

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 STOCHASTIC SYSTEMS AND FOKKER-PLANCK EQUATIONS	3
2.1 The Langevin Equation	3
2.1.1 The Kramers-Moyal expansion	3
2.1.2 The Ornstein-Uhlenbeck Process	5
2.2 Fokker-Planck for Homogeneous Sparse Networks	6
3 DYNAMICAL STATES OF RECURRENT NETWORKS	10
3.1 Introduction	10
4 STOCHASTIC COMPUTATION BY RECURRENT NETWORKS	11
4.1 Introduction	11
5 CONCLUSIONS	12
REFERENCES	13

ACKNOWLEDGMENTS

ABSTRACT

The primate cerebral cortex is a complex system estimated to harbor more than 25 billion neurons communicating via action potentials or ‘spikes’ and is responsible for many higher-order brain functions including memory and learning. Recent years have hosted many efforts to understand how such complex phenomena emerge from the communication of individual cells. Many studies have provided evidence that long term plasticity (LTP) in synapses permits a long-lasting alteration of network dynamics and, in turn, forms the basis of long-term memory and learning. However, understanding memory formation and learning in the brain is made difficult by the variability in the response of cortical neurons to stimuli. Therefore, capturing the apparent stochastic features of neural activity in computer based models, such as recurrent spiking neural networks (RSNNs), while explaining their manipulation of information mathematically has become the gold standard for computational neuroscience. Models of neural networks derived from statistical mechanics, such as those which assert that the membrane potential of a cortical neuron obeys a form of Langevin dynamics, can potentially account for stochastic network activity. Such models also provide the intriguing interpretation that neural activity represents sampling from a probability distribution - a technique central to statistical inference. Here, we apply a similiar mathematical treatment to the study of an RSNN by modeling the membrane potential statistics of an integrate and fire neuron using Fokker-Planck equations. With this statistical framework in hand, we can recast a network of neurons as a stochastic process in higher dimensions and explore the relationships between synaptic connectivity and its plasticity to the correlation structure of neural spike trains. This approach is also amenable to information theoretic analysis and is a step toward a mathematical relationship between neuroplasticity mechanisms and the emergent computational capabilities of cortical microcircuits.

CHAPTER 1

INTRODUCTION

Towards 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. A popular approach towards the end of 20th century simplified a network of neurons firing action potentials to an ensemble of coupled spins obeying an energy function over the space of states. , 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.

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 systems with high degrees of freedom. A Fokker-Planck equation allows us to solve for the time evolution of the probability distribution over such a variable, providing insights into the dynamics which cannot be seen from any one trajectory through the space of states.

CHAPTER 2

STOCHASTIC SYSTEMS AND FOKKER-PLANCK EQUATIONS

2.1 The Langevin Equation

In the context of neuroscience, the Fokker-Planck 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 general stochastic process in one dimension and apply this result to networks of integrate and fire neurons.

2.1.1 The Kramers-Moyal expansion

Given many instantiations of a stochastic variable V , we can construct a normalized histogram over all observations as a function of time $P(V, t)$. However, in order to systematically explore the relationship between the parameterization of the process and $P(V, t)$ we require an expression for $\dot{P}(V, t)$. If we make a fundamental assumption that the evolution of $P(V, t)$ follows a Markov process i.e. its evolution has the memoryless property, then we can write

$$P(V', t) = \int T(V', t|V, t - \tau)P(V, t - \tau)dV \quad (2.1)$$

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

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

If we use the Taylor expansion of the δ -function

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

Inserting this into the result from above, pulling out terms independent of u and swapping the order of the sum and integration gives

$$T(V', t|V, t - \tau) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial V} \right)^n \delta(V - V') \int (u - V)^n T(u, t|V, t - \tau) du \quad (2.2)$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial V} \right)^n \delta(V - V') M_n(V, t) \quad (2.3)$$

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

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

$$= P(V', t - \tau) + \sum_{n=1}^{\infty} \frac{1}{n!} \left(-\frac{\partial}{\partial V} \right)^n [M_n(V, t) P(V, t)] \quad (2.5)$$

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

$$\dot{P}(V, t) = \lim_{\tau \rightarrow 0} \left(\frac{P(V, t) - P(V, t - \tau)}{\tau} \right) \quad (2.6)$$

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

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*.

