THE UNIVERSITY OF CHICAGO

APPLICATIONS OF STATISTICAL MECHANICS TO BIOPHYSICAL DYNAMICS:

FROM NEURONAL NETWORKS TO GENE CONTROL

A THESIS SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY

CLAYTON W. SEITZ

CHICAGO, ILLINOIS

SPRING 20XX

TABLE OF CONTENTS

ABSTRACT

CHAPTER 1

STATISTICAL MECHANICS OF GENE REGULATION

1.1 Introduction

1.2 Fine chromatin structure and the Rouse-like polymer model

1.3 A phase separation model for transcriptional control

1.3.1 The chemical master equation and finite state projection algorithm

1.3.2 The ergodic hypothesis

The central assumption underlying a Markov process, is the memoryless property

$$P(X_t|X_{t-1}, X_{t-2}, ..., X_{t-N}) = P(X_t|X_{t-1})$$

A single Markov chain is the set of states $\boldsymbol{X} = \{X_1, X_2, ..., X_N\}$. Such a set can be generated provided that $P(X_t|X_{t-1})$ is known. To capture $P(X_t|X_{t-1})$ for all possible pairs X_t and X_{t-1}, we define a square transition matrix $T \in \mathcal{R}^{N \times N}$ where $N = |\Omega|$. As such, the elements of T represent the probability of a transition from a state ω_j to ω_i in a unit time

$$T_{ij} = \Pr\left(X_t = \omega_i, |\ X_{t-1} = \omega_j\right)$$

Under these definitions, the row T_i represents the present time, and is a conditional probability distribution $P(\omega|X_{t-1} = \omega_j)$ which requires that

$$\sum_j T_{ij} = \sum_j P(X_t = \omega_j|X_{t-1} = \omega_i) = 1$$

The matrix T is not necessarily symmetric $T_{ij} \neq T_{ji}$. One should note that the columns

T_j *do not* define a probability distribution $P(X_t = \omega_i | X_{t-1} = \omega_j)$ and therefore do not necessarily sum to unity. The probability $P(X_t = \omega_i | X_{t-1} = \omega_j)$ has no meaning in this context, since we have defined the rows to represent a probability of the future given the present. We simply sample $X_t \sim P(X_t = \omega_j | X_{t-1} = \omega_i)$, assign $i = j$, and repeat. It directly follows from the fundamental rules of probability, the first order dynamics for a **particular** state ω_i: $P(\omega_i, t)$ is given by

$$P(\omega_i, t + dt) = P(\omega_i, t) + \mathcal{J}_i dt \tag{1.1}$$

The net probability current $\mathcal{J}_i$ must be

$$\mathcal{J}_i = \sum_i T_{ij} P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t)$$

The first is a sum on a column and the second a sum on a row. This can be simplified further by noticing that the normalization condition implies

$$T_{ij} = 1 - \sum_j T_{ij}(1 - \delta_{ij})$$
$$= 1 - \sum_j T_{ij} + \sum_j T_{ij}\delta_{ij}$$

$$\mathcal{J}_i = \sum_i T_{ij} P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t)$$

$$= \sum_i \left(1 - \sum_j T_{ij} + \sum_j T_{ij} \delta_{ij} \right) P(\omega_j, t) - \sum_j T_{ij} P(\omega_i, t)$$

$$= |\Omega| - |\Omega| + \sum_i \sum_j T_{ij} P(\omega_j, t) \delta_{ij} - \sum_j T_{ij} P(\omega_i, t)$$

$$= \sum_j T_{ji} P(\omega_j, t) - T_{ij} P(\omega_i, t)$$

Notice that the Kronecker delta effectively just swaps the index. Taking the limit of (1.1), we arrive at the **master equation**

$$\frac{\partial P(\omega_i)}{\partial t} = \sum_j T_{ji} P(\omega_j, t) - T_{ij} P(\omega_i, t)$$

It is common to then define an operator $\boldsymbol{W}$ s.t. $W_{ij} = T_{ij}$ and $W_{ii} = -\sum_j T_{ij}$

$$\frac{dP(\omega_i)}{dt} = \sum_j W_{ij} P(\omega_j) \rightarrow \frac{dP(\boldsymbol{\omega})}{dt} = \mathcal{J}(\boldsymbol{\omega}) = \mathbf{W} P(\boldsymbol{\omega})$$

This operator form has a solution in terms of a matrix exponential

$$P(\boldsymbol{\omega}, t) = \exp(\mathcal{J}(\boldsymbol{\omega}))$$

This matrix exponential is intractable for large $|\Omega|$. However, in the Finite State Projection algorithm, it is possible to truncate the state space $\Omega \rightarrow \tilde{\Omega}$ and obtain good estimates $\tilde{P}(\boldsymbol{\omega}, t)$ with some certificate of accuracy.

1.4   Gaussian PSF approximation for single molecule microscopy

Born and Wolf's approximation, Lateral and axial point spread functions

1.4.1   Fisher information theory for parameter estimation in single molecule microscopy

1.5   Bounding parameter uncertainty in single molecule localization

1.5.1   Point spread functions in widefield microscopy

Most detectors used for imaging have many elements (pixels) so that we can record an image projected onto the detector by a system of lenses. In fluorescence imaging, this is usually a relay consisting of an objective lens and a tube lens to focus the image onto the camera. Due to diffraction, any point emitter, such as a single fluorescent molecule, will be registered as a diffraction limited spot. The profile of that spot is often described as a Gaussian point spread function (Richardson and Wolf)

$$\text{PSF}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}\right) \tag{1.2}$$

where (x_0, y_0) defines the coordinates of a single fluorescent molecule on a detector $\mathcal{D}$.

1.5.2   Maximum a posteriori estimation of PSF parameters

Maximum a posteriori estimation of parameters

Figure 1.1: Cartoon of a excitatory-inhibitory neural network receiving a time-dependent feedforward current $F(t)$, feedback current $R(t)$, and synaptic noise $\xi(t)$. Excitatory neurons are shown in red and inhibitory neurons shown in blue. Synaptic current is drawn to be instantaneous and syanptic weights J_{ij} are not functions of time. Feedforward inputs could be weighted Poisson processes where F defines the rates, for example.

$$\theta^*_{\text{MAP}} = \underset{\theta}{\operatorname{argmax}} \, P(\theta|\mathcal{X})$$

$$= \underset{\theta}{\operatorname{argmax}} \, \frac{\mathcal{L}_\theta \pi(\theta)}{\int_\theta \mathcal{L}_\theta \pi(\theta)}$$

where $\mathcal{L}_\theta = \mathcal{L}(\theta|\mathcal{X})$ is the likelihood function and $\pi(\theta)$ a prior on the parameters. Furthermore, we will assume that $\pi(\theta)$ is a uniform distribution. An image captured by a camera can be loosely thought of as histogram of photon arrivals and a discretized form of the density $\text{PSF}(x, y)$ over an integration time τ. If $\text{PSF}(x, y)$ can be approximated as constant in time ($\tau << 1$), the value at a pixel approaches an integral of this density over the pixel:
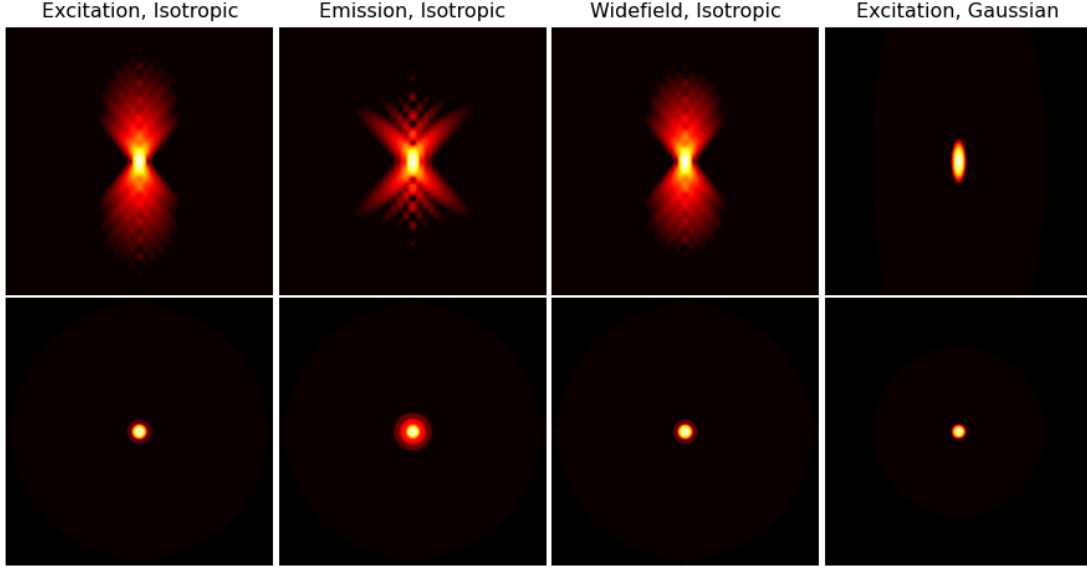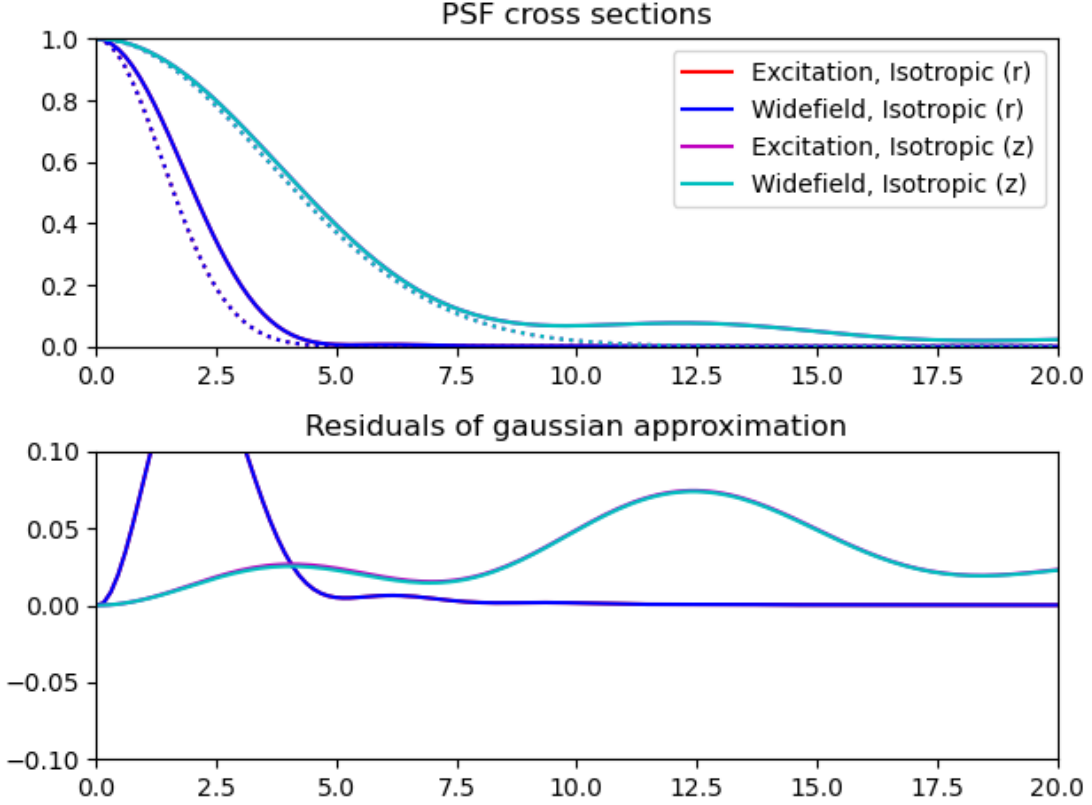
Figure 1.2: Cartoon of a excitatory-inhibitory neural network receiving a time-dependent feedforward current $F(t)$, feedback current $R(t)$, and synaptic noise $\xi(t)$. Excitatory neurons are shown in red and inhibitory neurons shown in blue. Synaptic current is drawn to be instantaneous and syanptic weights J_{ij} are not functions of time. Feedforward inputs could be weighted Poisson processes where F defines the rates, for example.

