

THE UNIVERSITY OF CHICAGO

STOCHASTIC COMPUTATION IN RECURRENT NETWORKS OF SPIKING
NEURONS

A THESIS SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL AND BIOLOGICAL
SCIENCES

IN CANDIDACY FOR THE DEGREE OF
MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOPHYSICS

BY

CLAYTON W. SEITZ

CHICAGO, ILLINOIS

WINTER 2021

Copyright © 2021 by Clayton W. Seitz
All Rights Reserved

Epigraph

TABLE OF CONTENTS

ACKNOWLEDGMENTS	v
ABSTRACT	vi
1 INTRODUCTION	1
2 STATISTICAL MECHANICS OF A POPULATION OF NEURONS	3
2.1 Formalism	3
2.2 The Fokker-Planck Equation	4
3 DYNAMICAL STATES OF RECURRENT NETWORKS	9
3.1 Introduction	9
4 STOCHASTIC COMPUTATION BY RECURRENT NETWORKS	10
4.1 Introduction	10
5 CONCLUSIONS	11
REFERENCES	12

ACKNOWLEDGMENTS

ABSTRACT

One of the most challenging questions in modern neuroscience is the mechanisms by which primate cortical circuits learn from example and perform inference based upon their past experience. The cerebral cortex is responsible for many higher-order brain functions including memory and . The search for principles of neural computation common to these phenomena is made extremely difficult by the diversity of cortical structures and their putative functions. Nevertheless, progress has been made in our understanding of the interplay between the organization of cortical neurons, network dynamics, and the resulting reorganization via synaptic plasticity during learning. Many deficiencies in our models of these processes can be attributed to our lack of a mathematical understanding of both the fundamental computational paradigm implemented by neural circuits and the roles of synaptic plasticity mechanisms observed *in vivo*. These problems are further aggravated by the apparent stochastic features of the firing activity of neurons and experimentally observed trial-to-trial variability in population recordings. However, there is, in principle, a very powerful framework for stochastic computation evident in the Boltzmann machine: probabilistic inference by sampling. Here, we implement a recurrent spiking neural network (RSNN) as a variant of the Boltzmann machine and propose a precise mathematical role for a set of synaptic plasticity mechanisms in the formation of a set of synaptic weights that successfully “embody” a Gaussian input statistic.

CHAPTER 1

INTRODUCTION

Recent decades have yielded a number of experimental and theoretical advances in our understanding of learning in biological neural networks through the introduction of as homosynaptic, heterosynaptic, and homeostatic plasticity. However, the precise mathematical relationship between the synaptic plasticity mechanisms or *learning rules* implemented by neurons and the formation of probabilistic models of their inputs, remains elusive.

Neural network models date back to McCulloch and Pitts [1] followed later by the Hopfield network [2] which spawned an entire field of research applying techniques from the statistical physics of spin glasses to the description of neural activity. By simplifying a network of neurons firing action potentials to an ensemble of coupled spins $\sigma \in \{-1, +1\}$, the Hopfield model related the patterns learned by a network to the energy landscape over the discrete space of states. The storage capacity of these networks and the geometry of this energy landscape were of particular interest and rigorous mathematical treatment has been used to show limits on the basins of attraction in such energy landscapes [4,10,11]. However, recent experimental evidence has suggested that networks of neurons may follow a stochastic, rather than deterministic, computational paradigm. There are many examples in the literature of trial-to-trial variability in the response of cortical neurons to identical stimuli, suggesting that computations in the brain are inherently stochastic. The origins of this stochasticity are hypothesized to originate in noise in synaptic conductances and infidelity in processes clearing neurotransmitters from the synaptic cleft. Interestingly, an extension of the Hopfield model - the Boltzmann machine actually leverages stochastic activity of Ising spins to perform powerful computations [14]. In such a model, the set of synaptic weights Φ “embody” the joint distribution $P_{\Phi}(X, R) = P_{\Phi}(X)P_{\Phi}(R|X)$ over network inputs and network response, respectively. Then, computations can be viewed as probabilistic inference after suitable transformations of the weights $\Delta\Phi$. Such physically inspired models have proven useful;

however, the models are sufficiently abstract to make experimental comparisons difficult. More recent endeavors appear to make useful predictions on network dynamics by using Fokker-Planck equations to compute distributions of the membrane voltage as a function of time [5]. Here, we apply a similar formalism to the description of network dynamics with heterogeneous and stochastic synaptic weights to probe the distribution over network states $P_{\Phi}(R)$ and describe how this framework can be used to provide insights on how the neural networks embody the distribution $P_{\Phi}(R|X)$ in statistical inference tasks. To conclude, we discuss the effect of synaptic plasticity.

