FORMATION OF STRUCTURE IN CORTICAL NETWORKS THROUGH SPIKE TIMING-DEPENDENT PLASTICITY

by

Gabriel Koch Ocker

B.A., Neuroscience and Applied Mathematics, Oberlin College 2010

Submitted to the Graduate Faculty of

the Kenneth P. Dietrich School of Arts and Sciences

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Gabriel Koch Ocker

It was defended on

September 18th 2015

and approved by

Brent Doiron, Dept. of Mathematics

Bard Ermentrout, Dept. of Mathematics

Jonathan Rubin, Dept. of Mathematics

Anne-Marie Oswald, Dept. of Neuroscience

Thanos Tzounopoulos, Dept. of Neurobiology

Kenneth D. Miller, Dept. of Neuroscience
& Center for Theoretical Neuroscience, Columbia University

Dissertation Director: Brent Doiron, Dept. of Mathematics

FORMATION OF STRUCTURE IN CORTICAL NETWORKS THROUGH SPIKE TIMING-DEPENDENT PLASTICITY

Gabriel Koch Ocker, PhD

University of Pittsburgh, 2015

The connectivity of mammalian brains exhibits structure at a wide variety of spatial scales, from the broad (which brain areas connect to which) to the extremely fine (where synapses form on the morphology of individual neurons). Two striking features of the neuron-to-neuron connectivity are 1) the strong over-representation of multi-synapse connectivity patterns compared to simple random-network models and 2) a strong relationship between neurons' local connectivity and their stimulus preferences, so that local network structure plays a large role in the computations neurons perform. A central question in systems neuroscience is how such structures emerge. Answers to this question are confounded by the mutual interactions of neuronal activity and neural network structure. Patterns of synaptic connectivity influence neurons' joint activity, while the synapses between neurons are plastic and strengthen or weaken depending on the activity of the pre- and postsynaptic neurons. In this thesis, I develop a self-consistent framework for the coevolution of network structure and spiking activity. Subsequent chapters leverage this to develop low-dimensional sets of equations that directly describe the plasticity of connectivity patterns in large spiking networks. I examine plasticity during spontaneous activity and then how the structure of external stimuli can shape network structure and subsequent spontaneous plasticity. These studies provide a step towards understanding how the structure of neuronal networks and neurons' joint activity interact to allow network computations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I owe thanks to many people for their support and help during my graduate school career. This work would not have been possible without the contributions of many others, nor would I be the person I am without the support, influence and example of my family, friends and peers. I would first like to thank my committee for their time and effort in shaping and supporting my work. I have learned a lot from conversations and journal clubs with Drs. Bard Ermentrout, Jon Rubin and Anne-Marie Oswald and from collaboration with Dr. Thanos Tzounopoulos. I would like to especially thank Dr. Ken Miller for taking the time to travel here in order to be a part of my committee.

I count myself extremely lucky to have found a home in Brent Doiron's group. The care he puts into his students' growth as scientists is, I believe, unmatched. I've learned a huge amount from him, and enjoyed sharing slightly less beer than that. He has been a true mentor and a good friend. I will remain grateful for the opportunities and advice he has given me.

The other members of the group have been hugely influential as well as good friends. Ashok and Rob, senior to me in the lab, were role models of careful thought and hard work. Both I and my work benefited enormously from our conversations. The same is true of Cheng, Zack, Tatjana, Jeff, Aubrey and Mike. I look forward to further conversations with all of them. Tatjana and Ken have been especially good friends, and I look forward to seeing them in Seattle in the near future.

My parents have been a constant source of support, encouragement and advise (even despite my decision to abandon the humanities they devoted their careers to). I count myself lucky to have had the opportunities they afforded me and their continued support. I still attempt to follow their writing advice to my high-school self. I always look forward

to seeing and speaking to my sisters and am grateful that they are able to be here for my defense. Finally, I cannot overstate my appreciation for Mona's love and support. She improves me immeasurably, and I look forward very much to the next chapter of our lives.

1.0 INTRODUCTION

Moment-by-moment, our nervous systems engender our thoughts and actions through their electrochemical activity. They allow us to detect, classify and respond to external stimuli. To do so appropriately, they must learn associations, for example between different stimuli (as in Pavlov's classical conditioning) or between internally produced behaviors and external stimuli (as in operant learning). Our immediate behaviors rely on neurons' ability to generate action potentials: millisecond-scale, 50-100 mV depolarizations that huge relative to the resting potential and generated through a positive feedback loop of depolarization-activated inward currents [1]. Generated at the axon hillock, action potentials propagate down the axon as well as back-propagating through the neuron's cell body and dendrites. When the action potential reaches boutons, the site of comunication with postsynaptic neurons, the depolarization triggers the release of neurotransmitter into the synaptic cleft. The neurotransmitter diffuses across the cleft and, binding to postsynaptic receptors, activates them.

Neurons are often divided broadly into two classes: excitatory and inhibitory. This division is based on the type of neurotransmitters they release and the effect of those neurotransmitters on postsynaptic neurons. There are many types of neurotransmitter. Excitatory neurons of the neocortex most often release glutamate, while most common inhibitory neurotransmitter in the cortex is γ-aminobutyric acid (GABA). Neurotransmitter receptors can be metabotropic, in which case they activate intracellular signaling cascades, or ionotropic - in which case they contain ion channels. When neurotransmitter binds to an ionotropic receptor, these channels open allowing ions to pass through and generating a postsynaptic potential (PSP). Glutamate receptors allow cations to flow inwards, generating depolarizing excitatory post-synaptic potentials (EPSPs), while GABAergic receptors allow anions (chlo-

ride) to enter the neuron, generating hyperpolarizing inhibitory post-synaptic potentials (IPSPs).

There are two main types of ionotropic glutamate receptor: N-methyl-D-aspartate receptors (NMDARs) and α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptors (AMPARs), each named for molecules that specifically bind to them. AMPARs have fast kinetics, generating currents with rising time constants under 1 ms and decay time constants up to 5 ms [2, 3, 4, 5, 6]. NMDARs have slower kinetics, generating currents with rising time constants of 5-10 ms [7, 8] and decay constants around 60 ms in postsynaptic excitatory neurons [3] (although early in development, the decay time constants are significantly longer[3, 9]). NMDARs also have a magnesium block: under resting conditions, the ion channel is blocked by magnesium. Depolarization for a few milliseconds removes the magnesium block, allowing glutamate binding to open the receptor [10, 11, 12].

1.1 LONG-TERM SYNAPTIC PLASTICITY: A BRIEF HISTORY

The search for the physiological basis of learning and memory has been a central aim of neuroscience since the field's birth [13]. The now-common term "plasticity", referring to short- or long-lasting changes in synaptic or cellular physiology in response to a stimulus, was apparently coined by the Italian physiologist Ernesto Lugaro in the early 20th century [14]. Electrical recordings showing long-lasting depolarization in response to high-frequency stimulation, by Ralf Gerard in frog nerves in 1930, provided the first evidence for long-lasting electrophysiological changes after stimulation. Donald Hebb, in his seminal book "The Organization of Behavior" (1949), hypothesized that "When an axon of cell A is near enough to excite 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 efficacy, as one of the cells firing B, is increased" [15]. He proposed this as a mechanism for the formation of engrams: long-term biological modifications subserving memory. Decades later in the 1960's and 70's, two sets of studies laid much of the groundwork for the modern study of long-term synaptic plasticity.

First, Eric Kandel and Ladislav Tauc provided what remains some of the most convincing evidence for synaptic plasticity as a mechanism for behavioral learning. Working in the sea slug *Aplysia depilans*, they developed an electrophysiological model of classical conditioning [16, 17] which allowed later work to directly link synaptic plasticity with behavioral learning of the *Aplysia* gill withdrawal reflex. Experiments pairing behavioral stimuli with electrical stimulation of neurons or neural populations remain a state-of-the-art tool for the study of synaptic mechanisms of behavioral learning (e.g. in various cortical regions, [18, 19, 20, 21, 22]).

Second, Terje Lømo and Timothy V.P. Bliss discovered that high-frequency stimulation of the perforant path from entorhinal cortex to the dentate gyrus led to long-term potentiation of post-synaptic potentials (LTP) [23, 24]. These results were confirmed by Douglas & Gerard (1975), who showed that repeated bursts of stimulation were more effective than a single high-frequency stimulus train [25]. Similar experiments also showed LTP of evoked field potentials in CA1 [26]. At the end of the 1970's McNaughton, Douglas & Gerard showed that multiple weakly-activated synaptic pathways could cooperate to induce associative LTP in both pathways' EPSPs [27]. Baranyi & Fehèr provided confirmation that EPSPs did not have to directly cause action potentials in order for their temporal coincidence to drive LTP [28]. Also at the end of the 1970's, Lynch, Dunwiddie & Gribkoff discovered long-term *depression* (LTD) in response to low-frequency stimulation, also in the hippocampus [29, 30]. Subsequent studies combining LTP and LTD protocols showed that potentiated synapses could subsequently depress [31, 32].

One of the first studies to examine the importance of the *relative* timing of pre- and post-synaptic activity was carried out by Levy & Steward. They found the first evidence of a temporal asymmetry in synaptic plasticity: repeated stimulation of a weak pathway, prior to repeated stimulation of a strong pathway, led to LTP of the weak pathway while the reverse stimulation order led to LTD [33]. In some of the first experiments investigating the relative timing of single presynaptic spikes and EPSPs, Gustafsson, Wigström & colleagues reported postsynaptic depolarization needed to occur within ∼100 ms of the conditioning stimulation [34, 35]. These studies presented trains of presynaptic and postsynaptic spikes in blocks, each block consisting entirely of pre- or of postsynaptic stimulation. It was not until

the late 1990's that experimentalists showed that repeated pairs of presynaptic spikes and EPSPs, depending on their relative timing, could lead to either potentiation or depression, discovering spike timing-dependent plasticity (STDP) [36, 37].

These studies highlighted how the relative timing of pairs of pre- and postsynaptic spikes can control plasticity using experimental protocols in which pre-post (or post-pre) spike pairs were elicited repeatedly at fixed time lags. This highly regular spiking stands in stark contrast to the variable activity of neurons *in vivo*. Further studies revealed that plasticity induced by larger groups of pre- and postsynaptic spikes cannot be accounted for by a solely pair-based effects [38, 39, 40, 41].

1.2 MECHANISMS OF STDP

The molecular machinery and emergent network mechanisms underlying the formation, consolidation and storage of long-term memories are complex, and their study is an exciting and very active area of research. Mechanistically, Hebbian STDP takes two main forms. These are 1) NMDA receptor-dependent LTP and LTD, and 2) the combination of NMDA receptor-dependent LTP and metabotropic glutamate receptor (mGluR) or cannabinoid receptor-dependent (CbR) LTD, both of which have been shown to take place at cortical synapses [42]. In NMDAR-dependent LTP and LTD, pre-post spike pairs relieve the magnesium block and allow strong calcium influx. Due to the kinetics of NMDARs [43], the coincidence of AMPA-mediated EPSPs with back-propagating postsynaptic action potentials [44, 45], and EPSPs activing inward currents and inactivating outward currents in the dendrites, enhancing the backpropagating depolarization [46, 47], pre-post spike pairs all enhance calcium influx. Post-pre spike pairs, in contrast, lead to weaker calcium influx [48, 49, 50, 51, 52]. In mGluR-dependent LTD, the combination of mGluRs and voltage-sensitive calcium channels leads to the release of the endocannabinoid transmitter 2-AG, which binds to presynaptic cannabinoid receptors and ultimately decreases the presynaptic release probability [53, 54, 55, 56, 57, 58, 59].

Anti-Hebbian STDP has also been observed, in cerebellar-like structure such as the dor-

sal cochlear nucleus [60] and electrosensory lateral line [61, 62]. The term broadly refers to STDP that does not respect the Hebbian restriction of potentiation only occurring for pre-post spike pairs and depression only occurring for post-pre spike pairs, but often takes the form of only LTD (for pre-post pairs) [63]. Anti-Hebbian LTD in pre-post spike pairs relies on endocannabinoid signaling in a similar manner to the Hebbian mGluR-dependent LTD [64]. Neuromodulators can control the sign of plasticity, converting potentiating stimulation patterns into depressing ones [65]. Mechanisms of short-term synaptic facilitation and depression, while acting on much faster timescales than long-term plasticity, can play a role in controlling the induction of LTP [66, 67, 68, 69]. How diverse synaptic and cellular mechanisms conspire to control long-term plasticity in different brain areas during in vivo activity remains an exciting area of research.

The

1.3 MECHANISTIC MODELS OF SYNAPTIC PLASTICITY

The modern study of learning and memory has been rife with mathematical theory since its early days [13]. Here, we will discuss two broad classes of synaptic plasticity model: mechanistic and phenomenological. The division of models into mechanistic and phenomenological classes could be carried out in many ways, depending on what we call a 'mechanism'. By a mechanistic model, we mean one that takes into account sub-cellular processes such as postsynaptic calcium concentrations and molecular signaling cascades [70]. We will briefly discuss calcium-based models, which have been especially influential.

The first is based on the calcium control hypothesis, which inspired experimental work showing that different postsynaptic calcium levels lead to different plasticity outcomes [71, 72, 73, 57]. Some of these models implement thresholds for the calcium concentration, θ_d and θ_p with $\theta_d < \theta_p$. If calcium levels are intermediate ($\theta_d < [Ca^{2+}] < \theta_p$) then the synapse depresses but if calcium levels are sufficiently elevated ($\theta_p < [Ca^{2+}]$) the synapse potentiates [49, 74, 75, 76]. The calcium concentration is increased by presynaptic spikes through activation of NMDA receptors. Back-propagating postsynaptic spikes can 1) increase

calcium levels through activation of voltage-dependent calcium channels [57, 37, 77] and 2) can remove the magnesium block of the NMDA receptors, allowing presynaptic spikes to induce calcium entry [10, 12]. Another type of calcium-based models is based on having multiple sensors for postsynaptic calcium [78, 79, 80] which compete to determine whether the synaptic efficacy potentiates, depresses, or does not change. Both of these classes of model can successfully reproduce the wide variety of STDP curves observed in different experimental paradigms [81, 42], although they can be differentiated by (for example) the effects of burst- or triplet-induced plasticity [80].

1.4 PHENOMENOLOGICAL MODELS OF SYNAPTIC PLASTICITY

1.4.1 Rate-based plasticity rules

Before the discovery of STDP, a number of rate-based plasticity rules revealed how Hebbian learning can allow computation in feedforward neural systems. We will briefly summarize three classic learning rules: those of Bienenstock, Cooper & Munro (BCM) [82], Oja [83] and Linsker [84]. These three models highlight how plasticity can allow neuronal circuits to learn to perform computations. The BCM model accounted for a wide range of experimental results pertaining to the development of ocular dominance columns in primary visual cortex. Importantly, they also introduced the notion of plasticity dependent on temporal, rather than spatial, relationships between pre- and postsynaptic activity. Notably, the BCM rule predicted the possibility of homosynaptic depression: synapses depressing not due to competition amongst synapses (heterosynaptic depression), but due to the joint activity of the pre- and post-synaptic neurons. In the BCM learning rule, the evolution of synaptic efficacy is a function of the pre- and postsynaptic firing rates:

$$\Delta \mathbf{W}_{\text{post}\leftarrow\text{pre}} = \eta \left(\rho_{\text{post}} \left(\rho_{\text{post}} - \theta\right) \rho_{\text{pre}}\right) - \epsilon \mathbf{W}_{\text{post}\leftarrow\text{pre}} \tag{1.1}$$

where ρ denotes the instantaneous firing rate. The last term, $-\epsilon \mathbf{W}_{\text{post}\leftarrow\text{pre}}$, stabilizes the weights in the face of certain stimulation patterns but is often neglected. θ is a threshold

separating depression at low rates from potentiation at high rates, and this threshold is itself a function of the postsynaptic firing rate, commonly taken as $\theta = \langle \rho_{\text{post}}/y_0 \rangle^2$ [85] where angular brackets denote the average output (over different inputs) and y_0 is a constant. The rate of learning is set by η.

Two other classical rate-based plasticity rules, also published in the 1980's, are those of Erkki Oja and Ralph Linsker. The Oja rule reads:

$$\Delta \mathbf{W}_{\text{post}\leftarrow\text{pre}} = \eta \left(\rho_{\text{pre}} \rho_{\text{post}} - \rho_{\text{post}}^2 \mathbf{W}_{\text{post}\leftarrow\text{pre}} \right) \tag{1.2}$$

Notably, this allows the postsynaptic neuron to learn to perform a principle components analysis of a group of inputs $\{\rho_{\text{pre}}\}$, providing a link between theories of Hebbian learning and neural-inspired computations [83]. Linsker's equation is:

$$\Delta \mathbf{W}_{\text{post}\leftarrow\text{pre}} = k + \eta \left(\rho_{\text{post}} - y_0 \right) \left(\rho_{\text{pre}} - x_0 \right) \tag{1.3}$$

where k, y_0 and x_0 are parameters. Linsker's equation, similarly to the BCM rule, was initially proposed as a Hebbian rule to explain the emergence of stimulus-selectivity in sensory areas, although Linsker focused on orientation selectivity rather than ocular dominance. While all of these plasticity rules are based on relationships between pre- and post-synaptic activity, their reliance on average firing rates as opposed to individual action potentials precludes discussion of how spiking activity on neuronal timescales affects plasticity.

1.4.2 Timing-based plasticity rules

In 1977, Terrence Sejnowski specifically discussed precise spike-time correlations, suggesting them as a requirement for synaptic plasticity when the overall amounts of depression and potentiation due to firing rates are balanced [86]. In 1993, Wulfram Gerstner & colleagues showed that plasticity based on spike times, rather than firing rates, was needed for a recurrent network to learn spatio-temporal patterns [87]. In 1996, Gerstner & colleagues proposed spike-timing plasticity models for sound localization [88]. Their timing-based plasticity rule specified the change in synaptic weight as a function of the time lag between

pre- and postsynaptic spikes, defining separate windows for potentiation and depression for positive/negative time lags:

$$\Delta \mathbf{W}_{\text{post}\leftarrow\text{pre}} = \begin{cases} f_+ \phi(t_{\text{post}} - t_{\text{pre}}), & \text{if } t_{\text{post}} - t_{\text{pre}} \geq 0 \\ -f_- \phi(t_{\text{post}} - t_{\text{pre}}), & \text{if } t_{\text{post}} - t_{\text{pre}} < 0, \end{cases} . \tag{1.4}$$

where $f_\pm$ sets the amplitude of synaptic weight changes and the function ϕ sets their time-dependence. Such models, by design, capture the results of pair-based plasticity protocols. This classic study is a notable example of theory predicting experiments in neuroscience.

When the synaptic amplitude changes are a function *only* of the time lag between pre- and postsynaptic spikes, synaptic weights can potentiate or depress unlimitedly, necessitating the imposition of bounds reflecting the finite resources available to a synapse. This occurs when the plasticity rule is Hebbian, so that pre-post spike pairs lead to potentiation and post-pre spike pairs lead to depression - rewarding presynaptic activity that effectively drives postsynaptic spikes [89]. This gives rise to bimodal distributions of synaptic weights, clustered near the upper and lower bounds. Intracellular recordings, however, reveals synaptic weights with unimodal distributions [90, 91, 92]. Introducing weight-dependence into the plasticity rule can resolve this issue [93, 94, 95], although potentially at the expense of learned stimulus specificity [96, 97]. Alternatively to examining the stability of the weight dynamics, supplementing the pair-based plasticity rule with additional dynamics driven by individual pre- or post-synaptic spikes can stabilize the resulting output firing rate [98, 99, 100]. Such non-Hebbian LTP/D has been observed at the neuromuscular junction [101] and in the hippocampus [102] and cerebellum & entorhinal cortex [103, 104], but is often not observed in neocortex [42].

Pair-based STDP rules capture the results of pair-based plasticity experiments, but how well do they generalize to predicting the results of more complicated training paradigms based off of triplets, quadruplets, or bursts or spikes? Unfortunately, the answer is: not well [39, 38, 41, 40]. In addition, the results of STDP experiments are sensitive to how frequently spike pairs are presented. Spike pairs presented at high frequency tend to cause potentiation, while low frequency spike pairs tend to cause depression [105]. Mechanistic models, such as those discussed above, can capture these complexities but are complex and

difficult to analyze. Recent studies have proposed phenomenological plasticity models based on spike times but explicitly considering beyond-pairwise interactions. Two popular models are the triplet-based plasticity rule of Pfister & Gerstner [106] and the spike-voltage rule of Clopath & Gerstner [107, 108]. Under certain assumptions, a BCM-style rate-based plasticity rule can be derived from each of these spike-based models.

Here, we briefly summarize one of these models, the triplet rule of [106]. This rule reads: $\Delta \mathbf{W}_{\text{post}\leftarrow\text{pre}} = L(s_1) + Q(s_1, s_2)$, where $L(s_1)$ is the pair-based STDP described above (Eq. (1.4) with $s_1 = t_{\text{post}} - t_{\text{pre}}$. $Q(s_1, s_2) = f_3\phi(s_1)\phi_2(s_2)$ described the contribution of spike triplets, with $s_2 = t_{\text{post }2} - t_{\text{post }1}$. The functions ϕ_1 and ϕ_2 are commonly taken to be exponential windows. Fitting the parameters of this rule to explain data from V1 slice experiments [40] gave Pfister & Gerstner a model composed of pair-based LTD and triplet-based LTP:

$$L(s) = -f_- e^{-\frac{s_1}{\tau_-}} \tag{1.5}$$

for $s_1 < 0$ and $L(s) = 0$ for $s_1 \geq 0$, and

$$Q(s_1, s_2) = f_3 e^{-\frac{s_1}{\tau_+}} e^{-\frac{s_2}{\tau_3}} \tag{1.6}$$

for $s_1 \geq 0$, $s_2 \geq 0$ (and $Q(s_1, s_2) = 0$ otherwise). This only takes into account post-pre-post and post-post-pre spike patterns. While there are other potential contributors (which the triplet rule can easily include), this simple form has been shown to well fit experimental data [106]. When individual changes in the synaptic weights are small, the synaptic weights evolve on a slower timescale than the spike-train correlations and are governed by [109]:

$$\frac{d\mathbf{W}_{ij}}{dt} = \int_{-\infty}^{\infty} L(s\left(r_i r_j + \mathbf{C}_{ij}(s)\right) ds + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} Q(s_1, s_2)\mathbf{\Gamma}(s_1, s_2) ds_1 ds_2 \tag{1.7}$$

where $\mathbf{C}_{ij}$ is the cross-covariance function of the pre- and postsynaptic spike trains and $\mathbf{\Gamma}(s_1, s_2)$ is the third-order correlation, $\Gamma(s_1, s_2) = \langle y_i(t) y_j(t - s_1) y_i(t - s_2)\rangle$ where y_i is the spike train of neuron i. The inclusion of this third-order correlation is a significant complication in developing theories for the triplet rule. Recent progress has been made on adapting methods of statistical field theory to estimate these triplet correlations [110, 111, 112, 113]. The simplest possible such theory examines the zeroth order contributions from the firing rates, $r_i^2 r_j$. The relative importance of the rate- and correlation-based contributions

to plasticity depend on the magnitude of the correlations, appropriately weighted by the plasticity rule. If the correlations are very weak, or the STDP rule strongly favors either potentiation or depression, then chance spike coincidences are the dominant contribution to plasticity. The latter argument is one we pursue in more depth later in this thesis. Here, for simplicity, we examine the former. If the pre- and post-synaptic spike trains are uncorrelated Poisson processes, then the rate-based terms are the only contribution to the pairwise and triplet terms of plasticity, yielding for the triplet rule:

$$\frac{d\mathbf{W}_{ij}}{dt} = -f_{-}\tau_{-}r_i r_j + f_3\tau_{+}\tau_3 r_i^2 r_j \tag{1.8}$$

Taking the amplitudes of potentiation and depression, f_{-}, to depend on the postsynaptic firing rate r_i allows these dynamics to be written exactly in the form of the BCM rule, linking spike-based plasticity to the computational capacity of rate-based learning [106]. This link requires, however, a strong assumption about the pre- and postsynaptic spikes trains (that they are uncorrelated Poisson processes). How spatial or temporal correlations in spike trains affect the computations they allow neural circuits to perform has only just begun to be studied [114].

Throughout the body of this thesis, we make use of pair-based STDP rules, neglecting higher-order spike interactions and rate-dependence of STDP. This corresponds to assuming that the dynamics of Eq. (1.8) are negligible. For a fixed input rate r_i, those dynamics is characterized by two fixed points (Figure 1.1). These dynamics is apparently unstable - high firing rates will lead to potentiation, which in a recurrent network would further increase firing rates, leading to blow-up potentiation and pathological activity. Other forms of heterosynaptic plasticity can stabilize the weight and firing rates, providing a stable fixed point for the synaptic weights at intermediate firing rates through non-Hebbian interactions [115]. If these rate-based dynamics is sufficiently close to a fixed point, the higher-order contributions from spike-train correlations will come into play. How close the rate-based terms must be depends on the magnitude of the correlations.

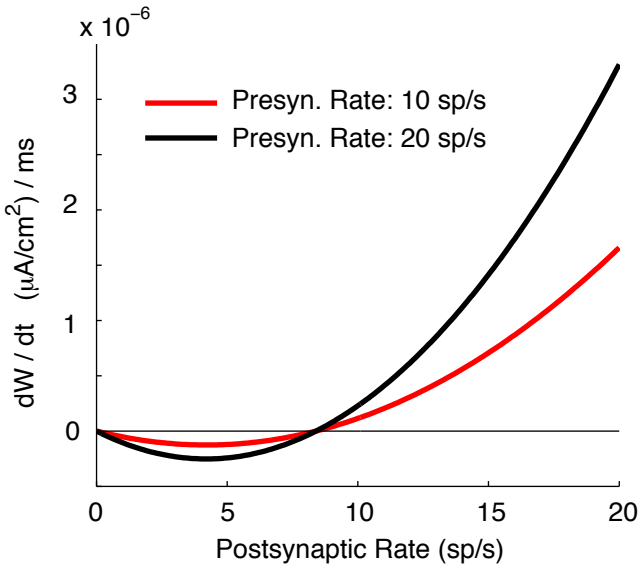

Figure 1.1: Rate-based contribution to triplet STDP [106] determined by solving Eq. (1.8) with $\tau_- = 30$ ms, $\tau_3 = 120$ ms. f_-, f_3 set as f_-, f_+ in Chap. 4.

Theories for internally-generated spike-time covariability in recurrent networks have only recently been developed. They provide an exciting avenue for studying how spike-time correlations contribute to synaptic plasticity in recurrent networks. Throughout this thesis, I will examine the effect of pair-based STDP rules on the coevolution of network structure and spiking activity. Further studies should be based on plasticity models that dynamically account for interactions between rate- and correlation-based effects, such as the triplet model - these will rely on advances in theories of spiking covariability in recurrent networks. Many questions remain to be answered; the frontier of network-level plasticity is wide open.

1.5 OUTLINE OF THIS THESIS

In this thesis I will present the results of two studies, one of which is published. In Chapter 2 I develop a self-consistent theory for STDP in recurrent networks, making use of recent developments in calculating spike-train correlations for recurrently connected, condunctace-based neuron models. The subsequent chapters leverage this work in order to develop simple theories predicting the plasticity of different types of connectivity pattern in networks. Chapter 3 discusses connectivity patterns composed of two synapses and how they are promoted or suppressed by plasticity during spontaneous activity. These two chapters have been previously published as [116] and were the result of a collaboration between myself, Ashok Litwin-Kumar and our advisor Brent Doiron. In Chapter 4, I present recent work examining how spontaneous and stimulus-driven activity can interact to produce strongly interconnected groups of neurons with shared stimulus preferences. This work was done in collaboration with my advisor.

2.0 SPIKE TIMING-DEPENDENT PLASTICITY IN RECURRENT NETWORKS

2.1 INTRODUCTION

While neural architecture undoubtedly plays a strong role in determining neuronal activity, the reverse is also true. Individual synapses can both potentiate (strengthen) and depress (weaken), and whether they do so depends on the relative timing of action potentials in the connected neurons [37, 36]. Such *spike timing-dependent plasticity* (STDP) has featured prominently in both experimental and theoretical studies of neural circuits [81, 117, 13]. Of particular interest, STDP provides a mechanism for Hebbian plasticity: synaptic potentiation occurs when a presynaptic neuron reliably drives spike responses from a postsynaptic neuron, while anti-causal spike pairs result in synaptic depression [89]. Hebbian plasticity provides a potential link between circuit structure and function through the formation of heavily wired assemblies of neurons, where assembly membership is associated with coordinated, elevated firing rates during a specific computation [118]. Evidence supporting this idea, originally proposed by Hebb [15], has been found in both hippocampus [119] and sensory cortex [120].

Despite the promise of STDP to provide insight into the functional wiring of large neural circuits, many studies of STDP have concentrated on the plasticity of synaptic connections between just a single pair of pre- and postsynaptic neurons, often focusing on the distribution of individual synaptic weights [89, 99, 121, 95, 93]. Other studies have shown that multiple temporally correlated inputs to a neuron will cooperate to potentiate, while uncorrelated inputs may depress [109, 89, 122, 123]. In this case STDP can generate feedforward circuits [124], which while important for the propagation of neural activity [125], are unlike the recurrent structure of the neocortex. Understanding the two-way interaction between plastic

recurrent network structure and spiking activity recruited in recurrent circuits is thus a central focus for theories of synaptic plasticity.

Due to this challenge, many studies have resorted to large-scale numerical simulations of cortical networks with plastic synapses [126, 127, 128, 129]. While intuition for the development of circuit structure can be gained using this approach, without a governing theoretical framework it is often difficult to extract generalized principles. Alternatively, mathematical analyses have been restricted to either small networks [130, 128], or have required the assumption that neurons fire as Poisson processes [131, 132, 123, 133]. These latter works assumed shared inputs from outside the network to be the only source of correlated spiking activity, neglecting covariance originating from recurrent coupling. Thus, there is a need for a coherent mathematical framework that captures how STDP drives self-organization of circuit structure in recurrent cortical networks.

To this end, we construct a self-consistent theory for the coevolution of spiking statistics and synaptic weights in networks with STDP. This theory makes use of a previously developed linear response framework for calculating joint spiking statistics [134, 135, 136] and a separation of timescales between spiking covariance and synaptic plasticity [109]. Most previous studies of plasticity in recurrent networks have focused on how they can be trained to represent an external stimulus. We focus on how spiking covariance generated by coupling within the network interacts with plasticity to shape network structure.

2.2 RESULTS

This chapter presents the development of a self-consistent theory for spike timing-dependent plasticity describing the plasticity of each synaptic weight in a recurrent network. It relies on two key results: 1) a theory for plasticity of individual synapses which links their potentiation or depression to the temporal correlations in their spike trains [109], and 2) a method for predicting those correlations in recurrent networks [136].

2.2.1 Spike train covariance determines synaptic plasticity

We begin by reviewing a well-studied phenomenological model of STDP [88], acting within a simple circuit of two reciprocally coupled neurons. Consider a pair of pre- and postsynaptic spike times with time lag $s = t_{\text{post}} - t_{\text{pre}}$. The evolution of the synaptic weight connecting presynaptic neuron j to postsynaptic neuron i obeys $\mathbf{W}_{ij} \to \mathbf{W}_{ij} + L(s)$, with the STDP rule $L(s)$ (Figure 2.2.1A) being Hebbian:

$$L(s) = \begin{cases} \mathcal{H}(W^{\max} - \mathbf{W}_{ij}) f_+ e^{-\frac{|s|}{\tau_+}}, & \text{if } s \geq 0 \\ \mathcal{H}(\mathbf{W}_{ij})\,(-f_-)\, e^{-\frac{|s|}{\tau_-}}, & \text{if } s < 0, \end{cases} . \tag{2.1}$$

Here $\mathcal{H}(x) = 1$ if $x > 0$ while $\mathcal{H}(x) = 0$ if $x \leq 0$, imposing bounds on the weights to prevent the magnitude of excitatory synapses from becoming negative or potentiating without bound (i.e. $0 \leq \mathbf{W}_{ij} \leq W^{\max}$). The coefficients $f_\pm$ scale the amplitude of weight changes induced by individual pre-post spike pairs and $\tau_\pm$ determine how synchronous pre- and postsynaptic spikes must be to drive plasticity.

The spike train from neuron i is the point process $\mathbf{y}_i(t) = \sum_k \delta(t - t_i^k)$, with t_i^k being the k^{th} spike time of neuron i. Following [109] we relate the joint statistics of $\mathbf{y}_i(t)$ and $\mathbf{y}_j(t)$ to the evolution of synaptic weights. We assume that individual pre-post spike pairs induce small changes in synaptic weights ($f_\pm \ll W^{\max}$). This makes synaptic weights evolve slowly, on a much longer timescale than the millisecond scale of pairwise spiking covariance due to network interactions. The separation of timescales between synaptic plasticity and spiking activity provides an approximation to the evolution of the synaptic weights (Methods: Learning dynamics, 2.4.2):

$$\frac{d\mathbf{W}_{ij}}{dt} = \mathbf{W}_{ij}^0 \int_{-\infty}^{\infty} L(s)\big(r_i r_j + \mathbf{C}_{ij}(s)\big) ds. \tag{2.2}$$

Here r_i is the time-averaged firing rate of neuron i, and $\mathbf{C}_{ij}(s) = \langle (\mathbf{y}_i(t) - r_i)(\mathbf{y}_j(t+s) - r_j) \rangle$ is the cross-covariance function of neuron i and j's spike trains. The separation of timescales allows us to calculate the equilibrium spiking statistics $\mathbf{C}$, taking $\mathbf{W}$ to be constant on the timescale of $\mathbf{C}(s)$. The term $r_i r_j$ in Eq. (2.2) captures the contribution of chance spike coincidences to STDP, while $\mathbf{C}_{ij}(s)$ models the sensitivity of STDP to spike time correlations.

Finally, $\mathbf{W}^0$ is the adjacency matrix of the network – a binary matrix with $\mathbf{W}^0_{ij} = 1$ denoting the presence of a synapse from neuron j to neuron i. Multiplying by $\mathbf{W}^0_{ij}$ ensures that synapses that do not exist cannot potentiate into existence. Eq. (2.2) requires only the first and second order joint spiking statistics. To facilitate calculations, many previous studies have used Poisson neuron models with a specified r_i and $\mathbf{C}_{ij}(s)$ to generate $\mathbf{y}_i(t)$. In contrast, we will use a white noise-driven exponential integrate-and-fire model [137] for the generation of spike times (Methods: Neuron and network model). While this complicates the calculation of the spike train statistics, it provides a more biophysically realistic model of neural dynamics [138, 139] that better captures the timescales and neuronal nonlinearities that shape r_i and $\mathbf{C}_{ij}(s)$. In total, the above theory determines synaptic evolution from the integrated combination of an STDP rule $L(s)$ and the spike train cross-covariance function $\mathbf{C}_{ij}(s)$. Thus, any mechanism affecting two neurons' spiking covariance is expected to shape network structure through STDP.

As a simple illustration of how spiking correlations can drive STDP, we examined the synaptic weight dynamics, $\mathbf{W}_{12}(t)$ and $\mathbf{W}_{21}(t)$, in a reciprocally coupled pair of neurons, both in the presence and absence of common inputs. Specifically, the fluctuating input to neuron i was the sum of a private and common term, $\sqrt{1-c}\xi_i(t) + \sqrt{c}\xi_c(t)$, with c being the fraction of shared input to the neurons. In the absence of common input ($c = 0$; Figure 2.2.1B), the two synapses behaved as expected with Hebbian STDP: one synapse potentiated and the other depressed (Figure 2.2.1C). The presence of common input ($c = 0.05$) was a source of synchrony in the two neurons' spike trains, inducing a central peak in the spike train cross-covariance function $\mathbf{C}_{ij}(s)$ (Figure 2.2.1B vs 1D). In response to this increased synchrony both synapses potentiated (Figure 2.2.1E), in contrast to the case with $c = 0$. This was because of the sharp potentiation side of the learning rule compared to the the depression side (Figure 2.2.1A), so that increased spike synchrony enhanced the degree of overlap between $\mathbf{C}_{ij}(s)$ and the potentiation component of $L(s)$. This overcame the effects of depression in the initially weaker synapse and promoted strong, bidirectional connectivity in the two-neuron circuit.

This example highlights how the temporal shape of the spike train cross-covariance function, $\mathbf{C}_{ij}(s)$, can interact with the shape of the learning rule, $L(s)$, to direct STDP. However,

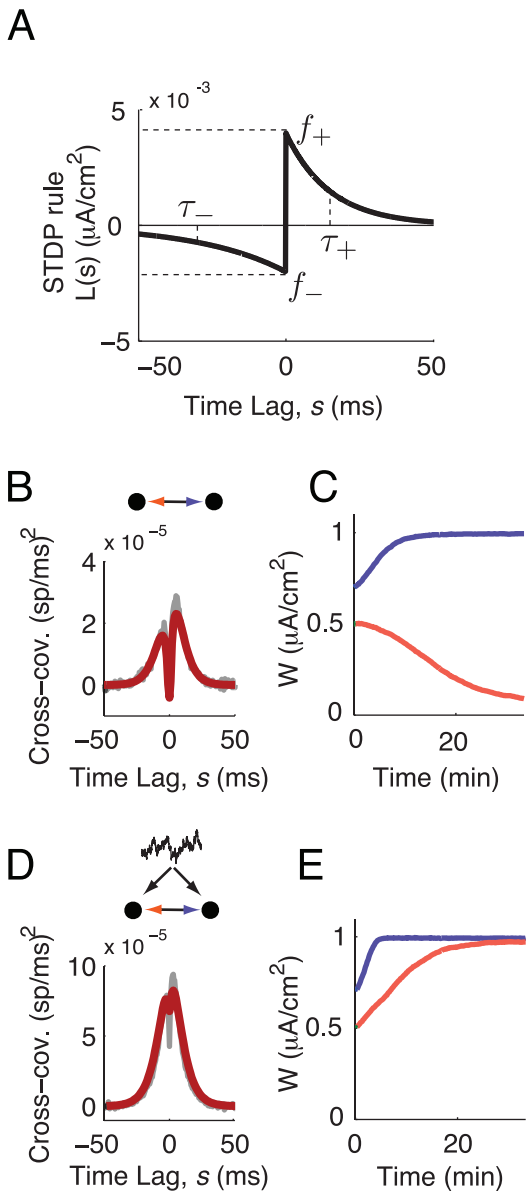

Figure 2.1: **Network structure shapes synaptic plasticity.** (A) The STDP rule, $L(s)$, is composed of exponential windows for depression (-) and potentiation (+). Each is defined by its amplitude $f_\pm$ and timescale $\tau_\pm$. (B) Spike train cross-covariance function for a pair of neurons with no common input, so that synapses between the two neurons are the only source of spiking covariance. Shaded lines: simulation, solid lines: theory (Eq. (2.4)). (C,E) Synaptic weight (peak EPSC amplitude) as a function of time in the absence (C) and presence (E) of common input. (D) Spike train cross-covariance function for a pair of neurons with common input, $c = 0.05$. Common input was modeled as an Ornstein-Uhlenbeck process with a 5 ms timescale.

this case only considered the effect of correlated inputs from outside of the modeled circuit (Figure 2.2.1). Our primary goal is to predict how spiking covariance due to internal network interactions combines with STDP to drive self-organized network structure. In order to do this, we first require a theory for predicting the spiking covariance between all neuron pairs given a static, recurrent connectivity. Once this theory has been developed, we will use it to study the case of plastic connectivity.

2.2.2 Network architecture determines spiking covariance in static networks

In this section we review approximation methods [134, 135] that estimate the pairwise spike train cross-covariances $\mathbf{C}_{ij}(s)$ using a static weight matrix $\mathbf{W}$ (see Methods: Spiking statistics for a more full description). The exposition is simplified if we consider the Fourier transform of a spike train, $\mathbf{y}_i(\omega) = \int_{-\infty}^{\infty} \mathbf{y}_i(t) e^{-2\pi i \omega t} dt$, where ω is frequency. Assuming weak synaptic connections $\mathbf{W}_{ij}$, we approximate the spike response from neuron i as:

$$\mathbf{y}_i(\omega) = \mathbf{y}_i^0(\omega) + \mathbf{A}_i(\omega) \left(\sum_{j=1}^{N} \mathbf{W}_{ij} J(\omega) \mathbf{y}_j(\omega) \right). \tag{2.3}$$

The function $\mathbf{A}_i(\omega)$ is the linear response [140] of the postsynaptic neuron, measuring how strongly modulations in synaptic currents at frequency ω are transferred into modulations of instantaneous firing rate about a background state $\mathbf{y}_i^0$. The function $J(\omega)$ is a synaptic filter. In brief, Eq. (2.3) is a linear ansatz for how a neuron integrates and transforms a realization of synaptic input into a spike train.

Following [134, 135, 136] we use this linear approximation to estimate the Fourier transform of $\mathbf{C}_{ij}(s)$, written as $\mathbf{C}_{ij}(\omega) = \langle \mathbf{y}_i(\omega) \mathbf{y}_j^*(\omega) \rangle$; here $\mathbf{y}^*$ denotes the conjugate transpose. This yields the following matrix equation:

$$\mathbf{C}(\omega) = \Big(\mathbf{I} - \big(\mathbf{W} \cdot \mathbf{K}(\omega) \big) \Big)^{-1} \mathbf{C}^0(\omega) \Big(\mathbf{I} - \big(\mathbf{W} \cdot \mathbf{K}(\omega) \big)^* \Big)^{-1}, \tag{2.4}$$

where $\mathbf{K}(\omega)$ is an interaction matrix defined by $\mathbf{K}_{ij}(\omega) = \mathbf{A}_i(\omega) \mathbf{J}_{ij}(\omega)$ and we use $\mathbf{W} \cdot \mathbf{K}$ to denote element-wise multiplication. The matrix $\mathbf{C}^0(\omega)$ is the covariance in the absence of synaptic coupling, with elements $\mathbf{C}_{ij}^0(\omega) = \langle \mathbf{y}_i^0(\omega) \mathbf{y}_j^{0*}(\omega) \rangle$, and $\mathbf{I}$ is the identity matrix. Using Eq. (2.4) we recover the matrix of spike train cross-covariance functions $\mathbf{C}(s)$ by inverse Fourier transformation. Thus, Eq. (2.4) provides an estimate of the statistics of pairwise spiking activity in the full network, taking into account the network structure. In Appendix C we present an alternative, constructive derivation of Eq. (2.4).

As a demonstration of the theory, we examined the spiking covariances of three neurons from a 1,000-neuron network (Figure 2.2A, colored neurons). The synaptic weight matrix $\mathbf{W}$ was static and had an adjacency matrix $\mathbf{W}^0$ that was randomly generated with Erdös-Rényi statistics (connection probability of 0.15). The neurons received no correlated input

from outside the network, making $\mathbf{C}^0(\omega)$ a diagonal matrix, and thus recurrent network interactions were the only source of spiking covariance. Neuron pairs that connected reciprocally with equal synaptic weights had temporally symmetric spike train cross-covariance functions (Figure 2.2C), while unidirectional connections gave rise to temporally asymmetric cross-covariances (Figure 2.2D). When neurons were not directly connected, their covariance was weaker than that of directly connected neurons but was still nonzero (Figure 2.2E). The theoretical estimate provided by Eq. (2.4) was in good agreement with estimates from direct simulations of the network (Figure 2.2C,D,E red vs. gray curves).

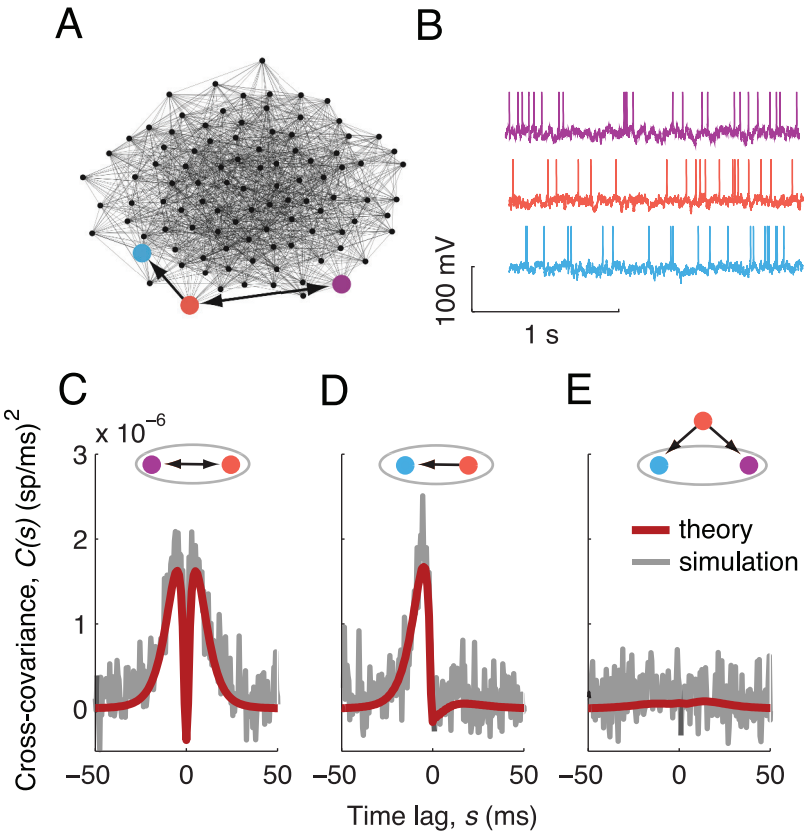

Figure 2.2: **Linear response theory for spiking covariances.** (A) Illustration of the network connectivity for a subset of 100 neurons. Three neurons, and the connections between them, are highlighted. Nodes are positioned by the Fruchterman-Reingold force algorithm. (B) Example voltage traces for the three highlighted neurons. (C-E) Spike train cross-covariance functions for the three combinations of labeled neurons. Top: A shaded ellipse contains the pair of neurons whose cross-covariance is shown. Shaded lines: simulations, red lines: linear response theory.