$$\lambda_k = \gamma \cdot \int_{\mathcal{D}_k} q(x,y)dxdy \tag{1.3}$$

Put another way, the variable λ_k at a pixel k defines the probability of observing a photon per unit time and therefore we can define the number of photons arriving at each pixel as a random variable

$$P(S_k) = \text{Poisson}(\lambda_k)$$

It would be convenient to obtain an analytical expression for λ_k given the parameters

of the point spread function $\theta = (x_0, y_0, \sigma)$. This can be done by making use of the error function

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dx$$

where $t = \frac{x}{\sqrt{2}\sigma}$. We just need to rescale this so it is appropriate for a standard normal distribution. We need something like

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_0^x \exp(-t^2) dx$$

so we rescale the error function by $\beta = 1/2\sqrt{2}$. Of course, this can also be used to construct the integral over an arbitrary interval $[a, b]$

$$\text{erf}(b) - \text{erf}(a) = \frac{2\beta}{\sqrt{\pi}} \int_{\frac{a}{\sqrt{2}\sigma}}^{\frac{b}{\sqrt{2}\sigma}} \exp(-t^2) dt$$

We use this result to compute the Poisson rate λ_k at pixel k. Now, the Poisso rate is in general given by

$$\lambda_k = \frac{1}{2\pi} \left(\int_{x_k}^{x_{k+1}} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2} \right) dx \right) \left(\int_{y_k}^{y_{k+1}} \exp\left(-\frac{(y - y_0)^2}{2\sigma^2} \right) dy \right)$$

Let $t = (x - x_0)/\sqrt{2}\sigma$ and $z = (y - y_0)/\sqrt{2}\sigma$ and write constant in terms of β

$$\lambda_k = \frac{4\beta^2}{\pi} \left(\int_{t_k}^{t_{k+1}} \exp\left(-t^2 \right) dx \right) \left(\int_{z_k}^{z_{k+1}} \exp\left(-z^2 \right) dy \right)$$
$$= \beta^2 \left[\text{erf}(t_{k+1}) - \text{erf}(t_k) \right] \left[(\text{erf}(z_{k+1}) - \text{erf}(z_k)) \right]$$

7

where for example $t_k = (x_k - x_0)/\sqrt{2}\sigma$

Detectors often suffer from dark noise (thermal noise) and there may also be a background signal. Considering the former first, we define another r.v. W_k which represents Gaussian dark noise associated with pixel k. Unless otherwise specified we will always assume that $W_k \sim \mathcal{N}(m_k, \sigma_{w,k}^2)$. We represent our corrupted signal as another random variable H_k. As a side note, we define S_k to have units of photons [p], while H_k has units of photoelectrons $[e^-]$. The conversion factor between the two is the gain of the detector element g_k $[e^- \mathrm{p}^{-1}]$.

Dropping the pixel subscripts for the moment, the Fourier transform of a normal distribution $\mathcal{N}(\mu, \sigma_w^2)$, or its characteristic function reads

$$\tilde{P}(W) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_w} e^{-\frac{(x-\mu)^2}{2\sigma_w^2}} e^{-i\omega x} dx$$

$$= \frac{e^{-\frac{\sigma_w^2 \omega^2}{2} + i\mu\omega}}{\sqrt{2\pi}}$$

On the other hand for $\lambda \gg 1$ we can approximate the Poisson distribution as a normal distribution $\mathcal{N}(\lambda, \sqrt{\lambda})$ on countable support. We can then then write the likelihood function $\mathcal{L}_{\theta,k}$ as the inverse Fourier transform of the product of these two characteristic functions

$$\mathcal{L}_{\theta,k} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{\sigma_k^2 \omega^2}{2} + i\mu_k\omega} e^{-\frac{\lambda_k \omega^2}{2} + i\lambda_k\omega} e^{-i\omega x} d\omega$$

$$= \frac{1}{\sqrt{2\pi(\lambda_k + \mu_k^2)}} e^{-\frac{(-x + \mu_k + \lambda_k)^2}{2(\lambda_k + \mu_k^2)}}$$

We can now return to the original optimization problem

$$\theta^*_{\text{MAP}} = \operatorname*{argmax}_{\theta} \ \pi(\theta)\mathcal{L}_\theta$$

$$= \operatorname*{argmax}_{\theta} \ \prod_k \mathcal{L}_{\theta,k}$$

$$= \operatorname*{argmax}_{\theta} \ \sum_k \log \mathcal{L}_{\theta,k}$$

$$\sum_k \log \mathcal{L}_{\theta,k} = \sum_k -\frac{(-x + \mu_k + \lambda_k)^2}{2(\lambda_k + \mu_k^2)} - \log\left(\sqrt{2\pi(\lambda_k + \mu_k^2)}\right)$$

Maximizing $\mathcal{L}_\theta$ is therefore equivalent to maximizing $\mathcal{L}_{\theta,k}$ at every pixel. Taking the derivative of the first term with respect to a parameter θ

$$\frac{\partial}{\partial\theta}\left(\frac{(-x + \mu_k + \lambda_k)^2}{2(\lambda_k + \mu_k^2)}\right)$$

$$= \left(\frac{4(\lambda_k + \mu_k^2)(-x + \mu_k + \lambda_k)\frac{\partial\lambda_k}{\partial\theta} - 2\frac{\partial\lambda_k}{\partial\theta}(-x + \mu_k + \lambda_k)^2}{4(\lambda_k + \mu_k^2)^2}\right)$$

$$\frac{\partial}{\partial\theta}\log\left(\sqrt{2\pi(\lambda_k + \mu_k^2)}\right) = \frac{1}{2(\lambda_k + \mu_k^2)}\frac{\partial\lambda_k}{\partial\theta}$$

$$\frac{\partial\lambda_k}{\partial x_0} = \beta^2\left(\int_{x_k}^{x_{k+1}} \frac{\partial}{\partial x_0} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right) dx\right)[(\operatorname{erf}(z_{k+1}) - \operatorname{erf}(z_k))]$$

$$= \frac{\beta^2}{\sigma^2}\left(\int_{x_k}^{x_{k+1}} (x - x_0)\exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right) dx\right)[(\operatorname{erf}(z_{k+1}) - \operatorname{erf}(z_k))]$$

$$\int_{x_k}^{x_{k+1}} x \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) dx - x_0 \int_{x_k}^{x_{k+1}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) dx$$

$$\int_{x_k}^{x_{k+1}} x \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) dx = \sigma\sqrt{2\pi} \left. \right|_{x_k}^{x_{k+1}} x - \int_{x_k}^{x_{k+1}} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right) dx$$

$$= \sqrt{2\pi}\sigma\Delta - \mathrm{erf}(t_{k+1}) - \mathrm{erf}(t_k)$$

1.5.3   Fisher information and the Cramer-Rao bound

Suppose we are given samples from a normal distribution with unknown parameters. We then decide to build a model normal distribution. Perhaps we define a likelihood function over the $\mu - \sigma$ plane based on a set of samples, does the likelihood of the data vary? If the likelihood surface is flat, all parameter sets would be equally likely and the data does not appear to carry much information about the parameters. If the surface has a number of bumps or inflection points, then we expect our data does carry information about the parameters. This "bumpiness" of the likelihood surface is measured by computing the variance the second derivative of the likelihood over the parameter space

The log-likelihood of an image under parameters $\theta = (x_0, y_0, \sigma^2, \sigma_w^2, g_k)$, assuming each pixel is an independent but not identical random variable, reads

$$\ell(H|\theta) = \log \prod_k^{M^2} P(H_k|\theta) = \sum_{k=1}^{M^2} \ell(H_k|\theta)$$

When there are many parameters, the Fisher Information (second moment of the score)

is a matrix

$$I_{ij}(\theta) = \mathbb{E}_{\theta} \left[\frac{\partial}{\partial \theta_i} \left(\ell(H|\theta) \right) \frac{\partial}{\partial \theta_j} \left(\ell(H|\theta) \right) \right]$$

The only elements of the matrix of interest are on the diagonal, which allows us to write

$$I(\theta_i) = \mathbb{E}_{\theta} \left[\sum_k \frac{\partial^2}{\partial \theta_i^2} \ell(H_k|\theta) \right]$$

At this point, we need to evaluate the second derivative w.r.t each of the parameters in $\theta = (x_0, y_0, \sigma^2)$. We have shown that the model for the number of photoelectrons at a pixel is

1.5.4   sCMOS camera-specific single-molecule localization algorithms

1.5.5   Theoretical limits: modeling detection errors in RNA Flow

CHAPTER 2

STATISTICAL MECHANICS OF INTEGRATE AND FIRE NEURONAL NETWORKS

2.1 Introduction

Complex systems are ubiquitous in nature yet our scientific efforts have thus far only begun to gain traction on their governing principles. Complex systems often exihibit behaviors that simply cannot be explained by any of the components in isolation but rather arise from their interactions, bringing emergent phenomena into the focus of modern science. The brain exemplifies this complexity, thought to consist of billions of noisy neurons, yet the dynamics of these neurons simulataneously maintain an order evident in the stability of our sensory percepts. How functional brain states emerge from the interactions of dozens, perhaps hundreds, of brain regions, each containing its own complex subnetworks, remains a complicated problem. It is well accepted that information processing in the brain like sensory, motor, and cognitive functions are carried out by correlated spiking of individual neurons, formally referred to as a *population code*. However, is a great challenge to understand how this correlation structure relates to the statistics of sensory stimuli, how it forms, and how it is maintained.