By solving the Fokker-Planck equations numerically for the time evolution of the distribution over membrane voltages $P_{\Phi}(V, t)$ simultaneously permits knowledge of the distribution $P_{\Phi}(\mathbf{z}, t)$. To see this, consider thresholding a sample from the joint distribution $\mathbf{z} \sim \Theta(\mathbf{v})$ where $\mathbf{v} \sim (P_{\Phi}(V, t))^N$. We might then ask how local synaptic plasticity rules permit the formation of cell assemblies which collectively encode stimuli.

CHAPTER 2

STATISTICAL MECHANICS OF A POPULATION OF NEURONS

In this section we make use of Fokker-Planck equations from statistical mechanics in the description of neural dynamics. A natural example is to describe the voltage dynamics of a single leaky integrate and fire (LIF) neuron that receives N feedforward Poisson input spike trains. The formalism presented can then be generalized to a population of LIF neurons.

2.1 Formalism

We consider two main populations of neurons: an external population and a recurrent population. The input neurons are described only by their spiking activity and are symbolized by the binary vector $\mathbf{s}(t) \in \mathbb{F}_2^n$. Furthermore, we describe the *state* variables of the recurrent population by a vector of voltages $\mathbf{v}(t) \in \mathbb{R}^{n \times 1}$ and a vector of so-called observable states denoted by $\mathbf{z}(t) = \Theta(\mathbf{v}(t))$ where $\Theta(v(t)) = H(v(t) - \theta)$ and H is the Heaviside step function. Input neurons communicate to the recurrent population through excitatory connections summarized by the input weight matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$. Similarly, a recurrent weight matrix $\mathbf{J} \in \mathbb{R}^{n \times n}$ specifies the recurrent network connectivity. The matrix \mathbf{J} may contain both excitatory and inhibitory connections. Therefore, it will prove useful to partition \mathbf{X} and \mathbf{J} into four quadrants as follows

$$\mathbf{X} = \begin{bmatrix} X_E \\ X_I \end{bmatrix} \qquad \mathbf{J} = \begin{bmatrix} J_{EE} & J_{IE} \\ J_{EI} & J_{II} \end{bmatrix} \qquad (2.1)$$

and an appropriate partitioning scheme is applied to the network state variables

$$\mathbf{v} = \begin{bmatrix} v_E \\ v_I \end{bmatrix} \qquad \mathbf{z} = \begin{bmatrix} z_E \\ z_I \end{bmatrix} \qquad (2.2)$$

which are defined such that, for example, the inputs from the external and recurrent population to an excitatory neuron $j \in \{1, 2, \dots, n' - 1, n'\}$ with $n' \leq n$ is simply $\xi_j(t) = J_{EE}z_E + J_{IE}z_I$. More generally, the voltage state variable for an arbitrary neuron evolves according to the update equation for the voltage

$$\begin{aligned} v_j(t + \Delta t) &= v_j(t) \exp\left(-\frac{\Delta t}{\tau}\right) + \xi_j(t) \\ &= v_j(t) \exp\left(-\frac{\Delta t}{\tau}\right) + \sum_{k \neq \ell} X_{k\ell} s_k(t) + \sum_{i \neq j} J_{ij} z_i(t) \end{aligned}$$

Letting the leakage term $\alpha = \exp\left(-\frac{\Delta t}{\tau}\right)$ this can be written for the entire population as $\mathbf{v}(t) = \alpha \mathbf{v}(t) + \mathbf{X}(t) \cdot \mathbf{s} + \mathbf{J}(t) \cdot \mathbf{z}$ where \cdot denotes standard matrix multiplication.

