

8-31-2024

## Large deviation theory in stochastic processes: applications to biological modeling

Moshe C. Silverstein

*New Jersey Institute of Technology*, [moshechayimsilverstein@gmail.com](mailto:moshechayimsilverstein@gmail.com)

Follow this and additional works at: <https://digitalcommons.njit.edu/dissertations>



Part of the [Biological Phenomena, Cell Phenomena, and Immunity Commons](#), [Neuroscience and Neurobiology Commons](#), [Other Applied Mathematics Commons](#), and the [Statistics and Probability Commons](#)

---

### Recommended Citation

Silverstein, Moshe C., "Large deviation theory in stochastic processes: applications to biological modeling" (2024). *Dissertations*. 1779.

<https://digitalcommons.njit.edu/dissertations/1779>

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at Digital Commons @ NJIT. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital Commons @ NJIT. For more information, please contact [digitalcommons@njit.edu](mailto:digitalcommons@njit.edu).

## **Copyright Warning & Restrictions**

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be “used for any purpose other than private study, scholarship, or research.” If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of “fair use” that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

**Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation**

Printing note: If you do not wish to print this page, then select “Pages from: first page # to: last page #” on the print dialog screen

The Van Houten library has removed some of the personal information and all signatures from the approval page and biographical sketches of theses and dissertations in order to protect the identity of NJIT graduates and faculty.

## ABSTRACT

### LARGE DEVIATION THEORY IN STOCHASTIC PROCESSES: APPLICATIONS TO BIOLOGICAL MODELING

by  
**Moshe C. Silverstein**

This dissertation delves into developing and applying stochastic models to analyze complex biological systems. It leverages Large Deviation Theory (LDT) to gain insights into these systems, focusing on two key examples: neural networks and calcium signaling dynamics. Traditional deterministic methods frequently fail to capture biological processes' randomness and inherent variability. Meanwhile, many stochastic approaches struggle to be mathematically tractable or provide accessible insights. The approach introduced in this study provides rigorous mathematical frameworks to enhance understanding of these stochastic behaviors while remaining tractable and insightful.

A stochastic model for a random biological neural network is constructed that addresses the dependencies and variabilities in neural connectivity. Applying LDT, significant theoretical results are derived from the large deviations in the system's dynamics, providing a deeper understanding of the probabilistic behaviors and events in neural activity.

The study next focuses on calcium signaling in biological cells, where a one-dimensional stochastic model is developed to simulate calcium dynamics. A Piecewise-deterministic Markov process (PDMP) model is implemented to capture the system's stochastic and deterministic nature. This model is validated by comparing experimental data from *in vitro* and *in vivo* studies via Maximum Likelihood Estimation and stochastic simulations. LDT is used to derive the Euler-Lagrange equations and identify optimal trajectories in calcium signaling, offering predictive insights into the system's behavior under stochastic mechanisms.

The findings of this study demonstrate the power of LDT in biological modeling, providing a robust framework for analyzing the probabilistic nature of complex biological systems. While the models incorporate several simplifications, such as one-dimensional assumptions in calcium signaling, they pave the way for more sophisticated and accurate representations of biological processes.

This work advances the application of stochastic processes and LDT in mathematical biology, offering enhancements to methodologies and insights that can be extended to other complex systems. The proposed approach opens up new avenues for understanding and predicting the behavior of stochastic biological systems, with potential applications in fields such as neuroscience, cell biology, and systems biology.

**LARGE DEVIATION THEORY IN STOCHASTIC PROCESSES:  
APPLICATIONS TO BIOLOGICAL MODELING**

by  
Moshe C. Silverstein

A Dissertation  
Submitted to the Faculty of  
New Jersey Institute of Technology and  
Rutgers, The State University of New Jersey—Newark  
in Partial Fulfillment of the Requirements for the Degree of  
Doctor of Philosophy in Mathematical Sciences

Department of Mathematical Sciences, NJIT  
Department of Mathematics and Computer Science, Rutgers-Newark

August 2024

Copyright © 2024 by Moshe C. Silverstein

ALL RIGHTS RESERVED

**APPROVAL PAGE**

**LARGE DEVIATION THEORY IN STOCHASTIC PROCESSES:  
APPLICATIONS TO BIOLOGICAL MODELING**

**Moshe C. Silverstein**

---

Dr. James MacLaurin, Dissertation Advisor Date  
Assistant Professor of Mathematics, NJIT

---

Dr. Amitabha Bose, Committee Member Date  
Professor of Mathematics, NJIT

---

Dr. Victor Matveev, Committee Member Date  
Professor of Mathematics, NJIT

---

Dr. David Shirokoff, Committee Member Date  
Associate Professor of Mathematics, NJIT

---

Dr. Etienne Tanre, Committee Member Date  
Researcher, Inria Center of Côte d'Azur University, Valbonne, France

## BIOGRAPHICAL SKETCH

**Author:** Moshe C. Silverstein

**Degree:** Doctor of Philosophy

**Date:** August 2024

### Undergraduate and Graduate Education:

- Doctor of Philosophy in Mathematical Sciences  
New Jersey Institute of Technology, Newark, NJ, 2024
- Master of Arts in Physics,  
Hunter College, CUNY, New York, NY, 2016
- Bachelor of Science, Physics  
Brooklyn College, CUNY, Brooklyn, NY, 2013

**Major:** Mathematical Sciences

### Publications:

- T. Asakura, K. Isobe, S. Kametani, O. T. Ukpebor, M. C. Silverstein, and G. S. Boutis. Characterization of water in hydrated bombyx mori silk fibroin fiber and films by 2h nmr relaxation and 13c solid state nmr. *Acta Biomaterialia*, 50:322–333, 2017.
- G. Barbet, J. MacLaurin, and M. Silverstein. Large deviations of piecewise-deterministic-markov-processes with application to stochastic calcium waves. (arXiv:2406.12493), June 2024.
- K. Bilici, S. W. Morgan, M. C. Silverstein, Y. Wang, H. J. Sun, Y. Zhang, and G. S. Boutis. Mechanical, structural, and dynamical modifications of cholesterol exposed porcine aortic elastin. *Biophysical Chemistry*, 218:47–57, 2016.
- D. J. B. Clarke, M. V. Kuleshov, B. M. Schilder, D. Torre, M. E. Duffy, A. B. Keenan, A. Lachmann, A. S. Feldmann, G. W. Gundersen, M. C. Silverstein, et al. Expression2kinases (x2k) web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Research*, 46(W1):W171–W179, 2018.
- D. J. Clarke, L. Wang, A. Jones, M. L. Wojciechowicz, D. Torre, K. M. Jagodnik, S. L. Jenkins, P. McQuilton, Z. Flamholz, M. C. Silverstein, et al. Fairshake: toolkit to evaluate the fairness of research digital resources. *Cell Systems*, 9(5):417–421, 2019.

- K. T. Downing, M. Billah, E. Raparia, A. Shah, M. C. Silverstein, A. Ahmad, and G. S. Boutis. The role of mode of delivery on elastic fiber architecture and vaginal vault elasticity: a rodent model study. *Journal of the Mechanical Behavior of Biomedical Materials*, 29:190–198, 2014.
- A. B. Keenan, S. L. Jenkins, K. M. Jagodnik, S. Koplev, E. He, D. Torre, Z. Wang, A. B. Dohlman, M. C. Silverstein, A. Lachmann, et al. The library of integrated network-based cellular signatures nih program: system-level cataloging of human cells response to perturbations. *Cell Systems*, 6(1):13–24, 2018.
- A. Lachmann, D. Torre, A. B. Keenan, K. M. Jagodnik, H. J. Lee, L. Wang, M. C. Silverstein, and A. Ma’ayan. Massive mining of publicly available rna-seq data from human and mouse. *Nature Communications*, 9(1):1–10, 2018.
- J. MacLaurin, M. Silverstein, and P. Vilanova. Population level activity in large random neural networks. *arXiv preprint arXiv:2401.15272*, 2024.
- M. C. Silverstein, K. Bilici, S. W. Morgan, Y. Wang, Y. Zhang, and G. S. Boutis. 13c, 2h nmr studies of structural and dynamical modifications of glucose-exposed porcine aortic elastin. *Biophysical Journal*, 108(7):1758–1772, 2015.
- Z. Wang, E. He, K. Sani, K. M. Jagodnik, M. C. Silverstein, and A. Ma’ayan. Drug gene budger (dgb): an application for ranking drugs to modulate a specific gene based on transcriptomic signatures. *Bioinformatics*, 35(7):1247–1248, 2019.

### **Presentations:**

- M. C. Silverstein. A piecewise-deterministic markov process model for calcium signaling. Society for Industrial and Applied Mathematics (SIAM) New York-New Jersey-Pennsylvania Section, 2023.
- M. C. Silverstein. Balanced neural fields. Society for Industrial and Applied Mathematics (SIAM) Conference on the Life Sciences, 2022.
- M. C. Silverstein. Model reduction of large random neural networks. International Conference on Mathematical Neuroscience, 2022.
- M. C. Silverstein. Wandering of coherent structures in stochastic neural fields. Society for Industrial and Applied Mathematics (SIAM) Conference on Nonlinear Waves and Coherent Structures, 2024.

## בְּסִיעָתָא דְשִׁמְיָא

*He has faith in his ability to perform the miracle of  
comprehending the world.*

*- R. Joseph B. Soloveitchik*

*Dedicated to my family, past, present, and future.*

## ACKNOWLEDGMENTS

Blessed are You, LRD, our GD, Sovereign of the Universe, who has granted us life, sustained us, and enabled us to reach this moment.

I am deeply grateful to my advisor, Prof. James MacLaurin, for his invaluable guidance, unwavering support, and profound insights throughout my academic journey. His mentorship has been instrumental in shaping my understanding of applied mathematics and research. I am particularly grateful for his patience and dedication, which have been crucial in navigating the challenges of my program.

I express my sincere appreciation to the members of my committee, Profs. Amitabha Bose, Victor Matveev, David Shirokoff, and Etienne Tanre for their valuable contributions and insights that have enriched the quality of my work. Their expertise has not only directly impacted my research but has also served as a model for the kind of excellence that I aspire to achieve as an applied mathematician and researcher.

I extend my heartfelt thanks to my friends and colleagues, coHRT-19, in no particular order, Austin Juhl, Atul Anurag, Jake Brusca, Nick Dubicki, Jose Pabon, Sam Evans, Prianka Bose, and John Wu, who have been more than just companions on this journey. Your camaraderie, encouragement, and shared experiences have not only made this academic journey bearable but also memorable. You have been the constant reminder that success is sweeter when shared. I am deeply grateful for the camaraderie that has defined our time together. You guys inspire me.

To my parents, Avrohom and Sara Chana Silverstein, I am at a loss for words to adequately express the depth of my gratitude for your endless love, encouragement, and sacrifices. Your unwavering support, the bedrock upon which my academic pursuits have flourished, has been instrumental in my success. I am forever indebted to you for your ceaseless belief in my abilities and the countless ways you have shaped me into who I am today.

A heartfelt thank you goes to my life partner, my wife, Sarah Nessah Silverstein. Her unwavering support, understanding, and encouragement, which have been my pillars of strength, have been instrumental in my success. Her belief in me and her sacrifices throughout this journey has made this achievement possible. I am truly fortunate to have her by my side.

I express my profound appreciation to all those who have contributed to my academic and personal growth. This accomplishment is as much yours as it is mine.

## TABLE OF CONTENTS

| Chapter  | Page |
|--|------|
| 1 INTRODUCTION . . . . .                                   | 1    |
| 1.1 Overview and Motivation . . . . .                      | 1    |
| 1.2 Structure of the Dissertation . . . . .                | 3    |
| 2 MATHEMATICAL FOUNDATIONS . . . . .                       | 5    |
| 2.1 Stochastic Processes . . . . .                         | 5    |
| 2.1.1 Brownian Motion . . . . .                            | 5    |
| 2.1.2 Poisson Process . . . . .                            | 6    |
| 2.2 Stochastic Differential Equation . . . . .             | 7    |
| 2.3 Skorohod Space . . . . .                               | 7    |
| 2.3.1 Definition . . . . .                                 | 7    |
| 2.3.2 Topology . . . . .                                   | 8    |
| 2.3.3 Key theorems . . . . .                               | 9    |
| 2.4 Empirical Measure . . . . .                            | 10   |
| 2.5 Piecewise Deterministic Markov Processes . . . . .     | 11   |
| 2.6 Radon-Nikodym Theorem and Derivative . . . . .         | 13   |
| 2.6.1 Radon-Nikodym Theorem . . . . .                      | 13   |
| 2.6.2 Properties of the Radon-Nikodym derivative . . . . . | 13   |
| 2.7 Relative Entropy . . . . .                             | 14   |
| 2.8 Large Deviation Theory . . . . .                       | 15   |
| 2.8.1 Large Deviation Principle . . . . .                  | 16   |
| 2.8.2 Cramer's Theorem . . . . .                           | 16   |
| 2.8.3 Varadhan's Theorem . . . . .                         | 17   |

**TABLE OF CONTENTS**  
(Continued)

| <b>Chapter</b>  | <b>Page</b> |
|---|-------------|
| 2.8.4 Bryc’s Theorem . . . . .                                    | 18          |
| 2.8.5 Sanov’s Theorem . . . . .                                   | 19          |
| 2.8.6 Freidlin-Wentzell Theory . . . . .                          | 19          |
| <b>3 BIOLOGICAL CONTEXTS AND APPLICATIONS . . . . .</b>           | <b>22</b>   |
| 3.1 Neuroscience . . . . .  | 22          |
| 3.1.1 Physiology . . . . .  | 22          |
| 3.1.2 Overview of historical mathematical modeling . . . . .      | 23          |
| 3.2 Calcium Signaling . . . . .                                   | 27          |
| 3.2.1 Physiology . . . . .  | 27          |
| 3.2.2 Overview of historical mathematical modeling . . . . .      | 29          |
| <b>4 NEURAL NETWORK MODEL . . . . .</b>                           | <b>31</b>   |
| 4.1 Notation and Definitions . . . . .                            | 32          |
| 4.2 Model Description . . . . .                                   | 33          |
| 4.3 Initial Condition Assumptions . . . . .                       | 34          |
| 4.4 Main Theoretical Results . . . . .                            | 34          |
| <b>5 LARGE DEVIATION PRINCIPLES FOR NEURAL NETWORKS . . . . .</b> | <b>36</b>   |
| 5.1 Large Deviations of the Uncoupled System . . . . .            | 36          |
| 5.2 Large Deviations of the Coupled System . . . . .              | 38          |
| 5.3 Regularity Estimates and Compactness . . . . .                | 39          |
| 5.4 Exponential Tightness . . . . .                               | 47          |
| 5.5 Proof of Large Deviations of the Uncoupled System . . . . .   | 48          |
| 5.6 Proof of Large Deviations of the Coupled System . . . . .     | 52          |

**TABLE OF CONTENTS**  
(Continued)

| <b>Chapter</b>   | <b>Page</b> |
|--|-------------|
| 6 CALCIUM SIGNALING MODEL . . . . .                          | 65          |
| 6.1 Model Description . . . . .                              | 65          |
| 6.2 Maximum Likelihood Estimation . . . . .                  | 69          |
| 6.3 Model Simulations . . . . .                              | 70          |
| 6.4 In Vitro Experimental Data Analysis . . . . .            | 71          |
| 6.5 In Vitro Numerical Results . . . . .                     | 78          |
| 6.6 In Vivo Experimental Data Analysis . . . . .             | 87          |
| 6.7 In Vivo Numerical Results . . . . .                      | 90          |
| 7 LARGE DEVIATION PRINCIPLES FOR CALCIUM SIGNALING . . . . . | 97          |
| 7.1 Large Deviation of PDMP . . . . .                        | 97          |
| 7.1.1 Notation and Definitions . . . . .                     | 97          |
| 7.1.2 Large Deviation Principle for PDMPs . . . . .          | 98          |
| 7.1.3 Topological Definitions . . . . .                      | 101         |
| 7.1.4 Euler-Lagrange Equations . . . . .                     | 103         |
| 7.1.5 Simplified Euler-Lagrange Equations . . . . .          | 107         |
| 7.2 Application to Calcium Signaling . . . . .               | 109         |
| 7.2.1 PDMP Model of Calcium Dynamics . . . . .               | 110         |
| 7.2.2 Large N Limiting Dynamics . . . . .                    | 111         |
| 7.2.3 Euler-Lagrange Equations . . . . .                     | 112         |
| 7.2.4 Optimal Trajectories . . . . .                         | 114         |
| 7.2.5 Numerical Results . . . . .                            | 115         |
| 8 CURRENT AND FUTURE WORK . . . . .                          | 118         |

**TABLE OF CONTENTS**  
(Continued)

| <b>Chapter</b>  | <b>Page</b> |
|---|-------------|
| 9 CONCLUSION . . . . .                                  | 120         |
| APPENDIX A BOUNDING FLUCTUATIONS OF THE NOISE . . . . . | 122         |
| APPENDIX B PROOFS FOR LDP FOR PDMP . . . . .            | 127         |
| REFERENCES . . . . .                                    | 136         |

## LIST OF TABLES

| <b>Table</b> |   | <b>Page</b> |
|--------------|---|-------------|
| 6.1          | Data Statistics for the In Vitro Data Analyzed . . . . .                | 72          |
| 6.2          | Parameter Values From MLE Results . . . . .                             | 78          |
| 6.3          | Statistics Comparing the Experimental and Simulated Data . . . . .      | 83          |
| 6.4          | 16th, 50th and 84th Percentile Parameter Values From the MCMC . . . . . | 84          |
| 6.5          | Data Statistics for the In Vivo Data Analyzed . . . . .                 | 87          |
| 6.6          | Parameter Values From MLE Results . . . . .                             | 91          |
| 6.7          | Statistics Comparing the Experimental and Simulated Data . . . . .      | 94          |
| 6.8          | 16th, 50th and 84th Percentile Parameter Values From the MCMC . . . . . | 95          |

## LIST OF FIGURES

| Figure  | Page |
|---|------|
| 6.1 Example of system events. . . . .   | 70   |
| 6.2 Fluorescent imaging. . . . .  | 71   |
| 6.3 Calcium concentration spiking in response to E.Coli DH5a. Left: Live pathogen. Right: HK pathogen. . . . .                      | 72   |
| 6.4 Calcium concentration spiking in response to E.Coli MC4100. Left: Live pathogen. Right: HK pathogen. . . . .                    | 72   |
| 6.5 Calcium concentration spiking in response to Listeria. Left: Live pathogen. Right: HK pathogen. . . . .                         | 73   |
| 6.6 Example of data preprocessing. . . . .  | 74   |
| 6.7 Example of peak selection. Spikes selected within 1, 2, and 3 standard deviations (SDV) from the mean. . . . .                  | 74   |
| 6.8 E.Coli DH5a. Left ISI distribution. Right, Spike counts across samples. .   | 75   |
| 6.9 E.Coli MC4100. Left ISI distribution. Right, Spike counts across samples.   | 75   |
| 6.10 Listeria. Left ISI distribution. Right, Spike counts across samples. . . .   | 76   |
| 6.11 Power spectral density (left), frequency distribution (center) and violin plot (right) for E.Coli DH5a. . . . .                | 76   |
| 6.12 Power spectral density (left), frequency distribution (center) and violin plot (right) for E.Coli MC4100. . . . .              | 77   |
| 6.13 Power spectral density (left), frequency distribution (center), and violin plot (right) for Listeria. . . . .                  | 77   |
| 6.14 Power spectral density (left), frequency distribution (center), and violin plot (right) for DH5a compared to Listeria. . . . . | 77   |
| 6.15 Power spectral density (left), frequency distribution (center), and violin plot (right) for MC4100 compared to DH5a. . . . .   | 78   |

**LIST OF FIGURES**  
(Continued)

| <b>Figure</b>  | <b>Page</b> |
|--|-------------|
| 6.16 Simulation (orange) versus Experimental (blue) ISI distributions for Dh5a. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters $\alpha_o$ and $\alpha_{cl}$ . . . . .     | 79          |
| 6.17 Simulation (orange) versus Experimental (blue) ISI distributions for MC4100. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters $\alpha_o$ and $\alpha_{cl}$ . . . . .   | 80          |
| 6.18 Simulation (orange) versus Experimental (blue) ISI distributions for Listeria. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters $\alpha_o$ and $\alpha_{cl}$ . . . . . | 81          |
| 6.19 Opening and closing probabilities for optimal parameters for DH5a. . . . .  | 85          |
| 6.20 Opening and closing probabilities for optimal parameters for MC4100. . . . .  | 86          |
| 6.21 Opening and closing probabilities for optimal parameters for Listeria. . . . .  | 86          |
| 6.22 Calcium concentration spiking in response to E.Coli DH5a. Left: Live pathogen. Right: HK pathogen. . . . .  | 88          |
| 6.23 Calcium concentration spiking in response to E.Coli MC4100. Left: Live pathogen. Right: HK pathogen. . . . .  | 88          |
| 6.24 E.Coli DH5a. Left ISI distribution. Right, Spike counts across samples. . . . .   | 89          |
| 6.25 Listeria. Left ISI distribution. Right, Spike counts across samples. . . . .  | 89          |
| 6.26 Power spectral density (left), frequency distribution (center) and violin plot (right) for E.Coli DH5a. . . . .   | 90          |
| 6.27 Power spectral density (left), frequency distribution (center), and violin plot (right) for Listeria. . . . .   | 90          |
| 6.28 Power spectral density (left), frequency distribution (center), and violin plot (right) for DH5a compared to Listeria. . . . .  | 90          |
| 6.29 Simulation (orange) versus Experimental (blue) ISI distributions for Dh5a. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters $\alpha_o$ and $\alpha_{cl}$ . . . . .     | 92          |

**LIST OF FIGURES**  
(Continued)

| <b>Figure</b>  | <b>Page</b> |
|--|-------------|
| 6.30 Simulation (orange) versus Experimental (blue) ISI distributions for Listeria. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters $\alpha_o$ and $\alpha_{cl}$ . | 93          |
| 6.31 Opening and closing probabilities for optimal parameters for DH5a. . . .  | 96          |
| 6.32 Opening and closing probabilities for optimal parameters for Listeria. . .  | 96          |
| 7.1 Example 1: Numerical results of optimal path trajectory. . . . .   | 116         |
| 7.2 Example 2: Numerical results of optimal path trajectory. . . . .   | 117         |

## LIST OF ACRONYMS

| <b>Notation</b> | <b>Description</b>                  | <b>Page<br/>List</b>   |
|-----------------|-------------------------------------|--|
| ER              | Endoplasmic Reticulum               | 28, 66,<br>67  |
| i.i.d.          | independent identically distributed | 17, 19   |
| IPR             | Inositol trisphosphate receptors    | 66   |
| ISI             | Interspike interval                 | 75, 78,<br>88, 91  |
| LDP             | Large Deviation Principle           | 2, 4,<br>16–20,<br>36, 38,<br>39, 97,<br>118–120,<br>127–129,<br>132 |
| LDT             | Large Deviation Theory              | 1–5,<br>15–17,<br>27, 115,<br>120, 121                               |
| MCMC            | Markov chain Monte Carlo            | 70, 120  |
| MLE             | Maximum Likelihood Estimate         | 69, 78,<br>91  |

| <b>Notation</b> | <b>Description</b>                                | <b>Page<br/>List</b>      |
|-----------------|---|---------------------------|
| PDMP            | Piecewise-deterministic Markov process            | 2–4,<br>11–13,<br>97, 120 |
| SERCA           | Sarcoplasmic/endoplasmic reticulum calcium ATPase | 66, 67                    |

# CHAPTER 1

## INTRODUCTION

### 1.1 Overview and Motivation

Rare events are often some of the most interesting. In any system, they can hold a special allure as they offer us a unique window into the inner workings of a complex system. These events can provide us with genuinely one-of-a-kind insights, which are not attainable from studying more common occurrences. By delving into these rare events, we can deepen our understanding of the underlying mechanisms and dynamics of the system. This knowledge can be invaluable for various fields such as science, engineering, and economics. Whether it is a rare natural phenomenon or an unexpected outcome in a controlled experiment, studying these events paves the way for discoveries and advancements in our understanding.

Biological systems are genuinely fascinating due to their intricate and sophisticated mechanisms. The interaction between dynamical and stochastic effects on different scales makes these systems incredibly complex and often chaotic, which presents a significant challenge for scientists and mathematicians aiming to model and comprehend them. These systems can present a delicate interplay between discrete events and continuous processes, requiring advanced mathematical tools to accurately capture and forecast their dynamics. Studying such systems provides insight into the fundamental workings of life and offers an opportunity to push the boundaries of scientific understanding.

In this work, we will study the stochastic modeling of biological systems. We will use Large Deviation Theory (LDT) to gain insight into the nature of these systems by analyzing aspects in light of rare events. We will focus on two examples: biological neural networks and calcium signaling.

While the literature extensively explores these systems, existing models are either primarily deterministic and overlook the observed stochastic nature of these systems or present stochastic models that may be too abstract for direct biological application or too complex for direct biological insight.

The present work hopes to contribute to this literature by presenting rigorous mathematical results of biologically motivated models that are mathematically tractable and readily give insight into the systems under study.

To this end, we present a balanced random neural network model to capture neuronal firing dynamics. We leverage Large Deviation Theory (LDT) to study this system's convergence of a probability measure (distribution). The derived Large Deviation Principle (LDP) allows us to transfer results from a simplified uncoupled network to a more realistic and complex coupled model. This technique rigorously facilitates studying complex and stochastic neurological models from simplified systems and fundamental biological principles.

For calcium cell signaling, we build a Piecewise-deterministic Markov process (PDMP) model to capture both the deterministic diffusion of calcium and the stochastic opening and closing of calcium channels. This hybrid model is validated by fitting the results to experimental data. A Large Deviation Principle (LDP) is derived in the context of Piecewise-deterministic Markov process (PDMP)s and specifically for the case of calcium signaling, which captures calcium spiking as a rare event within the context of rapidly and highly randomized opening and closing of calcium channels. Although this model is naive in its simplicity, its ability to accurately capture experimental dynamics and its rigorous derivation of a stochastic explanation for spiking phenomena is inspiring. It paves the way for more complex modeling and analysis of this nature.

The results confirm the effectiveness of LDT in analyzing the stochastic modeling of biological systems. This work offers new perspectives on existing models for these

complex systems, providing valuable insights into modeling and analyzing biological systems with broad applications to other fields.

## 1.2 Structure of the Dissertation

This dissertation is organized into nine chapters, each structured to progressively build the theoretical foundation and practical applications of stochastic modeling and LDT in biological systems.

Chapter 1, *Introduction*, begins with an *Overview and Motivation* section, introducing the primary problems addressed in this dissertation, their significance, and the novel approaches undertaken. This chapter also includes a brief outline of the dissertation's structure, providing a roadmap for the reader.

Chapter 2, *Mathematical Foundations*, lays the theoretical groundwork necessary for understanding the subsequent chapters. It discusses topics from probability theory and analysis, such as Skorohod Space, including its definitions, topology, and critical theorems relevant to the rest of this work. This chapter also covers empirical measures, PDMPs, and a brief exposition of LDT, introducing essential theorems such as Cramér's, Varadhan's, Bryc's, Sanov's, and the essentials of Freidlin-Wentzell theory.

Chapter 3, *Biological Contexts and Applications*, focuses on the application domains of the developed models. The first section provides an overview of neural physiology and historical mathematical models used to understand neural networks. The second section examines calcium physiology and reviews traditional models for calcium signaling in biological cells.

Chapter 4, *Neural Network Model*, presents the notation and detailed description of the neural network model. It concludes with the main theoretical results.

Chapter 5, *Large Deviation Principles for Neural Networks*, delves into the large deviation properties of the neural network model. It begins with analyzing large

deviations in the uncoupled system, followed by a discussion on exponential tightness and regularity estimates. It then extends the large deviation to the coupled system and finally establishes a convergence of measure.

Chapter 6, *Calcium Signaling Model*, develops a one-dimensional PDMP model for calcium signaling. It includes methods for maximum likelihood parameter estimation and simulation of the model. The chapter presents and analyzes experimental data in vitro and in vivo and corresponding numerical results.

Chapter 7, *Large Deviation Principles for Calcium Signaling*, applies LDT to PDMPs with a specific focus on calcium signaling. This chapter derives Euler-Lagrange equations and provides numerical results for optimal trajectories in calcium dynamics.

Chapter 8, *Current and Future Research Directions*, discusses ongoing research and potential future directions stemming from the findings of this dissertation.

Chapter 9, *Conclusion*, summarizes the key findings, their significance, and the implications of this research. It also reflects on the limitations and areas for further investigation.

The appendices provide supplementary material to support the main text. *Appendix A: Bounding Fluctuations of the Noise* includes supplementary proofs for the LDP analysis in neural networks. *Appendix B: Proofs for LDP for PDMP* contains detailed proofs for applying LDT to PDMPs.

The dissertation ends with a comprehensive list of all references cited throughout, offering a robust framework for further reading and context.

## CHAPTER 2

### MATHEMATICAL FOUNDATIONS

This chapter presents a brief overview of essential topics from probability theory, analysis, and Large Deviation Theory (LDT) that are relevant and necessary for the rest of the analysis.

#### 2.1 Stochastic Processes

A *stochastic process* is a collection of random variables indexed by a set  $T$ . Formally, let  $(\Omega, \mathcal{F}, \mathbb{P})$  be a probability space, where  $\Omega$  is the sample space,  $\mathcal{F}$  is a  $\sigma$ -algebra of events, and  $\mathbb{P}$  is a probability measure. A stochastic process is a family of random variables  $\{X_t : t \in T\}$  such that for each  $t \in T$ ,  $X_t$  is a measurable function from  $(\Omega, \mathcal{F})$  to a measurable space  $(S, \mathcal{S})$ . Here,  $T$  is called the *index set*, and  $S$  is called the *state space* of the process.

In other words, a stochastic process can be viewed as a function

$$X : T \times \Omega \rightarrow S$$

such that for each fixed  $t \in T$ , the mapping  $X_t : \Omega \rightarrow S$  defined by  $X_t(\omega) = X(t, \omega)$  is a random variable.

Typically, the index set  $T$  represents time and can be discrete, such as  $T = \{0, 1, 2, \dots\}$ , or continuous, such as  $T = [0, \infty)$ . The state space  $S$  can be various sets such as  $\mathbb{R}$ ,  $\mathbb{R}^d$ , or other more abstract spaces.

##### 2.1.1 Brownian Motion

A stochastic process  $W = (W_t, t \in [0, \infty))$  is called (standard) *Brownian motion* or a Wiener process if the following conditions are satisfied: (i) It starts at zero  $W_0 = 0$ .

(ii) It has stationary, independent increments. (iii) For every  $0 \leq s < t$ ,  $W_t - W_s$  is distributed as  $N(0, t - s)$ , and  $W_t - W_s$  is independent of  $W_r$  if  $r < s$ . (iv) It has continuous sample paths.

### 2.1.2 Poisson Process

A *Poisson process* is a stochastic process that models a series of events occurring randomly over time. A Poisson process with rate (or intensity)  $\lambda > 0$  is a counting process  $\{N(t) : t \geq 0\}$  that satisfies the following properties:

(i)  $N(0) = 0$ : The process starts at zero. (ii) *Independent increments*: The number of events occurring in any disjoint time intervals are independent. (iii) *Stationary increments*: The number of events occurring in any time interval of length  $t$  only depends on  $t$  and not on the specific position of the interval. Formally, for any  $s, t \geq 0$ ,

$$N(t + s) - N(s) \sim \text{Poisson}(\lambda t).$$

(iv) *No simultaneous events*: The probability of more than one event co-occurring is zero. Formally,

$$\mathbb{P}(N(t + h) - N(t) \geq 2) = o(h) \text{ as } h \rightarrow 0.$$

From these properties, it follows that the number of events  $N(t)$  in a Poisson process over the interval  $[0, t]$  is Poisson distributed with parameter  $\lambda t$ :

$$\mathbb{P}(N(t) = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, \quad k = 0, 1, 2, \dots$$

Additionally, the inter-arrival times  $T_i$  (the times between consecutive events) in a Poisson process are independent and identically distributed (i.i.d.) exponential random variables with parameter  $\lambda$ :

$$\mathbb{P}(T_i > t) = e^{-\lambda t}, \quad t \geq 0.$$

The Poisson process is widely used to model random events over time in various fields, such as queueing theory, telecommunications, and reliability engineering.

## 2.2 Stochastic Differential Equation

A stochastic differential equation (SDE) is an equation used to model systems that are influenced by white noise. The general form of an SDE is:

$$dX_t = a(X_t, t) dt + b(X_t, t) dW_t$$

where:

$X_t$  is the stochastic process,  $a(X_t, t)$  is the drift term, which represents the deterministic part of the system,  $b(X_t, t)$  is the diffusion term, which represents the random part of the system,  $dW_t$  is the increment of a Wiener process (also called a Brownian motion).

Classic introductory texts on the study of SDEs include Oksendahl [114], Evans [52], and Lawler [95].

## 2.3 Skorohod Space

The space of continuous function,  $C$ , is unsuitable for describing processes that must contain jumps, like the Poisson processes, in contrast to Brownian motion. This section introduces the Skorohod Space, which includes (certain) discontinuous functions. Proofs will be omitted for brevity. The following outline is based on Billingsley [17]; the interested reader is directed to that excellent resource for a complete treatment of this material.

### 2.3.1 Definition

Let  $\mathcal{D} = \mathcal{D}[0, 1]$  be the set of real functions on  $[0, 1]$  that are right-continuous and have left-hand limits. Such functions are termed *càdlàg* functions. Specifically, for  $0 \leq t < 1$ , the right-hand limit  $x(t+)$  exists and equals  $\lim_{s \downarrow t} x(s)$ , and for  $0 < t \leq 1$ ,

the left-hand limit  $x(t-)$  exists and equals  $\lim_{s \uparrow t} x(s)$ . A function has a discontinuity of the first kind at  $t$  if  $x(t-)$  and  $x(t+)$  exist but differ, with  $x(t)$  lying between them. Any discontinuities in a càdlàg function are of the first kind, and the requirement  $x(t) = x(t+)$  serves as a convention.

### 2.3.2 Topology

The space of càdlàg functions is usually endowed with the Skorohod topology. The Skorohod topology allows for small deformations in both the function values and the time scale. This topology is particularly useful when exact measurements of time are not possible. The Skorohod metric  $d(z, y)$  is defined using the space of strictly increasing continuous mappings of  $[0, 1]$  onto itself denoted by  $\Lambda$ , such that for  $\lambda \in \Lambda$  then  $\lambda 0 = 0$  and  $\lambda 1 = 1$ .  $d(z, y)$  quantifies the “inf-sup” difference between  $z(t)$  and  $y(\lambda t)$  and between  $t$  and  $\lambda t$ . Specifically,

$$d(x, y) := \inf_{\lambda \in \Lambda} \{ \|\lambda - I\| \vee \|x - \lambda y\| \} \quad (2.1)$$

Where  $I$  is the identity map on  $[0, 1]$ , the metric  $d$  satisfies properties such as non-negativity, symmetry, and triangle inequality, making it a valid metric. Convergence in the Skorohod topology requires that there exist functions  $\lambda_n$  such that  $z_n(\lambda_n t)$  converges uniformly to  $z(t)$  and  $\lambda_n t$  converges uniformly to  $t$ . The Skorohod topology coincides with the uniform topology when restricted to continuous functions.

We can define a new metric  $d_0$  in  $\mathcal{D}$ , equivalent to  $d$  and providing completeness to  $\mathcal{D}$ . This metric helps in characterizing compact sets. The definition of  $d_0$  involves a time-deformation function  $\lambda$  that must be near the identity function, with the slope of its chords close to 1, or equivalently, the logarithm of the slope close to 0. The norm

$$\|\lambda\|_t^\circ = \sup_{s < t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| \quad (2.2)$$

is introduced to measure this closeness. If  $\|\lambda\|^\circ$  is finite,  $\lambda$  is continuous, strictly increasing and belongs to  $\Lambda$ . The metric

$$d^\circ(x, y) := \inf_{\lambda \in \Lambda} \{ \|\lambda\|^\circ \vee \|x - y\lambda\| \} \quad (2.3)$$

is defined as the infimum of positive  $\epsilon$  for which there exists a  $\lambda \in \Lambda$  with  $\|\lambda\|^\circ < \epsilon$ .

### 2.3.3 Key theorems

The following are some important theorems that will be relevant in the forthcoming analysis.

The separability and completeness of  $\mathcal{D}$ .

**Theorem 2.3.1.** *The space  $\mathcal{D}$  is separable under  $d$  and  $d^\circ$  and is complete under  $d^\circ$ .*

Compactness in  $\mathcal{D}$ .

**Theorem 2.3.2.** *A necessary and sufficient condition for a set  $A$  to be relatively compact in the Skorohod topology is that*

$$\sup_{x \in A} \|x\| < \infty \quad (2.4)$$

and

$$\limsup_{\delta \rightarrow 0} \sup_{x \in A} w'_x(\delta) = 0. \quad (2.5)$$

where for  $\min_{1 \leq i \leq v} (t_i - t_{i-1}) > \delta$ , and  $0 \leq \delta \leq 1$ ,

$$w'_x(\delta) = \inf_{\{t_i\}} \max_{1 \leq i \leq v} w_x[t_{i-1}, t_i], \quad (2.6)$$

for the modulus of continuity of  $x$  given by,

$$w_x(\delta) = \sup_{0 \leq t \leq 1 - \delta} \sup_{s, t \in T} |x(s) - x(t)|, \quad \text{for } |t - s| \leq \delta. \quad (2.7)$$

The Poisson Limit.

**Theorem 2.3.3.** *Suppose that  $E \in \mathcal{D}$  and  $T_0$  is a countable, dense set in  $[0, 1]$ , and let  $P_n$  be the probability measure of a Poisson process with rate  $n$ . Suppose further that, if  $x, x_n \in E$  and  $x_n(t) \rightarrow x(t)$  for  $t \in T_0$ , then  $x_n \rightarrow x$  in the Skorohod topology.*

*If  $P_n(E) = P(E) = 1$  and  $P_n x_{t_1, \dots, t_k}^{-1} \rightarrow P x_{t_1, \dots, t_k}^{-1}$  for all  $k$ -tuples in  $T_0$ , then  $P_n \rightarrow P$ .*

These theorems allow us to define a sense of (weak) convergence in probability space. The space,  $\mathcal{D}$ , is named after the mathematician Anatoliy Skorokhod. It plays a significant role in the theory of random processes and will be used throughout this work.

## 2.4 Empirical Measure

The empirical measure is one of the most popular means of obtaining a low-dimensional representation of high-dimensional stochastic systems. Intuitively, it is a ‘discrete population density,’ and (in this dissertation) in the large  $N$  limit, it converges to a continuum population density.

Given a sample of  $n$  observations  $X_1, X_2, \dots, X_n$  from a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  with values in a measurable space  $(S, \mathcal{S})$ , the *empirical measure*  $\hat{P}_n$  is a random measure that assigns equal probability to each observed data point. Formally, the empirical measure  $\hat{P}_n$  is defined as:

$$\hat{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i},$$

where  $\delta_{X_i}$  denotes the Dirac delta measure centered at  $X_i$ .

For any measurable set  $A \in \mathcal{S}$ , the empirical measure  $\hat{P}_n$  assigns a probability given by:

$$\hat{P}_n(A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_A(X_i),$$

where  $\mathbf{1}_A$  is the indicator function of the set  $A$ , defined as:

$$\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$$

Thus,  $\hat{P}_n(A)$  represents the proportion of observations  $X_1, X_2, \dots, X_n$  that fall within the set  $A$ .

The empirical measure provides a non-parametric estimate of the underlying distribution from which the sample is drawn. Empirical measures are significant in mathematical statistics. The underlying probability measure is often unknown, so empirical measures are examined. By gathering observations and calculating relative frequencies, we can estimate the measure or a related distribution function using the empirical measure or the empirical distribution function. These estimates are uniformly reliable under certain conditions.

## 2.5 Piecewise Deterministic Markov Processes

PDMP are a general class of non-diffusion stochastic processes. They were first fully categorized by Davis in his seminal 1984 paper [40]. PDMP (also known as stochastic hybrid systems) are used to model systems with multiple timescales [6, 71]. They are also referred to as slow-fast systems. These processes enjoy numerous applications, particularly in biology [19, 20, 126]. For instance, they have been used to model excitable membranes in neuroscience [96, 121], population dynamics in ecology [73], run-and-tumble dynamics of bacteria [14], and stochastic models of calcium signals [88, 118, 119].

A PDMP is a stochastic process characterized by random jumps occurring at specific moments, with deterministic evolution governed by an ordinary differential equation (ODE) between these jumps. In other words, PDMPs contain two types of random variables: one with a discrete state space and one with a continuous state space.

The state of the process at time  $t$ , denoted by  $X_t = (v_t, \xi_t)$ , is determined by several components:

*Vector Fields:* Each subset  $M_v \in \mathbb{R}^d$  (i.e.,  $M_v$  represents the region in the state space where the process evolves deterministically under the influence of a vector field associated with  $v$ ) is associated with a vector field  $X_v$  that dictates the deterministic evolution of the process within  $M_v$ . The flow generated by these vector fields ensures unique integral curves, meaning that the solutions to the ODEs do not exhibit 'explosions' (i.e., they do not become unbounded in finite time).

