Rare Events and the Thermodynamic Action

A dissertation submitted to the Graduate School
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Author:
Patrick Malsom
B.S. Western Michigan University

Committee Chair:
Frank Pinski, Ph.D.

Department of Physics
College of Arts and Sciences
University of Cincinnati
Cincinnati, Ohio 45221, USA

October 8, 2015

Abstract

The purpose of this work is to develop improved mathematical methods to understand the dynamics of a system of atoms as it undergoes a conformational change. When the energy barrier separating the two stable configurations is large compared to the systems thermal energy, transitions rarely occur. Many of the standard methods, such as molecular dynamics, fail to sample such rare transitions efficiently, as the majority of the computational effort is exerted when the system is in one or the other stable configurations.

In order to create a more efficient approach, I consider a complete sequence of configurations to define a path, and constrain the path to start and end at predesignated states after a fixed time. The distribution of these double-ended paths are described by a functional that is based on Brownian dynamics. In this thesis, I show that a previously-used functional, which was derived in the continuous-time limit, is inconsistent with the equilibrium Boltzmann distribution. I then derive a series of accurate expressions for the thermodynamic distribution of double-ended paths. I also create a novel approach to efficiently sample paths described by these expressions and implement them for a one-dimensional model. As a result, I show how to produce an ensemble of thermodynamically consistent paths in an efficient manner.

Acknowledgments

I wish to express my deepest gratitude to my mentor and advisor, Professor Frank Pinski, for his guidance through this research project. Without his patience and encouragement, this work would not have been possible. I would also like to thank Professors Thomas Beck, Howard Jackson and Randy Johnson for providing guidance and carefully reviewing this work.

Special thanks should be given to Lincoln Bryant for his troubleshooting of technical problems and his thoughtful ear throughout my graduate career. My thanks are also extended to Professor Richard Gass and John Whitaker for their technical expertise and computational resources, and to John Markus for his careful revision of many manuscripts and presentations. The research shown in Section 6.2 was done using resources provided by the Open Science Grid.

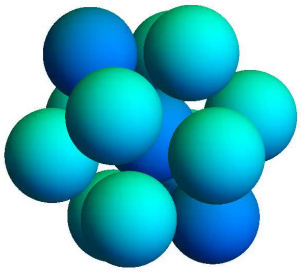
Finally, I wish to thank my wife Andrea, for her unwavering support and encouragement throughout my research endeavours.

Contents

List of Figures

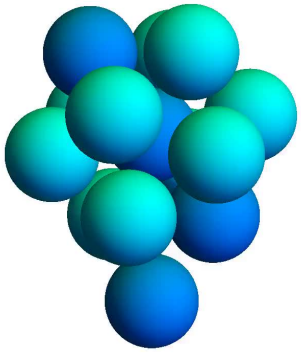
Animations

5.1 Motion of the LJ_{13} cluster at zero temperature corresponding to the energy landscape shown in Figure 5.5. This animation shows the symmetric breathing mode of the outside shell of particles, with the inner particles forming a chain.



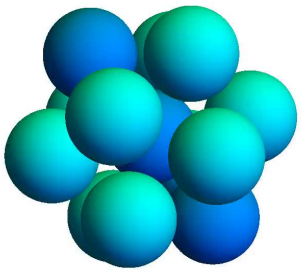
- `http://patrickmalsom.com/files/thesis/animation-5p1.webm`

5.2 Motion of the LJ_{14} cluster at zero temperature corresponding to the energy landscape in Figure 5.6. This transformation is perfectly symmetric between the initial and final configurations. The inner 4 particles form the chain and the outer shell symmetrically breathes to allow the chain particles to pass through the center of the cluster.



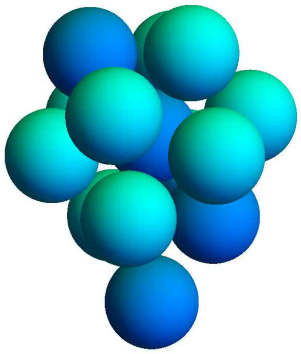
- `http://patrickmalsom.com/files/thesis/animation-5p2.webm`

5.3 Motion of the LJ_{13} cluster after many loops of the PSHMC algorithm at $\epsilon = 0.13$, corresponding to the energy landscape in 5.13. This motion clearly shows the broken instanton is healed, with the transition occurring very quickly. By observing, it is clear why this system is difficult to describe for the thermalized case.



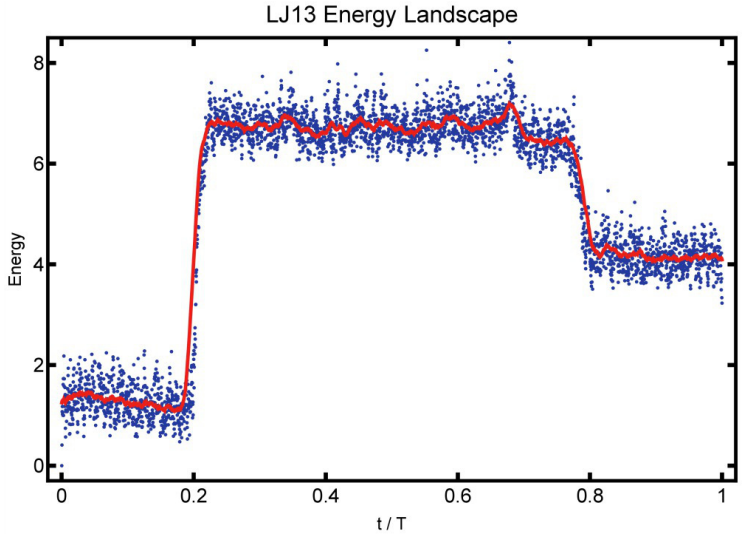
- `http://patrickmalsom.com/files/thesis/animation-5p3.webm`

5.4 Motion of the LJ_{14} cluster after many loops of the PSHMC algorithm at $\epsilon = 0.13$, corresponding to the energy landscape in 5.15. The symmetric mode is now destroyed and the transition occurs very quickly in the center region of the path.



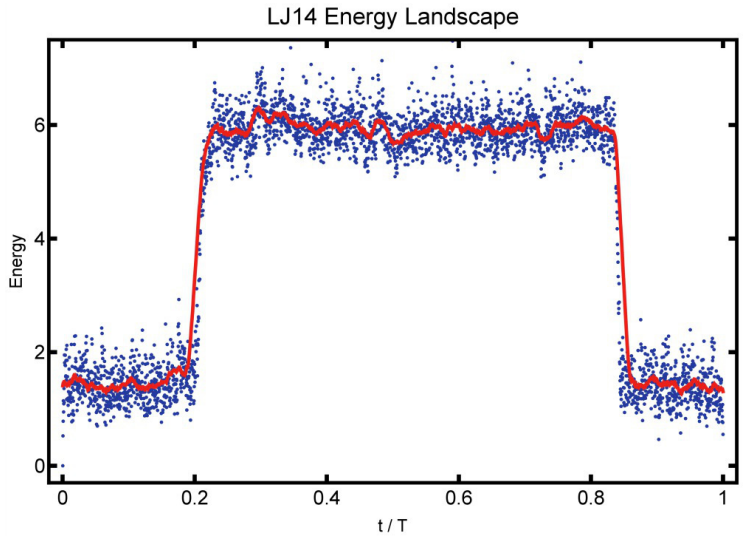
- http://patrickmalsom.com/files/thesis/animation-5p4.webm

5.5 Evolution of the energy landscape of the LJ_{13} as a function of the sampling index. The blue dots show the energy along the path and the red line serves as a guide to the eye. The initial landscape is simply the thermalized zero temperature configuration shown in Figure 5.9, and the ending configuration is shown in Figure 5.13.



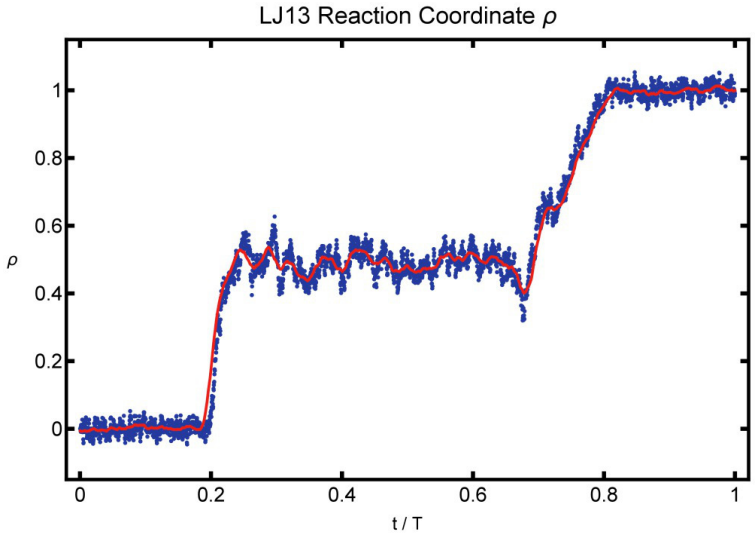
- http://patrickmalsom.com/files/thesis/animation-5p5.webm

5.6 Evolution of the energy landscape of the LJ_{14} as a function of the sampling index. The blue dots show the energy along the path and the red line serves as a guide to the eye. The initial landscape is simply the thermalized zero temperature configuration shown in Figure 5.10, and the ending configuration is shown in Figure 5.15.