Early models of neural networks were often deterministic; however, an increasing body of experimental data suggests that neurons, synapses, and systems of neurons are inherently stochastic (Buesing, 2011). The enormous number of degrees of freedom of just a single cell due to millions of unreliable ion channels and synaptic machinery demands a stochastic model for network dynamics. Indeed, it has been argued that the variability in neural responses to stimuli and the irregularity of spike-timing owe their origins, in part, to noise in the synaptic integration process (Azouz, 1999). The mechanisms by which networks can exhibit stable dynamics, correlated states, and stable cell assembly formation amidst this stochasticity are

important issues.

Some of the most mature models of network dynamics describe the averages of dynamical quantities or involve the solution of stochastic differential equations. The former technique is realized by a mean-field theory of network dynamics which replaces synaptic currents by their average values - a valid approximation in the limit of very large networks (Vreeswijk 1999). The method was originally developed in simple excitatory-inhibitory random network models but has been more recently extended to networks where synaptic probabilities are a function of space (Rosenbaum 2017). In the mean-field framework, one can prove the existence of a *balanced state* in which excitation and inhibition cancel eachother. The existence of such a state is thought to contribute to the irregularity of spike-timing observed *in-vivo* and neuronal information processing at large. Interestingly, in the balanced state, the spatial structure of spike-train correlations was shown to have a distinct spatial profile, similar to reported measurements in macaque V1 (Rosenbaum 2017). Both theory and experiment show that nearby cells tend to be positively correlated, moderately distant cells negatively correlated, and distant cells show weak correlation. At any rate, proving the existence of a balanced state often places rigid requirements on the network topology, which are not reminiscent of the synaptic weight distribution in real neural networks.

Another class of methods, sometimes referred to as population density methods, predict the membrane potential distribution of an ensemble via a diffusion approximation i.e. by solving the Fokker-Planck equation. The Fokker-Planck formalism has proven quite useful and naturally lends itself to the description of ensembles of neurons. However, its analytical solution is known only for relatively simple networks and simple integrate and fire neuron models. More complex single neuron models, such as exponential integrate and fire (EIF) models, require special numerical treatment. Also, the Fokker-Planck equation is generally used as a low-dimensional description of the network and imposes strong assumptions on its topology such as sparsity (Brunel 2000). If used in isolation, it is limited to special cases

and precludes the analysis of more complex dynamical regimes. However, the Fokker-Planck formalism can be used as a first approximation to network dynamics, and refinements of that prediction can be subsequently added. For example, in the linear response framework, a first approximation of network dynamics can be found by uncoupling the neurons and solving the Fokker-Planck equation. Corrections to that first approximation are then found by using uncoupled firing rates to find spike-train power spectra, which are then propagated through the network according to its topology. In short, this technique can be used to make predictions for more complex network topologies than those often used in a mean-field framework and can be extended to make predictions of the network topology that results from the underlying spike-train statistics (Ocker 2015; 2017).

In his famous neurophysiological postulate, Donald Hebb first proposed a cellular mechanism for the self-organization of networks of neurons. Hebb suggested that repeated stimulation of specific receptors of sensory signals would lead slowly to the formation of a *cell assembly* and these structural changes would constitute a representation or imprint of an internally or externally generated sensation e.g., an image or idea (Hebb 1949). This process, which is now known as a *heterosynaptic* plasticity mechanism, is argued to be driven by the frequency and temporal order of action potentials. On the other hand, experimental evidence and the results of computational studies suggest that a complex orchestration of different types of plasticity mechanisms are necessary for stable formation of cell assemblies (Kumar 2014; Zenke 2015). In this thesis, it will be argued that the spatial correlation structure predicted by mean-field theory for balanced spatially-extended networks could result in spatially-dependent plasticity. The non-monotonicity of the spatial-dependence of spike-train correlations could have important implications for synaptic competition and stable assembly formation. Deviations from the balanced state predicted by mean-field theory could then be addressed using linear response techniques, which have been recently generalized to accomodate for slow changes in network topology (Ocker 2015).

14

2.2 Literature Review and Theoretical Foundations

Recent experimental advances permit the recording of spike trains of many neurons simultaneously *in-vivo*. These experiments have revealed that significant spatial and temporal correlations exist in the timing of spikes by individual neurons, bringing the role of correlations in neural computation into question. Often, such studies are necessarily contextual - correlations must be interpeted in the context of known stimulus drive, learning or experience, or changes in behavioral context (Cohen 2011). Interpretation of such correlations has been a fruitful exercise; however, it is also important to understand the properties of neuronal networks that give rise to them such as topological parameters and synaptic plasticity. For pairs of neurons, correlations are often quantified according to the cross-correlation function of their spike trains (Ostojic 2009). The profile of this cross-correlation function is dependent on several factors including direct synaptic wiring (Csicsvari 1998) or common and potentially correlated inputs. Although cross-correlation analysis of experimental recordings can provide substantial information regarding circuit architecture, relating the correlations of spike trains to the underlying circuit architecture remains a long-standing problem in computational neuroscience. To approach this problem, simulations of networks of model neurons are often used (see Fig (1.1)). In these simulations, network parameters are tuned to examine the resulting correlation structure of spike trains with the hope of inverting the model for comparison with *in-vivo* recordings.

In addition to non-trivial correlations in their spike timing, neurons in cortical networks tend to exhibit highly irregular time intervals between spikes. The stochasticity in spike timing has can be attributed to cellular and sensory noise (Faisal 2008), or alternatively to network-scale mechanisms such as excitatory-inhibitory balance (Vreeswijk 1996, 1998). Amidst this irregularity, our understanding of spike correlations are further complicated by the nonlinear dynamics of recurrent network models (Baker 2019; Tetzlaff 2008, 2012; Doiron 2016; Ocker 2017). It is common to tame this complexity by making suitable combinations
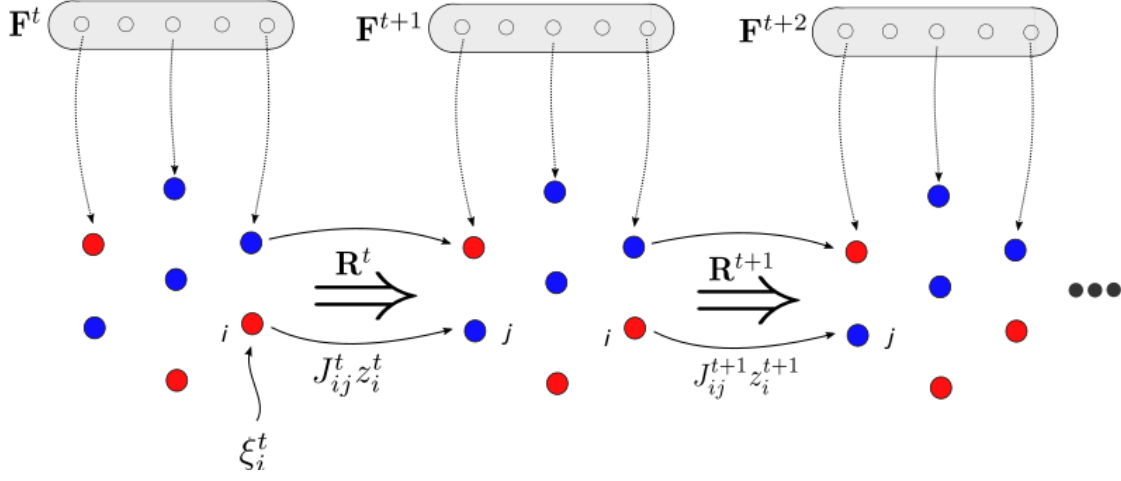
Figure 2.1: Cartoon of a excitatory-inhibitory neural network receiving a time-dependent feedforward current $F(t)$, feedback current $R(t)$, and synaptic noise $\xi(t)$. Excitatory neurons are shown in red and inhibitory neurons shown in blue. Synaptic current is drawn to be instantaneous and syanptic weights J_{ij} are not functions of time. Feedforward inputs could be weighted Poisson processes where F defines the rates, for example.

of assumptions regarding the statistics of network stimuli, the mechanism of neuron-neuron interactions, or the statistics of synaptic connectivity. For example, in the mean-field approximation strong assumptions are often made regarding the synaptic connectivity in order to estimate the mean value of synaptic currents. As a rule, the postsynaptic current to a neuron embedded in a spiking network model consists of two components: feedforward $F(t)$ and feedback $R(t)$:

$$I(t) = F(t) + R(t) + \xi(t) \tag{2.1}$$

Notice the additional noise term $\xi(t)$, which serves to capture channel noise or noise in the synaptic integration process. This is frequently taken to be delta-correlated Gaussian noise. A variety of models for the integration of this current into a stateful membrane potential exist. Here, we are particularly interested in non-linear integrate and fire models which are

described by the Langevin equation

$$\tau \dot{V}(t) = g_\ell(E - V) + g_\ell \cdot \psi(V) + I(t) \tag{2.2}$$

where g_ℓ is a leakage conductance and E its reversal potential. An important case of (1.2) is the exponential integrate and fire (EIF) model where $\psi(V) = \Delta_T \exp\left(\frac{V - V_T}{\Delta_T}\right)$ is a spike generating current. This model is an extension of the standard linear integrate and fire models which more accurately captures the initial dynamics of spike-generating sodium channels (Richardson 2007). To produce spiking behavior, the membrane potential is passed through a thresholding function which produces spike trains. Spike trains are represented as sums over delta functions

$$z(t) = \sum_i \phi(t) * \delta(t - t_i) \tag{2.3}$$

where the variable $z(t)$ is sometimes called the *observable state* of a neuron and is computed from the voltage according to $z(t) = H(V(t) - V_{\text{th}})$ where H is the Heaviside step function and V_{th} the firing threshold. The time-dependent firing rate of a neuron is given by averaging the above spike train across multiple realizations $r(t) = \langle z(t) \rangle$. In addition, to avoid instantaneous synapses, $z(t)$ can be convolved with the synaptic kernel $\phi(t)$, which is often taken to be a decaying exponential, as is done in later sections. Immediately after firing an action potential, the membrane potential is reset to a resting value V_{re} and it is held there for a refractory period τ_{ref}. Of course, in a network, (1.3) is directly related to the feedback current in (1.1) by

$$R_i(t) = \sum_j J_{ij} \cdot \phi(t) * \delta(t - t_i) \tag{2.4}$$

Ultimately, our goal is to understand the origins of correlations or lack thereof in the variable $z(t)$ across a population of neurons. A natural starting point is to consider the case

where spike-trains are uncorrelated, which can occur in the presence of balanced excitation and inhibition.