*Jump Rate Function:* The function  $\lambda : E \rightarrow \mathbb{R}_+$  specifies the rate at which jumps occur. For any state  $(v, \xi) \in E$ ,  $\lambda(v, \xi)$  represents the rate at which the process jumps out of the current deterministic trajectory.

The process  $X_t$  evolves according to the following rules: *Deterministic Flow:* Between jumps, the process follows the deterministic flow  $\Phi_t$ , defined by the vector fields  $X_v$ . Specifically, if  $X_t = (v, \xi)$  at time  $t$ , then for a short time interval  $[t, t + \delta t)$ , the state evolves according to the ODE:

$$\frac{d\xi_t}{dt} = X_v(\xi_t).$$

*Jumps:* A jump occurs at time  $\tau$ . The jump rate  $\lambda$  determines the likelihood of jumping in infinitesimally small intervals, ensuring that  $\lambda$  is integrable over small intervals.

In summary, a PDMP is fully defined by the following components: *Deterministic Flow*  $\Phi$ : Governs the continuous evolution between jumps. *Jump Rate*  $\lambda$ : Determines the frequency of jumps.

These elements combine to provide a rich framework for modeling systems exhibiting deterministic and stochastic behavior.

## 2.6 Radon-Nikodym Theorem and Derivative

Let  $(\Omega, \mathcal{F})$  be a measurable space, and let  $P$  and  $Q$  be two  $\sigma$ -finite measures on this space. The *Radon-Nikodym derivative* of  $P$  with respect to  $Q$ , denoted by  $\frac{dP}{dQ}$ , is a measurable function that satisfies the Radon-Nikodym theorem.

### 2.6.1 Radon-Nikodym Theorem

The Radon-Nikodym theorem states that if  $P$  is absolutely continuous with respect to  $Q$  (denoted  $P \ll Q$ ), then there exists a unique (up to  $Q$ -almost everywhere equivalence) non-negative measurable function  $\frac{dP}{dQ}$  such that for any measurable set  $A \in \mathcal{F}$ ,

$$P(A) = \int_A \frac{dP}{dQ} dQ.$$

Formally, the Radon-Nikodym derivative  $\frac{dP}{dQ}$  is the function that satisfies

$$P(A) = \int_A \frac{dP}{dQ} dQ \quad \text{for all } A \in \mathcal{F}.$$

The Radon-Nikodym derivative  $\frac{dP}{dQ}$  is often interpreted as the density of the measure  $P$  with respect to the measure  $Q$ .

### 2.6.2 Properties of the Radon-Nikodym derivative

*Uniqueness*: If there exist a  $f$  and  $g$  such that for any  $A$ ,  $P(A) = \int_A f dQ = \int_A g dQ$ , then  $f = g$  a.s. (almost surely) and you denote it  $dP/dQ$ .

*Linearity:* If  $P_1$  and  $P_2$  are absolutely continuous with respect to  $Q$ , and  $a, b$  are positive real numbers, then

$$\frac{d(aP_1 + bP_2)}{dQ} = a \frac{dP_1}{dQ} + b \frac{dP_2}{dQ}.$$

*Chain Rule:* If  $P \ll Q \ll R$ , then

$$\frac{dP}{dR} = \frac{dP}{dQ} \cdot \frac{dQ}{dR} \quad R\text{-almost everywhere.}$$

The existence and uniqueness of the Radon-Nikodym derivative are guaranteed by the Radon-Nikodym theorem.

## 2.7 Relative Entropy

The *relative entropy* (or Kullback-Leibler divergence) between two probability measures  $P$  and  $Q$  defined on the same probability space  $(\Omega, \mathcal{F})$  is a measure of how one probability distribution diverges from a second, expected probability distribution [27].

If  $P$  is absolutely continuous with respect to  $Q$  (denoted  $P \ll Q$ ), the relative entropy  $\mathcal{R}(P||Q)$  is defined as:

$$\mathcal{R}(P||Q) = \int_{\Omega} \log \left( \frac{dP}{dQ} \right) dP,$$

where  $\frac{dP}{dQ}$  is the Radon-Nikodym derivative of  $P$  with respect to  $Q$ .

Alternatively, if  $P$  and  $Q$  have probability density functions  $p$  and  $q$  respectively with respect to a common reference measure (e.g., the Lebesgue measure), then the relative entropy can be expressed as:

$$\mathcal{R}(P||Q) = \int_{\Omega} p(x) \log \left( \frac{p(x)}{q(x)} \right) dx.$$

Relative entropy is always non-negative, i.e.,  $\mathcal{R}(P\|Q) \geq 0$ , and it is zero if and only if  $P = Q$  almost everywhere. It is a fundamental concept in information theory, statistics, and LDT, where it quantifies the “distance” between two probability distributions.

## 2.8 Large Deviation Theory

In this section, I will briefly summarize topics from Large Deviation Theory (LDT) that are relevant and necessary for the rest of the analysis. Proofs will be omitted for brevity. The interested reader is directed to the works of Rassoul-Agha and Seppalainen [120], Dembo [45], and Freidlin and Wentzell [61] for a thorough treatment of this material. We also refer to an article by Touchette for a gentle introduction [134].

### Limit Superior and Limit Inferior

Large Deviation results are asymptotic and are expressed in terms of the limit superior and limit inferior. The lim sup (limit superior) of a sequence  $(a_n)$  is defined as follows:

$$\limsup_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \sup_{m \geq n} a_m$$

In words, it is the limit of the supremum of the tail ends of the sequence as  $n$  goes to infinity. It can be intuitively understood as the greatest value that the sequence gets arbitrarily close to infinitely often.

The lim inf (limit inferior) of a sequence  $(a_n)$  is defined as follows:

$$\liminf_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} \inf_{m \geq n} a_m$$

In words, it is the limit of the infimum of the tail ends of the sequence as  $n$  goes to infinity. It can be intuitively understood as the smallest value that the sequence gets arbitrarily close to infinitely often.

### 2.8.1 Large Deviation Principle

Large Deviation Theory estimates the probability of rare events [27, 89]. For example, one may be interested in the probability of extinction of a species in population dynamics or chemical reaction networks [23] or the spontaneous and stochastic production of waves or pulses [86].

The probability of system-wide rare events in high-dimensional stochastic systems typically decays exponentially in the size of the system. LDT determines the leading order coefficient for the exponential decay rate (the coefficient is known as a ‘rate function,’ defined below). This concept is referred to as Large Deviation Principle (LDP). A precise definition of a LDP is as follows [120]:

**Theorem 2.8.1.** *Let  $I : x \rightarrow [0, \infty]$  be a lower semicontinuous function and  $r_n \uparrow \infty$  a sequence of positive real constants. A sequence of probability measures  $\{\mu_n\} \subset \mathcal{P}(\mathcal{X})$  is said to satisfy a LDP with rate function  $I$  and normalization  $r_n$  if the following holds:*

$$\limsup_{n \rightarrow \infty} \frac{1}{r_n} \log \mu_n(F) \leq - \inf_{x \in F} I(x) \quad \forall \text{ closed } F \subset X, \quad (2.8)$$

$$\liminf_{n \rightarrow \infty} \frac{1}{r_n} \log \mu_n(G) \geq - \inf_{x \in G} I(x) \quad \forall \text{ open } G \subset X, \quad (2.9)$$

where  $X$  is a Hausdorff topological space and  $\mathcal{P}(\mathcal{X})$  is the space of probability measures on the  $(X, \mathcal{B})$  where  $\mathcal{B}$  is the Borel  $\sigma$ -algebra.

### 2.8.2 Cramer’s Theorem

One of the earliest results in LDT is Cramer’s Theorem. Introduced by Harald Cramér in 1938, this theorem characterizes the exponential decay of probabilities associated

with sums of independent identically distributed (i.i.d.) random variables deviating significantly from their expected value. Specifically, Cramer’s Theorem quantifies the rate at which the probability of such deviations decreases, using the rate function, or Cramér function, derived from the cumulant generating function of the random variables.

Let  $\{X_n\}_{n \geq 1}$  be i.i.d., real-valued random variables, and let  $X$  be another random variable with the same distribution. The moment generating function is written as  $M(\theta) = E[e^{\theta X}]$ , for  $\theta \in \mathbb{R}$ . Notice that  $M(\theta) > 0$  always and  $M(\theta) = \infty$  is possible. In Cramer’s Theorem, the rate function is the Fenchel-Legendre transform of the moment generating function, i.e.,

$$I(x) = \sup_{\theta \in \mathbf{R}} \{\theta x - \log M(\theta)\}. \quad (2.10)$$

Since  $M(0) = 1$ ,  $I : \mathbb{R} \rightarrow [0, \infty]$  is a well-defined function.

**Theorem 2.8.2** (Cramer’s Theorem). *Let  $\{X_n\}$  be a sequence of i.i.d. real-valued random variables. Let  $\mu_n$  be the distribution of the sample mean  $S_n/n$ . Then the LDP is satisfied with  $I$  defined in Equation 2.10.*

### 2.8.3 Varadhan’s Theorem

Varadhan’s theorem, a seminal result in LDT, extends the foundational principles laid out by earlier work in the field. This theorem, named after the mathematician S. R. Srinivasa Varadhan, provides a comprehensive framework for understanding the asymptotic probabilities of rare events in more complex settings, particularly for sequences of random variables with dependent structures. The practical application of this theorem is evident in its ability to evaluate the asymptotic value of exponential moments, providing a generalization of Laplace’s method to more abstract probability spaces.

**Theorem 2.8.3** (Varadhan's Theorem). *Suppose a LDP holds for  $\mu_n$ ,  $r_n$ , and  $I$ . Let  $f : X \rightarrow [-\infty, \infty]$  be a continuous function, and*

$$\lim_{b \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{r_n} \log \int_{f \geq b} e^{r_n f} d\mu_n = -\infty \quad (2.11)$$

*Then*

$$\lim_{n \rightarrow \infty} \frac{1}{r_n} \log \int e^{r_n f} d\mu_n = \sup_{x: f(x) \wedge I(x) < \infty} \{f(x) - I(x)\} \quad (2.12)$$

#### 2.8.4 Bryc's Theorem

Named after the mathematician Wlodzimierz Bryc, the following theorem offers an alternative approach to verifying the conditions under which a sequence of random variables satisfies a LDP. Bryc's theorem establishes that if a sequence of random variables has exponential moment generating functions that converge to a limiting function, which is convex and lower semicontinuous, then the sequence satisfies an LDP with a specific rate function. This theorem is especially valuable due to its practical applicability in various complex systems, where checking the standard conditions for an LDP can be challenging.

**Theorem 2.8.4** (Bryc's Theorem). *Let  $\{\mu_n\}$  be a sequence of probability measures on a metric space  $X$ . Assume  $\{\mu_n\}$  is exponentially tight with normalization  $r_n$ . Suppose the limit*

$$\Gamma(f) = \lim_{n \rightarrow \infty} \frac{1}{r_n} \log \int e^{r_n f} d\mu_n$$

*for all  $f \in C_b(X)$ . Then, a LDP holds with the tight rate function*

$$I(x) = \sup_{f \in C_b(X)} \{f(x) - \Gamma(f)\} \quad (2.13)$$

### 2.8.5 Sanov's Theorem

Named after the mathematician Ilya Sanov, the following theorem addresses the large deviation properties of the empirical measure (see Section 2.4) of a sequence of independent and i.i.d. random variables. Specifically, Sanov's Theorem quantifies the exponential rate at which the probability that the empirical distribution of these variables diverges from a given probability distribution decays. This rate is described by the relative entropy (or Kullback-Leibler divergence, see Section 2.7) between the empirical distribution and the true underlying distribution.

**Theorem 2.8.5** (Sanov's Theorem). *Let  $S$  be a Polish space (complete, metrizable, separable) and  $\rho_n$  the distribution of the empirical measure  $L_n$ , then a LDP holds on the space  $\mathcal{P}(S)$  with convex rate function  $\mathcal{R}(\nu) = \mathcal{R}(\nu|\lambda)$ , where  $\mathcal{R}$  is the relative entropy of  $\nu$  relative to  $\lambda$  given by*

$$\mathcal{R}(\nu|\lambda) = \begin{cases} \int \phi \log \phi d\lambda & \text{if } \nu \ll \lambda \text{ and } \phi = \frac{d\nu}{d\lambda} \\ \infty & \text{otherwise} \end{cases} \quad (2.14)$$

### 2.8.6 Freidlin-Wentzell Theory

Named after the mathematicians Mark Freidlin and Alexander Wentzell, the following theory extends the principles of large deviations to stochastic differential equations (SDEs) driven by vanishingly small noise. The central result, often referred to as the Freidlin-Wentzell Theorem, characterizes the exponential decay rates of the probabilities that the trajectories of these stochastic processes deviate significantly from their deterministic paths. The key theorem states that if  $X_t^\epsilon$  is a family of stochastic processes driven by small noise, then the probability that  $X_t^\epsilon$  deviates from the solution of the corresponding deterministic system can be described by a rate function  $I$ , derived from a variational principle. Formally, the fundamental theorem can be stated as follows:

Consider a family of stochastic processes  $\{X_t^\epsilon\}_{\epsilon>0}$  described by the stochastic differential equation (SDE)

$$dX_t^\epsilon = b(X_t^\epsilon) dt + \sqrt{\epsilon} \sigma(X_t^\epsilon) dW_t,$$

where  $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$  is the drift term,  $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  is the diffusion term,  $\epsilon$  is a small parameter, and  $W_t$  is a standard  $d$ -dimensional Wiener process.

The Freidlin-Wentzell Theorem provides a LDP for the trajectories of  $X_t^\epsilon$  as  $\epsilon \rightarrow 0$ .

Let  $C([0, T]; \mathbb{R}^d)$  denote the space of continuous functions from  $[0, T]$  to  $\mathbb{R}^d$  equipped with the uniform topology. The rate function  $I : C([0, T]; \mathbb{R}^d) \rightarrow [0, \infty]$  is defined as

$$I(\phi) = \begin{cases} \frac{1}{2} \int_0^T \left| \dot{\phi}(t) - b(\phi(t)) \right|_{\sigma^{-1}(\phi(t))}^2 dt & \text{if } \phi \in \mathcal{H}_1([0, T]; \mathbb{R}^d) \\ +\infty & \text{otherwise,} \end{cases}$$

where  $\mathcal{H}_1 := \left\{ \int_0^t f(s) ds : f \in L_2([0, 1]) \right\}$ , i.e., the space of all absolutely continuous functions with value 0 at 0 that possess a square integrable derivative, equipped with the norm  $\|g\|_{\mathcal{H}_1} = \left[ \int_0^1 |\dot{g}(t)|^2 dt \right]^{1/2}$ , and  $|\dot{\phi}(t) - b(\phi(t))|_{\sigma^{-1}} := (\dot{\phi}(t) - b(\phi(t)))' (\sigma(\phi(t)) \sigma'(\phi(t)))^{-1} (\dot{\phi}(t) - b(\phi(t)))$ , i.e., the norm induced by the inverse of the diffusion matrix  $\sigma$ . The theorem states that:

**Theorem 2.8.6** (Freidlin-Wentzell Theorem). *The family of measures  $\{\mathbb{P}_{X^\epsilon}\}_{\epsilon>0}$  on  $C([0, T]; \mathbb{R}^d)$  satisfies the LDP with the rate function  $I$ . That is, for any Borel set  $A \subset C([0, T]; \mathbb{R}^d)$ ,*

$$-\inf_{\phi \in A^\circ} I(\phi) \leq \liminf_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}(X_t^\epsilon \in A^\circ) \leq \limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbb{P}(X_t^\epsilon \in \bar{A}) \leq -\inf_{\phi \in \bar{A}} I(\phi), \quad (2.15)$$

where  $A^\circ$  and  $\bar{A}$  denote the interior and closure of  $A$ , respectively.

This theorem provides a rigorous framework for understanding the asymptotic behavior of the probabilities of rare events for the stochastic process  $X_t^\epsilon$ .

## CHAPTER 3

### BIOLOGICAL CONTEXTS AND APPLICATIONS

This chapter presents a brief but comprehensive summary of the critical biological and physiological aspects of each system that we will model and study in this dissertation. In addition, we will provide an overview of the historical research in the field that forms the basis for the work presented.

#### 3.1 Neuroscience

The first example system that we will explore is the biological neural network found in the brain.

##### 3.1.1 Physiology

The brain, a marvel of nature, is a vast and diverse network of neurons. The brain is intricately organized both anatomically and functionally, consisting of the spinal cord along with various regions, including the medulla oblongata, pons, cerebellum, midbrain, diencephalon, and cerebral hemispheres [117]. Fundamental structures include the thalamus and hypothalamus in the diencephalon and the basal ganglia, hippocampus, and amygdaloid nucleus within the cerebral hemispheres [81]. The cerebral cortex is divided into frontal, parietal, occipital, and temporal lobes, responsible for sensory, cognitive, and voluntary motor functions. Each lobe specializes in distinct functions: the frontal lobe in planning and organization, the occipital lobe in vision, the parietal lobe in sensory information, and the temporal lobe in hearing and language [90]. Each hemisphere primarily controls sensory and motor functions on the opposite side of the body, with sensory information organized topographically in the somatosensory cortex, allocating more space to sensitive regions like the fingers and

mouth. Cognitive functions such as language are localized in specific areas, including Wernicke’s area for understanding speech and Broca’s area for speech production [51].

Neurons are specialized cells with three main components: dendrites (inputs), soma (cell body), and axon (output). The axon can extend up to 1 meter and is insulated by a myelin sheath for faster signal propagation [9]. Neurons generate electrical signals called action potentials that travel down the axon in an all-or-none fashion with consistent amplitude and duration. Stronger stimuli produce higher firing frequencies to encode information, and neurons can exhibit various firing patterns such as bursting [51].

Neurons communicate at synapses, which can be chemical or electrical. Chemical synapses are most common in the mammalian brain, where neurotransmitters released from the pre-synaptic neuron bind to receptors on the postsynaptic neuron, causing either excitatory (depolarizing) or inhibitory (hyperpolarizing) effects. Synapses can be direct/fast or indirect/slow based on the receptor type. Each neuron receives inputs from approximately 1000 other neurons on average, forming complex synaptic architectures that enable specialized brain functions. Synapses can amplify or modulate signals, and their connections can be modified through processes such as learning [51].

Neurons collectively form an extensive and highly diverse interconnection network. The brain is estimated to contain around 86 billion neurons [74] and approximately 100 trillion nonuniform connections [145]. Furthermore, in addition to this immense complexity, a considerable degree of stochasticity is also observed [123]. Even if one focuses solely on examining a specific brain region, such as the visual cortex, the level of complexity remains beyond current computational capabilities [97].

### **3.1.2 Overview of historical mathematical modeling**

In 1963, Alan Lloyd Hodgkin and Andrew Fielding Huxley were awarded the Nobel Prize in Physiology or Medicine for their groundbreaking Hodgkin-Huxley model [76]. This mathematical framework offered a simplified portrayal of single-neuron dynamics

during an action potential, drawing upon principles from electrical circuits, thereby revolutionizing our comprehension of neuronal communication and paving the way for further research in neuroscience and computational neuroscience.

Their seminal work laid the groundwork for “bottom-up” approaches, which endeavor to construct neural function models based on individual neuron descriptions, beginning with a highly accurate microscopic model and then deriving macroscopic equations through analysis. However, incorporating the Hodgkin-Huxley equations directly into large-network models proves computationally impractical due to their intricate nature. It can potentially hinder insight into fundamental physiological mechanisms. Consequently, significant analytic efforts have focused on reducing complexity and deriving effective equations that faithfully represent large ensembles of neurons [21, 24, 25]. Noteworthy examples of such models include the Morris-Lecar model [110], the FitzHugh-Nagumo model [58], and the Integrate-and-fire model [28, 41]. This avenue of research remains vibrant, as exemplified by endeavors like the Human Brain Project [106].

However, despite the simplifications introduced by these models, the bottom-up approach still presents challenges and disadvantages. Computational complexity, data fitting, and the interpretability of results still present significant challenges. On the scale of modeling the number of neurons in the human brain, the complexity of even the most streamlined model approaches that of the brain itself. For a more detailed discussion of these triumphs and challenges, the reader is directed to [42] and [49].

An alternative method is commonly referred to as the “top-down” approach. Inspired by statistical mechanics, this approach aims to model neuronal behavior at a statistical level [43]. The resulting models produce what are called *neural field equations* (sometimes also ambiguously referred to as *neural mass models*, *mean field models*, and *neural population models*, but some make distinctions between them [34]). These models segregate cortical tissue en masse into populations with

shared statistical properties [143]. A typical partition would be into excitatory and inhibitory populations [79]. These models typically describe the average activity across a region of space, which significantly reduces the theoretical complexity [18]. Furthermore, these models tend to be well suited for interpretability with physiological measurements such as local field potentials (LFPs), electroencephalography (EEG), or magnetoencephalography (MEG) [122].

The origins of neural field equations can be traced back to Beurle [16] and Griffith [66, 67, 68]. However, modern canonical models of neural field theory are believed to truly begin with the work of Freeman [60], Wilson [143], Wilson and Cowan [142], Nunez [113], Lopes da Silva et al. [101] and Amari [2, 3].

A subset of models are known as *balanced network* models. It has been proposed that observations such as temporal variability in neuron firing can be produced from such equations when modeled as an approximately balanced state between excitatory and inhibitory inputs [64, 129]. Characteristics of such models are net excitation much greater than the firing threshold, balanced by inhibition input, and requiring substantial fluctuations above the long time mean to fire. Early exploration was due to Sompolinsky [135], see also [98, 124].

The complex and chaotic dynamics inherent in cellular and subcellular interactions among extensive neuron populations suggest the suitability of stochastic models, aligning with a prevalent theme in mathematical biology. Stochastic neural fields, a significant area of neural field modeling, stem from two primary methodologies: stochastic extensions of traditional neural field frameworks and population density approach emphasizing microscale neuron interactions [22]. These models excel in capturing randomness arising from neuron variability, providing a more phenomenological perspective conducive to quantifying neural activity [78, 107, 115, 132]. Notably, they elucidate phenomena related to finite-size population effects and rare events [23, 55, 92, 94]. By leveraging stochastic abstractions, these models

better match experimental data, fostering data-driven research in neural dynamics [13, 80, 128].

A popular means of studying high dimensional balanced neural networks is by taking the connections to be static random variables [133]. These models utilize a discrete matrix operator rather than a continuous spatial kernel to describe neuron connectivity, treating the dynamics as a directed graph with non-linearly coupled neural mass models. This approach uses random matrix theory to handle large-scale complexity and compute the statistical properties of the system. Early work by Van Vreeswijk, Crisanti, and Sompolinsky [135, 138] strongly suggested a phase transition to chaotic dynamics as the network size increases. Key findings include the critical role of connectivity parameters in phase transitions and using Dynamical Mean-field theory to simplify large network dynamics into a manageable stochastic process. Early work by Sompolinsky et al., [132] anticipated that low-dimensional population density type equations could accurately describe such systems. These models were originally developed for ‘spin glass’ systems [32, 108, 109, 116, 137]. Crisanti, Horner, and Sommers performed the initial derivation of correlation-response equations applicable to symmetric random neural networks [38], followed shortly afterward by Cugliandolo and Kurchan[39]. The correctness of these equations was subsequently confirmed by Ben Arous, Dembo, and Guionnet [12], who used concentration inequalities from probability theory to produce a thorough and detailed proof.

As  $N$  becomes large, a process usually applied to dynamical systems would be to calculate the fixed points and linearize the system around them to assess local stability; however, this becomes computationally intractable for random neural networks due to nested nonlinear coupling functions. Sompolinsky, Crisanti, and Sommers have made significant advances in the literature on large  $N$ -limiting equations for such networks. They proposed that Path Integral methods could yield limiting dynamical equations [132] and subsequently published their derivation [37]. Physicists use Path

Integral methods to derive population density equations by identifying where the probability measure for the N-dimensional system concentrates. In probability theory, Large Deviation Theory (LDT) is a potent tool for addressing this question [44]. The pioneering papers of Ben Arous and Guionnet [10, 11, 70] utilized LDT to analyze spin-glass dynamics. They obtained the first rigorous results regarding the large N limit of random neural networks. Following this work, Grunwald applied LDT to derive correlation/response equations for random neural networks with randomly flipping spins between discrete states [69]. Moynot and Samuelides explored the non-Gaussian case [111], while Faugeras and MacLaurin extended the work of Ben Arous and Guionnet to incorporate correlations in connectivity [53]. Touboul and Cabana determined the limiting equations for spatially extended systems [29, 30], and Faugeras, Soret, and Tanre [54] derived novel integral equations to describe the state of these systems. MacLaurin established limiting equations for jump-Markov spin glass systems [103].

For a comprehensive overview of the critical stages of the history and development of neural field theory, mean-field equations, and contemporary uses of this branch of mathematical neuroscience, the reader is directed to the article by Cook et al. [34] and the monograph by Helias and Dahmen [72].

## 3.2 Calcium Signaling

The second example system we will explore is the regulation of calcium signaling in biological cells.

### 3.2.1 Phsyiology

Approximately one percent of the calcium ( $\text{Ca}^{2+}$ ) in the human body is found within the cells. This calcium functions as an essential signaling molecule. It is necessary for excitation-contraction coupling in muscles [50], excitation-secretion coupling in

synapses [83], exocytosis [127], fluid transport elasticity in presynaptic and postsynaptic neurons [91], gene regulation and differentiation [15], and cell movement and cell death [33].

At equilibrium, most of the cells's calcium resides within inner cell structures, such as the endoplasmic reticulum. The concentration in the cytoplasm is several orders of magnitude lower. Energy-consuming active pumps pump calcium into the Endoplasmic Reticulum (ER). Channels on the surface of the ER can open to allow calcium to flow into the cytoplasm. More specifically, when an agonist binds to a G protein-coupled receptor (GPCR), it activates a G protein, activating phospholipase C (PLC). PLC then splits phosphatidylinositol bisphosphate (PIP<sub>2</sub>) into diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP<sub>3</sub>). IP<sub>3</sub> can then freely move within the cell cytoplasm. Upon binding to IP<sub>3</sub> receptors (IPR) located predominantly on the ER membrane, IP<sub>3</sub> triggers the release of calcium from the ER. The endoplasmic reticulum is the primary internal compartment for calcium dynamics, with contributions from mitochondrial stores and other inner cell structures playing a secondary role [47].

Calcium signaling is organized hierarchically [31, 62, 84]. Calcium dynamics in cells exhibit threshold behavior similar to other excitable systems, where small perturbations return to a steady state unless a significant disturbance causes a large transient response, known as a calcium spike. These spikes manifest as oscillations or waves across the cell, appearing smooth at a macroscopic level but resulting from stochastic events at a microscopic scale. Specifically, the random opening and closing of calcium channels (IPR channels) lead to localized releases referred to in the literature as blips. These blips can aggregate into larger releases known as puffs, which collectively generate cell-wide calcium waves. There is considerable evidence that calcium puffs and waves are nonlinear stochastic phenomena: it has been observed, for instance, that IP<sub>3</sub> channels can be highly active even when the average open probability is less than half its maximum [130]. The presence of abortive calcium waves indicates

the stochastic origins of the phenomena. Indeed, clusters have a 60–100 nanometers diameter, whereas the distance between clusters is much larger: 3–7 micrometers [84]. Thus, calcium signaling fundamentally relies on stochastic interactions of these channels influencing the overall calcium distribution in the cell’s cytoplasm.

One standard tool for measuring calcium concentration in a cell is via fluorescence data. A fluorescent agent that binds to calcium is introduced and observed under a microscope. Fluorescence ratio data is not a precise measurement of calcium concentration, and calculating calcium concentration from the fluorescence ratio is imprecise [77]. However, fluorescence data can give us an accurate measure of the increase or decrease of calcium and can, therefore, record calcium spikes within the cell.

### **3.2.2 Overview of historical mathematical modeling**

Calcium signaling models, like many real-world models, are categorized into four primary groups. The initial division distinguishes between deterministic and stochastic models, while the second separates spatially homogeneous models from spatially distributed ones. Classical models of calcium signaling are almost entirely deterministic [47]. These models typically assume that the calcium concentration (and signaling molecules such as IP3) within the cell can be well-approximated as homogeneous and the dynamics can be accurately described by ordinary differential equations for the evolution of their concentrations in time [84]. However, recent experimental evidence and mathematical analysis have called this into question [35, 65, 87, 96, 118, 121, 125]. Data shows that (i) the interspike interval can show significant variability, and (ii) intracellular calcium concentrations show steep gradients. It is, therefore, widely postulated that much of the emergent phenomena are, in fact, stochastic in nature [118, 125]. The well-established literature examines how spatially distributed calcium waves and patterns can arise from microscopic stochastic models.

Some of the earliest work is in the Fire-Diffuse-Fire model of Keizer and Smith [87, 88]. In this model, Ryanodine Receptors open (and release calcium) once the ambient calcium exceeds a threshold before closing, entering a refractory state, and then opening again. Keizer and Smith demonstrated that this model exhibits spatially distributed waves. Coombes, Hinch, and Timofeeva [35, 36] developed a similar model. They estimated the release probability for a cluster of channels and determined that it is approximately a sigmoidal function of the local calcium concentration. Keener [85] further extended these works by taking the opening and closing of individual channels to be stochastic.

There have been some efforts towards a detailed derivation of effective macroscopic equations from a microscopic model. Hinch and Chapman, for example, [75] used exponential asymptotic methods to determine the relative frequency of calcium sparks. Falcke et al. employed approximations to determine estimates for the approximate probability of calcium puffs (given the frequency of blips) and also estimates for the probability of a wave throughout the cell [84, 121].

## CHAPTER 4

### NEURAL NETWORK MODEL

This chapter concerns the high-dimensional dynamics of asymmetric random neural networks of the form, for  $j \in I_N = \{1, 2, \dots, N\}$

$$dx_t^j = \left( -x_t^j/\tau + \beta N^{-1/2} \sum_{k=1}^N J^{jk} \lambda(x_t^k) \right) dt + \sigma_t dW_t^j \quad (4.1)$$

Where  $\lambda$  is a Lipschitz function,  $\tau$  is a constant, and  $\{J^{jk}\}_{j,k \in I_N}$  are sampled independently from a centered normal distribution of variance 1,  $\{W_t^j\}_{j,k \in I_N}$  are Brownian Motions. We study the convergence of the double measure (a probability measure on the path space of the system)

$$N^{-1} \sum_{j \in I_N} \delta_{(\mathbf{z}_{[0,T]}^j, \mathbf{G}_{[0,T]}^j)} \quad (4.2)$$

where

$$G_t^j = N^{-1/2} \sum_{k=1}^N J^{jk} \lambda(x_t^k)$$

This work follows the approach of Ben Arous and Guionnet see Subsection [3.1.2](#). We employ the theory of Large Deviations to determine the large N limit of the empirical measure. The main novelty of our approach is as follows. We study the double empirical measure, which includes information on both the spins and the fields. Using a double empirical measure has several advantages: it facilitates accurate finite-dimensional approximations of the dynamics and a broader class of disorder-dependent initial conditions. Grunwald determined the Large Deviations of the Double Empirical Measure for Spin-Glass Dynamics of Jump-Markov Systems [[69](#)]. We also include Replicas (i.e., M copies of the system with the same connectivity but independent

Brownian Motions), which broadens the class of admissible disorder-dependent initial conditions. Taillefumier et al. have employed replicas to study mean field neural networks [7, 8, 144]. Lastly, the function  $\lambda$  can be unbounded, and the diffusion coefficient  $\sigma_t$  can vary over time. The time-varying nature of  $\sigma_t$  is essential for studying how periodic environmental noise in the brain shapes the dynamics of random neural networks. The novelty of implementing these various elements in unison allows us to gain unique insight, mathematical rigor, and tractability.

#### 4.1 Notation and Definitions

Let  $I_N = \{1, 2, \dots, N\}$  be the set of neuron indices. For any Polish space  $\mathcal{X}$ , let  $\mathcal{P}(\mathcal{X})$  denote all probability measures over  $\mathcal{X}$ . The space  $\mathcal{C}([0, T], \mathbb{R})$  is always endowed with the supremum topology (unless indicated otherwise), i.e.

$$\|x_{[0, T]}\| = \sup_{t \in [0, T]} |x_t|$$

For  $\mathbf{y} \in \mathbb{R}^N$ ,  $\|\mathbf{y}\|$  is the Euclidean norm. For any probability measures  $\mu$  and  $\nu$  over a Polish Space, let  $\mathcal{R}(\mu||\nu)$  denotes the relative entropy of measure  $\mu$  with respect to  $\nu$ . For any two measures on the same metric space with metric  $d$ ,  $d_W(\cdot, \cdot)$  indicates the Wasserstein distance, i.e.

$$d_W(\mu, \nu) = \inf_{\zeta} \mathbb{E}^{\zeta} [d(x, y)], \quad (4.3)$$

where the infimum is taken over all  $\zeta$  on the product space such that the marginal of the first variable is equal to  $\mu$  and the marginal of the second variable is equal to  $\nu$ . In the particular case that  $\mu, \nu \in \mathcal{C}([0, T], \mathbb{R}^M)$ , the distance is (unless otherwise indicated)  $d(x, y) = \sup_{t \in [0, T]} \sup_{p \in I_M} |x_t^p - y_t^p|$ .

For any  $\mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^2)$ , we write  $\mu^{(1)}, \mu^{(2)} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M))$  to be the marginals over (respectively) the first  $M$  variables and last  $M$  variables.

## 4.2 Model Description

We are going to rigorously determine the limiting dynamics of multiple replicas (with identical connections  $\mathbf{J}$ , but with independent initial conditions and independent Brownian Motions). We let the superscript  $a$  denote replica  $a \in I_M = \{1, 2, \dots, M\}$ , and consider the system

$$dz_t^{a,j} = (-z_t^{a,j}/\tau + G_t^{a,j})dt + \sigma_t dW_t^{a,j} \text{ where} \quad (4.4)$$

$$G_t^{a,j} = N^{-1/2} \sum_{k \in I_N} J^{jk} \lambda(z_t^{a,k}). \quad (4.5)$$

We assume that  $\lambda \in \mathcal{C}^2(\mathbb{R})$  and that the first derivative is uniformly bounded: this means, in particular, that there is a constant  $C_\lambda$  such that  $|\lambda(x) - \lambda(y)| \leq C_\lambda|x - y|$ . The noise intensity  $t \rightarrow \sigma_t$  is taken to be continuous and non-random, and such that for constants  $\underline{\sigma}$  and  $\bar{\sigma}$ ,

$$0 < \underline{\sigma} \leq \sigma_t \leq \bar{\sigma}. \quad (4.6)$$

Our primary motivation for time-varying diffusivity lies in neuroscience: often, synaptic noise exhibits particular rhythms. It has been of significant interest how these rhythms shape pattern formation [26].

The connectivities  $\{J^{jk}\}$  are taken to be independent centered Gaussian variables, with variance

$$\mathbb{E}[J^{jk} J^{lm}] = \delta(j, l)\delta(k, m).$$

Let  $\gamma^N \in \mathcal{P}(\mathbb{R}^{N^2})$  be their joint probability law.

### 4.3 Initial Condition Assumptions

One can assume that the initial conditions  $(z_0^j)_{j \in I_N}$  are (i) independent of the connectivity and (ii) sampled independently from a  $\mathbb{R}^M$ -valued probabilistic distribution of bounded variance. This distribution is written as  $\hat{\kappa} \in \mathcal{P}(\mathbb{R}^M)$ .

### 4.4 Main Theoretical Results

Our main result is that the empirical measure converges to a fixed point of a mapping  $\Phi : \mathcal{U} \rightarrow \mathcal{U}$ . Here  $\mathcal{U} \subset \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^2)$  will be defined to consist of (i) a broad class of measures with nice regularity properties, and (ii) such that the empirical measure inhabits  $\mathcal{U}$  with unit probability.

One first defines  $(G_t^p)_{p \in I_M, t \in [0, T]}$  to be a centered Gaussian system such that

$$\mathbb{E}[G_t^p G_s^q] = \mathbb{E}^\mu [\lambda(z_t^p) \lambda(z_s^q)].$$

$(z_0^p)_{p \in I_M}$  is independent of  $(G_t^p)_{p \in I_M, t \in [0, T]}$  and is distributed according to  $\hat{\kappa}$ . For Brownian Motions  $(W_{[0, T]}^p)_{p \in I_M}$ , that are independent of  $\mathbf{G}^\mu$ ,  $z_t^p$  is the strong solution of (4.4).

**Theorem 4.4.1.** *There exists a well-defined mapping  $\Phi$  for all  $\mu \in \mathcal{U}$ . Furthermore there exists a unique probability measure  $\xi \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^2)$  such that with unit probability,*

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{j \in I_N} \delta_{(\mathbf{z}_{[0, T]}^j, \mathbf{G}_{[0, T]}^j)} = \xi. \quad (4.7)$$

$\xi$  is the unique measure such that  $\Phi(\xi) = \xi$ . Furthermore,

$$\xi = \lim_{n \rightarrow \infty} \xi^{(n)}, \quad (4.8)$$

where  $\xi^{(n+1)} = \Phi(\xi^{(n)})$  and  $\xi^{(1)}$  is any measure in  $\mathcal{U}$ .

The proof of this theorem will be provided in Chapter 5. This theorem is valuable because it offers an efficient method to determine the large  $N$  limiting equations through the repeated application of the mapping  $\Phi$ . Given that the limiting system is Gaussian, solving for its covariance matrix suffices. For an alternative formulation of the limiting covariance function in terms of a PDE, refer to Helias and Dahmen [72]. In addition, Faugeras, Tanré, and Soret studied the asymptotic behavior of a network of linear Hopfield neurons with random synaptic connections. They demonstrated that the system converges in distribution to the sum of initial values and centered Gaussian processes. They showed that these processes, which are not Markovian, can be expressed using modified Bessel functions [56].

## CHAPTER 5

### LARGE DEVIATION PRINCIPLES FOR NEURAL NETWORKS

The main goal of this section is to prove Theorem 4.4.1 employing the theory of Large Deviations [45]. The method - similar to the original work by Ben Arous and Guionnet [10] - is to (i) prove a Large Deviations Principle for the uncoupled system and then (ii) perform an exponential change-of-measure using Girsanov's Theorem to obtain the Large Deviations Principle for the coupled system, before (iii) proving that the rate function has a unique zero.

The main differences between this work and the early papers of Ben Arous and Guionnet are that we (i) study the convergence of the double empirical measure (4.2) (whereas Ben Arous and Guionnet study the convergence of the 'annealed empirical measure' in their earlier papers [10]. In the later works [11, 12] quenched asymptotics are determined) and (ii) we employ replicas.

#### 5.1 Large Deviations of the Uncoupled System

We start by stating a LDP for the uncoupled system. Define the uncoupled dynamics,

$$y_t^{p,j} = z_0^{p,j} + \int_0^t \sigma_s dW_s^{p,j}, \quad (5.1)$$

and let  $P_{z_0}^N \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^N)$  be the law of  $\{y_{[0, T]}^j\}_{j \in I_N}$ , conditioned on  $y_0$  being equal to  $z_0$ .

We establish a LDP for the uncoupled system by locally freezing the dependence of the fields  $\{\tilde{G}_t^{p,j}\}$  on the empirical measure. To do this, we must first define a regular subset  $\mathcal{Q}_\alpha$  (for a positive integer  $\alpha \gg 1$ ) that is such that (i) the empirical measure  $\hat{\mu}^N(\mathbf{y}) = N^{-1} \sum_{j \in I_N} \delta_{\mathbf{y}_{[0, T]}^j} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M))$  inhabits with high probability

and (ii) there exist uniform bounds on the fluctuations in time. To this end, writing  $\mathcal{K}_a$  to be the compact set specified in Lemma A.2, define the set

$$\mathcal{Q}_a = \left\{ \mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)) : \mu \in \mathcal{K}_a \text{ and } \sup_{p \in I_M} \mathbb{E}^\mu \left[ \sup_{t \in [0, T]} (y_t^p)^2 \right] \leq a \text{ and} \right. \\ \left. \text{For all integers } m \geq a \text{ it holds that } \sup_{0 \leq i \leq m} \mathbb{E}^\mu \left[ \sup_{p \in I_M} (w_{t_{i+1}^{(m)}}^p - w_{t_i^{(m)}}^p)^2 \right] \leq \Delta_m^{1/4} \right\} \quad (5.2)$$

where  $\Delta_m = T/m$  and  $t_i^{(m)} = iT/m$ . Write

$$\Omega = \bigcup_{a \geq 1} \Omega_a. \quad (5.3)$$

**Lemma 5.1.1.** *For any  $L > 0$ , there exists  $a > 0$  such that*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{Q}_a) \leq -L \quad (5.4)$$

The above lemma is proved in the Appendix A. Next, for any  $\nu \in \mathcal{Q}$ , we define a centered Gaussian law  $\beta_\nu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M))$  as follows. We stipulate that  $\beta_\nu$  is the law of Gaussian random variables  $\{G_t^{\nu, p}\}_{t \in [0, T], p \in I_M}$  with covariance structure

$$\mathbb{E}^{\beta_\nu} [G_s^{\nu, p} G_t^{\nu, q}] = \mathbb{E}^\nu [\lambda(x_s^p) \lambda(x_t^q)] \quad (5.5)$$

This definition will be useful because for any  $j \in I_N$ , the law of  $\tilde{G}_{[0, T]}^j$  under  $\gamma^N$  is  $\beta_{\hat{\mu}^N(\mathbf{y})}$ . In the following Lemma we collect some regularity estimates for the Gaussian Law  $\beta_\nu$ .

