12-2023

Aspects of Stochastic Geometric Mechanics in Molecular Biophysics

David Frost
dlfrost@g.clemson.edu

Aspects of Stochastic Geometric Mechanics in Molecular Biophysics

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematics

---

by
David Frost
December 2023

Accepted by:
Dr. Peter Kiessler, Committee Chair
Dr. Keisha Cook
Dr. Hugo Sanabria
Dr. Brian Fralix
Dr. Martin Schmoll

Abstract

In confocal single-molecule FRET experiments, the joint distribution of FRET efficiency and donor lifetime distribution can reveal underlying molecular conformational dynamics via deviation from their theoretical Forster relationship. This shift is referred to as a dynamic shift. In this study, we investigate the influence of the free energy landscape in protein conformational dynamics on the dynamic shift by simulation of the associated continuum reaction coordinate Langevin dynamics, yielding a deeper understanding of the dynamic and structural information in the joint FRET efficiency and donor lifetime distribution. We develop novel Langevin models for the dye linker dynamics, including rotational dynamics, based on first physics principles and proper dye linker chemistry to match accessible volumes predicted by molecular dynamics simulations. By simulating the dyes' stochastic translational and rotational dynamics, we show that the observed dynamic shift can largely be attributed to the mutual orientational dynamics of the electric dipole moments associated with the dyes and not their accessible volumes.

Furthermore, using nonlinear semi-group convergence methods based on viscosity solutions to associated Hamilton-Jacobi equations developed by Feng and Kurtz and methods of verifying the comparison principle for viscosity solutions introduced by Versendaal et al., we prove a large deviations principle for dynamical systems on Riemannian manifolds perturbed by vanishing Markov noise. Further, using the correspondence between the aforementioned nonlinear semigroup and stochastic control theory, we find explicit representations of the rate function in several cases via a Legendre - Fenchel transform of a corresponding Hamiltonian functional. This provides a generalization of classical Friedlin-Wentzell theory to the case of degenerate general Markov perturbations on complete Riemannian manifolds.

Dedication

It is hard to write a dedication for such a monumental moment. Everyone in my life has played a role in helping me get here. I would like to thank several individuals specifically, however.

- Dani - Thank you for being my companion and helping me through difficult times. I will never forget you.
- Dr. Kiessler - Thank you for letting me explore and find my own topic. You have been so patient with my desire to study everything. I am very grateful to your help and support.
- Dr. Cook - Thank you for taking me on as a student and being so supportive and helpful.
- Dr. Sanabria - Thank you for showing me the world of Biophysics and helping me realize I can help.
- Anay - Thank you for helping me see the beauty around me again.
- Sara - I would not be here without you. Thank you for being who I needed.
- Ben - Your support all my life has helped get me here. You are an amazing brother. Thank you.
- The DnD Group - You all have been my friends before I could drive. Thank you for being such a good distraction throughout my PhD.
- To everyone else, thank you so much. I do not think there is a complete list of people who helped me, and I could not write out all the ways you helped.

Thank you to everyone. This is a dream come true. A goal finally met after struggling to get here my whole life. Through countless setbacks and discouragement, we made it and that is incredible.

Preface

It is my sincere belief that great insight is derived from diversity. Without diversity, knowledge becomes stale and rigid. Unfortunately, it seems that for some reason, the prominent belief is that a work containing multiple subjects is somehow *less* than the sum of its parts. I believe this reasoning is deeply flawed, and that especially now we need to be thinking less in rigid structures and use tools and ideas from many sources. It is due to this belief that I write the following dissertation and it is my hope that this document demonstrates not only my flexibility as a mathematician and scientist, but also the benefit of interdisciplinary thinking.

In this work I have attempted to bring together many subjects and it does indeed cover a wide breadth of material. In this document you will find mentions of several areas of science and mathematics. In the first project, biology, chemistry, and physics are studied using computational and statistical methods. The second contains geometry, analysis, algebra, and optimization. Both projects are connected with probability theory. In this way I hope to convey that I understand and can use each of the core areas of mathematics as presented by the department of mathematics at Clemson, and also that I am capable of learning and understanding many areas of science while applying these areas of mathematics to them.

One of my favorite mathematics textbooks is "Nonlinear Dynamics and Chaos" by Steven Strogatz. Not only is the subject interesting, but in the book Dr. Strogatz demonstrates the utility of the mathematics presented by applying it to several areas from population dynamics, chemical reactions and physics. This book exemplifies a fundamental idea - mathematics is useful. Moreover, the same mathematics is helpful in many different contexts. In the following work we will see that probability theory can be used for several things, from modeling the motion of molecules or fluorescence processes to being used to understand asymptotic behavior of measures on abstract path spaces.

Like Sgt. Colon above, I hope this work constitutes "A Contribution". Whether or not it does is not my decision. However, I stand by my decision to study these topics. In doing so I have broadened my understanding of both mathematics and the universe around me and for that, I will always be grateful.

Acknowledgments

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue duis dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea

commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis.

At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, At accusam aliquyam diam diam dolore dolores duo eirmod eos erat, et nonumy sed tempor et et invidunt justo labore Stet clita ea et gubergren, kasd magna no rebum. sanctus sea sed takimata ut vero voluptua. est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the following text, you will find a variety of different topics. The nature of this work is highly interdisciplinary. Throughout the career, mathematical objects are paired with an application from biophysics. The two main contributions of this work are as follows.

First, mathematical modeling of smFRET experiments. This contribution can be broken into two sub-parts. One is exploring the influence of the energy landscape geometry on the FRET vs lifetime distribution. This provided a fruitful discovery area, requiring SDE models to incorporate the theory of metastability to model protein models and the simulation of a random string polymer. In this study, features of the energy landscape can be readily seen in the marginal distributions of the FRET vs lifetime distribution. This provides numerical justification for further energy landscape analysis using this method. This is extremely important in molecular biology as an understanding of the energy landscape, and hence, proteins' structural and dynamic behavior may be understood. Moreover, a model for the fluorescence dynamics of the dyes used in smFRET is introduced in exploring smFRET. This model incorporates the rotational dynamics of the dipole moments and the dyes' chemical and physical properties. In these models, stochastic geometric mechanics is used to derive simple models for the dyes' rotational dynamics and translational motion.

The second contribution is a theorem about large deviations for dynamical systems perturbed by vanishing Markov noise on Riemannian manifolds. This work is motivated by understanding the fluctuations in the equilibrium that the dipole rotations may undertake. This requires understanding Friedlin-Wentzell's theory on Riemannian manifolds, which has yet to be developed. To keep generality and incorporate any anomalous diffusions or jumps caused by intrinsic anisotropy,

we prove the result in much greater generality than needed. Indeed, the result is stated for a general Markov process on a Riemannian manifold.

In each chapter new information is developed and discussed. New results are given in each chapter, typically in examples about the topic at hand. Rather than reviewing each new concept and then returning to the models, the idea is introduced and then the model is presented in parallel. Further, large deviation results are mentioned in many sections. This is to keep the reader aware of the connection to large deviation theory and hopefully motivate the studies in the final chapters to be more engaging and naturally occurring.

Chapter 2

Markov Processes

> "Chaos is found in greatest abundance wherever order is sought. It always defeats order because it is better organized."
>
> Terry Pratchett, Interesting Times: The Play

The main probabilistic object of interest in this work is the stochastic process, particularly the Markov process.

Definition 2.0.1. *Let T be a set and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and S be a set. A* ***stochastic process*** *is a function*

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

such that for $t \in T$, $X(t, \cdot)$ is a random variable and $\omega \in \Omega$, $X(\cdot, \omega)$ is a function of t.

The set T is usually descriptive of time but can take the form of a more general topological space; in that case, the stochastic process is termed a random field. In the following, we are only concerned with processes such that $T \subseteq \mathbb{N}$, or $T \subseteq \mathbb{R}_0^+ = [0, \infty)$. The set S is called the state space from which one observes the outcomes of the process.

The Markov process is a stochastic process central to the current endeavor. Markov processes are a class of stochastic processes with a restriction on the behavior of the process under conditional expectation concerning the history generated by the process; heuristically, a Markov process is one for which the future is independent of the past given the present. This means that the next step in the process only depends on the current location rather than the path it took to get there. The properties of Markov processes are numerous; hence, they are frequently used as modeling tools in

many disciplines such as physics, chemistry, biology, and finance. [43] [65]

Let $\sigma(X_s : s \leq t)$ be the sigma-algebra generated by the random variables $\{X_s\}_{s \leq t}$; this is frequently called the natural filtration associated with the process X, and can be thought of as the entire history of the process up to time t. Let $\sigma(X_t)$ be the sigma-algebra generated by the random variable X_t.

Definition 2.0.2. *A stochastic process is called a* ***Markov process*** *if for all $u > t$ the following holds [6] [49]*

$$\mathbb{E}[X_u|\sigma(X_s : s \leq t)] = \mathbb{E}[X_u|\sigma(X_t)]$$

this property is called the ***Markov property***[1].

When T and S are countable, the Markov process is called a **discrete time Markov chain (DTMC)**. When T is uncountable and S is countable, we call the Markov process a **continuous time Markov chain (CTMC)**.

Example 2.0.1. *[The Förster Resonance Energy Transfer (FRET) CTMC]*

Förseter Resonsance Energy Transfer (FRET) is a physical process by which energy is transferred between dipolar molecules nonradiatively. [13]

The FRET energy transfer rate, the rate at which energy is transferred nonradiatively, can be expressed as

$$k_{ET}(r) = k_D\left(\frac{R_0}{r}\right)^6.$$

Where $r > 0$ represents the distance between the electric dipoles, in this example it will be taken as static. The parameter R_0 is called the ***Förster radius*** *and will be taken as a collection of environmental terms for right now. [42]*

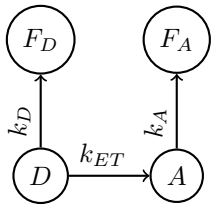
Define a state space $S = \{D, A, F_D, F_A\}$, such that the state of the CTMC keeps track of where the excitation energy is:

- *State D represents an excited donor*
- *State A represents an excited acceptor*
- *State F_D represents a fluoresced donor photon.*

[1] A useful way to think of conditional expectation $\mathbb{E}[X_t|\mathcal{F}_s]$ is as the projection of a measurable function, X_t in the space $L^2(\Omega, \mathcal{F}, \mathbb{P})$ to the subspace $L^2(\Omega, \mathcal{F}_s, \mathbb{P})$,where $\mathcal{F}_s$ is a sub-sigma algebra of $\mathcal{F}$. [9] This is not always true, but it is when X is square integrable.

- *State F_A represents a fluoresced acceptor photon.*

The transition rates k_D, k_A represent the fluorescence decay times for the donor and acceptor respectively. Now consider a process that transitions between these states according to the following directed graph:



Using the Markov property, one can observe that Markov processes can be described based on their transition probabilities. Moreover, one could view a Markov process as a set of measures indexed by time and space, which dictates the probability of transferring from one state to another. By investigating this idea further, one may find the Markov semigroup, a semigroup of operators acting on a space of functions.

Definition 2.0.3. *Let $\{X_t\}$ be a Markov process. Define the* ***transition or Markov semigroup****,*

$$P_t f(x) = \mathbb{E}[f(X_t)|X_0 = x]$$

for integrable f. [30] [18]

Markov semigroups completely characterize the process and are fundamental in the theory of Markov processes. The intuitive explanation of the semigroup is that it tells you the transition probabilities. Indeed, consider the case when $f(x) = 1_A(x)$, the indicator of a measurable set A, then $P_t f(x) = \mathbb{P}(X_t \in A|X_0 = x)$. Intuitively, the connection between Markov processes and semigroups provides a lot of information on the nature of Markov processes. Markov processes are those processes whose transition probabilities are only dependent on the location of the process in space-time and are independent of anything else. Understanding how one transitions from one state to another based on the location of the process is enough to completely characterize the process.

Following standard analytic semigroup theory, one can consider an operator associated to the semigroup and connect operator theory to the theory of Markov processes. In this way, one may use tools from more familiar differential equations to solve problems, such as deriving transition probabilities or hitting times, related to the Markov process at hand.

Definition 2.0.4. *Define the linear operator $\mathcal{A}$ by the following*

$$\mathcal{A}f := \lim_{t\to 0^+} \frac{P_t f - f}{t}.$$

*The domain of the operator $\mathcal{A}$ is $D(\mathcal{A}) = \{f : \lim_{t\to 0^+} \frac{P_t f - f}{t}$ exists$\}$. This linear operator is called the **infinitesimal generator** of the process $\{X_t\}$. [4] [30] [9]*

Usually, the domain of the generator is smaller than the set of functions that P_t acts on and may not even be a space of functions. However, we do have that the domain of the generator is densely contained in the domain of the semigroup. [18] One calls this the generator because $P_t f(x)$ can be viewed as the solution to the Kolmogorov backward equation, given by

$$\frac{\partial}{\partial t}u(x,t) = \mathcal{A}u(x,t), \quad u(x,0) = f(x)$$

[18] [21] [3]. In this way, one can formally view $P_t = e^{t\mathcal{A}}$. Notice that if one can find a $u(x,t)$ such that $P_t(u(x,0)) = u(x,t) = u(x,0)$, one will have isolated a stationary distribution for the process, and hence the equilibrium distribution. This is important in statistical mechanics, as when a system is in thermal equilibrium, it should be distributed according to the stationary distribution of the stochastic process describing the system.

Example 2.0.2. *[FRET CTMC Continued] Consider the FRET CTMC in Example 3.0.1. Using definition 3.0.4. one can find that the generator for a CTMC is a matrix containing the transition rates of each state. Using this one can then completely describe the FRET process with the rate matrix*

$$\begin{pmatrix} -(k_D + k_{ET}) & k_{ET} & k_D & 0 \\ 0 & -k_A & 0 & k_A \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now, if one wishes to determine the stationary distribution for this process the Kolmogorov equations may be solved. However, the resulting system of equations is underdetermined and hence cannot be solved. This is because the CTMC is absorbing, so it will not have a stationary distribution. Instead one can find the probability of absorption into a particular state.

Consider the probability that the absorbing state is F_D. Note that this can only happen if

energy transfer occurs, i.e., if, in the CTMC, we have a $D \to A$ transition. This means that in a race of exponential random variables, the $D \to A$ transition wins. Therefore, denoting τ_D and τ_{ET} as exponential random variables with rates k_D and k_{ET} respectively

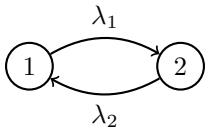
$$\mathbb{P}(\min(\tau_D, \tau_{ET}) = \tau_{ET}) = \frac{k_{ET}}{k_D + k_{ET}} = \frac{k_D\left(\frac{R_0}{r}\right)^6}{k_D + k_D\left(\frac{R_0}{r}\right)^6} = \frac{1}{\left(\frac{r}{R_0}\right)^6 + 1}.$$

*The probability of being absorbed into state F_D is the **FRET efficiency**. It is the main source of inquiry in FRET experiments.*

As seen in the previous example, when the state space is finite, the description of a Markov process becomes closely related to a linear algebra, wherein the infinitesimal generator of the process becomes the rate matrix for the CTMC. Then, the Kolmogorov equations can be stated as a system of linear ordinary differential equations, and many properties of the rate matrix can be exploited to further understand the process. This makes CTMCs an especially valuable modeling tool, and their simplicity can provide a way of clearing away conceptual clutter. Now consider a very simple case of a CTMC, that will yield insights later on.

2.1 Insights From the Two-State CTMC

Consider a system with two states, state 1 and state 2. If the system is in condition 1 it transitions to state 2 with rate λ_1. Similarly, if in state 2, it transitions to state 1 with rate λ_2. The rate diagram for such a system is shown below.



The rate matrix for such a process is

$$\begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}.$$

Knowing that the stationary distribution for such a Markov process can be found by solving the

stationary Kolmogorov equation, $\pi Q = 0$. This yields the system of equations:

$$\begin{pmatrix} \pi_1 & \pi_2 \end{pmatrix} \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix} = 0 \rightarrow \begin{cases} -\lambda_1\pi_1 + \lambda_2\pi_2 = 0 \\ \lambda_1\pi_1 - \lambda_2\pi_2 = 0. \end{cases}$$

The solution to this system is given by

$$\pi_1 = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad \pi_2 = \frac{\lambda_1}{\lambda_1 + \lambda_2}.$$

From here, we may take two directions of thought. One is that in equilibrium, a system randomly transitioning between two states will have a stationary distribution that we may use to determine the transition rates. The second is that we may find the equilibrium distribution by taking an ensemble average.

2.1.1 Protein Conformational Dynamics

The key application of the two-state CTMC that will be investigated in this section is its application to protein conformational dynamics. To understand why and how CTMCs are used in protein conformational dynamics, we will first need to investigate the nature of a protein. Proteins are one of the central classes of molecules investigated by molecular biology. They are sequences of amino acids. The primary structure of a protein is the sequence of amino acids. Depending on the sequence of amino acids, a protein will develop a secondary structure, a two-dimensional picture of the protein that categorizes loops and other topological properties. The full three-dimensional shape of the protein is called the tertiary structure. The structure of this protein dictates its function in the body, and understanding the shape of proteins is essential to understanding how they work in the body. However, due to thermal fluctuations and environmental effects, the protein will change shape randomly. The act of changing shape is termed changing conformation. The nature of these conformational changes is the subject of protein conformational dynamics. In the context of the two-state CTMC, one imagines a protein changing conformation between two stable states.

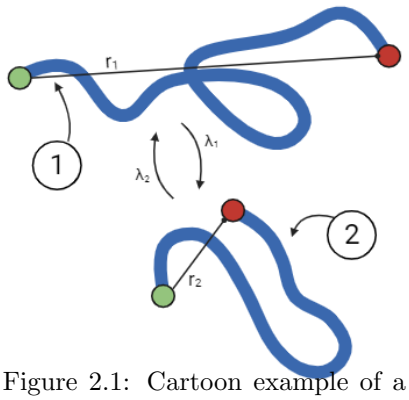


Figure 2.1: Cartoon example of a protein changing between two conformations according to the two-state CTMC

With this model under consideration, it becomes apparent that the most valuable information in studying the conformation change would be the transition rates of this CTMC and the equilibrium distribution and fluctuations about that distribution. The question now is how to estimate these rates and the distribution. To help answer the question experimentally, one can use the fluorescence spectroscopy technique known as Förster Resonance Energy Transfer, a phenomenon investigated in examples 2.0.1 and **??**.

2.1.1.1 The Idea of a smFRET Experiment

As mentioned in example 2.0.1, most FRET experiments aim to approximate the FRET efficiency, $\mathcal{E}$; a distance estimation may be made once this is estimated. [34] Experiments are delineated first by their sample preparation and composition. In smFRET, the sample, the molecules of interest, is diluted in a solution so that, on average, there is less than one molecule of interest in a volume the size of the confocal volume of the microscope used in the experiment. [31] In this way, one can see the absorbing state of individual FRET processes rather than a mixture of processes. When using smFRET, the FRET efficiency estimation can be done in two ways. The first is the experimental method of **intensity-based FRET**. [42] [32] This method repeatedly excites the donor and measures whether an acceptor or donor photon is observed. Effectively, this treats the experiment as calculating the probability of success of a binomial random variable with $p = \mathcal{E}$. It is well known that the estimator for this is given by the number of successes observed divided by the total number of trials. In this experiment, this becomes

$$\mathcal{E}_I = \frac{I_A}{I_A + I_D},$$

where I_A, I_D are the intensities (photon counts) for the acceptor and donor, respectively.

The second method, known as **lifetime based FRET**, comes from the following.

$$\begin{aligned}\mathbb{P}(\min(\tau_D, \tau_{ET}) = \tau_{ET}) + \mathbb{P}(\min(\tau_D, \tau_{ET}) = \tau_D) &= 1 \\ \rightarrow \mathcal{E} + \frac{k_D}{k_D + k_{ET}} &= 1 \\ \rightarrow \mathcal{E} &= 1 - \frac{k_D}{k_D + k_{ET}}.\end{aligned}$$

Noting that $\mathbb{P}(\tau_D > t | \min(\tau_D, \tau_{ET}) = \tau_D) \sim \exp(k_D + k_{ET})$ [51] we write the following

$$\mathcal{E} = 1 - \frac{\tau_D'}{\tau_D},$$

where $\tau_D' = (k_D + k_{ET})^{-1}$ is the lifetime of the donor in the presence of the acceptor and $\tau_D = k_D^{-1}$ is the lifetime in absence of acceptor. In this way, one may measure the mean lifetime of the donor and estimate the FRET efficiency by approximating the mean of an exponential random variable.

Out of the two, intensity-based is more prevalent due to cost constraints. [32] One only needs to measure photon counts and not their arrival times to conduct an intensity-based FRET experiment. Moreover, only FRET in a static sense has been considered so far. However, the energy transfer rate is dynamic.

2.1.2 The smFRET Dynamic Shift

In the previous section, the measurements presented assumed the underlying protein was static. However, we embarked on an effort to understand the conformational dynamics of a protein that transitions between two stable conformations. Consider a smFRET experiment in which the underlying protein changes conformation according to the two-state CTMC. Assume that during each fluorescence process, the protein does not change conformation. Then, the outcome of each fluorescence process is conditional on the state of the CTMC at the time of excitation. Now, let $k_D(r_i) = k_i, i = 1.2$ be the fluorescence rate of the donor when the protein is in conformation i during the fluorescence process. If the protein is in equilibrium, we can consider the probability of the protein being in conformation i at any given time is π_i according to the stationary distribution. [31] In this way, the distribution of the observed lifetime is a compound distribution of exponential

distributions combined using the stationary distribution of the two-state CTMC,

$$f(t) = (k_1 + k_2)(\pi_1 e^{-k_1 t} + \pi_2 e^{-k_2 t}).$$

Using this distribution one can see that if $\tau \sim f$ as above then

$$\begin{aligned}\mathbb{E}[\tau] &= \pi_1 \mathbb{E}[\tau'_D(r)] + \pi_2 \mathbb{E}[\tau'_D(r')] \\ &= \frac{\pi_1}{k_1} + \frac{\pi_2}{k_2}.\end{aligned}$$

In comparison, one may consider the lifetime distribution for the mean of the rates for when $\tau \sim \exp(\pi_1 k_1 + \pi_2 k_2)$. Noting that $\mathbb{E}[\tau] = \frac{1}{k}$ for exponential τ is a convex function in k it can said that the expected lifetime from a static state with fluorescence rate given by $\pi_1 k_1 + \pi_2 k_2$ will be less than the expected value from a dynamic state switching between rates k_1 and k_2. More specifically, if $\tau' \sim \exp(\pi_1 k_1 + \pi_2 k_2$ and $\tau \sim f$ as above, then we have the following

$$\mathbb{E}[\tau'] = \frac{1}{\pi_1 k_1 + \pi_2 k_2} \leq \frac{\pi_1}{k_1} + \frac{\pi_2}{k_2} = \mathbb{E}[\tau].$$

Consequently, the expected lifetime of the mixture of distance states is greater than that of a state, which is the average between the distances. Due to this property, one can detect the presence of dynamic switching by looking at FRET vs lifetime plots. Recall that the FRET efficiency is related to the lifetime distribution by

$$\mathcal{E}(\tau'_D) = 1 - \frac{\tau'_D}{\tau_D}$$

called the **static FRET line** [2] where $\tau'_D = k'^{-1}_D$ is the estimated lifetime of the donor in the presence of an acceptor. So, data from a static process will be centered on the static FRET line. If the data arises from a dynamic process, it will be shifted above the static FRET line. This shift is called the **dynamic shift**. [2, 47, 31] One can define a numerical value for the dynamic shift by considering the following. Let $\delta = (\mathcal{E}', \tau')$ be a point in the FRET vs Lifetime plane and consider $\mathcal{E}(\tau)$ as the static FRET

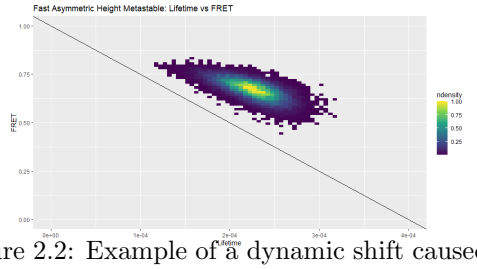


Figure 2.2: Example of a dynamic shift caused by changing conformations according to the two-state CTMC

line running through the plane. [2] The dynamic shift can be considered the magnitude of the orthogonal projection of δ to $\mathcal{E}(\tau)$. In other words, it is the minimum distance between the point $(\mathcal{E}', \tau')$ and the static line.

