Relationship between neuronal architecture and variability in cortical circuits

THESIS

Submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Neural Computation

Ashok Litwin-Kumar

B.S., Physics and Philosophy (Minor), California Institute of Technology

Carnegie Mellon University

Pittsburgh, PA

December 2013

Abstract

The connectivity of cortical neuronal networks is complex, exhibiting clustered network motifs and ensembles of neurons with high connection probability. However, the significance of these connectivity properties for computation and dynamics in cortex is unclear. In this thesis, I present several studies concerning the behavior of model cortical neurons receiving input from a surrounding network. I begin by studying pairs of neurons, investigating how overlapping excitatory and inhibitory inputs control the statistics of their outputs. I then study fully recurrent networks of neurons with nonuniform connection structures in the form of highly connected neuronal assemblies. These assemblies represent functionally related subsets of neurons, and I investigate their collective behavior in both spontaneously generated activity and evoked conditions. I show that the presence of assembly structure in recurrently coupled, balanced excitatory-inhibitory networks introduces slow timescales in the networks' dynamics and relate these modeling results to the experimental literature. Next, I present results on how these assemblies form and are maintained with realistic models of synaptic plasticity. In total, these results represent a step toward understanding how connectivity can be modified by sensory experience, and how these changes in turn shape cortical dynamics.

Acknowledgements

I am indebted to many people without whom completing this thesis would have been difficult or impossible. To begin, I would like to acknowledge the committee members who have provided guidance throughout this thesis: Drs. Brent Doiron (chair), Byron Yu, Marlene Cohen, Bard Ermentrout, Rob Kass, and Nicolas Brunel. I will also acknowledge funding from the NDSEG program and Richard King Mellon Foundation during my graduate studies.

During my education, I have often relied on my family. My parents are responsible for my ability to pursue science and for seeding my desire to do so. I regret that both are not able to see the completion of this thesis. My siblings share with me the curiosity our parents taught us and probably understand my motivations better than anyone else.

There are many friends to thank as well. Michael, Chris, Natasha, and Brandi have been friends since my time as an undergraduate, and I regularly rely on our common experience when I need advice. Laura, Deren, Annalisa, and Silver are responsible for some of my fondest memories of Pittsburgh.

Carolina has taught me a great deal of introspection, from which I have benefited as a student and as a person. Knowing she is involved magnifies my satisfaction from completing this thesis and the anticipation of my future afterward.

For organization's sake I will distinguish friends and colleagues, but all my colleagues are friends. Robert has been a wonderful resource, and conversations with him have led to many new ideas. The same is true for Zack, Guillaume, Gabe, and Tatjana. I learned many things because of collaborations with experimental colleagues: Nathan, Anne-Marie, and Maurice.

I have been lucky that my advisor has directed me closely and carefully, and who has been a friend throughout my studies. I have learned many things and shared many conversations over drinks with Brent. I have a hard time imagining a better situation for my graduate studies.

Contents

List of Figures

List of Tables

1. *Introduction*

The responses of cortical neurons to repeated stimulus presentations are irregular [1, 2], with spikes occurring at different times on each trial. What are the sources of this variability? This thesis focuses on network sources of collective fluctuations in neuronal activity, and in this introductory chapter I will describe prior work on this topic. I begin by providing an overview on the sources of neuronal variability, arguing that noise from bombardment by presynaptic action potentials is a major contributor to the irregularity of neuronal responses. I then discuss various network models of neuronal variability, culminating with the idea of the *balanced network* as a starting point for examining noise in cortical networks.

1.1 Intrinsic versus synaptic sources of variability

Cortical neurons receive inputs from thousands of other neurons, but they also have complex internal dynamics. These internal dynamics include stochastically opening ion channels, multiple ionic currents, and various molecular signaling cascades. Each of these processes is potentially noisy and may contribute to irregularity in output spiking. It is therefore important to estimate what proportion of spiking irregularity can be attributed to intrinsic versus network processes.

One source of noise is thermal noise, due to which the voltage across a resistor fluctuates, a phenomenon known as *Johnson noise*. For a resistor of resistance R, the Johnson noise causes fluctuations in voltage given by $\langle V^2 \rangle = 4k_B TRB$, where k_B is the Boltzmann constant, T is temperature, and B is the bandwidth over which one measures the noise [3]. Manwani & Koch [3] performed an analysis of the voltage fluctuations in a patch of membrane for a cortical pyramidal neuron, finding that fluctuations due to Johnson noise are negligible compared to those due to ionic or synaptic sources, so we focus on those other sources.

We next turn to the case of channel noise. Current flow across a neuron's membrane is due to charges flowing through permeable ion channels. These channels have some probability of being open or closed that depends on variables such as membrane voltage or the presence of bound ligands. In classic modeling work like that of Hodgkin & Huxley [4], the number of channels was assumed to be large enough that finite size fluctuations could be neglected, but in actual neurons the stochastic opening and closing of channels may contribute to response variability.

To directly assess the effect of ion channel stochasticity on neuronal firing, it is possible to study the responses of isolated neurons driven by current injections, which removes network fluctuations and synaptic effects from the response. Instead, ionic currents due to action potential generation will contribute to the majority of noise in spiking output. The spike trains generated under these conditions can then be analyzed and statistics compared to those produced *in vivo*. A simple way to quantify the *temporal* irregularity of spike trains is the coefficient of variation (CV) of the inter-spike interval (ISI) distribution (for a reference on the spike train statistics most commonly used in this thesis, see Appendix A):

$$\mathrm{CV}_{\mathrm{ISI}} = \frac{\sqrt{\langle (t_{i+1} - t_i)^2 \rangle - \langle t_{i+1} - t_i \rangle^2}}{\langle t_{i+1} - t_i \rangle} \tag{1.1}$$

where $\{t_i\}$ are the neuron's spike times. For a Poisson process with constant rate, $\mathrm{CV}_{\mathrm{ISI}}$ is 1. Cortical responses to visual stimuli have $\mathrm{CV}_{\mathrm{ISI}}$ values near 1, indicating irregular activity.

Another important view of variability is *trial-to-trial* variability, which focuses on the irregularity of responses to repeated stimulus presentations. A commonly used metric is the *Fano factor*, which measures the ratio of variance to mean of the number of spikes N_i^T emitted by neuron i in an interval of time T aligned with respect to stimulus onset:

$$\mathrm{FF}(i, T) = \frac{\mathrm{Var}\left(N_i^T\right)}{\langle N_i^T \rangle}. \tag{1.2}$$

For homogeneous Poisson processes, the Fano factor is 1. Observations from many cortical regions have demonstrated Fano factors near or above 1 with repeated stimulus presentations [1, 2].

Mainen & Sejnowski [5] studied isolated neurons in rat cortical slices driven by dc and fluctuating current injection. They found that the first spike in response to dc inputs was tightly locked (with a standard deviation of less than 1 ms across trials). Furthermore, CV_{ISI} was low on each trial, reflecting the fact that neurons fired in a temporally regular manner. Trial-to-trial spiking became less precise at later points in the trial, due to gradual desynchronization of the responses. The authors concluded that dc pulses produce reliable firing rates and regular firing. Temporally varying input currents produced, in addition, precise trial-to-trial spike timing. These results were reproduced in a Hodgkin-Huxley model with ion channels modeled stochastically [6]. In total, the results argue that, while some trial-to-trial variability in response to current injections, particularly dc inputs, may be attributed to non-network sources, isolated neurons do not have temporal variability consistent with *in vivo* recordings, as measured by CV_{ISI} [5, 7].

In addition to ion channel stochasticity, other sources of variability that are not due to network effects, such as fluctuations in intracellular and extracellular ion concentrations or neuromodulators, or the action of ion pumps, may also contribute to response variability. However, the above results argues that, even in the face of these local sources of noise, neurons are capable of producing regular responses on their own. Hence, it is necessary to examine the input from the surrounding network to understand neuronal irregularity.

1.2 Experimental observations of cortical spiking co-variability

Before discussing network models of variability, we review further experimental observations of spiking variability in cortex. In addition to the Fano factor, which depends on the spikes of a single neuron, trial-to-trial variability can also be characterized for multiple neurons. At the pairwise level, the correlation between the spike counts of neurons i and j can be quantified as:

$$\text{Corr}_{\text{signal}}(i,j,T) = \sum_{s\in\text{Stimuli}} \sum_{t\in\text{Trials}} \frac{\text{Cov}(N_i^T(s,t), N_j^T(s,t))}{\sqrt{\text{Var}(N_i^T(s,t))\text{Var}(N_i^T(s,t))}}. \tag{1.3}$$

The above quantity, known as *signal correlation*, measures the correlation between the responses of two neurons to multiple stimulus presentations. However, some of this correlation may simply be due to the fact that the neurons respond to the same stimuli, not that they are correlated

trial-to-trial because of common input, for example. To quantify this second type of correlation, the *noise correlation* is defined as:

$$\text{Corr}_{\text{noise}}(i,j,s,T) = \sum_{t \in \text{Trials}} \frac{\text{Cov}(N_i^T(s,t), N_j^T(s,t))}{\sqrt{\text{Var}(N_i^T(s,t))\text{Var}(N_i^T(s,t))}}. \tag{1.4}$$

Note that noise correlation depends on s, the stimulus condition. More generally, noise correlations can also be defined for spontaneous conditions, in which there is no stimulus and all correlations are due to shared input or coupling.

Correlations have been measured in many cortical regions, with visual cortex being the most studied. Most studies report noise correlations with a magnitude of approximately 0.1 on average [8], although some recent results argue that correlations are near zero [9]. These discrepancies may be attributed to a variety of factors. One is firing rate, as higher firing rates promote correlations by mitigating the decorrelating effect of the threshold nonlinearity [10]. Correlations also depend on the layer from which pairs are recorded [11]. Internal brain state fluctuations (which can be modulated by alertness, anesthetics, or other variables) can influence trial-to-trial covariability, as may methodological differences in correlation measurement such as spike sorting or choice of counting window, which increases correlation with T [8].

Recently, modulation of trial-to-trial variability with stimulus condition has attracted interest. Stimulus conditions [12, 13], adaptation [14], and attentional state [15, 16], all influence noise correlations. The effects of attention have been of particular interest, with results showing both reduced Fano factor and correlations in attended compared to unattended conditions [15, 16]. Evoked activity also exhibits a reduction in Fano factor and noise correlation compared to spontaneous activity in many cortical regions [17], a topic that has generated theoretical interest and will be discussed in later chapters.

1.3 Rate and spiking descriptions of neuronal variability

Several theories of the network source of fluctuations in neuronal activity have been proposed, with different levels of detail in their description. Early models considered only the firing rates of units in the network, making interpretations in terms of CV_{ISI}, for example, difficult. Later models examined full spiking networks.

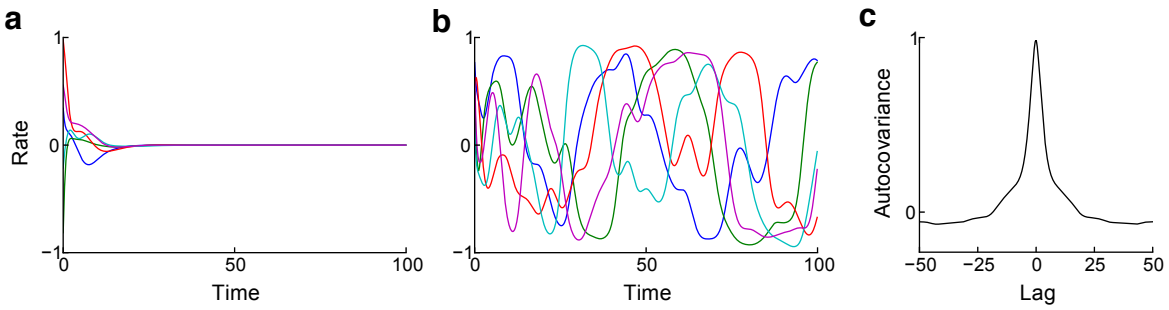

Figure 1.1: Simulations of equation (1.5) for **(a)** $N = 800$ and $g = 0.9$ and **(b)** $g = 1.5$. The trajectories of 5 units are shown. For $g < 1$, the system has a globally stable fixed point at the origin. For $g > 1$, trajectories exhibit chaos. **(c)** Normalized autocovariance function for units in a network with $g = 5$ and $N = 200$.

1.3.1 Chaotic firing rate networks

An early and influential model of neuronal activity fluctuations was proposed by Sompolinsky et al. [18]. In this model, the variables of interest are the activations of N rate-like units, each of whose activity is given by s_i. Each unit linearly sums its inputs according to some weight matrix J_{ij} and produces an output according to a sigmoidal nonlinearity along with an intrinsic decay, leading to the system of ordinary differential equations:

$$\frac{ds_i(t)}{dt} = -s_i(t) + \tanh\left(\sum_{j=1}^{N} J_{ij}s_j(t)\right). \tag{1.5}$$

Each entry of matrix J_{ij} is Gaussian distributed with a mean of 0 and a variance of g^2/N. The model thus features two common features of biological neuronal networks: nonlinearity in activation and disorder in connections.

Sompolinsky et al. [18] developed a dynamic mean field theory for (1.5), leading to an elegant calculation of the ensemble-averaged temporal autocorrelation of the activity $\langle s(t)s(t+\tau)\rangle$. This calculation led to a surprising result about the dependence of the activity on the weight matrix in the large system-size limit. When $g < 1$, the system has only one stable fixed point at the origin, and all trajectories eventually lead to it (Figure 1.1a). Above $g = 1$, however, the system exhibits extensive chaos, leading to irregular trajectories (Figure 1.1b). In the chaotic regime, trajectories have an autocorrelation time longer than that of an individual unit (Figure 1.1c).

This model has proven highly influential, both as a model of irregularity in neuronal networks and as a framework in which to study reservoir computing [19, 20]. However, it models neural activity only at the level of firing rates, not at the level of spiking. Whether or not it is a good model for spiking networks is unresolved.

1.3.2 How can threshold neurons generate irregular spikes?

We next turn to spiking neuron, which have two essential ingredients: a *threshold* in voltage above which spikes are likely to occur and *temporal integration* of inputs over a timescale specified by neuronal properties. With simplifying assumptions, the problem of understanding variability in spike output can be reduced to understanding the *first passage time* problem presented by the voltage integrating to threshold.

The simplest model of a neuron with a spiking threshold and an integration time constant τ is the *leaky integrate-and-fire* model, whose voltage obeys:

$$\frac{dV}{dt} = -V/\tau + I(t). \tag{1.6}$$

When V reaches 1, its threshold, the neuron emits a spike and V is instantaneously reset to 0. For this simple example, we neglect refractory effects. The input current $I(t)$ is given by the sum of the synaptic currents due to the firing of neurons presynaptic to the integrate-and-fire neuron. We take $I(t)$ to be a sum of instantaneous currents, so that:

$$I(t) = \sum_k J\delta(t - t_k^E) + \sum_k J\delta(t - t_k^I), \tag{1.7}$$

where δ is the Dirac delta function. In this simplified model, V is increased (decreased) by an amount J, whenever a presynaptic excitatory (inhibitory) input arrives. The times of these arrivals are given by $\{t_k^E\}$ for excitation and $\{t_k^I\}$ for inhibition. We assume the arrival times are given by Poisson processes with rates r^E for excitation and r^I for inhibition.

Softky & Koch [21] argued that integration of small excitatory postsynaptic potentials (EPSPs) without inhibition could not produce highly irregular output firing rates. First consider the limit of $\tau \to \infty$, corresponding to a perfect integrator. In this case, it takes $N = 1/J$ synaptic inputs to produce an output. Assuming these inputs come from a Poisson process with rate

r^E, the expected time between spikes is given by $\langle \text{ISI} \rangle = \frac{N}{r^E}$. Furthermore, the ISI variance is $\text{Var}(\text{ISI}) = \frac{N}{(r^E)^2}$, where we have used the fact that the time between synaptic inputs is given by an i.i.d. exponential random variable with variance $\frac{1}{(r^E)^2}$. Consequently, we find $\text{CV}_{\text{ISI}} = 1/\sqrt{N}$, so that if many Poisson inputs are integrated to produce a spike, CV_{ISI} is low.

There are several possible solutions to this problem, and we outline the discussion presented by Shadlen & Newsome [22]. First, neurons are not perfect integrators, meaning that the argument above is valid only in the limit of high firing rates, when the membrane time constant can be neglected. We simulated a more realistic situation in which $J = 0.05$ and $\tau = 10$ ms (Figure 1.2a). For an output firing rate of 20 Hz, $\text{CV}_{\text{ISI}} = 0.61$, which is still substantially lower than cortical neurons.

Two possible solutions are immediately apparent: J can be increased or τ can be decreased. For $J = 0.5$, $\text{CV}_{\text{ISI}} = 0.88$ at a firing rate of 20 Hz (Figure 1.2b). However, it is unlikely that on average only two inputs are needed to make a neuron spike. The fact that neurons make hundreds or thousands of contacts with presynaptic neurons, and the fact that those neurons exhibit spontaneous activity, would suggest that such high values of J would lead to unrealistically high firing rates.

If the time constant τ is decreased to 2 ms, the neuron becomes a coincidence detector, and $\text{CV}_{\text{ISI}} = 0.89$ at a firing rate of 20 Hz (Figure 1.2c). However, this time constant is substantially lower than estimates for cortical neurons even in high conductance states [23].

If we are forced to use realistic J and τ, how can integrate-and-fire neurons produce irregular spiking? Shadlen & Newsome [22] suggested that the solution is integration of inhibition. If neurons receive both positive and negative inputs of approximately equal magnitude, their voltages will exhibit dynamics similar to a random walk, so that the crossing times at the threshold will be highly variable. Indeed, for the neuron model above, excitation and inhibition of approximately equal magnitude lead to $\text{CV}_{\text{ISI}} = 0.95$ with realistic J and τ (Figure 1.2d).

The above results suggest *balanced* excitation and inhibition as a candidate mechanism for irregular cortical firing. However, how such a precise balance can be maintained is unclear. In later work, van Vreeswijk & Sompolinsky [24, 25] described how this balance occurs generically in large recurrent networks.

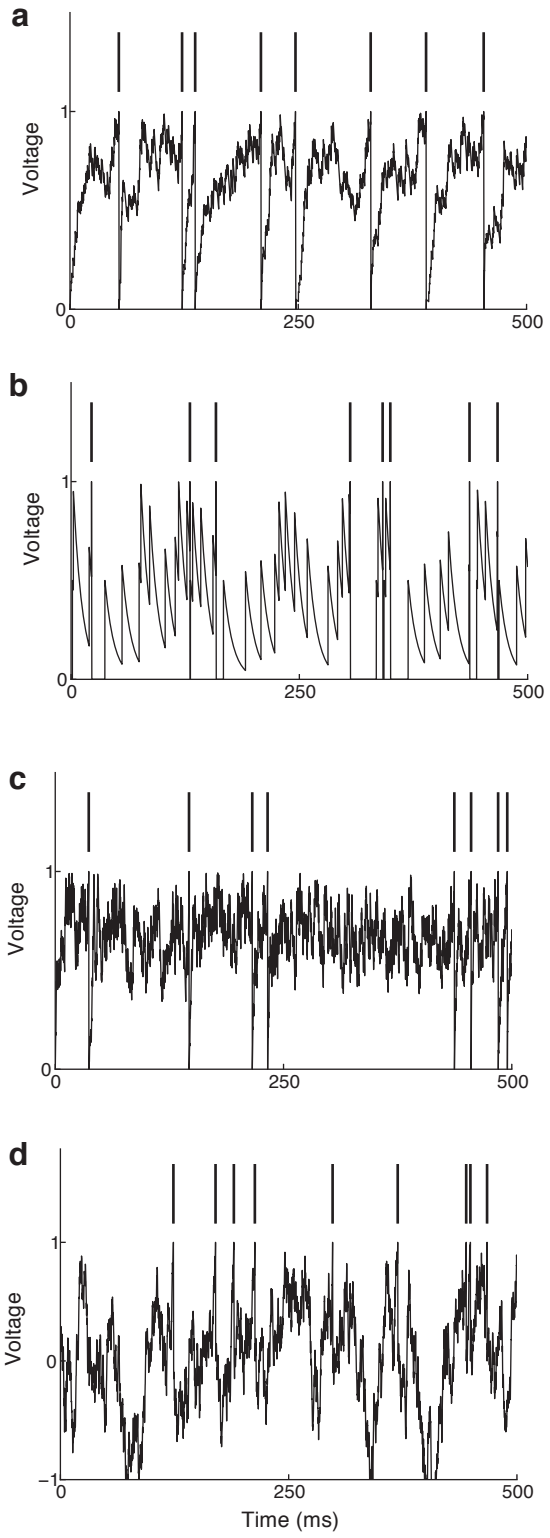

Figure 1.2: Simulations of leaky integrate-and-fire neuron defined by equation (1.6). In each case, the firing rate was approximately 20 Hz. Vertical bars above the voltage traces correspond to action potentials.
(a) $\tau = 10$ ms, $J = 0.05$, $r^E = 1.7$ kHz, $r^I = 0$ kHz. $\mathrm{CV}_{\mathrm{ISI}}$ was 0.61.
(b) $\tau = 10$ ms, $J = 0.5$, $r^E = 0.1$ kHz, $r^I = 0$ kHz. $\mathrm{CV}_{\mathrm{ISI}}$ was 0.88.
(c) $\tau = 2$ ms, $J = 0.05$, $r^E = 6.75$ kHz, $r^I = 0$ kHz. $\mathrm{CV}_{\mathrm{ISI}}$ was 0.89.
(d) $\tau = 10$ ms, $J = 0.05$, $r^E = 10.5$ kHz, $r^I = 10$ kHz. $\mathrm{CV}_{\mathrm{ISI}}$ was 0.95.

1.4 Balanced spiking networks

1.4.1 Derivation of the balance conditions

The balanced network of van Vreeswijk & Sompolinsky [24, 25] describes irregularity in networks of randomly coupled excitatory and inhibitory neurons. In the simplest formulation of the model, neuron i is described by the binary variable s_i, which is a simple threshold function of its inputs:

$$s_i(t) = H\left(\mu + \sum_{i=1}^{N} J_{ij} s_j(t) - \theta\right), \tag{1.8}$$

where $H(\cdot)$ is the Heaviside step function, μ is a constant bias to each neuron, and θ is the threshold. Neurons are partitioned into excitatory and inhibitory subpopulations, which we will denote by superscript E and I, respectively. Each neuron receives input from K excitatory and K inhibitory neurons.

A central aspect of balanced networks is the scaling of the strength of synaptic weights with the number of connections, K (if $K \propto N$, as is the case in studies of dense local networks, this is the same as the scaling with N). In studies of large networks, it is often assumed that connection strengths scale inversely with K, that is, $J \propto 1/K$. This ensures that the mean of the presynaptic input from population Y to population X, $I^{XY} \propto KJ$, remains order 1 as K approaches infinity. However, this necessarily implies, if each presynaptic input is uncorrelated, that $\text{Var}(I^{XY}) \propto K(1/K)^2 = 1/K$. Consequently, the variance of the input approaches zero for large K, and neurons fail to exhibit fluctuations.

In balanced networks, this is solved by assuming that the strength of synaptic weights from population Y to population X are given by $J^{XY} = \bar{J}^{XY}/\sqrt{K}$, where $X, Y \in \{E, I\}$ and $\bar{J}^{XY} = O(1)$ is a constant. In this case, the variance issue is solved, as $\text{Var}(I^{XY}) \propto K(1/\sqrt{K})^2 = O(1)$. A new apparent problem is introduced, however, since $\langle I^{XY} \rangle \propto K(1/\sqrt{K}) = \sqrt{K}$. In other words, the synaptic input from one population to another diverges with K. This problem is solved by assuming that the synaptic inputs from different populations (plus the mean neuronal bias μ) exactly cancel. We write the mean input to neurons in excitatory and inhibitory

subpopulations as:

$$\langle I^E \rangle = \mu^E + KJ^{EE}\langle s^E \rangle - KJ^{EI}\langle s^I \rangle = \sqrt{K}\left(\bar{\mu}^E + \bar{J}^{EE}\langle s^E \rangle - \bar{J}^{EI}\langle s^I \rangle\right) \tag{1.9}$$

$$\langle I^I \rangle = \mu^I + KJ^{IE}\langle s^E \rangle - KJ^{II}\langle s^I \rangle = \sqrt{K}\left(\bar{\mu}^I + \bar{J}^{IE}\langle s^E \rangle - \bar{J}^{II}\langle s^I \rangle\right), \tag{1.10}$$

where we have assumed that μ also scales as $\sqrt{K}$ and have thus defined $\bar{\mu} = \mu/\sqrt{K}$. In order for these equations to have a solution that is stable as K grows, we must have:

$$0 = \bar{\mu}^E + \bar{J}^{EE}\langle s^E \rangle - \bar{J}^{EI}\langle s^I \rangle \tag{1.11}$$

$$0 = \bar{\mu}^I + \bar{J}^{IE}\langle s^E \rangle - \bar{J}^{II}\langle s^I \rangle. \tag{1.12}$$

This immediately leads to a solution for $\langle s^E \rangle$ and $\langle s^I \rangle$, without any assumption about their dynamics. Without loss of generality, we consider the case of $J^{EE} = J^{IE} = 1$, which can be obtained by rescaling the thresholds and other weights. This leads to the following solution:

$$\langle s^E \rangle = \frac{\bar{J}^{II}\bar{\mu}^E - \bar{J}^{EI}\bar{\mu}^I}{\bar{J}^{EI} - \bar{J}^{II}} \tag{1.13}$$

$$\langle s^I \rangle = \frac{\bar{\mu}^E - \bar{\mu}^I}{\bar{J}^{EI} - \bar{J}^{II}}. \tag{1.14}$$

For $\langle s^E \rangle$ and $\langle s^I \rangle$ to be positive, we need one of the following to be true:

$$\bar{J}^{EI} > \bar{J}^{II},\ \frac{\bar{\mu}^E}{\bar{\mu}^I} > \frac{\bar{J}^{EI}}{\bar{J}^{II}} \tag{1.15}$$

$$\bar{J}^{EI} < \bar{J}^{II},\ \frac{\bar{\mu}^E}{\bar{\mu}^I} < \frac{\bar{J}^{EI}}{\bar{J}^{II}}. \tag{1.16}$$

The second solution can be rejected because it allows for an imbalanced solution in which $\langle I^E \rangle \to -\infty$ implying $\langle s^E \rangle = 0$. To see this, suppose $\langle s^E \rangle = 0$. Then by forcing equation (1.12) to be $O(1)$, we find $\langle s^I \rangle = \frac{\bar{\mu}^I}{\bar{J}^{II}}$. Indeed, in this case, $\langle I^E \rangle = \sqrt{K}\left(\bar{\mu}^E - \bar{J}^{EI}\frac{\bar{\mu}^I}{\bar{J}^{II}}\right) \to -\infty$ by the assumptions of equation (1.16). We therefore obtain equation (1.15) as the unique requirement for balanced activity. As long as these conditions are satisfied, a balanced solution will exist [25].

1.4.2 Dynamics in the balanced state

When equation (1.15) is satisfied, balanced networks produce irregularity in the spike times of each neuron (Figure 1.3a). This is reflected in CV_{ISI} (spikes here are defined as transitions from $s_i = 0$ to 1), which is 0.83 on average for the network of Figure 1.3. The excitatory and inhibitory components of the input current to any neuron are equal in magnitude, so the summed current is near zero and crosses threshold only occasionally (Figure 1.3b). This mechanism is identical to that of Figure 1.2d, but the network dynamics ensure that parameters do not need to be fine tuned for the currents to exhibit balance.

Another feature of the balanced state is a fast response to inputs. This can be understood by examining the dynamics of the mean rates of the two subpopulations in the mean field limit [25, 26]:

$$\dot{m}^X(t) = -m^X(t) + \langle H(I^X(t) - \theta)\rangle, \tag{1.17}$$

where $m^X(t)$ is the time-dependent firing rate of population X and $\langle H(I^X - \theta)\rangle$ represents an average over the distribution of input currents to neurons in the population. Assuming that the input current is Gaussian distributed, the second term can be expressed using a complementary error function:

$$\langle H(I^X - \theta)\rangle = \text{erfc}\left(\frac{\sqrt{K}\left(\bar{J}^{XE}m^E(t) + \bar{J}^{XI}m^I(t)\right)}{(\bar{J}^{XE})^2 m^E(t) + (\bar{J}^{XI})^2 m^I(t)}\right). \tag{1.18}$$

It is clear that derivatives of equation (1.17) with respect to $m^X(t)$ will be proportial to $\sqrt{K}$ because of the numerator of equation (1.18). In other words, perturbations recruit feedback of order $\sqrt{K}$. As a result, the magnitudes of the eigenvalues increase with K. Thus, for large systems, the system responds very quickly to input perturbations.

These fast dynamics are evident in the autocorrelation times of individual units (Figure 1.3c). Neurons in balanced networks behave like Poisson processes, which are memoryless. This can be contrasted with chaotic rate networks (Figure 1.1c), whose trajectories have autocorrelations longer than that of an individual unit. This distinction will be returned to in Chapter 3.

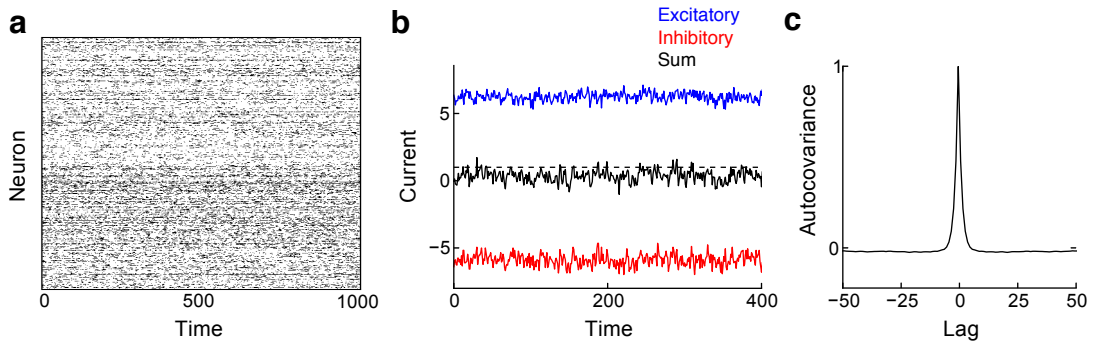

Figure 1.3: **(a)** Spike raster for a network with $N = 2000$ and $K = 500$. 250 excitatory and 250 inhibitory neurons are shown. **(b)** Input currents to one neuron in the network, compared to the threshold at $\theta = 1$ (dashed line). **(c)** Normalized autocovariance function for excitatory neurons in the network (compare with Figure 1.1c). Other parameters are $\theta = 1$, $\bar{\mu}^E = 0.15$, $\bar{\mu}^I = 0.1$, $\bar{J}^{EE} = \bar{J}^{IE} = \bar{J}^{II} = 1$, $\bar{J}^{EI} = 1.3$.

1.4.3 Asynchrony in densely connected balanced networks

A puzzling feature of Figure 1.3 is the lack of synchrony in the population. In fact, the distribution of correlation coefficients between s_i and s_j for 5000 randomly sampled excitatory neuron pairs in the network is centered at zero (0.0002 ± 0.04). An assumption of independence was necessary to obtain expressions for the variance of the input currents and to evaluate the dynamics of $m^X(t)$ in the mean field limit (equation (1.17)). But this assumption seems poorly motivated in the network of Figure 1.3, since, on average, neuron pairs shared 25% of their inputs. In van Vreeswijk & Sompolinsky [25], it was assumed that $K \ll N$ so that this overlap was negligible, thus guaranteeing independence of input currents. Why does the network continue to produce asynchronous activity even with high levels of presynaptic overlap?

This mystery was addressed by Renart et al. [27], who showed that balanced networks produce asynchronous activity even when they are densely connected. They demonstrated that, while neurons indeed receive correlated input currents due to shared presynaptic sources, shared excitatory and inhibitory fluctuations cancel. This can be seen by calculating separately the correlation of the excitatory and inhibitory input currents (which we define as $I_i^{XE} = \sum_j J_{ij}^{XE} s_j^E$ and $I_i^{XI} = \sum_j J_{ij}^{XI} s_j^I$, so that $I^X = \mu^X + I^{XE} + I^{XI}$), and comparing to the summed input current correlation. We have that:

$$\text{Corr}(I_i^X, I_j^X) = \text{Corr}(I_i^{XE}, I_j^{XE}) + \text{Corr}(I_i^{XI}, I_j^{XI}) + 2\text{Corr}(I_i^{XE}, I_j^{XI}). \tag{1.19}$$

The last term in this expression will be negative if the two currents are anticorrelated, thus canceling the correlating influence of common excitatory or inhibitory currents.

Examining the excitatory subpopulation of the network of Figure 1.3, Corr(I_i^{EE}, I_j^{EE}) was 0.56 ± 0.03, while Corr(I_i^{EI}, I_j^{EI}) was 0.43 ± 0.04. Thus, the components of the input current to neuron pairs were highly correlated. But Corr(I_i^E, I_j^E) was 0.10 ± 0.05, far lower than either of the two components that formed it (additional decorrelation due to spiking nonlinearities accounts for the difference between this correlation and Corr(s_i^E, s_j^E) [10]). This was because Corr(I_i^{EE}, I_j^{EI}) was negative, -0.44 ± 0.03.

Renart et al. [27] showed that this dynamic cancellation of current fluctuations is a consequence of the balanced state, and that correlations decrease with N. We will give a brief sketch of this result that demonstrates why a cancellation of correlations leads to a self-consistent solution. This derivation does not show that this solution is stable, only that it is possible.

We will start with the expression $\langle I_i^X s_k^Z \rangle$, the temporal correlation between the input current to neuron i in population X, and the activity of neuron k in population Z (we will abuse notation and assume that averages written as $\langle a_1 a_2 \rangle$ correspond to the covariance of a_1 and a_2; that is, the random variables are mean-subtracted). We can then write:

$$\begin{aligned}\langle I_i^X s_k^Z \rangle &= \sum_{j,Y} J_{ij}^{XY} \langle s_j^Y s_k^Z \rangle \\ &= J_{ik}^{XZ} \langle s_k^Z s_k^Z \rangle + \sum_{(j,E)\neq(k,Z)} J_{ij}^{XE} \langle s_j^E s_k^Z \rangle + \sum_{(j,I)\neq(k,Z)} J_{ij}^{XI} \langle s_j^I s_k^Z \rangle. \end{aligned} \tag{1.20}$$

The first term on the last line of the above equation represents the influence of a direct connection from neuron k in population Z to the postsynaptic neuron. The next terms represents the fact that, if s_j^Y and s_k^Z are correlated, then connections from the former to the postsynaptic neuron induce correlations between the postsynaptic neuron and s_k^Z.

We next derive a self-consistency condition for equation (1.20). We assume that activity-activity correlations are linearly related to current-activity correlations:

$$\langle s_i^X s_k^Z \rangle \approx A \langle I_i^X s_k^Z \rangle, \tag{1.21}$$

for some constant A. This is a *linear response assumption*, similar to the one we will use in Chapter 2. This allows us to write:

$$\langle s_i^X s_k^Z \rangle \propto J_{ik}^{XZ} \langle s_k^Z s_k^Z \rangle + \sum_{(j,E)\neq(k,Z)} J_{ij}^{XE} \langle s_j^E s_k^Z \rangle + \sum_{(j,I)\neq(k,Z)} J_{ij}^{XI} \langle s_j^I s_k^Z \rangle. \tag{1.22}$$

Now, we show that a self-consistent asynchronous state requires cancellation of correlations. In an asynchronous state, we require that $\langle s_i^X s_j^Y \rangle = O(1/N)$. If this were not the case, then the variance of the input current to a neuron would be dominated by covariances:

$$\langle I_i^X I_i^X \rangle = \sum_{j,Y} (J_{ij}^{XY})^2 \langle s_j^Y s_j^Y \rangle + \sum_{j,k,X,Y} J_{ij}^{XY} J_{ik}^{XZ} \langle s_j^Y s_k^Z \rangle. \tag{1.23}$$

The first term is $O(1)$, and the second term is $O(1)$ if $\langle s_i^X s_j^Y \rangle = O(1/N)$.

Hence, assuming asynchrony and using the linear response assumption (1.21), the l.h.s. of equation (1.20) is $O(1/N)$. The first term on the r.h.s. is $O(1/\sqrt{N})$, since $J = O(1/\sqrt{N})$ and $\langle s_k^Z s_k^Z \rangle = O(1)$. The other terms are a sum of N terms, each of which contains $J = O(1/\sqrt{N})$ and $\langle s_j^I s_k^Z \rangle = O(1/N)$. Hence, these terms are also $O(1/\sqrt{N})$. In other words, the l.h.s. of the equation is $O(1/N)$ while the r.h.s. is $O(1/\sqrt{N})$, unless individual terms on the r.h.s. cancel. This is the conclusion of Renart et al. [27], and demonstrates that a self-consistent solution can occur if the appropriate terms cancel.

1.4.4 Experimental evidence for the balanced state

Several studies have tested predictions of balanced network models, indicating that many cortical networks are in fact balanced. *In vitro* intracellular recordings of ferret prefrontal and occipital cortices found that neurons underwent transitions between periods of activity and quiescence (termed "UP" and "DOWN" states, respectively) [28]. During UP states, recurrent activity is generated by balanced barrages of EPSPs and IPSPs. This result was subsequently verified in anesthetized, *in vivo* conditions [29].

Tests of the stronger hypothesis that excitatory and inhibitory inputs balance not only in magnitude but are also temporally anticorrelated [27] have also been made. Indirectly, the lack of very synchronous activity in nearby cortical neurons that likely share many connections has been taken as evidence for the decorrelating effect of balanced inputs [9]. However, the

extent of this decorrelation is a subject of debate [8]. More direct evidence comes from intracellular measurements that indicate excitatory and inhibitory currents are indeed temporally anticorrelated [30].

In summary, balanced activity has been observed in many cortical regions, and much work on the statistics of cortical responses is influenced by the idea of co-varying excitatory and inhibitory currents [31, 32, 33]. Nonetheless, there is debate over its computational role [34]. Theoretical approaches are necessary to understand what types of computations such networks can support.

1.5 Outline of this thesis

In this thesis, I will present the results of several studies, some of which have been published. All of the results presented will involve the study of trial-to-trial variability in spiking neurons that receive balanced input. In Chapter 2, I study neuron pairs in isolation and assume that they receive balanced input from external sources. I then investigate how the statistics of the spike trains they produce depend on the input they receive. This work has been published previously [35] and involved a collaboration between myself, my advisor Dr. Brent Doiron, Dr. Anne-Marie Oswald, and Dr. Nathan Urban. In Chapter 3, I consider a fully recurrent network of spiking neurons in the balanced state, and investigate how its dynamics depend on the structure of its connections. Unlike Chapter 3, I therefore no longer need to assume any form for the statistics of a particular neuron's input. This work has also been published previously [36] and was done in collaboration with my advisor. In Chapter 4, I present results on balanced networks with plasticity and the emergence of the network structures that were assumed in Chapter 3. This work was done in collaboration with my advisor. Finally, in Chapter 5, I present a simplified model of interacting neuronal populations that qualitatively reproduces many of the features of the spiking networks studied in the previous chapters. This work was done in collaboration with an undergraduate student, Aubrey Thompson, and my advisor.

2. *Balanced synaptic input shapes the correlation between neural spike trains*

2.1 Abstract

Stimulus properties, attention, and behavioral context influence correlations between the spike times produced by a pair of neurons. However, the biophysical mechanisms that modulate these correlations are poorly understood. With a combined theoretical and experimental approach, we show that the rate of balanced excitatory and inhibitory synaptic input modulates the magnitude and timescale of pairwise spike train correlation. High rate synaptic inputs promote spike time synchrony rather than long timescale spike rate correlations, while low rate synaptic inputs produce opposite results. This correlation shaping is due to a combination of enhanced high frequency input transfer and reduced firing rate gain in the high input rate state compared to the low state. Our study extends neural modulation from single neuron responses to population activity, a necessary step in understanding how the dynamics and processing of neural activity change across distinct brain states.

2.2 Introduction

Correlations between the spike trains of neuron pairs are observed throughout the central nervous system [8]. The correlation between a pair of neurons' spike trains can change depending on the state of their neural circuit. For instance, correlated neural activity is altered by stimulus properties [12, 37], anesthetics [38, 39], stimulus adaptation [14], focus of spatial attention [15, 16], and the behavioral context of a task [40]. The level of spike train correlation between neuron pairs has implications for the accuracy of population codes [41], the formation

of neural assemblies [42], and the propagation of neural activity [43, 44]. Nonetheless, only recently has attention been given to the mechanisms by which correlated activity is modulated [45, 46, 47, 10, 48, 49, 50].

