivity seen in temporal changes of spike count correlations. We also showed that the effect of iSTDP is attenuated during rest due to a massive reduction of synaptic weights, which is in agreement with experimental findings. These results prove that the presented theoretical framework can serve as a powerful tool to make theoretical, testable predictions about the dynamics of real cortical networks.

The results from Chapter 4 also brings us back to the discussion on RCP's [149, 150]. Our modeling results, combined with experimental findings from Andrei et al. [6], provide a novel method to perturb networks and hence induce iSTDP in "fast" timescales similar to those in

RCP's that have not been reported before. Importantly, as mentioned earlier, the separation of timescales assumption still held, in agreement with previous work [14, 40, 74, 92, 150].

To conclude, in Chapter 2, we have presented a theoretical framework for balanced neural networks undergoing pairwise STDP. In Chapter 3, we have validated this theory with numerical simulations for a number of STDP rules. Consistent with previous studies [52, 100], we have shown that weight updates are mainly driven by firing rates rather than spike count covariances. We have shown that the theory does not capture the dynamics of networks under iSTDP, however we provide a semi–analytical extension to account for such cases. Further extensions, *e.g.*, correlated in–degrees and out–degrees, may eventually be necessary and may be accounted for in similar ways. This theory can also capture dynamics of plastic, balanced networks in the presence of external stimuli (*e.g.*, optogenetics). This allows for better guided experiments in real cortical networks, and this is exactly what we show in Chapter 4. We showed that our framework qualitatively captures temporal changes in functional connectivity due to repeated optogenetic input to a subset of pyramidal cells in visual cortex V1 of macaques monkeys. Our model robustly replicated experimental results when the network underwent iSTDP, hinting towards the existence of such RCP in real cortical networks. Therefore, this framework allows us to make testable predictions about the intricate dynamics of the cortex.

Bibliography

[1] Adesnik, H., and Scanziani, M. Lateral competition for cortical space by layer-specific horizontal circuits. *Nature 464*, 7292 (2010), 1155–60.

[2] Agnes, E., Luppi, A., and Vogels, T. Complementary inhibitory weight profiles emerge from plasticity and allow flexible switching of receptive fields. *Journal of Neuroscience 40*, 50 (2020), 9634–9649.

[3] Ahmadian, Y., and Miller, K. What is the dynamical regime of cerebral cortex? *arXiv 1908.10101* (2019).

[4] Ahmadian, Y., Rubin, D., and Miller, K. Analysis of the stabilized supralinear network. *Neural Computation 25*, 8 (2013), 1994–2037.

[5] Akil, A., Rosenbaum, R., and Josić, K. Balanced networks under spike–timing dependent plasticity. *biorxiv* (2020).

[6] Andrei, A., Akil, A., Kharas, N., Pojoga, S., Rosenbaum, R., Janz, R., Josić, K., and Dragoi, V. Rapid, state-dependent, compensatory plasticity revealed by functional connectivity dynamics in vivo. *Unpublished work* (2021).

[7] Ardid, S., Vinck, M., Raping, D., Marquez, S., Everling, S., and Womelsdorf, T. Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex. *Journal of Neuroscience 35* (2015), 2975–2991.

[8] Atallah, B., and Scanziani, M. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron 62*, 4 (2009), 566–77.

[9] Babadi, B., and Abbott, L. Pairwise analysis can account for network structures arising from spike-timing dependent plasticity. *PLOS Computational Biology 9*, 2 (2013), e1002906.

[10] Baker, C., Ebsch, C., Lampl, I., and Rosenbaum, R. Correlated states in balanced neuronal networks. *Physical Review E 99* (2019), 052414.

[11] Baker, C., Zhu, V., and Rosenbaum, R. Nonlinear stimulus representations in neural circuits with approximate excitatory–inhibitory balance. *PLOS Computational Biology 16*, 9 (2020), e1008192.

[12] Barral, J., and Reyes, A. D. Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience 19*, 12 (2016), 1690–96.