**Lemma 5.1.2.** *(i)  $\beta_\nu$  is a well-defined Gaussian probability law. (ii) Furthermore, the map  $t \rightarrow G_t^{\nu, p}$  is ‘uniformly continuous’ for all measures in  $\mathcal{U}_a$ , in the following*

sense. For any  $a > 0$ , and any  $\epsilon > 0$ , there exists  $\delta(a, \epsilon)$  such that for all  $\nu \in \mathcal{U}_a$ ,

$$\sup_{\nu \in \mathcal{U}_a} \sup_{p \in I_M} \mathbb{E}^{\beta_\nu} \left[ \sup_{s, t \in [0, T]; |s-t| \leq \delta(a, \epsilon)} |G_s^{\nu, p} - G_t^{\nu, p}| \right] \leq \epsilon \quad (5.6)$$

Write  $Q^N = \gamma^N \otimes P^N$  to be the law of the random variables  $(\mathbf{y}, \mathbf{G})$ , and for any  $\nu \in \mathfrak{Q}$ , define  $S_\nu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^2)$  to be  $S_\nu = P \otimes \beta_\nu$ .

We then arrive at the following LDP for the uncoupled system,

**Theorem 5.1.3.** *Let  $\mathcal{A}, \mathcal{O} \in \mathcal{B}(\mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^2))$ , such that  $\mathcal{O}$  is open and  $\mathcal{A}$  closed.*

*Then*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q^N(\hat{\mu}^N(\mathbf{y}_{[0, T]}, \mathbf{G}_{[0, T]}) \in \mathcal{A}) \leq - \inf_{\mu \in \mathcal{A}} \mathcal{R}(\mu || S_{\mu^{(1)}}) \quad (5.7)$$

$$\underline{\lim}_{N \rightarrow \infty} N^{-1} \log Q^N(\hat{\mu}^N(\mathbf{y}_{[0, T]}, \tilde{\mathbf{G}}_{[0, T]}) \in \mathcal{O}) \geq - \inf_{\mu \in \mathcal{O}} \mathcal{R}(\mu || S_{\mu^{(1)}}). \quad (5.8)$$

Here the rate function  $\mu \rightarrow \mathcal{R}(\mu || S_{\mu^{(1)}})$  is lower semi-continuous and has compact level sets.

## 5.2 Large Deviations of the Coupled System

We now specify the operator  $\tilde{\Phi} : \mathcal{U} \rightarrow \mathcal{U}$ . Fix  $\mu \in \mathcal{U}$  and define  $\tilde{\Phi}(\mu)$  to be the law of processes  $(z_{[0, T]}^p, G_{[0, T]}^p)_{p \in I_M; t \in [0, T]}$ . One first defines  $(G_t^p)_{p \in I_M; t \in [0, T]}$  to be a centered Gaussian system such that

$$\mathbb{E}[G_t^p G_s^q] = \mathbb{E}^\mu[\lambda(z_t^p) \lambda(z_s^q)].$$

$(z_0^p)_{p \in I_M}$  is defined to be independent of  $(G_t^p)_{p \in I_M; t \in [0, T]}$  and distributed according to  $\hat{\kappa}$ . Letting  $(W_{[0, T]}^p)_{p \in I_M}$  be Brownian Motions that are independent of  $\mathbf{G}^\mu$ , we define

$(z_t^p)_{p \in I_M; t \in [0, T]}$  to be the strong solution to the stochastic differential equation

$$dz_t^p = \left( -\tau^{-1} z_t^p + G_t^{\mu, p} \right) dt + \sigma_t dW_t^p. \quad (5.9)$$

We now arrive at the LDP for the coupled system and our main result on the convergence of the empirical measure for the system  $\hat{\mu}$  to a unique measure  $\xi$  as given in Theorem 4.4.1.

**Theorem 5.2.1.** *For any  $\epsilon > 0$ ,*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \xi) \geq \epsilon) < 0. \quad (5.10)$$

*Thanks to the Borel-Cantelli Lemma, this implies that with unit probability,*

$$\lim_{N \rightarrow \infty} \hat{\mu}^N(\mathbf{z}, \mathbf{G}) = \xi. \quad (5.11)$$

*Furthermore,*

$$\xi = \lim_{n \rightarrow \infty} \xi^{(n)}, \quad (5.12)$$

*where  $\xi^{(n+1)} = \tilde{\Phi}(\xi^{(n)})$  and  $\xi^{(1)}$  is any measure in  $\mathcal{U}$ .*

We have divided the proofs into four main sections. In Section 5.3, we prove general regularity properties of the stochastic processes. In Section 5.4, we prove that the empirical measure inhabits a compact set with arbitrarily high probability. In Section 5.5, we prove the LDP for the uncoupled system. In Section 5.6, we determine the limiting dynamics of the coupled system.

### 5.3 Regularity Estimates and Compactness

We first prove Lemma 5.1.2.

*Proof.* We first check that the covariance function is positive definite (when restricted to a finite set of times). Let  $\{t_i\}_{1 \leq i \leq m} \subset [0, T]$  be a finite set of times. Then evidently for any constants  $\{\alpha_i^p\}_{p \in I_M, 1 \leq i \leq m}$ , it must be that

$$\sum_{p, q \in I_M} \sum_{1 \leq i, j \leq m} \alpha_i^p \alpha_j^q \mathbb{E}^\nu [\lambda(x_{t_i}^p) \lambda(x_{t_j}^q)] = \mathbb{E}^\nu \left[ \left( \sum_{p \in I_M} \sum_{1 \leq i \leq m} \alpha_i^p \lambda(x_{t_i}^p) \right)^2 \right] \geq 0. \quad (5.13)$$

This means that there exists a finite set of centered Gaussian variables

$\{G_{t_i}^{\nu, p}\}_{p \in I_M, 1 \leq i \leq m}$  such that (5.5) holds. It then follows from the Komolgorov Extension Theorem that  $\beta_\nu$  is well-defined on any countably dense subset of times of  $[0, T]$ . It remains for us to demonstrate continuity, i.e., that a Gaussian probability law exists such that (5.5) holds for all time. We do this using standard theory for the continuity of Gaussian Processes (following Chapter 2 of [1]).

First, we notice that as  $G_t^{\nu, p}$  are given to be centered Gaussian variables, therefore they have bounded second moment (finite variance) by definition of the variance of Gaussian variables, so we have

$$\sup_{p \in I_M} \sup_{t \in [0, T]} \mathbb{E} \left[ (G_t^{\nu, p})^2 \right] < \infty. \quad (5.14)$$

Now define the canonical metric,

$$\bar{d}_p(s, t) = \mathbb{E} \left[ (G_s^{\nu, p} - G_t^{\nu, p})^2 \right]^{\frac{1}{2}} = \mathbb{E}^\nu \left[ (\lambda(x_s^p) - \lambda(x_t^p))^2 \right]^{\frac{1}{2}} \quad (5.15)$$

$$\leq \text{Const} \sup_{p \in I_M} \mathbb{E}^\nu \left[ |x_s^p - x_t^p|^2 \right]^{\frac{1}{2}} \leq a (t - s)^{\frac{1}{4}} \quad (5.16)$$

thanks to properties of the set  $\mathcal{Q}_a$ , for all  $s, t$  such that  $|s - t|$  is smaller than some constant depending on  $a$ . It follows from Theorem 1.4.1 of [1] that the Gaussian Process  $(G_t^{\nu, p})_{t \in [0, T]}$  is almost-surely continuous.

Write  $B_t(\epsilon) = \{s \in [0, T] : \bar{d}(s, t) \leq \epsilon\}$  to be the  $\epsilon$ -ball about  $t$ , and let  $\mathcal{N}(\epsilon)$  denote the smallest number of such balls that cover  $T$ . We see that there exists a

constant  $\mathbf{c}_a > 0$  such that

$$\mathcal{N}(\epsilon) \leq \mathbf{c}_a \epsilon^{-4}. \quad (5.17)$$

Writing  $H(\epsilon) = \log \mathcal{N}(\epsilon)$ , it follows from Theorem 1.3.5 in [1] that there exist  $M$  Gaussian Processes  $(G_t^{\nu,p})_{t \in [0,T]}$  such that  $t \rightarrow G_t^{\nu,p}$  is almost-surely continuous, and there exists a universal constant  $\mathfrak{K} > 0$  and a random  $\eta > 0$  such that for all  $\delta < \eta$ ,

$$\sup_{p \in I_M; s, t \leq T; \bar{d}(s,t) \leq \delta} |G_s^{\nu,p} - G_t^{\nu,p}| \leq \mathfrak{K} \int_0^\delta H^{1/2}(\epsilon) d\epsilon \quad (5.18)$$

$$\leq \mathfrak{K} \int_0^\delta (4 \log(\epsilon^{-1}) + \log \mathbf{c}_a)^{\frac{1}{2}} d\epsilon, \quad (5.19)$$

and we note that the above goes to 0 as  $\delta \rightarrow 0^+$ . This also implies (5.6).  $\square$

The following bound on the operator norm of the connectivity matrix is well-known (and the proof is omitted).

**Lemma 5.3.1.** *For any  $L > 0$ , there exists  $\ell$  such that*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \geq \ell) \leq -L, \quad (5.20)$$

where  $\mathcal{J}_N \in \mathbb{R}^{N \times N}$  has  $(j, k)$  entry

$$\mathcal{J}_{N,jk} = N^{-1/2} J^{jk}$$

**Lemma 5.3.2.** *For any  $\ell > 0$ , there exists  $L > 0$  such that for all  $p \in I_M$  and all  $N \geq 1$ ,*

$$N^{-1} \log \mathbb{P}(\mathcal{A}_c, \sup_{t \in [0,T]} \sum_{j \in I_N} (z_t^{p,j})^2 \geq N\ell) \leq -L \quad (5.21)$$

where

$$\mathcal{A}_c = \left\{ \|\mathcal{J}_N\| \leq c, \sup_{p \in I_M} \sum_{j \in I_N} (z_0^{p,j})^2 \leq N \mathbb{E}^\kappa[(z_0^p)^2] + N \right\}.$$

*Proof.* Write

$$u_t = N^{-1} \sum_{j \in I_N} (z_t^{p,j})^2.$$

If the event  $\mathcal{A}_c$  holds, then thanks to Ito's Lemma it must be that

$$du_t = \left\{ -2\tau^{-1}u_t + 1 + 2N^{-1} \sum_{j \in I_N} z_t^{p,j} G_t^{p,j} \right\} dt + 2N^{-1} \sum_{j \in I_N} z_t^{p,j} dW_t^{p,j} \quad (5.22)$$

$$\leq \left\{ -2\tau^{-1}u_t + 1 + cC_\lambda u_t \right\} dt + 2N^{-1} \sum_{j \in I_N} z_t^{p,j} dW_t^{p,j}, \quad (5.23)$$

since  $N^{-1} \sum_{j \in I_N} \lambda (z_t^{p,j})^2 \leq C_\lambda^2 u_t$ . Write

$$v_t = \sup_{s \in [0,t]} 2N^{-1} \left| \sum_{j \in I_N} \int_0^s z_s^{p,j} dW_s^{p,j} \right|, \quad (5.24)$$

and define the stopping time for a constant  $A > 0$ ,

$$\tau_A = \inf \{ t \geq 0 : v_t \geq \exp(At) + A \}. \quad (5.25)$$

Gronwall's Inequality implies that for all  $t \leq \tau_A$ ,

$$u_t \leq (A + u_0 + t) \exp(\tilde{c}t)$$

where  $\tilde{c} = A + cC_\lambda - 2\tau^{-1}$ . The quadratic variation of  $x(t) := N^{-1} \sum_{j \in I_N} \int_0^t z_s^{p,j} dW_s^{p,j}$  is

$$(QV)_t^N = N^{-2} \sum_{j \in I_N} \int_0^t (z_s^{p,j})^2 ds. \quad (5.26)$$

For all  $t \leq \tau_A$ ,

$$(QV)_t^N \leq N^{-1} \tilde{c}^{-1} (A + u_0 + t) \exp(\tilde{c}t) := N^{-1} h_t, \quad (5.27)$$

and notice that  $h_t$  is independent of the Brownian Motions. Now define the stochastic process  $w(t)$  to be such that

$$w(t) = x(\alpha_t^N) \text{ where} \quad (5.28)$$

$$\alpha_t^N = \inf \{s \geq 0 : (QV)_s^N = t\} \quad (5.29)$$

Thanks to the time-rescaled representation of a stochastic integral,  $w(t)$  is a Brownian Motion [82]. Writing  $f(t) = \exp(At) + A$ , it follows that

$$\begin{aligned} & \mathbb{P} \left( \text{There exists } s \leq T \text{ such that } |x(s)| \geq f(s) \right) \\ & \leq \mathbb{P} \left( \text{There exists } s \leq T \text{ such that } |w(N^{-1}h_s)| \geq f(s) \right) \\ & \leq \mathbb{P} \left( \text{There exists } s \leq T \text{ such that } |w(N^{-1}h_{s(m)})| \geq f(s(m)) \right) \end{aligned}$$

and we have written

$$s^{(m)} = \inf \{t_a^{(m)} : t_a^{(m)} \geq s\} \quad (5.30)$$

$$s_{(m)} = \sup \{t_a^{(m)} : t_a^{(m)} \leq s\}. \quad (5.31)$$

and we recall that  $t_a^{(m)} = Ta/m$ . Employing a union-of-events bound,

$$\begin{aligned} & \mathbb{P} \left( \text{There exists } s \leq T \text{ such that } |w(h_{s(m)})| \geq f(s(m)) \right) \leq \\ & \sum_{a=0}^{m-1} \left\{ \mathbb{P} \left( w(N^{-1}h_{t_{a+1}^{(m)}}) \geq f(t_a^{(m)}) \right) + \mathbb{P} \left( w(N^{-1}h_{t_{a+1}^{(m)}}) \leq -f(t_a^{(m)}) \right) \right\} \quad (5.32) \end{aligned}$$

Now since  $w(t)$  is centered and Gaussian, with variance of  $t$ ,

$$N^{-1} \log \mathbb{P} \left( w(N^{-1}h_{t_{a+1}^{(m)}}) \geq f(t_a^{(m)}) \right) = -\frac{N}{2} f(t_a^{(m)})^2 (h_{t_{a+1}^{(m)}})^{-1} + O(\log N) \quad (5.33)$$

$$N^{-1} \log \mathbb{P} \left( w(N^{-1}h_{t_{a+1}^{(m)}}) \leq -f(t_a^{(m)}) \right) = -\frac{N}{2} f(t_a^{(m)})^2 (h_{t_{a+1}^{(m)}})^{-1} + O(\log N). \quad (5.34)$$

We fix  $m = A$  and take  $A$  to be arbitrarily large. Then

$$\lim_{A \rightarrow \infty} \inf_{0 \leq a \leq m-1} f(t_a^{(m)})^2 (h_{t_{a+1}^{(m)}})^{-1} = \infty.$$

We thus find that, for large enough  $A$ ,

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P} \left( \mathcal{A}_c, \text{ There exists } s \leq T \text{ such that } |x(s)| \geq f(s) \right) \leq -L. \quad (5.35)$$

We have already demonstrated in the course of the proof that if the event  $\mathcal{A}_c$  holds, and  $\sup_{s \in [0, T]} |x(s)| \leq f(s)$ , then there exists a constant such that  $\sup_{t \in [0, T]} u_t \leq \text{Const.}$

We have thus established the Lemma.  $\square$

The following  $L^2$ -Wasserstein distance provides a very useful way of controlling the dependence of the fields  $(G_t^\nu)$  on the measure  $\nu$ . Define  $d_t^{(2)}(\cdot, \cdot)$  to be such that for any  $\mu, \nu \in \mathcal{U}$ ,

$$d_t^{(2)}(\mu, \nu) = \inf_{\zeta} \mathbb{E}^{\zeta} \left[ \sum_{p \in I_M} \int_0^t \{ (y_s^p - \tilde{y}_s^p)^2 + (G_s^p - \tilde{G}_s^p)^2 \} ds \right]^{1/2}, \quad (5.36)$$

where the infimum is over all  $\zeta \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^{2M}) \times \mathcal{C}([0, T], \mathbb{R}^{2M}))$ , such that the law of the first  $2M$  processes is given by  $\mu$ , and the law of the last  $2M$  processes is given by  $\nu$ . Let  $d^{(2)}(\mu, \nu) := d_T^{(2)}(\mu, \nu)$ .

**Lemma 5.3.3.** For any  $a > 0$ ,  $d^{(2)}(\cdot, \cdot)$  metrizes weak convergence in  $\mathcal{U}_a$ . Furthermore,

$$\lim_{\epsilon \rightarrow 0^+} \sup \{d_W(\mu, \nu) : \mu, \nu \in \mathcal{U}_a \text{ and } d^{(2)}(\mu, \nu) \leq \epsilon\} = 0. \quad (5.37)$$

*Proof.* Since  $\mathcal{U}_a$  is compact, Prokhorov's Theorem implies that for any  $\tilde{\epsilon} > 0$ , there exists a compact set  $\mathcal{D}_\epsilon \subset \mathcal{C}([0, T], \mathbb{R}^M)^2$  such that for all  $\mu \in \mathcal{U}_a$ ,

$$\mu(\mathcal{D}_\epsilon) \geq 1 - \tilde{\epsilon}. \quad (5.38)$$

Since  $\mathcal{D}_\epsilon$  is compact, it follows from the Arzela-Ascoli Theorem that for any  $\delta > 0$ , there exists  $v(\epsilon, \delta)$  such that for all  $f, g \in \mathcal{D}_\epsilon$  such that for all  $p \in I_{2M}$ ,

$$\int_0^T (f^p(t) - g^p(t))^2 dt \leq v(\epsilon, \delta), \quad (5.39)$$

it necessarily holds that

$$\sup_{p \in I_M} \sup_{t \in [0, T]} |f^p(t) - g^p(t)| \leq \delta. \quad (5.40)$$

Let  $\zeta$  be any measure that is within  $\eta \ll 1$  of realizing the infimum in (5.36). Then, writing

$$\mathcal{A}_\epsilon = \chi \{ \text{For each } p \in I_M, y^p, \tilde{y}^p, g^p, \tilde{g}^p \in \mathcal{D}_\epsilon \},$$

we have the bound

$$\begin{aligned} & \mathbb{E}^\zeta \left[ \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right] \\ & \leq \mathbb{E}^\zeta \left[ \left( \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right) \mathcal{A}_\epsilon \right] + \\ & \mathbb{E}^\zeta \left[ \left( \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right) (1 - \mathcal{A}_\epsilon) \right] \quad (5.41) \end{aligned}$$

Now we take  $d^{(2)}(\mu, \nu) \rightarrow 0^+$ , and  $\eta \rightarrow 0^+$  too. Since  $\mathcal{A}_\epsilon$  is closed, thanks to the Portmanteau Theorem, we thus find that for any  $\epsilon > 0$ ,

$$\mathbb{E}^\zeta \left[ \mathcal{A}_\epsilon \sum_{p \in I_M} \int_0^T \{ (y_s^p - \tilde{y}_s^p)^2 + (G_s^p - \tilde{G}_s^p)^2 \} ds \right] \rightarrow 0. \quad (5.42)$$

which in turn implies that (making use of the uniform convergence over  $\mathcal{D}_\epsilon$  in (5.40))

$$\mathbb{E}^\zeta \left[ \left( \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right) \mathcal{A}_\epsilon \right] \rightarrow 0. \quad (5.43)$$

For the other term on the RHS of (5.41), for  $b > 0$ , write

$$\mathcal{B}_b = \chi \left\{ \begin{array}{l} \text{For each } p \in I_M, \sup_{t \in [0, T]} |y_t^p| \leq b, \sup_{t \in [0, T]} |\tilde{y}_t^p| \leq b, \\ \sup_{t \in [0, T]} |g_t^p| \leq b, \sup_{t \in [0, T]} |\tilde{g}_t^p| \leq b \end{array} \right\}$$

Then,

$$\begin{aligned} & \mathbb{E}^\zeta \left[ \left( \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right) (1 - \mathcal{A}_\epsilon) \right] \\ & \leq \mathbb{E}^\zeta \left[ \left( \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right) (1 - \mathcal{A}_\epsilon) \mathcal{B}_b \right] \\ & + \mathbb{E}^\zeta \left[ \left( \sup_{p \in I_M} \sup_{t \in [0, T]} |y^p(t) - \tilde{y}^p(t)| + \sup_{p \in I_M} \sup_{t \in [0, T]} |g^p(t) - \tilde{g}^p(t)| \right) (1 - \mathcal{A}_\epsilon) (1 - \mathcal{B}_b) \right]. \end{aligned} \quad (5.44)$$

Thanks to the fact that, for all  $\mu \in \mathcal{U}_a$ ,

$$\sup_{p \in I_M} \mathbb{E}^\mu \left[ \sup_{t \in [0, T]} (y_t^p)^2 \right] \leq a,$$

one finds that the second term on the RHS of (5.44) goes to 0 as  $b \rightarrow \infty$ , uniformly for all  $\epsilon > 0$  and all  $\mu, \nu \in \mathcal{U}_a$ . For any fixed  $b \gg 1$ , the first term on the RHS of (5.44) must go to zero as  $\epsilon \rightarrow 0^+$ , thanks to (5.38). We have thus proved the Lemma.  $\square$

For  $\mu \in \mathfrak{Q}$ , we define  $d_t^{(2)}(\mu, \nu)$  analogously to (5.36).

**Lemma 5.3.4.** *There exists a constant  $\mathfrak{C} > 0$  such that for all  $\mu, \nu \in \mathfrak{Q}$  and all  $t \in [0, T]$ ,*

$$d_t^{(2)}(\beta_\nu, \beta_\mu) \leq \mathfrak{C} d_t^{(2)}(\nu, \mu). \quad (5.45)$$

*Also for all  $\mu, \nu \in \mathfrak{Q}$  such that for some  $b > 0$ ,  $\det(\mathfrak{Y}_{\mu,0}), \det(\mathfrak{Y}_{\nu,0}) \geq b > 0$ , there exists a constant  $\mathfrak{C}_b$  such that*

$$d_t^{(2)}(\beta_{\nu,\mathbf{g}}, \beta_{\mu,\mathbf{g}}) \leq \tilde{\mathfrak{C}}_b (1 + \|\mathbf{g}\|) d_t^{(2)}(\nu, \mu), \quad (5.46)$$

and  $\|\cdot\|$  is the Euclidean norm on  $\mathbb{R}^M$ .

## 5.4 Exponential Tightness

To prove a Large Deviation Principle, one requires that the empirical measure inhabits a compact set with arbitrarily high probability. For any  $\mathbf{y} \in \mathcal{C}([0, T], \mathbb{R}^M)^N$ , write  $\tilde{\gamma}_{\mathbf{y}}^N \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^N)$  to be the law of the random variables  $(\tilde{G}_t^{p,j})_{j \in I_N; p \in I_M; t \in [0, T]}$ .

The following lemmas are needed for this proof.

**Lemma 5.4.1.** *For any  $L > 0$ , there exists a compact set  $\tilde{\mathcal{C}}_L \subset \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M))$  such that the following holds. For any  $N \geq 1$ , and any  $\{\mathbf{y}_{[0, T]}^j\}_{j \in I_N}$  such that  $\hat{\mu}^N(\mathbf{y}) \in \mathcal{Q}_L$ ,*

$$N^{-1} \log \tilde{\gamma}_{\mathbf{y}}^N(\hat{\mu}^N(\mathbf{G}) \notin \tilde{\mathcal{C}}_L) \leq -L. \quad (5.47)$$

For  $\mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R})^M \times \mathcal{C}([0, T], \mathbb{R})^M)$ , write  $\mu^{(1)} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R})^M)$  to be the marginal of  $\mu$  over its first  $M$  variables, and  $\mu^{(2)}$  to be the marginal of  $\mu$  over its last  $M$  variables. Next, define the set

$$\mathcal{U}_a = \left\{ \mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R})^M \times \mathcal{C}([0, T], \mathbb{R})^M) : \mu^{(1)} \in \mathcal{Q}_a, \mu^{(2)} \in \tilde{\mathcal{C}}_a \text{ and} \right. \\ \left. \sup_{t \in [0, T]} \sup_{p \in I_M} \mathbb{E}^\mu[(G_t^p)^2] \leq C_\lambda^2 a, \text{ for all } 0 \leq s, t \leq T, \quad (5.48) \right. \\ \left. \sup_{p \in I_M} \mathbb{E}^\mu[(G_t^p - G_s^p)^2] \leq a C_\lambda^2 |t - s|^{1/2} \right\},$$

and let

$$\mathcal{U} = \bigcup_{a \geq 0} \mathcal{U}_a. \quad (5.49)$$

It follows immediately from the above definition that  $d_W(\mu, \nu) < \infty$  for any  $\mu, \nu \in \mathcal{U}$ .

We can now prove an ‘exponential tightness’ result.

**Lemma 5.4.2.** *For any  $a \geq 0$ ,  $\mathcal{U}_a$  is compact. For any  $L > 0$ , there exists  $a > 0$  such that*

$$\overline{\lim}_{N \rightarrow \infty} \sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N (\hat{\mu}^N \notin \mathcal{U}_a) \leq -L. \quad (5.50)$$

*Proof.* Since the sets  $\mathcal{Q}_a$  and  $\tilde{\mathcal{C}}_a$  are compact, this follows almost immediately from Lemma 5.4.1. □

## 5.5 Proof of Large Deviations of the Uncoupled System

In this section, we prove Theorem 5.1.3.

For some  $\nu \in \mathfrak{Q}$ , let  $Q_\nu^N \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^N \times \mathcal{C}([0, T], \mathbb{R}^M)^N)$  be the joint law of the uncoupled system, i.e.

$$Q_\nu^N = (\beta_\nu \otimes P_{\mathbf{z}})^{\otimes N}. \quad (5.51)$$

Define the empirical measure  $\tilde{\mu}^N \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R})^M \times \mathcal{C}([0, T], \mathbb{R})^M)$  to be

$$\tilde{\mu}^N = N^{-1} \sum_{j \in I_N} \delta_{\mathbf{y}_{[0, T]}^j, \tilde{\mathbf{G}}_{[0, T]}^{\nu, j}}, \quad (5.52)$$

where we recall that

$$y_t^p = z_0^p + \int_0^t \sigma_s dW_s^p. \quad (5.53)$$

**Lemma 5.5.1.** *Fix some  $\nu \in \mathcal{U}$ . Let  $\mathcal{A}, \mathcal{O} \subseteq \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)^2)$ , such that  $\mathcal{O}$  is open and  $\mathcal{A}$  closed. Then*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_\nu^N(\tilde{\mu}^N(\mathbf{y}_{[0, T]}, \mathbf{G}_{[0, T]}^\nu) \in \mathcal{A}) \leq - \inf_{\mu \in \mathcal{A}} \mathcal{R}(\mu || S_\nu) \quad (5.54)$$

$$\underline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_\nu^N(\tilde{\mu}^N(\mathbf{y}_{[0, T]}, \mathbf{G}_{[0, T]}^\nu) \in \mathcal{O}) \geq - \inf_{\mu \in \mathcal{O}} \mathcal{R}(\mu || S_\nu). \quad (5.55)$$

Furthermore  $\mu \rightarrow \mathcal{R}(\mu || S_\nu)$  is lower semi-continuous, and has compact level sets.

*Proof.* This is a consequence of Sanov's Theorem [44]. □

We will now prove Theorem 5.1.3 as follows.

*Proof.* We start with the upper bound (5.7). We write  $\hat{\mu}^N := \hat{\mu}^N(\mathbf{y}_{[0,T]}, \mathbf{G}_{[0,T]})$ . Using a union-of-events bound, for any  $a > 0$ ,

$$\begin{aligned} \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{A}) &\leq \\ &\max \left\{ \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{A} \cap \mathcal{U}_a), \right. \\ &\quad \left. \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \notin \mathcal{U}_a) \right\} \\ &\leq \max \left\{ \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{A} \cap \mathcal{U}_a), -L \right\}, \end{aligned} \quad (5.56)$$

for any  $L > 0$ , as long as  $a$  is sufficiently large, thanks to the exponential tightness proved in Lemma 5.4.2. By taking  $a \rightarrow \infty$ , it thus suffices that we prove that for arbitrary  $\mathcal{U}_a$  such that  $\mathcal{A} \cap \mathcal{U}_a \neq \emptyset$ ,

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{A} \cap \mathcal{U}_a) = - \inf_{\mu \in \mathcal{A} \cap \mathcal{U}_a} \mathcal{R}(\mu || S_{\mu^{(1)}}). \quad (5.57)$$

Since  $\mathcal{A} \cap \mathcal{U}_a$  is compact, for any  $\epsilon > 0$ , we can always find an open covering of the form, for some positive integer  $\mathcal{N}_\epsilon$ ,  $\{\zeta_i\}_{1 \leq i \leq \mathcal{N}_\epsilon} \subseteq \mathcal{A} \cap \mathcal{U}_a$ ,

$$\mathcal{A} \cap \mathcal{U}_a \subseteq \bigcup_{i=1}^{\mathcal{N}_\epsilon} B_\epsilon(\zeta_i). \quad (5.58)$$

We thus find that

$$\begin{aligned} \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{A} \cap \mathcal{U}_a) \\ \leq \sup_{1 \leq i \leq \mathcal{N}_\epsilon} \left\{ \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in B_\epsilon(\zeta_i)) \right\}. \end{aligned} \quad (5.59)$$

Thus, employing Lemma 5.1.2 in the third line below,

$$\begin{aligned}
& \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N (\hat{\mu}^N \in B_\epsilon(\zeta_i)) \\
&= \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{E}^{P_{\mathbf{z}_0}^{PN}} \left[ \gamma_{\mathbf{y}, \mathbf{g}_0}^N \left( \hat{\mu}^N \in B_\epsilon(\zeta_i) \right) \right] \\
&= \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{E}^{P_{\mathbf{z}_0}^{PN}} \left[ \tilde{\gamma}_{\hat{\mu}^N(\mathbf{y}), \mathbf{g}_0}^N \left( \hat{\mu}^N \in B_\epsilon(\zeta_i) \right) \right] \\
&\leq \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{E}^{P_{\mathbf{z}_0}^{PN}} \left[ \sup_{\nu \in B_\epsilon(\zeta_i)} \gamma_{\nu^{(1)}, \mathbf{g}_0}^N \left( \hat{\mu}^N \in B_\epsilon(\zeta_i) \right) \right] \\
&= - \inf_{\mu, \nu \in B_\epsilon(\zeta_i)} \mathcal{R}(\mu \| S_\nu),
\end{aligned} \tag{5.60}$$

thanks to Lemma 5.5.1. We thus find that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N (\hat{\mu}^N \in \mathcal{A} \cap \mathcal{U}_a) \leq - \inf_{1 \leq i \leq N_\epsilon} \inf_{\nu, \mu \in B_\epsilon(\zeta_i)} \mathcal{R}(\mu \| S_\nu). \tag{5.61}$$

Now, it is proved in Lemma 5.3.4 that  $\nu \rightarrow Q_{\nu, \mathbf{z}_0, \mathbf{g}_0}$  is continuous. Since the Relative Entropy is lower-semi-continuous in both of its arguments, we thus find that the following map is lower-semi-continuous,

$$(\nu, \mu) \rightarrow \mathcal{R}(\mu \| S_\nu).$$

Thus taking  $\epsilon \rightarrow 0^+$ , we obtain that

$$\lim_{\epsilon \rightarrow 0^+} \inf_{1 \leq i \leq N_\epsilon} \inf_{\nu, \mu \in B_\epsilon(\zeta_i)} \mathcal{R}(\mu \| S_\nu) = \inf_{\mu \in \mathcal{A} \cap \mathcal{U}_a} \mathcal{R}(\mu \| S_{\mu^{(1)}}), \tag{5.62}$$

and we have proved (5.57).

Turning to the lower bound (5.8), consider an arbitrary open set  $\mathcal{O}$ . If  $\mathcal{O} \cap \mathcal{U} = \emptyset$ , then

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N (\hat{\mu}^N \in \mathcal{O}) = -\infty = - \inf_{\mu \in \mathcal{O}} \mathcal{R}(\mu \| S_{\mu^{(1)}})(\mu),$$

since  $\mathcal{R}(\mu \| S_{\mu^{(1)}})$  is identically  $\infty$  outside of  $\mathcal{U}$ . In this case, its clear that (5.55) holds.

We can thus assume that  $\mathcal{O} \cap \mathcal{U} \neq \emptyset$ . Let  $\mu \in \mathcal{O}$  be such that  $\mu$  is in the interior of  $\mathcal{U}_a$ , for some  $a > 0$ . We can thus find a sequence of neighborhoods  $\{\mathcal{N}_i\}_{i \geq 1}$  of  $\mu$  such that  $\mathcal{N}_j \subseteq \mathcal{O} \cap \mathcal{U}_a \cap B_{j^{-1}}(\mu)$ . We thus find that for any  $j \geq 1$ ,

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{O}) \geq \overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{N}_j). \quad (5.63)$$

Similarly to the bound for the closed sets, we obtain that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{N}_j) \geq - \sup_{\nu \in \mathcal{N}_j} \inf_{\mu \in \mathcal{N}_j} \mathcal{R}(\mu \| S_\nu). \quad (5.64)$$

Taking  $j \rightarrow \infty$ , since  $(\nu, \mu) \rightarrow \mathcal{R}(\mu \| S_\nu)$  is lower semicontinuous, it must be that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{N}_j) \geq -\mathcal{R}(\mu \| S_{\mu^{(1)}}). \quad (5.65)$$

Since  $\mu \in \mathcal{O}$  is arbitrary, it must be that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N \in \mathcal{O}) \geq - \inf_{\mu \in \mathcal{O}} \mathcal{R}(\mu \| S_{\mu^{(1)}}). \quad (5.66)$$

□

## 5.6 Proof of Large Deviations of the Coupled System

Girsanov's Theorem implies that

$$\frac{dP_{\mathbf{J}, \mathbf{z}_0}^N}{dP_{\mathbf{z}_0}^N} \Big|_{\mathcal{F}_T}(\mathbf{y}) = \exp(N\Gamma_{\mathbf{J}, T}^N(\mathbf{y})) \quad (5.67)$$

where  $\Gamma_{\mathbf{J}, T}^N : \mathbb{R}^{MN} \rightarrow \mathbb{R}$  is

$$\Gamma_{\mathbf{J}, T}^N(\mathbf{y}) = N^{-1} \sum_{j \in I_N \# p \in I_M} \int_0^T \sigma_s^{-2} (\tilde{G}_s^{p,j} - \tau^{-1} y_s^{p,j}) dy_s^{p,j} - \frac{1}{2} \sigma_s^{-2} (\tilde{G}_s^{p,j} - \tau^{-1} y_s^{p,j})^2 ds. \quad (5.68)$$

We wish to specify a map  $\Gamma : \mathcal{U} \rightarrow \mathbb{R}$  with (i) as nice regularity properties as possible, and (ii) such that with unit probability

$$\Gamma_{\mathbf{J},T}^N(\mathbf{y}) = \Gamma(\hat{\mu}^N(\mathbf{y}, \mathbf{G})). \quad (5.69)$$

It is well-known that the stochastic integral is not a continuous function of the driving Brownian motion (considered as an object in  $\mathcal{C}([0, T], \mathbb{R})$ , and endowing this space with the topology of uniform convergence). Thus, we define the map  $\Gamma$  to be a limit of time-discretized approximations, and we will show that this limit must always converge for any measure in  $\mathcal{U}$ .

Our precise definition of  $\Gamma : \mathcal{U} \rightarrow \mathbb{R}$  is as follows. We first define a time-discretized approximation to  $\Gamma$ .  $\Gamma^{(m)} : \mathcal{U} \rightarrow \mathbb{R}^+$ ,

$$\begin{aligned} \Gamma^{(m)}(\mu) = \sum_{p \in I_M} \sum_{a=0}^{m-1} \mathbb{E}^\mu \left[ \sigma_{t_a^{(m)}}^{-2} (G_{t_a^{(m)}}^p - \tau^{-1} z_{t_a^{(m)}}^p) (z_{t_{a+1}^{(m)}}^p - z_{t_a^{(m)}}^p + \Delta_m \tau^{-1} z_{t_a^{(m)}}^p) \right. \\ \left. - \frac{1}{2} \sigma_{t_a^{(m)}}^{-2} \Delta_m (G_{t_a^{(m)}}^p - \tau^{-1} z_{t_a^{(m)}}^p)^2 \right]. \quad (5.70) \end{aligned}$$

where  $\Delta_m = T/m$  and  $t_i^{(m)} = iT/m$ . We now define  $\Gamma : \mathcal{U} \rightarrow \mathbb{R}$  to be such that (in the case that the following limit exists)

$$\Gamma(\mu) = \lim_{j \rightarrow \infty} \Gamma^{(m_{j,j})}(\mu), \quad (5.71)$$

where  $m_{j,j}$  is a positive integer defined further below in Lemma 5.6.2. If the above limit does not exist, then we define  $\Gamma(\mu) = 0$  (in fact, we will see that the limit always exists if  $\mu \in \mathcal{U}$ ). It may be observed that  $\Gamma$  is a well-defined measurable function.

**Lemma 5.6.1.** *For every  $N \geq 1$ , every  $(\mathbf{z}_0, \mathbf{g}_0) \in \hat{\kappa}$ , and for  $Q_{\mathbf{z}_0, \mathbf{g}_0}^N$  (the law of the random variables  $(\mathbf{y}, \mathbf{G})$  conditioned on  $\mathbf{z}_0, \mathbf{g}_0$ ) almost every  $(\mathbf{y}, \mathbf{G})$ , the following limit*

exists

$$\lim_{j \rightarrow \infty} \Gamma^{(m_{j,j})}(\hat{\mu}^N(\mathbf{y}, \mathbf{G})) \quad (5.72)$$

With unit probability, the Radon-Nikodym Derivative in (5.67) is such that

$$\left. \frac{dP_{\mathbf{J}, \mathbf{z}_0}^N}{dP_{\mathbf{z}_0}^N} \right|_{\mathcal{F}_T} = \exp(\Gamma(\hat{\mu}^N(\mathbf{y}, \mathbf{G}))) \quad (5.73)$$

Also for any  $\epsilon, L > 0$ , there exists  $k \in \mathbb{Z}^+$  such that for all  $N \geq 1$ ,

$$\sup_{j \geq k} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N \left( \left| \Gamma^{(m_{j,j})}(\hat{\mu}^N(\mathbf{y}, \mathbf{G})) - \Gamma(\hat{\mu}^N(\mathbf{y}, \mathbf{G})) \right| \geq \epsilon \right) \leq -L \quad (5.74)$$

*Proof.* Define the set

$$\mathcal{A}_j = \left\{ \mu \in \mathcal{U} : \left| \Gamma^{(m_{j,j})}(\mu) - \Gamma^{(m_{j+1,j+1})}(\mu) \right| \geq 2^{1-j} \right\} \quad (5.75)$$

Thanks to a union-of-events bound, for any  $N \geq 1$ , and using the bound in Lemma 5.6.2,

$$\sup_{\mathbf{z}_0, \mathbf{g}_0 \in \mathcal{Y}^N} Q_{\mathbf{z}_0, \mathbf{g}_0}^N \left( \hat{\mu}^N \in \bigcup_{j \geq k} \mathcal{A}_j \right) \leq \sum_{j=k}^{\infty} \exp(-N2^j). \quad (5.76)$$

It thus follows from the Borel-Cantelli Lemma [5, 57] that for each  $N \geq 1$ , there must exist a random  $k$  such that  $\hat{\mu}^N \notin \mathcal{A}_j$  for all  $j \geq k$ , and so the limit in (5.72) exists (almost surely). (5.74) follows analogously. As the time-discretization tends to 0, the summation must converge to the stochastic integral, hence (5.73) must be true too.

□

**Lemma 5.6.2.** (i)  $\Gamma^{(m)} : \mathcal{U} \rightarrow \mathbb{R}$  is continuous. (ii) Moreover, for any  $a, j \in \mathbb{Z}^+$ , there exists  $m_{a,j}$  such that for all  $m \geq m_{a,j}$  and all  $n \geq m$ ,

$$\sup_{\mathbf{z}_0, \mathbf{g}_0 \in \mathcal{Y}^N} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N (|\Gamma^{(m)}(\hat{\mu}^N(\mathbf{y}, \mathbf{G})) - \Gamma^{(n)}(\hat{\mu}^N(\mathbf{y}, \mathbf{G}))| \geq 2^{-j}) \leq -2^a. \quad (5.77)$$

*Proof.* (i) The continuity of  $\Gamma^{(m)}$  is immediate from the definition.