Cortical neurons receive a mixture of excitatory and inhibitory synaptic inputs, resulting in spiking activity that is driven by input fluctuations rather than the input mean [51, 23]. This state is often described as *balanced*, to denote that the mean excitatory and inhibitory inputs that neurons receive are approximately equal [22, 24]. Balanced activity is influenced by stimulus properties and history [32, 51], as well as internal brain state [28]. These changes can modulate the integration properties of single neurons, strongly influencing neuronal activity [23]. For example, increases in the firing rate of balanced pre-synaptic activity afferent to a neuron can reduce single neuron firing rate gain [52, 53, 54, 55, 56, 57]. Further, an increase in the temporal correlation between the arrival times of excitatory pre-synaptic inputs increases the firing rate of a post-synaptic target neuron [58, 59, 60], while correlations between excitatory and inhibitory inputs can reduce output activity [59, 61]. The impact of such shifts in the temporal structure of synaptic input is amplified when the post-synaptic cell has a small integration timescale, as expected for neurons in the high input rate, balanced state [23]. These examples deal with synaptic activity convergent to a single target cell. However, what is less studied is the role that the balanced state plays in modulating the responses of a pair of neurons subject to a common synaptic input. In this study, we consider this latter scenario and show that shifts in balanced pre-synaptic population activity modulate the magnitude and timescale of the correlations of spike trains from pairs of post-synaptic neurons.

We first explore a model system and show that output spike train correlations from a pair of neurons are modulated by varying the rate of fluctuating, balanced excitatory and inhibitory inputs. Specifically, we demonstrate that an increased synaptic input rate leads to an increase of short-timescale output correlation (i.e. precise spike synchrony) while correlation at long timescales (i.e firing rate co-variation) remains unaffected, or even decreases. Due to the differential effects of our mechanism on short and long timescale spiking activity we label the combined modulation *correlation shaping*. Correlation shaping has been observed in various sensory systems [12, 37, 13, 62], yet the core mechanisms underlying the modulation remain unknown. We present linear response analysis showing that the enhancement of output synchrony through an increase of input rate results from a shift in single neuron integration properties that favors

the transfer of high frequency inputs. Dynamic clamp recordings from cortical neurons verify our theoretical predictions. Finally, in a feedforward network model, we show how correlation shaping supports a selective propagation of network responses, so that activity can be gated by correlations in complex neuronal networks. In total, our work extends mechanisms of single neuron firing rate control include the control of pairwise correlations, thereby providing a bridge between single neuron and network state modulation.

2.3 Methods

2.3.1 Conductance-based neuron model

We modeled neurons as leaky integrate-and-fire units receiving conductance input [63]. Each neuron had an intrinsic timescale $\tau = 20$ ms and leak reversal potential $E_L = -65$ mV. Excitatory and inhibitory synaptic input caused conductance changes $g_e(t)$ and $g_i(t)$ with reversal potentials $E_e = 0$ mV and $E_i = -75$ mV so that the membrane potential dynamics followed:

$$\frac{dV}{dt} = \frac{1}{\tau}(E_L - V) + \frac{g_e(t)}{C}(E_e - V) + \frac{g_i(t)}{C}(E_i - V).$$

When V reached a threshold voltage $V_{th} = -55$ mV, the neuron spiked and the voltage was reset to $V_{re} = -65$ mV.

We modeled the excitatory and inhibitory synaptic conductances as Poisson processes with rates R_e and R_i consisting of series of δ-functions with heights $a_e = .01$ and $a_i = .02$. This framework was used for all of the simulations presented and provides a minimal model that captures our main results (for simulations of other models, see Figures 2.8, 2.10, and 2.12). These inputs consisted of independent processes private to each neuron as well as a shared component presynaptic to all neurons, yielding $R_{e/i} = cR^s_{e/i} + (1-c)R^i_{e/i}$ where superscripts i and s denote independent and shared components, respectively. For large rates, this input was approximated as a diffusion process [63, 64, 65, 66] (Figure 2.8):

$$a_{e/i}(E_{e/i} - V)\sum_n \delta(t - t^n_{e/i}) \approx a_{e/i}(E_{e/i} - V)\left(R_{e/i} + \sqrt{R_{e/i}}\xi_{e/i}(t)\right),$$

where $\xi_{e/i}(t) = \sqrt{1-c}\xi^i_{e/i}(t) + \sqrt{c}\xi^s_{e/i}(t)$ was a Gaussian white noise process with unit intensity. This allowed us to write our voltage equation in the form

$$\frac{dV}{dt} = \frac{1}{\tau_{eff}}(E_{eff} - V) + \sigma(V)\xi(t), \tag{2.1}$$

where $\tau_{eff} \equiv \frac{\tau}{1+\tau a_e R_e + \tau a_i R_i}$, $E_{eff} \equiv \frac{E_L + \tau R_e a_e E_e + \tau R_i a_i E_i}{1+\tau R_e a_e + \tau R_i a_i}$, and $\sigma^2(V) = a_e^2 R_e (E_e - V)^2 + a_i^2 R_i (E_i - V)^2$. Note that as the rates of excitation and inhibition R_e and R_i increase in a balanced manner, τ_{eff} decreases, σ increases, and E_{eff} does not change substantially because of the excitation and inhibition balance.

For our simulations and calculations, we set $\sigma(V) = \sigma(E_{eff})$. This approximation ignored the multiplicative nature of the noise, which in our simulations did not substantially change the results (Figure 2.8), since the change in τ_{eff} and $\sigma(E_{eff})$ were sufficient to modulate neuronal responses. To simulate pairs of neurons receiving correlated input, we set the fluctuating input to each neuron to be

$$\sigma(V)\xi(t) = \sigma(E_{eff})(\sqrt{c}\xi_s(t) + \sqrt{1-c}\xi_i(t)), \tag{2.2}$$

where $\xi_s(t)$ was shared across both neurons while $\xi_i(t)$ was independent for each neuron. We note that, although the correlation in output spike trains depended on the degree of pre-synaptic overlap, equation (2.2) shows that $\sigma(E_{eff})$, and hence the firing rate of neurons in our model, was independent of c. The rate of excitatory input in the low state was 1.50 kHz and 6.16 kHz in the high state, with the inhibitory rate chosen to elicit a firing rate of 15 Hz in both cases. Simulations were performed using an Euler-Maruyama numerical integration scheme with a simulation timestep of 0.005 ms.

2.3.2 Solving for transfer function and power spectrum with Fokker-Planck techniques

We next developed a theoretical framework to study the behavior of the above system and compared our theory against simulations of the stochastic system. For completeness, we write the governing equations used to calculate the single neuron power spectrum $\hat{C}_{ii}(f)$ and transfer function $\hat{A}(f)$; these techniques are fully presented in [66] and we refer the reader there for further details. Letting $h(V) = \frac{1}{\tau_{eff}}(E_{eff} - V)$, the voltage distribution $P(V,t)$ associated with

the stochastic differential equation (2.1) obeys the Fokker-Planck equation:

$$\frac{\partial P}{\partial t} = -\frac{\partial J}{\partial V} = -\frac{\partial}{\partial V}\left[h(V)P\right] + \frac{1}{2}\frac{\partial}{\partial V}\left(\sigma^2 \frac{\partial P}{\partial V}\right),$$

where $J(V,t)$ is the probability flux [67]. The boundary conditions for the probability distribution and flux at threshold are $P(V_{th}) = 0$ and $J(V_{th},t) = \nu(t)$, where $\nu(t)$ is the firing rate. Furthermore, the flux obeys $J(V,t) = \nu(t)$ for $V \in [V_{re}, V_{th}]$ and is 0 otherwise.

For time independent E_{eff} and σ the steady state distribution $P_0(V)$:

$$\begin{aligned}\frac{\partial P_0}{\partial V} &= -\frac{2}{\sigma^2}\left[J_0 - h(V)P_0\right], \\ \frac{\partial J_0}{\partial V} &= \nu_0\delta(V - V_{re}) - \nu_0\delta(V - V_{th}).\end{aligned}$$

Using the normalization condition $\int_{-\infty}^{Vth} P_0(V)dV = 1$, we can solve for the steady state firing rate ν_0.

In order to study the system's response to a correlated, fluctuating input, it is necessary to study the system's response to time-dependent inputs. This is done most effectively by writing a time-dependent Fokker-Planck equation in the Fourier domain:

$$\begin{aligned}\frac{\partial \hat{P}}{\partial V} &= -\frac{2}{\sigma^2}\left[\hat{J} - h(V)\hat{P}\right], \\ \frac{\partial \hat{J}}{\partial V} &= -2\pi i f\hat{P} - \hat{\nu}(f)\delta(V - V_{th}) + e^{-2\pi i f}\delta(V - V_{re}),\end{aligned}$$

where $\hat{X}$ denotes the Fourier transform of X with respect to t and $\hat{\nu}(f)$ is computed with initial condition $V = V_{re}$. Solving this equation yields the Fourier transform of the first passage time density $\hat{h}(f)$ [66]. The power spectrum $\hat{C}_{ii}(f) = \nu_0\left(1 + 2\Re[\hat{g}(f)]\right)$, where $\hat{g}(f)$ is calculated from the well known renewal relation $\hat{g}(f) = \hat{h}(f)/(1 - \hat{h}(f))$ [68].

Finally, we compute the transfer function $\hat{A}(f)$. Suppose that we add a time-varying periodic current $I(t) = I_0 e^{2\pi i f t}$ to the right hand side of equation (2.1). If we let I_0 be sufficiently small, we can compute the spike train response to these time-dependent modulations. Decomposing the probability density, flux, and firing rate into steady state and modulated components:

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

and then solving the Fokker-Planck equation for the time-dependent terms, we obtain a new set of equations:

$$-\frac{\partial \hat{P}_I}{\partial V} = \frac{2}{\sigma^2(V)}\left[\hat{J}_I + h(V) + I_0 P_0\right],$$

$$-\frac{\partial \hat{J}_I}{\partial V} = i\omega \hat{P}_I - \hat{r}_Q e^{-i\omega}\delta(V - V_{re}),$$

with boundary conditions

$$\hat{P}_I(V_{th}) = 0,\ \hat{J}_I(V_{th}) = \hat{A}.$$

These equations were solved numerically [66] obtaining a solution for the transfer function $\hat{A}(f)$.

2.3.3 Experimental techniques

Surgery: Somatosensory (S1) cortical slices were prepared from CBJ/Bl6 mice age P19-26. All surgical procedures followed the guidelines approved by the Carnegie Mellon Animal Welfare Committee. The mice were anesthetized with isoflourane and decapitated. The brain was exposed, removed from the skull and immersed, in ice cold oxygenated ($95\%O_2 - 5\%CO_2$) ACSF (in mM: 125 NaCl, 2.5 KCl, 25 $NaHCO_3$, 1.25 NaH2PO_4, 1.0 $MgCl_2$, 25 Dextrose, 2 $CaCl_2$) (all chemicals from Sigma, USA). Coronal slices (300 μm) of barrel cortex made using a vibratome (Leica, Place). The slices were maintained in ACSF at $37°$C for 30 min then rested at room temperature ($20 - 22°$C) for 1 hr prior to recording ($31 - 35°$C).

Electrophysiology: L2/3 pyramidal neurons were visualized using infrared-differential interference contrast microscopy (Olympus, Center Valley, PA). Whole cell, dynamic clamp recordings were performed using a MultiClamp 700B amplifier (Molecular Devices, Union City, CA). Data were low pass filtered (4 kHz) and digitized at 50 kHz using an ITC-18 (Instrutech, Mineola, NY) controlled by custom dynamic clamp software (R. Gerkin; http://rick.gerk.in/software/recording-artist/) written in IgorPro (Wavemetrics, Lake Oswego, OR). Pipettes were pulled from borosilicate glass (2.0 mm, outer diameter) on a Flaming/Brown micropipette puller (Sutter Instruments, Novato, CA) to a resistance of 6-10 MΩ. The intracellular solution consisted of (in mM) 130 K-gluconate, 5 KCl, 2 $MgCl_2$, 4 ATP-Mg, 0.3 GTP, 10 HEPES, and 10 phosphocreatine.

Stimulation: Pyramidal cells (n=8) were directly stimulated by a series (50-100 trials) of simulated noisy synaptic currents in dynamic clamp. Each trial was 4 s in duration with a 5 s inter-trial interval; the period of rest was used to ensure that stability of the recordings. For each trial, excitatory (E_e: 0 mV) or inhibitory (E_i: -60 mV) synaptic conductance inputs were simulated as Poisson distributed spike times convolved with alpha function $g_{e,i}(t) = \bar{g}_{e,i}\frac{t}{\tau_{e,i}}e^{1-t/\tau_{e,i}}$. ($\bar{g}_e = 1$ nS, $\bar{g}_i = 1$ nS, $\tau_e = 6$ ms, $\tau_i = 8$ ms). The Poisson rates for excitatory and inhibitory inputs were equal to one another ($R_e = R_i$), and were set to 3 kHz in the low state and 7.5 kHz in the high state. These rates were higher than in the simulations to ensure high spike time variability, since the input variability is attenuated by the finite temporal extent of the synaptic timescales. For each state, half of these inputs were common to all neurons stimulated and half were newly generated on each trial for every neuron. This produced an input correlation, c, of 0.5 between any given pair of neurons. This setup permitted $8 \cdot (8-1)/2 = 28$ pairwise comparisons. Since the synaptic drive was subthreshold, a bias current (0.3-0.7 nA) was added such that the balanced conductance fluctuations produced a mean cortical firing rate of 4-6 Hz in both the low and high states.

2.3.4 Feedforward network

We studied a layered network in which a population of 100 leaky integrate-and-fire neurons (Layer 2) received balanced input from a pre-synaptic layer (Layer 1) with $c = .2$ and provided excitatory input to two distinct downstream targets. Neurons in Layer 1 were assumed to be Poisson as in previous sections, and the total input to a Layer 2 neuron was therefore approximated by a diffusion process. In particular, the voltage dynamics of each Layer 2 neuron followed equations 2.1 and 2.2.

The downstream target was also modeled as leaky integrate-and-fire neuron. Because we wished to fix the timescale of the downstream target, we assumed delta-function, current-based synapses so that the voltage V of the downstream neuron followed:

$$\frac{dV}{dt} = \frac{1}{\tau_{downstream}}(E_L - V) + a_d \sum_{c=1}^{100} \sum_k \delta(t - t_c^k),$$

where $c = 1 \ldots 100$ indexes the neurons in Layer 2 and k indexes the spikes in each Layer 2 neurons' spike train. We compared $\tau_{downstream} = 20$ ms and $\tau_{downstream} = 3$ ms. For

$\tau_{downstream} = 20$ ms, we set $a_d = 0.12$ mV and for $\tau_{downstream} = 3$ ms, $a_d = 0.4$ mV so that the neurons fired at comparable rates given identical input. Other parameters, including leak, threshold, and reset voltages were identical to the model previously studied.

2.4 Results

2.4.1 Modulation of correlation susceptibility

In general, it is difficult to determine the specific changes in a neural system's dynamics that cause changes in spike train correlations. We studied a framework in which common inputs drive the correlations between the spike trains of a pair of neurons [69, 70, 71]. If the degree of input correlation, c, is small, a linear approximation relating c to the output spike correlation, ρ, is written as:

$$\rho \approx Sc.$$

Here the quantity S, termed the *correlation susceptibility*, determines the extent to which two neurons' spike trains will be correlated given a fixed level of correlation between the inputs they receive [10].

Throughout this study, we focused on a pair of neurons that shift their output correlation ($\rho_1 \to \rho_2$) due to a change in their pre-synaptic drive (Figure 2.1a). Under our linear model, two simple explanations for the shift in output correlation are possible. First, the shift may simply reflect a change in the correlation of the inputs that the neuron pair receives ($c_1 \to c_2$; Figure 2.1b). While this answer appears straightforward, understanding shifts in input correlation requires detailed anatomical knowledge of the network architecture, in the absence of which simplifying assumptions are required [27].

A second explanation for the shift in output correlation is a shift in correlation susceptibility ($S_1 \to S_2$), even when the input correlation remains fixed (Figure 2.1c). Because S relates the correlations in the spiking output of neurons to their common input, we expect S to be sensitive to how each neuron integrates its input. Indeed, single neuron response properties such as firing rate and neural excitability determine the extent to which neurons become synchronized

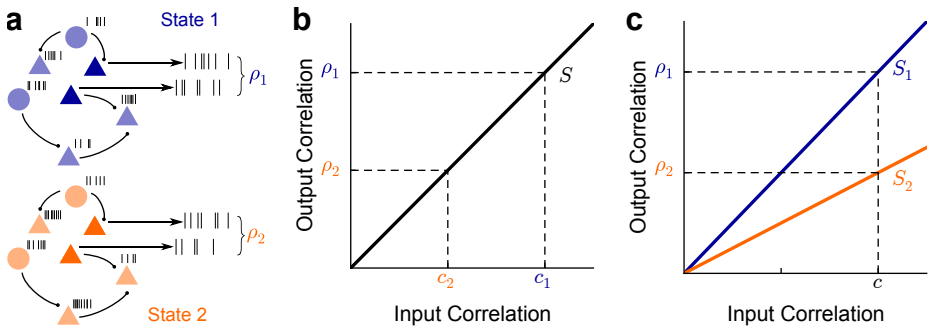

Figure 2.1: Mechanisms of correlation modulation. **(a)** The spike train correlation between a pair of neurons shifts from ρ_1 to ρ_2 as the state of pre-synaptic field shifts. **(b)** Mapping between input correlation c and output correlation ρ. The change in output correlation in panel A may be due to a change in input correlation from c_1 in state 1 to c_2 in state 2. **(c)** An alternative mechanism by which output correlation can change is that the correlation susceptibility S changes from S_1 in state 1 to S_2 in state 2, with input correlation c fixed throughout.

by shared input [46, 10, 48, 49, 50]. There has been substantial work on how single neuron properties, such as firing rates, are modulated [52, 53, 54, 55, 56, 57, 72, 73, 74, 75, 76], suggesting that S should also be open to modulation. We focused on this second mechanism and established how modulations of single neuron responses also modulated pairwise correlations in cortical populations.

2.4.2 Low and high rate synaptic input states

We first investigated the transfer of input correlations to output spike train correlations in a simplified two-neuron network. Each neuron received conductance-based, pre-synaptic inputs from a mixed population of excitatory and inhibitory neurons (Figure 2.2a). To model the stochastic nature of cortical activity, the arrival times of both excitation and inhibition were modeled as Poisson processes. We set the relative strengths and rates of excitation and inhibition so that the mean input was balanced [22, 24], and the average membrane potential was below spiking threshold. Balanced pre-synaptic activity results in large membrane fluctuations that trigger spikes in a random, aperiodic pattern, consistent with *in vivo* recordings from cortical neurons [51, 23].

Shifts in the activity level of a recurrent cortical population are observed in many neural systems and have been shown to affect the response properties of neurons *in vitro* and *in vivo* [23, 28, 56]. To explore the modulatory effects of balanced synaptic input, we considered the

neuron model in two states: a *low* state, in which pre-synaptic input arrived at a low rate, and a *high* state, in which pre-synaptic input arrived at a high rate (Figure 2.2a). While the level of balanced fluctuations may lie on a continuum, we compared two representative points, analogous to high and low activity states in a cortical network [77, 28]. A clear consequence of the shift from low to high states was an increase in the variability of both the input current and membrane potential response, due to greater fluctuating input (Figure 2.2b). This increase of input variability was reflected in an increase in spiking variability, with the coefficient of variation of the inter-spike intervals increasing from 0.73 in the low state to 0.91 in the high state. A second consequence of an increase in pre-synaptic rate was the reduction of the membrane time constant τ (Figure 2.2b). This was expected, since the membrane time constant $\tau \sim C/g$, with C the membrane capacitance and g the total membrane conductance [78]. As g is roughly proportional to the pre-synaptic rates, an increase in the rate of synaptic input lead to a decrease in τ. Taken together, the shift from the low to high state evoked a more stochastic and faster membrane potential response.

We first examined the effect of balanced synaptic input on firing rate gain, the slope of the firing rate curve when plotted as a function of excitatory input strength. When the rate of balanced excitatory and inhibitory synaptic input changed from low to high, the neuron's firing rate gain was substantially reduced (Figure 2.2c). This gain decrease in the high background state has been studied extensively in theoretical and *in vitro* work [52, 53, 54, 55, 56, 57] as well *in vivo* under specific stimuli conditions [56]. In the high state, larger membrane potential fluctuations increased firing rates for weak inputs. However, there was also a decrease of the net membrane input resistance, causing an increase in the rheobase current (minimum steady current required to recruit spiking). The combination of these two effects lead to an overall reduction in firing rate gain [54]. We next explored the consequences of gain modulation via balanced activity for correlation transfer by pairs of neurons.

2.4.3 Correlation shaping with synaptic activity

To study the effects of balanced excitatory and inhibitory inputs on pairwise spike train correlations, we extended our model to include a pair of post-synaptic neurons receiving overlapping pre-synaptic inputs (Figure 2.3a). Previous work has shown that the output firing rate affects

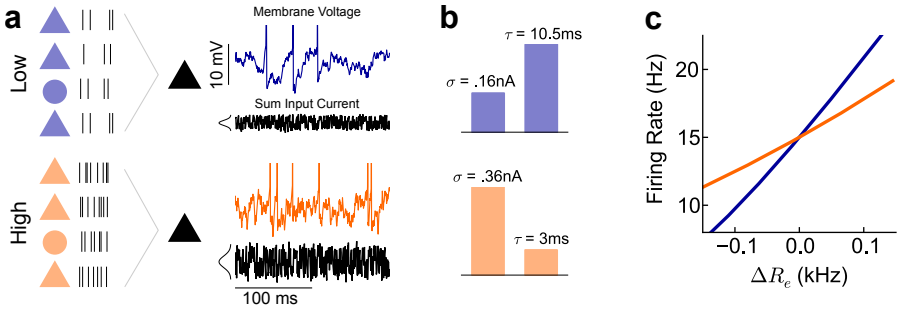

Figure 2.2: Single cell statistics in low and high synaptic input states. **(a)** Left: Schematic of low (top) and high (bottom) states. Excitatory inputs occurred at a rate of 1.5 kHz in the low state and 6.2 kHz in the high state, modeling the activity from a pool of pre-synaptic cells. Inhibitory inputs were chosen so that output firing rates were fixed at 15 Hz in both states. Right: Synaptic inputs converged onto a conductance-based leaky integrate-and-fire neuron model. Sample membrane potential traces of the neuron model in both the low (top) or high (bottom) states are shown. The total input current in either state is plotted below each membrane potential trace. **(b)** The input current variability (σ) and membrane potential time constant (τ) for both the low (top) and high (bottom) input states. **(c)** Firing rate of a neuron as the level of excitatory input is varied, showing decreased gain in the high input state compared to the low input state. The balanced condition in both low and high states resulted in an output firing rate of 15Hz. Curves were calculated using our theory (see Methods).

correlation susceptibility [10]. To preclude any firing rate-induced effects, the synaptic input was adjusted so that the average output firing rate of each neuron remained at 15 Hz in low and high states (Figure 2.2c). Furthermore, there was a fixed overlap in the input populations, so that the input correlation also remained constant in both network states (Figure 2.3a). Thus, any change in the output spike train correlation induced by changing synaptic input will be due exclusively to a shift in correlation susceptibility (Figure 2.1c).

We found that the timescale over which the two spike trains were correlated was dependent on the level of balanced synaptic activity (Figure 2.3a, right). When the synaptic rate increased from the low to high state, the magnitude of the peak of the cross-covariance function near zero lag increased, reflecting greater spike time synchrony between the neurons. However, this increase was not present for longer lags, and the spike train cross-covariance function was unchanged or reduced for sufficiently long lags (> 10ms).

To quantify this change in output correlation over a range of timescales, we first counted the number of spikes n_1^T and n_2^T that the two neurons emitted in intervals of T milliseconds. We

next computed the spike count correlation as a function of window size:

$$\rho_T = \frac{\text{Cov}(n_1^T, n_2^T)}{\sqrt{\text{Var}(n_1^T)\text{Var}(n_2^T)}}, \tag{2.3}$$

where Cov and Var denote covariance and variance, respectively. In the framework of our simple circuit (Figure 2.3a), correlation in output spike trains ρ_T was a consequence of a shared input correlation c. For small c, linear response theory [10] takes the output correlation to be a linear function of the input correlation (Figures 2.1b,c; 2.3b):

$$\rho_T = S_T c. \tag{2.4}$$

In our model, this linear relationship held for a range of c, in both low and high states and at both short and long T (Figure 2.3b). Further, the ρ_T values produced were, in magnitude, consistent with *in vivo* recordings from a variety of systems [79, 37, 12, 14]. When comparing ρ_T for the low and high states at fixed c, a differential change of correlation at different timescales was evident. Specifically, $\rho_T^{low} < \rho_T^{high}$ for small T (Figure 2.3b, T=3 ms), while $\rho_T^{low} > \rho_T^{high}$ for large T (Figure 2.3b, T=50 ms). This differential modulation of correlation occured over a broad range of timescales, with ρ_T^{low} and ρ_T^{high} intersecting only once (Figure 2.3c), and we label the modulation a *shaping* of correlation [62]. This substantial change in both the magnitude and timescale of correlation must involve a nontrivial change in how the neurons process their inputs, since the input correlation c and firing rate were the same in both low and high states. We note that the qualitative results of our study are also valid for larger c (Figure 2.10) and different synaptic strengths (Figure 2.12).

Since $\rho_T \to 0$ as $T \to 0$ [80], changes in ρ_T at small T are necessarily smaller in magnitude. However, synchrony at short timescales can have large effects on downstream targets sensitive to coincident pre-synaptic spikes [43] and indeed the peak of the cross-covariance function increased substantially in the high state (Figure 2.3a, right). To properly compare correlation shaping at small and large T we considered the ratio $S_T^{high}/S_T^{low} = \rho_T^{high}/\rho_T^{low}$, providing a relative measure across the low and high states. The ratio was a decreasing function of T, with substantial changes in correlation at both short and long timescales (Figure 2.3d). The negative slope of the curve indicates that increases in the rate of balanced synaptic activity

favor spike synchronization rather than long timescale correlation. Finally, the spectral measure of spike train coherence between the two spike trains in both states exhibited a decrease for low frequencies but a significant increase for high frequencies in the high state (Figure 2.3e). Here, the increase for high frequencies, which occurs over a broad range of frequency space, is related to the increase in short timescale synchrony, consistent with the spike count correlation shaping.

Correlation shaping is an unexpected feature of balanced synaptic activity. For subthreshold membrane potential dynamics (or any other linear system) the ratio $\rho_T^{high}/\rho_T^{low}$ is equal to 1 for all T assuming a fixed input correlation (Figure 2.3d, gray line). The mechanism that shapes correlation transfer so to promote spike train synchronization over long timescale correlation in the high state (Figure 2.3d) is the focus of the next section.

2.4.4 Relationship between correlation susceptibility and neuronal integration

Correlation shaping is a property of the joint statistics of a pair of neurons. However, since the input correlation was the same in the low and high states of our model, then the mechanism underlying the shaping is hypothesized to be related to changes in single neuron input integration and spike emission across the two synaptic states (Figure 2.1c rather than 2.1b). In this section, we show that correlation shaping is a consequence of a shift in the single neurons' frequency response across the low to the high input state.

The spike train autocovariance $C_{ii}(t)$ and cross-covariance $C_{i\neq j}(t)$ functions are written as:

$$C_{ij}(t) = \int_{-\infty}^{\infty} y_i(\tau) y_j(\tau - t) d\tau - \bar{y}_i \bar{y}_j, \tag{2.5}$$

where $y_i(t) = \sum_i \delta(t - t_{ik})$, with t_{ik} labeling the k^{th} spike time from neuron i ($i = 1, 2$). Here $\bar{y}_i$ is the mean firing rate of neuron i. We are interested in the joint spike count correlation for the neuron pair, where the spike count for neuron i over a window of length T is $n_i^T = \int_0^T y_i(t) dt$ (we take the neuron's stochastic dynamics to be in statistical equilibrium). The spike count variance and covariance are related to integrals of auto- and cross-covariance functions [68],

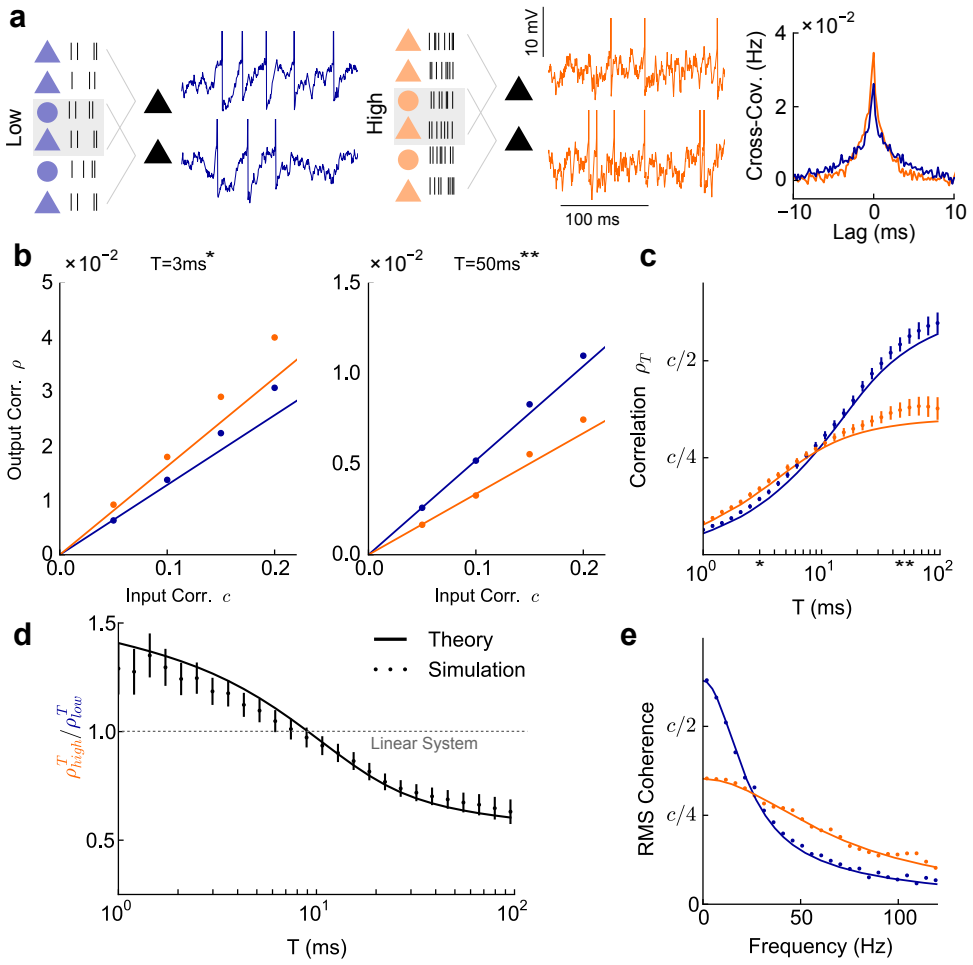

Figure 2.3: Pairwise cell statistics in low and high rate synaptic input regimes. **(a)** Schematic of low (left) and high (center) states with sample membrane traces. The marginal statistics of both cells are as reported in Figure 2.2, with a fixed overlap of excitatory and inhibitory pre-synaptic inputs for the cell pair. The input correlation is $c = 0.5$ for membrane traces and $c = 0.1$ otherwise, in both low and high states. Right: Spike train cross-covariance functions for the firing of the two neurons when receiving correlated input, showing state dependent shaping. **(b)** Relationship between spike count correlation ρ_T for windows of length T and input correlation c, showing linearity for small c and a dependence on T. **(c)** Output correlation as a function of window size in the high and low states. Asterisks mark the values of T that correspond to the plots in Figure 2.3b. **(d)** Ratio of correlations as a function of window size in the high and low states, showing favoring of short timescale synchrony in the high state. For comparison, the lack of correlation shaping for a purely linear neural transfer is indicated. **(e)** RMS coherence ($|\hat{C}_{12}(f)|/\sqrt{\hat{C}_{11}(f)\hat{C}_{22}(f)}$) between spike trains showing a decrease in low-frequency coherence and increase in high frequency coherence in the high state. The theoretical results (solid lines) shown in in panels (b) through (e) were derived from a linear response calculation valid in the small c limit (see Methods). Bars denote standard error in (b) through (d). In (b), standard error is smaller than the width of the dots.

yielding an alternate expresion for ρ_T:

$$\rho_T = \frac{\text{Cov}(n_1^T, n_2^T)}{\sqrt{\text{Var}(n_1^T)\text{Var}(n_2^T)}} = \frac{\int_{-T}^{T} C_{ij}(t)(T-|t|)dt}{\int_{-T}^{T} C_{ii}(t)(T-|t|)dt}. \tag{2.6}$$

The triangular weighting factor $T - |t|$ arises from the conversion from spike trains to spike counts and can be understood as the convolution of a boxcar function of width T with itself. In the second equality we have, for simplicity, assumed that $C_{11}(t) = C_{22}(t)$ (or equivalently $\text{Var}(n_1^T) = \text{Var}(n_2^T)$). These integrals can be transformed to the frequency domain, using the Wiener-Khinchin theorem [68] to relate covariance functions $C_{ij}(t)$ to their spectral analogues $\hat{C}_{ij}(f)$, yielding

$$\rho_T = \frac{\int_{-\infty}^{\infty} \hat{C}_{ij}(f) k_T(f) df}{\int_{-\infty}^{\infty} \hat{C}_{ii}(f) k_T(f) df}. \tag{2.7}$$

Here $k_T(f) \equiv \frac{1}{\pi^2 T f^2} \sin^2\left(\frac{2\pi f T}{2}\right)$ is the Fourier transform of the triangular weighting term in equation (2.6). Our strategy was to relate the cross spectrum between the spike trains, $\hat{C}_{ij}(f)$, to single neuron integration properties.

Single neuron input-output transfer is typically expressed through its spectral transfer function $\hat{A}(f)$. The transfer function measures the ratio of the amplitudes of a neuron's firing rate response and a small amplitude sinusoidal signal of frequency f (Figure 2.4a). For very slow inputs, the transfer function $|\hat{A}(0)|$ equals the firing rate gain, since this measures the sensitivity of firing responses to static ($f \approx 0$) inputs. For $f > 0$, $|\hat{A}(f)|$ is the susceptibility for a neuron's trial averaged response to be locked to a time varying signal. The transfer function $|\hat{A}(f)|$ is experimentally measurable [81], and is related to the more commonly reported spike triggered average [82]. In general, for neurons in the fluctuation-driven regime, $|\hat{A}(f)|$ is a decaying function of f (Figure 2.4b).

If each neuron receives a small shared signal $Q(t)$, then we can write the expectation of the Fourier transform of the spike train from neuron i as:

$$\langle \hat{y}_i(f) \rangle \approx \hat{A}_Q(f)\hat{Q}(f), \tag{2.8}$$

where the brackets denote an average over repeated frozen presentations of the shared signal $Q(t)$ with different realizations of the independent noise driving the neurons [83]. Here, $\hat{A}_Q(f)$

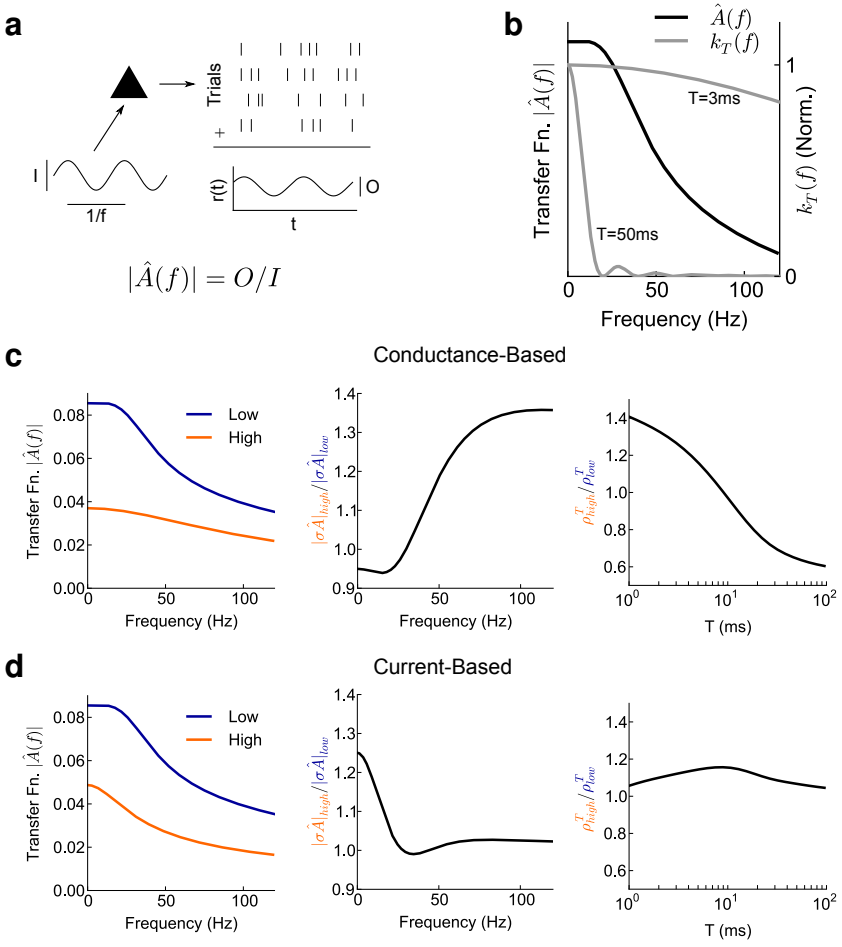

Figure 2.4: Relating correlation shaping to single neuron transfer. **(a)** Illustration of neuronal transfer function. A perturbing input of amplitude I and frequency f causes a modulation of the spike response of a fluctuation driven neuron. Averaging across stimulus presentation trials gives the average output firing rate $r(t)$ with amplitude O. The output-input ratio defines the neural transfer $\hat{A}(f) = O/I$. **(b)** Example $|\hat{A}(f)|$ for a fluctuation driven neuron (black curve). The weighting function $k_T(f)$ for $T = 3$ms and 50ms (grey curves). **(c)** Left: transfer function $|\hat{A}(f)|$ for neurons in the low and high background states. Center: Ratio of transfer functions in the two states (normalized by change in input strength $\sigma_{high}/\sigma_{low}$). Right: Ratio of correlations as in Figure 2.3d. **(d)** Same as (c), but for a current-based model in which τ_{eff} does not change in the high state. Note correlation shaping in (c), Right but not (d), Right.

is the linear response of the system to the perturbation $Q(t)$. Finally, averaging the quantity $\langle\hat{y}_1^*(f)\rangle\langle\hat{y}_2(f)\rangle$ over different realizations of the process $Q(t)$ yields the cross-spectrum between neurons 1 and 2 [10, 83, 84, 85]:

$$\hat{C}_{12}(f) = \langle\langle\hat{y}_1^*(f)\rangle\langle\hat{y}_2(f)\rangle\rangle_Q \approx |\hat{A}_Q(f)|^2\langle\hat{Q}^*\hat{Q}\rangle. \tag{2.9}$$

For the case of white noise input, we have that $\langle\hat{Q}^*\hat{Q}\rangle = c\sigma^2$. With equations (2.7) and (2.9) we calculated the spike count correlation coefficient between the two neurons receiving shared

white noise input as

$$\rho_T \approx S_T c = \left(\frac{\sigma^2 \int_{-\infty}^{\infty} |\hat{A}|^2(f) k_T(f) df}{\int_{-\infty}^{\infty} \hat{C}_{ii}(f) k_T(f) df} \right) c. \tag{2.10}$$

Our theory then relates single neuron transfer $|\hat{A}(f)|$ and power spectrum $\hat{C}_{ii}(f)$ to the joint pairwise response ρ_T.