[13] Beaman, C., Eagleman, S., and Dragoi, V. Sensory coding accuracy and perceptual performance are improved during the desynchronized cortical state. *Nature Communications 8*, 1 (2017), 1308.

[14] Bi, G., and Poo, M. Synaptic modification of correlated activity: Hebb's postulate revisited. *Annual Review of Neuroscience 24*, 1 (2001), 139–66.

[15] Bienenstock, E., Cooper, L., and Munro, P. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience 2*, 1 (1982), 32–48.

[16] Binzegger, T., Douglas, R., and Martin, K. A quantitative map of the circuit of cat primary visual cortex. *Journal of Neuroscience 24*, 39 (2004), 8441–53.

[17] Blumenfeld, B., Bibitchkov, D., and Tsodyks, M. Neural network model of the primary visual cortex: from functional architecture to lateral connectivity and back. *Physical Review Letter 20*, 2 (2006), 219–41.

[18] Bondy, A., Haefner, R., and Cumming, B. Feedback determines the structure of correlated variability in primary visual cortex. *Nature Neuroscience 21* (2018), 598–606.

[19] Bono, J., and Clopath, C. Synaptic plasticity onto inhibitory neurons as a mechanism for ocular dominance plasticity. *PLOS Computational Biology 15*, 3 (2019), e1006834.

[20] Boyden, E., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience 8*, 9 (2005), 1263–68.

[21] Bridi, M., Zong, F., Min, X., Luo, N., Tran, T., Qiu, J., Severin, D., Zhang, X., Wang, G., Zhu, Z., He, K., and Kirkwood, A. Daily oscillation of the excitation-inhibition balance in visual cortical circuits. *Neuron 105* (2020), 621–629.

[22] Burak, Y., and Fiete, I. Fundamental limits on persistent activity in networks of noisy neurons. *PNAS 109*, 43 (2012), 17645–17650.

[23] Burkitt, A., Gilson, M., and van Hemmen, J. Spike-timing dependent plasticity for neurons with recurrent connections. *Biological Cybernetics 96*, 5 (2007), 533–546.

[24] Chiu, C., Martenson, J., Yamazaki, M., Natsume, R., Sakimura, K., Tomita, S., Tavalin, S., and Higley, M. Input-specific nmdar-dependent potentiation of dendritic gabaergic inhibition. *Neuron 97*, 2 (2018), 368–377.

[25] Cohen, M. R., and Kohn, A. Measuring and interpreting neuronal correlations. *Nature Neuroscience 14*, 7 (2011), 811–9.

[26] Compte, A., Constantinidis, C., Tegner, J., Raghavachari, S., Chafee, M., Goldman-Rakic, P., and Wang, X. Temporally irregular mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *Journal of Neurophysiology 90*, 5 (2003), 3441–54.

[27] Darshan, R., van Vreeswijk, C., and Hansel, D. Strength of correlations in strongly recurrent neuronal networks. *Physical Review X 8* (2018), 031072.

[28] Deco, G., and Hugues, E. Neural network mechanisms underlying stimulus driven variability reduction. *PLOS Computational Biology 8*, 3 (2012), e1002395.

[29] Dehghani, N., Peyrache, A., Telenczuk, B., Quyen, M. L. V., Halgren, E., S Cash, N. H., and Destexhe, A. Dynamic balance of excitation and inhibition in human and monkey neocortex. *Nature 6*, 23176 (2016), 1–12.

[30] Doiron, B., and Litwin-Kumar, A. Balanced neural architecture and the idling brain. *Frontiers in Computational Neuroscience 8*, 56 (2014), 1–12.

[31] Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G. K., and Josić, K. The mechanics of state–dependent neural correlations. *Nature Neuroscience 19*, 3 (2016), 283–393.

[32] Dunwiddie, T., and Lynch, G. Long-term potentiation and depression of synaptic responses in the rat hippocampus: localization and frequency dependency. *Journal of Physiology 276*, 1 (1978), 353–67.

[33] Ebsch, C., and Rosenbaum, R. Imbalanced amplification: a mechanism of amplification and suppression from local imbalance of excitation and inhibition in cortical circuits. *PLOS Computational Biology 14*, 3 (2018), e1006048.