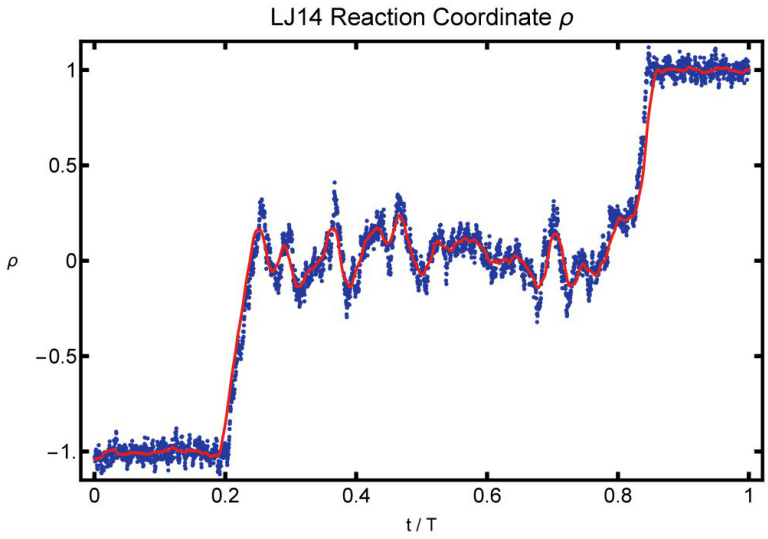
- http://patrickmalsom.com/files/thesis/animation-5p6.webm

5.7 Evolution of the reaction coordinate, ρ, for the LJ_{13} as a function of the sampling index. The blue dots show ρ along the path and the red line serves as a guide to the eye. The reaction coordinate for the initial path is shown in Figure 5.11 and the ending is shown in Figure 5.17.



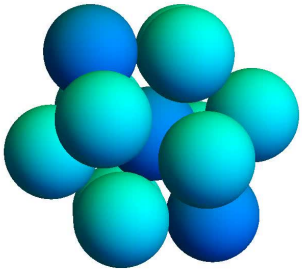
- http://patrickmalsom.com/files/thesis/animation-5p7.webm

5.8 Evolution of the reaction coordinate, ρ, for the LJ_{14} as a function of the sampling index. The blue dots show ρ along the path and the red line serves as a guide to the eye. The reaction coordinate for the initial path is shown in Figure 5.12 and the ending is shown in Figure 5.18.



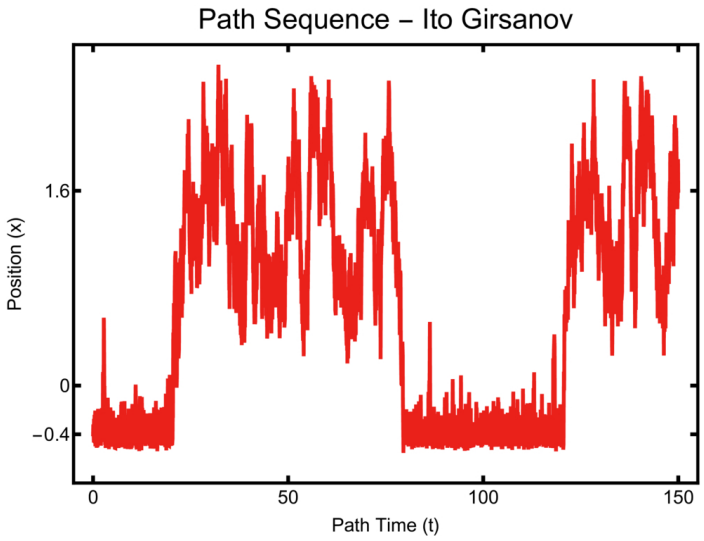
• http://patrickmalsom.com/files/thesis/animation-5p8.webm

5.9 The motion of the LJ_{13} cluster after quenching the fluctuations from the ending path, whose energy landscape and reaction coordinate is shown in Figures 5.19 and 5.20. The symmetry has now been broken in the outer shell particles, with a larger gap between some of the particles than its neighboring particles.



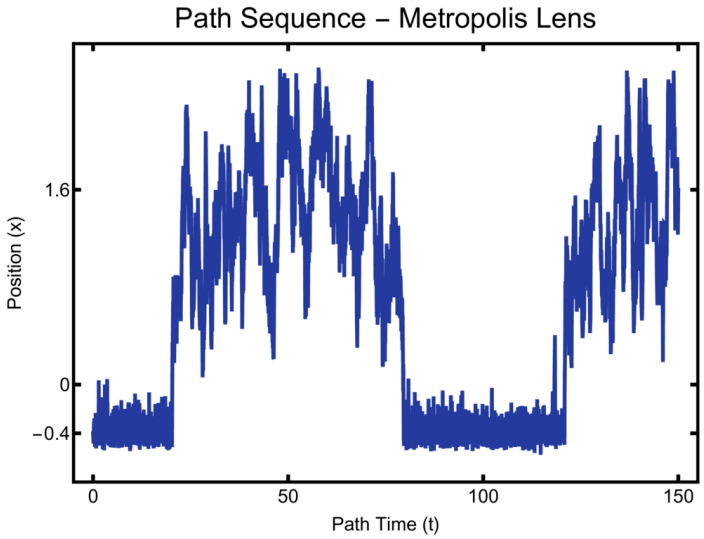
• http://patrickmalsom.com/files/thesis/animation-5p9.webm

6.1 The evolution of the sequence of the paths generated using the PSHMC path sampling method for the 1D potential from Equation 6.2. The sequence of paths are seen to quickly fall into the skinny well and never recover. This is the evidence that initiated the development of the novel HMC path sampling method based on the Metropolis algorithm.



- http://patrickmalsom.com/files/thesis/animation-6p1.webm

6.2 The evolution of the sequence of the paths generated using the novelHMC path sampling method for the 1D potential from Equation 6.2. Paths generated using this method are seen to be consistent with the underlying Boltzmann distribution.



- http://patrickmalsom.com/files/thesis/animation-6p2.webm

List of Symbols

x	position
t	time along path (Brownian)
h	time along path (Leap-Frog $h = \sqrt{2\Delta t}$)
τ	time between paths (Brownian)
η	time between paths (Leap-Frog $\eta = \sqrt{2\Delta\tau}$)
T	Lenth (time) of the path
$F(x)$	Force at position x
$V(x)$	Potential at position x
ρ	Reaction coordinate for LJ cluster
W_t	Wiener process (time dependent)
ϵ	configurational temperature
ξ	Gaussian distributed random number
J	Jacobian transformation
$\mathbb{J}$	Double Jacobian (symplecticness test)
$\Delta x_i = x_{i+1} - x_i$	positional change between time steps
$\bar{x}_i = (x_{i+1} + x_i)/2$	average position between time steps
v, p	velocity, momentum
$\mathbb{P}$	Probability
$\mathbb{P}_B$	Bolzmann probability
G	path potential
H	Hamiltonian
Λ	PSHMC log probability (Effective Hamiltonian)
$\boldsymbol{L}$	Laplacian operator
$\boldsymbol{L}^{-1}$	Inverse Laplacian operator
$\boldsymbol{M}$	Mass matrix
$\delta\Lambda$	change in PSHMC log probability
R	center of mass for LJ cluster
$B(s)$	broad well fraction for 1D system
δe	forward energy error
$\mathbb{I}$	Identity matrix

Notation

A'	Spacial derivative
$\dot{A}$	Time derivative

Chapter 1

Introduction

The motivation for this work lies in improving the tools available to study *rare events*. Such events, which occur extremely infrequently on average, many times dictate a system's most relevant and interesting dynamics. Everyday examples of such events include financial market crashes, earthquakes, or the failure of a mechanical part. In the natural sciences, the systems of interest are normally much smaller in scale, for example the study of activated chemical reactions, polymer dynamics or the configurational change of (relatively large) proteins in biological systems.

As the name "rare event" suggests, the dynamics driving these systems is dominated by the large chunks of time in which no abnormal events occur. The disparity in time scales is the main factor why experiment finds the study of these rare events so difficult. It is extremely inefficient to send a geologist to California in the hope that an earthquake will spontaneously happen in a short time period of time. In all likelihood, the scientist would give up and find a more time efficient method of measuring the event. In the case of a folding protein, the folding event from the random coil to the native folded state occurs on the order of microseconds to milliseconds, while the fluctuations of the individual atomic properties occur on the femtosecond scale.

In this thesis, I focus on collections of atoms fluctuating according to Brownian (over-damped Langevin) dynamics, which undergo a conformational change. In many cases, these changes are blocked by an energy barrier and the event occurs rarely when the thermal energy is small compared to the barrier height. The question now becomes: How can one efficiently study these events if they are so infrequently observed? Computer simulation is ideally suited to handle these problems. The resolution of these very small timescale events can be resolved by a computer easily, but the

computational effort involved in simulating the slower modes of transformation can be immense. Nevertheless, the simulation of these rare events is an extremely important problem to consider.

My interests lie in the development of an algorithm which can efficiently sample systems which exhibit such a rare event, rather than the study of a specific physical system. The familiarity of molecular systems to a physicist drives my interest to the field of molecular modeling, where the majority of rare events are often caused by an energy barrier which blocks the conformational change. These barriers exist in the Free Energy landscape which cannot be probed with straightforward techniques, such as gradient descent. The intent of developing an improved sampling method restricts this study to well-known systems, where solutions can be checked against results from other methods (or ideally, the exact answer). For this reason the systems of interest here are theoretical in nature.

