

MATHEMATICAL MODELING FOR TRANSCRIPTION OF DNA WITH
PAUSING: STOCHASTIC MODEL WITH TORQUE, AND DIFFUSIVE
TRANSPORT MODEL

by

Tamra Lindsey Heberling

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Mathematics

MONTANA STATE UNIVERSITY
Bozeman, Montana

April, 2016

©COPYRIGHT

by

Tamra Lindsey Heberling

2016

All Rights Reserved

ACKNOWLEDGEMENTS

My deepest thanks to Lisa Davis whose time, energy, attention to detail, and most of all, endless patience, made this work possible. I am indebted to Tomáš Gedeon who was never out of great ideas to guide this project and always, somehow, found time for a meeting with me. Thank you to Tianyu Zhang for taking the time to read this dissertation. Thank you as well to Lukas Geyer for helping me work out the details of a couple of the proofs. I would also like to thank Julia Platt, Marty Morgan, and Jakub Gedeon for putting their hard work into this project. Finally I am extremely grateful to Marcy Barge who was “just about to get bored” whenever I stopped by his office for help. I would not be where I am today without the love and support of my family. I am extremely fortunate to have kind and compassionate parents. They taught me that I could do anything if I set my mind to it. My sister is both a friend and a kindred spirit. I especially want to thank my husband for his unwavering support and commitment.

Funding Acknowledgment

This work was kindly supported by National Science Foundation grant Mathematical Biology DMS-1226213, and the Kopriva Graduate Student Fellowship. Thank you for funding this project.

TABLE OF CONTENTS

1. INTRODUCTION	1
2. A MECHANISTIC MODEL FOR COOPERATIVE BEHAVIOR OF CO-TRANSCRIBING RNA POLYMERASES	6
The Poisson Process	7
The Poisson Distribution	8
Exponential Distribution.....	10
Simulating Random Variables	11
General Introduction to Modeling Approach.....	13
TASEP Model.....	13
ETAM Model.....	15
Torque between RNAP and DNA	17
Incorporating Torque into Model.....	22
Physical Interpretation of Torque	24
3. SIMULATION AND RESULTS OF ETAM	28
Incorporating Experimental Data	28
Fitting the Data Linearly.....	31
Simulation.....	32
Simulation Events	32
Incorporation of Pauses	34
Incorporation of Collisions	35
Comparison to TASEP	38
Simulation Parameters.....	38
Results.....	39
Results from the Nonlinear Data Fit.....	40
Average Number and Average Duration of Pauses.....	40
Number and Duration of Collisions	42
Average Transcription Time	45
Average Transcriptional Delay	46
Results from the Linear Data Fit	50
The Importance of the Pause Duration	52
Different Pause Frequency Functions	55
Discussion	60

TABLE OF CONTENTS - CONTINUED

4. BROWNIAN RATCHETS.....	63
Equation of Motion.....	65
Important Quantities for a Brownian Ratchet on a Periodic Potential with Drift.....	67
Theorem for the Mean Escape Time	72
Probability of Escaping Left or Right	74
Proof of Theorem 4.0.1	83
Fourier Series Application	87
Quadratic Wave	93
Square Wave.....	99
Noise Intensity D for DNA Transcription	104
Discussion	106
5. CONCLUSION.....	108
REFERENCES CITED.....	111
APPENDICES	119
APPENDIX A: Derivation of Escape Time Quantities	120
APPENDIX B: Derivation for the Probabilities of Escape	134
APPENDIX C: Fourier Series Derivations	139

LIST OF TABLES

Table	Page
2.1 Parameters used in calculation of torque.....	20
3.1 Results for ETAM model: percent of the strand covered by polymerases, average transcription time per RNAP(s), average collision delay per RNAP (s), average pause delay per RNAP(s), and average total delay per RNAP (s).....	47
3.2 Results for TASEP model: percent of the strand covered by polymerases, average transcription time per RNAP(s), average collision delay per RNAP (s), average pause delay per RNAP (s), and average total delay per RNAP (s).	48
3.3 The average number of pauses per RNAP, average pause duration, and average pause delay per RNAP due to pauses experiences per RNAP for the different pause frequency functions. Frequency represents the value of the pause frequency function when torque is 11 pN·nm. All of the data reported is for $\alpha = 0.0115$	57

LIST OF FIGURES

Figure	Page
2.1	Polymerases P_{i-1} , P_i , and P_{i+1} in order on the DNA strand. When P_i translocates, the DNA between P_{i-1} and P_i will over-twist and the DNA between P_i and P_{i+1} will under-twist, increasing the elongation velocity of P_{i-1} and P_{i+1} 17
2.2	An elastic rod under torsion. The portion of the cylinder that is a dashed line represents the original distance L_0 and the current distance L . As the distance decreases from L_0 to L , the total amount of twist added to the rod is ϕ . The small increment $\Delta\ell_i$ represents a change in length due to RNAP motion by one nucleotide. 21
2.3	This figure depicts the four different cases of RNAP P_{i-1} , P_i , and P_{i+1} . Figure 2.3A shows the three RNAPs with the original distances between them. Then there are four different configurations; Figure 2.3B, $L_{0,i} > L_i$ and $L_{0,i+1} > L_{i+1}$, Figure 2.3C, $L_{0,i} > L_i$ and $L_{0,i+1} < L_{i+1}$, Figure 2.3D, $L_{0,i} < L_i$ and $L_{0,i+1} > L_{i+1}$, and Figure 2.3E, $L_{0,i} < L_i$ and $L_{0,i+1} < L_{i+1}$ 26
3.1	The data published by Ma et. al. is presented (red dots) as well as the curve fit to the data (black). The values that we choose based off of information from biology are given for very high and very low torque values (green dots). The equation for each curve can be found in Equations (3.1) - (3.3), respectively. 31
3.2	The data published by Ma et. al. is presented (red dots) as well as the piecewise linear curve fit to the data (green). The nonlinear curve fit is provided (blue) with our imposed values (green dots) for comparison purposes. The equation for each curve can be found in Equations (3.4) - (3.6), respectively. 33
3.3	A flowchart for our simulation of the elongation process. RNAPs that translocate follow the chart from elongation and continue in order depending on if that RNAP experiences a collision or a pause. If a neighboring RNAP translocates, updates for torque, velocity, and pause frequency are calculated for the affected RNAPs starting with that block of the flowchart. 37

LIST OF FIGURES - CONTINUED

Figure	Page
3.4	Over increasing initiation rates determined by α , the pause and collision results are presented for the ETAM model (blue triangles) and the TASEP model (magenta). The number of pauses and collisions are presented as the averaged number per RNAP. RNAPs in the ETAM model experience significantly fewer collisions and shorter pause durations than their TASEP counterparts. 43
3.5	Linear fit for the average number of collisions experienced per RNAP. The collisions in the results for TASEP (magenta) and compared to the linear fit (magenta stars) and similarly for the ETAM results (blue and blue stars). 44
3.6	Over increasing initiation rates $\alpha \cdot \beta$, we present our results for average transcription time and average total delay per RNAP. The ETAM model (blue triangles) and TASEP model (magenta) are both plotted for comparison. The dashed lines represent baseline simulations with no pauses. 48
3.7	Over increasing initiation rates $\alpha \cdot \beta$, we present our results for average transcription time and average total delay per RNAP in the ETAM model. The nonlinear model (blue triangles) and piecewise linear model (green dots), both detailed in previous sections, are plotted for comparison. 51
3.8	In addition to the nonlinear fit (blue triangles) and the piecewise linear fit (green dots), we also present our results for a piecewise linear fit except for the pause duration on the interval $[-10,5]$, “Nonlinear Left” (red stars), or on the interval $[5,10]$, “Nonlinear Right” (black stars). 53
3.9	The differences in the pause duration for the piecewise linear fit (green) and the nonlinear fit (blue) at low torque values is highlighted here. This difference accounts for the the very different results in average transcription time and average total delay depicted in Figure 3.7. 54

LIST OF FIGURES - CONTINUED

Figure	Page
3.10	Pause frequency function where the value at the end point, 11 pN·nm is raised. The data (red dots) is fit with one quadratic function up to 7.5 pN·nm. A second quadratic function is used to fit the end point, with the exception of the lowest function which is the original quadratic for the entire interval. 56
3.11	Over increasing initiation rates $\alpha \cdot \beta$, these results show average transcription time and total delay over different fits of the pause frequency function. The legend labels the pause frequency at the highest torque value. 56
3.12	The percentage of times a torque value is calculated in a baseline (blue) simulation and a pause (red) simulation is presented here as a histogram. The histogram bars at -10 and 10 is the percentage that those values were computed. Each of the other bars represents the percentage between (-10,-9], (-9,-8], etc., with the exception of the bars at the label < 10 . The histogram bars at < 10 represents the percentage of torque values computed in (9,10)..... 59
4.1	The sawtooth function is often employed in Brownian ratchets as it is an excellent example of an asymmetrical, periodic potential, $V(x)$ 64
4.2	The potential function $V(x)$ is divided into overlapping segments of length $2L$, here segment m_i spans the interval $[-L, L]$. When a particle exits from either end of the segment, it is placed at the position $x = 0$ of the neighboring segment..... 68
4.3	The multi-periodic potential function $V(x) = U_1(x, L/N) + U_2(x, L) - Fx$ is provided here for reference. $U_1(x, L/N)$ provides most of the oscillatory nature of this function, while $U_2(x, L)$ adds the higher amplitude wall that is more spaced out..... 73
4.4	The integration domain for μ segregated into N^2 parallelograms each with width L/N and height L/N 78
4.5	The tilted potential function $V(x)$ with $F = 0.2$ and $U(x, L)$ as the periodic extension given in Equation (4.88) on the interval $[0, 3L]$ 94

LIST OF FIGURES - CONTINUED

Figure	Page
4.6	The truncated Fourier series, $f_N(x)$ approaches the original function, $f(x)$, with each new term added. The original function (black) is shown compared to the Fourier series with one term, $f_1(x)$ (green stars). Another term is added until we consider the Fourier series with five terms, $f_5(x)$ (red dashes). 95
4.7	The MST for the original quadratic potential, μ , and the Fourier series μ_N are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96), $\tilde{\mu}_N$ is displayed with blue stars..... 97
4.8	The MST for the original quadratic potential μ , and the Fourier series μ_N are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96), $\tilde{\mu}_N$, is displayed with blue stars. This figure shows the results for different values of D with $N = 20$ as a fixed quantity..... 98
4.9	The tilted potential function $V(x)$ with $F = 0.95$ and $U(x, L)$ as the periodic extension of the square wave in Equation (4.97) on the interval $[0, 3L]$ 100
4.10	The truncated Fourier series approaches the original square wave function with each new term added. The original function, $f(x)$ (black), is shown compared to the Fourier series with one term, $f_1(x)$ (green dashes). Another term is added until we consider the Fourier series with five terms, $f_5(x)$ (red dashes). 101
4.11	The MST for the original square wave potential, μ , and the Fourier series, μ_N , are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96), $\tilde{\mu}_N$ is displayed with blue stars..... 103
4.12	The MST for the original quadratic potential μ and the Fourier series μ_N , are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96) $\tilde{\mu}_N$, is displayed with blue stars. This figure shows the results for different values of D with $N = 20$ as a fixed quantity..... 104

ABSTRACT

In fast-transcribing prokaryotic genes, like an *rrn* gene in *Escherichia coli*, many RNA polymerases (RNAPs) transcribe the DNA simultaneously. Active elongation of RNAPs is often interrupted by pauses, which has been observed to cause RNAP traffic jams; yet some studies indicate that elongation seems to be faster in the presence of multiple RNAPs than elongation by a single RNAP. We propose that an interaction between RNAPs via the torque produced by RNAP on helically twisted DNA can explain this apparent paradox. We have incorporated the torque mechanism into a stochastic model and simulated transcription both with and without torque. Simulation results illustrate that the torque causes shorter pause durations and fewer collisions between polymerases. Our results suggest that the torsional interaction of RNAPs is an important mechanism in maintaining fast transcription times, and that transcription should be viewed as a cooperative group effort by multiple polymerases. In an effort to further understand transcription, we investigate the Brownian ratchet model for nucleotide translocation. We model elongation as diffusive particle transport in a tilted periodic potential. To incorporate the RNAP pauses, a second periodic potential is added to the first. We present a formula for the mean escape time from a tilted, periodic potential composed of multiple periodic functions as the product of the mean escape time from each individual periodic function. This formula is extended to an arbitrary finite number of periodic functions. Two examples using truncated Fourier series are presented and analyzed.

INTRODUCTION

DNA stores the genetic information of all living organisms. DNA consists of a series of building blocks called nucleotides, and each nucleotide is itself constructed of a phosphate group, a sugar group, and one of four nitrogen bases: adenine (A), thymine (T), guanine (G), and cytosine (C). It is the order of these bases that encodes the differences in species and traits. DNA sequences are converted into proteins which carry out most of the functions performed by organisms. The production of proteins, when they are needed, is essential as it allows an organism to develop, reproduce, adapt and survive.

The process of converting DNA into a protein has two phases: transcription and translation. During transcription, an RNA polymerase (RNAP) reads the DNA sequence and produces a strand of RNA by complementary base pairing. When an RNAP reads T, it matches with A on the RNA strand, C matches with G, G matches with C, and A matches with uracil (U), a different nitrogen base. Therefore RNA is made up of a sequence of A's, U's, G's, and C's. The type of RNA that encodes a protein is called messenger RNA (mRNA).

During translation, the mRNA is read by a ribosome to produce the protein. Ribosomes are similar to RNAPs in that they read and build. The ribosome reads the bases of the mRNA in groups of 3. The order of that sequence tells the ribosome which amino acid, out of twenty options, will be next. This chain of amino acids is a protein. A different sequence of amino acids would encode a different protein.

There are three main types of RNA: mRNA as mentioned above, ribosomal RNA (rRNA), and transfer RNA (tRNA). The rRNA combines with a variety of proteins to produce a ribosome, which will then be used for translation of the mRNA. The tRNA brings the amino acids to the ribosome which chains them together to produce

the protein. Each of these RNAs are made during transcription. The type of RNA that is produced depends on which section of DNA, is being read.

Regulation plays a large role in protein production. The ability of an organism to adapt to its environment and to survive relates directly to the organism transcribing enough of each type of RNA to effectively translate into the needed proteins. However there is often not enough energy in a cell to allow the organism to produce a large surplus of these necessary materials. In vitro, experimenters can provide “high growth conditions” which means there is a surplus of food for an organism to convert to energy as needed. In vivo, the growth conditions can vary, and this variation forces an organism to perform a balancing act to produce enough, but not too much, of each necessary item. Because protein production is such a vital and complicated process, there are several regulatory networks in place.

In Eukaryotes, organisms whose cells have a nucleus, the DNA is transcribed inside of the nucleus. The RNA produced by this transcription then needs to be moved outside of the nucleus in order to be translated into a protein. There are regulatory steps involved at each part of this process. Prokaryotes, organisms whose cells lack a standard nucleus, perform both transcription and translation in the cytoplasm of the cell, often simultaneously. Therefore regulation of gene expression for bacteria, such as *Escherichia coli* (E. coli), happens largely during transcription [35, 74].

The process of transcription has several important phases. First, an RNA polymerase (RNAP) binds to a promoter sequence of a gene and initiates elongation. Next, the RNAP elongates down the DNA making a single-stranded copy of RNA, and finally terminates, releasing the nascent copy of RNA.

Elongation of RNAP along the DNA strand is not uniform, but is interrupted by frequent pauses. There are at least three different types of pauses: backtracking pauses, hairpin pauses, and ubiquitous pauses [26, 60]. Backtracking pauses, and

hairpin pauses have been shown to have a higher probability of occurring during transcription of specific sequences [22, 42, 44]. On the other hand, ubiquitous pauses are thought to have no dependence on DNA sequence and are equally likely to occur at any position along the DNA strand. These pauses are short (1 – 5 seconds) and occur approximately every 100 base pairs (bp) on average [60]. It has been theorized that ubiquitous pauses are caused by a restructuring of the polymerase [60], but the exact cause remains an open question.

There has been substantial interest to understand the effect of the presence of pauses on the average *transcription time* and therefore output of the RNA, for highly transcribed genes. The presence of pauses may lead to traffic jams of RNAPs when one polymerase stops, affecting the trailing polymerases [23, 24, 46]. According to Klumpp *et. al.* [46], in their stochastic model, RNAPs experienced a 40% reduction in the average elongation rate in dense traffic, amplifying the pause effect. This is similar to the results of a PDE model studied in recent years [23, 24]. In less traffic, the RNAPs in the model experienced only a 12% reduction of the average elongation rate [46].

The behavior of simple statistical models is at odds with experimental observations of highly transcribed genes. A prototypical example of a highly transcribed gene is an *rrn* operon in *E. coli*. The *E. coli* genome has seven *rrn* operons whose transcription produces ribosomal RNA (rRNA) which provides a scaffold for a ribosome [15, 16, 39]. During conditions of rapid growth there are as many as 70,000 ribosomes in a cell. To keep up with high demand for ribosomes, 90% of transcription in fast growing *E. coli* produces rRNA and tRNA, and only 10% produces mRNA [11]. As a result, there is a high density of RNAPs on all *rrn* operons and a high transcription completion rate is imperative. Experimental measurements have shown that approximately 31% of an *rrn* operon is covered by RNAPs (about 51 RNAPs) [25]

during high growth conditions, which strongly suggest that the polymerases interact either directly, or indirectly during transcription. This interaction appears to be cooperative. *In vivo* and *in vitro* experiments have demonstrated that the presence of multiple RNAPs in close proximity can assist in increasing the average elongation rate. A trailing RNAP can help a paused RNAP to re-enter translocation, thereby decreasing the delay caused by pauses [29]. The magnitude of the cooperativity effect has not been firmly established. However, it is worth noting that the average elongation rate of RNAPs on the *rrn* operon is 90 nucleotides per second (nt/s) [15, 46, 70, 82, 83], which is about double the *in vivo* elongation velocity on protein coding genes [12, 46, 57, 68].

While the elegant paper of Epstein *et. al.* [29] firmly established the cooperativity effect, it did not propose a mechanistic explanation for this phenomena. We propose that the torsional force between the elongating RNAP and the DNA, caused by the helical structure of the DNA, may provide the mechanical underpinning for the interaction between elongating polymerases. The basis for our model is a set of experimental measurements by Ma and co-authors [53]. Using *in vitro* single-molecule experiments, they measure both the magnitude of torque exerted by elongating RNAP on DNA, and the effect of supercoiled DNA on RNAP velocity, pause density and pause duration.

In Chapter 2, we detail the incorporation of torque into a basic stochastic model, named Elongation with Torque Assisted Motion (ETAM). A derivation of the torque formula and how to calculate the torque between two RNAPs is discussed in detail. Chapter 3 investigates several ways to fit the data published by Ma *et. al.* [53], describes the numerical simulation of the ETAM model, and presents the results of these simulation, which clearly show the cooperative effect of multiple RNAPs transcribing simultaneously. Finally, in Chapter 4 we present a theoretical motivation

for the motion of RNAPs by considering translocation of polymerases as diffusive particle transport. Chapter 4 details an approach to modeling transcription *with pausing* as an escape from a tilted periodic potential.

A MECHANISTIC MODEL FOR COOPERATIVE BEHAVIOR OF
CO-TRANSCRIBING RNA POLYMERASES

We construct a model for transcription that substantially extends a basic stochastic model referred to as a Totally Asymmetric Simple Exclusion Process (TASEP). In the TASEP model each particle, representing an individual RNAP, hops along the DNA strand at a uniform rate provided that the forward site is unoccupied. In our model, Elongation with Torque Assisted Motion (ETAM), the rate of hopping depends on the torque between the particle and its closest two neighboring particles. The amount of torque is, in turn, the result of the relative motion of particles on the DNA strand.

Our simulation results show that the torque-based interaction between RNAPs results in a substantial cooperation effect between RNAPs. As a trailing RNAP approaches a leading RNAP, the resulting torque increases the elongation rate and reduces the likelihood of and duration of pauses by the leading RNAP. At the same time, the elongation rate of the trailing polymerase decreases, while the likelihood and duration of pauses increases. As a result of this interaction, the duration of pauses decreases, and the average number of completed transcription occurrences increases when compared to the behavior of the standard TASEP under similar conditions. The effect of this interaction is not unlike that of autonomously driven and communicating vehicles (“google” cars) on the road. By automatically adjusting velocity and helping each other to maintain proper spacing and shorten pauses, the collective motion of polymerases becomes more efficient with an average transcription time that is 37.5% less than that of the TASEP model. In this sense, the RNAPs are collaborating in order to transcribe the strand more efficiently than by simply traveling at a constant rate.

Both TASEP and our constructed model are coupled Poisson processes. The first section of this chapter is devoted to the details of a basic Poisson process. This information and more can be found in these two excellent stochastic modeling books [79, 88].

The Poisson Process

A Poisson process is an example of a continuous time Markov chain. The distinguishing feature of a Markov chain is the Markov property which states that the future state of a system depends only on the present state and not on the past.

Definition 2.0.1. *A continuous time Markov chain with finite or countable state space \mathcal{X} is a family $\{X_t = X(t)\}_{t \geq 0}$ of \mathcal{X} -valued random variables such that for any set of times $t_i < t_{i+1}$ where $t_{i+1} = t_i + s_{i+1}$ and states $x_i \in \mathcal{X}$ with $t_0 = 0$,*

$$P(X(t_{k+1}) = x_{k+1} | X(t_i) = x_i, \forall i \leq k) = P(X(s_{k+1}) = x_{k+1} | X(0) = x_k)$$

This means that the probability of being in state x_{k+1} after the current state x_k does not depend on how state x_k was reached but is as if the current state is the initial state. In other words, the chain is “memoryless”.

A particular example of a continuous time Markov chain is the Poisson process which satisfies the above definition using the state space $\mathcal{X} = \{0, 1, 2, \dots\}$. With this state space, Poisson processes are often used to count the number of events that occur in an interval of time. Formally, the definition of a Poisson process is as follows,

Definition 2.0.2 (Taylor, 1984). *A Poisson process of intensity or rate $\lambda > 0$ is an integer-valued stochastic process $\{X(t); t \geq 0\}$ for which*

(i) for any time points $t_0 = 0 < t_1 < t_2 < \dots < t_n$, the process increments

$$X(t_1) - X(t_0), X(t_2) - X(t_1), \dots, X(t_n) - X(t_{n-1})$$

are independent random variables;

(ii) for $s \geq 0$ and $t > 0$, the random variable $X(s + t) - X(s)$ has the Poisson distribution

$$P\{X(s + t) - X(s) = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \quad \text{for } k = 0, 1, \dots$$

and

(iii) $X(0) = 0$.

The first piece of this definition demonstrates that the number of events that occur in an interval of time is independent of the number of events that occur in any other interval. This independence implies the Markov chain property and shows that the Poisson process is a particular instance of a continuous time Markov chain. Let's consider the Poisson distribution in more detail.

The Poisson Distribution

Consider the formula for the Poisson distribution where X is a random variable describing the number of events (successes) that occur in an interval,

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \lambda > 0, \quad k = 0, 1, 2, \dots \quad (2.1)$$

which is derived from the binomial distribution. In order to understand this process it is beneficial to see how the distribution is derived. Consider n Bernoulli trials

X_1, \dots, X_n where each X_i is a random variable with values 1 with probability $p = \lambda/n$ or a 0 with probability $1 - p = 1 - \lambda/n$ where n is the number of trials, λ/n is the probability of success, and λ is the expected value. The binomial distribution is

$$P(X = k) \approx \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \quad (2.2)$$

where $0 < \lambda < n$, and $k = 0, 1, 2, \dots, n$ represents the number of successes.

We obtain the Poisson distribution as a limit of binomial distributions as

$$\begin{aligned} P(X = k) &= \lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}, \quad k = 0, 1, 2, \dots \\ &= \lim_{n \rightarrow \infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \left(\frac{\lambda^k}{k!}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \left(\frac{\lambda^k}{k!}\right) \lim_{n \rightarrow \infty} \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{n^k} \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{-k} \\ &= \left(\frac{\lambda^k}{k!}\right) \cdot 1 \cdot e^{-\lambda} \cdot 1 \\ &= \frac{e^{-\lambda} \lambda^k}{k!} \quad k = 0, 1, 2, \dots \end{aligned} \quad (2.3)$$

which is the expression originally given by Equation (2.1). Here λ is the parameter of the Poisson process. It represents the rate of occurrences in a unit time interval. It also represents the expected value of the number of occurrences. To demonstrate

this we calculate

$$\begin{aligned}
 E[X] &= \sum_{k=1}^{\infty} k \frac{e^{-\lambda} \lambda^k}{k!} \\
 &= \sum_{k=1}^{\infty} \frac{e^{-\lambda} \lambda^k}{(k-1)!} \\
 &= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
 &= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\
 &= \lambda e^{-\lambda} e^{\lambda} \\
 &= \lambda.
 \end{aligned} \tag{2.4}$$

If the events occur over intervals of length other than unit length, for instance suppose the intervals are of length t , the rate of change is now λt and the mean number of occurrences in a time interval of length t is λt . Hence the probability becomes

$$P(X_t = k) = \frac{e^{-\lambda t} (\lambda t)^k}{k!} \quad k = 0, 1, 2, \dots \tag{2.5}$$

Notice in this distribution, the probability that one event occurs in an interval of length t , i.e. $k = 1$ in Equation (2.5), reduces to an exponential distribution. In fact, the time between two consecutive events is exponentially distributed as described in the following section.

Exponential Distribution

The Poisson distribution described in Equation (2.5) represents the probability that k events happen in a time interval of length t . The probability that an event occurs at a particular time is governed by an exponential distribution,

$$P(t) = \lambda e^{-\lambda t}. \tag{2.6}$$

This time distribution is derived using the Poisson distribution with $k = 1$.

Using the exponential distribution we can calculate the average time between events for a stochastic Poisson process. This is simply the expected value of our inter-arrival time distribution. Since the inter-arrival times, T , have an exponential distribution we can use the probability density function for this distribution, given by Equation (2.6). Then

$$\begin{aligned}
 E[T] &= \int_0^{\infty} tP(t) dt \\
 &= \int_0^{\infty} t\lambda e^{-\lambda t} dt \\
 &= -te^{-\lambda t} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda t} dt \quad \text{IBP} \\
 &= 0 - \frac{1}{\lambda} e^{-\lambda t} \Big|_0^{\infty} \\
 &= \frac{1}{\lambda}.
 \end{aligned} \tag{2.7}$$

Therefore the average time between events is

$$E[T] = \frac{1}{\lambda}. \tag{2.8}$$

Now that we have the exponential distribution for the time between events, we describe how to simulate a Poisson process by generating the time of the next event.

Simulating Random Variables

The most common way to choose the time of the next event, governed by an exponential distribution, is actually to generate a uniformly distributed random number. While this may seem counterintuitive, the Inverse Distribution Method provides everything needed to make this possible.

Theorem 2.0.1 (Inverse Distribution Method). *If U is a uniformly distributed random variable on the interval $[0,1]$, denoted $U(0,1)$, and $F(\cdot)$ is an invertible cumulative distribution function (CDF), then $X = F^{-1}(U)$ has CDF $F(\cdot)$.*

Proof. Consider

$$P(X \leq x) = P(F^{-1}(U) \leq x). \quad (2.9)$$

Since F is a CDF it is monotonically increasing and we can apply F to both sides of the argument on the right hand side,

$$P(X \leq x) = P(U \leq F(x)). \quad (2.10)$$

Now the right hand side of Equation (2.10) is the CDF for a uniform distribution, denoted F_U . Hence

$$P(X \leq x) = F_U(F(x)) = F(x) \quad (2.11)$$

since $F_U(u) = u$ on a unit interval. Therefore X has the cumulative distribution function F . □

We can use the Inverse Distribution Method to simulate an exponential distribution out of a uniform distribution which we will show as a corollary.

Corollary 2.0.1. *If $U \sim U(0,1)$ and $\lambda > 0$, then the random variable defined by*

$$X = -\frac{1}{\lambda} \ln(U)$$

has an exponential distribution with parameter λ .

Proof. Consider X , an exponentially distributed random variable which has the probability density function and CDF $f(x)$ and $F(x)$ respectively, given by the

equations

$$f(x) = \lambda e^{-\lambda x}, \quad F(x) = 1 - e^{-\lambda x}.$$

Inverting F leads to

$$F^{-1}(u) = -\frac{1}{\lambda} \ln(1 - u). \quad (2.12)$$

Since both U and $1 - U$ have $U(0, 1)$ distribution,

$$X = -\frac{1}{\lambda} \ln(U)$$

which has exponential distribution by the Inverse Distribution Method. □

Therefore to pick the time of the next event, a uniformly distributed random number on $[0,1]$ is generated. If the current time is t_i , the next event occurs at $t_{i+1} = t_i - \frac{1}{\lambda} \ln(u)$.

General Introduction to Modeling Approach

We build our model ETAM (Elongation with Torque Assisted Motion) off of a basic TASEP model, which we describe next.

TASEP Model

TASEP is a stochastic model that has been used to describe the process of both transcription and translation [9, 14, 23, 45, 46, 71, 74, 85, 90]. In TASEP, a particle representing an RNAP moves forward with a stochastic hopping motion with a mean hopping rate β on a one-dimensional strand with a discrete and finite number of sites. Particles cannot occupy the same site and therefore will cease active elongation if the next site is occupied (we refer to this as a collision). Only when the next site is vacated does elongation resume.

The elongation of a particle represents a transition from one state to another, where each state represents a nucleotide, and we model it as a Poisson process. Each particle has an elongation rate β , to transition from its current state to the translocated state and moves via its own Poisson process. However, as a result of the collisions, where one RNAP cannot elongate because of the positioning of another, the events are not independent of each other. In the case of DNA transcription, there are multiple Poisson processes that are coupled with one another. This type of process is referred to as a coupled Poisson Process.

Coupled Poisson processes can be simulated using a kinetic Monte Carlo algorithm. This algorithm chooses the next event based on the transition rate for each event (β for elongation using TASEP), with higher transition rates more likely to be selected. An RNAP that cannot elongate due to a collision would have its transition rate set to zero - which would eliminate the selection of an elongation event for that RNAP. After an event is chosen, the time the event occurs is generated next. The details of the kinetic Monte Carlo algorithm are provided in Algorithm 2.1 using r_j as the transition rate for the j th particle. For the case of TASEP, r_j is equal to the constant β for all actively elongating particles, however we prefer to use r_j in the algorithm for the sake of the ETAM model where these transition rates are variable as will be discussed in detail in the following section. It is important to note that the time of the next event does not depend on which event is chosen. The time of the next event is exponentially distributed with parameter R_N , the cumulative rate function for all transition rates.

We have implemented open boundary conditions for the TASEP model where the particles enter the strand at a given rate and exit the strand once they reach the opposite end. Specifically, a particle enters our simulation (initiates) with a rate of $(\alpha \cdot \beta)$, and leaves the simulation (terminates) with a rate of $(\gamma \cdot \beta)$. This is consistent

- 1: Set time to $t = 0$ to initiate
- 2: List all transition rates $r_j, j = 1, \dots, N$, from initial state to another state
- 3: Calculate cumulative rate function $R_i = \sum_{j=1}^i r_j$
- 4: Generate $u \in U(0, 1)$.
- 5: **if** $\frac{R_{i-1}}{R_N} < u < \frac{R_i}{R_N}$ **then**
- 6: Chose event i with transition rate r_i
- 7: Carry out event i
- 8: **end if**
- 9: Generate new $u' \in U(0, 1)$
- 10: Calculate $\Delta t = -\frac{\ln(u')}{R_N}$
- 11: Update time $t = t + \Delta t$
- 12: Recalculate all rates altered due to transition
- 13: Repeat from step 3.

Algorithm 2.1: This algorithm is a general kinetic Monte Carlo process.

with a differential equation model for transcription proposed originally in the late 1960s [34]. Therefore α and γ are scalars that multiply the elongation rate to obtain initiation and termination rates, respectively. After developing the particulars of the TASEP model, we can now introduce the ETAM model and the specifics of the derivation for this model.

ETAM Model

DNA double helix structure makes one full rotation in approximately 10.5 base pairs [51]. RNAPs are large molecules that translocate along this twisted structure, which places constraints on the mutual motion of DNA and RNAP. If the DNA strand were fixed in space, an RNAP would have to rotate around DNA during translocation. The size of the RNAP and the packed environment inside the cell precludes this motion and a localized rotation of DNA has been observed [17, 40]. If the DNA strand were free to rotate, it could spin along its long axis as it enters a

stationary RNAP. However, if DNA is fixed upstream of the elongating RNAP and the RNAP elongates without rotation, it applies torque to the DNA. This torque is stored within the portion of the DNA strand between the fixed end and the RNAP, and if the amount of torque is large enough, it can either preclude or facilitate the forward motion of itself and its neighboring RNAP.