2.2.3 Self-consistent theory for network structure and spiking covariance with plastic synapses

In general, it is challenging to develop theoretical techniques for stochastic systems with several variables and nonlinear coupling [140], such as in Eq. (2.2). Fortunately, in our model the timescale of spiking covariance in the recurrent network with static synapses is on the order of milliseconds (Figure 2.2C,D,E), while the timescale of plasticity is minutes (Figure 2.2.1C,E). This separation of timescales provides an opportunity for a self-consistent theory for the coevolution of $\mathbf{C}(s)$ and $\mathbf{W}(t)$. That is, so long as $f_{\pm}$ in Eq. (2.1) (the individual changes in synaptic weights) are sufficiently small, the spike-train covariances adjust instantaneously on the timescale of $\mathbf{W}$, so that we can insert Eq. (2.4) into Eq. (2.2). The resulting system yields a solution $\mathbf{W}(t)$ that captures the long timescale dynamics of the plastic network structure (Methods: Self-consistent theory for network plasticity).

As a first illustration of our theory, we focus on the evolution of three synaptic weights in a 1,000-neuron network (Figure 2.2.3A, colored arrows). The combination of Eqs. (2.2) and (2.4) predicted the dynamics of $\mathbf{W}(t)$, whether the weight decreased with time (Figure 2.2.3B left, red curve), increased with time (Figure 2.2.3C left, red curve), or remained approximately constant (Figure 2.2.3D left, red curve). In all three cases, the theory matched well the average evolution of the synaptic weight estimated from direct simulations of the spiking network (Figure 2.2.3B,C,D left, thick black curves). Snapshots of the network at three time points (axis arrows in Figure 2.2.3B,C,D, left), showed that $\mathbf{W}$ coevolved with the spiking covariance (Figure 2.2.3B,C,D right). We remark that for any realization of background input $\mathbf{y}^0(t)$, the synaptic weights $\mathbf{W}(t)$ deviated from their average value with increasing spread (Figure 2.2.3B,C,D left, thin black curves). This is expected since $\mathbf{C}(t)$ is an average over realizations of $\mathbf{y}^0(t)$, and thus provides only a prediction for the drift of $\mathbf{W}(t)$, while the stochastic nature of spike times leads to diffusion of $\mathbf{W}(t)$ around this drift [109]. In sum, the fast-slow decomposition of spiking covariance and synaptic plasticity provides a coherent theoretical framework to investigate the formation of network structure through STDP.

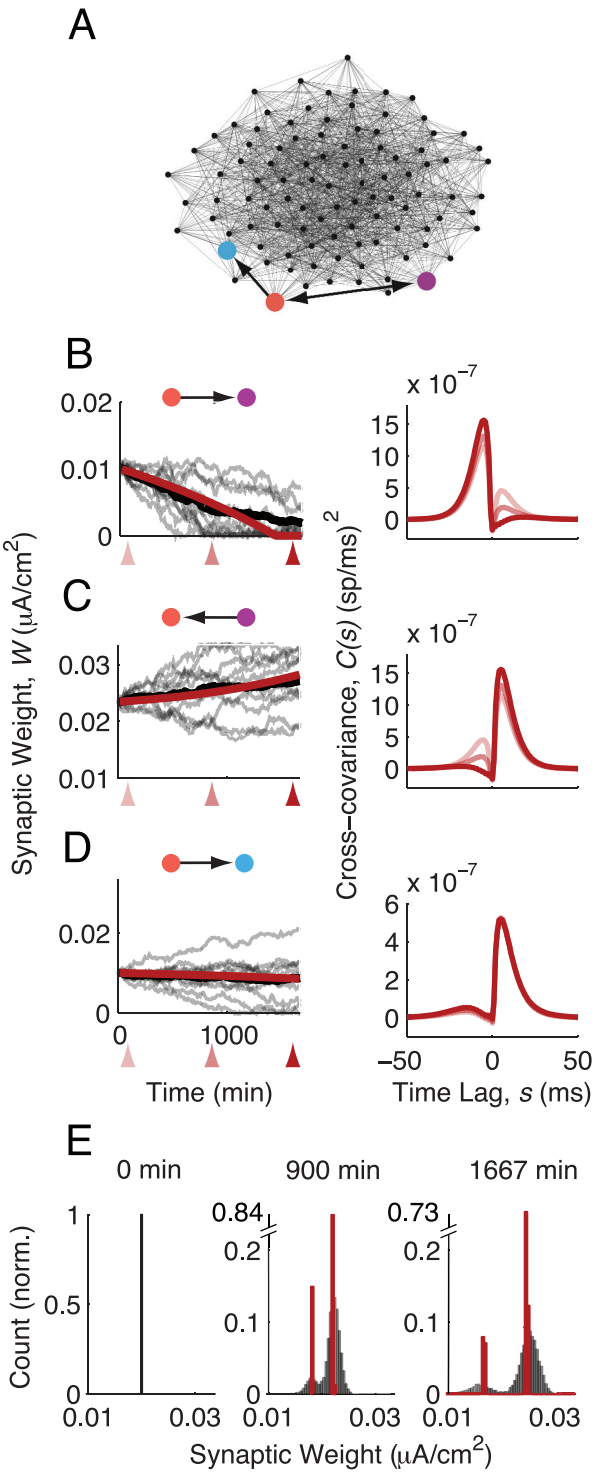

Figure 2.3: **STDP in recurrent networks with internally generated spiking covariance.** (A) As in Fig. 2.2A. (B-D) Left, Synaptic weight versus time for each of the three synapses in the highlighted network. Shaded lines: simulation, individual trials of the same initial network. Solid black lines: simulation, trial-average. Solid red lines: theory. Right, spike train cross-covariances at the three time points marked on the left (linear response theory). (E) Histogram of synaptic weights at three time points. Red, theory. Shaded: simulation.

2.3 DISCUSSION

Our treatment is complementary to past studies on STDP that focused on the development of architecture through external input [131, 132, 123, 141]. We restrict our analysis to networks with only internally generated correlations (i.e. $\mathbf{C}^0_{ij}(s) = \langle \mathbf{y}^0_i(t+s)\mathbf{y}^0_j(t)\rangle = 0$ for $i \neq j$), and thus focus on the formation of self-organized structure through STDP. A consequence of this modeling choice is low values of spiking correlations within the network: mean spike count correlation coefficients computed from all pairs within the network were approximately 5×10^{-4}, and when conditioned on cell pairs having a direct connection between them were 4×10^{-3} (S1 Text, 2.5.1). These low values agree with reports from unanethesized animals performing simple fixation task [142], or recordings restricted to cortical granule layers [143, 144], however a large number of other studies report significantly higher mean values of correlated activity [145].

There are several ways to increase the spiking correlations in our model. One is to assume weak external correlations in the background state (i.e. $\mathbf{C}^0_{ij}(s) = \langle \mathbf{y}^0_i(t+s)\mathbf{y}^0_j(t)\rangle > 0$ for $i \neq j$); this has been the focus of several past studies [134, 135]. Another is to reduce network size N to amplify any internally generated correlations within the network. When the network size was reduced from 1,000 to 100 the mean spike count correlation increased to 0.005 across all pairs and to 0.03 for directly coupled pairs (S1 Text, 2.5.1). Despite these larger correlations, our self-consistent theory (Eqs. (2.2) and (2.4)) predicted well the evolution of synaptic weights (S1 Fig, 2.5.2). This reduction in N also increased the speed of learning by a factor of 10, however the separation of timescales required was still valid.

A recent suite of studies derived a theory for how STDP shapes the full structure of networks of neurons whose spike trains are Poisson processes [131, 141, 132, 100, 123, 133]. The initial approach is similar to ours with a linear approximation to estimate spiking covariance (see Eq. (2.3)-(2.4)). However, these studies mostly focused on the effects of external input, considering how feedforward inputs entrain structure in recurrent synapses [141, 123, 132]. The only source of spiking covariance was inherited from external sources (meaning $\mathbf{C}_0(s)$ has off-diagonal structure), and subsequently filtered by the network to produce spiking covariance. Two previous studies by the same authors also examined STDP in networks without

external stimuli [131, 100]; these took a large system size limit ($N \to \infty$) and neglected the contribution of spiking covariance to STDP, focusing on the firing rate dependence due to an unbalanced learning rule.

While our theory gives an accurate description of plasticity in the network, it is nevertheless high-dimensional. Keeping track of every individual synaptic weight and spike train cross-covariance function involves $\mathcal{O}(N^2)$ variables. For large networks, this becomes computationally challenging. More importantly, this high-dimensional theory does not provide direct dynamical insights in the plasticity of connectivity patterns. In the next chapters, we will develop principled approximations of the high-dimensional system to closed low-dimensional theories describing the evolution of different types of structures within recurrent networks.

2.4 METHODS

2.4.1 Neuron and network model

We model a network of N neurons. The membrane dynamics of individual neurons obey the exponential integrate-and-fire (EIF) model [137], one of a class of models well-known to capture the spike initiation dynamics and statistics of cortical neurons [138, 139]. Specifically, the membrane voltage of neuron i evolves according to:

$$C\frac{dV_i}{dt} = g_L\left(V_L - V_i\right) + g_L\Delta\exp\left(\frac{V_i - V_T}{\Delta}\right) + I_{\mathrm{i}}(t) + \sum_{j=1}^{N}\mathbf{W}_{ij}\left(\mathbf{J}_{ij} * y_j.\right). \tag{2.5}$$

The first term on the right-hand side is the leak current, with conductance g_L and reversal potential V_L. The next term describes a phenomenological action potential with an initiation threshold V_T and steepness Δ: when the voltage reaches V_T, it diverges; this divergence marks an action potential. For numerical simulations, action potentials are thresholded at $V(t) = V_{\mathrm{th}}$, reset to a reset potential V_{re} and held there for an absolute refractory period τ_{ref}. The refractory period plays a role in shaping single-neuron input-output transfer, although in the noise-driven regime we consider it may have little qualitative effect.

Input from external sources not included in the model network is contained in $I_{\mathrm{i}}(t)$. We model this as a Gaussian white noise process: $I_i(t) = \mu + g_L\sigma D\xi_i(t)$. The mean of the the external input current is μ. The parameter σ controls the strength of the noise and $D = \sqrt{\frac{2C}{g_L}}$ scales the noise amplitude to be independent of the passive membrane time constant. With this scaling, the infinitesimal variance of the passive membrane voltage is $(g_L\sigma D)^2$.

The last term of Eq. (2.5) models synaptic interactions in the network. The $N\times N$ matrix $\mathbf{W}$ contains the amplitudes of each synapse's postsynaptic currents. It is a weighted version of the binary adjacency matrix $\mathbf{W}^0$, where $\mathbf{W}^0_{ij} = 1(0)$ indicates the presence (absence) of a synapse from neuron j onto neuron i. If a synapse ij is present then $\mathbf{W}_{ij}$ denotes its strength. Due to synaptic plasticity, $\mathbf{W}$ is dynamic: it changes in time as individual synapses potentiate or depress. The spike train from neuron j is the point process $y_j(t) = \sum_k \delta(t - t_j^k)$, where t_j^k denotes the k^{th} spike time from neuron j. The $N \times N$ matrix $\mathbf{J}(t)$ defines the shape of the postsynaptic currents. In this study, we use exponential synapses: $\mathbf{J}_{ij}(t-t_j^k) = \mathcal{H}(t-t_j^k)\exp\left(-\frac{t-t_j^k}{\tau_{\mathrm{S}}}\right)$, where $\mathcal{H}(t)$ is the Heaviside step function. Our theory is not exclusive to the EIF model or to the simple synaptic kernels we used; similar methods can be used with any integrate-and-fire model and arbitrary synaptic kernels. Model parameters are contained in Table 2.1 (unless specified otherwise in the text). In simulations, we took all synapses to initially have the same weight.

2.4.2 Learning dynamics

We now derive Eq. (2.2), summarizing a key result of [109]. Changes in a synaptic weight $\mathbf{W}_{ij}$ are governed by the learning rule $L(s)$, Eq. (2.1). We begin by considering the total change in synaptic weight during an interval of length T ms:

$$\Delta\mathbf{W}_{ij} = \mathbf{W}^0_{ij}\int_t^{t+T}\int_t^{t+T} L(t''-t')y_j(t'')y_i(t')dt''dt' \tag{2.6}$$

where multiplying by the corresponding element of the adjacency matrix ensures that nonexistent synapses do not potentiate into existence. Consider the trial-averaged rate of change:

$$\frac{\langle\Delta\mathbf{W}_{ij}\rangle}{T} = \mathbf{W}^0_{ij}\frac{1}{T}\int_t^{t+T}\int_{t-t'}^{t+T-t'} L(s)\langle y_j(t'+s)y_i(t')\rangle dsdt' \tag{2.7}$$

Table 2.1: **Model parameters**

Parameter	Description	Value
C	Membrane capacitance	1 μF/cm^2
g_L	Leak conductance	0.1mS/cm^2
V_L	Leak reversal potential	-72 mV
Δ	Action potential steepness	1.4 mV
V_T	Action potential initiation threshold	-48 mV
V_{th}	Action potential threshold	30 mV
V_{re}	Action potential reset	-72 mV
τ_{ref}	Action potential width	2 ms
μ	External input mean	1 μA/cm^2
σ	External input standard deviation	9 mV
N	Number of neurons	1000
p_0	Connection density	.15
$W^{\max}$	Maximum synaptic weight	5 μA/cm^2
τ_{S}	Synaptic time constant	5 ms

where $s = t'' - t'$ and $\langle\cdot\rangle$ denotes the trial average. We first note that this contains the definition of the trial-averaged spike train cross-covariance:

$$\mathbf{C}_{ij}(s) = \frac{1}{T}\int_t^{t+T} \langle y_j(t'+s) y_i(t')\rangle dt' - r_i r_j \tag{2.8}$$

where r_i is the time-averaged firing rate of neuron i and subtracting off the product of the rates corrects for chance spike coincidences. Inserting this definition into Eq. (2.7) yields:

$$\frac{\langle \Delta \mathbf{W}_{ij}\rangle}{T} = \mathbf{W}^0_{ij}\int_{t-t'}^{t+T-t'} L(s)\left(r_i r_j + \mathbf{C}_{ij}(s)\right) ds \tag{2.9}$$

We then take the amplitude of individual changes in the synaptic weights to be small: $f_+, f_- << W^{\max}$, where $\tau_\pm$ define the temporal shape of the STDP rule (see Eq. (2.1)). In

this case, changes in the weights occur on a slower timescale than the width of the learning rule. Taking $T >> \max(\tau_+, \tau_-)$ allows us to extend the limits of integration in Eq. (2.9) to $\pm\infty$, which gives Eq. (2.2). Note that in the results we have dropped the angle brackets for convenience. This can also be justified by the fact that the plasticity is self-averaging, since $\Delta \mathbf{W}_{ij}$ depends on the integrated changes over the period T.

2.4.3 Spiking statistics

In order to calculate $d\mathbf{W}_{ij}/dt$, we need to know the firing rates r_i, r_j and spike train cross-covariance $\mathbf{C}_{ij}(s)$ (Eq. (2.2)). We take the weights to be constant on the fast timescale of s, so that the firing rates and spike train cross-covariances are stationary on that timescale. We solve for the baseline firing rates in the network via the self-consistency relationship

$$\begin{aligned} r_i &= r_i(\mu_i^{\text{eff}}, \sigma), \text{where} \\ \mu_i^{\text{eff}} &= \mu + \sum_j \left(\int_{-\infty}^{\infty} \mathbf{J}_{ij}(t) dt \right) \mathbf{W}_{ij} r_j \end{aligned}$$

for $i = 1, \ldots, N$. This gives the equilibrium state of each neuron's activity. In order to calculate the spike train cross-covariances, we must consider temporal fluctuations around the baseline firing rates.

With sufficiently weak synapses compared to the background input, we can linearize each neuron's activity around the baseline state. Rather than linearizing each neuron's firing rate around r_i, we follow [134, 135, 136] and linearize each neuron's spike train around a realization of background activity, the uncoupled spike train $\mathbf{y}_i^0$ (Eq. (2.3)). The perturbation around the background activity is given by each neuron's linear response function, $\mathbf{A}_i(t)$, which measures the amplitude of firing rate fluctuations in response to perturbations of each neuron's input around the baseline μ_i^{eff}. We calculate $\mathbf{A}(t)$ using standard methods based on Fokker-Planck theory for the distribution of a neuron's membrane potential [146, 147].

This yields Eq. (2.3), approximating a realization of each neuron's spike train as a mixed point and continuous process. Spike trains are defined, however, as pure point processes. Fortunately, Eq. (2.2) shows that we do not need a prediction of individual spike train

realizations, but rather of the trial-averaged spiking statistics. We can solve Eq. (2.3) for the spike trains in the frequency domain as:

$$\mathbf{y}(\omega) = (\mathbf{I} - (\mathbf{W} \cdot \mathbf{K}(\omega)))^{-1} \mathbf{y}^0(\omega)$$

where as in the Results, $\mathbf{K}(\omega)$ is an interaction matrix defined by $\mathbf{K}_{ij}(\omega) = \mathbf{A}_i(\omega)\mathbf{J}_{ij}(\omega)$ and $\cdot$ denotes the element-wise product. Averaging this expression over realizations of the background spike trains yields a linear equation for the instantaneous firing rates. Averaging the spike trains $\mathbf{y}$ against each other yields the full cross-covariance matrix, Eq. (2.4). It depends on the coupling strengths $\mathbf{W}$, the synaptic filters $\mathbf{J}_{ij}$ and neurons' linear response functions $\mathbf{A}$, and the covariance of the baseline spike trains, $\mathbf{C}^0$.

We can calculate the baseline covariance in the frequency domain, $\mathbf{C}^0(\omega) = \langle \mathbf{y}^0 \mathbf{y}^{0*} \rangle$, by first noting that it is a diagonal matrix containing each neuron's spike train power spectrum. We calculate these using the renewal relationship between the spike train power spectrum $\mathbf{C}^0(\omega)$ and the first passage time density [148]; the first passage time density for nonlinear integrate and fire models can be calculated using similar methods as for the linear response functions [147].

2.4.4 Self-consistent theory for network plasticity

We solve the system Eqs. (2.2),(2.4) for the evolution of each synaptic weight with the Euler method with a time step of 100 seconds. At every time step of the plasticity, each neuron's activity is re-linearized and the firing rates and spike train covariances recomputed. A package of code for solving the self-consistent theory and running the spiking simulations, in MATLAB and C, is available at http://sites.google.com/site/gabrielkochocker/code. Additional code is available on request.

2.5 SUPPLEMENTARY INFORMATION

2.5.1 S1 Text: Magnitude of spike count correlations

Here we compute spike count correlations for the studied networks. A spike count from neuron i, $n_i^T(t)$ is the number of spikes occurring within the window $(t, t+T)$. The covariance of neuron i and j's spike counts is

$$\mathrm{Cov}(n_i^T, n_j^T) = \langle n_i^T n_j^T \rangle - \langle n_i^T \rangle \langle n_j^T \rangle, \tag{2.10}$$

where $\langle \cdot \rangle$ denotes an average over trials. The variance of neuron i's spike count is $\mathrm{Var}(n_i^T) = \mathrm{Cov}(n_i^T, n_i^T)$. The correlation coefficient of spike counts is

$$\mathrm{Corr}(n_i^T, n_j^T) = \frac{\mathrm{Cov}(n_i^T, n_j^T)}{\sqrt{\mathrm{Var}(n_i^T)\mathrm{Var}(n_j^T)}} \tag{2.11}$$

Here, we estimate spike count correlations by computing $\mathrm{Cov}(n_i^T, n_j^T)$ via the renewal relation [148]:

$$\mathrm{Cov}(n_i^T, n_j^T) = \int_{-T}^{T} \mathbf{C}_{ij}(s)\,(T - |s|)\,ds - r_i r_j. \tag{2.12}$$

and similar for $\mathrm{Var}(n_i^T)$. We examine this estimate for the average spike count correlation (over pairs of neurons in the network) as a function of window size. For the internally generated covariability in the main paper, spike count correlations are low (S1 Figure 1A, 2.5.2).

Here, we write the linear response theory with external input correlations, for completeness. Specifically the fluctuating external input to each neuron was the sum of a private term and a globally shared term, $g_L \sigma D \left(\sqrt{(1-c)}\xi_i(t) + \sqrt{c}\xi_c(t) \right)$ (here, $\xi_i(t)$ and $\xi_c(t)$ are Gaussian white noise of unit intensity and g_L, σ and D defined as in Methods). The covariance matrix of the external inputs was $\mathbf{C}^{\mathrm{ext}}$, with $\mathbf{C}^{\mathrm{ext}}_{ij} = c g_L \sigma D$ for $i \neq j$ and $\mathbf{C}^{\mathrm{ext}}_{ii} = 1$. With correlated external inputs, the full spike-train cross-covariance matrix is given (in the Fourier domain) by [136]

$$\mathbf{C}(\omega) = \Big(\mathbf{I} - \big(\mathbf{W} \cdot \mathbf{K}(\omega) \big) \Big)^{-1} \Big(\mathbf{C}^0(\omega) + \mathbf{A}(\omega) \mathbf{C}^{\mathrm{ext}}(\omega) \mathbf{A}^*(\omega) \Big) \Big(\mathbf{I} - \big(\mathbf{W} \cdot \mathbf{K}^*(\omega) \big) \Big)^{-1} \tag{2.13}$$

As in the main text, $\mathbf{W}$ is the weight matrix, $\mathbf{K}_{ij}(\omega) = \mathbf{A}_i(\omega)\mathbf{J}_{ij}(\omega)$ is the effective interaction matrix, $\mathbf{A}(\omega)$ is a diagonal matrix containing the linear response function of each neuron. In the main text, what we refer to as the "baseline correlation" here corresponds to $\mathbf{C}^0(\omega) + \mathbf{A}(\omega)\mathbf{C}^{\text{ext}}(\omega)\mathbf{A}^*(\omega)$.

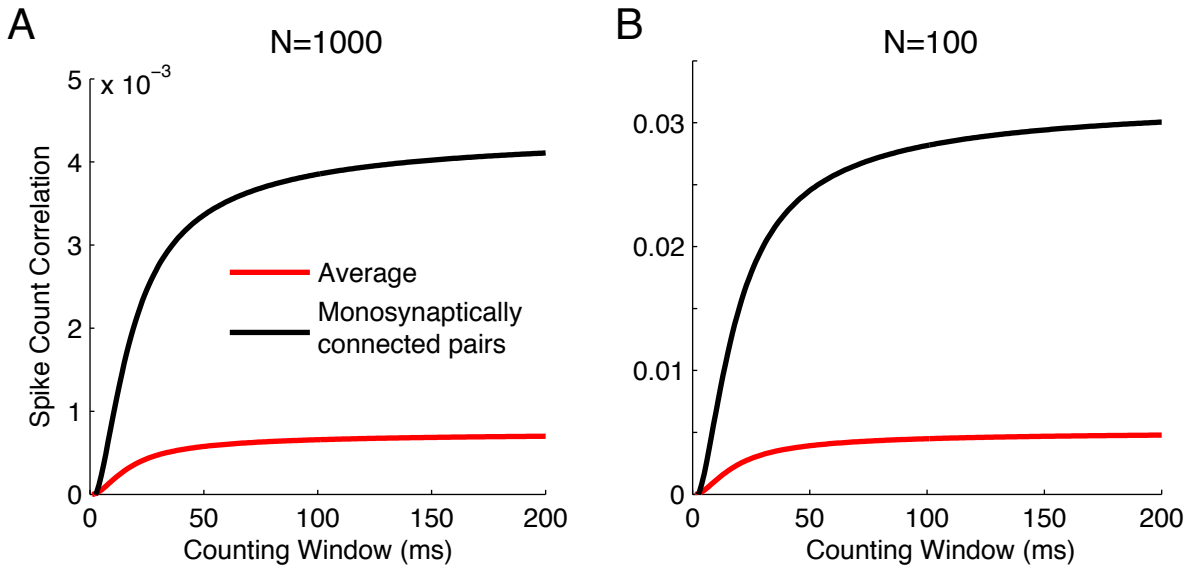

Figure 2.4: **Magnitude of spike count correlations.** Average spike count correlation as a function of the counting window size. (A) the network of Fig. 2.2. Spike count correlations are computed via the renewal relation, Eq. (2.12) using the first-order truncated approximation of the spike train cross-covariance function, Eq. (3.3). (B) A network of 100 neurons, with $W^{\max}$ increased by a factor of 10. Red, average spike count correlation across all pairs of neurons. Black, average spike count correlation in monosynaptically connected pairs.

Expanding the spike-train covariances in powers of the interactions $\mathbf{K}$ and truncating at first order yields the approximation Eq. (3.3). We compare the spike count correlations for the $N = 1000$ network of Figure 2 in the main text to a network with $N = 100$ and synaptic weights ($W^{\max}$) increased by a factor of 10 (Figure 2.5.1). As in Figs. 2.2 and 2.2.3, this network was generated with Erdös-Rényi adjacency matrix and all weights had the same initial value, $W^{\max} * .6$ In this smaller network, with stronger synaptic weights, spike count correlations were larger, with an average long-window correlation of .005, and .03 in monosynaptically connected pairs.

2.5.2 S1 Fig: Plasticity in networks with larger correlations

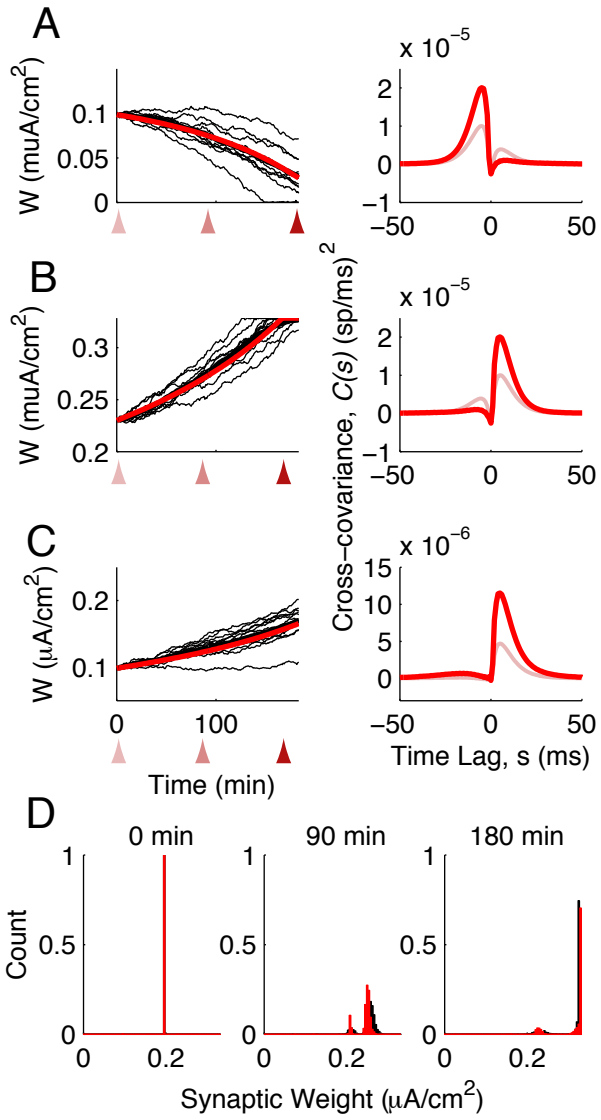

Figure 2.5: **Plasticity in a network of 100 neurons.** Here we examine synaptic plasticity in the $N = 100$ network of Figure 2.5.1B. As an illustration, we embed the same three-neuron microcircuit as in Figs 2 and 3 of the main text into this network and examine the evolution of its synaptic weights and spike-train covariances . (A-C) Left, Synaptic weight versus time for each of the three synapses in the highlighted microcircuit of Figs. 2,3. Thin lines: simulation, individual trials of the same initial network. Thick black lines: simulation, trial-average. Thick red lines: theory. Right, spike train cross-covariances at the beginning and endpoints (linear response theory). (D) Histogram of synaptic weights at three time points. Red, theory. Black: simulation.

3.0 SPIKE TIMING-DEPENDENT PLASTICITY OF TWO-SYNAPSE MOTIFS

3.1 INTRODUCTION

The wiring of neuronal networks exhibits structure across a broad range of spatial scales [149]. For example, patterns of connectivity among small groups of cortical neurons are over- or under-represented compared to random networks [150, 151, 91, 152]. The prevalence of these motifs is related to neurons' stimulus preferences and activity levels [153, 154]. Motivated in part by these observations, there is a growing body of theoretical work that discusses how wiring structure dictates the coordinated spiking activity of cortical neurons in recurrent networks [155, 156, 157, 158, 159, 160, 161, 136, 162, 163, 110].

In the previous chapter, we developed a method for predicting the plasticity of individual synapses in networks of simple model neurons. We now use that to derive a low-dimensional, closed theory for STDP of two-synapse connectivity motifs in recurrent networks. This reveals instabilities in the motif dynamics such that when potentiation and depression are approximately balanced, the dynamics is partitioned into regimes in which different motifs are promoted or suppressed depending on the initial network structure. It also highlights the circumstances in which spike time covariations, in contrast to firing rates, drive STDP. In total, we provide a consistent and general framework in which to study STDP in large recurrent networks.

3.2 RESULTS

While the theory of Chapter 2 gives an accurate description of plasticity in the network, it is nevertheless high-dimensional. Keeping track of every individual synaptic weight and spike train cross-covariance function involves $\mathcal{O}(N^2)$ variables. For large networks, this becomes computationally challenging. More importantly, this high-dimensional theory does not provide insights into the plasticity of the *connectivity patterns* or *motifs* that are observed in cortical networks [91, 151]. Motifs involving two or more neurons represent correlations in the network's weight matrix, which cannot be described by a straightforward application of mean-field techniques. In this chapter, we develop a principled approximation of the full weight dynamics to a closed low-dimensional theory for how the mean weight and the strength of two-synapse motifs evolve due to STDP.

3.2.1 Dynamics of mean synaptic weight

We begin by considering the simple case of a network with unstructured weights. Analogous to having an Erdös-Rényi *adjacency* matrix $\mathbf{W}^0$, we take there to be no second- or higher-order correlations in the *weight* matrix $\mathbf{W}$. In this case, we can consider only the mean synaptic weight, p:

$$p = \frac{1}{N^2} \sum_{i,j} \mathbf{W}_{ij}. \tag{3.1}$$

In order to calculate the dynamics of p, we insert the fast-slow STDP theory of Eq. (2.2) into Eq. (3.1):

$$\frac{dp}{dt} = \frac{1}{N^2} \sum_{i,j} \mathbf{W}^0_{ij} \int_{-\infty}^{\infty} L(s) \big(r_i r_j + \mathbf{C}_{ij}(s) \big) ds, \tag{3.2}$$

where the spiking covariances are calculated using linear response theory (Eq. (2.4)). This equation depends on the network structure in two ways. First, it depends on the full adjacency matrix $\mathbf{W}^0$. Multiplying by $\mathbf{W}^0_{ij}$ inside the average here prevents additional synapses from forming, so that we only consider the efficacy of synapses that exist, not the formation of new ones. Second, the spike train cross-covariances depend on the full weight matrix: $\mathbf{C}_{ij}(s) = \mathbf{C}_{ij}(s; \mathbf{W})$. This dependence of a first–order connectivity statistic on the network structure poses a challenge for the development of a closed theory.

The main steps in our approach here are two approximations. First, the matrix of spike train cross-covariances $\mathbf{C}(s)$ obtained from our linear ansatz (Eq. (2.4)) can be expanded in a power series around the background cross-covariances $\mathbf{C}^0(s)$ (see Eq. (C.4)). Powers of the interaction matrix $\mathbf{K}$ in this series correspond to different lengths of paths through the network [160, 136]. (As in Chapter 2, elements of $\mathbf{K}$ are given by postsynaptic neurons' linear response functions and the synaptic filters for pre-post transmission.) We truncate the spiking covariances at length one paths to obtain:

$$\mathbf{C}_{ij}(s) \approx \underbrace{\left(\mathbf{W}_{ij}\mathbf{K}_{ij} * \mathbf{C}^0_{jj}\right)(s)}_{\text{forward}} + \underbrace{\left(\mathbf{C}^0_{ii} * \mathbf{W}_{ji}\mathbf{K}^-_{ji}\right)(s)}_{\text{backward}} + \underbrace{\sum_k \left(\mathbf{W}_{ik}\mathbf{K}_{ik} * \mathbf{C}^0_{kk} * \mathbf{W}_{jk}\mathbf{K}^-_{jk}\right)(s)}_{\text{common}}, \tag{3.3}$$

where $*$ denotes convolution and $\mathbf{K}^-_{ji}(t') = \mathbf{K}_{ji}(-t')$. This truncation separates the sources of covariance between the spiking of neurons i and j into direct forward ($i \leftarrow j$) and backward ($i \rightarrow j$) connections, and common ($k \rightarrow i$ and $k \rightarrow j$) inputs. Nevertheless, after truncating $\mathbf{C}(s)$, the mean synaptic weight still depends on higher-order connectivity motifs (Eq. (3.23)). Fortunately, for weak connections, these higher-order terms do not contribute substantially to overall spiking covariance (S2 Fig, 3.6.2). This is especially true when we consider the covariance integrated against the plasticity rule $L(s)$ (difference of 6% between full and truncated covariance).

The second approximation is to ignore the bounds on the synaptic weight in Eq. (2.1). While this results in a theory that only captures the transient dynamics of $\mathbf{W}(t)$, it greatly simplifies the derivation of the low-dimensional dynamics of motifs, because dynamics along the boundary surface are not considered.

With these two approximations, the mean synaptic weight obeys:

$$\frac{dp}{dt} = r^2 S \frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij} + S_F\frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}\mathbf{W}_{ij} + S_B\frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}\mathbf{W}_{ji} + S_C\frac{1}{N^2}\sum_{i,j,k}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk}. \tag{3.4}$$

The first term on the right hand side of Eq. (3.4) is scaled by $S = \int_{-\infty}^{\infty} L(s)ds$, modeling the interaction between STDP and the mean firing rate, r, across the network. This captures STDP due to chance spiking coincidence and drives either net potentiation ($S > 0$) or depression ($S < 0$). The remaining terms capture how synaptic weights interact with the

temporal structure of spiking covariance. Because of the expansion in Eq. (3.3), these dependencies decompose into three terms, each scaled by the integral of the product of the STDP rule $L(s)$ and a component of the spike train cross-covariance $\mathbf{C}(s)$. Specifically, covariance due to forward connections is represented by S_F (Eq. (3.27); Figure 3.1A), covariance due to backward (reciprocal) connections is represented by S_B (Eq. (3.28); Figure 3.1B), and finally covariance due to common connections is represented by S_C (Eq. (3.29); Figure 3.1C).

For a network with unstructured weights, each sum in Eq. (3.4) can be simplified. Let $p_0 = \frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}$ be the connection density of the network. Since our theory for spiking covariances required weak synapses, we also explicitly scaled the weights, motifs, and amplitude of synaptic changes $f_\pm$ by $\epsilon = 1/(Np_0)$. This ensured that as the connection probability p_0 was varied, synaptic weights scaled to keep the total input to a neuron constant (neglecting plasticity). The first and second terms of Eq. (3.4) correspond to the definitions of p_0 and p. Since different elements of $\mathbf{W}^0$ and $\mathbf{W}$ are uncorrelated, the third term reduces to $\frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}\mathbf{W}_{ji} = \epsilon p p_0 + \mathcal{O}(\epsilon^{3/2})$ due to the central limit theorem. The last term can be similarly evaluated and the dynamics of p to first order in ϵ reduce to:

$$\frac{dp}{dt} = p_0 r^2 S + \epsilon\left(p\left(S_F + p_0 S_B\right) + p^2 S_C\right). \tag{3.5}$$

The contribution of forwards (monosynaptic) connections is $S_F p$ - it depends on the average strength of synapses. The contribution of reciprocal synapses is $p p_0 S_B$ because it is conditioned on the existence of a forwards connection. We next study this mean-field theory in two regimes, before examining the plasticity of networks that exhibit motif structure.

3.2.2 Unbalanced STDP of the mean synaptic weight

Eq. (3.5) contains one term proportional to the product of firing rates and the integral of the STDP rule, $r^2 S$, and additional terms proportional to the small parameter ϵ. When the learning rule, $L(s)$, is dominated by either depression or potentiation (so that $S \sim \mathcal{O}(1) \gg \epsilon$) the whole network either uniformly depresses (Figure 3.2A,C) or potentiates (Figure 3.2B,D) due to chance spike coincidences (the firing rate term dominates in Eq. (2.2)). These

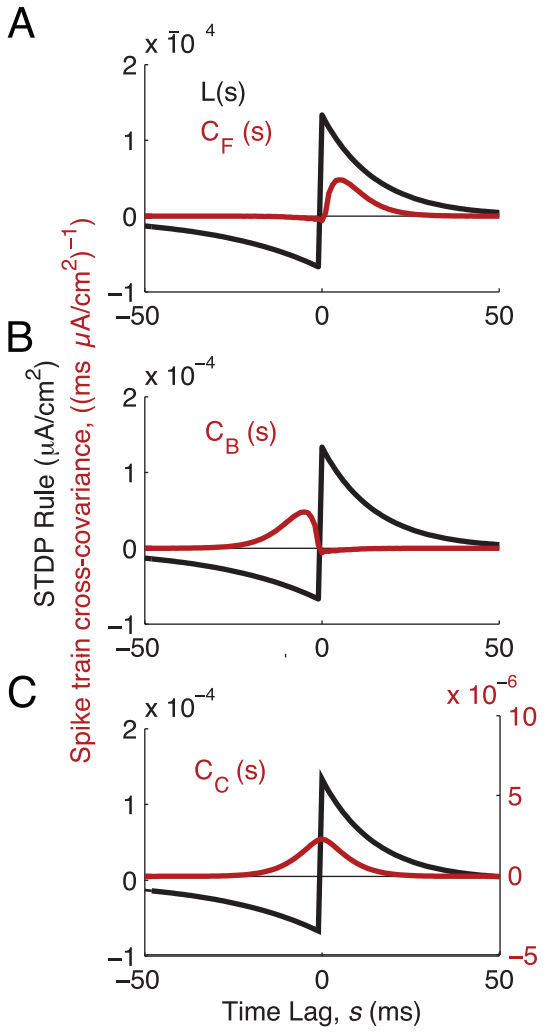

Figure 3.1: **Different sources of spiking covariance interact with different parts of the STDP rule.**] Black: STDP rule. Red: spike train cross-covariances, from Eq. (3.3). (A) Covariance from forward connections interacts with the potentiation side of the STDP rule. (B) Covariance from backward connections interacts with the depression side of the STDP rule. (C) Covariance from common input is temporally symmetric and interacts with both the potentiation and depression sides of the STDP rule.

dynamics is straightforward at the level of individual synapses and this intuition carries over to the mean synaptic weight. When the STDP rule is dominated by potentiation or depression, the $\mathcal{O}(\epsilon)$ terms in Eq. (3.5) are negligible; the average plasticity is solely determined by the firing rates, with spiking covariance playing no role. In this case, the

leading-order dynamics of p are:

$$p(t) = p_0 r^2 S t + p(0), \tag{3.6}$$

so that the mean synaptic weight either potentiates to its upper bound $p_0 W^{\max}$ or depresses to 0, depending on whether the integral of the STDP rule, S, is positive or negative. (We neglected bounds on individual synaptic weights to derive the dynamics. Here we reimpose bounds on the mean synaptic weight, although Eq. (3.5) only describes transient dynamics of $\mathbf{W}$.) For both depression- and potentiation-dominated STDP rules, our simple theory in Eq. (3.6) quantitatively matches $p(t)$ estimated from simulations of the entire network (Figure 3.2C,D, black vs. red curves).

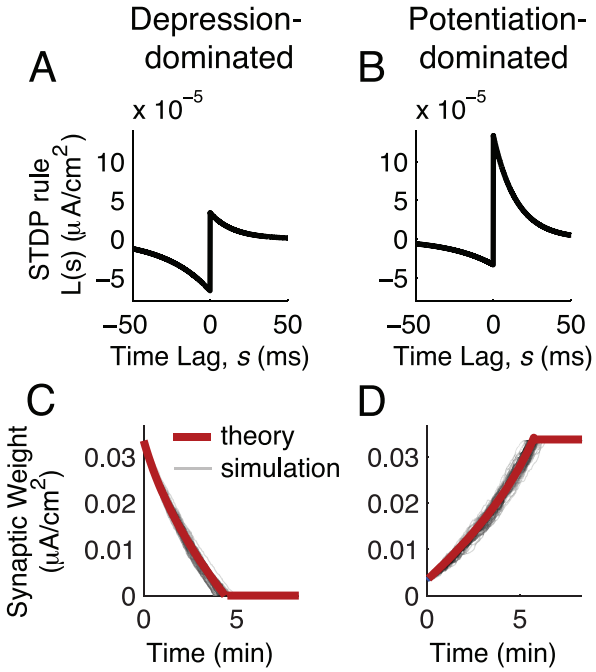

Figure 3.2: **Unbalanced plasticity gives rise to simple weight dynamics.** (A) Depression-dominated STDP rule: the amount of depression (integral of the depression side of the curve) is twice the amount of potentiation. (B) Potentiation-dominated STDP rule: the amount of potentiation is twice the amount of depression. (C) Evolution of synaptic weights with depression-dominated STDP: all weights depress. (D) Evolution of synaptic weights with potentiation-dominated STDP: all weights potentiate. Red lines: theory for mean synaptic weight. Shaded lines: simulation of individual synaptic weights.

3.2.3 Balanced STDP of the mean synaptic weight

On the other hand, if there is a balance between potentiation and depression in the STDP rule $L(s)$, then spiking covariance affects the average plasticity. In order to make explicit the balance between potentiation and depression, we write $S = \pm\delta\epsilon$ (with $+\delta\epsilon$ for STDP with the balance tilted in favor of potentiation and $-\delta\epsilon$ for balance tilted in favor of depression). The leading-order dynamics of p are then, for networks without motif structure,

$$\frac{1}{\epsilon}\frac{dp}{dt} = \pm\delta p_0 r^2 + p\left(S_F + p_0 S_B\right) + p^2 S_C. \tag{3.7}$$

This quadratic equation admits up to two fixed points for p. We begin by examining the dynamics of p for the case perfectly balanced potentiation and depression ($\delta = 0$) and a realistic shape of the STDP curve, and then consider the case of $\delta \neq 0$.

Experimentally measured STDP rules in cortex often show $f_+ > f_-$ and $\tau_+ < \tau_-$ [42, 38], making potentiation windows sharper and higher-amplitude than depression windows. In this case, the STDP-weighted covariance from forward connections, $S_F > 0$, is greater in magnitude than those from backward connections, $S_B < 0$ (Figure 3.1A,B). Since $0 \leq p_0 \leq 1$, $S_F + p_0 S_B > 0$. Furthermore, since the covariance from common input decays symmetrically around time lag $s = 0$ (Figure 3.1C), we have that $S_C > 0$. Consequently, when $\delta = 0$, all terms in Eq. (3.7) are positive and p potentiates to $p_0 W^{\max}$.

We next consider the case of imperfectly balanced STDP, with $\delta = .1$. For potentiation-dominated balanced STDP, $+\delta\epsilon$, again all terms in Eq. (3.7) are positive and p potentiates to $p_0 W^{\max}$ (Figure 3.3A). However, with depression-dominated balanced STDP ($-\delta\epsilon$ in Eq. (3.7)) p has two fixed points, at:

$$p = \frac{-\left(S_F + p_0 S_B\right) \pm \sqrt{\left(S_F + p_0 S_B\right)^2 + 4\delta p_0 r^2 S_C}}{2S_C}. \tag{3.8}$$

Since $S_F + p_0 S_B > 0$ and $S_C > 0$ because of our assumptions on $f_\pm$ and $\tau_\pm$, the term inside the square root is positive, making one fixed point is positive and the other negative. The positive fixed point is unstable and, if within $[0, p_0 W^{\max}]$, it provides a separatrix between potentiation and depression of p (Figure 3.3B). This separatrix arises from the competition

between potentiation (due to forward connections and common input) and depression (due to reciprocal connections and firing rates).

Examination of Eq. (3.8) shows competing effects of increasing the connection density p_0: the $S_F + p_0 S_B$ terms decrease, while the $4\delta p_0 r^2 S_C$ term increases. The latter effect dominates for the positive fixed point, raising the separatrix between potentiation and depression as p_0 increases. So the mean synaptic weight of sparsely connected networks (small p_0) have a propensity to potentiate, while more densely connected networks (large p_0) are more likely to depress (Figure 3.3B).

In total, we see that a slight propensity for depression can impose bistability on the mean synaptic weight. In this case, a network with an initially strong mean synaptic weight $p(0)$ can overcome depression and strengthen synaptic wiring, while a network with the same STDP rule and connection probability but with an initially weak mean synaptic weight will exhibit depression. In the next section we will show that similar separatrices exist in structured networks and govern the plasticity of different motifs.

3.2.4 Motif dynamics

We now consider networks that have structure at the level of motifs, so that different patterns of connectivity may be over- or under-represented compared to unstructured networks. We begin by defining the weighted two-synapse motif variables:

$$\begin{aligned} q^{\text{div}} &= \frac{1}{N^3} \sum_{i,j,k} \mathbf{W}_{ik} \mathbf{W}_{jk} - p^2, \\ q^{\text{con}} &= \frac{1}{N^3} \sum_{i,j,k} \mathbf{W}_{ik} \mathbf{W}_{ij} - p^2, \\ q^{\text{ch}} &= \frac{1}{N^3} \sum_{i,j,k} \mathbf{W}_{ij} \mathbf{W}_{jk} - p^2. \end{aligned} \tag{3.9}$$

The variables q^{div}, q^{con} and q^{ch}, respectively, measure the strength of divergent, convergent, and chain motifs. A divergent motif consists of one neuron, k, projecting to two others, i and j. A convergent motif consists of two neurons, k and j, projecting to a third, i. A chain consists of one neuron k, projecting to a second neuron j, which projects to a third neuron i.