The input spike train $\mathbf{s}_k(t)$ of neuron k is taken to be a heterogeneous Poisson process with firing rate $\nu_k(t)$. Given the nature of such a stochastic process, a fundamental description of network dynamics demands knowledge of the time evolution of the voltage distribution $P(v, t)$ over the population, which is the focus of the following sections.

2.2 The Fokker-Planck Equation

The quantitative discussion of the dynamics of many complex systems (systems where the number of interacting units N is large) in nature from networks of spiking neurons, geophysical systems, to excitable media, all necessarily require a statistical description. Indeed, enumerating the available states to such a system itself has proven intractable, even for small

systems over short time scales. As an example, consider the states of an interacting system of N binary variables denoted $\{z_i\}_{i=1}^N$ which might be physically realized as an ensemble of spins in a ferromagnet. Even for extremely small cases such as $N = 100$ the system can take on $2^{100} = 1.26 \times 10^{30}$ different configurations and by $N = 300$ the number of configurations exceed our best estimates for the number of atoms in the known universe. At the same time, we cannot hope to make enough measurements of such a system to estimate the probability distribution over the space of states and make inferences on the organization and interactions of the individual elements based on such a distribution. We can, however, develop model distributions over the available states based on stochastic interaction between the individual units that account for the degrees of freedom which we cannot approach analytically - a technique often employed in the description of the statistical physics of particles. This family of techniques, often formally referred to as Langevin dynamics, is defined by the use of stochastic differential equations to model the evolution of molecular systems. A common example is the description of molecular diffusion where a particle of interest is immersed in a liquid or a gas and is subject to stochastic thermal fluctuations as well as frictional forces. One can write down a statistical description of the motion of a particle in such a scenario, yielding insights into the possible spatial trajectories it may follow in time. An alternative view, pioneered by Adriaan Fokker and later by Max Planck, utilizes a probabilistic framework and yields the so-called Fokker-Planck equations which allow us to solve for the time evolution of the probability distribution of such a stochastic variable. In the context of neuroscience, this approach has been taken by many in describing the dynamics of the membrane potential distribution in a population of neurons. In the following paragraphs, we will sketch a derivation of the Fokker-Planck equation for a diffusion process in N dimensions and we will later see the power of this description in describing the time evolution of the joint distribution over the membrane voltage in a spiking neural network.

Our objective is to determine the time evolution of the joint probability distribution over

N variables written as $P(\{v_i\}, t)$. If we make a basic assumption that the evolution of P follows a Markov process i.e. its evolution has the memoryless property, then we can write

$$P(v, t) = \int P(v, t|v', t - \tau)P(v', t - \tau)dv \quad (2.3)$$

which is known as the Chapman-Kolmogorov equation. The term $P(\{v_i\}, t|\{v_i\}')$ is known as the *transition operator* in a Markov process. We can rewrite the transition operator as

$$\begin{aligned} P(v, t|v', t - \tau) &= \int \delta(u - v)P(u, t|v', t - \tau)du \\ &= \int \delta(v' - v + u - v')P(u, t|x', t - \tau)du \\ &= \int \sum_{n=0}^{\infty} \frac{(y - x')^n}{n!} \left(-\frac{\partial}{\partial x}\right)^n P(u, t|x', t - \tau)\delta(x' - x)du \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial x}\right)^n \delta(x' - x) \int (u - x')^n P(u, t|x', t - \tau)du \end{aligned}$$

Noticing that $\int (u - x')^n P(u, t|x', t - \tau)du$ is just the n th moment of the distribution $P(u, t|x', t - \tau)$, we have

$$P(v, t|v', t - \tau) = \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial v}\right)^n M_n(v, t, \tau)\right) \delta(v' - v) \quad (2.4)$$

Plugging this back in to (2.3) gives

$$P(v, t) = \int \left(1 + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial v}\right)^n M_n(v, t, \tau)\right) \delta(v' - v)P(v', t - \tau)dv \quad (2.5)$$

$$= P(v', t - \tau) + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial v}\right)^n M_n(v, t, \tau) \quad (2.6)$$