The effect of torque on DNA - RNAP interaction was experimentally quantified in recent work of Ma *et. al.* [53] where a single-molecule optical trap experiment was employed in order to measure the effect of twist in a DNA strand on transcribing RNAP. In particular, they first applied a predetermined value of torque to a strand of DNA and then measured the elongation rate, the pause frequency and the pause duration of an RNAP elongating on the strand. Two types of twisting mechanisms are used to describe the applied torque. The first is *over-twisting*, and it is characterized by applying twist in a manner that shortens the length of the full rotation of the helix (measured in base pairs). Likewise, a twist that increases the length of the full rotation of the helix is termed *under-twisting*. Ma and collaborators observed that over-twisting decreases the elongation rate of the RNAP and increases both the likelihood and the duration of pauses. On the other hand, under-twisting was found to increase the elongation rate and to decrease both the likelihood and the duration of the pauses.

To illustrate how these results can have an effect on the transcription of DNA by multiple polymerases, consider three consecutive polymerases on a DNA strand labeled as P_{i-1} , P_i and P_{i+1} in Figure 2.1. This notation will be defined in detail in the section titled *Incorporating Torque into Model*. With respect to the motion of P_i along the strand, the RNAP represented by P_{i-1} is referred to as the *leading* RNAP, and the RNAP labeled P_{i+1} is referred to as the *trailing* RNAP. If we assume that individual RNAPs never move at exactly the same time, when P_i moves forward, both

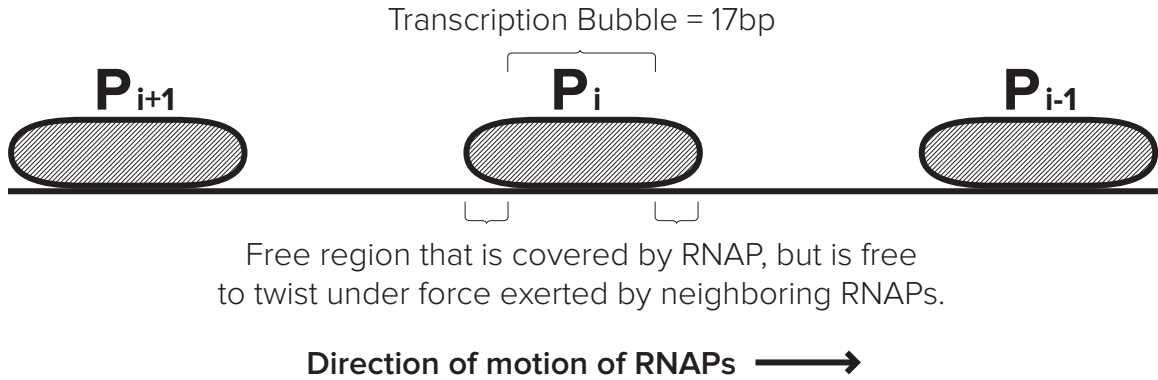


Figure 2.1: Polymerases P_{i-1} , P_i , and P_{i+1} in order on the DNA strand. When P_i translocates, the DNA between P_{i-1} and P_i will over-twist and the DNA between P_i and P_{i+1} will under-twist, increasing the elongation velocity of P_{i-1} and P_{i+1} .

P_{i-1} and P_{i+1} provide anchors for the DNA strand. This movement imparts a torque to the portion of the DNA strand between P_{i-1} and P_i , as well as to the DNA strand between P_i and P_{i+1} . The portion of the strand between P_{i-1} and P_i will over-twist, and the portion of the strand between P_i and P_{i+1} will under-twist. Note that the over-twist will increase the elongation rate of P_{i-1} (i.e. P_{i-1} receives a “push from the back”) and under-twist will also increase elongation rate of P_{i+1} (i.e. P_{i+1} receives a “pull from the front”). It is noted in a later section that both of these effects tend to synchronize the motion of all three polymerases. The following section makes this process more precise by deriving a mathematical expression for the amount of torque that RNAP imparts to DNA during translocation.

Torque between RNAP and DNA DNA consists of two strands in a double helix structure. In addition, DNA is a flexible structure that can experience bend. A polymer’s bend-persistence length is a mechanical property that characterizes its stiffness. For lengths less than the bend-persistence length, a polymer such as DNA, will exhibit behavior similar to that of an elastic rod. The bend-persistence length

of DNA is estimated to be 150 bp [58]. Therefore, on length scales shorter than the persistence length, the DNA strand is comparable to an elastic rod. Since the average distance between elongating polymerases on an *rrn* gene is 100 bp [89], it is reasonable to assume that, on average, the force exerted by one elongating RNAP on its neighbors occurs over a distance of approximately 100 bp. Since this distance is within the persistence length reported in the literature, we assume that the local behavior of the DNA strand connecting two adjacent RNAPs can be modeled as an elastic rod. In addition, the footprint of an RNAP is approximately 35 bp, of which 17 bp is occupied by the transcription bubble [32]. As depicted in Figure 2.1, we assume that the 17 bp where DNA is unwound inside of the RNAP are anchored and cannot be twisted but the other 18 bp (we assume that these base pairs are partitioned into 9 bp on each side of the bubble) are free to twist under an appropriate force. In this setting, we use classical elasticity theory to describe the interaction between DNA and elongating polymerases.

The torque $\hat{\tau}$ stored in an elastic rod under torsion is

$$\hat{\tau} = \mu \frac{\pi r^4}{2L} \Delta\phi \quad (2.13)$$

where r is the radius (≈ 1 nanometer (nm) for DNA [58]), μ is the shear modulus, L is the length of the rod, and $\Delta\phi$ is the angle of the total twist [30], see Figure 2.2. All quantities representing length are measured in base pairs and converted into nanometers using the conversion $1 \text{ bp} = 0.34 \text{ nm}$ [58].

In the context of our model, there is some flexibility in how one associates the length of the rod L with the distance between two neighboring polymerases as shown in Figure 2.1. First note that Equation (2.13) indicates that the torque goes to infinity as $L \rightarrow 0$. In terms of our model, this would preclude direct contact

between two polymerases. However, it has been observed experimentally [20] that RNAPs periodically exhibit direct contact with their neighbors during transcription. In addition, we recall that the footprint of elongating RNAP is approximately 35 bp, of which 17 bp is occupied by the transcription bubble [32]. The 17 bp inside of the transcription bubble cannot be twisted by an elongation force, but the remaining 18 bps are free to twist under an appropriate force. Therefore even when two RNAPs are directly adjacent to each other (without any empty base pairs separating them), there are still 18 bp between their corresponding transcription bubbles. Hence, we define the length L in (2.13) to be the distance between the transcription bubbles of adjacent RNAPs, where L has units of nucleotides. This also ensures that the underlying mathematical quantity in Equation (2.13) remains bounded.

The shear modulus, μ , is calculated using the formula

$$\mu = \frac{Y}{2(1 + \nu)} \quad (2.14)$$

where Y is Young's modulus and ν is Poisson's ratio [30]. The Poisson ratio of DNA has been reported to be anywhere in the interval $\nu \in [-0.7, 0]$, for more details see [7, 55, 56]. To calculate this parameter specifically we will follow the paper by Manning *et al* [56]. Using the formula

$$\nu = B/C - 1 \quad (2.15)$$

where B is the bending modulus and C is the twisting modulus [56], we calculate $\nu \approx -0.5$. Using these values for the Poisson ratio and Young's modulus in (2.14),

Table 2.1: Parameters used in calculation of torque.

Parameter	Symbol	Value	Reference
Shear Modulus	μ	300 pN/nm ²	calculated using (2.14)
Young's Modulus	Y	300 MPa	[58]
Poisson Ratio	ν	-0.5	calculated using (2.15)
Bending Modulus	B	230 pN nm ²	[56]
Twisting Modulus	C	460 pN nm ²	[56]

we have estimated the shear modulus for DNA to be

$$\mu \approx 300 \text{ pN/nm}^2.$$

We also simulated results using $\nu = 0$ and $\nu = -0.25$ and observed qualitatively similar behavior. These parameters and their references can be found in Table 2.1.

Next we derive a mathematical equation which describes the relationship between the length of the flexible segment of DNA strand between the transcription bubbles of successive RNAPs and the amount of torque in that segment of DNA. This segment of the strand is modeled using an elastic rod as sketched in Figure 2.2, and the notation is made precise below. Because there are 10.5 base pairs in one full helical rotation of DNA, the twist angle $\Delta\phi$ in (2.13) is proportional to the change in length $\Delta\ell$ as

$$\Delta\phi = \frac{2\pi}{10.5} \Delta\ell. \quad (2.16)$$

The torque between P_{i-1} and P_i (see Figure 2.1) due to the accumulation of applied torque is calculated by adding small increments of torque that correspond to motion by one nucleotide as in Equation (2.16). This measure of torque is always done in relation to the state of neutral twist of that segment of the strand. To be precise, we consider the segment of DNA strand between P_{i-1} and P_i as an elastic rod (see

Figure 2.1), and we assume that the DNA strand is in a state of neutral twist (not over-twisted or under-twisted) at the time of initiation and so the torque is zero. When the trailing RNAP (P_i) initiates, the distance between P_i and the nearest leading RNAP (P_{i-1}) downstream (in the direction of elongation) is defined to be L_0 . Assume that at some later time, the distance between P_{i-1} and P_i is L . In order to derive the total amount of torque stored in the rod at this instance, one notes that the amount of torque between P_{i-1} and P_i can be computed by treating the segment of the strand as an elastic rod with an original length of L_0 that has experienced a twist corresponding to angle ϕ , with a resulting length of the rod denoted by L , see Figure 2.2. The total torque is calculated by adding small increments of torque that correspond to elongation by one nucleotide. Applying (2.16), one calculates the

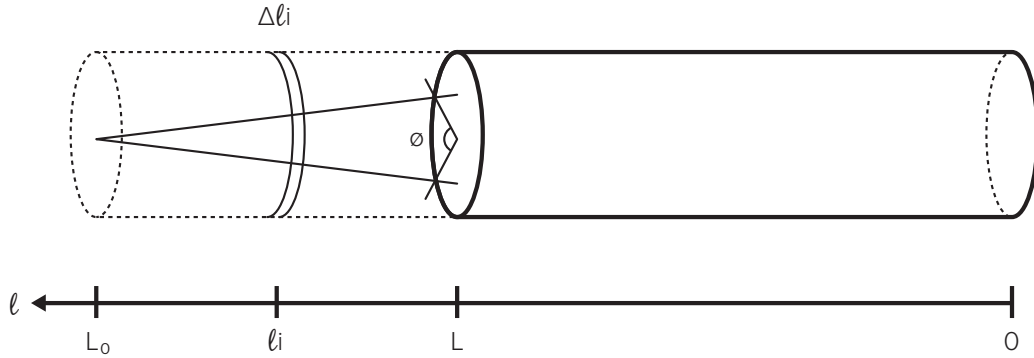


Figure 2.2: An elastic rod under torsion. The portion of the cylinder that is a dashed line represents the original distance L_0 and the current distance L . As the distances decrease from L_0 to L , the total amount of twist added to the rod is ϕ . The small increment Δl_i represents a change in length due to RNAP motion by one nucleotide.

torque produced by elongation of P_i by one nucleotide at the position of nucleotide i

to be

$$\begin{aligned}\Delta\tau_i &= \mu \frac{\pi r^4}{2\ell_i} \Delta\phi_i \\ &= \mu \frac{\pi r^4}{2\ell_i} \left(\frac{2\pi}{10.5} \Delta\ell_i \right).\end{aligned}$$

In order to construct a model that is efficient for extensive and repeated simulation, we use a continuous approximation of $\Delta\tau_i$ by $d\tau$, and lengths ℓ_i by s . The total torque in a segment of DNA strand with initial length L_0 and final length L is approximated by

$$\tau = \tau(L) = \int_L^{L_0} d\tau = \int_L^{L_0} \mu \frac{\pi r^4 \left(\frac{2\pi}{10.5} \right)}{2s} ds = \frac{\mu\pi^2 r^4}{10.5} \ln(L_0/L). \quad (2.17)$$

Note that for $L < L_0$ the DNA is *over-twisted* and the torque is positive, as in Figure 2.2. Conversely, if $L > L_0$ the DNA is *under-twisted* and the torque is negative. These correspond to the resisting and assisting torque found in [53]. If $L = L_0$ the torque is zero, since the DNA has the same length as when the trailing polymerase initiated onto the DNA. With this formula for the torque between two neighboring RNAPs, we next describe how this formula is used within the context of the stochastic elongation model.

Incorporating Torque into Model We will number RNAPs by index i which denotes the order of their initiation. The i -th polymerase will be characterized in the model by three numbers

$$P_i = (n_i, L_{0,i}, T_i) \quad (2.18)$$

where i represents a positive integer. This triple will be updated each time the RNAP translocates along the strand. The term n_i is an integer value denoting the furthest

downstream nucleotide number occupied by the i th RNAP, while T_i is the time of the next elongation of P_i . The calculation of this value is addressed in Chapter 3 in the section titled *Incorporation of Pauses*. Upon initiation of P_i onto the strand, define $L_{0,i}$ to be the distance (measured in nucleotides) between the transcription bubbles of P_i and that of its leading RNAP, P_{i-1} , at the time that P_i initiates transcription. This distance is calculated and stored in the triplet for each RNAP. When P_i initiates, the value of $L_{0,i}$ is fixed, however, this distance may be different for each RNAP as its initiation occurs at a randomly generated time and is independent of the distance traveled by any previously initiated RNAPs. We denote $L_i(t)$ to be the distance between the transcription bubbles of P_i and its leading RNAP P_{i-1} at time t . This distance is computed by accessing the variables n_{i-1} and n_i of P_{i-1} and P_i and measuring their difference. Using this distance, we can calculate the length of the DNA strand that is free to twist by subtracting the length of the transcription bubble, 17 nts. In other words,

$$L_i = (n_{i-1} - n_i) - 17. \quad (2.19)$$

In order to quantify the role of the torque calculation, we begin by considering elongation. First, suppose there are three RNAPs positioned on a segment of the DNA strand, and denote these RNAPs as P_k , for $k = i - 1, i, i + 1$ as labelled in Figure 2.1, and further suppose that P_i has just experienced elongation. The values L_i and L_{i+1} are calculated immediately following elongation. Using these distances, as well as $L_{0,i}$, and $L_{0,i+1}$, the torque that P_i is currently experiencing is calculated using Equation (2.17). Specifically, the torque has two components. The first is the component that corresponds to the torque between P_i and P_{i-1} , denoted $\tau(L_i)$, and the torque between P_i and P_{i+1} , denoted by $\tau(L_{i+1})$. Define τ_i to be the total torque

experienced by polymerase P_i . It is calculated as follows

$$\tau_i = \tau(L_i) - \tau(L_{i+1}) = \frac{\mu\pi^2 r^4}{10.5} \left[\ln \left(\frac{L_{0,i}}{L_i} \right) - \ln \left(\frac{L_{0,i+1}}{L_{i+1}} \right) \right]. \quad (2.20)$$

Analogous, calculations are repeated for RNAPs P_{i-1} and P_{i+1} , recalculating the torque for all three RNAPs after translocation of P_i . It is important to note, the calculations outlined in the previous paragraphs are only performed for the affected RNAPs; the elongating RNAP and its neighboring RNAPs.

During transcription, RNAPs first experience *initiation*, and the measurement of torque for initiation can be interpreted as a special case of the previous discussion. Upon initiation of an RNAP, labeled P_i in Figure 2.1, the distance between P_i and its leading RNAP, the quantity $L_{0,i}$, is set. Upon initiation and subsequent elongation of P_i , the torque behind this polymerase is zero until the next initiation. Once a trailing RNAP, P_i here, has initiated onto the strand, the calculation of the torque τ_i will have nonzero contributions from both neighbors of the polymerase.

We model RNAP termination as follows. Again referencing Figure 2.1, suppose P_{i-1} is the RNAP that is closest to the termination end of the DNA strand and therefore will be the next polymerase to terminate. Prior to termination, the torque measure τ_{i-1} has a nonzero contribution only from $\tau(L_i)$. That is, $\tau_{i-1} = 0 - \tau(L_i)$. Likewise, during this situation τ_i has contributions from both neighboring RNAPs as long as P_{i-1} is still transcribing the strand. Upon termination of P_{i-1} , the torque downstream of P_i is zero, hence $\tau_i = 0 - \tau(L_{i+1})$. Since P_i is now the last RNAP on the DNA strand, τ_i will be calculated as such until P_i itself terminates.

Physical Interpretation of Torque This formula for τ_i in (2.20) is based on the total amount of both assisting and resisting torque that is experienced by polymerase

P_i . Here we give some physical insights into the positioning of the RNAP relative to each of its neighbors and its effect on the torque calculation. Recall that when the distance L_i is compared to the original distance $L_{0,i}$, the DNA between them can be either under-twisted, over-twisted or neutral. *Over-twisting* in front of an RNAP provides a resisting torque, while over-twisting behind an RNAP provides an assisting torque (“push from the back”). On the other hand, under-twisting in front of an RNAP provides an assisting torque (“pull from the front”), while under-twisting in back of an RNAP is a resisting torque. Examining Equation (2.20) for several possible scenarios gives us insight into the influence of torque on movement of the polymerase.

For the neutral case when $L_{0,i} = L_i$ in Equation (2.20), $\tau(L_i)$ is zero and there is no contribution to τ_i from that component. The contribution to τ_i is similar for the case of $L_{0,i+1} = L_{i+1}$. Next we focus the discussion on the cases when $\tau(L_i)$ and $\tau(L_{i+1})$ are nonzero. There are four cases to consider which are depicted in Figure 2.3, scenarios (B) - (E) and are discussed below.

1. $L_{0,i} > L_i$ and $L_{0,i+1} > L_{i+1}$

The values of the two components of torque τ_i are $\tau(L_i) > 0$ and $\tau(L_{i+1}) > 0$. A positive value of $\tau(L_i)$ corresponds to a resisting torque being experienced by polymerase P_i relative to the position of its leading RNAP. A positive value of $\tau(L_{i+1})$ provides an assisting torque for P_i . Therefore the subtraction of $\tau(L_{i+1})$ implies that $\tau_i < \tau(L_i)$. In other words, RNAP P_{i+1} is assisting P_i to overcome the resisting torque in front of it at its current position.

2. $L_{0,i} > L_i$ and $L_{0,i+1} < L_{i+1}$

The values of the two components of torque τ_i are $\tau(L_i) > 0$ and $\tau(L_{i+1}) < 0$. These two factors indicate that the DNA between P_i and P_{i-1} is over-twisted

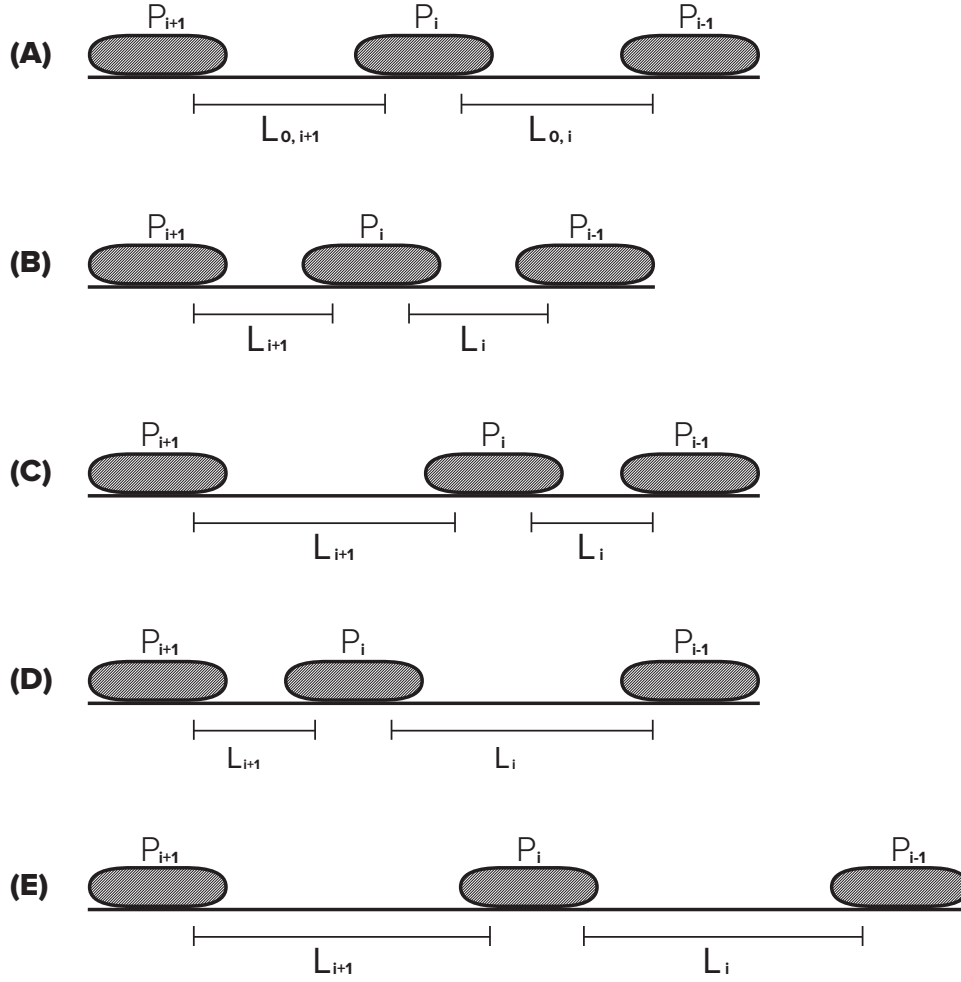


Figure 2.3: This figure depicts the four different cases of RNAP P_{i-1} , P_i , and P_{i+1} . Figure 2.3A shows the three RNAPs with the original distances between them. Then there are four different configurations; Figure 2.3B, $L_{0,i} > L_i$ and $L_{0,i+1} > L_{i+1}$, Figure 2.3C, $L_{0,i} > L_i$ and $L_{0,i+1} < L_{i+1}$, Figure 2.3D, $L_{0,i} < L_i$ and $L_{0,i+1} > L_{i+1}$, and Figure 2.3E, $L_{0,i} < L_i$ and $L_{0,i+1} < L_{i+1}$.

and the DNA between P_{i+1} and P_i is under-twisted resulting in P_i experiencing resisting torques from both sides. The negative value of $\tau(L_{i+1})$ is subtracted, increasing the value of τ_i , therefore the subtraction of $\tau(L_{i+1})$ implies that $\tau_i > \tau(L_i)$.

3. $L_{0,i} < L_i$ and $L_{0,i+1} > L_{i+1}$

The values of the two components of torque τ_i are $\tau(L_i) < 0$ and $\tau(L_{i+1}) > 0$.

The DNA between P_i and P_{i-1} is under-twisted and the DNA between P_{i+1} and P_i is over-twisted resulting in P_i experiencing an assisting torque from both sides.

4. $L_{0,i} < L_i$ and $L_{0,i+1} < L_{i+1}$

The values of the two components of torque τ_i are $\tau(L_i) < 0$ and $\tau(L_{i+1}) < 0$.

The DNA between P_i and its neighboring RNAPs is under-twisted. This leads to P_i experiencing an assisting torque from the front and a resisting torque from behind.

Next, in Chapter 3, we discuss the effects of torque on transcriptional pauses, specifically the ubiquitous pauses previously described. We will also detail the specifics of the simulation and compare the results of the model with torque, ETAM, versus the model without, TASEP.

SIMULATION AND RESULTS OF ETAM

Recent single molecule experiments by Ma and co-authors [53] attempt to advance our understanding of the relationship between the torque and the movement of the RNAP. The data obtained in those experiments suggests a relationship between torque and the elongation velocity of the RNAP. It has also been known for many years that during elongation, RNAPs experience short, frequent pauses where active elongation is stalled or arrested. Ma [53] also suggests that the nature of these pauses is tied to the amount of torque that an RNAP is experiencing at a given moment of time. The data reported in [53] was limited, as can be seen in Figures 3.1A - 3.1C, and we explored various ways to develop accurate representations of these relationships that could be used in a mathematical model. The following subsections give an overview of various approaches to incorporating the data into the current model, and simulation results using these different choices are discussed later.

Incorporating Experimental Data

In our model, we first fit the data from [53] to extract a functional relationship between the torque experienced by an individual polymerase and its elongation velocity, pause frequency and pause duration. Knowledge from various biological experiments describing physical properties of DNA are used. For instance, the full model employs a stall torque, a high torque value experienced by RNAPs where they are unable to elongate, reported in [53] to be on average 11 pN·nm. At values close to melting torque -10 pN·nm, the RNAPs would be experiencing a large amount of assistance, therefore we assumed a high velocity with no pauses. While there are many classes of functions to fit data, our goal was to fit the data using low order polynomials. We use the notation $V(\tau)$ to denote the polynomial function used to

describe the elongation velocity, $F(\tau)$ to denote the frequency of occurrence of a pause as a function of torque, and $D(\tau)$ to denote the duration time of a given pause.

When constructing the function $V(\tau)$ describing the elongation velocity, we assume that the velocity would decrease to 0 nt/s at the stall torque 11 pN·nm. To impose a value of velocity when torque was -10 pN·nm, we assume the polymerases would be traveling at rate that is significantly higher than $V(0)$. This assumption is due to the fact that negative torque is *assisting* torque which would aid efficient elongation. A large amount of assisting torque would lead to extremely fast velocities. For simplicity, we choose $V(-10) = 45$ nt/s, which is approximately twice the velocity at a torque of 0 pN·nm. Using these two extra data points we had a good agreement between the velocity data and a fifth order polynomial.

$$V(\tau) = -0.0002\tau^5 + 0.0008\tau^4 + 0.0041\tau^3 - 0.035\tau^2 - 0.2166\tau + 22.3574. \quad (3.1)$$

In order to develop equations for pause frequency and pause duration, we made assumptions that were consistent with the biology. At the stall torque we assume the RNAPs will pause with probability 1. Incorporating this information into the pause frequency function, an accurate fit is difficult to achieve using a single function since the data increases very quickly to 1 at a torque value of 11 pN·nm. Using an exponential fit severely underestimates the negative torque values, since the point (11, 1) dominates the fitting procedure. For this reason, a piecewise defined quadratic function is used to describe the functional relationship of the data.

$$F(\tau) = \begin{cases} 0.0001\tau^2 + 0.0022\tau + 0.0128 & \tau \leq 7.5 \\ 0.0453\tau^2 - 0.5621\tau + 1.7032 & \tau > 7.5 \end{cases} \quad (3.2)$$

As the torque approaches the stall torque, the pause duration should increase rapidly. RNAPs are able to reenter active elongation after reaching a stall torque if the torque is alleviated but will stay in a pause if the torque remains at the stall torque [53]. To capture both the rapid rise and infinite duration at 11 pN·nm, we wanted a function with a vertical asymptote at the stall torque. In order to produce the asymptotic behavior, we use a tangent function. For a torque of -10 pN·nm, we impose a pause duration of 0 seconds.

$$D(\tau) = 0.1914 \tan((\pi/23)\tau + 0.1469) + 0.5298 \quad (3.3)$$

The original data from [53] and the curve fits can be seen in Figure 3.1. Figures 3.1A - 3.1C correspond to qualitatively realistic relationships between torque and the parameters influencing the motion of a polymerase. The case of a zero value of torque is considered to be the case of no resistance. When an RNAP experiences a *resisting* torque, this corresponds to a positive torque value. The resulting elongation velocity (see Figure 3.1A) decreases from that of the zero torque case so that elongation is hampered. In addition, as the amount of torque increases, the velocity of the motion decreases, and the forward motion of the RNAP is resisted.

Examining Figures 3.1B - 3.1C, both the pause frequency and pause durations are strictly increasing functions of torque. Pauses are more likely to occur when the polymerase experiences a resisting torque, and when a pause occurs, it is likely to be longer in duration if the torque is resisting forward motion. The figures also reflect the case that experiencing an assisting torque is associated with a large elongation velocity along with small pause frequency and pause duration for an individual RNAP.

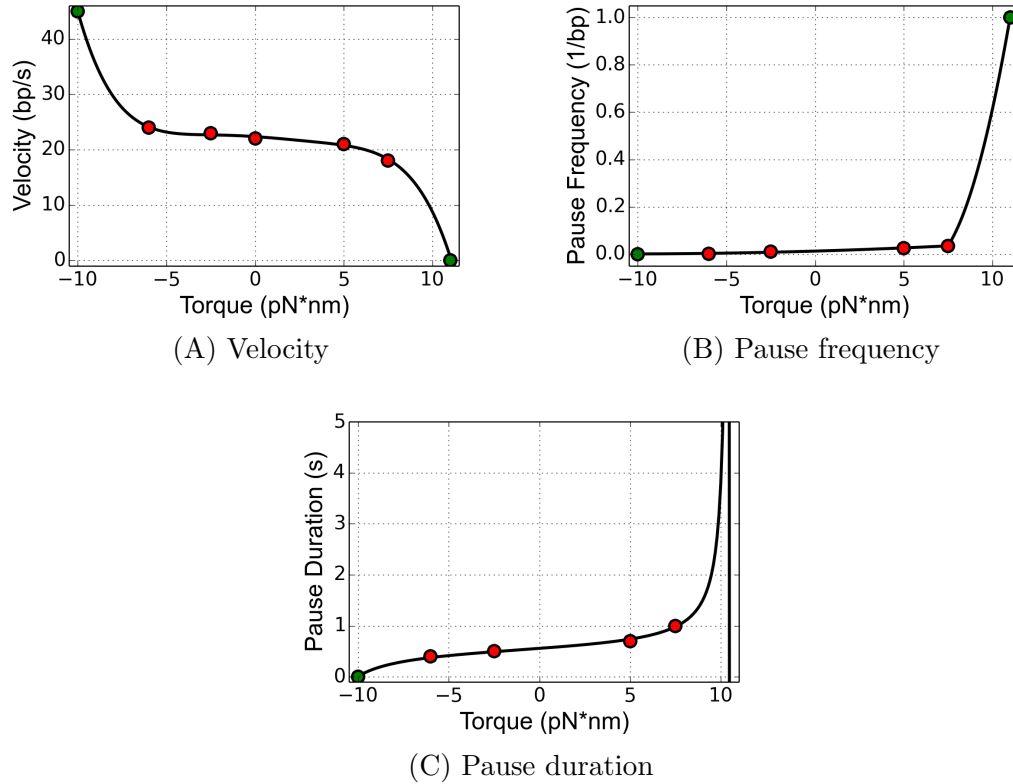


Figure 3.1: The data published by Ma et. al. is presented (red dots) as well as the curve fit to the data (black). The values that we choose based off of information from biology are given for very high and very low torque values (green dots). The equation for each curve can be found in Equations (3.1) - (3.3), respectively.

Fitting the Data Linearly

With limited data points, one may attempt a linear fit without using information such as stall torque. In addition to simulations of the data fits explained above, we also ran simulations using a piecewise linear curve fit. We will present the equations in this section and discuss the results of these simulations in a later section *Results from the Linear Data Fit*.

The data presented in [53] looks almost linear except for the highest torque value of 7.5 pN*nm. Therefore we decided to fit each equation with two linear fits,

one using the data points at -6 pN·nm and 5 pN·nm, the other uses the data points at 5 pN·nm and 7.5 pN·nm. These equations are

$$V(\tau) = \begin{cases} -0.2727\tau + 22.3636 & \tau \leq 5 \\ -1.2\tau + 27 & \tau \geq 5 \end{cases} \quad (3.4)$$

$$F(\tau) = \begin{cases} 0.002277\tau + 0.013864 & \tau \leq 5 \\ 0.004\tau + 0.005 & \tau \geq 5 \end{cases} \quad (3.5)$$

$$D(\tau) = \begin{cases} 0.027273\tau + 0.563636 & \tau \leq 5 \\ 0.12\tau + 0.1 & \tau \geq 5 \end{cases} \quad (3.6)$$

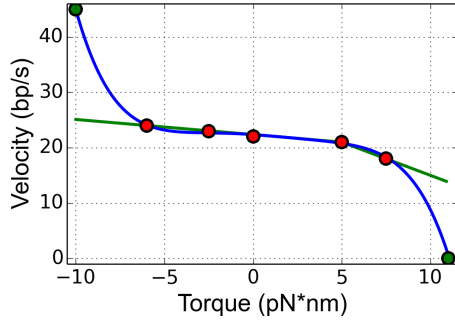
The data points and piecewise linear curve fits, as well as the nonlinear curve fits for comparison purposes, can be found in Figure 3.2.