[34] Ecker, A., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C., Smirnakis, S. M., Bethge, M., and Tolias, A. State dependence of noise correlations in macaque primary visual cortex. *Neuron 82*, 1 (2014), 235–248.

[35] Ecker, A., Berens, P., Keliris, G., Bethge, M., Logothetis, N., and Tolias, A. Decorrelated neuronal firing in cortical microcircuits. *Science 327*, 5965 (2010), 584–7.

[36] Erickson, M., Maramara, L., and Lisman, J. A single brief burst induces glur1-dependent associative short-term potentiation: a potential mechanism for short-term memory. *Journal of Cognitive Neuroscience 22* (2010), 2530–2540.

[37] Fourcaud-Trocme, N., Hansel, D., van Vreeswijk, C., and Brunel, N. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *Journal of Neuroscience 23*, 37 (2003), 11628–11640.

[38] Froemke, R., Merzenich, M., and Shreiner, C. A synaptic memory trace for cortical receptive field plasticity. *Nature 450*, 7168 (2007), 425–429.

[39] Froemke, R., Tsay, I., Raad, M., Long, J., and Dan, Y. Contribution of individual spikes in burst-induced long-term synaptic modification. *Journal of Neurophysiology 95*, 3 (2006), 1620–9.

[40] Froemke, R. C. Plasticity of cortical excitatory–inhibitory balance. *Annual Review of Neuroscience 38* (2015), 195–219.

[41] Galarreta, M., and Hestrin, S. Frequency-dependent synaptic depression and the balance of excitation and inhibition in the neocortex. *Nature Neuroscience 1*, 7 (1998), 587–94.

[42] Gerstner, W., Kempter, R., van Hemmen, J., and Wagner, H. A neuronal learning rule for sub-millisecond temporal coding. *Nature 383*, 1 (1996), 76–78.

[43] Gerstner, W., Lehmann, M., Liakoni, V., Corneil, D., and Brea, J. Eligibility traces and plasticity on behavioral time scales: experimental support of neoHebbian three–factor learning rules. *Frontiers in Neural Circuits 12*, 53 (2018), 1–16.

[44] Gilson, M., Burkitt, A., Grayden, D., Thomas, D., and van Hemmen, J. 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 (2009), 81–102.

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

[46] Gilson, M., Burkitt, A., Grayden, D., Thomas, D., and van Hemmen, J. 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 (2009), 411–26.

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

[48] Gilson, M., Burkitt, A., Grayden, D., Thomas, D., and van Hemmen, J. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. v: self-organization schemes and weight dependence. *Biological Cybernetics 103*, 5 (2010), 365–386.

[49] Gjorgjieva, J., Clopath, C., Audet, J., and Pfister, J. A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *PNAS 108*, 48 (2011), 19383–8.

[50] Goris, R., Movshon, J., and Simoncelli, E. Partitioning neuronal variability. *Nature Neuroscience 17*, 6 (2014), 858–865.

[51] Graupner, M., and Brunel, N. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *PNAS 109*, 10 (2012), 3991–6.

[52] Graupner, M., Wallisch, P., and Ostojić, S. Natural firing patterns imply low sensitivity of synaptic plasticity to spike timing compared with firing rate. *Journal of Neuroscience 36*, 44 (2016), 11238–58.

[53] Griffith, J., and Horn, G. An analysis of spontaneous impulse activity of units in the striate cortex of unrestrained cats. *Journal of Physiology 186*, 3 (1966), 516–34.

[54] Grytskyy, D., Tetzlaff, T., Diesmann, M., and Helias, M. A unified view on weakly correlated recurrent networks. *Frontiers in Computational Neuroscience 7* (2013), 131.

[55] Gütig, R., Aharonov, R., Rotter, S., and Sompolinsky, H. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *Journal of Neuroscience 23*, 9 (2003), 3697–714.

[56] Haider, B., Duque, A., Hasenstaub, A., and McCormick, D. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience 26*, 17 (2006), 4535–45.