The simulation of rare events has been studied extensively[1], and there are a great multitude of algorithms which aim to efficiently solve the problem. Standard Molecular Dynamics (MD) aims to use the laws of Newton (or similarly Lagrange and Hamilton) to calculate the forward evolution of a system under the influence of a set of forces[2]. Hamiltonian systems have been exhaustively studied[3] in the physics literature and well-behaved numerical methods exist that optimize the equations of motion, while leaving the conservation laws of Hamilton's equations intact. This method of simulation often leads to large collective motions of the system of study, due to the ballistic nature of MD. These advantages also lead to problems. The conservation laws create simulations which are restricted to a constant energy surface and the same volume of phase space. Escaping from an energy basin which is blocked by a barrier which is larger than the total energy of the system will not only be rare using this method, it will be impossible.

Stochastic simulation is an alternative which uses the underlying fluctuations as a driving parameter in the equations of motion. In the over-damped case, the motion mimics the Brownian nature of particles in a viscous medium. The randomly fluctuating nature often leads to a largely diffusive evolution, which can be extremely slow when a complex movement must be performed.

To help alleviate the problems with these techniques, various path sampling algorithms have been developed. These methods change the way in which classical sampling is performed by introducing a *path*, a sequence of atomic positions which is defined to traverse between two long-lived states which are known experimentally. Examples of these methods include Transition Path

Sampling[4, 5], Transition Interface Sampling[6] and Forward Flux Sampling[7, 8] as well as more specialized methods such as the string method[9, 10] and milestoning[11, 12].

In this work, I will focus on an alternative path sampling method which has roots in a hybridized version of MD and stochastic simulation, called the Path Space Hybrid Monte-Carlo (PSHMC) method[13]. This method differs from many of the above path sampling methods in that it makes the path the central focus instead of focusing on theoretical interfaces in a physical problem. I have worked on devising algorithms for generating ensembles of paths, and thus need to incorporate time changes in discrete steps, but have also studied the continuous-time limit of the formalism. This limit is mathematically challenging as infinities are introduced.

1.1 Outline of Thesis

The organization of this document is summarized as follows:

Chapter 2: Forward Integration

In this chapter, I will frame the basic underlying methods which will be utilized in the later chapters. I give a basic overview of some of the forward sampling methods focusing on stochastic simulation of a simple system, where a single particle moves diffusively via an externally imposed one-dimensional force field. These methods create what I will refer to as a *trajectory*, (which differs from the path which will be introduced later) which are a sequence of positions that evolve forward in time. I will then review the Metropolis-Hastings algorithm[14] and discuss the importance of the algorithm for these sampling methods. By using the 1D potential system, I will show why the Metropolis test is necessary for finite step size Stochastic simulation. Finally, I introduce a hybridized version of the basic stochastic sampling, called Hybrid (Hamiltonian) Monte Carlo[15], which introduces a Hamiltonian flow to the aforementioned Metropolis adjusted stochastic sampling.

Chapter 3: Path Sampling using Onsager-Machlup

This chapter serves as a bridge which joins the ideas of standard forward sampling methods from Chapter 2 and the path space methods of Chapters 4 and 7. The sampling problem is recast

using an object referred to as a *path*. This path differs from a trajectory in that it is defined for all time steps for a predetermined length of time T and the endpoint is predetermined. This path will be the fundamental object considered for the remainder of this thesis work.

In order to sample paths, I will employ the use of the Onsager-Machlup (OM) functional[16]. This result utilizes the underlying noise history of the path in order to calculate the probability of any path using the positions along the path itself.

Chapter 4: Path-Space Hybrid Monte-Carlo

I show how to obtain a continuous-time version of the OM functional, which can be interpreted as a *thermodynamic action*[17]. This new probability distribution takes an interesting form which resembles the probability of a free motion with an added *path potential*. I use the results of Ito calculus and the Girsanov theorem[18] to write the new path probability $\mathbb{P}_{IG}$, which is carefully constructed to avoid the infinities present in this continuous-time formulation. Using the method proposed in Reference [13], I develop an algorithm which can be used to efficiently sample this continuous-time path probability, called Path Space Hybrid Monte-Carlo (PSHMC).

Chapter 5: The Lennard-Jones System

A system composed of a cluster of 13 and 14 particles interacting via the Lennard-Jones cluster is studied using the PSHMC method proposed in Chapter 4. This example examines a low lying mode of conformational change of the cluster through a free energy barrier[19]. The PSHMC is shown to generate a sequence of paths which heal the unphysical initial state into a state which makes a quick transition over the barrier. Unfortunately, this complicated example ends raising more questions than it answers.

Chapter 6: Results in One Dimension

In order to gain a more thorough understanding of the probability measure from Chapter 4, a much simpler problem will be studied. The potential used in this chapter possess free energy basins which are degenerate in energy but disparate in entropy, with one well being thin and the other broad. The sequence of paths which are generated using the PSHMC method are shown to be driven to a non-physical regime and do not recover. Furthermore, using restricted sampling

procedure it is shown that the path probability should be equivalent in all cases at the same temperature, which is inconsistent with the PSHMC probability measure. With this evidence, it is clear that the Ito-Girsanov expression can not be interpreted as a path measure.

Chapter 7: Finite Method

In this chapter I will devise an alternative approach to the PSHMC method which uses the well understood Metropolis algorithm to elucidate any errors made to the numeric integration of Hamiltons's equations. This approach is specifically chosen to manage unknown faults in the derivation of the continuous-time formulation. This new HMC framework is constructed specifically to match the structure of the PSHMC method and an analog to the path potential, G, from Chapter 4 is derived. Importantly, the sequence of paths generated by this method is found to be consistent with the underlying (physical) Boltzmann distribution.

Chapter 8: Concluding Remarks

I conclude by presenting a set of arguments which question the validity of using the path-space probability $\mathbb{P}_{IG}$ as a way to sample physical paths. These arguments lie at the foundation of the interpretation of the Ito-Girsanov form of the path probability. I close with some future directions on where to continue this research.

1.2 Importance of this Work

The original goal of this work was to study rare transitions in molecular systems. To achieve this goal I used a method which was thought to be well understood. However, implementing the method led to surprising revelations.

I performed a thorough analysis of the Ito-Girsanov probability measure using the PSHMC method proposed by Beskos *et. al*[13]. Employing this method to sample paths generates sequences of paths which are unphysical. In order to understand why this method fails, I probed into the intricate details of the method.

To gain a more thorough understanding of the PSHMC algorithm, I developed an analogous path sampling method, which avoids the mathematical challenges of PSHMC, by reformulating

the analysis through the lens of the Metropolis algorithm. This novel framework generates paths that are consistent with the Boltzmann distribution. Furthermore, because all of the errors are fully exposed, this new perspective clearly shows limitations of the measure in which the PSHMC method is based. It is the use of the Ito-Girsanov expression as a path probability distribution which leads to the unphysical paths, as is discussed in Chapter 8.

While developing this path sampling method, I examined all assumptions made in the derivation. This illuminated some of the misunderstandings of the path probability. The analysis in Section 6.2 motivates the fact that all paths of equal length, have equal probability. This statement is incompatible with the Ito-Girsanov measure, which suggests some paths are more probable than others.

The result of greatest importance in this thesis is not necessarily the novel method of path sampling presented in detail in Chapter 7. The use of the Ito-Girsanov measure is engrained into the intuition of many of the experts in the field. This work uncovers a fundamental misunderstanding of this idea. This unphysical measure is used to calculate many of the results in the literature, many of which date back 35 years[20, 21]. Fields which build upon these ideas include the theory of entropy production in non-equilibrium thermodynamics[22], other path sampling methods[23, 1] and even protein folding simulations[24, 25]. These articles must be reevaluated in light of the findings of this thesis.

Chapter 2

Forward Integration

In this chapter, I present a study of the movement of a system of interacting particles which are spontaneously fluctuating in a thermal bath as the system evolves forward in time. The thermal background interjects random fluctuations into the motion of the particles, as originally seen by Brown in his work on dust particles floating in water[26], and later described theoretically by Einstein[27]. One of the main questions to be addressed is how to formally describe this seemingly random process mathematically. These random fluctuations have been put on a rigorous mathematical foundation by Wiener, and a calculus developed by Ito is used to describe the time-evolution of the system[28].

In many physical processes, the most interesting phenomena are observed only when studying the dynamics of the system. For many molecular systems, the interesting dynamics happen when a system transitions from one free energy basin to another. Free energy landscapes are often much too complex fully understand, even for relatively simple systems such as a cluster of Leonard-Jones particles. The free energy landscape of the systems, considered in this study, possess an energy barrier which impedes the system from moving freely between the two or more basins. To further complicate matters, transitions between these two free energy basins become exponentially rare with decreasing thermal energy, as shown for an example case in Figure 2.1*.

Historically, the simulation proceeds as an integration of the positions going forward in time. This can be handled by integrating a Hamiltonian flow (molecular dynamics). It is often convenient to average over several (or many) degrees of freedom, thereby introducing stochastic effects.

*The details of this simulation will be explained in the last section of this chapter.