(ii) For any  $t \in [0, T]$ , write  $t^{(m)} = \sup\{t_b^{(m)} : t_b^{(m)} \leq t\}$ . Starting with the discrete approximation to the stochastic integral, we can thus write

$$\sum_{b=0}^{m-1} \sigma_{t_a^{(m)}}^{-2} (G_{t_b^{(m)}}^p - \tau^{-1} z_{t_b^{(m)}}^p) (z_{t_{b+1}^{(m)}}^p - z_{t_b^{(m)}}^p) = \int_0^T \sigma_{t^{(m)}}^{-2} (G_{t^{(m)}}^p - \tau^{-1} z_{t^{(m)}}^p) dz_t^p. \quad (5.78)$$

Hence,

$$\begin{aligned} & \sum_{b=0}^{m-1} \mathbb{E}^\mu \left[ \sigma_{t_b^{(m)}}^{-2} (G_{t_b^{(m)}}^p - \tau^{-1} z_{t_b^{(m)}}^p) (z_{t_{b+1}^{(m)}}^p - z_{t_a^{(m)}}^p) \right] - \\ & \quad \sum_{ab=0}^{n-1} \mathbb{E}^\mu \left[ \sigma_{t_b^{(n)}}^{-2} (G_{t_b^{(n)}}^p - \tau^{-1} z_{t_b^{(n)}}^p) (z_{t_{b+1}^{(n)}}^p - z_{t_b^{(n)}}^p) \right] \\ &= \mathbb{E}^\mu \left[ \int_0^T \left\{ \sigma_{t^{(m)}}^{-2} (G_{t^{(m)}}^p - \tau^{-1} z_{t^{(m)}}^p) - \sigma_{t^{(n)}}^{-2} (G_{t^{(n)}}^p - \tau^{-1} z_{t^{(n)}}^p) \right\} dz_t^p \right] \\ &= \mathbb{E}^\mu \left[ \int_0^T \sum_{p \in I_M} (f_{t^{(m)}}^p - f_{t^{(n)}}^p) dz_t^p \right] \quad (5.79) \end{aligned}$$

where  $f_t^p = \sigma_t^{-2} (G_t^p - \tau^{-1} z_t^p)$ . Writing

$$f_t^{p,j} = \sigma_t^{-2} (G_t^{p,j} - \tau^{-1} z_t^{p,j}), \quad (5.80)$$

we obtain that

$$\mathbb{E}^{\hat{\mu}^N} \left[ \int_0^T \sum_{p \in I_M} (f_{t^{(m)}}^p - f_{t^{(n)}}^p) dz_t^p \right] = N^{-1} \sum_{j \in I_N} \sum_{p \in I_M} \int_0^T (f_{t^{(m)}}^{p,j} - f_{t^{(n)}}^{p,j}) dy_t^{p,j}. \quad (5.81)$$

The quadratic variation of this stochastic integral is

$$(QV)_t^{(m,n),N} = N^{-2} \sum_{j \in I_N; p \in I_M} \int_0^t (f_{s^{(m)}}^{p,j} - f_{s^{(n)}}^{p,j})^2 \sigma_s^2 ds \quad (5.82)$$

By definition of the set  $\mathcal{U}_a$ , if  $\hat{\mu}^N \in \mathcal{U}_a$ , then for any  $\delta > 0$ , one can find  $m_\delta$  such that as long as  $m, n \geq m_\delta$ , necessarily

$$(QV)_T^{(m,n),N} \leq N^{-1}\delta.$$

Then writing  $w(\cdot)$  to be a standard Brownian Motion, using the Dambin-Dubins-Schwarz [82] time-rescaled representation of the stochastic integral, as long as  $(m, n) \geq m_\delta$ ,

$$\mathbb{P}\left(\hat{\mu}^N \in \mathcal{U}_a, \left| \int_0^T \sum_{p \in I_M} (f_{t^{(m)}}^p - f_{t^{(n)}}^p) dz_t^p \right| \geq \frac{\epsilon}{2}\right) \leq \mathbb{P}(|w(N^{-1}\delta)| \geq \epsilon) \quad (5.83)$$

$$= \exp(-N\epsilon^2/(8\delta)) \leq \exp(-NL), \quad (5.84)$$

as long as we choose  $\delta$  sufficiently small.

The other terms in

$$\Gamma^{(m)}(\hat{\mu}^N(\mathbf{y}, \mathbf{G})) - \Gamma^{(n)}(\hat{\mu}^N(\mathbf{y}, \mathbf{G}))$$

are treated similarly (observe that they are just Riemann Sums, so it is straightforward to control their difference from the limiting integral).

□

We now prove Theorem 5.2.1.

*Proof.* We start by proving that for any  $\epsilon > 0$ , there must exist a measure  $\mu \in \mathcal{U}$  such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu) \leq \epsilon) = 0. \quad (5.85)$$

Write  $\mathfrak{U} = \mathcal{U}_a$ , where  $a$  is large enough that

$$\overline{\lim}_{N \rightarrow \infty} \sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \log Q_{\mathbf{z}_0, \mathbf{g}_0}^N(\hat{\mu}^N(\mathbf{y}, \tilde{\mathbf{G}}) \in \mathcal{U}_a) < -C.$$

where  $C$  is the upperbound for  $\Gamma$  in Lemma 5.6.3. This is possible thanks to the Exponential Tightness.

**Lemma 5.6.3.** *There exists a constant  $C > 0$  such that*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C) < 0. \quad (5.86)$$

*Proof.* For any  $\ell > 0$ ,

$$\begin{aligned} \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}\left(\Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C\right) &\leq \max \left\{ \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| > \ell), \right. \\ &\quad \left. \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \leq \ell, \Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C) \right\} \end{aligned} \quad (5.87)$$

Thanks to Lemma 5.3.1,  $\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| > \ell)$  converges to  $-\infty$  as  $\ell \rightarrow \infty$ . It thus suffices that we prove that, for arbitrary  $\ell > 0$ , there exists  $C_\ell > 0$  such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \leq \ell, \Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C_\ell) < 0. \quad (5.88)$$

Now, leaving out the negative-semi-definite terms, we find that

$$\Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \leq N^{-1} \sum_{j \in I_N \ni p \in I_M} \int_0^T \sigma_s^{-2} (\tilde{G}_s^{p,j} - \tau^{-1} y_s^{p,j}) dy_s^{p,j} \quad (5.89)$$

Furthermore, writing  $h_s^p = \sigma_s^{-2}(\tilde{G}_s^{p,j} - \tau^{-1}y_s^{p,j})$ , and assuming that  $\|\mathcal{J}_N\| \leq \ell$ , one finds that

$$\sum_{j \in I_N} (h_s^{p,j})^2 \leq 2\sigma_s^{-4} \sum_{j \in I_N} \{(\tilde{G}_s^{p,j})^2 + \tau^{-2}(y_s^{p,j})^2\} \quad (5.90)$$

$$\leq 2\sigma_s^{-4} \sum_{j \in I_N} \{\ell\lambda(y_s^{p,j})^2 + \tau^{-2}(y_s^{p,j})^2\} \quad (5.91)$$

$$\leq 2\sigma_s^{-4} \sum_{j \in I_N} \{\ell C_\lambda^2(y_s^{p,j})^2 + \tau^{-2}(y_s^{p,j})^2\}. \quad (5.92)$$

We thus find that, for any  $L > 0$  there exists a constant  $\bar{C}_L > 0$  such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}\left(\sup_{p \in I_M} \sum_{j \in I_N} (h_s^{p,j})^2 \geq N\bar{C}_L\right) \leq -L. \quad (5.93)$$

Write

$$\mathcal{H}_N = \left\{ \sup_{p \in I_M} \sum_{j \in I_N} (h_s^{p,j})^2 \leq N\bar{C}_L \right\}.$$

We thus find that,

$$\begin{aligned} & \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \leq \ell, \mathcal{H}_N, \Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C_\ell) \\ & \leq \max \left\{ \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\mathcal{H}_N^c), \right. \\ & \quad \left. \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \leq \ell, \mathcal{H}_N, \Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C_\ell) \right\} \\ & \leq \max \left\{ -L, \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \leq \ell, \mathcal{H}_N, \Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C_\ell) \right\} \quad (5.94) \end{aligned}$$

Furthermore, using the Dambins-Dubins Schwarz Theorem [82], and writing  $w(t)$  to be 1D Brownian Motion,

$$\begin{aligned} \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\|\mathcal{J}_N\| \leq \ell, \mathcal{H}_N, \Gamma_{\mathbf{J}, T}^N(\mathbf{z}) \geq C_\ell) \\ \leq \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}\left(\sup_{s \in [0, T]} |w(\bar{C}N^{-1}s)| \geq C_\ell\right) \leq -L, \end{aligned} \quad (5.95)$$

as long as  $C_\ell$  is sufficiently large, using standard properties of Brownian Motion.  $\square$

Thanks to the Radon-Nikodym derivative identity in (5.69), we thus find that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{z}, \tilde{\mathbf{G}}) \notin \mathfrak{U}) < 0. \quad (5.96)$$

Thus, for (5.85) to hold, it suffices that we prove that there exists  $\mu \in \mathfrak{U}$  such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{z}, \mathbf{G}) \in \mathfrak{U}, d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu) \leq \epsilon) = 0. \quad (5.97)$$

Since  $\mathfrak{U}$  is compact, for any  $\epsilon > 0$ , we can obtain a finite covering of  $\mathfrak{U}$  of the form

$$\mathfrak{U} \subseteq \bigcup_{i=1}^{\mathcal{N}_\epsilon} B_\epsilon(\mu_i), \quad (5.98)$$

where  $\mu_i \in \mathfrak{U}$ . By a union of events bound,

$$0 = \lim_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{z}, \tilde{\mathbf{G}}) \in \mathfrak{U}) \quad (5.99)$$

$$\leq \max_{1 \leq i \leq \mathcal{N}_\epsilon} \left\{ \overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{z}, \tilde{\mathbf{G}}) \in B_\epsilon(\mu_i)) \right\} \quad (5.100)$$

If our proposition in (5.97) were false, then (5.100) would be strictly negative, which would be a contradiction.

Write  $\mu_{(k)} \in \mathfrak{U}$  to be such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu_{(k)}) \geq k^{-1}) = 0. \quad (5.101)$$

Let  $\mu \in \mathfrak{U}$  be any measure such that for some subsequence  $(p_k)_{k \geq 1}$ ,  $\lim_{k \rightarrow \infty} \mu_{(p_k)} = \mu$  (this must be possible because  $\mathfrak{U}$  is compact).

We next claim that

$$\lim_{\epsilon \rightarrow 0^+} \underline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu) \leq \epsilon) = -\mathcal{R}(\mu || S_{\mu^{(1)}}) + \Gamma(\mu) \quad (5.102)$$

Indeed writing  $\mathcal{A}_\epsilon = \{d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu) \leq \epsilon\}$ ,

$$\begin{aligned} \mathbb{P}(d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu) \leq \epsilon) &= \mathbb{E}^\gamma \left[ \int_{\mathbb{R}^{MN}} P_{\mathbf{J}, \mathbf{x}}^N(\mathcal{A}_\epsilon) \rho_{\mathbf{J}}^N(\mathbf{x}) d\mathbf{x} \right] \\ &= \mathbb{E}^\gamma \left[ \int_{\mathbb{R}^{MN}} \mathbb{E}^{P_{\mathbf{x}}^N} [\exp(N\Gamma(\hat{\mu}^N)) \chi\{\mathcal{A}_\epsilon\}] \rho_{\mathbf{J}}^N(\mathbf{x}) d\mathbf{x} \right] \\ &= \int_{\mathbb{R}^{MN}} \mathbb{E}^\gamma \left[ \mathbb{E}^{P_{\mathbf{x}}^N} [\exp(N\Gamma(\hat{\mu}^N)) \chi\{\mathcal{A}_\epsilon\}] \rho_{\mathbf{J}}^N(\mathbf{x}) \right] d\mathbf{x} \\ &= \int_{\mathbb{R}^{MN}} \mathbb{E}^\gamma \left[ \mathbb{E}^\gamma \left[ \mathbb{E}^{P_{\mathbf{x}}^N} [\exp(N\Gamma(\hat{\mu}^N)) \chi\{\mathcal{A}_\epsilon\}] \rho_{\mathbf{J}}^N(\mathbf{x}) \mid \mathbf{G}_0 \right] \right] d\mathbf{x} \end{aligned} \quad (5.103)$$

and in this last step, we first perform the conditional expectation, for  $\gamma$  conditioned on the values of  $\{G_0^{p,j}\}_{j \in I_N; p \in I_M}$ .

Now, recall that

$$\rho_{\mathbf{J}}^N(\mathbf{z}_0) = (Z_{\mathbf{J}}^N)^{-1} \chi\{\hat{\mu}^N(\mathbf{z}_0, \mathbf{G}_0) \in B_{\delta_N}(\kappa)\}.$$

Furthermore, writing

$$u_N = N^{-1} \log \mathbb{E}[Z_{\mathbf{J}}^N],$$

our assumption on the initial condition dictates that for any  $\delta > 0$ ,

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(|N^{-1} \log Z_{\mathbf{J}}^N - u_N| \geq \delta) < 0. \quad (5.104)$$

Next, we claim that

$$\lim_{\epsilon \rightarrow 0^+} \inf_{\nu \in \mathfrak{U} \cap \mathcal{A}_\epsilon} \Gamma(\nu) = \Gamma(\mu). \quad (5.105)$$

Indeed (5.105) is a consequence of Lemma 5.6.1: this Lemma implies that  $\Gamma$  can be approximated arbitrarily well by continuous functions over  $\mathfrak{U}$ .

We thus obtain that

$$\begin{aligned} & \lim_{\epsilon \rightarrow 0^+} \underline{\lim}_{N \rightarrow \infty} \inf_{(\mathbf{z}_0, \mathbf{G}_0)} N^{-1} \log (d_W(\hat{\mu}^N(\mathbf{z}, \mathbf{G}), \mu) \leq \epsilon) = \\ & \Gamma(\mu) + \lim_{\epsilon \rightarrow 0^+} \underline{\lim}_{N \rightarrow \infty} \left\{ -u_N + N^{-1} \log \int_{\mathbb{R}^{MN}} \mathbb{E}^\gamma [Q_{\mathbf{x}, \mathbf{G}_0}^N(\mathcal{A}_\epsilon)] \chi \{ \hat{\mu}^N(\mathbf{z}_0, \mathbf{G}_0) \in B_{\delta_N}(\kappa) \} d\mathbf{x} \right\} \\ & = \Gamma(\mu) - \lim_{\epsilon \rightarrow 0^+} \inf_{\nu \in \mathcal{A}(\epsilon)} \mathcal{R}(\mu || S_\nu), \end{aligned} \quad (5.106)$$

since (by definition)

$$N^{-1} \log \int_{\mathbb{R}^{MN}} \mathbb{E}^\gamma [\chi \{ \hat{\mu}^N(\mathbf{z}_0, \mathbf{G}_0) \in B_{\delta_N}(\kappa) \}] d\mathbf{z}_0 = u_N,$$

and we have employed the uniform lower bound in (5.8). The lower semi-continuity of  $\mathcal{R}$  implies that

$$\lim_{\epsilon \rightarrow 0^+} \inf_{\nu \in \mathcal{A}(\epsilon)} \mathcal{R}(\mu || S_\nu) = \mathcal{R}(\mu || S_{\mu^{(1)}}).$$

We thus obtain (5.102), as required.

The theorem now follows from Corollary 5.1.3. Since the relative entropy is only zero when its two arguments are identical [27], any zero must be a fixed point of the operator  $\Phi$ . It is proved in the following Lemma that there is a unique zero.  $\square$

**Lemma 5.6.4.** *There exists a unique fixed point  $\xi$  of  $\Phi$  in  $\mathcal{U}$ . Furthermore  $\xi$  is such that for any  $\mu \in \mathcal{U}$ , writing  $\xi_{(1)} = \mu$  and  $\xi_{(n+1)} = \Phi(\xi_{(n)})$ , it holds that*

$$\xi = \lim_{n \rightarrow \infty} \xi_{(n)} \quad (5.107)$$

*Proof.* We start by considering the following restricted map  $\tilde{\Phi} : \mathfrak{Q} \rightarrow \mathfrak{Q}$  ( $\mathfrak{Q}$  is as defined in Equation 5.3). For some  $\mu \in \mathfrak{Q}$ , write  $\Phi(\mu)$  to be the law of the following random variables  $(\mathbf{z}, \mathbf{G})$ . First, it is stipulated that  $(\mathbf{z}_0, \mathbf{G}_0)$  have probability law  $\kappa$ . Second, conditionally on  $(\mathbf{z}_0, \mathbf{G}_0)$ , the distribution of  $(\mathbf{z}_{[0,T]}, \mathbf{G}_{[0,T]})$  is given by  $S_{\mu, \mathbf{z}_0, \mathbf{g}_0}$ .

Define  $d_t^{(2)} : \mathfrak{Q} \times \mathfrak{Q} \rightarrow \mathbb{R}^+$  to be such (as above in Equation 5.36) that for any  $\mu, \nu \in \mathfrak{Q}$ ,

$$d_t^{(2)}(\mu, \nu) = \inf_{\zeta} \mathbb{E}^{\zeta} \left[ \sum_{p \in I_M} \int_0^t \{(y_s^p - \tilde{y}_s^p)^2 + (G_s^p - \tilde{G}_s^p)^2\} ds \right]^{1/2}, \quad (5.108)$$

where the infimum is over all  $\zeta \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^{2M}) \times \mathcal{C}([0, T], \mathbb{R}^{2M}))$ , such that the law of the first  $2M$  processes is given by  $\mu$ , and the law of the last  $2M$  processes is given by  $\nu$ .

We are going to demonstrate that there is a constant  $c > 0$  such that for all  $\mu, \nu \in \mathfrak{Q}$ ,

$$d_t^{(2)}(\tilde{\Phi}_t(\mu), \tilde{\Phi}_t(\nu)) \leq c\sqrt{t}d_t^{(2)}(\mu, \nu). \quad (5.109)$$

That is, for any  $\mu, \nu \in \mathfrak{Q}$ , we construct a particular  $\zeta$  that is within  $\eta \ll 1$  of realizing the infimum in the definition of the Wasserstein distance in (5.36). To do this, we employ the construction of Lemma 5.3.4. Let  $\mathbf{G}^{\mu}, \mathbf{G}^{\nu}$  be  $\mathcal{C}([0, T], \mathbb{R}^M)$ -valued random variables (in the same probability space), with joint probability law  $\beta_{\mu, \nu}$ . Then for

Brownian motions  $(W_t^p)_{p \in I_M}$ , define

$$dz_t^{\nu,p} = (-\tau^{-1}z_t^{\nu,p} + G_t^{\nu,p})dt + \sigma_t dW_t^p \quad (5.110)$$

$$dz_t^{\mu,p} = (-\tau^{-1}z_t^{\mu,p} + G_t^{\mu,p})dt + \sigma_t dW_t^p. \quad (5.111)$$

The initial conditions are identical:  $z_0^{\nu,p} = z_0^{\mu,p}$ . We immediately see that

$$\frac{d}{dt}(z_t^{\nu,p} - z_t^{\mu,p}) = -\tau^{-1}(z_t^{\nu,p} - z_t^{\mu,p}) + G_t^{\nu,p} - G_t^{\mu,p}, \quad (5.112)$$

and hence

$$\frac{d}{dt}(z_t^{\nu,p} - z_t^{\mu,p})^2 = -2\tau^{-1}(z_t^{\nu,p} - z_t^{\mu,p})^2 + 2(z_t^{\nu,p} - z_t^{\mu,p})(G_t^{\nu,p} - G_t^{\mu,p}) \text{ and thus} \quad (5.113)$$

$$(z_t^{\nu,p} - z_t^{\mu,p})^2 = \int_0^t \exp(2(s-t)/\tau) 2(z_s^{\nu,p} - z_s^{\mu,p})(G_s^{\nu,p} - G_s^{\mu,p}) ds. \quad (5.114)$$

It follows from this that there exists a constant  $c > 0$  such that for all  $t \in [0, T]$ ,

$$d_t^{(2)}(\tilde{\Phi}_t(\mu), \tilde{\Phi}_t(\nu)) \leq ctd_t^{(2)}(\beta_\mu, \beta_\nu) \quad (5.115)$$

$$\leq cC_\lambda td_t^{(2)}(\mu, \nu), \quad (5.116)$$

using Lemma 5.3.4. Thus, for a small enough  $t$ , there is a unique fixed point of  $\tilde{\Phi}_t$  (the mapping up to time  $t$ ). Iterating this argument, we find a unique fixed point for  $\tilde{\Phi}$ . The uniqueness for  $\tilde{\Phi}$ , in turn, implies uniqueness for  $\Phi$ , thanks to the identity in Lemma 5.3.4.

To see why (5.107) holds. First, consider arbitrary  $\nu_{(1)} \in \mathfrak{Q}$ , and define  $\nu_{(n+1)} = \tilde{\Phi}(\nu_{(n)})$ . The above bound in (5.116) implies that necessarily  $(\nu_{(n)})_{n \geq 1}$  is Cauchy. It then immediate follows that for any  $\xi_{(1)} \in \mathcal{U}$  with first marginal equal to  $\nu_{(1)}$ , and writing  $\xi_{(n+1)} = \Phi(\xi_{(n)})$ , it must be that  $(\xi_{(n)})_{n \geq 1}$  is Cauchy.

Finally, we note that  $d^{(2)}$  metrizes weak convergence, thanks to Lemma [5.3.3](#).

□

## CHAPTER 6

### CALCIUM SIGNALING MODEL

One of our fundamental aims is to develop an accurate microscopic model of calcium signaling (which is thoroughly stochastic) and then use statistical mechanical techniques to determine effective macroscopic equations. The complex nature of cellular processes often makes it challenging to characterize them directly, requiring sophisticated modeling techniques to uncover the underlying mechanisms. This work investigates the analysis and modeling of calcium dynamics in the cytoplasm, a critical aspect of cell function. The focus is on a Piecewise-deterministic Markov process (PDMP), a modeling framework known for its deterministic flows between events, switching rates that control transitions between states, and a probability measure that defines these transitions. The deterministic flow of calcium is explained by a set of differential equations that distinguish between the open and closed states of the calcium channel. The components of the flow, such as calcium influx, active pumping, and passive leakage, are elucidated. We utilize a nondimensionalized version of the deterministic flow equation to make our model adaptable to various experimental measurements, especially fluorescence data from Dr Gaetan Barbet's lab. The model in this paper fully couples the calcium concentration to the opening/closing of the channel. It thus allows a more detailed estimation of how the feedback between the opening/closing of channels and the local calcium concentration leads to calcium waves.

#### 6.1 Model Description

A one-dimensional approach is well understood not to be biologically accurate; however, it serves as a useful approximation because calcium diffusion in the cytoplasm is

observed to occur rapidly over short distances [139]. Therefore, local variations are quickly evened out, making it sufficient to understand the average, or mean-field, behavior.

With this model, we aim to address whether we can accurately determine (through direct expression or simulation) the distribution of the time intervals between successive whole-cell calcium spikes by modeling the stochastic nature of individual calcium channels and the deterministic diffusion of calcium between spikes.

Our Piecewise-Deterministic Markov Process (PDMP) model is characterized by deterministic flow between events, a switching rate governing state transitions, and a probability measure defining state transitions.

Our model is a closed-cell model, meaning it does not account for calcium entering or exiting the cell. Calcium transport across the cell's plasma membrane seems to have minimal impact on calcium oscillations [131], primarily driven by release from internal structures like the ER [46, 47]. Furthermore, studying oscillations in closed-cell models allows for examining long-term oscillations since calcium is not lost from the cell.

Positive feedback of calcium concentration is crucial to the model's behavior. The flux through the open channel is proportional to the difference in calcium concentration between the cytoplasm and the ER. The proportionality constant,  $k_f$ , is a scaling factor that controls the maximum total flux and can be considered to represent the Inositol trisphosphate receptors (IPR) density. Its value is sourced from the literature [47]. The assumption that channel flux is linearly dependent on calcium concentration across the entire observed concentration range is dubious. However, this assumption is the simplest and is sufficiently accurate for most models [47].

Our expression for the Sarcoplasmic/endoplasmic reticulum calcium ATPase (SERCA) pump flux term is an example of a unidirectional model found in the literature [47, 99, 100, 119] and represents the consumption of one ATP molecule to

transfer two calcium ions from the cytoplasm to the ER. This expression incorporates a form of positive cooperation, where the binding of the first ion accelerates the binding of the second ion. This expression is a good approximation when the calcium concentration is small relative to the ratio of reaction rates given by  $K^2$ . Parameters for the SERCA pump expression are taken from the literature [47] and are based on Lytton et al. [102], who demonstrated that the pumping rate for various such pumps is well modeled by a Hill function with a coefficient of approximately 2.

In our current model, calcium buffering is not explicitly modeled; however, its effects are incorporated into the diffusion coefficients and scaling factors, assuming fast, linear buffering.

The leak term is necessary to achieve a steady state when the channel is closed when implementing a unidirectional SERCA pump [63, 112].

The dynamics of calcium concentration, denoted as  $c$ , is governed by the following differential equation:

$$\frac{dc}{dt} = f(c, \sigma) \quad (6.1)$$

Here,  $f(c, \sigma)$  describes the rate of change in calcium concentration, which depends on the state of the system  $\sigma \in (o, cl)$ . We consider two possible states of the system: the open state ( $o$ ) and the closed state ( $cl$ ). The expressions for  $f(c, \sigma)$  in these states are as follows:

$$\text{Open state: } f(c, o) = J_{\text{channel}}(c) - J_{\text{pump}}(c) + J_{\text{leak}}(c) \quad (6.2)$$

$$\text{Closed state: } f(c, cl) = -J_{\text{pump}}(c) + J_{\text{leak}}(c) \quad (6.3)$$

The terms  $J_{\text{channel}}(c)$ ,  $J_{\text{pump}}(c)$ , and  $J_{\text{leak}}(c)$  represent the fluxes associated with calcium channels, pumps, and leaks, respectively. These are defined as follows:

$$\begin{aligned}
J_{\text{channel}}(c) &= -k_f c \\
J_{\text{pump}}(c) &= \frac{V_s c^2}{K_s^2 + c^2} \\
J_{\text{leak}}(c) &= k_{\text{leak}}(-c) = -c k_{\text{leak}}
\end{aligned}$$

$J_{\text{channel}}(c)$  represents the calcium flux through the channels, where  $k_f$  is the rate constant for channel flux.  $J_{\text{pump}}(c)$  describes the calcium efflux due to pumps, modeled by a Michaelis-Menten-type kinetics with parameters  $V_s$  (maximum rate) and  $K_s$  (half saturation constant).  $J_{\text{leak}}(c)$  accounts for calcium leak, proportional to the concentration with rate constant  $k_{\text{leak}}$ .

All constants in these expressions are calibrated for the calcium concentration measured in micromolar ( $\mu\text{M}$ ).

To facilitate comparison with experimental data, particularly fluorescence measurements that may use different scales or units, we employed a nondimensionalized version of Equation 6.1. Introducing a nondimensional calcium concentration  $\tilde{c}$ , the transformed equation is now:

$$\frac{d\tilde{c}}{dt} = -k_f \tilde{c} - \left( \frac{V_s}{K_s} \right) \frac{\tilde{c}^2}{1 + \tilde{c}^2} - k_{\text{leak}} \tilde{c} \quad (6.4)$$

This nondimensionalized form ensures that the model can be applied to various experimental setups, allowing for a consistent analysis of calcium dynamics across different measurement techniques. Nondimensionalizing is justified due to the linear relationship between calcium concentration and fluorescence measurement [77].

The system switches between states  $o$  (open) and  $cl$  (closed) with rate  $\lambda$ . Where

$$\lambda_\sigma(c) = \alpha_\sigma c \quad (6.5)$$

That is, the system switches between the channels in the open and closed state at a rate proportional to the amount of calcium in the cytoplasm. The rate constant for going from closed to open is given by  $\alpha_o$  and for open to closed by  $\alpha_{cl}$ .

## 6.2 Maximum Likelihood Estimation

To validate our model with experimental data, the model parameters  $\alpha_o$  and  $\alpha_{cl}$  were fitted using the Maximum Likelihood Estimate (MLE) based on the data provided. The MLE for interspike times in our model aims to maximize the likelihood function  $L$ , defined as:

$$L = \prod_{j=1}^M \rho(\tau_i^{(j)}, \sigma_i^{(j)})_{i=0}^N \quad (6.6)$$

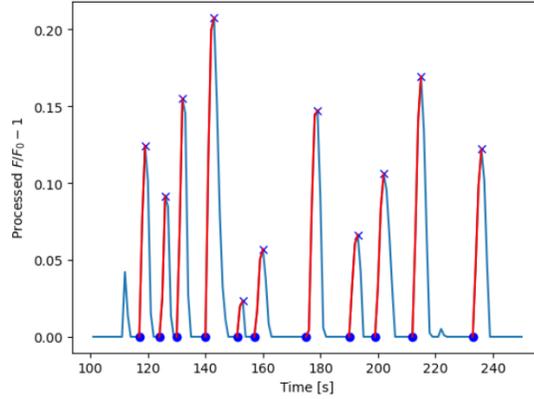
for  $M$  samples, where

$$\rho(\tau_i, \sigma_i)_{i=0}^N = \prod_{i=0}^N \lambda(\sigma, c) \exp\left(-\int_{\tau_i}^{\tau_{i+1}} \lambda(\sigma, s) ds\right) \quad (6.7)$$

for  $N$  events in a given sample.

In this context, *events* represents the opening and closing of the calcium channel. Figure 6.1 illustrates an example of these sample events. The circles represent the time of switching from a closed state to an open state, while the x-s represent the time of switching from an open state to a closed state. The red portion of the signal corresponds to the period when the channel is open, and the blue represents when the channel is closed. The y-axis is given in units of normalized fluorescence. Although acknowledged as an approximation to the true dynamics of the system, it is effective. Crucially, the number of events that occur in a time interval is independent of the number of events that occur in another interval that is disjoint from the first one.

The MLE-derived parameters are tabulated in Table 6.2.



**Figure 6.1** Example of system events.

### 6.3 Model Simulations

Once the parameters were fitted for each sample, a stochastic simulation was performed for the model with those parameters, based on [136]. For each run of the simulation in inter-event times,  $\tau_k$ , were sampled such that

$$\mathbb{P}(\tau_k \geq t) = \exp \left[ - \int_{r=0}^t \lambda \{ \Phi_r(z_{t_{k-1}}) \} dr \right]. \quad (6.8)$$

Where  $\Phi_r(z_{t_{k-1}})$  is the deterministic flow of the calcium, see Section 6.1.

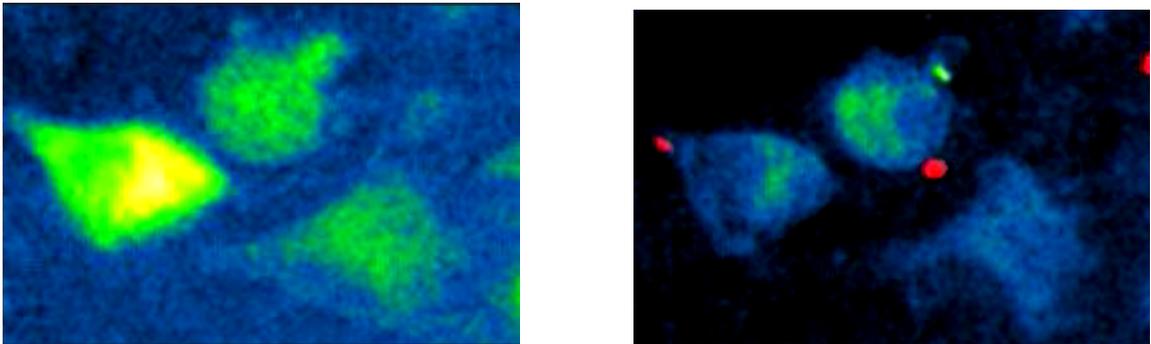
The calcium concentration was then found by solving the deterministic ODEs between events [136].

The results of these simulations are plotted along with the corresponding experimental data. In each case, the model fits well with the experimental data. Statistical metrics such as mean and skewness for experimental and simulated data are provided in Table 6.3.

Markov chain Monte Carlo (MCMC) simulations were performed using the Python package emcee (<https://emcee.readthedocs.io/en/stable/>) [59] in order to quantify the error in parameter fitting for each sample. Corner plots for each run are displayed below the corresponding sample, and the 16th, 50th, and 84th percentile values are tabulated in Table 6.4.

## 6.4 In Vitro Experimental Data Analysis

We collaborated with experimentalist Dr. Gaetan Barbet and the Barbet laboratory of Robert Wood Johnson Medical School, The Child Health Institute of NJ, Rutgers University, to analyze data derived from the average responses of multiple bone marrow-derived dendritic cells (BMDCs). Cells were exposed to two different strains of *Escherichia coli* (DH5 $\alpha$  and MC4100) and *Listeria innocua*. Both *E. coli* strains and *L. innocua* are non-virulent. The subsequent analysis focuses on in vitro calcium concentration fluorescence data in response to both *Live* and *Heat Killed* (HK) pathogens. Figure 6.2 shows examples of fluorescent data captured by the microscope of calcium spiking due to the stimulation of the pathogen in bone-marrow-derived dendritic cells, BMDCs (innate immune cells also referred to as mononuclear phagocytes), the activation of the three cells is due to the bacteria (red dots).



**Figure 6.2** Fluorescent imaging.

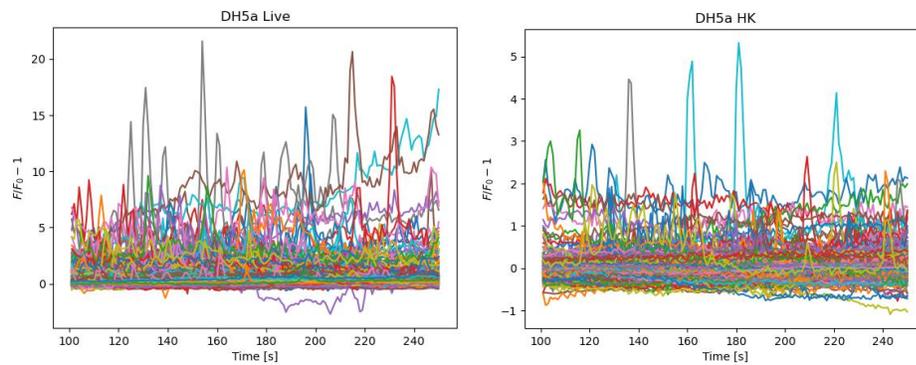
The data sets subjected to analysis are described in Table 6.1.

Figures 6.3, 6.4, and 6.5 present the spiking data for each sample, with the spiking time interval defined from 100 seconds to 250 seconds for each data set. This interval was carefully chosen through visual inspection, ensuring a consistent selection capturing the spiking response of each sample while mitigating the initial non-spiking

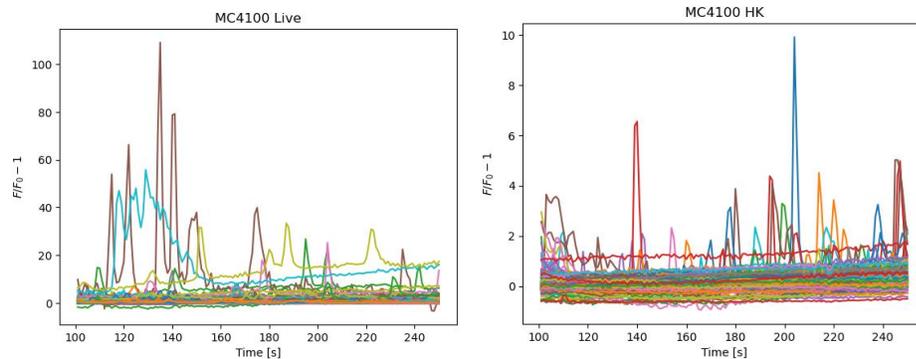
**Table 6.1** Data Statistics for the In Vitro Data Analyzed

| Sample Statistics   |                       |                    |                   |
|---------------------|-----------------------|--------------------|-------------------|
| Sample Name         | Duration of Recording | Sampling Frequency | Number of Samples |
| Ecoli DH5a (HK)     | 301 s                 | 1.0 Hz             | 169               |
| Ecoli DH5a (Live)   | 301 s                 | 1.0 Hz             | 223               |
| Ecoli MC4100 (HK)   | 301 s                 | 1.0 Hz             | 174               |
| Ecoli MC4100 (Live) | 301 s                 | 1.0 Hz             | 202               |
| Listeria (HK)       | 301 s                 | 1.0 Hz             | 212               |
| Listeria (Live)     | 301 s                 | 1.0 Hz             | 226               |

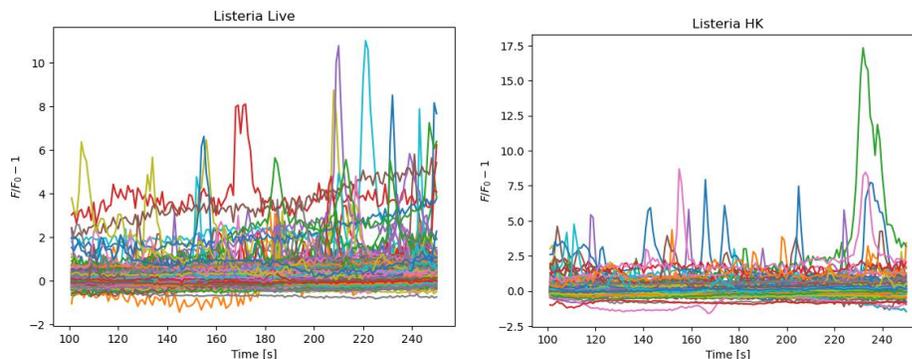
response and tail inaccuracies inherent in the sample measurement process. Observe the markedly stochastic nature of the resulting calcium signaling.



**Figure 6.3** Calcium concentration spiking in response to E.Coli DH5a. Left: Live pathogen. Right: HK pathogen.



**Figure 6.4** Calcium concentration spiking in response to E.Coli MC4100. Left: Live pathogen. Right: HK pathogen.



**Figure 6.5** Calcium concentration spiking in response to Listeria. Left: Live pathogen. Right: HK pathogen.

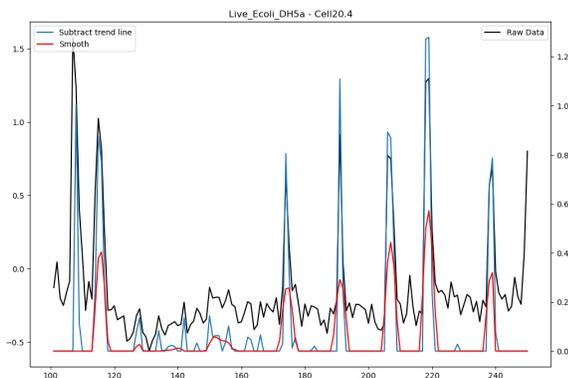
Data preprocessing was performed for each sample to ensure that the resulting stochasticity reflects the inherent variability of the biological process rather than the experimental measurement.

We began by mitigating the trend in the data by subtracting the rolling mean [140], using a window size of 0.1 times the sample length. This window size was chosen based on visual inspection to optimize data processing. Figure 6.6 presents an illustrative example in which the raw data are colored black, and the data with the subtracted trend line is colored blue.

Following this, the data was smoothed using a Hanning window [140] with a length of 9 data points. Similarly to rolling mean subtraction, the choice of this window size was made by visual assessment for optimal data processing. The smoothed data are red in Figure 6.6.

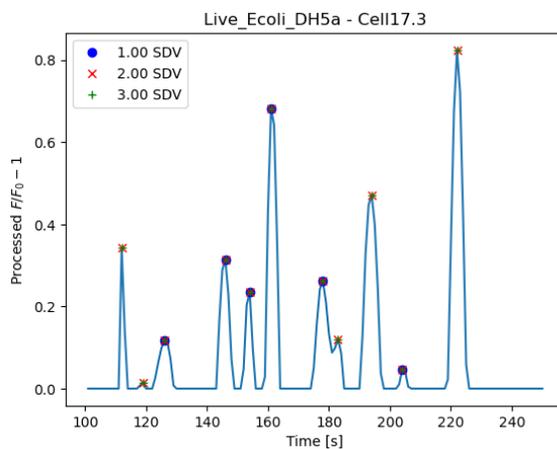
This preprocessing step facilitates accurately identifying calcium spikes and their corresponding timings.