[57] Haider, B., Häusser, M., and Carandini, M. Inhibition dominates sensory responses in the awake cortex. *Nature 493* (2013), 97–100.

[58] Han, X., Qian, X., Bernstein, J., Zhou, H., Talei, G., Stern, P., Bronson, R., Graybiel, A., Desimone, R., Boyden, E., and Franzesi, G. Millisecond-timescale optical control of neural dynamics in the nonhuman primate brain. *Neuron 62* (2009), 191–198.

[59] He, K., Huertas, M., Hong, S. Z., Tie, X. X., Hell, J. W., Shouval, H., and Kirkwood, A. Distinct eligibility traces for LTP and LTD in cortical synapses. *Neuron 88*, 3 (2015), 528–538.

[60] Hebb, D. *The organization of behavior*. New York: John Wiley and Sons, Inc., 1949.

[61] Heiss, J., Katz, Y., Ganmor, E., and Lampl, I. Shift in the balance between excitation and inhibition during sensory adaptation of s1 neurons. *Journal of Neuroscience 28*, 49 (2008), 13320–30.

[62] Helias, M., Tetzlaff, T., and Diesmann, M. The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. *PLOS Computational Biology 10*, 1 (2014), e1003428.

[63] Hennequin, G., Agnes, E. J., and Vogels, T. P. Inhibitory plasticity: balance, control, and codependence. *Annual Review of Neuroscience 40* (2017), 557–79.

[64] Hennequin, G., Ahmadian, Y., Rubin, D., Lengyel, M., and Miller, K. The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron 98*, 4 (2018), 846–860.

[65] Hertäg, L., Durstewitz, D., and Brunel, N. Analytical approximations of the firing rate of an adaptive exponential integrate-and-fire neuron in the presence of synaptic noise. *Frontiers in Computational Neuroscience 8* (2014).

[66] Hodgkin, A., and Huxley, A. A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology 117*, 4 (1952), 500–44.

[67] Houk, J., Adams, J., and Barto, A. *A model of how the basal ganglia generate and use neural signals that predict reinforcement*. Cambridge (MA): The MIT Press, 1995.

[68] Hromadka, T., DeWeese, M., and Zador, A. Sparse representation of sounds in the unanesthetized auditory cortex. *PLOS Biology 6*, 1 (2008), e16.

[69] Huang, C., Ruff, D. A., Pyle, R., Rosenbaum, R., Cohen, M. R., and Doiron, B. Circuit models of low–dimensional shared variability in cortical networks. *Neuron 101*, 2 (2019), 337–348.

[70] Izhikevich, E. Solving the distal reward problem through linkage of stdp and dopamine signaling. *Cerebral Cortex 17*, 10 (2007), 2443–52.

[71] Izhikevich, E., Gally, J., and Edelman, G. Spike-timing dynamics of neuronal groups. *Cerebral Cortex 14*, 8 (2004), 933–44.

[72] Kandel, E., and Tauc, L. Mechanism of prolonged heterosynaptic facilitation. *Nature 202*, 3 (1964), 145–147.

[73] Kandel, E., and Tauc, L. Heterosynaptic facilitation in neurones of the abdominal ganglion of aplysia depilans. *Journal of Physiology 181*, 1 (2012), 1–27.

[74] Kempter, R., Gerstner, W., and Van Hemmen, J. Hebbian learning and spiking neurons. *Physical Review E 59*, 4 (1999), 4498–4514.

[75] Kincaid, D., and Cheney, W. *Numerical analysis: mathematics of scientific computing.* Providence: American Mathematical Society, 2002.

[76] Klopf, A. *The hedonistic neuron: a theory of memory, learning, and intelligence.* New York: Hemisphere, 1982.

[77] Koch, K., and Fuster, J. Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Experimental Brain Research 76*, 2 (1989), 292–53064.

[78] Kohonen, T. *Self-organization and associative memory.* Berlin: Springer-Verlag, 1984.