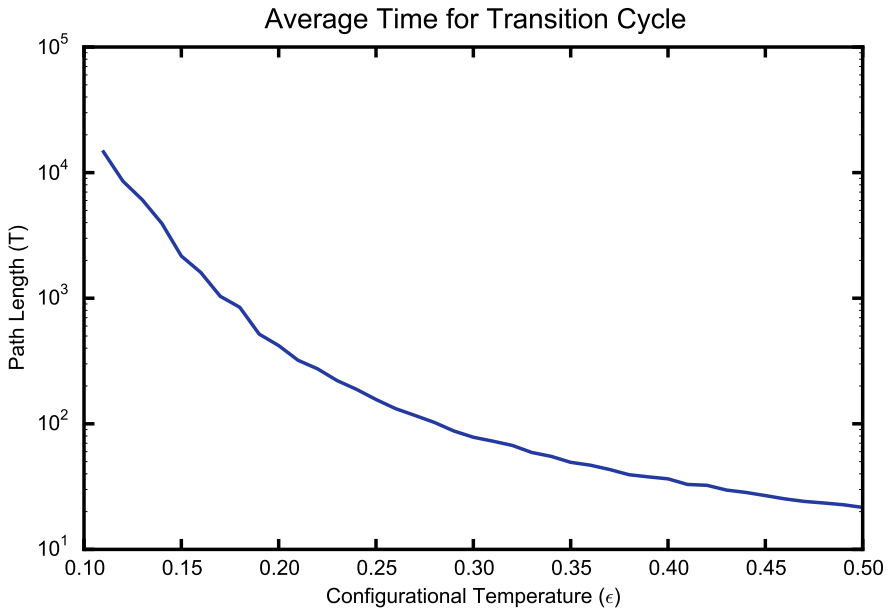


Figure 2.1: Estimated average length of trajectory required to observe a transition cycle as a function of temperature, ϵ. This length increases exponentially as the temperature decreases, leading to exponentially longer waiting times for the observation of a transition.

This generates a trajectory which starts at some initial state, given as an initial condition to the integrator, and end at some unknown position at a finite time. This process can be devised to correctly mimic physical nature of the time evolution, which includes very long waiting times when attempting to make rare transitions.

The goal of the work presented here is to develop an efficient method of sampling is rare transitions while preserving the physical nature of the stochastic system. This goal necessarily requires that the systems obey the correct underlying probability distribution. In the classical regime, this probability is given by the Boltzmann distribution. Thus, a method will be deemed successful only if the method of simulation produces trajectories which visit states consistent with the Boltzmann distribution.

In order to frame the following sections of this chapter, I will first need to explain some notation.

- $\epsilon \rightarrow$ temperature
- $x \rightarrow$ position
- $\Delta t \rightarrow$ time step
- $F(x) \rightarrow$ Force $(-dV(x)/dx = -V'(x))$

the forward evolution of these paths will start with an initial position, x_0, and using a constant time step, generate a proposed move to x_1. The proposed move is always accepted when using Brownian dynamics. However, as explained below, when using the Metropolis algorithm, the state may be rejected. The details of how to generate the updated position will depend on the specific method used.

2.1 Brownian Dynamics

Historically, the discussion of stochastic processes began with the problem of understanding the motion of molecules which exhibit Brownian motion. In this limit, all particles in the system are assumed to be at their terminal velocities in thermal equilibrium with the surrounding heat bath. The thermal bath imposes a randomly fluctuating element to the system that gives rise to the original observations of Brownian motion. This situation can be described using a damped driven equation of motion

$$\frac{dp}{d\tilde{t}} = F - \gamma\, p + \text{Noise} \tag{2.1}$$

The *Noise* is defined using the Wiener process, dW_t, which is scaled by the thermal factor set by the fluctuation-dissipation theorem. In the Brownian (over-damped) regime, the time change in momentum tends to zero ($dp/d\tilde{t} \to 0$). With the substitution of a new algorithmic time which is scaled by the mass and the frictional coefficient*, $\tilde{t} = \gamma m t$, the Langevin form of this equation of motion is readily seen

$$dx = F(x_t)dt + \sqrt{2\epsilon}\, dW_t \tag{2.2}$$

2.1.1 Euler-Maruyama method

This equation of motion is defined in the form of a continuous time process. In order to analyze this problem numerically, the equation of motion will need to be discretized. The Wiener process requires some special attention when performing this discretization. The sole condition on this process from mathematical standpoint, is that the variance of the noise must be finite. From a physical perspective, the noise must not prefer one direction over another, which means the first

*With the substitution from $\tilde{t} \to \gamma m t$, information about the actual experimental time in the simulation is lost.

moment (the mean) must be zero. Next, I invoke the central limit theorem, which states that the probability density of the sum of many (infinite) independent random variables will be close to Gaussian, and use the Gaussian distributed noise as the discrete form of the Wiener process which is compatible with the previous mathematical and physical constraints. These random Gaussian variables, with mean zero and variance unity, are labeled ξ_i for the i^{th} time step.

$$x_{i+1} = x_i + F(x_i)\Delta t + \sqrt{2\epsilon\Delta t}\,\xi_i \tag{2.3}$$

This approximate quadrature[29] forms the canonical procedure to generate a Brownian trajectory which evolves forward in time.

The discussion of the forward-Euler method of integrating the Brownian Stochastic Differential Equation (SDE) has so far avoided all mention of any errors that are overlooked in the approximate quadrature. The method of derivation of the SDE is a continuous time process, but the discretization will incorrectly sample this process for any finite time step, Δt. These errors are entirely hidden from view, with no clear mechanism of estimation. The text books discuss convergence of the path[30], but little attention is paid to how the errors affect the thermodynamic distribution.

2.1.2 Free Brownian Motion and the Quadratic Variation

Consider a special case of Brownian motion, where the force acting on the particles is zero. This corresponds to free Brownian motion which is defined entirely by the thermalized Wiener process

$$dx = \sqrt{2\,\epsilon}\,dW_t \tag{2.4}$$

Understandably, this process will lead to a RMS position which grows with time, as there is no restoring force. While this process does not seem to be of great importance physically, it does give rise to a very useful expression which I will call the quadratic variation sum rule. Summing the square of the change in position yields the following expression

$$\sum_{i=1}^{N} (x_{i+1} - x_i)^2 = 2\epsilon\Delta t \sum_{i=1}^{N} \xi_i^2 \tag{2.5}$$

The Gaussian distributed random numbers, ξ_i, are defined to have variance equal to unity, which allows us to rewrite the final sum as $\sum_{i=1}^{N} \xi_i^2 = N$. This expression is used to relate the fluctuations in the position of the particles to the underlying thermal bath and the time step

$$\sum_{i=1}^{N} (x_{i+1} - x_i)^2 \approx 2\epsilon \Delta t N = 2\epsilon T \tag{2.6}$$

If Δt is small enough, this sum rule will be approximately obeyed even when the force is non-zero.

2.2 The Metropolis-Hastings Algorithm

The Metropolis-Hastings Monte-Carlo (MHMC) algorithm[14] is a method for sampling a probability distribution function without knowing the normalization constant. The probability distribution function, $P\big(\{x_i\}\big)$, and the expected value of an operator O is given by $\int O\big(\{x_i\}\big)\, P\big(\{x_i\}\big)\Pi_i\, dx_i$. The Metropolis method can be visualized as a way of moving through the space of variables $X = \{x_i\}$ in a manner that reflects the underlying distribution P. The power of the Metropolis algorithm lies in its ability to exactly (in the infinite sample limit) sample a multidimensional probability density, that is within statistical errors.

Suppose the system is currently at a position X, and proposes a move to Y. Define the transition matrix $T(X \to Y)$ to provide the probability of choosing the proposed state Y given the current state X. The proposal probability, $Q(X \to Y)$, is the probability that a position Y will be chosen, given the current state is X. The acceptance probability, $\alpha(X \to Y)$, is the rate at which the move from X to Y will be accepted, according to some constraint.

The stationary property of the probability distribution requires the flux of particles into a state be balanced by the outward flux, namely

$$\sum_{Y} P(Y) T(Y \to X) = \sum_{Y} P(X) T(X \to Y). \tag{2.7}$$

This assures that the overall probability distribution is constant over time. For equilibrium processes, the stationarity of microstates is also ensured when all transition processes are equilibrated

by the reverse process. This is referred to as detailed balance where

$$P(Y)T(Y \to X) = P(X)T(X \to Y) \tag{2.8}$$

for all states X and Y, which is clearly consistent (term by term) with the stationarity condition.

Detailed balance implies that the Metropolis update leaves the canonical distribution for q and p invariant. To see this invariance explicitly, let $R(X)$ be the probability of state X being rejected as a proposal. The probability of the next state being in state Y is $P(Y)$. Note that the transformation will happen with probability one unless it is rejected, which can be restated as $\sum_X T(Y \to X) = 1 - R(Y)$.

There are two cases to be summed to give the probability that the proposed move ends in Y^{new}. Either the state is already in Y, rejection of the move is required, still in Y or the state is not in Y, require acceptance to go to Y.

$$\begin{aligned}
P(Y^{\text{new}}) &= P(Y)R(Y) + \sum_X P(X)T(X \to Y) \\
&= P(Y)R(Y) + \sum_X P(Y)T(Y \to X) \\
&= P(Y)R(Y) + P(Y)\sum_X T(Y \to X) \\
&= P(Y)R(Y) + P(Y)(1 - R(Y)) \\
&= P(Y)
\end{aligned}$$

Thus, the probability of the current and proposed states for moves which satisfy detailed balance will have equivalent probabilities. Detailed balance is a sufficient condition for a Markov process to reach a stationary posterior distribution, and will subsequently be required to hold for the MHMC algorithm.