The reasoning behind the dynamic shift extends beyond a discrete mixture of states, as shown. Indeed, let $k(r)$ be the energy transfer rate for a smFRET experiment with distribution $G(r)$. Then, by Jensen's inequality it is apparent that

$$\mathbb{E}[\tau'] = \frac{1}{\int k(r) dG(r)} \leq \int \frac{1}{k(r)} dG(r) = \mathbb{E}[\tau]$$

as above. Therefore, the dynamic shift can capture information from various sources, not just discrete transitions. In many contexts, one considers a discrete mixture of states and uses photon counts and hidden Markov models to determine the amount of time in each state and the number of states. [53] [54] [52] [10] A contribution of the work in future sections is the consideration of a continuum of states determined by the first principle physics using Langevin equations describing the conformation process, and other relevant processes. This way, one can examine the influence of the continuum dynamics on the dynamic shift and explore how slight differences in the structure of the continuous motion may impact the dynamic shift and FRET vs lifetime distribution. The topic of understanding continuous potentials using smFRET has recently become of interest. [10] Now, let the equilibrium distribution for the distance between the dyes be given by $G(r)$; then the lifetime distribution is

$$f(t) = \int_{\mathbb{R}^+} G(r) e^{-t k_{ET}(r)} dr.$$

However, the processes governing the dye motion and conformational process are multidimensional and, in some cases, may not be analytically tractable. For this reason, simulations are conducted to create synthetic data from the dynamics. Therefore, one can relate the underlying dynamics to the observed dynamic shift in FRET vs Lifetime plots. The most important aspect of this line of thought is that the lifetime distribution in smFRET experiments is susceptible to the structure of the stationary distribution of the inter-dye displacement.

2.1.3 Entropy and Large Deviations

One of the questions regarding the two-state CTMC model for protein conformation dynamics was the equilibrium distribution and how much the system fluctuates about this distribution. Since the process constantly evolves stochastically, the distribution for an ensemble of proteins will fluctuate about the equilibrium distribution. The nature of these fluctuations may be of concern in a biological application. In physics, one would typically investigate the concept of entropy; we will explore the equivalent notion [60] [58] of large deviations.

The central limit and the laws of large numbers are fundamental results in probability theory. Large deviations complement these results by giving insight into their rates of convergence. Heuristically, a sequence of probability measures satisfies a large deviation principle (LDP) with rate function I if,

$$\mathbb{P}_n(X \approx x) \approx e^{-nI(x)}$$

that is to say, these probabilities will decay exponentially with a rate I. This rate function is also closely related to the entropy of a process. [58] [19] It can be thought of as a negative entropy in that the most likely state is the one that minimized the rate function.

Let the state of the ith protein's conformation CTMC be denoted Y_i where $i = 1, ..., N$. In this situation, we wish to investigate the behavior of a sum of random variables,

$$S_N = \frac{1}{N} \sum_{i=1}^{N} X_i$$

where X_i is given by

$$X_i = \begin{cases} 1, & Y_i = 1 \\ 0, & Y_i = 0. \end{cases}$$

Hence, we are investigating the proportion of proteins in state 1 at any time in equilibrium. It is known by the law of large numbers that $S_N \to \pi_1$ is $N \to \infty$. However, for finite N, fluctuations around the equilibrium will represent large deviations in the collection of random variables X_i. Heuristically, a sequence of probability measures satisfies a large deviation principle (LDP) with rate function I if,

$$\mathbb{P}_n(X \approx x) \approx e^{-nI(x)}$$

that is to say, these probabilities will decay exponentially with a rate I. The rate function is also closely related to the entropy of a process [58] [19] [60] and can be thought of loosely as a negative entropy in that the most likely state is the one that minimizes the rate function.

To make this definition formal, consider the following. Let (E, d) be a complete separable metric space. Define a **rate function** $I : E \to [0, \infty]$ to be a lower semicontinuous function and a **good rate function** to be a rate function such that the level sets $\{x \in E | I(x) \leq \alpha\}$ are compact for each $\alpha \in \mathbb{R}$. [23]

Note, a lower semicontinuous function is a function $I : E \to [0, \infty]$ such that for all $x \in E$, we have

$$\liminf_{x_n \to x} I(x_n) \geq I(x).$$

Similarly, an upper semicontinuous function is one such that

$$\limsup_{x_n \to x} I(x_n) \leq I(x).$$

As we have seen, rate functions are always lower semicontinuous. Upper semicontinuity will play a role when we discuss viscosity solutions to the Hamilton - Jacobi equation.

Let $\{X_n\}$ be a sequence of random variables taking values in E. We say $\{X_n\}$ satisfies a **large deviations principle (LDP)** [23] with rate function $I : E \to [0, \infty]$ if the following two conditions hold:

$$\begin{cases} -\inf_{x \in \mathcal{A}} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \ln(\mathbb{P}\{X_n \in \mathcal{A}\}), \text{ for open } \mathcal{A} \subset E \\ \\ \limsup_{n \to \infty} \frac{1}{n} \ln(\mathbb{P}\{X_n \in B\}) \leq -\inf_{x \in B} I(x), \text{ for closed } B \subset E \end{cases} \tag{2.1}$$

In 1938, Harold Cramér proved the following result on these types of central limit sums,

Theorem 2.1.1 (Cramér, 1938)**.** *Let $S_N = \frac{1}{N} \sum_{j=1}^{N} X_j$ with X_j i.i.d. and define*

$$V(\theta) = \ln \mathbb{E}[e^{\theta X}].$$

Then, S_N satisfies an LDP with rate function

$$I(x) = \sup_{\theta}\{\theta x - V(\theta)\}.$$

This theorem relates the fluctuations of a system switching states in equilibrium to the Legendre-Fenchel transform of the log moment generating function of the stationary distribution. This fundamental idea will be explored more fully in chapter 9.

In the example of the two-state CTMC, we can see that the rate function will be given by

$$\begin{aligned} I(x) &= \sup_{\theta}\{\theta x - \ln(\pi_1 e^{\theta} + \pi_2)\} \\ &= x \ln\left(\frac{\pi_2}{\pi_1}\right) - \ln(\pi_1) + x \ln(x) + (1-x)\ln(1-x) \end{aligned}$$

Therefore, we may use $I(x)$ above to determine the most likely ensemble state we will see. Moreover, it provides a way to know the distribution around the equilibrium state. This rate function is the entropy for the system rotated about the maximum value. It describes how energy is spread out in the design and how fluctuations in the system behave. In this way, one can estimate the chances of seeing infrequent events. This becomes particularly important when considering the continuous case wherein transitions between stable points may be viewed from an entropic or large deviation perspective.

Chapter 3

Brownian motion and Stochastic Analysis

> "Individuals aren't naturally paid-up members of the human race, except biologically. They need to be bounced around by the Brownian motion of society, which is a mechanism by which human beings constantly remind one another that they are...well...human beings."
>
> Terry Pratchett, Men at Arms

3.1 Brownian Motion

The first accurate account of Brownian motion was by botanist Robert Brown in 1827 while viewing a spec of pollen drifting on a water surface. He is the namesake for Brownian motion. Louis Bachelier, in 1900, described Brownian motion in his Ph.D. thesis "On the Theory of Speculation" to analyze the behavior of stocks. Albert Einstein won the Nobel Prize for his description of Brownian motion in one of his famous 1905 papers. He did this by characterizing the motion by convergence of a jump process with normally distributed jumps. A similar work was done by Marian Schmoluchowski in 1906, and many equations in statistical mechanics bear his name. These theories were then given experimental proof by Jean Perrin in 1908. Later, in 1923, the first rigorous definition of Brownian motion was presented by Norbert Wiener, who defined a measure on the Banach space of continuous paths. Due to his contribution, Brownian motion is often called a Wiener process, which is why we

will denote Brownian motion by W in this work.

The basic definition of a Brownian motion in $\mathbb{R}$ is as follows.

Definition 3.1.1. *A Brownian motion on $\mathbb{R}$, starting at point $x \in \mathbb{R}$ is a stochastic process $\{W_t\}_{t\geq 0}$ such that the following properties hold[4][51]:*

1. *$W_0 = x$ with probability 1.*
2. *$\{W_t\}_{t\geq 0}$ has independent increments, i.e. $W_{t+u} - W_t$, $u \geq 0$ is independent of W_s for $s \leq t$.*
3. *The process has normally distributed increments, i.e. $W_{s+t} - W_s \sim N(0, t)$.*
4. *W_t is continuous in t, with probability 1.*

An intuitive and rigorous description of a Brownian motion comes from the physical idea of a particle randomly colliding with other particles. In this way the particle undergoes a simple random walk in which it moves according to a vector[1] X_i for a time, where i is the ith collision and then once another collision occurs it travels according to X_{i+1}. From this one can see that the path is given by the sum

$$\sum_{i=1}^{Nt} X_i$$

with N being the number of collisions. If we consider the number of collisions tending to infinity while the collision size goes to 0, in the scaling limit

$$\frac{1}{N}\sum_{i=1}^{Nt} X_i \Rightarrow B_t$$

will converge weakly to a Brownian motion. This is called Donsker's theorem. Using this idea it is possible to consider more general scaling limits allowing for stochastic differential equations, as well as providing a way in which to simulate Brownian motion as a simple random walk.

3.1.1 Path Entropy: Mogulskii's Theorem

Here, we will consider another large deviation theorem to understand the fluctuations about the limiting value of Brownian motion that a random walk approximation will take.

[1] One can also think that it travels along a geodesic in a random direction

Theorem 3.1.1 (Mogulskii)**.** *Let $\{X_i\}$ be a collection of i.i.d. random variables in $\mathbb{R}^d$ such that $\psi(\theta) = \ln \mathbb{E}[e^{\theta X_1}] < \infty$ then the sum*

$$\frac{1}{N}\sum_{i=1}^{Nt} X_i$$

satisfies a large deviation principle in $L^\infty([0,\infty), \mathbb{R}^d)$ with good rate function

$$I(\gamma) = \int_0^\infty \psi^*(\dot{\gamma}_t)dt$$

when γ is absolutely condinuous and infinite otherwise.

where $\psi^*(x) = \sup_{\theta\in\mathbb{R}^d}\{\langle\theta, x\rangle - \psi(\theta)\}$ is the Legendre-Fenchel transform of the log moment generating function ψ. We will see a similar idea present in the document's final part.

3.2 Stochastic Differential Equations

The main tool used in this work is the stochastic differential equation (SDE). The SDE will be introduced on a fairly abstract footing, but examples will be brought to attention quickly in hopes that either the extreme usefulness of the SDE can distract from their abstract footing or that the abstract nature of SDEs will make the application more interesting.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\mathcal{F}_t\}$ be a filtration, an increasing sequence of sigma algebras, of the sigma-algebra $\mathcal{F}$. A stochastic process $\{X_t\}$ is said to be **adapted** to $\{\mathcal{F}_t\}$ if for each $t \in T$, X_t is $\mathcal{F}_t$ measurable. [48]

Definition 3.2.1. *A real-valued adapted stochastic process $\{X_t\}$ is called a **martingale**[48][1] with respect to the filtration $\{\mathcal{F}_t\}$ if*

- *$X_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.*
- *If $s \le t$, $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ almost surely.*

Definition 3.2.2. *An $\mathbb{R}^d$ valued **semi-martingale** is a stochastic process $X = \{X_t\}$ such that for each t*

$$X_t = X_0 + M_t + A_t$$

where $M = \{M_t\}$ is a martingale and $A = \{A_t\}$ is an adapted process of finite variation. [1]

Note, if $f \in C^2(\mathbb{R}^d)$ and X_t is a semi-martingale then $f(X_t)$ is also a semi-martingale. This fact will be useful in the definition of manifold valued semi-martingales.

An important example of a semi-martingale and the most fundamental object of study in this work is the stochastic differential equation (SDE). The theory of SDEs has a rich history and a wealth of applications. These types of equations were first used by Langevin in 1908 to describe the motion of a particle submerged in a fluid subject to random fluctuations due to interaction with the surrounding particles. This is why in the physics literature SDEs are commonly referred to as Langevin equations. This type of equation can be formally written as

$$\dot{x}(t) = b(x(t)) + \xi(t)$$

where $b : \mathbb{R} \to \mathbb{R}$ is a function representing the drift, and ξ is a white noise process, meaning $\xi(t) \sim \mathcal{N}(0, \sigma)$ for each t and $\mathbb{E}[\xi(s)\xi(t)] = \delta(t - s)$. However, despite being useful and producing accurate results[2] the approach used by Langevin was not rigorously founded. Such a function ξ does not exist in the classical sense; it can be understood in the distributional sense and has been studied in Hida calculus in recent years. Moreover, due to the irregularity of such an object as ξ, the function $x(t)$ has no hope of being differentiable so the time derivative cannot exist. This provides conceptual problems for the notion of a velocity.

Itô is credited as the person who formalized the theory of SDEs, in his 1944 paper. In this approach, SDEs can be made using stochastic integral equations. Where we formally write

$$\int_0^t dx_s = \int_0^t b(x_s)ds + \int_0^t \sigma(x_s)dW_s$$

as

$$dx_t = b(x_t)dt + \sigma dW_t$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$ is a vector field, $\sigma : \mathbb{R}^d \to M(d \times d)$ is a matrix field[3] and W_s is a Brownian motion, and dW_t represents the white noise process above. However, it is well known that the paths of a Brownian motion are almost surely of nonfinite variation. So, the above integral cannot be defined in the standard Riemann - Steiljes sense. Due to this, a new integral needs to be defined.

[2] So much so that this formalism is still used

[3] One can view it as a tensor field and is often called the diffusion tensor. This can be helpful to remember when we transition to manifold valued SDEs.

Rigorous development of these types of integrals is too involved for the task at hand, so what will be said is that roughly, the stochastic integral is defined as a completion of sorts from the operator acting on simple functions in the following way. Let $f(t)$ be a simple function with $t \in [a, b] \subset \mathbb{R}$, and $\{t_j\}$ a partition of the interval $[a, b]$ with mesh size δ. Then,

$$\int_a^b f(t)dW_t = \lim_{\delta \to 0} \sum_i f(t_i^*)[W_{t_i} - W_{t_{i-1}}]$$

where t_i^* is a point that is to be decided.

Definition 3.2.3. *The two most common choices for t_i^* give the common stochastic integrals.*

- *If $t_i^* = t_i$ the stochastic integral is called the **Itô integral** and is denoted*

$$\int_a^b f(t)dW_t$$

- *If $t_i^* = (t_i - t_{i-1})/2$ the stochastic integral is called the **Stratonovich integral** and is denoted*

$$\int_a^b f(t) \circ dW_t$$

Both notions of integration can be defined as integration with respect to semi-martingales, but in general, they are not equivalent.[48] However, one can convert between the two integrals. Further, the Stratonovich integral is favored by physicists due to the fact that it obeys the standard rules of calculus, i.e. the chain rule and can be seen as a limit of Langevin-type equations with smooth approximations to the noise.[43] Due to this adherence to the regular rules of calculus it behaves well under coordinate transformations and is consequently the appropriate notion of stochastic integration on manifolds. The Itô integral is usually preferred by probabilists due to its nice probabilistic properties, namely if M is a martingale then

$$\int_0^t f dM_s$$

will also be a martingale - a property lost in the Stratonovich case. Note that in the case of constant noise, the Stratonovich and Itô integrals will yield the same results.

Example 3.2.1. *[The Noisy Spring] Consider a point mass, m, attached to a three-dimensional*

spring with spring force matrix K, and equilibrium vector X_{eq}. The equations of motion for this system follow Newton's second law in the form of a linear ordinary differential equation:

$$m\frac{d^2X}{dt^2} = K(X - X_{eq}).$$

This is a simple physical model with wide-reaching applications. However, at the scale we investigate, this equation will not work. Instead, one must consider the influence of random perturbations from surrounding molecules. This comes in the form of rewriting the equation above as a linear SDE

$$mdV_t = [K(X_t - X_{eq}) - \gamma V_t]dt + \sigma \circ dW_t \tag{3.1}$$

$$dX_t = V_t dt \tag{3.2}$$

An important consideration from the physical setting considerably reduces the model's complexity. In the setting we investigate the Reynolds number is very high. This means that the viscosity is large, and inertial effects may be ignored. This means that $dV_t = 0$. Therefore, we have

$$dX_t = V_t dt = \frac{1}{\gamma}K(X_t - X_{eq})dt + \frac{\sigma}{\gamma} \circ dW_t \tag{3.3}$$

This SDE can be solved using an integrating factor similar to an ODE, yielding the solution

$$X_t = e^{Kt}X_0 + \int_0^t \frac{\sigma}{\gamma}e^{K(t-s)}dW_s. \tag{3.4}$$

Since the solution is an integral of a deterministic function with respect to a Brownian motion, the process X_t is a Gaussian process. Moreover, the process will be stationary and the stationary distribution will be a Gaussian distribution.

3.2.1 Path Entropy: Freidlin - Wentzell Theory

As before we will consider fluctuations about the limiting value for our processes as a way of investigating the entropy of the system. Freidlin - Wentzell theory gives the rate function for a sequence of SDEs with vanishing noise

$$dX_t^n = V_0(X_t^n)dt + \frac{1}{\sqrt{n}}dW_t,\ n \to \infty \tag{3.5}$$

meaning that the system tends towards the deterministic part of the SDE, $V_0(x)$. In the case of only Brownian noise, the rate function can be found by using a change of measure on the Brownian rate function and hence is given by

$$I(\gamma) = \int_0^\infty \|V_0(\gamma_t) - \dot{\gamma}_t\|^2 dt$$

for absolutely continuous γ. In this way, it is shown that the entropy is related to the action of a particular path. This means that the more energy it takes to stray from the deterministic path, the less likely the path is to be taken. In the case of a deterministic process with several local equilibriums, this can be used to estimate the rate of transfer between each equilibrium.

Example 3.2.2 (The Noisy Spring)**.** *Consider a one-dimensional noisy spring with spring constant k. In the small noise limit using Friedlin-Wentzell theory we can investigate the rate function. It will be given by*

$$I(\gamma) = \int_0^\infty |k(\gamma_t - X_{eq}) + \dot{\gamma}_t|^2 dt. \tag{3.6}$$

To find the most likely path one must find the minimum of the rate function over all possible paths. In this case it is simple, notice that $I(\gamma) \geq 0$ for all γ. Furthermore, notice that if $\gamma_t = X_{eq}$ for all time, meaning that γ is a constant function with value X_{eq} then

$$I(\gamma) = \int_0^\infty |K(\gamma_t - X_{eq}) + \dot{\gamma}_t|^2 dt \tag{3.7}$$

$$= \int_0^\infty |K(X_{eq} - X_{eq}) + 0|^2 dt = 0 \tag{3.8}$$

and hence the constant function $\gamma_t = X_{eq}$ is the most likely path. However, it is apparent that small deviations off of this constant path are not unlikely, and quantification of how unlikely can be made using equation 4.6.

3.2.2 Metastable SDEs and CTMC Approximation

Consider now an SDE in the following form

$$dX_t = -\nabla\Phi(X_t)dt + \sigma_X dB_t$$

where $\Phi : \mathbb{R}^3 \to \mathbb{R}$ is a smooth function such that $\lim_{e\to\infty} \Psi(x) = \lim_{x\to\infty} |\nabla\Psi(x)| = \infty$. These types of SDEs are well studied. SDEs of this form have very nice properties, provided the potential

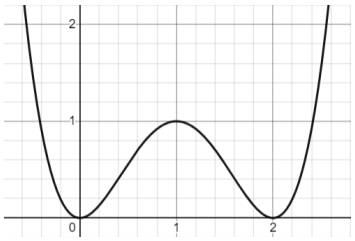
[H] 

Figure 3.1: The double well potential used in this example

meets some light conditions. These types of SDEs will be reversible, and will have stationary distributions related to their potential

$$P(x) = \frac{1}{Z} \int_{\mathbb{R}^3} e^{-\Phi(x)/\sigma_X} dx$$

where $Z \in \mathbb{R}$ is a normalizing constant. These types of distributions are called **Gibbs distributions**.

Now consider the minimum of Φ. In the deterministic setting, these minima would represent stable solutions to the differential equation. In the SDE case, the process will transition between the stable equilibria randomly, based on the noise and the structure of Φ. This phenomenon is called **metastability** and is ubiquitous in chemical physics and biophysics. The local minima are referred to as stable energy wells.

Example 3.2.3 (The Double Well Potential)**.** *This example discusses the properties of a gradient system with the following potential. Let* $\Psi(x) = E((x-W)^2 - \sqrt{r})^2 + y^2 + z^2$ *be a double well potential. This potential has minima at* $(x, y, z) = (r \pm W, 0, 0)$*, and a maximum height between the minima at* $x = r$*, a plot of the potential along the* x *axis is given in figure* **??** *below,*

Intuitively one imagines that the SDE should transition between $x = 0, 2$ *randomly based on the height of the well, the higher the barrier between the two minima, the more consecutive kicks up the hill one needs, making it more unlikely. A sample path demonstrating the switching behavior inherent in these types of systems is shown below.*

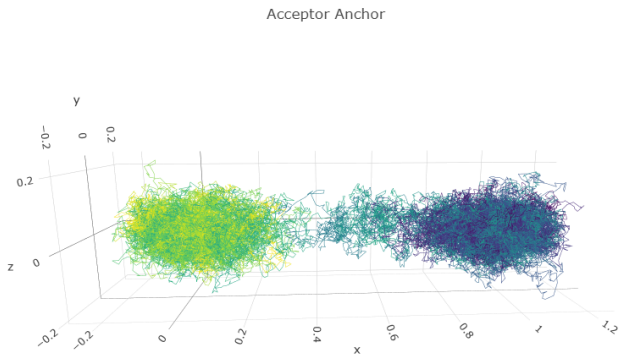


Figure 3.2: An example of a metastable transition between two equilibria.