[79] Kohonen, T. The self-organizing map. *Proceedings of the IEEE 78*, 9 (1990), 1464–80.

[80] Kohonen, T. Physiological interpretation of the self-organizing map algorithm. *Neural Networks 6*, 7 (1993), 895–905.

[81] Kuhn, A., Aersten, A., and Rotter, S. Higher order statistics of input ensembles and the response of simple model neurons. *Neural Computation 15*, 1 (2003), 67–101.

[82] Kullmann, D., Moreau, A., Bakiri, Y., and Nicholson, E. Plasticity of inhibition. *Neuron 75*, 6 (2012), 951–962.

[83] Landau, I., Egger, R., Dercksen, V., Oberlaender, M., and Sompolinsky, H. The impact of structural heterogeneity on excitation-inhibition balance in cortical networks. *Neuron 92*, 5 (2016), 1106–21.

[84] Landau, I., and Sompolinsky, H. Coherent chaos in a recurrent neural network with structured connectivity. *PLOS Computational Biology 14*, 12 (2018), e1006309.

[85] Lindner, B., Longtin, A., and Bulsara, A. Analytic expressions for rate and cv of a type i neuron driven by white Gaussian noise. *Neural Computation 15*, 8 (2003), 1761–1788.

[86] Litwin–Kumar, A., and Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience 15* (2012), 1498–1505.

[87] Litwin-Kumar, A., and Doiron, B. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications 5*, 5319 (2014).

[88] Lynch, G., Dunwiddie, T., and Gribkoff, V. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature 266* (1977), 737–739.

[89] Maimon, G., and Assad, J. Beyond Poisson: increased spike-time regularity across primate parietal cortex. *Neuron 62*, 3 (2009), 426–40.

[90] Malenka, R. Postsynaptic factors control the duration of synaptic enhancement in area ca1 of the hippocampus. *Neuron 6* (1991), 53–60.

[91] Markram, H., Gerstner, W., and Sjöström, P. A history of spike-timing-dependent plasticity. *Frontiers in Synaptic Neuroscience 3*, 4 (2011).

[92] Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. Regulation of synaptic efficacy by coincident postsynaptic aps and epsps. *Science 275*, 5297 (1997), 213–5.

[93] Mastrogiuseppe, F., and Ostojic, S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron 99*, 3 (2018), 609–23.

[94] Miller, K. Canonical computations of cerebral cortex. *Current Opinion in Neurobiology 37*, 3 (2012), 75–84.

[95] Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K. D., and de la Rocha, J. Stochastic transitions into silence cause noise correlations in cortical circuits. *PNAS 112*, 11 (2015).

[96] Mongillo, G., Rumpel, S., and Loewenstein, Y. Inhibitory connectivity defines the realm of excitatory plasticity. *Nature Neuroscience 21*, 10 (2018), 1463–1470.

[97] Montangie, L., Miehl, C., and Gjorgjieva, J. Autonomous emergence of connectivity assemblies via spike triplet interactions. *PLOS Computational Biology 16*, 5 (2020), e1007835.

[98] Morrison, A., Aertsen, A., and Diesmann, M. Spike timing–dependent plasticity in balanced random networks. *Neural Computation 19*, 6 (2007), 1437–1467.

[99] Ocker, G., Josić, K., SheaBrown, E., and Buice, M. Linking structure and activity in nonlinear spiking networks. *PLOS Computational Biology 13*, 6 (2017), e1005583.

[100] Ocker, G., Litwin-Kumar, A., and Doiron, B. Self–organization of microcircuits in networks of spiking neurons with plastic synapses. *PLOS Computational Biology 11*, 8 (2015), e1004458.

[101] O'Connor, D., Peron, S., Huber, D., and Svoboda, K. Neural activity in barrel cortex underlying vibrissa-based object localization in mice. *Neuron 67*, 6 (2010), 1048–61.

[102] Oja, E. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology 15*, 3 (1982), 267–273.

[103] Okun, M., and Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience 11*, 5 (2008), 535–7.