The theoretical predictions given in equation (2.10) gave a very good quantitative match to simulations of the leaky integrate-and-fire neuron pair (Figures 2.3b-e, compare solid curves to points), capturing the correlation shaping between the two states. Equation (2.10) has been previously derived [10, 48], however, the model neurons considered in those studies were current driven model neurons. We considered conductance driven model neurons, meaning that the calculation of $|\hat{A}(f)|$ and $\hat{C}_{ii}(f)$ must account for the linked shifts of the membrane time constant and membrane potential fluctuations from the low to the high state (Figure 2.2b). For our conductance based integrate-and-fire model neurons, the quantities $\hat{A}(f)$ and $\hat{C}_{ii}(f)$ were calculated by numerically integrating the Fokker-Planck equation associated with the stochastic differential equation expressed in equation (2.1) (see [66] and Methods). The distinction between current and conductance based neural integration will be shown to be critical for correlation shaping.

Before correlation shaping is related to the shifts in $\hat{A}(f)$ between the low and high states, we first discuss the dependence of susceptibility S_T on the window size T (Figure 2.3b). This dependence enters equation (2.10) through the weighting term $k_T(f)$, which determines the contribution of $\hat{A}(f)$ across frequency to S_T. For long timescales (large T), $k_T(f)$ is low-pass, so that only the neurons' response to low frequencies contributes to correlation susceptibility. In contrast, for short timescales (small T), $k_T(f)$ weighs the transfer function approximately equally across all frequencies. Hence, the neurons' high frequency response determines precise spike synchrony. Indeed, for $T \to \infty$ we have that $k_T(f) \to \delta(f)$, while $T \to 0$ limits $k_T(f)$ to a constant function on $(-\infty, \infty)$. Therefore, for large T, only the zero-frequency components of $\hat{A}(f)$ contribute to the integral, while for small T, all frequencies contribute.

A mechanistic understanding of correlation shaping (Figure 2.3d) requires knowledge of how the rate of balanced synaptic activity affects the transfer function. As discussed previously, the increase in synaptic input from the low to the high state decreased the effective membrane

time constant of the neuron τ while it increased the input variability σ (Figure 2.2b). The decrease in τ corresponded to a decrease in the timescale over which a neuron integrates inputs and hence an attenuation of the neuron's transfer function. For low frequency inputs, this reduction was precisely the firing rate gain control known to occur with increased synaptic input (Figure 2.2c). Increased variability and shunting due to heightened conductance reduced the neuron's ability to respond to slow depolarizing inputs. However, the reduction in the transfer function from the low to high state was not uniform across all frequencies (Figure 2.4c, left). This was because the smaller value of τ in the high state enhanced the tracking of fast inputs, mitigating the attenuation of the transfer function for high frequencies. The combination of the non-uniform attenuation of the transfer function and increase in σ from the low to high state determined the shaping of the correlation susceptibility S_T (see equation (2.10)).

To illustrate the shift in single neuron response between the low and high states, we considered the quantity $\sigma|\hat{A}(f)|$, the strength of the input fluctuations multiplied by the input transfer function. The ratio $|\sigma\hat{A}|_{high}/|\sigma\hat{A}|_{low}$ was an increasing function of frequency (Figure 2.4c, center), indicating that high frequency transfer is favored in the high state. In general, a favoring of high frequencies corresponds to a favoring of synchrony, measured over only small T (since $k_T(f)$ is nearly flat across f for small T). Thus, the high state is expected to favor small T correlation transfer compared to the low state (Figure 2.4c, right). In contrast, for large T which corresponds to low frequencies, correlation transfer was disfavored in the high state (since $k_T(f)$ only weights low f for large T). This ratio allowed us to intuitively link correlation shaping over different timescales to the shaping of the transfer function over different frequencies.

We argue above that a change in the effective membrane time constant is central to the correlation shaping we discuss. To demonstrate this fact, we computed the transfer function and correlations for a current-based model in which τ remained unchanged in the low and high state, although σ increased by the same amount. If firing rates were again fixed at 15 Hz, the transfer function was again reduced in the high state, but the ratio $|\sigma\hat{A}|_{high}/|\sigma\hat{A}|_{low}$ remained close to unity (Figure 2.4d, left and center). As a result, no substantial correlation shaping was observed (Figure 2.4d, right). The above comparison shows that this shaping requires the modulation of cellular properties that is allowed by a conductance-based model.

Finally, we note that, although our analysis has focused on the numerator of equation (2.7), the denominator also affects the correlation for large time windows (Figure 2.11). For these values of T, the denominator was increased in the high state, reflecting the higher variability of firing due to stronger input fluctuations. This further attenuated the value of ρ_T for large T in the high state.

2.4.5 Correlation shaping with different output firing rates

To avoid changes in correlation owing to firing rate [10], we chose the balance between excitation and inhibition in previous sections so that firing rate was fixed across both the low and high states (Figure 2.2c). However, it is unlikely that firing rates will remain fixed as a network shifts from a low conductance to a high conductance state. Thus, it is important to understand how correlation shaping via balanced excitatory and inhibitory inputs interacts with the correlation changes expected due to firing rate changes. In this section we show how the modulations of correlation due to balanced excitatory and inhibitory inputs and those due to firing rate changes from imbalanced inputs are distinct.

The firing rates of our output neurons were determined by the input rate of both the excitatory (R_e) and inhibitory (R_i) inputs. In fact, for any desired output rate, there was a curve in (R_i, R_e) space that achieved that rate (Figure 2.5a). For moderate input rates, a balanced shift in input (approximately linear in R_i and R_e) preserved output firing rate. A change in output firing rate (switching from one curve to another in Figure 2.5a), can occur from a shift in R_e, a shift in R_i, or some combination of the two. When we fixed R_i to its value in the low state and increase R_e so that the output rate increased, ρ_T increased over all timescales T (Figure 2.5b, top), as expected [10]. A similar effect occured if we repeat this in the high state (Figure 2.5b, bottom). Thus, the modulation of ρ_T by a rate change due to an imbalanced shift of R_e simply scales ρ_T for all T (collapsed blue and orange curves in Figure 2.5c). Nevertheless, after correcting for the rate scaling of ρ_T, the shaping of correlation between the low and high states remained clear (Figure 2.5c), demonstrating that correlation shaping due to a change from low to high states is distinct from correlation shifts due to arbitrary output firing rate changes.

To illustrate this, we considered a shift from 8 Hz in the low state to 35 Hz in the high state. In the shift from the low to high state, the effective membrane timescale τ_{eff} shifted

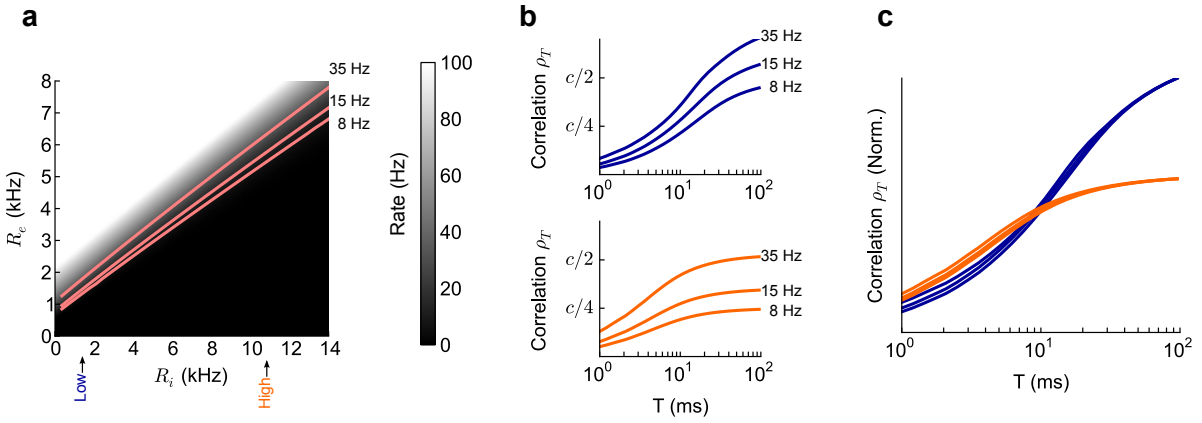

Figure 2.5: Comparing correlation shaping due to balanced excitatory and inhibitory inputs and correlation shifts due to shift in output rate. **(a)** Output firing rate as a function of R_i and R_e (firing rates above 100 Hz not shown). The curves in the space that yield output firing rates of 8,15, and 35 Hz are labeled. The lines lie in a small region of the full space, corresponding to the region where excitation and inhibition are balanced. **(b)** Top: Spike count correlation as a function of T for the three output rates, where rate changes are due to a change R_e, with R_i fixed at the low state value. Bottom: same as top, except that R_i is fixed at the high state value. **(c)** The curves in B for the low and high states scaled to match the center curve (15 Hz) at $T = 100$ ms.

from 10.8 to 2.9 ms and the amplitude of the input fluctuations σ from 0.16 to 0.37 nA. These shifts changed $\sigma|\hat{A}(f)|$ significantly (as discussed in the previous section), and changed the timescales over which the neuron pair was correlated. This was contrasted by a shift from 8 Hz to 35 Hz in the low state: a change in firing rate without a change between low and high states. Here, τ_{eff} shifted from 10.8 to 10.2 ms and the input fluctuations σ from .16 to .18 nA, having little influence on $\sigma|\hat{A}(f)|$ other than a uniform scaling due to the output rate change. In total, by changing both R_e and R_i, it was possible to not only change the output firing rate so as to amplify or attenuate ρ_T, but also to shape the timescales over which a neuron pair was correlated.

2.4.6 Experimental verification with dynamic clamp recordings

Our two-neuron framework for studying correlation transfer (Figure 2.1a) permited an experimental verification of correlation shaping with balanced, fluctuating conductance inputs. We performed *in vitro* patch clamp recordings from cortical pyramidal neurons receiving simulated excitatory and inhibitory inputs. Unlike past experimental studies of correlation transfer [86, 10], our model involved conductance-based, rather than current-based synapses. Therefore, we simulated synaptic input using dynamic clamp [87] (see Methods), which affected the

membrane integration timescale as well as membrane potential variability. We chose maximal excitatory and inhibitory conductances of 1 nS and synaptic timescales of 6 and 8 ms, respectively, producing a synaptic input that was more biophysically realistic than the diffusion process used in previous sections (Figure 2.6a). The shift from low to high state caused a near two-fold reduction in firing rate gain (Figure 2.6b), in qualitative agreement with our model simulations (Figure 2.2c) and past dynamic clamp studies [54]. Further, as was done in the model, we set the synaptic balance in the low and high states to produce approximately the same firing rate ($5.5 \pm .9$ Hz in the low state and 6.0 ± 1.5 Hz in the high state).

The correlated input for a given neuron pair was a mixture of shared and independent excitatory and inhibitory inputs, mimicking the input provided to the model (Figure 2.3a). The partial overlap in the synaptic input produced correlated membrane potential and spike dynamics for every neuron pair in both the low and high states. Our recorded spike trains showed a dependence of spike count correlation on T that was qualitatively similar to that of the model, apparent in the ratio of ρ_T in the high and low states (Figure 2.6c). The ratio was a decreasing function of T, indicating a bias toward synchrony in the high state compared to the low state. This shape was consistent with our model results (Figure 2.3d), although the ratio did not fall substantially below unity in the limit of large T. This suggested that the decrease in gain $|\hat{A}(0)|$ and the increase in variability σ from the low to high state were of similar magnitudes, since in the limit of large T correlation susceptibility is proportional to $\sigma^2|\hat{A}(0)|^2$ (see Methods). A conductance-based simulation using the same synaptic parameters used for dynamic clamp stimulation produced results in agreement with the experiment (Figure 2.9). The favoring of synchrony (T=2 ms) over long timescale correlation (T=200 ms) in the high state was statistically significant in a pairwise analysis across the dataset (Figure 2.6c, inset; $P < 3 \times 10^{-5}$, paired *t*-test). The experiments demonstrated that an increase in the rate of balanced conductance input shapes pairwise correlation so as to favor synchronization over long timescale correlation, thereby verifying the main theoretical predictions of our study.

Our theoretical treatment has ignored the timescale of synaptic input, and has associated all filtering to the membrane and spike properties of the model (Figure 2.4). Correlation transfer with realistic synaptic timescales did quantitatively differ from the case with instantaneous synaptic input (Figure 2.8b). Nevertheless, our theoretical work captured the main effects of

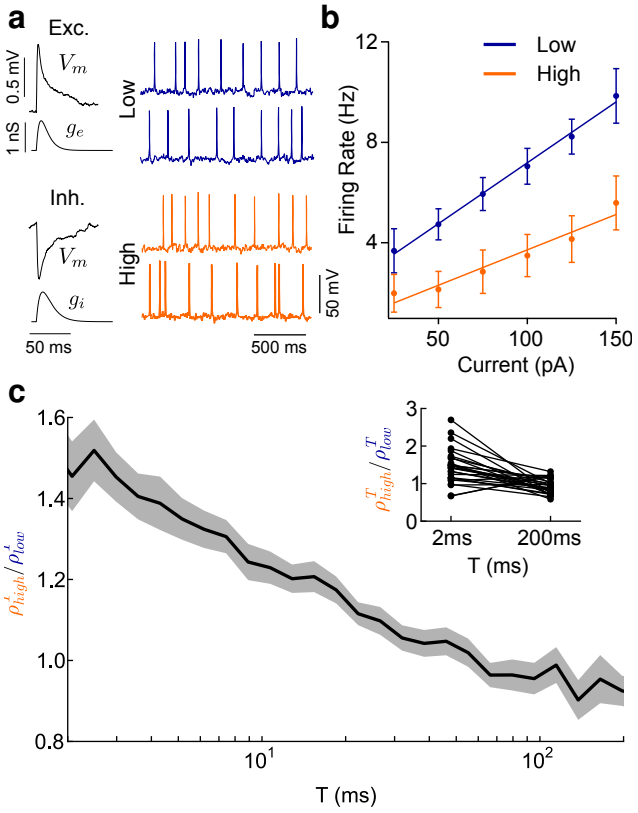

Figure 2.6: Correlation shaping in cortical dynamic clamp experiments. **(a)** Left: Recorded average EPSPs and IPSPs from resting neurons showing membrane voltage (V_m) deflections, with corresponding conductances $g_{e,i}$. Right: Voltage traces from example recorded neuron pairs in high and low states. The degree of synaptic overlap was 0.5 for both high and low states. The inter-spike interval coefficient of variation increased from 0.40 in the low state to 0.48 in the high state. **(b)** Firing rate versus mean input current curves for neurons in low and high states showing reduction in gain in the high background state. **(c)** The ratio of ρ_T in the high to low state as a function of window size T (compare to Figure 2.3d). Curves are population average results (n=28) with the shaded region denoting the standard error. Inset: Correlation ratio shown at T=2ms and 200ms for each recorded pair.

correlation shaping when synaptic timescales were realistic (Figures 2.8 and 2.12). This is because only the effective membrane time constant was sensitive to a shift in input firing rate, which our theory accounts for, while synaptic filtering did not change between low and high states. We remark that, for synapses with very long timescales, correlation shaping should only be present for large T, since correlations at small T will be negligible.

2.4.7 Consequences of correlation shaping for signal propagation

The spike train correlations between neuron pairs substantially influence the propagation of neural activity in feedforward architectures [44]. For example, while our study has so far

focused on the transfer of correlation for neuron pairs receiving common input, the firing rate of a single downstream neuron also depends on the correlation between neurons in its pre-synaptic pool [43]. If the integration timescale of the downstream target is small, only precise spike synchrony will effectively drive the neuron. In contrast, neurons that slowly integrate inputs will be sensitive to long timescale correlations. In our study, we demonstrated that an increase in the rate of synaptic input increases spike count correlation at small T while simultaneously decreasing the correlation at large T (Figure 2.3d). We therefore expected that this correlation shaping would influence the extent to which activity can be propagated to a downstream layer. Further, that the magnitude of this effect would depend on the integration timescales of the downstream targets.

As an illustration of this effect in a simplified system, we studied the firing rate of a downstream neuron receiving input from an upstream population of correlated neurons (Figure 2.7a; see Methods). The level of synaptic drive from layer 1 shaped the correlation of pairs of layer 2 neurons (Figure 2.7a, insets). The network was constructed so that the activity of any given pair of neurons in Layer 2 was equivalent to that of the neuron pairs studied in previous sections. As the correlation of layer 2 spike outputs was shaped, so too was the magnitude and timescale of the synaptic drive to the downstream target neuron (Figure 2.7b). For comparison, we show that downstream target's synaptic input when the layer 2 neurons were uncorrelated (Figure 2.7b, bottom), showing significantly reduced variability [43]. In the uncorrelated case, the firing rate of the downstream target was much less than 1 Hz, indicating that correlated input was necessary for its recruitment.

We study how correlation shaping of the layer 2 projections affected the recruitment of the downstream target neuron. In particular, we focused on how the changing timescale of correlation recruited downstream targets differentially, depending on their own integration properties. We varied the rate of balanced synaptic input from layer 1 to layer 2 in a smooth manner (following the R_e and R_i path for 15Hz output in Figure 2.5a), gradually shaping the covariance function between any given layer 2 neuron pair. The shaping included the low and high states described earlier as near endpoints on a continuum (Figure 2.7c). When the downstream target had a smaller time constant (3 ms), its firing rate was increased when the pre-synaptic population was in the high state (Figure 2.7c, dashed line). This contrasted with the decreased firing rate in the high state when the downstream target had a longer time constant (20 ms)

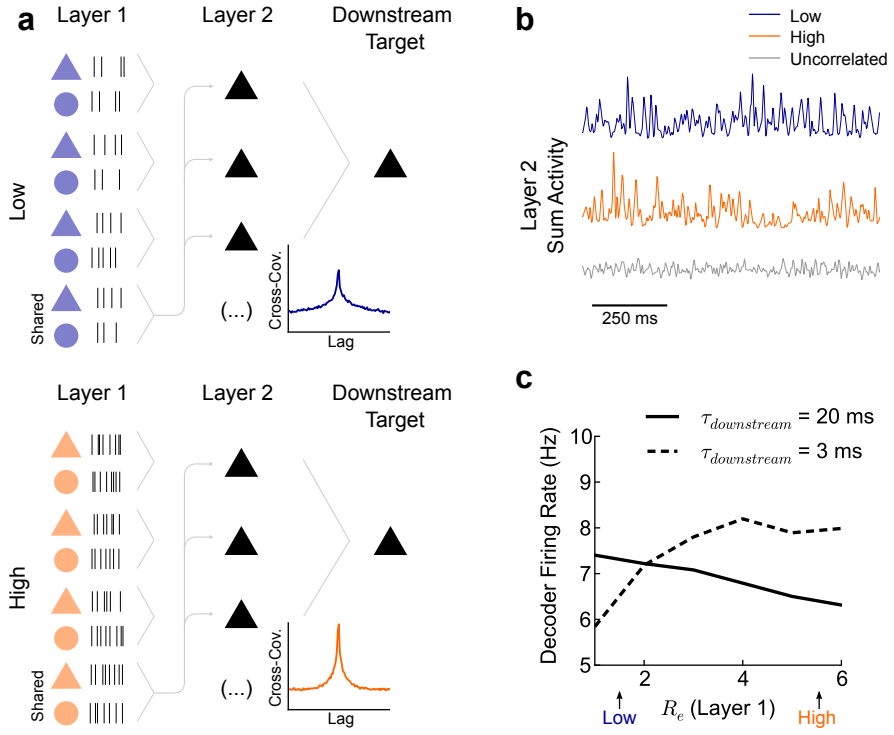

Figure 2.7: Effects of correlation shaping on the propagation of neural activity. **(a)** Schematic of layered network. Layer 1 neurons are modeled as Poisson processes and are either in the low (top) or high (bottom) state. Each layer 2 neuron receives a combination of private input and a globally common input from layer 1. The common input correlates each pair of layer 2 neurons, while the state of layer 1 shapes the correlations (cross-covariance function insets). The layer 2 neurons have marginal and pairwise statistics identical to the neurons in Figures 2.2 and 2.3. The spike outputs of the layer 2 neurons converge onto a downstream neuron with integration timescale $\tau_{downstream}$. **(b)** Example realization of the summed synaptic activity that drives the downstream target neuron in the low (top), high (middle), and, for comparison, when the layer 2 neurons are uncorrelated (bottom). **(c)** Effect of the state change on the downstream neuron's firing rate. The horizontal axis R_e shows the level of excitatory synaptic activity that the neurons in the second layer received from the first layer. R_i is adjusted in a balanced fashion so that the layer 2 neurons fire at 15 Hz (see Figure 2.5A). The downstream target neuron has either $\tau_{downstream} = 3$ ms or $\tau_{downstream} = 20$ ms. The neuron with the fast time constant was driven more strongly in the high state. However, the neuron with the slow time constant showed a decreased firing rate in the high state.

(Figure 2.7c, solid line). This differential effect was due to matching between the correlation timescale of layer 2 and the integration timescale of the downstream target. In the high state, synchrony drove the neuron with the short integration timescale, while, in the low state, long timescale correlations drove the slower neuron. Note that the firing rate of layer 2 neurons was unchanged in all cases studied. This simple example demonstrates that the structure of correlations between pre-synaptic neuron pairs can differentially drive downstream targets depending on their integration properties.

2.5 Discussion

We have demonstrated that the rate of balanced synaptic input changes the correlation timescale of spike trains of a pair of neurons receiving partially correlated input. High rate synaptic input promoted precise spike time synchrony, while low rate synaptic input enhanced long timescale correlation. This correlation shaping was independent of changes in input correlation or the output firing rate of the neuron pair. Rather, it required a thresholding nonlinearity between input and spike train response as well as a state-dependent integration timescale. Both of these are properties of many neurons in the central nervous system, and hence we expect that similar correlation shaping may occur in a variety of brain regions.

2.5.1 Correlation shaping compared to other forms of correlation modulation

Correlated neural activity continues to receive increasing attention [8], prompting investigations of the mechanisms that determine the transfer of correlation. Correlations are typically measured only at one timescale, but as we have shown, the magnitude of correlation depends on the timescale being considered, as does the likely significance of this correlation for activation of downstream neurons. Past studies have highlighted the dependence of spike train correlations on the magnitude of input correlation [46, 86], the form of spike excitability [49, 88, 89], or the firing rate of the neuron pair [10, 48]. However, how the timescale of correlations are modulated through plausible mechanisms had not been addressed. Changes in membrane conductance have been widely studied and strongly influence the dynamics of single neuron activity [23]. In our study, we found that timescale-specific changes in neural correlations are a necessary consequence of conductance based modulation schemes. Previous work that has examined how correlated activity is transferred has used linear response methods to examine the response of neurons to current fluctuations, thereby leaving membrane integration invariant [10, 48]. As a result, cellular properties such as timescale were not modulated (see Figure 2.4d). We showed that when synaptic conductance is considered, it is possible to shape both the magnitude and timescale of output spike train correlations. This is a novel result that is nevertheless consistent with, and complementary to, the observation that firing rate also modulates correlations (see Figure 2.5).

2.5.2 Noise correlation shaping in neural circuits

The widespread use of multi-unit recording techniques to study population activity has produced an increasingly clear picture of how neuronal spike trains are correlated in a variety of neural states. Recently, there has been particular interest in noise correlations, which are specific to within trial comparisons and cannot be directly attributed to a common signal [41].

Several groups have reported noise correlation measurements, ranging from small positive values [79, 12, 37, 14, 15, 16] to values that are, on average, zero, with positive and negative values equally represented [38, 9, 27]. Furthermore, in cases where significant noise correlation is measured, it can be modulated on distinct timescales. In the visual system, for example, noise correlation measured on timescales less than 100 ms is largest for cells with similar preferred stimulus orientations being driven at that orientation, observed in both spike responses [12] and synaptic input [13]. Further, while increasing stimulus contrast enhances short timescale correlation, it reduces long timescale (> 100 ms) correlation [12]. In primate area V4, stimulus attention reduces noise correlation when measured on timescales that are larger than 100 ms, yet has little influence on short timescale correlation [15, 16]. In contrast, other groups have shown that stimulus attention enhances spike synchrony measured at the gamma frequency timescale (20-40 ms) [90]. In the electrosensory system, long timescale noise correlation is reduced by recruitment of a non-classical receptive field, while synchrony is increased under the same conditions [37]. Thus, spike train noise correlations provide an excellent framework to study how the magnitude and timescale of correlations are shaped by neural state changes.

While a shaping of output correlation observed in these systems may be inherited from a state-dependence of input correlation (Figure 2.1b), single neuron response properties are often also modulated by network state. This suggests that a shift in correlation susceptibility may underlie a shift in pairwise correlation (as in Figure 2.1c). Indeed, firing rate gain is modulated by attention [91], stimulus contrast [51], and the recruitment of a non-classical receptive field [92]. In many cases, intracellular recordings have established that gain control is mediated by an increase in the rate of excitatory and inhibitory synaptic inputs [51, 56], in a fashion similar to the case presented in our study. Dual intracellular experiments that measure both input and output correlation across distinct neural states [69, 70, 13] are required to parcel the contribution of correlation inheritance and correlation transfer to the full shift in

noise correlations.

2.5.3 Connecting single neuron and network modulations

A central result of our paper is that changes in synaptic input rate shape the correlation between the output spike trains from a pair of neurons. This is a consequence of how synaptic input modulates the timescale of membrane integration and response sensitivity of the two neurons. Our theoretical analysis formalizes this concept by explicitly relating the spike train correlation coefficient to the single neuron transfer function. Though we focused on modulation by balanced synaptic inputs, the relationship between transfer function and correlation is general, requiring only that the input correlation be sufficiently small. Thus, we predict that any synaptic or cellular mechanism that modulates single neuron transfer will necessarily affect spike train correlations.

Modulation of single neuron transfer with the level of synaptic input rate is well studied [52, 53, 54, 55, 56, 57]. However, how other cellular processes affect neuronal transfer is equally well studied. For example, increases in the spike after-hyperpolarization [72] or decreases in the spike after-depolarization [73] reduce the gain of the firing rate response to static driving inputs. Sustained firing often recruits slowly activating adaptation currents that also reduce gain [74, 75]. We predict that these modulations will reduce long timescale spike rate correlations. In contrast, the presence of low threshold potassium currents in the auditory brainstem [93] promotes high frequency single neuron transfer and thus may also promote pairwise synchronization. In total, our result gives a general theory that links the modulation of single neuron and network responses, thereby expanding the applicability of studies of single neuron modulation.

2.5.4 Selective propagation of neural activity

How the brain selectively propagates signals is a basic question in systems neuroscience. One control mechanism is through an 'unbalancing' of feedforward excitation to inhibition, with disinhibited populations propagating activity and excessive inhibition silencing propagation [94].

Modulation of correlation is an alternative mechanism to control signal propagation. The correlation between spike trains from neurons in a population enhances the ability of that population's activity to drive downstream targets [43, 44]. We have shown that modulating the timescale of correlation in the upstream population to match the integration timescale of the downstream population improves signal propagation (Figure 2.7). Matching the integration dynamics of distinct neuronal populations to one another is a common theme in the binding of distributed activity [95]. In previous studies, the phase relationship between distinct neuronal populations both oscillating at some frequency gated the interaction between distinct brain regions. Our study did not assume rhythmic population dynamics, but rather only matched integration timescales.

The nonlinearity of spike generation allows for the transfer of shared input to multiple neurons to be controlled in complex ways. We have shown that well-studied mechanisms of single neuron response modulation, such as firing rate gain control, have direct relations to changes in correlation for neuron pairs. Thus, state dependent shifts in single neuron transfer also influence how populations of neurons coordinate their activity. Our results are a step in understanding how the collective behavior of neuronal networks can be controlled in different brain states.

2.6 Supplementary figures

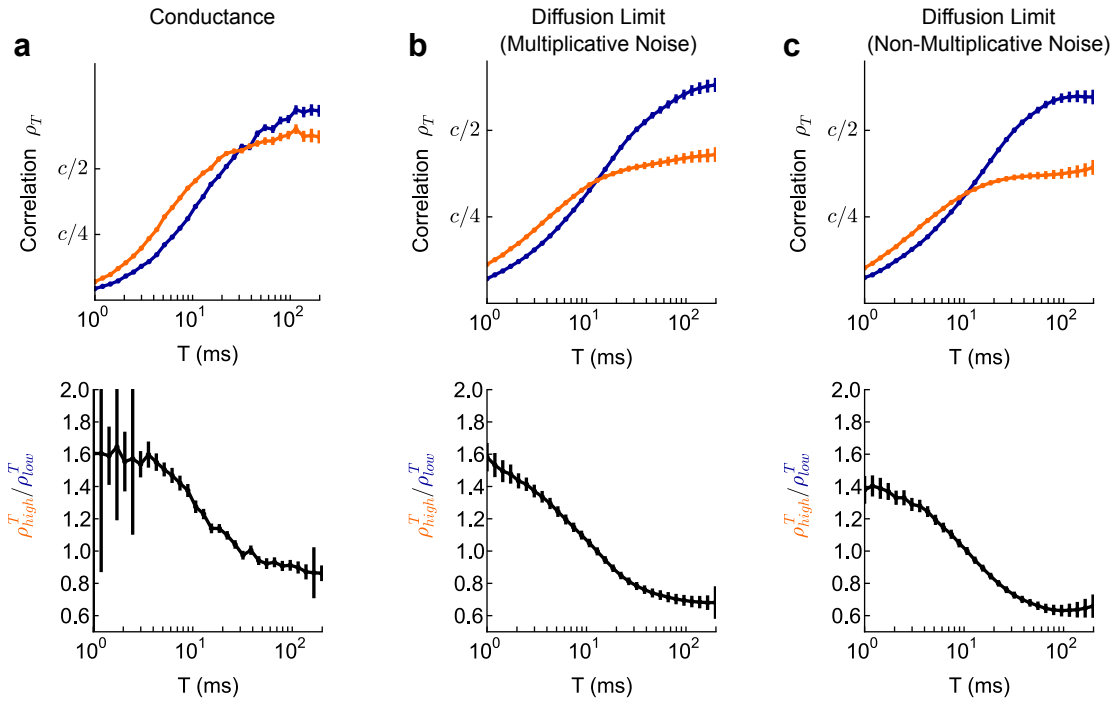

Figure 2.8: Diffusion limit shows qualitative effects of correlation shaping. **(a)** Top: Correlation in the low and high states for a conductance-based model with alpha-function synapses. The excitatory time constant was 2.5 ms and the inhibitory time constant 5 ms. The amplitude of the alpha function was taken so that it matched with the delta-function synapses described earlier. Other parameters were as before. Bottom: Ratio of correlations between the high and low states. **(b)** Same as (a), but after taking the diffusion approximation (see equation (2.1)). **(c)** Same as (b), but after taking $\sigma(V) = \sigma(E_{eff})$. The ratio $\rho^T_{high}/\rho^T_{low}$ exhibits similar correlation shaping in all cases.

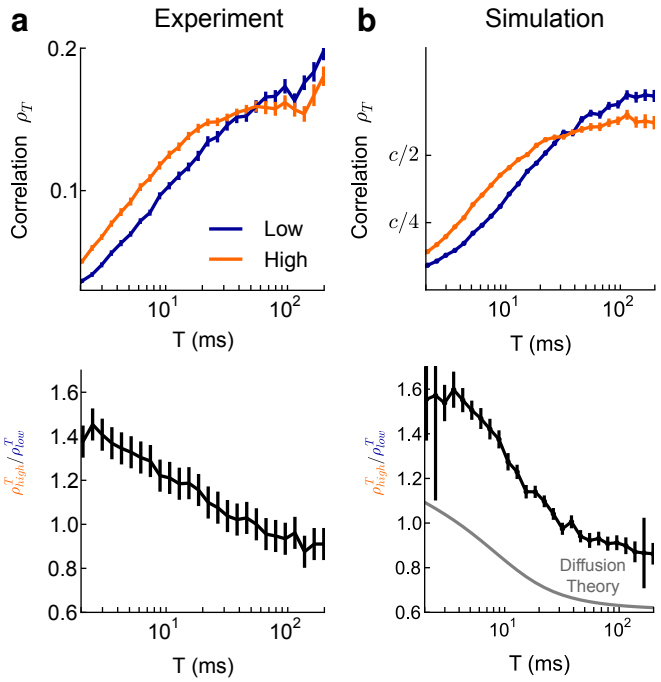

Figure 2.9: Comparison between simulation and experimental results. **(a)** Top: Correlation in the low and high states calculated from dynamic clamp experiments. Bottom: Ratio of correlations between the high and low states. **(b)** Similar to (a), showing results from a conductance-based model with alpha-function synapses. The excitatory time constant was 6 ms and the inhibitory time constant 8 ms. The firing rate was 5 Hz to match experiments. Other parameters were as before.

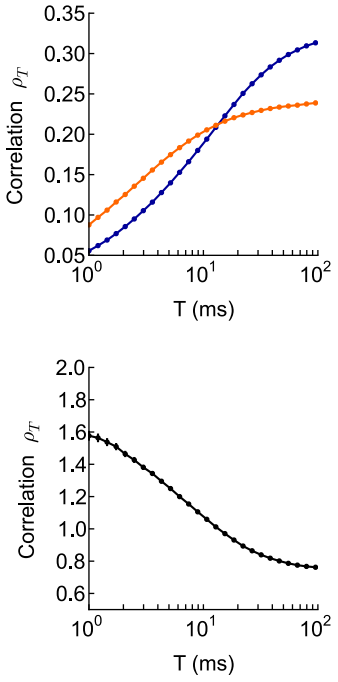

Figure 2.10: Results hold for large c. Top: Correlation in low and high states for $c = 0.5$, parameters otherwise identical to Figure 2.3. Bottom: Ratio of correlations in the low and high states.

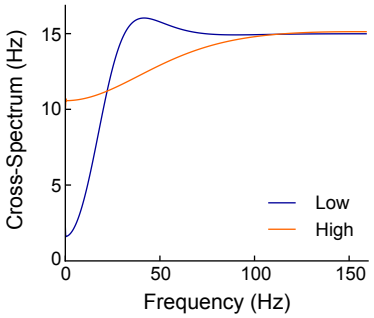

Figure 2.11: Change in power spectrum of the spike train $y(t)$ from low to high states. In both cases, the high-frequency limit of the power spectrum is equal to the firing rate of the neuron. For low frequencies, however, the power was increased in the high state, reflecting the increased variability of firing in the high state (note that as frequency $\to 0$, the power spectrum is equal to the firing rate multiplied by the square of the inter-spike interval CV). To determine the denominator of equation (2.7), we integrate the power spectrum by $k_T(f)$ to obtain Var(n^T). When T is small, Var(n^T) is identical in the two states, because the high frequency limits of the power spectrum are equal. When T is large, Var(n^T) is increased in the high state, because the low frequency limit of the power spectrum is enhanced.

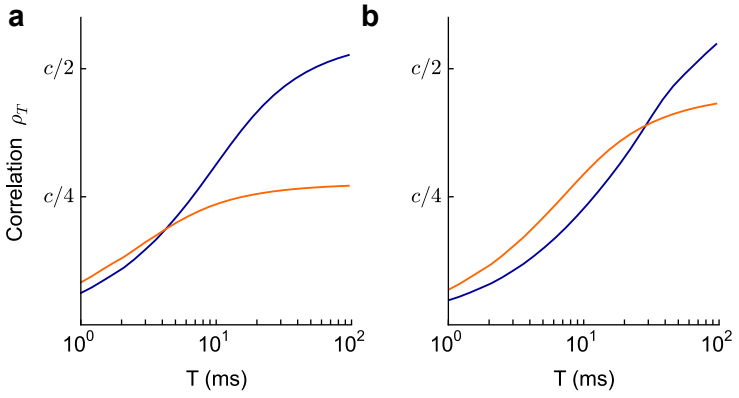

Figure 2.12: Correlation shaping occurs for different synaptic strengths. **(a)** Theoretically calculated correlation curve for $a_e = .02$, $a_i = .04$, $R_e = 1$ kHz in the low state and 4.08 kHz in the high state. The time constant decreased from 7.5 ms in the low state to 1.9 ms in the high state. Firing rates were 15 Hz in both states. **(b)** Theoretically calculated correlation curve for $a_e = .005$, $a_i = .01$, $R_e = 2$ kHz in the low state and 8 kHz in the high state. The time constant decreased from 15.2 ms in the low state to 4.8 ms in the high state. Firing rates were 15 Hz in both states.

3. *Slow dynamics and high variability in balanced cortical networks with clustered connections*

3.1 Abstract

Anatomical studies demonstrate that excitatory connections in cortex are not uniformly distributed across a network but instead exhibit clustering into groups of highly connected neurons. The implications of clustering for cortical activity are unclear. We study the effect of clustered excitatory connections on the dynamics of neuronal networks that exhibit high spike time variability due to a balance between excitation and inhibition. Even modest clustering substantially changes the behavior of these networks, introducing slow dynamics during which clusters of neurons transiently increase or decrease their firing rate. Consequently, neurons exhibit both fast spiking variability and slow firing rate fluctuations. A simplified model shows how stimuli bias networks toward particular activity states, thereby reducing firing rate variability as observed experimentally in many cortical areas. Our model thus relates cortical architecture to the reported variability in spontaneous and evoked spiking activity.

3.2 Introduction

Cortical neurons receive input with strong temporal fluctuations and produce spike trains with high variability [2, 96]. Models of recurrent cortical networks often employ balanced excitation and inhibition to account for this variability [22, 25, 97]. Consistent with these studies, balanced excitatory and inhibitory synaptic input has been measured in cortex [28] and is thought

to substantially influence cortical processing [23]. However, along with spike emission variability, cortical neurons show firing rate fluctuations over long timescales [12, 17, 98]. These stochastic dynamics reflect large trial-to-trial variability in cortical responses [99, 100, 101, 102]. Balanced cortical networks with simple uniform connection structures do not capture these dynamics [25, 27]. Our study's central aim is to uncover which network features, beyond balanced architecture, are responsible for slow firing rate fluctuations in cortical networks. Recent studies show that synaptic connections between excitatory neurons in cortex are clustered rather than uniform [103, 104]. In visual cortex, clusters are related to processing, with highly connected neurons within a cortical column receiving similar visual input [105, 106]. Clustering may also be related to activity level, with frequently firing neurons participating in clustered subnetworks [107]. In total, this suggests cortical wiring involving distinct, densely connected functional subnetworks.

Functional subnetworks are a popular architectural feature supporting attractor dynamics in network models [108, 109, 110, 111, 112]. Such models typically have neuronal subpopulations with increased local excitation, leading to a multitude of stable states in which particular subpopulations exhibit sustained firing. While attractor models have been studied in the context of working memory, their connection to fluctuations in spontaneous and stimulus-evoked cortical dynamics has not been examined.

We investigate the dynamical consequences of clustered excitatory connections in balanced networks. In these networks, specific neuronal clusters increase or decrease their firing rates over long timescales, promoting high trial-to-trial variability. Stimuli that excite particular clusters bias activity in those clusters while recurrent inhibition suppresses others, placing the network collectively into particular activity states. Thus, stimulation reduces firing rate variability while Poisson-like spiking variability remains, consistent with recent cortical recordings [17]. Our results show that even small perturbations from uniform connection structures – here a rewiring of only three percent of excitatory connections – can substantially change balanced network dynamics.

3.3 Methods

3.3.1 Spiking network simulations

Neurons were modeled as leaky integrate-and-fire units whose voltages obeyed:

$$\dot{V} = \frac{1}{\tau}(\mu - V) + I_{syn}(t). \tag{3.1}$$

When neurons reach a threshold $V_{th} = 1$, a spike is emitted and they are reset to $V_{re} = 0$ for an absolute refractory period of 5 ms. The membrane time constant τ was 15 ms and 10 ms for excitatory and inhibitory neurons, respectively. The bias μ was chosen according to a uniform random distribution between 1.1 and 1.2 for excitatory neurons and between 1 and 1.05 for inhibitory neurons. Although these values are suprathreshold, balanced dynamics ensured that the mean membrane potentials were subthreshold [25]. In Figure 3.1, the nondimensionalized voltages were transformed so that $V_{th} = -50$ mV and $V_{re} = -65$ mV.