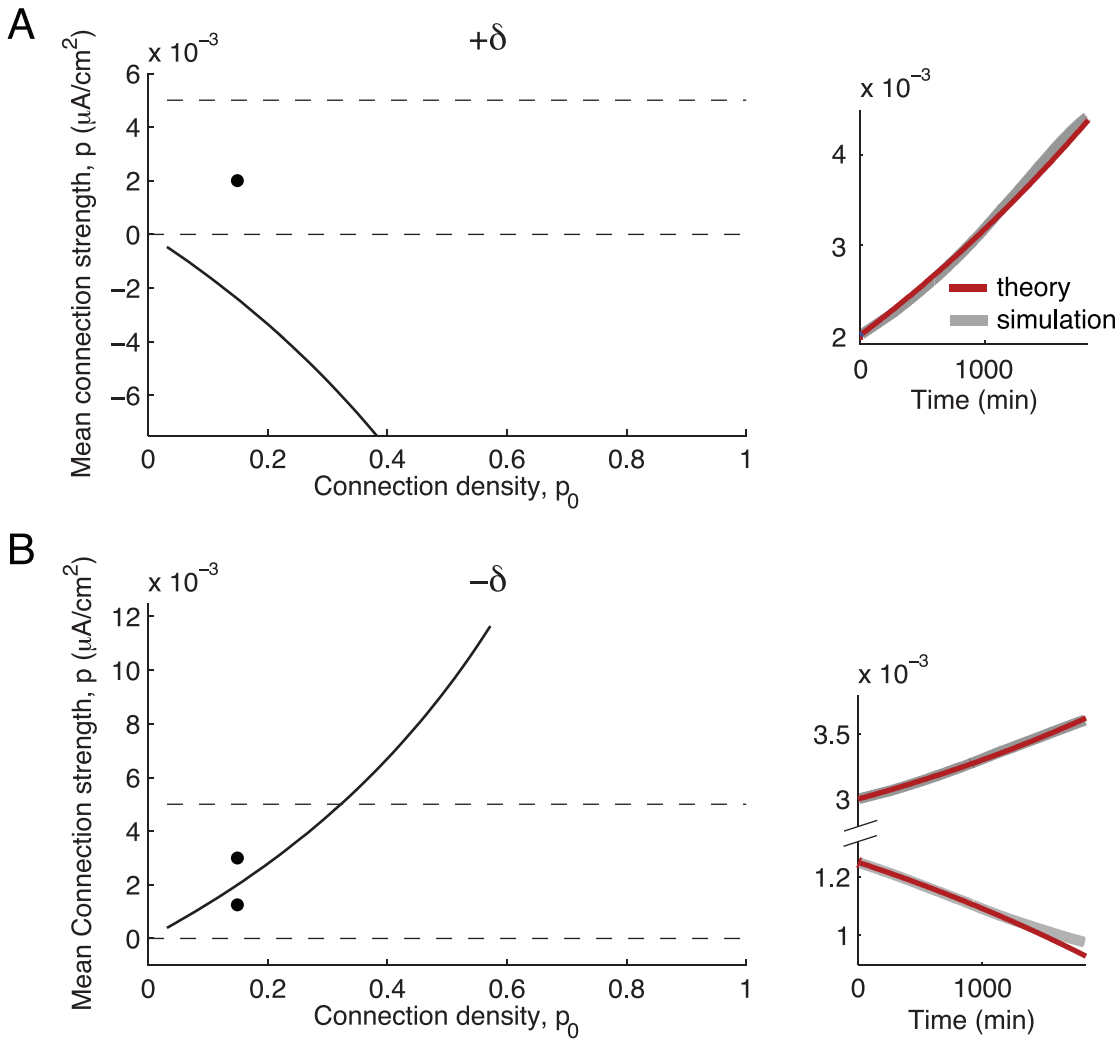

Figure 3.3: **Balanced plasticity of the mean synaptic weight.** (A) When the STDP rule is balanced and potentiation-dominated, the unstable fixed point for p is negative and decreases with the connection probability. (B) When the STDP rule is balanced and depression-dominated, the unstable fixed point is positive and increases with the connection probability. (A,B) Left: Dashed lines mark bounds for the mean synaptic weight, at 0 and $p_0 W^{\max}$. Black curves track the location of the unstable fixed point of p as the connection probability, p_0, varies. Black dots mark initial conditions for the right panels. (A,B) Right: Dynamics of the mean synaptic weight in each of the regimes of the left plots. Red lines mark the reduced theory's prediction (Eq. (3.5)) and shaded lines the result of simulating the full spiking network (10 trials are plotted individually; they lie within line thickness of each other). Note that the ordinate axis has different limits in the left and right sides of the figure.

For each variable, we subtract the expected value of the sum in a network with uncorrelated weights, p^2, so that the qs measure above- or below-chance levels of structure in the network. Since these variables depend on the strength of both synapses making up the motif, we will refer to them as *motif strengths*. Motif strengths are also related to neurons' (weighted) in- and out-degrees (the total strength of incoming or outgoing synapses for each neuron). The variables q^{div} and q^{con} are proportional to the variance of neurons' in- and out-degrees, while q^{ch}, on the other hand, is proportional to the covariance of neurons' in- and out-degrees. This can be seen by taking the definitions of these motifs, Eq. (3.9), and first summing over the indices i, j. This puts the sum in q^{div}, for example, in the form of a sum over neurons' squared out-degrees.

We remark that motif strengths (q) are separate from motif frequencies (q_0). Motif frequencies have analogous definitions to Eq. (3.9), but use the adjacency matrix $\mathbf{W}^0$ instead of the weight matrix $\mathbf{W}$ (Eq. (3.16)). It is clear that, for instance, $q^{\text{div}} \neq q_0^{\text{div}}$, although they would be proportional to one another if all weights $\mathbf{W}_{ij}$ were equal. An Erdös-Rényi network has an adjacency matrix $\mathbf{W}^0$ with negligible motif frequencies. To avoid confusion, we refer to a network with negligible motif *strengths* as an unstructured network.

We wish to examine the joint dynamics of the mean synaptic weight p and the motif strengths. We insert the fast-slow STDP theory of Eq. (2.2) into the definitions of p (Eq. (3.1)) and the three qs (Eq. (3.9)). Similarly to Eq. (3.5), the dynamics of motifs $q^{\text{div}}(t)$, $q^{\text{con}}(t)$, and $q^{\text{ch}}(t)$ then depend on the full network structure, $\mathbf{W}$. This dependence of first- and second-order connectivity statistics on the network structure poses a challenge for the development of a closed theory for the dynamics of motifs. The main steps in developing such a theory are the two approximations we used to develop Eq. (3.5), as well as one more.

As in the previous sections, our first approximation is to truncate the spike-train covariances at length one paths through the network. This removes the dependency of the dynamics on longer paths through the network. Nevertheless, after truncating $\mathbf{C}(s)$, the first- (p) and second-order (q^{div}, q^{con}, q^{ch}) motifs still depend on higher-order motifs (Eq. (3.4)). This is because of coupling between lower and higher-order moments of the connectivity matrix $\mathbf{W}$ (see Eqs. (3.23)-(3.25)) and presents a significant complication.

In order to close the dynamics at one- and two-synapse motifs, our new approximation

follows [162], and we rewrite higher-order motifs as combinations of individual synapses and two-synapse motifs (see Eqs. (3.30)-(3.31)). For the mean synaptic weight, for example, one third-order motif appears due to the common input term of the spike-train covariances (Eq. (3.4)). We break up this three-synapse motif into all possible combinations of two-synapse motifs and individual connections, estimating its strength as:

$$\frac{1}{N^3}\sum_{i,j,k}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk} \approx \left(p_0\left(q^{\text{div}}+p^2\right)+p\left(q_{\text{X}}^{\text{con}}+q_{\text{X}}^{\text{ch,B}}\right)\right). \tag{3.10}$$

This corresponds to assuming that there are no third- or higher-order correlations in the weight matrix beyond those due to second-order correlations; three- and more-synapse motifs are represented only as much as would be expected given the two-synapse motif strengths. (We assume that all of the third- and higher-order cumulants of the weight and adjancency matrices that we encounter are zero.) This allows us to close the motif dynamics at two-synapse motifs. However, two new motifs appear in Eq. (3.10), $q_{\text{X}}^{\text{con}}$ and $q_{\text{X}}^{\text{ch,B}}$. The subscript $_{\text{X}}$ denotes that these motifs are mixed between the weight and adjacency matrices, measuring the strength of individual connections conditioned on their being part of a particular motif. $q_{\text{X}}^{\text{con}}$ corresponds to the strength of connections conditioned on being part of a convergent motif and $q_{\text{X}}^{\text{ch,B}}$ to the strength of connections conditioned on the postsynaptic neuron making another synapse in a chain (Eq. (3.17)). As in previous sections, the final approximation is to ignore the bounds on the synaptic weight in Eq. (2.1), so that our theory only captures the transient dynamics of $\mathbf{W}(t)$.

These approximations allow us (see Eqs. (3.20), (3.23), and (3.32)) to rewrite the dynamics of the mean synaptic weight p as:

$$\frac{dp}{dt} = p_0 r^2 S + \epsilon\left[pS_F + \left(q_{\text{X}}^{\text{rec}}+p_0 p\right)S_B + \frac{1}{p_0}\left(p_0\left(q^{\text{div}}+p^2\right)+p\left(q_{\text{X}}^{\text{con}}+q_{\text{X}}^{\text{ch,B}}\right)\right)S_C\right]. \tag{3.11}$$

The parameters S, S_F, S_B and S_C are as defined in the previous section. Note that we recover Eq. (3.5) when all q's vanish (i.e. an unstructured network). When the network contains motif structure ($q \neq 0$), the dynamics of p contain new terms. In Eq. (3.11), the influence of forward connections through S_F is again proportional to the mean synaptic weight p. In contrast, the influence of backward connections S_B must interact with the new variable $q_{\text{X}}^{\text{rec}}$, which measures the mean strength of connections conditioned on their being

part of a reciprocal loop (i.e. the strength of a backwards connection, conditioned on the existence of the forward one). As described above (Eq. (3.10)), the covariance from common input S_C involves p, the divergent motif, q^{div}, as well as terms conditioned on weights being part of a convergent motif, $q_{\text{X}}^{\text{con}}$, or on the postsynaptic neuron making another synapse in a chain, $q_{\text{X}}^{\text{ch,B}}$. The definitions for the mixed motifs, the q_{X}s, are given in Eqs. (3.17). In total, the dynamics of mean synaptic weight cannot be written as a single closed equation, but also requires knowledge of how the second order motifs evolve.

Fortunately, using a similar approach dynamical equations can be derived for each of the two-synapse motifs q^{div}, q^{cov}, and q^{ch} (Eqs. (3.33)-(3.35)). To close the system we require dynamics for five mixed motifs, $q_{\text{X}}^{\text{con}}$, $q_{\text{X}}^{\text{div}}$, $q_{\text{X}}^{\text{rec}}$, $q_{\text{X}}^{\text{ch,A}}$, and $q_{\text{X}}^{\text{ch,B}}$ (Eqs. (3.36)-(3.40)). In total, this yields an autonomous 9-dimensional system of nonlinear differential equations describing the population-averaged plasticity of first- and second-order network structure. We have derived these equations in the absence of common external inputs to the neurons; the theory can easily be extended to this case by including external covariance in Eq. (3.3) (replacing $\mathbf{C}^0$ with $(\mathbf{C}^0 + \mathbf{C}^{\text{ext}})$, where $\mathbf{C}^{\text{ext}}$ is the covariance matrix of the inputs).

When the network structure $\mathbf{W}^0$ is approximately Erdös-Rényi, the motif frequencies q_0 are $\mathcal{O}\left(N^{-3/2}\right) = \mathcal{O}\left(\epsilon^{3/2}\right)$. (Since they are proportional to the variance of neurons' in- and out-degrees, which for Erdös-Rényi networks are binominally distributed.) If we further assume initial conditions for the motif strengths and the mixed motifs to be unstructured ($q(0) \sim \mathcal{O}\left(\epsilon^{3/2}\right)$ for all motifs), then we also have $dq_{\text{X}}/dt \sim \mathcal{O}\left(\epsilon^{3/2}\right)$ and $dq_{\text{X}}/dt \sim \mathcal{O}\left(\epsilon^{3/2}\right)$ for each motif. In this case we can neglect, to leading order, the motifs entirely. Here the leading order dynamics simplify tremendously, and are restricted to the set $\{p(t), q^{\text{div}} = q^{\text{con}} = q^{\text{ch}} = q_{\text{X}}^{\text{rec}} = q_{\text{X}}^{\text{con}} = q_{\text{X}}^{\text{div}} = q_{\text{X}}^{\text{ch,A}} = q_{\text{X}}^{\text{ch,B}} = 0\}$. Since the motif variables are zero the set corresponds to an unstructured network. Furthermore, since the leading order dynamics of the motif variables are zero this is an invariant set. The dynamics of $p(t)$ then collapse to those given by Eq. (3.5), which we have already examined (Figs 5 and 6).

The stability of that invariant set, however, remains to be determined. For finite N, the motif frequencies q_0 will be non-zero even for (approximately) Erdös-Rényi networks. In this case we may consider the full system Eqs. (3.32)-(3.40). In particular, the dynamics of the full system can be studied to determine the stability, or lack thereof, of the initial

unstructured synaptic weights.

We refer to the mean field theory of Eqs. (3.32)-(3.40) as the *motif dynamics* for a recurrent network with STDP. This theory accurately predicts the transient dynamics of the one- and two-synapse motifs of the full stochastic spiking network (Figure 3.4, compare red versus thin black curves), owing to significant drift compared to diffusion in the weight dynamics and these network-averaged motif strengths. The derivation and successful application of this reduced theory to a large spiking network is a central result of our study. However, recall that our theory requires the overall synaptic weights to be small so that our linear response ansatz remains valid. Thus, as expected, our theoretical predictions for the evolution of motif structure fail for sufficiently large initial mean synaptic weight $p(0)$ (S2 Text, 3.6.1). This is because for large recurrent weights the firing rate dynamics become unstable, and linearization about a background state is not possible.

Our theory captures several nontrivial aspects of the evolution of network structure. First, while the STDP rule is in the depression-dominated regime ($S < 0$ for the simulations in Figure 3.4), the mean synaptic weight p nevertheless grows (Figure 3.4A). Second, both divergent and convergent connections, q^{div} and q^{con}, grow above what is expected for an unstructured network (Figure 3.4B,C); however, at the expense of chain connections q^{ch} which decay (Figure 3.4G). The combination of these results show that for this STDP rule $L(s)$, the unstructured network is not stable, and spontaneous structure forms slowly over time. In the subsequent sections, we leverage the simplicity of our reduced theory to gain insight into how the STDP rule interacts with recurrent architecture to drive motif dynamics.

3.2.5 Unbalanced STDP of two-synapse motifs

When the STDP rule is dominated by potentiation or depression so that $S \sim \mathcal{O}(1) \gg \epsilon$, then the $\mathcal{O}(\epsilon)$ terms in Eqs. (3.33)-(3.40) are negligible. In this case each motif's plasticity is solely determined by the firing rates, with spiking covariance playing no role. Here the

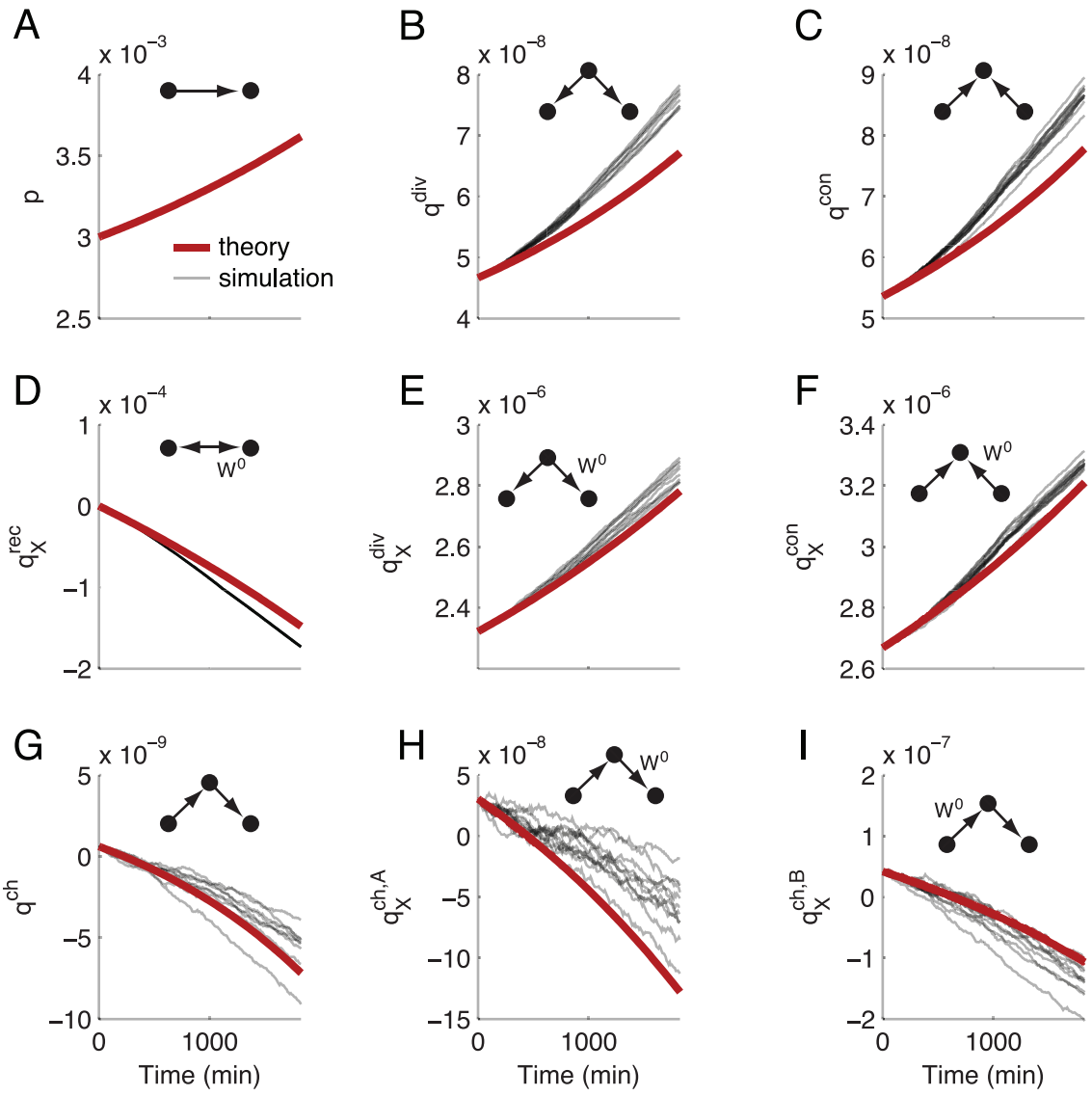

Figure 3.4: **Reduced theory for the plasticity of two-synapse motifs.** In each panel, the strength of a different motif or mixed motif is plotted as it evolves. Red: theoretical prediction (Eqs. (3.32)-(3.40)). Shaded lines: individual trials of the same initial network. (A) Mean synaptic weight. (B) Divergent motifs. (C) Convergent motifs. (D) Mixed recurrent motifs (strength of connections conditioned on their being part of a two-synapse loop). (E) Mixed divergent motifs (strength of individual synapses conditioned on their being part of a divergent motif). (F) Mixed convergent motifs. (G) Chain motifs. (H) Mixed chains type A (strength of individual synapses conditioned on their being the first in a chain). (I) Mixed chains type B (strength of individual synapses conditioned on their being the second in a chain). The STDP rule was in the depression-dominated balanced regime, as in Fig. 7B.

motif dynamics is simply:

$$\begin{aligned}\frac{dp}{dt} &= p_0 r^2 S + \mathcal{O}(\epsilon) \\ \frac{dq^\alpha}{dt} &= 2r^2 S q_{\mathrm{X}}^\alpha + \mathcal{O}(\epsilon) \\ \frac{dq_{\mathrm{X}}^\alpha}{dt} &= r^2 S q_0^\alpha + \mathcal{O}(\epsilon)\end{aligned} \tag{3.12}$$

for $\alpha = $ div, con, or ch (and taking $q_{\mathrm{X}}^{\mathrm{ch}} = \left(q_{\mathrm{X}}^{\mathrm{ch,A}} + q_{\mathrm{X}}^{\mathrm{ch,B}}\right)/2$ in the second equation - A and B refer to the strength of connections conditioned on, respectively, the presynaptic neuron receiving a connection or the postsynaptic neuron making another connection). The dynamics of p are the same here as for the unstructured case above; we include it for completeness. Dropping order ϵ terms gives the simple solutions:

$$\begin{aligned}p(t) &= p_0 r^2 S t + p(0) \\ q^\alpha(t) &= q^\alpha(0) + q_{\mathrm{X}}^\alpha(0) r^2 S t + \frac{1}{2} q_0^\alpha \left(r^2 S\right)^2 t^2\end{aligned} \tag{3.13}$$

for $\alpha =$ div, con, or ch (Methods: Unbalanced STDP). As stated previously, with $S \sim \mathcal{O}(1)$, individual synapses uniformly potentiate or depress (Figure 3.2). This is reflected in the linear decay or growth (for depression- or potentiation-dominated $L(s)$, respectively) of p with r^2 and quadratic amplification of baseline motif frequencies for the two-synapse motif strengths.

3.2.6 Balanced STDP of two-synapse motifs

Now we turn our attention to how internally generated spiking covariance interacts with balanced STDP to control motifs (examining the dynamics of Eqs. (3.32)-(3.40)). As before, we consider STDP rules with sharper windows for potentiation than depression ($\tau_+ < \tau_-$ and $f_+ > f_-$). Each two-synapse motif can have a nullcline surface in the nine-dimensional motif space. These nullclines define a threshold for the promotion or suppression of the corresponding motif, analogous to the case on the unstructured invariant set (Figure 3.4). We illustrate this by examining the dynamics in the $\left(q^{\mathrm{div}}, q^{\mathrm{con}}\right)$ plane. For STDP rules with a balance tilted towards depression ($-\delta\epsilon$), the nullclines provided thresholds for the promotion or suppression of divergent or convergent motifs (Figure 3.5A, blue lines). The flow in this slice of the motif space predicted the motif dynamics well (Figure 3.5A, compare individual

realizations of the full spiking network – thin black lines – to the flow defined by the vector field of the reduced motif system).

On the other hand, STDP rules with the balance tilted towards potentiation ($+\delta\epsilon$) have the nullclines at negative motif strengths (Fig. 8B). Can the motif strengths achieve negative values? As stated previously, q^{con} and q^{div} are proportional to the variances of neurons' in and out degrees, respectively. So, like the mean synaptic weight, $q^{\text{div}}, q^{\text{con}} \geq 0$, and these motifs always potentiated for $+\delta\epsilon$ STDP rules (Fig. 8B).

In examining the joint dynamics of divergent and convergent motifs, there is little evidence of interaction. The nullclines in the $(q^{\text{con}}, q^{\text{div}})$ plane are horizontal and vertical, so that whether divergent motifs potentiate or depress is independent of the dynamics of convergent motifs and vice versa (Figure 3.5A,B). This is reflected in the equations governing them. First, q^{div} does not depend directly on q^{con} (Eq. (3.33)). Second, q^{con} depends through q^{div} only through the re-summed approximation of a four-synapse motif and the STDP-weighted covariances from common inputs, $S_C q^{\text{div}} q_{\text{X}}^{\text{con}}$ (Eq. (3.34)) which due to the product $q^{\text{div}} q_{\text{X}}^{\text{con}}$ provides only weak dependency.

Chain motifs correspond to the covariance of neurons' weighted in- and out-degrees and so, in contrast to q^{div} and q^{con}, can achieve negative values. Indeed, the strength of chains can depress below zero even while the mean synaptic weight and other motifs potentiate (Figure 3.4A,G). Examining how $q^{\text{ch}}, q^{\text{div}}$ and q^{con} coevolve allowed us to see how in- and out-hubs developed in the network. With the $+\delta\epsilon$ STDP rule, q^{ch} increased along with q^{con} and q^{div} (Figure 3.5B). So, individual neurons tended to become both in- and out-hubs. With the $-\delta\epsilon$ STDP rule, however, q^{ch} could decrease while q^{div} and q^{con} increased (Figure 3.4, Figure 3.5D). In this case, neurons tended to become in- or out-hubs, but not both. In contrast to the vertical and horizontal nullclines in the $(q^{\text{con}}, q^{\text{div}})$ plane, q^{ch} directly depends on q^{con} and q^{div} (Eq. (3.35)). This is reflected in the nullcline structure of the $(q^{\text{ch}}, q^{\text{div}})$ plane: whether q^{ch} potentiates or depresses depends on the initial strength of q^{div} (Figure 3.5C,D). For these networks, q^{ch} exhibited similar dependencies on q^{con}.

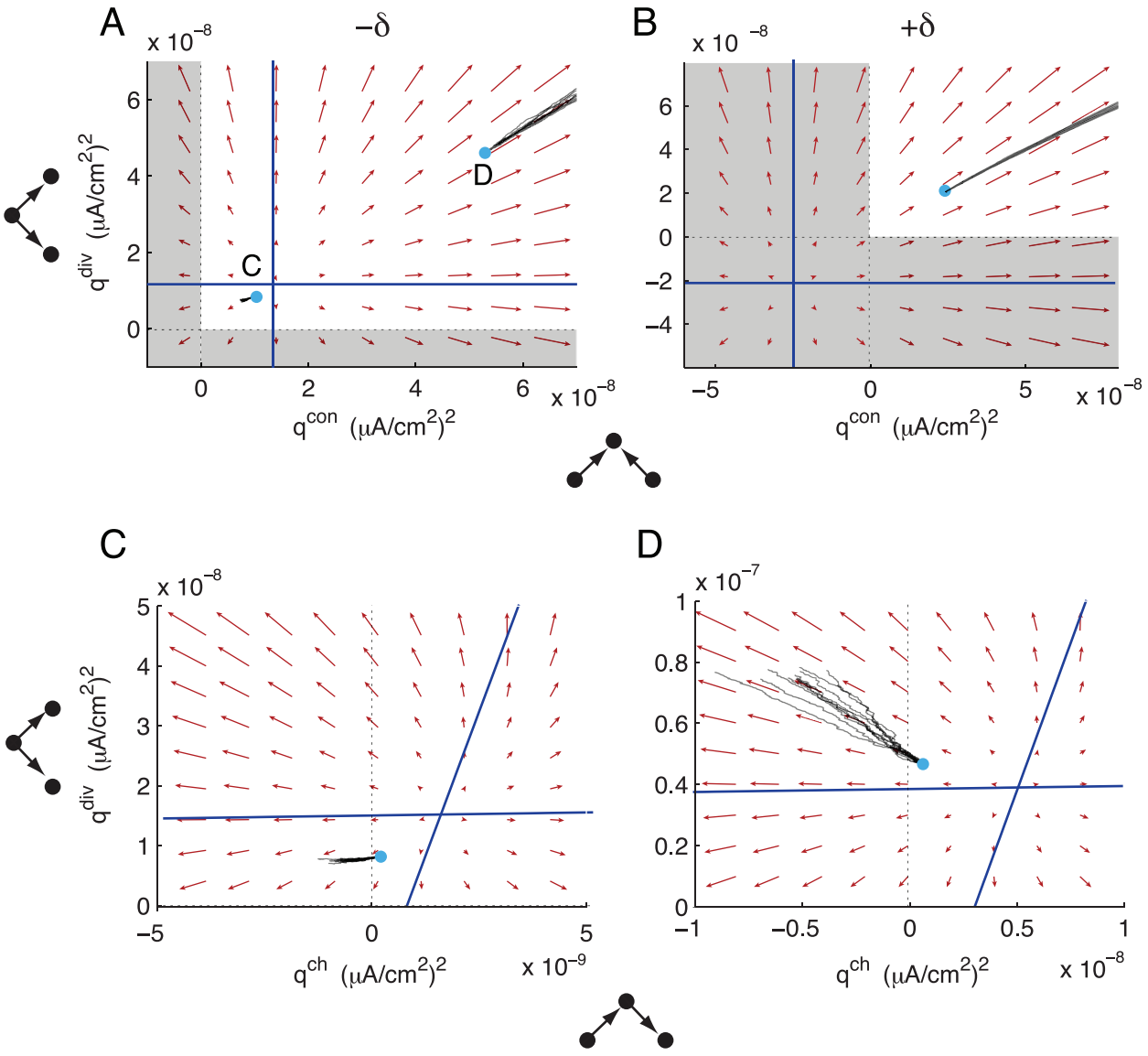

Figure 3.5: **Plasticity of convergent and divergent motifs with balanced STDP.** (A) Joint dynamics of convergent and divergent motifs when STDP is balanced and depression-dominated. Initial conditions as in Fig. 7A. (B) Joint dynamics of convergent and divergent motifs when STDP is balanced and potentiation-dominated. Initial conditions as in Fig. 7B. (C,D) Joint dynamics of divergent and chain motifs for the balanced, depression-dominated STDP rule. Initial conditions marked in panel A). Red: in all panels, the flow of the motif variables is projected into the corresponding plane, with all other motifs frozen at their initial conditions. Black: plasticity of the motifs in simulations of the full spiking network. Cyan dots mark initial conditions for the plotted variables. Each black trace is an individual realization of plasticity from the same initial network. For (A), the vector fields are indistinguishable, on the plotted scale, for both sets of initial conditions. In all panels, blue lines mark projections of each variable's nullcline into the plane defined by freezing the other motif variables at their initial conditions, and regions of unattainable negative motif strengths are shaded.

3.2.7 Co-evolution of open chains and reciprocal loops

Many studies have examined how STDP affects either feedforward or recurrent structure in neuronal networks, commonly showing that STDP promotes feedforward structure at the expense of recurrent loops [124, 164, 165]. This is consistent with the intuition gained from isolated pairs of neurons, where STDP can induce competition between reciprocal synapses and eliminate disynaptic loops [89]. Our theory provides a new way to examine how STDP regulates feedforward vs recurrent motifs by examining the dynamics of q^{ch}. This variable includes both recurrent loops (q^{rec}) and open chains (q^{op}). In order to understand the contribution of each of these to overall potentiation or depression of chains, we split the motif strength for chains into contributions from recurrent loops and open chains, rewriting q^{ch} as:

$$q^{\text{ch}} = \underbrace{\frac{1}{N^3}\sum_{i,j,k}\delta_{ik}\mathbf{W}_{ij}\mathbf{W}_{jk}}_{q^{\text{rec}}} + \underbrace{\frac{1}{N^3}\sum_{i,j,k}(1-\delta_{ik})\,\mathbf{W}_{ij}\mathbf{W}_{jk} - p^2}_{q^{\text{op}}}. \tag{3.14}$$

Similar to the case of other two-synapse motifs, the leading order dynamics of the recurrent motif are:

$$\frac{1}{2\epsilon}\frac{dq^{\text{rec}}}{dt} = r^2 S p_0\,(q_{\text{X}}^{\text{rec}} + p p_0) + S_F q^{\text{rec}} + S_B q_{\text{X2}}^{\text{rec}}. \tag{3.15}$$

We obtain the dynamics of the feedforward motif by subtracting dq^{rec}/dt from dq^{ch}/dt (Eq. (3.45)). In Eq. (3.14) we subtract p^2 from q^{op} because q^{op} is the dominant contributor to q^{ch}. This restricts q^{rec} to being non-negative. The new auxiliary variable $q_{\text{X2}}^{\text{rec}}$ is proportional to the conditional second moment of weights that are part of loops (Eq. (3.42)), and evolves according to Eq. (3.44). The replacement of q^{ch} by these variables expands the motif space to 11 dimensions.

We investigated the joint dynamics of open chains and recurrent loops by examining the $(q^{\text{op}}, q^{\text{rec}})$ plane. The q^{op} and q^{rec} nullclines divided this plane into regions where each motif potentiated or depressed. The shape of the STDP rule and the initial values of the other motif strengths affected the location of these nullclines. For the $+\delta\epsilon$ STDP rule, the q^{rec} nullcline was just below $q^{\text{rec}} = 0$ (Figure 3.6A, blue horizontal line). Since $q^{\text{rec}} \geq 0$, this forced q^{rec} to potentiate. The open chain motif, in contrast, could potentiate or depress above chance levels. In our spiking simulations, the initial conditions put q^{op} in the region of

depression, so that open chains depressed even while all other motifs were growing (Figure 3.6A, right panels).

These dynamics were the opposite of what would be expected from examining isolated pairs of neurons. With both the $+\delta\epsilon$ and $-\delta\epsilon$ balanced STDP rules, isolated pairs of neurons showed splitting of synaptic weights to eliminate the recurrent loop (S3 Fig, 3.6.4). Thus, with the $+\delta\epsilon$ STDP rule, the intuition gained from pairs of neurons did not predict the combined plasticity of open chains and recurrent loops. This is possible because our theory considers large networks that have both open chains and reciprocal loops in $\mathbf{W}^0$, and the motif plasticity takes both into account.

The locations of the q^{op} and q^{rec} nullclines were sensitive to the values of the other motif variables. Since the mean synaptic weight and q^{div} and q^{con} exhibited bistability under the $-\delta\epsilon$ STDP rule, we examined the $(q^{\text{op}}, q^{\text{rec}})$ slice through motif space when the other motifs were potentiating (Figure 3.6B, right panels) or depressing (Figure 3.6C, right panels). In both cases, the q^{rec} nullcline was above 0 so that the recurrent motif could either potentiate or depress, depending on its initial strength (Figure 3.6B,C blue horizontal lines). Similarly, the feedforward motif could either potentiate or depress.

In spiking simulations with $-\delta\epsilon$ STDP where p and the other motifs potentiated (Figure 3.6B, right), the initial conditions put $(q^{\text{op}}, q^{\text{rec}})$ in the region of phase space where they both depressed (Figure 3.6B, left). In spiking simulations with $-\delta\epsilon$ STDP where p and other motifs depressed (Figure 3.6C, right), the initial conditions put $(q^{\text{op}}, q^{\text{rec}})$ in the region where q^{op} potentiated and q^{rec} depressed. This region corresponded to what would be expected from examining isolated pairs of neurons (S3 Fig, 3.6.4): the loss of disynaptic loops and promotion of feedforward structure. So with the $-\delta\epsilon$ STDP rule, the region of phase space where the pair-based intuition was accurate at the network level was accessible. In most of the motif space, however, interactions between triplets of neurons played a strong role so that the theory developed here was necessary to predict the STDP of motif structure.

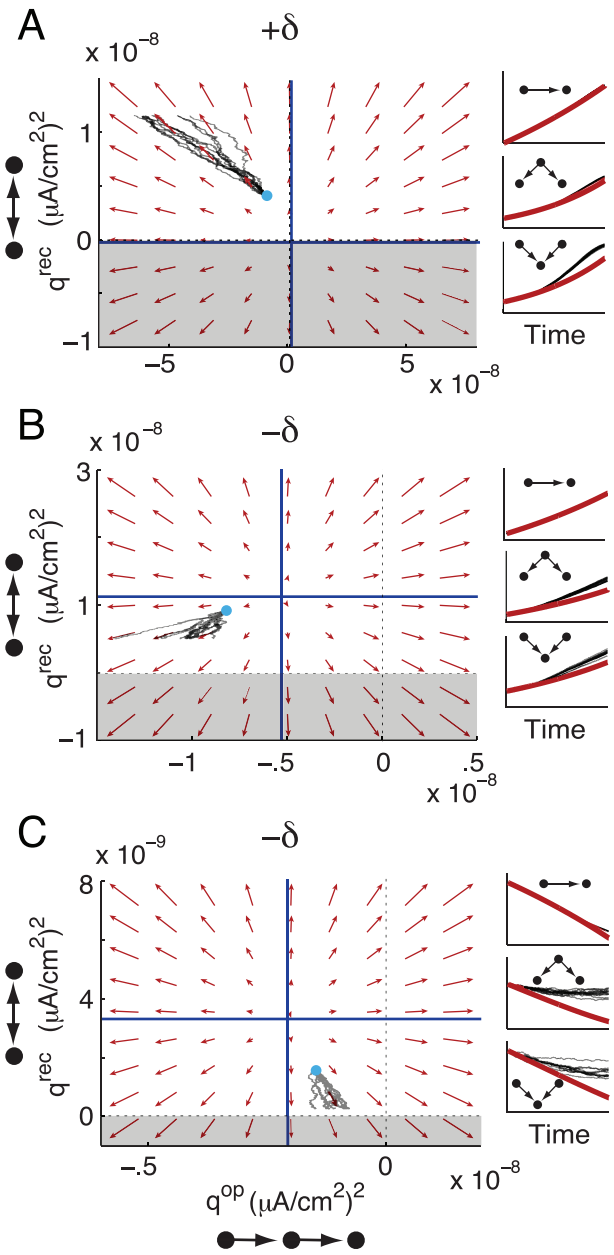

Figure 3.6: **Plasticity of recurrent loops and open chains with balanced STDP.** (A-C) The dynamics of loops and open chains, with all other variables fixed at their initial conditions. In all cases, the projections of the q^{op} and q^{rec} nullclines into this plane provide thresholds for the potentiation or depression of each motif. The shape of the STDP rule and the initial values of the other motif variables determine the locations of these nullclines. Color conventions are as in Fig. 8. In each panel, right insets show the time series of p (top), q^{div} (middle) and q^{con} (bottom), with spiking simulations in black and motif theory in red. A) The potentiation-dominated balanced STDP rule. B) The depression-dominated balanced STDP rule, in the region where p, q^{div} and q^{con} potentiate. C) The depression-dominated balanced STDP rule, in the region where p, q^{div} and q^{con} depress.

3.2.8 Motif dynamics in non-Erdös-Rényi networks

So far, we have examined the promotion or suppression of motif structure from initially unstructured networks with Erdös-Rényi $\mathbf{W}^0$. In order to check how well our theory applied to non-Erdös-Rényi networks, we examined networks with truncated power law in- and out-degree distributions (Methods: Neuron and network models). These networks exhibited much higher levels of divergent and convergent motif structure (Figure 3.7D,E). They also violated the approximation we made that three- and four-synapse motifs are only as represented as would be expected from the two-synapse motifs we measure (e.g. Eq. (3.10)).

For these networks, we varied the correlation of neurons' in- and out-degrees, thus changing the frequency and initial strength of chains (Figure 3.7B). In most cases, we saw that our motif plasticity theory still matched simulations of the full spiking network's evolution. This was true despite the motif variables being of several orders of magnitude larger compared to the Erdös-Rényi networks. In these networks, we see a similar bistability of the network structure to that observed earlier, both at the level of mean synaptic weights (Figure 3.7Cii-iv) and motifs (Figure 3.7 D,Eii-iv).

When chain motifs were sufficiently over-represented, however, the theory qualitatively mis-predicted the actual evolution of q^{ch} (Figure 3.7Fiv). Chain motifs play a large role in coupling various motifs to each other (Eqs. (3.32)-(3.40)). So, it is not surprising that although all these non-Erdös-Rényi networks violated the re-summing approximation, we only saw the theory qualitatively break down when chain motifs were sufficiently strong. Thus, for this type of non-Erdös-Rényi network the theory developed here holds surprising promise for the investigation of motif plasticity.

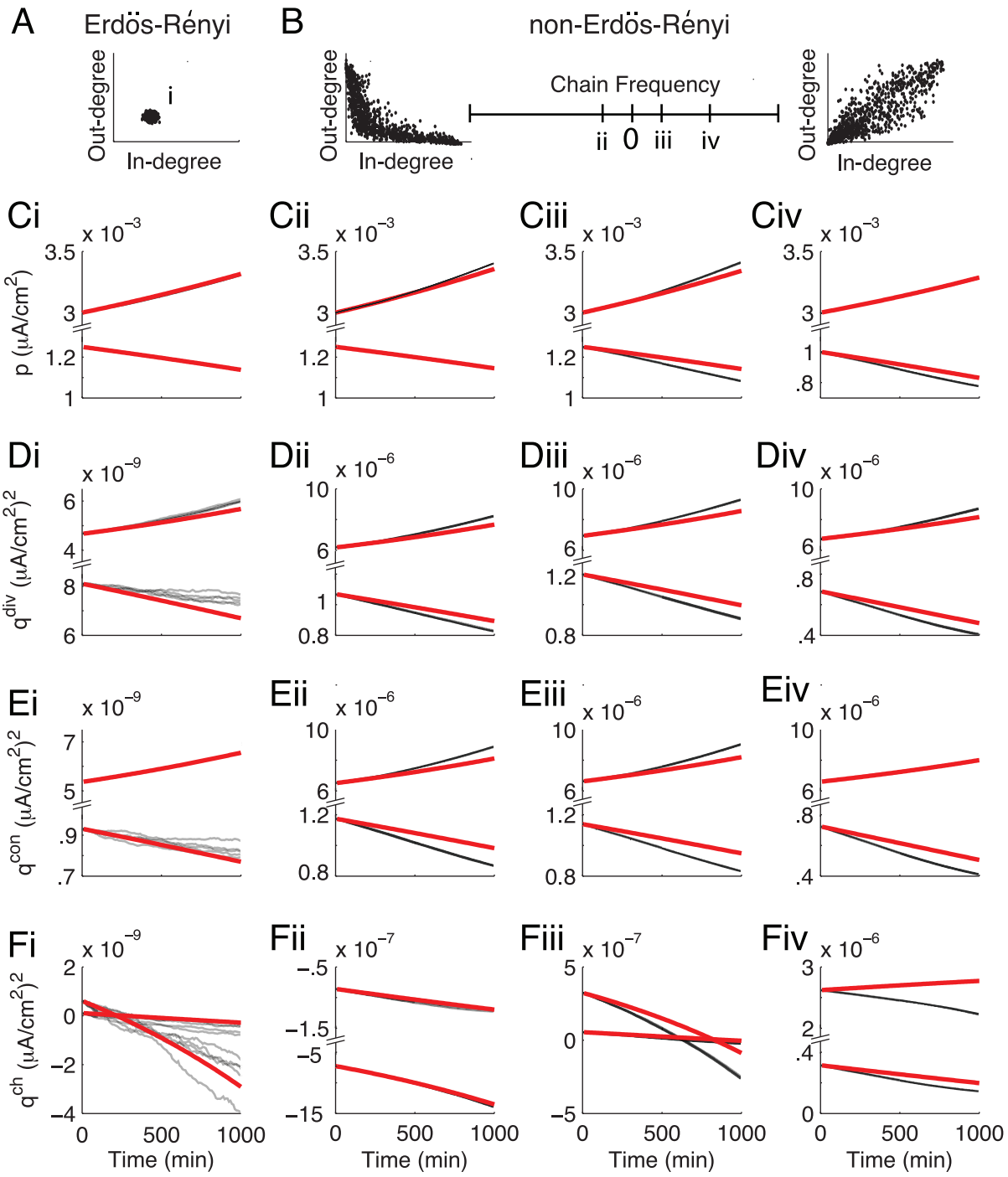

Figure 3.7: **Motif dynamics in non-Erdös-Rényi networks.** Caption on next page.

Motif dynamics in non-Erdös-Rényi networks. (A) Degree distribution of the finite-size Erdös-Rényi networks used in Figs. 2-9, showing each neuron's number of incoming synapses (in-degree) and outgoing synapses (out-degree). (B) Correlated, truncated power degree distributions. From left to right, the frequency of chain motifs increases. The degree distributions at either end of the "Chain Frequency" axis correspond to highly anti-correlated (left) and highly correlated (right) in- and out-degrees, with correlation coefficient $\rho = \pm.9$ (Methods: Neuron and network models). The networks in columns ii-iv are drawn from the labelled points on this axis, with $\rho = -.1$ (ii), $\rho = .1$ (iii) and $\rho = .5$ (iv). In each column, we sample networks from each side of threshold for potentiation of p. For the network in column ii, $q_0^{\text{div}} = .0149$, $q_0^{\text{con}} = .0163$, $q_0^{\text{ch}} = -.0012$. For the network in column iii, $q_0^{\text{div}} = .0165$, $q_0^{\text{con}} = .0157$, $q_0^{\text{ch}} = 7.7x10^{-4}$. For the potentiating network in column iv, $q_0^{\text{div}} = .0161$, $q_0^{\text{con}} = .0157$, $q_0^{\text{ch}} = .0062$. For the depressing network in column iv, $q_0^{\text{div}} = .0148$, $q_0^{\text{con}} = .0156$, $q_0^{\text{ch}} = .0068$. (C) Dynamics of the mean synaptic weight. (D) Dynamics of divergent motifs. (E) Dynamics of convergent motifs. (F) Dynamics of chain motifs. In all panels, the STDP rule is the balanced, depression-dominated one ($-\delta$ in Figs. 6-9).

3.3 SUMMARY OF MOTIF SYSTEMS DISCUSSED HERE

In this chapter, we discussed a number of different sets of motif dynamics. For simplicity, we first presented the motif dynamics under the condition where we can consider only the mean synaptic weight, p. When the network's adjacency matrix is Erdös-Rènyi, the mean synaptic weight is, to leading order, an invariant set for the full motif dynamics because the strengths are initially weak and furthermore, they evolve on a slower timescale because the motif frequencies are also weak.

We then considered the full dynamics by taking planar slices of the motif space, freezing other motifs at their initial conditions for an estimate of dynamics near those initial conditions. This showed that the invariant set of p is often unstable (albeit on a $1/\mathcal{O}(\epsilon^2)$ timescale). While we could examine the full motif dynamics when all except p are higher-order, and don't affect the dynamics of p, we also examined them in non-Erdös-Rènyi networks with strong motif structure (Figure 3.7). The full motif dynamics are derived in the Methods of this chapter, before restricting to lower-dimensional scenarios.

3.4 DISCUSSION

We have developed a theory for spike timing-dependent plasticity in weakly-coupled recurrent networks of exponential integrate-and-fire neurons. We used this framework to derive a low-dimensional dynamical system capturing the plasticity of two-synapse motifs. The resulting system naturally classifies STDP rules into two categories: 1) rules with an imbalance between potentiation and depression and plasticity dominated by the firing rates of neurons in the network, and 2) rules with balanced potentiation and depression in which different sources of spiking covariance interact with the STDP rule to determine network structure. In the latter case, the importance of spiking covariances due to forward connections, reciprocal connections, and common inputs creates new equilibrium points for the weighted motif structure of the network. For balanced, additive Hebbian STDP, these new equilibrium points are unstable. The nullcline manifolds that emanate from them divide the motif space into regions where different types of synaptic weight structure are either promoted or suppressed. When the balance in the STDP rule is tilted towards depression, regions where motifs are promoted or suppressed can both be accessible. For balanced STDP, any mechanism controlling spiking covariance in the network may affect how the network structure evolves. Thus, spike initiation dynamics [166, 167, 168, 169], spike-frequency adaptation [170, 171], synaptic inhibition [172, 173, 174] and passive membrane properties [175] could all, in addition to controlling firing rates, drive motif dynamics.