A natural question is to ask how fast these transitions occur and if there is a way of estimating the transition times based on the geometry of the potential landscape. The answer to this question comes from Kramer's law and Friedlin-Wentzell theory.

It is well known and shown rigorously in the weak noise limit in[29] and was originally estimated by Kramers[33] that the transition time between two local minima of the potential function will depend on the height of the barrier between them, called the activation energy. More specifically, if x_1, x_2 are two stable states and z is an index one saddle point between the stable states then the approximate transition rate between the stable point x_1 to x_2 is given by[65]

$$\frac{|\lambda_s|}{2\pi}\sqrt{\frac{\det(H(x_1))}{|\det(z)|}}\exp\left(-\frac{\delta\Psi}{\sigma}\right)$$

where $\det H(x)$ represents the determinant of the Hessian matrix of Ψ evaluated at x, λ_s is the unique negative eigenvalue of the Hessian matrix at the saddle point, σ is the noise present in the gradient system; lastly $\delta\Psi = \Psi(z) - \Psi(x_1)$ is the activation energy. Note that these types of systems can be modeled as CTMCs with states related to the stable points and transfer rates given by the transition rate above. This supports the idea of modeling the conformation processes as a CTMC.[5] However, while in an equilibrium state, the system is not static but deviates around the equilibrium according to the geometry of the potential landscape around it; quantifying the deviation and noise seen in each stable well needs to be calculated with care. While locally, the potential landscape can be approximated by a quadratic function, leading to Gaussian noise as an approximate perturbation

around the equilibrium, globally, the double well landscape does not have exact quadratic behavior; this means that the deviations in equilibrium will skew towards the transition state. In the current work, an approach of considering the continuum dynamics rather than a CTMC approximation is taken to investigate the role of the noise and its relationship to the potential function in each metastable state.

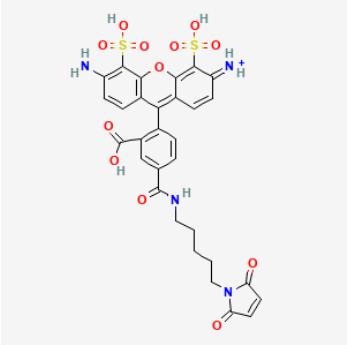
Chapter 4

Simulation of Confocal smFRET Experiments

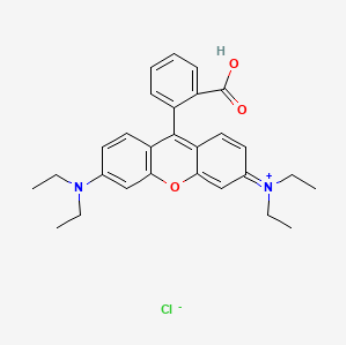
Following the ideas discussed in section 2.1.1.1 in this chapter a more thorough investigation into confocal smFRET experiments is conducted. In parallel a discussion of the simulation of such experiments is carried out. These simulations are used in later chapters to investigate how the structure of the free energy landscape influences the structure of the FRET vs lifetime joint distribution.

4.1 The Structure of a confocal smFRET Experiment

As mentioned in section 2.1.1.1, a smFRET experiment consists of a sample of interest and a solution in which it is diluted. The sample of interest is typically a protein with two fluorescent dyes attached in different locations. These are small molecules that are attached to the target molecule by a point mutation on the molecule and a methylene bridge of C-C links.



(a) Alexa 488



(b) Rhodamine B

Figure 4.1: Alexa 488 and Rhodamine B are two common fluorescent. These dyes are attached to a molecule via a chain of C-C bonds.[25][26]

The fluorescent dyes move as they are subject to the same thermal fluctuations as the protein. The nature of these dynamics is discussed in chapter 7.

Once the sample is diluted it is placed in a confocal microscope. This is a microscope that uses convex lenses to focus light on a specific area creating an illumination profile in the sample. The **confocal volume** is the region in which the illumination profile has, within a cut-off, the highest intensity; it represents the region that is "in focus". The confocal volume is well approximated by a truncated 3 dimensional Gaussian distribution given by[66]

$$I = I_0 \exp\left(\frac{-(x^2 + y^2 + z^2/k)}{w^2}\right)$$

where I_0 is the maximum intensity, z is the direction of propagation of the light, w is related to the beam radius in the xy plane and k is a constant dependent on index of refraction of the solution and the numerical aperture of the lens. This means that the confocal volume is symmetric along the plane perpendicular to the axis of propagation of the light and asymmetric along the axis of propagation.

It is important to note that the number of samples collected in a confocal smFRET experiment is random and dependent on the diffusion of the sample and the confocal volume. Moreover, due to the random diffusion of the sample, useful data is not always being recorded. The samples come in bursts of photon counts. To simulate the random nature of the number of bursts seen in a smFRET experiment, we define the time in which the experiment is conducted and generate a Poisson-distributed random number for each of these seconds. We use $\lambda = 1$ to parameterize the ar-

rival rates of bursts per second. The total number of bursts is defined as the sum of these outcomes. Let N_B denote the total number of bursts in an experiment.

Once in the microscope, a laser of a wavelength such that the photons emitted by this laser will excite the donor molecule is directed towards the diluted solution and pulsed according to a preset pulse time.[32] The pulse time is approximately every 50 nanoseconds or 5×10^{-8} seconds. In the simulations the iteration time of the protein motion is representative of a pulse time.

The light is then passed through a linear polarizer and directed towards the objective. The beam passes through a dichroic beamsplitter or dichroic. [1]

Provided a molecule is in the confocal volume it will be excited by an incoming photon with a probability based on its orientation relative to the polarization of the photon. Let A be the event that the photon is absorbed by the donor and let $\hat{\mu}_D$ be the electric dipole moment of the donor and $\hat{I}$ the unit vector corresponding to the polarization of the incident photon. Then

$$\mathbb{P}(A) = \langle \hat{\mu}_D, \hat{I} \rangle^2 = \cos^2(\theta)$$

where θ is the angle between the two vectors.[37][42] Notice that this distribution is symmetric about the planes perpendicular to the polarization of the incident photon. In this way, we can see that the absorption probability increases the more co-linear the dipole moment is with the polarization of the photon with a maximum probability of absorption when the two are parallel.

Once excited, the donor may transfer its energy via FRET, fluorescence, or decay nonradiatively. This nonradiative decay is representative of other energy pathways the absorbed energy may take, such as increasing vibration between two bonds in the molecule, increasing rotation, or loss of energy due to interactions with other molecules. The probability of emitting a photon once excited is called the **quantum yield** of the molecule.[42] Denote this by Q_D and Q_A for the donor and acceptor quantum yield respectively.

If either dye fluoresces the photon will be emitted in a direction with a distribution dependent on the electric dipole moment of the emitting dye. Let $\hat{e}$ be the unit vector in the direction of propagation of the emitted photon. Then the density of of the random vector $\hat{e}$ is given by

$$p(e) = \frac{1}{\pi^2}(1 - \langle \hat{\mu}, \hat{e} \rangle^2)$$

[1] Dichroic beamsplitters are types of lenses that allow certain wavelengths of light to pass through where others are reflected.

where $\hat{\mu}$ is the dipole moment of the emitting dye, either the acceptor or donor.[37] Note that the distribution for the direction of emission is symmetric about planes perpendicular to the electric dipole moment. This shows that the emitted photon is more likely to travel in a direction perpendicular to the dipole moment of the dye. In this way we can see that the emitted photon will have a polarization representative of the dipole moment of the fluorescing dye. This is the basis for fluorescence anisotropy.[42][37]

To see a photon it must travel in the direction of the detector. So, a great loss of photons occurs due to the stochastic nature of the emitted direction. Provided a photon travels in the appropriate direction, it is reflected by the dichroic mirror the incident light passed through earlier. Once reflected towards the detector it is split based on wavelength and polarization. Once split the photon travels towards a photon detector which may or may not register a photon based on the efficiency of the detector.[32] We refer to the **detector efficiency** by the following probabilities of detection, S_A and S_D for the acceptor and donor detectors respectively. Once a photon is registered is counted and time-stamped with the time after the pulse and the time in the overall experiment. This is then repeated to gather counts of acceptor, and donor photons as well as their lifetimes.

A schematic of the confocal set up is shown below in figure 1.2.

As one can see there is a great loss of data in these experiments, approximately 99% of the photons that should be seen are not due to the loss by quantum yield, orientational factors, or detection efficiency. However, these experiments do create a large pool of data with around $50,000$ photons counted during the course of an experiment. Once these photons are counted one can estimate the FRET efficiency in each burst and observe any dynamics that may occur in the underlying molecule of interest, such as conformational changes in a protein.

4.2 Algorithms and Simulations

For each of the N_B bursts, a sample path of the chosen model dynamics is simulated with an iteration step Δ, which is representative of the pulse time. Each sample path is simulated for a total time T where T is a Gamma-distributed random variable that approximates the burst time and is generated for each sample path. At each iteration (pulse time) of this sample path, a killed CTMC is simulated to represent the absorption, energy transfer, emission, and collection of a photon from a single pulse of the laser. The result of this CTMC gives whether or not the photon was collected,

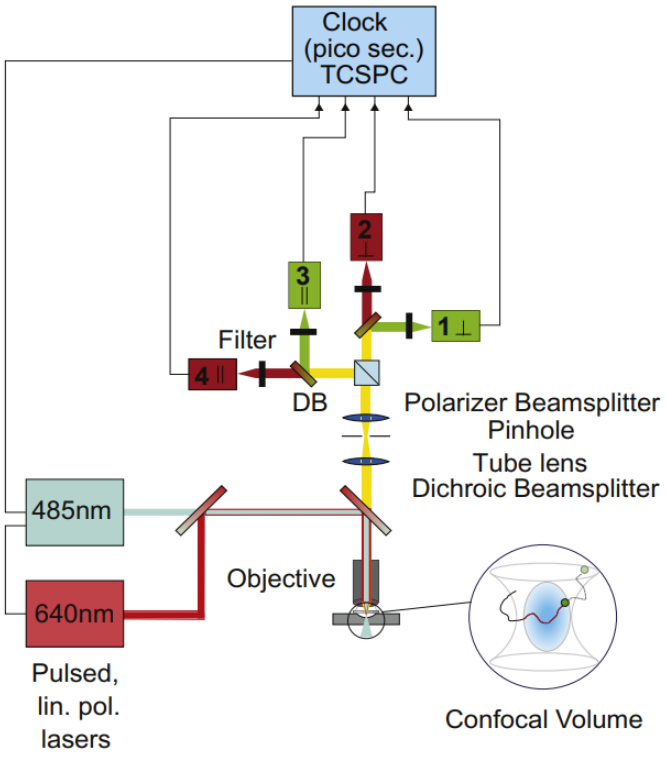


Figure 4.2: An example of a confocal microscopy set up for smFRET experiments. Taken from [32], credit to George Hamiltion.

the color of the photon, and the lifetime associated with the photon. The results from the CTMC for each pulse are collected for the total burst. Intensity-based FRET and lifetime-based FRET estimations can be conducted using the results from each burst.

To simulate the paths of an SDE the Euler-Maruyama method is used. A pseudocode for this algorithm is given below for simulating the general SDE of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t$$

Algorithm 1 Euler - Maruyama Method for SDEs

Require: Δ - Time step, T - Total time

$X(0) \leftarrow x$ ▷ $x \sim F$ for some distribution F is also acceptable.

$\delta W(0) \leftarrow 0$ ▷ Set to 0 since initial distribution is in the previous step.

for n from 1 to N **do**

$\delta W(n) \sim \mathcal{N}(0, \Delta)$

$X(n+1) \leftarrow X(n) + b(X(n))\Delta + \sigma(X(n))\delta W(n)$

end for

This method converges strongly, i.e., in mean square meaning,

$$\max_{t\in[0,T]} \mathbb{E}[|X_t^\Delta - X_t|^2] \leq C\Delta$$

where X_t^Δ is the approximating process.[65] This means that the algorithm converges in mean square with rate $\mathcal{O}(\Delta)$. Moreover the algorithm converges in expectation meaning,

$$\max_{t\in[0,T]} |\mathbb{E}[f(X_t^\Delta)] - \mathbb{E}[f(X_t)]| \leq C_f\Delta$$

for all $f \in C_b^\infty(\mathbb{R}^d)$. So, in both senses, the algorithm converges with a rate $\mathcal{O}(\Delta)$.[65]

An important problem to address is the initial conditions for the SDE simulations. The physical assumption is that the system is in equilibrium. This means that the initial conditions should be generated from the model's stationary distribution. To simulate random variables from these particular distributions the accept-reject sampling method is used.

Algorithm 2 Accept Reject Sampling [50]

Require: Distributions $f, g, \mathcal{U}_{[0,1]}$. Constant $M > 0$ such that $f(x) \leq Mg(x),\ \forall x$.

Generate $X \sim g$ and $U \sim \mathcal{U}_{[0,1]}$

Accept $Y = X$, if $U \leq \frac{f(X)}{Mg(X)}$

Return to state 1 otherwise

For models with non-standard stationary distributions, this algorithm is used with f given by the associated equilibrium distribution. A multivariate Gaussian simulation is used for spring models or the equilibrium disordered protein configuration.

With each iteration, we also calculate the processes dependent on these SDEs. Namely, at each iteration, calculate the vector displacement L_t^{Δ}, the scalar distance r_t^{Δ}, the energy transfer rate $k_{ET}^{\Delta}(t)$ and the FRET efficiency $\mathcal{E}_t^{\Delta}$. Using these processes, a simulation of the energy transfer and detection processes is conducted using the simulated k_{ET}. This process incorporates the loss of photons via the quantum yield, the detector efficiency and loss by orientational factors. The pseudocode for this is given below. This is a modified version of the CTMC code in [17]

Algorithm 3 Energy Transfer Simulation

Require: Δ, t $k_{ET}^{\Delta}(t)$ - time step in algorithm and energy transfer rate at a multiple of .

k_D, k_A - fluorescence rates for donor and acceptor resp.

Q_D, Q_A - Quantum yield of donor and acceptor resp.

S_D, S_A - Detector efficiency for donor and acceptor channels.

P_{ND} - a probability based on the photon loss due to orientation.

state $\leftarrow D$

while $t < \Delta$ state $\neq ND, F_D, F_A$ **do**[2]

if state $= D$ **then**

Generate two Bernoulli random variables with probabilities $(1 - Q_D)$ and

P_{ND} if either is 1 state $\leftarrow ND$ and $\tau = 0$.

Generate $\tau_{DF} \sim \exp(k_D)$ and $\tau_{ET} \sim \exp(k_{ET}(t))$

end if

if $min(\tau_D, \tau_{ET}) = \tau_D$ **then**

$\tau = \tau_D$

state $\leftarrow F_D$, break

else

$\tau = \tau_{ET}$

state $\leftarrow A$

end if

if state $= F_D$ **then**

Generate Bernoulli random variable with probability of success $(1 - S_D)$

if 1 then state $\leftarrow ND$.

end if

if state $= A$ **then**

Generate Bernoulli random variable with probability of success $(1 - Q_A)$

if 1 then state $\leftarrow ND$.

Generate $\tau_{AF} \sim \exp(k_A)$.

$\tau = \tau + \tau_{AF}$

State $\leftarrow F_A$, break

end if

if state $= F_A$ **then**

Generate Bernoulli random variable with probability of success $(1 - S_A)$

if 1 then state $\leftarrow ND$.

end if

For each time step t, the output is a lifetime and color for the associated photon with lifetimes of 0 corresponding to non-detection. From here, conduct intensity-based and lifetime-based FRET analysis on the data generated. This analysis consists of plotting the FRET vs lifetime distributions compared against the static line, as well as plotting the difference in the first and second moments of the lifetime distribution vs the FRET efficiency. Moreover, using the definition of the dynamic shift, dynamic shift histograms are generated for each set of data generated.

Chapter 5

Free Energy Landscape Geometry and smFRET

5.1 Metastable Proteins

Here, we consider introducing dynamics to the underlying protein by modeling the motion of the equilibrium position in the static model as a stochastic process. We simplify the complex protein dynamics by considering two reaction coordinates: the anchor points for the dyes. Hence, the protein dynamics are projected onto the dynamics of these two locations. The reaction coordinates, denoted P^d and P^A, evolve according to a nonlinear SDE with additive noise and drift according to the gradient of a potential function[28]. For more information on these types of SDEs review chapter 3 and section 3.2.2 in particular. The reaction coordinate processes are given by

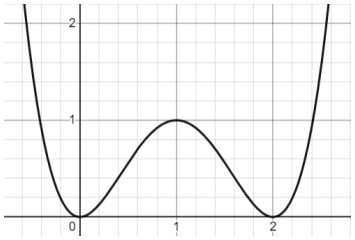
$$dP_t^D = -\nabla\Phi(P_t^D)dt + \sigma_{P^D} dB_t, \tag{5.1}$$

$$dP_t^A = -\nabla\Psi(P_t^A)dt + \sigma_{P^A} dB_t, . \tag{5.2}$$

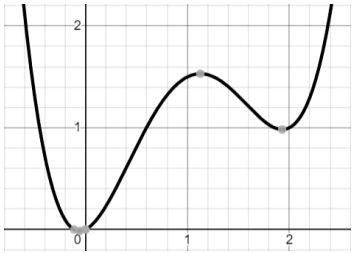
We consider several cases of a double well potential wherein the system exhibits transitions between two metastable states occurring only along the x - axis for the acceptor anchor only, without loss of generality. The donor behaves as a noisy spring similar to the dye model. Hence, there are only two possible stable locations for the system corresponding to the acceptor anchor transition

along the x-axis. The three cases of potential landscape geometry are:

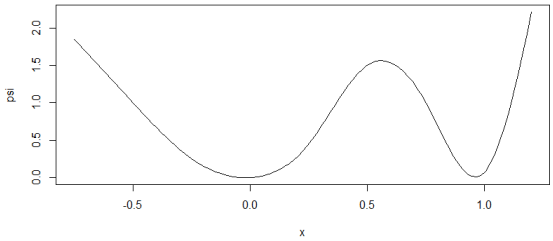
1. The symmetric Double Well: $\Psi(x, y, z) = E((x - W)^2 - W)^2 + y^2 + z^2$



2. Asymmetric Height Double Well: $\Psi(x, y, z) = E((x - W)^2 - W)^2 + Cx + y^2 + z^2$



3. Asymmetric Width Double Well: A cubic spline approximation to an asymmetric width double well.



Here we investigate the properties of the joint distribution of FRET and donor lifetime and see how the transition rate between states and the geometry of the energy landscape impacts this distribution.

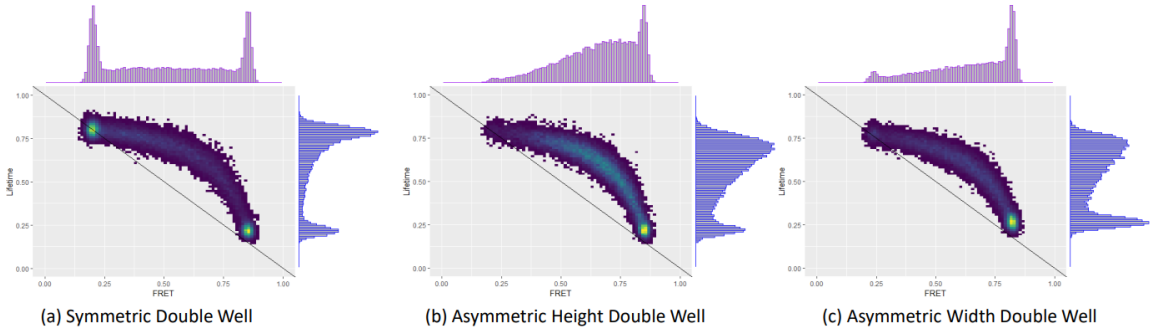


Figure 5.1: Example of moderate transition rates

In the above plot one can notice that each geometry has a different-looking distribution despite having approximately the same transition rates. Notice that the symmetric potential and the Asymmetric width potential are very different despite the fact that the equilibrium points are the same. The asymmetric width plot is reminiscent of the asymmetric height distribution, implying that the width of the energy landscape can influence the structure of the distribution. In the lifetime distribution there seems to be no real difference between having an asymmetry in the width or the hight of the energy landscape. The difference between the two can be seen in the FRET distribution wherein the asymmetric height presents a FRET distribution more skewed towards the lower equilibrium point. In the asymmetric width case the FRET efficiency peaks slightly at one well and then is mostly uniformly distributed as we approach the well with the sharper barrier.

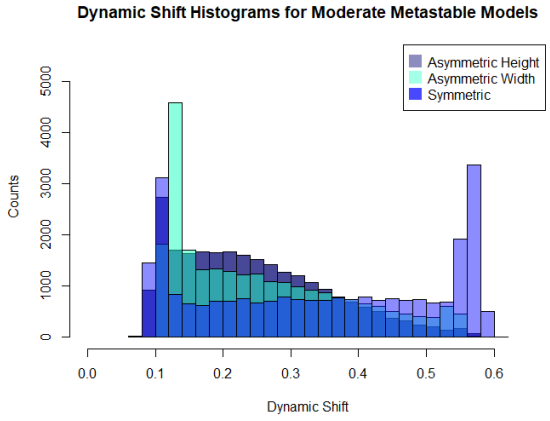


Figure 5.2: Dynamic shift histogram comparison between the different energy landscapes in the moderate transition regime.

In the case of a fast transition geometry still plays an important role. As seen in the figure

below the symmetric double well produces a distribution that has a higher average FRET and is oriented differently with respect to the static line.