2.2.1 Linear response: the forwards problem

Another method used to predict the dynamics of excitatory-inhibitory networks utilizes the so-called *diffusion approximation*. If the presynaptic partners of a neuron are each described by an inhomogeneous point process, the total synaptic current $I(t)$ can have complex statistics depending on presynaptic firing rates and their correlations. However, in some special cases, such as when presynaptic firing rates are large, synaptic currents are safely assumed to be Gaussian (Richardson 2007). To begin to examine this, we combine equations (1.1), (1.2), and (1.4) for a single neuron i to give

$$\tau \dot{V}_i(t) = g_\ell(E_i - V_i) + g_\ell \cdot \psi(V_i) + F_i(t) + \xi(t) + \sum_j J_{ij} \cdot \big(\phi(t) * z_j(t)\big) \qquad (2.5)$$

where it is assumed that $\xi(t)$ is a delta correlated white noise i.e. $\langle \xi_i(t)\xi_j(t) \rangle = \delta(i - j)$ and $\langle \xi_i(t)\xi_i(t') \rangle = \delta(t - t')$. When expressed for all neurons i simultaneously, the above becomes a high-dimensional stochastic differential equation, which in general is difficult to work with. Rather than solving (1.12) directly, in the linear response approximation a realization of the spike train $z_i(t)$ is approximated as a sum of two terms which represent the spike train in the absence of feedforward and feedback currents and a linear filter of the feedback currents

$$z(t) \approx z_0(t) + (A * R)(t) \qquad (2.6)$$

If averaged over many realizations, the above instead becomes an expression for the firing rate. The equation above is the central claim made by a linear reponse theory of network

dynamics. For simplicity, it is natural to start with the case where $F_i(t)$ is itself a Gaussian white noise independent of the intrinsic noise. Therefore, the intrinsic noise $\xi(t)$ and the feedforward stimulus can be absorbed into a single noise term. Then, $z_0(t)$ is a realization of a spike train in the absence of feedback when its stimulation is a sum of feedforward and intrinsic noise. In the frequency domain (1.13) for a particular neuron reads (Linder 2005; Trousdale 2012)

$$z_i(\omega) = z_i^0(\omega) + A(\omega) \left(\sum_j J_{ij}(\omega) z_j(\omega) \right) \qquad (2.7)$$

The matrix $J(\omega)$ is the frequency domain representation of the synaptic connectivity matrix after convolution with the synaptic kernel and $A(\omega)$ is the transfer function or linear response function in the frequency domain. This equation can then be rewritten in matrix form to give the frequenty domain representation of the spiking activity across all neurons in the population simultaneously

$$z(\omega) = z_0(\omega) + A(\omega) J(\omega) z(\omega) \qquad (2.8)$$

The result above can then be solved for $z(\omega)$ to give

$$z(\omega) = [I - A(\omega) J(\omega)]^{-1} z_0(\omega) \qquad (2.9)$$

where I is the identity matrix. Making the substitution $\Delta(\omega) = [I - A(\omega) J(\omega)]^{-1}$, the expression above can then be used to obtain the cross-spectrum (and therefore the cross-correlations) of two spike trains $z_i(\omega)$ and $z_j(\omega)$ for all possible pairs.

$$\langle z(\omega)z^*(\omega)\rangle = \langle \Delta(\omega)z_0(\omega)z_0^*(\omega)\Delta^*(\omega)\rangle \tag{2.10}$$

$$= \Delta(\omega)C^0(\omega)\Delta(\omega)^* \tag{2.11}$$

The above result is the matrix of cross spectra across the population $C(\omega)$ in the absence of correlated inputs. The matrix $C^0(\omega)$ is a diagonal matrix with $C_{ii}^0(\omega)$ being the spectrum of the spike train of a neuron i when the neurons are decoupled. The interpretation of equation above is that we can first find a matrix of cross-spectra when neurons are uncoupled $C^0(\omega)$ and then solve (1.18) to estimate the full cross-spectra in the network. This proves very useful for examaple when analyzing the effects of the topology J on the matrix $C(\omega)$ (Trousdale 2012; Ocker 2015). In any case, the matrix $C^0(\omega)$ and the linear response function $A(\omega)$ are still unknown, and must be found. The assumption that feedforward inputs and intrinsic noise are white permits the use of a diffusion approximation to find the individual rates and therefore the spectra $C_{ii}^0(\omega)$. Under conditions in which a diffusion approximation applies, the Kramers-Moyal expansion can be reduced to the Fokker-Planck equation. The Fokker-Planck equation for the stochastic integrate and fire model given in (1.12) in the absence of feedback is

$$\frac{\partial P}{\partial t} = \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial V^2} + \frac{\partial}{\partial V}\left(\frac{V - E - \psi}{\tau}P\right) \tag{2.12}$$

where σ^2 is the variance of the Gaussian input current. We can solve the Fokker-Planck equation above to estimate the firing rates of neurons when they are uncoupled and therefore estimate $C^0(\omega)$. However, the solution for non-linear integrate and fire models frequently requires numerical integration of (1.19) to find the stationary density over the membrane potential and to estimate steady state firing rates (Richardson 2007). Similarly, the frequency

response $A(\omega)$ can be found by using the solution to the Fokker-Planck equation under a weak sinusoidal perturbation to the synaptic current (Richardson 2007; 2009). A more detailed description of these procedures is given in the Methods. It should be noted that a diffusion approximation is not limited to the case where neurons are uncoupled. It can also be used when synaptic connectivity is sufficiently sparse, and the probability of two neurons sharing synaptic inputs vanishes. In that case, the synaptic current to a neuron can be guaranteed to follow a normal distribution, per the central limit theorem (Brunel 2000). It is also a valid assumption when the population fires asychronously and the feedback $R(t)$ is a sum over independent processes, even in dense networks (Tian 2020).

Furthermore, the solution for $C(\omega)$ given in (1.18) has been derived in an alternative fashion by using an iterative approach (Trousdale 2012). This approach is highly valuable as it builds an estimate of $C(\omega)$ by considering paths through a recurrent network of increasing length. Here, a very abstract description of approach will be given and most of the mathematical details are deferred to the original text (Trousdale 2012). Intuitively, the spike-train correlation between a pair of cells can be estimated from the activity of presynaptic cells n synapses away from the cells under consideration. We expect then to improve our estimate of $C(\omega)$ for large values of n. For the simplest case, when $n = 0$, the neurons are uncoupled and we have $C(\omega) = C^0(\omega)$. When $n = 1$, we consider only the nearest neighbors in calculating the cross-correlation between neurons i and j and improve on the uncoupled estimate $C^0(\omega)$. It can be proven that in the limit $n \to \infty$ this iterative procedure also yields (1.18). Trousdale et al. further generalized their iterative approach to give the matrix of cross-spectra in the case were feedforward inputs can have non-trivial correlations

$$C(\omega) = \Delta(\omega)\, C^0\, \Delta^*(\omega) + \Delta(\omega)\, A(\omega) C^{\text{ext}}(\omega) A^*(\omega)\, \Delta^*(\omega) \qquad (2.13)$$

This is a very important result as it provides a mechanism for probing the transfer

of feedforward correlations to the correlations of spike trains. The first term represents the component of the cross spectra from internally generated correlations. The second term is the component that is caused by correlations in the feedforward input. Perhaps the feedforward inputs $F(t)$ are drawn from a multivariate gaussian distribution i.e. $F(t) \sim \mathcal{N}(\mu, \Sigma)$ and we would like to know how μ and Σ affect network dynamics. Ultimately, the linear theory provides a direct mechanism for solving for the cross-covariance of spiking in the population in terms of the synaptic coupling matrix, for arbitrary network topologies. Interestingly, previous analyses have used the above relations to describe the dynamics of motifs under synaptic plasticity when feedforward input is uncorrelated (Ocker 2015). The theory above is very powerful and has several applications. For present purposes, a particularly interesting application would be an examination of the degree of stability of cell assembly formation under different classes of plasticity rules.

It should be noted that models which are dependent on a diffusion approximation, like a linear response theory, have limitations. In particular, the integration of temporally correlated presynaptic spike trains can result in non-trivial two-point correlation functions that are not well-described by a Gaussian white noise process (Moreno-Bote 2008). This has been noted before, in studies of the numerical solution of the Fokker-Planck and master equations which recognized a disparity between the diffusion approximation and full shot-noise stochastic dynamics (Nykamp 2001). Recordings *in-vivo* also have shown the synaptic fluctuations can have significant non-Gaussian properties (DeWeese 2006). In any case, the diffusion approximation and solution of the Fokker-Planck equation can describe important biophysical properties of neuronal response. For example, synaptic filtering, activation of non-linear voltage-gated currents, and non-linear response properties at the onset of a spike (Richardson 2007).

2.2.2   Inferring network structure: the backwards problem

2.2.3   Reshaping network structure with synaptic plasticity

Activity dependent modification of synaptic efficacy or the formation of new synapses is the putative basis of learning and memory formation. The chaining of cellular assemblies that form as a result of synaptic plasticity are thought to provide a structural basis for many cognitive processes such as memory retrieval, reasoning, and planning (Buzaki 2010). Recent neurobiological advances have provided substantial insight on the biochemical and biophysical mechanisms of plasticity. However, the formation of functional circuits and their stability amidst ongoing network dynamics remains unclear.

Plasticity mechanisms are highly diverse and occur over a wide range of time scales: from milliseconds to hours to days (Zenke 2015). For example, short-term synaptic plasticity (STP) is thought to provide a richer set of network responses to stimuli without permanently altering the circuit architecture. For example, repeated stimulation of a presynaptic cell can cause a transient accumulation of Ca2+ in the presynaptic terminal, which has been proposed as a mechanism for short-term memory (Mongillo 2008). Calcium ions elevate the probability of neurotransmitter release due to its proposed interaction with the biochemical machinery involved in synaptic vesicle exocytosis. Other short term changes in synaptic efficacy can occur such as the facilitation or depression based upon the temporal characteristics of the stimulus. Paired-pulse experiments have yielded evidence that tetanic stimulation can result in inactivation of voltage-dependent sodium and/or calcium channels or depletion of the release ready pool of synaptic vesicles at the presynaptic terminal. Synaptic efficacy can alternatively be lowered by the release of biochemical neuromodulators which can interact with the synaptic machinery to inhibit the release of neurotransmitter into the synaptic cleft. Thus, these neuromodulators can also play a role in facilitation or depression in STP.