Synapses between neurons were modeled as differences of exponentials, and the total synaptic input to neuron i in population x was:

$$I^x_{i,syn}(t) = \sum_{jy} J^{xy}_{ij} F^y * s^y_j(t), \tag{3.2}$$

where $x, y \in \{E, I\}$ denote excitatory or inhibitory populations of $N^E = 4000$ and $N^I = 1000$ neurons each, J^{xy}_{ij} is the strength of synaptic connections from neuron j in population y to neuron i in population x, $F^y(t)$ is the synaptic filter for projections from neurons in population y, $*$ denotes convolution, and $s^y_j(t)$ is the spike train of neuron j in population y, a series of delta-functions at time points where the neuron emits a spike.

$$F^y(t) = \frac{1}{\tau_2 - \tau_1}\left(e^{-t/\tau_1} - e^{-t/\tau_2}\right), \tag{3.3}$$

with $\tau_2 = 3$ ms for excitatory synapses and 2 ms for inhibitory synapses while $\tau_1 = 1$ ms. Connection probabilities p^{xy} from neurons in population y to x were $p^{EI} = p^{IE} = p^{II} = 0.5$; p^{EE} was on average 0.2 for all networks but depended on clustering as described in the Results. If a connection from neuron j in population y to neuron i in population x existed, $J^{xy}_{ij} = J^{xy}$ (unless

$x, y = E$ and neurons were in the same cluster, then connection strength was multiplied by 1.9 for homogeneous clustered networks and 1.8 for heterogeneous clustered networks); otherwise $J_{ij}^{xy} = 0$. Synaptic strengths were $J^{EE} = 0.024$, $J^{EI} = -0.045$, $J^{IE} = 0.014$, and $J^{II} = -0.057$. These parameters, multiplied by 15 mV, would give the deflection of the membrane potential of the postsynaptic target, neglecting leak, in our dimensionalized units. When clusters of excitatory neurons were stimulated (Figure 3.7), stimulation was accomplished by increasing μ for neurons in those clusters by 0.07. Simulations were performed using Euler integration with a timestep of 0.1 ms.

To generate an imbalanced network (Figure 3.2, dashed curve), we assumed weak synapses by reducing J^{EE} and J^{IE} by a factor of 20 and chose μ^E uniformly between 1.05 and 1.15. Consequently, excitatory neurons were mean-driven rather than fluctuation-driven.

Finally, to vary the size of the network (Figure 3.4e), we kept all network parameters the same except that we scaled the connection strengths J^{xy} by a factor proportional to $1/\sqrt{N}$, so that the J^{xy} were multiplied by $\sqrt{2}$ for the network with $N = 2,500$ and $1/\sqrt{2}$ for $N = 10,000$.

3.3.2 Visualizing and measuring clustering

Visualizations of connectivity for subsets of the excitatory networks in Figures 3.1 and 3.4 were created using NetworkX (`networkx.lanl.gov`). The layout of nodes was performed using the Fruchterman-Reingold force-directed algorithm [113], which places connected nodes nearby and yields layouts that respect the clustering of the network. For the purposes of these analyses, connections were taken to be undirected and unweighted.

3.3.3 Spike train statistics

Spike train statistics were computed for excitatory neurons. We denote the spike times of neuron i as $\{t_{i1}, t_{i2}, t_{i3}, \dots\}$. We can then define neuron i's spike train: $y_i(t) = \sum_k \delta(t_{ik})$. The number of spikes emitted by the neuron between times t and $t + \Delta t$ is

$$N_i(t, t + \Delta t) = \int_t^{t+\Delta t} y_i(t')dt'. \tag{3.4}$$

The firing rate of a neuron over an interval $(t, t + \Delta t)$ was defined as

$$r_i(t, t + \Delta t) = \frac{1}{\Delta t} N_i(t, t + \Delta t). \tag{3.5}$$

For the networks studied, firing rates and other statistics for the spontaneous state were calculated with t = 1.5s to prevent effects due to initial conditions and Δt = 1.5s.

We also computed the Fano factor $F_i(t, t + \Delta t)$ for neuron i by evaluating

$$F_i(t, t + \Delta t) = \frac{\text{Var}\left(N_i(t, t + \Delta t)\right)}{\langle N_i(t, t + \Delta t)\rangle}, \tag{3.6}$$

where the expectations are over repeated trials of the same network with random initial conditions. When computing the Fano factor as a function of time relative to stimulus onset, we computed the mean-matched Fano factor described in Churchland et al. [17] to control for changes in firing rate. Fano factors were computed over 100-ms windows.

We computed correlation coefficients for the spike counts of neuron pairs. The correlation between neurons i and j was given by

$$\rho_{ij} = \frac{\text{Cov}\left(N_i(t, t + \Delta t), N_j(t, t + \Delta t)\right)}{\sqrt{\text{Var}\left(N_i(t, t + \Delta t)\right)\text{Var}\left(N_j(t, t + \Delta t)\right)}}, \tag{3.7}$$

where the covariances and variances were computed over overlapping windows within each trial and then averaged across trials.

Finally, we computed the mean-subtracted spike train autocovariance and cross-covariance functions $C_{ii}(T)$ and $C_{i\neq j}(T)$. These were defined as

$$C_{ij}(T) = \int_t^{t+\Delta t} y_i(t')y_j(t' - T)dt' - r_i(t, t + \Delta t)r_j(t, t + \Delta t). \tag{3.8}$$

To estimate this quantity, spike trains were discretized with a timestep of δt = 2ms, so that $y_i(t) \to y_i(t_n)$ with $t_n = n\delta t$ and $y_i(t_n) = N_i(t_n, t_n + \delta t)$. Then $C_{ij}(T) = \sum_n y_i(t_n)y_j(t_n - T) - r_i(t, t + \Delta t)r_j(t, t + \Delta t)$ with $T = \ldots, -\delta t, 0, \delta t, \ldots$ and n such that $t_n \in (t, t + \Delta t)$.

3.3.4 Analysis of simplified network model

For our binary network model, the activity of neuron i in population x was given by $s_i^x(t) = \Theta\left(I_i^x(t)\right) \in \{0,1\}$, where $I_i^x(t)$ is the input to the neuron and Θ is the Heaviside step function. For simplicity, excitatory and inhibitory populations both consisted of $N^E = N^I = 4,000$ neurons with an average connection probability of 0.2 for all connection types, so that neurons received $K = 800$ connections on average from each population. The input was given by

$$I_i^x(t) = \sum_{jy} J_{ij}^{xy} s_j^y(t) + \mu^x, \tag{3.9}$$

where $\mu^E = 0.075\sqrt{K}$, $\mu^I = 0.05\sqrt{K}$. The presence of synaptic connections, including the effects of clustering, was determined in the same way as in the spiking networks, although for our initial analysis only one cluster was generated in the network.

In balanced networks, it is assumed that the synaptic strengths scale as $J \sim 1/\sqrt{K}$ so that the input variance remains constant as K grows [25]. In our network, the synaptic strengths were $J^{EE} = 1/\sqrt{K}$ (unless neurons belonged to the same cluster, in which case connection strength was multiplied by 1.5), $J^{EI} = -1.2/\sqrt{K}$, $J^{IE} = 1/\sqrt{K}$, and $J^{II} = -1/\sqrt{K}$. Simulations were performed using discrete time steps, on each of which $N^E + N^I$ neurons were chosen randomly and updated asynchronously.

For uniform networks, the average activity of neurons in each population $\langle s^E \rangle$, $\langle s^I \rangle$ can be calculated in the infinite K limit [25]:

$$\begin{aligned} \langle s^E \rangle &= \frac{J^{II}\mu^E - J^{EI}\mu^I}{\sqrt{K}(J^{EI} - J^{II})} \\ \langle s^I \rangle &= \frac{\mu^E - \mu^I}{K(J^{EI} - J^{II})}. \end{aligned} \tag{3.10}$$

Furthermore, by assuming that each neuron is independent and active randomly and sparsely, the input variances can be calculated:

$$\mathrm{Var}(I^E) = (J^{EE})^2 K \langle s^E \rangle + (J^{EI})^2 K \langle s^I \rangle \tag{3.11}$$

$$\mathrm{Var}(I^I) = (J^{IE})^2 K \langle s^E \rangle + (J^{II})^2 K \langle s^I \rangle. \tag{3.12}$$

Assuming the input distribution is Gaussian, we have

$$\langle s^x \rangle = \frac{1}{2}\text{erfc}\left(\frac{-\langle I^x \rangle}{\sqrt{2\text{Var}(I^x)}}\right), \tag{3.13}$$

which allows us to infer the mean input $\langle I^x \rangle = -\sqrt{2\text{Var}(I^x)}\text{erfc}^{-1}\left(2\langle s^x \rangle\right)$.

To analyze the effect of clustering, we calculate the deviation from the mean input in a uniform network induced by the presence of clustered connections. This approach has been taken in previous studies of memory networks [114, 110, 111]. In a uniform network, the mean input is given by

$$\langle I^E \rangle = KJ^{EE}\langle s^E \rangle + KJ^{EI}\langle s^I \rangle + \mu^E. \tag{3.14}$$

For a clustered network, we assume that the mean firing rate of neurons outside the cluster is given by $\langle s^E_{\text{out}} \rangle = \langle s^E \rangle$, while inside the cluster $\langle s^E_{\text{in}} \rangle$ may vary. Hence, the mean input to neurons in the cluster is

$$\langle I^E_{\text{in}} \rangle = K^{EE}_{\text{in}} J^{EE}_{\text{in}} \langle s^E_{\text{in}} \rangle + (K - K^{EE}_{\text{in}}) J^{EE}_{\text{out}} \langle s^E_{\text{out}} \rangle + KJ^{EI}\langle s^I \rangle + \mu^E, \tag{3.15}$$

where K^{EE}_{in} and denotes the average number of connections from inside the cluster. With $\mu \sim \sqrt{K}$, the final three terms in (3.15) are $O(\sqrt{K})$ but are balanced by the recurrent network dynamics, ensuring that their contribution does not diverge as K becomes large.

We next calculate $\Delta I^E_{\text{in}} = I^E_{\text{in}} - I^E$:

$$\Delta I^E_{\text{in}} = K^{EE}_{\text{in}}(J^{EE}_{\text{in}}\langle s^E_{\text{in}} \rangle - J^{EE}_{\text{out}}\langle s^E \rangle). \tag{3.16}$$

For cluster activity to have an effect on the mean input a neuron receives, this difference must be $O(1)$, which implies that $K^{EE}_{\text{in}} J^{EE}_{\text{in}} \sim O(1)$. Since $J^{EE}_{\text{in}} \sim 1/\sqrt{K}$ and $p^{EE}_{\text{in}} \sim O(1)$, this implies that $K^{EE}_{\text{in}} \sim \sqrt{K}$, so that the number of local cluster inputs scales as the square root of the total number of connections. In particular, for large K, $K^{EE}_{\text{in}} << K$.

We next analyze the time dynamics of the average cluster activity. The dynamics will follow [26]:

$$\dot{\langle s^E_{\text{in}} \rangle} = -\langle s^E_{\text{in}} \rangle + \frac{1}{2}\text{erfc}\left(\frac{-(\langle I^x \rangle + \Delta I^E_{\text{in}})}{\sqrt{2\text{Var}(I^x)}}\right) \equiv \frac{d}{d\langle s^E_{\text{in}} \rangle} U(\langle s^E_{\text{in}} \rangle), \tag{3.17}$$

where we have defined a potential U as the integral of $\langle \dot{s}^E_{\text{in}} \rangle$ with respect to $\langle s^E_{\text{in}} \rangle$. Minima of this potential correspond to self-consistent solutions of average cluster activity. This potential can be obtained using the equations defined above (for Figure 6b, empirically measured values of $\langle s^E \rangle$ and $\langle s^I \rangle$ were used in the calculations, to correct for deviations from (3.10) due to finite size effects).

Assuming that the fluctuations in $\langle \dot{s}^E_{\text{in}} \rangle$ can be modeled as Gaussian white noise and are small relative to well depth ΔU, the average transition time T out of a potential well should follow the Kramers formula [115]: $T \propto e^{k\Delta U/D}$, where D is the diffusion coefficient governing the fluctuations of the mean activity. While we can calculate ΔU using our theory above, the effective D depends on system size and correlations and cannot be easily evaluated. Nonetheless, we find a relationship between ΔU and T consistent with the Kramers formula (Figure 3.5).

3.4 Results

3.4.1 Clustering of connections yields new dynamics

Many real-world networks exhibit nonuniform connectivity structures, often including clustering of connections between different units [116]. Indeed, clustering of synaptic connections between pyramidal neurons in cortex has been observed in many studies [103, 104, 106]. However, inferring dynamical consequences from these architectural properties is challenging [117]. In cortex, the random nature of synaptic connectivity and individual neuron physiology adds to this difficulty. We used a simple model of excitatory clustering in a recurrent network to investigate its consequences for neural activity.

Our network consisted of 4000 excitatory (E) neurons and 1000 inhibitory (I) model neurons. Connections involving inhibitory neurons were non-specific and occurred with probability $p^{EI} = p^{IE} = p^{II} = 0.5$ [118, 119], where p^{xy} denotes the probability of a connection from a neuron in population y to a neuron in population x. Connections between excitatory neurons occurred with probability $p^{EE} = 0.2$. Synaptic dynamics were consistent with fast acting excitatory and inhibitory neurotransmitters (see Methods).

We began by replicating the known asynchronous dynamics of uniform (non-clustered) balanced networks (Figure 3.1, left). Neuronal membrane potentials exhibited fluctuations due to

a balance between excitation and inhibition [25] (Figure 3.1c, left), and spiking activity was asynchronous (Figure 3.1d, left). We note that our network models required no assumptions on the structure or timescale of external fluctuations, as variability arose solely from internal interactions [25, 97].

We next introduced clustered excitatory connections. The excitatory population was partitioned into clusters of 80 neurons each, and the connection probability p^{EE} was set to p^{EE}_{in} for neuron pairs in the same cluster or p^{EE}_{out} for neuron pairs in different clusters. The ratio $R^{EE} = p^{EE}_{\text{in}}/p^{EE}_{\text{out}}$ controlled neuronal clustering. Higher values of R^{EE} favored connections within a local cluster over non-local connections. The quantities p^{EE}_{in} and p^{EE}_{out} were chosen so that the connection probability between excitatory neurons remained 0.2 when averaged across all pairs. We also increased the synaptic strength for neurons in the same cluster [104].

For a clustered network with $R^{EE} = 2.5$, visualizing the connectivity of a subset of excitatory exposed clear divisions between distinct neuronal clusters (Figure 3.1a,b, right). However, on average only 38 of 800 excitatory connections that these neurons received came from within their local cluster, and others were randomly distributed. In clustered networks, neurons exhibited dynamic transitions between periods of higher and lower firing rate (Figure 3.1c,d, Right). These fluctuations reflected changes in the average firing rate of neurons in the same cluster and coexisted with randomness in the spike times of any individual neuron, yielding dynamics substantially different than those of the uniform network despite the small charge in architecture.

3.4.2 Spiking statistics for uniform and clustered networks

To quantify changes in network dynamics introduced by clustering, we calculated spike train statistics for excitatory neurons in uniform and clustered networks. The firing rate distribution was similar for both network types (3.3 ± 4.1 Hz for clustered networks and 2.0 ± 1.8 Hz for uniform networks; $\pm$ denotes standard deviation across the population). Although neurons in the clustered network exhibited transient high activity states, both networks exhibited low, broadly distributed time-averaged firing rates, consistent with recordings from spontaneously active cortex [120] (Figure 3.2a).

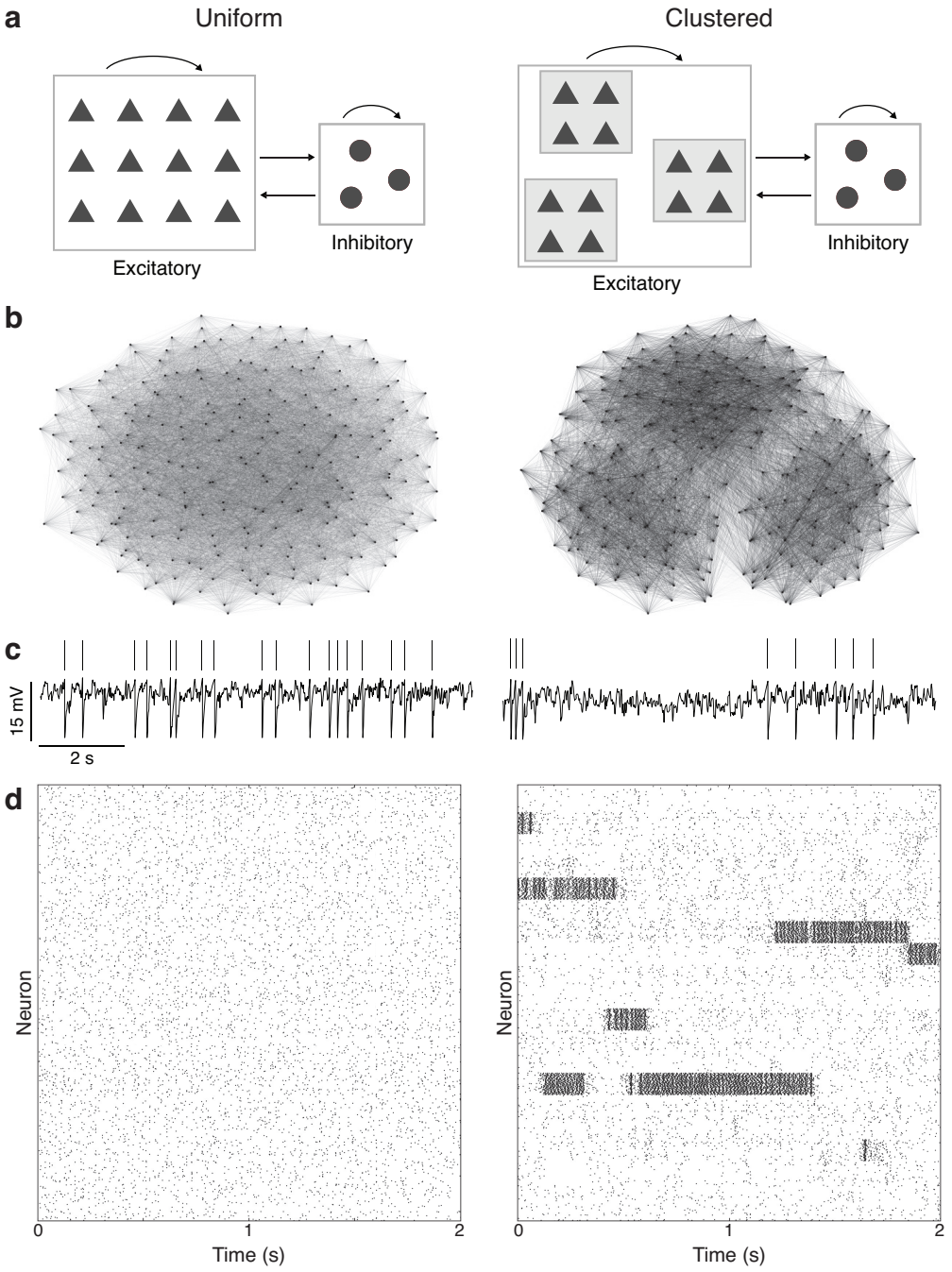

Figure 3.1: Comparison between uniform and clustered network connectivity and dynamics. **(a)** Schematic of recurrent network, showing excitatory (triangles) and inhibitory (circles) neurons. Connections within and between populations were non-specific for the uniform network. Shaded regions in the clustered network schematic indicate subpopulations with increased connection probability and strength for neurons belonging to the same cluster. The full network contained 4,000 excitatory neurons in 50 clusters of 80 neurons each and 1,000 inhibitory neurons. **(b)** Visualization of connectivity for a subpopulation of 240 excitatory neurons in three clusters. Nodes correspond to excitatory neurons and edges synaptic connections. Nodes are positioned according to the Fruchterman-Reingold force algorithm (see Methods). Edge widths reflect synaptic strength. **(c)** Example voltage trace for an excitatory neuron. **(d)** Spike raster showing the spike times of a subpopulation of 1,000 excitatory neurons.

To measure spiking variability, we calculated the spike count Fano factor for each excitatory neuron, defined as the ratio of the trial-to-trial variance to mean of the number of spikes a neuron emits in a fixed time window (here 100 ms; see Methods). For a Poisson process with a fixed firing rate, the Fano factor is 1, and Fano factors above 1 may be interpreted as evidence for variability in a neuron's underlying firing rate. High Fano factors prior to and during stimulus application have been observed in many cortical systems [1, 2, 17]. In uniform networks, the mean Fano factor was 0.78 (S.D. 0.09; Figure 3.2b). In clustered networks, it was substantially larger, with a mean of 1.4 (S.D. 0.7). This additional variability arose from firing rate fluctuations introduced by cluster state transitions from low to high activity (Figure 3.1b). Because these transitions were stochastic, the configuration of highly active clusters for any given moment across trials was variable.

Cortical variability is also reflected in the joint activity of neuron pairs [8]. Excitatory neurons in both network types exhibited near-zero average spike count correlations (0.001 ± 0.06 for clustered networks and 0.0005 ± 0.05 for uniform networks, Figure 3.2c; see Methods), consistent with theoretical studies of balanced networks [27]. However, differences emerged when the calculation of correlation was restricted to neurons in the same cluster. Clustered networks exhibited a distribution of correlation coefficients for intra-cluster pairs with a long positive tail and a mean of 0.13 (S.D. 0.18; Figure 3.2d). The distribution's tail reflected the slow correlated fluctuations in average firing rate that neurons experienced. Because intra-cluster pairs only accounted for 2% of all excitatory neuron pairs, this tail was not visible in the distribution of randomly sampled correlation coefficients (Figure 3.2c).

To analyze this long timescale spiking activity, we computed spike train autocovariance and cross-covariance functions for excitatory neurons (see Methods). Neurons in uniform networks exhibited no long timescales in their autocovariance or cross-covariance functions, while in clustered networks both exhibited long timescale decays (Figure 3.2e,f). Both functions indicated a weak oscillatory tendency in the gamma frequency range due to interactions between excitatory and inhibitory populations [121].

Finally, we explicitly related the system's long timescales to trial-to-trial variability by computing Fano factors as a function of time window. Compared to an imbalanced network in which

responses were mean-driven, Fano factors for balanced networks were substantially larger (Figure 3.2g), though sub-Poisson due to refractory effects (Figure 3.2e), saturating by $T < 100$ ms. However, Fano factors increased for longer time windows in clustered networks. In total, clustered networks had increased slow firing rate fluctuations while retaining the fast spiking variability characteristic of balanced networks, making neuronal activity effectively a "doubly-stochastic" process [122, 17, 98].

3.4.3 Heterogeneity in cluster membership

In previous sections, we studied an idealized model in which all clusters were identically sized and neurons belonged to only one cluster. This connectivity led to discrete and unambiguous transitions between high and low activity states (Figure 3.1). However, cortical assemblies may involve heterogeneity in size and overlap in membership. To determine whether our results hold for heterogeneous clustered networks, we generated a new network architecture in which cluster membership was random. Each excitatory neuron was assigned to two clusters out of 100 total, and connection probability and strength were increased for neuron pairs that shared membership in at least one cluster (Figure 3.3a). Because of the random assignment of neurons to clusters, cluster sizes were heterogeneous (Figure 3.3b).

Because neurons were not ordered by group membership, transitions were not immediately apparent in an unordered raster (Figure 3.3c). However, when rows in the raster were sorted by cluster, slow transitions appeared (Figure 3.3d). To compare homogeneous and heterogeneous clustered network dynamics, we calculated average firing rates for all neurons in individual clusters. Clusters in homogeneous networks exhibited discrete transitions between low and high activity states with average firing rates of 50-60 Hz (Figure 3.3c, top). In contrast, heterogeneous clustered networks exhibited more complicated temporal dynamics (Figure 3.3c, bottom). Hence, networks with clustered architectures can generate complex firing rate patterns reminiscent of spontaneous cortical recordings [101], although our study's primary results can be understood in the idealization of homogeneous clustered networks.

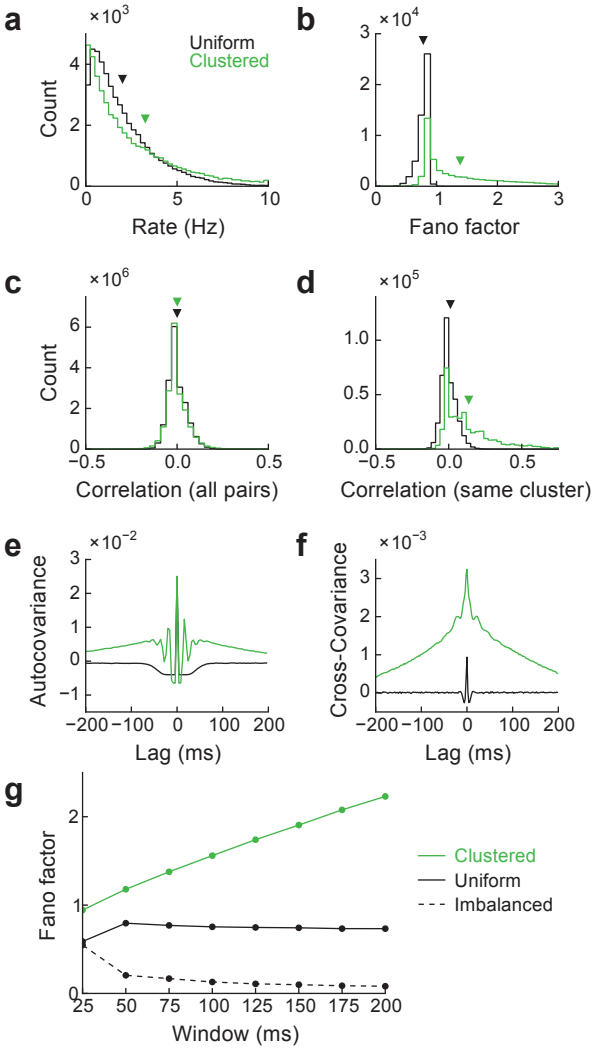

Figure 3.2: Marginal and pairwise spiking statistics for neurons in clustered and uniform networks. Statistics for each network type were computed over 12 realizations for each of the two connectivities, with 9 trials for each realization of connectivity. **(a)** Histogram of excitatory neuron firing rates in clustered and uniform networks. Triangles denote mean value. **(b)** Histogram of excitatory neuron Fano factors computed over 100 ms windows. **(c)** Histogram of correlation coefficients computed over 50 ms windows for all excitatory neuron pairs. **(d)** Same as c, but only computed for neuron pairs belonging to the same cluster. The distribution of correlation coefficients for a random subset of pairs from the uniform network is shown, since the uniform network lacked clusters. **(e)** Average autocovariance function for excitatory neurons. **(f)** Average cross-covariance function for neuron pairs belonging to the same cluster. Correlation functions are normalized to firing rate. **(g)** Fano Factors computed over different counting windows. Uniform, clustered, and imbalanced networks were compared. For imbalanced networks, a balance between excitation and inhibition was not obeyed and responses were mean-driven, rather than fluctuation-driven (see Methods).

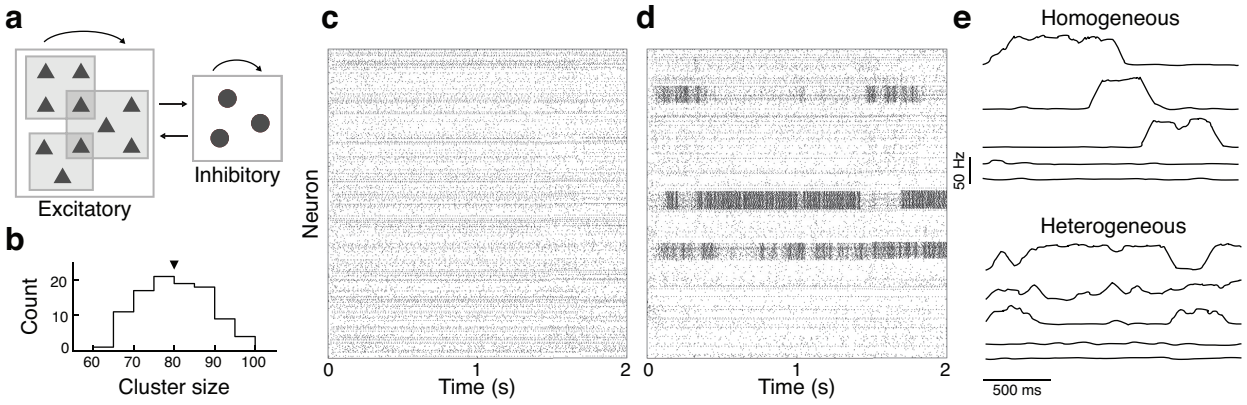

Figure 3.3: Heterogeneous clustered network. **(a)** Schematic of connectivity for heterogeneous network. Each excitatory neuron was assigned randomly to two clusters (schematic only shows a subset of neurons and clusters). Connection probability and strength was increased for any neuron pair that shared membership in at least one cluster. **(b)** Histogram of the number of neurons per cluster for the network shown in the following panels. Cluster size was heterogeneous since membership was assigned randomly. For this network, clusters ranged in size from 63 to 99 neurons. **(c)** Spike raster showing the spike times of 1,000 excitatory neurons. Since cluster membership was chosen randomly, the neurons are not ordered by cluster. **(d)** Same as c, but row indices reordered so that blocks of rows correspond to clusters. Some rows are repeated, since neurons belonged to multiple clusters. **(e)** Example average cluster firing rates for different clusters from the clustered network studied previously (Figure 3.1b) and the heterogeneous network. Rates were estimated from the number of spikes emitted by all neurons in a cluster in a sliding window of 100 ms.

3.4.4 Increased clustering lengthens rate fluctuation timescale

We have shown that clustered connections introduce new, slow dynamics not present in uniform networks (Figure 3.1). We next show that this timescale is related to the density of clustering.

We examined three different clustered network connectivities with R^{EE} increasing from 2 to 3. As R^{EE} was increased, neurons were more easily separated into distinct clusters (Figure 3.4a), and clusters exhibited longer periods of high activity (Figure 3.4b). To quantify the firing rate fluctuations' dependence on network connectivity, we calculated the average excitatory neuron Fano factor as a function of R^{EE} (Figure 3.4c). Fano factors increased for sufficiently high clustering. The degree of clustering R^{EE} can be recast in terms of the percentage of excitatory connections that need to be rewired to generate a clustered network from a uniform one. This quantity was below five percent for the range studied (Figure 3.4d). Thus only a small perturbation from uniform connectivity is required for substantial firing rate fluctuations.

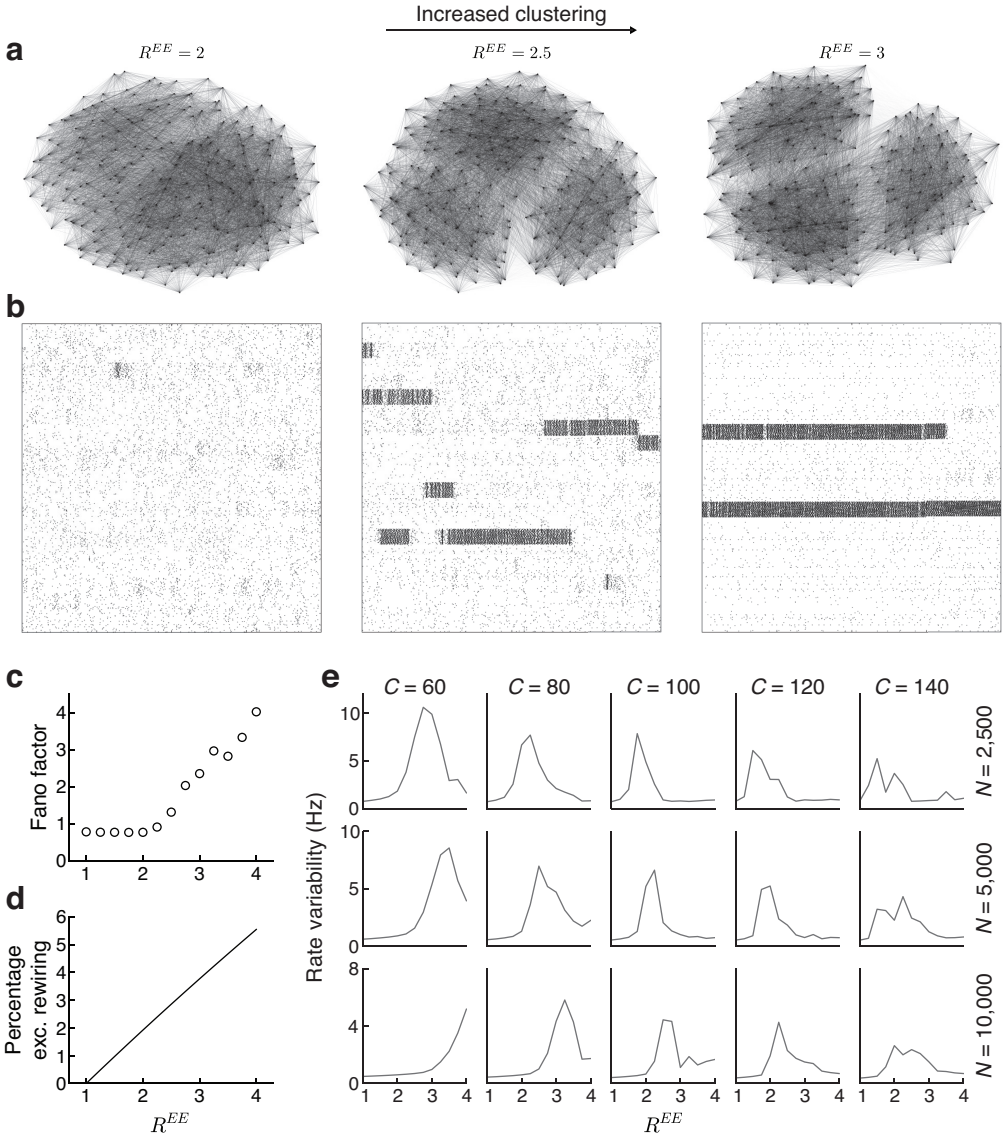

Figure 3.4: Effect of increased clustered connection probability. **(a)** Visualization of connectivity for a subset of 240 neurons in three different networks. The network for $R^{EE} = 2.5$ was identical to the network shown in Figure 3.1, right. **(b)** Spike raster for a subset of 1,000 excitatory neurons from the three different network configurations. As clustering was increased, the timescale of cluster activity state transitions lengthened. **(c)** Average Fano factor computed over 100 ms windows among excitatory neurons as a function of R^{EE} **(d)** Average percentage of excitatory connections rewired as a function of R^{EE}. Percent rewired was defined by calculating the average number of connections an excitatory neuron received from other neurons within its cluster, subtracting the average number of such connections it would have received if $R^{EE} = 1$, and dividing by the total number of connections. **(e)** Dependence of firing rate variability on clustering R^{EE}, for different values of cluster size C and total network size N. The firing rate variability for each cluster was estimated by computing the population firing rate for the C neurons in the cluster over 2 seconds, in 20 windows of 100 ms each. This was then averaged over all clusters and realizations.

3.4.5 Dependence of dynamics on cluster and network size

We next investigated whether changes in network dynamics were robust to different connectivity parameters. Previously, we saw that, above a critical level of clustering (above approximately $R^{EE} = 2.5$), the exhibited high firing rate variability. However, for large R^{EE}, transitions occurred over very long timescales (Figure 3.4b, right). Hence, variability on biophysically relevant timescales should be maximized for a fixed level of clustering, a prediction we investigated with varying network parameters.

We evaluated cluster firing rate variability as a function of R^{EE}, for different network sizes N and cluster sizes C (Figure 3.4e). For each combination of N and C, a critical value of R^{EE} above which firing rate variability increased existed. However, variability eventually decreased for higher R^{EE}, when the transition timescale was longer than the simulation time. This is in contrast to the dependence of the Fano factor on clustering (Figure 3.4c), since Fano factors were computed across trials with different initial conditions while firing rate variability was computed over time.

The shape of these curves depended both on N and C. As C increased, the firing rate variability peak location and height decreased, since more excitatory connections were present and cluster transition times diminished. However, the peak location increased as N grew, since, in balanced networks, synaptic strength scales inversely with network size (see Methods). Since a peak occurred for each combination of N and C, our results are robust to changes in network parameters. We further explore network scaling in the next section.

3.4.6 Theoretical analysis of clustered subnetworks in balanced networks

Our results indicate that network architecture, more so than individual neuron dynamics, gives rise to cluster state transitions. We therefore use a simplified binary neuron model to relate architecture to transitions. Similar analyses have been performed for memory networks [114, 110, 111], but the timescale of spontaneous transitions between different states has not been examined.

Our model simplifies the state of neuron i, denoted s_i, to either "inactive" ($s_i = 0$) or "active" ($s_i = 1$). The input to neuron i, denoted I_i, is a weighted sum of the activities of the neurons

to which it is connected plus a constant mean bias: $I_i = \sum_j J_{ij} s_j + \mu_i$, where μ_i is the bias and J_{ij} is the strength of a synaptic connection from neuron j to neuron i. If I_i is positive, then the neuron is active; otherwise, it is inactive.

We begin by studying a network whose connectivity is uniform except for a single cluster of C neurons. A neuron within the cluster receive $K_{\text{in}}^{EE} = p_{\text{in}}^{EE} C$ synaptic inputs from other neurons in this cluster. The mean and variance of the total input to such a neuron is:

$$\langle I_{\text{in}}^E \rangle = J_{\text{in}}^{EE} K_{\text{in}}^{EE} \langle s_{\text{in}}^E \rangle + J_{out}^{EE} (K - K_{\text{in}}^{EE}) \langle s_{\text{out}}^E \rangle + J^{EI} K \langle s^I \rangle + \mu^E, \tag{3.18}$$

$$\text{Var}(I_{\text{in}}^E) = (J_{\text{in}}^{EE})^2 K_{\text{in}}^{EE} \langle s_{\text{in}}^E \rangle + (J_{\text{out}}^{EE})^2 (K - K_{\text{in}}^{EE}) \langle s_{\text{out}}^E \rangle + (J^{EI})^2 K \langle s^I \rangle. \tag{3.19}$$

Here J^{xy} denotes the strength of synaptic connections from neurons in population y to those in population x, and $\langle s^x \rangle$ corresponds to the average activity of neurons in population x. Subscripts "in" and "out" refer to excitatory neurons within the cluster and outside of the cluster, respectively. Since $\text{Var}(I_{\text{in}}^E) \sim J^2 K$, we need $J \sim 1/\sqrt{K}$ for $\text{Var}(I_{\text{in}}^E)$ to be $O(1)$. As a consequence, I_{in}^E will diverge as $\sqrt{K}$ unless excitation and inhibition balance [25].

Under these conditions, we analyze the effect of the local cluster input. The first term in equation (3.18) is the average recurrent excitation from other neurons in the cluster:

$$I_{\text{rec}} = J_{\text{in}}^{EE} K_{\text{in}}^{EE} \langle s_{\text{in}}^E \rangle. \tag{3.20}$$

For I_{rec} to be $O(1)$ as K gets large, we need K_{in}^{EE} to be inversely proportional to J_{in}^{EE}. Because $J \sim 1/\sqrt{K}$, we therefore require

$$K_{\text{in}}^{EE} \sim \sqrt{K}. \tag{3.21}$$

In other words, for a balanced network, local cluster inputs may be much smaller in number (by a factor of $\sqrt{K}$) than total inputs K but still have a substantial effect. Indeed, as noted previously, neurons received on average only 38 of 800 total excitatory inputs from within their local cluster in our integrate-and-fire simulations. This sensitivity to rewiring is a consequence of the balance between excitation and inhibition; small deviations in balance due to local cluster activity strongly affect neurons' firing rates.

3.4.7 Transitions between low and high activity states

With our simplified model, we next study the dynamics of $\langle s_{\text{in}}^E \rangle$, the average firing rate of neurons within the cluster. We determined the potential energy landscape U that governed this quantity (see Methods). Minima in U corresponded to stable values of $\langle s_{\text{in}}^E \rangle$. For small R^{EE}, the only minimum was at a low firing rate, so no transitions could occur (Figure 3.5a). When R^{EE} increased sufficiently, a second minimum appeared, and $\langle s_{\text{in}}^E \rangle$ was therefore bistable, as is common in networks with local recurrent excitation [108, 109, 114, 110, 111]. Furthermore, because the cluster size was small, dynamic fluctuations due to finite size effects caused transitions between minima. Deeper potential wells corresponded to more attractive activity states, as quantified by average time spent in a well (Figure 3.5b).

Transitions between low and high activity states introduced long timescale variability to the cluster's activity (Figure 3.5c, top). Although transitions between wells were a source of variability in our model, we emphasize that they were not the only source of variability. Balanced excitation and inhibition ensured that neurons were bombarded with many synaptic inputs that produced high input variability relative to input mean, even when conditioned on their cluster being in one state (Figure 3.5c, bottom). Model neurons were therefore doubly stochastic, with spike time variability because of this broad input distribution and firing rate variability because of transitions between wells.