If the current state is X, what is the transition probability of going to any state which is consistent with this condition? The transition probability of the system actually making the final transition from state X to Y is simply the probability of proposing the new Y state, multiplied by

the probability of accepting the move

$$T(X \to Y) = Q(X \to Y)\alpha(X \to Y) \tag{2.9}$$

The acceptance rate will now need to be chosen to satisfy the detailed balance criteria in equation 2.8. * The Metropolis choice of

$$P(X)T(X \to Y) = \min(P(X)Q(X \to Y), P(Y)Q(Y \to X)) \tag{2.10}$$

$$P(Y)T(Y \to X) = \min(P(Y)Q(Y \to X), P(X)Q(X \to Y)) \tag{2.11}$$

are equivalent and thus obviously satisfy the detailed balance equation. Solving for the transition probability

$$\begin{aligned} T(X \to Y) &= \frac{1}{P(X)} \min(P(X)Q(X \to Y), P(Y)Q(Y \to X)) \\ &= Q(X \to Y) \min\left(1, \frac{P(Y)Q(Y \to X)}{P(X)Q(X \to Y)}\right) \end{aligned}$$

Comparing to equation 2.9 gives the acceptance rate

$$\alpha(X \to Y) = \min\left(1, \frac{P(Y)Q(Y \to X)}{P(X)Q(X \to Y)}\right) \tag{2.12}$$

In the algorithm, the acceptance rate is calculated by comparing α to η, a uniform variate on the interval $(0, 1]$, and rejecting the move from X to Y if $\alpha(X \to Y) < \eta$. Note that the normalization constant cancels in the definition of α. When the map $X \to Y$ is continuous and not symplectic, Green's name is added to the algorithm[31], and the Jacobian of the map is folded into α.

$$\alpha(X \to Y) = J \min\left(1, \frac{P(Y)Q(Y \to X)}{P(X)Q(X \to Y)}\right) \tag{2.13}$$

* There is also the case that the current state is X and the proposed state is X. The transition probability is equal to proposing X and accepting X, plus the probability of being in A, proposing state k, and being rejected.

$$T(X \to X) = Q(X \to X)\alpha(X \to X) + \int_k A(X \to k)(1 - \alpha(X \to k)dk$$

This transition probability is complicated, but will always satisfy detailed balance as $P(X)T(X \to X) = P(X)T(X \to X)$, and can thus be ignored for this discussion.

For the integration schemes in this section, this Jacobian term is as follows

$$\mathbb{J}_n = h\frac{\partial v_n^i}{\partial x_{n+1}}\left(\frac{\partial v_n^f}{\partial x_i}\right)^{-1} \tag{2.14}$$

where the symbols are explained below.

2.2.1 Outline of the Metropolis Hastings (Green) Algorithm

In order to proceed generating the sequence of integrated particle positions, I will first define the time step h and the thermal energy ϵ. The general algorithm will proceed as follows:

1. Initialize the starting position (x_0)
2. Generate a proposed position (x_1) using equation 2.16 with the velocity sampled from the Maxwell-Boltzmann distribution
3. Perform Metropolis-Hastings-Green on the updated position:
   - if $|\mathbb{J}|\exp(-\delta e/\epsilon) > \eta \Rightarrow$ accept proposed state
   - else $\Rightarrow$ reject state and set $x_1 = x_0$
4. Save state x_1 as the new current position (x_0)
5. Repeat from step 2

It is important to note that both accepted and rejected states are saved into memory at step 4. This will form a sequence of positions which are used to analyze the properties system.

2.3 Integrating Forward

Let us begin the discussion of forward integration with a method that is closely related to the Euler-Maruyama method. The leapfrog, or velocity-Verlet, method is widely used when performing molecular dynamics simulations. The integration method proceeds as follows. The half step velocity is evaluated using the force from the initial position. The positions are then evaluated at the full step, using the velocities at the half step. Finally, the velocity at the full step is calculated using

the force at the full step position.

$$v_{1/2} = v_0 + \frac{h}{2} F(x_0)$$

$$x_1 = x_0 + h\, v_{1/2}$$

$$v_1 = v_{1/2} + \frac{h}{2} F(x_1)$$

Combining these equations yields the leapfrog integration procedure:

$$x_1 = x_0 + h v_0 + \frac{h^2}{2} F(x_0)$$

$$v_1 = v_0 + h \left(\frac{F(x_0) + F(x_1)}{2} \right)$$

The reason that this method is important to this discussion becomes clear with the identification of the relation between the time step h in leapfrog, and Δt from Brownian dynamics.

$$h = \sqrt{2\,\Delta t} \tag{2.15}$$

If one would choose the initial velocity at each step to be a uniform Gaussian random number scaled by temperature, $v_0 = \sqrt{\epsilon}\,\xi_0$, the Euler-Maruyama form of the Brownian SDE (equation 2.3) is obtained. Thus, leapfrog can be viewed as an alternate perspective on the original Brownian SDE evaluated with the Euler-Maruyama method.

2.3.1 Generalizing the forward integration

This simple integration framework can be extended in order to apply to Metropolis Hastings algorithm. To achieve this, I will introduce a weighted force, F_w, which will be defined according to the specific integration method used. In the example of the Leapfrog scheme, this weighted force is simply the force evaluated at the initial position. This generalization will allowed the derivation of a general framework to be used with more complex integration schemes.

Note that some of the integration methods presented in this chapter require the position to be known at a future time (x_{i+1}) for full evaluation of the weighted force. While this seems like a counterproductive choice, in later chapters of this thesis I will focus on solving the path space

problem, and the implicit nature of the integration will not be an issue. Furthermore, in order to make the derivation of the path space methods concise, the integration schemes will need to be simple enough to understand the numerical complexities involved. For these reasons, I will not present these more sophisticated and well studied higher order methods of integrating Hamiltonian systems*.

In order to simplify the notation, I will introduce the average and the difference in the position between two subsequent time steps as $\bar{x}_n$ and Δx_n, respectively.

$$\bar{x}_n = \frac{x_{n+1} + x_n}{2} \qquad\qquad \Delta x_n = x_{n+1} - x_n$$

The average force between two subsequent positions is defined as

$$\overline{F}_w(x_n) = \frac{F_w(x_{n+1}, x_n) + F_w(x_n, x_{n+1})}{2}$$

Beginning with an analogous equation as was seen in section 2.3, I define this foreword evolution of the past positions the following equation:

$$x_{n+1} = x_n + v_n^i h + \frac{1}{2} h^2 F_w(x_n, x_{n+1}) \tag{2.16}$$

Recall that reversibility of the Markov chain is a requirement for satisfying detailed balance. The reversed evolution of the chain must then evolve according to the backward equation, where the velocities are negated and evaluated at the ending time:

$$x_n = x_{n+1} - v_n^f h + \frac{1}{2} h^2 F_w(x_{n+1}, x_n) \tag{2.17}$$

Here, only on time step is considered before being refreshed. It is important to remember that the velocities are not the normal rate of change of the positions ($v \neq dx/dt$). Rather, they drawn from the Maxwell Boltzmann distribution†. This means that the velocities derived from the forward evolution of positions will only be used for the calculation of the energy change when moving from the current state to the proposed state. In other words, these velocities will only be used in the

*Interested readers can refer to Leimkuhler and Reich[3] for a discussion of these higher order methods.

†This corresponds to the random Gaussian variates described in the Brownian SDE (Section 2.1)

Metropolis Hastings Monte Carlo portion of the algorithm.

$$v_n^f = \frac{1}{h}(x_{n+1} - x_n + \frac{h^2}{2}F_w(x_{n+1}, x_n)) \tag{2.18}$$

$$v_n^i = \frac{1}{h}(x_{n+1} - x_n - \frac{h^2}{2}F_w(x_n, x_{n+1})) \tag{2.19}$$

The sum and the difference of these two velocities are used in the calculation the change in energy between initial and final state.

$$v_n^f - v_n^i = \frac{h}{2}(F_w(x_{n+1}, x_n) + F_w(x_n, x_{n+1})) \tag{2.20}$$

$$v_n^f + v_n^i = \frac{h}{2}(x_{n+1} - x_n + \frac{h^2}{4}(F_w(x_{n+1}, x_n) - F_w(x_n, x_{n+1}))) \tag{2.21}$$

Error in energy between steps

$$\begin{aligned}
\delta e &= \Delta PE + \Delta KE \\
&= U(x_1) - U(x_0) + \frac{1}{2}\left(v_n^f\right)^2 - \left(v_n^i\right)^2 \\
&= U(x_1) - U(x_0) + \frac{1}{2}\left(v_n^f + v_n^i\right)\left(v_n^f - v_n^i\right) \\
&= U(x_1) - U(x_0) + \frac{1}{2}\left(\Delta x_n + \frac{h^2}{4}(F_w(x_{n+1}, x_n) - F_w(x_n, x_{n+1}))\right)\overline{F}_w(x_n) \\
&= U(x_1) - U(x_0) + \Delta x_n\,\overline{F}_w(x_n) + \frac{\Delta t}{4}\left(F_w(x_{n+1}, x_n)^2 - F_w(x_n, x_{n+1})^2\right)
\end{aligned} \tag{2.22}$$

The overall change in energy while moving between the starting and ending positions is required for Metropolis Hastings test. It also gives insight into what the Euler-Maruyama method ignores when integrating the Brownian SDE. While this equation might be complicated for complex systems, in simple cases these forces can be easily evaluated to find the overall magnitude of the error.