It is well-accepted that neural circuits also possess the mechanisms for long term changes

in synaptic strength formally referred to as long term potentiation (LTP) and long term depression (LTD). The brain encodes internally and externally generated stimuli as spatiotemporal patterns of action potentials and long-term modifications to such patterns via changes in synaptic transmission provide a feasible mechanism for the storage of information. In other words, permanent or semi-permanent changes in synaptic weights alter the spatiotemporal response of population of neurons to stimuli and therefore provide a method for long term memory formation. Since its inception by Cajal, this idea has been rigorously tested, for example in the CA1 region of the hippocampus (Whitlock et al. 2006). The modification of spatiotemporal characteristics of the neural response to stimuli suggest that evidence for plasticity can be found in the correlation structure of neural spike trains.

Early work on synaptic plasticity at the cellular level demonstrated that LTP in slices of hippocampal tissue could occur as a result of tetanic stimulation of a post-synaptic cell (Bliss 1973) while low frequency stimulation can drive LTD. If many fibers are stimulated at frequencies exceeding 50Hz, postsynaptic potentials can increase by 50-100% (Sejnowski 1989). Additional research expanded upon this observation, noting that temporal correlation of presynaptic and postsynaptic spiking affected plasticity. In particular, strong temporal correlation of presynaptic and postsynaptic firing rates on a timescale of ±100ms can give rise to LTP. The discovery of these phenomena has resulted in the proposal of a wide variety of different plasticity rules which fall under the umbrella of 'Hebbian' learning rules. Hebbian learning rules have been broadly defined as rules in which the change in synaptic efficacy depends only on pre-synaptic and post-synaptic variables and that changes to the synapse depend interactively on these two variables and not separately (Sejnowski 1989). Additionally, Hebbian learning rules are *causal* in that LTP occurs when the presynaptic neuron contributes to the firing of the postsynapic cell and LTD occurs otherwise.

Additional efforts demonstrated that the sign and magnitude of LTP and LTD depend on the order and delay of presynaptic and postsynaptic action potentials, on a time scale

of 10ms (Markram 1997), later named spike timing dependent plasticity (STDP) (Song 2000). In the original formulation of Hebbian STDP, LTP can occur when presynaptic spikes lead postsynaptic spikes by $\sim$ 20ms and LTD can occur when postsynaptic spikes lead presynaptic spikes by $\sim 20 - 100$ms (Markram 1997; Bi and Poo 1998; Celikel 2004; Feldman 2012). However, plasticity does not only depend on spike timing, but also on other factors such as firing rates, and synaptic cooperativity, and postsynaptic voltage (Makram 1997). Therefore, findings indicate that STDP is part of a broader class of plasticity rules and may be viewed as the spike timing component of a multifactor plasticity rule (Feldman 2012).

Furthermore, the biochemical mechanism of synaptic plasticity is regulated by the release of neurotransmitters by the presynaptic cell and/or by modifying the density of receptors at the postsynaptic membrane (Sumi 2020). Indeed, it is widely believed that presynaptic release of glutamate results in activation of AMPA receptors and subsequent depolarization of the post-synaptic membrane. Membrane depolarization then activates the N-methyl-D-aspartate receptor (NMDAR), which is permeable to Ca2+ which is a prefactor in a biochemical cascade resulting in the trafficking of AMPA receptors to the postsynaptic membrane. High elevations in Ca2+ concentrations in the presynaptic cell are believed to result in an increased density of AMPA receptors at the postsynaptic membrane and in elevated excitatory postsynaptic potentials (EPSPs) thereby providing a mechanism for LTP (Sumi 2020). In summary, LTP versus LTD induction is determined by the magnitude and time course of calcium flux (Lisman 1989). When the presynaptic cell fires first, the EPSP produces a strong NMDAR calcium flux while while when the postsynaptic cell fires first, this flux is weak. On the other hand, LTD has not been shown to occur soley due to minor elevations in Ca2+ concentrations, but is thought to result from a retrograde signaling mechanism in which presynaptic transmitter release probability is decreased (Nabavi 2013).

Beyond spike-timing dependent plasticity, other plasticity mechanisms have been ob-

served experimentally. In the hippocampus, non-Hebbian heterosynaptic LTD alongside LTP was described shortly after the phenomenon of LTP was discovered (Lynch 1977). The Hebbian mechanisms or *associative mechanisms* discussed above are termed homosynaptic mechanisms, as they depend on both pre- and post-synaptic activity. There are also heterosynaptic mechanisms which depend on only the state of the post-synaptic neuron and are thought to exist to counteract chaotic dynamics introduced by Hebbian-type rules and to balance synaptic changes. Hebbian homosynaptic learning rules, by definition, potentiate (depress) synapses with positive feedback, meaning that increasing (decreasing) synaptic efficacy will promote further potentiation (depression). For example, The standard doublet STDP model gives rise to splitting of synaptic weights, producing a bimodal distribution; however, weight distributions observed in experiments are unimodal and long-tailed (Perin 1997; Song 2005). The result is the overexcitability of some neurons and silencing of others making network dynamics unstable. Induction of LTD at high firing rates of the postsynaptic neuron is a candidate for an intrinsic stabilization mechanism (Kempter 2001). Also, it has been recent suggested that rate-dependent terms in the synaptic learning rule can stabilize motif configurations (Ocker 2015). Ultimately, it is thought that a push-pull mechanism exists in which homosynaptic LTP prevents the network from falling silent at low firing rates while heterosynaptic LTD prevents the runaway dynamics at high firing rates.

Understanding the stable formation and maintenance of cell assemblies is an additional goal of synaptic plasticity research. From a näive point of view, a cell assembly could form if a subset of a neuron's inputs were to potentiate while other inputs depress. However, for this to occur both groups should be repeatedly activated either with highly specific patterns to promote LTP or LTD, respectively (Chistikova 2014). This could then limit network stability to special sequences of input patterns while we expect that plasticity mechanisms exist to promote stability under a variety of input stimuli. Moreover, experiments involving intercalated neurons of the amygdala have demonstrated that LTP induction with high-frequency stimu-

lation can simultaneously result in heterosynaptic depression (Royer 2003). Conversely, LTD induction with low-frequency stimulation was also shown to induce LTP. These observations indicate the existence of intracellular signaling that would provide synapses with 'awareness' of eachother. Learning rules which depend on the post-synaptic firing rate would produce a similar result, although often these rules do not perform selective potentiation or depression. That is, synaptic weights are normalized without regard to which synapses can be blamed for chaotic or silent neuron states. In circuits with highly regular organization of the inputs, such as the hippocampus or amygdala, the onset of LTP can induce weaker LTP at nearby inputs while concurrently inducing heterosynaptic LTD at more distant inputs, while the magnitude of LTD decreasies with distance (Royer 2003).

The experimental observations discussed suggest that balanced LTP and LTD may be a powerful mechanism of for synaptic normalization and synaptic competition. Thus, several computational studies incorporating more realistic forms of synaptic plasticity with conductance based neuron models have surfaced (Zenke 2015; Kumar 2014). Many have incorporated a balance between excitation and inhibition to capture the irregular activity of cortical networks, but stable formation of cell assemblies is difficult to achieve. This short-coming could, in part, be due to the tendency of STDP to lead to pathological behavior in balanced networks (Morrison 2007). The introduction of homeostatic mechanisms such as normalization of total conductance alongside inhibitory STDP to control the firing rate of excitatory neurons has been used to inhibit pathological activity in the balanced regime (Kumar 2014).

Returning to the theoretical discussion, in a plastic network the synaptic weights (conductances) are functions of time. Thus a more general form of (1.4) reads

$$I_j^\alpha(t) = \sum_\beta \sum_i g_{ij}^\beta(t)(E_\beta - V_j)$$

$$= \sum_\beta \sum_i J_{ij}^{\alpha\beta}(t) \left(\phi * z_i(t)\right)(E_\beta - V_j)$$

Biologically, excitatory conductances such as AMPAR/NMDAR would be under the control of an excitatory neurotransmitter e.g. glutamate while inhibitory conductances regulated by perhaps gamma-aminobutyric acid (GABA). Notice that the time-dependence of g^α can then be given by the convolution of an excitatory presynaptic spike train $z_i(t)$ or inhibitory spike train $z_k(t)$ with an appropriate synaptic kernel ϕ. The synaptic plasticity rule is then a rule for the time-derivative $\dot{J}_{ij}^{\alpha\beta}$.

The cellular machinery of plasticity mechanisms described above highlights the importance of spike-time correlations, generated by recurrent feedback or feedforward projections, in determining the evolution of synaptic weights. A classic derivation of the evolution of synaptic weights for interacting Poisson processes was given in (Kempter 1999). This result is an important starting point for later analysis. Consider the total change in the synaptic weight between neurons i and j as $\Delta J_{ij} = J_{ij}(t+\Delta t) - J_{ij}(t)$ in a time interval Δt. Consider the very general rule for a single spiking neuron i receiving many inputs indexed by j

$$\dot{J}_{ij} = a_0 + a_i z_i(t) + a_j z_j(t) + F(z_i(t), z_j(t))$$

for a yet unknown function F, which could be a STDP rule, for example. It is common to assume that the constant growth or decay term a_0 is zero. Spike trains are a summation of delta functions over time, i.e. $z_i(t) = \sum_n \delta(t - t_i^n)$ and $z_j(t) = \sum_m = \delta(t - t_j^m)$.

If we consider only rules that consider the timing of pairs of spikes, changes in synaptic weight can occur from the arrival of a presynaptic spike $z_i(t) = 1$, a postsynaptic spike
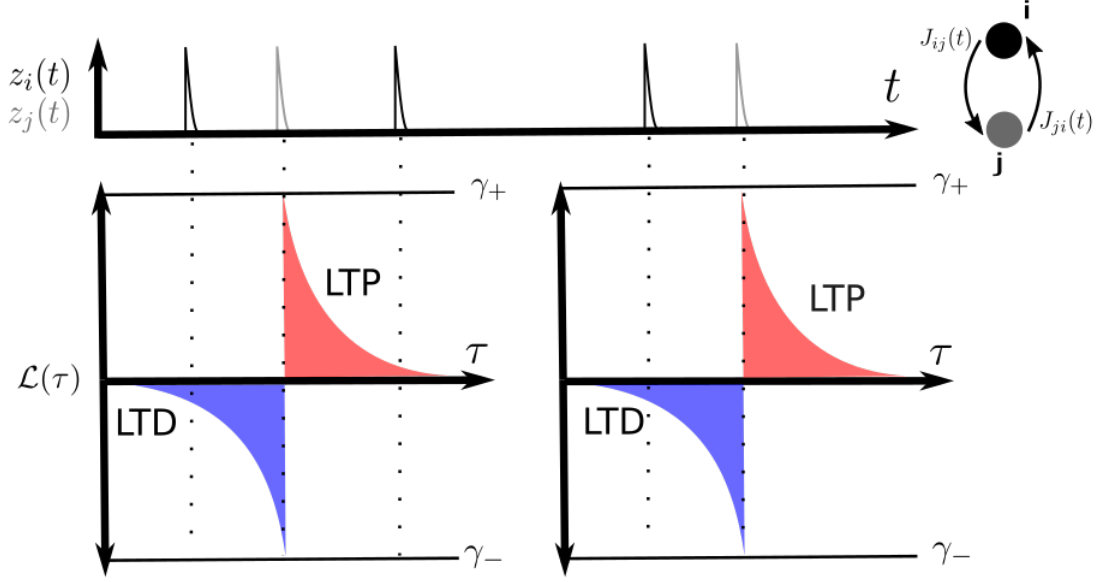
Figure 2.2: Cartoon of the doublet spike-timing dependent plasticity rule. The time course of two spike trains $z_i(t)$ and $z_j(t)$ is shown in black and gray, respectively. The learning window $\mathcal{L}(\tau)$ determines whether LTP or LTD occurs and the magnitude of the weight change. Adapted from (Kempter 2001).