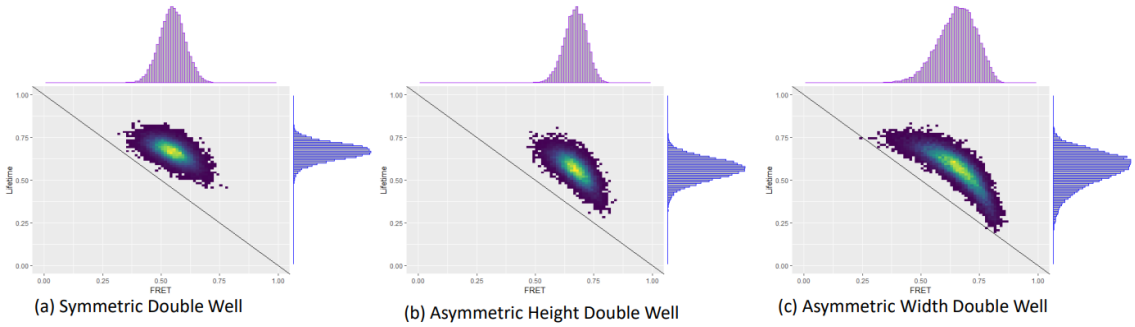


Figure 5.3: Fast Transitions

The main differences become apparent in the asymmetric double wells. The asymmetric height double well provides a slightly lower mean lifetime and an orientation that shows a tendency towards the deeper stable state. Interestingly the asymmetric width case provides the same mean lifetime as the asymmetric height but presents a skewing in both the lifetime and FRET distributions caused by the slower diffusion and larger volume of exploration in the wider well than can be seen in the other two models.

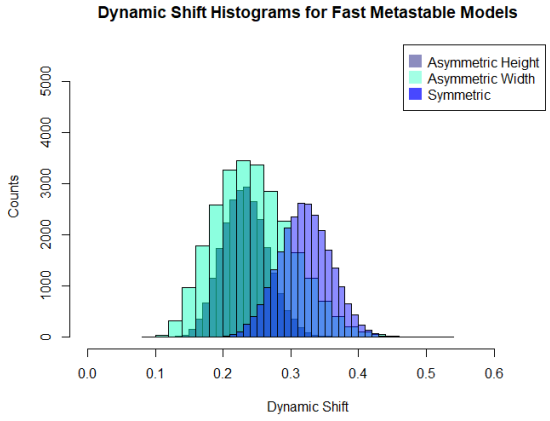


Figure 5.4: Dynamic shift histogram comparison between the different energy landscapes in the fast transition regime.

5.2 Disordered Landscapes

Here, the underlying protein is considered a flexible polymer modeled by a system of C overdamped springs with equal spring constants k. Each spring represents a part of the protein, and the location of this part is given by an SDE governed by the neighbors in the chain. We consider each spatial coordinate to evolve independently of the others. Consequently the i^{th} coordinate of each component of the chain is governed by a linear system of C SDEs $dU_t^i = \mathbf{K}U_t^i dt + \boldsymbol{\sigma} dB_t, \quad i = 1, 2, 3,$ where $\mathbf{K} \in \mathbb{R}^{C \times C}$ is the tridiagonal spring constant matrix with $-k$ on the diagonal and k on the off-diagonal. Independent thermal fluctuations drive the noise component. Hence, $\boldsymbol{\sigma} = \sigma I$, where σ is the standard thermal noise coefficient and I is the identity matrix.

Since each component is a linear SDE in $\mathbb{R}^C$ driven by Gaussian noise, the solution is readily available in terms of a matrix exponential and the stationary distribution will be a multivariate Gaussian with covariance given by the solution to the Lyaponov equation, $\mathbf{K}\boldsymbol{\omega} + \boldsymbol{\omega}\mathbf{K}^T = \boldsymbol{\sigma}\boldsymbol{\sigma}^T$. Since $\boldsymbol{K}$ is symmetric and $\boldsymbol{\sigma}\boldsymbol{\sigma}^T = \sigma^2 I$, the solution is given by $\boldsymbol{\omega} = -\sigma^2 \boldsymbol{K}$. Consequently, the stationary distribution for the system of equations is a multivariate Gaussian with mean vector 0 and covariance matrix $-\sigma \boldsymbol{K}$. Note that this is a C dimensional space; this does not mean that the mean length of the chain is 0.

In this section, we will only investigate how the placement of the linkers on the chain affects measurements. However, once some auxiliary information is discussed in regard to dye linker models and dipole moment dynamics, another discussion on disordered landscapes and the smFRET dynamic shift will be held.

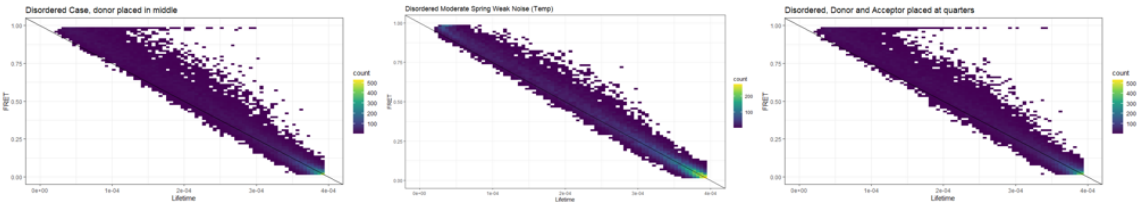


Figure 5.5: Samples of the above IDP model using the smFRET simulations

Notice here that the placement of the dye along the chain does not cause serious differences; moreover, the lifetime and FRET are both very uniformly distributed. However, there is a light tendency towards low FRET, as expected. Notice at a lower noise level, the strong dependence of the FRET lifetime distribution on the static line becomes more pronounced. In addition, with lower

noise, low FRETs are observed more frequently.

Chapter 6

Stochastic Geometric Mechanics

6.1 Rotational Dynamics

When considering the motion of a particle, such as a protein or a fluorescent dye, subject to thermal fluctuations, one considers the translational diffusion of the particle. This is how it moves in space, similar to how the reaction coordinates for the protein models moved in space. Mathematically, the state space for such models is $\mathbb{R}^n$. However, an important aspect of a particle's state is its orientation. The notion of orientational or rotational dynamics complicates the picture considerably.

To consider rotations of rigid bodies, one works with the Lie group $SO(3)$. This is the group of orthogonal matrices with determinant 1. This group represents all rigid rotations of Euclidean space. So, by considering a motion on this space, when starting with an initial configuration one obtains the rotational dynamics of a rigid body. Let $\omega \in so(3)$ be the rotational velocity of a curve in $SO(3)$, and the Lie bracket $[X, Y] = XY - YX$, and define $I \in M(3 \times 3)$ to be the inertia tensor of the rigid body. Then, the Euler equations tell us classically[64],

$$I\dot{\omega} = [I\omega, \omega] + TB_t \tag{6.1}$$

Where TB_t in the equation above is a Brownian motion added following the direction from section 6.2.1, this gives the stochastic equation of motion for a freely rotating body subject to white noise.[11] Notice now that if I is a multiple of the identity, then $[I\omega, \omega] = 0$, and hence $I\dot{\omega} = TB_t$. Therefore,

a Brownian motion will be the rotational dynamics of a particle with a multiple of the identity for its inertia tensor. Such rigid bodies are uniform spheres.

Example 6.1.1. *[Electric Dipole Dynamics] Consider a rigid body, Γ in three-dimensional space with inertia tensor I. Set an initial configuration for Γ, Γ_0. Let μ_0 be a unit vector centered in Γ_0. If one considers ω_t a process on $SO(3)$ given by the Euler equations above, then a problem of interest is to track the motion of $\omega_t\mu_0$. This is equivalent to tracking the diffusion of the electric dipole moment for a fluorophore.*

In FRET is particularly important. This is because the Förster radius is dependent on the mutual dipole orientations,

$$\kappa_t = \mu_t^A \cdot \mu_t^D - 3(\mu_t^A \cdot \hat{R})(\mu_t^D \cdot \hat{R}) \tag{6.2}$$

and $R_0^6 = \kappa_t^2 \bar{R}_0$. While the influence may not seem to be so large since the magnitude of R_0 only fluctuates by $\kappa_t^{1/3}$. However, the main issue to be concerned with is the fact that κ_t is often close to 0. This causes a skew in the energy transfer rate, and since κ dynamics happen in a fast time scale [59, 16], the rapid fluctuations in the Förster radius may cause additional variance in the lifetime distributions.

If we consider the fluorophore as a spherical molecule, then as shown above the Euler equations will give that the rotational diffusion of the dye is a Brownian motion. Let X_i be an orthonormal basis for the Lie algebra $so(3)$ of $SO(3)$. Then, the rotational Brownian motion can be expressed as

$$TB_t = \sum_{i=1}^{9} X_t \circ dB_t \tag{6.3}$$

with B_t being a Euclidean Brownian motion. More information on what these Stratonovich SDEs mean is shown in section 6.2. Now, the path of operators TB_t acts as a map $\mathcal{P}(SO(3)) \to \mathcal{P}(S^2)$ by considering the path $\mu_t = TB_t\mu_0 \in S^2$. Let $\hat{\mu}_0$ denote the operator taking the path TB_t and applying it to the vector μ_0. Then $\hat{\mu}_0$ is a diffeomorphism [40] and hence by proposition 1.2.4 in [35] we see that since TB_t solves the SDE for Brownian motion on $SO(3)$, then $\mu_t = \hat{\mu}_0(TB_t)$ solves the SDE for Brownian motion on S^2.

The equations of motion for a spherical Brownian motion are given in the following SDEs,

$$d\theta_t = \frac{\sigma_\theta^2}{\tan\theta_t} \circ dB_t^\sigma$$
$$d\phi_t = \frac{\sigma_\phi}{\sin(\theta_t)} \circ dB_t^\phi$$

This simplification is useful as a full simulation of more complicated dynamics on $SO(3)$ can be difficult. However, it can be done see for instance, [11, 12].

Further, in considering the rotational dynamics of a particle, one may also consider any constraints the particle is placed under. For example, the spherical pendulum is constrained to only move along the sphere. Hence, an understanding of the geometry of the sphere is needed. More generally, constraints and rotational dynamics highlight the need for a study on the interaction between geometries and processes

6.2 Differential Geometry

Differential geometry is a vast area of mathematics, and we will only review the tools needed to show the following results.

Heuristically speaking, a manifold is a topological space that locally looks like an Euclidean space. What is meant by this is that a topological space $\mathcal{M}$ is a manifold if for each $x \in \mathcal{M}$ there exists an open neighborhood $U_x \subset \mathcal{M}$ and a homeomorphism, called a **coordinate chart**, $\phi_{U_x} : U_x \to \mathbb{R}^d$. It is further required that these coordinate charts behave well on overlaps. Meaning if $U, V \subset \mathcal{M}$ such that $U \cap V \neq \emptyset$ then the map

$$\phi_V \circ \phi_U^{-1} : \phi_U(U \cap V) \to \mathbb{R}^d$$

is differentiable. [27][1]

[1]Note that $\phi_V \circ \phi_U^{-1}$ is a mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$, so we know what differentiability means.

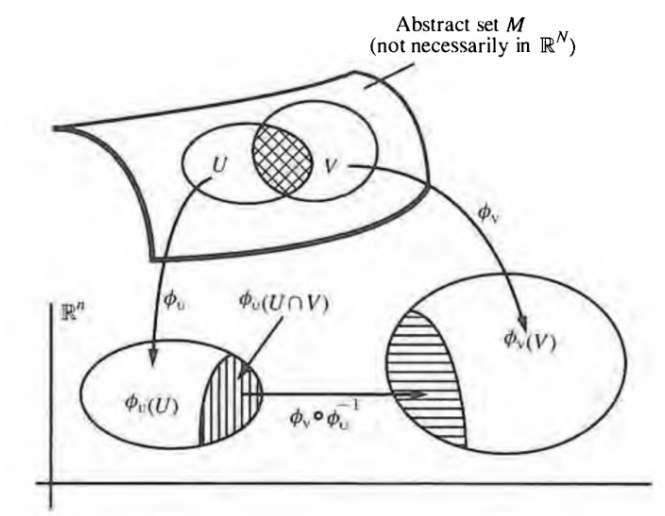


Figure 6.1: An Illustration of the Coordinate Charts on a Manifold. Taken from [27]

The manifold is called a C^K manifold if these transition maps are at least K times differentiable and **smooth** if $K = \infty$. The collection of coordinate charts is referred to as an **atlas** and if there exists an atlas that is countable, the manifold is called **second countable**. Moreover, the dimension of the manifold is the dimension of the Euclidean space it is locally homeomorphic to. For the purposes of this work we always consider the manifold $\mathcal{M}$ to be smooth and second countable. This eases our considerations.

A fundamental object in the study of differential geometry is the **tangent space**, $T_x\mathcal{M}$, at each point $x \in \mathcal{M}$. This is the set of all **tangent vectors** at the point x. Loosely, this means that for a smooth curve $p : [-1, 1] \to \mathcal{M}$ such that $p(0) = x$ the velocity vector of p at time t, $\dot{p}(0)$ is an element in the tangent space of x. This space can also be seen as a space of operators on smooth functions $f : \mathcal{M} \to \mathbb{R}$ by considering the directional derivative of f in the direction of $v \in T_x\mathcal{M}$.[27][44] In this way tangent vectors are often written in coordinates as

$$v = \sum_{i=1}^{d} v_i \frac{\partial}{\partial x_i}$$

and can be thought of as a first order differential operator. Elements of this space are sometimes called contravariant vectors or just vectors. The **tangent bundle** of $\mathcal{M}$ is denoted $T\mathcal{M}$ and is the

disjoint union of each tangent space,

$$T\mathcal{M} = \bigcup_{x \in \mathcal{M}} \{x\} \times T_x\mathcal{M}$$

meaning an element in the tangent bundle is expressed as (x, v) where $x \in \mathcal{M}$ and $v \in T_x\mathcal{M}$. Vector fields on $\mathcal{M}$ are smooth sections [2] of the tangent bundle. This means a vector field V on $\mathcal{M}$ is an assignment of a vector $v(x)$ at each point $x \in \mathcal{M}$. The space of all vector fields will be denoted $\Gamma(T\mathcal{M})$.

The tangent space at x is easily seen to be a finite-dimensional vector space. So, it is natural and useful to consider the dual to this space. This is the **cotangent space** and consists of all the linear functionals on the tangent space. This space is denoted $T_x^*\mathcal{M}$. Elements in this space are sometimes called 1-forms or covariant vectors. In a similar way to the tangent bundle, one can consider the cotangent bundle

$$T^*\mathcal{M} = \bigcup_{x \in \mathcal{M}} \{x\} \times T_x^*\mathcal{M}$$

and the set of covariant vector fields is denoted $\Gamma(T^*\mathcal{M})$.

An important class of covariant vectors are those of **differentials of functions**. These are those covariant vectors defined by the following; for a function $f : \mathcal{M} \to \mathbb{R}$ let $df(x) \in T_x^*\mathcal{M}$ be the covariant vector such that for any $v \in T_x\mathcal{M}$,

$$df(x)(v) := v(f)(x).$$

[27][44] Note, that there are many different ways of defining this idea, but this seems to be the simplest and most clear explanation for the current work.

6.2.0.1 Riemann Geometry

So far everything that has been mentioned works for any differentiable manifold. However, the setting of this work is in a Riemann manifold. A Riemann manifold is a manifold such that each tangent space $T_x\mathcal{M}$ has an inner product - symmetric, positive definite, bilinear form, g_x. We

[2] Note, we will not get involved with the existence of a global section or anything like this. It is assumed that these vector fields are defined as local sections that can be smoothly stitched together.

denote for $v, w \in T_x\mathcal{M}$,

$$g_x(v, w) = \langle v, w \rangle_{g(x)}.$$

If the bilinear form is smooth along $\mathcal{M}$, meaning for each $V, W \in \Gamma(T\mathcal{M})$ the function $\langle V(x), W(x) \rangle_{g(x)}$ is smooth, then it is called the **Riemann metric tensor**.[44] A manifold $\mathcal{M}$ equipped with a metric tensor is called a **Riemann manifold** and is denoted $(\mathcal{M}, g)$. Further using the Riemann metric tensor one can define an analogue of the Laplace operator on manifolds, the **Laplace - Beltrami** operator, $\Delta_{\mathcal{M}}$ written in coordinates it is given by[27][35]

$$\Delta_{\mathcal{M}} f = \sum_{i=1}^{d} \sum_{j=1}^{d} \frac{1}{\sqrt{det(g)}} \frac{\partial}{\partial x_i} \left(\sqrt{det(g)} (g^{-1})_{i,j} \frac{\partial f}{\partial x_j} \right).$$

It will be beneficial to identify the cotangent space and tangent space at a point. Since $T_x\mathcal{M}$ is a finite dimensional inner product space, we may identify it with its dual space, $T_x^*\mathcal{M}$. In this way we can define the so called musical isomorphisms.[44] Define for $p \in T_x^*\mathcal{M}$, $p^\sharp \in T_x\mathcal{M}$ to be the unique element in $T_x\mathcal{M}$ such that

$$p(v) = \langle p^\sharp, v \rangle_{g(x)}.$$

This is sometimes referred to as a sharp. Similarly, define for $w \in T_x\mathcal{M}$, $w^\flat \in T_x^*\mathcal{M}$ to be the unique element in $T_x^*\mathcal{M}$ such that

$$w^\flat(v) = \langle w, v \rangle_{g(x)}.$$

For Riemann manifolds there is a particular covariant derivative that works well with the metric tensor. This is called the Levi-Civita connection.

Definition 6.2.1. *The **Levi - Civita Connection** on $(\mathcal{M}, g)$ is the unique covariant derivative, ∇ that satisfies the following[44];*

- *For all $X, Y, Z \in \Gamma(T\mathcal{M})$,*

$$X\langle Y, Z \rangle_{g(x)} = \langle \nabla_X Y, Z \rangle_g(x) + \langle Y, \nabla_X Z \rangle_{g(x)}$$

- *For all $X, Y \in \Gamma(T\mathcal{M})$ we have for the lie bracket $[X,Y](f) = X(Y(f)) - Y(X(f))$,*

$$\nabla_X Y - \nabla_Y X = [X, Y]$$

The Levi-Civita connection is perferred and will be the connection we use from now on because it makes parallel transport an isometry. Meaning, that parallel translation of vectors perserves their inner product.[44] This is very important in the construction of Brownian motion and its infinitesimal generator on a Riemann manifold.

Using the Riemann metric tensor we can define the Riemann distance between two points on the manifold. First, note for $v \in T_x\mathcal{M}$ we can use the inner product to define a norm in the usual fashion,

$$|v|_{g(x)} = \langle v, v, \rangle_{g(x)}^{1/2}$$

and from this we may define the length, or total variation, of a differentiable curve γ such that $\gamma(0) = x$ and $\gamma(1) = y$ in the usual way to be[44]

$$\|\gamma\|_{TV} = \int_0^1 |\dot{\gamma}(t)|_{g(\gamma(t))} dt.$$

The **Riemann distance** is defined for $x, y \in \mathcal{M}$ to be

$$d(x,y) = \inf\{\|\gamma\|_{TV} : \gamma(0) = x, \gamma(1) = y\}.$$

It can be shown that the definition of geodesics being those paths that minimize the distance between two points coincides with the geodesics defined by the Levi-Civita connection.

Next we define the Riemann exponential map and define geodesic completness of a manifold. Properties that arise from the exponential map will also be important in proofs later on.

Definition 6.2.2. *Let γ_V be the geodesic starting at $x \in \mathcal{M}$ with velocity $V \in T_x\mathcal{M}$. Define $D(Exp_x) \subset T_x\mathcal{M}$ as the set $\{V \in T_x\mathcal{M} | \gamma_V$ is defined on the interval $[0,1]\}$. The **exponential map**[44] is the operator $Exp_x : D(Exp_x) :\to \mathcal{M}$ such that*

$$Exp_x(V) = \gamma_V(1)$$

If the domain of the exponential map for each $x \in \mathcal{M}$ is the entire tangent space $T_x\mathcal{M}$, then the manifold is called **geodesically complete**. By the Hopf - Rinow theorem, this is equivalent to the manifold being a complete metric space with respect to the Riemann metric, d. [44] Together with second countability, this makes the manifold a complete separable metric space. This is the type of space required for the theory in [23] to be applicable. From here on out, we assume that $\mathcal{M}$ is a geodesically complete manifold.

It is important to note that the exponential map need not be injective; many different geodesics can end at the same point. For example, there are infinite geodesics connecting antipodal points on the sphere. However, a ball exists around each $x \in \mathcal{M}$ such that inside this ball, the exponential map is injective and, in fact, a diffeomorphism. [44] Define the **injectivity radius** of Exp_x as

$$i(x) = \inf\{\delta > 0 | \text{Exp}_x \text{ is injective on } B(0, \delta)\}$$

note that $i(x) \in (0, \infty]$. Inside this ball, geodesics connecting points will be unique; this will be used later. Another important property gained from the exponential map is the scaling property of geodesics. The following proposition will be useful later.

Proposition 6.2.1. *Let $c \in \mathbb{R}^+$ and $V \in T_x\mathcal{M}$. Define γ_{cV} as the geodesic starting at x with velocity cV. Then*

$$\gamma_{cV}(t) = \gamma_V(ct).$$

In other words, scaling the velocity is equivalent to scaling the time. [44] In particular,

$$\lim_{c \to 0} \gamma_{cV}(t) = \gamma_V(0).$$

Moreover, observe that if $x_n, y_n \to z$ such that $d(x_n, y_n) \to 0$ then the sequence of minimal geodesics connecting x_n and y_n will converge to the constant geodesic. This implies that parallel transport along such a geodesic converges to the identity map on $T_z\mathcal{M}$. [44][8]

The following property was first defined in [61] and involves the notion of a measure being consistent along the manifold. It was initially developed in the context of a geodesic random walk to establish the independent increment properties and later extended to a condition to define a good containment function in [38].

As each tangent space is different depending on the point, each measure on each tangent

space is different. So, we need a way of translating a measure on those spaces in a way that does not alter this measure in any significant fashion, similar to the parallel transport of a vector along the manifold. Let $\{\mu_x\}_{x\in\mathcal{M}}$ be a collection of measures such that $\mu_x \in \mathcal{P}(T_x\mathcal{M})$. We say this collection satisfies the **consistency property** if for any smooth curve $\gamma : [a, b] \to \mathcal{M}$ such that $\gamma(a) = x$ and $\gamma(b) = y$ we have

$$\mu_y = \mu_x \circ \tau^{-1}_{x,y;\gamma} = (\tau_{x,y;\gamma})_{\#}\mu_x$$

i.e. for any smooth curve connecting two points on the manifold that the pushforward measure of the parallel transport map along that curve gives us the proper measure on the tangent space at the endpoint. Importantly, the normal distribution satisfied this property.

6.2.1 Stochastic Processes on Manifolds

In this section, we mostly follow [35] in the description of manifold-valued SDEs. We will discuss the construction of Brownian motion on a manifold by considering an SDE on the orthonormal frame bundle and the martingale problem for Bochner's Laplacian.

It is counterintuitive to find that the definition of a semi-martingale on a manifold is much simpler than that of a martingale.