3.4.1 STDP in recurrent networks

In contrast to previous studies of uncorrelated Poisson-spiking neurons [131, 100], we consider the case where the intrinsic variability of neurons' spike trains is the only source of spiking covariance, necessitating a finite sized network ($\epsilon = 1/(Np_0) > 0$). There is little difference between our results and those of past studies [131, 100] when the learning rule is unbalanced. If there is a balance between potentiation and depression, however, our theory shows how internally generated spiking covariances play a strong role in STDP-induced formation of self-organized structure. Furthermore, our use of integrate-and-fire models allows our theory

to predict the evolution of network structure without fixing the statistics of individual or joint spiking activity.

We have focused here on networks composed only of excitatory neurons, a clear oversimplification of actual neural systems. The inclusion of inhibitory neurons would not, however, qualitatively change any of the results shown. Their effect on the plasticity of motifs can be understood by first considering their effect on the spike train covariances: in the first-order truncation of the spiking covariances (Eq. 7), inhibitory neurons would provide additional common inputs to pairs of excitatory cells. If the inhibitory-excitatory projections are not plastic and have Erdös-Renyi connectivity, this would add a constant term to dp/dt. How the plasticity of inhibitory synapses [176, 177, 115, 129] interacts with excitatory plasticity to shape motif structure in neuronal networks remains an exciting open area of inquiry.

3.4.2 Stability of learned network structures

Early studies of long-term plasticity, which gave rise to the phenomenological plasticity model we used, focused on the relative timing of action potentials. More recent experiments have shown that neurons' firing rates and the postsynaptic membrane voltage and spike patterns all affect the shape of measured STDP curves [105, 39, 38, 40, 178]. More complicated models of long-term plasticity, based on spike-triplet- or voltage-dependent STDP [106, 108] or on calcium thresholds for the induction of depression and potentiation [49, 80, 76], can replicate many of these complexities. The observation that firing rates undergo large fluctuations over slow timescales [179, 180, 181, 182, 183] suggests that *in vivo* STDP may transition between unbalanced potentiation- and depression-dominated regimes. While long-term plasticity can be strongly affected by pre- and postsynaptic firing rates, connectivity motifs and spiking covariance could determine the direction of plasticity during transitions between potentiation- and depression-dominated regimes. While our paper provides an initial framework to study how STDP shapes structure in recurrent networks, a more realistic learning rule than that used here (Eq. (2.1)) will be needed to address these issues.

The additive, Hebbian STDP model we used here gives rise to splitting of synaptic weights: individual weights potentiate to some upper bound, or depress to a lower bound.

This produces a bimodal distribution of synaptic weights, while experimentally observed weight distributions tend to be unimodal and long-tailed [91, 151, 184, 185]. Modifications of this model, such as introducing axonal or dendritic delays or weight-dependence of plasticity, can yield weight distributions more closely resembling those observed in neural tissue [93, 186, 95, 121, 97]. Depending on the modification made (delays vs weight-dependence), either the same or similar theories for motif plasticity can be derived using the methods presented in our study. Strong weight dependence, however, forces every weight to the same value so that the baseline motif frequencies completely determine the structure of the weight matrix (S4 Text, 3.6.5). The dynamics of motifs under more realistic models of synaptic plasticity remain to be studied.

A major feature of STDP is that it can potentiate temporally correlated inputs [109]. Since synchronous inputs are effective at driving postsynaptic spiking, this can give rise to pathological activity in recurrent networks [127]. Synaptic depression driven by postsynaptic spikes, independent of presynaptic activity, can stabilize postsynaptic firing rates during STDP [99, 132]. Such additional rate-dependent terms of the plasticity rule can also stabilize the full weight matrix [123] and thus give rise to stable motif configurations. Recent work has focused on the necessity of homeostatic mechanisms, including synaptic scaling [187] or inhibitory plasticity [176], in stabilizing both the activity and structure of neural networks [188, 189, 190, 124, 191, 129]. Since balanced STDP can give rise to bistability of mean synaptic weights in a network (Figure 3.4B), it could also provide a mechanism for assembly formation (selected weights potentiate, while other weights depress). Mechanisms of metaplasticity [192], operating on a similar timescale to STDP, could give rise to such a balance. This suggests a novel role for metaplasticity in controlling not only single-neuron excitability but also the self-organization of microcircuits in recurrent networks.

3.4.3 Plasticity of motifs

Early studies on STDP focused on isolated pairs of reciprocally connected neurons, showing that the type of STDP we study tends to induce competition between reciprocal synapses (Figure 1B,C; [89]). Since then, many simulation studies have investigated how STDP affects

the structure and activity of recurrent networks [126, 193, 194, 195, 129, 115], commonly examining the emergence of highly connected clusters. Reduced theories exposing how STDP shapes network-level structure have, however, been difficult to obtain. Most have examined the average synaptic weight in a network [196, 197], focusing on the relationship between network-averaged firing rates and mean synaptic weights (p) but neglecting spiking covariance. Mean-field theories are accurate for fully homogenous networks, however if all neurons have the same weighted in- and out-degrees there is no plasticity of two-synapse motifs (S3 Text, 3.6.3).

The few reduced theories examining STDP of higher-order network structure have focused on the question of how STDP controls open chains versus recurrent loops. One study compared the mean strengths of feedforward versus recurrent inputs in a network receiving synchronous stimulation [165], but did so for a neuron that made no feedback connections to the network – effectively only taking into account the first term of Eq. (3.3). Another study examined the strength of loops in a network of linear excitatory neurons, showing that STDP tends to reduce the total number of loops (of all lengths) in a network [164]. Our theory is restricted to two-synapse loops. While we have shown that these can potentiate (as in Figure 3.6C), [164] predicts that longer loops would meanwhile be weakened. Whether this is the case with balanced STDP driven by more realistic neuron models remains to be seen.

There is a growing body of evidence that cortical networks exhibit fine-scale structure [150, 151, 91, 152]. Experimental studies have shown that such microcircuits depend on sensory experience [198, 199]. Our work provides an important advance towards explicitly linking the plasticity rules that control individual synapses and the emergent microcircuits of cortical networks. We have shown that synaptic plasticity based only on temporally precise spike-train covariance can give rise to a diversity and, under certain conditions, multistability of motif configurations. Motifs can have a strong influence on pairwise and population-level activity [155, 156, 157, 158, 159, 160, 161, 136, 162, 163, 110], suggesting that precise spike timing may play a role in how networks reorganize patterns of connectivity in order to learn computations.

3.5 METHODS

Neuron and network models are as described in Chapter 2.

3.5.1 Derivation of motif dynamics

The baseline structure of the network is defined by the adjacency matrix $\mathbf{W}^0$. The frequencies of different motifs are:

$$\begin{aligned} p_0 &= \frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}, \\ q_0^{\text{div}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}^0_{ik}\mathbf{W}^0_{jk} - p_0^2, \\ q_0^{\text{con}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}^0_{ik}\mathbf{W}^0_{ij} - p_0^2, \\ q_0^{\text{ch}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}^0_{ij}\mathbf{W}^0_{jk} - p_0^2. \\ q_0^{\text{rec}} &= \frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}\mathbf{W}^0_{ji} - p_0^2. \end{aligned} \tag{3.16}$$

Each of the q_0 parameters refers to a different two-synapse motif. In divergent motifs (q_0^{div}), one neuron k projects to two others, i and j. In convergent motifs (q_0^{con}), two neurons k and j project to a third, i. In chain motifs (q_0^{ch}), neuron k projects to neuron j, which projects to neuron i. Finally, in recurrent motifs (q_0^{rec}) two neurons connect reciprocally. In each of these equations, we subtract off p_0^2 to correct for the baseline frequencies expected in Erdös-Rényi random networks. So, these parameters measure above-chance levels of motifs in the adjacency matrix $\mathbf{W}^0$.

We extend this motif definition to a weighted version, given by Eqs. (3.9). Since our linear response theory for synaptic plasticity requires weak synapses, here we explicitly scale by the mean in-degree $\epsilon = \frac{1}{Np_0}$. In contrast to the motif frequencies, which depend only on

the adjacency matrix $\mathbf{W}^0$, the motifs here also depend on the weight matrix $\mathbf{W}$.

$$\begin{aligned}
\epsilon p &= \frac{1}{N^2}\sum_{i,j}\mathbf{W}_{ij}, \\
\epsilon^2 q^{\mathrm{div}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\mathbf{W}_{jk} - \epsilon^2 p^2, \\
\epsilon^2 q^{\mathrm{con}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\mathbf{W}_{ij} - \epsilon^2 p^2, \\
\epsilon^2 q^{\mathrm{ch}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}_{ij}\mathbf{W}_{jk} - \epsilon^2 p^2, \\
\epsilon q_{\mathrm{X}}^{\mathrm{rec}} &= \frac{1}{N^2}\sum_{i,j}\mathbf{W}_{ij}\mathbf{W}^0_{ji} - \epsilon p p_0, \\
\epsilon q_{\mathrm{X}}^{\mathrm{div}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\mathbf{W}^0_{jk} - \epsilon p p_0, \\
\epsilon q_{\mathrm{X}}^{\mathrm{con}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\mathbf{W}^0_{ij} - \epsilon p p_0, \\
\epsilon q_{\mathrm{X}}^{\mathrm{ch,A}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}_{ij}\mathbf{W}^0_{jk} - \epsilon p p_0, \\
\epsilon q_{\mathrm{X}}^{\mathrm{ch,B}} &= \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}^0_{ij}\mathbf{W}_{jk} - \epsilon p p_0
\end{aligned} \tag{3.17}$$

Here we have defined the two-synapse motifs, as well as five auxiliary variables, $\{q_{\mathrm{X}}\}$. These mixed motifs, defined by products of the weight and adjacency matrices, measure the strength of synapses *conditioned* on their being part of a motif. The motifs $\{q\}$, on the other hand, measure the total strength of the motifs. While the variables $\{q_{\mathrm{X}}\}$ are not of direct interest, we will see that they are required in order to close the system of equations. In comparison to the motif *frequencies* $\{q_0\}$, which measure motif frequencies in comparison to an independently *connected* network, the motif *strengths* are defined relative to an independently *weighted* network.

We also scale the amplitude of individual synaptic changes, $L(s)$, by ϵ. We now go through the derivation of dp/dt, dq^{div}/dt and $dq_{\mathrm{X}}^{\mathrm{div}}/dt$ as examples; the other six variables follow the same steps. First, note that the spike train cross-covariance matrix of the network,

Eq. (2.4), can be expanded in the Fourier domain around the baseline covariance $\mathbf{C}^0(\omega)$:

$$\mathbf{C}(\omega) = \left(\sum_{i=0}^{\infty} (\mathbf{W}\cdot\mathbf{K})^i\right)\mathbf{C}^0(\omega)\left(\sum_{j=0}^{\infty}((\mathbf{W}\cdot\mathbf{K})^*)^i\right) \tag{3.18}$$

where the interaction matrix $\mathbf{W}\cdot\mathbf{K}$ is the element-wise product of the weight matrix $\mathbf{W}$ and the matrix of filters, $\mathbf{K}$. Elements of $\mathbf{K}$ are defined by the postsynaptic neuron's linear response function and the synaptic filter, as in Chapter 2. Powers of $\mathbf{W}\cdot\mathbf{K}$ represent lengths of paths through the network. Only taking into account up to length one paths yields (for $i \neq j$):

$$\mathbf{C}_{ij}(s) \approx \underbrace{\left(\mathbf{W}_{ij}\mathbf{K}_{ij} * \mathbf{C}^0_{jj}\right)(s)}_{\text{forwardconnection}} + \underbrace{\left(\mathbf{C}^0_{ii} * \mathbf{W}_{ji}\mathbf{K}^-_{ji}\right)(s)}_{\text{backwardconnection}} + \underbrace{\sum_k \left(\mathbf{W}_{ik}\mathbf{K}_{ik} * \mathbf{C}^0_{kk} * \mathbf{W}_{jk}\mathbf{K}^-_{jk}\right)(s)}_{\text{commoninputs}}. \tag{3.19}$$

where we have inverse Fourier transformed for convenience in the following derivation and $\mathbf{K}^-(t) = \mathbf{K}(-t)$.

Differentiating each motif with respect to time, using the fast-slow STDP theory Eq. (2.2) and inserting the first-order truncation of the cross-covariance functions, Eq. (3.3), yields:

$$\begin{aligned}\epsilon\frac{dp}{dt} = \frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}\int_{-\infty}^{\infty}\epsilon L(s)\Big(r_i r_j + \delta_{ij}\mathbf{C}^0_{ij}(s) + \left(\mathbf{W}_{ij}\mathbf{K}_{ij} * \mathbf{C}^0_{jj}\right)(s) \\ + \left(\mathbf{C}^0_{ii} * \mathbf{W}_{ji}\mathbf{K}^-_{ji}\right)(s) + \sum_k \left(\mathbf{W}_{ik}\mathbf{K}_{ik} * \mathbf{C}^0_{kk} * \mathbf{W}_{jk}\mathbf{K}^-_{jk}\right)\Big)ds\end{aligned} \tag{3.20}$$

$$\begin{aligned}\epsilon^2\frac{dq^{\text{div}}}{dt} = \frac{2}{N^3}\sum_{i,j,k}\Big[\mathbf{W}_{ik}\mathbf{W}^0_{jk}\int_{-\infty}^{\infty}\epsilon L(s)\Big(r_j r_k + \delta_{jk}\mathbf{C}^0_{jk}(s) + \left(\mathbf{W}_{jk}\mathbf{K}_{jk} * \mathbf{C}^0_{kk}\right)(s) \\ + \left(\mathbf{C}^0_{jj} * \mathbf{W}_{kj}\mathbf{K}^-_{kj}\right)(s) + \sum_l \left(\mathbf{W}_{jl}\mathbf{K}_{jl} * \mathbf{C}^0_{ll} * \mathbf{W}_{kl}\mathbf{K}^-_{kl}\right)\Big)ds\Big] - 2\epsilon^2 p\frac{dp}{dt}\end{aligned} \tag{3.21}$$

$$\begin{aligned}\epsilon\frac{dq^{\text{div}}_{\text{X}}}{dt} = \frac{1}{N^3}\sum_{i,j,k}\mathbf{W}^0_{jk}\mathbf{W}^0_{ik}\int_{-\infty}^{\infty}\epsilon L(s)\Big(r_i r_k + \delta_{ik}\mathbf{C}^0_{ik}(s) + \left(\mathbf{W}_{ik}\mathbf{K}_{ik} * \mathbf{C}^0_{kk}\right)(s) \\ + \left(\mathbf{C}^0_{ii} * \mathbf{W}_{ki}\mathbf{K}^-_{ki}\right)(s) + \sum_l \left(\mathbf{W}_{il}\mathbf{K}_{il} * \mathbf{C}^0_{ll} * \mathbf{W}_{kl}\mathbf{K}^-_{kl}\right)\Big)ds\Big] - \epsilon p_0\frac{dp}{dt}\end{aligned} \tag{3.22}$$

We now define the network-averaged firing rate r, spike train autocovariances C^0 and linear response function. Since we model all postsynaptic currents with the same shape, this makes

the matrix $\mathbf{K}$ a constant matrix; we replace its elements with the scalar K. Also neglecting the weight bounds in $L(s)$ allows us to write:

$$\frac{dp}{dt} = r^2 S \frac{1}{N^2} \sum_{i,j} \mathbf{W}^0_{ij} + S_F \frac{1}{N^2} \sum_{i,j} \mathbf{W}^0_{ij} \mathbf{W}_{ij} + S_B \frac{1}{N^2} \sum_{i,j} \mathbf{W}^0_{ij} \mathbf{W}_{ji} + S_C \frac{1}{N^2} \sum_{i,j,k} \mathbf{W}^0_{ij} \mathbf{W}_{ik} \mathbf{W}_{jk} \tag{3.23}$$

$$\begin{aligned} \epsilon \frac{dq^{\text{div}}}{dt} &= r^2 S \frac{2}{N^3} \sum_{i,j,k} \mathbf{W}_{ik} \mathbf{W}^0_{jk} + S_F \frac{2}{N^3} \sum_{i,j,k} \mathbf{W}_{ik} \mathbf{W}^0_{jk} \mathbf{W}_{jk} \\ &\quad + S_B \frac{2}{N^3} \sum_{i,j,k} \mathbf{W}_{ik} \mathbf{W}^0_{jk} \mathbf{W}_{kj} + S_C \frac{2}{N^3} \sum_{i,j,k,l} \mathbf{W}_{ik} \mathbf{W}^0_{jk} \mathbf{W}_{jl} \mathbf{W}_{kl} - 2\epsilon p \frac{dp}{dt} \end{aligned} \tag{3.24}$$

$$\begin{aligned} \frac{dq^{\text{div}}_{\text{X}}}{dt} &= r^2 S \frac{1}{N^3} \sum_{i,j,k} \mathbf{W}^0_{jk} \mathbf{W}^0_{ik} + S_F \frac{1}{N^3} \sum_{i,j,k} \mathbf{W}^0_{jk} \mathbf{W}^0_{ik} \mathbf{W}_{ik} \\ &\quad + S_B \frac{1}{N^3} \sum_{i,j,k} \mathbf{W}^0_{jk} \mathbf{W}^0_{ik} \mathbf{W}_{ki} + S_C \frac{1}{N^3} \sum_{i,j,k,l} \mathbf{W}^0_{jk} \mathbf{W}^0_{ik} \mathbf{W}_{il} \mathbf{W}_{kl} - p_0 \frac{dp}{dt} \end{aligned} \tag{3.25}$$

where we have cancelled off an ϵ from the left and right-hand sides. We have absorbed the integrals over the STDP rule and the spiking covariances into $r^2 S$, S_F, S_B and S_C. These correspond, respectively, to the total STDP-weighted spiking covariances from chance coincidence, forward connections, backward connections, and common input:

$$S = \int_{-\infty}^{\infty} L(s)\, ds \tag{3.26}$$

$$S_F = \int_{-\infty}^{\infty} L(s) \left(K(t) * C^0(s) \right) ds \tag{3.27}$$

$$S_B = \int_{-\infty}^{\infty} L(s) \left(C^0(s) * K^-(t) \right) ds \tag{3.28}$$

$$S_C = \int_{-\infty}^{\infty} L(s) \left(K(t) * C^0(s) * K^-(t) \right) ds \tag{3.29}$$

These parameters depend on the spike train auto-covariance $C^0(s)$ and interaction kernel $K(t; A)$ of neurons. As the mean synaptic weight p changes, the average firing rate r will change and this will also affect $C^0(s)$ and $K(t; A)$. So r , S_F, S_B and S_C are implicitly functions of p and thus evolve with the network. We have assumed weak synapses, so we expect small changes in firing rates and thus fix these at their value at $p = p_0 W^{\max}/2$,

making r , S_F, S_B and S_C constant parameters. In order to determine the impact of this approximation on our results, we compared the evolution of motifs in the reduced theory while re-calculating r , S_F, S_B and S_C at every time-step. The approximation introduced negligible errors in calculating the evolution of the weighted motifs (S4 Fig, 3.6.6).

Each dynamical equation now contains four different sums of products of the weight and adjacency matrices. First examining dp/dt, we see that the first three sums correspond to defined motifs: $1/N^2 \sum_{i,j} \mathbf{W}^0_{ij} = p_0$, $1/N^2 \sum_{i,j} \mathbf{W}^0_{ij}\mathbf{W}_{ij} = p$ and $1/N^2 \sum_{i,j} \mathbf{W}^0_{ij}\mathbf{W}_{ji} = q_{\mathrm{X}}^{\mathrm{rec}} + pp_0$. The last term in Eq. (3.23), however, corresponds to a third-order motif mixed between the weight and adjacency matrices. Similarly, third- and fourth-order mixed motifs appear in Eqs. 3.24 and 3.25. In order to calculate these, we extend a re-summing technique developed in [162]. We assume that there are no third- or higher-order correlations between elements of the weight and/or adjacency matrices, and approximate the frequency of each of these higher-order motifs by the number of ways it can be composed of one and two-synapse motifs. For a third order motif, this corresponds to adding up the likelihoods that all three synapses occur by chance and that each possible combination of one synapse and a two-synapse motif occur. In Eq. (3.23),

$$\sum_{i,j,k} \mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk} \approx \epsilon^2 N^3 \left(p_0 \left(q^{\mathrm{div}} + p^2 \right) + p \left(q_{\mathrm{X}}^{\mathrm{con}} + q_{\mathrm{X}}^{\mathrm{ch,B}} \right) \right). \tag{3.30}$$

and for the four-synapse motif in Eq. (3.24),

$$\sum_{i,j,k,l} \mathbf{W}_{ik}\mathbf{W}^0_{jk}\mathbf{W}_{jl}\mathbf{W}_{kl} \approx \epsilon^3 N^4 \left(p^3 p_0 + p^2 \left(q_{\mathrm{X}}^{\mathrm{div}} + q_{\mathrm{X}}^{\mathrm{con}} + q_{\mathrm{X}}^{\mathrm{ch,B}} \right) + pp_0 \left(q^{\mathrm{div}} + q^{\mathrm{ch}} \right) + q^{\mathrm{div}} q_{\mathrm{X}}^{\mathrm{div}} + q^{\mathrm{ch}} q_{\mathrm{X}}^{\mathrm{con}} \right) \tag{3.31}$$

This re-summing, along with the inclusion of the mixed motifs $\{q_{\mathrm{X}}\}$, is what allows us to close the motif dynamics. Re-summing each third- and fourth-order motif in our system in terms of two-synapse motifs yields, after simplification, the final motif dynamics:

$$\frac{dp}{dt} = p_0 r^2 S + \epsilon \left[pS_F + \left(q_{\mathrm{X}}^{\mathrm{rec}} + p_0 p \right) S_B + \frac{1}{p_0} \left(p_0 \left(q^{\mathrm{div}} + p^2 \right) + p \left(q_{\mathrm{X}}^{\mathrm{con}} + q_{\mathrm{X}}^{\mathrm{ch,B}} \right) \right) S_C \right] \tag{3.32}$$

$$\frac{dq^{\mathrm{div}}}{dt} = 2r^2 S q_{\mathrm{X}}^{\mathrm{div}} + 2\epsilon \left[q^{\mathrm{div}} S_F + \left(p_0 q^{\mathrm{ch}} + p q_{\mathrm{X}}^{\mathrm{div}} \right) S_B + \frac{1}{p_0} \left(q^{\mathrm{ch}} \left(q_{\mathrm{X}}^{\mathrm{con}} + p p_0 \right) + q_{\mathrm{X}}^{\mathrm{div}} \left(q^{\mathrm{div}} + p^2 \right) \right) S_C \right] \tag{3.33}$$

$$\frac{dq^{\mathrm{con}}}{dt} = 2r^2 S q_{\mathrm{X}}^{\mathrm{con}} + 2\epsilon \left[q^{\mathrm{con}} S_F + \left(p_0 q^{\mathrm{ch}} + p q_{\mathrm{X}}^{\mathrm{con}} \right) S_B + \frac{1}{p_0} \left(q^{\mathrm{con}} \left(q_{\mathrm{X}}^{\mathrm{ch,B}} + p p_0 \right) + q_{\mathrm{X}}^{\mathrm{con}} \left(q^{\mathrm{div}} + p^2 \right) \right) S_C \right] \tag{3.34}$$

$$\begin{aligned} \frac{dq^{\mathrm{ch}}}{dt} = {} & r^2 S \left(q_{\mathrm{X}}^{\mathrm{ch,A}} + q_{\mathrm{X}}^{\mathrm{ch,B}} \right) + \epsilon \Big[2 q^{\mathrm{ch}} S_F + \left(p_0 \left(q^{\mathrm{con}} + q^{\mathrm{div}} \right) + p \left(q_{\mathrm{X}}^{\mathrm{ch,A}} + q_{\mathrm{X}}^{\mathrm{ch,B}} \right) \right) S_B \\ & + \frac{1}{p_0} \left(\left(q^{\mathrm{div}} + p^2 \right) \left(q_{\mathrm{X}}^{\mathrm{ch,A}} + q_{\mathrm{X}}^{\mathrm{ch,B}} \right) + q^{\mathrm{ch}} \left(q_{\mathrm{X}}^{\mathrm{ch,B}} + p p_0 \right) + q^{\mathrm{con}} \left(q_{\mathrm{X}}^{\mathrm{con}} + p p_0 \right) \right) S_C \Big] \end{aligned} \tag{3.35}$$

$$\begin{aligned} \frac{dq_{\mathrm{X}}^{\mathrm{rec}}}{dt} = {} & r^2 S q_0^{\mathrm{rec}} + \epsilon \Big[q_{\mathrm{X}}^{\mathrm{rec}} S_F + (1 - p_0) \left(q_{\mathrm{X}}^{\mathrm{rec}} + p p_0 \right) S_B \\ & + \frac{1}{p_0} \left(q_0^{\mathrm{rec}} \left(q^{\mathrm{div}} + p^2 \right) + q_{\mathrm{X}}^{\mathrm{ch,B}} \left(q_{\mathrm{X}}^{\mathrm{ch,B}} + p p_0 \right) + q_{\mathrm{X}}^{\mathrm{con}} \left(q_{\mathrm{X}}^{\mathrm{con}} + p p_0 \right) \right) S_C \Big] \end{aligned} \tag{3.36}$$

$$\frac{dq_{\mathrm{X}}^{\mathrm{div}}}{dt} = r^2 S q_0^{\mathrm{div}} + \epsilon \left[q_{\mathrm{X}}^{\mathrm{div}} S_F + \left(p q_0^{\mathrm{div}} + p_0 q_{\mathrm{X}}^{\mathrm{ch,B}} \right) S_B + \frac{1}{p_0} \left(q_0^{\mathrm{div}} \left(q^{\mathrm{div}} + p^2 \right) + q_{\mathrm{X}}^{\mathrm{ch,B}} \left(q_{\mathrm{X}}^{\mathrm{con}} + p p_0 \right) \right) S_C \right] \tag{3.37}$$

$$\frac{dq_{\mathrm{X}}^{\mathrm{con}}}{dt} = r^2 S q_0^{\mathrm{con}} + \epsilon \left[q_{\mathrm{X}}^{\mathrm{con}} S_F + \left(p q_0^{\mathrm{con}} + p_0 q_{\mathrm{X}}^{\mathrm{ch,A}} \right) S_B + \frac{1}{p_0} \left(q_0^{\mathrm{con}} \left(q^{\mathrm{div}} + p^2 \right) + q_{\mathrm{X}}^{\mathrm{con}} \left(q_{\mathrm{X}}^{\mathrm{ch,B}} + p p_0 \right) \right) S_C \right] \tag{3.38}$$

$$\frac{dq_{\mathrm{X}}^{\mathrm{ch,A}}}{dt} = r^2 S q_0^{\mathrm{ch}} + \epsilon \left[q_{\mathrm{X}}^{\mathrm{ch,A}} S_F + \left(p q_0^{\mathrm{ch}} + p_0 q_{\mathrm{X}}^{\mathrm{con}} \right) S_B + \frac{1}{p_0} \left(q_0^{\mathrm{ch}} \left(q^{\mathrm{div}} + p^2 \right) + q_{\mathrm{X}}^{\mathrm{con}} \left(q_{\mathrm{X}}^{\mathrm{con}} + p p_0 \right) \right) S_C \right] \tag{3.39}$$

$$\frac{dq_{\mathrm{X}}^{\mathrm{ch,B}}}{dt} = r^2 S q_0^{\mathrm{ch}} + \epsilon \left[q_{\mathrm{X}}^{\mathrm{ch,B}} S_F + \left(p q_0^{\mathrm{ch}} + p_0 q_{\mathrm{X}}^{\mathrm{div}} \right) S_B + \frac{1}{p_0} \left(q_0^{\mathrm{ch}} \left(q^{\mathrm{div}} + p^2 \right) + q_{\mathrm{X}}^{\mathrm{ch,B}} \left(q_{\mathrm{X}}^{\mathrm{ch,B}} + p p_0 \right) \right) S_C \right] \tag{3.40}$$

Examination of these equations reveals how different types of joint spiking activity affect motif dynamics. Chance spiking coincidence (the $r^2 S$ terms) couple each motif to the mixed

version of itself, and each mixed motif to the baseline structure of the adjacency matrix. With Hebbian STDP and excitatory synapses, $S_F > 0$ and $S_B < 0$. So, spiking covariance from forward connections provide positive feedback, reinforcing the current network structure. Spiking covariance from backward connections and common input couple divergent, convergent and chain motifs to each other.

The dynamics on the invariant set (Results: Balanced STDP of the mean synaptic weight, Figure 3.3) were plotted in MATLAB. The vector fields of Figs 8 and 9 were calculated in XPPAUT [200]. For those figures, results from simulations of the full spiking network were plotted in MATLAB and then overlaid on the vector fields from XPPAUT.

3.5.1.1 Plasticity of loops and open chains The chain variable q^{ch} includes both open chains and recurrent loops. (open chains correspond to $k \neq i$ in the definition of q^{ch}, Eq. (3.17), and recurrent loops to $k = i$.) As in the main text, we break q^{ch} into these two cases: $q^{\text{ch}} = q^{\text{rec}} + q^{\text{op}}$, where

$$\begin{aligned} \epsilon^2 q^{\text{rec}} &= \frac{1}{N^3} \sum_{i,j,k} \delta_{ik} \mathbf{W}_{ij} \mathbf{W}_{jk} = \frac{1}{N^3} \sum_{i,j} \mathbf{W}_{ij} \mathbf{W}_{ji} \\ \epsilon^2 q^{\text{op}} &= \frac{1}{N^3} \sum_{i,j,k} (1 - \delta_{ik}) \mathbf{W}_{ij} \mathbf{W}_{jk} - \epsilon^2 p^2 \end{aligned} \tag{3.41}$$

We also define an auxiliary variable which we will require in the dynamics of q^{rec}:

$$\epsilon^2 q^{\text{rec}}_{\text{X2}} = \frac{1}{N^3} \sum_{i,j} \mathbf{W}^2_{ij} \mathbf{W}^0_{ji} \tag{3.42}$$

which is proportional to the conditioned second moment of weights that are part of disynaptic loops. The dynamics of q^{rec} are calculated exactly as for the other motifs and are:

$$\frac{1}{2\epsilon} \frac{dq^{\text{rec}}}{dt} = r^2 S p_0 \left(q^{\text{rec}}_{\text{X}} + p p_0 \right) + S_F q^{\text{rec}} + S_B q^{\text{rec}}_{\text{X2}} \tag{3.43}$$

where the new auxiliary variable obeys

$$\frac{1}{2\epsilon} \frac{dq^{\text{rec}}_{\text{X2}}}{dt} = r^2 S p_0 \left(q^{\text{rec}}_{\text{X}} + p p_0 \right) + S_F q^{\text{rec}}_{\text{X2}} + S_B q^{\text{rec}} \tag{3.44}$$

We can then recover the dynamics of open chains as:

$$\begin{aligned}\frac{dq^{\text{op}}}{dt} &= \frac{dq^{\text{ch}}}{dt} - \frac{dq^{\text{rec}}}{dt} \\ &= r^2 S\left(q_{\text{X}}^{\text{ch,A}} + q_{\text{X}}^{\text{ch,B}}\right) + \epsilon\Big[-2r^2 S p_0\left(q_{\text{X}}^{\text{rec}} + pp_0\right) + 2S_F q^{\text{op}} \\ &\quad + \left(p_0\left(q^{\text{con}} + q^{\text{div}}\right) + p\left(q_{\text{X}}^{\text{ch,A}} + q_{\text{X}}^{\text{ch,B}}\right) - 2q_{\text{X2}}^{\text{rec}}\right) S_B \\ &\quad \frac{1}{p_0}\left(\left(q^{\text{div}} + p^2\right)\left(q_{\text{X}}^{\text{ch,A}} + q_{\text{X}}^{\text{ch,B}}\right) + q^{\text{ch}}\left(q_{\text{X}}^{\text{ch,B}} + pp_0\right) + q^{\text{con}}\left(q_{\text{X}}^{\text{con}} + pp_0\right)\right) S_C\Big]\end{aligned} \tag{3.45}$$

3.5.1.2 Unbalanced STDP When there is an imbalance between the net amounts of potentiation and depression in the STDP rule, the motif dynamics is governed by simpler equations. If $S \sim \mathcal{O}(1)$, the $\mathcal{O}(\epsilon)$ terms in Eqs. 3.32-3.40 are negligible. For each mixed motif,

$$q_{\text{X}}(t) = r^2 S q_0 t + q_{\text{X}}(0) \tag{3.46}$$

so that

$$p(t) = p_0 r^2 S t + p(0) \tag{3.47}$$

$$q^{\text{div}}(t) = q^{\text{div}}(0) + q_{\text{X}}^{\text{div}}(0) r^2 S t + \frac{1}{2} q_0^{\text{div}}\left(r^2 S\right)^2 t^2 \tag{3.48}$$

$$q^{\text{con}}(t) = q^{\text{con}}(0) + q_{\text{X}}^{\text{con}}(0) r^2 S t + \frac{1}{2} q_0^{\text{con}}\left(r^2 S\right)^2 t^2 \tag{3.49}$$

$$q^{\text{ch}}(t) = q^{\text{ch}}(0) + \left(q_{\text{X}}^{\text{ch,A}}(0) + q_{\text{X}}^{\text{ch,B}}(0)\right) r^2 S t + \frac{1}{2} q_0^{\text{ch}}\left(r^2 S\right)^2 t^2 \tag{3.50}$$

Writing $q_{\text{X}}^{\text{ch}} = q_{\text{X}}^{\text{ch,A}} + q_{\text{X}}^{\text{ch,B}}$ puts the dynamics for all the motifs in the same form. The motifs expand from the initial conditions and baseline structure of the network. Note that since the quadratic term is proportional to S^2, even when STDP is depression-dominated the long-term dynamics is expansive rather than contractive.

3.6 SUPPLEMENTARY INFORMATION

3.6.1 S2 Text: Strength of synaptic weights and stability of firing rates

Here, we varied the initial synaptic strength $p(0)$ and compared our theoretical predictions to empirical measurements from simulations for the mean synaptic strength $p(T)$, divergent $q^{\text{div}}(T)$, convergent $q^{\text{con}}(T)$, and chain $q^{\text{ch}}(T)$ motifs, where $T = 1000$ min. For $p(0) \leq 5 \times 10^{-3}$ there was very good agreement between our reduced theory and spiking network simulations. When $p(0) \geq 7 \times 10^{-3}$ we observed a clear quantitative disagreement between theory and simulations. Increasing $p(0)$ or $p(T)$ above 8×10^{-3} caused an instability in the recurrent network, with runaway excitation destabilizing the equilibrium state, preventing any perturbative theory. The stability boundary for the stationary firing rates is computed as in [201], by computing the eigenvalues of the network firing rate from a Fokker-Planck theory [146]. Throughout the manuscript we always used $p(0) \leq 5 \times 10^{-3}$ to ensure agreement between simulations and theory.

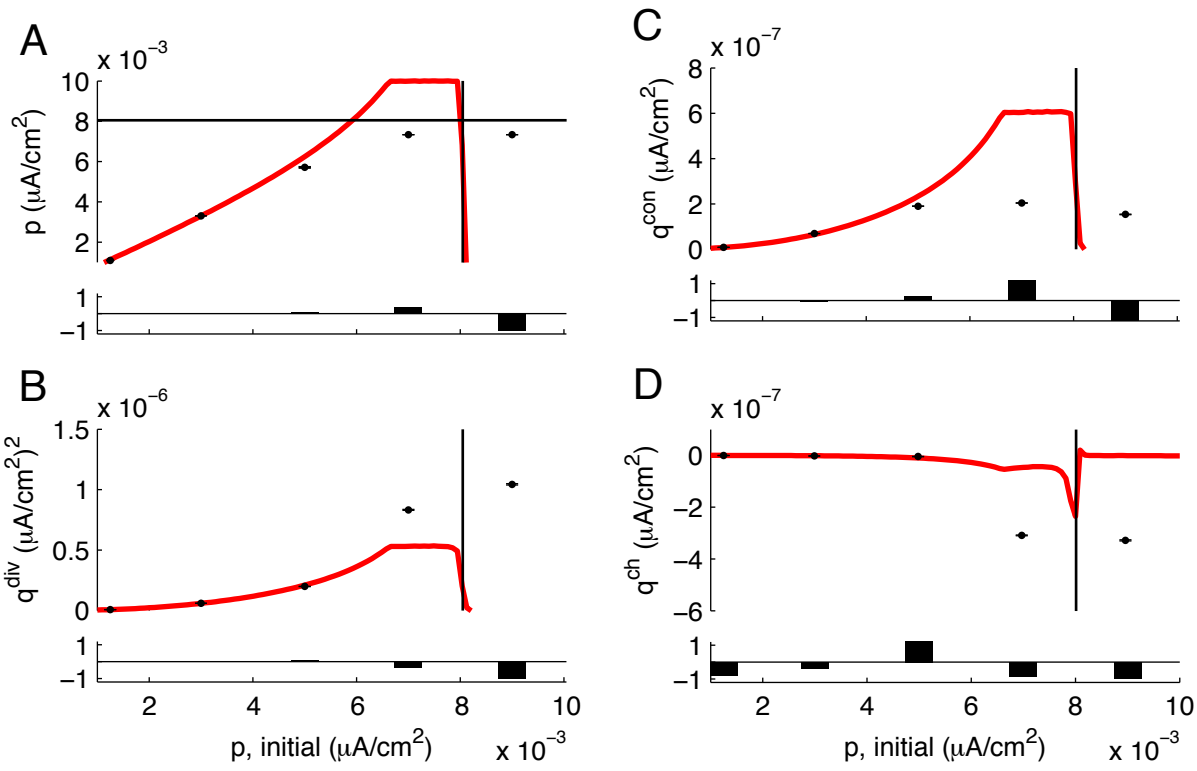

Figure 3.8: **STDP in networks with larger synaptic weights.** All panels: Red, motif theory of Eqs. (42)-(50) in main text. Black dots: result of spiking simulations. Error bars (smaller than marker size) are SEM. Black bars in lower part of each panel are the relative error of the theory to the simulations, (theory-sims)/(sims). (A) Final mean synaptic weight as a function of initial synaptic weight. Vertical and horizontal black lines mark the stability boundary for the stationary firing rates, where the network-averaged instantaneous firing rate loses stability [201]. For p beyond that boundary, we cannot use linear response theory for spike-train covariability (since there is no fixed point to linearize around) and so the theory will not match simulations. (B) Final strength of divergent motifs versus initial mean synaptic weight. (C) Final strength of convergent motifs versus initial mean synaptic weight. (D) Final strength of chain motifs versus initial mean synaptic weight.

3.6.2 S2 Fig: Truncated vs full spike train cross-covariances

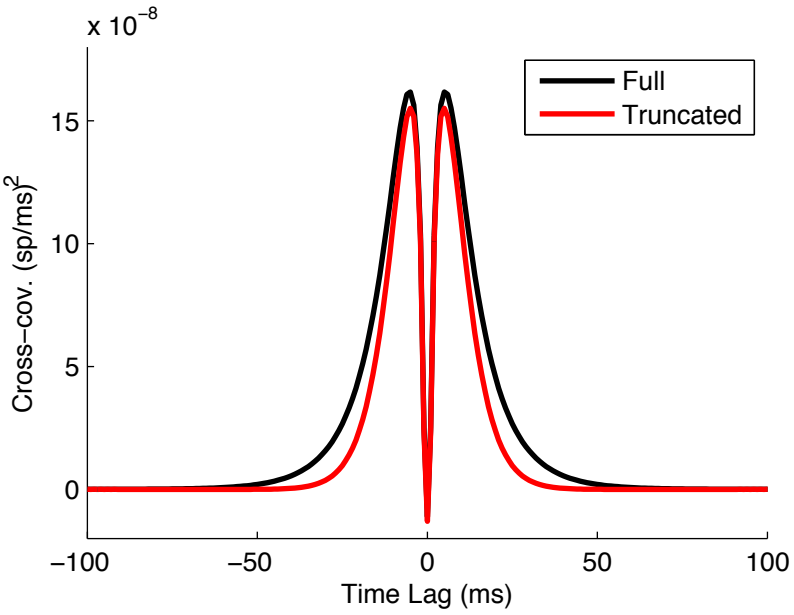

Figure 3.9: **Truncated vs full spike train cross-covariance functions.** Black, average spike train cross-covariance function of the network of Fig. 2. Red, truncated approximation (Eq. 7 of main text) of the average spike train cross-covariance function. When integrated against the balanced, depression dominated STDP rule, the full theory is 1.06 times as large as the truncated.

3.6.3 S3 Text: Motif plasticity in homogenous networks

If all neurons have the same weighted in- and out-degrees, then the motif dynamics simplify considerably. In such a homogenous network, divergent motifs obey:

$$\begin{aligned}\frac{dq^{\text{div}}}{dt} &= \frac{2}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\frac{d\mathbf{W}_{jk}}{dt} - 2p\frac{dp}{dt} \\ &= \frac{2}{N^3}\sum_{k}\left(\sum_{i}\mathbf{W}_{ik}\sum_{j}\frac{d\mathbf{W}_{jk}}{dt}\right) - 2p\frac{dp}{dt} \\ &= \frac{2}{N^3}Np\sum_{j,k}\frac{d\mathbf{W}_{jk}}{dt} - 2p\frac{dp}{dt} \\ &= 2p\frac{dp}{dt} - 2p\frac{dp}{dt} = 0\end{aligned} \tag{3.51}$$

Convergent and chain motifs are also neutrally stable. So, it is inhomogeneities in the network structure that give rise to drift of the motif structure.

3.6.4 S3 Fig: Balanced STDP in isolated pairs of neurons

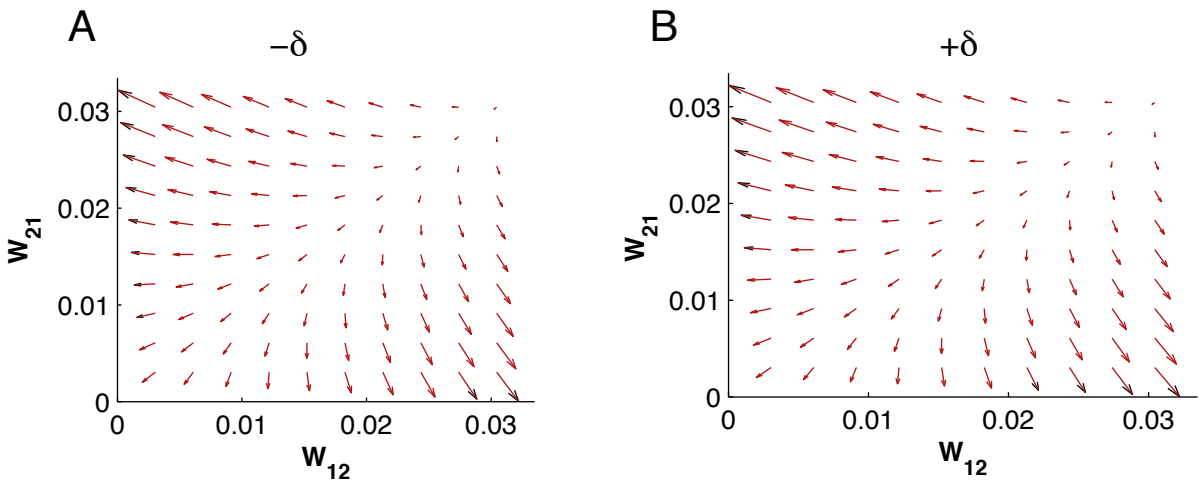

Figure 3.10: **Both types of balanced STDP lead to splitting of synaptic weights.** We take an isolated pair of neurons with the same intrinsic and synaptic parameters as in the full network. The neurons are reciprocally connected and we plot the (A) The depression-dominated balanced STDP rule. (B) The potentiation-dominated balanced STDP rule. Both are exactly the same as in the main paper. We show the dynamics of the weights in a pair of neurons isolated from the network. In both cases, the reciprocal loop is eliminated and initial conditions determine which synaptic weight is potentiated and which depressed.