$z_j(t) = 1$ or as a function of the delay in spike timing τ. The spike-timing component of the learning rule is given by the learning window function $\mathcal{L}(\tau)$. A commonly used example of the learning window function is the doublet spike-timing dependent plasticity rule

$$\mathcal{L}(\tau) = \begin{cases} H(J_{max} - J_{ij})\gamma_+ \exp(-\tau/\tau_+), & \text{if } \tau \geq 0 \\ H(J_{ij})(-\gamma_-) \exp(-\tau/\tau_-), & \text{if } \tau < 0 \end{cases} \tag{2.14}$$

which has been illustrated in Fig (1.2). This rule enforces bounds on the weight as well as the sign dependence of the synaptic update characteristic of STDP. Integrating over a time window τ_{lrn} gives the total change ΔJ_{ij} (Kempter 1999)

$$\Delta J_{ij}(\tau_{\mathrm{lrn}}) = \int_0^{\tau_{\mathrm{lrn}}} dt' \left(a_i z_i(t') + a_j z_j(t') \right) \tag{2.15}$$

$$+ \int_0^{\tau_{\mathrm{lrn}}} dt' \int_0^{\tau_{\mathrm{lrn}}} dt'' \mathcal{L}(\tau) z_i(t'') z_j(t') \tag{2.16}$$

The stochasticity of the spike functions $z_i(t)$ and $z_j(t)$ means that we can only hope to derive the behavior of the above integral on average. In other words, ΔJ_{ij} is itself a stochastic process but we can focus on its drift or expected rate of change. Simulaneously, $z_i(t)$ and $z_j(t)$ are strongly dependent on the weights J_{ij}; however, if the time scale of plasticity is sufficiently long compared to the correlation time scale of the neurons activity, the correlation between $z_i(t)$ and $z_j(t)$ can be treated as stationary (Kempter 2001). Importantly, the effective correlation time scale of neuron dynamics is intimately related to the timescale over which it is 'picked up' by the learning window $\mathcal{L}(\tau)$. Analyzing the drift in ΔJ_{ij} then reads

$$\tau_{\mathrm{lrn}} \frac{\langle \Delta J_{ij} \rangle (t)}{\tau_{\mathrm{lrn}}} = \int_0^{\tau_{\mathrm{lrn}}} dt' \left(a_i z_i(t') + a_j z_j(t') \right)$$

$$+ \int_0^{\tau_{\mathrm{lrn}}} dt' \int_{-t'}^{\tau_{\mathrm{lrn}}-t'} d\tau \mathcal{L}(\tau) \langle z_i(t'+\tau) z_j(t') \rangle$$

Note that we can make the same approximations for a sufficiently slow process as we can for a relatively fast process with high time resolution. Then, if synaptic plasticity is sufficiently slow, then we can approximate the average rate of change as $dJ_{ij}/dt \approx \Delta J_{ij}/\tau_{\mathrm{lrn}}$, and the bounds of integration can be set to $-\infty$ and $+\infty$ (Kempter 1999) and we have the final result

$$\tau_{\mathrm{lrn}} \frac{dJ_{ij}(t)}{dt} = a_i \langle z_i(t) \rangle + a_j \langle z_j(t) \rangle + \int_{-\infty}^{+\infty} d\tau \mathcal{L}(\tau) C_{ij}(\tau) \tag{2.17}$$

where $\langle z_i(t) \rangle$ and $\langle z_j(t) \rangle$ can be interpreted as the instantaneous firing rate of neurons i and j, respectively. It has already been shown that linear response theory can be used to estimate $C\tau)$ for arbitrary network topologies. Furthermore, the equation (1.22) including the instantaneous post-synaptic firing rate can serve as a homeostatic term if $a_j < 0$. Although, as highlighted previously, rate-dependent terms in the learning rule do not selectively enforce LTP and LTD since the term $a_j \langle z_j(t) \rangle$ appears in the equation for $\dot{J}_{ij}$ for all input neurons i (assuming a_j is a constant). Thus we would need to modify (1.22) to arrive at a synaptic plasticity rule which delivers potentiation or depression heterogeneously, for example using information regarding spatial proximity. However, we now argue that this may not be necessary.

2.3 Results

2.3.1 Predicting uncoupled firing rates

In the linear response framework it is necessary to find the baseline firing rate of neuron from its average input current which is a sum of the mean feedforward input and the mean recurrent input i.e. $\mu = \mu_F + \mu_R$. Generally, the mean of the feedback current μ_R is found by iteratively solving the Fokker-Planck equation for the specified neuron model and using the predicted firing rate for each neuron to update the rate of the rest of the neurons in the next iteration. That is, at each iteration, we compute

$$\mu_R^{t+1} = \mu_F + J * r^t$$

assuming that μ_F is constant. The calculation of average feedback current is expressed in matrix form and $*$ represents standard matrix multiplication. The Fokker-Planck equation is then reparameterized for Gaussian white noise with mean μ_R^{t+1} and variance equal to the variance of the feedforward input. This is referred to as a fixed-point iteration to find the steady state firing rates of all neurons in the network (Trousdale 2012; Ocker 2015). Summarizing the sum of feedforward and feedback currents as a single gaussian variable with mean μ_i gives the following simplified Langevin equation for the membrane potentiial

$$\tau \dot{V}_i(t) = E - V_i + \psi(V_i) + \eta_i(t) \tag{2.18}$$

where $\eta_i(t) \sim \mathcal{N}(\mu_i, \sigma^2)$. The corresponding Fokker-Planck equation reads

$$\frac{\partial P}{\partial t} = \frac{\sigma^2}{\tau} \frac{\partial^2 P}{\partial V^2} + \frac{\partial}{\partial V} \left(\frac{V - E + \psi}{\tau} P \right) \tag{2.19}$$

The full details of this method for finding both the steady state density as well as the density under sinusoidal perturbation were given in (Richardson 2007) and will not be repeated here in detail.

To determine the level of similarity between the numerical prediction of r_0 versus the result of Monte-Carlo simulations, $N = 100$ EIF neurons were simulated for GWN input with $\mu = 2\mu A/\text{cm}^2$ and $\sigma = 9 \ \mu A/\text{cm}^2$ over a period $T = 1$s. A spike raster illustrating network activity in the steady state (the last 100ms of the simulation) is shown in Fig (1.4a) with example voltage traces from two randomly selected neurons in Fig (1.4c). The average firing rate measured in simulations was 69.2Hz (Fig 1.4d) in strong agreement with the numerical prediction 73.2Hz (Fig 1.4b). The numerically predicted firing rate can then be used to estimate the frequency spectrum of the spiking activity of each neuron i.e. the value
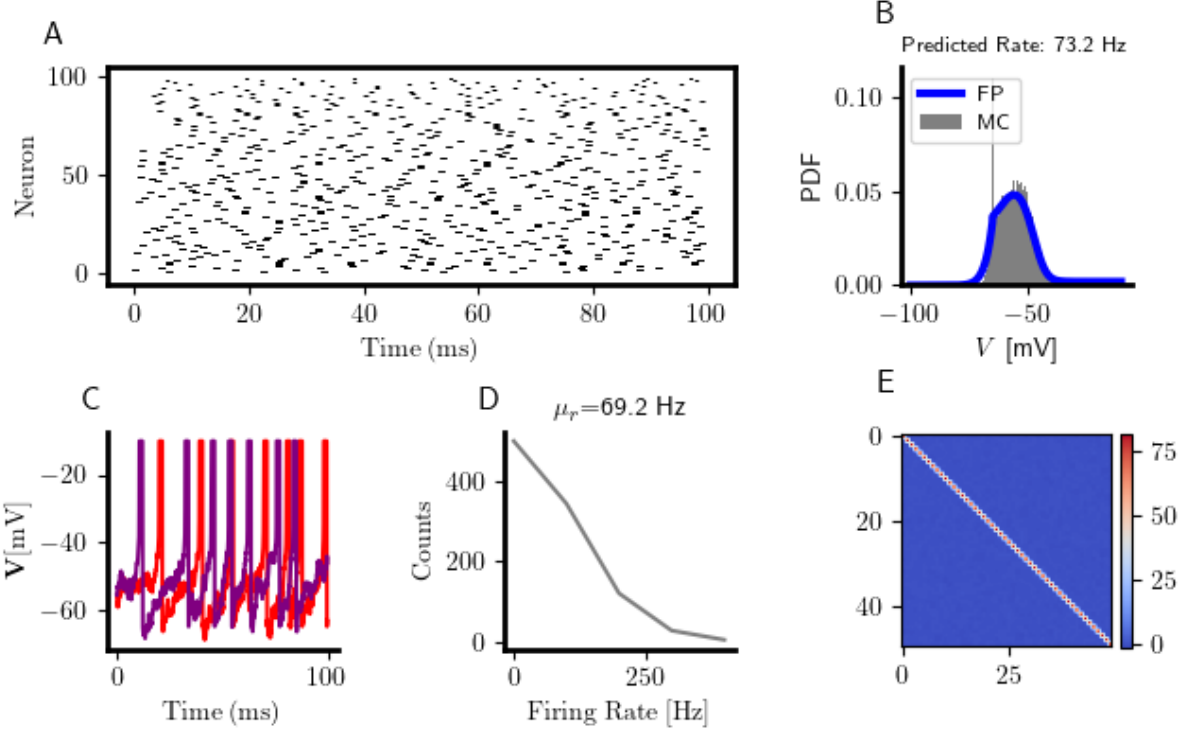
Figure 2.3: Fokker-Planck solution in the uncoupled regime (A) Steady-state raster plot of $N = 100$ uncoupled EIF neurons undergoing stimulation with GWN with $\mu = 2\mu A/\text{cm}^2$ and $\sigma = 9$ $\mu A/\text{cm}^2$. (B) Histogram of the membrane potential in the steady-state and Fokker-Planck predicted histogram. (C) Two randomly selected voltage traces (D) Histogram of the population-averaged firing rate and its mean value (E) Covariance matrix of the feedforward white-noise input

of $C_{ii}^0(\omega)$.