What controls the timescale of these transitions? For deep enough potential wells and assuming that the effective noise driving average cluster activity is Gaussian and white, the transition time out of a well of depth ΔU will scale as:

$$T \propto e^{k\Delta U/D}, \tag{3.22}$$

where k is a constant and D is the amplitude of the effective diffusion driving the fluctuations in mean activity [115] (see Methods). D depends both on system size, with larger clusters exhibiting diminished fluctuations due to averaging, and on correlations between neurons, which amplify fluctuations in mean activity.

As increased clustering increases the high activity state's well depth (Figure 3.5a), we expect highly clustered networks to take longer to transition out of this state. To verify this prediction,

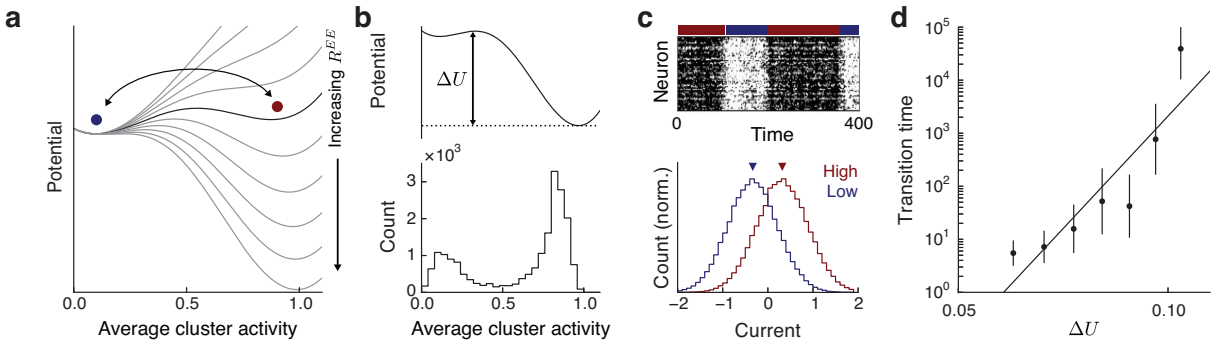

Figure 3.5: Emergence of bistability in simplified model. **(a)** Potential governing dynamics of average cluster activity $\langle s^E_{\text{in}} \rangle$, for different values of clustering R^{EE}. As clustering is increased beyond a critical value, bistability emerges and switches between the two stable states can occur (black curve). **(b)** Potential (top) and corresponding histogram for average cluster activity (bottom). **(c)** Top: Example activity raster for the parameters in **b**. Transitions between the high activity (red) and low activity (blue) states are evident. Bottom: Histogram of input currents for neurons in the cluster conditioned on the cluster being in the low or high activity state (defined as average activity less than or greater than 0.5, respectively) and normalized in peak height. **(d)** Average timescale of transitions from high activity to low activity state as a function of well depth ΔU. Each point corresponds to an average over four networks simulated for 200,000 timesteps. The line corresponds to a linear regression of well depth against the logarithm of the transition time ($R^2 = 0.89$).

we measured the average transition time out of the high activity state for networks with different degrees of clustering and plotted the result as a function of ΔU (Figure 3.5d). The logarithm of T increased linearly with ΔU as predicted by equation (3.5). In total, architectural clustering controls U and hence the presence of bistability and the attractiveness of high activity states.

We next seek to understand how clusters interact. So far, our analysis has focused on a single cluster's dynamics, neglecting its influence on other clusters or the possibility of other clusters changing state. Indeed, if every cluster was independent and all were governed by the potential in Figure 3.5b, we would expect most clusters to remain in the high activity state and average activity to be unrealistically high. In the fully recurrent network, this is not the case. Recurrent inhibition mediates cluster-cluster interaction, limiting the number of clusters simultaneously in the high activity state and thus promoting transitions. We examine this more closely as we next consider the relationship between stimulation and clustered network dynamics.

3.4.8 Biasing activity state with stimulation

Recently, the relationship between spontaneous and evoked trial-to-trial variability has attracted attention [17, 123]. Throughout the cortex, trial-to-trial spiking variability in sensory systems is

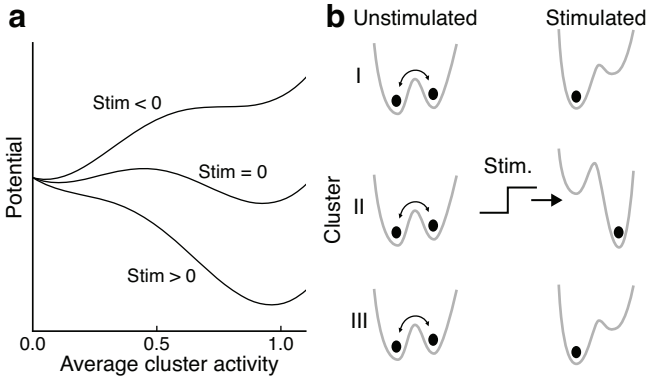

Figure 3.6: Effects of stimulation. **(a)** Potentials for $R^{EE} = 2$ and different values of stimulus applied to the cluster. A stimulus can bias activity toward the low or high activity state depending on its sign. **(b)** Consequences of stimulation and recurrent inhibition. Left: When all clusters are symmetric, any cluster can transition to the high activity state. Right: When Cluster 2 receives a stimulus to bias it to the high activity state, other clusters are suppressed by recurrent inhibition. Hence, the network remains in one particular activity configuration.

reduced when a stimulus is presented [17]. Despite its generality, the mechanisms responsible for this variability reduction are undetermined. With our theory, we examined stimulation's effects on variability in clustered networks.

We calculated U for different stimulus values (Figure 3.6a). Negative stimuli made the high activity state less attractive, while positive stimuli increased its attractiveness. This, along with previous observations about recurrent inhibition, leads to a prediction about network dynamics under stimulation. Without stimulation, clusters are symmetric and any cluster may transition between states (Figure 3.6b, left). When a fraction of clusters are stimulated, this symmetry is broken and stimulated clusters are more likely to have high activity. Recurrent inhibition discourages other clusters from transitioning to high activity states, forcing the system into a single configuration and thus suppressing long timescale variability (Figure 3.6b, right). We next investigate this prediction, returning to a full spiking network.

3.4.9 Effect of stimulation on spiking variability

Because our spiking model exhibited high trial-to-trial variability consistent with cortical data, we investigated stimulation's effect, using a simple depolarizing stimulus applied to 5 of the 50 clusters for 400 ms. During spontaneous activity, clusters exhibited transitions. When the stimulus was applied, stimulated clusters became highly active, while unstimulated clusters

were suppressed (Figure 3.7a,b, compare shaded and unshaded region in a). After stimulus application ceased, the system eventually relaxed to spontaneous dynamics.

To quantify reductions in variability, we calculated the mean matched Fano factor over 100 ms windows as a function of time. Mean matching controlled for stimulus-induced changes in firing rate, ensuring that observed changes in Fano factor were due to a change in underlying rate variability [17]. For clustered networks, Fano factors were above 1 during spontaneous activity but dropped slightly below 1 during stimulation (Figure 3.7c). Similar results were obtained for heterogeneous clustered networks (Figure 3.8), clustered networks with more complex stimulus tuning (Figure 3.9), and clustered networks with ring or feed forward structure (Figure 3.10). In contrast, uniform networks exhibited no noticeable decrease in variability when the stimulus was applied (Figure 3.7c). The Fano factor restricted to either stimulated or unstimulated neurons also exhibited a reduction in Fano factor only for clustered networks (Figure 3.7d). These results are consistent with cortical recordings in which neurons whose preference did not match a presented stimulus nevertheless showed a stimulus-induced drop in variability (Figure 5 in Churchland et al. [17]), a feature not present in other models of cortical variability [112].

Because our spiking model exhibits high trial-to-trial spiking variability consistent with cortical data, we investigated the effect of simple external stimuli to a subset of clusters in our spiking networks.

3.4.10 Variability of inhibitory neurons

To this point we have been interested in excitatory neuron spiking variability, but inhibitory neuron variability may also be stimulus dependent. In visual area V4, a reduction in inhibitory neuron Fano factor with stimulation has been reported [124], but firing rates increased substantially during stimulation, and no rate matching was performed to control for this modulation. We show that inhibitory neurons in clustered networks may also exhibit decreased Fano factors due to stimulation, an effect that depends on the specificity of their afferent connections from excitatory neurons.

In the above model the connections involving inhibitory neurons were uniform. Anatomical evidence suggests that inhibitory innervation of pyramidal neurons is nonspecific [119],

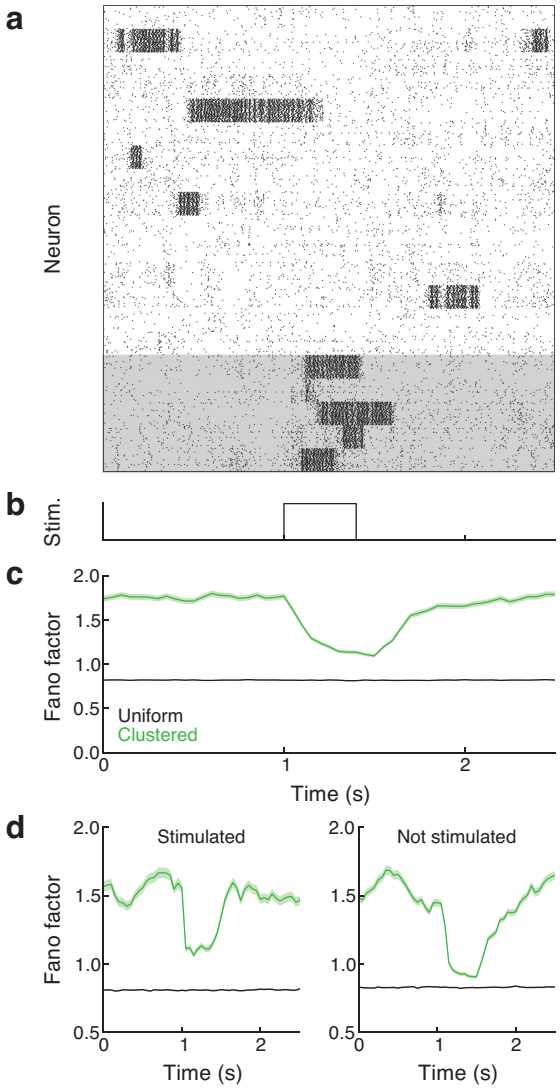

Figure 3.7: Effect of stimulation on spiking variability and dynamics. **(a)** Spike raster showing the activity of 1,000 excitatory neurons on one trial. Neurons in the shaded region received the stimulus when applied. **(b)** Time course of stimulus, a depolarizing step of current that lasted for 400 ms. The stimulus caused the clusters that received it to transition into a high activity state (**a**, shaded region). **(c)** Fano factor computed over 100 ms windows as a function of time for excitatory neurons. For comparison, the average Fano factor for uniform networks receiving an identical stimulus is shown. Only clustered networks exhibited a noticeable decrease in variability when stimulated. Shaded regions denote 95% confidence intervals. **(d)** Left: Same as **c**, but restricted to neurons in stimulated clusters (corresponding to **a**, shaded region). Right: Same as left, but for neurons not in stimulated clusters. Fano factors were computed using mean matching techniques presented in Churchland et al. [17].

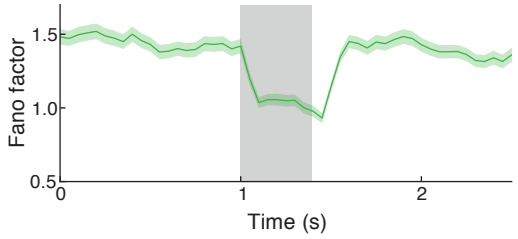

Figure 3.8: Effect of stimulation on spiking variability in the heterogeneous clustered network. The mean-matched Fano factor computed over 100 ms windows as a function of time, averaged over excitatory neurons in the heterogeneous clustered networks, is plotted. Green shaded region denotes 95% confidence interval. Gray shaded region denotes the interval during which the stimulus was applied.

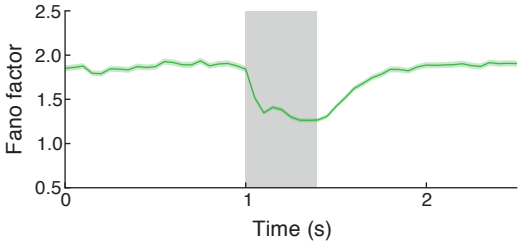

Figure 3.9: Effect of stimulation on spiking variability in a clustered network with randomly assigned stimulus strengths. Instead of stimulating five clusters with a fixed stimulus strength, all excitatory neurons received a stimulus input. For each cluster, the stimulus strength was chosen independently from a Gaussian distribution with zero mean and standard deviation of 0.05. All neurons in the same cluster received the same stimulus strength. The curve shows the mean-matched Fano factor computed over 100 ms windows as a function of time, averaged over excitatory neurons. Green shaded region denotes 95% confidence interval. Gray shaded region denotes the interval during which the stimulus was applied.

but there is some evidence for selectivity in excitatory input to certain classes of inhibitory interneurons [125]. We therefore studied the dependence of inhibitory neuron variability on excitatory-to-inhbitory input specificity.

We compared clustered networks presented previously to ones in which inhibitory neurons were more likely to receive input from excitatory neurons in a particular cluster (Figure 3.11a). In clustered E+I networks, inhibitory neurons were assigned to 50 clusters of 20 neurons each. Excitatory to inhibitory connections occurred with probability p_{in}^{IE} if both neurons belonged to the ith cluster of their respective population and otherwise with probability p_{out}^{IE}. $R^{IE} = p_{\text{in}}/p_{\text{out}}$ was set to 1.5.

Excitatory neurons in both networks had high variability and exhibited a reduction in Fano factor with stimulus (Figure 3.11b). However, inhibitory neuron Fano factors in clustered E+I

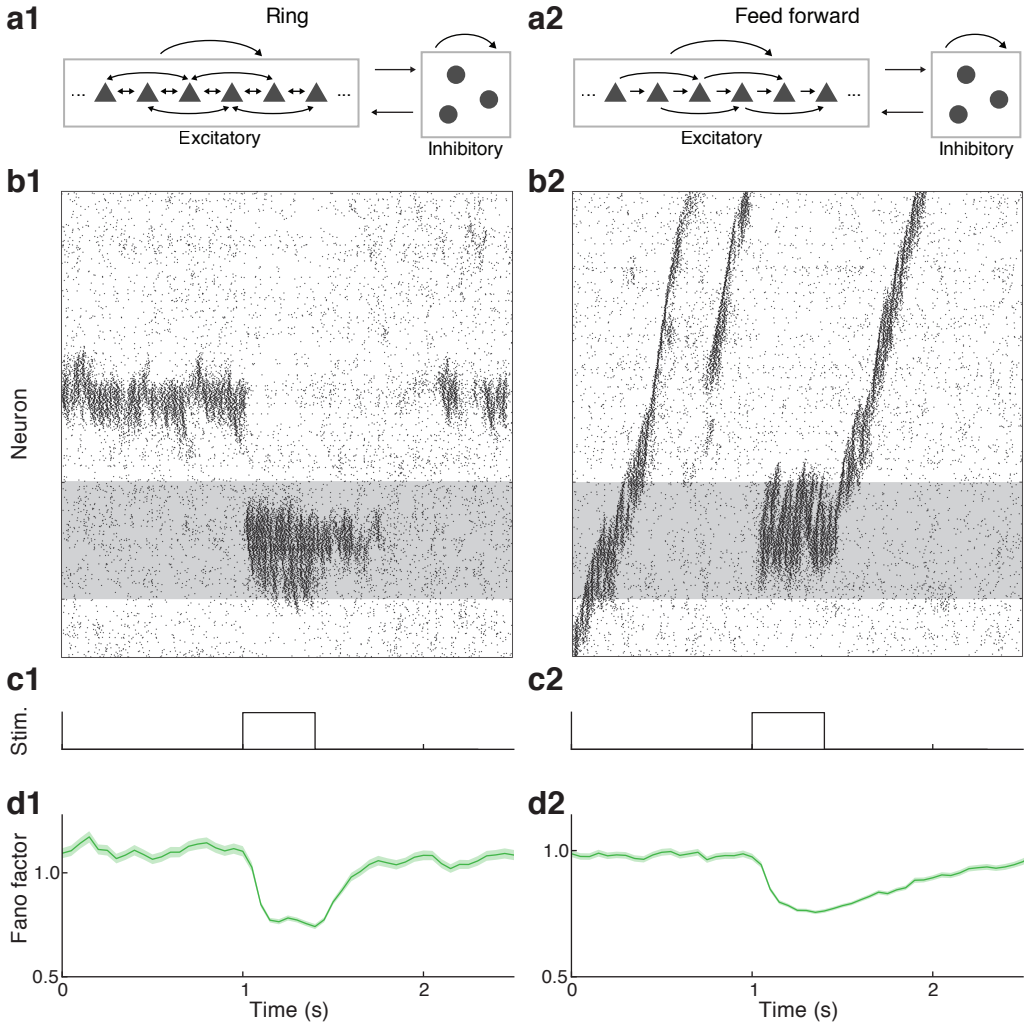

Figure 3.10: Effect of stimulation on spiking variability in ring and feed forward networks. **(a1)** Schematic of ring network connectivity. To generate the ring network, excitatory neurons i and j were connected with probability $p_{\text{in}}^{\text{EE}}$ if $|i-j| < 40$ (with periodic boundary conditions); otherwise, they were connected with probability $p_{\text{out}}^{\text{EE}}$. Hence, the effective cluster size was again 80 neurons, but the clusters overlapped throughout the domain. R^{EE} was set to 2.4. All other parameters were identical to the original clustered network. As in the original clustered networks, J^{EE} was also increased for those neurons that had an increased probability of connection. **(b1)** Raster plot showing a subset of excitatory neurons in the ring network. Gray region denotes those neurons that received the stimulus. **(c1)** Time course of stimulus, a depolarizing current step that lasted for 400 ms. **(d1)** Mean matched Fano factor computed over 100 ms windows as a function of time, averaged over excitatory neurons in the networks. Shaded region denotes 95% confidence interval. **(a2)-(d2)**: Same as (a1)-(d1) but for feed forward networks. In the feed forward network, a connection from excitatory neuron i to j occurred with probability $p_{\text{in}}^{\text{EE}}$ if $i-j \in [-35, 45]$ (with periodic boundary conditions); otherwise, it occurred with probability $p_{\text{out}}^{\text{EE}}$. This resulted in an architecture similar to the ring network, but with a directional bias in the connections, leading to propagating waves which were pinned by stimuli. R^{EE} was set to 2.5. All other parameters were identical to the original clustered network. As in the original clustered networks, J^{EE} was also increased for those neurons that had an increased probability of connection.

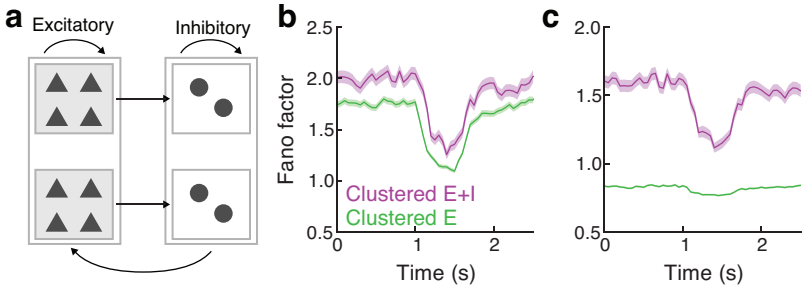

Figure 3.11: Effect of stimulation on inhibitory neuron spiking variability. **(a)** Schematic of network connectivity for the clustered E+I network. Excitatory projections were cluster-specific while inhibitory projections were uniform. **(b)** Mean matched Fano factor computed over 100 ms windows as a function of time for excitatory neurons in the clustered E+I network and the clustered E network (Figure 3.7). Shaded region denotes 95% confidence interval. **(c)** Same as **b** but for inhibitory neurons.

networks increased substantially compared to the clustered excitatory network, with a correspondingly greater reduction with stimulation (Figure 3.11c). Thus, excitatory input specificity can drive stimulus-induced variability reduction in inhibitory neurons. Studies have shown that the specificity of excitatory input to different inhibitory cell classes varies [125], with fast-spiking interneurons receiving more specific input than adapting interneurons. We predict adapting interneurons should therefore exhibit a smaller effect of stimulus on variability.

3.5 Discussion

We have shown that clustering of excitatory connections [103, 104, 106] substantially changes balanced network dynamics. Small perturbations of connectivity that introduced excitatory clustering yielded dynamical phenomena observed in cortex experimentally but not present in simple uniform networks: slow firing rate fluctuations in spontaneous conditions and stimulus-induced reductions in trial-to-trial rate variability with spike time variability remaining intact [17].

3.5.1 Anatomical realism and attractor dynamics

The network connectivity in this study was motivated by anatomical evidence for clustering of connections between pyramidal neurons in cortex. Bidirectional and clustered three-neuron connection motifs occur with frequencies significantly above chance in the visual system [103]. These subnetworks also receive similar feedforward input, suggesting that clustering may be

related to functional maps [105], an idea with recent *in vivo* support [106]. Inhibitory projections were not similarly clustered in our model, consistent with studies showing that inhibitory neurons connect densely and nonspecifically to pyramidal neurons [119] and that inhibitory neuron connectivity is less tuning-specific than pyramidal neurons [126].

The dynamics we investigated are reminiscent of persistent state activity, often studied in attractor networks with highly segregated subnetworks [109, 112]. Spatially dependent connection probabilities can also lead to persistent activity when excitation is more local than inhibition [127]. Our result for clustered networks, although it involves excitation to a local subset of neurons, assumes neither a spatial distribution of connections nor an excessive reorganization from an unstructured network. Similar results are obtained in networks with spatially dependent connection structures, as long as underlying multistability is present (Figure 3.10).

3.5.2 Robustness and dependence of cluster size

Our results depend on cluster size, and our parameter choices are consistent with anatomical evidence for fine-scale clusters of tens, rather than thousands of neurons [104]. Large scale, biophysically realistic network models have shown clusters of this size emerging with spike time dependent plasticity [128]. For a variety of cluster and network sizes, high firing rate variability emerged in our model (Figure 3.4). Because larger clusters have smaller fluctuations in mean firing rate due to averaging, very large clusters were unlikely to exhibit frequent transitions and also tended to suppress smaller clusters (Figure 3.3). However, large scale correlated fluctuations, adaptation, or other mechanisms not modeled in our study could promote cluster state transitions even for very large clusters.

We also note that our model exhibits sharp transitions between attractor states with large differences in firing rate (Figure 3.3e). These shifts are consistent with rapid state transitions observed in behaving animals [129]. Although average firing rates in our model are consistent with spontaneous data (Figure 3.2a), the activity within active clusters are more regular than typically recorded in spontaneous activity. This observation has been made in previous studies of bistability in balanced networks [108, 109, 110]. Matching exactly higher order spike train statistics is difficult, yet the agreement for time-averaged firing rate and spike train variability is promising. As such, clustered networks provide a simplified framework for studying

multistability and the spiking variability it produces.

3.5.3 Models of cortical variability and co-variability

ecently, it has been shown that uniform, balanced networks evolve to an asynchronous state in which correlations are negligible even when connections are dense [27], consistent with some experimental data [9]. The distribution of correlation coefficients in clustered networks has a near-zero mean, similar to these studies (Figure 3.2c). However, neuron pairs within the same cluster do exhibit nonzero correlations (Figure 3.2d). Assuming that clusters are related to functional processing, this observation is consistent with experimental studies showing that correlations are enhanced for similarly tuned neurons [12, 8] and predicts that measured correlations will depend on the subset of neurons to which recording is restricted.

Our modeling studies have examined networks exhibiting high variability from conductance modulations or other biophysical properties [97, 130], but long timescale firing rate variability (Figure 3.2g) has not been addressed. Our minimal model captures the distribution of firing rates [120] (Figure 3.2a), pairwise spiking correlations [9, 8] (Figure 3.2c,d,f), and rate variability [17] (Figure 3.2b,e) characteristic of spontaneous unanesthetized cortex. However, dual whole-cell recordings from excitatory cells in awake, spontaneous rodent somatosensory cortex show prominent, low frequency correlated membrane potential fluctuations [131]. Our model does not capture this correlation, nor does any other spiking model (to our knowledge) lacking externally imposed variability. Imaging studies suggest that the correlated fluctuations are likely due to traveling waves between motor and sensory cortex [132], and spatial effects would need to be introduced into a model to properly account for these dynamics (Figure 3.10). Nonetheless, this membrane potential correlation does not transfer to synchronous excitatory neuron spiking (Figure 6b in Gentet et al. [131]), presumably due to sparse firing in the spontaneous state. Understanding the source of these fluctuations is an important topic for future study.

3.5.4 Relationship between spontaneous and evoked activity

Patterns of neuronal activity evoked by sensory stimulation are thought to be a subset of those sampled during spontaneous activity [100, 102]. Clustered network dynamics are consistent with this idea, with activity states sampled stochastically in spontaneous conditions and biased

toward particular states with stimuli. The relationship between strong local connectivity and increased attractiveness of high activity states (Figure 3.4) suggests that Hebbian plasticity may wire together stimulated clusters so that they are sampled more frequently. Thus, the spontaneous state's statistics would reflect past stimulus input statistics, as reported in visual cortex [133].

Observations of reductions in single neuron trial-to-trial variability from spontaneous to evoked conditions complement the above data [17]. In describing these results, spiking activity is typically characterized as a "doubly stochastic" process, in which variability in spike emission and rate dynamics can be separated. While this simplification is attractive, the two are not easily segregated in mechanistic models. Stimulus-induced reductions in variability have been studied in firing rate models [123]; however, these models implicitly separate firing rate and spiking variability. In contrast, a recent attractor-based study [112] also modeled variability reduction but assumed doubly stochastic external inputs. We studied generation of both rate and spike time variability by internal network interactions, providing mechanistic plausibility for the intuition presented in Churchland et al. [17].

Finally, our work begins to link cortical architecture, variability, and computation. There is growing evidence that probabilistic inference is central to neural coding [134, 98]. Rich, stochastic spontaneous dynamics may reflect sampling that supports probabilistic inference [135]. Merging attractor theories of computation with probabilistic coding schemes is an important direction for systems neuroscience, for which a clear theory of neural variability in cortical circuits will be needed.

4. *Formation of neuronal assemblies and modification of spontaneous activity through plasticity in a model cortical network*

4.1 Abstract

The patterns of fluctuating spontaneous activity cortical neurons exhibit are related to the architecture in which they are embedded. This architecture may be modified by sensory experience, but the mechanisms that lead to the formation and persistence of structured connectivity in large recurrent networks are unclear. We demonstrate a network model that can be reorganized by stimuli through realistic plasticity and homeostatic mechanisms. Stimulus presentation leads to the formation of neuronal assemblies composed of neurons that receive common input. After training, spontaneous activity reflects prior evoked activity and stabilizes the learned network architecture. Sufficiently strong novel stimuli, however, can remap this architecture to reflect the new stimulus set. The model makes several predictions concerning the effect of sensory experience on spiking activity in spontaneous and evoked conditions. It also suggests that spontaneous activity fluctuations may consolidate learned connectivity patterns, rather than simply being a source of noise.

4.2 Introduction

Cortical neurons undergo complex and variable activity fluctuations during spontaneous dynamics [99, 102, 17], but how this activity is related to functional processing is poorly understood.

Several studies have shown that reactivation of evoked activity patterns can occur without stimulus application [133, 136, 137, 138]. Noise correlations, which reflect shared fluctuations from underlying network architecture, tend to co-vary with signal correlations, which reflect similar evoked stimulus preferences [139, 12, 140]. Together, these observations suggest that spontaneous dynamics are related to previously experienced stimuli. One explanation for the relationship between spontaneous and evoked activity is that synaptic plasticity reorganizes a network to reflect past patterns of activation [141, 142, 143]. This view is consistent with observations from the visual system that neurons with similar stimulus preference are more likely to be connected [106] and that this specificity emerges after eye opening [144].

Determining which plasticity mechanisms are needed for stimuli to reorganize neuronal architecture requires a realistic model of both plasticity and spontaneous dynamics. Recurrent model networks with a balance between excitation and inhibition have been successful in reproducing the high spike time variability consistent with the spontaneous state of cortex [24, 27, 36]. Models of attractor circuits in such networks have related simplified plasticity dynamics to the appearance of working memory states [108, 145]. However, embedding realistic spike time dependent plasticity (STDP) rules into balanced networks has proven difficult, often leading to pathological behavior [130]. Recently, detailed and biologically plausible STDP models have been proposed [146, 147, 148], but so far their application has been restricted to small systems, making it difficult to assess their effects on collective dynamics and trial-to-trial variability. Further, spontaneous activity in plastic networks can potentially destabilize a learned network architecture, especially if the magnitude of spontaneous fluctuations is comparable to that of evoked responses [102]. Many models assume that plasticity is inactivated after learning or do not study the persistence of learned architectures, underscoring the necessity of models that incorporate these features.

We study the dynamics of a balanced network with realistic excitatory STDP [146], showing that stimulus application leads to strongly connected neuronal assemblies. After training, spontaneous dynamics reflect experienced stimuli, with transient activations of assemblies that have been activated in the past. Our study demonstrates the importance of homeostatic mechanisms in maintaining this learned architecture [149, 150]. Furthermore, we show that spontaneous reactivation of assemblies promotes the stability of well-learned patterns, mitigating the destabilizing effect of random firing on structured connectivity. The model provides a framework in

which to study the effect of stimulus presentation on neuronal architecture, stimulus representation, and trial-to-trial variability.

4.3 Methods

4.3.1 Membrane potential dynamics

The model network consisted of N^E excitatory (E) and N^I inhibitory (I) neurons. Throughout, we will denote population (E or I) with superscripts and neuron index (1 through $N^{E/I}$) by subscripts. The equation for the voltage dynamics of neuron i in population X was:

$$\frac{d}{dt}V_i^X(t) = \frac{1}{\tau^X}\left(E_L^X - V_i^X(t) + \Delta_T^X \exp(\frac{V_i^X(t) - V_{T,i}^X(t)}{\Delta_T^X})\right) + \frac{g_i^{XE}(t)}{C}\left(E^E - V_i^X(t)\right) + \frac{g_i^{XI}(t)}{C}\left(E^I - V_i^X(t)\right) - \frac{w_i^X(t)}{C}. \tag{4.1}$$

Parameter values are summarized in Table 4.1. Excitatory units were modeled as exponential integrate-and-fire neurons with an adaptation current and adaptive threshold [151, 152]. Inhibitory units were modeled as simple non-adapting integrate-and-fire neurons ($\Delta_T^I, w_i^I(t) \to 0$). The dynamics of the neuronal threshold for excitatory neurons, $V_{T,i}^E$, were given by:

$$\frac{d}{dt}V_{T,i}^E(t) = \frac{1}{\tau_T}\left(V_T - V_{T,i}^E(t)\right). \tag{4.2}$$

When neuron i spiked due to its voltage diverging (in simulation, determined by voltage exceeding 20 mV), its potential was reset to V_{re} and clamped for an absolute refractory period of τ_{abs}. If it was excitatory, its threshold $V_{T,i}^E$ was set to $V_T + A_T$. For inhibitory neurons, $V_{T,i}^I = V_T$ at all times. We denote the spike train of neuron i in population X as $s_i^X(t) = \sum_{k=1}^n \delta(t - t_{i,k}^X)$, where $t_{i,1}^X \ldots t_{i,n}^X$ are the times when the neuron spiked and δ is the Dirac delta function.

The adaptation current for excitatory neuron i, w_i^E, was given by:

$$\frac{d}{dt}w_i^E(t) = \frac{1}{\tau_w}\left(a_w(V_i^E(t) - E_L^E) - w_i^E(t)\right). \tag{4.3}$$

When excitatory neuron i spiked, w_i^E was increased by b_w. Simulations were performed using code written in Python and C++ implementing Euler integration with a timestep of 0.1 ms.

4.3.2 Dynamics of synaptic conductances

Connections occurred with probability p, and the strength of a connection from neuron j in population Y to neuron i in population X was denoted J_{ij}^{XY}. If a connection did not exist, $J_{ij}^{XY} = 0$. Recurrent excitatory connection weights were bounded by J_{min}^{EE} and J_{max}^{EE}, while weights from inhibitory to excitatory neurons were bounded by J_{min}^{EI} and J_{max}^{EI}. J^{IE} and J^{II} were fixed and constant for all connected i, j.

The total excitatory or inhibitory conductance of neuron i in population X was given by:

$$g_i^{XY}(t) = F^Y(t) * \left(J_{\text{ext}}^{XY} s_{i,\text{ext}}^{XY}(t) + \sum_j J_{ij}^{XY} s_j^Y(t) \right), \tag{4.4}$$

where $Y \in (E, I)$, $F^Y(t)$ is the synaptic kernel for input from population Y, and $*$ denotes convolution. Synaptic kernels were given by a difference of exponentials:

$$F^Y(t) = \frac{1}{\tau_d^Y - \tau_r^Y} \left(e^{-t/\tau_d^Y} - e^{-t/\tau_r^Y} \right), \tag{4.5}$$

for t positive. In addition to recurrent input, neurons also received external excitatory input in the form of a spike train $s_{i,\text{ext}}^{XE}(t)$, which was an independent homogeneous Poisson process for each neuron with rate r_{ext}^{XE}. For external excitatory input, the synaptic strength was set to the minimim recurrent synaptic strength ($J_{\text{ext}}^{XE} = J_{\text{min}}^{XE}$). Neurons did not receive external inhibitory input ($s_{i,\text{ext}}^{XI} = 0$). Parameters were chosen such that a fully potentiated excitatory or inhibitory synapse onto an excitatory neuron caused a postsynaptic potential of approximately 2.5 mV if all other inputs were blocked. Parameter values can be found in Table 4.2.

4.3.3 Excitatory synaptic plasticity

We implemented a voltage based STDP rule [146] that modified J_{ij}^{EE} within the bounded range (J_{min}^{EE},J_{max}^{EE}). Furthermore, we imposed a homeostatic normalization of the total excitatory synaptic weights that a neuron received by scaling each row of J_{ij}^{EE} every 20 ms to maintain a constant row sum [153, 154]. This was accomplished by subtracting a constant amount from each nonzero entry in the row: $J_{ij}^{EE}(t) \leftarrow J_{ij}^{EE}(t) - \frac{\sum_j J_{ij}^{EE}(t) - J_{ij}^{EE}(0)}{N_i^E}$, where N_i^E is the number of nonzero entries [155].

The dynamics of the synapse from excitatory neuron j to excitatory neuron i were given by:

$$\frac{d}{dt}J_{ij}^{EE}(t) = -A_{\text{LTD}}s_j^E(t)R\left(u_i^E(t)-\theta_{\text{LTD}}\right)+A_{\text{LTP}}x_j^E(t)R\left(V_i^E(t)-\theta_{\text{LTP}}\right)R\left(v_i^E(t)-\theta_{\text{LTD}}\right), \quad (4.6)$$

where R is a linear rectifying function ($R(x) = 0$ if $x < 0$, $R(x) = x$ otherwise), u_i^E and v_i^E represent the membrane voltage V_i^E low-pass filtered with time constants τ_u and τ_v respectively, and x_j^E represents the spike train s_j^E low-pass filtered with time constant τ_x. Parameter values can be found in Table 4.3.

4.3.4 Inhibitory synaptic plasticity

We implemented an inhibitory STDP rule [150] that modified J_{ij}^{EI} within the bounded range (J_{min}^{EI},J_{max}^{EI}). Weight changes depended on y_i^X, which represents the spike train s_i^X low-pass filtered with time constant τ_y. Upon a spike from either the presynaptic or postsynaptic neuron, the weight was modified according to the following:

$$\begin{aligned} J_{ij}^{EI} &\leftarrow J_{ij}^{EI} + \eta\left(y_i^E(t) - 2r_0\tau_y\right) \text{ if the presynaptic inhibitory neuron fired,} \\ J_{ij}^{EI} &\leftarrow J_{ij}^{EI} + \eta y_j^I(t) \text{ if the postsynaptic excitatory neuron fired.} \end{aligned} \quad (4.7)$$

The rate r_0 represents the target firing rate to which the inhibitory plasticity attempts to balance the postsynaptic neuron. During unstimulated conditions, neurons fired on average less frequently than their target rate due to synaptic bounds (mean of 1.7 Hz rather than 3 Hz; Figure 4.8), allowing the system to exhibit a distribution of firing rates because of the variability in the connectivity matrix. Parameter values can be found in Table 4.4.

4.3.5 Stimulation protocols

For training (Figure 4.1), each stimulus $i = 1 \ldots 20$ was activated sequentially for a period of 1 second, with 3 second gaps in between stimulus presentations. This was repeated until each stimulus had been repeated 20 times. For retraining (Figure 4.3), each stimulus was presented 100 times. During training, stimuli increased r_{ext}^{XE} by 8 kHz. These simulations began with a 10 second period without STDP to allow transients to die out. To probe the dependence of Fano factor and correlation before and after training (Figures 4.5 and 4.6), we used a weaker

stimulus, with a duration of 500 ms and an increase in r_{ext}^{XE} of 2 kHz. These simulations were repeated 50 times for each of 10 stimuli to assess trial-to-trial variability. We disabled STDP for these short protocols.

For the non-recurrent network (Figure 4.2a,b), we chose the inhibitory STDP target firing rate for each neuron to be equal to its spontaneous firing rate in the trained system and removed synaptic bounds for inhibitory STDP. We also increased r_{ext}^{EE} by 4 kHz and r_{ext}^{IE} by 2 kHz to compensate for the loss of recurrent excitation. We then allowed the system to run for 200 s for firing rates to reach their target values before activating excitatory STDP. This allowed for a matched firing rate comparison between the two networks.

Symbol	Description	Value
τ^E	E neuron resting membrane time constant	20 ms
τ^I	I neuron resting membrane time constant	20 ms
E_L^E	E neuron resting potential	-70 mV
E_L^I	I neuron resting potential	-62 mV
Δ_T^E	E neuron EIF slope factor	2 mV
C	Capacitance	300 pF
E^E	E reversal potential	0 mV
E^I	I reversal potential	-75 mV
V_T	Threshold potential	-52 mV
A_T	Post spike threshold potential increase	10 mV
V_{re}	Reset potential	-60 mV
τ_{abs}	Absolute refractory period	1 ms
a_w	Subthreshold adaptation	4 nS
b_w	Spike triggered adaptation	0.805 pA

Table 4.1: Parameters for neuronal membrane dynamics

Symbol	Description	Value
N^E	Number of E neurons	4,000
N^I	Number of I neurons	1,000
p	Connection probability	0.2
τ_r^E	Rise time for E synapses	1 ms
τ_d^E	Decay time for E synapses	6 ms
τ_r^I	Rise time for I synapses	0.5 ms
τ_d^I	Decay time for I synapses	2 ms
r_{ext}^{EE}	Rate of external input to E neurons	4.5 kHz
r_{ext}^{IE}	Rate of external input to I neurons	2.25 kHz
J_{min}^{EE}	Minimum E to E synaptic weight	1.78 pF
J_{max}^{EE}	Maximum E to E synaptic weight	21.4 pF
J_0^{EE}	Initial E to E synaptic weight	2.76 pF
J_{min}^{EI}	Minimum I to E synaptic weight	48.7 pF
J_{max}^{EI}	Maximum I to E synaptic weight	243 pF
J_0^{EI}	Initial I to E synaptic weight	48.7 pF
J^{IE}	Synaptic weight from E to I	1.27 pF
J^{II}	Synaptic weight from I to I	16.2 pF

Table 4.2: Parameters for recurrent coupling

Symbol	Description	Value
A_{LTD}	Long-term depression (LTD) strength	.0008 mV^{-1}
A_{LTP}	Long-term potentiation (LTP) strength	.0014 mV^{-2}
θ_{LTD}	Threshold to recruit LTD	-70 mV
θ_{LTP}	Threshold to recruit LTP	-49 mV
τ_u	Time constant of low-pass filtered membrane voltage (for LTD)	10 ms
τ_v	Time constant of low-pass filtered membrane voltage (for LTP)	7 ms
τ_x	Time constant low-pass filtered spike train (for LTP)	15 ms

Table 4.3: Parameters for excitatory synaptic plasticity

Symbol	Description	Value
τ_y	Time constant of low-pass filtered spike train	20 ms
η	Synaptic plasticity learning rate	1
r_0	Target firing rate	3 Hz

Table 4.4: Parameters for inhibitory synaptic plasticity

4.4 Results

4.4.1 Formation of neuronal assemblies

While many studies have investigated the dynamics of balanced networks with fixed connection weights, few have investigated the case of plastic synapses. We simulated a model cortical network of excitatory and inhibitory neurons, supplemented with plasticity rules that modified the connections onto excitatory neurons (see Methods). The network was composed of 4,000 adaptive exponential integrate-and-fire excitatory neurons and 1,000 integrate-and-fire inhibitory neurons [151, 152] coupled randomly with a connection probability of 0.2. Excitatory to excitatory synapses were governed by a voltage based STDP rule (146; Figure 4.1a, top) and inhibitory to excitatory synapses by a symmetric Hebbian STDP rule [150]. Both synapse types were bounded between minimum and maximum values. The excitatory STDP rule captured the dependence of long-term depression and potentiation (LTD and LTP) on firing rate (156; Figure 4.1a, bottom), so that neurons that exhibited correlated rate fluctuations tended to recruit LTP. The total excitatory conductance onto any neuron was normalized [153, 154], thus imposing competition between synapses. Inhibitory synaptic plasticity limited the firing rate of excitatory neurons, as frequently firing neurons recruited more inhibition [150].