Definition 6.2.3. *Let $Z = \{Z_t\}$ be an $\mathbb{R}^d$ valued continuous semi-martingale. Further, define $V_0, V_1, ..., V_d$ to be smooth vector fields on a d dimensional manifold $\mathcal{M}$. We call $X = \{X_t\}$ a solution to the SDE*

$$dX_t = V_0(X_t)dt + \sum_{i=1}^{d} V_i(X_t) \circ dZ^i_t$$

if for each $f \in C^\infty(\mathcal{M})$, $f(X_t)$ the following holds

$$f(X_t) = f(X_0) + V_0 f(X_t) + \sum_{i=1}^{d} \int_0^t V_i f(X_t) \circ dZ^i_t.$$

Note that these processes are only defined up to a (possibly finite) stopping time, τ, called an explosion time. If $\mathbb{P}(\tau = \infty) = 1$, i.e., if this stopping time is almost surely infinite, the manifold is called **stochastically complete**. There are many sufficient conditions [3] for a Manifold to be stochastically complete with respect to Brownian motion[4] Hence, SDE is driven by Brownian motion.

[3] As far as I am aware there are no necessary conditions known.

[4] Note, we have yet to define Brownian motion on $\mathcal{M}$ yet. However, starting this next theorem will relieve some technicalities when we do.

The following criteria is a work of Grigorian as the following[35]

Proposition 6.2.2. *Let $\mathcal{M}$ be a Riemannian manifold with distance d. Define the radial function $r(x) = d(x, o)$ for an arbitrary point $o \in \mathcal{M}$. Further, let $Vol(B(r))$ be the volume of a geodesic ball around o of radius r. If*

$$\int_c^\infty \frac{r}{|ln(Vol(B(r)))|} dr = \infty$$

then $\mathcal{M}$ is stochastically complete.

We can see from this that stochastic completeness relies on the manifold's curvature. Indeed, this condition can be related to the Ricci curvature of the manifold, and it is true (but is a much more strict assumption) that if $Ric_M \geq L$ for some $L \in \mathbb{R}$ then the manifold is stochastically complete.[35][38] However, the claim above provides much more freedom in our choice of geometry. Essentially, it requires that the volume of a geodesic ball is bounded by De^{Ar^2} for some constants A, D. From now on, we assume we work on a stochastically complete manifold.

Having the definition of a $\mathcal{M}$ valued SDE, it is natural to wonder about the case when $Z = W$ is a Brownian motion. In this case, we return to the theory of diffusion processes. It is well known there is a deep connection between second-order elliptic operators and diffusion processes, so one would like this to extend to the manifold setting. Indeed, this is the case. It can be shown [55][35] that the $\mathcal{M}$ valued diffusion process

$$dX_t = V_0(X_t)dt + \sum_{i=1}^d V_i(X_t) \circ dW_t^i$$

solves the martingale problem for the second order elliptic operator

$$\mathcal{A} = V_0 + \frac{1}{2}\sum_{i=1}^d V_i^2$$

where $V_i^2(f) = V_i(V_i(f))$. Such operators are called Hörmander form operators[55][35] due to their relation to Hörmander's theorem.

Note that this type of operator is a different notation for the second-order elliptic differential operator we expect from working on $\mathbb{R}^d$. We can investigate V_i^2 in coordinates to see how this is the case. Let

$$V_i(x) = \sum_{j=1}^d a_j(x)\frac{\partial}{\partial x_j}$$

then we have

$$V_i(V_i(f)) = \sum_{k=1}^{d} a_k(x)\frac{\partial}{\partial x_k}\left(\sum_{j=1}^{d} a_j(x)\frac{\partial}{\partial x_j}\right)$$
$$= \sum_{k=1}^{d}\sum_{j=1}^{d}\left(a_k(x)\frac{\partial a_j(x)}{\partial x_k}\frac{\partial}{\partial x_j}\right) + a_k(x)a_j(x)\frac{\partial^2}{\partial x_k \partial x_j}$$

So, adding V_i^2 in our generator slightly augments the drift term while including the diffusion term. Note that this augmentation of the drift is similar to the augmenting of the drift when converting from Stratonovich to Ito SDEs. Since the Stratonovich integral is used, the noise should be expected to be reflected in the drift term. Further, the geometry of the space can present a "virtual" drift in the generator.

This operator is technically an element of the double tangent bundle of $\mathcal{M}$ and is sometimes called a "diffusor". These types of vector fields are a vital element in the theory of second-order differential geometry.[20] This is an area of geometry that is inspired by stochastic differential geometry and, in many ways, generalizes stochastic differential geometry. However, this topic is beyond the scope of the current work and will not be discussed further.

One of the main difficulties in defining Brownian motion on a manifold is that one would expect a Riemann Brownian motion to be a process X_t defined by some Euclidean-driven SDE as above. However, this would include some constant vector fields on $\mathcal{M}$. It is well known that if $\mathcal{M}$ is not parallelizable, such vector fields do not exist. This presents a problem. Moreover, one could expect to characterize Brownian motion by the Lévy characterization, but this leads to the problem of defining quadratic variation for Manifold-valued processes. It turns out that for each $(0, 2)$ tensor field on $\mathcal{M}$ one can define a quadratic variation.[35][20]

6.3 Path Entropy: Schilder's Theorem

The last part of this section is to recall a large deviation result in the context of understanding the fluctuations of a randomly rotating electric dipole with a spherical molecule. We know that this molecule rotates according to a spherical Brownian motion. In a law of large numbers type scaling, we can ask what the rate function for such a process is. This generalizes Schilder's theorem to manifolds. The extension was recently carried out by Versendaal et al. in [62]. We see that the

rate function for a Brownian motion under the scaling $W(nt)$ as $n \to \infty$ is given by

$$I(\gamma) = \int_0^\infty |\dot{\gamma}_t|^2_{g(\gamma_t)} dt. \tag{6.4}$$

However, it turns out that the extension of the same theory for stochastic differential equations or general Markov processes has not been completed yet. So, analysis of the rate function for rotational dynamics of nonspherical molecules needs to be done. This is the topic of chapter 10

Chapter 7

Fluorescence Dynamics in smFRET

The term fluorescence dynamics has been coined recently to describe the study of processes influencing fluorescent measurements.[45] As mentioned in previous sections, FRET is very sensitive to changes in inter-dye displacement; as such, the motion of the fluorescent probes will influence FRET measurements. As the field currently stands, most uncertainty quantification for FRET measurements involves molecular dynamics (MD) simulations, in which a deterministic simulation is conducted considering all known forces on each atom. These simulations are costly and do not simulate an extended amount of time; usually, the time is less than a burst time in a smFRET experiment. The next chapter will outline a modeling approach using SDEs and ideas from stochastic geometric mechanics as an alternative to all-atom MD simulations.

The first section concerns different mechanical models for the dye and how they are derived and considered. The next chapter details simulations conducted to investigate the model behavior in a smFRET experiment. It discusses the simulation methods and what variables are considered when simulating the smFRET experiment. The final chapter examines different results from the dye models using smFRET as a benchmark by which to study the dye models. As we have seen in previous sections, when one collects FRET data, the data is altered by the underlying system being in a mixture of states rather than just one. This section considers the mix of states imposed on the system by the dye motion. Therefore, an understanding of the dye dynamics is needed.

7.1 Spring Models

The first model considered is the noisy spring, as in example 3.2.1. The spring force is modeled by considering the linker's chemical structure, with effective spring forces from the $C-C$ bonds calculated by their vibrational bond frequency and interpreting the linker as a series of springs with this spring constant.

As mentioned in example 3.2.1 as a linear SDE this process is a Gaussian process and hence will produce Gaussian inter-dye displacement distributions. Further, the model shows an exponential correlation decay, allowing for almost independent sampling from the steady-state Gaussian distribution during a burst time. These match assumptions that would be beneficial in a smFRET environment. Gaussian inter-dye displacements are a common assumption, moreover influences from noise are easier to manage if one can consider them uncorrelated to some degree. This provides a simple mechanical model. However, in general, the dye linker dynamics could not be considered as a single isotropic potential well. This produces a spherical equilibrium distribution about the mean vector $\mathbf{X}_{eq}$. Asymmetry in the distribution will lead to more of a dynamic shift. We also consider an anisotropic spring to account for possible anisotropic behavior in the dye linker spring force. The model is similar to the isotropic spring above. Still, the spring force matrix is not a scalar multiple of the identity but instead a diagonal matrix with different weights on the diagonal.

In other words the matrix K is no longer taken to be diagonal, but instead $K = [k_{i,j}]_{i,j=1}^{3}$ is a 3×3 matrix with positive coefficients representing the spring force in the i^{th} direction when pulled in the j^{th} direction. In this case, the stationary distribution is still a multivariate Gaussian distribution with mean X_{eq} and covariance Σ given by the solution to the Lyapunov equation

$$\omega K + K^T \omega = 2\sigma I_{3\times 3}. \tag{7.1}$$

Weakening the tether resistance in different coordinates introduces an asymmetry in the AV that is representative of the excluded volume introduced by the underlying molecule. The damping in the spring constant along certain directions will be proportional to the force imparted on the molecule via the underlying molecule, either by electronic interactions or thermodynamic forces. This way, certain directions may be more favorable and others more constrained.

This model will still produce Gaussian inter-dye displacement distributions with a more complicated covariance. This allows for greater flexibility in the dye motion while keeping the

original assumptions. Based on the matrix **K**, the stationary distribution can be altered to produce many different elliptic geometries; moreover, the spring constant matrix should depend on the dye linker composition, and thus, the coefficients should be readily found.

7.2 Elastic Pendulum

The second model we consider is the case of an elastic pendulum subject to noise. This model accounts for the greater angular flexibility of $C-C$ bonds versus their radial flexibility. This is done in a different manner than in the anisotropic spring model. Indeed, this model has a restoring force on the polar angle to the equilibrium angle where the line between the linker and the dye is perpendicular to the surface below. However, there is no restorative force on the azimuthal angle, allowing the dye to freely rotate around the linker axis. The model views the dye motion in spherical coordinates with coupled stochastic differential equations evolving on the angles and radius,

$$\begin{aligned} dr_t &= -k_r(r_t - r_{eq}) + \frac{1}{r_t}dt + \sigma_r \circ dB_t^r \\ d\theta_t &= -k_\theta \sin(\theta_t) + \frac{\sigma_\theta^2}{r_t^2 \tan\theta_t} \circ dB_t^\sigma \\ d\phi_t &= \frac{\sigma_\phi}{r_t \sin(\theta_t)} \circ dB_t^\phi \end{aligned}$$

where σ is the polar angle from the axis along the linker and attachment point, and ϕ is the azimuthal angle. Each component of the model is subject to a different amount of noise and the radial spring constant is different from the polar spring constant. By adjusting each of the parameters one can produce a motion similar to the classical wobble in a code model in which the dye linker rotates about the linker axis.

In our model, the linker will transition across the wobble stochastically as the polar component is subject to a force and random perturbations. Further, the radial component may change. Due to the interaction between Brownian motion on a sphere and the sphere's radius, a change in the radial component of the diffusion will result in changes in the angular parts. More specifically, if the radial com-

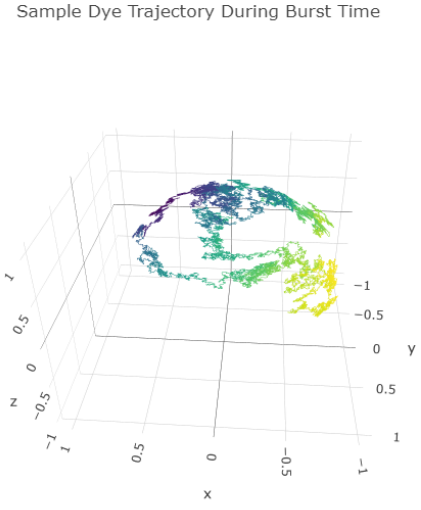


Figure 7.1: Sample path of the elastic pendulum model

ponent is particularly noisy, so will the angular components. The elastic pendulum model does not have Gaussian inter-dye displacements and, depending on the parameters will show dependence between excitation times. Still, as shown in section 8.1, these features are essential to producing an accurate model for dye motion.

7.3 κ^2 and Dipole Orientation Dynamics

The last section of this chapter considers the dynamics of the electric dipole moments of the two fluorescent dyes. As seen in example 6.1.1 FRET is sensitive to the orientational dynamics of the electric dipoles via the κ^2 parameter. Typically, this value is taken as its mean value in the case of isotropically distributed moments. However, the approach taken is the approach seen in example 6.1.1.

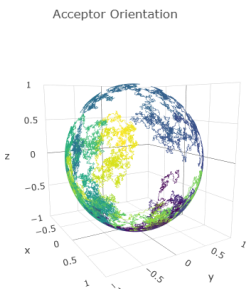


Figure 7.2: Example of a spherical Brownian motion

It is important to mention that for other shapes of molecules, the Brownian motion approximation will not be true, and more delicate considerations will need to be taken. Further, if the interaction between the rigid body orientation and its translational motion is considered, a new approach must be taken - the equivalent process on the sphere may not be a Brownian motion.

In light of the above, we only consider the electric dipole moments to evolve as spherical Brownian motions seen in example 6.1.1

$$d\theta_t = \frac{\sigma_\theta^2}{\tan\theta_t} \circ dB_t^\sigma$$
$$d\phi_t = \frac{\sigma_\phi}{\sin(\theta_t)} \circ dB_t^\phi$$

Using this motion the process

$$\kappa_t = \mu_t^A \cdot \mu_t^D - 3(\mu_t^A \cdot \hat{R})(\mu_t^D \cdot \hat{R})$$

to monitor the change of κ^2 over time. Then, using the κ^2 dynamics, the fluorescence CTMC representing FRET is no longer time-homogeneous, and the energy transfer rate is conditional on κ_t during the energy exchange process. The fluorescence CTMC no longer being time-homogeneous changes the simulation method as well.

Algorithm 4 Non-homogeneous FRET CTMC

Require: R_T, $\{\kappa_t\}$, k_D

Generate $X, Y \sim \mathcal{U}([0,1])$.

Define $\tau_X = \inf\{t > 0 : \int_0^t k_D \left(\frac{\kappa_t^2 \bar{R}_0}{R_T} dt \right) = X\}$ and $\tau_Y = \inf\{t > 0 : k_D t = Y\}$.

If $\tau_X > \tau_Y$ then $D \to F_D$ in FRET CTMC otherwise $D \to A$.

Using this algorithm an exploration of the effect of dipole moment dynamics on smFRET measurements can be conducted.

Chapter 8

Influence of Dye Motion on smFRET Dynamic Shift

"It doesn't stop being magic just because you know how it works."

Terry Prachett, Wee Free Men

8.1 Dye Models and Properties

Now to understand which dye model provides the best and most physically accurate picture of the underlying dynamics. To do this, we consider the dynamic shift induced by dye motion in a DNA standard experiment. In this type of experiment, the DNA can be taken as static along the burst time duration. Therefore the only dynamics seen will be those from the fluorescent dye motion.

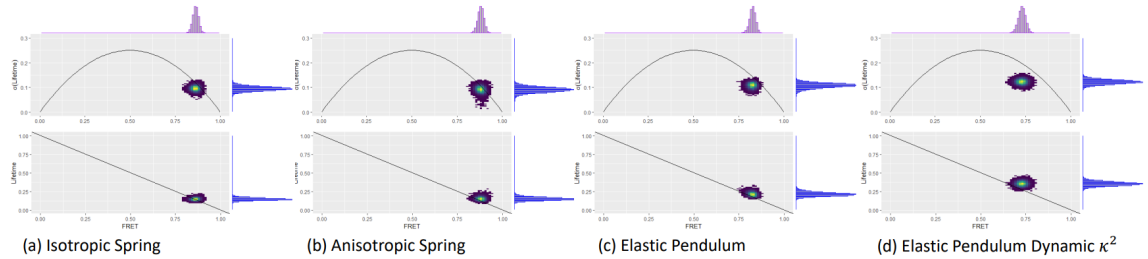


Figure 8.1: Comparison of dye models

From the above plot, one can notice that the isotropic spring shows the least amount of a dynamic shift, whereas the elastic pendulum with a dynamic κ shows the most dynamic shift. This is due to the fact that the isotropic spring model oscillates too frequently, providing an averaging effect during the burst time. From this the lifetime and FRET distributions are averaged in each burst, meaning that there is no real deviation between the average of the lifetimes or the lifetime of the averages. However, a dynamic shift can be seen to form in the anisotropic spring case, wherein the stationary distributions are ellipses with large cross-sectional areas in a plane and small perpendicular to that, like a coin. The stationary distributions are oriented, so the large areas are on perpendicular planes. This asymmetry and slower diffusion times provide the mixture needed for a dynamic shift. However, the volume of the stationary distribution can be quite large, with the longer axes being roughly 3 orders of magnitude greater than that of the shorter. The elastic pendulum model allows for enough freedom and slower times to produce a dynamic shift while keeping with appropriate volumes. However, the final push for an experimentally accurate dynamic shift comes from the inclusion of the κ^2 parameter.

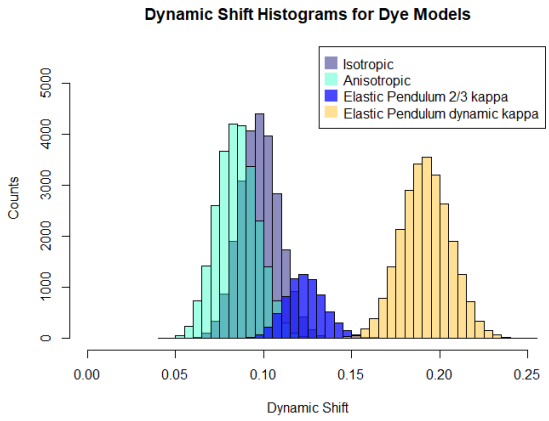


Figure 8.2: Dynamic shift histogram comparison between the different dye models

In the dynamic shift histogram above one can see that the elastic pendulum model with κ^2 dynamics included provides the highest dynamic shift. Moreover, it also provides the highest spread in the dynamic shift distribution.

8.1.1 Impact of Linker Length and Rigidity

In this section, we take the elastic pendulum model and compare different dye linker compositions to determine how the composition will influence the resulting dynamic shift. This is useful to understand so that appropriate linkers can be chosen for each experiment.

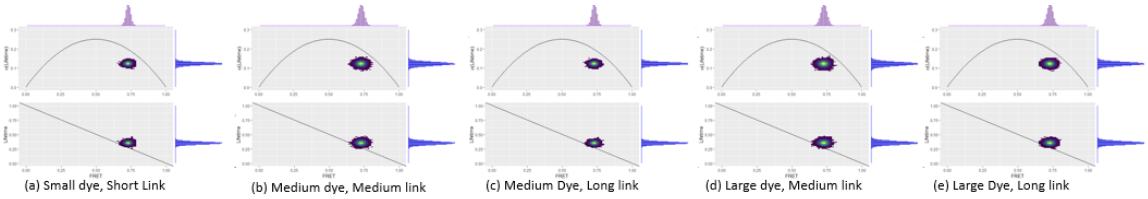


Figure 8.3: Dye linker compositions arraged from smallest dye size and shortest length to largest dye size and longest length

One can see that the dynamic shift in each example remains mostly the same; this is due to the dynamic shift from the dipole orientations since, at this stage, it is not linked to the linker composition.

TABLE II. Parameter Values for the Effective Spring Constant and Local Friction From Hydrodynamic Radius and Number of C-C Bonds in Linker

Number of Links	Hydrodynamic Radius
3	0.15 nm
7	0.3 nm
11	0.3 nm
7	1.1 nm
11	1.1 nm

It can be noted that the variance in the distribution does change depending on linker composition. When the dye is medium size with a long linker, the variance is smaller than with a short linker. This is due to the extra freedom of movement the long linker gives the dye. Note that this is not a freedom of greater area to explore, but a freedom in exploration speed. With the smaller spring constant from the longer linker and the lower drag from being a medium dye vs a large dye, this provides enough freedom to effectively lower the relaxation time and average out some of the noise. This is not the case in the large dye setting due to the drag on the dye overpowering the lowering of the spring constant.

8.1.2 The Role of Azimuthal Diffusion

Finally considering the influence of the spring constant in the dye linker composition above we consider the role of the Azimuthal diffusion of the dyes. This is due to the fact that the dye is most free to move along the azimuthal coordinate; there is no force there. From the speed of the wobble along this coordinate, averaging could take place. If the dye wobbles rapidly during a burst time, then the distance between the dyes will average out to be the equilibrium distance and reduce

the dynamic shift or variance of the distribution.[1]

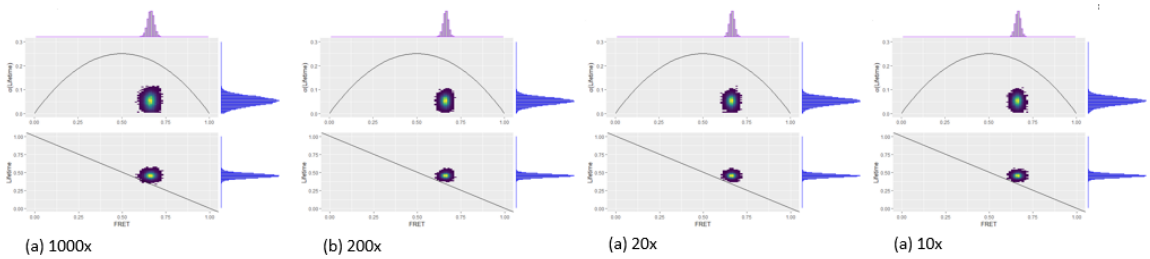


Figure 8.4: Azimuthal diffusion relationships

One can see that when the angular diffusion is increased, there is a slight increase in the dynamic shift, but the moment difference plot shows the most change in behavior. Notice that the variance is greatly increased in the very fast angular diffusion. However, at slower diffusion, the variance does not change much depending on the parameter. This result is most likely due to the parallel lowering of the polar spring constant. In this plot, we are seeing the results from having complete angular freedom rather than just azimuthal freedom.

[1]It should be noted in the above plot that the same factor also reduces the polar spring constant due to an old simulation method. A newer plot is being prepared.

Chapter 9

Large Deviations Theory

"THAT'S MORTALS FOR YOU, Death continued. THEY'VE ONLY GOT A FEW YEARS IN THIS WORLD AND THEY SPEND THEM ALL IN MAKING THINGS COMPLICATED FOR THEMSELVES. FASCINATING."

Terry Pratchett, Mort

Recall the definition of a rate function and a large deviation principle from section 3.1.3. The conditions given in (3.2) are similar to conditions in the Portmanteau theorem from the theory of weak convergence. [7] A third condition states that a sequence of random elements $\{X_n\}$ converges in distribution to a random element X if and only if for every bounded continuous function f, $E[f(X_n)] \to E[f(X)]$. The approach taken by Feng and Kurtz, as well as this dissertation begins with a large deviation version of this result known as the Bryc formula. Building on the Bryc formula, Feng and Kurtz develop a methodology for developing large deviations results for a broad class of Markov processes.