[104] Okun, M., and Lampl, I. Balance of excitation and inhibition. *Scholarpedia 4*, 8 (2009), 7467.

[105] Ostojic, S. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature Neuroscience 17*, 4 (2014), 594–600.

[106] Petersen, C., Malenka, R., Nicoll, R., and Hopfield, J. All-or-none potentiation at ca3-ca1 synapses. *PNAS 95*, 8 (1998), 4732–7.

[107] Petrenau, L., Huber, D., Sobczyk, A., and Svoboda, K. Channelrhodopsin-2–assisted circuit mapping of long-range callosal projections. *Nature Neuroscience 10*, 5 (2007), 663–8.

[108] Pfister, J., and Gerstner, W. Triplets of spikes in a model of spike-timing dependent plasticity. *Journal of Neuroscience 26*, 38 (2006), 9673–82.

[109] Pouille, F., Burgin, A. M., Adesnik, H., Atallah, B., and Scanziani, M. Input normalization by global feedforward inhibition expands cortical dynamic range. *Nature Neuroscience 12*, 12 (2009), 1577–85.

[110] Pyle, R., and Rosenbaum, R. Highly connected neurons spike less frequently in balanced networks. *Physical Review E 93* (2016), 040302.

[111] Ravid Tannenbaum, N., and Burak, Y. Shaping neural circuits by high order synaptic interactions. *PLOS Computational Biology 12*, 8 (2016), e1005056.

[112] Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Praga, N., Reyes, A., and Harris, K. D. The asynchronous state in cortical circuits. *Science 327*, 2 (2010), 587–590.

[113] Richardson, M. J. Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive. *Physical Review E 76*, 2 (2007), 021919.

[114] Rosenbaum, R. A diffusion approximation and numerical methods for adaptive neuron models with stochastic inputs. *Frontiers in Computational Neuroscience 10* (2016).

[115] Rosenbaum, R., and Doiron, B. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X 4*, 2 (2014), 021039.

[116] Rosenbaum, R., Smith, M. A., Kohn, A., Rubin, J. E., and Doiron, B. The spatial structure of correlated neuronal variability. *Nature Neuroscience 20*, 1 (2017), 107–114.

[117] Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience 31*, 45 (2011), 16217–16226.

[118] Sanzeni, A., Histed, M., and Brunel, N. Emergence of irregular activity in networks of strongly coupled conductance-based neurons. *biorxiv* (2020).

[119] Seeman, S., Mogen, B., Fetz, E., and Perlmutter, S. Paired stimulation for spike-timing dependent plasticity in primate sensorimotor cortex. *Journal of Neuroscience 37* (2017), 1935–1949.

[120] Shadlen, M., and Newsome, W. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology 4*, 4 (1994), 569–79.

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

[122] Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., and Bodner, M. Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience 146*, 3 (2007), 1082–108.

[123] Shaham, N., and Burak, Y. Slow diffusive dynamics in a chaotic balanced neural network. *PLOS Computational Biology 13*, 5 (2017), e1005505.

[124] Shouval, H., Bear, M., and Cooper, L. A unified model of nmda receptor-dependent bidirectional synaptic plasticity. *PNAS 99*, 16 (2002), 10832–6.

[125] Smith, M., and Kohn, A. Spatial and temporal scales of neuronal correlation in primary visual cortex. *Journal of Neuroscience 28* (2008), 12591–12603.

[126] Smith, M. A., Jia, X., Zandvakili, A., and Kohn, A. Laminar dependence of neuronal correlations in visual cortex. *Journal of Neurophysiology 109*, 4 (2013), 940–947.

[127] Softky, W., and Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience 13*, 1 (1993), 334–350.

[128] Sompolinsky, H., Crisanti, A., and Sommers, H. Chaos in random neural networks. *Physical Review Letter 61* (1988), 259.

[129] Song, S., Sjöström, P., Reigl, M., Nelson, S., and Chklovskii, D. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLOS Biology 3*, 3 (2005), e68.

[130] Sprekeler, H. Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. *Current Opinion in Neurobiology 49* (2017), 198–203.