2.3.2 Leapfrog with MHMC

Let us begin with the simplest case, where the weighted force is simply the force at the initial position

$$F_w(x_i, x_{i\pm1}) = F(x_i) \tag{2.23}$$

This definition of the force corresponds directly to the leapfrog and Euler-Maruyama case in the previous sections. The velocities are chosen from the Maxwell-Boltzmann distribution, $v_i = \sqrt{\epsilon}\,\xi_i$. Equation 2.16 yields the following procedure

$$\begin{aligned} x_{n+1} &= x_n + \frac{1}{2}h^2 F(x_n) + \sqrt{\epsilon}\,\xi_i\, h \\ &= x_n + F(x_n)\Delta t + \sqrt{2\,\epsilon\,\Delta t}\,\xi_i \end{aligned} \tag{2.24}$$

It is important to note that this method is explicit, the updated position can be calculated directly from information given by the current position.

The absolute value of the Jacobian transformation shown in below is unity, which implies the integrator is symplectic.

$$|\mathbb{J}_n| = \left| \frac{\partial v_n^i}{\partial x_{n+1}} \left(\frac{\partial v_n^f}{\partial x_n} \right)^{-1} \right| = \left| \frac{1}{h} \left(-\frac{1}{h} \right)^{-1} \right| = 1 \tag{2.25}$$

Thus, the Metropolis Hastings procedure is sufficient to drive the integration to the stationary probability, without requiring Green's correction.

All that is left to compute to complete Metropolis Hastings procedure is the error made between the current and proposed positions, as stated in equation 2.22.

$$\delta e = U(x_{i+1}) - U(x_i) + (x_{i+1} - x_i) \left(\frac{F(x_{i+1}) + F(x_i)}{2} \right) + \frac{\Delta t}{4} \left(F(x_{i+1})^2 - F(x_i)^2 \right) \tag{2.26}$$

2.3.3 Mid Point with MHMC

Let us now consider a refinement in the calculation of the weighted force, that continues to be a symplectic method, but which eliminates a term and the energy error. This weighted force is evaluated at the average of the current and proposed positions.

$$F_w(x_i, x_{i+1}) = F_w(x_{i+1}, x_i) = F\left(\frac{x_i + x_{i+1}}{2} \right) \tag{2.27}$$

Using this so called midpoint force, I write the forward integration

$$\begin{aligned} x_{n+1} &= x_n + \frac{1}{2}h^2 F(\bar{x}_n) + \sqrt{\epsilon}\,\xi_i\,h \\ &= x_n + F\left(\frac{x_n + x_{n+1}}{2}\right)\Delta t + \sqrt{2\,\epsilon\,\Delta t}\,\xi_i \end{aligned} \tag{2.28}$$

This weighted force will rely on both the current and proposed position, and is thus an implicit integrator. For a very simple force this equation as a possibility of being inverted to explicitly solved for the proposed move. When the force is more complicated, the solution to the proposed position will need to be carefully constructed such that the numerical error is minimized. In the following analysis, this is implemented via an iterative method, but the numerical error is only of the machine precision of the underlying variable type (approximately 13 digits using double precision variables).

This integrator is symplectic, which is again seen by evaluating the Jacobian transformation (equation 2.25), which has a more complicated form than the leapfrog method

$$|\mathbb{J}_n| = \left|\frac{\partial v_n^i}{\partial x_{n+1}}\left(\frac{\partial v_n^f}{\partial x_n}\right)^{-1}\right| = \left|\left(1 - \frac{\Delta t}{2}F'(\bar{x}_i)\right)\left(-1 + \frac{\Delta t}{2}F'(\bar{x}_i)\right)^{-1}\right| = 1 \tag{2.29}$$

In analogy to the leapfrog method, Metropolis-Hastings-Green is not needed and Metropolis Hastings will suffice.

The advantage of the midpoint method is seen when looking at the integration error. The last part of the equation for the energy error (equation 2.22) is zero, which will turn out to be integral to the understanding of the path space methods explained in later chapters.

$$\delta e = U(x_{i+1}) - U(x_i) + (x_{i+1} - x_i)\,F(\bar{x}_i) \tag{2.30}$$

2.3.4 Simpson's Method with MHMC

The last, and most sophisticated, integration method which will be discussed in this chapter is based on Simpson's method. The method can be thought of as a weighted average between the midpoint quadrature explained above and the trapezoidal quadrature which will not be considered

in the following analysis*. It can also be derived using a quadratic polynomial as an interpolating function between the two endpoints. The advantage of this method is that the energy error in the quadrature is of order $\Delta t^2 \propto h^4$.

The weighted force for Simpson's method is defined to be

$$F_w(x_i, x_{i+1}) = F_w(x_{i+1}, x_i) = (F(x_n) + 4F(\bar{x}_n) + F(x_{n+1}))/6 \tag{2.32}$$

This more complicated expression for the force again leads to an implicit method of integration. The prescription for the forward integration scheme is as follows

$$\begin{aligned} x_{n+1} &= x_n + \frac{h^2}{12}(F(x_n) + 4F(\bar{x}_n) + F(x_{n+1})) + \sqrt{\epsilon}\,\xi_i\, h \\ &= x_n + (F(x_n) + 4F(\bar{x}_n) + F(x_{n+1}))\frac{\Delta t}{6} + \sqrt{2\,\epsilon\,\Delta t}\,\xi_i \end{aligned} \tag{2.33}$$

It is very unlikely that an inverse can be found for this somewhat complicated expression, thus the only option for solving for the proposed position is by an iterative method.

In contrast to the previous two integration schemes, Simpson's method breaks the symplectic nature of the Hamiltonian flow. This can be seen by evaluating Jacobian

$$\begin{aligned} |\mathbb{J}_n| = \left| \frac{\partial v_n^i}{\partial x_{n+1}} \left(\frac{\partial v_n^f}{\partial x_n} \right)^{-1} \right| &= \left| \left(1 \frac{\Delta t}{6} \left(2F'(\bar{x}_n) + F'(x_{n+1}) \right) \right) \left(-1 \frac{\Delta t}{6} \left(2F'(\bar{x}_n) + F'(x_n) \right) \right)^{-1} \right| \\ &= \left| \frac{6 - \Delta t \left(2F'(\bar{x}_n) + F'(x_{n+1}) \right)}{6 - \Delta t \left(2F'(\bar{x}_n) + F'(x_n) \right)} \right| \end{aligned} \tag{2.34}$$

This expression in general, cannot be simplified. In order to deal with this non-unitary Jacobian, must resort to using the Metropolis Hastings Green form of Monte Carlo.

The error incurred from this quadrature is

$$\delta e = U(x_{i+1}) - U(x_i) + (x_{i+1} - x_i)\,(F(x_n) + 4F(\bar{x}_n) + F(x_{n+1}))/6 \tag{2.35}$$

*The trapezoidal method follows a similar derivation as shown in this section, but the below weighted force. This method is not symplectic and has the same order error in time as the midpoint method, so it falls out of favour here.

$$F_w(x_i, x_{i+1}) = (F(x_i) + F(x_{i+1}))/2 \tag{2.31}$$

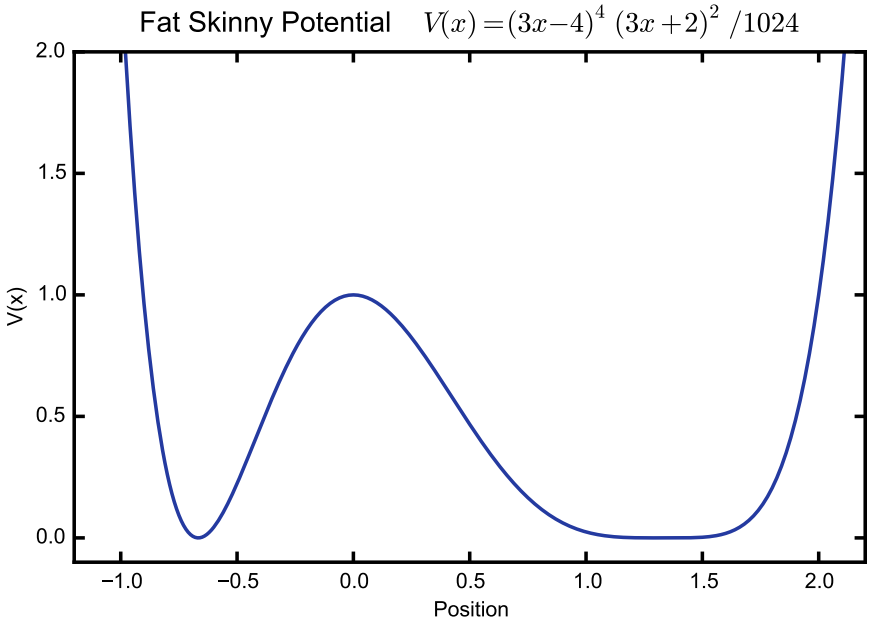


Figure 2.2: Plot of the fat-skinny potential $V(x)$

as stated earlier, the advantage of this method is that the error is the order Δt^2.