3.6.5 S4 Text: Motif plasticity with weight-dependent STDP

The multiplicative STDP rule [93, 94] is:

$$\epsilon L(s) = \begin{cases} (\epsilon W^{\max} - \mathbf{W}_{ij}) f_+ e^{-\frac{|s|}{\tau_+}}, & \text{if } s \geq 0 \\ (\mathbf{W}_{ij})\,(-f_-)\, e^{-\frac{|s|}{\tau_-}}, & \text{if } s \leq 0, \end{cases} . \tag{3.52}$$

Each weight has a stable fixed point:

$$\mathbf{W}^*_{ij} = \mathbf{W}^0_{ij} \frac{f_+\tau_+}{f_+\tau_+ + f_-\tau_-} W^{\max} \tag{3.53}$$

Now the mean weight and two-synapse motifs evolve according to:

$$\epsilon\frac{dp}{dt} = \frac{1}{N^2}\sum_{i,j}\mathbf{W}^0_{ij}\int_{-\infty}^{\infty}\epsilon L(s)\left(r_i r_j + \delta_{ij}\mathbf{C}^0_{ij}(s)\right)ds \tag{3.54}$$

$$\epsilon^2\frac{dq^{\text{div}}}{dt} = \frac{2}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\mathbf{W}^0_{jk}\int_{-\infty}^{\infty}\epsilon L(s)\left(r_j r_k + \delta_{jk}\mathbf{C}^0_{jk}(s)\right)ds - 2\epsilon^2 p\frac{dp}{dt} \tag{3.55}$$

$$\epsilon^2\frac{dq^{\text{con}}}{dt} = \frac{2}{N^3}\sum_{i,j,k}\mathbf{W}_{ik}\mathbf{W}^0_{ij}\int_{-\infty}^{\infty}\epsilon L(s)\left(r_i r_j + \delta_{ij}\mathbf{C}^0_{ij}(s)\right)ds - 2\epsilon^2 p\frac{dp}{dt} \tag{3.56}$$

$$\epsilon^2\frac{dq^{\text{div}}}{dt} = \frac{2}{N^3}\sum_{i,j,k}\mathbf{W}_{ij}\mathbf{W}^0_{jk}\int_{-\infty}^{\infty}\epsilon L(s)\left(r_j r_k + \delta_{jk}\mathbf{C}^0_{jk}(s)\right)ds - 2\epsilon^2 p\frac{dp}{dt} \tag{3.57}$$

where we only examine the contribution of firing rates to the plasticity, assuming that the pre- and post-synaptic neurons' spike trains are uncorrelated. This corresponds to the observation that with multiplicative STDP, the weight-dependence of $L(s)$ dominates the dynamics of the weights. Inserting Eq. (3.52) and the motif definitions and assuming homogenous firing rates yields:

$$\frac{1}{\epsilon}\frac{dp}{dt} = r^2\left(p_0 W^{\max} f_+\tau_+ - p\left(f_+\tau_+ + f_-\tau_-\right)\right) \tag{3.58}$$

$$\frac{1}{\epsilon}\frac{dq^{\text{div}}}{dt} = 2r^2\left(f_+\tau_+ W^{\max} q_{\text{X}}^{\text{div}} - \left(f_+\tau_+ + f_-\tau_-\right)q^{\text{div}}\right) \tag{3.59}$$

$$\frac{1}{\epsilon}\frac{dq^{\text{con}}}{dt} = 2r^2\left(f_+\tau_+ W^{\max} q_{\text{X}}^{\text{con}} - \left(f_+\tau_+ + f_-\tau_-\right)q^{\text{con}}\right) \tag{3.60}$$

$$\frac{1}{\epsilon}\frac{dq^{\text{ch}}}{dt} = r^2\left(f_+\tau_+ W^{\max}\left(q_{\text{X}}^{\text{ch,A}} + q_{\text{X}}^{\text{ch,B}}\right) - 2\left(f_+\tau_+ + f_-\tau_-\right)q^{\text{ch}}\right) \tag{3.61}$$

The mixed divergent motifs obey:

$$\frac{dq_{\text{X}}^{\text{div}}}{dt} = r^2\left(f_+\tau_+ W^{\max} q_0^{\text{div}} - \left(f_+\tau_+ + f_-\tau_-\right)q_{\text{X}}^{\text{div}}\right) \tag{3.62}$$

and $q_{\text{X}}^{\text{con}}$, $q_{\text{X}}^{\text{ch,A}}$ and $q_{\text{X}}^{\text{ch,B}}$ obey exactly analogous equations. Defining $q_{\text{X}}^{\text{ch}} = q_{\text{X}}^{\text{ch,A}} + q_{\text{X}}^{\text{ch,B}}$ puts the dynamics of q^{ch} into the same form as those for q^{div} and q^{con}. Dropping the motif labels, since they obey the same dynamics, yields a three-dimensional system for (p, q, q_{X}) with steady state

$$\begin{pmatrix} p \\ q \\ q_{\text{X}} \end{pmatrix}^* = \begin{pmatrix} \frac{f_+\tau_+}{f_+\tau_+ + f_-\tau_-}p_0 W^{\max} \\ \left(\frac{f_+\tau_+}{f_+\tau_+ + f_-\tau_-}W^{\max}\right)^2 q_0 \\ \frac{f_+\tau_+}{f_+\tau_+ + f_-\tau_-}W^{\max} q_0 \end{pmatrix} \tag{3.63}$$

and Jacobian:

$$\begin{pmatrix} -(f_+\tau_+ + f_-\tau_-) & 0 & 0 \\ 0 & -(f_+\tau_+ + f_-\tau_-) & f_+\tau_+ W^{\max} \\ 0 & 0 & -(f_+\tau_+ + f_-\tau_-) \end{pmatrix} \tag{3.64}$$

The eigenvalue of the three-dimensional multiplicative STDP system is $-(f_+\tau_+ + f_-\tau_-)$ which is always negative, so the steady state is linearly stable. So, multiplicative STDP simply stabilizes whatever motif structure is embedded in the adjacency matrix.

3.6.6 S4 Fig: Rate-dependence of spike train covariability

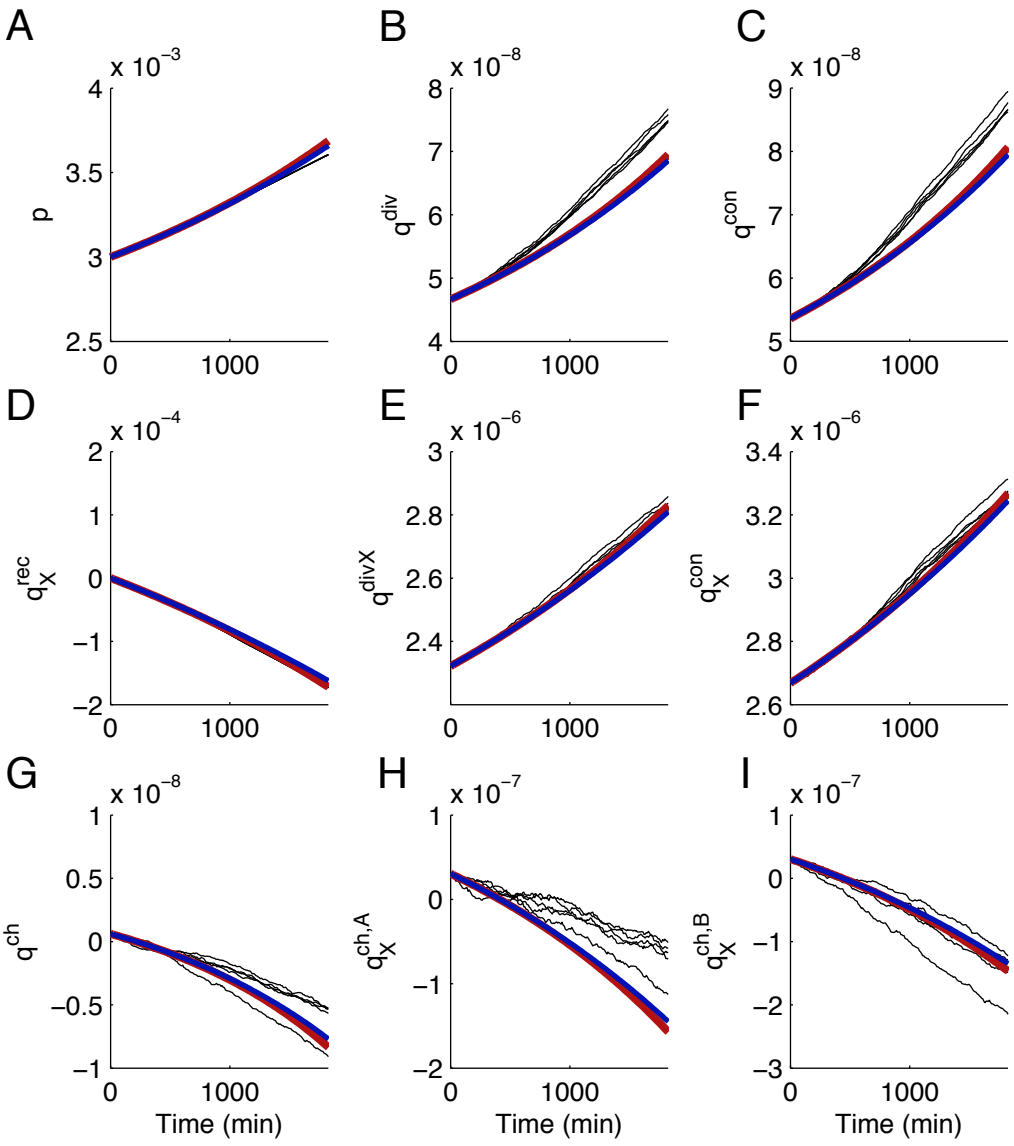

Figure 3.11: **Motif plasticity does not depend on firing rates.** In the main text, we examined the motif dynamics when the spike-train statistics are fixed. That is, for the motif theory we linearize the spike-train statistics around the initial state $r(t) = r(0)$, and similarly for the STDP-weighted spike train covariances S_F, S_B and S_C. Firing rates do, however, evolve with the mean synaptic weight. For $p = 0$, $r = 7.6$ sp/s while for $p = p_0 W^{\max}$, $r = 12.5$ sp/s. In order to check whether the approximation of using r, S_F, S_B and S_C calculated at initial conditions was accurate, we re-linearized the system at every time step of the slow evolution of synaptic weights so that the spiking statistics were not fixed. For the networks and parameters used here, this did not have a strong effect on the motif dynamics. We illustrate this by comparing the results of Fig. 7 to the motif dynamics with evolving spiking statistics. Red: motifs dynamics with fixed r, S_F, S_B, S_C - the spiking statistics are calculated by linearizing around the initial condition and held fixed. Blue: The spiking statistics (r, S_F, S_B, S_C) are re-linearized at every time step, so that the theory accounts for how firing rates and spike-train correlations depend on the net synaptic input neurons' receive.

4.0 SPIKE TIMING-DEPENDENT PLASTICITY OF HEBBIAN ASSEMBLIES

4.1 INTRODUCTION

In the previous sections, we developed a theory describing the plasticity of connectivity patterns composed of two synapses, motivated by results that such patterns are over-represented compared to simple random-network models [91, 151]. Recent studies have also shown that the recurrent connectivity between cortical neurons depends on their tuning preferences. In mouse primary visual cortex, neurons with similar stimulus preferences connect more frequently and with stronger synapses than neurons with dissimilar stimulus preferences. Similar trends hold with respect to reciprocal connectivity between pairs of similarly or dissimilarly tuned neurons [153, 92]. Additionally, synaptically connected neurons tend to receive more common inputs than would be expected by chance, suggesting a clustered architecture [152, 151]. While this enhanced recurrent connectivity between similarly tuned neurons is present even at eye opening, it is enhanced during development and especially by visual experience [198, 199]. This suggests long-term synaptic plasticity as a key mechanism for the organization of cortical circuits into Hebbian assemblies: groups of strongly recurrently connected neurons with shared stimulus preferences.

The mechanisms of synaptic plasticity remain an active area of investigation. The precise timing of pre- and post-synaptic spikes can be a crucial determinant of plasticity [36, 37]. STDP can allow the development of feedforward structures [202] allowing temporally precise [88] and tuned [203] responses and that can allow sequential activity [87, 124]. Such feedforward architecture stands in contrast to the strongly recurrent organization of Hebbian assemblies.

The overall firing rates of pre- and post-synaptic cells, as well as slow postsynaptic depolarization, play a large role in shaping measured STDP curves [105]. Recent theoretical studies have often focused on this rate-dependence, for example by reducing models of STDP to rate-based plasticity models (e.g. [132, 108, 129]). Such rate-dependent plasticity can allow the development of Hebbian assemblies [129, 115, 194]. Recent studies of assembly formation have relied on plasticity models that transition between strongly potentiation- and depression-dominated regimes depending on the pre- and post-synaptic firing rates [129, 115]. This rate-dependent plasticity relies on the integral of the STDP curve being large; $S \gg 0$, allowing the negligence of spike-time correlations.

Models that can produce such unbalanced STDP curves also, however, give rise to approximate balance between potentiation and depression for many pre- and postsynaptic firing rates (Figure 4.1, starred regions). Furthermore, spike trains in diverse cortical areas do exhibit covariable trial-by-trial fluctuations (noise correlations). These noise correlations covary with neurons' stimulus preferences (signal correlations) [204, 180, 205] and synaptically connected neurons have higher noise correlations [153, 92]. Furthermore, these correlations exhibit temporal structure on the millisecond scale of measured STDP rules. This raises the question of whether fast spike-time correlations could play a role in the learning of assembly structure.

Here, we show that such spike-time correlations can, in the absence of rate-based plasticity mechanisms, allow the formation of Hebbian assemblies in response to correlated external inputs to groups of neurons. Combining linear response theory for spike-train covariances in recurrent networks [136] with a slow-fast theory of STDP [109], we develop low-dimensional theories describing the evolution of the network structure. These reveal that training can promote strong connectivity, as well as strong *reciprocal* connectivity, within commonly-stimulated groups. We further show that after training and in the absence of any external input correlations, internally-generated spike time correlations reinforce learned architectures during spontaneous activity. This study reveals a potential role for precise spike-time correlations in the formation of Hebbian assemblies in response to correlated external inputs, as well as their active maintenance during spontaneous activity.

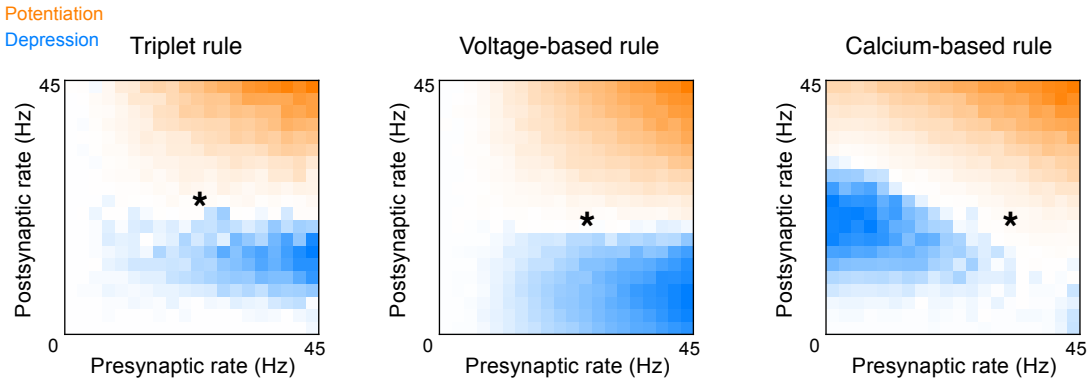

Figure 4.1: **Balance between potentiation and depression in different plasticity models.** Adapted from Litwin-Kumar & Doiron, [129]. Average change in synaptic weight as a function of pre- and post-synaptic firing rates, assuming the pre- and postsynaptic spike trains are uncorrelated Poisson processes. Left, the triplet rule of Pfister & Gerstner [106]. Middle, the voltage-based rule of Clopath & Gerstner [107]. Right, the calcium-based model of Graupner & Brunel [76].

4.2 RESULTS

4.2.1 Plasticity of partially symmetric networks during spontaneous activity

One striking feature of cortical networks is the over-representation of reciprocally connected pairs of excitatory neurons, compared to a simple randomly-wired (Erdös-Rényi) network [91, 151]. In order to reflect this structure we took the baseline excitatory-excitatory connectivity of our network, $\mathbf{W}^0_{EE}$, to be composed of two parts: $\mathbf{W}^0_{EE} = \mathbf{W}^0_{\text{sym}} + \mathbf{W}^0_{\text{asym}}$, where $\mathbf{W}^0_{\text{sym}}$ is a symmetric random binary matrix with connection probability Ωp_0 and $\mathbf{W}^0_{\text{asym}}$ a random binary matrix with connection probability $(1-\Omega)p_0$ (without any symmetry constraint). Both had Erdös-Rényi statistics. The parameter Ω thus determined the frequency of bidirectionally connected pairs of excitatory neurons in $\mathbf{W}^0_{EE}$. We modeled networks of 1500 excitatory neurons and 300 inhibitory neurons, both types following exponential integrate-and-fire dynamics [137]. The overall connection probability between excitatory neurons was $p_0 = 0.15$, with $\Omega = 0.4$. Excitatory-inhibitory, inhibitory-excitatory and inhibitory-inhibitory connectivity were asymmetric ($\Omega = 0$), with connection probability

0.4.

We begin by studying plasticity in these networks during spontaneous activity. In order to focus on learning due to precise spike-time correlations, we used a classical spike pair-based STDP rule [88]. In recurrent networks, excitatory plasticity can lead to the destabilization of asynchronous activity [206] and the development of pathological synchrony [127]. Recent studies have suggested plasticity of inhibition as a stabilizing mechanism [207], preventing runaway activity in networks with [129, 115] and without [177] excitatory plasticity. In order to prevent runaway excitation due to potentiation of excitatory synapses, inhibitory $\rightarrow$ excitatory synapses underwent homeostatic inhibitory plasticity (iSTDP): pairs of pre- and post-synaptic spikes cause potentiation of inhibitory-excitatory synapses, while individual presynaptic spikes cause depression, as in [177]. The strength of this depression is determined by the homeostatic target rate of the excitatory population, $\bar{r}_E$ (4.5.1). We will look at the dynamics of the excitatory connectivity before examining the role of the inhibitory plasticity.

We begin with a simple characterization of the network excitatory-excitatory structure in terms of two variables:

$$\begin{aligned} p &= \frac{1}{N_E^2} \sum_{i,j \in E} \mathbf{W}_{ij} \\ q &= \frac{1}{N_E^2} \sum_{i,j \in E} \mathbf{W}^0_{ij} \mathbf{W}_{ji} - p_0 p \end{aligned} \tag{4.1}$$

These characterize the mean weight of excitatory-excitatory synapses (p) and the mean weight of *reciprocal* excitatory-excitatory synapses (q) above what would be expected in an unstructured network. Here, q corresponds to $q_{\mathrm{X}}^{\mathrm{rec}}$ in Chapter 3. Note that with asymmetric connectivity, $\Omega = 0$, q becomes weak ($\mathcal{O}(N^{-3/2})$) so that the network connectivity can be described (to leading order) only by p. This corresponds to the unstructured invariant set of Chapter 3. The structure we impose on the network by setting $\Omega \neq 0$ enforces that the variables p, q form an invariant set for the full higher-order motif dynamics discussed in Chapter 3.

We derive dynamics for these variables following the same steps as we took in Chapter 3. We approximate the average spike train covariance from the contributions of length one paths in the network (Eq. (4.23)) and neglect the bounds on synaptic weights in the eSTPD rule, so that this only takes in to account the transient plasticity and not the steady-state

of the weights. The network structure p, q then obeys (4.5.3):

$$\frac{dp}{dt} = \left(\bar{r}_E^2 S + c_{EE}\sigma^2 S_\eta\right) p_0 + \epsilon \left[S_F p + S_B \left(q + p_0 p\right) + S_C p^2 + S_C^I \gamma (p_{EI}^*)^2\right] \tag{4.2}$$

$$\frac{dq}{dt} = \left(\bar{r}_E^2 S + c_{EE}\sigma^2 S_\eta\right) q_0 + \epsilon \left[S_F q + S_B \left(1 - p_0\right)\left(q + p_0 p\right) + S_C \frac{q_0}{p_0} p^2 + S_C^I \gamma \frac{q_0}{p_0} (p_{EI}^*)^2\right] \tag{4.3}$$

The first terms on the right-hand side of Eq. (4.2) describe the contributions of chance spike coincidences ($\bar{r}_E^2 S$) and correlations induced by external inputs ($c_{EE}\sigma^2 S_\eta$). S is the integral of the eSTDP rule, while S_η is the integral of the STDP rule against the average susceptibility of two neurons to externally-induced correlations (4.5.3). The latter terms describe the contribution of correlations induced by coupling within the network, weighted by the eSTDP rule. We use a Hebbian STDP rule with potentiation from pre-post spike pairs and depression from post-pre spike pairs (Figure 4.2B). The effect of correlations due to direct connections is measured by S_F, and those due to reciprocal connections are measured by S_B. The final terms arise from correlations due to common inputs from excitatory (S_C) or inhibitory (S_C^I) neurons.

We take there to be a balance between potentiation and depression, so that $S \sim \mathcal{O}(\epsilon)$, tilted slightly in favor of depression. In this case the dynamics is governed by different sources of spiking covariability, each interacting with the eSTDP rule $L(s)$. Spiking covariations from direct connections mainly contribute at positive time lags, interacting with the potentiation side of the eSTDP rule. This is reflected in the average spike train covariance between monosynaptically connected neurons. Reciprocal connections, in contrast, contribute spiking covariations at negative time lags, interacting with the depression side of the eSTDP rule. This is reflected in the average spike train covariance between reciprocally connected pairs, which includes the contributions from both direct and reciprocal connections. Finally, the contributions from common inputs are temporally symmetric around zero time lag, interacting with both the potentiation and depression windows (Figure 4.2F).

The competition between these sources of spiking covariability imposes thresholds for potentiation and depression of the mean-field variables p and q (Figure 4.2D). If either is initially stronger than that threshold, it will potentiate. Our theory provides a good

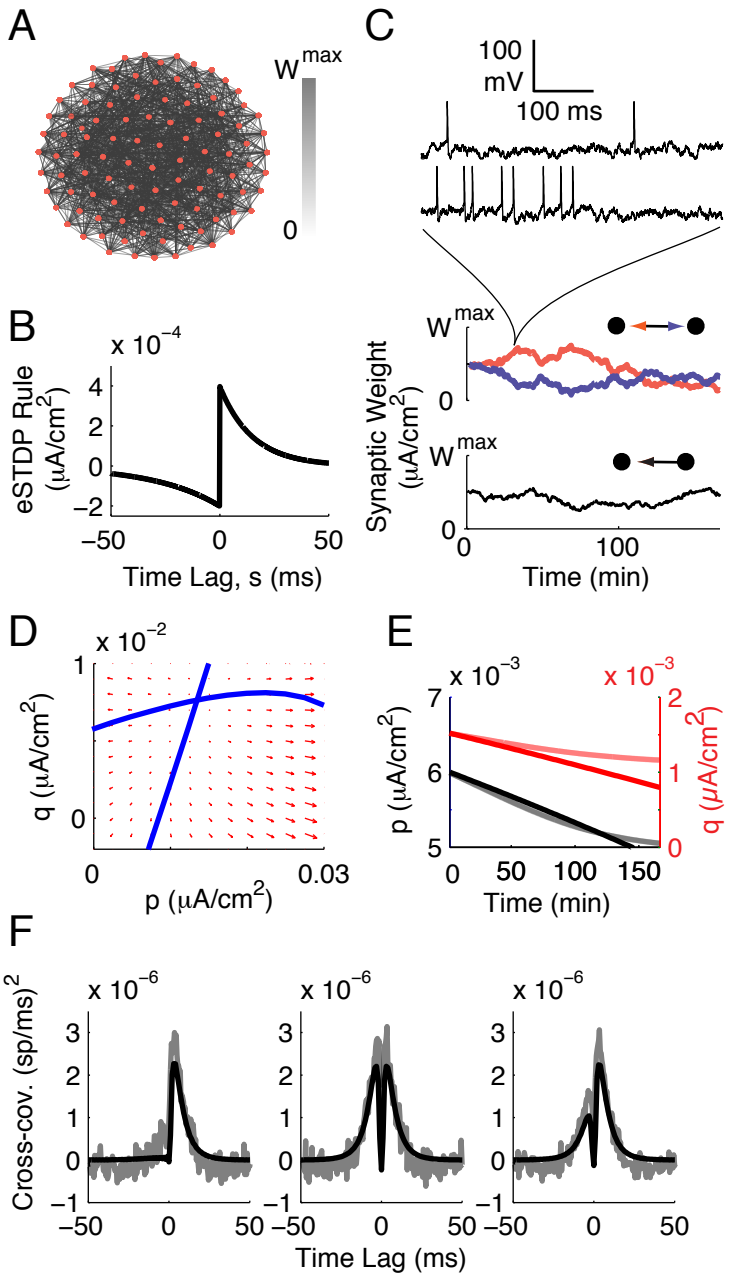

Figure 4.2: **Network structure shapes synaptic plasticity.** (A) Visualization of a random subset of the excitatory-excitatory connectivity. (B) The eSTDP rule, $L(s)$, is composed of exponential windows for depression (-) and potentiation (+). Each is defined by its amplitude $f_\pm$ and timescale $\tau_\pm$. (C) Synaptic weights evolve on a slow timescale. Individual synaptic weights are governed by the relative timing of spikes in the pre- and post-synaptic neurons' spike trains. (D) Dynamics of the mean synaptic weight (p) and the mean above-chance strength of *reciprocal* synapses, q. There is a threshold for potentiation of each given by its nullcline (blue lines). (E) Time course of p and q in the case where both are depressing. Solid lines: theory, Eqs. (4.2),(4.3). Shaded lines: simulation of the spiking network. (F) Average spike train covariance between monosynaptically connected pairs (left), reciprocally connected pairs (right) and all pairs (right). Shaded lines: simulation. Solid lines: linear response theory (first-order truncation, Eq. (4.23)).

prediction of the plasticity of these average levels of connectivity in the absence of external input correlations (Figure 4.2D,E). Before examining how external input correlations can train the network into different macroscopic structures, we first examine the role of inhibition

and inhibitory plasticity in this network.

4.2.2 Inhibition and homeostatic inhibitory STDP maintain stable activity

Inhibitory feedback recruited within the network reduces the spiking activity of excitatory neurons. With fixed inhibitory-excitatory projection strengths, potentiation of excitatory weights can lead to runaway activity if the recurrent excitation becomes strong enough to overcome the inhibitory feedback [127, 129]. In order to prevent this, inhibitory-excitatory synapses obeyed an inhibitory STDP (iSTDP) rule [177, 207, 176]:

$$\epsilon L_I(s) = \mathcal{H}(\mathbf{W}_{ij} - W^{\max,\mathrm{I}}) f_I e^{-\frac{|s|}{\tau_I}} \tag{4.4}$$

where f_I sets the maximum amplitude of individual potentiations and τ_I determines their dependence on relative spike timing (parameter values in Table 4.1). In addition, presynaptic (inhibitory) spikes cause depression by $\mathcal{H}(-\mathbf{W}_{ij})d_I$, with $d_I = -2f_I\bar{r}_E\tau_I$. We have assumed that the excitatory STDP rule is balanced between potentiation and depression, but made no such assumption about the inhibitory plasticity rule. Indeed, if the excitatory firing rates are far from the target rate $\bar{r}_E$ then the inhibitory plasticity becomes unbalanced and its leading-order dynamics is (Methods, 4.5.5):

$$\frac{dp_{EI}}{dt} = \left(r_I\left(r_E - \bar{r}_E\right)S^I + c_{EI}\sigma^2 S_\eta^{EI}\right) p_0^{EI} \tag{4.5}$$

Together with the dynamics of the firing rates r_E, r_I, these dynamics occur on a faster timescale than the balanced plasticity of excitatory connectivity. Examining the fixed points and stability of (p_{EI}, r_E, r_I) on this unbalanced timescale reveals that the inhibitory plasticity stabilizes the firing rates so that $r_E - \bar{r}_E \sim \mathcal{O}(\epsilon)$ (4.5.5). Indeed, in simulations we saw that as p increases (decreases), p_{EI} potentiates (depresses) and maintains $r_E = \bar{r}_E$ (e.g. Figure 4.3B). The location of the homeostatic inhibitory weight, p_{EI}^*, is given by solving the leading-order dynamics of the unbalanced inhibitory plasticity for $dp_{EI}/dt = 0, dr_E/dt = 0, dr_I/dt = 0$. Due to the separation of timescales between the homeostatic iSTDP and the balanced eSTDP, we could predict the location of the homeostatic inhibitory weight p_{EI}^* through a quasi-static approximation of p (4.5.5). We tracked the location of the homeostatic inhibitory weight

p^*_{EI} a function of p. As expected, strong recurrent excitation required stronger inhibitory-excitatory feedback to maintain the excitatory rates at $r_E = \bar{r}_E$ (Figure 4.3C). In order to investigate the conditions under which inhibition is able to maintain stable activity at that homeostatic fixed point, we compared the cases of plastic and non-plastic inhibition. With nonplastic inhibition, firing rates increased with p. If the excitatory feedback became strong enough to overcome the nonplastic recurrent inhibition, the stationary firing rates lost stability (Figure 4.3D). We predicted the location of that stability boundary by numerically computing the eigenvalues of the Fokker-Planck equation associated with the single-neuron voltage distribution and examining how activity is recurrently filtered through the network [201]. This instability was reflected in the development of hypersynchronous spiking (Figure 4.3E), in contrast to the asynchronous activity in the network with plastic inhibition. In total, we saw that recurrent inhibition stabilized stationary, asynchronous spiking in our networks, and iSTDP allowed the inhibitory-excitatory feedback to homeostatically maintain that stability.

4.2.3 Stimulus-induced noise correlations drive the formation of Hebbian assemblies

The thresholds for potentiation and depression suggests a mechanism for the formation of Hebbian assemblies through spike timing. Namely, if we define p and q variables for within- and cross-assembly connectivity, each should obey similar dynamics to Eqs. (4.2), (4.3). Particularly, each should have a threshold for potentiation and futhermore, these should depend on the spatial correlation of the external inputs within- or cross-cluster pairs of neurons. We begin by studying the simpler case of networks with asymmetric baseline connectivity ($\Omega = 0$). We divided the excitatory neurons into M putative assemblies, of κ neurons each, based on their stimulus preferences. Each assembly contained neurons that received spatially correlated synaptic inputs due to an external stimulus. For simplicity, we assume that the assemblies are symmetric so that the connectivity within and between

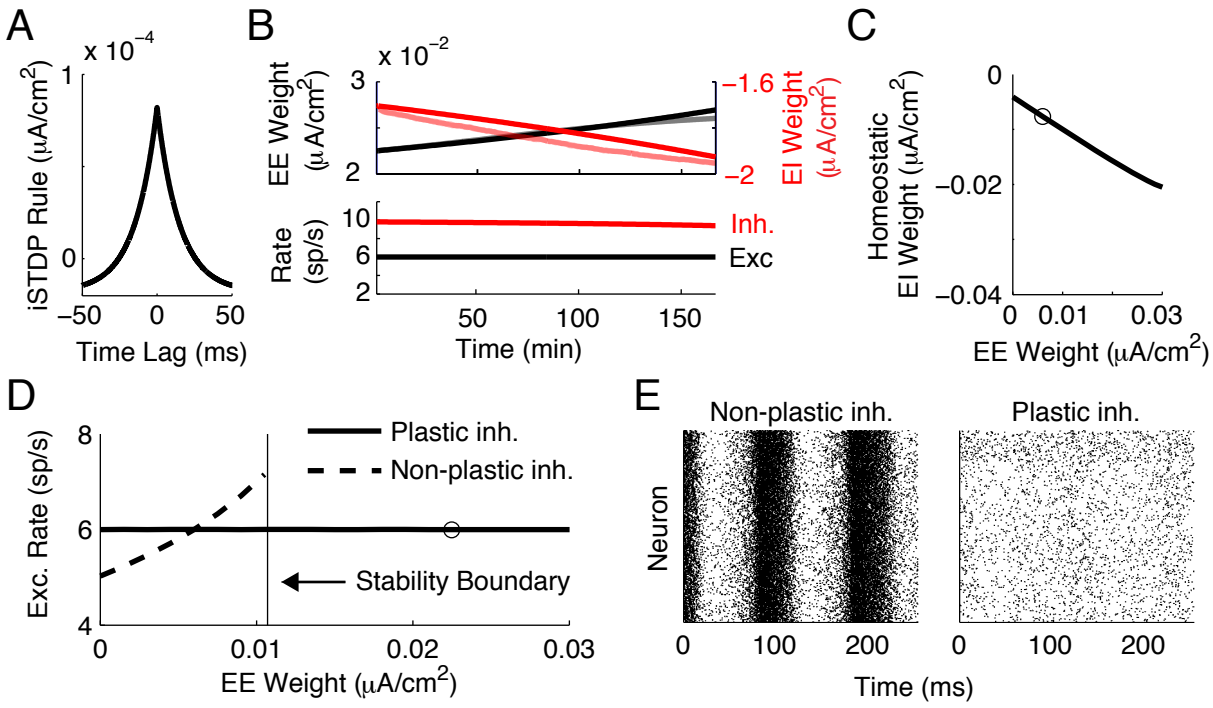

Figure 4.3: **Homeostatic inhibitory plasticity dynamically stabilizes firing rates.** (A) STDP rule for inhibitory-excitatory synapses. (B) Top: Co-evolution of mean excitatory-excitatory weight p (black) and mean inhibitory-excitatory p_{EI} (red). Bottom: Firing rates during plasticity. (C) The fixed point for p_{EI} as a function of the mean excitatory strength p. Open circle marks the inhibitory weight used for the non-plastic inhibition in later panels. (D) Firing rates as a function of excitatory weight in the cases of plastic and nonplastic inhibition. (E) Raster plots of the network activity. In both bases the excitatory weight is at the value marked by the circle in panel D. For the right raster, p_{EI} is at its homeostatic fixed point.

assemblies is characterized by:

$$\begin{aligned} \epsilon p_{AA} &= \frac{1}{\kappa^2} \sum_{i,j \in A} \mathbf{W}_{ij} \\ \epsilon p_{AB} &= \frac{1}{\kappa(N_E - \kappa)} \sum_{i \in A} \sum_{j \notin A} \mathbf{W}_{ij} \end{aligned} \tag{4.6}$$

where p_{AA} is the mean strength of connections within an assembly (across all assemblies), and p_{AB} is the mean strength of all cross-assembly connections. The inhibitory-excitatory, excitatory-inhibitory and inhibitory-inhibitory part of the network remained unstructured and asymmetric. Following the same steps as for p, we derived dynamical equations for the

mean within- and cross-cluster connectivity. The mean within- and cross-cluster synaptic weights obeyed (Methods, 4.5.6):

$$\frac{dp_{AA}}{dt} = \left(\bar{r}_E^2 S + c_{AA}\sigma^2 S_\eta\right) p_0 + \epsilon\left[S_F p_{AA} + S_B p_0 p_{AA} + S_C\left(p_{AA}^2 + (M-1)\, p_{AB}^2\right) + S_C^I \gamma (p_{EI}^*)^2\right] \tag{4.7}$$

$$\frac{dp_{AB}}{dt} = \left(\bar{r}_E^2 S + c_{AB}\sigma^2 S_\eta\right) p_0 + \epsilon\left[S_F p_{AB} + S_B p_0 p_{AB} + S_C\left(2 p_{AA} p_{AB} + (M-2)\, p_{AB}^2\right) + S_C^I \gamma (p_{EI}^*)^2\right] \tag{4.8}$$

Due to our approximation of the spike train covariances, Eq. (4.23), the dynamics of the mean synaptic weight within and across assemblies are coupled to each other only through correlations due to common inputs (the S_C terms). The nullclines of p_{AA}, q_{AB} predict their thresholds for potentiation.

The form of the dynamics suggests that the p_{AA} and p_{AB} nullclines could be simply computed by solving those quadratic polynomial equations. This procedure is complicated, however, by the fact that the homeostatic inhibitory weight depends on both p_{AA} and p_{AB}. In order to numerically compute (for example) the p_{AA} nullcline, we use a root-finding algorithm to find for each p_{AA} the p_{AB} which, in combination with the induced inhibitory weight p_{EI}^*, yields $dp_{AA}/dt = 0$. We can compute the nullclines explicitly by making a simplifying assumption: that the STDP rule is temporally symmetric ($\tau_- \sim \tau_+ + \mathcal{O}(\epsilon)$). This allows us to neglect the term containing the homeostatic inhibitory weight in Eqs. (4.7),(4.8). While this assumption is inaccurate for our STDP rule (which has $\tau_- = 2\tau_+$), its result reveals the main effect of external input correlations. In this case, the nullclines are given by (Methods, 4.5.7):

$$p_\alpha^* = -\frac{\left(\bar{r}_E^2 S + c_\alpha \sigma^2 S_\eta\right) p_0}{\epsilon\left(S_F + p_0 S_B\right)} \tag{4.9}$$

for $\alpha \in \{AA, AB\}$. Note that while the small parameter ϵ appears in the denominator, both terms of the numerator are also $\mathcal{O}(\epsilon)$ due to the balance between potentiation and depression in the eSTDP rule. In the absence of external input correlations ($c_\alpha = 0$), this nullcline is positive. This is because 1) we took $S < 0$ and 2) the eSTDP rule has a sharper potentiation window than depression window, so that $S_F + p_0 S_B > 0$. With $c_\alpha > 0$, the nullcline p_α^* is shifted to the left by an amount proportional to c_α. This prediction was borne

out by the actual nullclines (Figure 4.4D-F). In particular, $c_{AA} > 0$ reduces the threshold for potentiation of within-cluster connectivity.

We tested this prediction in simulations of the full system of spiking neurons, divided into $M = 3$ assemblies. External stimulation induced correlations in the spike trains of neurons in each assembly (Figure 4.4A). After 20 min of stimulation, we observed the formation of strongly connected assemblies of neurons (Figure 4.4H). The connectivity between assemblies was not potentiated; the assemblies did not fuse. We contrast this to the same network after 20 min of spontaneous activity: structure did not form spontaneously (Figure 4.4I). Our theory explains this by the initial symmetry in connectivity and spiking covariability between within- and cross-assembly pairs of neurons. The formation of assemblies and their separation is similarly explained by the stimulus-induced break in this symmetry.

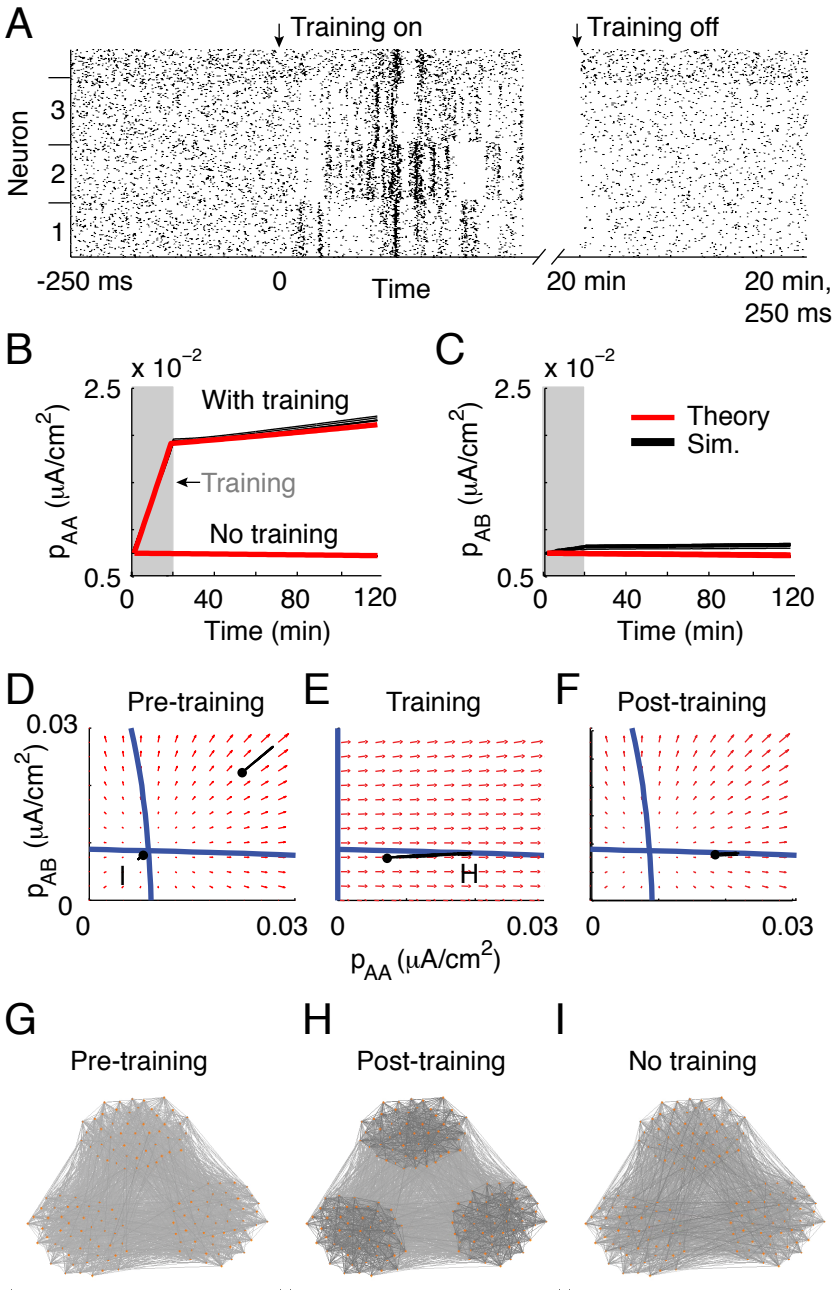

Figure 4.4: **Training and spontaneous reinforcement of assembly structure.** (A) Raster of the network activity. Excitatory eurons are ordered by assembly membership (numbers on the ordinate axis). (B,C) Dynamics of the mean strength of within-assembly synapses (B) and of cross-assembly synapses (C). Red, theory. Black, spiking simulation. Grey boxes mark the stimulus duration. (D-F) Phase planes governing the mean synaptic weights before the training signal (D), during the training period (E) and after training (F). (G-I) Visualization of a subset of the excitatory-excitatory connectivity. Nodes positioned by the Fruchterman-Reingold force algorithm.

4.2.4 Spontaneous spiking covariability after learning reinforces learned structure

In the absence of a training signal, assemblies did not spontaneously form. Our theory predicted, however, that a threshold for assembly formation did exist under spontaneous spiking conditions (Figure 4.4D). This suggested that if the assemblies potentiate sufficiently during the stimulus presentation to cross that spontaneous threshold, then internally-generated spiking covariances will reinforce the learned architecture after stimulus removal. Internally-generated spiking covariability after training was much larger within assemblies than between (Figure 4.5). We predicted that this asymmetry in spiking covariability, due to the learned network structure, would reinforce that structure after the end of training.

We tested this prediction by examining the evolution of (p_{AA}, p_{AB}) for 100 min after stimulus removal in our spiking simulations. The learned assembly structure was stable, and further reinforced by internally-generated spiking covariability after the removal of the training signal (Figure 4.4B,D,F). This reinforcement was also reflected by the further increase of spiking correlations between within-assembly neuron pairs (Figure 4.5). Internally generated correlations, because they reflected the trained structure of the network, further enforced that architecture.

4.2.5 Reciprocal excitatory connectivity is preferentially promoted between similarly-tuned neurons

In the previous section, we discussed how spatial correlations in external signals can promote the formation of Hebbian assemblies. We discussed this only at the level of mean synaptic weights, the simplest measure of connectivity between neuron pairs. Recent data have revealed another striking feature of pair-based connectivity: pairs of neurons with similar stimulus preferences have strong reciprocal connectivity [92]. We next examined whether plasticity driven by precisely correlated spike times could contribute to the development of such structure.

To that end, we consider networks with partially symmetric baseline connectivity ($\Omega = 0.4$). This reciprocal structure is reflected in the weight matrix $\mathbf{W}$. To measure it in a

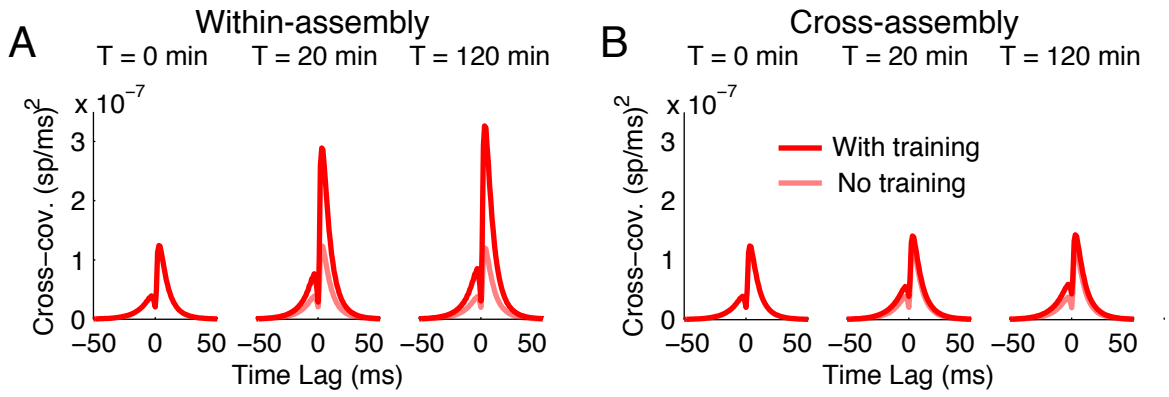

Figure 4.5: **Spike train covariability reflects and reinforces learned network structure.** (A) Average spike train cross-covariance between within-assembly pairs of neurons. (B) Average spike train cross-covariance between cross-assembly pairs of neurons. Cross-covariances estimated by the truncated linear response theory, Eq. (4.23). Solid: with training. Shaded: without training. Left: before training. Middle: end of stimulus presentation. Right: after spontaneous activity following stimulus presentation (as in Fig. 4.4).

way that allows us to take into account the development of stimulus-driven assemblies, we consider two new metrics of the network structure in addition to p_{AA} and p_{AB}:

$$\begin{aligned} q_{AA} &= \frac{1}{\kappa^2} \sum_{i,j \in A} \mathbf{W}^0_{ij} \mathbf{W}_{ji} - p_0 p_{AA} \\ q_{AB} &= \frac{1}{\kappa \left(N_E - \kappa \right)} \sum_{i \in A} \sum_{j \notin A} \mathbf{W}^0_{ij} \mathbf{W}_{ji} - p_0 p_{AB} \end{aligned} \tag{4.10}$$

These measure the average strength of reciprocal connections either within (q_{AA}) or between (q_{AB}) assemblies, above what would be expected by chance. As before, we assume symmetry between different assemblies. The inclusion of the mean reciprocal weights expands our description of the network structure to four dimensions: $(p_{AA}, p_{AB}, q_{AA}, q_{AB})$. Furthermore, the mean synaptic weights p_{AA} and p_{AB}, in addition to depending on each other, depend on q_{AA}, q_{AB} through the STDP-weighted covariances due to reciprocal connections (Methods, 4.5.6).

In order to obtain a simpler description, we consider the transformation:

$$\begin{aligned} p_\Delta &= p_{AA} - p_{AB} \\ q_\Delta &= q_{AA} - q_{AB}. \end{aligned} \tag{4.11}$$

These measure the relative strength of assembly structure in the network, at the levels of mean connection strength (p_Δ) and above-chance reciprocal connection strength (q_Δ). In order for a network to respect the structure observed in mouse V1 by Cossell et al. [92], it should have $p_\Delta > 0, q_\Delta > 0$.