Simulation

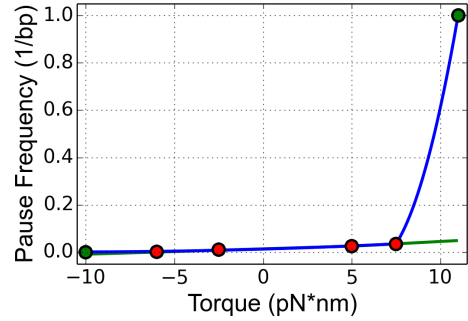
We simulate the transcription process using a Kinetic Monte Carlo algorithm [84, 88] which is often used for simulation of coupled Poisson processes. First, we provide an outline of the computations performed at each step of the simulation following an elongation event. We describe how we incorporated the pauses and recorded the collisions. We will also detail the basic structure and order of the simulation algorithm.

Simulation Events

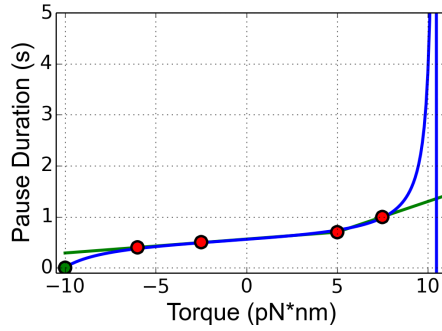
The simulations consist of a series of events that occur for each particle within the system. These events are *elongation* and *initiation*. We assume that the time between



(A) Piecewise linear velocity



(B) Piecewise linear pause frequency



(C) Piecewise linear pause duration

Figure 3.2: The data published by Ma et. al. is presented (red dots) as well as the piecewise linear curve fit to the data (green). The nonlinear curve fit is provided (blue) with our imposed values (green dots) for comparison purposes. The equation for each curve can be found in Equations (3.4) - (3.6), respectively.

two consecutive events for a given polymerase is exponentially distributed. Therefore, given that an event for polymerase P_i just occurred, the amount of time that will elapse prior to the next event for P_i has a probability distribution $P(t) = \lambda e^{-\lambda t}$, where $\lambda > 0$ is an arbitrary rate parameter. The expected value of an exponential distribution is $1/\lambda$, which means that, on average, the time between events is $1/\lambda$ seconds. The greater the value of λ , the more quickly the next event occurs, while events with smaller rate parameters will happen less frequently. Each of the two main particle events has its own unique rate parameter. The first parameter of interest is

related to elongation, and it is denoted by β , where β is the average rate at which a single RNAP elongates from one base pair to the next on the DNA strand in the absence of any pauses or collisions with other RNAPs. This parameter has units of nucleotide per second. The basic elongation event has a rate parameter β , which implies that an elongation will happen every $1/\beta$ seconds on average.

The second rate parameter that governs the basic transcription process is the average initiation rate, $\lambda = (\alpha * \beta)$, where $\alpha \in (0, 1]$ is a constant value and β is the average elongation rate as described above. Here α is used to scale the elongation rate to a lower value causing initiation events to happen less frequently, on average every $1/(\alpha * \beta)$ seconds. The value of α is fixed for each simulation. The incorporation of pauses into the model is achieved through modifications made dynamically to the elongation rate parameter. Next we describe the details of the simulation events, and for each pause event, we describe the adjustments that are made to the various rate parameters.

Incorporation of Pauses Continuing the notation used in previous sections, suppose that an elongation event has occurred for P_i . Once the torques τ_{i+1} , τ_i , and τ_{i-1} are recalculated for RNAPs in Figure 2.1, these values are used to update both the velocity and the pause frequency of all three RNAPs using Equations (3.1) and (3.2) respectively. The velocity is used as the rate parameter, λ in the exponential distribution defined above. Specifically, once τ_i , the torque experienced by P_i , is calculated, the velocity and pause frequency, $V(\tau_i)$ and $F(\tau_i)$ are calculated using Equations (3.1) and (3.2). The time of the next movement of P_i is now governed by the probability distribution $P(t) = V(\tau_i)e^{-V(\tau_i)t}$ with an expected value of $1/V(\tau_i)$. The time of next movement by the RNAP is drawn using this exponential distribution

and recorded as T_i into the triple P_i . That is, from Equation (2.18), the variable T_i is assigned by sampling this exponential distribution using the current value of $V(\tau_i)$.

The likelihood of P_i entering a pause depends on the pause frequency under the calculated torque, $F(\tau_i)$, and once this quantity is calculated, a uniformly distributed random number, $u \in [0, 1]$ is drawn. If $u \geq F(\tau_i)$, the RNAP continues the elongation process. However, if $u < F(\tau_i)$, the RNAP enters a pause. When P_i enters a pause, then the pause duration, $D(\tau_i)$, is computed using Equation (3.3). The velocity of the RNAP is then reset to $V(\tau_i) = 1/D(\tau_i)$, and the time of the next movement, T_i is drawn based on this velocity in the same manner that was described in the preceding paragraph. Therefore the expected value of the time of the next elongation for P_i is $1/V(\tau_i) = D(\tau_i)$. Note here that this duration is updated when a neighboring RNAP translocates (because the value of τ_i changes when that motion occurs), therefore the time of next elongation for the paused RNAP P_i is recalculated, and the pause duration changes accordingly. The simulation scheme also monitors other pieces of information that are crucial for our analysis, in particular the furthest downstream nucleotide occupied by a paused RNAP (position) as well as both the start time and the end time of each pause are recorded for a posteriori analysis purposes.

Incorporation of Collisions The other event that can affect elongation is a collision event. As mentioned previously, a collision occurs when the trailing RNAP, P_i must cease active elongation because the next nucleotide is already occupied by the leading RNAP, P_{i-1} . Since RNAPs cannot overlap nucleotides, or “hop over” each other as seen in specific cases with ribosomes during translation [41, 49, 86], P_i will halt elongation until the next nucleotide is free. When a collision occurs, $V(\tau_i)$ is set to zero and the time of next movement is not drawn until P_i is free to move. Essentially, elongation of P_i is put on hold. This ensures that an elongation event

for P_i , when it is directly behind P_{i-1} , will not be chosen for execution; therefore no polymerases will overlap. When P_{i-1} translocates, the rate of P_i is subsequently updated, and the next time of elongation for P_i is drawn. We consider the end of the collision to be when P_i translocates, as opposed to when the next nucleotide becomes available. The position, start time and end time of each collision is also recorded for later use in our analysis.

To summarize, upon elongation of P_i , the process given in Algorithm 3.2 is performed for P_i and the neighboring RNAPs, P_{i-1} and P_{i+1} . A flowchart detailing the elongation process can be seen in Figure 3.3. The following subsections describe how we can simplify the model to recover the standard TASAP model. We will use this TASAP model for comparison with the ETAM model. A discussion of the parameters used for the numerical results is also presented below.

```

1:  $L_i, L_{i+1} \leftarrow$  using  $n_{i-1}, n_i, n_{i+1}$ 
2: if  $L_i = 0$  then
3:  $V(\tau_i) = 0$ 
4: end if
5:  $\tau_i \leftarrow$  using  $L_i, L_{i+1}$ , and  $L_{0,i}, L_{0,i+1}$ 
6:  $V(\tau_i), F(\tau_i) \leftarrow$  using  $\tau_i$ 
7:  $u \leftarrow$  random(0, 1)
8: if  $u < F(\tau_i)$  Pause then
9:  $D(\tau_i) \leftarrow$  using  $\tau_i$ 
10:  $V(\tau_i) = 1/D(\tau_i)$ 
11: else
12:  $u \geq F(\tau_i)$  No Pause, Do not update  $V$ 
13: end if
14:  $T_i \leftarrow V(\tau_i)exp(-V(\tau_i)t)$ 

```

Algorithm 3.2: This algorithm occurs after the elongation of P_i .

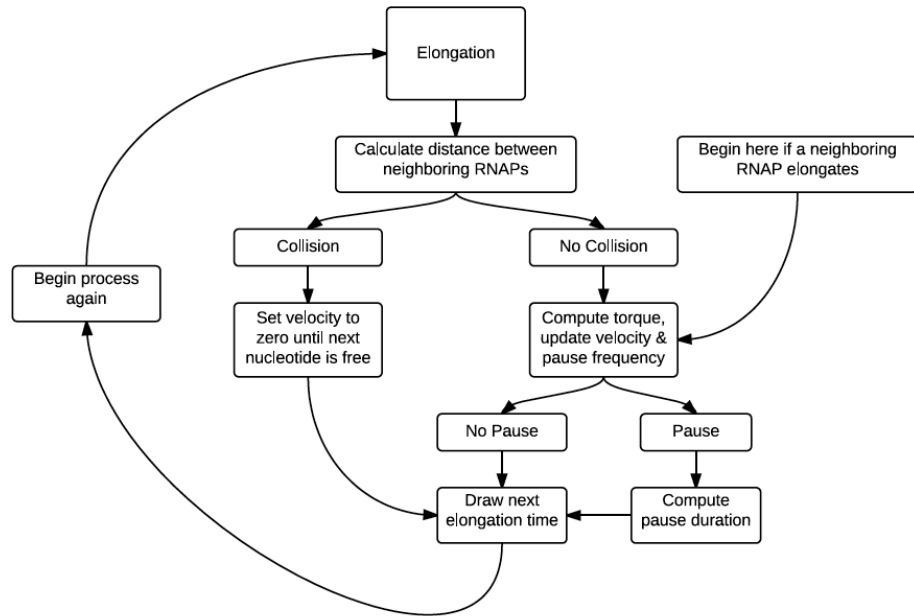


Figure 3.3: A flowchart for our simulation of the elongation process. RNAPs that translocate follow the chart from elongation and continue in order depending on if that RNAP experiences a collision or a pause. If a neighboring RNAP translocates, updates for torque, velocity, and pause frequency are calculated for the affected RNAPs starting with that block of the flowchart.

Remark. *In order to calculate the transcription time of each RNAP, the time of initiation and termination is recorded. If an RNAP initiates but does not complete termination within the simulation time, then the initiation time is not saved, and only the transcription times for the RNAPs that terminate during the time of simulation are used for the subsequent analysis. While all pauses and collisions that occur are recorded, for the analysis presented here we only include the information from pauses and collisions that occur before the final recorded termination time during the simulation.*

Comparison to TASEP

In order to compare the effect of the torque on the RNAP elongation, we will compare ETAM model to the TASEP model which has the same stochastic structure as ETAM, with the key exception that the torque effects are not included. In particular, we use the same flowchart as in Figure 3.3 but the velocity, pause frequency, and pause duration are all constant values that correspond to Equations (3.1), (3.2), and (3.3) with $\tau = 0$. This is equivalent to RNAPs experiencing zero torque for the entire simulation. We use a constant velocity of $\beta = V(0)$ which also gives a constant initiation rate of $\alpha \cdot \beta$. Since the RNAPs are elongating with a constant rate of β and we have boundary conditions corresponding to initiation with rate $\alpha \cdot \beta$ and termination with rate $\gamma \cdot \beta$, this simulation describes the typical TASEP simulation described previously.

Pauses are simulated similarly to the pauses in ETAM. Since the torque value is always zero, the velocity and pause frequency is $V(0)$ and $F(0)$ respectively. These values are held constant throughout the simulation. As described before, upon elongation a uniformly distributed random number, u , is drawn. If $u < F(0)$ the RNAP will experience a pause. The duration of the pause $D(0)$ is employed to recalculate the velocity as $1/D(0)$. Upon completion of the pause, the velocity is reset to $V(0)$, and the time of the next elongation calculated and stored.

Simulation Parameters

The two most important parameters governing the simulation events are the average elongation rate, denoted by β and the scalar that affects the rate of initiation, denoted by α . In the literature, β can vary from about 20 nt/s [53] up to about 90 nt/s [25] depending on several factors including the gene the RNAP is transcribing, environmental conditions, and the specific structure of the gene [82, 83]. The average

elongation rate for RNAPs on an *rrn* operon gene has been shown experimentally to be approximately 90 nt/s [15, 70, 82, 83]. Note that this is different from the data calculated by Ma *et. al.* [53] where they measured the velocity for an RNAP under no torque at approximately 22 bp/s. This measurement is consistent with single-molecule experiments on certain *E. coli* gene sequences [2, 80]. In order to simulate transcription on an *rrn* operon, we multiply the velocity function described by Equation (3.1) by an appropriate a scaling factor, K . The value of K is chosen so that an RNAP under no torque travels at 90 nt/s, which gives $K \cdot V(0) = \beta = 90$ nt/s. Therefore, when drawing the next time for elongation of an RNAP, we use the exponential distribution $KV(\tau)e^{-KV(\tau)t}$, with the expected value $(KV(\tau))^{-1}$.

While mathematically, the initiation parameter α can be as high as 1, in reality the values of α are much less than 1. Simulations were conducted using a variety of initiation rates ranging from $\alpha = 0.0001$, which corresponds to an initiation every 109 seconds on average, to $\alpha = 0.0115$, which corresponds to an initiation every 1.6 seconds on average. According to the literature, an *rrn* operon has approximately 31% coverage of the DNA strand by RNAPs [25]. Accounting for this percentage of coverage on a strand of DNA that is 5450 nts long occupied by RNAPs, each of length 35 nts, corresponds to approximately 50 RNAPs on the strand at any given time. Hence, the choice of the parameter range $\alpha \in [0.0001, 0.0115]$ is meant to mimic conditions ranging from light coverage of the DNA strand to a density of RNAPs that is beyond the experimentally observed conditions described above.

Results

In order to examine the effect of the torque on the transcription simulation we compare several quantities of interest for both the ETAM and TASEP models.

By comparing ETAM and TASEP, we isolate the effect of the torque mechanism. We report the following quantities: the average transcription time, the average pause duration, the average collision duration time, the number of pauses and collisions, and the average transcriptional delay time experienced by an RNAP. Each of the above quantities is calculated over a range of initiation rates $\alpha \in [0.0001, 0.0115]$ using 11 discrete values within this interval. For each value of α , we performed 50 simulations of both ETAM and TASEP and ran the simulation for 10,000 seconds. In each of the fifty simulations we record the start and end time of each RNAP transcription time, and we also record both the beginning and the end time of each pause and collision. When analyzing the results of our simulations, all of the data for each of the runs is compiled and averaged together. The details of these computations and results are given below.

Results from the Nonlinear Data Fit

For comparison purposes, we provide results for both the nonlinear data fit discussed in the section *Incorporating Experimental Data* and the piecewise linear data fit in detailed in the section *Fitting the Data Linearly*. The results for the piecewise linear data fit are included to illustrate different behavior of models that both fit the limited data well. Because the results of both data fits are vastly different from one another, the following sections also illustrate both the value of and the need for more data of the type given in [53], especially description of pause duration, pause frequency and elongation velocity in the regions for extreme values of torque.

Average Number and Average Duration of Pauses In the TASEP model simulations, each RNAP experienced, on average, 70 pauses with an average duration of 0.55 seconds. These results were constant for the TASEP model over the entire

range of initiation rates as seen for the magenta curves in Figure 3.4A and 3.4B respectively. The constant results were expected since the pause frequency, $F(0)$, and pause duration, $D(0)$, were constant throughout the TASEP simulations.

For the ETAM model, the average number of pauses experienced by an RNAP varied significantly over the range of initiation rates included in the simulations. The number of pauses increased monotonically from 202.69 for $\alpha = 0.0001$ to 2169.32 for $\alpha = 0.0115$. This is a 970.3% increase over the range of initiation rates; moreover, for the entire collection of these simulations, the average number of pauses was consistently larger than that of the TASEP model, see Figure 3.4A. For the largest value of α , RNAPs in ETAM experienced 2999% more pauses than RNAPs in TASEP. This result is physically intuitive based on the construction of the ETAM model. Recall that the pause frequency is a strictly increasing function of torque, see Figure 3.1B. If there are more RNAPs on a DNA strand, not only are those RNAPs more likely to interact but the absolute value of the torque they experience is likely to be larger since the distance between RNAPs is smaller, on average, for a strand with a higher percentage of RNAP coverage. Therefore one would expect RNAPs to experience more pauses in an environment with more coverage than they would in an environment with less coverage.

An important and interesting observation that accompanies the preceding results is the data for the average duration time of these transcriptional pauses. Our results indicate that in the ETAM model, the average duration time of the pauses decreases significantly for higher initiation rates, see Figure 3.4B. The duration time decreased 91.7% over the range of initiation rates, from 0.24 seconds when $\alpha = 0.0001$ to 0.02 seconds when $\alpha = 0.0115$. Note that the values of pause duration on the order of 0.02s will not be detectable in the experiments, and so very likely the motion of polymerases at these coverages will experience fewer observable pauses. At the highest value for α ,

the average pause duration time is 96.4% lower in ETAM than in TASEP. We propose that the effects of the torque mechanism on the average duration time of pauses can be summarized as follows. While an RNAP is more likely to pause in regimes with higher initiation rates, the effects of torque, that is, the “push from the back” and the “pull from the front,” are stronger when the neighboring elongating RNAPs are closer to the paused RNAP. Therefore, the paused RNAP can be pushed or pulled “out” of its pause state and into active elongation by means of these torsional effects much more quickly than in a regime where the pause duration time is determined by purely stochastic effects.

Number and Duration of Collisions We propose that the torque mechanism of the ETAM model allows transcribing RNAPs to maintain their spacing relative to their neighbors and to decrease the number of collisions (as described in the section *Incorporation of Collisions*) that occur among polymerases. In order to investigate this hypothesis, we monitor and record the average number of collisions that occur and the time durations of those collisions for both the TASEP and the ETAM models. The initial intuitive expectation is that one would observe an increase in the number of collisions as the initiation rates increase (or as the percent of coverage increases). That is, for a larger coverage of RNAPs on the DNA strand, collisions should become more likely. Likewise, if there are very few RNAPs on the DNA simultaneously, collisions are unlikely. As shown in Figure 3.4C, the data for the average number of collisions behaves as expected. However, the number of collisions increases much more rapidly in the case of the TASEP model than for the ETAM model. For the highest initiation rate simulated, RNAPs in the TASEP model experienced 1026.5% more collisions than RNAPs in the ETAM model. In particular, results show that approximately 550 more collisions per RNAP occur within TASEP; we observe an

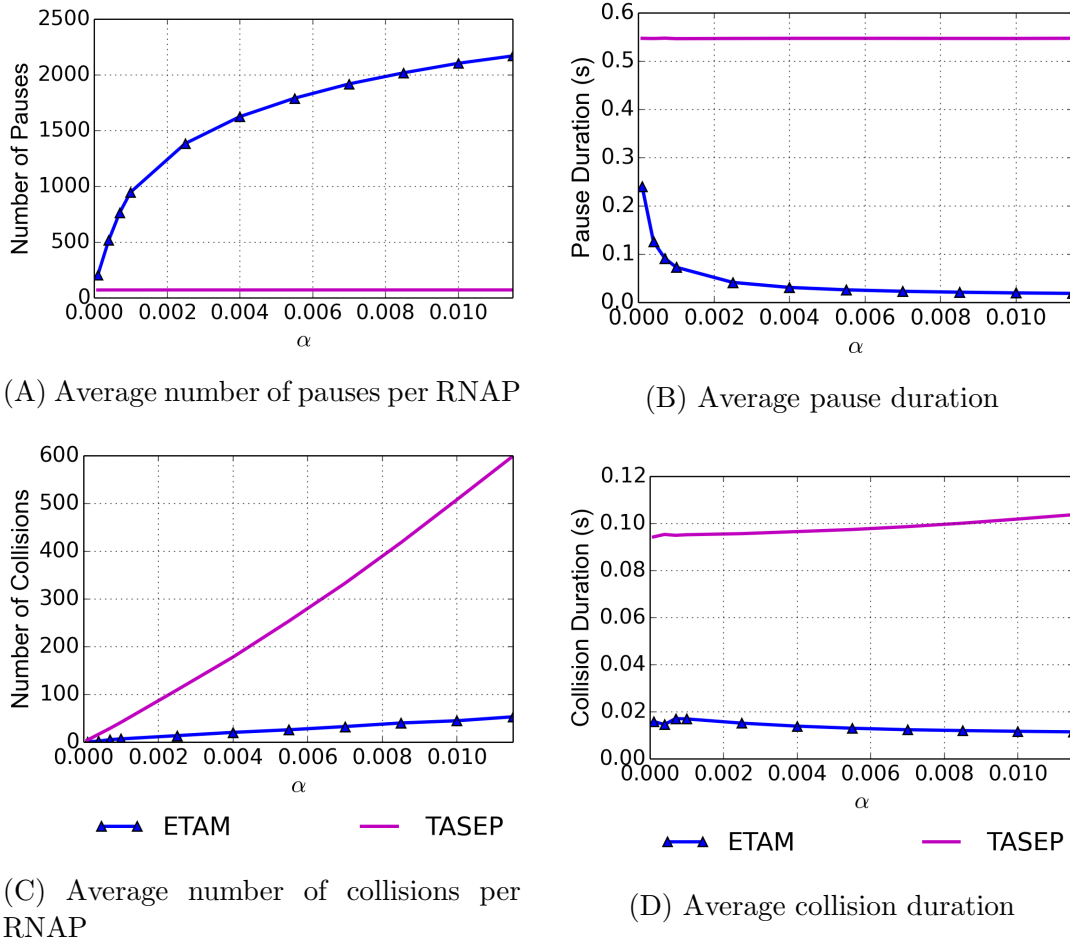


Figure 3.4: Over increasing initiation rates determined by α , the pause and collision results are presented for the ETAM model (blue triangles) and the TASEP model (magenta). The number of pauses and collisions are presented as the averaged number per RNAP. RNAPs in the ETAM model experience significantly fewer collisions and shorter pause durations than their TASEP counterparts.

average of 53.18 collisions per RNAP in the ETAM model and an average of 599.09 collisions per RNAP in the TASEP model. For each of the models, the data sets shown in Figure 3.4C were fit with a linear least squares model, and the average number of collisions experienced by an RNAP has approximately linear growth over the range of α values for both cases. The linear fits can be seen graphically in 3.5,

and the equations for the lines $C_\tau(\alpha)$ for ETAM and $C(\alpha)$ for TASEP are given by

$$C_\tau(\alpha) = 4609.65\alpha$$

$$C(\alpha) = 52175.4\alpha$$

Computing a ratio of the slopes of these two lines, we observe that the number of

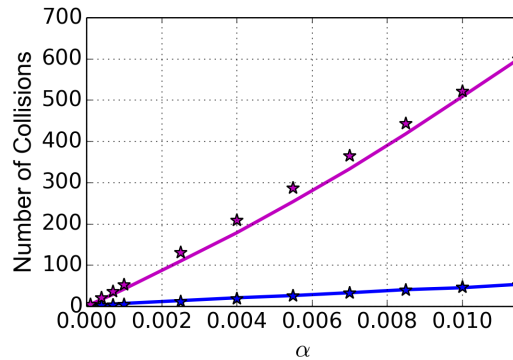


Figure 3.5: Linear fit for the average number of collisions experienced per RNAP. The collisions in the results for TASEP (magenta) and compared to the linear fit (magenta stars) and similarly for the ETAM results (blue and blue stars).

collisions for the ETAM results is growing at approximately 9% of the rate of the number of collisions of the TASEP results.

Given that RNAPs in the TASEP model collide much more often, the average duration time of each collision becomes very crucial. Figure 3.4D shows that for the largest initiation rate, $\alpha = 0.0115$, the average duration time of a collision is significantly longer in TASEP, 0.104 seconds, than in ETAM, 0.011 seconds. We believe these results indicate that the torque is contributing to RNAP's ability to maintain a degree of spacing and distance with their neighbors, thereby reducing the number of collisions that occur for the ETAM case. In addition, when collisions occur in the ETAM model, they tend to be very short. Figure 3.4D indicates that RNAPs

in the TASEP model tend to experience collisions that last approximately five times as long as those in the ETAM model. Moreover, this effect is consistent over the entire range of initiation rates included in the study.

To summarize the inclusion of torque effects into the stochastic model seems to allow the RNAPs to dynamically manage elongation velocity and spacing so as to avoid collisions and traffic jams. While the RNAPs experienced more pauses in the ETAM simulations, these pauses were significantly shorter in duration than those of the TASEP simulations, with far fewer collisions and shorter collision durations. Naturally, the preceding results lead to the main question of how torque affects the average transcription time of the gene – arguably the most important question for the gene’s survival and proliferation.

Average Transcription Time Experimental results from physicists and biologists give an average transcription time for the *rrn* gene of approximately 60 seconds [25,60], and in [25], the authors also assert that the *rrn* gene is, on average, approximately 31% covered. This corresponds to an average velocity of 90 nt/s. With the physical parameters used for both the TASEP and ETAM model simulations, we attempt to mimic the biological case of transcription of this gene. We now examine the average transcription time per RNAP for each of the models. In order to obtain the average transcription time per RNAP, the transcription time for each RNAP within the simulation is recorded, and these values are averaged over all of the RNAPs with successful transcriptions for each specific initiation rate. The results can be seen in Figure 3.6A where data is presented for two situations. In addition to the simulations for both TASEP and ETAM models with pauses, we also performed numerical simulations for both models in the case of no pauses for comparison purposes. We refer to the case without pauses as the “baseline” case for each model. While the

RNAPs could still experience collisions for the baseline case, the number of collisions was significantly lower, see the curves with the dotted lines in Figure 3.6A. This baseline model allows us to calculate the transcription time for an RNAP without any transcriptional delays caused by pauses, and we note that for both models, the average transcription time is close to 60 seconds and agrees well with numbers reported in the literature. For $\alpha = 0.0115$, we see a 61.23 second transcription time for the baseline TASEP model and slightly faster transcription time of 54.7 seconds for the baseline ETAM model.

Examining the case where transcriptional pauses are introduced into each of the models, we see very different effects, and these are shown in the solid curves of Figure 3.6A. For $\alpha = 0.0115$, the average transcription time for TASEP is approximately 156.25 seconds, which for a DNA strand of 5450 nucleotides corresponds to an average velocity of 34.88 nt/s. This rate is significantly lower than the 90 nt/s resulting from an experimental average transcription time of 60 seconds for the *rrn* gene reported in [25,60]. In contrast, for the ETAM model at $\alpha = 0.0115$, the average transcription time was approximately 97.73 seconds, which corresponds to an average velocity of 55.77 nt/s. This average velocity agrees much better with the experimental values. We believe that the significant difference in average transcription times between the two models is attributable to the many collisions and subsequent delays experienced by RNAPs in the TASEP model. This is carefully quantified in the following section.

Average Transcriptional Delay One calculation of particular interest is the average transcriptional delay experienced by an RNAP when it is subject to transcriptional pausing. During the transcription process, an RNAP experiences both pauses and collisions, each with a certain time duration. These interruptions cause active elongation of the RNAP to cease until the RNAP is able to move again. The

Table 3.1: Results for ETAM model: percent of the strand covered by polymerases, average transcription time per RNAP(s), average collision delay per RNAP (s), average pause delay per RNAP(s), and average total delay per RNAP (s).

α	% Coverage	Time	Collision Delay	Pause Delay	Total Delay
0.0001	1.26	109.12	0.01	48.62	48.63
0.0004	3.42	124.44	0.03	64.70	64.73
0.0007	5.51	128.88	0.09	69.59	69.68
0.001	7.52	127.91	0.12	68.89	69.01
0.0025	14.88	115.36	0.21	57.12	57.33
0.004	20.55	108.04	0.29	50.17	50.46
0.0055	25.42	103.99	0.34	46.33	46.67
0.007	29.61	101.47	0.41	43.90	44.31
0.0085	33.31	99.85	0.48	42.31	42.79
0.01	36.81	98.59	0.53	41.08	41.61
0.0115	39.70	97.73	0.61	40.20	40.81

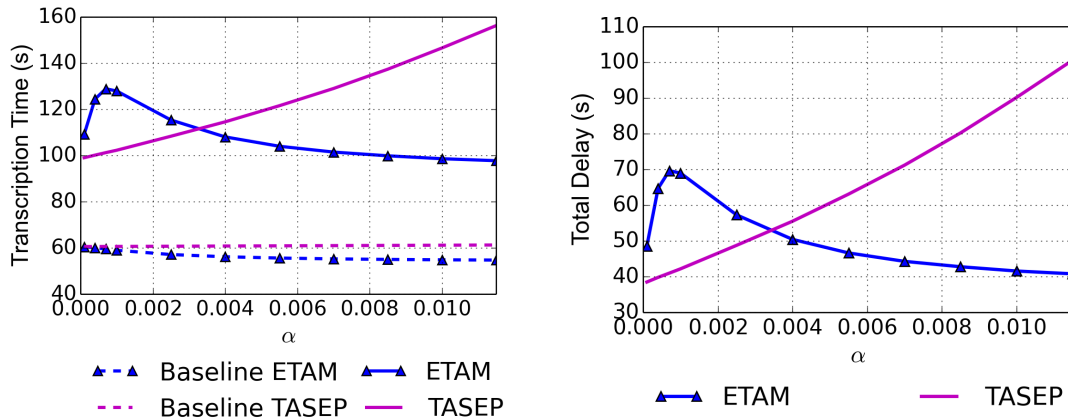
amount of time that an RNAP is unable to translocate contributes to the delay that the RNAP experiences. To quantify this concept, we define the average total delay to be sum of both the average delay due to pauses and the average delay due to collisions. The average total delay is computed with the formula

$$\begin{aligned}
 \text{Delay}_{\text{total}} &= \text{Delay}_{\text{pause}} + \text{Delay}_{\text{collision}} \\
 &= (\text{ave \# of pauses per RNAP}) \cdot (\text{ave pause duration}) \\
 &+ (\text{ave \# of collisions per RNAP}) \cdot (\text{ave collision duration})
 \end{aligned}$$

the results of which can be seen in Figure 3.6B. Specific values of the total delay over the range of initiation rates for each of the ETAM and TASEP models can be found in Tables 3.1 and 3.2, respectively. The average delay experienced by an RNAP in the TASEP model increased from 38.52 seconds when $\alpha = 0.0001$, to 100.61 seconds

Table 3.2: Results for TASEP model: percent of the strand covered by polymerases, average transcription time per RNAP(s), average collision delay per RNAP (s), average pause delay per RNAP (s), and average total delay per RNAP (s).

α	% Coverage	Time	Collision Delay	Pause Delay	Total Delay
0.0001	1.21	99.02	0.40	38.12	38.52
0.0004	2.90	100.15	1.57	38.26	39.83
0.0007	4.55	101.23	2.71	38.30	41.01
0.001	6.20	102.27	3.95	38.26	42.21
0.0025	14.26	108.33	10.45	38.33	48.78
0.004	21.75	114.56	17.20	38.34	55.54
0.0055	29.05	121.54	24.72	38.41	63.13
0.007	36.08	129.00	32.82	38.41	71.23
0.0085	42.89	137.35	41.85	38.41	80.26
0.01	49.40	146.55	51.75	38.43	90.18
0.0115	55.26	156.25	62.14	38.47	100.61



(A) Average transcription time

(B) Average delay per RNAP

Figure 3.6: Over increasing initiation rates $\alpha \cdot \beta$, we present our results for average transcription time and average total delay per RNAP. The ETAM model (blue triangles) and TASEP model (magenta) are both plotted for comparison. The dashed lines represent baseline simulations with no pauses.

when $\alpha = 0.0115$. Conversely for the ETAM model, the average delay *decreased* from 48.63 seconds to 40.82 seconds over the same range of initiation rates.

For the ETAM model, the eventual decrease in delay values for increasing initiation rates (and thus increasing coverage), is evidence to suggest that the torque contributes to an increase in efficiency with multiple RNAPs transcribing simultaneously. Moreover, the comparison of these two models allows us to observe that, in the TASEP model, the phenomena that drives the total delay is that of the delays due to collisions (as opposed to the delays due to transcriptional pauses experienced). Table 3.2 clearly demonstrates that, for the TASEP model, the delays which the RNAPs experience due to pauses remains virtually constant over the entire range of initiation parameters and that the increasing behavior of the total delay is almost exclusively attributable to the delays due to collisions. In contrast, Figure 3.4C and Table 3.1 demonstrate that the torque mechanism included in the ETAM model prevents many collisions from happening, and it also decreases the amount of delay that RNAPs incur from those few collisions that actually do occur. This is especially apparent with the higher initiation rates. The torsional interaction between RNAPs leads to far more efficient transcription, especially in the case of high coverage of the DNA strand.

The torque provides a mechanism for the RNAPs to interact and cooperatively prevent collisions from occurring. If an RNAP pauses, the trailing RNAP will slow down and/or enter a pause with a much higher probability due to the increasing torque applied to it as its elongation continues. The trailing RNAP will likely enter a pause or it may push the leading RNAP into active elongation before a collision occurs. The evidence of this interaction can be seen in the difference in the number of collisions experienced per RNAP in the two models in Figure 3.4C.