2.4 Results of forward-time methods

In the previous sections, I have explained a selection of methods of sampling a system of particles in the Brownian regime. In this section, will focus on the results obtained when sampling with these methods. In order to concisely present these results, I have used a very simple system as a test for the above proposed methods. The system consists of a single particle that resides in an externally imposed potential which is in thermal equilibrium with the underlying heat bath. The potential chosen for this analysis will be called the *fat-skinny* potential, which is shown in Figure 2.2, and has the following form

$$V(x) = \frac{(3x-4)^4(3x+2)^2}{1024} \tag{2.36}$$

This potential is used to define the force used in the forward integration methods.

$$F(x) = -\frac{dV(x)}{dx} \tag{2.37}$$

Among other effects, I am interested in understanding the role of entropy. This simple potential was carefully constructed to have degenerate energy basins of zero energy with a barrier of unit height separating these basins. These energy basins have different widths, the well on the left being narrow, the well on the right well, wide.

It is instructive to ask the following question: "What am I expecting to obtain from performing this analysis on the above potential?" This being a thesis in physics, the ultimate goal of the analysis is to produce a physical quadrature. If the method is to be considered a physical integration scheme, the particles should visit positions in these basins according to the Boltzmann distribution. At a given temperature ϵ, the probability of the particle residing between positions $x_a < x < x_b$ is

$$P(a,b) = \frac{\int_{x_a}^{x_b} \exp(-V(x)/\epsilon)}{\int_{-\infty}^{\infty} \exp(-V(x)/\epsilon)} \tag{2.38}$$

Remembering the original goal of sampling rare events, the thermal energy of the system will be set at $\epsilon = 0.15$, which is much less than the barrier height. At this temperature, the particle should spend approximately 0.81% of the trajectory in the positive wider well. The accuracy of the integrator turns out to be an important factor when trying to recover this probability distribution.

I have used the quadrature methods outlined in above, to generate very long trajectories, which are used to gather statistical information about the positional probability. The simulation necessarily depends upon the path being extremely long, and therefore, each trajectory possesses a large number of transitions. As can be seen in figure 2.3, the probability of the particle spending the correct amount time in each well is strongly dependent on integrator and the step size used. This time step size is especially important when using lower order methods such as leapfrog. In fact, even for relatively small time step sizes this integration method fails to sample the exact Boltzmann distribution. Note that this incorrect sampling is a systematic error which is overlooked without using a Monte-Carlo method. It is an unavoidable artifact of the approximate quadrature.

The appropriate way to handle this error in sampling is to apply the Metropolis-Hastings algorithm. A specific realization of this correction is shown in figure 2.4. This simulation uses the leapfrog quadrature with a time step $\Delta t = 0.035$, and bins the positions to form a histogram. The results of using leapfrog without the Metropolis Hastings correction leads to a grossly incorrect sampling of the distribution. Using the same step size as the leapfrog case, sampling with Metropolis

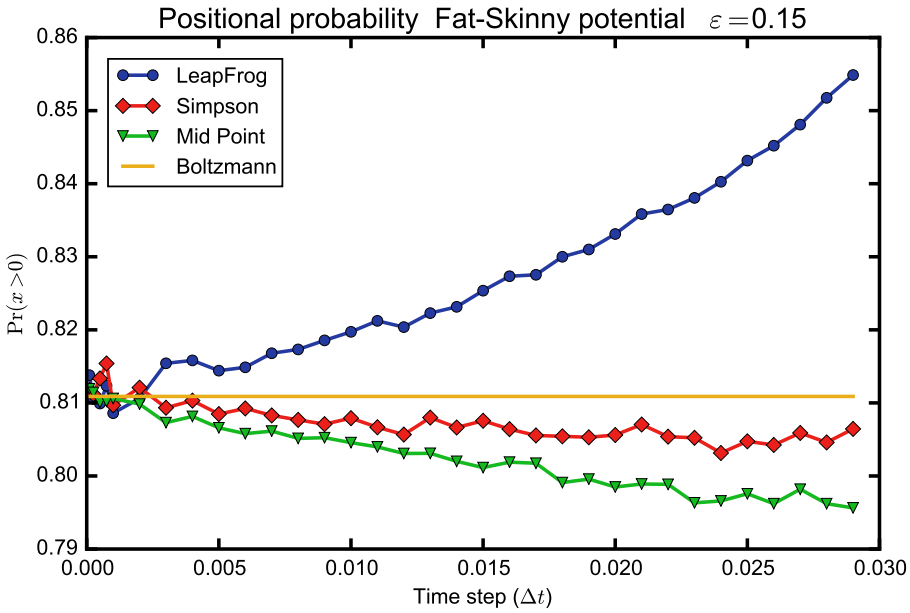


Figure 2.3: Probability of the path being in the positive well, for varying time steps, shown for each of the forward integration methods. This error results directly from the error made in integration, and can be corrected by applying the Metropolis-Hastings(-Green) algorithm. Note as step size increases leapfrog overestimates the probability of the path in the narrow well, while the other two methods underestimate this probability.

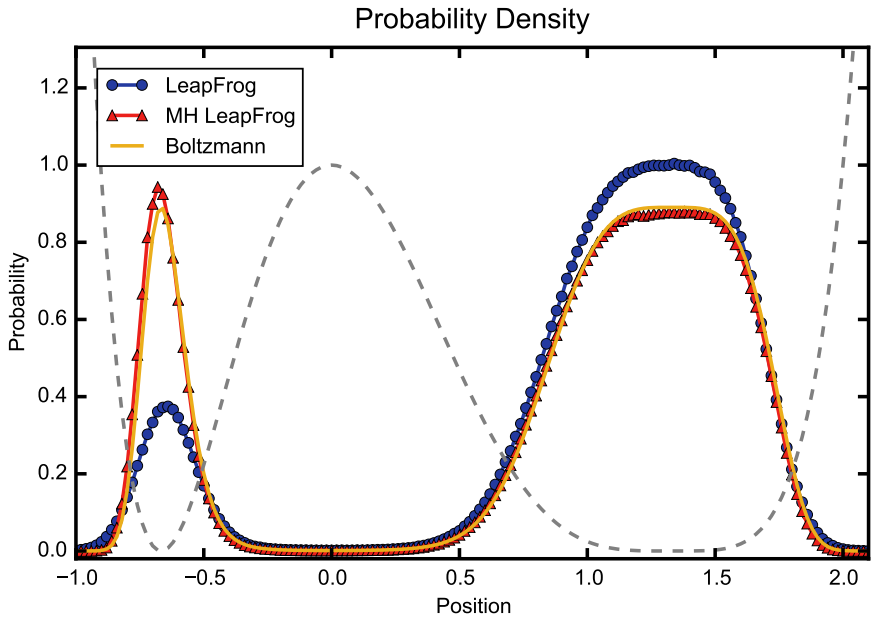


Figure 2.4: The probability of the position of the particle for a very long path ($T = 3.5 \times 10^4$) with $\Delta t = 0.035$ at a temperature $\epsilon = 0.15$. The Metropolis Hastings adjusted sampling, shown in red triangles, closely matches the expected Boltzmann probability. The integration using the leapfrog sampling quadrature drastically underestimates the probability of the particle in the narrow well.

Hastings drives (exactly in the long time limit) the sampling to the posterior probability distribution. Understanding where these errors originate is a very subtle problem, which will turn out to be a very useful exercise when trying to understand path space methods in later chapters.

On closer inspection of figure 2.4, one can see that the positional probability in the narrow well is much broader than it should be according to the Boltzmann distribution. A renormalization of the probability in the left well will still not lead to physical sampling. (I'm trying to convey the fact that the probability at the center of the narrow well underestimates the probability while just outside, specifically near $x \approx -0.7$, the probability is overestimated. Renormalizing this probability to match the Boltzmann weight will still yield an unphysical distribution.)

This is not the only test which shows physical properties of this integration. Another test is to look at the quadratic variation of the particle for specific positions in the potential, as can be seen in figure 2.5. Observe that the thermal fluctuations corresponding to the correct thermal background in the simulation is exaggerated when the curvature of the potential is large. This curvature becomes very large in the narrow well, and is close to zero and the wide well. The Metropolis sampling attempts to correct for this thermal overestimation. This can be understood by looking at the rejection rate (shown in figure 2.6) for the Metropolis Hastings sampling. At positions where there is a large overestimation of the thermal fluctuation for the vanilla leapfrog sampling, the rejection rate is large. Therefore, the rejections in these regions are responsible for bringing the thermal fluctuations back in check.

2.5 Hybrid Monte Carlo

An issue often overlooked with SDE sampling is the diffusive nature of the motion defined by the process. The motion often seems only to be randomly fluctuating in its local neighborhood, and collective motion of the system is rare. This behaviour is created by the underlying thermal fluctuations as modeled by the Wiener process, which appears as the random Gaussian variates in the discrete case, and creates a diffusive path. This randomness leads to a large inefficiency in the sampling, where most of the computational time is spent simulating the background noise, rather than sampling the interesting dynamics of the system.