[131] Tan, A., Chen, Y., Scholl, B., Seidemann, E., and Priebe, N. J. Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature 509*, 7499 (2014), 226–229.

[132] Tetzlaff, T., Helias, M., Einevoll, G., and Diesmann, M. Decorrelation of neural network activity by inhibitory feedback. *PLOS Computational Biology 8*, 8 (2012), e1002596.

[133] Trainito, C., von Nicolai, C., Miller, E., and Siegel, M. Extracellular spike waveform dissociates four functionally distinct cell classes in primate cortex. *Current Biology 29*, e5 (2019), 2973–2982.

[134] Trousdale, J., Hu, Y., Shea-Brown, E., and Josic, K. Impact of network structure and cellular response on spike time correlations. *PLOS Computational Biology 8*, 3 (2012), e1002408.

[135] Trousdale, J., Hu, Y., Shea-Brown, E., and Josic, K. A generative spike train model with time-structured higher order correlations. *Frontiers in Computational Neuroscience 7*, 84 (2013), 1–21.

[136] Tsodyks, M., and Sejnowski, T. Rapid state switching in balanced cortical network models. *Network: Computation in Neural Systems 6*, 2 (1995), 111–124.

[137] van Vreeswijk, C., and Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science 274*, 5293 (1996), 1724–1726.

[138] van Vreeswijk, C., and Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural Computation 10*, 6 (1998), 1321–1371.

[139] Vegue, M., and Roxin, A. Firing rate distributions in spiking networks with heterogeneous connectivity. *Physical Review E 100*, 2 (2019).

[140] Vogels, T. P., Froemke, R. C., Doyon, N., Gilson, M., Haas, J. S., Liu, R., Maffei, A., Miller, P., Wierenga, C. J., Woodin, M. A., Zenke, F., and Sprekeler, H. Inhibitory synaptic plasticity: spike timing–dependence and putative network function. *Frontiers in Neural Circuits 7*, 119 (2013).

[141] Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science 334*, 6062 (2011), 1569–1573.

[142] Vyazovskiy, V., Cirelli, C., Pfister-Genskow, M., Faraguna, U., and Tononi, G. Molecular and electrophysiological evidence for net synaptic potentiation in wake and depression in sleep. *Journal of Neuroscience 11* (2008), 2975–2991.

[143] White, J., Rubinstein, J., and Kay, A. Channel noise in neurons. *Trends in Neurosciences 23*, 3 (2000), 131–7.

[144] Wilmes, K., and Clopath, C. Inhibitory microcircuits for top-down plasticity of sensory representations. *Nature Communications 10*, 5055 (2019).

[145] Wimmer, K., Compte, A., Roxin, A., Peixoto, D., Renart, A., and de la Rocha, J. Sensory integration dynamics in a hierarchical network explains choice probabilities in cortical area mt. *Nature Communications 6*, 6177 (2015).

[146] Xue, M., Atallah, B., and Scanziani, B. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature 511*, 7511 (2014), 596–600.

[147] Yizhar, O., Fenno, L., Prigge, M., Schneider, F., Davidson, T., O'Shea, D., Sohal, V., Goshen, I., Finkelstein, J., Paz, J., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J., Hegemann, P., and Deisseroth, K. Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature 477*, 7363 (2011), 171–8.

[148] Zenke, F., Agnes, E., and Gerstner, W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications 6*, 6922 (2015), 1–13.

[149] Zenke, F., and Gerstner, W. Hebbian plasticity requires compensatory processes on multiple timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences 372*, 1715 (2017), 20160259.

[150] Zenke, F., Gerstner, W., and Ganguli, S. The temporal paradox of Hebbian learning and homeostatic plasticity. *Current Opinion in Neurobiology 43* (2017), 166–176.

[151] Zhou, M., Liang, F., Xiong, X., Li, L., Li, H., Xiao, Z., Tao, H., and Zhang, L. Scaling down of balanced excitation and inhibition by active behavioral states in auditory cortex. *Nature Neuroscience 17*, 6 (2014), 841–50.