The comparison of the ETAM and TASEP models allows us to propose that torque is an important mechanism in regulation of transcription. Neighboring RNAPs may interact with each other using torque to optimize their elongation efficiency as a

group effort as opposed to an individual process as suggested in [29]. In this paper, Epshtein *et. al.* show that transcription times are faster with multiple RNAPs as opposed to a single molecule. Our ETAM simulations for $\alpha = 0.0001$ had an average transcription time per RNAP of 109 seconds, with an average time between initiations of 109 seconds. This simulation was essentially transcription of a single molecule. The average transcription time for $\alpha = 0.0115$ decreased nearly 12 seconds from the lower value of α . This is largely do to the paused RNAPs being pushed back into active elongation by their neighboring polymerases, a phenomenon suggested by Epshtein *et. al.*.

Results from the Linear Data Fit

The small number of data points given in [53] require us to extrapolate in order to characterize the relationships between the torque and the various physical quantities (elongation velocity, pause frequency and pause duration) for the ETAM model in the extreme cases where the torque values were near the melting or the stall cases, that is, where the absolute value of the torque is large. For those two cases, the functional relationships are constructed based on experimental biological data from [53] as well as other literature. The previous section focused on that combination of information to inform the mathematical model; however, the choices made for the functional relationships were still somewhat arbitrary. The results of this section illustrate that the choices made for the case of torque values near either melting or stall are very crucial to the results of the mathematical model. The most important result shows that the cooperative effect discussed in the previous section is observed for some choices of data fit constructions but not for others. Hence, the results reported in this section demonstrate the need for more experimental data over a larger range of torque values in order to produce a realistic mathematical model.

Though the average transcription time and total delay functions displayed for the ETAM model in the previous section are largely concave up, the results are very different when using the piecewise linear curve fit discussed in the section *Fitting the Data Linearly* for ETAM simulations. With the piecewise linear fit, the average transcription time and total delay have the same initial behavior as the nonlinear fit but the results continue to increase and stay concave down for the duration of the simulations. This can be seen in Figure 3.7.

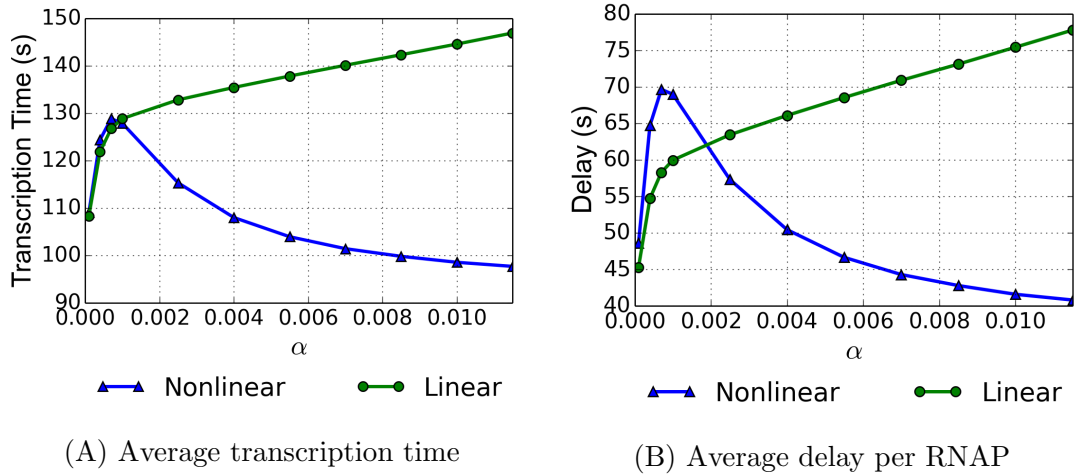


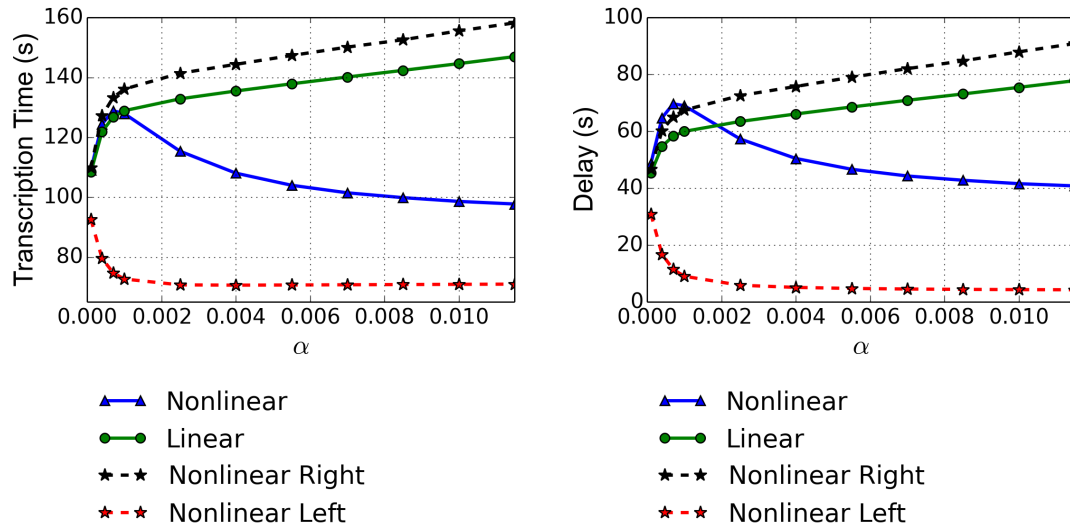
Figure 3.7: Over increasing initiation rates $\alpha \cdot \beta$, we present our results for average transcription time and average total delay per RNAP in the ETAM model. The nonlinear model (blue triangles) and piecewise linear model (green dots), both detailed in previous sections, are plotted for comparison.

For $\alpha = 0.0001$ the average transcription time was approximately 108 seconds with a total delay of 45 seconds. This increased over the range of α , with a final average transcription time of 147 seconds and a delay of 78 seconds. This is a 36% increase in transcription times and a 73% increase in total delay. If we compare these to the values calculated for its nonlinear counterpart, we see that for the largest value of $\alpha = 0.0115$, a 50% increase in transcription time and a 91% increase in

delay for the piecewise linear fit. While the result with the piecewise linear fit is much closer to a 60 second average transcription time than TASEP, the cooperative effect of the nonlinear fit from the previous section is diminished for this particular choice of piecewise linear fit. The cooperative effect of the ETAM in the previous section shows a decrease in transcription time as the coverage of the DNA strand increases to biologically relevant situations, and that effect is dramatically different than both the piecewise linear fit for ETAM and the TASEP case where transcription time monotonically increases as the coverage of the DNA strand increases.

The Importance of the Pause Duration

Because of the drastic difference in results from a nonlinear fit to a piecewise linear fit, we investigate which function or combination of functions drives the difference. Below we explore several possible combinations of piecewise linear and nonlinear data fit choices, and we find that the behavior is clearly being driven by the pause duration function, specifically the pause duration for very low torque values near the melting threshold, see Figure 3.8. This figure compares the results from the nonlinear fit (Figure 3.1), and the piecewise linear fit (Figure 3.2) with two other curves. The curve labeled “Nonlinear Left” are the results of the piecewise linear fit for all functions except the pause duration. The pause duration instead is a combination of the nonlinear fit from $[-10,5]$ and the piecewise linear fit from $[5,11]$. Similarly the curve labeled “Nonlinear Right” are the results for the nonlinear fit from $[5,11]$ and the piecewise linear fit from $[-10,5]$ in the pause duration function. As you can see, the results for the piecewise linear fit and the nonlinear right fit are qualitatively similar. The most surprising are the results for the nonlinear left fit. For this curve fit, under large values of α , the RNAPs experience nearly 60 second transcription



(A) Average transcription time

(B) Average delay per RNAP

Figure 3.8: In addition to the nonlinear fit (blue triangles) and the piecewise linear fit (green dots), we also present our results for a piecewise linear fit except for the pause duration on the interval $[-10,5]$, “Nonlinear Left” (red stars), or on the interval $[5,10]$, “Nonlinear Right” (black stars).

times with virtually no delay. The reason for this difference in results can be seen as we investigate the graph of pause duration for torque values near -10 pN·nm.

Figure 3.9 depicts the nonlinear pause duration and the piecewise linear pause duration for torque values in the interval $[-10,5]$. It’s important to note that pause durations are not fixed, but are calculated upon entering a pause and then recalculated when neighboring RNAPs elongate. However, when an RNAP is paused, the recalculated pause duration will always be smaller than the original because it is either being pulled from in front or pushed from behind by the elongation of a neighbor. Therefore even if the original pause duration is large, the recalculated pause duration is likely to be small, as evidenced by the extremely small average pause durations in high coverage environments. Hence, the range of pause duration values for the case of small torque is very important. In the case of the nonlinear

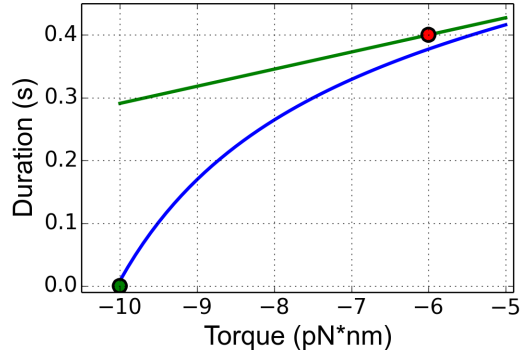


Figure 3.9: The differences in the pause duration for the piecewise linear fit (green) and the nonlinear fit (blue) at low torque values is highlighted here. This difference accounts for the the very different results in average transcription time and average total delay depicted in Figure 3.7.

data fit with low torque values, the pause duration can be set as low as zero, and the RNAP is released from the pause state and is free to elongate. However for the case of the piecewise linear data fit, the lowest pause duration possible is approximately 0.3 seconds. With the large amount of pauses experienced by an RNAP, the difference in pause duration for low torque values is driving the difference in behavior between the nonlinear and piecewise linear data fit.

To finish our discussion on pause duration we again consider Figure 3.8, we would like to highlight the difference in the results of the “Nonlinear” data fit and the “Nonlinear Left” data fit whose pause duration is linear for $\alpha \in [5,10]$ and nonlinear for $\alpha \in [-10,5]$. If the behavior of the results is driven by the pause duration for low torque values, how do we account for this difference, as they have the same pause duration for those torque values? This difference can be attributed to the pause frequency function. Recall, the “Nonlinear Left” data fit has the piecewise linear fit for velocity and pause frequency. The piecewise linear fit for large torque values would give a pause frequency of 0.049 when the torque is 11 pN·nm, as opposed to a

pause frequency of 1 for the nonlinear fit. The RNAPs will experience significantly fewer pauses with the piecewise linear pause frequency, encountering on average 146 pauses when $\alpha = 0.0115$, as opposed to 2169 pauses. Regardless of how short the duration is, these minor interruptions in elongation can have a large effect on the overall transcription time of an RNAP. With this in mind, we investigate how the value of the pause frequency function when torque is 11 pN·nm can affect the results of the model.

Different Pause Frequency Functions

As mentioned earlier, the choices for data fit in the cases of very high torque values and very low torque values were somewhat arbitrary with the limited data points. One choice which has proven to be crucial is to use a pause frequency of 1 when torque is 11 pN·nm, the approximate value for stall torque. Here we illustrate the the impact of that choice on our results. Using the nonlinear data fit for pause duration and velocity, as in Equations (3.3), and (3.1), we perform a set of calculations for various choices for the function value when torque is 11 pN·nm. We continue to fit the pause frequency function using quadratic functions similar to Equation (3.2); however the value of the pause frequency function when the torque value is 11 pN·nm is allowed to range over a variety of values smaller than 1 in order to compare the results, see Figure 3.10. The values for pause frequency used when torque was 11 pN·nm are $\{0.2, 0.5, 0.6, 0.7, 0.8, 0.9\}$. We also report the results for the original pause frequency function that takes on the value of 1 at the stall torque. Also included are the results for the case where the pause frequency function data is fit with one quadratic function which has a value of approximately 0.05 when at stall torque. The average transcription time and total delay for all of these different choices for data fit can be seen in Figures 3.11A and 3.11B respectively. Pause frequency increasing

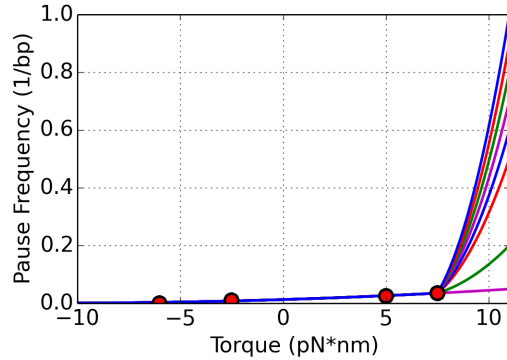


Figure 3.10: Pause frequency function where the value at the end point, 11 pN·nm is raised. The data (red dots) is fit with one quadratic function up to 7.5 pN·nm. A second quadratic function is used to fit the end point, with the exception of the lowest function which is the original quadratic for the entire interval.

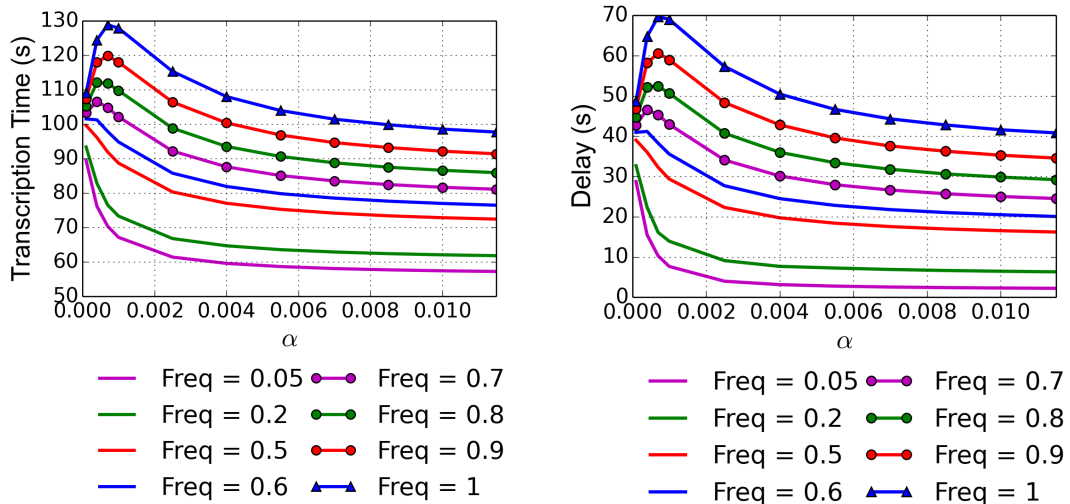


Figure 3.11: Over increasing initiation rates $\alpha \cdot \beta$, these results show average transcription time and total delay over different fits of the pause frequency function. The legend labels the pause frequency at the highest torque value.

to 0.05 and 0.2 have the fastest transcription times, both being very close to a 60 second transcription time and in the case of the lowest pause frequency, actually being faster than a 60 second transcription time. These two pause frequency functions give

Table 3.3: The average number of pauses per RNAP, average pause duration, and average pause delay per RNAP due to pauses experiences per RNAP for the different pause frequency functions. Frequency represents the value of the pause frequency function when torque is 11 pN·nm. All of the data reported is for $\alpha = 0.0115$.

Frequency	# of Pauses	Duration (s)	Delay (s)
0.05	162	0.013	2.106
0.2	433	0.014	6.062
0.5	1013	0.016	16.208
0.6	1219	0.016	19.504
0.7	1443	0.017	24.531
0.8	1668	0.017	28.356
0.9	1909	0.018	34.362
1	2169	0.018	39.042

results that agree extremely well with experimental data for large values of α . If we consider the difference in the average number of pauses per RNAP under these different pause frequency functions, we can understand the difference in delay. Table 3.3 shows the average number of pauses per RNAP, average pause duration, and the corresponding delay due to pauses for the difference pause frequencies, when $\alpha = 0.0115$. By examining the delay values in Table 3.3 and comparing these to Figure 3.11B, we can see that the pause delay contributes to the vast majority of the total delay each polymerase experiences. The pause delay is influenced mostly by the number of pauses per polymerase which is a direct result of the pause frequency function.

Another interesting result is the shift in behavior from frequency 0.5 to frequency 0.6. The results for lower frequency values are concave up over the range of α . However with pause frequency up to 0.6 we begin to see a global maximum when $\alpha = 0.0004$. For pause frequency greater than or equal to 0.8 the global maximum is when $\alpha = 0.0007$. For values of α larger than 0.0007 the coverage of the DNA

strand is large enough that the RNAPs feel a substantial effect from their neighboring polymerases and begin to experience a cooperative effect. It is in this range that while the polymerases experience more pauses, the decrease in the pause duration is enough to shorten the total delay. We believe this is the range of initiation values where the RNAPs are now close enough for the torsional interaction to push a paused RNAP back into elongation, as proposed by Epshtein *et. al.* [29], substantially quicker than they had been previously. As the values of α continue to increase from here, this cooperation becomes even more pronounced. Regardless of pause frequency, the overall cooperative behavior is clear by the decrease in delay and transcription times for the larger coverages.

Mathematically, choosing a value for pause frequency close to 1 or equal to 1 when torque is 11 pN·nm is the natural choice as this is the stall torque. However, if we want to fit our model to the experimental data, we would choose a pause frequency close to 0.2. Here we have 30% coverage of the DNA strand with an average transcription time of just under 62 seconds at the highest initiation rate. This agrees very well with the results presented by Neuman *et. al.* [60]. In order to properly tune our ETAM model, more data is necessary near the stall torque and melting torque.

In order to determine, just how important accurate measurements near the stall torque and the melting torque could be, we investigated the how often these torque values were calculated over our range of α . Figure 3.12 depicts the number of times each torque value is calculated in a simulation as a percentage out of the total number of torque values that were computed, displayed as a histogram. We show the results for both the baseline simulation (no pauses), and the pause simulation for $\alpha = 0.0001$ (Figure 3.12A) and $\alpha = 0.0115$ (Figure 3.12B). In these simulations, we counted the number of times the torque values of 10 pN·nm and -10 pN·nm were computed and

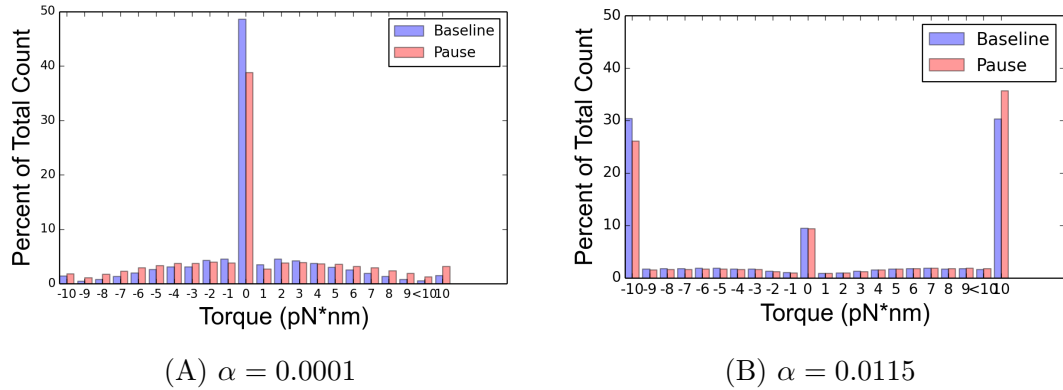


Figure 3.12: The percentage of times a torque value is calculated in a baseline (blue) simulation and a pause (red) simulation is presented here as a histogram. The histogram bars at -10 and 10 is the percentage that those values were computed. Each of the other bars represents the percentage between $(-10,-9]$, $(-9,-8]$, etc., with the exception of the bars at the label < 10 . The histogram bars at < 10 represents the percentage of torque values computed in $(9,10)$.

plotted this percentage as histogram bars at -10 and 10 respectively. In between these values we counted the number of times a torque value fell in the interval $(-10,-9]$ and plotted that percentage in the histogram bars at -9. The percentage of torque values in $(-9, -8]$ were plotted in the histogram bars at -8, and so on, up to the torque value of 9. The percentage of values in the interval $(9,10)$ were plotted under the label of < 10 . As you can see from Figure 3.12, for our lowest initiation rate ($\alpha = 0.0001$) computed a torque of 0 pN·nm approximately 50 percent of the time in the baseline simulation and just under 40 percent for the pause simulation. This is to be expected as many of the RNAPs for this value of α are transcribing as single molecules and therefore would not generate torque values away from 0 pN·nm. For $\alpha = 0.0115$, the results are very different. The interaction between RNAPs causes the 0 pN·nm to be computed only 10% of the time. The extreme values of -10 pN·nm and 10 pN·nm are calculated the most. In the pause simulation -10 pN·nm is computed approximately 25% of the time and 10 pN·nm about 35% of the time. This high percentage is due

to the fact that the RNAPs are very close together at this high density and therefore the interactions between them are extremely strong. As a result, these torque values and how we model the velocity, pause duration, and pause frequency at these values, has a large impact on the ETAM results.

Discussion

By incorporating the torque mechanism into a basic TASEP model of transcription we are able to see a cooperative effect among transcribing RNAPs. This effect was noted experimentally in 2003 by Epshtein and Nudler [29]. At the time, the mechanism causing this behavior was unclear. After the recent developments by Ma *et al* [53], and results from our model simulations, we propose that the torsion on the DNA caused by RNAP transcription is allowing the RNAPs to communicate with each other in order to maintain proper spacing and to increase the rate of transcription.

The cooperation between RNAPs is clearly seen from the results of our stochastic model, ETAM, which incorporates torque into a basic TASEP model. We compare this model with the TASEP model to isolate the effect of the torque. There are two results that clearly demonstrate this cooperative effect: the average number of collisions each RNAP experiences and the average pause duration.

With a high initiation rate of RNAPs onto the DNA, each polymerase experiences on average 550 fewer collisions with neighboring RNAP during the course of transcription when the elongation is regulated by torque. This is a direct result of the torque allowing the RNAPs to communicate with the polymerases that are closest to them. During a simulation of TASEP, an RNAP will elongate until it either pauses or is stopped because the next nucleotide downstream is occupied by a paused polymerase. With ETAM, as an RNAP elongates close to a paused RNAP,

the resisting torque experienced by the elongating RNAP makes it much more likely to pause. At the same time, the paused RNAP experiences an assisting torque from the elongating RNAP which can push out of a pause and into active elongation. This interaction prevents an average of 550 extra collisions from occurring.

The average pause duration is the other quantity most affected by the torque. With ETAM the pause duration is not a fixed quantity but can be dynamically recalculated to account for the actions of neighboring RNAPs. Pause durations can be shortened when the polymerases surrounding the paused RNAP elongate. Simulation of ETAM produced on average 0.02 second pauses as opposed to 0.55 second pauses simulated in TASEP.

When comparing the average transcription time per polymerase we see an even more striking difference. The ETAM model shows polymerases experiencing an average transcription time of 97.73 seconds as opposed to 156.25 seconds in TASEP. This is a difference of nearly a minute! The delay a polymerase experiences as a result of collisions and pauses has the largest effect on the overall transcription time. With fewer collisions and shorter pause durations, the RNAPs simulated in ETAM have significantly less delay resulting in far more efficient transcription.

As promising as these results are, they depend on how we fit limited data for velocity, pause frequency, and pause duration into functions of torque. With a small amount of data points available and no data for values near the stall torque and melting torque, the functional modeled used for these functions at those end points is somewhat arbitrary. These high and low values for torque are calculated quite often with a high density of RNAPs on a DNA strand since the torsional interactions are so strong. As a result, the results of ETAM depend on the choice of these functions of torque in the regions where no experimental data is currently available.

As nanotechnology continues to improve, our hope is that data will become available for velocity, pause duration, and pause frequency at both very low and very high torque values. This will allow us to better fit our model to the data without needing to make assumptions for the extreme cases. Even with the limited data, the cooperation effect is evident in the ETAM results, with shorter transcription times in the simulations for the range of high initiation rates. With more polymerases transcribing DNA simultaneously, each RNAP experiences less delay than RNAPs transcribing with a smaller amount of polymerases. Transcription should be viewed as a group effort, with torsional interactions allowing all of the RNAPs to transcribe more efficiently.

BROWNIAN RATCHETS

RNAP elongation can be thought of as having two different phases: translocation, and pausing. During translocation, the RNAP moves forward one basepair while adding one ribonucleotide triphosphate (NTP) to the nascent RNA chain [1,21]. The order of these steps has been under debate in the past. One proposed model is the power stroke model where the NTP binds to the RNA chain releasing pyrophosphate which triggers translocation. However, it is now accepted that RNAPs act as a Brownian ratchet, where RNAPs translocate as a result of thermal energy and this movement is stabilized with the addition of NTP [1,5,6,21,36,37,50,75]. Specifically, the RNAP oscillates between pre-translocated and post-translocated states until the binding of NTP stabilizes the movement favoring forward motion [1,21].

During elongation, RNAPs experience pauses: long pauses (> 25 s), and short ubiquitous pauses (< 25 s). The long pauses belong to two different classes. The first class is caused by hairpins in the nascent RNA. During a hairpin pause, the RNAP is thought to remain in the pre-translocation state hindering further nucleotide addition [81]. The second class of long pauses is backtracking pauses, thought to be caused by weakness of the RNA-DNA hybrid in the transcription elongation complex [47,62]. During backtracking pauses, the RNAP slides upstream on the DNA approximately 5 bp then recovers its original position [73]. Unlike the longer pauses, the origin of short ubiquitous pauses is unknown. Some researchers believe they are due to a rearrangement of the structure of the RNAP [4,61]. Others believe that a significant portion of short pauses are due to backtracking [27]. Regardless of the origin, the ubiquitous pauses are short in duration and occur approximately every 100 bp [61]. Since the elongation of RNAPs can be modeled as a Brownian ratchet, we have decided to model both translocation *and* pausing in this manner.

RNAP translocation as a Brownian ratchet can be viewed as an example of a one dimensional random walk [10, 38, 43]. We model RNAP motion as the motion of a particle in a periodic potential that describes the energy that is both added and subtracted during translocation with the translocated state being at the minima of the potential, as described in [78]. RNAP motion is largely downstream on the DNA strand with the rare upstream movement (as in backtracking pauses). In a Brownian ratchet, a particle experiences symmetric Brownian motion which is then stabilized in the preferred direction, as is the case with the addition of NTP for RNAP translocation. This asymmetry can be achieved in a Brownian ratchet model by employing an asymmetrical, periodic “ratchet” potential, such as a sawtooth potential shown in Figure 4.1. With this potential, a particle is more likely to escape the “well”,

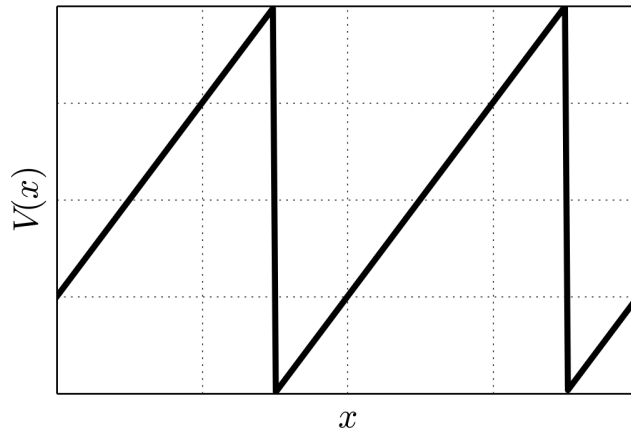


Figure 4.1: The sawtooth function is often employed in Brownian ratchets as it is an excellent example of an asymmetrical, periodic potential, $V(x)$.

or minima, to the right up the sloped ramp than it is to exit left over the “wall”. An alternative to the asymmetrical potential that will still produce a preferred direction of motion is to subtract a tilting force term from the potential [66]. In this case the

potential is of the form

$$V(x) = U(x) - Fx \quad (4.1)$$

where $U(x)$ is periodic with period L and F is the tilting force. If $F < 0$ the force would pull particles to the left and if $F > 0$ the force pulls particles to the right. In the context of DNA, the RNAPs would drift upstream on DNA if $F < 0$, or downstream on DNA if $F > 0$. In order to keep the potential arbitrary yet still maintain biased (asymmetrical) motion, we choose to employ a tilted periodic potential.

Equation of Motion

The concept of a Brownian ratchet was first introduced in the early 1900s by Smoluchowski and later developed by Feynman [30, 77]. Suppose there is a particle that moves via a Brownian ratchet in one dimension with spatial coordinate $x(t)$. Newton's law of motion then gives the formula

$$m \frac{d^2 x(t)}{dt^2} + \frac{dV(x)}{dx} = -\gamma \frac{dx(t)}{dt} + \xi(t) \quad (4.2)$$

where m is the mass of the particle, $V(x)$ is a periodic potential with period L (i.e. $[V(x) = V(x+L)]$), and γ is the friction coefficient. The variable $\xi(t)$ is Gaussian white noise which is an independent and identically distributed random variable with a normal distribution and zero mean. For small systems where thermal fluctuations are important, the motion is over-damped and inertia is negligible. Therefore an excellent approximation to Equation (4.2) is to set the inertia force term, $m \frac{d^2 x(t)}{dt^2}$, to zero. Hence the model becomes

$$\gamma \frac{dx(t)}{dt} = -\frac{dV(x(t))}{dx} + \xi(t). \quad (4.3)$$

Using the tilted periodic potential in Equation (4.1), we rescale the force term and the noise to omit the friction coefficient and arrive at our final equation for *biased* motion of a Brownian particle

$$\frac{dx}{dt} = -\frac{d}{dx}(U(x) - Fx) + \sqrt{2D}\xi(t) \quad (4.4)$$

with noise intensity, D . Often the parameter D is given by the Einstein relation [28],

$$D = \frac{k_\beta T}{\gamma} \quad (4.5)$$

where k_β is the Boltzmann constant, and T is temperature. This is consistent with the derivation of Equation (4.4).

Since the equation (4.4) describes a stochastic process, different realizations of $\xi(t)$ will produce different solutions to equation (4.4). The collection of such solutions is characterized by a probability density function, $P(x, t|x_0, t_0)$, used to describe the distribution of Brownian particles starting at position x_0 at time t_0 . The notation $P(x, t|x_0, t_0)$, denotes the probability that the particle is located at the spatial position x at time t , given that the particle started at position x_0 at time t_0 . We will use the notation $P(x, t)$ when $x_0 = 0$ and $t_0 = 0$. The time evolution of a statistical ensemble is captured by the time evolution of $P(x, t)$ which satisfies the Fokker-Planck equation [52, 64, 67]

$$\partial_t P(x, t) = \partial_x(V'(x) + D\partial_x)P(x, t) \quad (4.6)$$

with initial and boundary conditions to be explained in the following section. While the Fokker-Planck equation is not solvable in general, some key properties of this evolution can be described by considering both *drift* and *dispersion*. In the case

of biased motion, over time $P(x, t)$ will drift in the direction of the bias and spread (disperse) over many periods of the potential given in Equation (4.1). For more information on the history of Brownian ratchets and the details of our model derivation, please see the comprehensive review article [64].

Important Quantities for a Brownian Ratchet on a Periodic Potential with Drift

Transport of particles under a Brownian ratchet can be well described by three main quantities: drift, dispersion, and efficiency. The drift is given by the mean velocity of a particle,

$$v = \lim_{t \rightarrow \infty} \frac{\langle x(t) - x(0) \rangle}{t} \quad (4.7)$$

where $\langle \cdot \rangle$ denotes the mean value. Dispersion is characterized by the effective diffusion coefficient,

$$D_{eff} = \frac{1}{2} \lim_{t \rightarrow \infty} \frac{\langle [x(t) - \langle x(t) \rangle]^2 \rangle}{t}. \quad (4.8)$$

Efficiency can be quantified by the Péclet number,

$$Pe = \frac{vL}{D_{eff}}. \quad (4.9)$$

A large Péclet number indicates a large mean velocity with a small effective diffusion, resulting in efficient particle transport. Analysis of these three quantities is often the focus of diffusion research.

In an elegant paper written by Lindner *et. al.* [52], the model for diffusion transport in a tilted periodic potential and the corresponding Fokker-Planck equation, Equations (4.4) and (4.6) respectively, are analyzed. While many papers have addressed this problem [18, 54, 64–66], Lindner *et. al.* employed a different technique by dividing the tilted potential into overlapping segments, and analyzing an escape

from one of these segments. They then aggregated the process with an asymptotic limit as $t \rightarrow \infty$. This approach allowed for succinct formulas and analysis.