The dynamics of (p_Δ, q_Δ) can be simply calculated from those of $(p_{AA}, p_{AB}, q_{AA}, q_{AB})$ and are:

$$\frac{dp_\Delta}{dt} = c_\Delta \sigma^2 S_\eta p_0 + \epsilon \left[S_F p_\Delta + S_B (q_\Delta + p_0 p_\Delta) + S_C p_\Delta^2 \right] \tag{4.12}$$

$$\frac{dq_\Delta}{dt} = c_\Delta \sigma^2 S_\eta q_0 + \epsilon \left[S_F q_\Delta + S_B \left(1 - p_0\right) \left(q_\Delta + p_0 p_\Delta\right) + S_C \frac{q_0}{p_0} p_\Delta^2 \right] \tag{4.13}$$

where $c_\Delta = c_{AA} - c_{AB}$. Notably, these are decoupled from the overall strengths of excitation and inhibition in the network. The homeostatic inhibitory STDP, for example, is still in effect and maintains excitatory firing rates at $\bar{r}_E$. Common inputs from inhibitory neurons, however, contribute equally to the spike train covariance of within- and cross-assembly excitatory neuron pairs. Because of that symmetry, they do not affect the differential plasticity of assembly structure, p_Δ or q_Δ. Similarly, the contribution of chance spike coincidences, $\bar{r}_E^2 S$, cancels out because neurons in each assembly have the same average firing rate.

Similarly to the case of asymmetric networks in the previous section, these dynamics admit nullclines that represent thresholds for potentiation/depression. With $c_\Delta = 0$, these are given by:

$$\begin{aligned} p_\Delta^* &= \frac{-(S_F + p_0 S_B) \pm \sqrt{(S_F + p_0 S_B)^2 - 4 S_C S_B q_\Delta}}{2 S_C} \\ q_\Delta^* &= -\frac{S_B (1 - p_0) p_0 p_\Delta + S_C \frac{q_0}{p_0} p_\Delta^2}{S_F + S_B (1 - p_0)} \end{aligned} \tag{4.14}$$

with an unstable fixed point at $(0, 0)$. (The other unstable point is at inaccessible values of (p_Δ, q_Δ).) The phase plane is divided into four regions, containing each potential combination of potentiation/depression of p_Δ and q_Δ (Figure 4.6C left). We take the network to be initially unstructured, so that before training $p_\Delta \approx q_\Delta \approx 0$. If external input correlations are higher for within-assembly pairs than cross-assembly pairs ($c_\Delta > 0$), this unstable point is shifted to negative (p_Δ, q_Δ). This pushes the network towards having assemblies of strongly reciprocally-connected neurons (Figure 4.6C).

This network structure is reflected by the magnitude of spike train covariances within and between assemblies. Due to the higher levels of reciprocal connectivity, the average spike train covariances at negative time lag are larger than for the $\Omega = 0$ networks discussed above (compare Figure 4.6D vs Figure 4.5). The training of assembly structure into the network leads to a greater than 2x increase in spike train covariability for within-assembly neurons compared to cross-assembly neurons (Figure 4.6D). In sum, these results suggest that spontaneously-generated spike-train correlations, in addition to providing a signature of learned network structure, can actively reinforce it. This holds true both at the level of average connection strengths and for reciprocal connectivity.

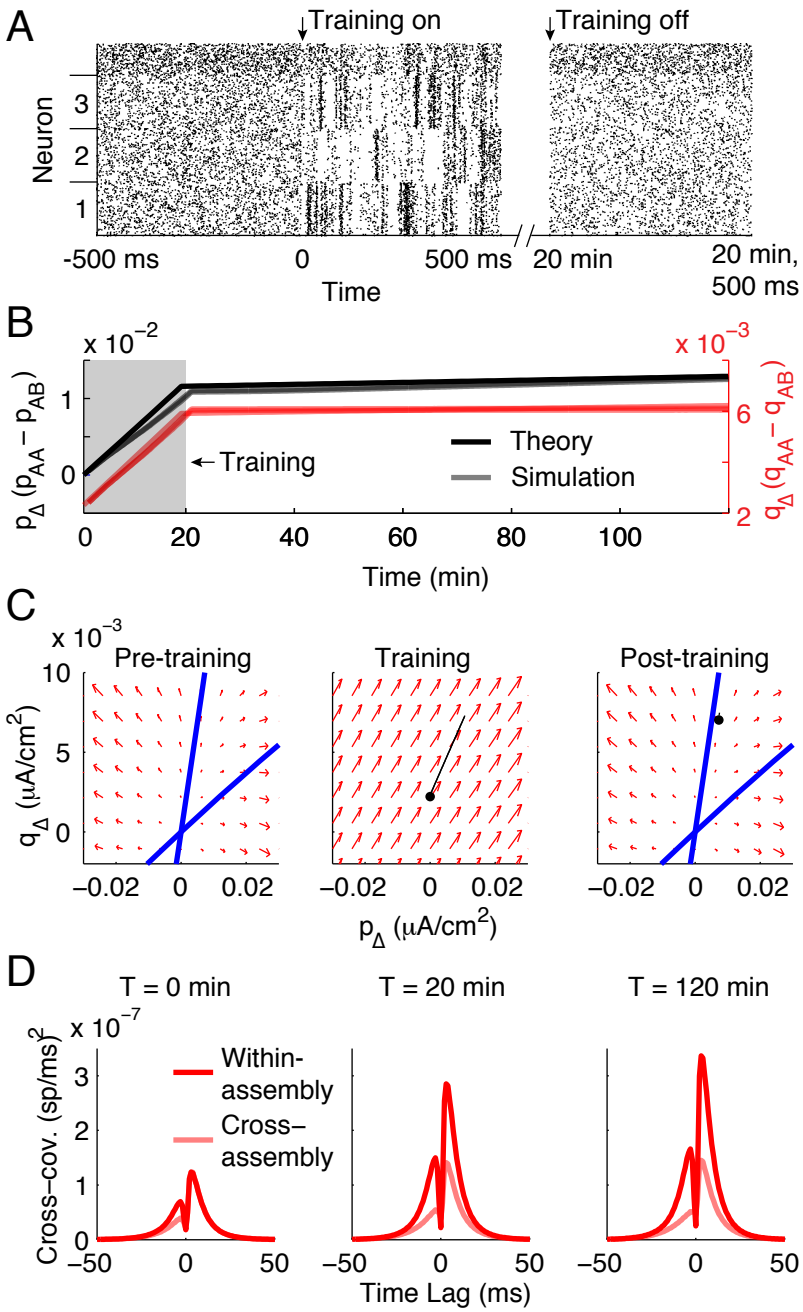

Figure 4.6: **Reciprocal connectivity is preferentially promoted within assemblies.** We consider networks with partially symmetric baseline connectivity. (A) Raster of network activity during pre-training, training and post-training phases. Excitatory neurons ordered by assembly membership (labeled on ordinate axis). (B) Time course of the relative strength of within-assembly synapses, p_Δ (black) and within-assembly reciprocal synapses q_Δ (red). Both have units of μA/cm^2. (C) Phase plane of p_Δ, q_Δ in the absence (pre- and post-training) or presence (training) of external input correlations. Blue: nullclines. Black traces: simulation of the spiking network. (D) Average spike train cross-covariances (truncated approximation, Eq. (4.23)). Left, before training. Middle, immediately at end of training. Right, 100 min post-training.

4.3 RELATION TO MOTIF DYNAMICS

In Chapter 3, we derived a 9-dimensional set of equations which described the dynamics of the mean synaptic weight and mean strengths of two-synapse motifs. When the network had Erdös-Rènyi connectivity and unstructured weights, the mean synaptic weight (p) was, to leading order, an invariant set of the dynamics. That invariant set could, however, be unstable - for many initial conditions, we saw that the two-synapse motifs, while not affecting the leading-order dynamics of p, increased so that eventually the system would move off of the invariant set of $(p, 0, 0, 0, 0, 0, 0, 0, 0)$.

In this chapter, we focus on the dynamics on an invariant set. We examine, however, networks with excitatory-excitatory connectivity that is composed of a symmetric part and an asymmetric part. The symmetric part gives rise to an over-representation of bidirectional connectivity: $q_0^{\text{rec}} \sim \mathcal{O}(1)$. Without special conditions on the weight matrix, this makes $q^{\text{rec}} \sim \mathcal{O}(1)$. So, rather than having an invariant set of only p, the invariant set of the dynamics is p, q^{rec}. Since the recurrent motif is the only one we discuss in this chapter, we use $q \equiv q^{\text{rec}}$. Note that if the symmetric part of the connectivity is sufficiently weak, the invariant set (p, q) becomes just p.

The main focus of this chapter is networks divided into Hebbian assemblies. This corresponds to block structure in the weight matrix: pairs of neurons with similar stimulus preferences are strongly connected. In order to describe this we separate the mean synaptic weight into the mean weight for synapses within an assembly (p_{AA}) and the mean weight for synapses between assemblies (p_{AB}). Similarly, the strength of reciprocal synapses is divided into q_{AA} and q_{AB}. Similar separations for the other two-synapse motifs lead to an explosion in the total number of motif variables describing the network structure.

In order to side-step this, we take the networks to have (partially symmetric) Erdös-Rènyi connectivity, conditioned on the adjacency matrix's block structure. Similar to in Chapter 3, this yields a lower-dimensional invariant set for the dynamics. Due to the partial symmetry of the network connectivity, that invariant set is (p, q) for networks without assembly structure, or $(p_{AA}, p_{AB}, q_{AA}, q_{AB})$ for networks with assembly structure.

4.4 DISCUSSION

4.4.1 Roles of noise correlations

Many theoretical studies have asked how the trial-by-trial fluctuations in neurons' stimulus responses -noise correlations- impact population coding [208, 209, 145]. The answer to this question depends on many factors. In particular, their impact on coding depends on how the structure of noise correlations in the population relates to neurons' stimulus preferences (e.g. [210, 211, 212, 213, 214]). Noise correlations are often related to neurons' stimulus tuning [215, 216, 217, 180, 218]. In the absence of sensory stimulation, patterns of activity across cortical populations are often similar to those observed during sensory stimulation [183, 182, 219, 220, 221, 222, 223]. Our work suggests that in addition to reflecting neurons' learned stimulus preferneces [216, 224], internally-generated spike train covariances can contribute to the active maintenance of learned structures.

4.4.2 Inhibitory plasticity and inhibitory stabilization

Inhibitory feedback plays two main roles in this study. The first is to modulate excitatory plasticity by contributing to spike-train covariability amongst excitatory neurons. The strength of this contribution is governed by the strength of the inhibitory feedback, which is in turn governed by inhibition's second role: homeostatic control of firing rates. Inhibition's role in stabilizing network activity in the face of strong recurrent excitation has been the focus of much recent work in theoretical neuroscience. Notably, strong inhibitory feedback provides dynamical explanations for the generation of variable activity in balanced networks ([225, 226, 172]) and can also account for paradoxical responses to external inhibitory inputs [227] and diverse features of tuning in visual cortex [228, 229].

In the absence of inhibition, potentiation of excitatory synapses in our networks leads to runaway excitation, so that the network exists in an inhibitory-stabilized regime (Figure 4.3). In contrast to other recent studies [129, 115], inhibitory STDP alone is sufficient to stabilize the network activity in our work without imposing synaptic scaling or other compensatory mechanisms. This is due to the relationship between the timescales of excitatory

and inhibitory plasticity. We take the excitatory plasticity to be balanced between potentiation and depression. This sets the dynamics of the mean excitatory weight p to occur on an $\mathcal{O}(\epsilon)$ timescale set by the eSTDP rule and the magnitude of spike-train correlations. When the firing rates are maintained at their stable fixed points, the inhibitory STDP is similarly governed by a timescale set by the iSTDP rule and the magnitude of spike-train correlations. If the firing rates are outside an $\mathcal{O}(\epsilon)$ neigborhood around their fixed point, this causes the iSTDP rule to become unbalanced, so that it is governed by an $\mathcal{O}(1)$ timescale (Methods, 4.5.5). This feature - that the inhibitory STDP can become unbalanced in order to maintain stable activity - guarantees that it can dynamically stabilize the network activity in the face of the balanced excitatory plasticity.

How neurons can undergo associative, Hebbian learning while maintaining stable activity has long been studied [230, 231]. While homeostasis is often thought of as a slower process than learning, recent work has highlighted the necessity of homeostatic mechanisms operating on a comparable timescale to excitatory plasticity [191]. A more specific example is yielded by the study of ocular dominance plasticity: the analysis of simple rate-based models reveal that homeostatic plasticity mechanisms are necessary for ocular dominance plasticity [232]. Homeostatic regulation acting alone, however, can paradoxically destabilize network activity, inducing oscillations in neurons' firing rates [233]. Homeostatic regulation mediated by diffusive transmitters like nitrous oxide can have different effects than that mediated by synaptic mechanisms [234]. The study of how homeostatic regulation and mechanisms for associative learning interact to allow stable memories and stable activity remains an exciting area of open inquiry.

4.5 METHODS

4.5.1 Plasticity models

Synapses between excitatory neurons undergo additive Hebbian STDP:

$$\epsilon L(s) = \begin{cases} \mathcal{H}(W^{\max} - \mathbf{W}_{ij}) f_+ e^{-\frac{|s|}{\tau_+}}, & \text{if } s \geq 0 \\ \mathcal{H}(\mathbf{W}_{ij})\,(-f_-)\,e^{-\frac{|s|}{\tau_-}}, & \text{if } s < 0, \end{cases} . \tag{4.15}$$

where $s = t_{post} - t_{pre}$ is the time lag between spikes. $f_\pm$ give the amplitude of individual changes in synaptic weights due to potentiation (f_+) or depression (f_-), and the time constants $\tau_\pm$ determine how synchronous spike pairs must be to cause plasticity. When $f_\pm \ll W^{\max}$, so that the timescale of plasticity is much longer than that of the STDP rule, individual weights undergo diffusion (Kempter et al., 1999) and their drift can be calculated as:

$$\frac{d\mathbf{W}_{ij}}{dt} = \mathbf{W}^0_{ij} \int_{-\infty}^{\infty} \epsilon L(s)\big(r_i r_j + \mathbf{C}_{ij}(s)\big) ds. \tag{4.16}$$

Here, r_i is the time-averaged firing rate of neuron i and $\mathbf{C}_{ij}(s)$ is the spike train cross-covariance function of neurons i and j. We will assume that the integral of $L(s)$ is small enough ($\mathcal{O}(\epsilon)$) so that firing rates do not dominate the plasticity.

The inhibitory STDP rule is

$$\epsilon L_I(s) = \mathcal{H}(\mathbf{W}_{ij} - W^{\max,\mathrm{I}}) f_I e^{-\frac{|s|}{\tau_I}} \tag{4.17}$$

In addition to this pair-based rule, each presynaptic (inhibitory) spike drives depression of the inhibitory synapses by $\mathcal{H}(-\mathbf{W}_{ij})d_I = -2f_I\bar{r}_E\tau_I$. This gives inhibitory $\rightarrow$ excitatory synapses a drift of

$$\frac{d\mathbf{W}_{ij}}{dt} = \mathbf{W}^0_{ij}\Big(\int_{-\infty}^{\infty} \epsilon L_I(s)\big(r_i r_j + \mathbf{C}_{ij}(s)\big) ds - 2f_I\tau_I\bar{r}_E r_j \Big) \tag{4.18}$$

4.5.2 Network model

We consider a network of N neurons, N_E of which are excitatory and divided into M clusters of size κ. There are $N_I = \gamma\kappa$ inhibitory neurons. Model parameters are in Table 4.1. We take the excitatory-excitatory block of the adjacency matrix $\mathbf{W}^0$ to be partially symmetric: $\mathbf{W}^0_{EE} = (p_0 - p_{0\text{sym}})\,\mathbf{W}^0_{\text{ER}} + p_{0\text{sym}}\mathbf{W}^0_{\text{sym}}$ where $\mathbf{W}^0_{\text{ER}}$ has (directed, i.e. non-symmetric) Erdös-Rényi statistics and $\mathbf{W}^0_{\text{sym}}$ is symmetric with Erdös-Rényi statistics (i.e. as in an undirected graph). Additionally we exclude autapses ($\mathbf{W}^0_{ii} = 0\ \forall i \in 1, \ldots, N$).

This means that the excitatory-excitatory connectivity is characterized by its empirical connection density p_0 and the frequency of loops q_0

$$\begin{aligned} p_0 &= \frac{1}{N_E^2}\sum_{i,j=1}^{N_E}\mathbf{W}^0_{ij} \\ q_0 &= \frac{1}{N_E^2}\sum_{i,j=1}^{N_E}\mathbf{W}^0_{ij}\mathbf{W}^0_{ji} - p_0^2 \end{aligned} \tag{4.19}$$

We assume that the statistics of the adjacency matrix for within- and between-assembly connectivity are the same (and equal to p_0 and q_0). The synaptic weight matrix, $\mathbf{W}$ is initially generated from $\mathbf{W}^0$ by giving each synapse the same initial weight. We consider the mean strength of E-E synapses within one cluster A and from other clusters into cluster A, p_{AA} and p_{AB} respectively:

$$\begin{aligned} \epsilon p_{AA} &= \frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}_{ij} \\ \epsilon p_{AB} &= \frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}_{ij} \end{aligned} \tag{4.20}$$

The small parameter $\epsilon = (\kappa p_0)^{-1}$ scales the synaptic weights. We take the mean strength of connections within each cluster to be symmetric, and the strength of connections into any one cluster from outside to be the same as into the others (so for all clusters A and B, $p_{AA} = p_{BB}$ and $p_{AB} = p_{BA}$). Similarly, we measure the strength of *reciprocal* connections within a cluster, q_{AA} or between clusters, q_{AB}:

$$\begin{aligned} \epsilon q_{AA} &= \frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}_{ij}\mathbf{W}^0_{ji} - \epsilon p_0 p_{AA} \\ \epsilon q_{AB} &= \frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}_{ij}\mathbf{W}^0_{ji} - \epsilon p_0 p_{AB} \end{aligned} \tag{4.21}$$

Table 4.1: **Model parameters**

Parameter	Description	Value
C	Membrane capacitance	1 μF/cm^2
g_L	Leak conductance	0.1mS/cm^2
V_L	Leak reversal potential	-72 mV
Δ	Action potential steepness	1.4 mV
V_T	Action potential initiation threshold	-48 mV
V_{th}	Action potential threshold	30 mV
V_{re}	Action potential reset	-72 mV
τ_{ref}	Action potential width	2 ms
μ	External input mean	1 μA/cm^2
σ	External input standard deviation	9 mV
$W^{\max,\mathrm{E}}$	Maximum synaptic weight	15ϵ μA/cm^2
$W^{\max,\mathrm{I}}$	Maximum synaptic weight	-7.5ϵ μA/cm^2
τ_{sE}	Excitatory synaptic time constant	2 ms
τ_{sI}	Inhibitory time constant	10 ms

By subtracting off $p_0 p_{AA}$ in the definition of q_{AA} (and likewise for q_{AB}), we measure the mean strength of reciprocal connections above what would be expected in a network with no correlations between synapses. Note: if the network is asymmetric ($\mathbf{W}^0_{\mathrm{sym}} = \mathbf{0}$) then q_0 is negligible ($\mathcal{O}(\epsilon^{-3/2})$) and so are the initial values of q_{AA} and q_{AB}.

We take the connectivity in between inhibitory and excitatory neurons, and within inhibitory neurons, to have (asymmetric) Erdös-Rényi statistics, so that these are characterized by their mean synaptic weights: p_{EI} for inhibitory $\to$ excitatory connections,

$$\epsilon p_{EI} = \frac{1}{N_E N_I} \sum_{i=1}^{N_E} \sum_{j=N_E+1}^{N} \mathbf{W}_{ij}, \tag{4.22}$$

and likewise p_{IE} and p_{II}.

4.5.3 Derivation of assembly dynamics

Here we will derive the dynamics of the assembly structure in networks of integrate-and-fire neurons undergoing STDP. We will begin by consider the dynamics of mean synaptic weights and mean reciprocal synaptic weights both within and between assemblies. The dynamics of (p, q) and (p_Δ, q_Δ) considered in the main text will then be recovered at the end. The derivation follows the same steps as the derivation of the motif dynamics in Chapter 3. We begin by expanding the covariance matrix $\mathbf{C}$ in path lengths through the network (Trousdale et al., 2012) and truncating at first order in the interactions to obtain:

$$\mathbf{C}_{ij}(s) \approx \overbrace{\left(\mathbf{A}_i * \mathbf{C}^\eta * \mathbf{A}_j\right)(s)}^{\text{external inputs}} + \overbrace{\left(\mathbf{W}_{ij}\mathbf{K}_{ij} * \mathbf{C}^0_{jj}\right)(s)}^{\text{forwards connections}} + \overbrace{\left(\mathbf{C}^0_{ii} * \mathbf{W}_{ji}\mathbf{K}^-_{ji}\right)(s)}^{\text{backwards connections}} + \underbrace{\sum_{k=1}^{N_E}\left(\mathbf{W}_{ik}\mathbf{K}_{ik} * \mathbf{C}^0_{kk} * \mathbf{W}_{jk}\mathbf{K}^-_{jk}\right)(s)}_{\text{common E inputs}} + \underbrace{\sum_{k=N_E+1}^{N}\left(\mathbf{W}_{ik}\mathbf{K}_{ik} * \mathbf{C}^0_{kk} * \mathbf{W}_{jk}\mathbf{K}^-_{jk}\right)(s)}_{\text{common I inputs}}. \tag{4.23}$$

As can be seen from Eqs. (4.16),(4.18), these cross-covariances will control plasticity through their integral against the STDP rule. We define variables measuring these STDP-weighted covariances (factoring out their amplitude, given by the $\mathbf{W}_{ij}$ factors in Eq. (4.23)):

$$\begin{aligned}
S &= \int_{-\infty}^{\infty} L(s)ds \\
S_\eta &= \int_{-\infty}^{\infty} L(s)\left(A_E(t) * A_E(-t)\right)ds \\
S_F &= \int_{-\infty}^{\infty} L(s)\left(K_{EE}(t) * C^0_E(s)\right)ds \\
S_B &= \int_{-\infty}^{\infty} L(s)\left(C^0_E(s) * K_{EE}(-t)\right)ds \\
S_C &= \int_{-\infty}^{\infty} L(s)\left(K_{EE}(t) * C^0_E(s) * K_{EE}(-t)\right)ds \\
S^I_C &= \int_{-\infty}^{\infty} L(s)\left(K_{EI}(t) * C^0_I(s) * K_{EI}(-t)\right)ds
\end{aligned} \tag{4.24}$$

and

$$\begin{aligned}
S^I &= \int_{-\infty}^{\infty} L_I(s)ds = 2f_I\tau_I \\
S_\eta^{EI} &= \int_{-\infty}^{\infty} L_I(s)\big(A_E(t) * A_I(-t)\big)ds \\
S_F^{EI} &= \int_{-\infty}^{\infty} L_I(s)\big(K_{EI}(t) * C_I^0(s)\big)ds \\
S_B^{EI} &= \int_{-\infty}^{\infty} L_I(s)\big(C_E^0(s) * K_{IE}(-t)\big)ds \\
S_C^{EIE} &= \int_{-\infty}^{\infty} L_I(s)\big(K_{EE}(t) * C_E^0(s) * K_{IE}(-t)\big)ds \\
S_C^{EII} &= \int_{-\infty}^{\infty} L_I(s)\big(K_{EI}(t) * C_I^0(s) * K_{II}(-t)\big)ds
\end{aligned} \tag{4.25}$$

In each of these definitions, $A_\alpha(t)$ corresponds to the mean linear response function of neurons of type α, $\alpha \in \{E, I\}$. $K_{\alpha\beta}(t)$ is the convolution of $A_\alpha(t)$ and the synaptic filter for synapses from β neurons to α neurons ($\alpha, \beta \in \{E, I\}$). We also define r_E and r_I, the average excitatory and inhibitory firing rates. Note that each of these are implicitly functions of the mean synaptic drive onto excitatory and inhibitory neurons. Note that for the iSTDP rule, each presynaptic spike causes depression by $-S^I\bar{r}_E$.

We want the dynamics of the connectivity variables $p_{AA}, p_{AB}, q_{AA}, q_{AB}$, so we differentiate these with respect to time. Then, inserting Eq. (4.23) into Eq. (4.16) and this into dp_{AA}/dt yields:

$$\begin{aligned}
\frac{dp_{AA}}{dt} &= \left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right)\frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}_{ij}^0 + S_F\frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}_{ij}^0\mathbf{W}_{ij} + S_B\frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}_{ij}^0\mathbf{W}_{ji} \\
&\quad + S_C\frac{1}{\kappa^2}\sum_{i,j\in A}\sum_{k=1}^{N_E}\mathbf{W}_{ij}^0\mathbf{W}_{ik}\mathbf{W}_{jk} + S_C^I\frac{1}{\kappa^2}\sum_{i,j\in A}\sum_{k=N_E+1}^{N}\mathbf{W}_{ij}^0\mathbf{W}_{ik}\mathbf{W}_{jk}
\end{aligned} \tag{4.26}$$

and similar for p_{AB}:

$$\begin{aligned}
\frac{dp_{AB}}{dt} &= \left(r_E^2 S + c_{AB}\sigma^2 S_\eta\right)\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}_{ij}^0 + S_F\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}_{ij}^0\mathbf{W}_{ij} \\
&\quad + S_B\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}_{ij}^0\mathbf{W}_{ji} + S_C\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\sum_{k=1}^{N_E}\mathbf{W}_{ij}^0\mathbf{W}_{ik}\mathbf{W}_{jk} \\
&\quad + S_C^I\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\sum_{k=N_E+1}^{N}\mathbf{W}_{ij}^0\mathbf{W}_{ik}\mathbf{W}_{jk}
\end{aligned} \tag{4.27}$$

The mean bidirectional connection strengths similarly evolve according to:

$$\frac{dq_{AA}}{dt} = \left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right)\frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}^0_{ij}\mathbf{W}^0_{ji} + S_F\frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}^0_{ij}\mathbf{W}_{ij}\mathbf{W}^0_{ji} + S_B\frac{1}{\kappa^2}\sum_{i,j\in A}\mathbf{W}^0_{ij}\mathbf{W}_{ji}\mathbf{W}^0_{ji} + S_C\frac{1}{\kappa^2}\sum_{i,j\in A}\sum_{k=1}^{N_E}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk}\mathbf{W}^0_{ji} + S_C^I\frac{1}{\kappa^2}\sum_{i,j\in A}\sum_{k=N_E+1}^{N}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk}\mathbf{W}^0_{ji} - p_0\frac{dp_{AA}}{dt} \tag{4.28}$$

$$\frac{dq_{AB}}{dt} = \left(r_E^2 S + c_{AB}\sigma^2 S_\eta\right)\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}^0_{ij}\mathbf{W}^0_{ji} + S_F\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}^0_{ij}\mathbf{W}_{ij}\mathbf{W}^0_{ji} + S_B\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\mathbf{W}^0_{ij}\mathbf{W}_{ji}\mathbf{W}^0_{ji} + S_C\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\sum_{k=1}^{N_E}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk}\mathbf{W}^0_{ji} + S_C^I\frac{1}{\kappa(N_E-\kappa)}\sum_{i\in A}\sum_{j\notin A}\sum_{k=N_E+1}^{N}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk}\mathbf{W}^0_{ji} - p_0\frac{dp_{AB}}{dt} \tag{4.29}$$

The mean inhibitory-to-excitatory synaptic weight obeys:

$$\begin{aligned}\frac{dp_{EI}}{dt} &= \frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\frac{d\mathbf{W}_{ij}}{dt}\\ &= \frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\mathbf{W}^0_{ij}\Big(\int_{-\infty}^{\infty}\epsilon L_I(s)\big(r_i r_j + \mathbf{C}_{ij}(s)\big)ds - 2f_I\tau_I\bar{r}_E r_I\Big)\end{aligned}$$

Inserting the first-order truncation of spike-train covariances yields:

$$\frac{dp_{EI}}{dt} = \left(r_I\big(r_E-\bar{r}_E\big)S^I + c_{EI}\sigma^2 S_\eta^{EI}\right)\frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\mathbf{W}^0_{ij} + S_F^{EI}\frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\mathbf{W}^0_{ij}\mathbf{W}_{ij} + S_B^{EI}\frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\mathbf{W}^0_{ij}\mathbf{W}_{ji} + S_C^{EIE}\frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\sum_{k=1}^{N_E}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk} + S_C^{EII}\frac{1}{N_E N_I}\sum_{i=1}^{N_E}\sum_{j=N_E+1}^{N}\sum_{k=N_E+1}^{N}\mathbf{W}^0_{ij}\mathbf{W}_{ik}\mathbf{W}_{jk} \tag{4.30}$$

The next step in writing down dynamics for each of the p and q variables of interest is to evaluate the sums over $\mathbf{W}$ and $\mathbf{W}^0$ in Eqs. (4.26)-(4.30). Recalling that the adjacency

matrix is Erdös-Rényi except for the partial symmetry of the excitatory-excitatory block, this yields (neglecting higher-order motif contributions):

$$\frac{dp_{AA}}{dt} = \left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right) p_0 + S_F \epsilon p_{AA} + S_B \epsilon \left(q_{AA} + p_0 p_{AA}\right) + S_C \epsilon^2 p_0 \left(\kappa p_{AA}^2 + \left(N_E - \kappa\right) p_{AB}^2\right) + S_C^I \epsilon^2 N_I p_0 p_{EI}^2 \tag{4.31}$$

$$\begin{aligned}\frac{dp_{AB}}{dt} = &\left(r_E^2 S + c_{AB}\sigma^2 S_\eta\right) p_0 + S_F \epsilon p_{AB} + S_B \epsilon (q_{AB} + p_0 p_{AB}) \\ &+ S_C \epsilon^2 p_0 \left(2\kappa p_{AA} p_{AB} + \left(N_E - 2\kappa\right) p_{AB}^2\right) + S_C^I \epsilon^2 N_I p_0 p_{EI}^2\end{aligned} \tag{4.32}$$

$$\begin{aligned}\frac{dq_{AA}}{dt} = &\left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right) q_0 + S_F \epsilon q_{AA} + S_B \epsilon (1 - p_0)(q_{AA} + p_0 p_{AA}) \\ &+ S_C \epsilon^2 q_0 \left(\kappa p_{AA}^2 + (N_E - \kappa) p_{AB}^2\right) + S_C^I \epsilon^2 N_I q_0 p_{EI}^2\end{aligned} \tag{4.33}$$

$$\begin{aligned}\frac{dq_{AB}}{dt} = &\left(r_E^2 S + c_{AB}\sigma^2 S_\eta\right) q_0 + S_F \epsilon q_{AB} + S_B \epsilon \left(1 - p_0\right) \left(q_{AB} + p_0 p_{AB}\right) \\ &+ S_C \epsilon^2 q_0 \left(2\kappa p_{AA} p_{AB} + \left(N_E - 2\kappa\right) p_{AB}^2\right) + S_C^I \epsilon^2 q_0 N_I p_{EI}^2\end{aligned} \tag{4.34}$$

$$\begin{aligned}\frac{dp_{EI}}{dt} = &\left(r_I \big(r_E - \bar{r}_E\big) S^I + c_{EI}\sigma^2 S_\eta^{EI}\right) p_0^{EI} + S_F^{EI} \epsilon p_{EI} + S_B^{EI} \epsilon p_0^{EI} p_{IE} \\ &+ S_C^{EIE} \epsilon^2 p_0^{EI} p_{IE} \left(\kappa p_{AA} + \left(N_E - \kappa\right) p_{AB}\right) + S_C^{EII} \epsilon^2 p_0^{EI} N_I p_{EI} p_{II}\end{aligned} \tag{4.35}$$

Finally, we recall that $\epsilon = (\kappa p_0)^{-1}$ and $N_I = \gamma\kappa$, revealing that the dynamics above stop at $\mathcal{O}(\epsilon)$:

$$\frac{dp_{AA}}{dt} = \left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right) p_0 + \epsilon \left[S_F p_{AA} + S_B \left(q_{AA} + p_0 p_{AA}\right) + S_C \left(p_{AA}^2 + \left(M - 1\right) p_{AB}^2\right) + S_C^I \gamma p_{EI}^2\right] \tag{4.36}$$

$$\begin{aligned}\frac{dp_{AB}}{dt} = &\left(r_E^2 S + c_{AB}\sigma^2 S_\eta\right) p_0 + \epsilon \big[S_F p_{AB} + S_B (q_{AB} + p_0 p_{AB}) + S_C \left(2 p_{AA} p_{AB} + \left(M - 2\right) p_{AB}^2\right) \\ &+ S_C^I \gamma p_{EI}^2\big]\end{aligned} \tag{4.37}$$

$$\begin{aligned}\frac{dq_{AA}}{dt} = &\left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right) q_0 + \epsilon \big[S_F q_{AA} + S_B (1 - p_0)(q_{AA} + p_0 p_{AA}) \\ &+ S_C \frac{1}{p_0} q_0 \left(p_{AA}^2 + (M - 1) p_{AB}^2\right) + S_C^I \frac{\gamma}{p_0} q_0 p_{EI}^2\big]\end{aligned} \tag{4.38}$$

$$\frac{dq_{AA}}{dt} = \left(r_E^2 S + c_{AA}\sigma^2 S_\eta\right) q_0 + \epsilon\big[S_F q_{AA} + S_B(1-p_0)(q_{AA}+p_0 p_{AA}) + S_C \frac{1}{p_0} q_0 \left(p_{AA}^2 + (M-1)p_{AB}^2\right) + S_C^I \frac{\gamma}{p_0} q_0 p_{EI}^2\big] \tag{4.39}$$

$$\frac{dq_{AB}}{dt} = \left(r_E^2 S + c_{AB}\sigma^2 S_\eta\right) q_0 + \epsilon\big[S_F q_{AB} + S_B\left(1-p_0\right)\left(q_{AB}+p_0 p_{AB}\right) + S_C \frac{1}{p_0} q_0 \left(2p_{AA}p_{AB} + \left(M-2\right)p_{AB}^2\right) + S_C^I \frac{\gamma}{p_0} q_0 p_{EI}^2\big] \tag{4.40}$$

$$\frac{dp_{EI}}{dt} = \left(r_I\big(r_E - \bar{r}_E\big) S^I + c_{EI}\sigma^2 S_\eta^{EI}\right) p_0^{EI} + \epsilon\big[S_F^{EI} p_{EI} + S_B^{EI} p_0^{EI} p_{IE} + S_C^{EIE} \frac{p_0^{EI}}{p_0} p_{IE}\left(p_{AA} + \left(M-1\right)p_{AB}\right) + S_C^{EII} \frac{p_0^{EI}}{p_0} \gamma p_{EI} p_{II}\big] \tag{4.41}$$

4.5.4 Firing rate dynamics

Here we have written the dynamics in terms of the average firing rates r_E, r_I and STDP-weighted spiking covariances as if those were parameters. As the mean excitatory and inhibitory weights change, so will neurons' firing rates. We now supplement the dynamics of the connectivity by examining the evolution of the population-averaged firing rates r_α, with $\alpha \in \{E, I\}$. The quasi-stationary firing rates obey:

$$r_\alpha(t) = f_\alpha(\mu_\alpha(t), \sigma^2) \tag{4.42}$$

where f_α is the rate-current function of an EIF neuron belonging to population α and

$$\begin{aligned}
\mu_E &= \mu_{\text{ext},E} + \epsilon N_E p \tau_E r_E + \epsilon N_I p_{EI} \tau_I r_I \\
&= \mu_{\text{ext},E} + \epsilon\left(\kappa p_{AA} + (N_E - \kappa) p_{AB}\right)\tau_E r_E + \epsilon N_I p_{EI}\tau_I r_I \\
&= \mu_{\text{ext},E} + \frac{1}{p_0}\left(p_{AA} + \left(M-1\right)p_{AB}\right)\tau_E r_E + \frac{\gamma}{p_0} p_{EI}\tau_I r_I \\
\mu_I &= \mu_{\text{ext},I} + \epsilon N_E p_{IE}\tau_E r_E + \epsilon N_I p_{II}\tau_I r_I \\
&= \mu_{\text{ext},I} + \frac{M}{p_0} p_{IE}\tau_E r_E + \frac{\gamma}{p_0} p_{II}\tau_I r_I
\end{aligned} \tag{4.43}$$

is the average external input to one those neurons and we assume that a sufficient combination of low firing rates and weak/slow synapses keeps recurrent connectivity from contributing significantly to the effective variance of inputs to a neuron. τ_E and τ_I are the integrals of excitatory and inhibitory synaptic kernels (these are described by single exponentials, so the integral is their decay time constant).

The dynamics of the quasi-stationary firing rates is then given by:

$$\frac{dr_\alpha}{dt} = \frac{df_\alpha}{d\mu_\alpha}\frac{d\mu_\alpha}{dt} \tag{4.44}$$

Recalling that $\left.\frac{df_\alpha}{d\mu}\right|_{\mu_\alpha} = \int_0^\infty A_\alpha(t)dt$, where $A_\alpha(t)$ is the average linear response of neurons of type α, we define

$$S_A^\alpha \equiv \int_0^\infty A_\alpha(t)dt \tag{4.45}$$

Assuming that $\mu_{\text{ext},\alpha}$ is constant in time, we obtain:

$$\frac{dr_E}{dt} = S_A^E\left(\frac{\tau_E}{p_0}\left(\left(p_{AA}+(M-1)\,p_{AB}\right)\frac{dr_E}{dt} + \left(\frac{dp_{AA}}{dt}+(M-1)\frac{dp_{AB}}{dt}\right)r_E\right) + \frac{\gamma\tau_I}{p_0}\left(\frac{dp_{EI}}{dt}r_I + p_{EI}\frac{dr_I}{dt}\right)\right) \tag{4.46}$$

and since the excitatory $\to$ inhibitory and inhibitory $\to$ inhibitory weights are not plastic, r_I tracks r_E:

$$\begin{aligned}\frac{dr_I}{dt} &= S_A^I\left(\frac{M}{p_0}p_{IE}\tau_E\frac{dr_E}{dt} + \frac{\gamma}{p_0}p_{II}\tau_I\frac{dr_I}{dt}\right)\\ &= \left(\frac{S_A^I\frac{M}{p_0}p_{IE}\tau_E}{1-\frac{\gamma}{p_0}p_{II}\tau_I}\right)\frac{dr_E}{dt}\end{aligned} \tag{4.47}$$

Inserting Eq. (4.47) into Eq. (4.46) then yields

$$\frac{dr_E}{dt} = \frac{S_A^E\left(\frac{\tau_E}{p_0}\left(\frac{dp_{AA}}{dt}+(M-1)\frac{dp_{AB}}{dt}\right)r_E + \frac{\gamma\tau_I}{p_0}\frac{dp_{EI}}{dt}r_I\right)}{\left(1-S_A^E\frac{\tau_E}{p_0}\left(p_{AA}+(M-1)\,p_{AB}\right) - S_A^E\frac{\gamma\tau_I}{p_0}p_{EI}\left(\frac{S_A^I\frac{M}{p_0}p_{IE}\tau_E}{1-\frac{\gamma}{p_0}p_{II}\tau_I}\right)\right)} \tag{4.48}$$

4.5.5 Stability of firing rates

The iSTDP rule imposes a form of rate homeostasis on the dynamics, keeping r_E within $\mathcal{O}(\epsilon)$ of $\bar{r}_E$. Indeed, this was one major motivation for its theoretical proposal (Sprekeler & Vogels et al, 2011). We now check how this affects the dynamics of the weights. If there is a balance between potentiation and depression in the eSTDP rule $L(s)$ so that $S \sim \mathcal{O}(\epsilon)$, then the dynamics of mean excitatory weights have an $\mathcal{O}(1/\epsilon)$ timescale. There is a different condition for balance between potentiation and depression of inhibitory $\rightarrow$ excitatory synapses. This balance occurs when the excitatory rate is close to $\bar{r}_E$, requiring $(r_E - \bar{r}_E) \sim \mathcal{O}(\epsilon)$. If the eSTDP rule is balanced but $(r_E - \bar{r}_E) \sim \mathcal{O}(1)$ then the leading order dynamics of the firing rates and p_{EI} become $\mathcal{O}(1)$ and obey Eq. (4.5):

$$\frac{dr_E}{dt} = \underbrace{\left(\frac{S_A^E \frac{\gamma \tau_I}{p_0}}{1 - S_A^E \frac{\tau_E}{p_0} \left(p_{AA} + (M-1)\, p_{AB} \right) - S_A^E \frac{\gamma \tau_I}{p_0} p_{EI} \left(\frac{S_A^I \frac{M}{p_0} p_{IE} \tau_E}{1 - \frac{\gamma}{p_0} p_{II} \tau_I} \right)} \right)}_{X(p_{EI})} r_I \frac{dp_{EI}}{dt} \tag{4.49}$$

$$\frac{dr_I}{dt} = \underbrace{\left(\frac{S_A^I \frac{M}{p_0} p_{IE} \tau_E}{1 - \frac{\gamma}{p_0} p_{II} \tau_I} \right)}_{Y} \frac{dr_E}{dt} \tag{4.50}$$

with fixed points (p_{EI}^*, r_I^*, r_E^*) obeying:

$$\begin{aligned} 0 &= \left(r_I^* \left(r_E^* - \bar{r}_E \right) S^I + c_{EI} \sigma^2 S_\eta^{EI} \right) \\ 0 &= X(p_{EI}^*) \cdot r_I^* \cdot \left(r_I^* \left(r_E^* - \bar{r}_E \right) S^I + c_{EI} \sigma^2 S_\eta^{EI} \right) \\ 0 &= Y \cdot X(p_{EI}^*) \cdot r_I^* \cdot \left(r_I^* \left(r_E^* - \bar{r}_E \right) S^I + c_{EI} \sigma^2 S_\eta^{EI} \right) \end{aligned} \tag{4.51}$$

In order for the first condition to hold ($dp_{EI}/dt = 0$), the fixed point rates must lie on the hyperbola given by

$$r_E^* = -\frac{c_{EI} \sigma^2 S_\eta^{EI}}{S^I} \left(\frac{1}{r_I^*} \right) + \bar{r}_E \tag{4.52}$$

This also satisfies $dr_E/dt = 0$ and $dr_I/dt = 0$. If $c_{EI} = 0$, this reduces to $r_E^* = \bar{r}_E$.

We next examine the linear stability of this solution. The Jacobian for Eqs. (4.5)-(4.50) is:

$$\begin{pmatrix} 0 & r_I S^I p_0^{EI} & (r_E - \bar{r}_E) S^I p_0^{EI} \\ r_I \left(r_I (r_E - \bar{r}_E) S^I + c_{EI} \sigma^2 S_\eta^{EI} \right) p_0^{EI} \frac{\partial X}{\partial p_{EI}} & X r_I^2 S^I p_0^{EI} & 2X p_0^{EI} (r_E - \bar{r}_E) S^I r_I + X c_{EI} \sigma^2 S_\eta^{EI} \\ Y r_I \left(r_I (r_E - \bar{r}_E) S^I + c_{EI} \sigma^2 S_\eta^{EI} \right) p_0^{EI} \frac{\partial X}{\partial p_{EI}} & Y X r_I^2 S^I p_0^{EI} & 2Y X p_0^{EI} (r_E - \bar{r}_E) S^I r_I + Y X c_{EI} \sigma^2 S_\eta^{EI} \end{pmatrix}$$

where

$$\frac{\partial X}{\partial p_{EI}} = \frac{\left(S_A^E \frac{\gamma \tau_I}{p_0} \right)^2 \left(\frac{S_A^I \frac{M}{p_0} p_{IE} \tau_E}{1 - \frac{\gamma}{p_0} p_{II} \tau_I} \right)}{\left(1 - S_A^E \frac{\tau_E}{p_0} \left(p_{AA} + (M-1) p_{AB} \right) - S_A^E \frac{\gamma \tau_I}{p_0} p_{EI} \left(\frac{S_A^I \frac{M}{p_0} p_{IE} \tau_E}{1 - \frac{\gamma}{p_0} p_{II} \tau_I} \right) \right)^2} \tag{4.53}$$

The eigenvalues of the Jacobian, evaluated at $p_{EI}^*, r_E^* = \bar{r}_E, r_I^*$ with $c_{EI} = 0$, are:

$$\begin{aligned} \lambda_1 &= \lambda_2 = 0, \\ \lambda_3 &= \frac{\gamma \left(r_I^* \right)^2 S_A^E S^I \tau_I}{1 - \frac{S_A^E \tau_E (p_{AA} + (M-1) p_{AB})}{p_0} - \frac{\gamma M p_{EI} p_{IE} S_A^E S_A^I \tau_E \tau_I}{p_0^2 - \gamma p_0 p_{II} \tau_I}} \end{aligned} \tag{4.54}$$

Below, we plot these eigenvalues (with $c_{EI} = 0$) as a function of the total excitation $p_{AA} + (M-1)p_{AB}$ with $c_{EI} = 0$ so that $r_E^* = \bar{r}_E$. For each $p_{AA} + (M-1)p_{AB}$, we use bisection to find $p_{EI}^* \in [0, \mathrm{W}_{\max}^I]$ that minimizes $|(r_E - \bar{r}_E)|$ (for the particular cellular and network parameters used). Fortunately, the inhibition is strong enough to achieve $r_E = \bar{r}_E$ - it would be possible for this not to be the case, for example with weak $\mathrm{W}_{\max}^I$.