2.1.2 The Ornstein-Uhlenbeck Process

To illustrate the power of the Fokker-Planck equations, we will consider a generic Ornstein-Uhlenbeck process - a natural starting point for its similarity to simplified neuron models. Consider the following Langevin equation representing a stochastic process with drift

$$\dot{V}(t) = -\frac{V(t) - V_r}{\tau} + \sigma \eta(t)$$

with linear drift back to a resting state V_r according to the rate constant τ and noise amplitude σ . To illustrate the applicability of (2.10) in predicting the distribution $P(V, t)$, we consider the case where $\eta(t)$ is a 1D Gaussian white noise. We have the following Fokker-Planck equation

$$\dot{P}(V, t) = \frac{1}{\tau} \frac{\partial}{\partial V} [(V(t) - V_r) P(V, t)] + \frac{\sigma^2}{2} \frac{\partial^2}{\partial V^2} [P(V, t)]$$

The transition density $T(V', t' | V, t)$ has a known analytical form:

$$T(V', t + \Delta t | V, t) = \sqrt{\frac{1}{\tau\pi\sigma^2(1 - \exp(-2\Delta t/\tau))}} \exp\left(-\frac{1}{\tau\sigma^2} \frac{(V' - V \exp(-\Delta t/\tau))^2}{1 - \exp(-2\Delta t/\tau)}\right)$$

Plugging this into (2.1) and setting the initial condition $P(V, t) = \delta(V - V_R)$ gives

$$\begin{aligned} P(V', t) &= \sqrt{\frac{1}{\tau\pi\sigma^2(1 - \exp(-2\Delta t/\tau))}} \exp\left(-\frac{1}{\tau\sigma^2} \frac{(V' - V \exp(-\Delta t/\tau))^2}{1 - \exp(-2\Delta t/\tau)}\right) \delta(V - V_R) dV \\ &= \sqrt{\frac{1}{\tau\pi\sigma^2(1 - \exp(-2\Delta t/\tau))}} \exp\left(-\frac{1}{\tau\sigma^2} \frac{(V' - V_R \exp(-\Delta t/\tau))^2}{1 - \exp(-2\Delta t/\tau)}\right) \delta(V - V_R) dV \end{aligned}$$

$$\begin{aligned} P(V, t + \Delta t) &= P(V, t) + \Delta t \left[\alpha t \cdot \frac{P(V + \Delta V, t) - P(V, t)}{\Delta V} \right] \\ &\quad - \Delta t \left[\frac{1}{2} \frac{P(V + \Delta V, t) - 2P(V, t) + P(V - \Delta V, t)}{\Delta V^2} \right] \end{aligned}$$

2.2 Fokker-Planck for Homogeneous Sparse Networks

The moments $M_n(t)$ derived above in the Kramers-Moyal expansion are dependent on the connectivity of the network and the statistics of the input. We will first consider a classic case, where we have a sparse directed network with constant synaptic efficacy between all presynaptic and postsynaptic pairs of cells. This primary pool of neurons is subject to stimulation by N_{in} input neurons with a connection probability γ_{in} giving $C_{\text{in}} = \gamma_{\text{in}} N_{\text{rec}}$ unique connections between the input population and a single neuron in the recurrent pool. Within the recurrent pool, we have a connection probability of $\gamma_{\text{rec}} \ll 1$ for any pair giving