We began with a network with a homogeneous weight structure. The network exhibited irregular and asynchronous firing due to a balance between excitatory and inhibitory currents [24, 27]. No structured connectivity emerged during spontaneous activity (Figure 4.7). We investigated how this connectivity could be modified by the application of external stimuli. To do so, we defined a set of twenty stimulus patterns, which when active corresponded to increased excitatory drive to neurons targeted by that pattern. Neurons were targeted by each of these patterns with a probability of 0.05. As a result, the number of stimuli targeting each neuron was binomially distributed and approximately 36% of neurons were not targeted, 38% were targeted by one stimulus, and 26% were targeted by multiple stimuli. For each stimulus $i = 1, 2, \ldots, 20$, we refer to those neurons targeted by the stimulus as neuronal assembly i.

Stimuli were presented sequentially in repeated presentation blocks. We recorded $\bar{W}_{\text{in}}$, the average synaptic strength of connections in each of the twenty neuronal assemblies. During the

training period, this quantity increased as stimuli were presented, until the synapses between neurons in the same assembly were strongly potentiated (Figure 4.1b,c).

Besides modifying synaptic weights, training also changed the network's spontaneous dynamics. The clustered excitatory connectivity led to transient activations of previously stimulated neuronal assemblies (Figure 4.1d,e). These activations occurred over a timescale of hundreds of milliseconds. We conclude that the network is capable of reorganizing its connections in response to stimulus presentation, and that this reorganization substantially modifies spontaneous dynamics to reflect prior stimuli. We refer to these new dynamics as *structured spontaneous activity*, in contrast to the uncorrelated activity in the uniformly connected network.

4.4.2 Stability of learned connectivity

We next examined the stability of the network architecture during these spontaneous fluctuations. Many studies have examined how different *synaptic* dynamics can affect the stability of learned connectivity in the face of random spontaneous firing [157], employing models that include, for example, cascades of synaptic states [158] or metaplasticity [159]. We took a complementary approach, asking whether the modification of *network* dynamics after training could be beneficial for the persistence of the learned architecture compared to the random firing case. We reasoned that structured spontaneous activity (Figure 4.1d) could consolidate the learned assembly structure.

We therefore compared the trained recurrent network to a hypothetical non-recurrent network without structured spontaneous activity. Both networks began with the same initial trained weight matrix and exhibited the same firing rates (see Methods). However, in the non-recurrent network, synaptic plasticity modified the weight matrix depending on neurons' spike times, but excitatory spikes did not cause EPSPs in postsynaptic targets (Figure 4.2a). Therefore, excitatory inputs were purely external and uncorrelated, which prevented coordinated assembly activation. We measured $\bar{W}_{\text{in}}$ over a period of spontaneous activity in the two models. In the non-recurrent network, $\bar{W}_{\text{in}}$ decayed faster than in the recurrent network with structured spontaneous activity (Figure 4.2b), suggesting that this activity prevents degradation of learned connectivity.

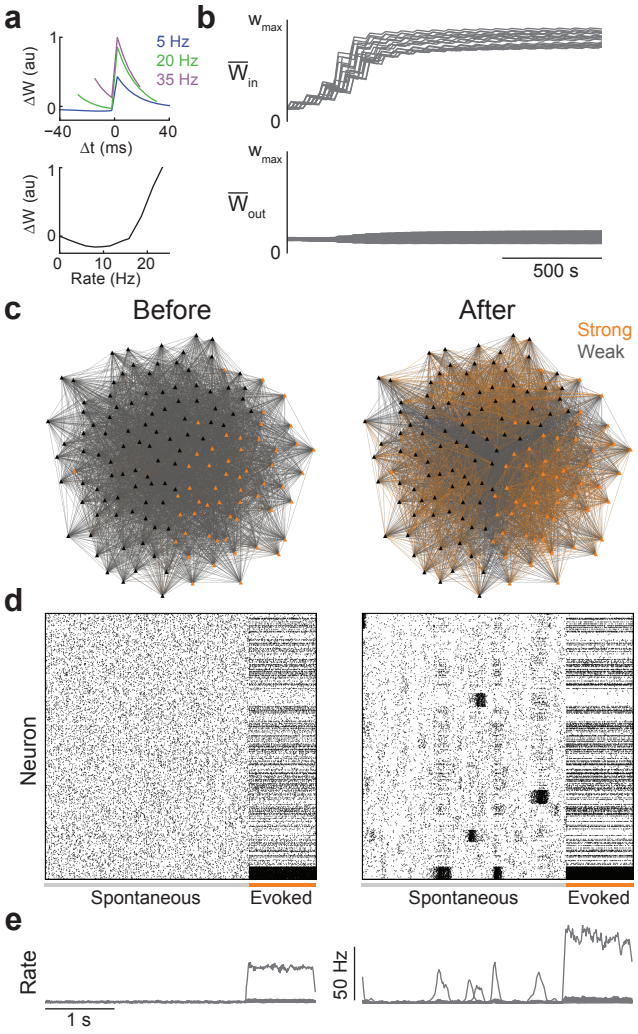

Figure 4.1: Modification of weights and spontaneous activity by training. **(a)** STDP curve for different pairing frequencies (top) and weight change as a function of firing rate assuming both neurons fire as Poisson processes with the same rate (bottom). **(b)** Average synaptic weight $\bar{W}_{\text{in}}$ for synapses between neurons within an assembly (top) and $\bar{W}_{\text{out}}$ for synapses between neurons in different assemblies (bottom). **(c)** Graph showing connection strength between neuron pairs for 50 neurons sampled from three assemblies before (left) and after (right) training. Orange lines correspond to strong excitatory connections. Neurons in the same assembly are placed nearby. Colored nodes indicate neurons targeted by stimulus 1. Due to overlap, some of these targeted neurons are also in the assemblies corresponding to stimulus 2 and 3. **(d)** Excitatory neuron spike rasters before and after training. Adjacent rows in the raster correspond to neurons in the same assembly. Some rows are repeated since neurons can belong to multiple assemblies. Neurons not belonging to assemblies are not shown. **(e)** Average firing rate for neurons in each assembly, corresponding to the activity in D.

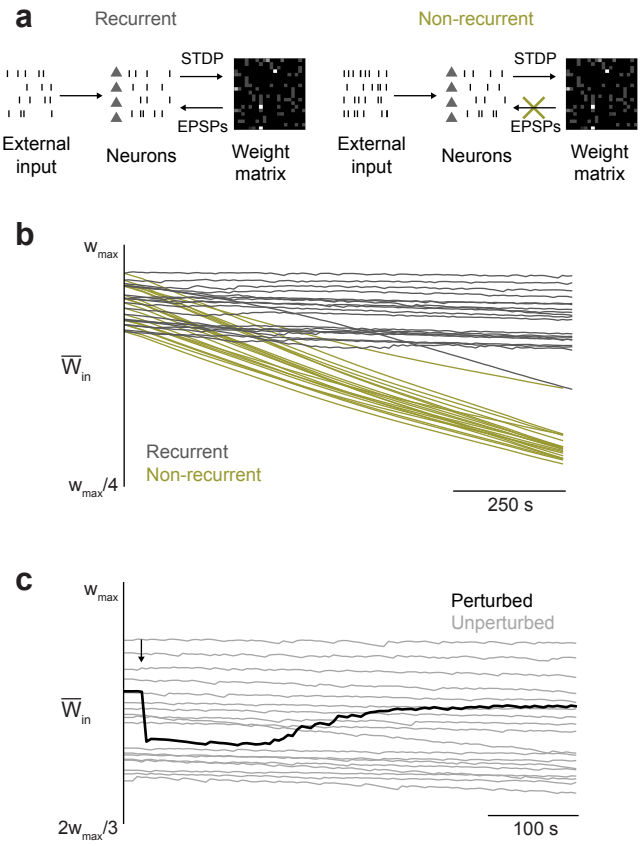

Figure 4.2: Stability of trained architecture. **(a)** Schematic of recurrent and non-recurrent networks. In both networks, neurons fire leading to STDP that modifies the weight matrix. In the non-recurrent network, recurrent EPSPs that would be caused by excitatory neuron spiking are blocked. **(b)** Average synaptic weight within assemblies $\bar{W}_{\text{in}}$ during spontaneous activity after training for the two networks. **(c)** Dynamics of $\bar{W}_{\text{in}}$ in response to a perturbation that reduced $\bar{W}_{\text{in}}$ for one assembly at $t = 15$ s (arrow).

To further probe the effect of structured spontaneous activity, we perturbed $\bar{W}_{\text{in}}$, reducing it by 10% for one assembly, and analyzed the resulting dynamics. After the perturbation, $\bar{W}_{\text{in}}$ increased for that cluster, showing that structured spontaneous activity can retrain a network if its architecture is disrupted (Figure 4.2c). Hence, this activity both lengthens the timescale over which a network can retain its learned connectivity and provides an error-correcting mechanism given disruptions of this connectivity. This error-correcting property is distinct from mechanisms that operate at the synaptic level, as such mechanisms have no knowledge of learned network-level structures. Instead, collective activations of entire assemblies underlie this process.

4.4.3 Remapping architecture with novel stimuli

We have shown that a set of stimulus patterns can be embedded in a recurrent network with STDP. How flexible is this network; that is, can it respond to the presentation of novel stimulus sets? We examined this by defining a new set of twenty stimulus patterns and repeatedly presenting them to the network. We then compared $\bar{W}_{\text{in}}$ for the new stimulus-defined assemblies and the old assemblies throughout this protocol.

During this retraining protocol, $\bar{W}_{\text{in}}$ increased for the new assemblies and decayed for the old assemblies (Figure 4.3a). Fully potentiating the new set of assemblies required more stimulus presentations than the original training protocol, indicating that trained stimulus-specific assemblies can interfere with the formation of new ones. Retraining also shifted spontaneous activity. Prior to retraining, only the original assemblies were activated during spontaneous activity (Figure 4.3b). Afterwards, the new assemblies were activated, as well as some of the old assemblies that remained potentiated (Figure 4.3c). Given the stability of the trained assemblies (Figure 4.2) and these results, we conclude that the network is stable to weak perturbations of its input statistics, but sufficiently strong changes in its inputs cause the network to reorganize itself.

4.4.4 Effects of homeostatic inhibitory plasticity

So far, we have addressed the dynamics of connections between excitatory neurons. We also modeled plasticity of inhibitory synapses onto excitatory neurons (iSTDP) using a recently proposed rule [150]. Under the iSTDP rule, inhibitory synapses were on average potentiated when the postsynaptic excitatory neuron fired above a target firing rate and depressed otherwise. As a result, the rule homeostatically regulated the distribution of excitatory firing rates in the network toward this target value.

We next show that this homeostatic mechanism is beneficial for assembly dynamics. Under the iSTDP rule, strongly interconnected assemblies will receive more inhibition than weakly interconnected ones (Figure 4.4a,b). This is because stronger assemblies are more likely to become active and hence have higher average firing rates. Without iSTDP, imbalances in $\bar{W}_{\text{in}}$ or assembly size can lead to winner-take-all dynamics in which the strongest cluster remains

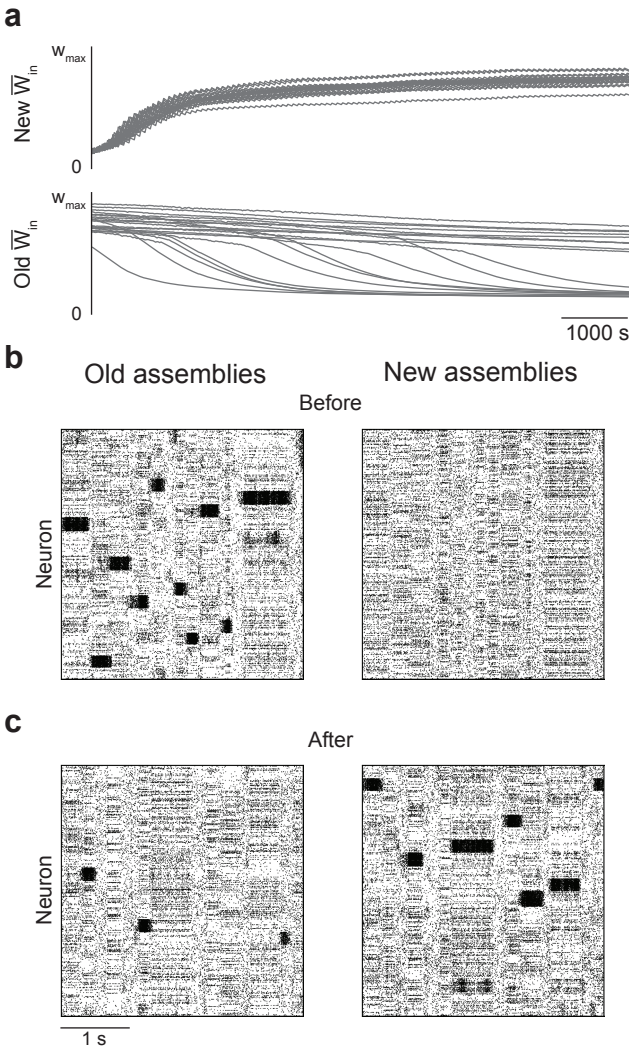

Figure 4.3: Remapping of architecture with presentation of new stimulus ensemble. **(a)** Average synaptic weight within assemblies $\bar{W}_{\text{in}}$ for neurons in the new set of assemblies (top) and the original set of assemblies (bottom), during repeated presentations of the new stimulus set. **(b)** Plot of network activity for rows arranged according to the original assemblies (left) and new assemblies (right). The activity is the same for both rasters, but the rows are permuted according to assembly membership. **(c)** Same as B, but after training on the new stimulus set.

active for long periods of time. In general, fine-tuning of parameters is needed to avoid these pathological states [153]. Indeed, if the training protocol from Figure 4.1 was applied to a network without iSTDP, only a subset of assemblies achieved a high $\bar{W}_{\text{in}}$ (Figure 4.4c). This lack of training occurred because the strongest assembly, once sufficiently potentiated, dominated the spontaneous activity of the network, unlike the network with iSTDP (Figure 4.4d). Homeostatic regulation of excitatory activity therefore prevents winner-take-all dynamics and promotes the formation of multiple heterogeneous assemblies. Such regulation is critical for the network to robustly exhibit structured spontaneous activity.

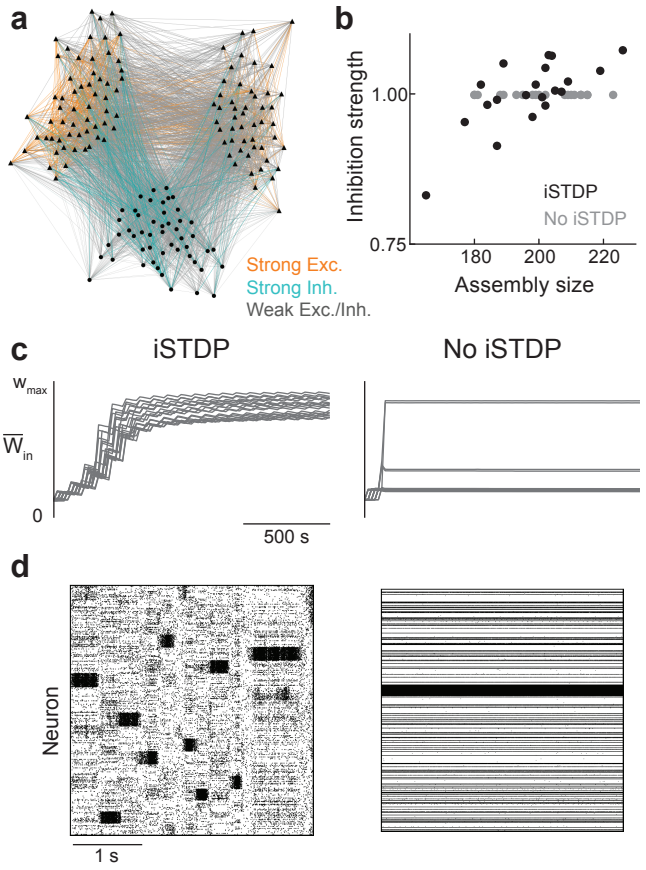

Figure 4.4: Effects of inhibitory STDP (iSTDP). **(a)** Schematic showing connections between two assemblies and inhibitory neurons. One strongly connected and one weakly connected assembly is shown. Orange lines correspond to strong excitatory connections and blue lines to strong inhibitory connections. **(b)** Average inhibitory connection strength onto assemblies of different sizes for networks with and without iSTDP. **(c)** Average connection strength within clusters $\bar{W}_{\text{in}}$ during training for networks with (left) and without (right) iSTDP. **(d)** Spike rasters during spontaneous activity for networks with (left) and without (right) iSTDP.

4.4.5 Spike statistics before and after training

We have investigated the mechanisms that cause the network to respond differently to familiar and unfamiliar stimulus patterns. Next, we quantify the changes in spike train statistics due to training. We examined the responses of neurons to the new set of stimulus patterns (Figure 4.3) before and after being trained on those patterns. This allowed us to make predictions about spontaneous and evoked dynamics before and after training.

We first considered evoked firing rates. Consistent with increased recurrent connectivity, the gain of stimulated neurons was higher for trained versus untrained inputs (Figure 4.5a). In addition to amplification, structured excitatory connectivity is often proposed as a mechanism for pattern completion [160]. To test whether the network was capable of pattern completion,

we presented stimuli that targeted half of the neurons in assemblies. We then compared the firing rates of stimulated neurons and non-stimulated neurons that were in targeted assemblies to baseline firing rates. Before training, stimulated neurons fired at increased rates, but non-stimulated neurons within a stimulated assembly fired at baseline levels (Figure 4.5b). After training, both stimulated and non-stimulated neurons had increased firing rates. Hence, the presence of a stimulus could be read out from neurons that did not directly receive it through feedforward projections, as long as they were part of the corresponding assembly.

We also examined trial-to-trial variability of excitatory neurons in the network. We computed the Fano factor of spike counts in 100 ms windows over repeated presentations of stimuli. Stimulation caused a reduction in Fano factor as is frequently observed in cortex [17] (Figure 4.5b). This reduction was due to the suppression of spontaneous fluctuations by the stimulus [36, 112]. This decrease was greater for trained versus untrained stimuli (Figure 4.5c). The combination of this decrased trial-to-trial variability and increased gain (Figure 4.5a) suggests that training improves the reliability of stimulus representation. To test this, we attempted to detect the presence of a stimulus using the spike count of single neurons in 200 ms intervals after stimulus onset. The resulting receiver operating characteristic (ROC) curve was higher after training (Figure 4.5d). Hence, training led to a measurable improvement in stimulus encoding, even at the single neuron level.

We also quantified collective activity using the spike count correlation between neuron pairs measured in 100 ms windows across trials. On average, noise correlations during spontaneous activity were near zero consistent with an asynchronous state (27; Figure 4.6a). However, neuron pairs in the same cluster had positive average spike count correlation after training (Figure 4.6b). This correlation reflected collective fluctuations due to structured spontaneous activity. Notably, the contribution of these positive correlations was suppressed when the network was stimulated, with greater suppression for trained stimuli (Figure 4.6c). We conclude that, after training, networks exhibit collective spontaneous fluctuations consistent with their previous inputs. However, when external input is applied, these fluctuations are suppressed and reliable responses are produced.

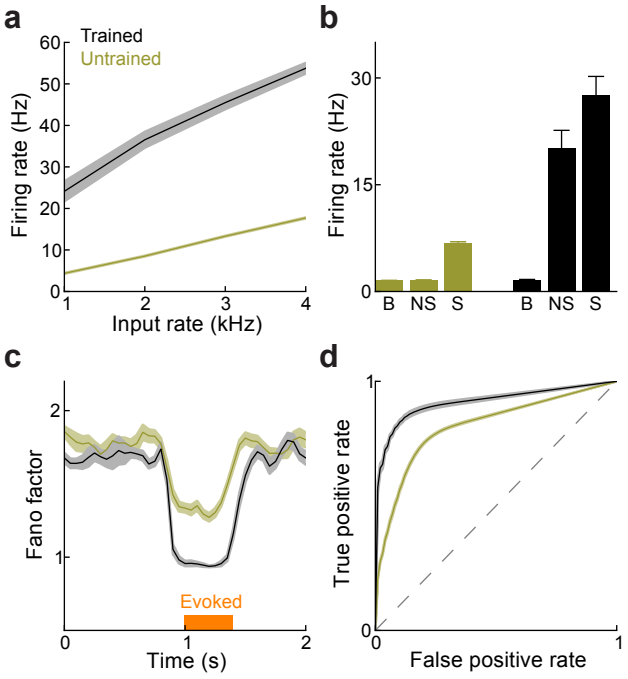

Figure 4.5: Changes in evoked spike statistics after training on a new stimulus ensemble. **(a)** Average evoked firing rate in stimulated neurons as a function of stimulus strength, defined here as the change in the firing rate of the external excitatory input to stimulated neurons. All error bars denote standard error of the mean across different stimuli. **(b)** Pattern completion before and after training. Average firing rates are shown in baseline spontaneous conditions (b), for non-stimulated neurons within a stimulated assembly (NS) and for stimulated neurons (S). **(c)** Fano factor as a function of time for all excitatory neurons. Evoked stimulation interval is denoted by the orange bar. Fano factor was computed using the mean matching techniques presented in 17. **(d)** Average receiver operating characteristic (ROC) curve for detecting the presence of a stimulus using the spike count of single neurons in the first 200 ms after stimulus onset.

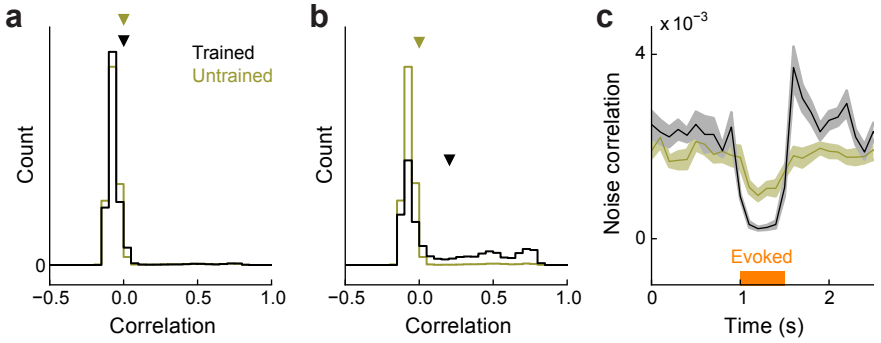

Figure 4.6: Spike train co-variability before and after training. **(a)** Distribution of correlation coefficients during spontaneous activity for excitatory neuron pairs. Triangles denote means of distributions. **(b)** Same as A for neurons that shared membership in at least one assembly. **(c)** Average noise correlation as a function of time for all excitatory neurons. Evoked stimulation interval is denoted by the orange bar. Shaded region denotes standard error of the mean across different stimuli.

4.5 Discussion

4.5.1 Predictions of the model

Clearly identifying neuronal assemblies is difficult because accessing a large fraction of the connections in a local circuit is experimentally challenging, although some progress has been made [104]. Our model, however, makes predictions that can be tested with single neurons or pairs (Figures 4.5 and 4.6). The first prediction, that recurrent connectivity is responsible for a large proportion of stimulus-tuned excitatory input, is consistent with findings from visual and auditory cortex that tuned recurrent excitation amplifies responses to thalamic input [161, 162, 163]. The fact that these responses are associated with slow dynamics [36] is also consistent with results from auditory cortex suggesting that recurrent excitation prolongs response duration [163].

Stimulus evoked reductions in Fano factor are seen in many cortical regions [17]. We further predict that training can reduce Fano factors. This observation has been made in prefrontal cortex, in which an increase in firing rate and decrease in variability was found after training on a working memory task [164]. However, in that study, Fano factor was reduced during the fixation period as well. This result would be consistent with our model assuming that task relevant assemblies were trained and remained active during the performance of the task and hence suppressed spontaneous fluctuations.

Our network also makes predictions about spike train co-variability. Structured spontaneous activity leads to increased noise correlation for neurons with similar stimulus preference, as is commonly reported [139, 12]. Further, because connectivity is related to stimulus preference in our network, this implies that neurons that share strong connections have higher noise correlations (Figure 4.6b). This is consistent with findings from visual cortex showing noise correlations are higher for connected pairs [106].

4.5.2 Homeostatic mechanisms and synaptic dynamics

We have shown that homeostatic regulation of firing rates in the form of inhibitory synaptic plasticity is crucial for robust training of the network (Figure 4.4). We do not claim that this

particular form of plasticity is necessary, but rather that regulation of average firing rate is likely an important mechanism in maintaining reasonable activity patterns. Firing-rate modulated scaling of total excitatory input weight, for example, could accomplish similar effects. Indeed, miniature excitatory postsynaptic current (mEPSC) sizes and firing rates increase following sensory deprivation that initially causes a decrease in visual cortex activity [165]. Previous theoretical studies have shown that homeosynaptic scaling can be beneficial for heterogeneous working memory circuits, similar to the way inhibitory synaptic plasticity prevents the largest of a heterogeneous set of assemblies from dominating network activity (153; Figure 4.4).

Our network also relied on heterosynaptic competition, which maintained the total excitatory synaptic strength onto a neuron, and bounds on synaptic strength. Heterosynaptic competition has been studied experimentally and in model networks [166, 149, 154]. In a previous computational model, normalization of both presynaptic and postsynaptic weights along with a simple STDP rule led to the spontaneous development of feedforward chains [154]. Our model, in contrast, includes a dependence of STDP on firing rate and fewer constraints on synaptic weights, leading to qualitatively different dynamics. Bounds on synaptic strength were also necessary to curb the instrinsic instability of excitatory STDP. Without them, strong stimuli could cause assemblies' recurrent weights to increase without limit, or, conversely, parts of the network to become completely disconnected. Such pathological behavior has been observed in previous models of balanced excitatory-inhibitory networks with STDP that lacked the homeostatic mechanisms in our study [130].

4.5.3 Other forms of plasticity and network dynamics

We made several assumptions on plasticity in our network. For simplicity and due to lack of experimental studies, we did not model plasticity of synapses onto inhibitory neurons. Given recent studies focusing on the diversity of interneuron subtypes [167, 168], future computational studies should explore the dynamics of networks with multiple inhibitory populations. We also assumed that the feedforward inputs to neurons were fixed during training. In development, thalamic input specificity occurs prior to the establishment of specific recurrent connectivity [144]. On the other hand, simultaneous plasticity of feedforward and recurrent pathways may yield new principles. For example, inhibitory plasticity of feedforward pathways may compete

with the increases in gain seen in our study (Figure 4.5a), transforming the primarily feedforward untrained system to a primarily recurrent trained system with similar response characteristics. This shift from a feedforward to a pattern-completing network is somewhat reminiscent of predictive coding theories, in which higher regions predict the activity in lower regions and feedforward connections serve primarily to transmit errors in this prediction [169].

A critical aspect of our model was the dependence of potentiation and depression on the firing rate of the postsynaptic neuron (Figure 4.1a, bottom). Such dependence has been shown experimentally and is a feature of multiple biophysically motivated STDP rules [156, 170, 146, 148]. For an analysis of these other rules in the context of assembly formation, see Appendix B. Unlike previous studies in which spike timing was the only determinant of synaptic changes, rate, not spike timing, was the most important quantity in our network. Indeed, our reduced model captured many features of the spiking network using a plasticity rule similar to the rate-based BCM rules [171]. However, while we have shown that neuronal assemblies constitute stable configurations of the synaptic weights for our system, we have not shown that they are the only ones. Other structures, such as feedforward chains, may also be stabilized by the voltage based STDP rule we studied or other rules. Such structures may be important for explaining spontaneous activation of temporal activity sequences [137]. Relating biophysically motivated plasticity rules to the stable weight configurations they can support remains an open area of inquiry.

4.6 Supplementary Figures

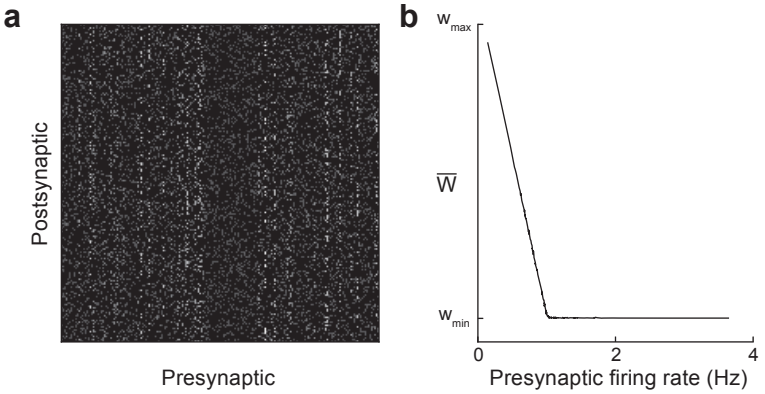

Figure 4.7: No development of structure during spontaneous activity without stimulation. **(a)** 200×200 subset of the excitatory weight matrix after 500 s of spontaneous activity. Vertical bars indicate that the output weights of individual neurons are correlated. **(b)** Average output weight $\bar{W}$ of neurons as a function of the presynaptic neuron's firing rate. Lower firing rate neurons tend to have higher output weights, as would be expected for depression-dominated plasticity.

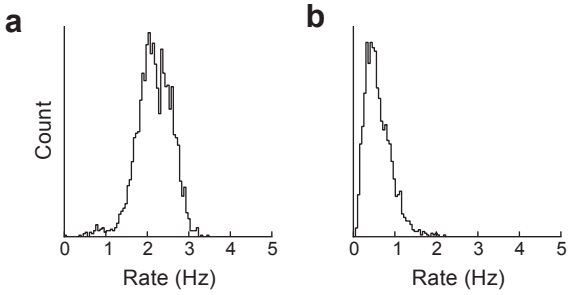

Figure 4.8: Firing rate distributions for excitatory neurons after training. **(a)** Neurons belonging to one or more assemblies. **(b)** Neurons not belonging to any assembly.

5. *Analysis of simplified birth-death process network model*

5.1 Introduction

In the previous two chapters, we presented results from simulations of spiking systems with multiple interacting subpopulations. These subpopulations could exhibit bistability (Chapter 3). When plasticity was added, this bistability promoted the stability of clustered weight configurations (Chapter 4). To understand the mechanisms behind these dynamics, we developed a simple model of plastic neuronal populations. By reducing the system to a small number of populations, this model allowed us to formally demonstrate how rich spontaneous activity leads to stability of trained weight configurations. It also indicates that this result is not specific to the particular spiking implementation of the previous chapter.

5.2 Escape rates for a simplified bistable neuronal population model

To obtain a tractable model of interacting neuronal populations, we use techniques of Bressloff [172]. We consider simplified dynamics in which each population i is characterized by $n_i(t)$, the number of active units within the population at time t. $n_i(t)$ takes nonnegative integer values but is not bounded above. It evolves according to a birth-death process characterized by $T_i^+(n)$ and $T_i^-(n)$, which represent the rate of transitions from the state of n neurons active to $n+1$ or $n-1$ neurons active, respectively. Hence, the probability distribution over the states of $n_i(t)$ is given by:

$$\frac{dP_i(n,t)}{dt} = T_i^+(n-1)P_i(n-1) + T_i^-(n+1)P_i(n+1) - \left(T_i^+(n) + T_i^-(n)\right) P_i(n). \qquad (5.1)$$

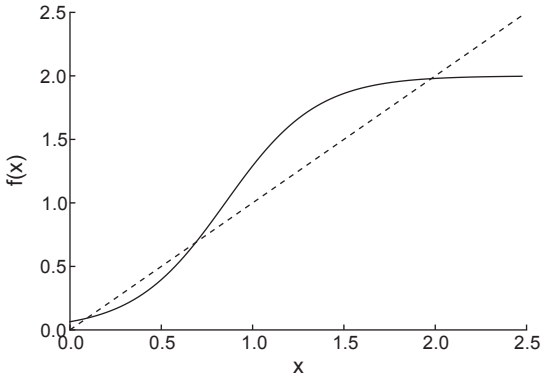

Figure 5.1: Plot of $f(x)$ with $x_b = 0$, $f_0 = 2$, $\gamma = 4$, and $\theta = 0.85$. Intersections with the line $y = x$ (dashed curve) correspond to equilibrium points of the Wilson-Cowan equations (5.4) in the limit of large N.

Following Bressloff [172], we construct T^+ and T^- so that, in the limit of large population size, we recover the deterministic Wilson-Cowan equations. This can be seen by using the master equation to calculate the time evolution of the expectation of n, neglecting the influence of higher order moments. To that end, we define an abstract size parameter N (note that this parameter is *not* the upper bound on the number of active neurons per population). We then define:

$$T_i^+(n) = Nf(n/N), \;\; T_i^-(n) = n, \tag{5.2}$$

where f is a standard sigmoidal function of its inputs:

$$f(x) = x_b + \frac{f_0}{1 + e^{-\gamma(x-\theta)}}. \tag{5.3}$$

Here, γ controls the sharpness of the sigmoid, θ controls the offset, x_b the baselane, and f_0 controls the maximum value. As $N \to \infty$, the birth-death process (5.1) obeys the equation:

$$\frac{dx}{dt} = -x + f(x), \tag{5.4}$$

if n/N is identified as x.

To begin, we consider the case of a single neuronal population and omit the subscript i. We first note that, for sufficiently strong activation f, the deterministic equation (5.4) exhibits bistability (Figure 5.1). Simulations of the birth-death processes described in this section can be performed using the Gillespie algorithm. Simulating a finite-sized system with $N = 20$ confirms

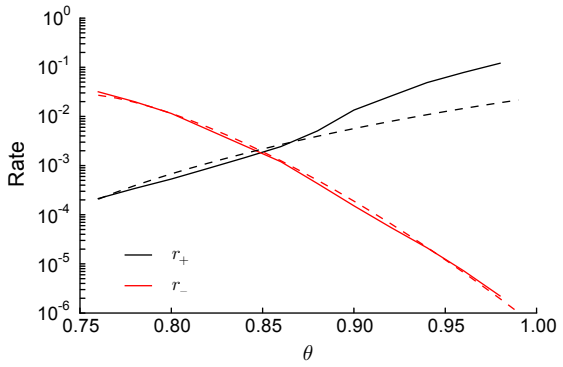

Figure 5.2: Transition rates for a single population with $x_b = 0$, $f_0 = 2$, $\gamma = 4$, $N = 20$, and $\theta = 0.85$. Solid lines represent transition rates estimated from simulations using the Gillespie algorithm. Dashed lines represent the theoretical estimates from equations (5.5).

that n/N spends most of its time near two equilibria, which we label x_- and x_+. We seek to analyze the transitions between these two states.

By approximating the solution in the vicinity of the two equilibria x_- and x_+ and then matching the two solutions near the saddle point x_0, we obtain expressions for r_- and r_+, the rate of escape from the lower and higher equilibrium, respectively [172] (see Appendix C):

$$\begin{aligned} r_- &= \frac{\Omega_+(x_-)}{2\pi}\sqrt{|S''(x_0)|S''(x_-)}e^{-N(S(x_0)-S(x_-))}, \\ r_+ &= \frac{\Omega_+(x_+)}{2\pi}\sqrt{|S''(x_0)|S''(x_+)}e^{-N(S(x_0)-S(x_+))}. \end{aligned} \tag{5.5}$$

where $\Omega^{\pm}(x_i) = T_i^{\pm}(n_i/N)$ and $S(x) = \int^x \ln \frac{\Omega^-(y)}{\Omega^x(y)} dy$. This approximation provides an excellent match of simulations of the birth-death process (Figure 5.2).

5.3 Transition rates for multiple populations

5.3.1 Identical populations

We extended the results of Bressloff [172] to the case of M identical bistable neuronal populations with inhibition, forming a competitive network. To implement recurrent inhibition, we assume that, for each neuron in population i, the rate of transitions from quiescent to spiking is given by:

$$T_i^+(n) = Nf\left(\sum_j (W_{ij} - w_{\text{inh}})n_j/N\right). \tag{5.6}$$

This corresponds to homogeneously coupled populations in which every neuron in population i receives input of strength $(W_{ij} - w_{\text{inh}})/N$ from every neuron in population j. We assume that W_{ij} is positive for $i = j$ reflecting self-excitation and negative for $i \neq j$ reflecting inhibition between populations. We are interested in the probability distribution $P(m)$, where m is the number of populations in the high activity state. To calculate the probability distribution, we make several simplifying assumptions:

1. Each population is composed of N neurons.
2. $W_{ii} = W_{\text{in}}\ \forall i$.
3. $W_{ij} = W_{\text{out}}\ \forall i \neq j$.
4. Transitions between different network states (that is, between states of different m) occur only as transitions of a single population at a time. In other words, we neglect the possibility of two or more populations transitioning simultaneously.
5. In between a transition, the input from other clusters can be approximated as constant.

With these assumptions, populations can be treated as indistinguishable and hence the system is described as a birth-death process with respect to m. In this situation, $P(m)$ is determined solely by $r_+(m)$ and $r_-(m)$, the rate at which populations transition to or from the active state given that m are currently active. Once these rates are calculated, the dynamics of the system can be analyzed by studying the matrix of transition probabilities:

$$\mathbf{Q} = \begin{pmatrix} 1 - Mr_+(0) & Mr_+(0) & 0 & \cdots \\ r_-(1) & 1 - r_-(1) - (M-1)r_+(1) & (M-1)r_+(1) & \cdots \\ 0 & 2r_-(2) & 1 - 2r_-(2) - (M-2)r_+(2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{5.7}$$

In particular, the steady-state distribution of $P(m)$ is given by the eigenvector of $\mathbf{Q}$ corresponding to an eigenvalue of 1 (the Perron-Frobenius eigenvector).

We now describe the calculation of $r_{\pm}(m)$. First, we calculate $x_{\pm}(m)$, which now depend on m. This is done by numerically solving the two-dimensional system of equations:

$$0 = -x_- + f\left((W_{\text{in}} - w_{\text{inh}})x_- + (W_{\text{out}} - w_{\text{inh}})\left[mx_+ + (M - m - 1)x_-\right]\right), \tag{5.8}$$

$$0 = -x_+ + f\left((W_{\text{in}} - w_{\text{inh}})x_+ + (W_{\text{out}} - w_{\text{inh}})\left[(m-1)x_+ + (M - m)x_-\right]\right). \tag{5.9}$$

for each m. We then calculate r_+ and r_- using the same procedure as in the single population case, with equation 5.5, but with:

$$\gamma \leftarrow (W_{\text{in}} - w_{\text{inh}})\gamma, \tag{5.10}$$

and

$$\theta \leftarrow \frac{1}{W_{\text{in}} - w_{\text{inh}}}(\theta - (W_{\text{out}} - w_{\text{inh}})[mx_+ + (M - m - 1)x_-]) \tag{5.11}$$

for r_-, and

$$\theta \leftarrow \frac{1}{W_{\text{in}} - w_{\text{inh}}}(\theta - (W_{\text{out}} - w_{\text{inh}})[(m-1)x_+ + (M - m)x_-]) \tag{5.12}$$

for r_+. These redefinitions become apparent when observing that $f((W_{\text{in}} - w_{\text{inh}})x_- + (W_{\text{out}} - w_{\text{inh}})\left[mx_+ + (M - m - 1)x_-\right])$ (see equation (5.3)) is equal to $\tilde{f}(x)$ if $\tilde{f}(x)$ corresponds to $f(x)$ with the redefined γ and θ. The redefinition of θ amounts to approximating the sum $\sum_{i\neq j}(W_{ij} - w_{\text{inh}})x_j$ as $\sum_{i\neq j}(W_{\text{out}} - w_{\text{inh}})x_{\pm}$; that is, assuming all other populations' activities are fixed at the low or high activity local minimum, making this external input equivalent to a constant bias. We computed $P(m)$ for $M = 2$ using these assumptions, finding a good approximation between theory and simulations (Figure 5.3).