Originally, techniques in large deviations theory involved change of measure techniques combined with weak convergence results, such as in [19][29]. In this work, we consider processes on Riemannian manifolds and we wish to consider these objects globally. To date there is no nice transformation for diffusion-type processes on general Riemann manifolds, only for compact manifolds as seen in [68]. This limitation leads us to consider a different route. Namely, the route of nonlinear semigroup convergence that was introduced in [23]. As shown in [38] this approach greatly simplifies the task of analyzing LDP on Riemann manifolds.

Proving a large deviation principle for Markov processes is similar in spirit to showing the convergence of a sequence of Markov processes $\{X_t^{(n)}\}$ to a Markov process X_t. To show the latter, let A_n be the generator of $X_t^{(n)}$ and let A be that of X_t. Then A_n generates the semigroup T_t^n, where $T_t^n f(x) = E_x[f(X_t^{(n)})]$ and similarly for A. If A_n converges to A, then one can show that T_t^n converges to T_t. And hence one has $E_x[f(X_t^{(n)})] \to E_x[f(X_t)]$. Thus, for fixed t, $X_t^{(n)}$ converges in distribution to X_t. Convergence in distribution of the finite-dimensional distributions will follow from the semigroup or Markov property. Convergence in distribution of the processes follows once the sequence of processes $\{X_t^{(n)}\}$ is shown to be tight.

In demonstrating the former, one starts with a sequence of nonlinear dissipative operators

$$H_n f = \frac{1}{n} e^{-nf} \mathcal{A}_n e^{nf}.$$

We refer to this generator as the **exponential tilt** of A_n.[1] Each H_n generates a nonlinear semigroup

$$V_t^n f(x) = \frac{1}{n} \ln(\mathbb{E}(e^{nf(X_t)})|X_0 = x) = \frac{1}{n} \ln(T_t e^{nf}), \quad f \in B(\mathcal{M}).$$

Suppose that there exists a nonlinear operator H for which H_n converges to H. At this point, one would like to proceed as in the case of convergence in distribution. First, we would like H to be the generator of a semigroup $V(t)$. If $V^n t$ converges to V_t, one almost has the Bryc formula and a one-dimensional LDP result. From here one can extend to a finite dimensional LDP and then a functional LDP.

While all of this seems reasonable enough it is fraught with difficulties. The operator H may not have a domain that is large enough to generate the necessary limiting semigroup. For this to be the case, H must satisfy a range condition: there exists an $\alpha_0 > 0$ such that

$$D(H) \subset R(I - \alpha H)$$

for all $0 < \alpha < \alpha_0$. Then, from the Crandall - Liggett Theorem [15] the operator H will generate a nonlinear semigroup and one can show that the semigroups generated by H_n will converge to this

[1] This is because of the similarity between martingale exponential tilting and the nonlinear semigroup defined above.

semigroup. [23] However, this requires solutions to the equation

$$(I - \lambda H)f = g$$

which in general require regularity results that are not available. [23] Due to this concern we consider a weaker notion of solution to this equation. Namely, a viscosity solution and a comparison principle. Having $V_t^n f(x)$ converge to $V_t f(x)$ is not enough for the Bryc formula to hold.

Definition 9.0.1. *A sequence of probability measures $\mathbb{P}_n$ is called* ***exponentially tight****[23] if for each $a > 0$ there exists a compact set $K_a \subset \mathcal{M}$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \ln(\mathbb{P}_n(K_a^c)) \leq -a.$$

One can proceed to develop finite dimensional LDP and finally function. and then functional LDP. However, the obtained rate function will typically not be in a convenient form. At this point, some ideas from control theory are used to put the rate function in a more tractable form.

Feng and Kurtz [23] demonstrate how to overcome these issues and present a roughly four-part approach to finding to LDPs breaks:

- Verify the convergence of a sequence of nonlinear operators H_n to an operator H.
- Verify exponential tightness of the sequence of measures.
- Verify the comparison principle for viscosity solutions of the Hamiltion-Jacobi equation for the limiting operator H.
- Construct a variational representation of H and find the rate function by the Legendre-Fenchel transform of H.

9.0.1 Viscosity Solutions and Stochastic Optimal Control

Viscosity solutions are weak solutions to operator equations and are mainly applied to nonlinear equations. [14] These types of solutions have appeared in the context of calculus of variations and control theory. [24][22] These results were only applied to nonlinear differential operators, but in [23], this notion of solution was generalized for general operators.

Definition 9.0.2. *Viscosity Solution Let $H : D(H) \subset C(\mathcal{M}) \to B(\mathcal{M})$. Fix $h \in C(\mathcal{M})$, and $\alpha > 0$. Let $f \in B(\mathcal{M})$ and consider the equation*

$$(I - \lambda H)f = h. \tag{9.1}$$

- *f is a **viscosity subsolution** of (9.1) if and only if f is upper semicontinuous and for each $f_0 \in D(H)$ there exists $x_0 \in \mathcal{M}$ such that*

$$f(x_0) - f_0(x_0) = \|f - f_0\|_\infty$$

and

$$\frac{1}{\lambda}(f(x_0) - \lambda H f(x_0) - h(x_0)) \leq 0$$

- *f is a **viscosity supersolution**of (9.1) if and only if f is upper semicontinuous and for each $f_0 \in D(H)$ there exists $x_0 \in \mathcal{M}$ such that*

$$f_0(x) - f(x_0) = \|f - f_0\|_\infty$$

and

$$\frac{1}{\lambda}(f(x_0) - \lambda H f(x_0) - h(x_0)) \geq 0$$

*A function f is a **viscosity solution** if it is both a subsolution and a supersolution.*

To provide a unique viscosity solution knowing the subsolutions and supersolutions one verifies the comparison principle.

Definition 9.0.3. *The Hamilton - Jacobi equation (9.1) satisfies the **comparison principle** if $\overline{f}$ is a subsolution and $\underline{f}$ is a supersolution implies $\overline{f} \leq \underline{f}$.[23]*

As we have seen in the outline, if the sequence of measures is exponentially tight, and the sequence of generators converge to a limiting operator that satisfies the comparison principle then the sequence of processes satisfies an LDP.[23] What remains is to find a representation for the rate function of this LDP. This can be done by considering a variational form for the limiting operator H.

In many cases one can show that this operator is the generator for a Nisio semigroup, a semigroup arising in stochastic control theory. [46][23] If we have this representation we are able to represent the rate function in terms of the Legendre - Fenchel transform of the variational form of H. The variational form of H, in this work, is a function $\mathcal{H} : T^*\mathcal{M} \to \mathbb{R}$ such that $\mathcal{H}$ is continuous and fixing $x \in \mathcal{M}$ the map $\mathcal{H} : T_x^*\mathcal{M} \to \mathbb{R}$ is convex and for all $x \in \mathcal{M}$ and $f \in C_c^\infty(\mathcal{M})$ we have $Hf(x) = \mathcal{H}(x, df(x))$.[38] We often refer to the variational form of H as the **Hamiltonian** in reference to its connection to statistical mechanics.

The main result used is the following[38][63][23],

Theorem 9.0.1. *Let $\{\mathbb{P}_n\}$ be a sequence of measures on $D([0,\infty), \mathcal{M})$. Suppose this sequence is exponentially tight and for each n, there exists an operator A_n that satisfies the martingale problem for $\mathbb{P}_n$. Define $H_n f = \frac{1}{n} e^{-nf} A_n e^{nf}$ and suppose $H_n \to H$ uniformly on compact sets. Further assume for each $\lambda > 0$ and $h \in C_b(\mathcal{M})$ the comparison principle is satisfied for $(I - \lambda H)f = h$. Then $\mathbb{P}_n$ satisfies an LDP in $D([0,\infty), \mathcal{M})$ with good rate function*

$$I(\gamma) = \int_0^\infty \mathcal{L}(\gamma_t, \dot{\gamma}_t) dt \tag{9.2}$$

for $\gamma \in H_0^1(\mathcal{M})$ and ∞ otherwise. Where $\mathcal{L} : T\mathcal{M} \to [0,\infty]$ is given by

$$\mathcal{L}(x, v) = \sup_{p \in T_x^*\mathcal{M}} \{p(v) - \mathcal{H}(x, p)\}$$

is the Lengendre - Fenchel transform of the variational form of H.

The function $\mathcal{L}$ is sometimes referred to as a **Lagrangian**.

9.0.2 Useful Facts

Now we will discuss some results that simplify the task of determining exponential tightness and the comparision principle for processes on manifolds considerably. The following two propositions will be instrumental to our task.

First we define a compact containment function.

Definition 9.0.4. *First we define a compact containment function. A **good containment function** is a function $\Upsilon : \mathcal{M} \to \mathbb{R}_0^+$ satisfying the first three properties below and a **good containment function for** H if it satisfies all of the properties.*

- *There exists a point* $x_0 \in \mathcal{M}$ *such that* $\Upsilon(x_0) = 0$.
- $\Upsilon \in C^2(\mathcal{M})$.
- $\forall c \in \mathbb{R}_0^+$ *the set* $\{x \in \mathcal{M} : \Upsilon(x) \leq c\}$ *is compact.*
- $\sup_{z \in \mathcal{M}}(\mathcal{H}(z, d\Upsilon(z))) < \infty$.

It is shown in [38][39] that if H has a good containment function then the sequence $\mathbb{P}_n$ is exponentially tight. Further, this containment function is used in the following condition to show the comparision principle. This is again shown in [38][39].

Proposition 9.0.1. *If* H *has a good containment function that satisfies the following condition, then the viscosity solutions satisfy the comparison principle.*

Let $\overline{f}$ *and* $\underline{f}$ *be sub and super viscosity solutions to* $(I - \lambda H)f = h$ *for fixed* $\lambda > 0$ *and* $h \in C_b(\mathcal{M})$ *and let* Υ *be a good containment function for* H. *Define the sequences* $\{y_{\alpha,\epsilon}\}$ *and* $\{x_{\alpha,\epsilon}\}$ *for* $\alpha, \epsilon > 0$ *by the following*

$$\begin{aligned} &\frac{\overline{f}(x_{\alpha,\epsilon})}{1-\epsilon} - \frac{\underline{f}(y_{\alpha,\epsilon})}{1+\epsilon} - \frac{\alpha}{2}d(x_{\alpha,\epsilon}, y_{\alpha,\epsilon}) - \frac{\epsilon}{1-\epsilon}\Upsilon(x_{\alpha,\epsilon}) - \frac{\epsilon}{1+\epsilon}\Upsilon(y_{\alpha,\epsilon}) \\ &\quad = \sup_{x,y \in \mathcal{M}} \left(\frac{\overline{f}(x)}{1-\epsilon} - \frac{\underline{f}(y)}{1+\epsilon} - \frac{\alpha}{2}d(x,y) - \frac{\epsilon}{1-\epsilon}\Upsilon(x) - \frac{\epsilon}{1+\epsilon}\Upsilon(y) \right) \end{aligned} \tag{9.3}$$

Then, if

$$\liminf_{\epsilon \to 0} \liminf_{\alpha \to \infty} \left\{ \mathcal{H}\left(x_{\alpha,\epsilon}, \frac{\alpha}{2} d(d^2(\cdot, y_{\alpha,\epsilon})(x_{\alpha,\epsilon})) \right) - \mathcal{H}\left(y_{\alpha,\epsilon}, \frac{-\alpha}{2} d(d^2(x_{\alpha,\epsilon}, \cdot)(y_{\alpha,\epsilon}) \right) \right\} \leq 0 \tag{9.4}$$

then H *satisfies the comparison principle.*

This is a modification of conditions 9.10 and 9.11 for verifying the comparison principle in Feng and Kurtz [23] adapted to manifolds. These conditions are based upon Lemma 9.3 in Feng and Kurtz [23] where it is essentially shown that

$$\begin{aligned} &\sup_{x,y \in \mathcal{M}} \left(\frac{\overline{f}(x)}{1-\epsilon} - \frac{\underline{f}(y)}{1+\epsilon} - \frac{\alpha}{2}d(x,y) - \frac{\epsilon}{1-\epsilon}\Upsilon(x) - \frac{\epsilon}{1+\epsilon}\Upsilon(y) \right) \\ &\leq \liminf_{\alpha \to \infty} \left\{ \mathcal{H}\left(x_{\alpha,\epsilon}, \frac{\alpha}{2} d(d^2(\cdot, y_{\alpha,\epsilon})(x_{\alpha,\epsilon})) \right) - \mathcal{H}\left(y_{\alpha,\epsilon}, \frac{-\alpha}{2} d(d^2(x_{\alpha,\epsilon}, \cdot)(y_{\alpha,\epsilon}) \right) \right\}. \end{aligned}$$

Letting $\epsilon \to 0$ and applying (9.3) and (9.4) implies that

$$\overline{f} - \underline{f} \leq 0$$

which is the comparison principle. A similar condition can be found in [39] where the penalizing function is introduced. Here the penalizing function is the Riemannian distance. By [38][39], we know that $d(x_{\alpha,\epsilon}, y_{\alpha,\epsilon}) \to 0$ as $\alpha \to \infty$. It will become of interest to know how fast this happens.

Proposition 9.0.2.

$$d(x_{\alpha,\epsilon}, y_{\alpha,\epsilon}) \leq \frac{C}{\alpha}$$

for any $\epsilon > 0$ and a constant $C \in \mathbb{R}_0^+$.

Proof. To see the above, let $y \in \mathcal{M}$ such that $\Upsilon(y) = 0$ and $0 < \epsilon < 1$ define

$$\begin{aligned} L(y) &= \frac{\overline{f}(y)}{1-\epsilon} - \frac{\underline{f}(y)}{1+\epsilon} - \frac{\epsilon}{1-\epsilon}\Upsilon(y) - \frac{\epsilon}{1+\epsilon}\Upsilon(y) \\ &= \frac{\overline{f}(y)}{1-\epsilon} - \frac{\underline{f}(y)}{1+\epsilon} \end{aligned}$$

Then, since $L(y)$ is a special case of the function

$$\frac{\overline{f}(x)}{1-\epsilon} - \frac{\underline{f}(y)}{1+\epsilon} - \frac{\alpha}{2}d(x,y) - \frac{\epsilon}{1-\epsilon}\Upsilon(x) - \frac{\epsilon}{1+\epsilon}\Upsilon(y)$$

when $x = y$ it must be that the supremum value attained at $x_{\alpha,\epsilon}$ and $y_{\alpha,\epsilon}$ is greater than $L(y)$. Therefore

$$\begin{aligned} L(y) &\leq \frac{\overline{f}(x_{\alpha,\epsilon})}{1-\epsilon} - \frac{\underline{f}(y_{\alpha,\epsilon})}{1+\epsilon} - \frac{\alpha}{2}d(x_{\alpha,\epsilon}, y_{\alpha,\epsilon}) - \frac{\epsilon}{1-\epsilon}\Upsilon(x_{\alpha,\epsilon}) - \frac{\epsilon}{1+\epsilon}\Upsilon(y_{\alpha,\epsilon}) \\ &\to \frac{\alpha}{2}d(x_{\alpha,\epsilon}, y_{\alpha,\epsilon}) \leq \frac{\overline{f}(x_{\alpha,\epsilon})}{1-\epsilon} - \frac{\underline{f}(y_{\alpha,\epsilon})}{1+\epsilon} - \frac{\epsilon}{1-\epsilon}\Upsilon(x_{\alpha,\epsilon}) - \frac{\epsilon}{1+\epsilon}\Upsilon(y_{\alpha,\epsilon}) - L(y) \end{aligned}$$

Since sub and super viscosity solutions are bounded and the distance cannot be less than 0, the sequences $\Upsilon(x_{\alpha,\epsilon})$ and $\Upsilon(y_{\alpha,\epsilon})$ are also bounded one can take the supremum of the right hand side over alpha and calling this value $\frac{C}{2}$ we see that

$$d(x_{\alpha,\epsilon}, y_{\alpha,\epsilon}) \leq \frac{C}{\alpha}$$

hence as $\alpha \to \infty$ the sequences of distances must also go to 0 at a rate proportional to $\frac{1}{\alpha}$. $\square$

Chapter 10

Fluctuations of Vanishing Markov Perturbations

"Questions don't have to make sense, Vincent," said Miss Susan. "But answers do."

Terry Pratchett, Theif of Time

Let $(\mathcal{M}, g)$ be a d dimensional Riemannian manifold. First, we will define several required objects. Let $c : \mathcal{M} \to [0, \infty)$ be a measurable function, $b = (b_i)_{i=1}^d \in \mathbb{R}^d$, $\sigma = (\sigma_{i,j})_{i,j=1}^d$ be a positive semi-definite $d \times d$ matrix, and $\mathbf{V} = \{V_k\}_{k=0}^d$ a collection of complete vector fields on $\mathcal{M}$ such that any linear combination of the vector fields is also complete. Further, we will make use of the exponential map from complete vector fields to diffeomorphisms defined by

$$Exp(V)(x) = \gamma_V(x, 1)$$

where $\gamma_V(x, t)$ is the integral curve associated to the one-parameter semigroup of diffeomorphisms generated by the vector field V starting at $x \in \mathcal{M}$ and evaluated at time $t \in [0, T]$. To simplify notation, we denote

$$Exp\bigg(\sum_{i=1}^d y_i V_i\bigg)(x) = \gamma(y, \mathbb{V})(x)$$

for vectors $y \in \mathbb{R}$ and collections of complete vector fields $\mathbf{V}$, in a similar way to Lee and Kunita. [44, 41] Finally, let η be a Lévy kernel such that for each $x \in \mathcal{M}$, $\eta(x, \cdot)$ is a Lévy measure on $\mathbb{R}^d$.

Consider a large deviation principle for a sequence of Feller processes $X_n(t)$ such that the

generator for $X_n(t)$ is given by

$$A_n f(x) = -c(x) + V_0(f)(x) + \sum_{i=1}^{d} b_i V_i(f)(x) + \frac{1}{2n}\sum_{i,j=1}^{d} \sigma_{i,j} V_i(V_j(f))(x)$$
$$+n\int_{\mathbb{R}^d} \big[f(\gamma(z, n^{-1}\mathbf{V})(x)) - f(x) - \frac{1}{n} I_{B_1}(z)\sum_{i=1}^{d} z_i V_i(f)(x)\big]\eta(x, dz).$$

Notice that in the Euclidean case, this is equivalent to the central limit scaling $\frac{1}{n}X(nt)$. However, since Markov processes generally do not have a uniform scaling property, there is no direct analog for this type of space-time scaling in the general Riemannian case. As we will see in later sections, specific cases such as Brownian motion, Riemannian valued Marcus SDEs, or geodesic random walks will allow for a process-level scaling interpretation. Usually, this scaling will lead to a deterministic limit and be interpreted as a dynamical system perturbed by a general Markov process.

Theorem 10.0.1. *Let X_n be a sequence of Markov processes generated by the operators A_n such that for any $g \in C^\infty(\mathcal{M})$ such that $r_0(x)$ we have $\|g - r\| \leq 1$ and $|dg| < 2$ we have*

$$gV_j(g)(x) \leq K_j(1 + g^2(x)),$$

for all $x \in \mathcal{M}$ and $j = 0, 1, ..., d$. Further, let $\eta(x, dz)$ be a Lévy measure on $\mathbb{R}^d$ indexed by $\mathcal{M}$ such that there exists a measure μ on $\mathbb{R}^d$ such that for any $x \in \mathcal{M}$ $\eta(x, dz) = K(x, z)\mu(dz)$ with $K(x, \cdot)$ continuous on $\mathcal{M}$. Further, for any $x \in \mathcal{M}$ and $c \in \mathbb{R}^d$ the Lévy measures satisfy

$$\int_{\mathbb{R}^d} \exp\big(\sum_{i=1}^{d} c_i z_i\big)\eta(x, dz) < \infty.$$

. Then, X_n satisfied a LDP with good rate function

$$I(\gamma) = \int_0^\infty \mathcal{L}(\gamma(t), \dot{\gamma}(t))dt$$

We will prove the result as a consequence of several lemmas.

Lemma 10.0.1 (Operator Convergence)**.** *The sequence of operators $H_n f = \frac{1}{n} e^{-nf}\mathcal{A}_n e^{nf}$ converges*

to the operator

$$Hf(x) = V_0(f)(x) + \sum_{i=1}^{d} b_i V_i(f)(x) + \frac{1}{2}\sum_{i,j=1}^{d} \sigma_{i,j} V_i(f)(x) V_j(f)(x) + \int_{\mathbb{R}^d} \Big[\exp\big(\sum_{i=1}^{d} z_i V_i(f)(x)\big) - 1 - I_{B_1}(z) \sum_{i=1}^{d} z_i V_i(f)(x) \Big] \eta(x, dz)$$

in the buc - LIM sense.

Proof. Let $H_n f = \frac{1}{n} e^{-nf} A_n e^{nf}$. Then,

$$\begin{aligned} H_n f(x) &= -\frac{c(x)}{n} + \frac{1}{n} e^{-nf(x)} \big(V_0(e^{nf(x)}) + \sum_{i=1}^{d} b_i V_i(e^{nf(x)})\big) + \frac{e^{-nf(x)}}{2n^2} \sum_{i,j=1}^{d} \sigma_{i,j} V_i(V_j(e^{nf(x)})) \\ &+ \frac{1}{n} \int_{\mathbb{R}^d} \Big[\exp(nf(\gamma(z, n^{-1}\mathbb{V}) - nf(x)) - 1 - \frac{e^{-nf(x)}}{n} I_{B_1}(z) \sum_{i=1}^{d} z_i V_i(e^n f(x)) \Big] \eta(x, dz) \\ &= \frac{-c(x)}{n} + V_0(f)(x) + \sum_{i=1}^{d} b_i V_i(f)(x) + \frac{1}{2} \sum_{i,j=1}^{d} \Big(\sigma_{i,j} V_i(f)(x) V_j(f)(x) + \frac{1}{n} V_i(V_j(f))(x) \Big) \\ &+ \int_{\mathbb{R}^d} \Big[\exp(n(f(\gamma(z, n^{-1}\mathbb{V}) - f(x))) - 1 - I_{B_1}(z) \sum_{i=1}^{d} z_i V_i(f(x)) \Big] \eta(x, dz) \end{aligned}$$

Now we show $\lim_{n\to\infty} H_n f(x) = Hf(x)$ where

$$Hf(x) = V_0(f)(x) + \sum_{i=1}^{d} b_i V_i(f)(x) + \frac{1}{2}\sum_{i,j=1}^{d} \sigma_{i,j} V_i(f)(x) V_j(f)(x) + \int_{\mathbb{R}^d} \Big[\exp\big(\sum_{i=1}^{d} z_i V_i(f)(x)\big) - 1 - I_{B_1}(z) \sum_{i=1}^{d} z_i V_i(f)(x) \Big] \eta(x, dz)$$

□

Lemma 10.0.2 (Variational Form and Compact Containment)**.** *The operator H as above has a variational form $\mathcal{H} : T\mathcal{M} \to \mathbb{R}$ such that for each f, $Hf(x) = \mathcal{H}(x, df)$ and further, the function $\Upsilon(x) = \ln(1 + g^2(x))$ is a good containment function.*

Proof. For clarity, we will show that the function is a good containment function in two parts. The

first for the continuous part of the generator and the second for the discontinuous part.

$$\begin{aligned}\mathcal{H}_c(x, d\Upsilon) &= d\Upsilon(V_0)(x) + \sum_{i=1}^{d} b_i d\Upsilon(V_i)(x) + \frac{1}{2}\sum_{i=1}^{d} \sigma_{i,j} d\Upsilon(V_i)(x) d\Upsilon(V_j)(x) \\ &= \frac{2g(x)}{1+g^2(x)} V_0(g)(x) + \sum_{i=1}^{d} b_i \left(\frac{2g(x)}{1+g^2(x)} V_i(g)(x) \right) \\ &+ \frac{1}{2}\sum_{i=1}^{d} \sigma_{i,j} \left(\frac{g(x)}{1+g^2(x)} V_i(g)(x) \right) \left(\frac{g(x)}{1+g^2(x)} V_j(g)(x) \right) \\ &\leq K_0 + \sum_{i=1}^{d} b_i K_i + \frac{1}{2}\sum_{i=1}^{d} \sigma_{i,j}(x) K_i K_j < \infty\end{aligned}$$

by the assumptions on the vector fields V_i. For the discontinuous part,

$$\begin{aligned}\mathcal{H}_J(x, d\Upsilon) &= \int_{\mathbb{R}^d} \left[\exp\big(\sum_{i=1}^{d} z_i d\Upsilon(V_i) \big) - 1 - I_{B_1}(z) (\sum_{i=1}^{d} z_i d\Upsilon(V_i) \right] \eta(x, dz) \\ &\leq \int_{\mathbb{R}^d} \exp(\sum_{i=1} z_i K_i) \eta(x, dz) < \infty\end{aligned}$$

Therefore, Υ is a good containment function for H, and so the sequence of measures is exponentially tight.