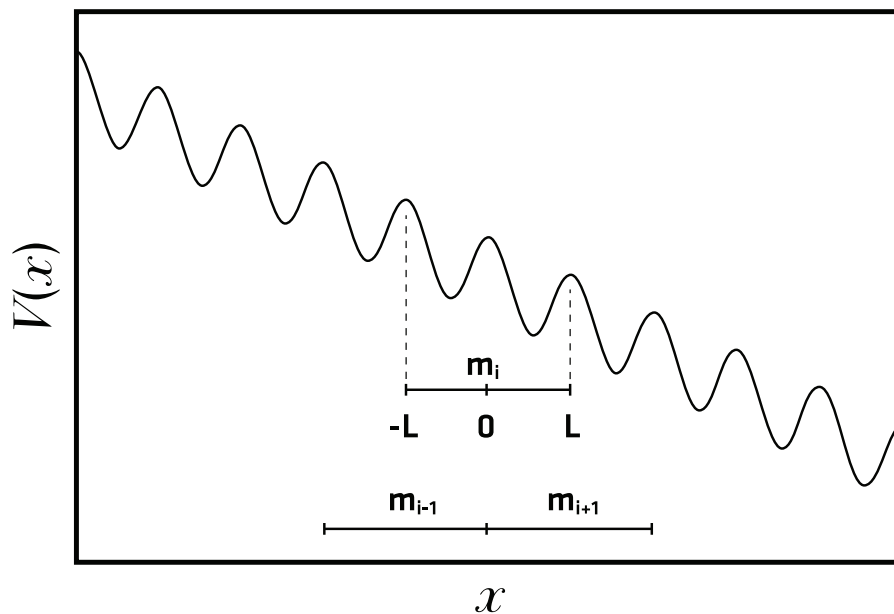


Figure 4.2: The potential function $V(x)$ is divided into overlapping segments of length $2L$, here segment m_i spans the interval $[-L, L]$. When a particle exits from either end of the segment, it is placed at the position $x = 0$ of the neighboring segment.

First we describe the specific division of the potential into segments m_i , depicted in Figure 4.2. The potential $V(x) = U(x) - Fx$, where $U(x)$ is periodic with period L , can be divided into overlapping segments of length $2L$. An escape from one segment is independent of all other segments. Therefore each segment can be analyzed individually on the one dimensional spatial interval $[-L, L]$. Segments are constructed with absorbing boundary conditions at the endpoints, $\pm L$. When a particle reaches either $\pm L$ of segment m_i it is removed from that segment and placed at $x = 0$ of the segment m_{i+1} if it exited right, or m_{i-1} if the exit was on the left. In this way they reduced the problem to a Fokker-Planck equation, Equation (4.6), by prescribing the

initial condition and boundary conditions

$$P(x, 0) = \delta(x), \quad P(-L, t) = P(L, t) = 0, \quad (4.10)$$

where the initial condition is given by the dirac delta function, $\delta(x)$, with point source at $x = 0$. The probability that a particle starts in a segment at $x = 0$ is one, and the probability that the particle is still in the segment at $x = \pm L$ is zero. When the particle reaches $\pm L$ it is now at the next segment's location of $x = 0$ with probability one. In this way, as described by the Fokker-Planck equation, the particles move over the segments.

Moving throughout these segments can be thought of as a one dimensional random walk. The net movement at time t , denoted n_t , is therefore

$$n_t = \sum_{i=1}^{N_t} W_i. \quad (4.11)$$

This is a stochastic process with jumps left and right depicted as realizations of $W_i = \{-1, 1\}$ for each $i = 1, 2, \dots, N_t$ with probabilities p_- and p_+ respectively. N_t is the total number of jumps before time t . Since W_i is a random variable describing a one dimensional random walk, we obtain the mean and variance

$$\langle W \rangle = p_+ - p_-, \quad \langle \Delta W^2 \rangle = 4p_+p_-. \quad (4.12)$$

Hence the asymptotic mean and variance of n_t are given by [19] as

$$\langle n_t \rangle = E[N_t] \langle W \rangle = (p_+ - p_-) \frac{t}{\mu} \quad (4.13)$$

and

$$\langle \Delta n_t^2 \rangle = E[N_t] \text{var}(W) + \text{var}(N_t) E[W]^2 = 4p_+p_- \frac{t}{\mu} + (p_+ - p_-)^2 \frac{t\sigma^2}{\mu^3} \quad (4.14)$$

In these equations μ and σ^2 are the mean and variance of the escape time density, and Equation (4.14) is given by the conditional variance identity. Using the Central Limit Theorem, N_t is found asymptotically to have a normal distribution with mean t/μ and variance $\sigma^2 t/\mu^3$.

The process n_t allows us to approximate the spatial location as $x(t) \approx Ln_t$ with an uncertainty of $\Delta x = \pm L$ at any time t . So for any time t , the position of $x(t) \in [Ln_t - L, Ln_t + L]$ with the specific position approximated as Ln_t . Under the limit as $t \rightarrow \infty$ the mean and variance of the particle position is

$$\langle x(t) \rangle = L(p_+ - p_-) \frac{t}{\mu}, \quad \langle \Delta x(t)^2 \rangle = L^2 \left(4p_+p_- \frac{t}{\mu} + (p_+ - p_-)^2 \frac{t\sigma^2}{\mu^3} \right). \quad (4.15)$$

Using Equation (4.15) with Equations (4.7), (4.8), and (4.9), as outlined in [52], we arrive at our final equations

$$v = L \frac{p_+ - p_-}{\mu}, \quad (4.16)$$

$$D_{eff} = \frac{L^2}{2} \left(\frac{1}{\mu} + \frac{(p_+ - p_-)^2}{\mu^3} (\sigma^2 - \mu^2) \right), \quad (4.17)$$

and

$$Pe = 2 \frac{p_+ - p_-}{1 - (1 - \frac{\sigma^2}{\mu^2})(p_+ - p_-)^2}. \quad (4.18)$$

In each of these equations, μ is of central importance and is therefore critical when describing diffusive transport of Brownian particles.

The last piece of these equations that needs to be addressed is to calculate μ and σ^2 . These expressions can be computed, using the techniques outlined in [31], as

$$\mu = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{V(x)}{D}} e^{-\frac{V(y)}{D}} dy dx, \quad (4.19)$$

and

$$\sigma^2 = \frac{2}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{V(x)}{D}} e^{-\frac{V(y)}{D}} \left(T_1(y) - \frac{1}{2}\mu\right) dy dx, \quad (4.20)$$

where $T_1(x)$ is the mean first passage time of a particle starting at x . This expression is

$$T_1(x) = \frac{1}{D} \left[\frac{\left(\int_{-L}^x \Phi(y) dy\right) \int_x^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \left(\int_x^L \Phi(y) dy\right) \int_{-L}^x \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy}{\int_{-L}^L \Phi(y) dy} \right] \quad (4.21)$$

using the simplifying $\Phi(x) = e^{\frac{V(x)}{D}}$ for clarity. The derivations of these formulas are detailed in Appendix A.

The focus of this chapter is μ , the mean escape time (MST) from a segment in a tilted periodic potential. In this chapter, we add an arbitrary number of periodic functions to the potential, referred to as a multi-periodic potential. These periodic functions have decreasing periods $L, L/N_1, L/N_2, \dots$ for $N_i \leq N_{i+1}$ where each $N_i \in \mathbb{N}$. Notice that for all of these periodic functions there exists a common period L . A new formula is then derived for the MST, the original formula being given by Equation (4.19). The derived formula expresses the total escape time as a function of the MST of each individual periodic function in the potential. Two examples using truncated Fourier series will be provided.

Theorem for the Mean Escape Time

Now that all of the equations and background information have been outlined, we proceed by constructing a new potential function which is a sum of periodic functions. For the tilted, multi-periodic potential, we use

$$V(x) = U_1(x, L/N_1) + U_2(x, L/N_2) + \cdots + U_n(x, L/N_n) + \widehat{U}(x, L) - Fx. \quad (4.22)$$

The notation, $U_i(x, L/N_i)$, denotes a periodic function with period L/N_i , where $N_i \in \mathbb{N}$ for all i . The function $\widehat{U}(x, L)$ is the periodic function with the largest period L . All other periodic functions in Equation (4.22) have a period that divides L and therefore each $U_i(x, L/N_i)$ is also periodic with period L . Therefore $U(x) = U_1(x, L/N_1) + U_2(x, L/N_2) + \cdots + U_n(x, L/N_n) + \widehat{U}(x, L)$ is periodic with period L .

To model RNAP transcription and pausing, we implement the tilted multi-periodic potential

$$V(x) = U_1(x, L/N) + U_2(x, L) - Fx. \quad (4.23)$$

The original motivation was for one potential to capture RNAP elongation, while the other potential captures pausing. In this equation, we can express the translocation of an RNAP one nucleotide at a time with $U_1(x, L/N)$. To describe the pauses we add the periodic function $U_2(x, L)$. This potential is meant to represent a pause every L nucleotides. For a visual representation of this multi-periodic potential see Figure 4.3. Using a periodic potential to describe a specific fixed spatial location of a pause does strip the RNAP model of the random aspect of pausing. While there is a lot to explore from having a periodic pause, it is not as biologically relevant as a stochastic pause model. However, it represents one step towards the integration of experimental knowledge into a model framework. As reported in [61], on average

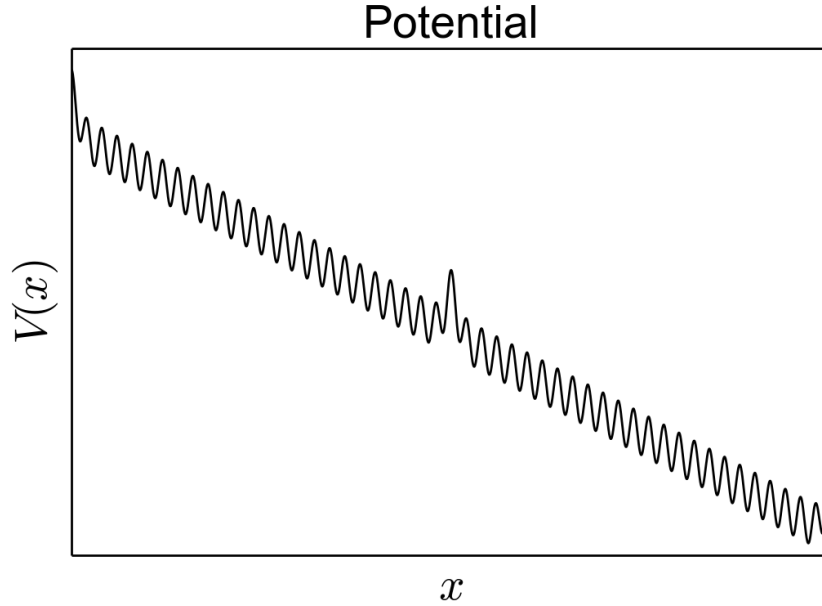


Figure 4.3: The multi-periodic potential function $V(x) = U_1(x, L/N) + U_2(x, L) - Fx$ is provided here for reference. $U_1(x, L/N)$ provides most of the oscillatory nature of this function, while $U_2(x, L)$ adds the higher amplitude wall that is more spaced out.

an RNAP pauses every 100 basepairs. Therefore using $L = 100$ basepairs for the pause and $L/N = 1$ for translocation of one nucleotide, one could analyze a diffusion model for elongation incorporating periodic pauses. While RNAP motion is the motivation for this project, the investigation is mathematically worthwhile in its own right regardless of the relevance to biology. As such for the remainder of this thesis, the focus is on diffusive particle transport in general, as opposed to as a model for DNA transcription specifically.

The MST out of a segment is of central importance in the analysis of diffusive transport. Using results from [52], namely Equation (4.19), we have an expression for the MST out of the segment spanning the interval $[-L, L]$ with the potential function $V(x) = \hat{U}(x, L) - Fx$. It stands to reason since several periodic functions are added together in the potential, that μ should be influenced by all periodic functions

contained in the potential. We develop a multiplicative formula for μ in terms of the MST for each periodic function in the potential.

Let $\mu_i(1)$ be the MST for the periodic potential $V_i(x) = U_i(x, 1)$ with no tilt and period one, and let $\hat{\mu}(L)$ be the MST for the tilted periodic potential $\hat{V}(x) = \hat{U}(x, L) - Fx$.

Theorem 4.0.1. *Suppose $V(x) = U_1(x, L/N_1) + U_2(x, L/N_2) + \dots + U_n(x, L/N_n) + \hat{U}(x, L) - Fx$ is a $C^1[-L, L]$, tilted multi-periodic potential where $N_i \in \{2, 3, 4, \dots\}$, $N_1 \leq N_2 \leq \dots \leq N_n$. Then*

$$\mu = (2D)^n \mu_1(1) \mu_2(1) \cdot \dots \cdot \mu_n(1) \hat{\mu}(L) + \mathcal{O}(L/N_1) \quad (4.24)$$

The proof for this theorem is provided in the following section. For the remainder of this section we provide several lemmas that aid in the proof. First, we present a theorem from [31] that shows the probability of escaping left or right is the same regardless of the what the periodic potential looks like so long as the period L and the tilt F are fixed.

Probability of Escaping Left or Right

The first thing to highlight, as it pertains to the multi-periodic potential in Equation (4.22), is the probability of escaping. Specifically, the probability of exiting to the right from position L , denoted by p_+ , and the probability of escaping left from the position $-L$, p_- , remain unchanged with the addition of more periodic functions. The potential $V(x) = \hat{U}(x, L) - Fx$ and the potential given in Equation (4.22) would have the exact same probability of escaping right or left. The probability p_- is related

to the probability p_+ through the equation

$$p_- + p_+ = 1, \quad (4.25)$$

since a particle will leave the segment with probability one. Using formulas from [31], another relationship between these probabilities is shown.

Theorem 4.0.2. *Suppose $V(x) = U(x, L) - Fx$ is a tilted periodic potential function.*

Then

$$p_+ = e^{\frac{FL}{D}} p_-$$

Proof. Using well known formulas given in Section 5.2.8 of [31] and derived in Appendix B, the probability of escaping the segment out of either $\pm L$ when beginning at $x = 0$ is

$$p_- = \frac{\int_0^L e^{\frac{V(y)}{D}} dy}{\int_{-L}^L e^{\frac{V(y)}{D}} dy}, \quad p_+ = \frac{\int_{-L}^0 e^{\frac{V(y)}{D}} dy}{\int_{-L}^L e^{\frac{V(y)}{D}} dy}. \quad (4.26)$$

Consider $\int_{-L}^0 e^{\frac{V(x)}{D}} dx$ under the change of variables $y = x + L$. Then

$$\int_{-L}^0 e^{\frac{V(x)}{D}} dx = \int_0^L e^{\frac{V(y-L)}{D}} dy$$

Now,

$$V(y - L) = U(y - L) - F(y - L) = U(y) - Fy + FL = V(y) + FL$$

since U is L periodic, therefore

$$\begin{aligned} \int_{-L}^0 e^{\frac{V(x)}{D}} dx &= \int_0^L e^{\frac{V(y-L)}{D}} dy \\ &= e^{\frac{FL}{D}} \int_0^L e^{\frac{V(y)}{D}} dy \end{aligned}$$

Hence

$$\int_{-L}^0 e^{\frac{V(x)}{D}} dx = e^{\frac{FL}{D}} \int_0^L e^{\frac{V(x)}{D}} dx$$

Plugging this into Equation (4.26), we have

$$p_+ = \frac{\int_{-L}^0 e^{\frac{V(y)}{D}} dy}{\int_{-L}^L e^{\frac{V(y)}{D}} dy} = \frac{e^{\frac{FL}{D}} \int_0^L e^{\frac{V(y)}{D}} dy}{\int_{-L}^L e^{\frac{V(y)}{D}} dy} = e^{\frac{FL}{D}} p_-$$

□

Remark. In this theorem $U(x, L)$ is an arbitrary function with period L . Therefore, using formulas derived in [52], the potential function $V(x) = U(x, L) - Fx$ has the probability of escaping to the right or to the left, respectively as

$$p_+ = \frac{1}{1 + e^{\frac{-FL}{D}}} \quad (4.27)$$

and

$$p_- = \frac{e^{\frac{-FL}{D}}}{1 + e^{\frac{-FL}{D}}}. \quad (4.28)$$

Since the potential in Equation (4.22) has periodic function $U_1(x, L/N_1) + U_2(x, L/N_2) + \dots + U_n(x, L/N_n) + \widehat{U}(x, L)$ with combined period L , the multi-periodic potential has the same probability of escaping right or left, in Equation (4.27) and (4.28) respectively.

The proof of Theorem 4.0.1 employs the Weighted Mean Value Theorem for Double Integrals. For the reader's convenience, the statement of this theorem is provided below.

Theorem 4.0.3 (Weighted Mean Value Theorem for Double Integrals). *Let f and g be continuous functions on a closed and bounded domain Ω . If g is positive then*

there exists a point $(s_0, t_0) \in \Omega$ such that

$$\iint_{\Omega} f(x, y)g(x, y) dy dx = f(x_0, y_0) \iint_{\Omega} g(x, y) dy dx.$$

A proof of this theorem for single integrals is presented in [3].

There are a few simplifying notations we utilize throughout the remainder of this chapter. First denote the domain of integration by Ω , where

$$\Omega = \{(x, y) \mid 0 \leq x \leq L, \quad x - L \leq y \leq x\}.$$

Often, Ω will be partitioned into several smaller parallelograms with height and width L/N . This partitioned domain will be referred to as

$$\Omega_{k,s} = \{(x, y) \mid (k-1)L/N \leq x \leq kL/N, \quad x - L + (s-1)L/N \leq y \leq x - L + sL/N\}$$

for each choice of (k, s) , $k = 1, \dots, N$, $s = 1, \dots, N$, see Figure 4.4. We also employ $\Psi(x_0, L) = e^{\frac{\bar{V}(x_0(k,s))}{D}}$ where we express the multi-periodic potential

$$V(x) = \bar{V}(x) + U_n(x, L/N_n)$$

by separating the potential with the shortest period. Hence

$$\bar{V}(x) = U_1(x, L/N_1) + \dots + U_{n-1}(x, L/N_{n-1}) + \hat{U}(x, L) - Fx. \quad (4.29)$$

Now we provide a series of lemmas for use in the proof.

Lemma 4.0.1. *Let $V(x) = U_1(x, L/N_1) + U_2(x, L/N_2) + \dots + U_n(x, L/N_n) + \hat{U}(x, L) - Fx = \bar{V}(x) + U_n(x, L/N_n)$ be a tilted multi-periodic potential where*

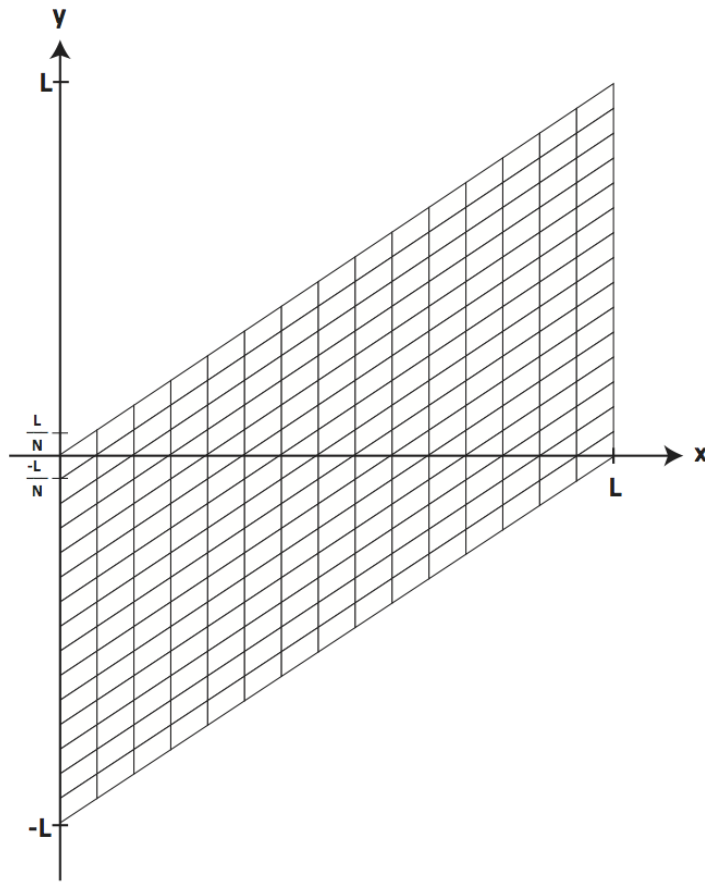


Figure 4.4: The integration domain for μ segregated into N^2 parallelograms each with width L/N and height L/N .

$N_i \in \{2, 3, 4, \dots\}$, $N_1 \leq N_2 \leq \dots \leq N_n$ and $V(x)$ is $C^1[-L, L]$. Then there exists a point $(x_0(k, s), y_0(k, s)) \in \Omega_{k,s}$ such that the MST given by equation (4.19) is

$$\mu = \mu_n(L/N_n) \frac{2D}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n} e^{\frac{\bar{V}(x_0(k,s))}{D}} e^{\frac{-\bar{V}(y_0(k,s))}{D}}. \quad (4.30)$$

Proof. From equation (4.19), an expression for the MST from the tilted, multi-periodic potential $V(x) = \bar{V}(x) + U_n(x, L/N_n)$ is

$$\mu = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{\bar{V}(x)}{D}} e^{\frac{-\bar{V}(y)}{D}} e^{\frac{U_n(x, L/N_n)}{D}} e^{\frac{-U_n(y, L/N_n)}{D}} dy dx. \quad (4.31)$$

Next divide domain Ω into subdomains $\Omega_{k,s}$ as described above and shown in Figure 4.4. Then

$$\mu = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n} \iint_{\Omega_{k,s}} \frac{e^{\frac{\bar{V}(x)}{D}} e^{\frac{U_n(x, L/N_n)}{D}}}{e^{\frac{\bar{V}(y)}{D}} e^{\frac{U_n(y, L/N_n)}{D}}} dy dx. \quad (4.32)$$

Using Theorem 4.0.3, for each (k, s) there exists a point $(x_0(k, s), y_0(k, s)) \in \Omega_{k,s}$ such that

$$\mu = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n} \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \iint_{\Omega_{k,s}} \frac{e^{\frac{U_n(x, \frac{L}{N_n})}{D}}}{e^{\frac{U_n(y, \frac{L}{N_n})}{D}}} dy dx. \quad (4.33)$$

To transform the integral domain, we define a simple linear area preserving change of variables $G_{k,s}$ that maps arbitrary domain $\Omega_{k,s}$ to the canonical domain Ω_0 , where $\Omega_0 = \{(x, y) \mid 0 \leq x \leq L, \quad x - L \leq y \leq x\}$. Specifically, this mapping is defined as

$$G_{k,s} : (x, y) \rightarrow (x - (k - 1)L/N_n, y - (s - N_n)L/N_n - (k - 1)L/N_n). \quad (4.34)$$

Multiplying and dividing by $2D$,

$$\mu = \frac{2D}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n} \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \left(\frac{1}{2D} \iint_{\Omega_0} \frac{e^{-\frac{U_n(x, L/N_n)}{D}}}{e^{-\frac{U_n(y, L/N_n)}{D}}} dy dx \right). \quad (4.35)$$

In Equation (4.35), the term within the parentheses is the MST, $\mu_n(L/N_n)$, for the periodic potential $U_n(x, L/N_n)$ given by Equation (4.19) with no tilt. Therefore

$$\mu = \frac{2D\mu_n(L/N_n)}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_n} \sum_{s=1}^{N_n} \frac{\Psi(x_0, L)}{\Psi(y_0, L)}. \quad (4.36)$$

□

Lemma 4.0.2. *Let the MST for the periodic potential $V(x) = U(x, L/N)$ with period L/N be $\mu(L/N)$. Then*

$$\mu(L/N) = \left(\frac{L}{N}\right)^2 \mu(1). \quad (4.37)$$

Proof. First recall

$$\mu(L/N) = \frac{1}{2D} \int_0^{L/N} \int_{x-L/N}^x e^{\frac{U(x, L/N)}{D}} e^{-\frac{U(y, L/N)}{D}} dy dx. \quad (4.38)$$

To change the period we substitute L/N by qL/N where $q \in \mathbb{Q}^+$. Hence,

$$\mu(qL/N) = \frac{1}{2D} \int_0^{qL/N} \int_{x-qL/N}^x e^{\frac{U(x, qL/N)}{D}} e^{-\frac{U(y, qL/N)}{D}} dy dx. \quad (4.39)$$

Now let's return to Equation (4.38) under the change of variables $X = qx$, $Y = qy$. To see how this change of variables affects a periodic function, let's consider a prototypical periodic function of period L/N , $\cos\left(\frac{2\pi Nx}{L}\right)$. Under the change of

variables $X = qx$, we compute

$$\cos\left(\frac{2\pi Nx}{L}\right) = \cos\left(\frac{2\pi NX}{qL}\right)$$

which is periodic with period qL/N . Using this relationship for the periodic function in Equation (4.38), we have

$$\mu(L/N) = \frac{1}{2Dq^2} \int_0^{qL/N} \int_{X-qL/N}^X e^{\frac{U(X,qL/N)}{D}} e^{-\frac{U(Y,qL/N)}{D}} dY dX. \quad (4.40)$$

Then using Equations (4.39) and (4.40), we have

$$\mu(qL/N) = q^2 \mu(L/N) \quad (4.41)$$

for any $q \in \mathbb{Q}^+$. Therefore from Equation (4.41),

$$\mu(L/N) = (L/N)^2 \mu(1) \quad (4.42)$$

where

$$\mu(1) = \frac{1}{2D} \int_0^1 \int_{x-1}^x e^{\frac{U_1(x,1)}{D}} e^{-\frac{U_1(y,1)}{D}} dy dx. \quad (4.43)$$

□

Lemma 4.0.3. *Let the multi-periodic potential be defined as in Lemma 4.0.1. Then the difference, ϵ , between the double sum in Equation (4.30) and the double integral for the MST of the multi-periodic potential $\bar{V}(x)$, i.e.*

$$\epsilon = \left| \iint_{\Omega} e^{\frac{\bar{V}(x)}{D}} e^{-\frac{\bar{V}(y)}{D}} dy dx - \sum_{k=1}^{N_n} \sum_{s=1}^{N_n} \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \left(\frac{L}{N_n}\right)^2 \right| \quad (4.44)$$

is on the order of $\mathcal{O}\left(\frac{L}{N_n}\right)$.

Proof. In order to do an estimation of the error accrued, begin by looking at the error, $\epsilon_{k,s}$, over an arbitrary parallelogram $\Omega_{k,s}$. For simplicity the point $(x_0(k, s), y_0(k, s))$ is denoted as (x_0, y_0) . The error for each $\Omega_{k,s}$ is

$$\epsilon_{k,s} = \left| \iint_{\Omega_{k,s}} e^{\frac{\bar{V}(x)}{D}} e^{-\frac{\bar{V}(y)}{D}} dy dx - \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \left(\frac{L}{N_n}\right)^2 \right|. \quad (4.45)$$

The Taylor expansion of the integrand, $f(x, y)$, in a neighborhood of (x_0, y_0) is

$$f(x, y) = \frac{\Psi(x_0, L)}{\Psi(y_0, L)} + \frac{\bar{V}'(a) \Psi(a, L)}{D \Psi(b, L)}(x - x_0) - \frac{\bar{V}'(b) \Psi(a, L)}{D \Psi(b, L)}(y - y_0) \quad (4.46)$$

where (a, b) is a specific point in the neighborhood of (x_0, y_0) . Using the first order Taylor expansion in the integral gives

$$\epsilon_{k,s} = \left| \iint_{\Omega_{k,s}} \left[\frac{\Psi(x_0, L)}{\Psi(y_0, L)} + \frac{\bar{V}'(a) \Psi(a, L)}{D \Psi(b, L)}(x - x_0) - \frac{\bar{V}'(b) \Psi(a, L)}{D \Psi(b, L)}(y - y_0) \right] dy dx - \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \left(\frac{L}{N_n}\right)^2 \right|. \quad (4.47)$$

We replace $(x - x_0)$ by the upper bound L/N_n , and $(y - y_0)$ the upper bound $2L/N_n$.

Using these bounds, integrating over $\Omega_{k,s}$ gives the error estimation

$$\epsilon_{k,s} \leq \left| \left(\frac{\bar{V}'(a)}{D} \right) \frac{\Psi(a, L)}{\Psi(b, L)} \left(\frac{L}{N_n}\right) \left(\frac{L}{N_n}\right)^2 - \left(\frac{\bar{V}'(b)}{D} \right) \frac{\Psi(a, L)}{\Psi(b, L)} \left(\frac{2L}{N_n}\right) \left(\frac{L}{N_n}\right)^2 \right|. \quad (4.48)$$

Simplifying this expression

$$\epsilon_{k,s} \leq \left| \frac{\Psi(a, L)}{\Psi(b, L)} \left(\frac{L}{N_n}\right)^3 \left(\frac{\bar{V}'(a)}{D} - \frac{2\bar{V}'(b)}{D} \right) \right|. \quad (4.49)$$

Since $\bar{V}(x)$ is C^1 on the compact set $[0, L]$, $\bar{V}(x)$ is bounded and as mentioned in the earlier section the derivative will also be bounded on $[0, L]$. Hence there exists a finite constant C such that

$$C \geq \left| \frac{\Psi(a, L)}{\Psi(b, L)} \left(\frac{\bar{V}'(a)}{D} - \frac{2\bar{V}'(b)}{D} \right) \right|. \quad (4.50)$$

Hence from Equations (4.49) and (4.50), we have

$$\epsilon_{k,s} \leq C \left(\frac{L}{N_n} \right)^3. \quad (4.51)$$

Summing the error over the N_n^2 parallelograms we determine

$$\epsilon \leq C \left(\frac{L}{N_n} \right)^3 N_n^2 = C(L)^2 \frac{L}{N_n} = \hat{C} \frac{L}{N_n} \quad (4.52)$$

since L is fixed. Therefore the error is of the order L/N_n . \square

Now we prove Theorem 4.0.1.

Proof of Theorem 4.0.1

Proof. The proof is by induction. We begin with the case of $n = 1$. Suppose the effective potential is given by

$$V(x) = U_1(x, L/N) + \hat{U}(x, L) - Fx \quad (4.53)$$

where $N \in \mathbb{N}$.

With our multi-periodic potential $V(x)$, we can plug this into Equation (4.19) and rearrange to get the MST

$$\mu = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{\hat{U}(x,L)-Fx}{D}} e^{-\frac{\hat{U}(y,L)-Fy}{D}} e^{\frac{U_1(x,L/N)}{D}} e^{-\frac{U_1(y,L/N)}{D}} dy dx. \quad (4.54)$$

Now we segregate our integration domain, a parallelogram of width L and height L , into N^2 parallelograms of width L/N and height L/N , see Figure 4.4. According to Lemma 4.0.1, for each (k, s) there exists a point $(x_0(k, s), y_0(k, s)) \in \Omega_{k,s}$ such that the MST in Equation (4.19) can be written as

$$\mu = \mu_1(L/N) \frac{2D}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^N \sum_{s=1}^N \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \quad (4.55)$$

where $\Psi(x_0, L) = e^{\frac{\hat{U}(x_0(k,s),L)-Fx}{D}}$ and $\mu_1(L/N)$ is the MST of the potential $U_1(x, L/N)$ with no tilt, i.e $F = 0$. By Lemma 4.0.2, we have from Equation (4.37),

$$\mu_1(L/N) = (L/N)^2 \mu_1(1). \quad (4.56)$$

Hence the MST becomes

$$\mu = \mu_1(1) \frac{2D}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^N \sum_{s=1}^N \frac{\Psi(x_0, L)}{\Psi(y_0, L)} (L/N)^2. \quad (4.57)$$

From here we would like to express μ in terms of the MST integral over the domain Ω given by Equation (4.19), which would produce a small amount of error. From Lemma 4.0.3, we compute this error to be on the order of $\mathcal{O}\left(\frac{L}{N}\right)$. Returning

to Equation (4.57) we have the MST

$$\begin{aligned}\mu &= 2D\mu_1(1) \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{\widehat{U}(x,L)-Fx}{D}} e^{-\frac{(\widehat{U}(y,L)-Fy)}{D}} dy dx + \mathcal{O}(L/N) \\ &= 2D\mu_1(1)\widehat{\mu}(L) + \mathcal{O}(L/N),\end{aligned}\tag{4.58}$$

and the induction case of $n = 1$ is proven.