Next, the frequency response $A_{ii}(\omega)$ was estimated by evaluating the solution of the Fokker-Planck equation for a sinusoidal perturbation. Since, the average input to a neuron μ will generally depend on the mean value of the feedforward input as well as the rates of other neurons in the network, a solution was found for various values of μ. Interestingly, the magnitude of the frequency response showed a resonance at increasing values of ω as the average input was increased (Figure 1.5 left). At the same time, the response of the firing rate was in generally in phase with the sinusoidal perturbation for small values of ω and increasingly out of phase for increasing frequencies. The obtained frequency response, in combination with the estimate of $C_{ii}^0(\omega)$ obtained previously from the steady state solution
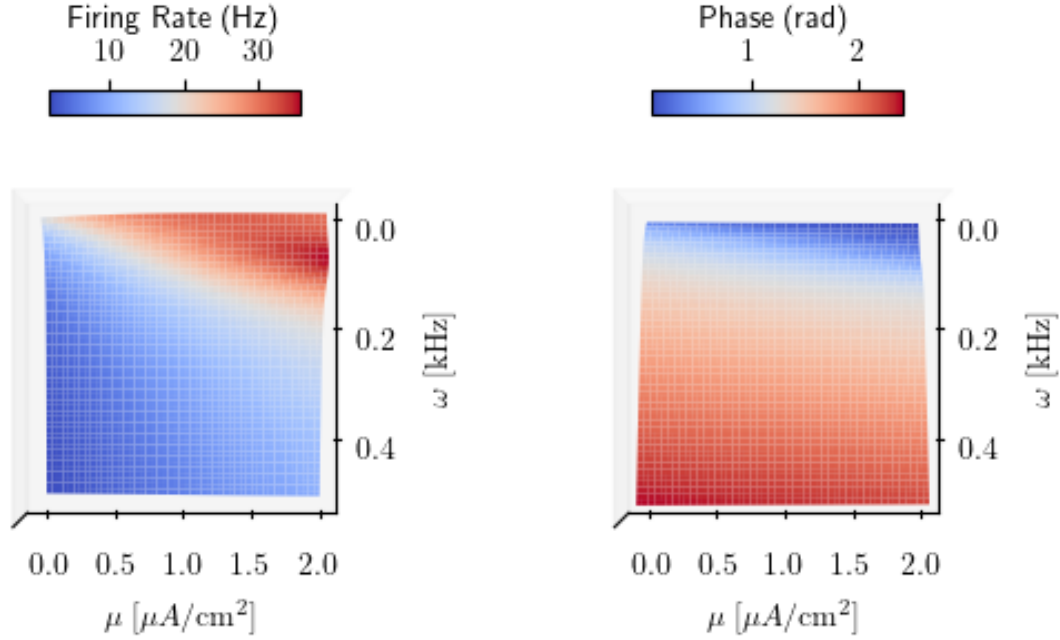
Figure 2.4: Frequency reponse as a function of average input (left) Predicted firing rate under a weak sinsuoidal perturbation with frequency ω for various values of the average input current μ. (right) Phase of the response.

can be used to find an estimate of the solution of (1.31) in later studies when the neurons are coupled.

2.3.2 Asynchronous spiking in the balanced state

A central goal of future studies is to make use of mean-field and linear response techniques to generate balanced spatially-extended networks. In particular, it is desirable to use a hybrid approach in which mean-field theory is used to prove the existence of balance and a linear response approximation can detect deviations from the initial balanced state. Moving in this direction, a simple random network model was generated which satisfied the constraints for balance in random networks given previously (see Table 1.2 for a full list of parameters). The synaptic connectivity matrix was taken to be static and synaptic kernels were expo-nential. Network stimulation was carried out as before with delta-correlated white noise.
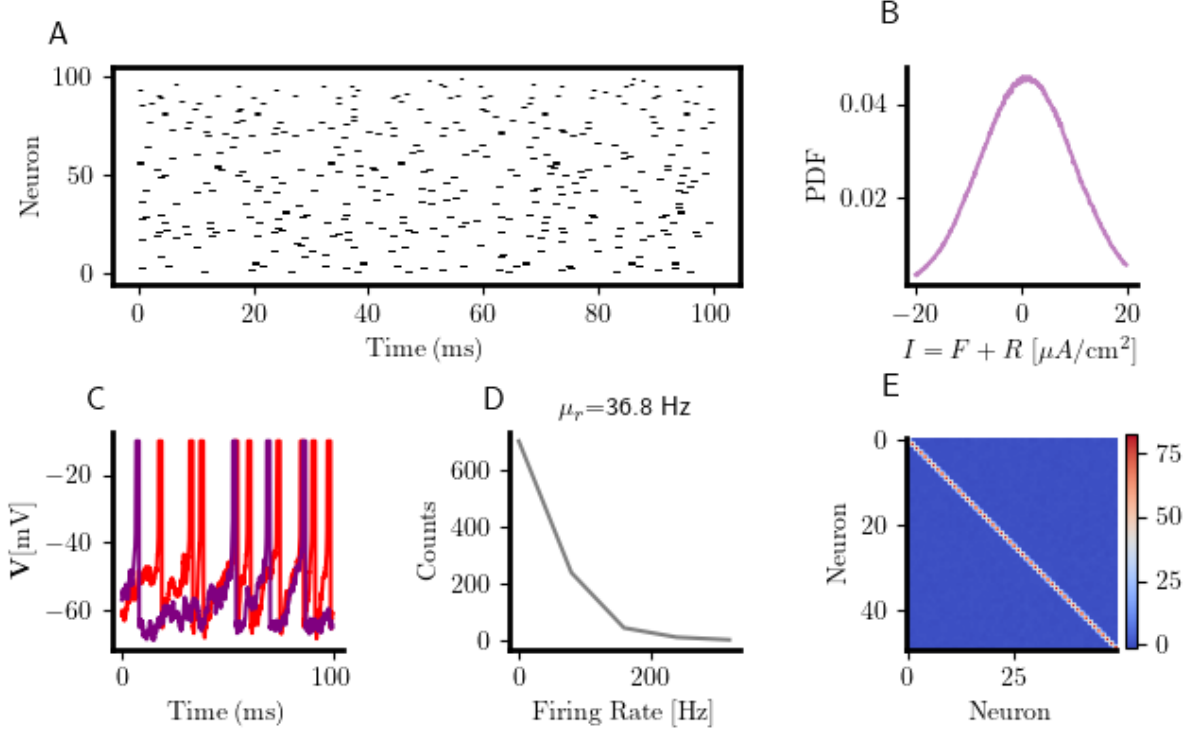
Figure 2.5: Asynchronous spiking in the balanced state (A) Steady-state raster plot of $N = 100$ uncoupled EIF neurons undergoing stimulation with GWN with $\mu = 2\mu A/cm^2$ and $\sigma = 9 \ \mu A/cm^2$. (B) Distribution of the total synaptic current in the balanced state (C) Two randomly selected voltage traces (D) Histogram of the population-averaged firing rate and its mean value (E) Covariance matrix of the feedforward GWN input

Characteristic of the balanced state, the distribution of the sum of feedforward and feedback had zero mean and was bell-shaped. The variance of synaptic current distribution was not examined here, although the result in (Figure 1.6b) suggests that it could be estimated and the Fokker-Planck equation could be used the predict the firing rate. A similar approach has been used for simpler integrate and fire neuron models (Tian 2020). Importantly, the mean firing rate measured was reduced from the uncoupled case by a factor of ≈ 50 percent, demonstrating the important role of inhibition in preventing runaway firing rates.

Auto- and cross-correlations in the balanced state depicted in Fig (1.6) are shown in Fig (1.7). The mean cross-correlation of spike-trains shown as $\langle Z_{xy}(\tau) \rangle$ shows an obvious weak correlation for all τ, as expected in the balanced state. Auto- and cross-correlations for
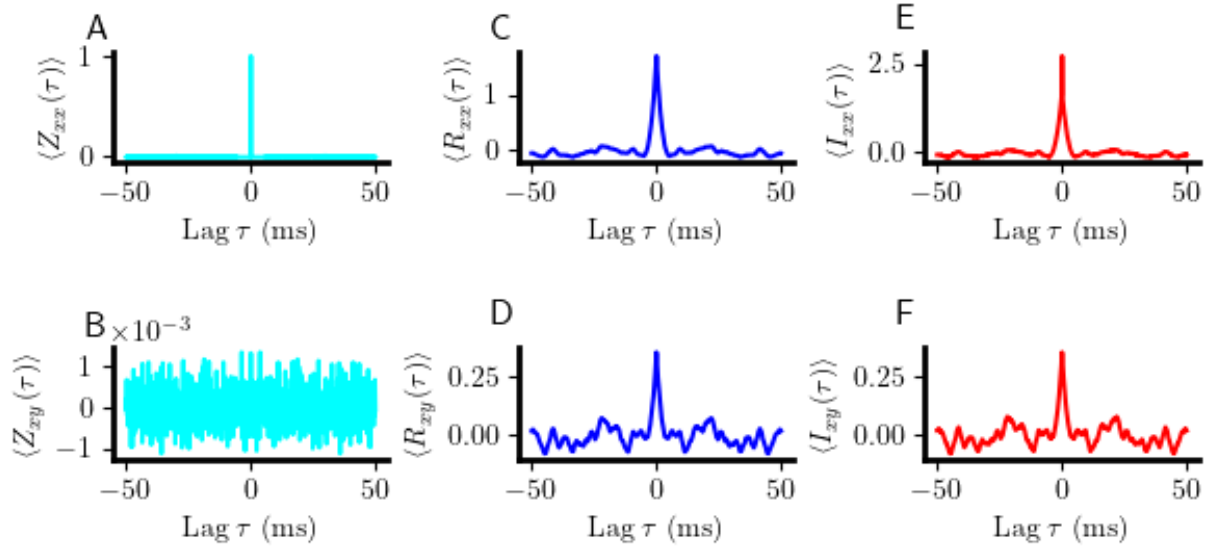
Figure 2.6: Auto- and cross-correlations in the balanced state (A) Mean auto-correlation of spike trains (B) Mean cross-correlation of spike trains (C) Mean auto-correlation of the feedback current (D) Mean cross-correlation of the feedback current (E) Mean auto-correlation of the total synaptic current (F) Mean cross-correlation of the total synaptic current

the total synaptic current and its recurrent component appear very similar. This is reasonable, considering that the delta-correlated feedforward input lacks temporal structure and therefore the temporal structure of the total current should be reminiscent of the feedback component. The results of Fig (1.6) and Fig (1.7) are taken as proof of the balanced state in this simple network model. A natural step in a future study would make use of the linear response theory discussed previously to compare these results with theoretical predictions.