□

Lemma 10.0.3 (Comparison Principle)**.** *The operator H satisfies the comparision principle for viscosity solutions.*

Proof. This is done by showing $\mathcal{H}$ satisfies proposition 12.0.1 Since $\mathcal{H}(x,p) = \mathcal{H}_c(x,p) + \mathcal{H}_J(x,p)$ we will show proposition 2.2 for each operator separately.

Using the fact that

$$\tau_{yx} d(d^2(x,\cdot))(y) = -d(d^2(\cdot, y))(x)$$

and vice versa we can see the following,

$$\begin{aligned}
&\mathcal{H}_c\Big(x, \frac{\alpha}{2}d(d^2(\cdot,y)(x))\Big) - \mathcal{H}_c\Big(y, \frac{-\alpha}{2}d(d^2(x,\cdot)(y)\Big) \\
&\leq \frac{\alpha}{2}d^2(x,y)|V_0(x) - \tau_{yx}V_0(y)|^2 - \frac{\alpha^2}{8}d^2(x,y)\sum_{i=1}^{d}|(|V_i(x)|^2 - |\tau_{yx}V_i(y)|^2) \\
&= \frac{\alpha}{2}d^2(x,y)|V_0(x) - \tau_{yx}V_0(y)|^2 - \frac{\alpha^2}{8}d^2(x,y)\sum_{i=1}^{d}(|V_i(x)| - |\tau_{yx}V_i(y)|)(|V_i(x)| + |\tau_{yx}V_i(y)|) \\
&\leq \frac{\alpha}{2}d(x,y)|V_0(x) - \tau_{yx}V_0(y)| - \frac{\alpha^2}{8}d^2(x,y)\sum_{i=1}^{d}(|V_i(x) - \tau_{yx}V_i(y)|)(|V_i(x)| + |\tau_{yx}V_i(y)|)
\end{aligned}$$

Where the second line denotes the use of the Cauchy-Schwarz inequality and the combination of terms from the parallel transport. Note that $V_i(x) \in T_X\mathcal{M}$ as the evaluation of the vector field V_i at $x \in \mathcal{M}$. So, since each $V_j,\ j = 0, 1, ..., d$ is smooth and $d(x_\alpha, y_\alpha) \to 0$ as in proposition 9.0.1 and $\alpha d(x_\alpha, y_\alpha) \to C$ as $\alpha \to \infty$, we have that the above also goes to 0 as $\alpha \to \infty$.

Now,

$$\begin{aligned}
&\mathcal{H}_{\mathcal{J}}\Big(x, \frac{\alpha}{2}d(d^2(\cdot,y)(x))\Big) - \mathcal{H}_J\Big(y, \frac{-\alpha}{2}d(d^2(x,\cdot)(y)\Big) \\
&= \int_{\mathbb{R}^d}\Big[\exp\big(\frac{\alpha}{2}\sum_{i=1}^{d} z_i d(d^2(\cdot,y)(V_i)(x)\big) - 1 - I_{B_1}(z)(\frac{\alpha}{2}\sum_{i=1}^{d} b_i d(d^2(\cdot,y)(V_i)(x)\Big]\eta(x,dz) \\
&- \int_{\mathbb{R}^d}\Big[\exp\big(-\frac{\alpha}{2}\sum_{i=1}^{d} z_i d(d^2(x,\cdot)(V_i)(y)\big) - 1 + I_{B_1}(z)(\frac{\alpha}{2}\sum_{i=1}^{d} b_i d(d^2(x,\cdot)(V_i)(y)\Big]\eta(y,dz) \\
&= \int_{\mathbb{R}^d}\Big[\exp\big(\frac{\alpha}{2}\sum_{i=1}^{d} z_i d(d^2(\cdot,y)(V_i)(x)\big) - 1 - I_{B_1}(z)(\frac{\alpha}{2}\sum_{i=1}^{d} b_i d(d^2(\cdot,y)(V_i)(x)\Big]\eta(x,dz) \\
&- \int_{\mathbb{R}^d}\Big[\exp\big(\frac{\alpha}{2}\sum_{i=1}^{d} z_i d(d^2(\cdot,y)(V_i)(x)\big) - 1 - I_{B_1}(z)(\frac{\alpha}{2}\sum_{i=1}^{d} b_i d(d^2(\cdot,y)(V_i)(x)\Big]\eta(y,dz)
\end{aligned}$$

Now substituting $\eta(\cdot, dz)$ with $K(\cdot, z)\mu(dz)$ and factoring the expression we obtain,

$$\begin{aligned}
&\int_{\mathbb{R}^d}\Big[\exp\big(\frac{\alpha}{2}\sum_{i=1}^{d} z_i d(d^2(\cdot,y)(V_i)(x)\big) - 1 - I_{B_1}(z)(\frac{\alpha}{2}\sum_{i=1}^{d} b_i d(d^2(\cdot,y)(V_i)(x)\Big](K(x,z) - K(y,z))\mu(dz) \\
&\leq \int_{\mathbb{R}^d}\Big[\exp\big(\frac{C}{2}\sum_{i=1}^{d} z_i\|V_i\|_{g(x)}\big) - 1 - I_{B_1}(z)(\frac{C}{2}\sum_{i=1}^{d} b_i\|V_i\|_{g(x)})\Big](K(x,z) - K(y,z))\mu(dz).
\end{aligned}$$

Letting $d(x,y) \to 0$ using dominated convergence and noting that $K(\cdot, z)$ is continuous, we see that the integral above converges to 0. Consequently, the comparison principle holds. $\square$

Proof of Theorem. The result follows from combining the three lemmas above and theorem 9.0.1. □

10.1 A Corollary

"Coming back to where you started is not the same as never leaving."

Terry Pratchett, A Hat Full of Sky

When investigating the variational form $\mathcal{H}$ of the operator H in the theorem one can notice that there is a similarity between the H operator and the symbol of a Markov process in $\mathbb{R}^d$. Recall that the generator of any Markov process in $\mathbb{R}^d$ can be expressed as a psuedodifferential operator,

$$Ag(x) = \int_{\mathbb{R}^d} e^{i\langle x,\xi\rangle}\phi(\xi)\hat{g}(\xi)d\xi$$

with a symbol,

$$\psi(\xi) = i\sum_{i=1}^{d} b_i\xi_i + \frac{1}{2}\sum_{i,j=1}^{d}\sigma_{i,j}\xi_i\xi_j + \int_{\mathbb{R}^d}\left[\exp\big(i\sum_{i=1}^{d} z_i\xi_i\big) - 1 - iI_{B_1}(z)\sum_{i=1}^{d} z_i\xi_i\right]\eta(\xi, dz)$$

Using the fact that the above theorem requires that the integral

$$\int_{\mathbb{R}^d}\left[\exp\big(\sum_{i=1}^{d} z_i\xi_i\big) - 1 - I_{B_1}(z)\sum_{i=1}^{d} z_i\xi_i\right]\eta(\xi, dz) < \infty$$

then we know that $\psi(\xi)$ exists for $\psi \in i\mathbb{R}^d$. Hence, by defining,

$$\mathbb{V}(f)(x)_i = (V_i(f)(x)),\ i = 1, ..., d$$

we can express the variational form of the operator H as

$$\mathcal{H}(x, df) = \psi(-i\mathbb{V}(f)(x))$$

10.2 Examples

"When you really need them the most," he said, "million-to-one chances always crop up. Well-known fact."

Terry Pratchett, Guards! Guards!

10.2.1 Brownian Motion

The generator for a Brownian motion is given by the Laplace Beltrami operator. In $\mathbb{R}^d$ with inner product given by the Riemann metric, the Laplace-Beltrami operator can be expressed, at $x \in \mathcal{M}$ as a pseudodifferential operator with symbol $\|\xi\|^2_{g(x)}$. Therefore, the Lagrangian is given by the Legendre-Fenchel transform

$$\begin{aligned}\mathcal{L}(x,q) &= \sup_{p \in T^*_x \mathcal{M}} \{p(q) - \mathcal{H}(x,p)\} \\ &= \sup_{p \in T^*_x \mathcal{M}} \{p(q) - \|p\|^2_{g(x)}\} \\ &= \|q\|^2_{g(x)}\end{aligned}$$

which then gives us Schilder's theorem with rate function

$$I(\gamma) = \int_0^\infty \|\dot{\gamma}_t\|^2_{g(\gamma_t)} dt$$

10.2.2 Stochastic Differential equations

Consider the sequence of SDEs

$$dX_t = V_0(X_t)dt + \frac{1}{\sqrt{n}} \sum_{i=1}^d V_i(X_t) \circ dW_t$$

we know that the symbol for the generator for such a process will be given by $\phi(\xi) = i\langle V_0, \xi \rangle_{g(x)} - \sum_{i=1}^d \langle V_i, \sigma_{i,j} V_j \rangle$. Hence the variational form is

$$\mathcal{H}(x, df) = \psi(-i\mathbb{V}(f)(x)) = V_0(f)(x) - \sum_{i,j=1}^d \sigma_{i,j} V_i(f)(x) V_j(f)(x)$$

To finally calculate the rate function

$$I(\gamma) = \int_0^\infty \mathcal{L}(\gamma_t, \dot{\gamma}_t)dt$$

where

$$\mathcal{L}(x, v) = \sup_{p \in T_x^*\mathcal{M}} (\langle p, v\rangle - \mathcal{H}(x, p))$$

when $\sigma = I_{d\times d}$. Let $A(x) : T_x\mathcal{M} \to T_x\mathcal{M}$ be the linear operator and is defined by the sum of the tensor products $V_i(x) \in T_x\mathcal{M}$ with themselves,

$$A(x) = \sum_{i=1}^{d} V_i(x) \otimes V_i(x).$$

This operator is then extended to a tensor field in the sane way the Riemann metric tensor is extended. Let A be this tensor field on $\mathcal{M}$. Since the vector fields V_i are smooth, A will also be smooth, and it is defined for each $x \in \mathcal{M}$.

The next step is to consider $A^{-1}(x)$ for each $x \in \mathcal{M}$ as the inverse of the linear operator $A(x)$ in $T_x\mathcal{M}$. For this to exist conditions on $A(x)$ are needed. Namely, we need the linear operator to be invertible and to extend $A^{-1}(x)$ to the whole of $\mathcal{M}$ we need $A(x)$ to be invertible for each x. This is the case if for each $x \in \mathcal{M}$ the set of vectors $V_i(x)$ span the tangent space $T_x\mathcal{M}$. In other words, if for each $x \in \mathcal{M}$, the linear operator $A(x)$ is positive definite. Let $A(x)$ be positive definite for each $x \in \mathcal{M}$. Define $A^{-1}(x)$ as above.

Then, with that in mind consider

$$\begin{aligned}
\mathcal{L}(x, v) &= \sup_{p \in T_x^*\mathcal{M}} (\langle p, v\rangle - \mathcal{H}(x, p)) \\
&= \sup_{p \in T_x^*\mathcal{M}} (\langle p, v\rangle - \langle p, V_0\rangle_{g(x)} - \frac{1}{2}\sum_{i=1}^{d}(p(V_i))^2) \\
&= \sup_{p \in T_x^*\mathcal{M}} (\langle p, v - V_0\rangle_{g(x)} - \langle p, Av\rangle_{g(x)})
\end{aligned}$$

this is a common Legendre - Fenchel transform in linear algebra and using this the Lagrangian is give by

$$\mathcal{L}(x, v) = \frac{1}{2}\langle A^{-1}(x)(v - V_0), (v - V_0)\rangle_{g(x)}.$$

Therefore by theorem 2.1.1, the rate function is given by

$$I(\gamma) = \frac{1}{2}\int_0^\infty \langle A^{-1}(\gamma_t)(\dot{\gamma}_t - V_0(\gamma_t)), \dot{\gamma}_t - V_0(\gamma_t))\rangle_{g(\gamma_t)} dt.$$

This is the form of the rate function we would expect from classical Freidlin- Wentzell theory, now the inner product is changed to the Riemann inner product and the matrix is instead given by the tensor field defined by A^{-1}.

When the linear operator $A(x)$ is not invertible for all $x \in \mathcal{M}$ we need to consider a different approach. Namely, we need to delve into the variational representation for $\mathcal{H}$. Note that if $\mathcal{L}(x,v) = \frac{|v|^2_{g(x)}}{2}$ and we define $\sigma(x) : T_x\mathcal{M} \to T_x\mathcal{M}$ as the square root of $A(x)$, i.e. $\sigma(x)\sigma^T(x) = A(x)$ then if we define

$$Af(x,u) = (\sigma(x)u)(f)(x) + V_0(f)(x)$$

where $u \in T_x\mathcal{M}$ then we have that for all $f \in C^\infty(\mathcal{M})$,

$$Hf(x) = \sup_{u \in T_x\mathcal{M}} \{Af(x,u) - \mathcal{L}(x,u)\}$$

and $A : f \to Af(x,u)$ satisfies conditions 8.9, 8.10 and 8.11 in Feng and Kurtz. So,

$$I(x) = \inf_{(x,\lambda)\in\mathcal{J}^t} \left\{ \int_{T\mathcal{M}\times[0,\infty)} \mathcal{L}(x,u)\lambda(du \times ds) \right\}$$

$(x,\lambda) \in \mathcal{J}^t$ implies $\forall f \in C^\infty(\mathcal{M})$ we have

$$f(X_t) - f(X_0) = \int_{T\mathcal{M}\times[0,t]} (\sigma(X_s)u)(f)(X_s) - V_0(f)(X_s)\lambda(du \times ds)$$

if we then decompose $\lambda(du \times ds) = \mu_s(du) \times ds$ (using the admissible controls such that μ_s is only supported on $\{T_{X_s}\mathcal{M}\} \subset T\mathcal{M}$) we can let $u_s = \int_{T\mathcal{M}} u\mu_s(du)$ so that the above then becomes

$$f(X_t) - f(X_s) = \int_0^t [\sigma(X_s)u_s](f)(X_s) - V_0(f)(X_s)ds \quad \star$$

and therefore setting $\Lambda = \{\eta \in H_1^2(\mathcal{M}) : \dot{\eta} \text{ satisfies } \star\}$ we have

$$I(\gamma) = \frac{1}{2}\inf_{\eta\in\Lambda} \int_0^\infty |\dot{\eta}_s|^2_{g(\gamma_s)} ds$$

Chapter 11

Conclusion

It is hard to close such a document. In reality, it will never quite be done. It will definitely never quite be the document that was planned. Throughout the document the interplay between mathematics, physics, and biology has been shown. However, there are still many open problems in this area. Still projects to be worked on and ideas to be had. Here I will mention a few areas that this work did not cover.

The first area is the nature of more than two energy wells in the metastable protein transitions. Moreover, more general energy landscapes could be constructed, and more detailed information on the relationship between their geometry and the FRET vs lifetime distribution could be understood. Further, a full $SE(3)$ model for the IDPs could be implemented [12]. This would provide a better approximation to the IDP movement while also incorporating rotational dynamics.

There are several areas not covered in the fluorescent dye models. One big issue is the interplay between the rotational dynamics and the translational dynamics. It is known that classically the rigid body attached to an elastic pendulum exhibits nontrivial coupling between the rotational motion and translational motion. In this work we took the motions to be independent. Moreover, the dye is taken to be spherical, so that the dipole motion evolves according to a spherical Brownian motion. More elaborate simulations detailing the motion of non-spherical dyes still need to be implemented to study the impact of dye molecular geometry.

The interplay between the previous two topics is also not investigated very much. The nature of rotation of the underlying protein is seen to be important as the dynamic shift in the IDP models demonstrates. However, the coupling between dye motion and protein rotation has yet to

be explored outside of the IDP case mentioned in this work.

While there is a lot of work to be done in the field of stochastic differential geometry, it is the applications need to be explored. It is with the hope that this document has given a contribution to the scientific community as a whole, and has shown the benefit and interest of interdisciplinary thinking.

Appendices

Appendix A Physics of FRET

First recall that atoms and molecules possess electron orbitals with discrete energy levels. Upon interaction with photons of specific energy electrons can change energy levels and bring the molecule to an "excited" state. This electron can then "relax" and the molecule emits a photon - the molecule fluoresces. The mechanism of FRET is stating that a molecule, the donor, may enter an excited state, but instead of fluorescing the energy is transferred to another nearby molecule, the acceptor making this molecule enter an excited state. This energy exchange is due to the interaction between the electric fields of the molecules, which is based on the interaction between electric dipoles. The electric dipole is a vector which measures charge distribution $\rho : \mathbb{R}^3 \to \mathbb{R}$ to be[67]

$$\mu = \int_{\mathbb{R}^3} \rho(r) r dr.$$

In probabilistic terms, if ρ was a multivariate probability distribution, the dipole would be the mean vector. The electric dipole moment is the direction of the electric dipole[67],

$$\hat{\mu} = \frac{\mu}{\|\mu\|_{\mathbb{R}^3}}.$$

Here we are concerned with molecular dipole-dipole interaction.

Quantum mechanical FRET can be described using the dipole operator[13]:

$$\hat{H} = \frac{\kappa \|\mu_A\|_{\mathbb{R}^3} \|\mu_D\|_{\mathbb{R}^3}}{R^3}$$

where $R = \|D - A\|_{\mathbb{R}^3}$ is the scalar distance between the donor and acceptor position vectors and

$$\kappa = (\hat{\mu}_D \cdot \hat{\mu}_A) - 3(\hat{R} \cdot \hat{\mu}_D)(\hat{R} \cdot \hat{\mu}_A)$$

is an orientation factor for the electric field induced by the electric dipoles for the acceptor and donor, μ_A and μ_D respectively.[13] From this one can use Fermi's "golden rule",

$$k_{ET} = \frac{2\pi}{\hbar} \beta^2 \tilde{\rho}(E),$$

to describe the transition rate. Here, $\hat{\rho}$ represents the density of states with energy E, β is an

interaction term between states ψ_{D^*A} and ψ_{AD^*} given by the dipole operator

$$\beta = \langle \psi_{D^*A} | \hat{H} | \psi_{AD^*} \rangle.$$

The states ψ_{D^*A} and ψ_{AD^*} represent the donor excited state and the acceptor excited state respectively. From this one arrives at the energy transition rate

$$k_{ET} = C\tilde{\rho}(E)\kappa^2 \left(\frac{2\pi}{\hbar}\right) \left(\frac{\|\mu_A\|_{\mathbb{R}^3} \|\mu_D\|_{\mathbb{R}^3}}{R^3}\right)^2 \left(\langle \chi_{D^*} | \chi_D \rangle \langle \chi_A | \chi_{A^*} \rangle\right)^2$$

where the last terms are the Frank Condon factors [13] representing the intensity of vibrational states in each molecule when excited or relaxed.

Appendix B Martingale Problem

A natural question to ask is if starting with an operator is it possible to define a process with such that the operator is its generator? The answer is yes. [1] This relationship is given by the martingale problem. Martingales can be thought of as stochastic constants and solving the martingale problem can be thought of as the stochastic equivalent to finding integral curves of a first-order differential operator.[55] Where in the case of first-order differential operators integral curves are paths y_t such that for each $f \in C_b^\infty(\mathbb{R}^d)$ we have

$$f(y_t) - \int_0^t \sum_{i=1}^d v_i(x) \frac{\partial}{\partial x_i} f(y_t),$$

which is constant. The martingale problem extends this reasoning to more general operators by replacing the constant requirement with a stochastic constant, or a martingale.[55]

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\mathcal{F}_t\}$ be a filtration, an increasing sequence of sigma algebras, of the sigma-algebra $\mathcal{F}$. A stochastic process $\{X_t\}$ is said to be **adapted** to $\{\mathcal{F}_t\}$ if for each $t \in T$, X_t is $\mathcal{F}_t$ measurable. [48]

Definition B.1. *A real-valued adapted stochastic process $\{X_t\}$ is called a **martingale**[48][1] with respect to the filtration $\{\mathcal{F}_t\}$ if*

- $X_t \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.
- *If $s \leq t$, $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ almost surely.*

Notice that deterministic constants are martingales, so finding integral curves of a first-order differential operator can be viewed as solving the martingale problem for that operator. Now, the martingale problem is defined as follows.[56][55]

Definition B.2. *Let $\mathcal{A}$ be a linear operator. A stochastic process $X = \{X_t\}$ is said to solve the **martingale problem** for $\mathcal{A}$ if for all $f : S \to \mathbb{R}$ such that $f \in D(\mathcal{A})$,*

$$f(X_t) - f(X_0) - \int_0^t \mathcal{A}f(X_s)ds$$

is a martingale with respect to the natural filtration of X.

[1] Note that usually one wants to verify that an operator will generate a semigroup using the Hille - Yosida theorem, we will not mention this.