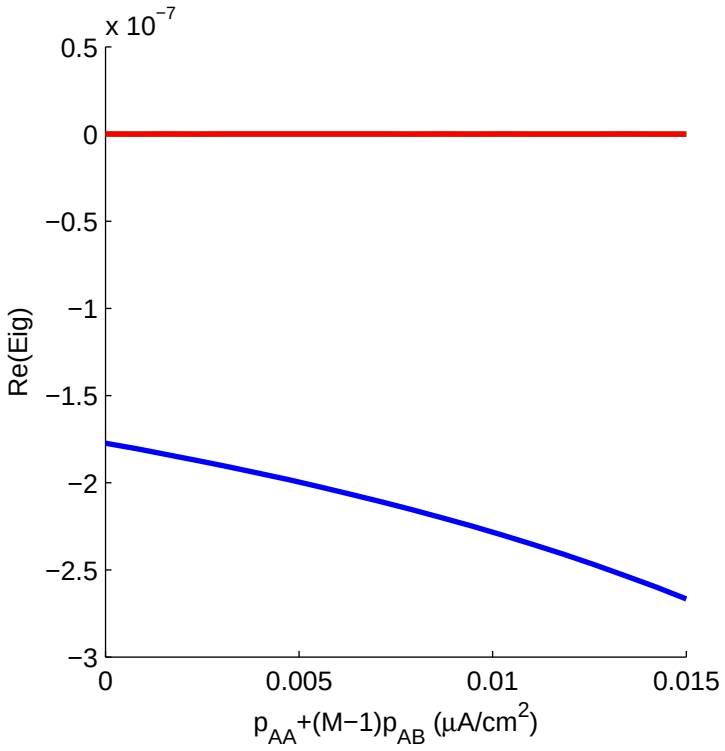

Figure 4.7: **Eigenvalues of the system Eqs.** (4.5)-(4.50) **with unbalanced iSTDP.**

4.5.6 Final dynamics of network structure: mind your p's and q's

The above analysis of unbalanced iSTDP reveals that there is a $\mathcal{O}(\epsilon)$ neighborhood around $p^*_{EI}, \bar{r}_E, r^*_I$ which is attracting along those dimensions, so that $r_E = \bar{r}_E + \mathcal{O}(\epsilon), r_I = r^*_I + \mathcal{O}(\epsilon), p_{EI} = p^*_{EI} + \mathcal{O}(\epsilon)$. (If $c_{EI} \neq 0$ then $\lambda_2 \neq 0$ and the dynamics could be different, a potential subject for future study.) Inserting these yields the following equations, up to $\mathcal{O}(\epsilon)$ and for balanced eSTDP (so $S \sim \mathcal{O}(\epsilon)$:

$$\frac{dp_{AA}}{dt} = \left(\bar{r}_E^2 S + c_{AA}\sigma^2 S_\eta\right) p_0 + \epsilon \left[S_F p_{AA} + S_B \left(q_{AA} + p_0 p_{AA}\right) + S_C \left(p_{AA}^2 + (M-1)\, p_{AB}^2\right) + S_C^I \gamma (p^*_{EI})^2\right] \tag{4.55}$$

$$\frac{dp_{AB}}{dt} = \left(\bar{r}_E^2 S + c_{AB}\sigma^2 S_\eta\right) p_0 + \epsilon\big[S_F p_{AB} + S_B(q_{AB} + p_0 p_{AB}) + S_C\left(2p_{AA}p_{AB} + (M-2)\,p_{AB}^2\right) + S_C^I \gamma (p_{EI}^*)^2\big] \tag{4.56}$$

$$\frac{dq_{AA}}{dt} = \left(\bar{r}_E^2 S + c_{AA}\sigma^2 S_\eta\right) q_0 + \epsilon\big[S_F q_{AA} + S_B(1-p_0)(q_{AA} + p_0 p_{AA}) + S_C \frac{1}{p_0} q_0 \left(p_{AA}^2 + (M-1)p_{AB}^2\right) + S_C^I \frac{\gamma}{p_0} q_0 (p_{EI}^*)^2\big] \tag{4.57}$$

$$\frac{dq_{AB}}{dt} = \left(\bar{r}_E^2 S + c_{AB}\sigma^2 S_\eta\right) q_0 + \epsilon\big[S_F q_{AB} + S_B\,(1-p_0)\,(q_{AB} + p_0 p_{AB}) + S_C \frac{1}{p_0} q_0 \left(2p_{AA}p_{AB} + (M-2)\,p_{AB}^2\right) + S_C^I \frac{\gamma}{p_0} q_0 (p_{EI}^*)^2\big] \tag{4.58}$$

Note that the location of $(\bar{r}_E, r_I^*, p_{EI}^*)$ depends on the net excitation, $p_{AA} + (M-1)p_{AB}$ and so will evolve on the slow timescale of the balanced eSTDP. We compute the nullclines of these equations in asymmetric networks by bisection. For example, for each p_{AA} we find the p_{AB} for which the homeostatic p_{EI}^* associated with (p_{AA}, p_{AB}) gives $dp_{AA}/dt = 0$.

4.5.7 Temporally symmetric eSTDP

When the timescales of potentiation and depression in the excitatory STDP rule are similar, $\tau_+ \sim \tau_- + \mathcal{O}(\epsilon)$ then the dynamics of the network structure simplify considerably. Since the correlations from common inputs (both from excitatory and inhibitory neurons) are temporally symmetric around 0 lag, this makes $S_C, S_C^I, S_C^{EIE}, S_C^{EII} \sim \mathcal{O}(\epsilon)$. The dynamics then reduce to:

$$\frac{dp_\alpha}{dt} = \left(\bar{r}_E^2 S + c_\alpha \sigma^2 S_\eta\right) p_0 + \epsilon\left[S_F p_\alpha + S_B\left(q_\alpha + p_0 p_\alpha\right)\right] \tag{4.59}$$

$$\frac{dq_\alpha}{dt} = \left(\bar{r}_E^2 S + c_\alpha \sigma^2 S_\eta\right) q_0 + \epsilon\big[S_F q_\alpha + S_B(1-p_0)(q_\alpha + p_0 p_\alpha)\big] \tag{4.60}$$

for $\alpha = AA$ or AB.

4.5.8 Separable dynamics of assembly formation and segregation

The dynamics of the network structure simplify if we take a linear transformation of our p and q variables:

$$\begin{aligned} p &= \frac{Mp_{AA} + M(M-1)p_{AB}}{M^2} \\ q &= \frac{Mq_{AA} + M(M-1)q_{AB}}{M^2} \\ p_\Delta &= p_{AA} - p_{AB} \\ q_\Delta &= q_{AA} - q_{AB} \end{aligned} \tag{4.61}$$

The first two, p, q measure the total mean synaptic weight and the mean weight of reciprocal connections overall in the network. The second two measure the formation of structure. The dynamics of these transformed variables are:

$$\frac{dp}{dt} = \left(\bar{r}_E^2 S + c_{EE}\sigma^2 S_\eta\right) p_0 + \epsilon\left[S_F p + S_B\left(q + p_0 p\right) + S_C M p^2 + S_C^I \gamma (p_{EI}^*)^2\right] \tag{4.62}$$

$$\frac{dq}{dt} = \left(\bar{r}_E^2 S + c_{EE}\sigma^2 S_\eta\right) q_0 + \epsilon\left[S_F q + S_B\left(1-p_0\right)\left(q + p_0 p\right) + S_C M \frac{q_0}{p_0} p^2 + S_C^I \gamma \frac{q_0}{p_0}(p_{EI}^*)^2\right] \tag{4.63}$$

$$\frac{dp_\Delta}{dt} = c_\Delta \sigma^2 S_\eta p_0 + \epsilon\left[S_F p_\Delta + S_B(q_\Delta + p_0 p_\Delta) + S_C p_\Delta^2\right] \tag{4.64}$$

$$\frac{dq_\Delta}{dt} = c_\Delta \sigma^2 S_\eta q_0 + \epsilon\left[S_F q_\Delta + S_B\left(1-p_0\right)\left(q_\Delta + p_0 p_\Delta\right) + S_C \frac{q_0}{p_0} p_\Delta^2\right] \tag{4.65}$$

where c_{EE} is defined, analogously to p, as the average correlation of external inputs and $c_\Delta = c_{AA} - c_{AB}$. Here we see that the spontaneous dynamics of overall potentation/depression (p, q) are separable from the dynamics of structure formation (p_Δ, q_Δ).

The nullclines are given by solving each equation for the steady-state, and are:

$$\begin{aligned} p^* &= \frac{-\epsilon(S_F + p_0 S_B) \pm \sqrt{\left(\epsilon(S_F + p_0 S_B)\right)^2 - 4\epsilon S_C M\left(\left(\bar{r}_E^2 S + c_{EE}\sigma^2 S_\eta\right) p_0 + \epsilon(S_B q^* + S_C^I \gamma (p_{EI}^*)^2)\right)}}{2\epsilon S_C M} \\ q^* &= -\frac{\left(\bar{r}_E^2 S + c_{EE}\sigma^2 S_\eta\right) q_0 + \epsilon\left(S_B(1-p_0)p_0 p^* + S_C M \frac{q_0}{p_0}(p^*)^2 + S_C^I \gamma \frac{q_0}{p_0}(p_{EI}^*)^2\right)}{\epsilon(S_F + (1-p_0)S_B)} \end{aligned} \tag{4.66}$$

$$
\begin{aligned}
p^*_\Delta &= \frac{-\epsilon(S_F + p_0 S_B) \pm \sqrt{\epsilon^2 \left(S_F + p_0 S_B\right)^2 - 4\epsilon S_C \left(c_\Delta \sigma^2 S_\eta p_0 + \epsilon S_B q^*_\Delta\right)}}{2\epsilon S_C} \\
q^*_\Delta &= -\frac{c_\Delta \sigma^2 S_\eta q_0 + \epsilon \left(S_B(1-p_0) p_0 p^*_\Delta + S_C \frac{q_0}{p_0} (p^*_\Delta)^2\right)}{\epsilon \left(S_F + S_B(1-p_0)\right)}
\end{aligned}
\tag{4.67}
$$

In the spontaneous case ($c_{EE} = c_\Delta = 0$) and defining $S = -\delta\epsilon$ these simplify to:

$$
\begin{aligned}
p^* &= \frac{-(S_F + p_0 S_B) \pm \sqrt{\left((S_F + p_0 S_B)\right)^2 - 4 S_C M \left(-\bar{r}_E^2 \delta p_0 + S_B q^* + S_C^I \gamma (p^*_{EI})^2\right)}}{2 S_C M} \\
q^* &= -\frac{-\bar{r}_E^2 \delta q_0 + S_B(1-p_0) p_0 p^* + S_C M \frac{q_0}{p_0} (p^*)^2 + S_C^I \gamma \frac{q_0}{p_0} (p^*_{EI})^2}{S_F + (1-p_0) S_B}
\end{aligned}
\tag{4.68}
$$

$$
\begin{aligned}
p^*_\Delta &= \frac{-(S_F + p_0 S_B) \pm \sqrt{\left(S_F + p_0 S_B\right)^2 - 4 S_C \left(S_B q^*_\Delta\right)}}{2 S_C} \\
q^*_\Delta &= -\frac{S_B(1-p_0) p_0 p_\Delta + S_C \frac{q_0}{p_0} (p^*_\Delta)^2}{S_F + (1-p_0) S_B}
\end{aligned}
\tag{4.69}
$$

5.0 CONCLUSIONS

5.1 SUMMARY

In this thesis I have presented the results of two studies on synaptic plasticity in recurrent networks. In Chapter 2, I developed a self-consistent framework for the coevolution of network structure and joint spiking statistics. In Chapter 3, I leveraged this to develop theories directly describing the plasticity of two-synapse motifs during spontaneous activity. This chapter used networks of solely excitatory neurons, necessitating weak coupling in order to prevent runaway activity. In Chapter 4, I used similar techniques to examine the formation of Hebbian assemblies due to external input correlations, and their active reinforcement by spontaneous activity. This chapter also examined the effect of inhibitory recurrence in affecting excitatory neurons' plasticity and the role of inhibitory plasticity in dynamically stabilizing the network's activity.

The two largest simplifications we have relied on have been 1) considering STDP only due to spike pairs and 2) only considering the likelihood of spike pairs as controlled by direct interactions between neurons. This second approximation, despite neglecting spike-train correlations due to indirect paths through the network, yielded surprisingly accurate predictions for the results of the full spiking simulations. We expect that this accuracy would degrade as we increased the upper bound for synaptic weights, even while remaining within the regime of stable activity [235]. More accurate predictions for the spike train correlations could be obtained by truncating the spike-train correlations, Eq. (C.4) at $n >$ 1st order in the connectivity. This accounts for the correlations to due paths up to length n through the network. For example, truncating at length 2 paths would also account for the spike-train correlations between neurons i and j due to any paths $j \to k \to i$, which

we neglected. An even greater improvement can be obtained be taking into account all higher-order patterns of connectivity, estimating them by how they are composed of up to length n paths [162, 110]. (Mathematically, this corresponds to assuming that all higher-than-nth-order cumulants of the network structure are negligible. The studies of this thesis, on the other hand, neglected all 2nd- and higher-order *moments* of the connectivity in estimating spike-train correlations.) This simple approximation yielded accurate predictions of plasticity in the full spiking networks. Improved estimation of the spiking correlations that drive plasticity could uncover new dynamics in parameter regimes we did not study.

5.2 BEYOND-PAIRWISE SPIKE CORRELATIONS IN PLASTICITY

While there is of course room for future improvement in this thesis' theory of spike pair-based plasticity, a more significant limitation of this work as it applies to actual plasticity in the nervous system is that we only consider how pairs of pre- and postsynaptic spikes contribute to plasticity. There are a number of experimental results that cannot be accounted for by pair-based plasticity, in particular its dependence on the presentation frequency and larger spike patterns [105, 39, 38, 41, 40]. A general theory of synaptic plasticity can be obtained by considering the weight changes induced by pre- and postsynaptic spikes as unknown functionals of the two spike trains [236, 237, 238]:

$$\frac{d\mathbf{W}_{ij}}{dt} = H_0 + y_i(t)F\big(y_i(t), y_j(t)\big) + y_j(t)G\big(y_i(t), y_j(t)\big) \tag{5.1}$$

where $y_i(t) = \sum_{k:t^k<t} \delta(t - t_i^k)$ is the spike history of neuron i up to time t. The functionals F and G model the mapping from pre- and postsynaptic spikes, through calcium signaling or other biophysical mechanisms, into effective weight changes. As long these mappings are

time-invariant, they can each be expanded in a Volterra series:

$$\begin{aligned} F(y_i(t), y_j(t)) =& F_1 + \int_0^\infty F_2^{ii}(s) y_i(t-s) ds + \int_0^\infty F_2^{ij}(s) y_j(t-s) ds \\ &+ \int_0^\infty \int_0^\infty F_3^{iii}(s, s') y_i(t-s) y_i(t-s') ds' ds \\ &+ \int_0^\infty \int_0^\infty F_3^{iij}(s, s') y_i(t-s) y_j(t-s') ds' ds \\ &+ \int_0^\infty \int_0^\infty F_3^{ijj}(s, s') y_j(t-s) y_j(t-s') ds' ds + \cdots \end{aligned} \tag{5.2}$$

where each term in the expansion of F accounts for the effect of spike pairs, triplets, etc. that include a postsynaptic spike. The functional accounting for weight changes triggered by presynaptic spikes can be similarly expanded in a Volterra series where each term accounts for the effect of spike pairs, triplets, etc. that include a presynaptic spike. The first term F_1 accounts for weight changes driven by individual postsynaptic spikes. The superscripts in subsequent terms denote the combination of pre- (j) and postsynaptic (i) spikes accounted for by that term. Specific phenomenological models of plasticity can be produced by truncating these expansions at n-th order spike interactions and specifying the necessary functions relating action potentials to weight changes. The pair-based STDP rule have used, for example, corresponds to truncating F and G at pairwise spike interactions, taking $F_1 = G_1 = 0$ and $F_2^{ii} = G_2^{jj} = 0$ and F_2^{ij} and G_2^{ji} to be exponential functions of the pre-post time lag.

After truncating at a certain order, n, of spike interactions and specifying the form of the plasticity model, the crucial last step in developing a theory for synaptic plasticity is to compute the n-th order spike train correlations. For rate-based plasticity models such as those discussed in the Introduction, these are simply products of the firing rates. In order to account for spike-driven plasticity, spike-train correlations must be computed. In feedforward networks this is comparatively simple. One approach towards computing these in recurrent networks is to consider the neurons' spike trains to be linearly-interacting Poisson processes (Hawkes processes), in which case the second-order correlations can be explicitly computed [239, 160]. Using a conductance-based model complicates the transfer of synaptic inputs to spiking outputs, and thus the calculation of spike-train correlations. The approach we have used is to take recurrently connected conductance-based neurons receiving external Poisson

inputs. This allows a diffusion approximation for the external inputs and the application of Fokker-Planck techniques to calculate the dynamical transfer of individual neurons (Appendix B), and then the use of linear response techniques to calculate spike-train correlations [136]. In related work, Ermentrout & colleagues have examined how neurons with fundamentally oscillatory spiking activity can synchronize in the presence of correlated external noise [240, 241, 242, 243, 244].

While these techniques allow the development of plasticity theories for conductance-based recurrent networks, the requirement of Poisson external inputs restricts the approach's generality. Spike trains in many brain regions exhibit non-Poisson variability [245]. The external Poisson requirement also limits the description of internally-generated spike time variability, a striking feature of networks with balanced excitation and inhibition [225, 226, 173, 235]. Recent work has adapted techniques from the kinetic theory of plasmas to allow the calculation of arbitrary-order spike-train correlations in networks of either binary- or phase-neuron models [246, 247, 248, 111, 249]. The usage of these techniques to describe conductance-based neurons relies on mapping a conductance-based model onto, usually, a phase-neuron model [249, 111]. For example, the quadratic integrate-and-fire model can be mapped simply onto a widely-popular phase model, the theta neuron (or Ermentrout-Kopell canonical model) [250, 251]. These powerful techniques for calculating higher-order spike-train correlations provide an exciting path towards theories of plasticity that self-consistently describe both rate- and timing-based effects control the evolution of network structure. Another, possibly simpler approach, would be to estimate triplet- or higher-order spike train correlations by examining how they can be composed of spike pairs, similar to the approach taken by [162] for correlations in the network structure. This type of approach would rely on whether higher-order joint cumulants of neurons' spike trains are negligible, a question which remains to be tested.

5.3

5.4 MEMORY IN RECURRENT NETWORKS

Spontaneous activity is usually viewed as a problem for plasticity: learned weight changes must be stable in the face of spontaneous activity. Some previous studies have addressed this issue by endowing individual synapses with dynamical bistability between weak and strong weights [252, 76, 253, 115]. These models potentially provide a rich dynamical explanation of the observation that individual CA3-CA1 synapses tend to occupy either strong or weak states and switch between them, rather than having a continuum of weights [254, 255]. In contrast the STDP rule we have used here, as with other recently proposed models [256, 257], relies on imposed upper and lower bounds for the synaptic weight to prevent synaptic weights from potentiating or depressing to unphysiological values.

The plasticity model we have used here is motivated by experiments demonstrating changes in synaptic efficacy lasting for tens of minutes, mediated by post-translational synaptic modifications. Long-term potentiation (and possibly also depression), however, can last for months or longer[258]. This dichotomy inspired experimental studies on the changes in gene expression and protein synthesis mediating long-term synaptic changes. Models of long-term plasticity based on synaptic tagging (by pre- and postsynaptic activity) and capture (of plasticity-mediating proteins, synthesized postsynaptically, in response to overall stimulation levels) can explain a range of experimental results, exhibiting both early and late-phase potentiation as well as depotentiation [259, 260], as well as linking depotentiation protocols to behavioral memory [261]. Complementary to these studies, our results suggest that in a recurrent network the patterns of activity that drive synaptic tagging could, after learning, be re-generated intrinsically by the network. Indeed, a number of past studies have shown that spontaneously patterns of activity resemble stimulus-evoked activity [183, 182, 219, 220, 221, 222, 223]. Previous work has shown that in rate-based plasticity, spontaneously-generated rate fluctuations due to finite-size heterogeneity in the network structure can reflect trained structures, actively reinforcing them [129]. In Chapter 4, we show that this can occur with timing-based plasticity even despite the fact that our

spontaneously generated patterns are fast-timescale spike-train correlations which are weak compared to the noisy background activity.

We did not examine how presentation of further stimuli after a training period would shape the network - whether the network would be at risk of unlearning, or able to reconfigure in response to new stimuli. Our theory does provide predictions for these. Spontaneous unlearning of trained network structure will occur if the training protocol is insufficient to bring the network structure across the spontaneous thresholds for promotion of assembly structure (Figure 4.4). After training of one network structure, the networks studied here would response to new training protocols in a way that depends on the structure of the correlations those protocols induce in neurons' external inputs. For example, take a network trained into an assembly configuration, as in (Figure 4.4). The subsequent presentation of anti-assembly input correlations (i.e. $c_{AB} > 0$) would decrease the threshold for potentiation of cross-assembly connectivity, pushing the network into a fused configuration without preferential connectivity within or between assemblies (large p_{AA} and large p_{AB}). Even if the network has previously reached a steady state $d\mathbf{W}/dt = 0$, if subsequent stimulation pushes weights off of their equilibrium manifold then the low-dimensional theories developed here should describe the evolution of the network's structure. The results of Chapter 4 also extend naturally to assemblies with overlapping membership, as long as the assumption of symmetry between assemblies is respected.

In this thesis we investigated the formation of certain types of structure due to spike timing. We defined structures by measure the average levels of different connectivity motifs (Chap 3, or of connectivity within or between groups of neurons defined by their feedforward stimulus preferences 4. A complementary line of investigation, which has received much theoretical attention, is the question of how many different patterns of connectivity can be maintained in a network despite ongoing activity - what is the memory capacity of a network? And, how does it depend on the plasticity of the network's constituent synapses? These questions have been part of theoretical neuroscience since at least the early work of Rosenblatt on the perceptron, a one-layer feedforward system which could learn to classify linearly separable inputs [262]. The perceptron learns via a simple learning rule which linearly modifies the input weights to match the inputs. Hopfield's seminal study, examining

memories encoded in the activity of recurrently connected neurons, brought the topic of memory in recurrent networks into prime view [263, 264]. Early work on memory storage in networks with Hebbian-inspired plasticity, dependent on the covariance of neurons' activities, showed memory capacities that scaled linearly with the network size, N [86, 265, 266, 267]. These models relied, however, on synapses being able to potentiate or depress unboundedly. Their enormous capacity came, furthermore, with a price: the existence of a critical number of memories that could be stored, past which all would be forgotten.

A fundamental problem of memory is how to, with a finite capacity, maintain memories in the face of ongoing activity. When there are bounds on the possible synaptic weights, simple synaptic models characterized by binary switches between weak or strong states exhibit a memory capacity of at most $\mathcal{O}(\sqrt{N})$, achieved when training stimuli are presented rapidly compared to the timescale of synaptic switching. On the other hand if training stimuli arrive slowly, the memory capacity is as low as $\mathcal{O}(\ln N)$ [268, 269, 270, 271]. These networks are palimpsests: rather than exhibiting critical forgetting, old memories become rewritten by new ones [272, 273]. Because of this, a natural measure for their memory performance is how long a memory lasts in the face of ongoing stimulation. One measure of this is the memory lifetime: how long a randomly selected memory will last. This is typically estimated as a signal-to-noise ratio, measuring the strength of the stored memory compared to the ongoing fluctuations in synaptic weights.

There have been two main approaches to estimating the memory performance populations in the face of ongoing activity: 1) to examine the dynamics of the constituent synapses, and 2) to examine the distribution of memories amongst synapses. Enriching the description of individual synapses to multi-state cascades, with each state characterized by different levels of plasticity, can increase the performance [274]. Interestingly, when hard bounds are imposed on the synaptic weights, the memory lifetime depends crucially on the balance between potentiation and depression. If they are unbalanced, then the memory lifetime tends to a constant set by the rates of potentiating and depressing events. If they are balanced, however, the lifetime of individual memories is $\mathcal{O}(\alpha^{-2})$, where α is the size of individual potentiations or depressions [275]. While this work used a different synaptic model than ours, the notion of a (possibly imperfect) balance between potentiation and depression cor-

responds exactly. Plasticity with dynamics that depend on the current value of the synaptic weight can also improve memory lifetimes, memory lifetimes scale as $\mathcal{O}(\alpha^{-1})$ independently of whether potentiation and depression are balanced or not [275]. The interaction of short-term facilitation or depression could also be studied, pending the development of linear response theory for spike-train correlations with short-term plasticity. How the stochastic dynamics of synapses determines their memory capacity remains an open area of inquiry [276].

Examining the distribution of memories amongst synapses typically involves a statement about the coding fraction f: the number of neurons involved in individual memories. Sparse coding (f goes to zero as $N \to \infty$) reduces the interference between old and new memories. This can extend memory capacity up to $\mathcal{O}(N)$ (as long as f decreases only up to $\mathcal{O}(\sqrt{N})$, in which case critical forgetting becomes a danger again) [270]. The actual memory capacity of finite-size networks remains a difficult question, dependent on the specifics of the network [277, 278].

In the sparse coding limit, memories are uncorrelated with each other; each involves only very few randomly and independently selected neurons. Similarly, theories of palimpsest networks typically consider plasticity that is uncorrelated between synapses. There are several reasons to further examine this assumption. One mechanism of correlations between synapses' plasticity is heterosynaptic plasticity, jointly affecting the dynamics of all synapses onto a particular neuron. For example, work by Turrigiano & colleagues shows that neocortical neurons homeostatically scale synaptic weights, on a timescale of days, in order to maintain stable firing activity [90]. This can have a strong stabilizing effect on activity in recurrent networks (e.g. [129, 115]) and is by definition highly correlated between a neuron's afferents. The effect of heterosynaptic plasticity on memory capacity, to our knowledge, remains to be examined.

The structure of neural networks are characterized by correlations between different neurons' inputs and outputs [91, 151], which can strongly shape the correlation between different neurons' activity [162, 110, 155, 161, 160]. We showed in Chapter 3 that such structural correlations can be reinforced whenever synapses are, on average, potentiating. Correlations between neurons' activity during stimulus-driven or spontaneous activity, potentially re-

flecting previously learned stimulus structures, will induce correlations in the plasticity of different synapses - as we studied in Chapter 4. Temporal correlations within an individual pre- or postsynaptic neurons' activity can increase the storage capacity of a perceptron-like learning rule [279]. Senn & Fusi have shown that the presence of inhibition and realistically balanced synapses can allow a perceptron-like feedforward network to discriminate correlated, non-linearly-separable inputs [280]. The effect of spatial correlations on memory capacity in recurrent networks remains an important open question. Spatial correlations between neurons' activity should, in a recurrent network, induce correlations in the plasticity of different synapses. Motif statistics represent a general way to account for different types of structural correlation in networks [110]; the development of a motif-based theory of memory capacity would be an exciting development for this field.

APPENDIX A

MATHEMATICAL NOTATION

Table A1: **Mathematical Notation - Spiking Activity**

Variable	Description	Units
r_i	Stationary firing rate of neuron i	sp/ms
r, r_E	Population-averaged excitatory firing rate	sp/ms
$\mathbf{y}(t)$	Spike train vector	1/ms
$\mathbf{A}(t)$	Matrix of neurons' linear response functions (diagonal)	1/ms
$\mathbf{y}^0(t)$	Fictional spike train in the absence of synaptic coupling	1/ms
$\mathbf{C}^0(s)$	Spike train auto-covariance matrix (diagonal)	$(\text{sp/ms})^2$
$J(t)$	Synaptic filter	1/ms
$\mathbf{K}(t)$	Effective interaction matrix with $K_{ij}(t) = A_i(t) * J(t)$	$(\text{ms } \mu\text{A/cm}^2)^{-1}$
$\mathbf{C}(s)$	Spike train covariance matrix	$(\text{sp/ms})^2$

Table A2: **Mathematical Notation - Plasticity**

Variable	Description	Units
$L(s)$	Excitatory STDP (eSTDP) rule	μA/cm^2
S	Integral of $L(s)$	ms μA/cm^2
S_η	eSTDP-weighted susceptibility of excitatory neuron pairs to external input correlations	(ms μA/cm^2)$^{-1}$
S_F	eSTDP-weighted covariance from monosynaptic connections	1/ms
S_B	eSTDP-weighted covariance from monosynaptic connections	1/ms
S_C	eSTDP-weighted covariance from reciprocal connections	(ms μA/cm^2)$^{-1}$
S_C^I	eSTDP-weighted covariance from common inhibitory inputs to excitatory neuron pairs	(ms μA/cm^2)$^{-1}$
r_I	Population-averaged inhibitory firing rate	sp/ms
$L^I(s)$	Inhibitory STDP (iSTDP) rule	μA/cm^2
S^I	Integral of $L^I(s)$	(ms μA/cm^2)$^{-1}$
S_η^{EI}	iSTDP-weighted susceptibility of excitatory-inhibitory neuron pairs to external input correlations	(ms μA/cm^2)$^{-1}$
S_F^{EI}	iSTDP-weighted covariance from inhibitory-excitatory synapses	(ms μA/cm^2)$^{-1}$
S_B^{EI}	iSTDP-weighted covariance from reciprocal excitatory-inhibitory synapses	(ms μA/cm^2)$^{-1}$
S_C^{EIE}	iSTDP-weighted covariance to excitatory-inhibitory pairs from common excitation	(ms μA/cm^2)$^{-1}$
S_C^{EII}	iSTDP-weighted covariance to excitatory-inhibitory pairs from common inhibition	(ms μA/cm^2)$^{-1}$

Table A3: **Mathematical Notation - Network connectivity**

Variable	Description	Units
$\mathbf{W}$	Synaptic weight matrix	μA/cm^2
$\mathbf{W}^0$	Connectivity (adjacency) matrix	none
p_0	connection density	none
q_0^{div}	frequency of two-synapse divergent motifs above Erös-Rènyi levels	none
q_0^{con}	frequency of two-synapse convergent motifs above Erös-Rènyi levels	none
q_0^{ch}	frequency of two-synapse chain motifs above Erös-Rènyi levels	none
q_0^{rec}	frequency of reciprocal loops above (non-symmetric) Erös-Rènyi levels	none

Table A4: **Mathematical Notation - Motif strengths in networks without macroscopic structure (Chapter 3)**

Variable	Description	Units
p	Mean excitatory synaptic weight	μA/cm^2
q^{div}	Mean strength of excitatory two-synapse divergent motif above chance	$(\mu\text{A/cm}^2)^2$
q^{con}	Mean strength of excitatory two-synapse convergent motif above chance	$(\mu\text{A/cm}^2)^2$
q^{ch}	Mean strength of excitatory two-synapse chain motif above chance	$(\mu\text{A/cm}^2)^2$
q^{con}	Mean strength of excitatory two-synapse convergent motif above chance	$(\mu\text{A/cm}^2)^2$
$q_{\text{X}}^{\text{rec}}$	Mean strength of excitatory synapses, conditioned on being part of a reciprocal loop	$(\mu\text{A/cm}^2)$
$q_{\text{X}}^{\text{div}}$	Mean strength of excitatory synapses conditioned on being part of a divergent motif, above chance	$(\mu\text{A/cm}^2)$
$q_{\text{X}}^{\text{con}}$	Mean strength of excitatory synapses conditioned on being part of a convergent motif, above chance	$(\mu\text{A/cm}^2)$
$q_{\text{X}}^{\text{ch,A}}$	Mean strength of excitatory synapses conditioned on being the second of a two-synapse chain, above chance	$(\mu\text{A/cm}^2)$
$q_{\text{X}}^{\text{ch,B}}$	Mean strength of excitatory synapses conditioned on being the first of a two-synapse chain, above chance	$(\mu\text{A/cm}^2)$
q^{rec}	Mean strength of excitatory two-synapse loops	$(\mu\text{A/cm}^2)^2$
q^{op}	Mean strength of excitatory non-reciprocal (open) two-synapse chains, above chance	$(\mu\text{A/cm}^2)^2$
$q_{\text{X2}}^{\text{rec}}$	Mean squared excitatory synaptic weight, conditioned on being part of a two-synapse loop	$(\mu\text{A/cm}^2)^2$

Table A5: **Mathematical Notation - Motif strengths in networks with macroscopic structure (Chapter 4)**

Variable	Description	Units
p	Mean excitatory synaptic weight	μA/cm^2
q	Mean strength of excitatory synapses, conditioned on being part of a reciprocal loop (same as $q_{\mathrm{X}}^{\mathrm{rec}}$ in Chapter 3)	μA/cm^2
p_{EI}	Mean inhibitory-excitatory synaptic weight	μA/cm^2
p_{AA}	Mean strength of excitatory synapses within assemblies	μA/cm^2
p_{AB}	Mean strength of excitatory synapses between assemblies	μA/cm^2
q_{AA}	Mean strength of excitatory synapses within assemblies conditioned on being part of a reciprocal loop, above chance	μA/cm^2
q_{AB}	Mean strength of excitatory synapses between assemblies conditioned on being part of a reciprocal loop, above chance	μA/cm^2

APPENDIX B

FOKKER-PLANCK METHODS FOR CALCULATING SINGLE-NEURON SPIKING STATISTICS

In this section we review a set of methods for calculating single-neuron spiking statistics. These are based off of one main idea: that a neuron's membrane voltage is noisy, and should be described by a probability distribution $P(V,t)$. There are multiple sources of noise in a neuron's membrane voltage [281]. The definitively stochastic sources are the probabilistic opening and closing of single ion channels and Johnson noise. The main source we consider, however, is the apparently stochastic spiking of presynaptic neurons. We examine the case where a neuron receives a large number of presynaptic inputs, which can in sum be approximated as a diffusion process (i.e. Gaussian white noise).

In this case, the membrane potential distribution $P(V,t)$ obeys a partial differential equation called the Fokker-Planck equation. While the exact form of the equation depends on the membrane potential dynamics, general techniques exist for obtaining solutions under the condition that the membrane potential is statistically stationary. We will review the calculation for the transfer function first developed in [146] and extended to models with voltage-activated conductances in [282]. We refer the reader to these past studies for any details excluded here. For earlier treatments of spiking responses in stochastic neural network activity, see [283, 284], and for a useful exposition of Fokker-Planck techniques in neural networks, see [285]. We begin by considering a single neuron receiving inputs from pools of

presynaptic excitatory and inhibitory neurons:

$$C\frac{dV}{dt} = g_L\left(V_L - V\right)) + \psi(V) + g_E(t)\left(V_E - V\right) + g_I(t)\left(V_I - V\right) + \mu(t) \tag{B.1}$$

where g_L and V_L are the leak conductance and reversal potential and $\mu(t)$ is some applied current. $\psi(V)$ is a spike-generating current. For the exponential integrate-and-fire model used in the studies of this thesis, $\psi(V) = g_L\Delta\exp\frac{V_i - V_T}{\Delta}$ where V_T is the spike initiation threshold and Δ controls the spike width. The exponential term gives rise to a divergence in the membrane voltage: an action potential is marked by $V \to \infty$ [137]. For numerical purposes the voltage dynamics are supplemented by a spike peak threshold V_{th} and a reset potential: $V(t_i) \geq V_{\text{th}} \to V(t_+) = V_{\text{re}}$ where t_i is the time of the ith postsynaptic spike. The voltage is then held at V_{re} for an absolute refractory period τ_{ref}

We take the synaptic conductances, $g_E(t)$ and $g_I(t)$ to be driven by presynaptic spikes, each of which elicits the same amount of conductance (a_E and a_I for presynaptic excitatory or inhibitory spikes, respectively). Here we treat the synaptic inputs as conductances. We will soon approximate them as currents, as was done in the work composing this thesis. We take the presynaptic spike trains to be homogenous Poisson processes with rates R_E and R_I. For large presynaptic rates, we can approximate them as a diffusion process using Campbell's theorem (which relates integrals over functions of a point process to the intensity measure of that process):

$$g_E(t)\left(V_E - V\right) + g_I(t)\left(V_I - V\right) \approx \Big(R_E a_E\left(V_E - V\right) + R_I a_I\left(V_I - V\right)\Big) + \sigma(V)\xi(t) \tag{B.2}$$

where $\left(g_L\sqrt{2C/g_L}\right)^2\sigma^2(V) = R_E\left(a_E\left(V_E - V\right)\right)^2 + R_I\left(a_I\left(V_I - V\right)\right)^2$ and $\xi(t)$ is a Gaussian white noise process with zero mean and unit amplitude. Here, we have scaled σ by $g_L\sqrt{2C/g_L}$ so that σ^2 is the infinitesimal variance of the passive membrane voltage - this is purely for convenience. This allows us to write the membrane dynamics as:

$$C\frac{dV}{dt} = g_{\text{eff}}\left(V_{\text{eff}} - V\right) + \psi(V) + \mu + \sigma(V)g_L\sqrt{2C/g_L}\xi(t) \tag{B.3}$$

where $g_{\text{eff}} = g_L + R_E a_E + R_I a_I$ and $V_{\text{eff}} = \left(g_L V_L + R_E a_E V_E + R_I a_I V_I\right)\left(g_L + R_E a_E + R_I a_I\right)^{-1}$. We then make the approximation that $\sigma^2(V) = \sigma^2(V_{\text{eff}}) \equiv \sigma^2$, thus ignoring the multiplicative nature of the synaptic noise.

The probability distribution associated with the membrane voltage dynamics then obeys a Fokker-Planck equation:

$$\frac{\partial P(V,t)}{\partial t} = -\frac{\partial J(V,t)}{\partial V} = -\frac{\partial}{\partial V}\left(I(V)P\right) + \frac{\sigma^2}{2}\frac{\partial}{\partial^2 V}P \tag{B.4}$$

where J is the probability flux [286] and $I(V) = g_{\text{eff}}\left(V_{\text{eff}} - V\right) + \psi(V) - \mu(t)$. When V_{eff}, σ^2 and μ are time-independent, the voltage has a steady state distribution $P_0(V)$ which obeys the continuity and flux equations:

$$-\frac{\partial J_0}{\partial V} = r_0\Big(\delta\left(V - V_{th}\right) - \delta\left(V - V_{re}\right)\Big) \tag{B.5}$$

$$-\frac{\partial P_0}{\partial V} = \frac{1}{g_L\sigma^2}\big(CJ_0 + IP_0\big) \tag{B.6}$$

In order to solve these, we first scale out the firing rates (i.e. replacing (P_0, J_0) by (p_0, j_0), where $P_0 = r_0(p_0 + p_{\text{ap}})$ and similar for J_0) and then integrate them backwards from the spike recording threshold V_{th} using the boundary conditions $p_0(V_{\text{th}}) = 0$ and $j_0(V_{\text{th}}) = 1$. We then recover the firing rate from the normalization condition $\int_{-\infty}^{\infty} P_0 dV + r_0\tau_{ref} = 1$. The Fokker-Planck equation allows the accurate prediction of the membrane potential distribution (Fig B1).

We then investigate the response of the system to inputs that fluctuate in time by considering the time-varying responses to a periodic input as a perturbation from the equilibrium state: $\mu(t) = \mu_0 + \hat{\mu}_1 e^{2\pi i f t}$. Here μ_0 is the time-independent component of the input and $\hat{\mu}_1$ is the amplitude of the input modulation, taken to be small. To first order in $\hat{\mu}_1$, the periodic input induces periodic modulations in the system at the same frequency f. We decompose the probability density, probability flux and firing rate into steady-state and modulated components:

$$P = P_0 + \hat{P}_1 e^{2\pi i f t},\ J = J_0 + \hat{J}_1 e^{2\pi i f t},\ r = r_0 + \hat{A} e^{2\pi i f t}.$$

We solve for the modulated components in the Fourier domain after obtaining the equilibrium solution to Eqs. (B.5)-(B.6). They obey the first order continuity and flux equations:

$$-\frac{\partial \hat{J}_1}{\partial V} = 2\pi i f \hat{P}_1 + \hat{A}\Big(\delta\left(V - V_{th}\right) - e^{-2\pi i f \tau_{ref}}\delta\left(V - V_{re}\right)\Big) \tag{B.7}$$

$$-\frac{\partial \hat{P}_1}{\partial V} = \frac{1}{g_L\sigma^2}\Big(\mu_0\hat{P}_1 + C\hat{J}_1 - \hat{\mu}_1 P_0\Big) \tag{B.8}$$

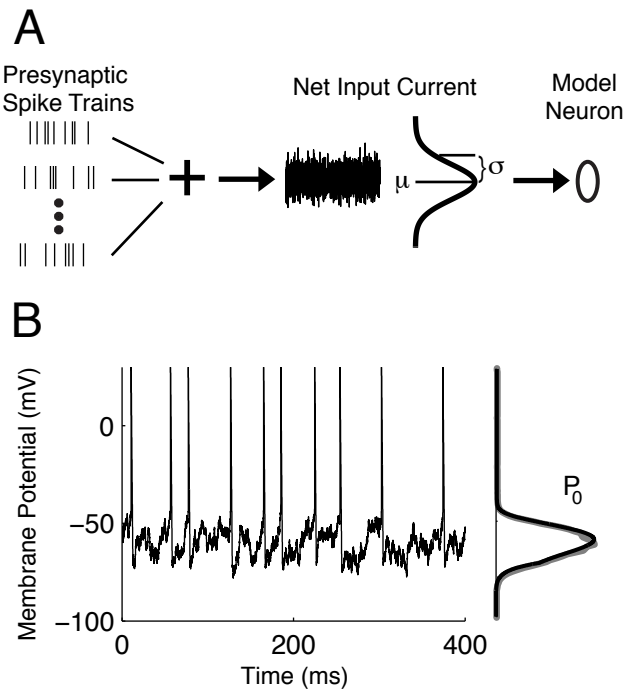

Figure B1: Adapted from Ocker & Doiron 2014, [170]. (A) Schematic of white-noise approximation of Poisson inputs. (B) Left, time series of membrane potential. Right, distribution of the membrane potential. Solid, Fokker-Planck theory. Shaded, simulation.)

In total, Fokker-Planck theory provides an accurate prediction of the responses of model neurons receiving Poisson inputs (Fig B2).

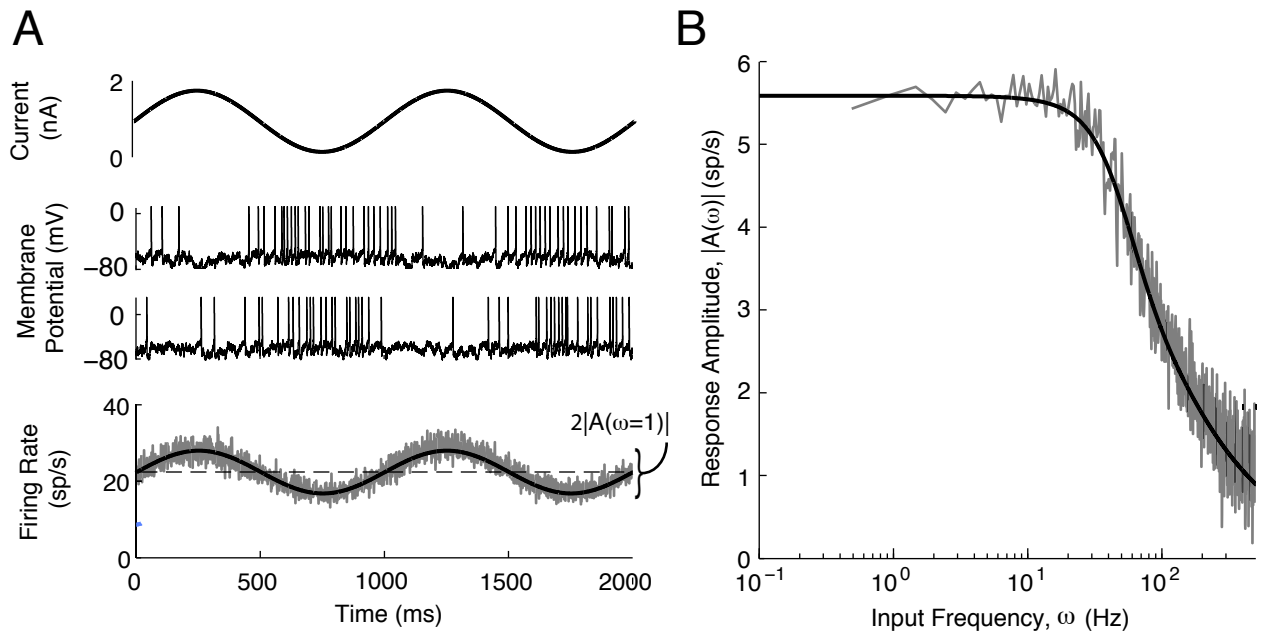

Figure B2: Adapted from Ocker & Doiron 2014 [170]. (A) Injecting a sinusoidal input induces sinusoidal modulations of the membrane potential on individual trials and of the instantaneous firing rate, averaged across trials. (B) Amplitude of the linear response (modulation of the instantaneous firing rate) as a function of input frequency. Solid, Fokker-Planck theory. Shaded, simulation.

APPENDIX C

LINEAR RESPONSE FOR SPIKE TRAIN COVARIANCE IN RECURRENT NETWORKS

In this appendix, we present a distinct derivation of Eq. (2.4) through a constructive approach towards the approximation in Eq. (2.3). This is a summary of work from the Josič and Shea-Brown groups [136]. We start with a realization of each neuron's spike train, taking into account only the mean (time-averaged) synaptic input to each cell, calling these baseline spike trains $\mathbf{y}^0$. Importantly, $\mathbf{y}^0$ neglects any time-varying synaptic interactions. The spiking output of the postsynaptic neuron, i, in response to the time-varying inputs from presynaptic cells can be approximated as:

$$\mathbf{y}_i^{(1)}(\omega) = \mathbf{y}_i^0(\omega) + \mathbf{A}_i(\omega)\left(\sum_{j=1}^{N} \mathbf{W}_{ij} J(\omega)\mathbf{y}_j^0(\omega)\right) \tag{C.1}$$

The difference between this approximation and the previous approximation Eq. (2.3) is that here, we examine the response to the baseline presynaptic spike trains $\mathbf{y}_j^0$, rather than the full presynaptic spike trains $\mathbf{y}_j$. The cross-covariance of the processes $\mathbf{y}^1$ yields a first approximation to the full cross-covariance matrix:

$$\begin{aligned}\mathbf{C}_{ij}(\omega) \approx \mathbf{C}_{ij}^{(1)}(\omega) = \delta_{ij}\mathbf{C}_{ij}^0 + \left(\mathbf{W}_{ij}\mathbf{K}_{ij}(\omega)\mathbf{C}_{jj}^0(\omega)\right) + \left(\mathbf{W}_{ji}\mathbf{K}_{ij}^*(\omega)\mathbf{C}_{ii}^0(\omega)\right) \\ + \sum_k \left(\mathbf{W}_{ik}\mathbf{K}_{ik}(\omega)\mathbf{W}_{jk}\mathbf{K}_{kj}^*(\omega)\mathbf{C}_{kk}^0(\omega)\right)\end{aligned} \tag{C.2}$$

This accounts for spiking covariability due to direct synaptic interactions but does not measure the contribution of larger network structures containing paths of multiple synapses.

Such larger structures can strongly affect spike-train covariability [156, 155, 162, 110, 160, 161, 158]. Further approximations can be built, again starting from $\mathbf{y}^0$. Each approximation takes into account longer paths through the network [287, 288]:

$$\mathbf{y}^{(n+1)} = \mathbf{y}^0 + \mathbf{A}_i(\omega)\left(\sum_{j=1}^{N} \mathbf{W}_{ij} J(\omega)\mathbf{y}_j^{(n)}(\omega)\right) \tag{C.3}$$

The first step in this iterative approximation assumes that the inputs are generated by neurons spiking in isolation so only monosynaptic paths contribute. In the second step, neurons in the network influence each other through disynaptic paths. Likewise, at step n the activity of isolated neurons is filtered through paths up to length n to approximate the spiking response.