Spike identification was performed using the `find_peaks` function from Python's Scipy library's signal module (<https://docs.scipy.org/doc/scipy/reference/signal.html>). Initially, we determined the prominence and width for all spikes across all samples within a given treatment condition (e.g., Live E.Coli DH5a). Then, the mean and



**Figure 6.6** Example of data preprocessing.

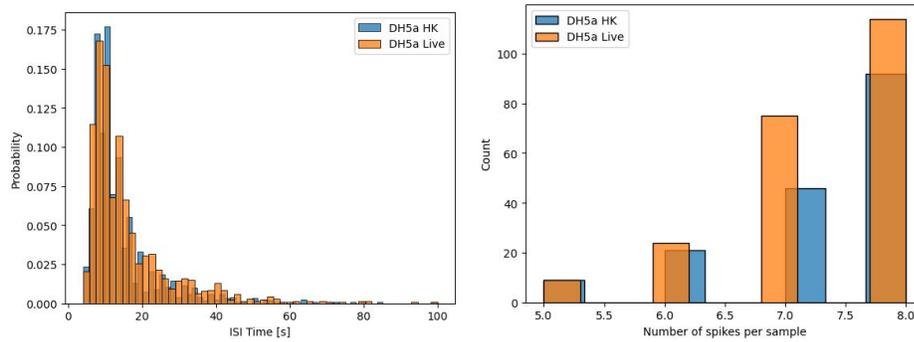
standard deviation of width and prominence were calculated. Subsequently, spikes within  $X$  standard deviations of the mean were selected, where  $X$  was varied for different analyses, as described below. This rigorous selection minimizes the potential influence of aberrant signals within the dataset. See Figure 6.7 for an example using 1, 2, and 3 standard deviations.



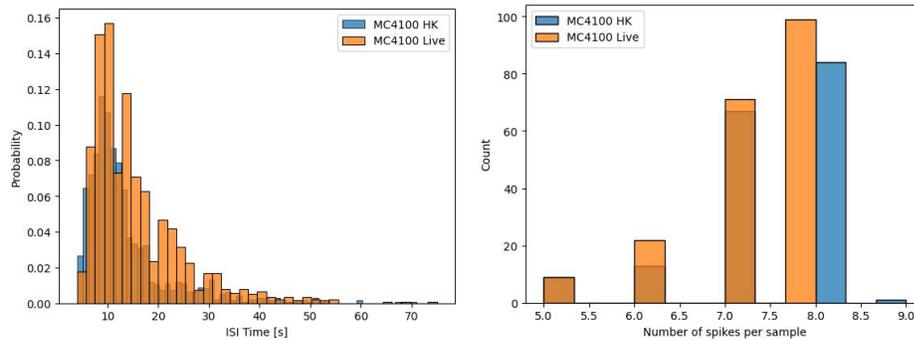
**Figure 6.7** Example of peak selection. Spikes selected within 1, 2, and 3 standard deviations (SDV) from the mean.

To assess the distribution of the number of spikes across the samples, refer to Figures 6.8, 6.9, and 6.10. The selection process ensures consistent spike counts across samples, contributing to the reliability of subsequent analyses.

The Interspike interval (ISI) was calculated for each pair of spikes, i.e., the time between subsequent spikes. The resulting distribution is plotted in Figures 6.8, 6.9, and 6.10 alongside the distribution of the number of spikes across each sample. Spikes that fell within one standard deviation of the mean width and height of the spikes in each sample were selected. A difference between the *Live* and *Heat Killed* appears consistent across the samples.



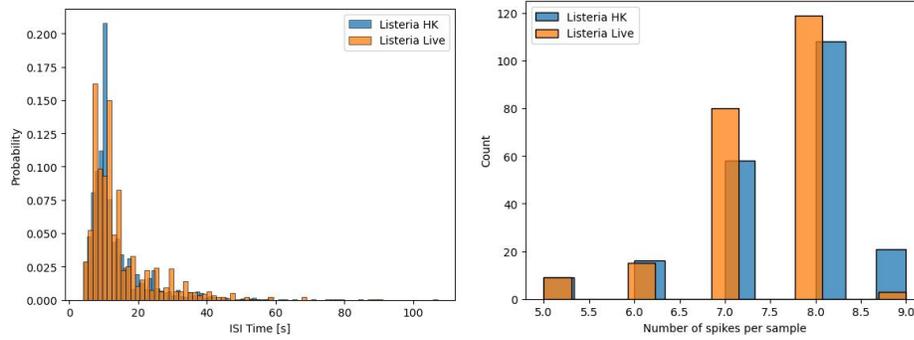
**Figure 6.8** E.Coli DH5a. Left ISI distribution. Right, Spike counts across samples.



**Figure 6.9** E.Coli MC4100. Left ISI distribution. Right, Spike counts across samples.

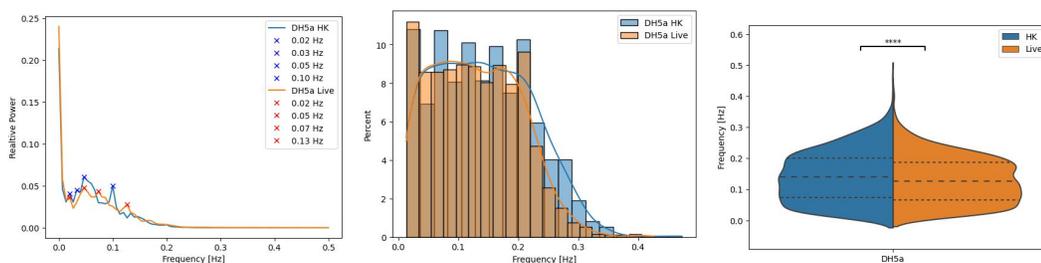
A comprehensive frequency analysis was performed to determine potential differences in signal patterns between data sets; for each sample, the time domain data were transformed into the frequency domain using Python’s rfft function (<https://numpy.org/doc/stable/reference/generated/numpy.fft.rfft.html>).

Subsequently, the average power spectral density and frequencies were computed across multiple samples within each condition. The power spectra convey the distribution of



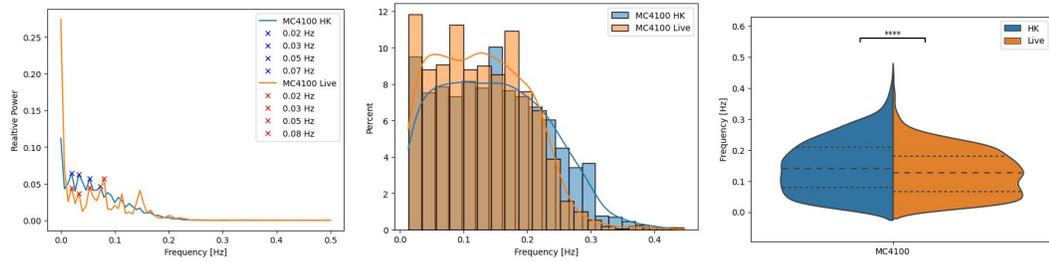
**Figure 6.10** Listeria. Left ISI distribution. Right, Spike counts across samples.

signal power across various frequencies, while the frequency distributions portray the occurrence of specific frequencies within each data set. The resulting distributions and power spectral densities for specific datasets are presented in Figures 6.11, 6.12, and 6.13. In addition, violin plots of the frequency distributions show the statistical significance of the difference between distributions. To get the significance level, a Mann-Whitney U rank test for two independent samples was performed using Python's Scipy Stats library (<https://docs.scipy.org/doc/scipy/reference/stats.html>). There appears to be a consistent difference between the *Live* and *HK* samples, with the *HK* samples comprising a wide range of significant frequencies, which can be associated with greater randomness in the spiking.

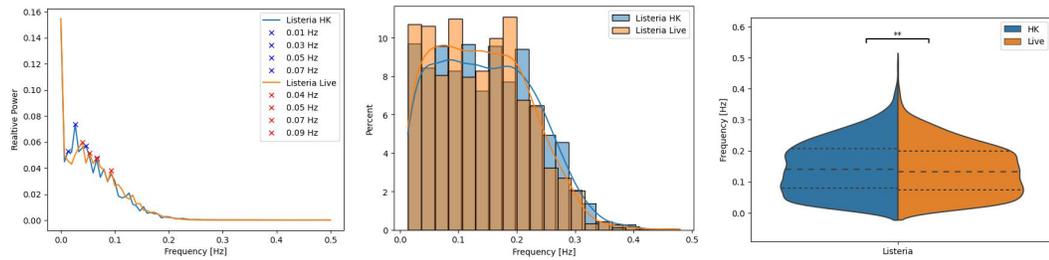


**Figure 6.11** Power spectral density (left), frequency distribution (center) and violin plot (right) for E.Coli DH5a.

A comparison between the frequency analysis for the DH5a versus Listeria *Live* samples indicates a statistically significant, although smaller, difference between the

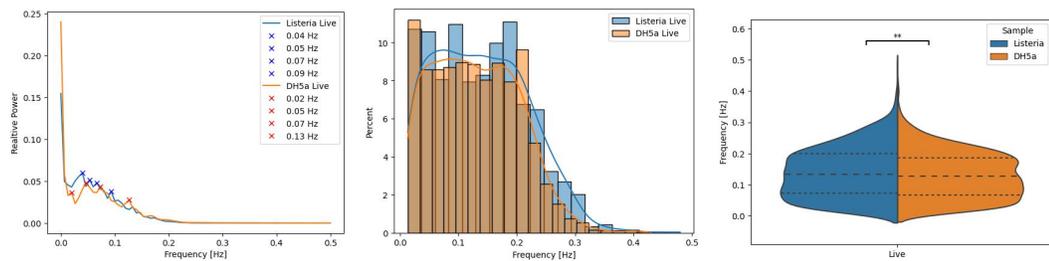


**Figure 6.12** Power spectral density (left), frequency distribution (center) and violin plot (right) for E.Coli MC4100.



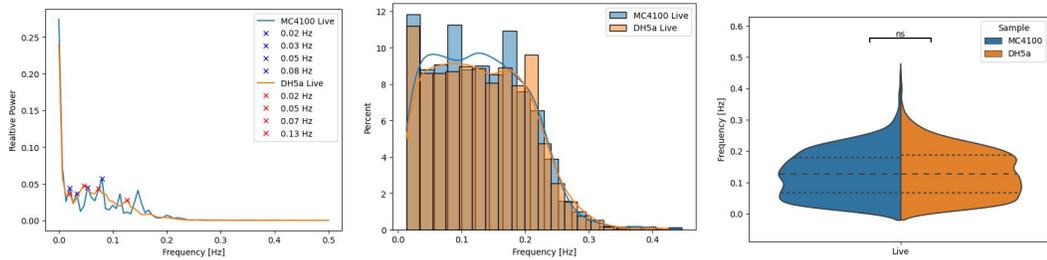
**Figure 6.13** Power spectral density (left), frequency distribution (center), and violin plot (right) for Listeria.

two samples, as can be seen in Figure 6.14



**Figure 6.14** Power spectral density (left), frequency distribution (center), and violin plot (right) for DH5a compared to Listeria.

In contrast, no significant differences are observed between the frequency distribution of the MC4100 and DH5a *Live* samples frequency distribution as seen in Figure 6.15.



**Figure 6.15** Power spectral density (left), frequency distribution (center), and violin plot (right) for MC4100 compared to DH5a.

## 6.5 In Vitro Numerical Results

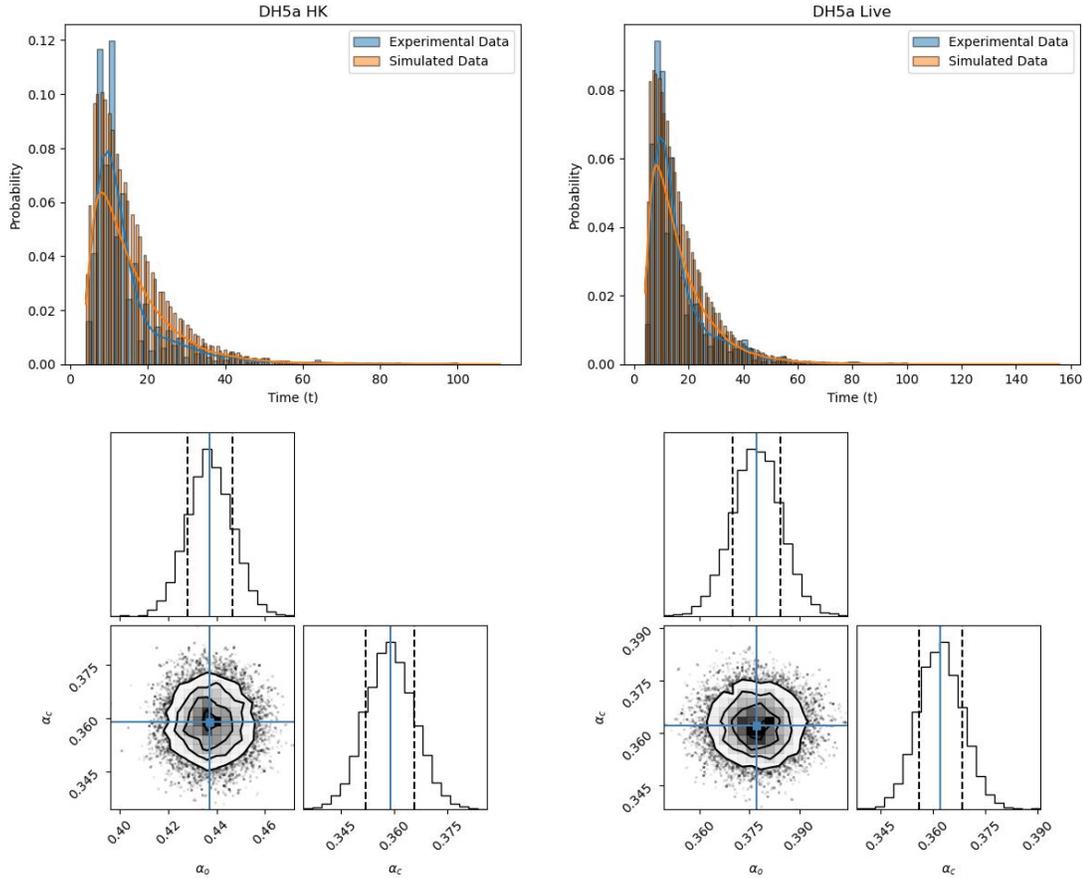
The model’s parameters  $\alpha_o$  (the rate constant for the opening of channels) and  $\alpha_{cl}$  (the rate constant for the closing of channels) were determined numerically by calculating the MLE as described in Subsection 6.2. Table 6.2 summarizes the parameter values obtained.

**Table 6.2** Parameter Values From MLE Results

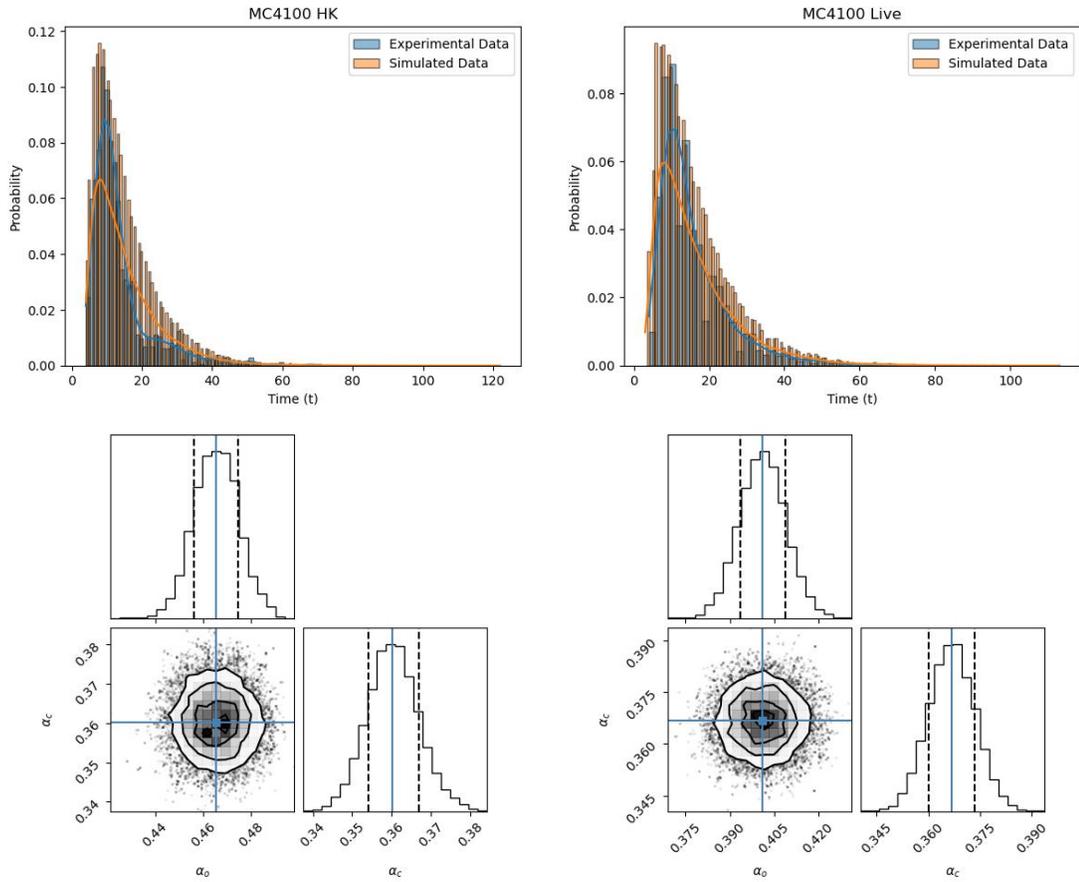
|               | $\alpha_o \left[ \frac{1}{s} \right]$ | $\alpha_{cl} \left[ \frac{1}{s} \right]$ |
|---------------|---------------------------------------|--|
| DH5a HK       | 0.436932                              | 0.358845                                 |
| DH5a Live     | 0.377017                              | 0.362093                                 |
| MC4100 HK     | 0.465579                              | 0.360245                                 |
| MC4100 Live   | 0.400927                              | 0.366744                                 |
| Listeria HK   | 0.478608                              | 0.373902                                 |
| Listeria Live | 0.410899                              | 0.361924                                 |

Stochastic simulations were performed once the parameters were obtained as outlined in Subsection 6.3 above. The results of ISI obtained from these simulations are plotted along with the distribution of the ISI obtained directly from the experimental data in Figures 6.16, 6.17, and 6.18. The goodness of fit indicates that the model captures, at least, the first-order dynamics of the system. Table 6.3 presents a statistical comparison between the experimental data and the simulated data generated by the stochastic model for each sample, key statistical measures such as the number of data points, minimum, maximum, mean, variance, skewness, and kurtosis are provided.

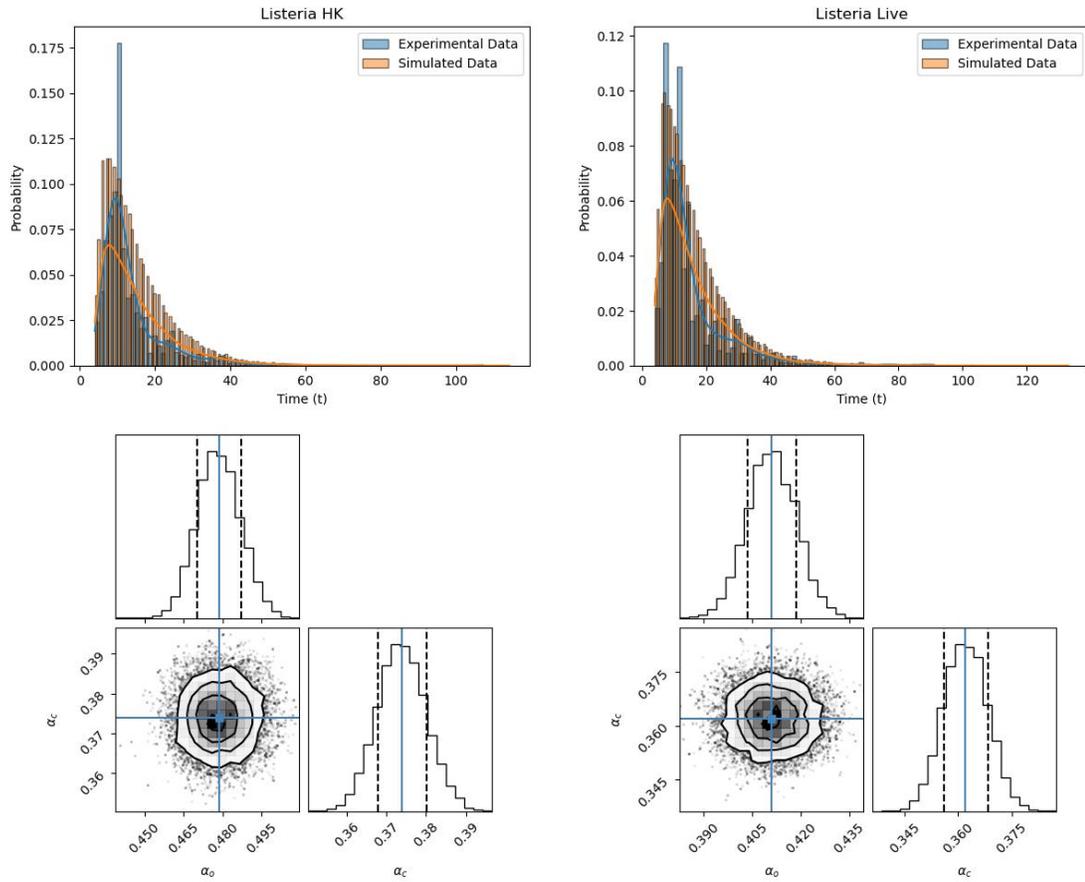
Markov Chain Monte Carlo (MCMC) simulations assessed the uncertainty in the parameter estimates. The 16th, 50th, and 84th percentile values of the parameters  $\alpha_o$  and  $\alpha_{cl}$  are reported in Table 6.4. These percentiles provide a measure of the variability and reliability of the parameter estimates obtained from the simulations. The results from the MCMC simulations indicate that the parameter estimates are reasonably well-constrained, with narrow intervals between the 16th and 84th percentiles (i.e., a standard deviation). This suggests a good confidence level in the estimated parameters derived from the MLE.



**Figure 6.16** Simulation (orange) versus Experimental (blue) ISI distributions for Dh5a. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters  $\alpha_o$  and  $\alpha_{cl}$ .



**Figure 6.17** Simulation (orange) versus Experimental (blue) ISI distributions for MC4100. Left: HK. Right: Live. Below each distribution are the results of the MCMC showing the distribution for the parameters  $\alpha_o$  and  $\alpha_{cl}$ .



**Figure 6.18** Simulation (orange) versus Experimental (blue) ISI distributions for Listeria. Left: HK. Right: Live. Below each distribution are the results of the MCMC showing the distribution for the parameters  $\alpha_o$  and  $\alpha_{cl}$ .

The comparison between experimental and simulated data (Table 6.3) demonstrates that the stochastic model reasonably approximates the experimental observations across different bacterial strains and conditions. Notably, the means and variances of the simulated data are in close agreement with those of the experimental data. However, there are discrepancies in the skewness and kurtosis values, which may indicate areas for further model refinement but may also be inherent to the stochastic nature of the simulation, the amount of experimental data, and the overall stochasticity of the system. The model fits for the *Live* sample appears better than for the *HK* samples, which may be further indicative of less structured dynamics for the *HK* samples.

**Table 6.3** Statistics Comparing the Experimental and Simulated Data

|               |              | Data points | Minimum | Maximum | Mean | Variance | Skewness | Kurtosis |
|---------------|--------------|-------------|---------|---------|------|----------|----------|----------|
| DH5a HK       | Experimental | 1231        | 4       | 100     | 14   | 106      | 3        | 13       |
|               | Simulated    | 50014       | 4       | 111     | 16   | 106      | 2        | 5        |
| DH5a Live     | Experimental | 1399        | 4       | 100     | 16   | 136      | 2        | 8        |
|               | Simulated    | 50012       | 4       | 156     | 17   | 133      | 2        | 6        |
| MC4100 HK     | Experimental | 1347        | 4       | 71      | 14   | 77       | 2        | 8        |
|               | Simulated    | 50009       | 4       | 122     | 15   | 94       | 2        | 5        |
| MC4100 Live   | Experimental | 1369        | 4       | 75      | 16   | 83       | 2        | 5        |
|               | Simulated    | 50006       | 3       | 113     | 16   | 122      | 2        | 6        |
| Listeria HK   | Experimental | 1673        | 4       | 107     | 13   | 72       | 3        | 14       |
|               | Simulated    | 50017       | 4       | 100     | 15   | 93       | 2        | 5        |
| Listeria Live | Experimental | 1552        | 4       | 91      | 15   | 119      | 3        | 10       |
|               | Simulated    | 50007       | 4       | 143     | 16   | 118      | 2        | 6        |

The parameter estimates from the MLE and their respective uncertainties from the MCMC (Tables 6.2 and 6.4) show consistency across different strains and conditions. These results validate the robustness of the stochastic model in capturing the dynamics of calcium concentration in cells, providing a solid foundation for future studies and potential model improvements.

**Table 6.4** 16th, 50th and 84th Percentile Parameter Values From the MCMC

|               | $\alpha_o$ |          |          | $\alpha_{cl}$ |          |          |
|---------------|------------|----------|----------|---------------|----------|----------|
|               | 16th       | 50th     | 84th     | 16th          | 50th     | 84th     |
| DH5a HK       | 0.428141   | 0.437171 | 0.446455 | 0.352003      | 0.358821 | 0.365775 |
| DH5a Live     | 0.369815   | 0.377209 | 0.384409 | 0.355974      | 0.362125 | 0.368510 |
| MC4100 HK     | 0.456408   | 0.465706 | 0.474710 | 0.354025      | 0.360308 | 0.366852 |
| MC4100 Live   | 0.393421   | 0.401310 | 0.408892 | 0.360105      | 0.366815 | 0.373528 |
| Listeria HK   | 0.470267   | 0.478335 | 0.487188 | 0.367825      | 0.373851 | 0.380103 |
| Listeria Live | 0.403597   | 0.411012 | 0.418652 | 0.356017      | 0.362072 | 0.368369 |

Overall, the statistical analysis and parameter estimation underscore the effectiveness of the stochastic modeling approach in replicating experimental calcium concentration data, paving the way for more detailed investigations into the underlying biological processes.

To determine whether there is a significant difference between the *Live* and *HK* samples, we compared the parameter values  $\alpha_o$  and  $\alpha_{cl}$  obtained from the Maximum Likelihood Estimation (MLE) and their respective uncertainties from the Markov Chain Monte Carlo (MCMC) simulations across all samples.

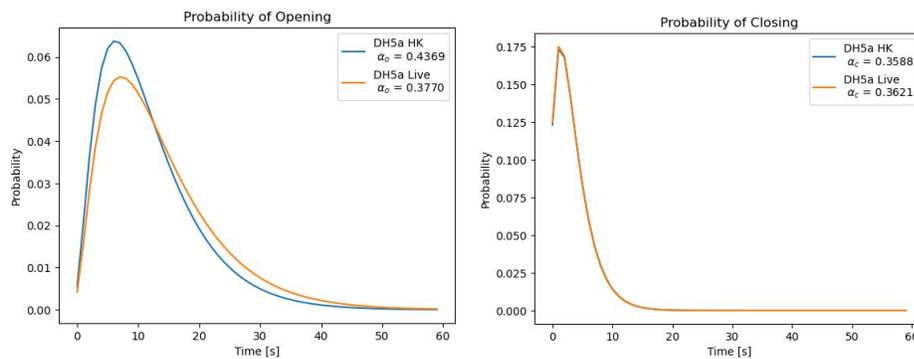
For the parameter  $\alpha_o$ , which represents the rate constant for calcium channels opening, the MLE results show that *Live* samples consistently have lower values compared to *HK* samples across all strains. DH5a: 0.377017 (Live) vs. 0.436932 (HK). MC4100: 0.400927 (Live) vs. 0.465579 (HK). Listeria: 0.410899 (Live) vs. 0.478608 (HK).

The MCMC results further support this observation, indicating non-overlapping or minimally overlapping 16th to 84th percentile ranges for the 'Live' and 'HK' samples, suggesting a statistically significant difference.

For the parameter  $\alpha_{cl}$ , which represents the rate constant for calcium channels closing, the differences between *Live* and *HK* samples are less pronounced. DH5a: 0.362093 (Live) vs. 0.358845 (HK). MC4100: 0.366744 (Live) vs. 0.360245 (HK). *Listeria*: 0.361924 (Live) vs. 0.373902 (HK).

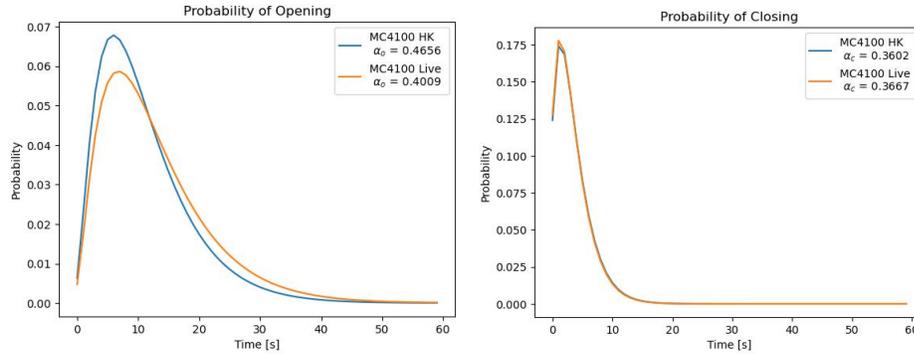
The MCMC results show overlapping a standard deviation for *Live* and *HK* samples, indicating that the differences in  $\alpha_{cl}$  are not statistically significant. This leads us to conclude that there is a significant difference in the parameter  $\alpha_o$  between 'Live' and 'HK' samples, with 'HK' samples consistently showing higher  $\alpha_o$  values. There is no significant difference in the parameter  $\alpha_{cl}$  between 'Live' and 'HK' samples, as the values are comparable and their uncertainty ranges overlap.

To further highlight this difference, Figures 6.19, 6.20, and 6.21 plot the probability of a channel opening or closing over time for the respective parameters. The consistent difference between *Live* and *HK* across all samples is observed.

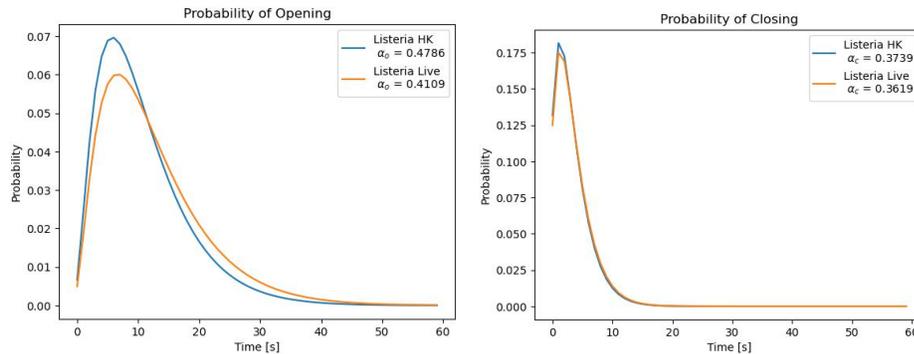


**Figure 6.19** Opening and closing probabilities for optimal parameters for DH5a.

This analysis indicates that cell viability significantly impacts the dynamics of calcium channel opening, specifically  $\alpha_o$ , across different bacterial strains. In contrast, the dynamics of calcium channel closing,  $\alpha_{cl}$ , remain relatively unaffected.



**Figure 6.20** Opening and closing probabilities for optimal parameters for MC4100.



**Figure 6.21** Opening and closing probabilities for optimal parameters for Listeria.

Similar to what we saw in the frequency analysis, the MLE values and the percentile ranges from MCMC suggest that  $\alpha_o$  is higher in Listeria Live samples compared to DH5a Live samples. The 16th to 84th percentile ranges for  $\alpha_o$  do not overlap significantly, indicating a statistically significant difference between the two samples. The MLE values and the percentile ranges from MCMC indicate that  $\alpha_{cl}$  values for DH5a Live and Listeria Live are very close, overlapping the 16th to 84th percentile ranges. This overlap suggests no statistically significant difference in the rate constants for calcium channel closing between the two samples. This leads to the conclusion that there is a statistically significant difference in the parameter  $\alpha_o$  between Live DH5a and Live Listeria samples, with Listeria Live having a higher rate constant for calcium channel opening. There is no statistically significant difference in the parameter  $\alpha_{cl}$  between Live DH5a and Live Listeria samples, as the values are comparable and their uncertainty ranges overlap.

In contrast to the frequency analysis, the MLE values and the percentile ranges from MCMC suggest that  $\alpha_o$  is slightly higher in MC4100 Live samples compared to DH5a Live samples. The 16th to 84th percentile ranges for  $\alpha_o$  do not overlap significantly, indicating a slight statistically significant difference between the two samples. The MLE values and the percentile ranges from MCMC indicate that  $\alpha_{cl}$  values for DH5a Live and MC4100 Live overlap the 16th to 84th percentile ranges. This overlap suggests no statistically significant difference in the rate constants for calcium channel closing between the two samples. There is a statistically significant difference in the parameter  $\alpha_o$  between Live DH5a and Live MC4100 samples, with MC4100 Live having a higher rate constant for calcium channel opening. Although this difference is less significant than that of the Listeria sample, There is no statistically significant difference in the parameter  $\alpha_{cl}$  between Live DH5a and Live MC4100 samples, as the values are comparable and their uncertainty ranges overlap.

## 6.6 In Vivo Experimental Data Analysis

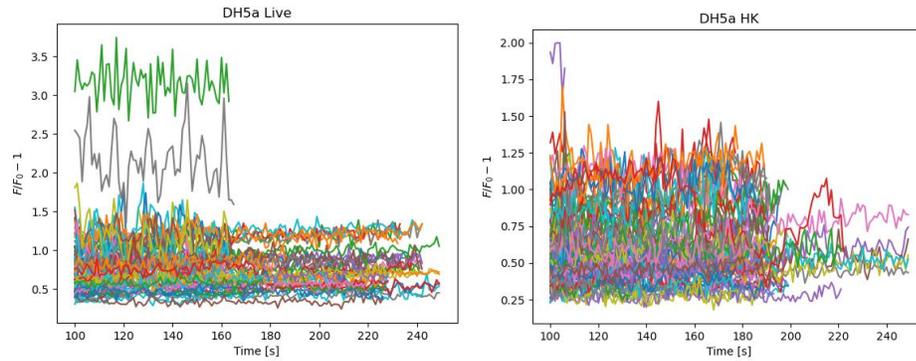
In addition to the above dataset, we were fortunate to receive an *in vivo* dataset from our collaborators. Calcium variation over time was measured in dendritic cells within the spleen of mice after injection of bacteria. Once again, we have *Live* and *HK* samples.

The data sets subjected to analysis are described in Table 6.5:

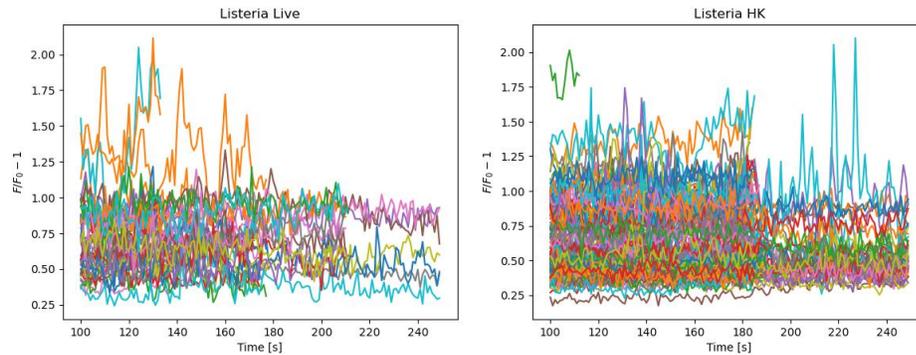
**Table 6.5** Data Statistics for the In Vivo Data Analyzed

| Sample Statistics |                       |                    |                   |
|-------------------|-----------------------|--------------------|-------------------|
| Sample Name       | Duration of Recording | Sampling Frequency | Number of Samples |
| Ecoli DH5a (HK)   | 270 s                 | 1.0 Hz             | 128               |
| Ecoli DH5a (Live) | 418 s                 | 1.0 Hz             | 179               |
| Listeria (HK)     | 690 s                 | 1.0 Hz             | 164               |
| Listeria (Live)   | 276 s                 | 1.0 Hz             | 162               |

Figures 6.22 and 6.23 present the spiking data for each sample, with the spiking time interval defined from 100 seconds to 250 seconds for each data set. This is consistent with what was done for the in vitro data. Observe the markedly stochastic nature of the resulting calcium signaling. We note that a time shift mapping was performed on these datasets to align them with the in vitro data.



**Figure 6.22** Calcium concentration spiking in response to E.Coli DH5a. Left: Live pathogen. Right: HK pathogen.

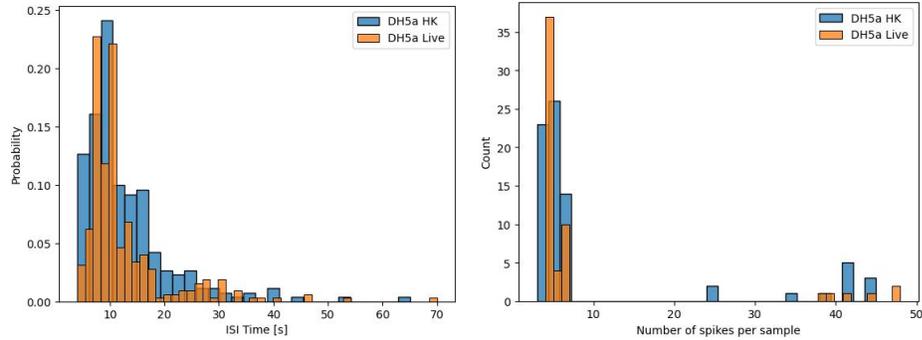


**Figure 6.23** Calcium concentration spiking in response to E.Coli MC4100. Left: Live pathogen. Right: HK pathogen.

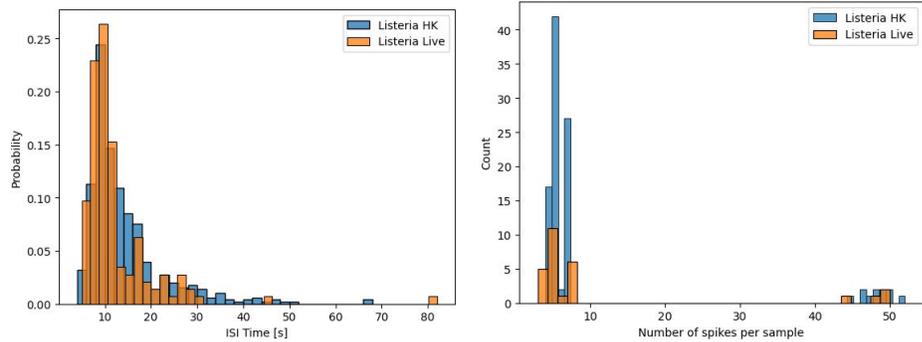
The preprocessing and spike selection were performed identically to that performed for the in vitro data.

The ISI was calculated for each pair of spikes, i.e., the time between subsequent spikes. The resulting distribution is plotted in Figures 6.24 and 6.25 alongside the distribution of the number of spikes across each sample. Spikes that fell within one

standard deviation of the mean width and height of the spikes in each sample were selected. A difference between the *Live* and *Heat Killed* appears consistent across the samples. We note that there is less consistency in the number of spikes across the samples for the in vivo samples when compared to the in vitro samples.

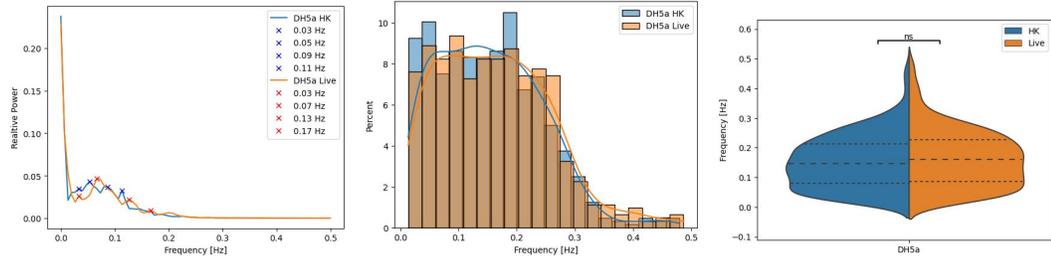


**Figure 6.24** E.Coli DH5a. Left ISI distribution. Right, Spike counts across samples.

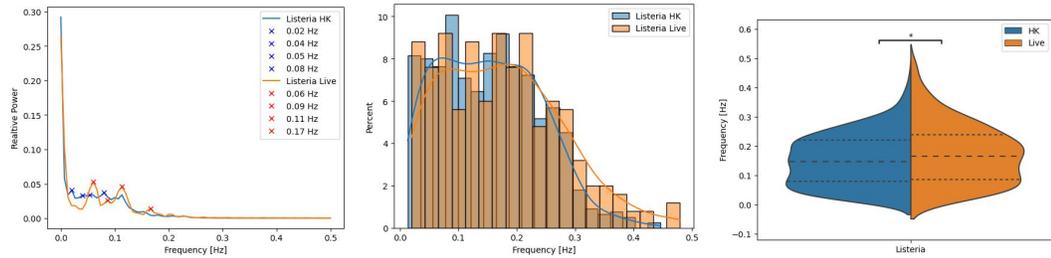


**Figure 6.25** Listeria. Left ISI distribution. Right, Spike counts across samples.

Similarly, as with the in vitro data, a comprehensive frequency analysis was performed to determine potential differences in signal patterns between data sets. The resulting distributions and power spectral densities for specific datasets are presented in Figures 6.26 and 6.27. violin plots of the frequency distributions show the statistical significance of the difference between distributions. There appears to be no statistical difference between the Dh5a *Live* and *HK* samples. A small significant difference is observed between the Listeria *Live* and *HK* samples.