5.3.2 Non-identical populations

If the populations are not identical (that is, W_{ij} is arbitrary), it is no longer possible to reduce the 2^M possible states to M states as before. To find the stable states of the system, we solve numerically the set of M nonlinear equations:

$$0 = -x_i + f\left((W_{ii} - w_{\text{inh}})x_i + \sum_{i\neq j}(W_{ij} - w_{\text{inh}})x_j\right). \tag{5.13}$$

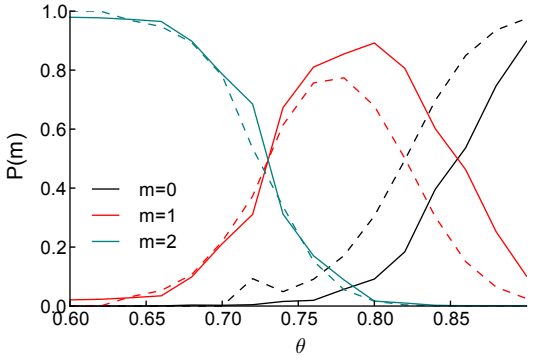

Figure 5.3: Estimation of the equilibrium state distribution $P(m)$ for two populations with $x_b = 0$, $f_0 = 2$, $\gamma = 4$, $N = 16$, $\theta = 0.85$, $w_{\text{in}} = 1.05$, $w_{\text{out}} = 0$, and $w_{\text{inh}} = 0.05$. Solid lines represent transition rates estimated from simulations using the Gillespie algorithm. Dashed lines represent the theoretical estimates from the Perron-Frobenius eigenvector of $\mathbf{Q}$.

We assume that the stable solutions of this system are of the form $x_i = x_\pm$; that is, the populations may be bistable. Note that depending on W, any given population may have two or only one stable states. We denote the possible states of the system as:

$$\mathbf{S} = \{(0, 0, \ldots 0), (0, 0, \ldots, 1), \ldots (1, 1, \ldots, 1)\}, \tag{5.14}$$

where 0 corresponds to the low activity state x_-, 1 corresponds to the high activity state x_+, and the ith term in the M dimensional ordered pair corresponds to the ith population. Not all 2^M states may be stable solutions of equations (5.13). Assuming that population i is bistable, we next compute $r_{i+}(s)$ or $r_{i-}(s)$ for state $s \in \mathbf{S}$. Define $x_i(s)$ as the equilibrium solution for x_i obtained from equations (5.13) in states s. Then we again apply equation (5.5) which we derived for the one population case, but with the substitutions:

$$\gamma \leftarrow (W_{\text{in}} - w_{\text{inh}})\gamma, \tag{5.15}$$

and

$$\theta \leftarrow \frac{1}{W_{\text{in}} - w_{\text{inh}}}(\theta - \sum_{i \neq j}(W_{ij} - w_{\text{inh}})x_j). \tag{5.16}$$

Using the computed $r_{\pm i}$, the matrix of transition probabilities between states $\mathbf{Q}$ can again be developed. In this case, $\mathbf{Q}$ will be 2^M dimensional, but for any state $s_k \in \mathbf{S}$ that is not a stable solution of equations (5.13), we set $Q_{kl} = Q_{lk} = 0 \; \forall l$. Otherwise, for any transition rate $r_{i\pm}(s_k)$ that corresponds to a transition from state s_k to state s_l, we set $\mathbf{Q}_{kl} = r_{i\pm}(s_l)$. The diagonal

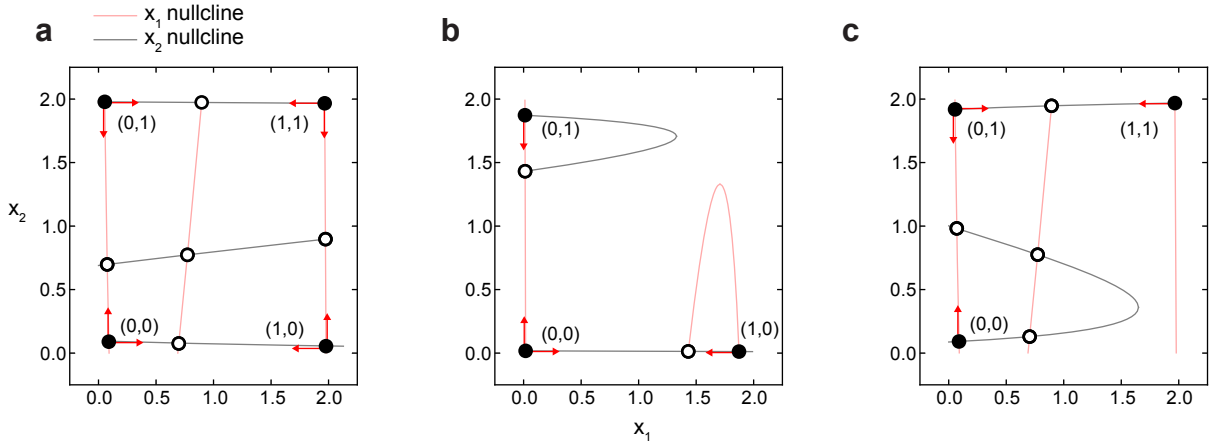

Figure 5.4: Possible equilibria of the set of nonlinear equations (5.13) with $x_b = 0$, $f_0 = 2$, $\gamma = 4$, $N = 16$, and $w_{\text{inh}} = 0.05$. Solid circles correspond to stable equilibria while unfilled circles represent unstable equilibria. Lines correspond to nullclines. Each stable equilibrium is labeled with its state $s \in \mathbf{S}$. Red arrows indicate the direction of possible transitions between stable states that can be computed using our theory. **(a)** Symmetric system with $w_{\text{in}} = 1.05$, $w_{\text{out}} = 0$, and $\theta = 0.85$. Four stable solutions exist. **(b)** Same as (a) but with a higher threshold $\theta = 1.2$ so that the $s = (1, 1)$ state is no longer a solution. **(c)** Asymmetric system with $W_{11} = 0.9$, $W_{12} = 0.15$, $W_{22} = 1.05$, $W_{21} = 0$, and $\theta = 0.85$. Population 1 can only be in the high activity state when it receives excitation from population 2.

elements for all stable states s_k are then given by $\mathbf{Q}_{kk} = 1 - \sum_l r_{kl}$. Similar to the identical population case, the equilibrium distribution $P(s)$ can be determined from the Perron-Frobenius eigenvector of $\mathbf{Q}$.

As an illustration of this technique, we plot the possible steady states of the system for a variety of W_{ij} (Figure 5.4). The fixed point structure of such Wilson-Cowan systems has been characterized previously [173]. Depending on the parameters of the system, the number of stable equilibria may change, and for asymmetric W_{ij} some populations may be bistable while others may not (Figure 5.4c).

5.4 Activity-dependent plasticity in a multiple-population model

5.4.1 Timescale separation of activity dynamics and weight dynamics

We next extend the framework we developed above to the case of slowly varying weights W_{ij}. We assume that the weights change according to some function g of the activity of populations

i and j:

$$\tau_W \frac{dW_{ij}}{dt} = g(x_i, x_j, W_{ij}). \tag{5.17}$$

In the limit of $\tau_W \to \infty$, $\frac{dW_{ij}}{dt}$ can be expressed in terms of the average weight change for fixed weights:

$$\tau_W \frac{dW_{ij}}{dt} \approx \langle g(x_i, x_j, W_{ij}) \rangle_{x_i, x_j | W_{ij}}, \tag{5.18}$$

where the expectation is over the stationary distribution of x_i and x_j given W_{ij}. We find this stationary distribution by determining the stable states of the system using equations (5.13). This yields the set of states $\mathbf{S}$. By assuming the system spends most of its time near these equilibria, we can write:

$$\tau_W \frac{dW_{ij}}{dt} \approx \sum_{s \in \mathbf{S}} P(s) g(x_i(s), x_j(s), W_{ij}). \tag{5.19}$$

The above amounts to a double timescale separation. We neglect the fast dynamics of individual neurons and focus only on switches in the activity of the neuronal populations, which in turn drive the much slower plasticity dynamics.

5.4.2 Application to two population model

We now apply this ansatz to the case of two interacting populations to study how spontaneous activation of neuronal populations influences the weight dynamics. We consider two populations $i = 1, 2$, each of which contains N neurons. The weight matrix is given by

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{pmatrix}. \tag{5.20}$$

The change in weight from population j to i is given by:

$$\frac{dW_{ij}}{dt} = \eta x_j x_i (x_i - \mu) \tag{5.21}$$

with the threshold $\mu = 1.8$ and $\eta = 0.001$. This learning rule is similar to the classic Bienenstock-Cooper-Munro rule, without the sliding threshold [171]. When the postsynaptic population i

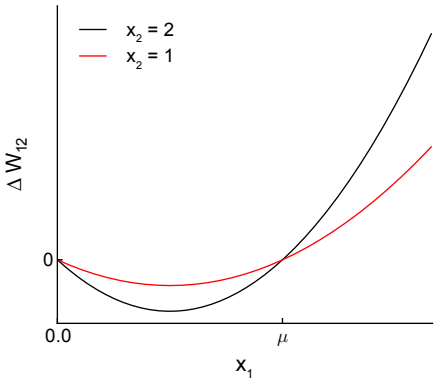

Figure 5.5: Plot of the plasticity rule $\Delta W_{12} \propto g(x_1, x_2)$ as a function of the postsynaptic population's activity x_1. The curve is shown for two values of x_2.

is in the high activity state, potentiation occurs, while it is in the low activity state, depression occurs.

In analogy with the spiking network simulations, we impose hard weight bounds so that $W_{ij} \in (w_{\text{min}}, w_{\text{max}})$. We also constrain the summed presynaptic input to each population, so that $W_{i1} + W_{i2} = W_{\text{sum}}$. The constraint is implemented by adding a constant to W_{i1} and W_{i2} on each timestep:

$$W_{ij} \leftarrow W_{ij} - (W_{i1} + W_{i2} - W_{\text{sum}})/2. \tag{5.22}$$

Because of the weight normalization, there are only two degrees of freedom among the four weight variables W_{ij}. Synaptic parameters were $w_{\text{min}} = 0$, $w_{\text{max}} = 1$, $w_{\text{inh}} = 0.2$, and $W_{\text{sum}} = 1$. N was set to 16.

When all synaptic weights were equal, both populations exhibited low levels of activity (Figure 5.6a, left). However, when recurrent excitation within population 1 and 2 was strong ($W_{11} = W_{22} = w_{\text{max}}$), the populations transitioned between low and high activity states, similar to the trained spiking network (Figure 5.6a, right). These different dynamics led to differential recruitment of plasticity. In particular, little plasticity was recruited when both populations had low rates, but the within-population weight W_{ii} was strongly potentiated whenever population i was in a high activity state (Figure 5.6a, bottom).

As described above, and similar to the spiking model of the previous chapter, we assumed that each row of the 2×2 matrix defining the weights for the two populations summed to a constant (synaptic normalization). This led to competition between the synapses onto each population. This assumption also allowed us to reduce the four-dimensional dynamics of the

weights to a two-dimensional plane. The axes corresponded to W_{11} and W_{22}, the strength of the synaptic weight within each of the two populations. When both W_{ii} were not potentiated, the network had only one stable activity level, and W_{ii} did not change substantially (Figure 5.6a, left and 5.6b, bottom left). But when the W_{ii} were sufficiently large, the system exhibited structured spontaneous activity and activation of populations drove the W_{ii} to w_{max}, their fully potentiated value (Figure 5.6a, right and 5.6b, top right). In total, the model illustrated that when W_{ii} is potentiated, structured spontaneous activity appears and causes W_{ii} to approach w_{max}.

5.4.3 Four population model

Next, we extended the model to study the effects of remapping. We defined a four-population model in which each population was half the size of the two-population case ($N = 8$). Further, synaptic strengths were halved ($w_{\text{max}} = 0.5$, $w_{\text{inh}} = 0.1$) so that pairs of populations could together form assemblies with identical dynamics to Figure 5.6a. When populations 1 and 3 were strongly connected and populations 2 and 4 were strongly connected, each pair (1,3) and (2,4) formed an assembly that underwent correlated fluctuations (Figure 5.6c, left). Further, if the connectivity was changed such that the pairs (1,2) and (3,4) were strongly connected, the new pairs underwent correlated fluctuations (Figure 5.6c, right).

We examined the dynamics of the synaptic weights in the subspace spanned by W_{12} and W_{34}. We found that both the (1,3) and (2,4) pairing (Figure 5.6c, left and 5.6d, bottom left) and the (1,2) and (3,4) pairing (Figure 5.6c, right and 5.6d, top right) were stable weight configurations. The model therefore shows that learned synaptic weight matrices can be stabilized by structured spontaneous activity, and that there may be multiple such stable weight matrices. As the number of possible combinations of subpopulations that can form assemblies increases, the number of stable weight matrices will grow exponentially.

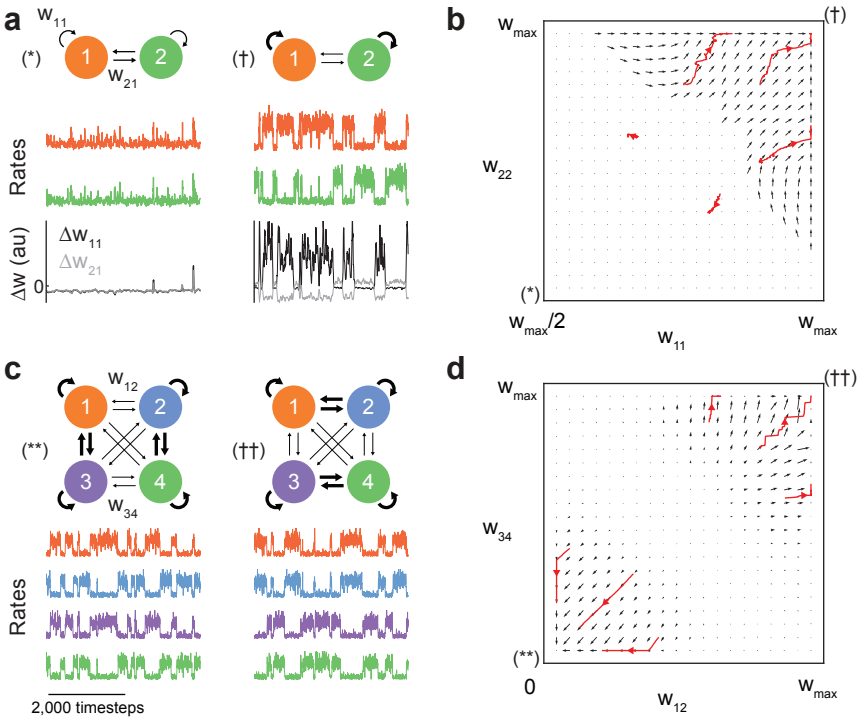

Figure 5.6: Weight dynamics in reduced model. **(a)** Two population model in the untrained (left) and trained (right) states. Top: Schematic illustrating weights. Thicker lines correspond to stronger weights. Middle: Dynamics of population firing rates. Bottom: Change in synaptic weights from population 1 onto itself (ΔW_{11}) and population 1 to 2 (ΔW_{21}) due to the rate fluctuations in the middle subpanel. **(b)** Vector field showing the weight dynamics for W_{11} and W_{12}. The top right is stable and corresponds to the right side of (a). Red trajectories correspond to individual simulations of the model. **(c)** Similar to (a) but for the four population model. Left: State in which populations (1,3) and populations (2,4) form a strongly connected assembly. Right: State in which populations (1,2) and populations (3,4) form a strongly connected assembly. **(d)** Vector field for the four population model showing the weight dynamics for W_{12} and W_{34}. The bottom left corresponds to the left side of (c), and the top right corresponds to the right side of (c).

6. *Conclusion*

In this thesis, I presented several studies on collective behavior of neurons in balanced networks. In Chapter 2, this took the form of correlations driven by common presynaptic input. In Chapters 3, 4, and 5, I focused primarily on the dynamics of neuronal assemblies. In this concluding chapter, I will discuss the relationship between these results and other models and discuss possible extensions of the work.

6.1 Uniform versus structured connectivity

A recent study demonstrated that networks with balanced excitation and inhibition can exhibit tuning for stimulus features such as orientation even if both inputs from tuned neurons in a previous layer and recurrent connections are random [174]. The mechanism relies on an amplification of the quenched component of the feedforward projections due to a balance between excitation and inhibition. The authors argue that this mechanism helps explain the presence of orientation tuning in rodents, whose visual cortices lack orientation columns.

On the other hand, it is known that recurrent connections in mouse visual cortex are not random [106]. Neurons are orientation-tuned at eye opening and seemingly random synaptic connections are refined through experience so that similarly tuned neurons are more likely to be connected [144]. The model of Hansel & van Vreeswijk [174], therefore, may be more appropriate for the immature visual cortex or other cortical regions.

However, the study raises the question: why is specific recurrent connectivity present if stimulus tuning can be accomplished through random connections? Our study provides a few possible reasons. First, recurrent circuitry can amplify responses and perform pattern completion if a partial stimulus is presented. This amplification is observed in several primary sensory

regions [161, 162, 163]. Second, we show that recurrent inhibitory circuitry can suppress fluctuations when stimuli are presented (Figure 4.5). Third, this recurrent connectivity may serve to temporally prolong the response to stimuli, as observed in auditory cortex [163]. This result may be beneficial for neuronal computations that rely on dynamical "traces" of past inputs [175].

6.2 Balanced attractor networks and scaling

Since cortical networks exhibit specific recurrent connectivity, we developed the models examined in this thesis, composed of many neuronal assemblies embedded in a sea of random connections. These networks can be seen as a merging of ideas from the balanced network literature [24, 27], which proposes models of trial-to-trial variability, and the Hopfield network literature [160], which proposes models of neuronal assemblies. This connection has been made in many studies previously [108, 114, 111]. Our contribution is specifically to study the implications of such attractor structures for trial-to-trial variability (Chapter 3) and how such attractors could form in networks with realistic plasticity mechanisms (Chapter 4). Below, we discuss more thoroughly the plausibility of merging these ideas in the context of scaling arguments.

One of the requirements to merge these two lines of research was a specific scaling of the size of the neuronal assemblies. In balanced networks, if neurons receive K connections, then the connection strength J must be proportional to $1/\sqrt{K}$ (see Chapter 1). We consider the case of *dense connectivity*, so that $K \propto N$. We want the perturbation away from a homogeneous balanced network due to the attractor structure to be $O(1)$, to prevent large attractors from dominating the network activity. We let the size of a neuronal assembly be C neurons and define the *coding fraction* as C/N. Then we must have:

$$C(J_{\text{in}}p_{\text{in}} - J_{\text{out}}p_{\text{out}}) = O(1), \tag{6.1}$$

where $J_{\text{in}} - J_{\text{out}}$ is how much stronger connections within the assembly are and $p_{\text{in}} - p_{\text{out}}$ is how much more probable they are. There are three scenarios:

1. $C \propto K$, that is, each assembly is composed of a macroscopic number of neurons and the coding fraction is constant. This is the case of a classic Hopfield network [160]. However,

to satisfy equation (6.1), we must have $J_{\text{in}}p_{\text{in}} - J_{\text{out}}p_{\text{out}} \propto 1/K$. This presents a fine-tuning problem, as the J's themselves are proportional to $1/\sqrt{K}$, much larger than the desired difference between J_{in} and J_{out}. Classic Hopfield networks avoided this problem as J was proportional to $1/K$.

2. $C \propto \sqrt{K}$, so the coding fraction is small. This was the case considered in our studies and does not require fine-tuning as setting $J_{\text{in}} = aJ_{\text{out}}$ for some constant a ensures equation (6.1) is satisfied. In such networks, the number of neurons in a local assembly is smaller by a factor of $\sqrt{K}$ than the total number of inputs received. If stimuli are assumed to excite a few assemblies at once, this corresponds to a network exhibiting *sparse coding*, which is frequently encountered in cortex [120].

3. $C \propto 1$. This corresponds to a case of extremely sparse coding, as assembly size does not grow at all with K. Further, in this case we must have $J_{\text{in}} = O(1)$ while $J_{\text{out}} = O(1/\sqrt{K})$, corresponding to a few extremely strong local connections. For these reasons, we do not consider this case.

Typically, spiking implementations of attractor networks fall into the second category, as was true for our study [108, 114, 110]. Roudi & Latham [111] took an alternative approach and studied the first category, differentiating between "background weights" that were $O(\sqrt{K})$ and "foreground weights" (connections between neurons in the same assembly) that were $O(1/K)$. In order to obtain stable assembly activation, these foreground weights needed to be tuned to within 6% in a network of 10,000 neurons, and it is unclear that such fine-tuning is plausible.

It is therefore an open question how to robustly merge balanced and attractor networks with a nonvanishing coding fraction. Unlike primary sensory regions, higher-order associative regions such as prefrontal cortex have a large degree of "mixed selectivity," with high coding fractions and assembly overlap [176]. The fine-scale architecture of such regions has yet to be determined.

6.3 Future work

6.3.1 Modulations of correlation in different states

In Chapter 2, we showed that a change in the cellular properties of neuron pairs (in that case, a change in membrane conductance via increased balanced input) can change the correlation between their spike trains. In showing this, we assumed that the neurons received external, correlated input (Figure 2.3). This assumption of external correlating input is often made in modeling studies concerned with pairwise statistics [86, 46, 10, 49, 177, 178]. However, in recurrent balanced networks, correlations are often weak [27]. As a result, modulations of pairwise statistics that require external correlating inputs can be difficult to observe in randomly coupled networks.

In Chapter 3 we demonstrated that particularly connectivity structures can induce correlated fluctuations in networks without assuming correlated input. This and other recent studies that focus on the generation of collective activity in balanced networks [112, 179] open the possibility of studying the modulation of correlations that emerge internally from network interactions, instead of being imposed externally. For example, Deco & Hugues [112] modeled the influence of attention as a signal that biased activity in particular states of a multistable system, similar to the influence of a signal in our clustered network (Figure 3.5). It was found that this signal decreased correlation, as observed in visual cortex [15]. Recently, it was shown that burstiness and action-potential height are modulated with attention [180]. Studies that include recurrently generated correlated activity along with such modulations of cellular properties may shed light on the mechanisms of correlation modulation.

6.3.2 Sequence generation

Several experimental studies have shown that spontaneous activity patterns in cortical networks can exhibit repeatable spatiotemporal structure. For example, stereotyped waves of activity in visual cortex occur spontaneously after stimulation [133, 138, 137]. The fact that these activity patterns reflect previously experienced stimuli suggest that the visual cortex may be sampling from a statistical model of the external world [136, 135].

The attractor networks studied in this thesis demonstrate complex temporal dynamics, consisting of activations of neuronal assemblies that persist for timescales longer than the neuronal time constant. However, the sequence of activations of different assemblies is unpredictable. This is because between-cluster connectivity is uniform, so that no cluster is preferentially activated by another. We briefly investigated sequential neuronal activation in the form of feedforward networks (Figure 3.10). However, the plasticity mechanisms that are necessary to stabilize such structures are unclear. A recent study used simple STDP with normalization of both presynaptic and postsynaptic weights to produce feedforward chains in a simplified network [154]. Whether these results can be applied to balanced networks with more realistic plasticity rules remains to be studied.

Klampfl & Maass [181] showed that a purely timing-based STDP rule can lead to spontaneous activity that includes patterns of spikes that reflect past inputs. This rule also did not account for the firing rate dependence of STDP [156]. Nonetheless, the authors demonstrated that this plasticity mechanism allows the network to integrate long-term memory with current inputs, a potentially useful feature for reservoir computing [175].

6.3.3 Diversity of interneuron subtypes

Modeling studies often segregate neurons into two classes: excitatory and inhibitory. Indeed, this partition is a central assumption in the derivation of the balance conditions (Section 1.4.1). However, cortical circuits actually consist of a diverse set of neurons, each with different cellular and connectivity properties. Recently, the functional and architectural properties of local inhibitory interneurons in cortical sensory areas have been probed experimentally [119, 168, 167, 182, 183, 184]. Future theoretical models should be informed by these advances.

In our results concerning neuronal assemblies, connections involving inhibitory neurons was assumed to be nonspecific (Figure 3.1). This assumption led to a model in which excitatory neuronal assemblies contained strong excitatory recurrence and projected lateral inhibition to other assemblies. If, on the other hand, inhibitory neurons were highly localized and projected strongly only within an assembly, the activation of any assembly would be discouraged. In this situation, multistability of the network may be lost. Hence, inhibitory connectivity can modulate activity fluctuations.

The functional and connectivity properties of somatostatin-expressing (SOM) and parvalbumin-expressing (PV) neurons, two important classes of cortical inhibitory neurons, have recently been probed using optogenetics. In mouse frontal cortex, SOM neurons target local pyramidal neurons densely and nonspecifically [119], an observation that motivated the assumption of unstructured inhibition in Chapter 3. SOM neurons have been shown to receive long-range horizontal excitation, suggesting that they may be responsible for lateral inhibition in visual cortex [167]. On the other hand, PV neurons appear to target pyramidal neurons with similar orientation preference more specifically than SOM neurons [168].

These distinctions offer rich possibilities for future studies of functionally clustered networks. Our work predicts that selectively activating PV or SOM neuron subpopulations will differentially effect ongoing dynamics, with PV neurons suppressing assembly dynamics and SOM neurons promoting it. PV neurons may also be important for modulating the timescale of evoked responses within an assembly. Future studies should construct models that incorporate nonuniform connectivity among pyramidal neurons and multiple inhibitory neuron classes.

6.4 Final remarks

Balanced excitation and inhibition in large recurrent networks is ubiquitous in the modeling literature. In addition to its consistency with experimental data, it is a seemingly necessary component in model networks that exhibit substantial trial-to-trial variability. However, the dynamics of randomly connected balanced networks are relatively uninteresting, at least at the level of the mean-field. This thesis attempts to extend the balanced network approach to nonuniform connection structures and demonstrates novel dynamical phenomena in these networks. However, much work remains to understand how balanced networks can perform complex computations. How do systems with so much seeming randomness, both in connection structure and dynamics, compute? Further work that merges ideas from disordered systems, machine learning, signal processing, and neuroscience will be necessary to answer this question.

A. *Spike train statistics*

In this appendix we define some elementary quantities for measuring properties of spike trains that are used throughout the text. Suppose a neuron produces spikes at times $t_1, t_2, t_3, \ldots$. We refer to these times $\{t_k\}$ as the neuron's *spike train*. Sometimes, we will be interested in situations in which we record spike times over multiple trials, or from multiple neurons simultaneously. In these cases, we will use additional indices to denote the trial or neuron index.

A.1 Firing rate

The *firing rate* of a neuron is the rate at which it emits spikes. Typically, the firing rate is estimated by counting the number of spikes in some time window, say between t and $t+T$, and dividing by T. Therefore, the rate r is

$$r(t, t+T) = \frac{\#\{t_k | t_k \in (t, t+T)\}}{T}. \tag{A.1}$$

A.2 Inter-spike interval distribution

One measure of the variability of a spike train is how regularly spaced the spikes are. We can look at the *inter-spike intervals*, or ISIs of the spike train to evaluate this. If $\{t_k\}$ are the neuron's spikes, define

$$d_k = t_{k+1} - t_k \tag{A.2}$$

as the kth ISI. We can examine the distribution of these intervals $P(d)$, which is the *inter-spike interval distribution*. A frequently calculated measure is the *coefficient of variation* of this

distribution, which is the ratio of the standard deviation of the distribution to the mean.

$$\text{CV}_{\text{ISI}} = \frac{\text{STD}(d)}{\text{E}(d)} = \frac{\sqrt{\frac{1}{N}\sum_k d_k^2 - (\frac{1}{N}\sum_k d_k)^2}}{\frac{1}{N}\sum_k d_k}. \tag{A.3}$$

where N is the number of ISIs. For a homogeneous Poisson process, CV_{ISI} is 1.

A.3 Fano factor

The above quantities relate to the *temporal* properties of spike train; that is, the statistics of one spike train over time. We are also often interested in the *trial-to-trial* variability of a spike train. Trial-to-trial variability tells us how reliably a neuron produces spikes when presented with the same stimulus multiple times on different trials. Let $s = 1, 2, 3, \ldots$ be the stimulus presentation index and let $\{t_k^s\}$ be the neuron's spike times on the sth stimulus presentation.

An important measure of trial-to-trial variability is the *Fano factor*, which is defined as the ratio of the variance of the spike counts across trials to the mean in some time window. Let $N^s(t, t+T) = \#\{t_k^s | t_k^s \in (t, t+T)\}$ be the number of spikes on trial s that happened between t and $t+T$. Then the Fano factor is

$$\text{FF}(t, t+T) = \frac{\text{Var}\left(N(t, t+T)\right)}{\text{E}\left(N(t, t+T)\right)} = \frac{\frac{1}{N}\sum_s (N(t, t+T))^2 - \left(\frac{1}{N}\sum_s N(t, t+T)\right)^2}{\frac{1}{N}\sum_s N(t, t+T)}, \tag{A.4}$$

where N is the number of trials. For a homogeneous Poisson process, the Fano factor, like the CV of the ISI distribution, is 1.

A.4 Spike count correlation

So far we have only been considering one neuron at a time, but we may wish to ask how related the spikes of two neurons are. For neurons that are recorded under multiple stimulus conditions, we can compute the *signal correlation*:

$$\text{Corr}_{\text{signal}}(i, j, T) = \sum_{s \in \text{Stimuli}} \sum_{t \in \text{Trials}} \frac{\text{Cov}(N_i^T(s,t), N_j^T(s,t))}{\sqrt{\text{Var}(N_i^T(s,t))\text{Var}(N_i^T(s,t))}}. \tag{A.5}$$

In order to isolate correlations due to common stimulus preference from those due to network architecture, we define the *noise correlation*:

$$\text{Corr}_{\text{noise}}(i,j,s,T) = \sum_{t \in \text{Trials}} \frac{\text{Cov}(N_i^T(s,t), N_j^T(s,t))}{\sqrt{\text{Var}(N_i^T(s,t))\text{Var}(N_i^T(s,t))}}. \tag{A.6}$$

Throughout this thesis, we are primarily interested in noise correlations, as we measure correlations either when neurons receive a fixed stimulus across trials, or during spontaneous conditions.

A.5 Covariance functions

Finally, we may be interested in how spikes are correlated in time. The temporal *covariance function* (sometimes called correlation function) of two spike trains tells us the likelihood of a spike in spike train 2 at time $t + \tau$, given that there was a spike in spike train 1 at time t. Specifically, let $\{t_k^i\}$ be the spike times of neuron i. Also, let $y_i(t) = \sum_k \delta(t - t_k^i)$. Then the covariance function between spike trains i and j is

$$C_{ij}(\tau) = \text{E}\left[y_i(t)y_j(t+\tau)\right] - \text{E}\left[y_i(t)\right]\text{E}\left[y_j(t)\right]. \tag{A.7}$$

The expectation is taken over time, and the variable τ is known as the *lag*. When $i = j$, the above function is called the *autocovariance function*. Otherwise, it is called the *cross-covariance function*.

B. *Models of synaptic plasticity*

In recent years, several biophysically motivated models of spike timing dependent plasticity (STDP) have been developed [146, 148]. In this thesis, we used one of them, the voltage-based model of Clopath et al. [146], to study the formation of neuronal assemblies. In this appendix, we review a few models of plasticity and discuss their application to the formation of assemblies.

B.1 The BCM Rule

The Bienenstock-Cooper-Munro (BCM) rule [171] is a theory for the change in synaptic weight w_{ij} from neuron j to neuron i. It depends only on the firing rates of the two neurons, r_i and r_j. It was developed to explain the formation of orientation selectivity and binocular tuning in cat visual cortex. In its typical form, weights evolve according to:

$$\frac{dw_{ij}}{dt} = r_i(r_i - \theta)r_j - \epsilon w_{ij}. \tag{B.1}$$

The variable θ represents the threshold in firing rate of the postsynaptic neuron above which potentiation is recruited. In the original formulation of the model, θ is the expectation of r_i raised to some power and scaled by a constant, $\theta = cE^p[r_i]$. This "sliding threshold" allows the ratio of potentiation to depression to be modulated by the level of activity of the postsynaptic neuron, making potentiation more difficult to recruit for high firing rate neurons.

B.2 Classic pair based STDP

Classic experimental studies have demonstrated that the recruitment of long-term potentiation (LTP) and long-term depression (LTD) depends on the timing of action potentials [185, 186].

Typically, the dependence of w_{ij} on a pair of spikes t_i and t_j separated by a time lag $\Delta t = t_i - t_j$ is modeled as $\Delta w_{ij} \propto F(\Delta t)$, with $F(\Delta t)$ defined as the following [187]:

$$F(\Delta t) = \begin{cases} A_+ e^{-|\Delta t|/\tau_+} & \text{if } \Delta t > 0, \\ A_- e^{-|\Delta t|/\tau_-} & \text{if } \Delta t < 0. \end{cases} \tag{B.2}$$

Under such a rule, timing relationships between the two neurons contribute most of the weight changes. Pairings with positive Δt ("pre-post" pairings) tend to recruit potentiation, while negative Δt ("post-pre" pairings) recruits depression. If the integral of $F(\Delta t)$ is zero, then two neuron firing independently will recruit on average zero weight change, regardless of firing rate, in contrast to the BCM rule.

B.3 Triplet rules

Motivated by the dependence of LTP and LTD on firing rate [156], several authors have attempted to extend the classic pair based STDP model above. An obvious approach is to consider interactions between more than two spikes. This has motivated "triplet rules" in which weight changes depend on three spikes [170, 147]. We review the triplet rule of Pfister & Gerstner [170]. The dynamics of the model depend on four dynamical variables:

$$\begin{aligned} \frac{dr_1}{dt} &= -r_1/\tau_+ + \sum_k \delta(t - t^k_{\text{pre}}) \\ \frac{dr_2}{dt} &= -r_2/\tau_x + \sum_k \delta(t - t^k_{\text{pre}}) \\ \frac{do_1}{dt} &= -o_1/\tau_- + \sum_k \delta(t - t^k_{\text{post}}) \\ \frac{do_1}{dt} &= -o_1/\tau_y + \sum_k \delta(t - t^k_{\text{post}}). \end{aligned} \tag{B.3}$$

The sums increment the r and o variables by 1 whenever a pre- or postsynaptic spike, respectively, occurs. Upon a presynaptic spike, w_{ij} is updated according to:

$$\Delta w_{ij} = -o_1 \left(A_- + A_3^- r_2 \right), \tag{B.4}$$

while upon a postsynaptic spike, the update is:

$$\Delta w_{ij} = -r_1 \left(A_+ + A_3^+ o_2 \right). \tag{B.5}$$

If $A_3^- = A_3^+ = 0$, the rule reduces to the pair based rule above. Hence, these two terms determine the extra influence of spike triplets. Positive A_3^- yields additional LTD when a pre-post-pre pairing occurs compared to a classic post-pre pairing. Similarly, positive A_3^+ yields additional LTP when a post-pre-post pairing occurs compared to a classic pre-post pairing.

B.4 Voltage based rule of Clopath et al.

Pair and triplet rules depend only on spike times, while the cellular processes involved in plasticity likely involve other quantities. It has been shown that postsynaptic voltage influences LTP and LTD [188], and, based on this, Clopath et al. [146] developed a voltage based STDP rule. According to the rule, the dynamics of w_{ij} were given by:

$$\frac{dw_{ij}}{dt} = -A_{\text{LTD}} s_j(t) R\left(u_i(t) - \theta_{\text{LTD}}\right) + A_{\text{LTP}} x_j(t) R\left(V_i(t) - \theta_{\text{LTP}}\right) R\left(v_i(t) - \theta_{\text{LTD}}\right), \tag{B.6}$$

where R is a linear rectifying function ($R(x) = 0$ if $x < 0$, $R(x) = x$ otherwise), u_i and v_i represent the membrane voltage V_i low-pass filtered with time constants τ_u and τ_v respectively, and x_j represents the spike train $s_j(t) = \sum_k \delta(t - t_j^k)$ low-pass filtered with time constant τ_x.

Under this rule, LTD occurs when the low-pass filtered postsynaptic voltage $u_i(t)$ exceeds θ_{LTD}. Similarly, LTP occurs when the low-pass filtered postsynaptic voltage $v_i(t)$ exceeds θ_{LTD} *and* the instantaneous voltage $V_i(t)$ exceeds θ_{LTP}.

B.5 Calcium based rule of Graupner & Brunel

Calcium concentration in the dendritic spine has been proposed as a determinant of synaptic plasticity [189], and models have been created based on this observation [190, 191, 148]. We review the model of Graupner & Brunel [148]. According to the model, the efficacy of a synapse

is given by:

$$\tau \frac{dw_{ij}}{dt} = -w_{ij}(1 - w_{ij})(w^* - w_{ij}) + \gamma_p(1 - w_{ij})H(c(t) - \theta_p) - \gamma_d w_{ij} H(c(t) - \theta_d), \tag{B.7}$$

where H is the Heaviside step function, γ_p, γ_d are the strengths of LTP and LTD, and θ_p, θ_d are the calcium concentration thresholds for recruitment, respectively. The first term implements bistability in the synaptic weight, with a saddle point at w^*. For large γ_p, γ_d, this term can be neglected. The calcium concentration $c(t)$ is given by $\tau_{\text{Ca}} \frac{dc(t)}{dt} = -c(t)$ and is incremented by C_{pre} or C_{post} when a pre- or postsynaptic spike occurs, respectively. Depending on model parameters, several types of STDP curves can be produced. When $C_{\text{post}} > C_{\text{pre}}$, curves qualitatively similar to that of classic pair based rules occur at low frequency.

B.6 STDP rules and formation of neuronal assemblies

In Chapter 4, we used the voltage based rule [146] in a network that was capable of forming neuronal assemblies. What were the necessary aspects of this rule that allowed this to occur, and could other rules reproduce this result? In this section, we provide a non-rigorous discussion of the requirements, based on extensive simulations. Since homeostasis and synaptic bounds were imposed through other mechanisms, the requirements of a synaptic rule to reproduce the results of Chapter 4 are:

1. Depression dominates at low firing rates, so that assemblies do not form spontaneously,
2. Potentiation is recruited when pre- and postsynaptic neurons fire at elevated firing rates,
3. Strong potentiation is *not* recruited when only the presynaptic neuron is firing.

Assuming the network can be approximated as asynchronous and Poisson, the above requirements depend only on the dependence of the Δw_{ij} on r_i and r_j. Thus, a BCM-type analysis is appropriate. In Figure B.1a, we plot Δw_{ij} as a function of pre- and postsynaptic firing rate for the three modern rules described above. Note that for traditional pair-based rules with $\int F(\Delta t) d\Delta t = 0$, the corresponding plot would have no potentiation or depression.

All three rules have a strong dependence of potentiation or depression on postsynaptic firing rate, with intermediate rates recruiting net depression and higher rates recruiting net potentiation. Thus, all three rules satisfy requirements 1 and 2. The only qualitative difference is in the calcium based rule's dependence on presynaptic firing rate: for sufficiently strong presynaptic firing rates, potentiation can be recruited even if the postsynaptic neuron is firing at a low rate. This can be easily seen in the definition of the model: if the presynaptic neuron fires enough, $c(t)$ will eventually spend substantial time above θ_p, recruiting potentiation. In the context of assembly formation, this has a deleterious effect: when assembly A is active, synapses from assembly A to assembly B will potentiate while synapses from assembly B to assemby B will not. Thus, stimulation of assembly A interferes with the training of assembly B. In numerical simulations, the qualitative results of Chapter 4 were reproduced with the triplet and calcium based rules, but the calcium based rule had fewer assemblies remain potentiated than the other rules due to this interference (Figure B.1b,c). Nonetheless, these results show that the effects of structured spontaneous activity occur similarly for a variety of STDP rules.