As we iteratively approximate the spike trains, so can we successively approximate the spike train covariances. At each step we obtain the approximation:

$$\mathbf{C}^{(n)}(\omega) = \left(\sum_{k=0}^{n} (\mathbf{W}\cdot\mathbf{K})^k\right)\mathbf{C}^0(\omega)\left(\sum_{l=0}^{n} ((\mathbf{W}\cdot\mathbf{K})^*)^l\right) \tag{C.4}$$

In order to take into account the full network structure, we take the limit $n \to \infty$ in Eq. (C.4), yielding a power series for $\mathbf{C}(\omega)$. If the spectral radius of the interaction matrix $\mathbf{W}\cdot\mathbf{K}$ is less than one, this series converges to Eq. (2.4). This convergence condition must hold at each frequency ω. Interestingly, at $\omega = 0$ this is the same condition as for linear stability of the stationary firing rates in the network.

BIBLIOGRAPHY

[1] Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. The Journal of Physiology 117: 500–544.

[2] Spruston N, Jonas P, Sakmann B (1995) Dendritic glutamate receptor channels in rat hippocampal CA3 and CA1 pyramidal neurons. The Journal of Physiology 482: 325–352.

[3] Bellingham MC, Lim R, Walmsley B (1998) Developmental changes in EPSC quantal size and quantal content at a central glutamatergic synapse in rat. The Journal of Physiology 511: 861–869.

[4] Jonas P, Major G, Sakmann B (1993) Quantal components of unitary EPSCs at the mossy fibre synapse on CA3 pyramidal cells of rat hippocampus. The Journal of Physiology 472: 615–663.

[5] Angulo MC, Rossier J, Audinat E (1999) Postsynaptic Glutamate Receptors and Integrative Properties of Fast-Spiking Interneurons in the Rat Neocortex. Journal of Neurophysiology 82: 1295–1302.

[6] Taschenberger H, Engert F, Grantyn R (1995) Synaptic current kinetics in a solely AMPA-receptor-operated glutamatergic synapse formed by rat retinal ganglion neurons. Journal of Neurophysiology 74: 1123–1136.

[7] Sah P, Hestrin S, Nicoll RA (1990) Properties of excitatory postsynaptic currents recorded in vitro from rat hippocampal interneurones. The Journal of Physiology 430: 605–616.

[8] Kinney GA, Peterson BW, Slater NT (1994) The synaptic activation of N-methyl-D-aspartate receptors in the rat medial vestibular nucleus. Journal of Neurophysiology 72: 1588–1595.

[9] Flint AC, Maisch US, Weishaupt JH, Kriegstein AR, Monyer H (1997) NR2a Subunit Expression Shortens NMDA Receptor Synaptic Currents in Developing Neocortex. The Journal of Neuroscience 17: 2469–2476.

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

[11] Vargas-Caballero M, Robinson HPC (2004) Fast and Slow Voltage-Dependent Dynamics of Magnesium Block in the NMDA Receptor: The Asymmetric Trapping Block Model. The Journal of Neuroscience 24: 6171–6180.

[12] Jahr CE, Stevens CF (1990) A quantitative description of NMDA receptor-channel kinetic behavior. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 10: 1830–1837.

[13] Markram H, Gerstner W, Sjöström PJ (2011) A history of spike-timing-dependent plasticity. Frontiers in Synaptic Neuroscience 3: 4.

[14] Lugaro, Ernesto (1909) Modern Problems in Psychiatry. Manchester: The University Press.

[15] Hebb DO (1949) The organization of behavior: a neuropsychological theory. Mahwah, N.J.: L. Erlbaum Associates.

[16] Kandel ER, Tauc L (1964) Mechanism of Prolonged Heterosynaptic Facilitation. Nature 202: 145–147.

[17] Kandel ER, Tauc L (1965) Heterosynaptic facilitation in neurones of the abdominal ganglion of Aplysia depilans. The Journal of Physiology 181: 1–27.

[18] Meliza CD, Dan Y (2006) Receptive-field modification in rat visual cortex induced by paired visual stimulation and single-cell spiking. Neuron 49: 183–189.

[19] Jacob V, Brasier DJ, Erchova I, Feldman D, Shulz DE (2007) Spike Timing-Dependent Synaptic Depression in the In Vivo Barrel Cortex of the Rat. The Journal of neuroscience : the official journal of the Society for Neuroscience 27: 1271–1284.

[20] Nishimura Y, Perlmutter SI, Eaton RW, Fetz EE (2013) Spike-Timing-Dependent Plasticity in Primate Corticospinal Connections Induced during Free Behavior. Neuron .

[21] Conde V, Vollmann H, Taubert M, Sehm B, Cohen LG, et al. (2013) Reversed timing-dependent associative plasticity in the human brain through interhemispheric interactions. Journal of Neurophysiology 109: 2260–2271.

[22] Koch G, Ponzo V, Lorenzo FD, Caltagirone C, Veniero D (2013) Hebbian and Anti-Hebbian Spike-Timing-Dependent Plasticity of Human Cortico-Cortical Connections. The Journal of Neuroscience 33: 9725–9733.

[23] Lømo T (1966) Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. Acta Physiol Scand 68.

[24] Bliss TVP, Lømo T (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. The Journal of Physiology 232: 331–356.

[25] Douglas RM, Goddard GV (1975) Long-term potentiation of the perforant path-granule cell synapse in the rat hippocampus. Brain Research 86: 205–215.

[26] Schwartzkroin PA, Wester K (1975) Long-lasting facilitation of a synaptic potential following tetanization in thein vitro hippocampal slice. Brain Research 89: 107–119.

[27] McNaughton BL, Douglas RM, Goddard GV (1978) Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. Brain Research 157: 277–293.

[28] Baranyi A, Fehér O (1981) Intracellular studies on cortical synaptic plasticity. Conditioning effect of antidromic activation on test-EPSPs. Experimental Brain Research 41: 124–134.

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

[30] Dunwiddie T, Lynch G (1978) Long-term potentiation and depression of synaptic responses in the rat hippocampus: localization and frequency dependency. The Journal of Physiology 276: 353–367.

[31] Levy WB, Steward O (1979) Synapses as associative memory elements in the hippocampal formation. Brain Research 175: 233–245.

[32] Barrionuevo G, Schottler F, Lynch G (1980) The effects of repetitive low frequency stimulation on control and "potentiated" synaptic responses in the hippocampus. Life Sciences 27: 2385–2391.

[33] Levy WB, Steward O (1983) Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. Neuroscience 8: 791–797.

[34] Gustafsson B, Wigstrom H (1986) Hippocampal long-lasting potentiation produced by pairing single volleys and brief conditioning tetani evoked in separate afferents. The Journal of Neuroscience 6: 1575–1582.

[35] Gustafsson B, Wigstrom H, Abraham WC, Huang YY (1987) Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. The Journal of Neuroscience 7: 774–780.

[36] Markram H (1997) Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. Science 275: 213–215.

[37] Bi Gq, Poo Mm (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of Neuroscience 18: 10464–10472.

[38] Froemke RC, Dan Y (2002) Spike-timing-dependent synaptic modification induced by natural spike trains. Nature 416: 433–438.

[39] Bi GQ, Wang HX (2002) Temporal asymmetry in spike timing-dependent synaptic plasticity. Physiology & Behavior 77: 551–555.

[40] Wang HX, Gerkin RC, Nauen DW, Bi GQ (2005) Coactivation and timing-dependent integration of synaptic potentiation and depression. Nature Neuroscience 8: 187–193.

[41] Froemke RC, Tsay IA, Raad M, Long JD, Dan Y (2006) Contribution of individual spikes in burst-induced long-term synaptic modification. Journal of Neurophysiology 95: 1620–1629.

[42] Feldman DE (2012) The Spike-Timing Dependence of Plasticity. Neuron 75: 556–571.

[43] Kampa BM, Clements J, Jonas P, Stuart GJ (2004) Kinetics of Mg2+ unblock of NMDA receptors: implications for spike-timing dependent synaptic plasticity. The Journal of Physiology 556: 337–345.

[44] Fuenzalida M, Fernández de Sevilla D, Couve A, Buño W (2010) Role of AMPA and NMDA receptors and back-propagating action potentials in spike timing-dependent plasticity. Journal of Neurophysiology 103: 47–54.

[45] Holbro N, Grunditz Å, Wiegert JS, Oertner TG (2010) AMPA receptors gate spine Ca2+ transients and spike-timing-dependent potentiation. Proceedings of the National Academy of Sciences 107: 15975–15980.

[46] Hoffman DA, Magee JC, Colbert CM, Johnston D (1997) K+ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. Nature 387: 869–875.

[47] Stuart GJ, Häusser M (2001) Dendritic coincidence detection of EPSPs and action potentials. Nature Neuroscience 4: 63–71.

[48] Karmarkar UR, Najarian MT, Buonomano DV (2002) Mechanisms and significance of spike-timing dependent plasticity. Biological Cybernetics 87: 373–382.

[49] Shouval HZ, Bear MF, Cooper LN (2002) A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. Proceedings of the National Academy of Sciences of the United States of America 99: 10831–10836.

[50] Rosenmund C, Feltz A, Westbrook GL (1995) Calcium-dependent inactivation of synaptic NMDA receptors in hippocampal neurons. Journal of Neurophysiology 73: 427–430.

[51] Tong, G, Shepherd, D, Jahr, CE (1995) Synaptic desensitization of NMDA receptors by calcineurin. Science 267: 1510–1512.

[52] Froemke RC, Poo Mm, Dan Y (2005) Spike-timing-dependent synaptic plasticity depends on dendritic location. Nature 434: 221–225.

[53] Hashimotodani Y, Ohno-Shosaku T, Tsubokawa H, Ogata H, Emoto K, et al. (2005) Phospholipase Cbeta serves as a coincidence detector through its Ca2+ dependency for triggering retrograde endocannabinoid signal. Neuron 45: 257–268.

[54] Nakamura T, Barbara JG, Nakamura K, Ross WN (1999) Synergistic release of Ca2+ from IP3-sensitive stores evoked by synaptic activation of mGluRs paired with backpropagating action potentials. Neuron 24: 727–737.

[55] Sjöström PJ, Turrigiano GG, Nelson SB (2003) Neocortical LTD via coincident activation of presynaptic NMDA and cannabinoid receptors. Neuron 39: 641–654.

[56] Bender VA, Bender KJ, Brasier DJ, Feldman DE (2006) Two Coincidence Detectors for Spike Timing-Dependent Plasticity in Somatosensory Cortex. The Journal of Neuroscience 26: 4166–4177.

[57] Nevian T, Sakmann B (2006) Spine Ca2+ signaling in spike-timing-dependent plasticity. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 26: 11001–11013.

[58] Seol GH, Ziburkus J, Huang S, Song L, Kim IT, et al. (2007) Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. Neuron 55: 919–929.

[59] Fino E, Paille V, Cui Y, Morera-Herreras T, Deniau JM, et al. (2010) Distinct coincidence detectors govern the corticostriatal spike timing-dependent plasticity. The Journal of Physiology 588: 3045–3062.

[60] Tzounopoulos T, Kim Y, Oertel D, Trussell LO (2004) Cell-specific, spike timing-dependent plasticities in the dorsal cochlear nucleus. Nature neuroscience 7: 719–725.

[61] Bastian J (1996) Plasticity in an electrosensory system. I. General features of a dynamic sensory filter. Journal of Neurophysiology 76: 2483–2496.

[62] Bell CC, Han VZ, Sugawara Y, Grant K (1997) Synaptic plasticity in a cerebellum-like structure depends on temporal order. Nature 387: 278–281.

[63] Requarth T, Sawtell NB (2011) Neural mechanisms for filtering self-generated sensory signals in cerebellum-like circuits. Current Opinion in Neurobiology 21: 602–608.

[64] Tzounopoulos T, Rubio ME, Keen JE, Trussell LO (2007) Coactivation of pre- and postsynaptic signaling mechanisms determines cell-specific spike-timing-dependent plasticity. Neuron 54: 291–301.

[65] Zhao Y, Tzounopoulos T (2011) Physiological Activation of Cholinergic Inputs Controls Associative Synaptic Plasticity via Modulation of Endocannabinoid Signaling. The Journal of Neuroscience 31: 3158–3168.

[66] Markram H, Tsodyks M (1996) Redistribution of synaptic efficacy between neocortical pyramidal neurons. Nature 382: 807–810.

[67] Buonomano DV (1999) Distinct Functional Types of Associative Long-Term Potentiation in Neocortical and Hippocampal Pyramidal Neurons. The Journal of Neuroscience 19: 6748–6754.

[68] Selig DK, Nicoll RA, Malenka RC (1999) Hippocampal Long-Term Potentiation Preserves the Fidelity of Postsynaptic Responses to Presynaptic Bursts. The Journal of Neuroscience 19: 1236–1246.

[69] Oswald AMM, Lewis JE, Maler L (2002) Dynamically Interacting Processes Underlie Synaptic Plasticity in a Feedback Pathway. Journal of Neurophysiology 87: 2450–2463.

[70] Graupner M, Brunel N (2010) Mechanisms of induction and maintenance of spike-timing dependent plasticity in biophysical synapse models. Frontiers in Computational Neuroscience 4: 136.

[71] Yang SN, Tang YG, Zucker RS (1999) Selective induction of LTP and LTD by postsynaptic [Ca2+]i elevation. Journal of Neurophysiology 81: 781–787.

[72] Mizuno T, Kanazawa I, Sakurai M (2001) Differential induction of LTP and LTD is not determined solely by instantaneous calcium concentration: an essential involvement of a temporal factor. The European Journal of Neuroscience 14: 701–708.

[73] Ismailov I, Kalikulov D, Inoue T, Friedlander MJ (2004) The kinetic profile of intracellular calcium predicts long-term potentiation and long-term depression. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 24: 9847–9861.

[74] Shouval HZ, Kalantzis G (2005) Stochastic properties of synaptic transmission affect the shape of spike time-dependent plasticity curves. Journal of Neurophysiology 93: 1069–1073.

[75] Cai Y, Gavornik JP, Cooper LN, Yeung LC, Shouval HZ (2007) Effect of stochastic synaptic and dendritic dynamics on synaptic plasticity in visual cortex and hippocampus. Journal of Neurophysiology 97: 375–386.

[76] Graupner M, Brunel N (2012) Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. Proceedings of the National Academy of Sciences 109: 3991–3996.

[77] Magee JC, Johnston D (1997) A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. Science (New York, NY) 275: 209–213.

[78] Abarbanel HDI, Gibb L, Huerta R, Rabinovich MI (2003) Biophysical model of synaptic plasticity dynamics. Biological Cybernetics 89: 214–226.

[79] Karmarkar UR, Buonomano DV (2002) A model of spike-timing dependent plasticity: one or two coincidence detectors? Journal of Neurophysiology 88: 507–513.

[80] Rubin JE, Gerkin, RC, Bi, G-Q, Chow, C (2005) Calcium Time Course as a Signal for Spike-Timing-Dependent Plasticity. Journal of Neurophysiology 93: 2600–2613.

[81] Abbott LF, Nelson SB (2000) Synaptic plasticity: taming the beast. Nature neuroscience 3: 1178–1183.

[82] Bienenstock EL, Cooper LN, Munro PW (1982) Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. The Journal of Neuroscience 2: 32–48.

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

[84] Linsker R (1986) From basic network principles to neural architecture: emergence of spatial-opponent cells. Proceedings of the National Academy of Sciences 83: 7508–7512.

[85] Intrator N, Cooper LN (1992) Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. Neural Networks 5: 3–17.

[86] Sejnowski TJ (1977) Storing covariance with nonlinearly interacting neurons. Journal of Mathematical Biology 4: 303–321.

[87] Gerstner W, Ritz R, van Hemmen JL (1993) Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. Biological Cybernetics 69: 503–515.

[88] Gerstner W, Kempter R, van Hemmen JL, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature 383: 76–78.

[89] Song S, Miller KD, Abbott LF (2000) Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nature Neuroscience 3: 919–926.

[90] Turrigiano GG, Leslie KR, Desai NS, Rutherford LC, Nelson SB (1998) Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature 361: 892–896.

[91] Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB (2005) Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol 3: e68.

[92] Cossell L, Iacaruso MF, Muir DR, Houlton R, Sader EN, et al. (2015) Functional organization of excitatory synaptic strength in primary visual cortex. Nature 518: 399–403.

[93] Rubin J, Lee D, Sompolinsky H (2001) Equilibrium Properties of Temporally Asymmetric Hebbian Plasticity. Physical Review Letters 86: 364–367.

[94] Van Rossum MC, Bi GQ, Turrigiano GG (2000) Stable Hebbian learning from spike timing-dependent plasticity. The Journal of Neuroscience 20: 8812–8821.

[95] Gütig R, Aharonov R, Rotter S, Sompolinsky H (2003) Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 23: 3697–3714.

[96] Billings G, Rossum MCWv (2009) Memory Retention and Spike-Timing-Dependent Plasticity. Journal of Neurophysiology 101: 2775–2788.

[97] Gilson M, Fukai T (2011) Stability versus Neuronal Specialization for STDP: Long-Tail Weight Distributions Solve the Dilemma. PLoS ONE 6: e25339.

[98] Kistler WM, Van Hemmen JL (2000) Modeling synaptic plasticity in conjunction with the timing of pre-and postsynaptic action potentials. Neural Computation 12: 385–405.

[99] Kempter R, Gerstner W, Van Hemmen JL (2001) Intrinsic stabilization of output rates by spike-based Hebbian learning. Neural Computation 13: 2709–2741.

[100] Gilson M, Burkitt AN, Grayden DB, Thomas DA, Hemmen JL (2009) 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: 411–426.

[101] Nelson PG, Fields RD, Yu C, Liu Y (1993) Synapse elimination from the mouse neuromuscular junction in vitro: a non-Hebbian activity-dependent process. Journal of Neurobiology 24: 1517–1530.

[102] Urban NN, Barrionuevo G (1996) Induction of Hebbian and Non-Hebbian Mossy Fiber Long-Term Potentiation by Distinct Patterns of High-Frequency Stimulation. The Journal of Neuroscience 16: 4293–4299.

[103] Salin PA, Malenka RC, Nicoll RA (1996) Cyclic AMP Mediates a Presynaptic Form of LTP at Cerebellar Parallel Fiber Synapses. Neuron 16: 797–803.

[104] Alonso A, de Curtis M, Llinás R (1990) Postsynaptic Hebbian and non-Hebbian long-term potentiation of synaptic efficacy in the entorhinal cortex in slices and in the isolated adult guinea pig brain. Proceedings of the National Academy of Sciences of the United States of America 87: 9280–9284.

[105] Sjöström PJ, Turrigiano GG, Nelson SB (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron 32: 1149–1164.

[106] Pfister JP, Gerstner W (2006) Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. The Journal of Neuroscience 26: 9673–9682.

[107] Clopath C, Gerstner W (2010) Voltage and Spike Timing Interact in STDP – A Unified Model. Frontiers in Synaptic Neuroscience 2.

[108] Clopath, Claudia, Büsing, Lars, Vasilaki, Eleni, Gerstner, Wulfram (2010) Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nat Neurosci 13: 344–352.

[109] Kempter R, Gerstner W, Van Hemmen JL (1999) Hebbian learning and spiking neurons. Physical Review E 59: 4498.

[110] Hu Y, Trousdale J, Josić K, Shea-Brown E (2014) Local paths to global coherence: Cutting networks down to size. Physical Review E 89: 032802.

[111] Buice MA, Chow CC (2013) Beyond mean field theory: statistical field theory for neural networks. Journal of Statistical Mechanics (Online) 2013: P03003.

[112] Buice MA, Chow CC (2013) Generalized activity equations for spiking neural network dynamics. Frontiers in Computational Neuroscience 7: 162.

[113] Hu, Yu, Josić, Krešimir, Shea-Brown, Eric, Buice, Michael A (2015) From structure to dynamics: origin of higher-order spike correlations in network motifs. Cosyne Abstracts .

[114] Gjorgjieva J, Clopath C, Audet J, Pfister JP (2011) A triplet spike-timing–dependent plasticity model generalizes the Bienenstock–Cooper–Munro rule to higher-order spatiotemporal correlations. Proceedings of the National Academy of Sciences 108: 19383–19388.

[115] Zenke F, Agnes EJ, Gerstner W (2015) Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nature Communications 6.

[116] Ocker GK, Litwin-Kumar A, Doiron B (2015) Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses. PLoS Comput Biol 11: e1004458.

[117] Caporale N, Dan Y (2008) Spike Timing–Dependent Plasticity: A Hebbian Learning Rule. Annual Review of Neuroscience 31: 25–46.

[118] Harris KD (2005) Neural signatures of cell assembly organization. Nature Reviews Neuroscience 6: 399–407.

[119] Buzsáki G (2010) Neural syntax: cell assemblies, synapsembles, and readers. Neuron 68: 362–385.

[120] Harris KD, Mrsic-Flogel TD (2013) Cortical connectivity and sensory coding. Nature 503: 51–58.

[121] Babadi B, Abbott LF (2010) Intrinsic Stability of Temporally Shifted Spike-Timing Dependent Plasticity. PLoS Comput Biol 6: e1000961.

[122] Meffin H, Besson J, Burkitt AN, Grayden DB (2006) Learning the structure of correlated synaptic subgroups using stable and competitive spike-timing-dependent plasticity. Physical Review E 73: 041911.

[123] Gilson M, Burkitt AN, Grayden DB, Thomas DA, Hemmen JL (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity–strengthening correlated input pathways. Biological Cybernetics 101: 81–102.

[124] Fiete IR, Senn W, Wang CZH, Hahnloser RHR (2010) Spike-Time-Dependent Plasticity and Heterosynaptic Competition Organize Networks to Produce Long Scale-Free Sequences of Neural Activity. Neuron 65: 563–576.

[125] Kumar A, Rotter S, Aertsen A (2010) Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. Nat Rev Neurosci 11: 615–627.

[126] Izhikevich EM, Gally JA, Edelman GM (2004) Spike-timing Dynamics of Neuronal Groups. Cerebral Cortex 14: 933–944.

[127] Morrison A, Aertsen A, Diesmann M (2007) Spike-Timing-Dependent Plasticity in Balanced Random Networks. Neural Computation 19: 1437–1467.

[128] Babadi B, Abbott LF (2013) Pairwise Analysis Can Account for Network Structures Arising from Spike-Timing Dependent Plasticity. PLoS Computational Biology 9: e1002906.

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

[130] Karbowski J, Ermentrout G (2002) Synchrony arising from a balanced synaptic plasticity in a network of heterogeneous neural oscillators. Physical Review E 65.

[131] Burkitt AN, Gilson M, Hemmen JLv (2007) Spike-timing-dependent plasticity for neurons with recurrent connections. Biological Cybernetics 96: 533–546.

[132] Gilson M, Burkitt AN, Grayden DB, Thomas DA, Hemmen JL (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks IV: Structuring synaptic pathways among recurrent connections. Biological Cybernetics 101: 427–444.

[133] Gilson M, Burkitt AN, Grayden DB, Thomas DA, Hemmen JL (2010) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks V: self-organization schemes and weight dependence. Biological Cybernetics 103: 365–386.

[134] Doiron B, Lindner B, Longtin A, Maler L, Bastian J (2004) Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. Phys Rev Let 93.

[135] Lindner B, Doiron B, Longtin A (2005) Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. Phys Rev E 72.

[136] Trousdale J, Hu Y, Shea-Brown E, Josić, Krešimir (2012) Impact of Network Structure and Cellular Response on Spike Time Correlations. PLoS Computational Biology 8: e1002408.

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

[138] Jolivet R, Lewis TJ, Gerstner W (2004) Generalized Integrate-and-Fire Models of Neuronal Activity Approximate Spike Trains of a Detailed Model to a High Degree of Accuracy. Journal of Neurophysiology 92: 959–976.

[139] Jolivet R, Schürmann F, Berger TK, Naud R, Gerstner W, et al. (2008) The quantitative single-neuron modeling competition. Biological Cybernetics 99: 417–426.

[140] Gardiner C (2009) Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer Berlin Heidelberg.

[141] Gilson M, Burkitt AN, Grayden DB, Thomas DA, Hemmen JL (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity—symmetry breaking. Biological Cybernetics 101: 103–114.

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

[143] Smith MA, Jia X, Zandvakili A, Kohn A (2013) Laminar dependence of neuronal correlations in visual cortex. Journal of Neurophysiology 109: 940–947.

[144] Hansen BJ, Chelaru MI, Dragoi V (2012) Correlated variability in laminar cortical circuits. Neuron 76: 590–602.

[145] Cohen MR, Kohn A (2011) Measuring and interpreting neuronal correlations. Nat Neurosci 14: 811–9.

[146] Richardson M (2007) Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive. Phys Rev E 76: 021919.

[147] Richardson M (2008) Spike-train spectra and network response functions for non-linear integrate-and-fire neurons. Biol Cybern 99: 381–392.

[148] Cox D, Isham V (1980) Point Processes. Monographs on Statistics and Applied Probability. CRC Press.

[149] Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10: 186–198.

[150] Markram H (1997) A network of tufted layer 5 pyramidal neurons. Cerebral Cortex 7: 523–533.

[151] Perin R, Berger TK, Markram H (2011) A synaptic organizing principle for cortical neuronal groups. Proceedings of the National Academy of Sciences 108: 5419–5424.

[152] Yoshimura Y, Dantzker JLM, Callaway EM (2005) Excitatory cortical neurons form fine-scale functional networks. Nature 433: 868–873.

[153] Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, et al. (2011) Functional specificity of local synaptic connections in neocortical networks. Nature 473: 87–91.

[154] Yassin L, Benedetti BL, Jouhanneau JS, Wen JA, Poulet JFA, et al. (2010) An embedded subnetwork of highly active neurons in the neocortex. Neuron 68: 1043–1050.

[155] Zhao L, Beverlin BI, Netoff T, Nykamp DQ (2011) Synchronization from second order network connectivity statistics. Frontiers in Computational Neuroscience 5: 28.

[156] Roxin A (2011) The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. Frontiers in Computational Neuroscience 5: 8.

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

[158] Gaiteri C, Rubin JE (2011) The interaction of intrinsic dynamics and network topology in determining network burst synchrony. Frontiers in Computational Neuroscience 5: 10.

[159] Kriener B, Helias M, Aertsen A, Rotter S (2009) Correlations in spiking neuronal networks with distance dependent connections. Journal of Computational Neuroscience 27: 177–200.

[160] Pernice V, Staude B, Cardanobile S, Rotter S (2011) How structure determines correlations in neuronal networks. PLoS Comput Biol 7: e1002059.

[161] Pernice V, Deger M, Cardanobile S, Rotter S (2013) The relevance of network microstructure for neural dynamics. Frontiers in Computational Neuroscience 7.

[162] Hu Y, Trousdale J, Josić K, Shea-Brown E (2013) Motif statistics and spike correlations in neuronal networks. Journal of Statistical Mechanics: Theory and Experiment 2013: P03012.

[163] Helias M, Tetzlaff T, Diesmann M (2014) The Correlation Structure of Local Neuronal Networks Intrinsically Results from Recurrent Dynamics. PLoS Comput Biol 10: e1003428.

[164] Kozloski J, Cecchi GA (2010) A Theory of Loop Formation and Elimination by Spike Timing-Dependent Plasticity. Frontiers in Neural Circuits 4.

[165] Kunkel S, Diesmann M, Morrison A (2011) Limits to the Development of Feed-Forward Structures in Large Recurrent Neuronal Networks. Frontiers in Computational Neuroscience 4.

[166] Galan R, Fourcaud-Trocme N, Ermentrout B, Urban NN (2006) Correlation-induced synchronization of oscillations in olfactory bulb neurons. J Neurosci 26: 3646–3655.

[167] de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. Nature 448: 802–6.

[168] Shea-Brown E, Josić, Krešimir, de la Rocha J, Doiron B (2008) Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. Phys Rev Let 100.

[169] Hong S, Ratte S, Prescott SA, De Schutter E (2012) Single neuron firing properties impact correlation-based population coding. J Neurosci 32: 1413–1428.

[170] Ocker GK, Doiron B (2014) Kv7 channels regulate pairwise spiking covariability in health and disease. Journal of Neurophysiology 112: 340–352.

[171] Deger M, Schwalger T, Naud R, Gerstner W (2013) Dynamics of interacting finite-sized networks of spiking neurons with adaptation. arXiv:13114206 [q-bio] .

[172] Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, et al. (2010) The asynchronous state in cortical circuits. Science 327: 587–590.

[173] Brunel N (2000) Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons. Journal of Computational Neuroscience 8: 183–208.

[174] Litwin-Kumar A, Chacron MJ, Doiron B (2012) The spatial structure of stimuli shapes the timescale of correlations in population spiking activity. PLoS Comput Biol .

[175] Litwin-Kumar A, Oswald AMM, Urban NN, Doiron B (2011) Balanced synaptic input shapes the correlation between neural spike trains. PLoS Comput Biol 7: e1002305.

[176] D'amour JA, Froemke RC (2015) Inhibitory and Excitatory Spike-Timing-Dependent Plasticity in the Auditory Cortex. Neuron 86: 514–528.

[177] Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W (2011) Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. Science 334: 1569–1573.

[178] Wittenberg GM, Wang SSH (2006) Malleability of Spike-Timing-Dependent Plasticity at the CA3–CA1 Synapse. The Journal of Neuroscience 26: 6610–6617.

[179] Churchland AK, Kiani R, Chaudhuri R, Wang XJ, Pouget A, et al. (2011) Variance as a signature of neural computations during decision making. Neuron 69: 818–31.

[180] Kohn A, Smith MA (2005) Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. The Journal of Neuroscience 25: 3661–3673.

[181] Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nature Neuroscience 13: 369–378.

[182] Arieli A, Sterkin A, Grinvald A, Aertsen A (1996) Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Science (New York, NY) 273: 1868–1871.

[183] Tsodyks M, Kenet T, Grinvald A, Arieli A (1999) Linking Spontaneous Activity of Single Cortical Neurons and the Underlying Functional Architecture. Science 286: 1943–1946.

[184] Lefort S, Tomm C, Sarria JCF, Petersen CCH (2009) The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex. Neuron 61: 301–316.

[185] Ikegaya Y, Sasaki T, Ishikawa D, Honma N, Tao K, et al. (2013) Interpyramid Spike Transmission Stabilizes the Sparseness of Recurrent Network Activity. Cerebral Cortex 23: 293–304.

[186] Rubin JE (2001) Steady states in an iterative model for multiplicative spike-timing-dependent plasticity. Network: Computation in Neural Systems 12: 131–140.

[187] Royer S, Paré D (2003) Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature 422: 518–522.

[188] Renart A, Song P, Wang XJ (2003) Robust Spatial Working Memory through Homeostatic Synaptic Scaling in Heterogeneous Cortical Networks. Neuron 38: 473–485.

[189] Lazar A, Pipa G, Triesch J (2009) SORN: a self-organizing recurrent neural network. Frontiers in Computational Neuroscience 3: 23.

[190] Zheng P, Dimitrakakis C, Triesch J (2013) Network Self-Organization Explains the Statistics and Dynamics of Synaptic Connection Strengths in Cortex. PLoS Comput Biol 9: e1002848.

[191] Zenke F, Hennequin G, Gerstner W (2013) Synaptic Plasticity in Neural Networks Needs Homeostasis with a Fast Rate Detector. PLoS Comput Biol 9: e1003330.

[192] Abraham WC (2008) Metaplasticity: tuning synapses and networks for plasticity. Nature Reviews Neuroscience 9: 387–387.

[193] Levy N, Horn D, Meilijson I, Ruppin E (2001) Distributed synchrony in a cell assembly of spiking neurons. Neural Networks: The Official Journal of the International Neural Network Society 14: 815–824.

[194] Mongillo G, Curti E, Romani S, Amit DJ (2005) Learning in realistic networks of spiking neurons and spike-driven plastic synapses. The European Journal of Neuroscience 21: 3143–3160.

[195] Liu JK, Buonomano DV (2009) Embedding multiple trajectories in simulated recurrent neural networks in a self-organizing manner. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29: 13172–13181.

[196] Chen CC, Jasnow D (2010) Mean-field theory of a plastic network of integrate-and-fire neurons. Physical Review E 81: 011907.

[197] Mayer J, Ngo HVV, Schuster HG (2012) Dynamical Mean-Field Equations for a Neural Network with Spike Timing Dependent Plasticity. Journal of Statistical Physics 148: 677–686.

[198] Ko H, Cossell L, Baragli C, Antolik J, Clopath C, et al. (2013) The emergence of functional microcircuits in visual cortex. Nature 496: 96–100.

[199] Ko H, Mrsic-Flogel TD, Hofer SB (2014) Emergence of Feature-Specific Connectivity in Cortical Microcircuits in the Absence of Visual Experience. The Journal of Neuroscience 34: 9812–9816.

[200] Ermentrout B (2002) Simulating, Analyzing, and Animating Dynamical Systems. Software, Environments and Tools. Society for Industrial and Applied Mathematics. URL http://epubs.siam.org/doi/book/10.1137/1.9780898718195.

[201] Ledoux E, Brunel N (2011) Dynamics of networks of excitatory and inhibitory neurons in response to time-dependent inputs. Frontiers in Computational Neuroscience 5: 25.

[202] Takahashi YK, Kori H, Masuda N (2009) Self-organization of feed-forward structure and entrainment in excitatory neural networks with spike-timing-dependent plasticity. Physical Review E 79: 051904.

[203] Song S, Abbott LF (2001) Cortical development and remapping through spike timing-dependent plasticity. Neuron 32: 339–350.

[204] Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. J Neurosci 21: 1676–1697.

[205] Rothschild G, Nelken I, Mizrahi A (2010) Functional organization and population dynamics in the mouse primary auditory cortex. Nature Neuroscience 13: 353–360.

[206] Lubenov EV, Siapas AG (2008) Decoupling through synchrony in neuronal circuits with propagation delays. Neuron 58: 118–131.

[207] Vogels TP, Froemke RC, Doyon N, Gilson M, Haas JS, et al. (2013) Inhibitory synaptic plasticity: spike timing-dependence and putative network function. Frontiers in Neural Circuits 7.

[208] Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370: 140–143.

[209] Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7: 358–66.

[210] Abbott LF, Dayan P (1999) The Effect of Correlated Variability on the Accuracy of a Population Code. Neural Computation 11: 91–101.

[211] Sompolinsky H, Yoon H, Kang K, Shamir M (2001) Population coding in neuronal systems with correlated noise. Physical Review E 64: 051904.

[212] Ecker AS, Berens P, Tolias AS, Bethge M (2011) The effect of noise correlations in populations of diversely tuned neurons. J Neurosci 31: 14272–14283.

[213] Moreno-Bote R, Beck J, Kanitscheider I, Pitkow X, Latham P, et al. (2014) Information-limiting correlations. Nature Neuroscience 17: 1410–1417.

[214] Hu Y, Zylberberg J, Shea-Brown E (2014) The Sign Rule and Beyond: Boundary Effects, Flexibility, and Noise Correlations in Neural Population Codes. PLoS Comput Biol 10: e1003469.

[215] Aertsen A, Gerstein G, Habib M, Palm G (1989) Dynamics of neuronal firing correlation: modulation of "effective connectivity". J Neurophysiol 61: 900–917.

[216] Ahissar E, Vaadia E, Ahissar M, Bergman H, Arieli A, et al. (1992) Dependence of cortical plasticity on correlated activity of single neurons and on behavioral context. Science (New York, NY) 257: 1412–1415.

[217] Espinosa IE, Gerstein GL (1988) Cortical auditory neuron interactions during presentation of 3-tone sequences: effective connectivity. Brain Research 450: 39–50.

[218] Ruff DA, Cohen MR (2014) Attention can either increase or decrease spike count correlations in visual cortex. Nature Neuroscience 17: 1591–1597.

[219] Kenet T, Bibitchkov D, Tsodyks M, Grinvald A, Arieli A (2003) Spontaneously emerging cortical representations of visual attributes. Nature 425: 954–956.

[220] Luczak A, Barthó P, Harris KD (2009) Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. Neuron 62: 413–425.

[221] Han F, Caporale N, Dan Y (2008) Reverberation of recent visual experience in spontaneous cortical waves. Neuron 60: 321–327.

[222] Eagleman SL, Dragoi V (2012) Image sequence reactivation in awake V4 networks. Proceedings of the National Academy of Sciences 109: 19450–19455.

[223] Xu S, Jiang W, Poo Mm, Dan Y (2012) Activity recall in a visual cortical ensemble. Nature Neuroscience 15: 449–455.

[224] Komiyama T, Sato TR, O'Connor DH, Zhang YX, Huber D, et al. (2010) Learning-related fine-scale specificity imaged in motor cortex circuits of behaving mice. Nature 464: 1182–1186.

[225] van Vreeswijk C, Sompolinsky H (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science (New York, NY) 274: 1724–1726.

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

[227] Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL (1997) Paradoxical effects of external modulation of inhibitory interneurons. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 17: 4382–4388.

[228] Ozeki H, Finn IM, Schaffer ES, Miller KD, Ferster D (2009) Inhibitory Stabilization of the Cortical Network Underlies Visual Surround Suppression. Neuron 62: 578–592.

[229] Rubin DB, Van Hooser SD, Miller KD (2015) The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. Neuron 85: 402–417.

[230] Miller KD, MacKay DJC (1994) The Role of Constraints in Hebbian Learning. Neural Computation 6: 100–126.

[231] Miller KD (1996) Synaptic economics: competition and cooperation in synaptic plasticity. Neuron 17: 371–374.

[232] Toyoizumi T, Kaneko M, Stryker MP, Miller KD (2014) Modeling the dynamic interaction of Hebbian and homeostatic plasticity. Neuron 84: 497–510.

[233] Harnack D, Pelko M, Chaillet A, Chitour Y, van Rossum MC (2015) Stability of Neuronal Networks with Homeostatic Regulation. PLoS Comput Biol 11: e1004357.

[234] Sweeney Y, Hellgren Kotaleski J, Hennig MH (2015) A Diffusive Homeostatic Signal Maintains Neural Heterogeneity and Responsiveness in Cortical Networks. PLoS Comput Biol 11: e1004389.

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

[236] Gerstner W, Kistler WM (2002) Mathematical formulations of Hebbian learning. Biological Cybernetics 87: 404–415.

[237] Gerstner W, Kistler W (2002) Spiking Neuron Models - single neurons, populations, plasticity. Cambridge University Press.

[238] Pfister, Jean-Pascal (2006) Theory of non-linear spike-time-dependent plasticity. Ph.D. thesis, Swiss Federal Institute of Technology in Lausanne, EPFL.

[239] Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. Biometrika 58: 83–90.

[240] Marella S, Ermentrout GB (2008) Class-II neurons display a higher degree of stochastic synchronization than class-I neurons. Physical Review E 77: 041918.

[241] Abouzeid A, Ermentrout B (2009) Type-II phase resetting curve is optimal for stochastic synchrony. Physical Review E 80: 011911.

[242] Marella S, Ermentrout B (2010) Amplification of Asynchronous Inhibition-Mediated Synchronization by Feedback in Recurrent Networks. PLoS Computational Biology 6.

[243] Smeal RM, Ermentrout GB, White JA (2010) Phase-response curves and synchronized neural networks. Philosophical Transactions of the Royal Society B: Biological Sciences 365: 2407–2422.

[244] Zhou P, Burton SD, Urban NN, Ermentrout GB (2013) Impact of neuronal heterogeneity on correlated colored noise-induced synchronization. Frontiers in Computational Neuroscience 7.

[245] Goris RLT, Movshon JA, Simoncelli EP (2014) Partitioning neuronal variability. Nature Neuroscience 17: 858–865.

[246] Hildebrand EJ, Buice MA, Chow CC (2007) Kinetic Theory of Coupled Oscillators. Physical review letters 98: 054101.

[247] Buice MA, Chow CC (2007) Correlations, fluctuations, and stability of a finite-size network of coupled oscillators. Physical Review E 76: 031118.

[248] Buice MA, Chow CC, Cowan JD (2010) Systematic fluctuation expansion for neural network activity equations. Neural Comp 22.

[249] Buice MA, Chow CC (2013) Dynamic Finite Size Effects in Spiking Neural Networks. PLoS Comput Biol 9: e1002872.

[250] Ermentrout GB, Kopell N (1986) Parabolic Bursting in an Excitable System Coupled with a Slow Oscillation. SIAM J Appl Math 46: 233–253.

[251] Ermentrout B (1996) Type I Membranes, Phase Resetting Curves, and Synchrony. Neural Computation 8: 979–1001.

[252] Graupner M, Brunel N (2007) STDP in a Bistable Synapse Model Based on CaMKII and Associated Signaling Pathways. PLoS Comput Biol 3: e221.

[253] Higgins D, Graupner M, Brunel N (2014) Memory Maintenance in Synapses with Calcium-Based Plasticity in the Presence of Background Activity. PLoS Comput Biol 10: e1003834.

[254] Petersen CCH, Malenka RC, Nicoll RA, Hopfield JJ (1998) All-or-none potentiation at CA3-CA1 synapses. Proceedings of the National Academy of Sciences 95: 4732–4737.

[255] O'Connor DH, Wittenberg GM, Wang SSH (2005) Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proceedings of the National Academy of Sciences of the United States of America 102: 9679–9684.

[256] Boustani SE, Yger P, Frégnac Y, Destexhe A (2012) Stable Learning in Stochastic Network States. The Journal of Neuroscience 32: 194–214.

[257] Yger P, Harris KD (2013) The Convallis Rule for Unsupervised Learning in Cortical Networks. PLoS Comput Biol 9: e1003272.

[258] Abraham WC (2003) How long will long-term potentiation last? Philosophical Transactions of the Royal Society B: Biological Sciences 358: 735–744.

[259] Clopath C, Ziegler L, Vasilaki E, Buesing L, Gerstner W (2008) Tag-Trigger-Consolidation: A Model of Early and Late Long-Term-Potentiation and Depression. PLoS Comput Biol 4: e1000248.

[260] Barrett AB, Billings GO, Morris RGM, van Rossum MCW (2009) State Based Model of Long-Term Potentiation and Synaptic Tagging and Capture. PLoS Comput Biol 5: e1000259.

[261] Ziegler L, Zenke F, Kastner DB, Gerstner W (2015) Synaptic Consolidation: From Synapses to Behavioral Modeling. The Journal of Neuroscience 35: 1319–1334.

[262] Rosenblatt F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 65: 386–408.

[263] Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 79: 2554–2558.

[264] Hopfield JJ (1984) Neurons with graded response have collective computational properties like those of two-state neurons. Proceedings of the National Academy of Sciences 81: 3088–3092.

[265] Amit DJ, Gutfreund H, Sompolinsky H (1985) Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks. Physical Review Letters 55: 1530–1533.

[266] Amit DJ, Gutfreund H, Sompolinsky H (1985) Spin-glass models of neural networks. Physical Review A 32: 1007–1018.

[267] Tsodyks MV, Feigel'man MV (1988) The Enhanced Storage Capacity in Neural Networks with Low Activity Level. EPL (Europhysics Letters) 6: 101.

[268] Tsodyks M (1990) Associative memory in neural networks with binary synapses. Modern Physics Letters B 04: 713–716.

[269] Amit DJ, Fusi S (1993) Constraints on Learning in Dynamic Synapses. In: Gielen S, Kappen B, editors, ICANN '93, Springer London. pp. 730–733. URL http://link.springer.com/chapter/10.1007/978-1-4471-2063-6_203.

[270] Amit DJ, Fusi S (1994) Learning in Neural Networks with Material Synapses. Neural Comput 6: 957–982.

[271] Fusi S (2002) Hebbian spike-driven synaptic plasticity for learning patterns of mean firing rates. Biological Cybernetics 87: 459–470.

[272] Nadal JP, Toulouse G, Changeux JP, Dehaene S (1986) Networks of Formal Neurons and Memory Palimpsests. EPL (Europhysics Letters) 1: 535.

[273] Parisi G (1986) A memory which forgets. Journal of Physics A: Mathematical and General 19: L617.

[274] Fusi S, Drew PJ, Abbott LF (2005) Cascade Models of Synaptically Stored Memories. Neuron 45: 599–611.

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

[276] Lahiri S, Ganguli S (2013) A memory frontier for complex synapses. In: NIPS Proceedings. pp. 1034–1042. URL http://machinelearning.wustl.edu/mlpapers/papers/NIPS2013_4872.

[277] Huang Y, Amit Y (2010) Capacity analysis in multi-state synaptic models: a retrieval probability perspective. Journal of Computational Neuroscience 30: 699–720.

[278] Dubreuil AM, Amit Y, Brunel N (2014) Memory Capacity of Networks with Stochastic Binary Synapses. PLoS Comput Biol 10: e1003727.

[279] Clopath C, Nadal JP, Brunel N (2012) Storage of Correlated Patterns in Standard and Bistable Purkinje Cell Models. PLoS Comput Biol 8: e1002448.

[280] Senn W, Fusi S (2005) Learning Only When Necessary: Better Memories of Correlated Patterns in Networks with Bounded Synapses. Neural Computation 17: 2106–2138.

[281] Faisal AA, Selen LPJ, Wolpert DM (2008) Noise in the nervous system. Nature Reviews Neuroscience 9: 292–303.

[282] Richardson M (2009) Dynamics of neurons and networks of neurons with voltage-activatd and calcium-activated current. Phys Rev E 80: 021928.

[283] Abbott LF, van Vreeswijk C (1993) Asynchronous states in networks of pulse-coupled oscillators. Physical Review E 48: 1483–1490.

[284] Ginzburg I, Sompolinsky H (1994) Theory of correlations in stochastic neural networks. Physical Review E 50: 3171–3191.

[285] Fourcaud N, Brunel N (2002) Dynamics of the Firing Probability of Noisy Integrate-and-Fire Neurons. Neural Computation 14: 2057–2110.

[286] Risken H (1996) The Fokker-Planck equation: methods of solution and applications. Springer-Verlag, 3 edition.

[287] Rangan AV (2009) Diagrammatic Expansion of Pulse-Coupled Network Dynamics. Physical Review Letters 102: 158101.

[288] Rangan AV (2009) Diagrammatic expansion of pulse-coupled network dynamics in terms of subnetworks. Physical Review E 80: 036101.