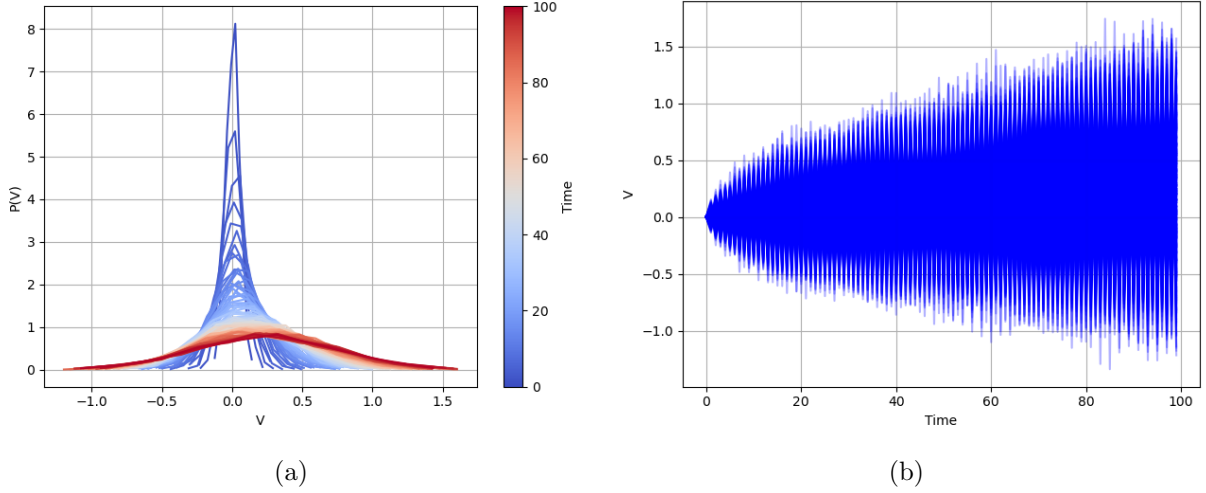


Figure 2.1: Results of numerical integration of a Langevin equation for a gaussian white noise process with $J = 1$, $\tau = 1$, $\alpha = 0.01$, and $V(0) = 0$. (a) Distribution of the stochastic variable over $N = 1000$ units as a function of time. (b) Time course for $N = 1000$ units

$C_{\text{rec}} = \gamma_{\text{rec}} N_{\text{rec}}$ recurrent inputs per postsynaptic cell. We assume that a presynaptic action potential invokes a post synaptic potential (PSP) with magnitude J_0 in the postsynaptic cell with $J_0 \ll \theta$ for both input and recurrent projections. The first moment of the transition operator $T(V', t|V, t - \tau)$ is given by

$$\begin{aligned}
 \mu(t) &= \mu_{\text{in}} + \mu_{\text{rec}}(t) \\
 &= (C_{\text{in}} \nu_{\text{in}}(t) + C_{\text{rec}} \nu_{\text{rec}}(t)) \tau \langle J \rangle \\
 &= (C_{\text{in}} \nu_{\text{in}}(t) + C_{\text{rec}} \nu_{\text{rec}}(t)) \tau J_0
 \end{aligned}$$

where $\nu_{\text{in}}(t)$ is the rate parameter for the input Poisson process. The second moment

$$\begin{aligned}
\sigma^2(t) &= \sigma_{\text{rec}}^2 + \sigma_{\text{ext}}^2 \\
&= (C_{\text{in}}\nu_{\text{in}}(t) + C_{\text{rec}}\nu_{\text{rec}}(t)) \tau \langle J^2 \rangle \\
&= (C_{\text{in}}\nu_{\text{in}}(t) + C_{\text{rec}}\nu_{\text{rec}}(t)) \tau J_0^2
\end{aligned}$$

After inserting the first two moments into (2.10) we arrive at the following Fokker-Planck equation

$$\dot{P}(V, t) = -\frac{\partial}{\partial V}[\mu(t)P(V, t)] + \frac{1}{2}\frac{\partial^2}{\partial V^2}[\sigma^2(t)P(V, t)]$$

At this point, it is necessary to impose the appropriate boundary conditions on the above Fokker-Planck equation as in so as to maintain biological realism. The Fokker-Planck equation can be written as the continuity equation

$$\frac{\partial P(v, t)}{\partial t} = -\frac{\partial S}{\partial V}$$

(Risken, 1984) with

$$S(V, t) = -\frac{v - V_L - \mu}{\tau}P(V, t) - \frac{\sigma^2(t)}{2\tau}\frac{\partial P(V, t)}{\partial V}$$

which is known as the *probability current* through voltage V at a time t . The instantaneous firing rate is equivalent to the probability current through the threshold i.e. $\nu(t) = S(\theta, t)$. Furthermore, we require that the probability current through the firing threshold $P(\theta, t) = 0$ and that instead this probability emerges at the resting potential after a refractory period of τ_{ref} . This condition gives the following boundary condition for the

derivative of the probability with respect to voltage

$$\frac{\partial P(\theta, t)}{\partial V} = -\frac{2\tau\nu(t)}{\sigma^2(t)}$$

To account for the refractory period τ_{ref} , we define an auxiliary distribution

$$p_r(t) = \int_{t-\tau_{\text{ref}}}^t \nu(t) dt$$

which together with the distribution $P(V, t)$ satisfy the normalization condition:

$$\int P(v, t) dV + p_r(t) = 1$$

$$\frac{\partial P(v, t)}{\partial t} = \left(\mu(t) - \frac{v - v_L}{\tau} \right) \frac{\partial}{\partial v} P(v, t) - \frac{\sigma^2(t)}{2\tau} \frac{\partial^2}{\partial v^2} P(v, t) + \nu(t - \tau_{\text{ref}}) \delta(v - V_R)$$

which we approximate by central finite differences

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

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