Suppose for the potential

$$\overline{V}(x) = U_1(x, L/N_1) + U_2(x, L/N_k) + \cdots + U_K(x, L/N_K) + \widehat{U}(x, L) - Fx \tag{4.59}$$

we assume that the MST for this potential is

$$\mu = (2D)^K \mu_1(1)\mu_2(1) \cdots \mu_K(1)\widehat{\mu}(L) + \mathcal{O}(L/N_1).\tag{4.60}$$

To prove the case of $n = K + 1$ let the potential equal

$$\begin{aligned}V(x) &= U_1(x, L/N_1) + \cdots + U_K(x, L/N_K) + U_{K+1}(x, L/N_{K+1}) + \widehat{U}(x, L) - Fx \\ &= \overline{V}(x) + U_{K+1}(x, L/N_k).\end{aligned}\tag{4.61}$$

Applying Lemma 4.0.1, there exists a point $(x_0(k, s), y_0(k, s)) \in \Omega_{k,s}(L/N_{K+1})$ for each (k, s) such that the MST can be expressed as

$$\mu = \frac{2D\mu_{K+1}(L/N_{K+1})}{D \left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_{K+1}} \sum_{s=1}^{N_{K+1}} \frac{\Psi(x_0, L)}{\Psi(y_0, L)}.\tag{4.62}$$

From Lemma 4.0.2,

$$\mu_{K+1}(L/N_{K+1}) = \left(\frac{L}{N_{K+1}}\right)^2 \mu_{K+1}(1).\tag{4.63}$$

Hence

$$\mu = \frac{2D\mu_{K+1}(1)}{D\left(1 + e^{-\frac{FL}{D}}\right)} \sum_{k=1}^{N_{K+1}} \sum_{s=1}^{N_{K+1}} \frac{\Psi(x_0, L)}{\Psi(y_0, L)} \left(\frac{L}{N_{K+1}}\right)^2. \quad (4.64)$$

Next, we examine the error when comparing the double sum to the double integral.

By Lemma 4.0.2, the error is on the order of $\mathcal{O}\left(\frac{L}{N_{K+1}}\right)$. Replacing the summation in

Equation (4.64) by the integral considered above, μ becomes,

$$\mu = 2D\mu_{K+1}(1) \left(\frac{1}{D\left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x \frac{e^{\frac{\bar{V}(x)}{D}}}{e^{\frac{\bar{V}(y)}{D}}} dy dx \right) + \mathcal{O}\left(\frac{L}{N_{K+1}}\right). \quad (4.65)$$

From Equation (4.60), we have

$$\mu = 2D\mu_{K+1}(1) \left[(2D)^K \mu_1(1)\mu_2(1) \cdot \dots \cdot \mu_K(1)\widehat{\mu}(L) + \mathcal{O}\left(\frac{L}{N_1}\right) \right] + \mathcal{O}\left(\frac{L}{N_{K+1}}\right) \quad (4.66)$$

from the induction assumption in the case of $n = K$. Multiplying this expression out gives

$$\mu = (2D)^{K+1} \mu_1(1) \cdot \dots \cdot \mu_K(1)\mu_{K+1}(1)\widehat{\mu}(L) + 2D\mu_{K+1}(1)\mathcal{O}\left(\frac{L}{N_1}\right) + \mathcal{O}\left(\frac{L}{N_{K+1}}\right). \quad (4.67)$$

Consider the $\mathcal{O}\left(\frac{L}{N_1}\right)$ term in the equation above. The specific formulation for $\mu_{K+1}(1)$

is

$$\mu_{K+1}(1) = \frac{1}{2D} \int_0^1 \int_{x-1}^1 e^{\frac{U_{K+1}(x,1)}{D}} e^{-\frac{U_{K+1}(y,1)}{D}} dy dx, \quad (4.68)$$

where $U_{K+1}(x)$ and $-U_{K+1}(y)$ are continuous on a compact set and therefore bounded.

Hence $2D\mu_{K+1}(1)$ is a finite constant that is independent of K and we can replace

$2D\mu_{K+1}(1)\mathcal{O}\left(\frac{L}{N_1}\right)$ with simply $\mathcal{O}\left(\frac{L}{N_1}\right)$. Finally $\mathcal{O}\left(\frac{L}{N_1}\right) + \mathcal{O}\left(\frac{L}{N_{K+1}}\right)$ is of the order

$\mathcal{O}\left(\frac{L}{N_1}\right)$ which leads to the final equation

$$\mu = (2D)^{(K+1)} \mu_1(1) \mu_2(1) \cdot \dots \cdot \mu_K(1) \mu_{K+1}(1) \hat{\mu}(L) + \mathcal{O}\left(\frac{L}{N_1}\right). \quad (4.69)$$

By induction, we have proven Theorem 4.0.1. \square

Fourier Series Application

It is common when using a periodic potential, to approximate the potential with a truncated Fourier series approximation. Typically with a truncated Fourier series, one can achieve a close approximation to the original potential assuming a sufficient number of terms are present in the truncated series.

For a function $f(x)$ that is periodic with period L on the interval $[0, L]$, the Fourier series representation of $f(x)$ is

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{2\pi nx}{L}\right). \quad (4.70)$$

This is a rewriting of the function $f(x)$ using orthonormal basis functions,

$$\sqrt{\frac{2}{L}} \cos\left(\frac{2\pi nx}{L}\right), \quad \sqrt{\frac{2}{L}} \sin\left(\frac{2\pi nx}{L}\right)$$

for $n \in \mathbb{N}$. The coefficients of the Fourier series are governed by the following equations

$$a_0 = \frac{2}{L} \int_0^L f(x) dx, \quad (4.71)$$

$$a_n = \frac{2}{L} \int_0^L f(x) \cos\left(\frac{2\pi nx}{L}\right) dx, \quad (4.72)$$

and

$$b_n = \frac{2}{L} \int_0^L f(x) \sin\left(\frac{2\pi nx}{L}\right) dx. \quad (4.73)$$

The most common Fourier series expansions are defined on the interval $[-L, L]$, see Appendix C for details on the transformation to the interval $[0, L]$.

Given Equation (4.70), the truncated Fourier expansion is then

$$f_N(x) = \frac{a_0}{2} + \sum_{n=1}^N a_n \cos\left(\frac{2\pi nx}{L}\right) + \sum_{n=1}^N b_n \sin\left(\frac{2\pi nx}{L}\right). \quad (4.74)$$

According to the Fourier Convergence Theorem, if $f(x)$ and $f'(x)$ are piecewise continuous the Fourier series is convergent and converges pointwise to $f(x)$ for all values x where $f(x)$ is continuous. At points of discontinuity, $x = x_0$, the Fourier series converges to the average of the left and right limits of $f(x)$ as $x \rightarrow x_0$. The convergence of the Fourier series becomes uniform if we require that $f(x)$ is continuous with $f'(x)$ piecewise continuous [13]. The Fourier Convergence Theorem allows us to look at a broad class of periodic functions without requiring C^1 , only piecewise continuity of the function and its derivative.

In this section we are comparing several different MST calculations. For this purpose, we introduce the following notation: the MST given by Equation (4.19) of the original potential, and the truncated Fourier series potential, is denoted as μ and μ_N respectively, while the multiplicative approximation of the MST given by Equation (4.24) using the truncated Fourier series potential is denoted $\tilde{\mu}_N$. Given the convergence of the truncated Fourier series to the original periodic function, a theorem for the convergence of the MST for the truncated Fourier series potential, μ_N , to the MST for the original potential, μ , is presented below.

Theorem 4.0.4. *Let $f(x)$ be a continuous periodic function with a piecewise continuous derivative, with $x \in [-L, L]$. If $f_N(x)$ represents the truncated Fourier series approximation of $f(x)$ given in Equation (4.74) then*

$$\lim_{N \rightarrow \infty} \mu_N = \mu. \quad (4.75)$$

Proof. Consider the difference between MSTs of the periodic potential $f(x) - Fx$ and the truncated Fourier series potential with N terms $f_N(x) - Fx$ both with period L ,

$$\begin{aligned} |\mu - \mu_N| &= \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \left| \iint_{\Omega} \frac{e^{\frac{f(x)-Fx}{D}}}{e^{\frac{f(y)-Fy}{D}}} dy dx - \iint_{\Omega} \frac{e^{\frac{f_N(x)-Fx}{D}}}{e^{\frac{f_N(y)-Fy}{D}}} dy dx \right| \\ &= C \left| \iint_{\Omega} e^{-\frac{Fx}{D}} e^{\frac{Fy}{D}} \left(e^{\frac{f(x)}{D}} e^{-\frac{f(y)}{D}} - e^{\frac{f_N(x)}{D}} e^{-\frac{f_N(y)}{D}} \right) dy dx \right| \end{aligned} \quad (4.76)$$

where $\Omega = \{(x, y) \mid 0 \leq x \leq L, \quad x - L \leq y \leq x\}$ and $C = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)}$. Factoring out an $e^{\frac{f_N(x)}{D}}$ and an $e^{-\frac{f(y)}{D}}$, the error becomes

$$\begin{aligned} |\mu - \mu_N| &= C \left| \iint_{\Omega} e^{-\frac{Fx}{D}} e^{\frac{Fy}{D}} e^{-\frac{f(y)}{D}} e^{\frac{f_N(x)}{D}} \left(e^{\frac{f(x)-f_N(x)}{D}} - e^{\frac{f(y)-f_N(y)}{D}} \right) d\Omega \right| \\ &= C \left| \iint_{\Omega} e^{\frac{f_N(x)-Fx}{D}} e^{-\frac{(f(y)-Fy)}{D}} \left(e^{\frac{f(x)-f_N(x)}{D}} - e^{\frac{f(y)-f_N(y)}{D}} \right) d\Omega \right| \end{aligned} \quad (4.77)$$

Since $f(y)$ is continuous on a compact set, the quantity $e^{-\frac{(f(y)-Fy)}{D}}$ is bounded, i.e. $e^{-\frac{(f(y)-Fy)}{D}} < M_1$ for all $y \in \Omega$. Consider $e^{\frac{f_N(x)-Fx}{D}}$. For each N , $f_N(x)$ is continuous on the compact set $[0, L]$ and is therefore bounded. Since $f(x)$ is continuous with a piecewise continuous derivative, by the Fourier Convergence Theorem $f_N(x)$ converges uniformly to $f(x)$ [13]. Every uniformly convergent sequence of bounded functions is uniformly bounded [69], therefore, $f_N(x)$ is uniformly bounded and $e^{\frac{f_N(x)-Fx}{D}}$ is uniformly bounded on $[0, L]$. Hence $e^{\frac{f_N(x)-Fx}{D}} < M_2$. Using this in Equation (4.77),

we have

$$|\mu - \mu_N| < CM_1M_2 \left| \iint_{\Omega} \left(e^{\frac{f(x)-f_N(x)}{D}} - e^{\frac{f(y)-f_N(y)}{D}} \right) d\Omega \right|. \quad (4.78)$$

Note, the function $g(x) = e^{\frac{f(x)-f_N(x)}{D}}$ is continuous on the compact set $[0, L]$ and is therefore uniformly continuous [59]. As a result of the uniform convergence of $f_N(x)$ to $f(x)$ and the uniform continuity of the function $g(x)$ on $[0, L]$, for any arbitrarily small δ there exists an N such that for all $n \geq N$,

$$\left| e^{\frac{f(x)-f_n(x)}{D}} - e^{\frac{f(y)-f_n(y)}{D}} \right| < \frac{\delta}{CM_1M_2L^2}.$$

Therefore

$$|\mu - \mu_N| < cM_1M_2 \iint_{\Omega} \left| e^{\frac{f(x)-f_n(x)}{D}} - e^{\frac{f(y)-f_n(y)}{D}} \right| d\Omega \quad (4.79)$$

and

$$|\mu - \mu_N| < \frac{\delta}{L^2} \iint_{\Omega} d\Omega = \delta. \quad (4.80)$$

Since δ is arbitrary,

$$\lim_{N \rightarrow \infty} \mu_N = \lim_{N \rightarrow \infty} C \iint_{\Omega} e^{\frac{f_N(x)-x}{D}} e^{\frac{-(f_N(y)-Fy)}{D}} d\Omega = C \iint_{\Omega} e^{\frac{f(x)-Fx}{D}} e^{\frac{-(f(y)-Fy)}{D}} d\Omega = \mu \quad (4.81)$$

and the MST of the potential with the truncated Fourier series converges to the MST of the potential with the original function $f(x)$. \square

Given the convergence of μ_N to μ as N goes to infinity, ideally we would like to quantify the error of the MST of the truncated series based on the number of terms, N . While this may be done for the error given by the L^2 norm [33], $\|f - f_N\|_2^2$, we have not been able to do so using the error given by the difference $|f - f_N|$ as found in Equation (4.78). In fact, as shown in [48] the Fourier series of a continuous function

will converge, though it may converge arbitrary slowly. The Theorem from [48] is presented below using $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ to denote the real line mod 2π , and $S_n(g, t)$ being the first n terms of the Fourier expansion of the function $g(t)$.

Theorem 4.0.5 (Körner, 1989). *Given a decreasing sequence $\delta_0, \delta_1, \dots$, with $\delta_n \rightarrow 0$ as $n \rightarrow \infty$ we can find a continuous function $g : \mathbb{T} \rightarrow \mathbb{C}$ with $S_n(g, t) \rightarrow g(t)$ uniformly, but $\sup_{t \in \mathbb{T}} |g_n(t) - S_n(g, t)| \geq \delta_n$ for all $n \geq 0$.*

Although the difference of the exponentials is used in Equation (4.78), we are not yet able to quantify the error in terms of N within the difference $|f(x) - f_N(x)|$, or implement the L^2 norm into calculation of the error between μ and μ_N .

Remark. *It is important to note that this proof for the convergence of the MSTs only holds when the original function $f(x)$ is continuous with a piecewise continuous derivative. The convergence of the MSTs for a piecewise continuous $f(x)$ is not proven in this chapter and remains a question for future work.*

We now examine the limit as N goes to infinity, and let us also examine what occurs to Equation (4.24) under this limit. Rewriting Equation (4.24) as a product, $\tilde{\mu}_N$ becomes

$$\tilde{\mu}_N = (2D)^N \left(\prod_{i=1}^N \mu_i(1) \right) \hat{\mu}(L) + \mathcal{O}(L/N_1). \quad (4.82)$$

Consider the form of an individual $\mu_i(1)$ given by

$$\mu_i(1) = \frac{1}{2D} \int_0^1 \int_{x-1}^x e^{\frac{U_i(x,1)}{D}} e^{-\frac{U_i(y,1)}{D}} dy dx \quad (4.83)$$

and now rewritten in the form

$$\mu_i(1) = \frac{\eta_i(1)}{2D} \quad (4.84)$$

where $\eta_i(1)$ represents the double integral in Equation (4.83). Then the MST becomes

$$\begin{aligned}\tilde{\mu}_N &= (2D)^N \left(\prod_{i=1}^N \frac{\eta_i(1)}{2D} \right) \hat{\mu}(L) + \mathcal{O}(L/N_1) \\ &= \left(\prod_{i=1}^N \eta_i(1) \right) \hat{\mu}(L) + \mathcal{O}(L/N_1).\end{aligned}\tag{4.85}$$

Under the limit, the MST for infinitely many periodic functions in the tilted multi-periodic potential is

$$\begin{aligned}\tilde{\mu} &= \lim_{N \rightarrow \infty} \tilde{\mu}_N \\ &= \lim_{N \rightarrow \infty} \left[\left(\prod_{i=1}^N \eta_i(1) \right) \hat{\mu}(L) + \mathcal{O}(L/N_1) \right] \\ &= \left(\prod_{i=1}^{\infty} \eta_i(1) \right) \hat{\mu}(L) + \mathcal{O}(L/N_1)\end{aligned}\tag{4.86}$$

if the limit exists. Specifically what occurs in the infinite product, would depend on the periodic functions used in the expression for $\eta_i(1)$ and thus the value computed as a result of the integral in Equation (4.83).

In the following two sections, we explore two examples of periodic potentials, a quadratic wave and the square wave. We derive the Fourier series on $[0, L]$ for each and investigate the truncated series as the periodic function $U(x)$ in the potential $V(x) = U(x) - Fx$. Of particular interest is the MST, particularly calculating the multiplicative approximation of the MST for the truncated Fourier series potential using Equation (4.24). Employing Equation (4.24) can be particularly advantageous where there is a large gap in Fourier series frequencies as is the case in a periodic function $U_1(x, L) + U_2(x, L/N)$ where N is quite large. This type of multi-periodic potential is highly oscillatory over the interval $[0, L]$ which makes a

numerical integration for the MST in Equation (4.19) challenging, difficult, and often inefficient. With N large, the $\mathcal{O}\left(\frac{L}{N}\right)$ error in Equation (4.24) is small, and the numerical integration for the MST of the potential $V(x) = U_1(x, L) - Fx$ and the potential $V(x) = U_2(x, 1)$ are both computed easily.

Quadratic Wave

We will begin with the quadratic function

$$f(x) = -\left(x - \frac{L}{2}\right)^2 \quad (4.87)$$

on the interval $[0, L]$. The periodic function is the periodic extension

$$f(x) = -\left((x - nL) - \frac{L}{2}\right)^2, \quad x \in [nL, (n+1)L], \quad n \in \mathbb{N}. \quad (4.88)$$

Essentially, $f(x)$ in Equation (4.87) becomes $f(x - nL)$ on the interval $[nL, (n+1)L]$.

We consider the tilted periodic potential

$$V(x) = U(x, L) - Fx \quad (4.89)$$

where $U(x, L)$ is the periodic extension given in Equation (4.88), and $F = 0.2$. The graph of this potential on the interval $[0, 3L]$ is shown in Figure 4.5.

Since $f(x)$ is a continuous function, by the Fourier Convergence Theorem the Fourier series converges uniformly to $f(x)$ for all x . The Fourier series for $f(x)$ is given by

$$f(x) = \frac{-L}{12} - \sum_{n=1}^{\infty} \left(\frac{L}{n\pi}\right)^2 \cos\left(\frac{2\pi nx}{L}\right), \quad (4.90)$$

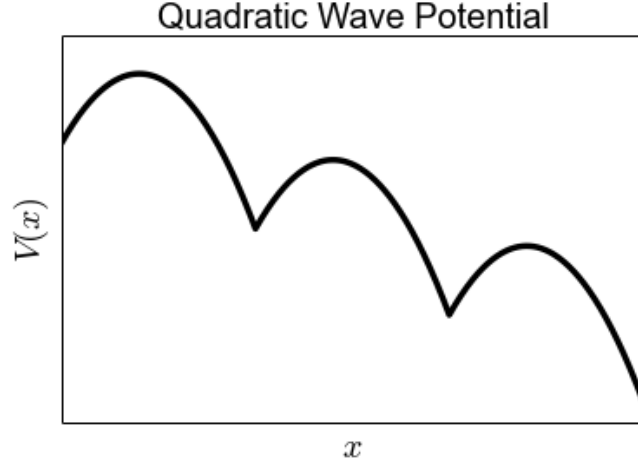


Figure 4.5: The tilted potential function $V(x)$ with $F = 0.2$ and $U(x, L)$ as the periodic extension given in Equation (4.88) on the interval $[0, 3L]$.

see Appendix C for details. The truncated Fourier series has a finite sum as opposed to an infinite sum

$$f_N(x) = \frac{-L}{12} - \sum_{n=1}^N \left(\frac{L}{n\pi} \right)^2 \cos \left(\frac{2\pi nx}{L} \right). \quad (4.91)$$

The error between the truncated Fourier series and the original function decreases with the addition of more terms, see Figure 4.6.

To frame the truncated Fourier series in the context of a tilted multi-periodic potential consider

$$V(x) = - \sum_{n=2}^N \left(\frac{L}{n\pi} \right)^2 \cos \left(\frac{2\pi nx}{L} \right) + \left(\frac{-L}{12} - \left(\frac{L}{\pi} \right)^2 \cos \left(\frac{2\pi x}{L} \right) \right) - Fx. \quad (4.92)$$

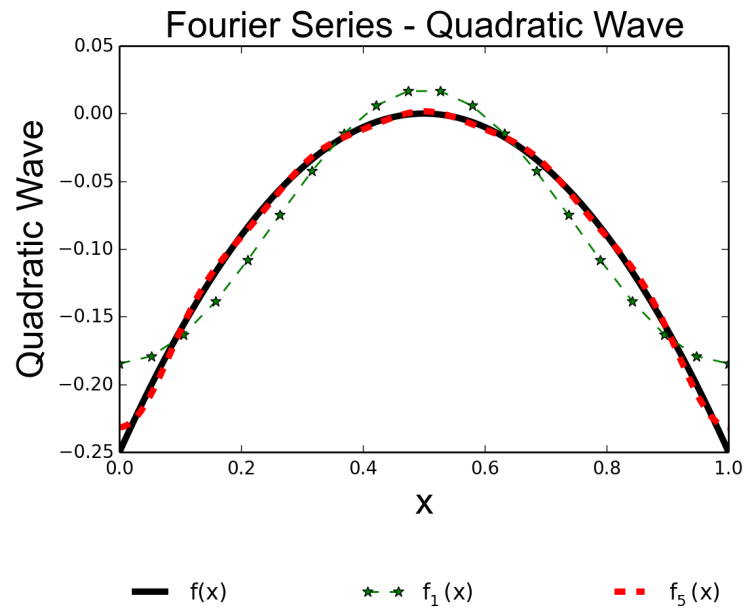


Figure 4.6: The truncated Fourier series, $f_N(x)$ approaches the original function, $f(x)$, with each new term added. The original function (black) is shown compared to the Fourier series with one term, $f_1(x)$ (green stars). Another term is added until we consider the Fourier series with five terms, $f_5(x)$ (red dashes).

The function with the largest period which corresponds to the first term in Equation (4.91) is removed from the series and labeled $\widehat{U}(x, L)$

$$\widehat{U}(x, L) = \left(\frac{-L}{12} - \left(\frac{L}{\pi} \right)^2 \cos \left(\frac{2\pi x}{L} \right) \right). \quad (4.93)$$

Each periodic function with a smaller potential remains in the summation, that is,

$$V(x) = \sum_{i=2}^N U_{i-1}(x, L_i) + \widehat{U}(x, L) - Fx \quad (4.94)$$

where

$$U_{i-1}(x, L/i) = - \left(\frac{L}{i\pi} \right)^2 \cos \left(\frac{2\pi i x}{L} \right), \quad i = 2, 3, \dots, N. \quad (4.95)$$

Using the multiplicative approximation from Theorem 4.0.1 we calculate

$$\tilde{\mu}_N = (2D)^{(N-1)} \mu_1(1) \mu_2(1) \cdot \dots \cdot \mu_{N-1}(1) \widehat{\mu}(L) + \mathcal{O}(L/2). \quad (4.96)$$

The MST for this example can be calculated three different ways, as μ , μ_N , and $\tilde{\mu}_N$. For all three calculations we used the parameters $D = 2$ and $L = 1$. The results of all three, for increasing values of N are given in Figure 4.7. Notice, in Figure 4.7, $\tilde{\mu}_N$ appears to be identical to μ_N . This is because the difference between the two results is calculated to be on the order of 10^{-8} , which is quite a bit smaller than the $\mathcal{O}(\frac{1}{2})$ computed in Equation (4.96). This difference is invisible in the figure. Now if we shift our attention to μ we see a noticeable difference. This difference μ and μ_N is 0.051734 for the truncated function with one term and decreases to 0.051707 for the truncated function with twenty terms. From Theorem 4.0.4, the error will approach zero as the number of terms in the truncated series, N , approaches infinity though the speed of that convergence is unknown.

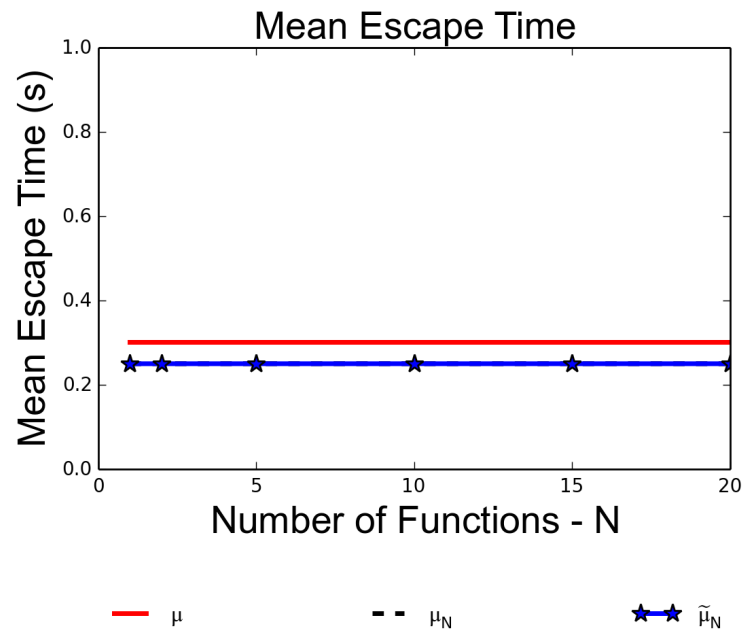


Figure 4.7: The MST for the original quadratic potential, μ , and the Fourier series μ_N are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96), $\tilde{\mu}_N$ is displayed with blue stars.

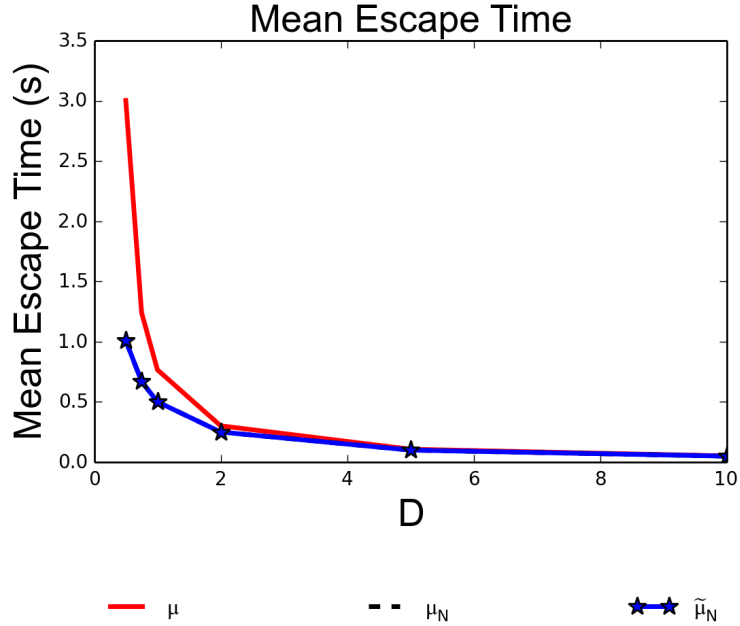


Figure 4.8: The MST for the original quadratic potential μ , and the Fourier series μ_N are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96), $\tilde{\mu}_N$, is displayed with blue stars. This figure shows the results for different values of D with $N = 20$ as a fixed quantity.

Next consider the MST for a fixed $N = 20$ with different values of the noise intensity D . From our discussion of the error estimate, $\mathcal{O}\left(\frac{L}{N_1}\right)$, in the proof for Lemma 4.0.3, we derived Equation (4.50) for the finite constant C . From Equation (4.50) it is expected that the error would increase as D decreases. The results of these calculations are displayed in Figure 4.8. The difference in μ_N and $\tilde{\mu}_N$ for $D = 0.5$ is on the order of 10^{-5} . This quantity increases to 10^{-2} for $D = 0.1$ (not pictured in Figure 4.8) however μ , for $D = 0.1$ is 997 seconds. While the difference between μ and μ_N does increase dramatically for small values of D , $\tilde{\mu}_N$ remains close to μ_N . The difference between μ and μ_N is also expected to increase as D decreases, from our proof of Theorem 4.0.4. While $\mu_N \rightarrow \mu$ as $N \rightarrow \infty$ from Theorem 4.0.4, $|\mu - \mu_N|$

will take more terms to go to zero for each smaller value of D . As we are holding $N = 20$ fixed, the error will continue to increase as D decreases.

For values of $D > 1$, Equation (4.96) provides excellent agreement to Equation (4.19). From here we will look at another example that is not as well behaved as a continuous function.

Square Wave Consider a square wave given by

$$f(x) = 2 \left[H \left(\frac{2x}{L} \right) - H \left(\frac{2x}{L} - 1 \right) \right] - 1 \quad (4.97)$$

with the same periodic extension $f(x - nL)$ as discussed in the previous section. In this equation, $H(x)$ is the Heaviside step function

$$H(x) = \begin{cases} 0 & x < 0 \\ 1/2 & x = 0 \\ 1 & x > 0 \end{cases} \quad (4.98)$$

Therefore $f(x)$ may also be written as a piecewise function,

$$f(x) = \begin{cases} 0 & x < 0 \\ 1/2 & x = 0 \\ 1 & 0 < x < L/2 \\ 0 & x = L/2 \\ -1 & L/2 < x < L \end{cases} \quad (4.99)$$

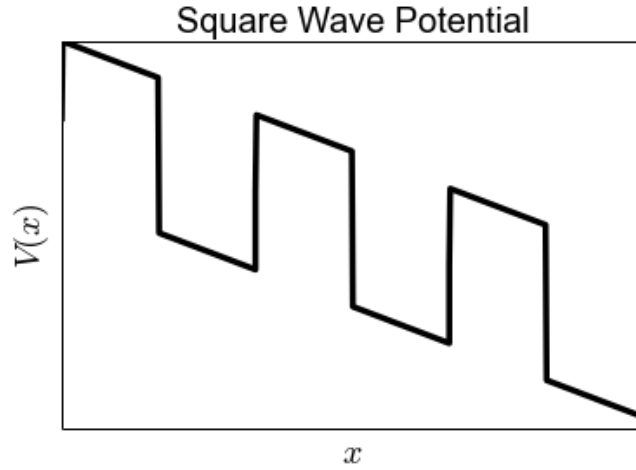


Figure 4.9: The tilted potential function $V(x)$ with $F = 0.95$ and $U(x, L)$ as the periodic extension of the square wave in Equation (4.97) on the interval $[0, 3L]$.

The potential function $V(x)$ for the square wave is then

$$V(x) = U(x, L) - Fx \quad (4.100)$$

where $U(x, L)$ is the periodic extension of Equation (4.97), and $F = 0.95$. The graph of this potential on the interval $[0, 3L]$ can be seen in Figure 4.9.

The square wave function is not everywhere continuous or differentiable as there are jump discontinuities at $x = nL/2$ for $n \in \mathbb{N}$. It is however piecewise continuous, hence the Fourier Convergence Theorem applies and we can look at the Fourier series for the square wave as derived in Appendix C. This Fourier series is as follows

$$f(x) = \sum_{n=1}^{\infty} \frac{4}{\pi(2n-1)} \sin\left(\frac{2\pi(2n-1)x}{L}\right). \quad (4.101)$$

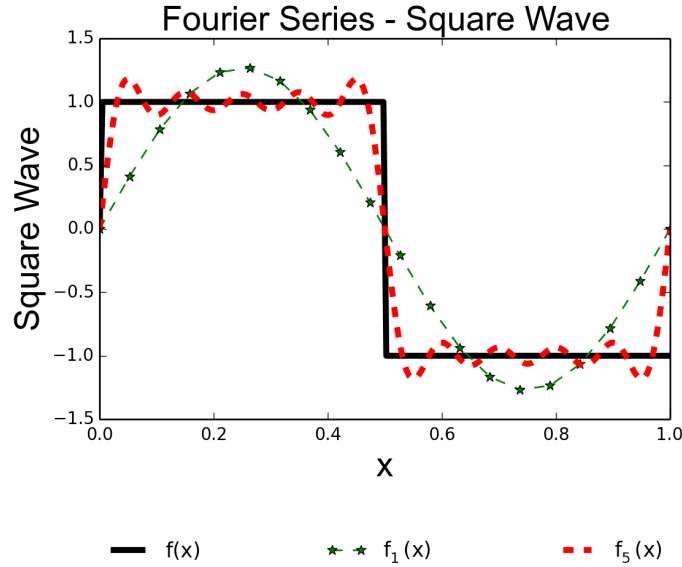


Figure 4.10: The truncated Fourier series approaches the original square wave function with each new term added. The original function, $f(x)$ (black), is shown compared to the Fourier series with one term, $f_1(x)$ (green dashes). Another term is added until we consider the Fourier series with five terms, $f_5(x)$ (red dashes).

The truncated Fourier series is a finite sum

$$f_N(x) = \sum_{n=1}^N \frac{4}{\pi(2n-1)} \sin\left(\frac{2\pi(2n-1)x}{L}\right). \quad (4.102)$$

The difference between the truncated Fourier series and the original function can be observed from the plot of each in Figure 4.10.

Notice from Figure 4.10, we do not have the close agreement between the truncated Fourier series and the original function at five terms as we had in the previous example. This is a direct result of the jump discontinuities explained by Gibbs phenomenon [34, 87]. To briefly summarize Gibbs phenomenon, the truncated Fourier series will overshoot the original function at jump discontinuities. Furthermore, this discrepancy persists for increasing values of N . The overshoot

will approach a finite value in the limit as N goes to infinity. This discrepancy is important to keep in mind in this example as we would expect higher error values between the MST of the original square wave and its truncated Fourier series than we would experience with a continuous function.