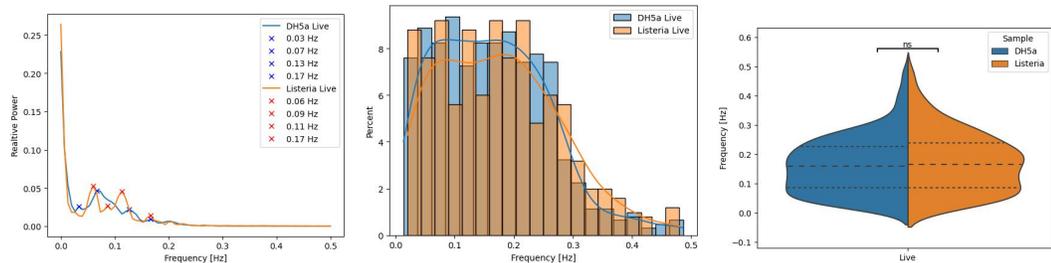


**Figure 6.26** Power spectral density (left), frequency distribution (center) and violin plot (right) for E.Coli DH5a.



**Figure 6.27** Power spectral density (left), frequency distribution (center), and violin plot (right) for Listeria.

A comparison between the frequency analysis for the DH5a versus Listeria *Live* can be seen in Figure 6.28. No statistically significant difference is observed between the samples.



**Figure 6.28** Power spectral density (left), frequency distribution (center), and violin plot (right) for DH5a compared to Listeria.

## 6.7 In Vivo Numerical Results

The model's parameters  $\alpha_o$  (the rate constant for the opening of channels) and  $\alpha_{cl}$  (the rate constant for the closing of channels) were determined numerically by calculating

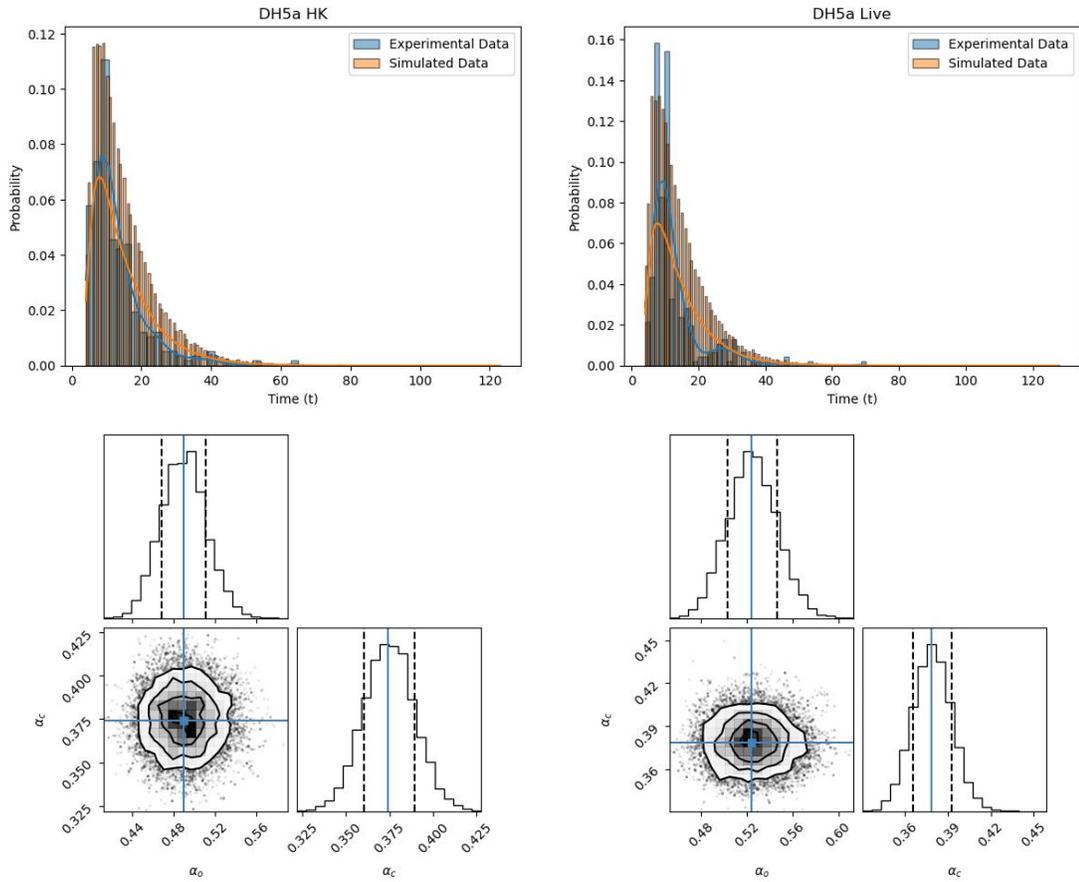
the MLE as described in Subsection 6.2. Table 6.6 summarizes the parameter values obtained.

**Table 6.6** Parameter Values From MLE Results

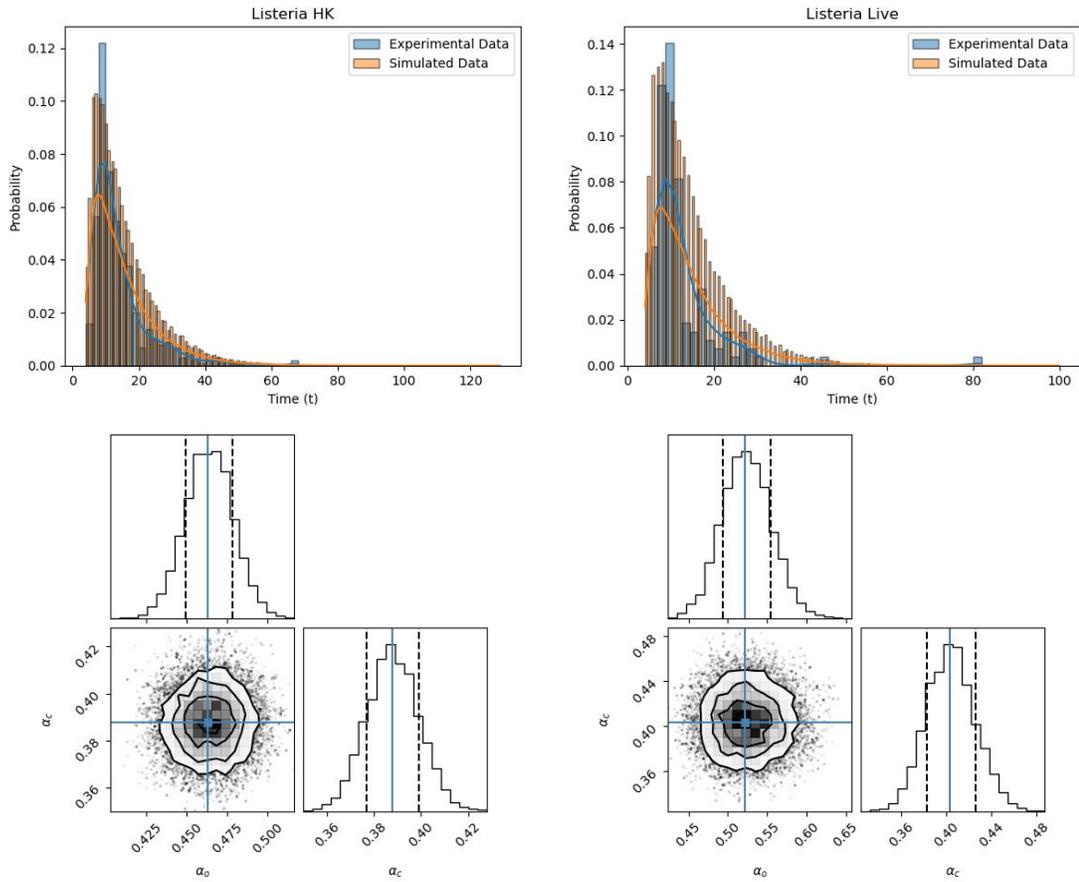
|               | $\alpha_o \left[\frac{1}{s}\right]$ | $\alpha_c \left[\frac{1}{s}\right]$ |
|---------------|-------------------------------------|-------------------------------------|
| DH5a HK       | 0.489308                            | 0.374295                            |
| DH5a Live     | 0.523938                            | 0.378663                            |
| Listeria HK   | 0.462825                            | 0.387598                            |
| Listeria Live | 0.522040                            | 0.403095                            |

Stochastic simulations were performed once the parameters were obtained as outlined in Subsection 6.3 above. The results of ISI obtained from these simulations are plotted along with the distribution of the ISI obtained directly from the experimental data in Figures 6.29 and 6.30. The goodness of fit indicates that the model captures, at least, the first-order dynamics of the system. Table 6.7 presents a statistical comparison between the experimental data and the simulated data generated by the stochastic model for each sample, key statistical measures such as the number of data points, minimum, maximum, mean, variance, skewness, and kurtosis are provided.

Markov Chain Monte Carlo (MCMC) simulations assessed the uncertainty in the parameter estimates. The 16th, 50th, and 84th percentile values of the parameters  $\alpha_o$  and  $\alpha_{cl}$  are reported in Table 6.8. These percentiles provide a measure of the variability and reliability of the parameter estimates obtained from the MCMC simulations. The results from the MCMC simulations indicate that the parameter estimates are reasonably well-constrained, with narrow intervals between the 16th and 84th percentiles (i.e., a standard deviation). This suggests a good confidence level in the estimated parameters derived from the MLE.



**Figure 6.29** Simulation (orange) versus Experimental (blue) ISI distributions for Dh5a. Left: HK. Right Live. Below each distribution are the results of the MCMC showing the distribution for the parameters  $\alpha_o$  and  $\alpha_{cl}$ .



**Figure 6.30** Simulation (orange) versus Experimental (blue) ISI distributions for Listeria. Left: HK. Right: Live. Below each distribution are the results of the MCMC showing the distribution for the parameters  $\alpha_o$  and  $\alpha_{cl}$ .

**Table 6.7** Statistics Comparing the Experimental and Simulated Data

|               |              | Data points | Minimum | Maximum | Mean | Variance | Skewness | Kurtosis |
|---------------|--------------|-------------|---------|---------|------|----------|----------|----------|
| DH5a HK       | Experimental | 261         | 4       | 65      | 13   | 67       | 3        | 9        |
|               | Simulated    | 50019       | 4       | 123     | 15   | 89       | 2        | 5        |
| DH5a Live     | Experimental | 321         | 4       | 70      | 12   | 65       | 3        | 12       |
|               | Simulated    | 50016       | 4       | 128     | 14   | 81       | 2        | 5        |
| Listeria HK   | Experimental | 504         | 4       | 68      | 14   | 75       | 2        | 8        |
|               | Simulated    | 50012       | 4       | 129     | 15   | 101      | 2        | 5        |
| Listeria Live | Experimental | 144         | 5       | 82      | 12   | 73       | 5        | 30       |
|               | Simulated    | 50001       | 4       | 100     | 15   | 86       | 2        | 5        |

**Table 6.8** 16th, 50th and 84th Percentile Parameter Values From the MCMC

|               | $\alpha_o$ |          |          | $\alpha_{cl}$ |          |          |
|---------------|------------|----------|----------|---------------|----------|----------|
|               | 16th       | 50th     | 84th     | 16th          | 50th     | 84th     |
| DH5a HK       | 0.467998   | 0.489768 | 0.511232 | 0.360290      | 0.374866 | 0.389398 |
| DH5a Live     | 0.503293   | 0.524894 | 0.546767 | 0.365501      | 0.378915 | 0.392719 |
| Listeria HK   | 0.449229   | 0.463869 | 0.478328 | 0.376750      | 0.387855 | 0.399115 |
| Listeria Live | 0.493171   | 0.523821 | 0.554447 | 0.382678      | 0.403914 | 0.425749 |

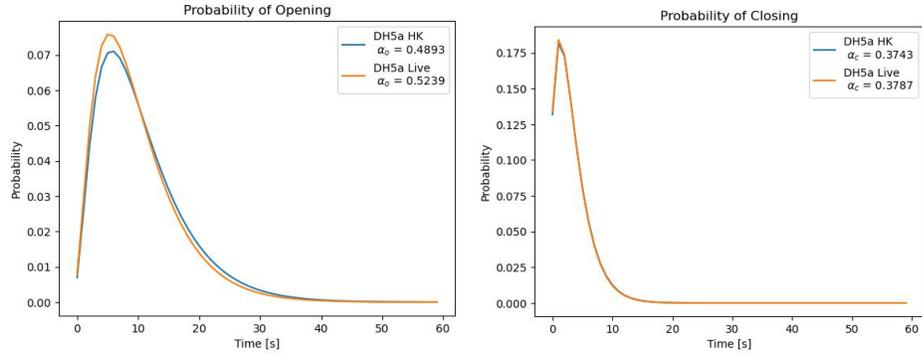
Comparison of the rate constants for the opening ( $\alpha_o$ ) and closing ( $\alpha_{cl}$ ) of the calcium channel between the DH5a Live and DH5a HK samples show no significant difference. For  $\alpha_o$ , DH5a Live has a higher MLE value (0.523938) than DH5a HK (0.489308), but the 16th to 84th percentile ranges overlap slightly, suggesting no statistically significant difference. For  $\alpha_{cl}$ , the MLE values for DH5a Live (0.378663) and DH5a HK (0.374295) are very close, with overlapping percentile ranges, indicating that there is no statistically significant difference in the rate constants for the closing of the calcium channel between the two samples.

Comparison of the rate constants for the opening ( $\alpha_o$ ) and closing ( $\alpha_{cl}$ ) of calcium channels between the Live and HK Listeria and Listeria samples has revealed significant differences. For  $\alpha_o$ , Listeria Live has a higher rate constant (0.522040 MLE) than Listeria HK (0.462825 MLE), with non-overlapping 16th to 84th percentile ranges, indicating a statistically significant difference. Similarly, for  $\alpha_{cl}$ , Listeria Live also shows a higher rate constant (0.403095 MLE) than Listeria HK (0.387598 MLE), again with non-overlapping percentile ranges, suggesting a significant difference in the rate constant for channel closing between the two samples.

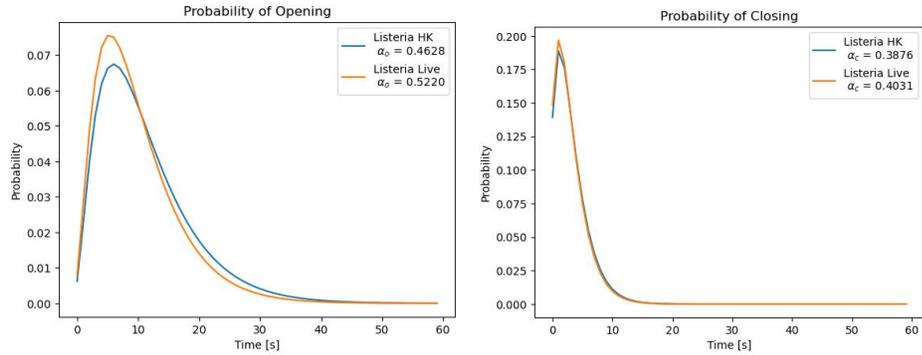
Based on the analysis, the following conclusions can be drawn from the in vivo data. For the DH5a sample, we did not observe a significant increase in the rate constant for the calcium channel opening ( $\alpha_o$ ) in the Live samples compared to the HK samples. In addition, there is no significant difference in the rate constant for the calcium channel closure ( $\alpha_{cl}$ ). For the Listeria sample, there are statistically significant

increases in the rate constants for calcium channel opening ( $\alpha_o$ ) and calcium channel closing ( $\alpha_{cl}$ ) in Live samples compared to HK samples.

To highlight the resulting rate functions further, Figures 6.31 and 6.32 plot the probability of a channel opening or closing over time for the respective parameters.



**Figure 6.31** Opening and closing probabilities for optimal parameters for DH5a.



**Figure 6.32** Opening and closing probabilities for optimal parameters for Listeria.

## CHAPTER 7

### LARGE DEVIATION PRINCIPLES FOR CALCIUM SIGNALING

#### 7.1 Large Deviation of PDMP

This section establishes a LDP for Piecewise-deterministic Markov process (PDMP). We derive specific Euler-Lagrange equations to calculate the most likely trajectory by which the system produces a stochastic calcium puff.

The simplicity of the new proof lies in its original approach. By transforming the PDMP system into a group of homogeneous Poisson Processes using a time-rescaling technique, we directly link to the known LDP for the Poisson Processes through the Inverse Contraction Principle. Furthermore, we identify the specific form of the Euler-Lagrange equations, which is essential for determining the optimal first-hitting times. These equations allow us to formulate the problem as an optimization problem over the period  $[0, T]$  under the constraint imposed by the slow processes.

##### 7.1.1 Notation and Definitions

Let  $\|\cdot\|_t$  denote the supremum norm over  $[0, t]$ , i.e.

$$\|z\|_t = \sup_{s \in [0, t]} |z_s|.$$

Let  $\mathcal{D}([0, t], \mathbb{R})$  be the càdlàg space of real functions from  $[0, t]$  to  $\mathbb{R}$  that are right-continuous and have left-hand limits. Define  $\Lambda_t$  as the class of strictly increasing continuous mappings of  $[0, t]$  onto itself (so that any  $\lambda \in \Lambda_t$  is such that  $\lambda(0) = 0$  and  $\lambda(t) = t$ ). Following Billingsley [17], for any  $\lambda \in \Lambda$ , define

$$\|\lambda\|_t^\circ = \sup_{s < t} \left| \log \frac{\lambda(t) - \lambda(s)}{t - s} \right| \quad (7.1)$$

Define the Skorohod Metric on  $D_t$ ,

$$d_t^\circ(x, y) = \inf_{\lambda \in \Lambda_t} \left\{ \|\lambda\|_t^\circ \wedge \sup_{s \in [0, t]} |x(s) - y(\lambda_j(s))| \right\}. \quad (7.2)$$

It is proved in Billingsley [17] that  $d_t^\circ$  is a metric and that  $D_t$  is complete and separable with respect to the topology induced by  $d_t^\circ$ . Next, for  $a \in \mathbb{Z}^+$  define the pseudometric  $\tilde{d}_a^\circ : \mathcal{D}([0, \infty), \mathbb{R}) \times \mathcal{D}([0, \infty), \mathbb{R}) \rightarrow \mathbb{R}^+$  to be

$$\tilde{d}_a^\circ(x, y) = d_a^\circ(g_a x, g_a y) \text{ where} \quad (7.3)$$

$$g_a(t) = \begin{cases} 1 & \text{if } t \leq a - 1 \\ 0 & \text{if } t \geq a \\ a - t & \text{if } a - 1 \leq t \leq a \end{cases} \quad (7.4)$$

Then define the metric  $d_\infty^\circ : \mathcal{D}([0, \infty), \mathbb{R}) \times \mathcal{D}([0, \infty), \mathbb{R}) \rightarrow \mathbb{R}^+$  to be

$$d_\infty^\circ(x, y) = \sum_{a=1}^{\infty} \inf \{ 2^{-a}, \tilde{d}_a^\circ(x, y) \}. \quad (7.5)$$

It is proved in Billingsley [17] that  $d_\infty^\circ$  metrizes convergence over  $\mathcal{D}([0, \infty), \mathbb{R})$ .

### 7.1.2 Large Deviation Principle for PDMPs

A chemical reaction network [4] consists of the following triple: *Species*  $\mathcal{S}$ , which are the chemical components whose counts we wish to model dynamically. *Complexes*  $\mathcal{C}$  which are nonnegative linear combinations of species that describe how the species can interact, and *Reactions*  $\mathcal{R}$  which describe how to convert one complex to another.

Let  $\xi_\alpha \in \mathbb{R}^d$  be the reaction vector corresponding to the  $\alpha^{\text{th}}$  reaction (it takes values in  $\mathbb{Z}$ ), assuming that there are  $M$  reactions overall. The intensity of the  $n^{\text{th}}$  reaction is defined to be

$$N\lambda_n(x(t), u(t)), \quad (7.6)$$

where  $\lambda_n : \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^+$  is continuous and bounded. It is assumed that reactions cannot produce a negative concentration: that is, it is assumed that if  $x \in \mathbb{R}_+^d$  and  $u \in \mathbb{R}^m$ , and  $\alpha \in \mathcal{R}$  are such that  $x + N^{-1}\xi_\alpha \notin \mathbb{R}_+^d$ , then necessarily  $\lambda_\alpha(x, u) = 0$ .

The ordinary differential equation takes values in a state space  $\mathbb{R}^m$  and is such that

$$\frac{du}{dt} = A(u(t), x(t)), \quad (7.7)$$

where  $A : \mathbb{R}^m \times \mathbb{R}^d \rightarrow \mathbb{R}^m$  is continuous. The initial conditions are taken to be non-random constants, i.e.

$$u(0) := \hat{u}_0^N \text{ and } x(0) := \hat{x}_0^N. \quad (7.8)$$

and it is assumed that the following limits exist

$$\lim_{N \rightarrow \infty} \hat{u}_0^N = \hat{u}_0 \quad (7.9)$$

$$\lim_{N \rightarrow \infty} \hat{x}_0^N = \hat{x}_0. \quad (7.10)$$

Finally, it is assumed that the initial jump rates are always non-zero in the large  $N$  limit, i.e., it is assumed that

$$\inf_{\alpha \in \mathcal{R}} \lambda_\alpha(\hat{x}_0, \hat{u}_0) > 0. \quad (7.11)$$

We assume the following uniform growth bounds: for all  $t \geq 0$ , and  $u \in \mathbb{R}^m$ , and all  $1 \leq n \leq M$ ,

$$\|A(u, X)\| \leq K(1 + \|u\| + \|X\|) \quad (7.12)$$

$$|\lambda_n(X, u)| \leq K. \quad (7.13)$$

Second, we assume that  $A$  and  $\lambda_n$  are both uniformly Lipschitz in their arguments: that is,

$$\left\| A(u, X) - A(\tilde{u}, \tilde{X}) \right\| \leq C \{ \|u - \tilde{u}\| + \|X - \tilde{X}\| \} \quad (7.14)$$

$$\sup_{n \in \mathcal{R}} \|\lambda_n(u, X) - \lambda_n(\tilde{u}, \tilde{X})\| \leq C \{ \|u - \tilde{u}\| + \|X - \tilde{X}\| \}. \quad (7.15)$$

Standard ODE theory dictates that for any particular value of  $x$ , there is a unique solution to the ODE in (7.7).

Let  $Z_\alpha(t)$  count the total number of  $\alpha \in \mathcal{R}$  reactions across the network. The linearity of Poisson Processes ensures that  $Z_\alpha(t)$  can be represented as, writing  $\{Y_\alpha(t)\}_{\alpha \in \mathcal{R}}$  to be independent unit-intensity Poisson Processes [4],

$$Z_\alpha(t) = Y_\alpha \left( N \int_0^t \lambda_\alpha^N(x(s), u(s)) ds \right) \quad (7.16)$$

$$z_\alpha(t) = N^{-1} Z_\alpha(t) \text{ with initial condition} \quad (7.17)$$

$$z_\alpha(0) = 0. \quad (7.18)$$

Writing  $y_\alpha(t) = N^{-1} Y_\alpha^N(Nt)$ , it may be observed that

$$z_\alpha(t) = y_\alpha \left( \int_0^t \lambda_\alpha(x(s), u(s)) ds \right) \quad (7.19)$$

and we have that

$$X(t) = \sum_{\alpha \in \mathcal{R}} Z_\alpha(t) \xi_\alpha \text{ and we define the scaled number of reactions} \quad (7.20)$$

$$x(t) = N^{-1} X(t) = \sum_{\alpha \in \mathcal{R}} z_\alpha(t) \xi_\alpha. \quad (7.21)$$

The main result of this paper is the following theorem. Further below in Equation 7.25, we specify  $\Upsilon$  to be the state space in which the triple  $(z, x, u)$  lives (and we also

define a topology). The sequence of probability laws of  $(z, x, u)$  satisfies the following asymptotic estimate on the space  $\Upsilon$  as  $N \rightarrow \infty$ . The proof of this theorem is given in Appendix B.

**Theorem 7.1.1.** *Suppose that  $\mathcal{O}, \mathcal{A} \subset \Upsilon$ , with  $\mathcal{O}$  open and  $\mathcal{A}$  closed, are such that for any  $T > 0$ ,*

$$\inf_{(z,x,u) \in \mathcal{A} \cup \mathcal{O}} \inf_{\alpha \in \mathbb{R}} \inf_{s \in [0, T]} \lambda_\alpha(x(s), u(s)) > 0. \quad (7.22)$$

*Then there exists a function  $\mathcal{J} : \Upsilon \rightarrow \mathbb{R}$  (specified in Subsection 7.1.3) that is (i) lower-semicontinuous and (ii) has compact level sets such that*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log P((z, x, u) \in \mathcal{A}) \leq - \inf_{\beta \in \mathcal{A}} J(\beta) \quad (7.23)$$

$$\underline{\lim}_{N \rightarrow \infty} N^{-1} \log P((z, x, u) \in \mathcal{O}) \geq - \inf_{\beta \in \mathcal{O}} J(\beta) \quad (7.24)$$

### 7.1.3 Topological Definitions

Before we state our main result, we must briefly note the topological space to which our variables belong. Write the state space for  $(z, x, u)$  as  $\Upsilon$ , i.e.,

$$\Upsilon := \left\{ (z, x, u) \in \mathcal{D}([0, \infty), \mathbb{R}^+)^M \times D([0, \infty), \mathbb{R}^+)^d \times C^1([0, \infty), \mathbb{R})^m \right.$$

where

$$\begin{aligned} \text{(i)} \quad & \sum_{\alpha \in \mathbb{R}} z_\alpha(t) \zeta_\alpha = x(t), \\ \text{(ii)} \quad & \text{for all } t \geq 0, \frac{du}{dt} = A(u(t), x(t)) \text{ and } u(0) = \hat{u} \end{aligned} \quad (7.25)$$

and endow  $\Upsilon$  with the product topology. Since the values of  $x$  and  $u$  are effectively determined by  $z$ , we only need to metrize the convergence of  $z$ . The following lemma captures this.

**Lemma 7.1.2.**  *$\Upsilon$  is a closed subset of  $\mathcal{D}([0, \infty), \mathbb{R}^+)^M \times D([0, \infty), \mathbb{R}^+)^d \times C^1([0, \infty), \mathbb{R})^m$ . In addition,  $\Upsilon$  is separable and metrizable. Furthermore for any  $t > 0$ , there exists*

$c_t > 0$  such that for all  $\delta > 0$ ,

$$\begin{aligned}
& \left\{ (z, x, u), (\tilde{z}, \tilde{x}, \tilde{u}) \in \Upsilon \times \Upsilon : \sup_{\alpha \in \mathbb{R}} d_t^\circ(z_\alpha, \tilde{z}_\alpha) \leq \delta \right\} \\
& \subseteq \left\{ (z, x, u), (\tilde{z}, \tilde{x}, \tilde{u}) \in \Upsilon \times \Upsilon : d_t^\circ(z, \tilde{z}) \leq \delta \right. \\
& \text{and } \sup_{0 \leq s \leq t} |x(s) - \tilde{x}(s)| \leq c_t \delta \\
& \left. \text{and } \sup_{0 \leq s \leq t} |u(s) - \tilde{u}(s)| \leq c_t \delta \right\}
\end{aligned} \tag{7.26}$$

We can now define the rate function  $J(z, x, u) : \Upsilon \rightarrow \mathbb{R}^+$ ,

$$\mathcal{J}(z, x, u) : \Upsilon \rightarrow \mathbb{R}^+, \tag{7.27}$$

starting by stipulating that it is infinite in the case that  $z_\alpha : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is not absolutely continuous for any  $\alpha \in \mathcal{R}$ . Next, we stipulate that  $\mathcal{J}(z, x, u) = \infty$  in the case that  $(z, x, u) \notin \bar{\Upsilon}$ , where  $\bar{\Upsilon}$  is the closure of  $\hat{\Upsilon} \subset \Upsilon$ , that is,

$$\hat{\Upsilon} = \{(z, x, u) \in \Upsilon : \text{There exists } y \in \mathcal{H}_1 \text{ such that Equation 7.19 is satisfied}\}. \tag{7.28}$$

Otherwise, if

- (i)  $(z, x, u) \in \bar{\Upsilon}$  and
- (ii)  $z_\alpha$  is absolutely continuous for every  $\alpha \in \mathbb{R}$  (which means that it must have a derivative  $\dot{z}_\alpha$  for Lebesgue almost every time),

for any realization of  $z$ , define the set

$$\mathfrak{W}(z, x, u) = \{t \in \mathbb{R}^+ : \text{For any } \alpha \in \mathbb{R}, \dot{z}_\alpha(t) \neq 0 \text{ and } \lambda_\alpha(x(t), u(t)) = 0\} \tag{7.29}$$

and define the rate function  $J(z, x, u) = \infty$  if  $\mathfrak{V}(z, x, u)$  is of nonzero Lebesgue Measure, else otherwise

$$\mathcal{J}(z, x, u) = \sum_{\alpha \in \mathbb{R}} \int_{\mathbb{R}^+ \setminus \mathfrak{V}(z, x, u)} \ell \left( \frac{\dot{z}_\alpha(r)}{\lambda_\alpha(x(r), u(r))} \right) \lambda_\alpha(x(r), u(r)) dr \quad (7.30)$$

where

$$\ell(a) = a \log a - a + 1. \quad (7.31)$$

#### 7.1.4 Euler-Lagrange Equations

In this section, we determine the structure of the Euler-Lagrange equations that determine the most likely trajectory followed by the system in attaining a specified state. Computing estimates for first-hitting times is one of the most important applications of Large Deviations Theory. For the application of Large Deviations Theory to first-hitting-time estimates, see, for example, [48, 65, 104]. The ‘fast processes’ (i.e., the ODE in  $u(t)$ ) are implemented as a constraint (with Lagrange multipliers). We start by describing the problem in terms of reaction fluxes; that is, we wish to determine the most likely trajectory followed by the system to reach a target reaction flux  $z^* \in (\mathbb{R}^+)^{|\mathcal{R}|}$  at time  $T$ . We must determine the trajectory,

$$(z, x, u) \in \mathcal{C}^1(\mathbb{R})^M \times \mathcal{C}^1(\mathbb{R})^d \times \mathcal{C}^1(\mathbb{R})^m$$

such that  $\mathcal{J}_T$  is minimized, where

$$\mathcal{J}_T = \sum_{\alpha \in \mathcal{R}} \int_0^T L(\dot{z}(s), x(s), u(s)) ds, \quad (7.32)$$

where  $L : \mathbb{R}^M \times \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^+$  is the integrand of the rate function in (7.30), that is,

$$L(q, x, u) = \sum_{\alpha \in \mathcal{R}} \ell(q_\alpha / \lambda_\alpha(x, u)) \lambda_\alpha(x, u) \quad (7.33)$$

$$\ell(a) = a \log a - a + 1.$$

subject to the following constraints and boundary conditions

$$\begin{aligned} \frac{du}{dt} &= A(u, x) \\ x(t) &= x(0) + \sum_{\alpha \in \mathcal{R}} z_\alpha(t) \xi_\alpha. \\ u(0) &= \hat{u} \\ z_\alpha(0) &= 0 \\ z_\alpha(T) &= z_\alpha^*. \end{aligned} \quad (7.34)$$

We introduce Lagrange Multipliers  $\{\eta^i(t)\}_{1 \leq i \leq m}$  corresponding to the  $m$  constraints

$$\frac{du_t^i}{dt} = A^i(u_t, x_t). \quad (7.35)$$

Our task is now to find the critical points of the functional

$$\tilde{\mathcal{J}}_T(z, u, \eta) = \int_0^T \tilde{L}(\dot{z}(s), x(s), u(s)) ds \text{ where} \quad (7.36)$$

$$\tilde{L}(\dot{z}(s), x(s), u(s)) = L(\dot{z}(s), x(s), u(s)) - \sum_{i=1}^m \eta^i(s) \{ \dot{u}^i(s) - A^i(u(s), x(s)) \} \text{ such that} \quad (7.37)$$

$$x(t) = \hat{x}(0) + \sum_{\alpha \in \mathcal{R}} z_\alpha(t) \xi_\alpha. \quad (7.38)$$

We now take the Frechet Derivative of (7.36), that is, for  $w \in \mathcal{C}^1([0, T], \mathbb{R})^M$  and  $v \in \mathcal{C}^1([0, T], \mathbb{R})^m$ , and  $\kappa \in \mathcal{C}^1([0, T], \mathbb{R})^m$ , with boundary conditions  $w(0) = 0, w(T) = 0, v(0) = 0$ , and the boundary conditions for  $\kappa$  are not yet clear. Thanks to the

constraint in (7.38), we have that

$$\frac{d}{dz_\alpha} = \sum_{i=1}^d \frac{\partial}{\partial x^i} \xi_\alpha^i. \quad (7.39)$$

We, therefore, find that

$$\begin{aligned} D\tilde{\mathcal{J}}_T(z, u, \eta) \cdot (w, v, \kappa) &:= \lim_{\epsilon \rightarrow 0^+} \epsilon^{-1} \{ \tilde{\mathcal{J}}_T(z + \epsilon w, u + \epsilon v, \eta + \epsilon \kappa) - \tilde{\mathcal{J}}_T(z, u, \eta) \} \\ &= \sum_{\alpha \in \mathcal{R}} \int_0^T w_\alpha(t) \left\{ \sum_{i=1}^d \left( \frac{\partial L}{\partial x^i} \xi_\alpha^i + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial x^i} \xi_\alpha^i \right) - \frac{d}{dt} \frac{\partial L}{\partial \dot{z}_\alpha} \right\} dt \\ &+ \sum_{i=1}^m \int_0^T v^i(t) \left\{ \frac{\partial L}{\partial u^i} + \frac{d\eta^i}{dt} + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial u^i} \right\} - \kappa^i \left\{ \frac{du^i}{dt} - A^i(u(t), x(t)) \right\} dt \\ &+ \sum_{\alpha \in \mathcal{R}} \left\{ w_\alpha(T) \frac{\partial}{\partial \dot{z}_\alpha} L(\dot{z}(T), x(T), u(T)) - w_\alpha(0) \frac{\partial}{\partial \dot{z}_\alpha} L(\dot{z}(0), x(0), u(0)) \right\} \\ &\quad - \sum_{i=1}^m \{ \eta^i(T) v^i(T) - \eta^i(0) v^i(0) \}, \end{aligned}$$

and to obtain the above expression, we have used integration by parts to write

$$\begin{aligned} \sum_{\alpha \in \mathcal{R}} \int_0^T \dot{w}_\alpha(t) \frac{\partial L}{\partial \dot{z}_\alpha} dt &= \sum_{\alpha \in \mathcal{R}} \left\{ w_\alpha(T) \frac{\partial}{\partial \dot{z}_\alpha} L(\dot{z}(T), x(T), u(T)) \right. \\ &\quad \left. - w_\alpha(0) \frac{\partial}{\partial \dot{z}_\alpha} L(\dot{z}(0), x(0), u(0)) - \int_0^T w_\alpha(t) \frac{d}{dt} \frac{\partial L}{\partial \dot{z}_\alpha} dt \right\} \quad (7.40) \end{aligned}$$

and

$$\sum_{i=1}^m \int_0^t \eta^i(t) \dot{v}^i(t) dt = \sum_{i=1}^m \{ \eta^i(T) v^i(T) - \eta^i(0) v^i(0) - \int_0^t \dot{\eta}^i(t) v^i(t) dt \} \quad (7.41)$$

Setting the coefficients of  $w_\alpha$ ,  $v$  and  $\kappa^i$  to zero, we obtain the equations

$$\begin{aligned}
\sum_{i=1}^d \left( \frac{\partial L}{\partial x^i} \xi_\alpha^i + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial x^i} \xi_\alpha^i \right) - \frac{d}{dt} \frac{\partial L}{\partial \dot{z}_\alpha} &= 0 \\
\frac{\partial L}{\partial u^i} + \frac{d\eta^i}{dt} + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial u^i} &= 0 \\
\frac{du^i}{dt} - A^i(u(s), x(s)) &= 0 \\
\eta^i(T) &= 0.
\end{aligned} \tag{7.42}$$

Noting that  $\dot{\ell}(a) = \log a$ , we compute that  $\frac{\partial L}{\partial \dot{z}_\alpha} = \log \frac{\dot{z}_\alpha}{\lambda_\alpha(x(t), u(t))}$ , and therefore

$$\begin{aligned}
\frac{d}{dt} \frac{\partial L}{\partial \dot{z}_\alpha} &= \frac{d}{dt} \left( \log \frac{\dot{z}_\alpha}{\lambda_\alpha(x(t), u(t))} \right) \\
&= \frac{\ddot{z}_\alpha}{\dot{z}_\alpha} - \frac{1}{\lambda_\alpha(x(t), u(t))} \frac{d}{dt} \lambda_\alpha(x(t), u(t)).
\end{aligned} \tag{7.43}$$

Thus in the case that  $\lambda_\alpha(x, u) > 0$ , the Euler Lagrange equations are such that for  $1 \leq j \leq m$ ,  $\alpha \in \mathcal{R}$  and  $1 \leq i \leq d$ ,

$$\begin{aligned}
\frac{\ddot{z}_\alpha}{\dot{z}_\alpha} &= \frac{1}{\lambda_\alpha} \frac{d\lambda_\alpha}{dt} + \sum_{i=1}^d \xi_\alpha^i \frac{\partial L}{\partial x^i} - \sum_{k=1}^d \sum_{j=1}^m \eta^j(t) \frac{\partial A^j}{\partial x^k} \xi_\alpha^k \\
\frac{d\eta^j}{dt} &= - \frac{\partial L}{\partial u^j} - \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial u^j} \\
x^i(t) &= x^i(0) + \sum_{\alpha \in \mathcal{R}} z_\alpha(t) \xi_\alpha^i \text{ for all } t \geq 0
\end{aligned} \tag{7.44}$$

$$\eta^j(T) = 0$$

$$z_\alpha(0) = 0$$

$$z(T) = z_*$$

For the sake of completeness, we note that

$$\frac{d\lambda_\alpha}{dt} = \sum_{\beta \in \mathcal{R}} \sum_{i=1}^d \frac{\partial \lambda_\alpha}{\partial x^i} \xi_\beta^i \dot{z}_\beta + \sum_{i=1}^m \frac{\partial \lambda_\alpha}{\partial u^i} A^i(u(t), x(t)) \quad (7.45)$$

$$\frac{\partial L}{\partial u^i} = \sum_{\alpha \in \mathcal{R}} \frac{\partial \lambda_\alpha}{\partial u^i} \left( 1 - \frac{\dot{z}_\alpha}{\lambda_\alpha} \right) \quad (7.46)$$

It can be seen that this system amounts to an ODE boundary value problem. One way one may attempt to solve this problem is via the shooting method [48]. At time 0, the unknowns are  $\{\dot{z}_\alpha(0)\}_{\alpha \in \mathcal{R}}$  and  $\{\eta^j(0)\}_{1 \leq j \leq m}$ . The shooting method requires guessing the values of unknowns at time 0, (ii) integrate the equations forward to time  $T$ , and then (iii) implement the constraints  $\{\dot{z}_\alpha(T) = \dot{z}_{*,\alpha}\}_{\alpha \in \mathcal{R}}$  and  $\{\eta^j(T) = 0\}_{1 \leq j \leq m}$ .

### 7.1.5 Simplified Euler-Lagrange Equations

Usually, one wants to know the first-hitting-time for the concentrations  $\{x^i(T)\}_{1 \leq i \leq d}$ , rather than the first hitting time for the reaction fluxes. That is, for a fixed  $x_* \in \mathbb{R}^d$ , one wishes to estimate the probability that  $x_T \simeq x_*$  and also determine the optimal (most likely) path followed by the system in attaining this point. For this reason, it can be simpler to eliminate the reaction fluxes and reduce the problem to one purely in terms of concentrations and ODE variables  $\{u^i(T)\}_{1 \leq i \leq m}$ . To this end, for any  $\dot{x} \in \mathbb{R}^d$ ,  $x \in (\mathbb{R}^+)^d$ , and  $u \in \mathbb{R}^m$ , define the set

$$\Xi(\dot{x}, x, u) = \left\{ \dot{z} : \dot{x}^p = \sum_{\alpha \in \mathcal{R}} \xi_\alpha^p \dot{z}_\alpha \right\} \subset (\mathbb{R}^+)^M \quad (7.47)$$

Then, define the function

$$\hat{L}(\dot{x}(s), x(s), u(s)) = \inf \left\{ L(\dot{z}, x(s), u(s)) : \dot{z} \in \Xi(\dot{x}, x, u) \right\} \quad (7.48)$$

In the case that  $\Xi(\dot{x}, x, u) = \emptyset$ , define  $\hat{L}(\dot{x}(s), x(s), u(s)) = \infty$ . Thus, one can define the contracted rate function,

$$\begin{aligned} \tilde{\mathcal{J}}_T : \mathcal{C}_{ac}([0, T], (\mathbb{R}^+)^d) \times \mathcal{C}_{ac}([0, T], \mathbb{R}^m) &\rightarrow \mathbb{R} \\ \tilde{\mathcal{J}}_T(x, u) &= \int_0^T \tilde{L}(\dot{x}(s), x(s), u(s)) ds, \end{aligned} \quad (7.49)$$

and it is immediate from the Contraction Principle [45] that  $\tilde{\mathcal{J}}_T$  governs the Large Deviations of  $(x, u)$ . The following convexity property makes it easier to compute optimal trajectories.