Approximating the derivative as a finite difference and sending $\tau \rightarrow 0$ gives

$$\frac{\partial P(v, t)}{\partial t} = \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial v} \right)^n M_n(v, t, \tau) P(v, t) \quad (2.7)$$

$$= \mathcal{L}_{\text{KM}} P(v, t) \quad (2.8)$$

with $\mathcal{L}_{\text{KM}} = \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial v} \right)^n M_n$ which is the Kramers-Moyal (KM) expansion. The Fokker-Planck equation is an approximation of the KM expansion in (2.7) which neglects terms $n \geq 3$ in the so-called diffusion approximation. Recalling that M_n represents the moments of the distribution over changes in the voltage of the population suggests that we can write the first two moments as

$$M_1 = \mu(t) = NA(t)\langle J \rangle - \frac{v - v_L}{\tau} + \mu_{ext}$$

$$M_2 = NA(t)\langle J^2 \rangle = \frac{\sigma^2(t)}{\tau} + \sigma_{ext}^2$$

with $A(t) = \int_{\theta}^{\infty} P(v, t) dv$ which represent a drift term and diffusion term and we arrive at the following Fokker-Planck equation

$$\frac{\partial P(v, t)}{\partial t} = \left(\mu(t) - \frac{v - v_L}{\tau} \right) \frac{\partial}{\partial v} P(v, t) - \frac{\sigma^2(t)}{2\tau} \frac{\partial^2}{\partial v^2} P(v, t)$$

which we approximate by central finite differences

$$\begin{aligned} \frac{p(v, t + \Delta t) - p(v, t)}{\Delta t} = & \left(N\xi - \frac{v - v_L}{\tau} + \mu_{ext} \right) \frac{p(v + \Delta v, t) - p(v, t)}{\Delta v} \\ & - \frac{1}{2} \left(\frac{N\xi(1 - \xi)}{\tau} + \sigma_{ext}^2 \right) \frac{p(v + \Delta v, t) - 2p(v, t) + p(v - \Delta v, t)}{\Delta v^2} \end{aligned}$$

with $\xi = \sum_{p_i > \theta} p_i(v, t)$

CHAPTER 3

DYNAMICAL STATES OF RECURRENT NETWORKS

3.1 Introduction

CHAPTER 4

STOCHASTIC COMPUTATION BY RECURRENT NETWORKS

4.1 Introduction

CHAPTER 5

CONCLUSIONS

- [1] Mculloch and Pitts. *A logical calculus of the ideas imminent in nervous activity*. Journal of Mathematical Biophysics. 1943.
- [2] J.J. Hopfield. *Neural networks and physical systems with emergent collective computational abilities*. 1982.
- [3] D.J. Amit. *Spin-glass models of neural networks*. Physical Rev A. 1985.
- [4] E. Gardner. *The space of interactions in neural network models*. Journal of Physics A: Mathematical and General. 1988.
- [5] N. Brunel. *Dynamics of sparsely connected networks of excitatory and inhibitory neurons*. Journal of Computational Neuroscience. 2000.
- [6] Rosenbaum and Doiron. *Balanced Networks of Spiking Neurons with Spatially Dependent Recurrent Connections*. Physical Review X. 2014.
- [7] Zenke et al. *Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks*. Nature Communications. 2015.
- [8] Williams et al. *Nonnegative decomposition of multivariate information*. arXiv. 2010.
- [9] Amit et al. *Storing infinite numbers of patterns in a spin-glass model of neural networks*. Physical Rev. Letters. 1985
- [10] Nishimori. *Statistical physics of spin glasses and information processing: an introduction*. Clarendon Press. 2010
- [11] Buesing et al. *Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons*. PLOS Computational Biology. 2011
- [12] Pecevski et al. *Learning Probabilistic Inference through Spike-Timing-Dependent Plasticity*. eNeuro. 2016
- [13] Pecevski et al. *Formation and maintenance of neuronal assemblies through synaptic plasticity*. Nature Communications. 2014
- [14] Pecevski et al. *A learning algorithm for Boltzmann Machines*. Cognitive Science. 1985