Now consider the truncated Fourier series in the context of the tilted periodic potential,

$$V(x) = \sum_{n=2}^N \frac{4}{\pi(2n-1)} \sin\left(\frac{2\pi(2n-1)x}{L}\right) + \frac{4}{\pi} \sin\left(\frac{2\pi x}{L}\right) - Fx. \quad (4.103)$$

The first term is removed from the summation as before for $\widehat{U}(x, L)$, the function with the largest period,

$$\widehat{U}(x, L) = \frac{4}{\pi} \sin\left(\frac{2\pi x}{L}\right). \quad (4.104)$$

The periodic functions with smaller potentials are in the summation,

$$U_{i-1}(x, L/(2i-1)) = \frac{4}{\pi(2i-1)} \sin\left(\frac{2\pi(2i-1)x}{L}\right), \quad i = 2, 3, \dots, N. \quad (4.105)$$

To be clear, even though the original function $f(x)$ is not C^1 , the truncated Fourier series is a finite number of well behaved sine functions. Therefore the truncated Fourier series is not only C^1 but is in fact analytic, making $V(x)$ infinitely differentiable. Since our assumptions are satisfied we may use our result from Theorem 4.0.1, to calculate the MST as in Equation (4.96).

Once again using the parameters $D = 2$ and $L = 1$, the results of all three calculations, for increasing values of N are given Figure 4.11. The results for the potential in Equation (4.96) in Figure 4.7 are again very close. The difference is on the order of 10^{-3} . Even the difference between μ and μ_N is on average 0.0576 for $N = 1$ and decreased to 0.0453 for $N = 20$. The differences begin to emerge when

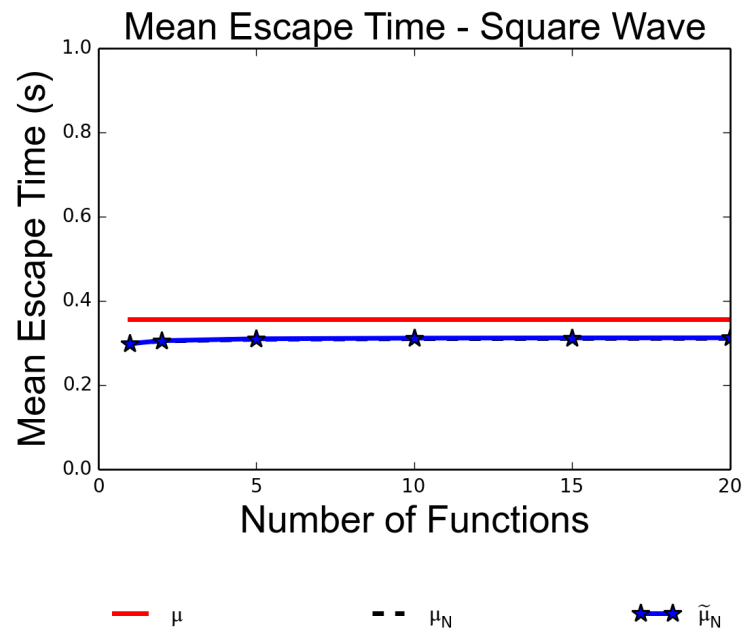
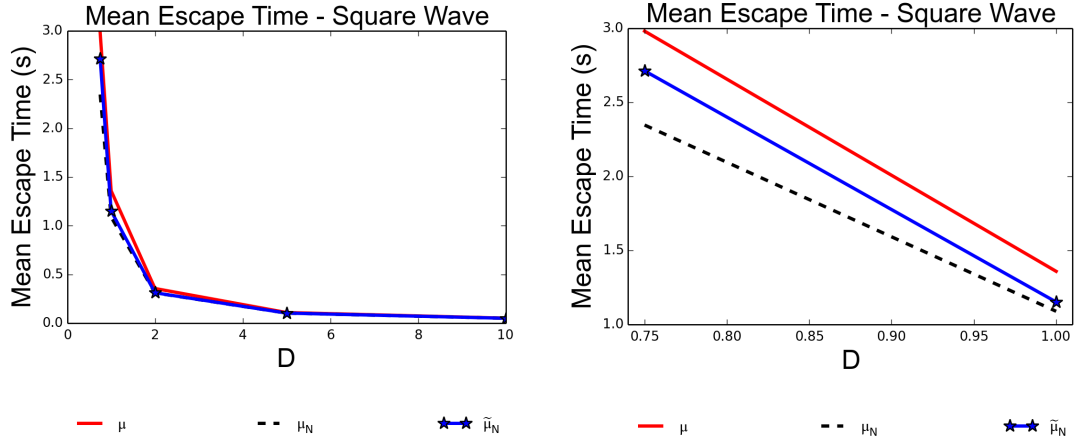


Figure 4.11: The MST for the original square wave potential, μ , and the Fourier series, μ_N , are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96), $\tilde{\mu}_N$ is displayed with blue stars.



(A) The results for decreasing values of D . (B) For clarity $D = 0.75$ and $D = 1$ are plotted here.

Figure 4.12: The MST for the original quadratic potential μ and the Fourier series μ_N , are shown in red and black dashes respectively. The MST for the Fourier series calculated with Equation (4.96) $\tilde{\mu}_N$, is displayed with blue stars. This figure shows the results for different values of D with $N = 20$ as a fixed quantity.

considering different values of the noise intensity D for $N = 20$ fixed. These results are shown in Figure 4.12. Since the lower values of D are difficult to see in Figure 4.12A, a plot of the two lowest values of D is provided in Figure 4.12B. For $D = 0.75$, μ is approximately 2.98 seconds, and μ_N is 2.34 seconds, while $\tilde{\mu}_N$ is 2.71 seconds. This difference continues to increase for lower values of D .

Again we reiterate that μ_N and $\tilde{\mu}_N$ have excellent agreement, especially for $D \geq 1$. For values of $D \in [0, 1)$ the MST calculations begin to diverge from each other for a fixed value of N .

Noise Intensity D for DNA Transcription

The parameter D , the noise intensity described in Equation (4.5), is clearly an important quantity in diffusive transport models. One of the difficulties of employing

a diffusive transport model for DNA transcription is approximating the correct parameter value for D . This parameter is an inherent quantity of the RNAP and many of the other estimated parameters when tuning the model depend on the value of D . The Boltzmann constant, $k_{\beta}T$, is found experimentally to be 4.1 pN·nm [53,58]. The friction coefficient, γ , is given by the sum of the rotational and translational friction. By approximating the enzyme as a sphere of radius R , approximately 6.5 nm [63], we have

$$\gamma = 6\pi\eta R + \left(\frac{2\pi}{10BP}\right)^2 8\pi\eta R^3 \quad (4.106)$$

as developed in [72] where η is the viscosity of the medium, typically water, which is estimated to be 10^{-9} pN·s/nm², and BP is the distance between basepairs, i.e. 0.34 nm. Therefore, according to Equation (4.106), $\gamma = 2.4 \times 10^{-5}$ pN·s/nm. Using these values for $k_{\beta}T$ and γ , we calculate D to be approximately 1.7×10^5 nm²/s, which agrees with the experimental value of 1.5×10^5 nm²/s from [76] and is close to the one dimensional noise intensity for a lac repressor protein moving along DNA, 3×10^5 nm²/s [8]. Converting D from nanometers into nucleotides,

$$D = 1.7 \times 10^5 \text{nm}^2/\text{s} \times \frac{1\text{nt}^2}{0.34^2\text{nm}^2} = 1.47 \times 10^6 \text{nt}^2/\text{s}. \quad (4.107)$$

This value of D is very large as the units for D are in terms of nucleotides and seconds. One possibility is to use this value of D in the diffusive transport model with the period $L/N = 1$ nucleotide for translocation and a pause period $L = 100$ nucleotides [61]. The other possibility is to scale the value of D to change the units from nucleotides into a different measurement length. For instance, the value of D given in nucleotides in Equation (4.107), can be changed in to centimeters,

$$D = 1.7 \times 10^{-9} \text{cm}^2/\text{s}. \quad (4.108)$$

This is an extremely small value of D and the periods for translocation, L/N , and pausing, L , would be scaled accordingly in the model to reflect nucleotide values expressed in centimeters. According to preliminary calculations, scaling D , L/N and L to be in centimeters instead of nucleotides maintains excellent agreement between μ in Equation (4.19) and the multiplicative approximation in Equation (4.24). However, the MST calculated by these formulas should remain fixed regardless of the units and in our preliminary investigation, the MST varied under scaling. More investigation is needed on how to properly scale the parameters in the MST calculations to model DNA transcription, and nondimensionalization will be explored.

Discussion

Diffusive transport in a tilted periodic potential, $V(x)$, is characterized by three important quantities: drift velocity, effective diffusion, and the Péclet number. The equations for each can be found in Equation (4.16), (4.17) and (4.18) respectively. There are two quantities that are common in each equation, the probability of leaving a segment on the left and right, p_- and p_+ respectively, and the MST, μ . We examined both of these quantities in this chapter.

The focus of our analysis has been the effect on the MST when adding functions to the periodic potential, without changing the period of the function. For $V(x) = U(x) - Fx$ with $U(x)$ periodic with period L , we have shown the the probability of escaping a segment on the right or the left is unchanged with the addition of more periodic functions so long as the period of $U(x)$ is fixed. The escape probability does depend on the tilt parameter F and the main period L , as well as the parameter D . Two different potential functions, $V_1(x) = U(x, L) - Fx$, and $V_2(x) = U_1(x, L/N_1) + \dots + U_n(x, L/N_n) + \widehat{U}(x, L) - Fx$, using the same parameters F , L , and D , have the

same probability of escaping right and left. However, the MST for $V_1(x)$ and $V_2(x)$ would be altered.

The MST calculated by Equation (4.19), (and derived in Appendix A), depends explicitly on the function $U(x)$ and how it changes with the addition of subsequent periodic functions. For example, if $V_1(x) = U_1(x, L) - Fx$ and $V_2(x) = U_1(x, L) + U_2(x, L/N) - Fx$ have the same probability of escape for $N \in \mathbb{N}$, but a different MST even though the period remains the same. We derived an expression for the MST of a tilted periodic potential composed of several periodic functions, in terms of each of the periodic functions. This formula given by Equation (4.24) in Theorem 4.0.1 is not limited to two functions, but is valid for any finite number of periodic functions. Theorem 4.0.1 is easily proven with induction.

One very clear application for Equation (4.24) is for Fourier series as shown by our two examples. For the range of noise intensity $D \geq 1$ there is excellent agreement between the two MST estimates, μ_N and the multiplicative approximation of μ_N , even with a highly oscillatory potential as is produced by the Fourier series of a function with jump discontinuities. Though we have not explicitly analyzed the error between the MST of a function and its truncated Fourier series, convergence of the MST of the Fourier series was proven for continuous functions. Quantifying the error of the MST for the truncated Fourier series as well as proving convergence of the MST for Fourier series of piecewise continuous functions will be of primary interest for future work.

CONCLUSION

DNA transcription is a complicated process with several aspects that remain under investigation by researchers. Some examples include interactions of RNAPs, and specific mechanisms for movement of polymerases. Even RNAP translocation of one basepair involves several steps that could be studied in minute detail. With newer technological advances, more is known about transcription than ever before. The challenge for mathematicians is to incorporate new information into models for transcription that more accurately reflect what is occurring biologically. This thesis investigates the mathematical modeling of two recent breakthroughs in DNA transcription.

In 2003, it was discovered that RNAPs cooperate with each other during transcription. Polymerases experience faster transcription times with multiple RNAPs reading a gene simultaneously than when compared to single molecule transcription [29]. Throughout the next decade researchers found that RNAPs introduce torsion into a DNA strand as they translocate. Then in 2013, Ma *et. al.* published experimental results that quantified this torque with actual data [53]. To incorporate interactions of RNAPs during transcription, we added this torque data to a basic TASEP (Totally Asymmetric Single Exclusion Process) model. This new model is referred to as ETAM (Elongation with Torque Assisted Motion). The torque allows RNAPs to communicate with each other during transcription and maintain a degree of spacing and separation throughout the transcription process. With the influence of torque on velocity, pause frequency and pause duration, as published in [53], we are able to simulate faster transcription times that are more biologically realistic than the transcription times obtained using the basic TASEP model. The faster transcription times are a direct result of the torque governing the interactions

between RNAPs. The RNAPs are able to avoid hundreds of collisions and experience significantly shorter pause durations on average. Both of these results consequently led to less average delay experienced per RNAP during transcription.

While we are encouraged by the results of ETAM, we would like to develop the model further. We make some assumptions with respect to how best to fit the data that greatly influence the model's behavior. In other words, our results are not robust under changes to this data fit. With limited data available at this time, some assumptions are necessary. However we are confident that more data will become available as nanotechnology continues to improve.

Another relatively new development in DNA transcription is the differentiation between the power stroke model for RNAP translocation and the Brownian ratchet model. Nanotechnology allowed researchers to investigate the specific movements of RNAPs and most researchers now agree that RNAP translocation is an example of a Brownian ratchet. Specifically, a polymerase will move back and forth from a pre-translocation state and a post-translocation state several times before stabilizing one nucleotide downstream. The RNAP uses its own thermal energy to produce these oscillations in an example of Brownian motion.

We were inspired to explore RNAP elongation and pausing further by using equations detailed by Lindner *et. al.* [52]. By adding two periodic functions together in a tilted periodic potential, it is shown that one can model translocation with one function and pausing with the other. Tilted periodic potentials are used often in diffusive particle transport and have become an interesting research topic in physics over the past 20 years. While modeling DNA transcription and pausing is the inspiration for our focus on diffusive transport, we have been focused largely on the addition of potentials and its effect on the mean escape time of a particle from a potential segment. To this end we derived a formula for mean escape time that

explicitly incorporates the mean escape time of each individual periodic function in the potential. While this formula has a lot of applications in a variety of fields, we are excited about the application in Fourier series as mentioned in two such examples.

While we were able to prove the validity of our mean escape time formula for any finite number of functions, investigating the error between the mean escape time of a periodic function and the function's truncated Fourier series is desired. One future challenge will be to research the quantification of the error as a function of the number of terms in the series, especially if the original periodic function is piecewise continuous as opposed to C^1 . These are interesting challenges yet to be addressed in the field of diffusive transport.

REFERENCES CITED

- [1] E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block. Direct observation of base-pair stepping by RNA polymerase. *Nature*, 438(7067):460–465, 2005.
- [2] K. Adelman, A. La Porta, T. J. Santangelo, J. T. Lis, J. W. Roberts, and M. D. Wang. Single molecule analysis of rna polymerase elongation reveals uniform kinetic behavior. *Proc. Natl. Acad. Sci.*, 99(21):13538–13543, 2002.
- [3] T. Apostol. *Calculus. One-Variable Calculus with an Introduction to Linear Algebra*, volume 1. John Wiley and Sons, 1966.
- [4] I. Artsimovitch and R. Landick. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci.*, 97(13):7090–7095, 2000.
- [5] L. Bai, A. Shundrovsky, and M. D. Wang. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *Journal of molecular biology*, 344(2):335–349, 2004.
- [6] G. Bar-Nahum, V. Epshtein, A. E. Ruckenstein, R. Rafikov, A. Mustaev, and E. Nudler. A ratchet mechanism of transcription elongation and its control. *Cell*, 120(2):183–193, 2005.
- [7] C. Baumann, S. Smith, V. Bloomfield, and C. Bustamante. Ionic effects on the elasticity of single DNA molecules. *Proc. Natl. Acad. Sci.*, 94(12):6185–6190, June 1997.
- [8] O. G. Berg and C. Blomberg. Association kinetics with coupled diffusional flows: special application to the lac repressor-operator system. *Biophysical chemistry*, 4(4):367–381, 1976.
- [9] C. Brackley, M. Romano, C. Grebogi, and M. Thiel. Limited resources in a driven diffusion process. *Physical Review Letters*, 105, 2010.
- [10] B. P. Bratton, R. A. Mooney, and J. C. Weisshaar. Spatial distribution and diffusive motion of RNA polymerase in live Escherichia coli. *Journal of bacteriology*, 193(19):5138–5146, 2011.
- [11] H. Bremer and P. Dennis. Modulation of chemical composition and other parameters in the cell by growth rate. In F. C. Neidhardt, editor, *Escherichia coli and Salmonella*, pages 1553 – 1569. ASM Press, 1996.
- [12] H. Chen, K. Shiroguchi, H. Ge, and X. S. Xie. Genome-wide study of mrna degradation and transcript elongation in escherichia coli. *Molecular systems biology*, 11(1):781, 2015.

- [13] R.-V. Churchill. Fourier series and boundary value problems. 1961.
- [14] L. Ciandrini, I. Stansfield, and M. Romano. Role of the particle's stepping cycle in an asymmetric exclusion process: model of mRNA translation. *Physical Review E*, 81(2), 2010.
- [15] C. Condon, S. French, C. Squires, and C. Squires. Depletion of functional ribosomal RNA operons in escherichia coli causes increased expression of the remaining intact copies. *The EMBO Journal*, 12(11):4305 – 4315, 1993.
- [16] C. Condon, D. Liveris, C. Squires, I. Schwartz, and C. Squires. rRNA operon multiplicity in Escherichia coli and the physiological implications of rrn inactivation. *Journal of Bacteriology*, 177(14):4152 – 4156, 1995.
- [17] P. R. Cook. The organization of replication and transcription. *Science*, 284(5421):1790 – 1795, 1999.
- [18] G. Costantini and F. Marchesoni. Threshold diffusion in a tilted washboard potential. *EPL (Europhysics Letters)*, 48(5):491, 1999.
- [19] D. R. Cox. *Renewal Theory*. Methuen and Co LTD, 1962.
- [20] N. Crampton, W. Bonass, J. Kirkham, C. Rivetti, and N. Thomson. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Research*, 34(19):5416 – 5425, 2006.
- [21] M. Dangkulwanich, T. Ishibashi, L. Bintu, and C. Bustamante. Molecular mechanisms of transcription through single-molecule experiments. *Chemical reviews*, 114(6):3203–3223, 2014.
- [22] J. Davenport, J. Wuite, R. Landick, and C. Bustamante. Single-molecule study of transcriptional pausing and arrest by E. coli RNA polymerase. *Science*, 287(5462):2497 – 2500, 2000.
- [23] L. Davis, T. Gedeon, J. Gedeon, and J. Thorenson. A traffic flow model for bio-polymerization processes. *Journal of Mathematical Biology*, 68(3):667 – 700, 2014.
- [24] L. Davis, T. Gedeon, and J. Thorenson. Discontinuous Galerkin calculations for a nonlinear PDE model of DNA transcription with short, transient and frequent pausing. *Journal of Computational Mathematics*, 32(6):601 – 629, 2014.
- [25] P. Dennis, M. Ehrenberg, D. Fange, and H. Bremer. Varying rate of RNA chain elongation during rrn transcription in Escherichia coli. *Journal of bacteriology*, 191(11):3740–3746, 2009.

- [26] M. Depken, E. Galburt, and S. Grill. The origin of short transcriptional pauses. *Biophysical Journal*, 96(6):2189 – 2193, 2009.
- [27] M. Depken, E. A. Galburt, and S. W. Grill. The origin of short transcriptional pauses. *Biophysical journal*, 96(6):2189–2193, 2009.
- [28] A. Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annalen der physik*, 4, 1905.
- [29] V. Epshtein and E. Nudler. Cooperation between RNA polymerase molecules in transcription elongation. *Science*, 300(5620):801 – 805, 2003.
- [30] R. Feynman, R. Leighton, and M. Sands. *The Feynman Lectures On Physics*, volume II. Addison-Wesley Publishing Company, Reading, Massachusetts, 1964.
- [31] C. W. Gardiner. *Handbook of stochastic methods*, volume 3. Springer, 2004.
- [32] K. Geszvain and R. Landick. The structure of bacterial RNA polymerase. *The Bacterial Chromosome*, pages 283 – 296, 2005.
- [33] C. Giardina. Bounds on the truncation error of periodic signals. *Circuit Theory, IEEE Transactions on*, 19(2):206–207, 1972.
- [34] J. W. Gibbs. Fourier’s series. *Nature*, 59:606, 1899.
- [35] A. Griffiths, W. Gelbart, J. Miller, and R. Lewontin. *Modern Genetic Analysis*. W. H. Freeman and Co., New York, 1999.
- [36] R. Guajardo and R. Sousa. A model for the mechanism of polymerase translocation. *Journal of molecular biology*, 265(1):8–19, 1997.
- [37] Q. Guo and R. Sousa. Translocation by T7 RNA polymerase: a sensitively poised brownian ratchet. *Journal of molecular biology*, 358(1):241–254, 2006.
- [38] M. Guthold, X. Zhu, C. Rivetti, G. Yang, N. H. Thomson, S. Kasas, H. G. Hansma, B. Smith, P. K. Hansma, and C. Bustamante. Direct observation of one-dimensional diffusion and transcription by Escherichia coli RNA polymerase. *Biophysical journal*, 77(4):2284–2294, 1999.
- [39] Z. Gyorfy, G. Draskovits, V. Vernyik, F. Blattner, T. Gaal, and G. Posfai. Engineered ribosomal RNA operon copy-number variants of E. coli reveal the evolutionary trade-offs shaping rRNA operon number. *Nucleic Acid Research*, 2015.
- [40] Y. Harada, O. Ohara, A. Takatsuki, H. Itoh, N. Shimamoto, and K. Kinosita. Direct observation of DNA rotation during transcription by *escherichia coli* RNA polymerase. *Nature*, 409:113 – 155, 2000.

- [41] A. Herr, J. Atkins, and R. Gesteland. Coupling of open reading frames by translational bypassing. *Annual Review Biochemistry*, 69:343 – 372, 2000.
- [42] G. Kassavetis and M. Chamberlin. Pausing and termination of transcription within the early region of bacteriophage T7 DNA in vitro. *Journal of Biological Chemistry*, 256:2777 – 2786, 1981.
- [43] J. H. Kim and R. G. Larson. Single-molecule analysis of 1D diffusion and transcription elongation of T7 RNA polymerase along individual stretched DNA molecules. *Nucleic acids research*, 35(11):3848–3858, 2007.
- [44] R. Kingston and M. Chamberlin. Pausing and attenuation of in vitro transcription in the *rrnB* operon of *E. coli*. *Cell*, 27(3 Pt 2):523531, 1981.
- [45] S. Klumpp. Pausing and backtracking in transcription under dense traffic conditions. *Journal of Statistical Physics*, 142:1252 – 1267, 2011.
- [46] S. Klumpp and T. Hwa. Stochasticity and traffic jams, in transcription of ribosomal RNA: Intriguing role of termination and antitermination. *Proc. Natl. Acad. Sci.*, 105(47):18159 – 18164, 2008.
- [47] N. Komissarova and M. Kashlev. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *Journal of Biological Chemistry*, 272(24):15329–15338, 1997.
- [48] T. W. Körner. *Fourier analysis*. Cambridge university press, 1989.
- [49] B. Langa, M. Jakubkovab, E. Hegedusovab, R. Daouda, T. Vinarc, G. Burgera, P. Kosab, D. Fricovab, M. Nebohacovab, and J. Nosekb. Massive programmed translational jumping in mitochondria. *Proc. Natl. Acad. Sci.*, 111(16), 2014.
- [50] M. H. Larson, R. Landick, and S. M. Block. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Molecular cell*, 41(3):249–262, 2011.
- [51] M. Levitt. How many base-pairs per turn does DNA have in solution and in chromatin? Some theoretical calculations. *Proc. Natl. Acad. Sci. USA*, 75(2):640 – 644, 1978.
- [52] B. Lindner, M. Kostur, and L. Schimansky-Geier. Optimal diffusive transport in a tilted periodic potential. *Fluctuation and Noise Letters*, 1(1), 2001.
- [53] J. Ma, L. Bai, and M. Wang. Transcription under torsion. *Science*, 340(1580), 2013.

- [54] X.-g. Ma, P.-Y. Lai, B. J. Ackerson, and P. Tong. Colloidal transport and diffusion over a tilted periodic potential: dynamics of individual particles. *Soft matter*, 11(6):1182–1196, 2015.
- [55] G. S. Manning. Correlation of polymer persistence length with euler buckling fluctuations. *Physical Review A*, 34(5), 1986.
- [56] G. S. Manning. Poissons ratio for a polyelectrolyte. *Soft Matter*, 8(9334), 2012.
- [57] M. Manor, D. Goodman, and G. Stent. RNA chain growth rates in Escherichia coli. *Journal of Molecular Biology*, 39(1):1 – 29, 1969.
- [58] J. Marko and E. Siggia. Stretching DNA. *Macromolecules*, 28:8759–8770, 1995.
- [59] J. Munkres. *Topology*. 2000.
- [60] K. Neuman, E. Abbondanzieri, R. Landick, J. Gelles, and S. Block. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell*, 115(4):437 – 447, 2003.
- [61] K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell*, 115(4):437–447, 2003.
- [62] E. Nudler, A. Mustaev, A. Goldfarb, and E. Lukhtanov. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, 89(1):33–41, 1997.
- [63] N. Opalka, J. Brown, W. Lane, K. Twist, and R. Landick. Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. *PLOS Biology*, 2010.
- [64] P. Reimann. Brownian motors: noisy transport far from equilibrium. *Physics reports*, 361(2):57–265, 2002.
- [65] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. Rubi, and A. Pérez-Madrid. Giant acceleration of free diffusion by use of tilted periodic potentials. *Physical review letters*, 87(1):010602, 2001.
- [66] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. Rubi, and A. Pérez-Madrid. Diffusion in tilted periodic potentials: Enhancement, universality, and scaling. *Physical Review E*, 65(3):031104, 2002.
- [67] H. Risken. *Fokker-Planck Equation*. Springer, 1984.
- [68] J. Rose, R. Mosteller, and C. Yanofsky. Tryptophan messenger ribonucleic acid elongation rates and steady-state levels of tryptophan operon enzymes under various growth conditions. *Journal of Molecular Biology*, 51(3):541 – 550, 1970.

- [69] W. Rudin. *Principles of mathematical analysis*, volume 3. McGraw-Hill New York, 1964.
- [70] J. Ryals, R. Little, and H. Bremer. Temperature dependence of RNA synthesis parameters in *Escherichia coli*. *Journal of Bacteriology*, 151(2):879 – 887, 1982.
- [71] M. Sahoo and S. Klumpp. Backtracking dynamics of rna polymerase: pausing and error correction. *Journal of Physics: Condensed Matter*, 25(37):374104, 2013.
- [72] J. Schurr. The one-dimensional diffusion coefficient of proteins absorbed on DNA. hydrodynamic considerations. *Biophysical Chemistry*, 9:413 – 414, 1979.
- [73] J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, and S. M. Block. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, 426(6967):684–687, 2003.
- [74] K. Shaw. Negative transcription regulation in prokaryotes. *Nature Education*, 1(1):122, 2008.
- [75] D.-A. Silva, D. R. Weiss, F. P. Avila, L.-T. Da, M. Levitt, D. Wang, and X. Huang. Millisecond dynamics of RNA polymerase ii translocation at atomic resolution. *Proc. Natl. Acad. Sci.*, 111(21):7665–7670, 2014.
- [76] P. Singer and C. Wu. Promoter search by *Escherichia coli* RNA polymerase on a circular DNA template. *Journal of Biological Chemistry*, 262(29):14178 – 14189, 1987.
- [77] M. Smoluchowski. Experimentell nachweisbare, der üblichen thermodynamik widersprechende molekularphänomene. *Pisma Mariana Smoluchowskiego*, 1(2):226–251, 1927.
- [78] V. R. Tadigotla, D. Ó. Maoiléidigh, A. M. Sengupta, V. Epshtein, R. H. Ebright, E. Nudler, and A. E. Ruckenstein. Thermodynamic and kinetic modeling of transcriptional pausing. *Proc. Natl. Acad. Sci.*, 103(12):4439–4444, 2006.
- [79] H. M. Taylor and S. Karlin. *An introduction to stochastic modeling*. Academic press, 2014.
- [80] S. Tolić-Norrelykke, A. Engh, R. Landick, and J. Gelles. Diversity in the rates of transcript elongation by single RNA polymerase molecules. *Journal of Biological Chemistry*, 279:3292 – 3299, 2004.
- [81] I. Touloukhonov, J. Zhang, M. Palangat, and R. Landick. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Molecular cell*, 27(3):406–419, 2007.

- [82] U. Vogel and K. Jensen. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *Journal of Bacteriology*, 176(10):2807 – 2813, 1994.
- [83] U. Vogel and K. Jensen. Effects of the antiterminator boxA on transcription elongation kinetics and ppGpp inhibition of transcription elongation in *Escherichia coli*. *Journal of Biological Chemistry*, 270:18335 – 18340, 1995.
- [84] A. Voter. Introduction to the Kinetic Monte Carlo method. In K. Sickafus and E. Kotomin, editors, *Radiation Effects in Solids*. Springer, NATO Publishing Unit, Dordrecht, The Netherlands, 2005.
- [85] H. Wang. Effective diffusion and effective drag coefficient of a Brownian particle in a periodic potential. *Acta Mathematica Scientia*, 31B(6):2323 – 2342, 2011.
- [86] R. Weiss, W. Huang, and D. Dunn. A nascent peptide is required for ribosomal bypass of the coding gap in bacteriophage T4 gene 60. *Cell*, 62(1), 1990.
- [87] H. Wilbraham. On a certain periodic function. *Cambridge and Dublin Math. J.*, 3:198–201, 1848.
- [88] D. Wilkinson. *Stochastic Modelling for Systems Biology*. CRC Press, London, 2 edition, 2011.
- [89] S. Zhang and H. Bremer. Effects of Fis on ribosome synthesis and activity and on rRNA promoter activities in *escherichia coli*. *Journal of Molecular Biology*, 259:27 – 40, 1996.
- [90] R. Zia, J. Dong, and B. Schmittmann. Modeling translation in protein synthesis with TASEP: a tutorial and recent developments. *Journal of Statistical Physics*, 144(2):405 – 428, 2011.

APPENDICES

APPENDIX A

DERIVATION OF ESCAPE TIME QUANTITIES

There are three escape time quantities that we will derive. The mean escape time, μ , is the average time it takes a particle to escape a segment. The variance of the escape time, σ^2 , indicates how far the values of escape time are from the mean. With a small variance the values of escape time are very close to the mean, a larger variance indicates that the escape time values are spread out from the mean. The third escape time quantity is the mean first passage time. This quantity is a function of the starting position of the particle in the segment and calculates how long on average the particle takes to escape the segment.

Mean First Passage Time

We will begin by deriving $T_1(x)$ which represents the mean first passage time of a particle beginning in the segment m_i at position x and exiting the segment through either the right end L or the left end $-L$, see Figure 4.2. The segment has absorbing boundary conditions as discussed in Chapter 4, therefore the particle is removed from m_i when it reaches either end. Let the probability that the particle is at x' at time t given that it started at x at time $t = 0$ be denoted as a C^1 function of x' , $p(x', t|x, 0)$. Then the probability that the particle is still in the segment m_i at time t is given by

$$\int_{-L}^L p(x', t|x, 0) dx' \equiv G(x, t). \quad (\text{A.1})$$

Suppose the time that the particle leaves m_i is T . We can rewrite Equation (A.1) as

$$\text{Prob}(T \geq t) = \int_{-L}^L p(x', t|x, 0) dx = G(x, t) \quad (\text{A.2})$$

Rewriting the Fokker-Planck equation (FPE) given by Equation (4.6) using the conditional probability $p(x', t|x, 0)$, we have

$$\partial_t p(x', t|x, 0) = \mathcal{L}_x p(x', t|x, 0) \quad (\text{A.3})$$

where $\mathcal{L}_x p(x', t|x, 0) = \partial_x(V'(x)p(x', t|x, 0)) + D\partial_x^2(p(x', t|x, 0))$. The domain of \mathcal{L}_x is

$$\text{Dom}(\mathcal{L}_x) = \{p \in C^2(-L, L) \mid p(-L) = p(L) = 0\}$$

since the probability of being in the segment at the endpoints is 0. Equation (A.3), often referred to as the backward FPE, has the initial condition

$$p(x', 0|x, 0) = \delta(x - x').$$

Since $G(x, t)$ is the probability that the particle is in segment m_i at time t , the probability of escaping by time t is $1 - G(x, t)$. The rate of change of $1 - G(x, t)$ at a fixed time T , gives the distribution function for the first passage time

$$g(x, T) = -\frac{\partial G(x, T)}{\partial T} = -\int_{-L}^L \frac{\partial p(x', T|x, 0)}{\partial T} dx'. \quad (\text{A.4})$$

Hence the formula for the mean first passage time is

$$\begin{aligned} T_1(x) &= \int_0^\infty t g(x, t) dt \\ &= -\int_0^\infty t \int_{-L}^L \frac{\partial p(x', t|x, 0)}{\partial t} dx' dt \\ &= -\int_{-L}^L \int_0^\infty t \frac{\partial p(x', t|x, 0)}{\partial t} dt dx', \end{aligned} \quad (\text{A.5})$$

we exchange the order of integration because the mean first passage time is a finite number and therefore the integrals are convergent. Integration by parts gives

$$T_1(x) = \int_{-L}^L \int_0^\infty p(x', t|x, 0) dt dx' = \int_{-L}^L p_1 dx' \quad \text{where } p_1 = \int_0^\infty p(x', t|x, 0) dt \quad (\text{A.6})$$

since $\lim_{t \rightarrow 0} tp(x', t|x, 0) = 0$ and $\lim_{t \rightarrow \infty} tp(x', t|x, 0) = 0$. The boundary conditions for $T_1(x)$ are

$$T_1(-L) = T_1(L) = 0 \quad (\text{A.7})$$

since a particle starting at an endpoint would exit at $t = 0$, a result of the absorbing boundary conditions of the segment. By integrating Equation (A.3) over all time, and noting $p(x', \infty|x, 0) = 0$, we can derive a differential equation for p_1 ,

$$\int_0^\infty \partial_t p(x', t|x, 0) dt = \int_0^\infty (\partial_x(V'(x)p(x', t|x, 0)) + D\partial_x^2 p(x', t|x, 0)) dt.$$

Since $p(x', t|x, 0)$ is C^1 , we can exchange the order of integration and differentiation,

$$-p(x', 0|x, 0) = \partial_x(V'(x)p_1) + D\partial_x^2 p_1. \quad (\text{A.8})$$

Which corresponds to the equation

$$-\delta(x - x') = \mathcal{L}_x p_1 \quad (\text{A.9})$$

We can express Equation (A.8) as a differential equation for T_1 by using the adjoint operator \mathcal{L}_x^\dagger of \mathcal{L}_x . This will allow us to solve for $T_1(x)$.