**Lemma 7.1.3.**  *$\tilde{L}$  is convex in its first argument.*

*Proof.* This is because  $\Xi$  is linear in  $\dot{x}$ , and  $L$  is convex in its first argument.  $\square$

Next, we outline the Euler-Lagrange equations for  $\tilde{\mathcal{J}}_T$ . The derivation parallels that of the previous section. Our task is to find the critical points of the functional

$$\begin{aligned} \dot{\mathcal{J}}_T(x, u, \eta) &= \int_0^T \dot{L}(\dot{x}(s), x(s), u(s)) ds, \quad \text{where} \\ \dot{L}(\dot{x}(s), x(s), u(s), \eta(s)) &= \tilde{L}(\dot{x}(s), x(s), u(s)) - \sum_{i=1}^m \eta^i(s) \{ \dot{u}^i(s) - A^i(u(s), x(s)) \}, \end{aligned} \quad (7.50)$$

and  $\eta^i \in \mathcal{C}^1([0, T], \mathbb{R})$  is the Lagrange Multiplier.

We now take the Frechet Derivative of (7.50), i.e. for  $w \in \mathcal{C}^1([0, T], \mathbb{R})^M$  and  $v \in \mathcal{C}^1([0, T], \mathbb{R})^m$ , and  $\kappa \in \mathcal{C}^1([0, T], \mathbb{R})^m$ , with boundary conditions  $w(0) = 0, w(T) =$

$0, v(0) = 0$ , and the boundary conditions for  $\kappa$  are not yet clear,

$$\begin{aligned}
D\tilde{\mathcal{J}}_T(x, u, \eta) \cdot (w, v, \kappa) &:= \lim_{\epsilon \rightarrow 0^+} \epsilon^{-1} \{ \tilde{\mathcal{J}}_T(x + \epsilon w, u + \epsilon v, \eta + \epsilon \kappa) - \tilde{\mathcal{J}}_T(x, u, \eta) \} \\
&= \int_0^T \sum_{i=1}^d w_i(t) \left\{ \left( \frac{\partial \hat{L}}{\partial x^i} + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial x^i} \right) - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}^i} \right\} dt \\
&+ \sum_{i=1}^m \int_0^T v^i(t) \left\{ \frac{\partial \hat{L}}{\partial u^i} + \frac{d\eta^i}{dt} + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial u^i} \right\} \\
&- \kappa^i \left\{ \frac{du^i}{dt} - A^i(u(t), x(t)) \right\} dt \\
&- \sum_{i=1}^m \{ \eta^i(T) v^i(T) - \eta^i(0) v^i(0) \}
\end{aligned}$$

The Euler-Lagrange equations are thus, for  $1 \leq i \leq d$  and  $1 \leq j \leq m$ ,

$$\begin{aligned}
\frac{\partial \hat{L}}{\partial x^i} + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial x^i} - \frac{d}{dt} \frac{\partial \hat{L}}{\partial \dot{x}^i} &= 0 \\
\frac{\partial \hat{L}}{\partial u^j} + \frac{d\eta^j}{dt} + \sum_{k=1}^m \eta^k \frac{\partial A^k}{\partial u^j} &= 0 \\
\frac{du^i}{dt} - A^i(u(s), x(s)) &= 0 \\
\eta^i(T) &= 0.
\end{aligned} \tag{7.51}$$

For the sake of completeness, we note that the derivative assumes the form

$$\frac{d}{dt} \frac{\partial \hat{L}}{\partial \dot{x}^i} = \sum_{j=1}^d \left\{ \frac{\partial^2 \tilde{L}}{\partial \dot{x}^i \partial \dot{x}^j} \ddot{x}^j + \frac{\partial^2 \hat{L}}{\partial \dot{x}^i \partial x^j} \dot{x}^j \right\} + \sum_{j=1}^m \frac{\partial^2 \hat{L}}{\partial \dot{x}^i \partial u^j} \frac{du^j}{dt} \tag{7.52}$$

## 7.2 Application to Calcium Signaling

In this section, we apply the theory of Section 7.1 to determine estimates for the typical time it takes for cells with stochastically opening and closing calcium channels to change from most of the channels being closed to most being open [47]. To begin with, we must outline a simple microscopic model of stochastic calcium effects. Following,

for example, [36, 85, 118, 119], we employ a hybrid model: with the calcium diffusion modeled deterministically, the IP3 concentration constant throughout the cell, and the channel opening and closing modeled stochastically; see Chapter 6 for more details. Our primary aim is to determine the probability of calcium puffs and waves by studying a stochastic model similar to [36, 85, 118, 119]. The calcium and IP3 concentration throughout the cluster is assumed to be homogeneous to calculate the probability of a calcium puff in a cluster. However, there is a feedback effect on the opening/closing of the channels: when these bind or unbind the calcium / IP3, they alter the local concentration. Thus, the opening and closing of channels within a cluster are not independent because the channels communicate via the calcium / IP3 concentrations. We assume the system requires a critical number of channels to open together to cause a spike in the cell's overall calcium. In that case, we aim to find the path the system will likely take through the probability space for such an event.

### 7.2.1 PDMP Model of Calcium Dynamics

The following model is based on the one derived in Chapter 6 fitted for the subsequent formulation. We assume that there are  $N$  channels distributed on the cell membrane. The spatial effects are assumed to be negligible, so calcium concentrations can be modeled as approximately spatially homogeneous using ODEs. Let  $u_1$  be the calcium concentration in the cellular cytosol and  $u_2$  be the (explicate) calcium concentration in the endoplasmic reticulum.

We take the number of stochastic variables corresponding to each channel to be 1. That is,  $Z_t^i = 1$  if the channel is open and  $Z_t^i = 0$  if the channel is closed. Employing a standard model [47] for the calcium dynamics, we obtain (as before) that

$$\begin{aligned} \frac{du_1}{dt} &= k_f x(t)(u_2 - u_1) - J_{serca}(u) + J_{leak} := A^1(u, x) \\ \frac{du_2}{dt} &= \gamma J_{serca}(u_1) - \gamma k_f x(t)(u_2 - u_1) - \gamma J_{leak} := A^2(u, x), \end{aligned} \tag{7.53}$$

Where

$$J_{serca}(u_1) = \frac{V_s(u_1)^2}{K_s^2 + (u_1)^2}$$

$$J_{leak}(u) = k_{leak}(u_2 - u_1)$$

One may observe that

$$\frac{d}{dt}(\gamma u_1 + u_2) = 0, \quad (7.54)$$

which means that we can eliminate  $u_2$  from the dynamics, and write  $u := u_1(t)$ , and  $A(t) := A^1(u, x)$ .

Now

$$x(t) = \frac{1}{N} \sum_{j=1}^N Z_t^j. \quad (7.55)$$

The Markovian switching of the channel that goes from 1 to 0 with intensity  $\lambda_1(u, x) = \alpha_{-1}x$  (a constant) and stoichiometric constant is  $\zeta_1 = -1$ , and the switching from 0 to 1 with intensity  $\lambda_2(u, x) = \alpha_1 u_1(1 - x)$  with stoichiometric constant is  $\zeta_2 = 1$ . Let  $z_1(t)$  count the number of times that reaction 1 occurs and  $z_2(t)$  count the number of times that reaction 2 occurs. Then it must be that

$$x(t) = x(0) + z_2(t) - z_1(t). \quad (7.56)$$

### 7.2.2 Large N Limiting Dynamics

As  $N \rightarrow \infty$ , it is a classical result that the concentration of open channels converges to the following ODE [93].

$$\frac{dz}{dt} = \lambda_2(u, x) - \lambda_1(u, x). \quad (7.57)$$

We compute the fixed point analytically. That is, solve (7.57) and (7.53). Take this fixed point as the starting point for your Euler-Lagrange equations.

### 7.2.3 Euler-Lagrange Equations

Working with the contracted rate function in (7.48) seems easier. We first find a much simpler form for the contracted rate function.

**Lemma 7.2.1.**

$$\hat{L}(\dot{x}, x, u) = \ell\left(\frac{\dot{z}_1}{\lambda_1(x, u)}\right)\lambda_1(x, u) + \ell\left(\frac{\dot{x} + \dot{z}_1}{\lambda_2(x, u)}\right)\lambda_2(x, u) \quad (7.58)$$

where  $\dot{z}_1$  is such that

$$\dot{z}_1(\dot{x} + \dot{z}_1) = \lambda_1(x, u)\lambda_2(x, u). \quad (7.59)$$

One must choose the root of the above quadratic that is such that  $\dot{z}_1 \geq 0$ .

*Proof.* It must be,

$$\dot{x} = \dot{z}_2 - \dot{z}_1.$$

Thus, upon substitution, the Lagrangian assumes the form in (7.58). We must choose  $\dot{z}_1$  to minimize (7.58), subject to the constraint  $0 \leq \dot{z}_1 \leq \dot{x}$ . We obtain the following by differentiating (7.58) with respect to  $\dot{z}_1$  and setting the derivative to zero.

$$\log\left(\frac{\dot{z}_1}{\lambda_1(x, u)}\right) + \log\left(\frac{\dot{z}_1 + \dot{x}}{\lambda_2(x, u)}\right) = 0. \quad (7.60)$$

Exponentiating both sides, we obtain (7.59). □

Adapted to the calcium signaling model of this section, the original Euler-Lagrange equations in (7.51) are thus, for  $i \in \{1, 2\}$ ,

$$\begin{aligned}
\frac{\partial \hat{L}}{\partial x} + \sum_{k=1}^2 \eta^k \frac{\partial A^k}{\partial x} &= \frac{d}{dt} \frac{\partial \hat{L}}{\partial \dot{x}} \\
\frac{\partial \hat{L}}{\partial u} + \frac{d\eta^i}{dt} + \sum_{k=1}^2 \eta^k \frac{\partial A^k}{\partial u} &= 0 \\
\frac{du}{dt} - A(u(t), x(t)) &= 0 \\
\eta^i(T) &= 0.
\end{aligned} \tag{7.61}$$

The following lemmas are needed to continue the calculation. The first-order derivatives are,

**Lemma 7.2.2.** *First-order derivatives*

$$\begin{aligned}
\frac{\partial \hat{L}}{\partial \dot{x}} &= \log \left( \frac{\dot{z}_1}{\lambda_1(x, u)} \right) \frac{\partial \dot{z}_1}{\partial \dot{x}} + \log \left( \frac{\dot{x} + \dot{z}_1}{\lambda_2(x, u)} \right) \left( 1 + \frac{\partial \dot{z}_1}{\partial \dot{x}} \right) \\
&\text{where} \\
\frac{\partial \dot{z}_1}{\partial \dot{x}} &= -\frac{\dot{z}_1}{2\dot{z}_1 + \dot{x}}.
\end{aligned} \tag{7.62}$$

Notice also that the derivative assumes the form

$$\frac{d}{dt} \frac{\partial \hat{L}}{\partial \dot{x}} = \frac{\partial^2 \hat{L}}{\partial \dot{x}^2} \ddot{x} + \frac{\partial^2 \hat{L}}{\partial \dot{x} \partial x} \dot{x} + \sum_{j=1}^2 \frac{\partial^2 \hat{L}}{\partial \dot{x} \partial u_j} \frac{du_j}{dt} \tag{7.63}$$

and the second-order derivatives are,

**Lemma 7.2.3.** *Second-order derivatives*

$$\begin{aligned}
\frac{\partial^2 \dot{z}_1}{\partial \dot{x}^2} &= \frac{\dot{z}_1}{(2\dot{z}_1 + \dot{x})^2} \left( 1 + \frac{\dot{x}}{2\dot{z}_1 + \dot{x}} \right) \\
\frac{\partial^2 \hat{L}}{\partial \dot{x}^2} &= 2 \frac{\partial^2 \dot{z}_1}{\partial \dot{x}^2} \log \left( \frac{\dot{z}_1(\dot{x} + \dot{z}_1)}{\lambda_1(x, u)\lambda_2(x, u)} \right) + \frac{\lambda_1(x, u)}{\dot{x} + \dot{z}_1} \left( 1 + \frac{\partial \dot{z}_1}{\partial \dot{x}} \right) \\
\frac{\partial^2 \hat{L}}{\partial \dot{x} \partial x} &= -\frac{\partial \lambda_1(x, u)}{\partial x} \frac{1}{\lambda_1(x, u)} \frac{\partial \dot{z}_1}{\partial \dot{x}} - \frac{\partial \lambda_2(x, u)}{\partial x} \frac{1}{\lambda_2(x, u)} \left( 1 + \frac{\partial \dot{z}_1}{\partial \dot{x}} \right) \\
\frac{\partial^2 \hat{L}}{\partial \dot{x} \partial u_i} &= -\frac{\partial \lambda_1(x, u)}{\partial u_i} \frac{1}{\lambda_1(x, u)} \frac{\partial \dot{z}_1}{\partial \dot{x}} - \frac{\partial \lambda_2(x, u)}{\partial u_i} \frac{1}{\lambda_2(x, u)} \left( 1 + \frac{\partial \dot{z}_1}{\partial \dot{x}} \right)
\end{aligned} \tag{7.64}$$

Equations (7.61) now lead to second-order systems of ODEs,

$$\begin{aligned}
\frac{du_1}{dt} &= k_f x (u_2 - u_1) - \frac{V_s u_1^2}{k_s^2 + u_1^2} + k_{leak} (u_2 - u_1) \\
\frac{du_2}{dt} &= -\gamma \frac{du_1}{dt} \\
\frac{d\eta_1}{dt} &= \frac{\dot{x} + \dot{z}_1}{u_1} + (\eta_1 - \gamma \eta_2) \left( k_f x + \frac{2K_s V_s u_1}{(K_s^2 + u_1^2)^2} + k_{leak} \right) - \alpha_1 (1 - x) \\
\frac{d\eta_2}{dt} &= (\gamma \eta_2 - \eta_1) (k_f x + k_{leak}) \\
\frac{d\dot{x}}{dt} &= (2\dot{z}_1 + \dot{x}) \times \\
&\quad \left( \alpha_{-1} - \frac{\dot{z}_1}{x} - \alpha_1 u_1 + \frac{(\dot{x} + \dot{z}_1)}{(1 - x)} + \eta_1 k_f (u_2 - u_1) \right. \\
&\quad \left. - \eta_2 \gamma k_f (u_2 - u_1) - \left( \frac{1}{(1 - x)} \right) \dot{x} + \left( \frac{1}{u_1} \right) \frac{du_1}{dt} \right)
\end{aligned} \tag{7.65}$$

#### 7.2.4 Optimal Trajectories

We wish to predict the probability for the proportion of open calcium channels being significantly greater than its equilibrium value at some time  $T$ . We write the 'target' proportion of the channels as  $\hat{x} \in (x^*, 1]$ . We thus wish to compute the trajectory  $(x, u) \in A \subset C^2([0, T], \mathbb{R}) \times C^1([0, T], \mathbb{R}^2)$  such that

$$\tilde{J}_T(x, u) = \inf \left\{ \tilde{J}_T(y, v) : (y, v) \in A \right\} \tag{7.66}$$

and  $A \subset C^2([0, T], \mathbb{R}) \times C^1([0, T], \mathbb{R}^2)$  consists of all  $(x, u)$  such that

$$\begin{aligned} x(0) &= x^*(0), \\ x(T) &= \hat{x}, \\ u(0) &= u^*, \end{aligned} \tag{7.67}$$

and for all  $t \in [0, T]$

$$\frac{du}{dt} = A(u(t), x(t)). \tag{7.68}$$

The trajectory that minimizes  $\tilde{J}_T$  indicates the most likely means that a significant proportion of calcium channels open over the time interval  $[0, T]$  due to stochastic effects. The second order ODE for  $x(t)$  is stated previously in Subsection 7.2.3. The dynamics of the Lagrange multiplier  $\eta(t)$  is given by, and the boundary value at time  $T$  is

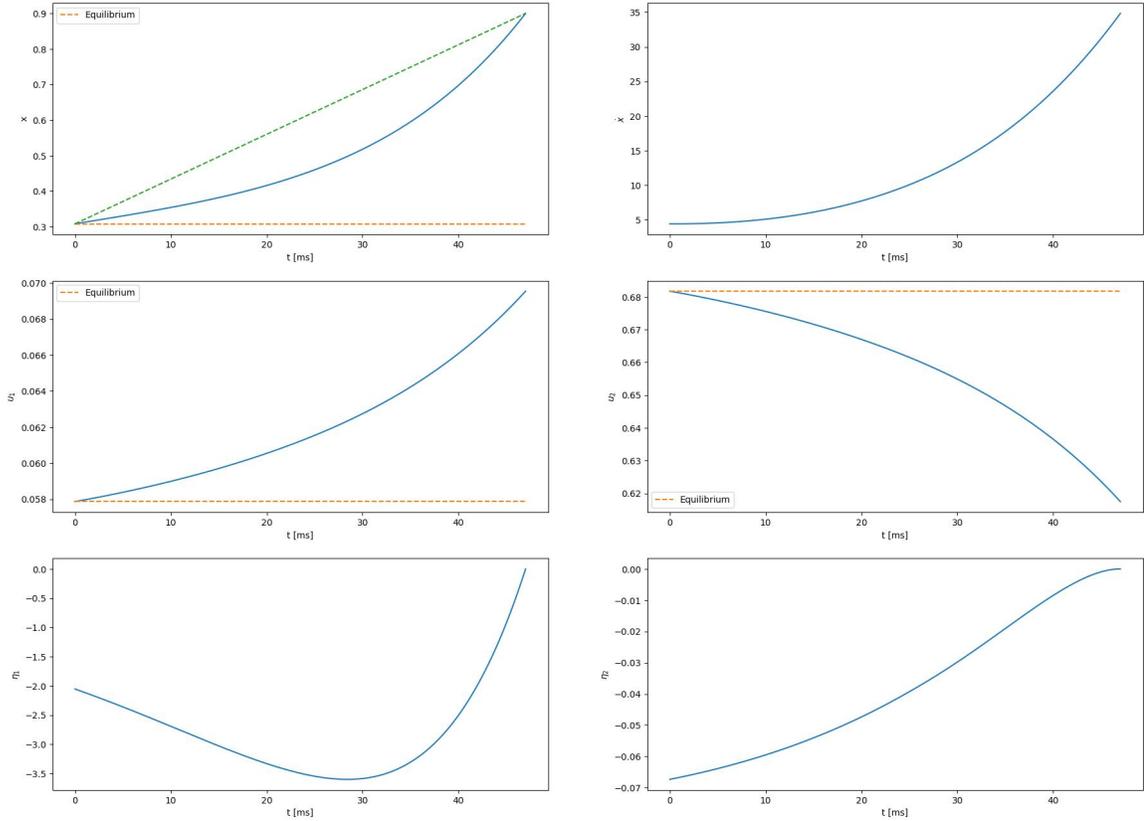
$$\begin{aligned} \eta_1(T) &= 0, \\ \eta_2(T) &= 0. \end{aligned} \tag{7.69}$$

Our system thus amounts to a boundary value ODE. It is solved using the shooting method. The unknown derivatives at time 0 (which must be solved for using the shooting method) are  $\dot{\eta}_1(0)$ ,  $\dot{\eta}_2(0)$ ,  $\dot{x}(0)$ . Three constraints at time  $T$  are used to solve for the unknowns at time 0 and for the boundaries given in Equation 7.67 and Equation 7.69.

### 7.2.5 Numerical Results

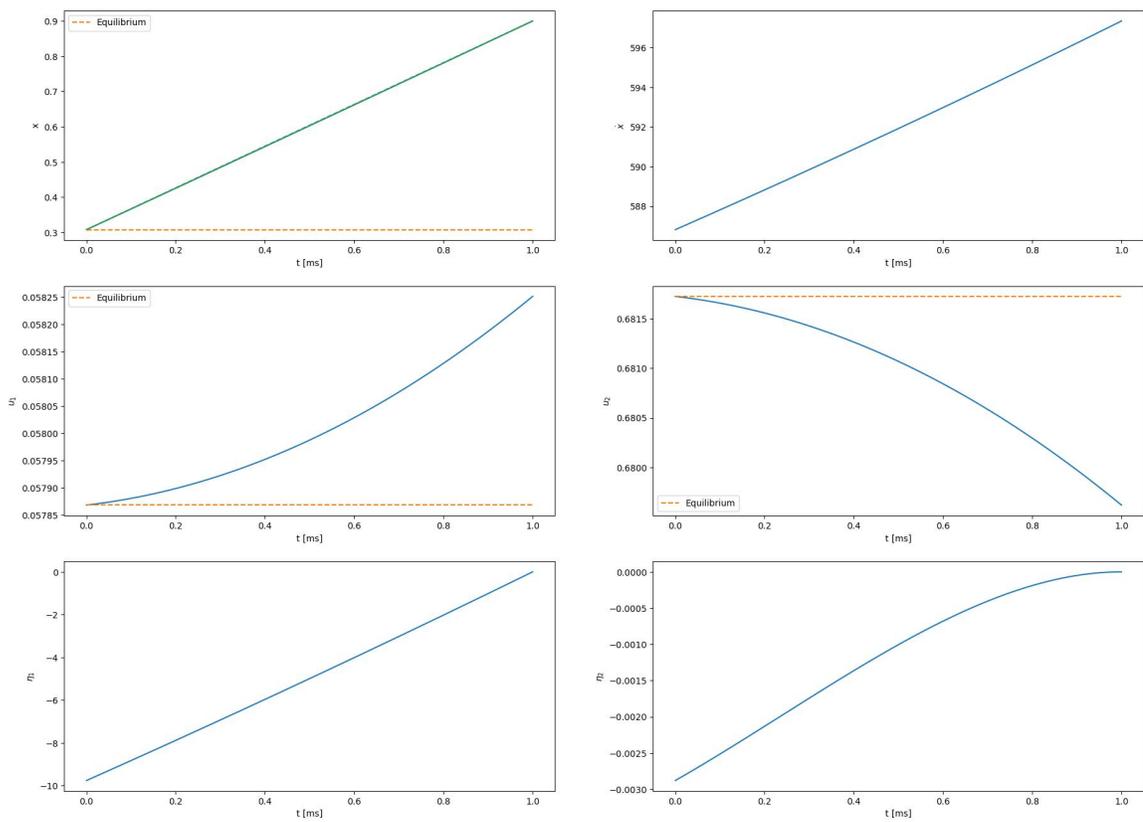
Figure 7.1 shows an example of the numerical solution for a path trajectory. We have demonstrated that LDT can be used to compute the most likely trajectory followed by  $N$  stochastic channels / diffusing calcium to make the spark-to-wave transition [47]. The final time,  $T_f = 46$  ms and  $\hat{x} = 0.9$  were chosen to clearly show the system

finding a path for any (even exaggerated) given conditions. The orange lines represent the equilibrium trajectories. The green line shows the *shortest* path to contrast with the *optimal* path.



**Figure 7.1** Example 1: Numerical results of optimal path trajectory.

To further highlight that the optimal path is calculated for a given set of boundary conditions, Figure 7.2 shows another solution for the optimal path, this time for  $T_f = 1$  ms. Note the difference in the system's path trajectory to achieve the desired result.



**Figure 7.2** Example 2: Numerical results of optimal path trajectory.

## CHAPTER 8

### CURRENT AND FUTURE WORK

One notable advantage of working with biological systems is the perpetual opportunity to enhance the complexity and biological relevance of the models. This inherent flexibility provides numerous avenues for further research and model refinement, critical for advancing our understanding of complex biological phenomena.

This dissertation presented the neural network model under connectivity-independent initial conditions. This simplification has allowed us to develop foundational insights into the model's behavior. However, it is also desirable to consider connectivity-dependent initial conditions, as the primary motivation is that over longer timescales (that diverge in  $N$ ), the system will attain other parts of the random energy landscape, so this would be necessary to understand better how this affects the dynamics. This scenario is inherently more complex, necessitating a uniform Large Deviation Principle (LDP) for the conditioned probability laws. Our forthcoming paper [105] (submitted to *Annals of Applied Probability* at the time of this writing) will address this challenge, comprehensively analyzing connectivity-dependent initial conditions, in addition to those mentioned in this work. This extension is expected to yield more profound insights into the neural network's dynamics, mainly how initial connectivity influences long-term behavior and emergent properties. Addressing these complexities will enhance the accuracy and applicability of our predictive models, making them more relevant to real-world neural systems.

Our current efforts focus on developing a specially extended version of the calcium signaling model that captures a broader range of biological phenomena and interactions. The existing one-dimensional model has laid the groundwork for understanding fundamental signaling dynamics. Nevertheless, extending the model

will allow us to incorporate additional biological details, such as spatial heterogeneity and temporal dynamics. The Barbet laboratory's experimental data reveal that the wave has a clear spatial aspect and is also very spatially heterogeneous, highlighting the value of this work.

Concurrently, we are establishing a LDP for this extended model. The numerical validation process is critical to our ongoing research. This process is computationally intensive for the spatially extended model and requires significant resources, but it is essential to verify the robustness and applicability of our model. This validation will ensure the model accurately reflects biological reality and can be used confidently in further studies and applications.

These ongoing and future projects are expected to significantly contribute to the field by providing more precise and comprehensive models that better reflect the complexities of biological systems. These models' continuous development and validation will facilitate discoveries and a deeper understanding of the underlying biological processes. Ultimately, this work will contribute to the broader goal of developing predictive, mechanistic models that can be used to design experiments, interpret data, and generate new hypotheses in studying biological systems.

By pushing the boundaries of current modeling approaches and addressing the complexities inherent in biological systems, we aim to provide tools and insights to drive future research and innovation in the field. The progress outlined here underscores the importance of an iterative approach to model development, where theoretical advancements and numerical validations work in tandem to enhance our understanding of complex biological phenomena.

## CHAPTER 9

### CONCLUSION

This dissertation focused on developing stochastic models and applying the Large Deviation Theory (LDT) to them, with a particular emphasis on biological systems. Exploring two key examples, random neural networks and calcium signaling in biological cells, provided significant insights into the behavior of complex stochastic processes.

The primary contributions included establishing a Large Deviation Principle (LDP) for a spin glass type system and developing one for a PDMP modeled system. In the first example, a random neural network was analyzed using LDT. By examining systems with independent Brownian Motions, the large  $N$  limit of the empirical measure was determined, and convergence to a fixed point was shown. This analysis afforded a relatively concise macroscopic description of the behavior of a large ensemble of neurons. Additionally, the dissertation demonstrated a uniform LDP for uncoupled systems and derived an LDP for the coupled system by performing an exponential change of measure using Girsanov's Theorem.

The second example focused on calcium signaling in biological cells. A PDMP model for calcium concentration dynamics was developed and characterized by deterministic flow, switching rates, and probability measures. This model was validated through parameter fitting and Markov chain Monte Carlo (MCMC) analysis, accurately representing observed calcium dynamics. Applying LDT to this model enabled the estimation of probabilities for cell-wide calcium waves, providing detailed insights into the spark-to-wave transition and identifying the most likely trajectories for this critical biological process. The transformation of the PDMP system into homogeneous Poisson processes via time-rescaling was linked to the general LDP through the

Inverse Contraction Principle. This technique framed the problem as a constrained optimization task and allowed the derivation of Euler-Lagrange equations for optimal trajectories, advancing the theoretical framework of LDT.

In summary, this dissertation advanced the theoretical framework of stochastic modeling and large deviation theory, applying these concepts to complex biological systems. Integrating theoretical models with experimental data underscored the practical relevance of this work and its potential impact across various scientific domains. The insights gained from analyzing random neural networks and calcium signaling provided a solid foundation for future research in computational neuroscience and cellular biology, highlighting the utility of mathematical modeling in understanding intricate stochastic processes.

## APPENDIX A

### BOUNDING FLUCTUATIONS OF THE NOISE

This appendix provides a detailed proof of the Lemma 5.1.1 stated in the main body of the text in Subsection 5.1, which is crucial for our subsequent analysis. The lemma asserts that for any given  $L > 0$ , there exists a positive constant  $a$  such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{Q}_a) \leq -L. \quad (\text{A.1})$$

Here,  $\hat{\mu}^N(\mathbf{y})$  represents an empirical measure associated with a sample  $\mathbf{y}$  of size  $N$ , and  $\mathcal{Q}_a$  denotes a specified set parameterized by  $a$ . This Lemma essentially provides a large deviation bound indicating that the probability of the empirical measure  $\hat{\mu}^N(\mathbf{y})$  falling outside the set  $\mathcal{Q}_a$  decays exponentially with rate at least  $L$ .

For the processes  $(\mathbf{y}_{[0,T]}^j)_{j \in I_N}$  that are defined in (5.53), define the empirical measure

$$\hat{\mu}^N(\mathbf{y}) = N^{-1} \sum_{j \in I_N} \delta_{\mathbf{y}_{[0,T]}^j} \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)). \quad (\text{A.2})$$

Next, we bound the probability of the empirical being in the set  $\mathcal{Q}_a$ , defined in (5.2), which we recall

$$\mathcal{Q}_a = \left\{ \mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)) : \sup_{m \geq \mathbf{a}} \sup_{0 \leq i \leq m} \mathbb{E}^\mu \left[ \sup_{M \in I_M} (w_{t_{i+1}^{(m)}}^p - w_{t_i^{(m)}}^p)^2 \right] > \Delta_m^{1/4} \text{ and} \right. \\ \left. \mu \in \mathcal{K}_a \text{ and } \sup_{p \in I_M} \mathbb{E}^\mu \left[ \sup_{t \in [0, T]} (y_t^p)^2 \right] \leq \mathbf{a} \right\} \quad (\text{A.3})$$

where  $\Delta_m = T/m$  and  $t_i^{(m)} = iT/m$ . The main result of this section is the following.

**Lemma A.1.** For any  $L > 0$ , there exists  $\mathbf{a} \in \mathbb{Z}^+$  such that for all  $N \geq 1$ ,

$$\sup_{(\mathbf{z}_0, \mathbf{g}_0)} N^{-1} \log P_{\mathbf{z}_0}^N(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{Q}_{\mathbf{a}}) \leq -L. \quad (\text{A.4})$$

*Proof.* Employing a union-of-events bound, or any  $(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N$ ,

$$\begin{aligned} N^{-1} \log P_{\mathbf{z}_0}^N(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{Q}_{\mathbf{a}}) &\leq N^{-1} \log \left\{ P_{\mathbf{z}_0}^N \left( \sup_{p \in I_M} N^{-1} \sum_{j \in I_N} \sup_{t \in [0, T]} (y_t^{p,j})^2 > \mathbf{a} \right) \right. \\ &\left. + P_{\mathbf{z}_0}^N \left( \sup_{0 \leq t \leq \Delta_m} \sum_{j \in I_N} \sup_{0 \leq i \leq m-1} \sup_{p \in I_M} |y_{t+t_i^{(m)}}^{p,j} - y_{t_i^{(m)}}^{p,j}|^2 \geq Na\Delta_m \right) + P_{\mathbf{z}_0}^N(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{K}_{\mathbf{a}}) \right\}. \end{aligned} \quad (\text{A.5})$$

To bound the first term on the RHS, since  $y_0^{p,j} = z_0^{p,j}$ ,

$$(y_t^{p,j})^2 \leq 2(y_t^{p,j} - y_0^{p,j})^2 + 2(z_0^{p,j})^2.$$

Thus, for a positive constant  $b > 0$ ,

$$\begin{aligned} \mathbb{E}_{\mathbf{z}_0}^{P_{\mathbf{z}_0}^N} \left[ \exp \left( b \sup_{p \in I_M} \sum_{j \in I_N} \sup_{t \in [0, T]} (y_t^{p,j})^2 \right) \right] \\ \leq \mathbb{E}_{\mathbf{z}_0}^{P_{\mathbf{z}_0}^N} \left[ \exp \left( 2b \sup_{p \in I_M} \sum_{j \in I_N} (z_0^{p,j})^2 + 2b \sum_{j \in I_N} \sup_{t \in [0, T]} (y_t^{p,j} - z_0^{p,j})^2 \right) \right]. \end{aligned} \quad (\text{A.6})$$

Thus, thanks to Chernoff's Inequality,

$$N^{-1} \log P_{\mathbf{z}_0}^N \left( \sup_{p \in I_M} N^{-1} \sum_{j \in I_N} \sup_{t \in [0, T]} (y_t^{p,j})^2 > \mathbf{a} \right) \leq \frac{2b}{N} \sup_{p \in I_M} \sum_{j \in I_N} (z_0^{p,j})^2 \quad (\text{A.7})$$

$$+ N^{-1} \log \mathbb{E}_{\mathbf{z}_0}^{P_{\mathbf{z}_0}^N} \left[ \exp \left( 2b \sum_{j \in I_N} \sup_{t \in [0, T]} (y_t^{p,j} - z_0^{p,j})^2 \right) \right] - b\mathbf{a}. \quad (\text{A.8})$$

The first term on the RHS is bounded for all  $N$  and all  $(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N$ . For the second term on the RHS, standard theory on stochastic processes implies that the exponential

moment exists as long as  $b$  is small enough. Thus, taking  $\mathbf{a} \rightarrow \infty$ , the RHS can be made arbitrarily small. We thus find that

$$\lim_{\mathbf{a} \rightarrow \infty} \sup_{N \geq 1} \sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \log P_{\mathbf{z}_0}^N \left( \sup_{p \in I_M} N^{-1} \sum_{j \in I_N} \sup_{t \in [0, T]} (y_t^{p,j})^2 > \mathbf{a} \right) = -\infty. \quad (\text{A.9})$$

The Lemma now follows from applying (A.9), Lemma A.3 and Lemma A.2 to (A.5).  $\square$

The following result is well-known. Nevertheless, we sketch a quick proof for clarity.

**Lemma A.2.** *For any  $L > 0$ , there exists a compact set  $\mathcal{K}_L$  such that for all  $N \geq 1$ ,*

$$\sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \log P_{\mathbf{z}_0}^N (\hat{\mu}^N(\mathbf{y}) \notin \mathcal{K}_L) \leq -L \quad (\text{A.10})$$

*Proof.* The following property follows straightforwardly from properties of the stochastic integral (noting that the diffusion coefficient is uniformly bounded): for any  $\epsilon > 0$ , there exists a compact set  $\mathcal{C}_\epsilon \subset \mathcal{C}([0, T], \mathbb{R}^M)$  such that for all  $j \in I_N$  such that  $\|z_0^j\| \leq \epsilon^{-1}$ ,

$$\sup_{j \in I_N} P_{\mathbf{z}_0}^N (y_{[0, T]}^j \notin \mathcal{C}_\epsilon) \leq \epsilon. \quad (\text{A.11})$$

Write

$$u_\epsilon^N = \sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \sum_{j \in I_N} \chi_{\{\|y_0^j\| \geq \epsilon^{-1}\}}, \quad (\text{A.12})$$

and note that our assumptions on  $\mathcal{Y}^N$  dictates that

$$\lim_{\epsilon \rightarrow 0^+} \lim_{N \rightarrow \infty} u_\epsilon^N = 0. \quad (\text{A.13})$$

For any  $m \in \mathbb{Z}^+$ , define the set  $\mathcal{L}_{m,\delta} \subset \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M))$  to be such that

$$\mathcal{L}_{m,\delta} = \{\mu \in \mathcal{P}(\mathcal{C}([0, T], \mathbb{R}^M)) : \mu(\mathcal{C}_{m-1}) \geq \delta\} \quad (\text{A.14})$$

We claim that for any  $m \geq 1$ , there exists  $\delta_m > 0$  such that

$$\sup_{N \geq 1} \sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \log P_{\mathbf{z}_0}^N(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{L}_{m,\delta_m}) \leq -m \quad (\text{A.15})$$

To see this, employing a Chernoff Inequality, for a constant  $b > 0$ , for any  $(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N$ ,

$$N^{-1} \log P_{\mathbf{z}_0}^N(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{L}_{m,\delta}) \leq \mathbb{E}^{P_{\mathbf{z}_0}^N} \left[ \exp \left( b \sum_{j \in I_N} \chi\{\mathbf{y}_{[0,T]}^j \notin \mathcal{C}_{m-1}\} - Nb\delta \right) \right] \quad (\text{A.16})$$

$$\leq -b\delta + N^{-1} \log \{(\epsilon + u_\epsilon^N)(\exp(b) - 1) + 1\}^N \quad (\text{A.17})$$

$$= -b\delta + \log \{(\epsilon + u_\epsilon^N)(\exp(b) - 1) + 1\} \quad (\text{A.18})$$

Taking  $\epsilon$  to be sufficiently small and  $b$  sufficiently large, we obtain (A.15).

Now, for an integer  $m_L$  to be specified further below, define  $\mathcal{K}_L = \bigcap_{m \geq m_L} \mathcal{L}_{m,\delta_m}$ . Prokhorov's Theorem implies that  $\mathcal{K}_L$  is compact. Employing a union-of-events bound, we obtain that

$$P_{\mathbf{z}_0}^N(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{K}_L) \leq \sum_{m \geq m_L} \exp(-mN) \quad (\text{A.19})$$

$$\leq \exp(-m_L N) \sup_{n \geq 1} \sum_{j=0}^{\infty} \exp(-jN). \quad (\text{A.20})$$

We thus find that, for large enough  $m_L$ ,

$$\sup_{N \geq 1} N^{-1} \log \mathbb{P}(\hat{\mu}^N(\mathbf{y}) \notin \mathcal{K}_L) \leq -L,$$

as required. □

**Lemma A.3.** *There exists a constant  $\mathfrak{C}$  such that for any positive integer  $m$  and any  $a > 0$ , writing  $\Delta_m = Tm^{-1}$  and  $t_i^{(m)} = Ti/m$ , for any  $N \geq 1$ ,*

$$\sup_{(\mathbf{z}_0, \mathbf{g}_0) \in \mathcal{Y}^N} N^{-1} \log P_{\mathbf{z}_0}^N \left( \sup_{0 \leq t \leq \Delta_m} \sum_{j \in I_N} \sup_{0 \leq i \leq m-1} \sup_{p \in I_M} |y_{t+t_i^{(m)}}^{p,j} - y_{t_i^{(m)}}^{p,j}|^2 \geq Na\Delta_m \right) \leq \mathfrak{C} + \log m - \frac{a}{4} \quad (\text{A.21})$$

*Proof.* Define, for  $t \in [0, \Delta_m)$ ,

$$f_t^N = \sum_{j \in I_N} \sup_{0 \leq i \leq m-1} \sup_{p \in I_M} (y_{t+t_i^{(m)}}^{p,j} - y_{t_i^{(m)}}^{p,j})^2$$

Notice that  $t \rightarrow f_t^N$  is a submartingale. Thus, writing  $a = (4\Delta_m)^{-1}$ ,  $\exp(af_t^N)$  is a submartingale. Therefore, thanks to Doob's Submartingale Inequality,

$$\mathbb{P}(f_t^N \geq Nx) \leq \mathbb{E} \left[ \exp(af_t^N - aNx) \right] \quad (\text{A.22})$$

$$\leq \{mM(1 - 2\Delta_m \bar{\sigma} a)^{-1/2}\}^N \exp(-aNx) \quad (\text{A.23})$$

$$= \{mM2^{1/2}\}^N \exp(-Nx/(4\Delta_m)) \quad (\text{A.24})$$

□

## APPENDIX B

### PROOFS FOR LDP FOR PDMP

In this appendix, we prove Theorem 7.1.1 found in the main body of the text in Subsection 7.1.2. The statement of the theorem is as follows:

Suppose that  $\mathcal{O}, \mathcal{A} \subset \Upsilon$ , with  $\mathcal{O}$  open and  $\mathcal{A}$  closed, are such that for any  $T > 0$ ,

$$\inf_{(z,x,u) \in \mathcal{A} \cup \mathcal{O}} \inf_{\alpha \in \mathbb{R}} \inf_{s \in [0,T]} \lambda_\alpha(x(s), u(s)) > 0.$$

Then there exists a function  $\mathcal{J} : \Upsilon \rightarrow \mathbb{R}$  (specified in Subsection 7.1.3) that is (i) lower-semicontinuous and (ii) has compact level sets such that

$$\begin{aligned} \overline{\lim}_{N \rightarrow \infty} N^{-1} \log P((z, x, u) \in \mathcal{A}) &\leq - \inf_{\beta \in \mathcal{A}} J(\beta) \\ \underline{\lim}_{N \rightarrow \infty} N^{-1} \log P((z, x, u) \in \mathcal{O}) &\geq - \inf_{\beta \in \mathcal{O}} J(\beta) \end{aligned}$$

We apply the contraction principle to the Large Deviation Principle (LDP) for the independent system. In other words, we will perform a change of variable, which will transfer known results for the Large Deviations of independent Poisson Processes [141, Chapter 6]. We first recount the Large Deviations for independent Poisson Processes (recalling the definition of  $\{y_\alpha(t)\}_{\alpha \in \mathcal{A}}$  in (7.19)).