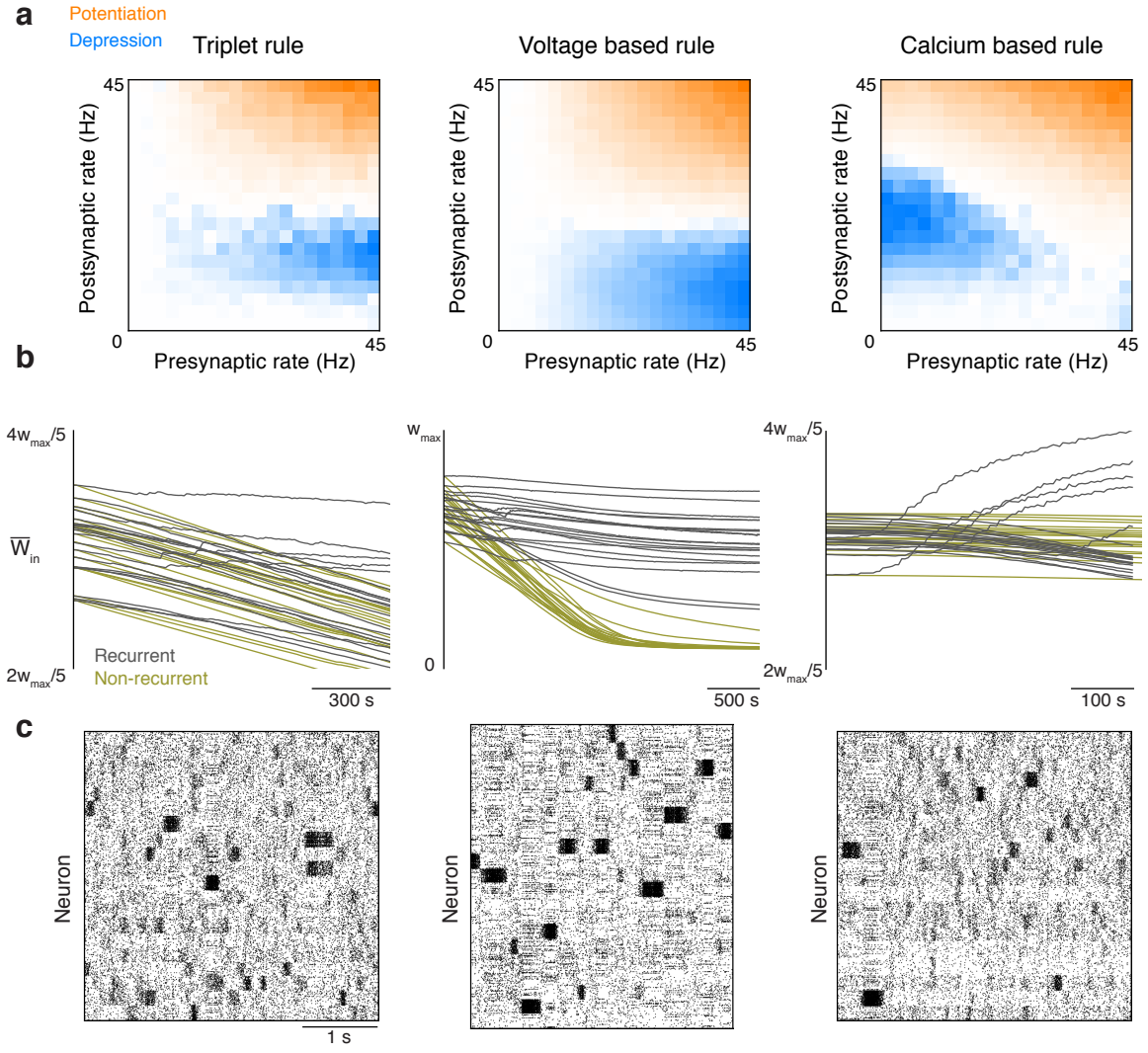

Figure B.1: Results for different STDP rules: the triplet STDP rule of 170, the voltage based STDP rule of 146, and the calcium based rule of 148. **(a)** Schematic showing the average change in synaptic strength as a function of pre- and postsynaptic firing rate, assuming both neurons fire as Poisson processes. **(b)** Comparison between recurrent and external networks as in Figure 2b of the main text. Although the number of stored assemblies depends on the specific rule, structured spontaneous activity improves the stability of assemblies in all cases. **(c)** Structured spontaneous dynamics during spontaneous activity after training. For the triplet based rule, the following parameters were used (following the notation of 170): $\tau_+ = 16.8$ ms, $\tau_- = 33.7$ ms, $\tau_x = 101$ ms, $\tau_y = 125$ ms, $A_2^+ = 7.5 \times 10^{-10}$, $A_3^+ = 9.3 \times 10^{-3}$, $A_2^- = 7 \times 10^{-3}$, $A_3^- = 2.3 \times 10^{-4}$. The voltage based rule is described in Chapter 4. For the calcium based rule, the following parameters were used: $\gamma_p = 7.25$, $\gamma_d = 3.31$, $\theta_p = 1.3$, $\theta_d = 1$, $\tau_{\mathrm{Ca}} = 22.7$ ms, $C_{\mathrm{pre}} = 0.56$, $C_{\mathrm{post}} = 1.24$.

C. *Derivation of escape rates for bistable population*

In this appendix we follow closely the derivation and notation of Bressloff [172] to obtain estimates of the escape rate between bistable potentials in a finite-sized population model governed by a birth-death process.

As in Chapter 5, we will work with a system that in the large system size limits obeys the birth-death process:

$$\frac{dP(n,t)}{dt} = T^+(n-1)P_i(n-1) + T^-(n+1)P(n+1) - \left(T^+(n) + T^-(n)\right)P(n), \tag{C.1}$$

with

$$T^+(n) = Nf(n/N), \;\; T^-(n) = n. \tag{C.2}$$

We then define the normalized variable $x = n/N$ and treat it as a continuous variable. We also define the scaled transition rates $N\Omega^{\pm} = T^{\pm}(Nx)$, allowing us to rewrite our master equation as:

$$\frac{dP(x,t)}{dt} = \Omega^+(x-1/N)P(x-1/N) + \Omega^-(x+1/N)P(x+1/N) - \left(\Omega^+(n) + \Omega^-(n)\right)P(n), \tag{C.3}$$

We next assume that the system has two metastable states, whose locations we write as x_+ and x_-, separated by a saddle point x_0. We seek to find the rate of transitions from $x_{\pm}$ to the other metastable state, which we define as $r_{\pm}$. We will take advantage of our knowledge that $r_{\pm}$ decrease exponentially with the system size N.

Suppose that $\Pi(x)$ is a quasistationary solution of the master equation corresponding to the system spending time in one of the basins of attraction of the fixed points $x_{\pm}$, which we denote

x^*. We assume that the solution takes the form:

$$\Pi(x) = K(x)e^{-NS(x)}, \tag{C.4}$$

with $K(x^*) = 1$ and $S(x^*) = 0$. We then substitute this into the master equation (C.3) with $P(x,t) = \Pi(x)$ and $\frac{dP}{dt} = 0$. Then, we perform a Taylor expansion about x^* which leads to the equation:

$$\Omega^-(x)\left(e^{-S'(x)} - 1\right) + \Omega^+(x)\left(e^{S'(x)} - 1\right) = x\left(e^{-S'(x)} - 1\right) + f(x)\left(e^{S'(x)} - 1\right). \tag{C.5}$$

This equation can be interpreted as a stationary Hamilton-Jacobi equation $H(x, S'(x)) = 0$ where the Hamiltonian is given by equation (C.5). So we can think of the dynamics of x as being governed by that of a particle, whose position is x and momentum is $p = S'(x)$, governed by the corresponding Hamiltonian dynamical system. Hamilton's equations yield:

$$\dot{x} = \frac{\partial H}{\partial p} = -xe^{-p} + f(x)e^{p} \tag{C.6}$$

$$\dot{p} = -\frac{\partial H}{\partial x} = xe^{-p} - 1 + f'(x)[e^{p} - 1] \tag{C.7}$$

Here $S(x)$ is interpreted as classical action along the *least-action trajectory* from x^* to x (assuming $S(x^*) = 0$). This least action trajectory is the one that minimizes the integral of the Lagrangian $L = p\dot{x} - H(x,p)$:

$$S(x) = \inf_{x(t_0)=x^*, x(T)=x} \int_0^T L(x, \dot{x})dt. \tag{C.8}$$

Along this path, $p = S'(x)$, so we can find $S(x) = \int_{x^*}^{x} p(x')dx'$ where the integral is taken along the path. So the leading order term in the WKB approximation corresponds to zero-energy solutions such that $H(x, p(x)) = 0$. It then remains to determine the multiplicative factor $K(x)$ in the WKB approximation. This can be accomplished by plugging the WKB ansatz into the master equation (C.3), which leads to the following expression that can be used to obtain $K(x)$:

$$\frac{\partial H}{\partial p}\frac{K'}{K} = -\frac{1}{2}p'\frac{\partial^2 H}{\partial p^2} - \frac{\partial^2 H}{\partial p \partial x}. \tag{C.9}$$

There are two types of solutions corresponding to $H(x, p(x)) = 0$. The first is the $p = 0$ solution, corresponding to the deterministic Wilson-Cowan equation. In other words, this solution represents the deterministic relaxation back to x^* given a perturbation away from it. The second solution is the so-called *activation* solution, $p(x) = p^*(x) = \ln \frac{\Omega_-(x)}{\Omega_+(x)}$.. We can solve equation (C.9) to obtain $\Pi(x)$ for activation trajectories:

$$\Pi(x) = \frac{A}{\sqrt{\Omega_+(x)\Omega_-(x)}} e^{-NS(x)}, \tag{C.10}$$

with $S(x) = \int^x \ln \frac{\Omega_-(x')}{\Omega_+(x')} dx'$, and for relaxation trajectories:

$$\Pi(x) = \frac{B}{\sqrt{\Omega_+(x)\Omega_-(x)}}. \tag{C.11}$$

There is one remaining calculation before we can determine $r_\pm$: we must match the two quasistationary solutions at the saddle point x_0. This is done by using a diffusion approximation of the master equation around x_0 and matching the flux from the relaxation and activation trajectories.

The constant flux through the saddle is given by:

$$J = (\Omega_+(x) - \Omega_-(x))\Pi(x) - \frac{1}{2N}\frac{\partial}{\partial x}[(\Omega_+(x) + \Omega_-(x))\Pi(x)]. \tag{C.12}$$

Taylor expading this equation and integrating, we find:

$$\Pi(x) = \frac{JN}{\Omega_+(x_0)} e^{(x-x_0)^2/\sigma^2} \int_x^\infty e^{-(x'-x_0)^2/\sigma^2} dx', \tag{C.13}$$

with:

$$\sigma = \sqrt{\frac{2\Omega_+(x_0)}{N[\Omega'_x(x_0) - \Omega'_-(x_0)]}}. \tag{C.14}$$

The asymptotic behavior of this solution is given by:

$$\Pi(x) = \begin{cases} \frac{NJ\sigma^2}{(x-x_0)\Omega_+(x_0)}, & x - x_0 \gg \sigma \\ \frac{NJ\sigma\sqrt{\pi}}{\Omega_+(x_0)} e^{(x-x_0)^2}, & x - x_0 \ll \sigma. \end{cases} \tag{C.15}$$

Matching the first case to the relaxation trajectory leads to $B = J$ while matching the second case to the activation trajectory requires a Taylor expansion of $S(x)$ around x_0, leading to:

$$J = \frac{A\Omega_+(x_0)}{\sqrt{\Omega_+(x_0)\Omega_-(x_0)}}\sqrt{\frac{|S''(x_0)|}{2\pi N}}e^{-NS(x_0)}. \tag{C.16}$$

The escape rate r_- is given by the integral $r_-^{-1} = \frac{1}{J}\int_0^{x_0}\Pi(x')dx'$.. This integral is strongly peaked around x_-, and a Gaussian approximation of $\Pi(x)$ around that point yields:

$$r_- = \frac{\Omega_+(x_-)}{2\pi}\sqrt{|S''(x_0)|S''(x_-)}e^{-N(S(x_0)-S(x_-))}, \tag{C.17}$$

and similar arguments for r_+ yield:

$$r_+ = \frac{\Omega_+(x_+)}{2\pi}\sqrt{|S''(x_0)|S''(x_+)}e^{-N(S(x_0)-S(x_+))}. \tag{C.18}$$

Bibliography

[1] D. Tolhurst, J. Movshon & A. Dean. The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Res.* **23**(8), 775–785 (1983).

[2] K. H. Britten, M. N. Shadlen, W. T. Newsome & J. A. Movshon. Responses of neurons in macaque MT to stochastic motion signals. *Vis. Neurosci.* **10**(06), 1157–1169 (1993).

[3] A. Manwani & C. Koch. Detecting and estimating signals in noisy cable structure, i: neuronal noise sources. *Neural Comput.* **11**(8), 1797–1829 (1999).

[4] A. L. Hodgkin & A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**(4), 500–544 (1952).

[5] Z. F. Mainen & T. J. Sejnowski. Reliability of spike timing in neocortical neurons. *Science* **268**(5216), 1503–1506 (1995).

[6] E. Schneidman, B. Freedman & I. Segev. Ion channel stochasticity may be critical in determining the reliability and precision of spike timing. *Neural computation* **10**(7), 1679–1703 (1998).

[7] G. R. Holt, W. R. Softky, C. Koch & R. J. Douglas. Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *J. Neurophysiol.* **75**(5), 1806–1814 (1996).

[8] M. R. Cohen & A. Kohn. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* **14**(7), 811–819 (2011).

[9] A. S. Ecker et al. Decorrelated neuronal firing in cortical microcircuits. *Science* **327**(5965), 584–587 (2010).

[10] J. de la Rocha, B. Doiron, E. Shea-Brown, K. Josic & A. Reyes. Correlation between neural spike trains increases with firing rate. *Nature* **448**(7155), 802–806 (2007).

[11] B. Hansen, M. Chelaru & V. Dragoi. Correlated variability in laminar cortical circuits. *Neuron* **76**(3), 590–602 (2012).

[12] A. Kohn & M. A. Smith. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25**(14), 3661 –3673 (2005).

[13] J. Yu & D. Ferster. Membrane potential synchrony in primary visual cortex during sensory stimulation. *Neuron* **68**(6), 1187–1201 (2010).

[14] D. A. Gutnisky & V. Dragoi. Adaptive coding of visual information in neural populations. *Nature* **452**(7184), 220–224 (2008).

[15] M. R. Cohen & J. H. R. Maunsell. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**(12), 1594–1600 (2009).

[16] J. F. Mitchell, K. A. Sundberg & J. H. Reynolds. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**(6), 879–888 (2009).

[17] M. M. Churchland et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat. Neurosci.* **13**(3), 369–378 (2010).

[18] H. Sompolinsky, A. Crisanti & H. J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.* **61**(3), 259–262 (1988).

[19] D. Sussillo & L. Abbott. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**(4), 544–557 (2009).

[20] R. Laje & D. V. Buonomano. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**(7), 925–933 (2013).

[21] W. Softky & C. Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J Neurosci* **13**(1), 334 –350 (1993).

[22] M. N. Shadlen & W. T. Newsome. Noise, neural codes and cortical organization. *Curr. Opin. Neurobiol.* **4**(4), 569–579 (1994).

[23] A. Destexhe, M. Rudolph & D. Pare. The high-conductance state of neocortical neurons in vivo. *Nat. Rev. Neurosci.* **4**(9), 739–751 (2003).

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

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

[26] R. Glauber. Time-dependent statistics of the Ising model. *J. Math. Phys.* **4**(2), 294–307 (1963).

[27] A. Renart et al. The asynchronous state in cortical circuits. *Science* **327**(5965), 587–590 (2010).

[28] Y. Shu, A. Hasenstaub & D. A. McCormick. Turning on and off recurrent balanced cortical activity. *Nature* **423**(6937), 288–293 (2003).

[29] B. Haider, A. Duque, A. R. Hasenstaub & D. A. McCormick. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* **26**(17), 4535 –4545 (2006).

[30] M. Graupner & A. D. Reyes. Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. *J. Neurosci.* **33**(38), 15075–15085 (2013).

[31] M. J. Higley & D. Contreras. Balanced excitation and inhibition determine spike timing during frequency adaptation. *J. Neurosci.* **26**(2), 448–457 (2006).

[32] J. E. Heiss, Y. Katz, E. Ganmor & I. Lampl. Shift in the balance between excitation and inhibition during sensory adaptation of S1 neurons. *J. Neurosci.* **28**(49), 13320–13330 (2008).

[33] J. Zhu, M. Jiang, M. Yang, H. Hou & Y. Shu. Membrane potential-dependent modulation of recurrent inhibition in rat neocortex. *PLoS Biol.* **9**(3), e1001032 (2011).

[34] M. Wehr & A. M. Zador. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**(6965), 442–446 (2003).

[35] A. Litwin-Kumar, A.-M. M. Oswald, N. N. Urban & B. Doiron. Balanced synaptic input shapes the correlation between neural spike trains. *PLoS Comput. Biol.* **7**(12), e1002305 (2011).

[36] A. Litwin-Kumar & B. Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nat. Neurosci.* **15**(11), 1498–1505 (2012).

[37] M. J. Chacron & J. Bastian. Population coding by electrosensory neurons. *J. Neurophysiol.* **99**(4), 1825–1835 (2008).

[38] D. S. Greenberg, A. R. Houweling & J. N. D. Kerr. Population imaging of ongoing neuronal activity in the visual cortex of awake rats. *Nat. Neurosci.* **11**(7), 749–751 (2008).

[39] A. Kohn, A. Zandvakili & M. A. Smith. Correlations and brain states: from electrophysiology to functional imaging. *Curr. Opin. Neurobiol.* **19**(4), 434–438 (2009).

[40] E. Vaadia et al. Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature* **373**(6514), 515–518 (1995).

[41] B. B. Averbeck, P. E. Latham & A. Pouget. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**(5), 358–366 (2006).

[42] K. D. Harris. Neural signatures of cell assembly organization. *Nat. Rev. Neurosci.* **6**(5), 399–407 (2005).

[43] E. Salinas & T. J. Sejnowski. Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* **2**(8), 539–550 (2001).

[44] A. Kumar, S. Rotter & A. Aertsen. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. *Nat. Rev. Neurosci.* **11**(9), 615–627 (2010).

[45] I. Ginzburg & H. Sompolinsky. Theory of correlations in stochastic neural networks. *Phys. Rev. E* **50**(4), 3171 (1994).

[46] R. Moreno-Bote & N. Parga. Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett.* **96**(2), 028101 (2006).

[47] A. Gal et al. Dynamics of excitability over extended timescales in cultured cortical neurons. *J. Neurosci.* **30**(48), 16332–16342 (2010).

[48] E. Shea-Brown, K. Josić, J. de la Rocha & B. Doiron. Correlation and synchrony transfer in integrate-and-fire neurons: Basic properties and consequences for coding. *Phys. Rev. Lett.* **100**(10), 108102 (2008).

[49] S. Marella & G. B. Ermentrout. Class-II neurons display a higher degree of stochastic synchronization than class-I neurons. *Phys. Rev. E* **77**(4), 041918 (2008).

[50] S. Ostojic, N. Brunel & V. Hakim. How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci.* **29**(33), 10234–10253 (2009).

[51] J. S. Anderson, I. Lampl, D. C. Gillespie & D. Ferster. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science* **290**(5498), 1968–1972 (2000).

[52] N. Hô & A. Destexhe. Synaptic background activity enhances the responsiveness of neocortical pyramidal neurons. *J. Neurophysiol.* **84**(3), 1488 –1496 (2000).

[53] B. Doiron, A. Longtin, N. Berman & L. Maler. Subtractive and divisive inhibition: Effect of voltage-dependent inhibitory conductances and noise. *Neural Comput.* **13**(1), 227–248 (2001).

[54] F. S. Chance, L. Abbott & A. D. Reyes. Gain modulation from background synaptic input. *Neuron* **35**(4), 773–782 (2002).

[55] S. J. Mitchell & R. Silver. Shunting inhibition modulates neuronal gain during synaptic excitation. *Neuron* **38**(3), 433–445 (2003).

[56] J. A. Cardin, L. A. Palmer & D. Contreras. Cellular mechanisms underlying stimulus-dependent gain modulation in primary visual cortex neurons in vivo. *Neuron* **59**(1), 150–160 (2008).

[57] C. Ly & B. Doiron. Divisive gain modulation with dynamic stimuli in integrate-and-fire neurons. *PLoS Comput. Biol.* **5**(4), e1000365 (2009).

[58] E. Salinas & T. J. Sejnowski. Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *J. Neurosci.* **20**(16), 6193–6209 (2000).

[59] R. Moreno, J. de la Rocha, A. Renart & N. Parga. Response of spiking neurons to correlated inputs. *Phys. Rev. Lett.* **89**(28), 288101 (2002).

[60] M. Rudolph & A. Destexhe. Correlation detection and resonance in neural systems with distributed noise sources. *Phys. Rev. Lett.* **86**(16), 3662–3665 (2001).

[61] J. Cafaro & F. Rieke. Noise correlations improve response fidelity and stimulus encoding. *Nature* **468**(7326), 964–967 (2010).

[62] S. Giridhar, B. Doiron & N. N. Urban. Timescale-dependent shaping of correlation by olfactory bulb lateral inhibition. *PNAS* (2011).

[63] C. Koch. *Biophysics of computation: information processing in single neurons* (Oxford University Press, 2005).

[64] M. J. E. Richardson. Effects of synaptic conductance on the voltage distribution and firing rate of spiking neurons. *Phys. Rev. E* **69**(5), 051918 (2004).

[65] M. J. E. Richardson & W. Gerstner. Synaptic shot noise and conductance fluctuations affect the membrane voltage with equal significance. *Neural Comput.* **17**(4), 923–947 (2005).

[66] M. Richardson. Spike-train spectra and network response functions for non-linear integrate-and-fire neurons. *Biol. Cybern.* **99**(4), 381–392 (2008).

[67] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, 1996).

[68] D. R. Cox & V. Isham. *Point processes* (CRC Press, 1980).

[69] J. F. A. Poulet & C. C. H. Petersen. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**(7206), 881–885 (2008).

[70] M. Okun & I. Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci.* **11**(5), 535–537 (2008).

[71] P. K. Trong & F. Rieke. Origin of correlated activity between parasol retinal ganglion cells. *Nat. Neurosci.* **11**(11), 1343–1351 (2008).

[72] M. H. Higgs, S. J. Slee & W. J. Spain. Diversity of gain modulation by noise in neocortical neurons: Regulation by the slow afterhyperpolarization conductance. *J. Neurosci.* **26**(34), 8787–8799 (2006).

[73] W. H. Mehaffey, B. Doiron, L. Maler & R. W. Turner. Deterministic multiplicative gain control with active dendrites. *J. Neurosci.* **25**(43), 9968–9977 (2005).

[74] B. Ermentrout. Linearization of F-I curves by adaptation. *Neural Comput.* **10**(7), 1721–1729 (1998).

[75] F. R. Fernandez & J. A. White. Gain control in CA1 pyramidal cells using changes in somatic conductance. *J. Neurosci.* **30**(1), 230–241 (2010).

[76] B. Doiron, Y. Zhao & T. Tzounopoulos. Combined LTP and LTD of modulatory inputs controls neuronal processing of primary sensory inputs. *J. Neurosci.* **31**(29), 10579–10592 (2011).

[77] A. Hasenstaub, R. N. S. Sachdev & D. A. McCormick. State changes rapidly modulate cortical neuronal responsiveness. *J. Neurosci.* **27**(36), 9607–9622 (2007).

[78] A. N. Burkitt. Balanced neurons: analysis of leaky integrate-and-fire neurons with reversal potentials. *Biol. Cybern.* **85**(4), 247–255 (2001).

[79] E. Zohary, M. N. Shadlen & W. T. Newsome. Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature* **370**(6485), 140–143 (1994).

[80] R. E. Kass & V. Ventura. Spike count correlation increases with length of time interval in the presence of trial-to-trial variation. *Neural Comput.* **18**(11), 2583–2591 (2006).

[81] H. Köndgen et al. The dynamical response properties of neocortical neurons to temporally modulated noisy inputs in vitro. *Cereb. Cortex* **18**(9), 2086 –2097 (2008).

[82] P. Dayan & L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (The MIT Press, 2001), 1st edition.

[83] R. D. Vilela & B. Lindner. Comparative study of different integrate-and-fire neurons: Spontaneous activity, dynamical response, and stimulus-induced correlation. *Phys. Rev. E* **80**(3), 031909 (2009).

[84] B. Doiron, B. Lindner, A. Longtin, L. Maler & J. Bastian. Oscillatory activity in electrosensory neurons increases with the spatial correlation of the stochastic input stimulus. *Phys. Rev. Lett.* **93**(4), 048101 (2004).

[85] B. Lindner, B. Doiron & A. Longtin. Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys. Rev. E* **72**(6), 061919 (2005).

[86] R. F. Galan, N. Fourcaud-Trocme, G. B. Ermentrout & N. N. Urban. Correlation-induced synchronization of oscillations in olfactory bulb neurons. *J. Neurosci.* **26**(14), 3646–3655 (2006).

[87] A. A. Prinz, L. F. Abbott & E. Marder. The dynamic clamp comes of age. *Trends Neurosci.* **27**(4), 218–224 (2004).

[88] A. K. Barreiro, E. Shea-Brown & E. L. Thilo. Time scales of spike-train correlation for neural oscillators with common drive. *Phys. Rev. E* **81**(1), 011916 (2010).

[89] A. Abouzeid & B. Ermentrout. Correlation transfer in stochastically driven neural oscillators over long and short time scales. *Phys. Rev. E* **84**, 061914 (2011).

[90] P. Fries, J. H. Reynolds, A. E. Rorie & R. Desimone. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science* **291**(5508), 1560–1563 (2001).

[91] J. H. Reynolds & D. J. Heeger. The normalization model of attention. *Neuron* **61**(2), 168–185 (2009).

[92] M. J. Chacron, B. Doiron, L. Maler, A. Longtin & J. Bastian. Non-classical receptive field mediates switch in a sensory neuron's frequency tuning. *Nature* **423**(6935), 77–81 (2003).

[93] S. J. Slee, M. H. Higgs, A. L. Fairhall & W. J. Spain. Two-dimensional time coding in the auditory brainstem. *J. Neurosci.* **25**(43), 9978–9988 (2005).

[94] T. P. Vogels & L. F. Abbott. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat. Neurosci.* **12**(4), 483–491 (2009).

[95] T. Womelsdorf et al. Modulation of neuronal interactions through neuronal synchronization. *Science* **316**(5831), 1609–1612 (2007).

[96] M. London, A. Roth, L. Beeren, M. Hausser & P. E. Latham. Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature* **466**(7302), 123–127 (2010).

[97] T. P. Vogels & L. F. Abbott. Signal propagation and logic gating in networks of integrate-and-fire neurons. *J. Neurosci.* **25**(46), 10786 –10795 (2005).

[98] A. K. Churchland et al. Variance as a signature of neural computations during decision making. *Neuron* **69**(4), 818–831 (2011).

[99] A. Arieli, A. Sterkin, A. Grinvald & A. Aertsen. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. *Science* **273**(5283), 1868–1871 (1996).

[100] M. Tsodyks, T. Kenet, A. Grinvald & A. Arieli. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**(5446), 1943–1946 (1999).

[101] M. M. Churchland, B. M. Yu, S. I. Ryu, G. Santhanam & K. V. Shenoy. Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* **26**(14), 3697 –3712 (2006).

[102] A. Luczak, P. Barthó & K. D. Harris. Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron* **62**(3), 413–425 (2009).

[103] S. Song, P. J. Sjöström, M. Reigl, S. Nelson & D. B. Chklovskii. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol.* **3**(3), e68 (2005).

[104] R. Perin, T. K. Berger & H. Markram. A synaptic organizing principle for cortical neuronal groups. *PNAS* **108**(13), 5419 –5424 (2011).

[105] Y. Yoshimura, J. L. M. Dantzker & E. M. Callaway. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433**(7028), 868–873 (2005).

[106] H. Ko et al. Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**(7345), 87–91 (2011).

[107] L. Yassin et al. An embedded subnetwork of highly active neurons in the neocortex. *Neuron* **68**(6), 1043–1050 (2010).

[108] D. Amit & N. Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex* **7**(3), 237–252 (1997).

[109] X.-J. Wang. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36**(5), 955–968 (2002).

[110] A. Renart, R. Moreno-Bote, X.-J. Wang & N. Parga. Mean-driven and fluctuation-driven persistent activity in recurrent networks. *Neural Comput.* **19**(1), 1–46 (2006).

[111] Y. Roudi & P. E. Latham. A balanced memory network. *PLoS Comput. Biol.* **3**(9), e141 (2007).

[112] G. Deco & E. Hugues. Neural network mechanisms underlying stimulus driven variability reduction. *PLoS Comput. Biol.* **8**(3), e1002395 (2012).

[113] T. M. J. Fruchterman & E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience* **21**(11), 1129–1164 (1991).

[114] H. Sompolinsky & C. van Vreeswijk. Irregular activity in large networks of neurons. In C. Chow, B. Gutkin, D. Hansel, C. Meunier & J. Dalibard, editors, *Les Houches Lectures LXXX: Methods and models in neurophysics*, 341–402 (Elsevier, 2005).

[115] P. Hänggi, P. Talkner & M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–342 (1990).

[116] R. Albert & A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**(1), 47 (2002).

[117] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez & D.-U. Hwang. Complex networks: Structure and dynamics. *Phys. Rep.* **424**(4-5), 175–308 (2006).

[118] A.-M. M. Oswald, B. Doiron, J. Rinzel & A. D. Reyes. Spatial profile and differential recruitment of GABAB modulate oscillatory activity in auditory cortex. *J. Neurosci.* **29**(33), 10321 –10334 (2009).

[119] E. Fino & R. Yuste. Dense inhibitory connectivity in neocortex. *Neuron* **69**(6), 1188–1203 (2011).

[120] T. Hromádka, M. R. Deweese & A. M. Zador. Sparse representation of sounds in the unanesthetized auditory cortex. *PLoS Biol.* **6**(1), e16 (2008).

[121] N. Brunel & X.-J. Wang. What determines the frequency of fast network oscillations with irregular neural discharges? I. Synaptic dynamics and excitation-inhibition balance. *J. Neurophysiol.* **90**(1), 415 –430 (2003).

[122] M. C. Teich, C. Heneghan, S. B. Lowen, T. Ozaki & E. Kaplan. Fractal character of the neural spike train in the visual system of the cat. *Journal of the Optical Society of America A* **14**(3), 529–546 (1997).

[123] K. Rajan, L. F. Abbott & H. Sompolinsky. Stimulus-dependent suppression of chaos in recurrent neural networks. *Phys. Rev. E* **82**(1), 011903 (2010).

[124] J. F. Mitchell, K. A. Sundberg & J. H. Reynolds. Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* **55**(1), 131–141 (2007).

[125] Y. Yoshimura & E. M. Callaway. Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity. *Nat. Neurosci.* **8**(11), 1552–1559 (2005).

[126] S. B. Hofer et al. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nat. Neurosci.* **14**(8), 1045–1052 (2011).

[127] S. Coombes. Waves, bumps, and patterns in neural field theories. *Biol. Cybern.* **93**, 91–108 (2005).

[128] E. M. Izhikevich, J. A. Gally & G. M. Edelman. Spike-timing dynamics of neuronal groups. *Cereb. Cortex* **14**(8), 933 –944 (2004).

[129] M. Abeles et al. Cortical activity flips among quasi-stationary states. *PNAS* **92**(19), 8616 –8620 (1995).

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

[131] L. J. Gentet, M. Avermann, F. Matyas, J. F. Staiger & C. C. Petersen. Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron* **65**(3), 422–435 (2010).

[132] I. Ferezou et al. Spatiotemporal dynamics of cortical sensorimotor integration in behaving mice. *Neuron* **56**(5), 907–923 (2007).

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

[134] W. J. Ma, J. M. Beck, P. E. Latham & A. Pouget. Bayesian inference with probabilistic population codes. *Nat. Neurosci.* **9**(11), 1432–1438 (2006).

[135] L. Buesing, J. Bill, B. Nessler & W. Maass. Neural dynamics as sampling: A model for stochastic computation in recurrent networks of spiking neurons. *PLoS Comput. Biol.* **7**(11), e1002211 (2011).

[136] P. Berkes, G. Orbán, M. Lengyel & J. Fiser. Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science* **331**(6013), 83–87 (2011).

[137] S. L. Eagleman & V. Dragoi. Image sequence reactivation in awake V4 networks. *PNAS* **109**(47), 19450–19455 (2012).

[138] S. Xu, W. Jiang, M.-m. Poo & Y. Dan. Activity recall in a visual cortical ensemble. *Nat. Neurosci.* **15**(3), 449–455 (2012).

[139] W. Bair, E. Zohary & W. T. Newsome. Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J. Neurosci.* **21**(5), 1676–1697 (2001).

[140] G. Rothschild, I. Nelken & A. Mizrahi. Functional organization and population dynamics in the mouse primary auditory cortex. *Nat. Neurosci.* **13**(3), 353–360 (2010).

[141] L. C. Katz & C. J. Shatz. Synaptic activity and the construction of cortical circuits. *Science* **274**(5290), 1133–1138 (1996).

[142] D. V. Buonomano & M. M. Merzenich. Cortical plasticity: From synapses to maps. *Annu. Rev. Neurosci.* **21**(1), 149–186 (1998).

[143] R. C. Froemke et al. Long-term modification of cortical synapses improves sensory perception. *Nat. Neurosci.* **16**(1), 79–88 (2013).

[144] H. Ko et al. The emergence of functional microcircuits in visual cortex. *Nature* **496**(7443), 96–100 (2013).

[145] G. Mongillo, E. Curti, S. Romani & D. J. Amit. Learning in realistic networks of spiking neurons and spike-driven plastic synapses. *Eur. J. Neurosci.* **21**(11), 3143–3160 (2005).

[146] C. Clopath, L. Büsing, E. Vasilaki & W. Gerstner. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat. Neurosci.* **13**(3), 344–352 (2010).

[147] J. Gjorgjieva, C. Clopath, J. Audet & J.-P. Pfister. A triplet spike-timing–dependent plasticity model generalizes the Bienenstock–Cooper–Munro rule to higher-order spatiotemporal correlations. *PNAS* **108**(48), 19383–19388 (2011).

[148] M. Graupner & N. Brunel. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *PNAS* **109**(10), 3991–3996 (2012).

[149] G. G. Turrigiano & S. B. Nelson. Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.* **5**(2), 97–107 (2004).

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

[151] R. Brette & W. Gerstner. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J. Neurophysiol.* **94**(5), 3637–3642 (2005).

[152] L. Badel et al. Dynamic I-V curves are reliable predictors of naturalistic pyramidal-neuron voltage traces. *J. Neurophysiol.* **99**(2), 656–666 (2008).

[153] A. Renart, P. Song & X.-J. Wang. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**(3), 473–485 (2003).

[154] I. R. Fiete, W. Senn, C. Z. Wang & R. H. Hahnloser. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**(4), 563–576 (2010).

[155] J. Karbowski & G. Ermentrout. Synchrony arising from a balanced synaptic plasticity in a network of heterogeneous neural oscillators. *Phys. Rev. E* **65**(3), 031902 (2002).

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

[157] W. C. Abraham & A. Robins. Memory retention – the synaptic stability versus plasticity dilemma. *Trends Neurosci.* **28**(2), 73–78 (2005).

[158] S. Fusi, P. J. Drew & L. Abbott. Cascade models of synaptically stored memories. *Neuron* **45**(4), 599–611 (2005).

[159] S. E. Boustani, P. Yger, Y. Frégnac & A. Destexhe. Stable learning in stochastic network states. *J. Neurosci.* **32**(1), 194–214 (2012).

[160] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *PNAS* **79**(8), 2554–2558 (1982).

[161] A. D. Lien & M. Scanziani. Tuned thalamic excitation is amplified by visual cortical circuits. *Nat. Neurosci.* **16**(9), 1315–1323 (2013).

[162] Y.-t. Li, L. A. Ibrahim, B.-h. Liu, L. I. Zhang & H. W. Tao. Linear transformation of thalamocortical input by intracortical excitation. *Nat. Neurosci.* **16**(9), 1324–1330 (2013).

[163] L.-y. Li, Y.-t. Li, M. Zhou, H. W. Tao & L. I. Zhang. Intracortical multiplication of thalamocortical signals in mouse auditory cortex. *Nat. Neurosci.* **16**(9), 1179–1181 (2013).

[164] X.-L. Qi & C. Constantinidis. Variability of prefrontal neuronal discharges before and after training in a working memory task. *PLoS ONE* **7**(7), e41053 (2012).

[165] T. Keck et al. Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. *Neuron* **80**(2), 327–334 (2013).

[166] S. Royer & D. Paré. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* **422**(6931), 518–522 (2003).

[167] H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang & M. Scanziani. A neural circuit for spatial summation in visual cortex. *Nature* **490**(7419), 226–231 (2012).

[168] N. R. Wilson, C. A. Runyan, F. L. Wang & M. Sur. Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature* **488**(7411), 343–348 (2012).

[169] R. P. N. Rao & D. H. Ballard. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**(1), 79–87 (1999).

[170] J.-P. Pfister & W. Gerstner. Triplets of spikes in a model of spike timing-dependent plasticity. *J. Neurosci.* **26**(38), 9673–9682 (2006).

[171] E. L. Bienenstock, L. N. Cooper & P. W. Munro. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**(1), 32–48 (1982).

[172] P. C. Bressloff. Metastable states and quasicycles in a stochastic Wilson-Cowan model of neuronal population dynamics. *Phys. Rev. E* **82**(5), 051903 (2010).

[173] G. Ermentrout. Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys.* **61**, 353–430 (1998).

[174] D. Hansel & C. van Vreeswijk. The mechanism of orientation selectivity in primary visual cortex without a functional map. *J. Neurosci.* **32**(12), 4049–4064 (2012).

[175] W. Maass, T. Natschläger & H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Comput.* **14**(11), 2531–2560 (2002).

[176] M. Rigotti et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**(7451), 585–590 (2013).

[177] A. Litwin-Kumar, M. J. Chacron & B. Doiron. The spatial structure of stimuli shapes the timescale of correlations in population spiking activity. *PLoS Comput. Biol.* **8**(9), e1002667 (2012).

[178] J. W. Middleton, C. Omar, B. Doiron & D. J. Simons. Neural correlation is stimulus modulated by feedforward inhibitory circuitry. *J. Neurosci.* **32**(2), 506–518 (2012).

[179] R. Rosenbaum & B. Doiron. Balanced networks in space. arXiv e-print 1308.6014 (2013).

[180] E. B. Anderson, J. F. Mitchell & J. H. Reynolds. Attention-dependent reductions in burstiness and action-potential height in macaque area V4. *Nat. Neurosci.* **16**(8), 1125–1131 (2013).

[181] S. Klampfl & W. Maass. Emergence of dynamic memory traces in cortical microcircuit models through STDP. *J. Neurosci.* **33**(28), 11515–11529 (2013).

[182] S.-H. Lee et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**(7411), 379–383 (2012).

[183] B. Atallah, W. Bruns, M. Carandini & M. Scanziani. Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron* **73**(1), 159–170 (2012).

[184] C. K. Pfeffer, M. Xue, M. He, Z. J. Huang & M. Scanziani. Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* **16**(8), 1068–1076 (2013).

[185] H. Markram, J. Lübke, M. Frotscher & B. Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science* **275**(5297), 213–215 (1997).

[186] G.-q. Bi & M.-m. Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**(24), 10464–10472 (1998).

[187] S. Song, K. D. Miller & L. F. Abbott. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. Neurosci.* **3**(9), 919–926 (2000).

[188] A. Artola, S. Bröcher & W. Singer. Different voltage-dependent thresholds for inducing long-term depression and long-term potentiation in slices of rat visual cortex. *Nature* **347**(6288), 69–72 (1990).

[189] T. Nevian & B. Sakmann. Spine Ca2+ signaling in spike-timing-dependent plasticity. *J. Neurosci.* **26**(43), 11001–11013 (2006).

[190] U. R. Karmarkar & D. V. Buonomano. A model of spike-timing dependent plasticity: One or two coincidence detectors? *J. Neurophysiol.* **88**(1), 507–513 (2002).

[191] J. E. Rubin, R. C. Gerkin, G.-Q. Bi & C. C. Chow. Calcium time course as a signal for spike-Timing–Dependent plasticity. *Journal of Neurophysiology* **93**(5), 2600–2613 (2005).