It is shown in [56][18] that if X solves the martingale problem for $\mathcal{A}$ then X is a Markov process with infinitesimal generator $\mathcal{A}$. It is of some interest to note that many times one is interested in the flow of diffeomorphisms that are induced by first-order differential operators. This line of reasoning can also be extended to more general settings by considering stochastic flows of diffeomorphisms, see [40][41] for more information. However, as we have seen Markov processes do not need to have a state space that admits differentiation, and can exist on very general structures. A key note in this work is the exploration of the behavior of Markov chains on different state spaces with differing properties.

Appendix C The Orthonormal Frame Bundle

The last tool needed from differential geometry is the orthonormal frame bundle. This topic is required as it is used in the definition of a Riemannian Brownian motion through stochastic development, and will also be used to define the process studied in section 2.2.

A **frame** is a is a linear isometry $u : \mathbb{R}^d \to T_x\mathcal{M}$ from Euclidean space to the tangent space at a point $x \in \mathcal{M}$.[35][55] In this way an orthogonal basis $\{e_i\} \subset \mathbb{R}^d$ can be taken to an orthogonal basis $\{ue_i\} \subset T_x\mathcal{M}$ in the tangent space at x. Denote the set of all frames at x by $\mathcal{O}(\mathcal{M})_x$. Define the **orthonormal frame bundle** by

$$\bigcup_{x\in\mathcal{M}} \{x\} \times \mathcal{O}(\mathcal{M})_x.$$

Notice that the set $\mathcal{O}(\mathcal{M})_x$ is acted on by the orthogonal group $\mathcal{O}(d)$ by right multiplication i.e. if $g \in \mathcal{O}(d)$ and $u \in \mathcal{O}(\mathcal{M})_x$ then $ug \in \mathcal{O}(\mathcal{M})_x$. This makes the orthonormal frame bundle a principle fiber bundle on $\mathcal{M}$ with fibers given by the Lie group $\mathcal{O}(d)$.[35][55] Let $\pi : \mathcal{O}(\mathcal{M}) \to \mathcal{M}$ be the canonical surjection given by $\pi(x, u) = x$.

The orthonormal frame bundle is a $d + d^2$ dimensional manifold, which means that one can consider the tangent space at a frame $u \in \mathcal{O}(\mathcal{M})$, $T_u\mathcal{O}(\mathcal{M})$.[35][55] A tangent vector $v \in T_u\mathcal{O}(\mathcal{M})$ is called **vertical** if it is tangent to the fiber $\mathcal{O}(\mathcal{M})_{\pi u}$. [35] Denote this space as $V_u\mathcal{O}(\mathcal{M})$. Note that moving in a vertical direction amounts to a group action by $\mathcal{O}(d)$, in this way vertical vectors are like infinitesimal rotations of the frame u.[55] Vertical vectors do not comprise the entirety of the tangent space $T_u\mathcal{O}(\mathcal{M})$, indeed the space of vertical vectors has only d^2 dimensions.[35] However there is not canonical assignment of a space to the other d dimensions. This choice of linear complement to $V_u\mathcal{O}(\mathcal{M})$ actually constitutes a connection on the manifold.[35][27][55] However, working in Riemannian geometry one already has the Levi-Civita connection so let us see how this manifests itself as the linear complement to $V_u\mathcal{O}(\mathcal{M})$.

Let u_t be a curve in $\mathcal{O}(\mathcal{M})$, meaning it is a smooth choice of frames along πu_t a smooth curve in $\mathcal{M}$. The curve u_t is called **horizontal** if for each $e \in \mathbb{R}^d$ the vector field $u_t e \in T_{\pi u_t}\mathcal{M}$ is parallel along the curve $\pi u_t \in \mathcal{M}$.[35][55] Then a tangent vector in $h \in T_u\mathcal{O}(\mathcal{M})$ is called a **horizontal vector** if it is tangent to a horizontal curve at u. Denote the space of all horizontal vectors at u as $H_u\mathcal{O}(\mathcal{M})$. This induces an isomorphism $\pi_* : H_u\mathcal{O}(\mathcal{M}) \to T_{\pi u}\mathcal{M}$.[35][55] Moreover,

for each $V \in T_{\pi u}\mathcal{M}$ there is a unique horizontal vector $V^* \in H_u\mathcal{O}(\mathcal{M})$ given by $\pi_* V^* = V$, this is called the **horizontal lift** of V. Horizontal vectors describe a parallel transport in the sense that horizontal curves are parallel transports of frames along curves in $\mathcal{M}$. Parallel transport along a curve πu_t can then be expressed very cleanly as

$$\tau_{t_0,t_1} V = u_{t_1} u_{t_0}^{-1} V$$

noting that $u_{t_0}^{-1} : T_{\pi u_{t_0}}\mathcal{M} \to \mathbb{R}^d$ and $u_{t_1} : \mathbb{R}^d \to T_{\pi u_{t_1}}\mathcal{M}$.[35] In this way it is very clear that if $t_1 \to t_0$ then $\tau_{t_0,t_1} \to I$.

Now given a vector $e \in \mathbb{R}^d$ one can construct an entire horizontal vector field, H_e, on $\mathcal{O}(\mathcal{M})$. This is done by horizontal lifts,

$$H_e(u) = (ue)^*.$$

In this way one can define the **fundamental horizontal vector fields** on $\mathcal{O}(\mathcal{M})$.[35]

Definition C.1. *Let $\{e_i\}_{i=1}^d \subset \mathbb{R}^d$ be an orthonormal basis for $\mathbb{R}^d$. Define the fundamental horizontal vector fields by*

$$H_i = H_{e_i},\ i = 1, .., d.$$

The fundamental horizontal vector fields define an orthonormal basis on the horizontal tangent space at each $u \in \mathcal{O}(\mathcal{M})$. [35]. Using these tangent vectors one can define an object that is central to the study of Brownain motion on manifolds, **Bochner's horizontal Laplacian**,

$$\Delta_{\mathcal{O}(\mathcal{M})} = \sum_{i=1}^{d} H_i^2$$

While there is much theory behind the study of Riemann Brownian motion, we will only briefly outline how it is constructed. Riemann Brownian motion is defined on the orthonormal frame bundle of $\mathcal{M}$. This is because one generalizes the notion of development of a curve to stochastic paths. The intuitive idea of the development of a curve to a manifold a "rolling without slipping" proceedure. Imagine rolling a sphere over a curve of wet ink, the ink will trace a path on the sphere, the path traced is the development of the path to the sphere. This process is conducted for smooth paths by solving an ordinary differential equation in the orthonormal frame bundle. However, it can be generalized to nonsmooth paths by considering a SDE in the orthonormal frame bundle.

This idea was first recommeded by Bochner and carried out in the case of Brownian motion and a sphere by Itô. [36] Figure 2.1 gives a picture of the process. One raises the continuous Euclidean semi-martingale Z to $\mathcal{O}(\mathcal{M})$ by solving the SDE

$$dU_t = \sum_{i=1} H_i(U_t) \circ dZ_t^i$$

where H_i are the fundamental vector fields in $H(\mathcal{O}(\mathcal{M}))$.[35] The process, U_t defined by this SDE in $\mathcal{O}(\mathcal{M})$ can then be projected on to the manifold by using the natural surjection, π on $\mathcal{O}(\mathcal{M})$ so that $X_t = \pi U_t$ defines a process on $\mathcal{M}$.[35][55]

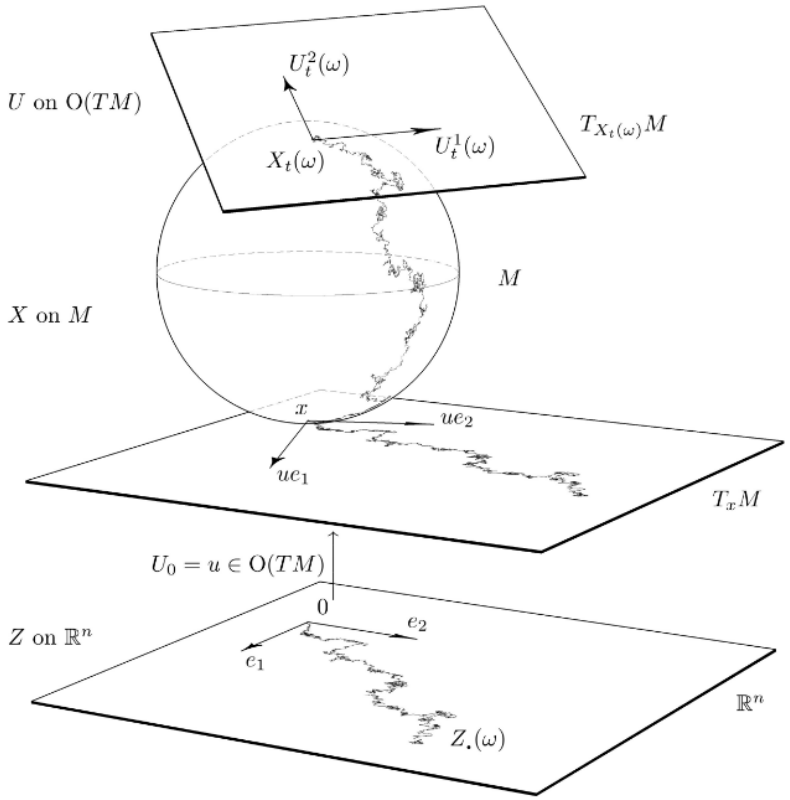


Figure 1: A representation of Stochastic Development. Here $T(OM)$ is used instead of $\mathcal{O}(\mathcal{M})$. This picture was created by Anton Thalmaier in his lecture notes on Stochastic Riemannain Geometry. [57]

When $Z = W$ is a Euclidean Brownian motion, one can carry out this procedure and construct a Riemann Brownian motion. Consequently, an $\mathcal{O}(\mathcal{M})$ valued Brownian motion is given

by the solution to

$$dU_t = \sum_{i=1}^{d} H_i(U_t) \circ dW_t^i$$

and thus the generator of an $\mathcal{O}(\mathcal{M})$ valued Brownian motion is given by

$$\frac{1}{2}\sum_{i=1}^{d} H_i^2 = \frac{1}{2}\Delta_{\mathcal{O}(\mathcal{M})}.$$

So, the generator for a orthonormal frame bundle valued Brownian motion is Bochner's horizontal Laplacian and a Riemann Brownian motion, X_t is given by πU_t.[35] To relate this to the generator of a Riemann Brownian motion on $\mathcal{M}$, one uses the following proposition

Proposition C.1. *Let $x = \pi U$ and $f \in C^\infty(\mathcal{M})$. Note that $f \circ \pi : \mathcal{O}(\mathcal{M}) \to \mathbb{R}$ is a smooth function on $\mathcal{O}(\mathcal{M})$. Further, let $\Delta_{\mathcal{M}}$ be the Laplace - Beltrami operator on $\mathcal{M}$. Then,*

$$\Delta_{\mathcal{O}}(\mathcal{M})(f \circ \pi)(U) = \Delta_{\mathcal{M}} f(x).$$

Using this proposition one relate a Riemann Brownian motion to the Laplace-Beltrami operator via the martingale problem, this is taken as the definition of a Riemann Brownian motion in this work.

Definition C.2. *An semi-martingale $X = \{X_t\}$ on $\mathcal{M}$ is called a **Riemann Brownian Motion** if it solves the martingale problem for $\Delta_{\mathcal{M}}$. In other words, a Riemann Brownian motion is a diffusion process on $\mathcal{M}$ with infinitesimal generator $\Delta_{\mathcal{M}}$. [35][55]*

It is worth mentioning that general $\mathcal{M}$ valued SDE can be formulated as an SDE on the orthonormal frame bundle similarly to Brownian motion. Define the $\mathcal{O}(\mathcal{M})$ valued SDE

$$dU_t = V_0^*(U_t)dt + \sum_{k=1}^{d} V_k^*(U_t) \circ dW_t^k$$

where V_i^* is the horizontal lift of the vector fields V_i. Then $X_t = \pi U_t$ will be the solution to the SDE

$$dX_t = V_0(X_t)dt + \sum_{k=1}^{d} V_k(X_t) \circ dW_t^i$$

and consequently have the same generator. [35]

This gives an interesting comparison between the processes and a glimpse into the nature of

horizontal vector fields. There are no vector fields in $T\mathcal{M}$ such that their horizontal lift will become H_i. The fundamental horizontal vector fields, while being defined as horizontal lifts, change along the fibers of $\mathcal{O}(\mathcal{M})$, meaning $\pi_* H_i(u) \neq \pi_* H_i(\tilde{u})$ for $u \neq \tilde{u}$ with both in $\mathcal{O}(\mathcal{M})_x$ i.e. the horizontal vector fields are not invariant under the group action of $\mathcal{O}(\mathcal{M})$. Whereas the horizontal lift of a vector field V will remain constant along these fibers, meaning for any $x \in \mathcal{M}$ the projection of the horizontal lift to $T_x\mathcal{M}$ is invariant under choice of u in the fiber above x, i.e. $\pi_* V_x^*(u) = V_x$ no matter what $u \in \mathcal{O}(\mathcal{M})_x$ the vector V_x is lifted to - this is a part of the definition of the horizontal lift. In this way the fundamental horizontal vector fields can only live in $\mathcal{O}(\mathcal{M})$, outside of this bundle they lose their structure. Vertical lifts, however, are "embedded" into a bigger space and given more structure. This speaks to the fundamental difference between the two processes. While they are very similar they come from two subtly different contexts. Brownian motion needs the frame bundle to exist while other processes do not. In this way, defining a process that is augmented by solely white noise requires more tools than defining a process that has a nontrivial diffusion tensor.

Bibliography

[1] David Applebaum. *Lèvy Processes and Stochastic Calculus*. Cambridge University Press, 2004.

[2] Anders Barth, Oleg Opanasyuk, Thomas-Otavio Peulen, Suren Felekyan, Stanislav Kalinin, Hugo Sanabria, and Claus AM Seidel. Unraveling multi-state molecular dynamics in single-molecule fret experiments. i. theory of fret-lines. *The Journal of Chemical Physics*, 156(14):141501, 2022.

[3] Richard F. Bass. *Diffusions and Elliptic Operators*. Springer -Verlag, 1998.

[4] Richard F. Bass. *Stochastic Processes*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 2011.

[5] Roux Benoît. Transition rate theory, spectral analysis, and reactive paths. *The Journal of Chemical Physics*, 156(13):134111, 2022.

[6] Patrick Billingsley. *Probability and Measure*. John Wiley and Sons, Inc., 111 River Street, Hoboken, NJ 07030, 1995.

[7] Patrick Billingsley. *Convergence of Probability Measures*. John Wiley and Sons, Inc., 111 River Street, Hoboken, NJ 07030, 1999.

[8] Richard L Bishop and Samuel I Goldberg. *Tensor analysis on manifolds*. Courier Corporation, 2012.

[9] Adam Bobrowski. *Functional Analysis for Probability and Stochastic Processes*. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK, 2005.

[10] J Shepard Bryan IV and Steve Pressé. Learning continuous potentials from smfret. *Biophysical journal*, 2022.

[11] Gregory S Chirikjian. *Stochastic Models, Information Theory, and Lie Groups, Volume 1: Classical Results and Geometric Methods*. Springer Science & Business Media, 2009.

[12] Gregory S Chirikjian. *Stochastic models, information theory, and Lie groups, volume 2: Analytic methods and modern applications*, volume 2. Springer Science & Business Media, 2011.

[13] Robert M Clegg. Fluorescence resonance energy transfer. *Fluorescence imaging spectroscopy and microscopy*, 137:179–251, 1996.

[14] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.

[15] Michael G Crandall and Thomas M Liggett. Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, 93(2):265–298, 1971.

[16] Robert E Dale, J Eisinger, and WE Blumberg. The orientational freedom of molecular probes. the orientation factor in intramolecular energy transfer. *Biophysical journal*, 26(2):161–193, 1979.

[17] Robert P Dobrow. *Introduction to stochastic processes with R*. John Wiley & Sons, 2016.

[18] Richard Durret. *Stochastic Calculus: A Practical Introduction*. CRC Press, 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431, 1996.

[19] Richard S Ellis. *Entropy large deviations and statistical mechanics*. Taylor & Francis, 2006.

[20] Michel Emery. *Stochastic calculus in manifolds*. Springer Science & Business Media, 2012.

[21] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

[22] Lawrence C Evans. *Partial differential equations*, volume 19. American Mathematical Society, 2022.

[23] Jin Feng and Thomas G. Kurt. *Large Deviations for Stochastic Processes*. American Mathematical Society, 2006.

[24] Wendell H Fleming and Halil Mete Soner. *Controlled Markov processes and viscosity solutions*, volume 25. Springer Science & Business Media, 2006.

[25] National Center for Biotechnology Information. Pubchem compound summary for cid 102384756, alexa-488. 2023.

[26] National Center for Biotechnology Information. Pubchem compound summary for cid 6694, rhodamine b. 2023.

[27] Theodore Frankel. *The Geometry of Physics: An Introduction*. Cambridge University Press, 2004.

[28] Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254(5038):1598–1603, 1991.

[29] Mark Freidlin and Alexander Wentzell. *Random Perturbations of Dynamical Systems*. Springer-Verlag, 1979.

[30] Dominique Bakry Ivan Gentil and Michel Ledoux. *Analysis and Geometry of Markov Diffusion Operators*. Springer International Publishing - Switzerland, 2014.

[31] Irina V Gopich and Attila Szabo. Single-molecule fret with diffusion and conformational dynamics. *The Journal of Physical Chemistry B*, 111(44):12925–12932, 2007.

[32] George Hamilton and Hugo Sanabria. Multiparameter fluorescence spectroscopy of single molecules. In *Spectroscopy and dynamics of single molecules*, pages 269–333. Elsevier, 2019.

[33] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, Apr 1990.

[34] Björn Hellenkamp, Sonja Schmid, Olga Doroshenko, Oleg Opanasyuk, Ralf Kuhnemuth, Soheila Rezaei Adariani, Benjamin Ambrose, Mikayel Aznauryan, Anders Barth, Victoria Birkedal, et al. Precision and accuracy of single-molecule fret measurements—a multi-laboratory benchmark study. *Nature methods*, 15(9):669–676, 2018.

[35] Elton P. Hsu. *Stochastic Analysis on Manifolds*. American Mathematical Society, 2002.

[36] Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes*. Elsevier, 2014.

[37] David S Kliger and James W Lewis. *Polarized light in optics and spectroscopy*. Elsevier, 2012.

[38] Richard C. Kraaij, Frank Redig, and Rik Versendaal. Classical large deviation theorems on complete riemannian manifolds. *Stochastic Processes and their Applications*, 129(11):4294–4334, 2019.

[39] Richard C. Kraaijl. Dynamical moderate deviations for the curie–weiss model. *Stochastic Processes and their Applications*, 127(9):2900–2925, 2017.

[40] Hiroshi Kunita. *Stochastic flows and stochastic differential equations*, volume 24. Cambridge university press, 1997.

[41] Hiroshi Kunita. *Stochastic flows and jump-diffusions*. Springer, 2019.

[42] Joseph R Lakowicz. *Principles of fluorescence spectroscopy*. Springer, 2006.

[43] Bernard H Lavenda. *Nonequilibrium statistical thermodynamics*. Courier Dover Publications, 2019.

[44] John M Lee. *Riemannian manifolds: an introduction to curvature*, volume 176. Springer Science & Business Media, 2006.

[45] Eitan Lerner, Anders Barth, Jelle Hendrix, Benjamin Ambrose, Victoria Birkedal, Scott C Blanchard, Richard Börner, Hoi Sung Chung, Thorben Cordes, Timothy D Craggs, et al. Fret-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *Elife*, 10:e60416, 2021.

[46] Makiko Nisio. Stochastic control theory. *ISI Lecture Notes*, 9, 2015.

[47] Oleg Opanasyuk, Anders Barth, Thomas-Otavio Peulen, Suren Felekyan, Stanislav Kalinin, Hugo Sanabria, and Claus AM Seidel. Unraveling multi-state molecular dynamics in single-molecule fret experiments. ii. quantitative analysis of multi-state kinetic networks. *The Journal of Chemical Physics*, 157(3):031501, 2022.

[48] Philip E Protter. *Stochastic differential equations*. Springer, 2005.

[49] Sidney I. Resnick. *A Probability Path*. Springer, 2014.

[50] C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York, 2005.

[51] Sheldon M. Ross. *Introduction to Probability Models*. Academic Press, 2014.

[52] Matthew Safar, Ayush Saurabh, Bidyut Sarkar, Mohamadreza Fazel, Kunihiko Ishii, Tahei Tahara, Ioannis Sgouralis, and Steve Pressé. Single-photon smfret. iii. application to pulsed illumination. *Biophysical Reports*, 2(4):100088, 2022.

[53] Ayush Saurabh, Mohamadreza Fazel, Matthew Safar, Ioannis Sgouralis, and Steve Pressé. Single-photon smfret. i: Theory and conceptual basis. *Biophysical Reports*, 3(1):100089, 2023.

[54] Ayush Saurabh, Matthew Safar, Mohamadreza Fazel, Ioannis Sgouralis, and Steve Pressé. Single-photon smfret: Ii. application to continuous illumination. *Biophysical Reports*, 3(1):100087, 2023.

[55] Daniel W. Stroock. *An Introduction to the Analysis of Paths on a Riemannain Manifold*. American Mathematical Society, 2000.

[56] Daniel W. Stroock and S.R.Srinivasa Varadhan. *Multidimensional Diffusion Processes*. Springer - Verlag, 1997.

[57] Anton Thalmaier. Stochastic riemannian geometry. 2021.

[58] Hugo Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, 2009.

[59] BW Van der Meer. Kappa-squared: from nuisance to new sense. *Reviews in Molecular Biotechnology*, 82(3):181–196, 2002.

[60] SRS Varadhan. Large deviations and entropy. *Entropy*, 47:199, 2003.

[61] Rik Versendaal. Large deviations for geodesic random walks, 2019.

[62] Rik Versendaal. Large deviations for brownian motion in evolving riemannian manifolds. *arXiv preprint arXiv:2004.00358*, 2020.

[63] Rik Versendaal. *Large Deviations for Stochastic Processes on Riemannian Manifolds*. PhD thesis, Technical University of Delft, Netherlands, 2020.

[64] K. Vogtmann, A. Weinstein, and V.I. Arnol'd. *Mathematical Methods of Classical Mechanics*. Graduate Texts in Mathematics. Springer New York, 1997.

[65] E Weinan, Tiejun Li, and Eric Vanden-Eijnden. *Applied stochastic analysis*, volume 199. American Mathematical Soc., 2021.

[66] David E. Wolf. What is the confocal volume? FCS.

[67] Andrew Zangwill. *Modern electrodynamics*. Cambridge University Press, 2013.

[68] Jingxiao Zhang and D. Kannan. A girsanov type theorem on the path space over a compact riemannian manifold. *Stochastic Analysis and Applications*, 25(3):667–678, 2007.