**Theorem B.1.** *The sequence of probability laws of  $(y_\alpha)_{\alpha \in \mathcal{A}} \in \tilde{\mathcal{D}}([0, \infty), \mathbb{R}^+)^M$  satisfy a LDP with good rate function.*

*That is, for  $\mathcal{O}, \mathcal{A} \subset \tilde{\mathcal{D}}([0, \infty), \mathbb{R}^+)^M$ , with  $\mathcal{O}$  open and  $\mathcal{A}$  closed,*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(y \in \mathcal{A}) \leq - \inf_{\alpha \in \mathcal{A}} \mathcal{I}(\alpha) \tag{B.1}$$

$$\underline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(y \in \mathcal{O}) \geq - \inf_{\alpha \in \mathcal{O}} \mathcal{I}(\alpha), \tag{B.2}$$

and the so-called ‘local rate function’ is

$$\mathcal{I}(y) = \begin{cases} \infty & \text{in the case that } y_\alpha \text{ is not absolutely continuous for some } \alpha \in \mathcal{R} \\ \sum_{\alpha \in \mathcal{R}} \int_0^\infty \ell(\dot{y}_\alpha(r)) dr & \text{otherwise,} \end{cases}$$

where

$$\ell(r) = r \log r - r + 1.$$

(B.3)

Furthermore, the level sets of  $\mathcal{I}$  are compact.

*Proof.* The Large Deviations Principle for arbitrarily long finite time intervals is proved in [141, Chapter 5], with corresponding rate function

$$\mathcal{I}(y) = \sum_{\alpha \in \mathcal{R}} \int_0^\infty \ell(\dot{y}_\alpha(r)) dr. \quad (\text{B.4})$$

□

The first result we must prove is that the Large Deviations Principle must hold for sets with reaction rates bounded away from zero.

**Lemma B.2.** *Suppose that  $\mathcal{O}, \mathcal{A} \in \mathcal{B}(\Upsilon)$ , with  $\mathcal{O}$  open and  $\mathcal{A}$  closed, are such that for any  $T > 0$ ,*

$$\inf_{(z,x,u) \in \mathcal{A} \cup \mathcal{O}} \inf_{\alpha \in \mathcal{R}} \inf_{s \in [0,T]} \lambda_\alpha(x(s), u(s)) > 0. \quad (\text{B.5})$$

*Then the sequence of probability laws of  $(z, x, u)$  satisfy a LDP on the space  $\Upsilon$ , i.e.*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\beta \in \mathcal{A}) \leq - \inf_{\beta \in \mathcal{A}} \mathcal{J}(\beta) \quad (\text{B.6})$$

$$\underline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\beta \in \mathcal{O}) \geq - \inf_{\beta \in \mathcal{O}} \mathcal{J}(\beta). \quad (\text{B.7})$$

*$\mathcal{J}$  is lower-semi-continuous and the level sets of  $\mathcal{J}$  are compact.*

*Proof.* We will prove the LDP using a contraction principle. To this end, define a mapping  $\Psi : \Upsilon \rightarrow \tilde{\mathcal{D}}([0, \infty), \mathbb{R}^+)^M$  as follows. For any  $(z, x, u) \in \Upsilon$ , define

$$\Lambda_\alpha(t) = \int_0^t \lambda_\alpha(x(s), u(s)) ds. \quad (\text{B.8})$$

Let  $\Lambda_\alpha^{-1} \in \tilde{\mathcal{D}}([0, \infty), \mathbb{R}^+)$  be such that

$$\Lambda_\alpha^{-1}(t) = \inf \{s \geq 0 : \Lambda_\alpha(s) = t\}. \quad (\text{B.9})$$

Note that if  $\inf_{s \in [0, t]} \lambda_\alpha(x(s), u(s)) > 0$ , then  $\Lambda_\alpha^{-1}$  is the function inverse of  $\Lambda_\alpha$ . Write

$$\tau_\alpha = \lim_{t \rightarrow \infty} \Lambda_\alpha(t), \quad (\text{B.10})$$

and note that  $\tau_\alpha$  could be  $\infty$ . We define  $\Psi(z, x, u) = (w_\alpha)_{\alpha \in \mathcal{R}}$ , where for any  $t < \tau_\alpha$ ,

$$w_\alpha(t) = z_\alpha(\Lambda_\alpha^{-1}(t)). \quad (\text{B.11})$$

In the case that  $\tau_\alpha < \infty$ , for all  $t \geq \tau_\alpha$ , we define  $w_\alpha(t) = \lim_{s \rightarrow \tau_\alpha^-} w_\alpha(s)$ . One observes that over sets of the form (B.5),  $\Psi$  is continuous and one-to-one. Furthermore,  $w_\alpha(t) = y_\alpha(t)$ . The lemma follows from the Inverse Contraction Principle [45], and Lemmas B.1 and B.4.

□

**Lemma B.3.** *Over any set  $A_\epsilon$  of the form*

$$A_\epsilon = \inf_{(z, x, u) \in \Upsilon} \inf_{\alpha \in \mathbb{R}} \inf_{s \in [0, T]} \lambda_\alpha(x(s), u(s)) \geq \epsilon. \quad (\text{B.12})$$

$\Psi$  is one-to-one.

*Proof.* Let  $(z, x, u), (\hat{z}, \hat{x}, \hat{u}) \in \Upsilon$ . Write

$$\Lambda_\alpha(t) = \int_0^t \lambda_\alpha(x(s), u(s)) ds \quad (\text{B.13})$$

$$\hat{\Lambda}_\alpha(t) = \int_0^t \lambda_\alpha(\hat{x}(s), \hat{u}(s)) ds \quad (\text{B.14})$$

$$w_\alpha(t) = z_\alpha(\Lambda_\alpha^{-1}(t)) \quad (\text{B.15})$$

$$\hat{w}_\alpha(t) = \hat{z}_\alpha(\hat{\Lambda}_\alpha^{-1}(t)). \quad (\text{B.16})$$

$$(\text{B.17})$$

It is noted in Lemma 7.1.2 that there is a constant  $C_t > 0$  such that

$$\sup_{\alpha \in \mathbb{R}} \Lambda_\alpha(t) \leq C_t. \quad (\text{B.18})$$

We are going to demonstrate that for any  $t \geq 0$ , there exists a constant  $c_{\epsilon, t}$  such that for all  $s \leq t$ ,

$$\tilde{d}_s^\circ(z_\alpha, \hat{z}_\alpha) \leq \tilde{d}_{C_s}^\circ(w_\alpha, \hat{w}_\alpha) + c_{\epsilon, t} \tilde{d}_s^\circ(z_\alpha, \hat{z}_\alpha). \quad (\text{B.19})$$

An application of Gronwall's Inequality to Equation B.19 implies that if  $w_\alpha = \hat{w}_\alpha$  for all  $\alpha \in \mathbb{R}$ , then necessarily  $z_\alpha = \hat{z}_\alpha$ . This implies the Lemma.

Lemma the aim of proving Equation B.19, it follows from Calculus that

$$\Lambda_\alpha^{-1}(t) = \int_0^t \lambda_\alpha^{-1}(x(s), u(s)) ds \quad (\text{B.20})$$

$$\hat{\Lambda}_\alpha^{-1}(t) = \int_0^t \lambda_\alpha^{-1}(\hat{x}(s), \hat{u}(s)) ds. \quad (\text{B.21})$$

Our assumptions on the set  $A_\epsilon$  imply that

$$\Lambda_\alpha^{-1}(t) \leq t\epsilon^{-1} \quad (\text{B.22})$$

$$\hat{\Lambda}_\alpha^{-1}(t) \leq t\epsilon^{-1}. \quad (\text{B.23})$$

Define

$$\tilde{z}_\alpha(t) = w_\alpha(\hat{\Lambda}(t)). \quad (\text{B.24})$$

The triangle inequality implies that

$$\tilde{d}_t^\circ(z_\alpha, \hat{z}_\alpha) \leq \tilde{d}_t^\circ(z_\alpha, \tilde{z}_\alpha) + \tilde{d}_t^\circ(\tilde{z}_\alpha, \hat{z}_\alpha). \quad (\text{B.25})$$

The definition of the Skorohod metric implies that

$$\tilde{d}_t^\circ(z_\alpha, \tilde{z}_\alpha) \leq \sup_{s \leq t} \left| \int_0^s \frac{\lambda_\alpha(x(r), u(r))}{\lambda_\alpha(\hat{x}(r), \hat{u}(r))} dr - s \right|. \quad (\text{B.26})$$

Our assumption that the reaction rates and functions are Lipschitz, together with Lemma 7.1.2 implies that there is a constant  $\bar{C}_{\epsilon,t} > 0$  such that

$$\begin{aligned} \sup_{s \in [0,t]} \sup_{\alpha \in \mathbb{R}} |\lambda_\alpha(x(s), u(s)) - \lambda_\alpha(\hat{x}(s), \hat{u}(s))| &\leq \bar{C}_{\epsilon,t} \sup_{s \in [0,t]} \sup_{\alpha \in \mathbb{R}} |z_\alpha(s) - \hat{z}_\alpha(s)| \\ &\leq \bar{C}_{\epsilon,t} \sup_{\alpha \in \mathbb{R}} \tilde{d}_t^\circ(z_\alpha, \hat{z}_\alpha). \end{aligned} \quad (\text{B.27})$$

Since the reaction rates are (by assumption in the statement of the Lemma) bounded from below, we thus obtain that there is a constant  $\ell_{C_{\epsilon,t}}$  such that for all  $s \leq \bar{t}$ ,

$$\sup_{\alpha \in \mathbb{R}} \tilde{d}_s^\circ(z_\alpha, \tilde{z}_\alpha) \leq \ell_{C_{\epsilon,\bar{t}}} \sup_{\alpha \in \mathbb{R}} \tilde{d}_s^\circ(z_\alpha, \hat{z}_\alpha). \quad (\text{B.28})$$

The definition of the Skorokhod metric, together with the bounds in Equation B.22, also implies

$$\tilde{d}_t^\circ(\tilde{z}_\alpha, \hat{z}_\alpha) \leq \tilde{d}_{C_t}^\circ(w_\alpha, \hat{w}_\alpha). \quad (\text{B.29})$$

□

The following exponential tightness result is a standard requirement for LDPs.

**Lemma B.4.** *For any  $L > 0$ , there exists a compact set  $\mathcal{K}_L \subset \Upsilon$  such that*

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}((z, x, u) \notin \mathcal{K}_L) \leq -L. \quad (\text{B.30})$$

*Proof.* We can equivalently formulate the system in terms of an empirical measure, and the exponential tightness is an immediate consequence. In more detail, write

$$\hat{\mu}^N = N^{-1} \sum_{j \in I_N} \delta_{w^j} \in \mathcal{P}(\mathcal{D}([0, T], \mathbb{Z}^+)^{|\mathcal{R}|}). \quad (\text{B.31})$$

Here  $w^j := (w_\alpha^j)_{\alpha \in \mathcal{R}} \subset \mathcal{D}([0, T], \mathbb{Z}^+)^{|\mathcal{R}|}$  are inhomogeneous counting processes, i.e. they are such that

$$\mathbb{P}(w_\alpha^j(t + \Delta) = w_\alpha^j(t) + 1 \mid \mathcal{F}_t) \simeq \Delta \lambda_\alpha(x(t), u(t)) + O(\Delta^2) \quad (\text{B.32})$$

$$\mathbb{P}(w_\alpha^j(t + \Delta) = w_\alpha^j(t) \mid \mathcal{F}_t) \simeq 1 - \Delta \lambda_\alpha(x(t), u(t)) + O(\Delta^2). \quad (\text{B.33})$$

Here

$$x(t) = x(0) + N^{-1} \sum_{j \in I_N} \sum_{\alpha \in \mathcal{R}} \xi_\alpha w_\alpha^j(t) \quad (\text{B.34})$$

$$\frac{du}{dt} = A(u(t), x(t)) \quad (\text{B.35})$$

We then find that substituting  $z_\alpha(t) = N^{-1} \sum_{j=1}^N w_\alpha^j(t)$ , the above system has the same probability law as the original system. Write  $\mathcal{D}_*([0, T], \mathbb{Z}^+)^{|\mathcal{A}|} \subseteq \mathcal{D}([0, T], \mathbb{Z}^+)^{|\mathcal{A}|}$  to consist of all processes that are (i) equal to 0 at time 0, (ii) non-decreasing.

More precisely, we see that there exists a continuous mapping  $\Psi : \mathcal{P}(\mathcal{D}_*([0, T], \mathbb{Z}^+)^{|\mathcal{A}|})$  such that  $(z, x, u) = \Psi(\hat{\mu}^N)$ , with unit probability. For a positive number  $L > 0$ , we are going to define a compact set  $\mathcal{K}_L \subseteq \mathcal{P}(\mathcal{D}_*([0, T], \mathbb{Z}^+)^{|\mathcal{A}|})$  such that

$$\overline{\lim}_{N \rightarrow \infty} N^{-1} \log \mathbb{P}(\hat{\mu}^N \notin \mathcal{K}_L) \leq -L. \quad (\text{B.36})$$

This suffices for the lemma because the continuity of  $\Psi$  then implies that  $\Psi(\mathcal{K}_L)$  is compact.

For a positive integer  $p \geq 1$ , write  $\mathcal{U}_{p,T} \subseteq \mathcal{D}_*([0, T], \mathbb{Z}^+)^{|\mathcal{A}|}$  to consist of all paths that are less than or equal to  $p$  at time  $T$ . It is easy to check that  $\mathcal{U}_p$  is compact with respect to the Skorohod Topology (there are at most  $p$  ‘spike times,’ and these spike times must be in the compact time interval  $[0, T]$ ). Write  $Y_\alpha^{p,j}$  to be independent counting processes of unit intensity. Thus, time-rescaled representation of Poisson Processes [4] means that we can write

$$w_\alpha^j(t) = Y_\alpha^j \left( \int_0^t \lambda_\alpha(x(s), u(s)) ds \right) \quad (\text{B.37})$$

For another number  $b_p > 0$ , we find that

$$\mathbb{P}(\hat{\mu}^N(\mathbf{w})(\mathcal{U}_{p,T}) < 1 - b_p) \leq \mathbb{P}(\hat{\mu}^N(\mathbf{Y})(\mathcal{U}_{p,KT}) < 1 - b_p). \quad (\text{B.38})$$

since  $\lambda \leq K$  uniformly. Thanks to Chernoff's Inequality, for a constant  $c > 0$ ,

$$\begin{aligned}
\mathbb{P}(\hat{\mu}^N(\mathbf{Y})(\mathcal{U}_{p,KT}) < 1 - b_p) &\leq \mathbb{P}\left(\sup_{\alpha \in \mathcal{R}} N^{-1} \sum_{j=1}^N \chi\{w_\alpha^j(KT) > p\} \geq b_p\right) \\
&\leq |\mathcal{R}| \mathbb{E}\left[\exp\left(c \sum_{j=1}^N \chi\{w_\alpha^j(KT) > p\} - Ncb_p\right)\right] \\
&= |\mathcal{R}| \left\{1 + \mathbb{P}(w_\alpha^j(KT) > p)(\exp(c) - 1)\right\}^N \exp(-Ncb_p)
\end{aligned} \tag{B.39}$$

Through taking  $p$  large enough, and  $1 \ll c \ll -\log \mathbb{P}(w_\alpha^j(KT) > p)$ , we find that

$$\mathbb{P}(\hat{\mu}^N(\mathbf{w})(\mathcal{U}_{p,T}) < 1 - b_p) \leq \exp(-pN). \tag{B.40}$$

Now define  $\mathcal{K}_L$  to consist of all measures  $\mu$  such that for all  $p \geq p_L$  (for an integer  $p_L$  to be specified below),

$$\mu(\mathcal{U}_{p,KT}) \geq b_p. \tag{B.41}$$

We thus find through a union of events bound that

$$\begin{aligned}
\mathbb{P}(\hat{\mu}^N \notin \mathcal{K}_L) &\leq \sum_{p=p_L}^{\infty} \mathbb{P}(\hat{\mu}^N(\mathbf{Y})(\mathcal{U}_{p,KT}) < 1 - b_p) \\
&\leq \sum_{p=p_L}^{\infty} \exp(-pN).
\end{aligned} \tag{B.42}$$

Thus for large enough  $p_L$ , it must be that for all  $N \geq 1$ ,

$$N^{-1} \log \mathbb{P}(\hat{\mu}^N \notin \mathcal{K}_L) \leq -L. \tag{B.43}$$

□

We finally note the proof of Lemma 7.1.2.

*Proof.* It is clear from the definition of  $x$  that there is a universal constant such that

$$\|x(s) - \tilde{x}(s)\| \leq C \sup_{\alpha \in \mathbb{R}} |z_\alpha(s) - \tilde{z}_\alpha(s)|. \quad (\text{B.44})$$

It is demonstrated in Billingsley [17] that convergence in the Skorohod metric implies convergence in the supremum norm, i.e., as  $d_\infty^\circ(z, \tilde{z}) \rightarrow 0$ , it must be that

$$\sup_{\alpha \in \mathbb{R}} \sup_{0 \leq s \leq t} |z_\alpha(s) - \tilde{z}_\alpha(s)| \rightarrow 0 \quad (\text{B.45})$$

The Lemma now follows from a standard application of Gronwall's Inequality.  $\square$

## REFERENCES

- [1] R. Adler and J. Taylor. *Random Fields and Geometry 53*. Springer, New York, NY, USA, 1st edition, 2019.
- [2] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27(2):77–87, 1977.
- [3] S. Amari. Homogeneous nets of neuron-like elements. *Biological Cybernetics*, 17(4):211–220, 1975.
- [4] D. F. Anderson and T. G. Kurtz. *Stochastic Analysis of Biochemical Systems*. Springer, New York, NY, USA, 1st edition, 2015.
- [5] R. B. Ash and C. A. Doléans-Dade. *Probability and Measure Theory*. Academic Press, San Diego, CA, USA, 1st edition, 2000.
- [6] R. Azaïs and A. Muller-Gueudin. Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic markov processes. *Electronic Journal of Statistics*, 10(2):3648–3692, 2016.
- [7] F. Baccelli and T. Taillefumier. Replica-mean-field limits for intensity-based neural networks. *SIAM Journal on Applied Dynamical Systems*, 18(4):1756–1797, 2019.
- [8] F. Baccelli and T. Taillefumier. The pair-replica-mean-field limit for intensity-based neural networks. *SIAM Journal on Applied Dynamical Systems*, 20(1):165–207, 2021.
- [9] M. Bear, B. Connors, and M. A. Paradiso. *Neuroscience: Exploring the Brain, Enhanced Edition: Exploring the Brain*. Jones & Bartlett Learning, Burlington, MA, USA, 1st edition, 2020.
- [10] G. Ben Arous and A. Guionnet. Large deviations for langevin spin glass dynamics. *Probability Theory and Related Fields*, 102:455–509, 1995.
- [11] G. Ben Arous and A. Guionnet. Symmetric langevin spin glass dynamics. *The Annals of Probability*, 25(3):1367–1422, 1997.
- [12] G. Ben Arous, A. Dembo, and A. Guionnet. Cugliandolo-kurchan equations for dynamics of spin-glasses. *Probability Theory and Related Fields*, 136:619–660, 2006.
- [13] O. Benjamin, T. H. Fitzgerald, P. Ashwin, K. Tsaneva-Atanasova, F. Chowdhury, M. P. Richardson, and J. R. Terry. A phenomenological model of seizure initiation suggests network structure 5 explain seizure frequency in idiopathic generalised epilepsy. *The Journal of Mathematical Neuroscience*, 2:1–30, 2012.

- [14] H. C. Berg. *Random Walks in Biology*. Princeton University Press, Princeton, NJ, USA, 1st edition, 1993.
- [15] M. J. Berridge, P. Lipp, and M. D. Bootman. The versatility and universality of calcium signalling. *Nature Reviews Molecular Cell Biology*, 1(1):11–21, 2000.
- [16] R. L. Beurle. Properties of a mass of cells capable of regenerating pulses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*:55–94, 1956.
- [17] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons, Hoboken, NJ, USA, 1st edition, 2013.
- [18] M. Breakspear. Dynamic models of large-scale brain activity. *Nature Neuroscience*, 20(3):340–352, 2017.
- [19] P. C. Bressloff and J. N. Maclaurin. Stochastic hybrid systems in cellular neuroscience. *Journal of Mathematical Neuroscience*, 8(1):12, 2018.
- [20] P. Bressloff. *Stochastic Processes in Cell Biology*. Springer, New York, NY, USA, 2nd edition, 2021.
- [21] P. C. Bressloff. From invasion to extinction in heterogeneous neural fields. *The Journal of Mathematical Neuroscience*, 2:1–27, 2012.
- [22] P. C. Bressloff. Stochastic neural field theory. *Neural Fields: Theory and Applications*:235–268, 2014.
- [23] P. C. Bressloff and J. M. Newby. Path integrals and large deviations in stochastic hybrid systems. *Physical Review E*, 89(4):042701, 2014.
- [24] P. C. Bressloff and M. A. Webber. Front propagation in stochastic neural fields. *SIAM Journal on Applied Dynamical Systems*, 11(2):708–740, 2012.
- [25] P. C. Bressloff and J. Wilkerson. Traveling pulses in a stochastic neural field model of direction selectivity. *Frontiers in Computational Neuroscience*, 6:90, 2012.
- [26] N. Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, 8:183–208, 2000.
- [27] A. Budhiraja and P. Dupuis. Analysis and approximation of rare events. *Representations and Weak Convergence Methods. Series Prob. Theory and Stoch. Modelling*, 94, 2019.
- [28] A. N. Burkitt. A review of the integrate-and-fire neuron model: i. homogeneous synaptic input. *Biological Cybernetics*, 95:1–19, 2006.

- [29] T. Cabana and J. D. Touboul. Large deviations for randomly connected neural networks: i. spatially extended systems. *Advances in Applied Probability*, 50:983–1004, 2018.
- [30] T. Cabana and J. D. Touboul. Large deviations for randomly connected neural networks: ii. state-dependent interactions. *Advances in Applied Probability*, 50:983–1004, 2018.
- [31] P. Cao, G. Donovan, M. Falcke, and J. Sneyd. A stochastic model of calcium puffs based on single-channel data. *Biophysical Journal*, 105(5):1133–1142, 2013.
- [32] P. Charbonneau, E. Marinari, M. Mezard, G. Parisi, F. Ricci-Tersenghi, G. Sicuro, and F. Zamponi, editors. *Spin Glass Theory and Far Beyond*. World Scientific, Hackensack, NJ, USA, 1st edition, 2023.
- [33] D. E. Clapham. Calcium signaling. *Cell*, 131(6):1047–1058, 2007.
- [34] B. J. Cook, A. D. Peterson, W. Woldman, and J. R. Terry. Neural field models: a mathematical overview and unifying framework. *Mathematical Neuroscience and Applications*, 2(1):1–54, 2022.
- [35] S. Coombes, R. Hinch, and Y. Timofeeva. Receptors, sparks and waves in a fire-diffuse-fire framework for calcium release. *Progress in Biophysics and Molecular Biology*, 85(2-3):197–216, 2004.
- [36] S. Coombes and Y. Timofeeva. Sparks and waves in a stochastic fire-diffuse-fire model of ca<sup>2+</sup> release. *Physical Review E*, 68(2):021915, 2003.
- [37] A. Crisanti and H. Sompolinsky. Path integral approach to random neural networks. *Physical Review E*, 98(6):062120, 2018.
- [38] A. Crisanti, H. Horner, and H. J. Sommers. The spherical p-spin interaction spin-glass model. *Zeitschrift fur Physik B Condensed Matter*, 92:257–271, 1993.
- [39] L. F. Cugliandolo and J. Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173–176, 1993.
- [40] M. H. Davis. Piecewise-deterministic markov processes: a general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.
- [41] P. Dayan and L. F. Abbott. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. MIT Press, Cambridge, MA, USA, 1st edition, 2005.

- [42] H. De Garis, C. Shuo, B. Goertzel, and L. Ruiting. A world survey of artificial brain projects, part i: large-scale brain simulations. *Neurocomputing*, 74(1-3):3–29, 2010.
- [43] G. Deco, V. K. Jirsa, P. A. Robinson, M. Breakspear, and K. Friston. The dynamic brain: from spiking neurons to neural masses and cortical fields. *PLoS Computational Biology*, 4(8):e1000092, 2008.
- [44] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer, New York, NY, USA, 2nd edition, 1998.
- [45] A. Dembo. *Large Deviations Techniques and Applications*. Springer, New York, NY, USA, 1st edition, 2009.
- [46] G. Dupont, L. Combettes, G. S. Bird, and J. W. Putney. Calcium oscillations. *Cold Spring Harbor Perspectives in Biology*, 3(3):a004226, 2011.
- [47] G. Dupont, M. Falcke, V. Kirk, and J. Sneyd. *Models of Calcium Signalling*, volume 43. Springer, New York, NY, USA, 1st edition, 2016.
- [48] W. E and E. Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61:391–420, May 2010.
- [49] C. Eliasmith and O. Trujillo. The use and abuse of large-scale brain models. *Current Opinion in Neurobiology*, 25:1–6, 2014.
- [50] M. Endo. Calcium release from the sarcoplasmic reticulum. *Physiological Reviews*, 57(1):71–108, 1977.
- [51] B. Ermentrout and D. H. Terman. *Foundations of Mathematical Neuroscience*. Springer, Princeton, NJ, USA, 1st edition, 2010.
- [52] L. C. Evans. *An Introduction to Stochastic Differential Equations*, volume 82. American Mathematical Society, Providence, RI, USA, 1st edition, 2012.
- [53] O. Faugeras and J. Maclaurin. Asymptotic description of neural networks with correlated synaptic weights. *Entropy*, 17:4701–4743, 2015.
- [54] O. Faugeras, E. Soret, and E. Tanré. Asymptotic behaviour of a network of neurons with random linear interactions. *Preprint HAL Id: hal-01986927*, 2019.
- [55] O. Faugeras and J. Maclaurin. Asymptotic description of stochastic neural networks. i. existence of a large deviation principle. *Comptes Rendus. Mathématique*, 352(10):841–846, 2014.

- [56] O. Faugeras, E. Soret, and E. Tanré. Asymptotic behavior of a network of neurons with random linear interactions. *arXiv preprint arXiv:2006.03083*, 2020.
- [57] W. Feller. *An Introduction to Probability Theory and Its Applications, Volume 2*, volume 81. John Wiley & Sons, New York, NY, USA, 1st edition, 1991.
- [58] R. FitzHugh. Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, 1(6):445–466, 1961.
- [59] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. Emcee: the MCMC hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306, 2013.
- [60] W. J. Freeman III. Waves, pulses, and the theory of neural masses. *Progress in Theoretical Biology*, 2(1):1–10, 1972.
- [61] M. I. Freidlin, A. D. Wentzell, M. Freidlin, and A. Wentzell. *Random Perturbations*. Springer, Berlin, Germany, 1st edition, 1998.
- [62] V. N. Friedhoff, L. Ramlow, B. Lindner, and M. Falcke. Models of stochastic  $Ca^{2+}$  spiking: established approaches and inspirations from models of neuronal spikes. *European Physical Journal: Special Topics*, 230(16-17):2911–2928, Aug. 2021.
- [63] D. D. Friel.  $Ca^{2+}$  oscillations in sympathetic neurons: an experimental test of a theoretical model. *Biophysical Journal*, 68(5):1752–1766, 1995.
- [64] E. Gelenbe. Stability of the random neural network model. *Neural Computation*, 2(2):239–247, 1990.
- [65] T. Grafke and E. Vanden-Eijnden. Non-equilibrium transitions in multiscale systems with a bifurcating slow manifold. *Journal of Statistical Mechanics: Theory and Experiment*, 2017, 2017.
- [66] J. S. Griffith. A field theory of neural nets: i: derivation of field equations. *The Bulletin of Mathematical Biophysics*, 25:111–120, 1963.
- [67] J. S. Griffith. A field theory of neural nets: ii. properties of the field equations. *The Bulletin of Mathematical Biophysics*, 27:187–195, 1965.
- [68] J. S. Griffith. On the stability of brain-like structures. *Biophysical Journal*, 3(4):299–308, 1963.
- [69] M. Grunwald. Sanov results for glauher spin-glass dynamics. *Probability Theory and Related Fields*, 106:187–232, 1996.
- [70] A. Guionnet. Averaged and quenched propagation of chaos for spin glass dynamics. *Probability Theory and Related Fields*, 109:183–215, 1997.

- [71] M. Hasler, V. Belykh, and I. Belykh. Dynamics of stochastically blinking systems. part i: finite time properties. *SIAM Journal on Applied Dynamical Systems*, 12(2):1007–1030, 2013.
- [72] M. Helias and D. Dahmen. *Statistical Field Theory for Neural Networks*, volume 970. Springer, New York, NY, USA, 1st edition, 2020.
- [73] A. Hening, D. H. Nguyen, N. Nguyen, and H. Watts. Random switching in an ecosystem with two prey and one predator. *SIAM Journal on Mathematical Analysis*, 55(1):347–366, 2023.
- [74] S. Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*:31, 2009.
- [75] R. Hinch and S. J. Chapman. Exponentially slow transitions on a markov chain: the frequency of calcium sparks. *European Journal of Applied Mathematics*, 16(4):427–446, 2005.
- [76] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500, 1952.
- [77] K. L. Hyrc, Z. Rzeszotnik, B. R. Kennedy, and M. P. Goldberg. Determining calcium concentration in heterogeneous model systems using multiple indicators. *Cell Calcium*, 42(6):576–589, 2007.
- [78] J. R. Ipsen and A. D. Peterson. Consequences of dale’s law on the stability-complexity relationship of random neural networks. *Physical Review E*, 101(5):052412, 2020.
- [79] B. H. Jansen and V. G. Rit. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. *Biological Cybernetics*, 73(4):357–366, 1995.
- [80] S. N. Kalitzin, D. N. Velis, and F. H. L. da Silva. Stimulation-based anticipation and control of state transitions in the epileptic brain. *Epilepsy & Behavior*, 17(3):310–323, 2010.
- [81] E. R. Kandel, J. H. Schwartz, T. M. Jessell, S. Siegelbaum, A. J. Hudspeth, S. Mack, et al. *Principles of Neural Science*, volume 4. McGraw-Hill, New York, NY, USA, 1st edition, 2000.
- [82] I. Karatzas and S. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, New York, NY, USA, 2nd edition, 1991.
- [83] B. Katz. *The Release of Neural Transmitter Substances*, volume 10. Charles C Thomas Pub Limited, Springfield, IL, USA, 1st edition, 1969.

- [84] J. Keener and J. Sneyd. *Mathematical Physiology: II: Systems Physiology*, volume 8. Springer Science & Business Media, New York, NY, USA, 1st edition, 2008.
- [85] J. P. Keener. Stochastic calcium oscillations. *Mathematical Medicine and Biology: A Journal of the IMA*, 23(1):1–25, 2006.
- [86] J. P. Keener and J. M. Newby. Perturbation analysis of spontaneous action potential initiation by stochastic ion channels. *Physical Review E*, 84(1):011918, July 2011.
- [87] J. Keizer and G. D. Smith. Spark-to-wave transition: saltatory transmission of calcium waves in cardiac myocytes. *Biophysical Chemistry*, 72(1-2):87–100, 1998.
- [88] J. Keizer, G. D. Smith, S. Ponce-Dawson, and J. E. Pearson. Saltatory propagation of  $ca^{2+}$  waves by  $ca^{2+}$  sparks. *Biophysical Journal*, 75(2):595–600, 1998.
- [89] H. Kim, Y. Mori, and J. B. Plotkin. Optimality of intercellular signaling: direct transport versus diffusion. *Physical Review E*, Apr. 2022.
- [90] B. Kolb, I. Q. Whishaw, G. C. Teskey, I. Q. Whishaw, and G. C. Teskey. *An Introduction to Brain and Behavior*. Worth Publishers, New York, NY, USA, 1st edition, 2001.
- [91] P. G. Kostyuk. Calcium ions in nerve fiber processes. *Annual Review of Neuroscience*, 4(1):107–126, 1981.
- [92] C. Kuehn and M. G. Riedler. Large deviations for nonlocal stochastic neural fields. *The Journal of Mathematical Neuroscience*, 4:1–33, 2014.
- [93] T. G. Kurtz. Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356, 1971.
- [94] E. Lang and W. Stannat. Finite-size effects on traveling wave solutions to neural field equations. *The Journal of Mathematical Neuroscience*, 7:1–35, 2017.
- [95] G. F. Lawler. *Introduction to Stochastic Processes*. Chapman and Hall/CRC, Boca Raton, FL, USA, 1st edition, 2018.
- [96] S. D. Lawley, J. A. Best, and M. C. Reed. Neurotransmitter concentrations in the presence of neural switching in one dimension. *Discrete and Continuous Dynamical Systems - Series B*, 21:2255–2273, 2016.
- [97] P. Lennie. The cost of cortical computation. *Current Biology*, 13(6):493–497, 2003.
- [98] S. Lim and M. S. Goldman. Balanced cortical microcircuitry for maintaining information in working memory. *Nature Neuroscience*, 16(9):1306–1314, 2013.

- [99] W. Liu and F. Tang. An output feedback controller for store-operated calcium entry and extracellular calcium sensing in yeast cells. In *2010 Chinese Control and Decision Conference*, pages 3748–3753. IEEE, 2010.
- [100] W. Liu, F. Tang, and J. Chen. Designing dynamical output feedback controllers for store-operated  $ca^{2+}$  entry. *Mathematical Biosciences*, 228(1):110–118, 2010.
- [101] F. Lopes da Silva, A. Hoeks, H. Smits, and L. Zetterberg. Model of brain rhythmic activity: the alpha-rhythm of the thalamus. *Kybernetik*, 15:27–37, 1974.
- [102] J. Lytton, M. Westlin, S. E. Burk, G. E. Shull, and D. H. MacLennan. Functional comparisons between isoforms of the sarcoplasmic or endoplasmic reticulum family of calcium pumps. *Journal of Biological Chemistry*, 267(20):14483–14489, 1992.
- [103] J. Maclaurin. An emergent autonomous flow for mean-field spin glasses. *Probability Theory and Related Fields*, 2021.
- [104] J. MacLaurin and J. M. Newby. Extreme first passage times for populations of identical rare events. *arXiv preprint arXiv:2309.01827*, 2023.
- [105] J. MacLaurin, M. Silverstein, and P. Vilanova. Population level activity in large random neural networks. *arXiv preprint arXiv:2401.15272*, 2024.
- [106] A. C. Marreiros, J. Daunizeau, S. J. Kiebel, and K. J. Friston. Population dynamics: variance and the sigmoid activation function. *NeuroImage*, 42(1):147–157, 2008.
- [107] F. Mastrogiuseppe and S. Ostojic. Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *PLoS Computational Biology*, 13(4):e1005498, 2017.
- [108] R. May. Will a large complex system be stable? *Nature*, 238:413–414, 1972.
- [109] M. Mézard, G. Parisi, N. Sourlas, G. Toulouse, and M. Virasoro. Replica symmetry breaking and the nature of the spin glass phase. *Journal de Physique*, 45(5):843–854, 1984.
- [110] C. Morris and H. Lecar. Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal*, 35(1):193–213, 1981.
- [111] O. Moynot and M. Samuelides. Large deviations and mean-field theory for asymmetric random recurrent neural networks. *Probability Theory and Related Fields*, 123:41–75, 2002.
- [112] W. Nayler, I. McInnes, D. Chipperfield, V. Carson, and J. Kurtz. Ventricular function and the calcium-accumulating activity of the sarcoplasmic reticulum. *Journal of Molecular and Cellular Cardiology*, 1(3):307–324, 1970.

- [113] P. L. Nunez. The brain wave equation: a model for the eeg. *Mathematical Biosciences*, 21(3-4):279–297, 1974.
- [114] B. Oksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer Science & Business Media, Berlin, Germany, 1st edition, 2013.
- [115] S. Ostojic. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature Neuroscience*, 17(4):594–600, 2014.
- [116] G. Parisi and G. Toulouse. A simple hypothesis for the spin glass phase of the infinite-ranged SK model. *Journal de Physique Lettres*, 41(15):361–364, 1980.
- [117] D. Purves. *Brains as Engines of Association: An Operating Principle for Nervous Systems*. Oxford University Press, Oxford, UK, 1st edition, 2019.
- [118] L. Ramlow, M. Falcke, and B. Lindner. An integrate-and-fire approach to ca2+ signaling. part i: renewal model. *Biophysical Journal*, 122:713–736, Feb. 2023.
- [119] L. Ramlow, M. Falcke, and B. Lindner. An integrate-and-fire approach to ca2+ signaling. part ii: cumulative refractoriness. *Biophysical Journal*, 122:4710–4729, Dec. 2023.
- [120] F. Rassoul-Agha and T. Seppäläinen. *A Course on Large Deviations with an Introduction to Gibbs Measures*, volume 162. American Mathematical Society, Providence, RI, USA, 1st edition, 2015.
- [121] M. G. Riedler, M. Thiullen, and G. Wainrib. Limit theorems for infinite-dimensional piecewise deterministic markov processes. applications to stochastic excitable membrane models. *Electronic Journal of Probability*, 17, 2012.
- [122] P. A. Robinson, C. J. Rennie, and J. J. Wright. Propagation and stability of waves of electrical activity in the cerebral cortex. *Physical Review E*, 56(1):826, 1997.
- [123] E. T. Rolls and G. Deco. *The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function*, volume 34. Oxford University Press, Oxford, UK, 1st edition, 2010.
- [124] R. Rosenbaum and B. Doiron. Balanced networks of spiking neurons with spatially dependent recurrent connections. *Physical Review X*, 4(2):021039, 2014.
- [125] S. Rüdiger. Stochastic models of intracellular calcium signals. *Physics Reports*, 534:39–87, 2014.
- [126] R. Rudnicki and M. Tyran-Kaminska. *Piecewise Deterministic Processes in Biological Models*. Springer Briefs in Applied Sciences and Technology, Cham, Switzerland, 1st edition, 2017.

- [127] T. C. Südhof. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron*, 80(3):675–690, 2013.
- [128] H. Schmidt, W. Woldman, M. Goodfellow, F. A. Chowdhury, M. Koutroumanidis, S. Jewell, M. P. Richardson, and J. R. Terry. A computational biomarker of idiopathic generalized epilepsy from resting state eeg. *Epilepsia*, 57(10):e200–e204, 2016.
- [129] M. N. Shadlen and W. T. Newsome. Noise, neural codes and cortical organization. *Current Opinion in Neurobiology*, 4(4):569–579, 1994.
- [130] I. Siekmann, L. E. Wagner, D. Yule, E. J. Crampin, and J. Sneyd. A kinetic model for type i and ii ip3r accounting for mode changes. *Biophysical Journal*, 103:658–668, 2012.
- [131] J. Sneyd, K. Tsaneva-Atanasova, D. Yule, J. Thompson, and T. Shuttleworth. Control of calcium oscillations by membrane fluxes. *Proceedings of the National Academy of Sciences*, 101(5):1392–1396, 2004.
- [132] H. Sompolinsky, A. Crisanti, and H.-J. Sommers. Chaos in random neural networks. *Physical Review Letters*, 61(3):259, 1988.
- [133] M. Stern, H. Sompolinsky, and L. F. Abbott. Dynamics of random neural networks with bistable units. *Physical Review E*, 90(6):062710, 2014.
- [134] H. Touchette. A basic introduction to large deviations: theory, applications, simulations. *arXiv preprint arXiv:1106.4146*, 2011.
- [135] C. Van Vreeswijk and H. Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.
- [136] P. Vanetti, A. Bouchard-Côté, G. Deligiannidis, and A. Doucet. Piecewise-deterministic markov chain monte carlo. *arXiv preprint arXiv:1707.05296*, 2017.
- [137] J. Vannimenus, G. Toulouse, and G. Parisi. Study of a simple hypothesis for the mean-field theory of spin-glasses. *Journal de Physique*, 42(4):565–571, 1981.
- [138] C. van Vreeswijk and H. Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural Computation*, 10(6):1321–1371, 1998.
- [139] C. Villarruel, P. S. Aguilar, and S. P. Dawson. High rates of calcium-free diffusion in the cytosol of living cells. *Biophysical Journal*, 120(18):3960–3972, 2021.
- [140] A. E. Waugh. *Elements of Statistical Method*. McGraw-Hill Book Co., New York, NY, USA, 2nd edition, 1943.

- [141] A. Weiss and A. Shwartz. *Large Deviations For Performance Analysis: Queues, Communication and Computing*. Routledge, Milton Park, Abingdon, Oxfordshire, UK, 1st edition, 2019.
- [142] H. R. Wilson and J. D. Cowan. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, 13(2):55–80, 1973.
- [143] H. R. Wilson and J. D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12(1):1–24, 1972.
- [144] L. Yu and T. O. Tallefumier. Metastable spiking networks in the replica-mean-field limit. *PLoS Computational Biology*, 18(6):e1010215, 2022.
- [145] C. Zimmer. 100 trillion connections. *Scientific American*, 304(1):58–61, 2011.