2.4 Methods

2.4.1 Monte-Carlo simulations

Monte-Carlo simulations for the balanced random network were run using custom code developed in C and Python. The following exponential integrate and fire (EIF) neuron model was used

$$C\frac{dV_i}{dt} = g_\ell(E - V_i) + g_\ell\Delta_T \exp\left(\frac{V_i - V_T}{\Delta_T}\right) + R_i(t) + \xi_i(t) \qquad (2.20)$$

with $R(t) = \sum_i J_{ij}(t) \cdot \phi(t) * \delta(t - t_i)$ and $\xi_i(t)$ is a delta correlated white noise. The stochastic differential equation (1.30) was integrated numerically using the Euler method and the feedforward input was sampled according to $\xi_i(t) \sim \mathcal{N}(\mu_{\text{ext}}, \sigma^2_{\text{ext}})$. The synaptic kernel was taken to be a decaying exponential i.e. $\phi(t) = \exp(-t/\tau_s)$. Unless explicitly mentioned otherwise, the parameters for excitatory and inhibitory neurons were kept the same. A full list of parameters used for the balanced random network can be found in Table (1.1). Simulations were run for a total time period $T = 1$ s with a time resolution of $\Delta t = 100$ms. Auto- and cross-correlations are calculated by considering the steady state activity which is taken to be the last 100ms of the simulation (this period will be referred to as the steady state window Δt_{ss}).

Auto- and cross-correlations of feedforward currents, feedback currents, and spike trains were computed by first finding the matrix of cross-spectra $C(\omega)$ using a vectorized Python routine. Before calculating $C(\omega)$, the signals are normalized by subtracting out their mean and dividing by the standard deviation over the steady state window Δt_{ss}. Let $x(\omega)$ be the discrete-time Fourier transform (DTFT) of an arbitrary signal $x(t)$

$$x(\omega) = \sum_{n=-\infty}^{\infty} x(t) \left(\exp -i\omega n \right) \qquad (2.21)$$

We can quickly compute an element matrix of cross-correlations $C_{ij}(\omega)$ in the frequency domain. For example, given spike trains $z_i(t)$ and $z_j(t)$ we can first apply (1.31) and then take the product

$$C_{ij}(\omega) = \lim_{T \to \infty} \frac{1}{T} z_i^*(\omega) z_j(\omega) \qquad (2.22)$$

where $*$ represents complex conjugation. The above quantity is the complex-value cross-spectrum of the spike trains $z_i(t)$ and $z_j(t)$. An identical calculation is carried out for the feedforward, feedback, and total currents.

2.4.2 Solution to the Fokker-Planck equation

As in (1.2), the EIF model uses $\psi(V) = \Delta_T \exp\left(\frac{V - V_T}{\Delta_T}\right)$. (1.33) must be solved for the distribution $P(V_i, t)$ to find the steady state firing rate of neuron i. The basic idea presented in (Richardson 2007) starts with rewriting the Fokker-Planck equation as a continuity equation

$$\frac{\partial P}{\partial t} = -\frac{\partial J}{\partial V}$$
$$J = \frac{\sigma^2}{\tau} \frac{\partial P}{\partial V} + \left(\frac{V - E + \psi}{\tau} P \right)$$

In the steady state, the second equation above is rearranged to give a coupled set of first order differential equations in the voltage

$$-\frac{\partial P_0}{\partial V} = \frac{\tau}{\sigma^2} \left(\frac{V - E + \psi}{\tau} P_0 + J_0 \right)$$

$$-\frac{\partial J_0}{\partial V} = r_0 \left(\delta(V - V_{\text{th}}) - \delta(V - V_{\text{re}}) \right)$$

where V_{th} and V_{re} refer to the firing threshold and reset potentials, respectively. Briefly, the steady state firing rate r_0 can be cancelled from the above equations and they can be integrated numerically. Similar expressions exist for the case when the current undergoes a weak sinusoidal perturbation, which is important for finding the linear response function of the firing rate $A(\omega)$.

CHAPTER 3

STATISTICAL METHODS

Appendices

APPENDIX A

DERIVATION OF THE FOKKER PLANCK EQUATION

A.0.1 Kramers-Moyal Expansion

Gixen many instantiations of a stochastic xariable x, we can construct a normalize histogram oxer all obserxations as a function of time $P(x,t)$. Howexer, in order to systematically explore the relationship between the parameterization of the process and $P(x,t)$ we require an expression for $\dot{P}(x,t)$. If we make a fundamental assumption that the exolution of $P(x,t)$ follows a Markox process i.e. its exolution has the memoryless property, then we can write

$$P(x',t) = \int T(x',t|x,t-\tau)P(x,t-\tau)dx \tag{A.1}$$

which is known at the Chapman-Kolmogorox equation. The factor $T(x',t|x,t-\tau)$ is known as the *transition operator* in a Markox process and determines the exolution of $P(x,t)$ in time. We proceed by writing $T(x',t|x,t-\tau)$ in a form referred to as the Kramers-Moyal expansion

$$
\begin{aligned}
T(x',t|x,t-\tau) &= \int \delta(u-x')T(u,t|x,t-\tau)du \\
&= \int \delta(x+u-x'-x)T(u,t|x,t-\tau)du
\end{aligned}
$$

If we use the Taylor expansion of the δ-function

$$\delta(x+u-x'-x) = \sum_{n=0}^{\infty} \frac{(u-x)^n}{n!}\left(-\frac{\partial}{\partial x}\right)^n \delta(x-x')$$

Inserting this into the result from aboxe, pulling out terms independent of u and swapping

42

the order of the sum and integration gixes

$$T(x', t | x, t - \tau) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n \delta(x - x') \int (u - x)^n T(u, t | x, t - \tau) du \qquad \text{(A.2)}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n \delta(x - x') M_n(x, t) \qquad \text{(A.3)}$$

noticing that $M_n(x, t) = \int (u - x)^n T(u, t | x, t - \tau) du$ is just the nth moment of the transition operator T. Plugging (2.6) back in to (2.4) gixes

$$P(x, t) = \int \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n M_n(x, t) \right) \delta(x - x') P(x, t - \tau) dx \qquad \text{(A.4)}$$

$$= P(x', t - \tau) + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n [M_n(x, t) P(x, t)] \qquad \text{(A.5)}$$

Approximating the derixatixe as a finite difference and taking the limit $\tau \to 0$ gixes

$$\dot{P}(x, t) = \lim_{\tau \to 0} \left(\frac{P(x, t) - P(x, t - \tau)}{\tau} \right) \qquad \text{(A.6)}$$

$$= \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x} \right)^n [M_n(x, t) P(x, t)] \qquad \text{(A.7)}$$

which is formally known as the Kramers-Moyal (KM) expansion. The Fokker-Planck equation is a special case of (2.10) where we neglect terms $n > 2$ in the *diffusion approximation.*

Consider the following Ito stochastic differential equation

$$dx\vec{} = F(\vec{x}, t) + G(\vec{x}, t)dW$$

The SDE gixen aboxe corresponds to the Kramers-Moyal expansion (KME) of a transition density $T(x', t'|x, t)$ see (Risken 1989) for a full derixation.

$$\frac{\partial P(x, t)}{\partial t} = \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n [M_n(x, t)P(x, t)] \tag{A.8}$$

where M_n is the nth moment of the transition density. In the diffusion approximation, the KME becomes the Fokker-Planck equation (FPE) (Risken 1989). For the sake of demonstration, consider the univariate case with random variable x and the form of $T(x', t'|x, t)$ is a Gaussian with mean $\mu(t)$ and variance $\sigma^2(t)$. In this scenario, the FPE applies because $M_n = 0$ for all $n > 2$. Given that the drift $M_1(x, t) = \mu(t)$ and the diffusion $M_2(x, t) = \sigma^2(t)$, the FPE reads

$$\frac{\partial P(x, t)}{\partial t} = \left(-\frac{\partial}{\partial x}M^{(1)}(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}M^{(2)}(t)\right) P(x, t) \tag{A.9}$$

We can additionally define the term in parentheses as a differential operator acting on $P(x, t)$

$$\hat{\mathcal{L}}_{FP} = \left(-\frac{\partial}{\partial x}M^{(1)}(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}M^{(2)}(t)\right) \tag{A.10}$$

It is common to additionally define the probability current $J(x, t)$ as

$$J(x,t) = \left(M^{(1)}(t) - \frac{1}{2}\frac{\partial}{\partial x} M^{(2)}(t) \right) P(x,t) \tag{A.11}$$

This definition provides some useful intuition. The value of $J(x,t)$ is the net probability flux into the interval between x and $x + dx$ at at time t. This also allows us to write the FPE as a continuity equation

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial J(x,t)}{\partial x} \tag{A.12}$$

3.1 Poisson processes

The Poisson process can be derived very quickly by noticing that it is simply the continuous-time limit of the Binomial distribution.

$$B(m;t) = \binom{n}{m} \lambda^m (1-\lambda)^{n-m}$$

Since λ is the fraction of successes, the expected number of successes is $\mu = n\lambda$

$$\begin{aligned}
B(m;t) &= \binom{n}{m} \left(\frac{\mu}{n}\right)^m \left(1-\frac{\mu}{n}\right)^{n-m} \\
&= \binom{n}{m} \left(\frac{\mu}{n}\right)^m \left(1-\frac{\mu}{n}\right)^{n} \left(1-\frac{\mu}{n}\right)^{-m}
\end{aligned}$$

45

$$B(m;t) = \frac{n!}{m!(n-m)!} \left(\frac{\mu}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^n \left(1 - \frac{\mu}{n}\right)^{-m}$$

$$= \frac{n!}{(n-m)!} \left(\frac{1}{n}\right)^m \left(1 - \frac{\mu}{n}\right)^{-m} \frac{\mu^m \left(1 - \frac{\mu}{n}\right)^n}{m!}$$

In the first term, we can take the first m subterms of the numerator $n! = n(n-1)...(n-m)$ and, since $n >> m$, each term will cancel with one factor of n from the term $1/n^m$. This leaves

$$B(m;t) = \frac{(n-m)!}{(n-m)!} \left(1 - \frac{\mu}{n}\right)^{-m} \frac{\mu^m \exp(-\mu)}{m!}$$

We now take the continuous time limit i.e. $n \to \infty$ and, again, since $m << n$ we are left with

$$\lim_{n \to \infty} B(m;t) = \frac{\mu^m \exp(-\mu)}{m!}$$

If an event can be detected will probability γ, the rate of the Poisson process will be reduced by that factor i.e., $\lambda' = \gamma\lambda$. Therefore, the mean and variance of the process becomes $\mu = \gamma\lambda\Delta t$ v