The adjoint operator, \mathcal{L}^\dagger , is defined by the property

$$\int_{-L}^L u(\mathcal{L}v) dx = \int_{-L}^L (\mathcal{L}^\dagger u)v dx \quad (\text{A.10})$$

for any functions $u(x)$ and $v(x)$ which are zero at the boundaries. First, we apply the inverse operator to Equation (A.9) to get

$$p_1 = -\mathcal{L}_x^{-1}\delta(x - x').$$

By plugging this into (A.6), we get

$$T_1(x) = -\int_{-L}^L \mathcal{L}_x^{-1}\delta(x - x') dx' \quad (\text{A.11})$$

This is the adjoint property in Equation (A.10) when $u(x) = 1$ and $v(x) = \delta(x - x')$.

Therefore,

$$T_1(x) = -\int_{-L}^L (\mathcal{L}_x^{-\dagger}1)\delta(x - x') dx' = -\mathcal{L}_x^{-\dagger}(1) \quad (\text{A.12})$$

and $T_1(x)$ is a solution to the differential equation $\mathcal{L}_x^\dagger T_1(x) = -1$.

To calculate the expression for the adjoint operator, we use Equation (A.10) and integrate by parts twice,

$$\begin{aligned} \int_{-L}^L u(\mathcal{L}_x v) dx &= \int_{-L}^L u(\partial_x(V'(x)v) + D\partial_x^2 v) dx \\ &= -\int_{-L}^L (\partial_x(u)V'(x)v + D\partial_x(u)\partial_x(v)) dx \\ &= \int_{-L}^L (-V'(x)\partial_x(u)v + D\partial_x^2(u)v) dx \\ &= \int_{-L}^L (\mathcal{L}_x^\dagger u)v dx \end{aligned}$$

Hence, the adjoint operator is

$$\mathcal{L}_x^\dagger = -V'(x)\partial_x + D\partial_x^2. \quad (\text{A.13})$$

Employing Equations (A.12) and (A.13) the expression for the differential equation of the mean first passage time as

$$\mathcal{L}_x^\dagger T_1(x) = -V'(x)\partial_x T_1(x) + D\partial_x^2 T_1(x) = -1. \quad (\text{A.14})$$

We redefine $T_1(x) = T(x)$ for simplicity.

Now we need to solve

$$T'' + \frac{-V'(x)}{D}T' = -\frac{1}{D}. \quad (\text{A.15})$$

Let the integrating factor $\psi(x) = e^{\frac{-V(x)}{D}}$, so $\psi'(x) = -\frac{V'(x)}{D}e^{\frac{-V(x)}{D}}$. This implies $\psi'(x)/\psi(x) = \frac{-V'(x)}{D}$. Therefore Equation (A.15) becomes

$$T'' + \frac{\psi'(x)}{\psi(x)}T' = -\frac{1}{D}. \quad (\text{A.16})$$

Multiplying through by $\psi(x)$, we have

$$\psi(x)T'' + \psi'(x)T' = \frac{d}{dx}(\psi(x)T') = -\frac{\psi(x)}{D} \quad (\text{A.17})$$

Integrate both sides of Equation (A.17)

$$\begin{aligned} \int_{-L}^x \frac{d}{dx}(\psi(y)T') dy &= \int_{-L}^x -\frac{\psi(y)}{D} dy \\ \psi(x)T'(x) - \psi(-L)T'(-L) &= -\frac{1}{D} \int_{-L}^x \psi(y) dy \\ T'(x) &= \frac{\psi(-L)T'(-L)}{\psi(x)} - \frac{-\frac{1}{D} \int_{-L}^x \psi(y) dy}{\psi(x)} \end{aligned}$$

Integrating one more time gives,

$$T(x) - T(-L) = \psi(-L)T'(-L) \int_{-L}^x \frac{1}{\psi(y)} dy - \frac{1}{D} \int_{-L}^x \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy \quad (\text{A.18})$$

where $T(-L) = 0$ by Equation (A.7). We can use Equation (A.7) and Equation (A.18) to solve for $T'(-L)$ as well. Substituting $x = L$ into Equation (A.18) we have

$$T(L) = 0 = \psi(-L)T'(-L) \int_{-L}^L \frac{1}{\psi(y)} dy - \frac{1}{D} \int_{-L}^L \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy \quad (\text{A.19})$$

Therefore

$$T'(-L) = \frac{\frac{1}{D} \int_{-L}^L \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy}{\psi(-L) \int_{-L}^L \frac{1}{\psi(y)} dy}.$$

Plugging this into Equation (A.18), we have

$$\begin{aligned}
T(x) &= \psi(-L) \frac{\frac{1}{D} \int_{-L}^L \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy}{\psi(-L) \int_{-L}^L \frac{1}{\psi(y)} dy} \int_{-L}^x \frac{1}{\psi(y)} dy - \frac{1}{D} \int_{-L}^x \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy \\
&= \frac{1}{D} \left[\frac{\left(\int_{-L}^x \frac{1}{\psi(y)} dy \right) \int_{-L}^L \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy - \left(\int_{-L}^L \frac{1}{\psi(y)} dy \right) \int_{-L}^x \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy}{\int_{-L}^L \frac{1}{\psi(y)} dy} \right] \\
&= \frac{1}{D} \left[\frac{\left(\int_{-L}^x \frac{1}{\psi(y)} dy \right) \left(\int_{-L}^x \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy + \int_x^L \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy \right)}{\int_{-L}^L \frac{1}{\psi(y)} dy} \right. \\
&\quad \left. - \frac{\left(\int_{-L}^x \frac{1}{\psi(y)} dy + \int_x^L \frac{1}{\psi(y)} dy \right) \int_{-L}^x \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy}{\int_{-L}^L \frac{1}{\psi(y)} dy} \right] \\
&= \frac{1}{D} \left[\frac{\left(\int_{-L}^x \frac{1}{\psi(y)} dy \right) \int_x^L \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy - \left(\int_x^L \frac{1}{\psi(y)} dy \right) \int_{-L}^x \int_{-L}^y \frac{\psi(z)}{\psi(y)} dz dy}{\int_{-L}^L \frac{1}{\psi(y)} dy} \right]
\end{aligned}$$

Letting $\Phi(x) = e^{\frac{V(x)}{D}} = \frac{1}{\psi(x)}$, we arrive at Equation (4.21)

$$T_1(x) = \frac{1}{D} \left[\frac{\left(\int_{-L}^x \Phi(y) dy \right) \int_x^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \left(\int_x^L \Phi(y) dy \right) \int_{-L}^x \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy}{\int_{-L}^L \Phi(y) dy} \right]. \tag{A.20}$$

Mean Escape Time

Now that we know the formula for the mean first passage time we can use it to calculate the mean escape time (MST). $T_1(x)$ the mean first passage time, is the average amount of time it takes to escape from the segment given that the particle was located at position x at time $t = 0$. Aside from the particle's first escape, each subsequent segment will have the particle starting at position $x = 0$ of that segment. Position $x = 0$ is where a particle is inserted into the next segment after it reaches the

boundary of the segment that the particle was in previously. Given that the particle will always start at $x = 0$ in every subsequent segment, we compute the MST as $T_1(0)$, the mean first passage time given the starting position of the particle is $x = 0$.

Plugging $x = 0$ into $T_1(x)$, we compute

$$\mu = T_1(0) = \frac{\left(\int_{-L}^0 \Phi(y) dy\right) \int_0^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \left(\int_0^L \Phi(y) dy\right) \int_{-L}^0 \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy}{D \int_{-L}^L \Phi(y) dy} \quad (\text{A.21})$$

Recall that $\Phi(x) = e^{V(x)/D}$, where $V(x) = U(x) - Fx$ and $U(x)$ is L periodic. Let $y \in [-L, 0]$ and $x \in [0, L]$ with the relationship

$$x = y + L. \quad (\text{A.22})$$

Then

$$\begin{aligned} V(x) &= V(y + L) \\ &= U(y + L) - Fy - FL \\ &= U(y) - Fy - FL \\ &= V(y) - FL. \end{aligned}$$

Hence,

$$V(y) = V(x) + FL \quad (\text{A.23})$$

Using this relationship, we have

$$\begin{aligned}
\int_{-L}^0 e^{\frac{V(y)}{D}} dy &= \int_0^L e^{\frac{V(x)+FL}{D}} dx \\
&= \int_0^L e^{\frac{V(x)}{D}} e^{\frac{FL}{D}} dx \\
&= e^{\frac{FL}{D}} \int_0^L e^{\frac{V(x)}{D}} dx \\
&= \frac{1}{e^{\frac{-FL}{D}}} \int_0^L e^{\frac{V(x)}{D}} dx \\
&= \frac{1}{e^{\frac{-FL}{D}}} \int_0^L e^{\frac{V(y)}{D}} dy
\end{aligned}$$

Employing this in Equation A.21 gives,

$$\begin{aligned}
\mu &= \frac{1}{D} \frac{\left(\frac{1}{e^{\frac{-FL}{D}}} \int_0^L \Phi(y) dy \right) \int_0^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \left(\int_0^L \Phi(y) dy \right) \int_{-L}^0 \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy}{\int_{-L}^L \Phi(y) dy} \\
&= \frac{1}{D} \left(\frac{\int_0^L \Phi(y) dy}{\int_{-L}^L \Phi(y) dy} \right) \left(\frac{1}{e^{\frac{-FL}{D}}} \int_0^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \int_{-L}^0 \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy \right)
\end{aligned}$$

Note:

$$\begin{aligned}
\frac{\int_0^L \Phi(y) dy}{\int_{-L}^L \Phi(y) dy} &= \frac{\int_0^L \Phi(y) dy}{\int_{-L}^0 \Phi(y) dy + \int_0^L \Phi(y) dy} \\
&= \frac{\int_0^L \Phi(y) dy}{\frac{1}{e^{\frac{-FL}{D}}} \int_0^L \Phi(y) dy + \int_0^L \Phi(y) dy} \\
&= \frac{1}{1 + \frac{1}{e^{\frac{-FL}{D}}}} \\
&= \frac{e^{\frac{-FL}{D}}}{1 + e^{\frac{-FL}{D}}}
\end{aligned}$$

Plugging this back into μ , we have,

$$\mu = \frac{e^{-\frac{FL}{D}}}{D \left(1 + e^{-\frac{FL}{D}}\right)} \left(\frac{1}{e^{-\frac{FL}{D}}} \int_0^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \int_{-L}^0 \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy \right). \quad (\text{A.24})$$

Employing our change of variables in Equation (A.22) to the second double integral in Equation (A.24),

$$\int_{-L}^0 \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy = \frac{1}{e^{-\frac{FL}{D}}} \int_0^L \int_{-L}^{x-L} \frac{\Phi(x)}{\Phi(z)} dz dx = \frac{1}{e^{-\frac{FL}{D}}} \int_0^L \int_{-L}^{y-L} \frac{\Phi(y)}{\Phi(z)} dz dy.$$

Going back to Equation (A.24) to substitute this expression, we have

$$\begin{aligned} \mu &= \frac{e^{-\frac{FL}{D}}}{D \left(1 + e^{-\frac{FL}{D}}\right)} \left(\frac{1}{e^{-\frac{FL}{D}}} \int_0^L \int_{-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy - \frac{1}{e^{-\frac{FL}{D}}} \int_0^L \int_{-L}^{y-L} \frac{\Phi(y)}{\Phi(z)} dz dy \right) \\ &= \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{y-L}^y \frac{\Phi(y)}{\Phi(z)} dz dy \end{aligned}$$

Therefore, we have derived Equation (4.19),

$$\mu = \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x \frac{\Phi(x)}{\Phi(y)} dy dx. \quad (\text{A.25})$$

Variance of the Escape Time

The derivation for the escape time variance follows extremely closely to the derivation for both the mean first passage time and the MST, detailed in the above sections. For this reason, several of the details will be left out of this derivation. To calculate the escape time variance we will begin with Equation (4.20) and rewrite it

to a different form.

$$\begin{aligned}
\sigma^2 &= \frac{2}{D \left(1 + e^{-\frac{FL}{D}}\right)} \left(\int_0^L \int_{x-L}^x \frac{e^{\frac{V(x)}}{D}}}{e^{\frac{V(y)}}{D}} T_1(y) dy dx - \int_0^L \int_{x-L}^x \frac{e^{\frac{V(x)}}{D}}}{e^{\frac{V(y)}}{D}} \frac{1}{2} \mu dy dx \right) \\
&= \frac{2}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{V(x)}{D}} e^{-\frac{V(y)}{D}} T_1(y) dy dx \\
&\quad - \mu \frac{1}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{V(x)}{D}} e^{-\frac{V(y)}{D}} dy dx
\end{aligned}$$

Using Equation (A.25), we have

$$\sigma^2 = \frac{2}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{V(x)}{D}} e^{-\frac{V(y)}{D}} T_1(y) dx dy - \mu^2. \quad (\text{A.26})$$

This is the equation that we will derive.

The equation for the variance in terms of the mean is

$$\sigma^2 = \langle T^2 \rangle - \mu^2. \quad (\text{A.27})$$

So we need to show

$$\langle T^2 \rangle = T_2(0) = \frac{2}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x e^{\frac{V(x)}{D}} e^{-\frac{V(y)}{D}} T_1(y) dx dy \quad (\text{A.28})$$

where

$$T_2(x) = \int_0^\infty t^2 g(x, t) dt \quad (\text{A.29})$$

Using Equation (A.4),

$$T_2(x) = - \int_0^\infty t^2 \partial_t G(x, t) dt$$

which following integration by parts, we have

$$T_2(x) = 2 \int_0^\infty tG(x, t) dt \quad (\text{A.30})$$

since $\lim_{t \rightarrow 0} t^2 G(x, t) = 0$ and $\lim_{t \rightarrow \infty} t^2 G(x, t) = 0$. Since $G(-L, t) = G(L, t) = 0$ we have $T_2(-L) = T_2(L) = 0$. Using $G(x, t)$ in the FPE with the adjoint operator, we have the formula

$$\partial_t G(x, t) = -V'(x) \partial_x G(x, t) + D \partial_x^2 G(x, t).$$

Multiplying by t and integrating over all time, gives us

$$\int_0^\infty t \partial_t G(x, t) dt = -V'(x) \partial_x \int_0^\infty t G(x, t) dt + D \partial_x^2 \int_0^\infty t G(x, t) dt.$$

From Equation (A.5) we know $\int_0^\infty t \partial_t G(x, t) dt = -T_1(x)$ so our formula becomes

$$\begin{aligned} -T_1(x) &= -V'(x) \partial_x \frac{1}{2} T_2(x) + D \partial_x^2 \frac{1}{2} T_2(x) \\ -2T_1(x) &= -V'(x) \partial_x T_2(x) + D \partial_x^2 T_2(x) \end{aligned}$$

Now that we have a differential equation for $T_2(x)$, the derivation follows exactly as for $T_1(x)$ beginning from Equation (A.15). Omitting these details we arrive at the formula

$$T_2(x) = \frac{2}{D} \left[\frac{\left(\int_{-L}^x \Phi(y) dy \right) \int_x^L \int_{-L}^y \frac{\Phi(y) T_1(z)}{\Phi(z)} dz dy}{\int_{-L}^L \Phi(y) dy} - \frac{\left(\int_x^L \Phi(y) dy \right) \int_{-L}^x \int_{-L}^y \frac{\Phi(y) T_1(z)}{\Phi(z)} dz dy}{\int_{-L}^L \Phi(y) dy} \right]. \quad (\text{A.31})$$

From here we can plug in $x = 0$ and follow exactly the derivation for μ detailed above.

We calculate

$$\langle T^2 \rangle = \frac{2}{D \left(1 + e^{-\frac{FL}{D}}\right)} \int_0^L \int_{x-L}^x \frac{\Phi(x)T_1(y)}{\Phi(y)} dy dx \quad (\text{A.32})$$

which is what we wanted to show in Equation (A.28). Therefore we have derived the expression for σ^2 given by Equation (4.20).

APPENDIX B

DERIVATION FOR THE PROBABILITIES OF ESCAPE

Here we derive the probability that a particle escapes a potential segment, see Figure 4.2, through the right boundary L , or the left boundary $-L$. We denote the probability of escaping through L when a particle begins at location x at time $t = 0$, $p_+(x)$, and the probability of escaping left as $p_-(x)$.

Consider the probability current $J(x, t)$ defined using the FPE as

$$\partial_t p(x, t) + \partial_x J(x, t) = \partial_t p(x, t) + \partial_x [-(V'(x)p(x', t|x, 0) + D\partial_x p(x', t|x, 0))] = 0.$$

The total probability that the particle exits through $x = L$ after time t is given by the time integral of the probability current at $x = L$, $J(L, t|x, 0)$, given that the particle is located at x at time $t = 0$. This quantity is

$$g_L(x, t) = \int_t^\infty J(L, t'|x, 0) dt'.$$

Therefore $p_+(x) = g_L(x, 0)$. Since $J(L, t'|x, 0)$ satisfies the backward FPE (explained in Appendix A), using the adjoint operator, we have

$$\begin{aligned} -V'(x)\partial_x g_L(x, t) + D\partial_x^2 g_L(x, t) &= \int_t^\infty \partial_{t'} J(L, t'|x, 0) dt' \\ &= -J(L, t|x, 0) \\ &= -\partial_t g_L(x, t). \end{aligned} \tag{B.1}$$

In order to derive a formula for $p_+(x)$ we would like to use Equation (B.1) to derive an ordinary differential equation for $p_+(x)$. Hence we take $t \rightarrow 0$ on both sides of Equation (B.1). Please notice that $J(L, 0|x, 0) = 0$ for $x \neq L$. Therefore in the limit

we calculate

$$\begin{aligned} \lim_{t \rightarrow 0} (-V'(x) \partial_x g_L(x, t) + D \partial_x^2 g_L(x, t)) &= \lim_{t \rightarrow 0} (-\partial_t g_L(x, t)) \\ -V'(x) \partial_x p_+(x) + D \partial_x^2 p_+(x) &= -J(L, 0|x, 0) \\ &= 0. \end{aligned} \tag{B.2}$$

The boundary value problem describing $p_+(x)$ becomes

$$p_+''(x) - \frac{V'(x)}{D} p_+'(x) = 0 \tag{B.3}$$

with the boundary conditions

$$p_+(L) = 1, \quad p_+(-L) = 0. \tag{B.4}$$

Let $\psi(x) = e^{-\frac{V(x)}{D}}$ which gives the derivative $\psi'(x) = \frac{-V'(x)}{D} e^{-\frac{V(x)}{D}} = \frac{-V'(x)}{D} \psi(x)$.

Multiplying Equation (B.3) on both sides by $\psi(x)$ we compute

$$\begin{aligned} \psi(x) p_+''(x) + \psi'(x) p_+'(x) &= 0 \\ \partial_x [\psi(x) p_+'(x)] &= 0. \end{aligned}$$

Integrating both sides leads to

$$\begin{aligned} \int_x^L \partial_z [\psi(z) p_+'(z)] dz &= 0 \\ \psi(L) p_+'(L) - \psi(x) p_+'(x) &= 0 \end{aligned}$$

From here we are able to solve for $p'_+(x)$

$$p'_+(x) = \frac{\psi(L)p'_+(L)}{\psi(x)}. \quad (\text{B.5})$$

Integrating Equation (B.5) gives

$$\begin{aligned} \int_x^L p'_+(x') dx' &= \int_x^L \frac{\psi(L)p'_+(L)}{\psi(x')} dx' \\ p_+(L) - p_+(x) &= \psi(L)p'_+(L) \int_x^L \frac{1}{\psi(y)} dy. \end{aligned} \quad (\text{B.6})$$

We solve Equation (B.6) for $p_+(x)$

$$p_+(x) = 1 - \psi(L)p'_+(L) \int_x^L \frac{1}{\psi(y)} dy. \quad (\text{B.7})$$

Plugging in $x = -L$, we calculate

$$p_+(-L) = 0 = 1 - \psi(L)p'_+(L) \int_{-L}^L \frac{1}{\psi(y)} dy,$$

hence

$$\psi(L)p'_+(L) = \frac{1}{\int_{-L}^L \frac{1}{\psi(y)} dy}.$$

Using this in Equation (B.7),

$$p_+(x) = 1 - \frac{1}{\int_{-L}^L \frac{1}{\psi(y)} dy} \int_x^L \frac{1}{\psi(y)} dy.$$

Combining with a common denominator and simplifying

$$p_+(x) = \frac{\int_{-L}^x \frac{1}{\psi(y)} dy}{\int_{-L}^L \frac{1}{\psi(y)} dy}.$$

Finally, since $\pi_{-L}(x) = 1 - p_+(x)$ we compute the remaining probability

$$p_-(x) = \frac{\int_x^L \frac{1}{\psi(y)} dy}{\int_{-L}^L \frac{1}{\psi(y)} dy}.$$

APPENDIX C

FOURIER SERIES DERIVATIONS

The Fourier series for two periodic functions with period L on the interval $[0, L]$ were used as examples in Chapter 4. In this appendix, we include the details of the computation of the Fourier series for each function. Before providing these details, first we derive the Fourier series expansion on the interval $[0, L]$ given by Equation (4.70) with coefficients in Equations (4.71), (4.72), and (4.73).

The most common interval for a Fourier expansion to be defined on is $[-L, L]$. This expansion is given by the following equation

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(\frac{n\pi x}{L}\right) + \sum_{n=1}^{\infty} b_n \sin\left(\frac{n\pi x}{L}\right) \quad (\text{C.1})$$

with coefficients

$$a_0 = \frac{1}{L} \int_{-L}^L f(x) dx, \quad (\text{C.2})$$

$$a_n = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{n\pi x}{L}\right) dx, \quad (\text{C.3})$$

and

$$b_n = \frac{1}{L} \int_{-L}^L f(x) \sin\left(\frac{n\pi x}{L}\right) dx. \quad (\text{C.4})$$

To change this expansion on the interval $[-L, L]$ to the Fourier expansion on the interval $[0, L]$, apply a change of variables $x' = \frac{x+L}{2}$ where $x \in [-L, L]$ and $x' \in [0, L]$.

Consider the basis functions $\cos\left(\frac{n\pi x}{L}\right)$ and $\sin\left(\frac{n\pi x}{L}\right)$ under this transformation,

$$\begin{aligned} \cos\left(\frac{n\pi x}{L}\right) &= \cos\left(\frac{n\pi(2x' - L)}{L}\right) \\ &= \cos\left(\frac{2n\pi x'}{L} - n\pi\right) \\ &= \cos\left(\frac{2n\pi x'}{L}\right) \cos(n\pi) + \sin\left(\frac{2n\pi x'}{L}\right) \sin(n\pi) \\ &= (-1)^n \cos\left(\frac{2n\pi x'}{L}\right) \end{aligned} \quad (\text{C.5})$$

and

$$\begin{aligned}
 \sin\left(\frac{n\pi x}{L}\right) &= \sin\left(\frac{n\pi(2x' - L)}{L}\right) \\
 &= \sin\left(\frac{2n\pi x'}{L} - n\pi\right) \\
 &= \sin\left(\frac{2n\pi x'}{L}\right)\cos(n\pi) - \cos\left(\frac{2n\pi x'}{L}\right)\sin(n\pi) \\
 &= (-1)^n \sin\left(\frac{2n\pi x'}{L}\right).
 \end{aligned} \tag{C.6}$$

Using this in Equation (C.1), the Fourier expansion on $[0, L]$ is

$$f(x') = \frac{a'_0}{2} + \sum_{n=1}^{\infty} (-1)^n a'_n \cos\left(\frac{2n\pi x'}{L}\right) + \sum_{n=1}^{\infty} (-1)^n b'_n \sin\left(\frac{2n\pi x'}{L}\right). \tag{C.7}$$

Now the coefficients are calculated under the change of variables

$$a'_0 = \frac{2}{L} \int_0^L f(x') dx', \tag{C.8}$$

$$a'_n = \frac{2(-1)^n}{L} \int_0^L f(x') \cos\left(\frac{2n\pi x'}{L}\right) dx', \tag{C.9}$$

and

$$b'_n = \frac{2(-1)^n}{L} \int_0^L f(x') \sin\left(\frac{2n\pi x'}{L}\right) dx'. \tag{C.10}$$

Substituting these coefficients into Equation (C.7), the $(-1)^n$ terms in the sum and the coefficients cancel giving the expansion and coefficients originally presented in Chapter 4 as Equations (4.70), (4.71), (4.72), and (4.73). The remainder of this appendix is used to calculate the specific Fourier series expansion for the quadratic wave and square wave as defined in Chapter 4.

Quadratic Wave

We derive the Fourier series of the function

$$f(x) = -\left(x - \frac{L}{2}\right)^2.$$

We begin with a_0 using Equation (4.71),

$$\begin{aligned} a_0 &= \frac{2}{L} \int_0^L -\left(x - \frac{L}{2}\right)^2 dx \\ &= \frac{2}{L} \left[\frac{-(x - \frac{L}{2})^3}{3} \Big|_0^L \right] \\ &= \frac{2}{L} \left[\frac{-(\frac{L}{2})^3}{3} - \frac{-(-\frac{L}{2})^3}{3} \right] \\ &= \frac{-2}{L} \left(\frac{L^3}{12} \right) \\ &= -\frac{L^2}{6}. \end{aligned}$$

Using Equation (4.72) we calculate a_n with several integration by parts,

$$a_n = \frac{2}{L} \int_0^L -\left(x - \frac{L}{2}\right)^2 \cos\left(\frac{2\pi nx}{L}\right) dx.$$

With the first integration by parts,

$$\begin{aligned} a_n &= \frac{-2}{L} \left[\left(x - \frac{L}{2}\right)^2 \frac{L}{2\pi n} \sin\left(\frac{2\pi nx}{L}\right) \Big|_0^L \right. \\ &\quad \left. - \frac{L}{\pi n} \int_0^L \left(x - \frac{L}{2}\right) \sin\left(\frac{2\pi nx}{L}\right) dx \right]. \end{aligned}$$

The first term is zero, and we integrate the second term by parts again,

$$a_n = \frac{2}{\pi n} \left[\left(x - \frac{L}{2} \right) \frac{-L}{2\pi n} \cos \left(\frac{2\pi n x}{L} \right) \Big|_0^L + \frac{L}{2\pi n} \int_0^L \cos \left(\frac{2\pi n x}{L} \right) dx \right].$$

The integral in the second term is zero. Evaluating the first term,

$$a_n = \frac{-L}{(\pi n)^2} \left[\frac{L}{2} + \frac{L}{2} \right] = - \left(\frac{L}{\pi n} \right)^2.$$

Finally we calculate b_n with Equation (4.73),

$$b_n = \frac{2}{L} \int_0^L - \left(x - \frac{L}{2} \right)^2 \sin \left(\frac{2\pi n x}{L} \right) dx.$$

The first integration by parts gives,

$$b_n = \frac{-2}{L} \left[\left(x - \frac{L}{2} \right)^2 \frac{-L}{2\pi n} \cos \left(\frac{2\pi n x}{L} \right) \Big|_0^L + \frac{L}{\pi n} \int_0^L \left(x - \frac{L}{2} \right) \cos \left(\frac{2\pi n x}{L} \right) dx \right].$$

Again, the first term is zero, and we integrate the second term by parts,

$$b_n = \frac{-2}{\pi n} \left[\left(x - \frac{L}{2} \right) \frac{L}{2\pi n} \sin \left(\frac{2\pi n x}{L} \right) \Big|_0^L - \frac{L}{2\pi n} \int_0^L \sin \left(\frac{2\pi n x}{L} \right) dx \right].$$

The first term is zero. Evaluating the second integral,

$$\begin{aligned} b_n &= \frac{-L}{(\pi n)^2} \left[\frac{L}{2\pi n} \cos\left(\frac{2\pi n x}{L}\right) \Big|_0^L \right] \\ &= 0. \end{aligned}$$

With these coefficients, using Equation (4.70), we have the Fourier series given by Equation (4.90),

$$f(x) = \frac{-L^2}{12} - \sum_{n=1}^{\infty} \left(\frac{L}{\pi n}\right)^2 \cos\left(\frac{2\pi n x}{L}\right). \quad (\text{C.11})$$

Square Wave

Now we will derive the Fourier series for the square wave given by Equation (4.97),

$$f(x) = 2 \left[H\left(\frac{2x}{L}\right) - H\left(\frac{2x}{L} - 1\right) \right] - 1.$$

As before, we begin with a_0 using Equation (4.71),

$$\begin{aligned} a_0 &= \frac{2}{L} \int_0^L \left(2 \left[H\left(\frac{2x}{L}\right) - H\left(\frac{2x}{L} - 1\right) \right] - 1 \right) dx \\ &= \frac{2}{L} \left[\int_0^{\frac{L}{2}} 1 dx - \int_{\frac{L}{2}}^L 1 dx \right]. \end{aligned}$$

Evaluating these integrals,

$$a_0 = \frac{2}{L} \left[\frac{L}{2} - \frac{L}{2} \right] = 0.$$

Using Equation (4.72), we calculated a_n ,

$$\begin{aligned} a_n &= \frac{2}{L} \left[\int_0^{\frac{L}{2}} \cos\left(\frac{2\pi nx}{L}\right) dx - \int_{\frac{L}{2}}^L \cos\left(\frac{2\pi nx}{L}\right) dx \right] \\ &= \frac{2}{L} \left[\frac{L}{2\pi n} \sin\left(\frac{2\pi nx}{L}\right) \Big|_0^{\frac{L}{2}} - \frac{L}{2\pi n} \sin\left(\frac{2\pi nx}{L}\right) \Big|_{\frac{L}{2}}^L \right] \\ &= 0. \end{aligned}$$

Finally, we calculate the last coefficient b_n using Equation (4.73),

$$\begin{aligned} b_n &= \frac{2}{L} \left[\int_0^{\frac{L}{2}} \sin\left(\frac{2\pi nx}{L}\right) dx - \int_{\frac{L}{2}}^L \sin\left(\frac{2\pi nx}{L}\right) dx \right] \\ &= \frac{2}{L} \left[\frac{-L}{2\pi n} \cos\left(\frac{2\pi nx}{L}\right) \Big|_0^{\frac{L}{2}} + \frac{L}{2\pi n} \cos\left(\frac{2\pi nx}{L}\right) \Big|_{\frac{L}{2}}^L \right] \\ &= \frac{2}{\pi n} [1 - \cos(\pi n)]. \end{aligned}$$

Therefore

$$b_n = \begin{cases} 0 & n \text{ even} \\ \frac{4}{\pi n} & n \text{ odd} \end{cases}.$$

Putting together these coefficients, we have the Fourier series given in Equation (4.101),

$$f(x) = \sum_{n=1}^{\infty} \frac{4}{\pi(2n-1)} \sin\left(\frac{2\pi(2n-1)x}{L}\right). \quad (\text{C.12})$$