This randomness can also be viewed as the method's main advantage, as choosing velocities at

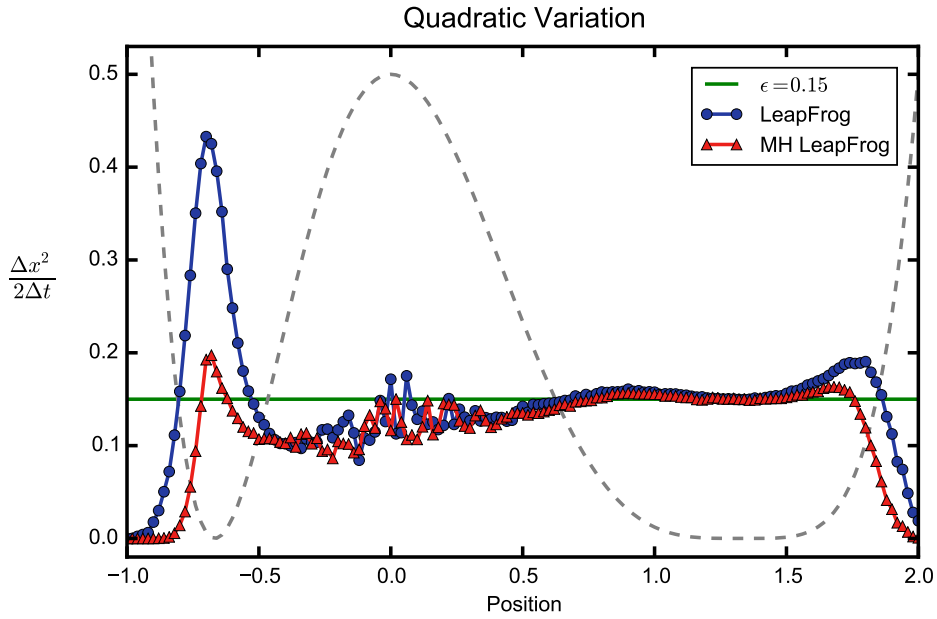


Figure 2.5: Figure showing how the quadratic variation sum rule is satisfied at different parts of potential. Note the large overestimation in particle fluctuation near the left, very stiff, potential wall. The very noisy structure in the vicinity of barrier is due to the low sampling weight at the top of the barrier.

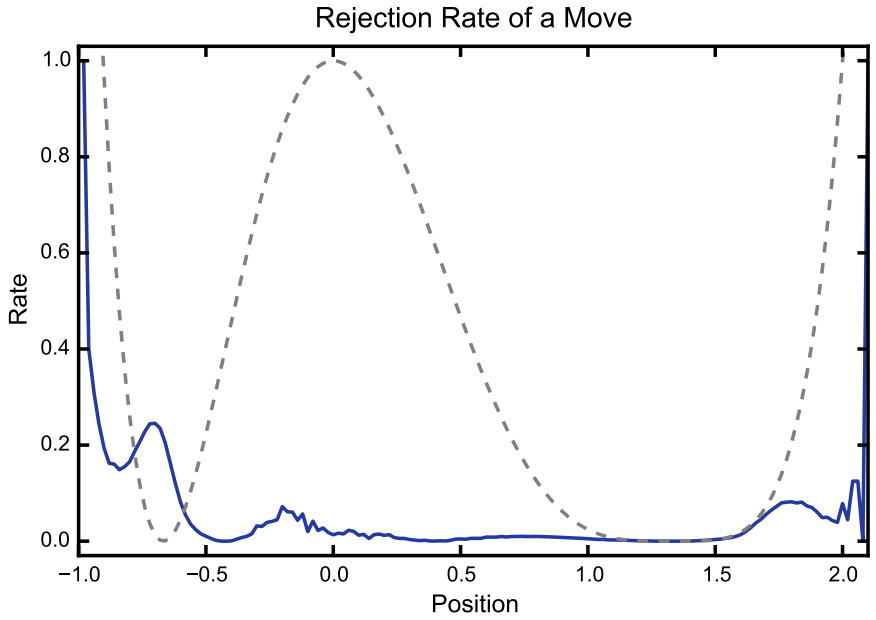


Figure 2.6: Rejection rate of the Metropolis Hastings leapfrog sampling which gives insight into why the quadratic variation is better behaved for the Metropolis Hastings algorithm where the potential is very stiff (see Figure 2.5). There is a relatively large rejection rate (≈ 0.2) then this vicinity.

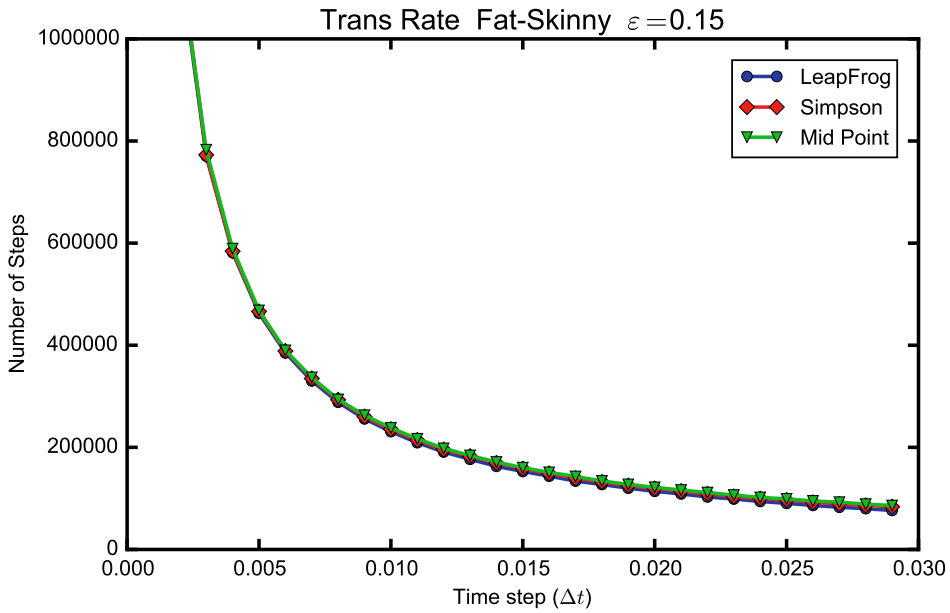


Figure 2.7: The length of path required to make the transition from one basin, to another, and back again. Observe the exponential nature of increasing at to observe as the time step, Δt, decreases.

random leads to sampling of the entirety of the available phase space. At very large integration times, when combined with the Metropolis adjustment, this sampling is very robust and will almost surely converge to the correct posterior probability distribution. The problem now is how to increase the sampling efficiency.

One possible way to improve the efficiency of the sampling is to use the Hybrid Monte Carlo (HMC) method[15] which uses a Hamiltonian flow to greatly increases the amount of phase space explored compared to the Brownian sampling. The previously discussed methods, like leapfrog or Metropolis Adjusted Langevin Algorithm (and other simple methods such as random walk metropolis), inefficiently sample the state space as the overall motion is at best diffusive and at worst random. The Brownian sampling draws states from the entire distribution of possible states given for the configurational temperature and thus imposes no restrictions on the possible phase space accessible. This is a powerful sampling property that stochastic simulation provides. The diffusive movement however, means that sampling is quite slow to move around the available phase space. This is in accordance with standard physical intuition, where particles are bouncing around very sporadically and collective movement is rare.

Roughly, the Brownian method generates moves in a region of phase space that grows propor-

tional to the square root of the sampling time. On the other hand, since the Hamiltonian flow is ballistic, the region probed is linear. The Hamiltonian flow conserves energy, possesses a time reversal symmetry and conserves the amount of phase space available at any point (known as Louville's Theorem), all of which are shown in Appendix 2.A. These properties lead to an overall collective motion of the system over short periods of time, which increases the overall configurational change along the constant energy surface. This can be understood by imagining a classical ballistic system; once the system is moving in a direction in phase space, (q,p), it will continue around barriers. This phase space conservation restricts the total phase space available to the stochastic simulation. This means that the initial conditions on the flow need to be modified in order to sample the entirety of phase space.

The final part of the sampling is a Metropolis-Hastings test which drives the probability distribution to stationarity. This algorithm has a fairly strict set of conditions, relying on detailed balance and the ergodic theorem. The HMC method uses a single step of the Brownian dynamics followed by many deterministic steps and finally tests the proposals with the Metropolis-Hastings test. For this analysis it is critical that the stationarity of the probability distribution be preserved. The detailed balance criteria of Metropolis-Hastings ensures this behaviour, with the acceptance probability given in equation 2.13. This test is necessary only because of the errors made in the numerical integration.

2.5.1 Constructing the Markov Chain

This particular analysis is restricted to the classical canonical ensemble.

$$\mathbb{P}(x) = \frac{1}{\mathcal{Z}} \exp\left(\frac{-E(x)}{\varepsilon}\right) \tag{2.39}$$

In particular, I am interested in studying systems which are described by a classical Hamiltonian which depends on momentum (p) and positions (q). The probability distribution can the be defined in terms of the Hamiltonian joint distribution as follows:

$$\mathbb{P}(q,p) = \frac{1}{\mathcal{Z}} \exp\left(\frac{-H(q,p)}{\varepsilon}\right) \tag{2.40}$$

This choice drives the dynamics to move within a shell of constant probability. Defining the Hamiltonian to be $H(p,q) = V(q) + K(p)$ allows the separation of the spacial from the momenta components in the probability.

$$\mathbb{P}(q,p) = \frac{1}{\mathcal{Z}} \exp\left(\frac{-U(q)}{\varepsilon}\right) \exp\left(\frac{-K(p)}{\varepsilon}\right) \tag{2.41}$$

This becomes an important consideration, as it allows the generation of axillary variables in the path space implementation of the algorithm, to be discussed in chapters 4 and 7. Each of these terms must have distributions which are consistent with the background temperature bath.

2.5.2 HMC algorithm

There are two steps that transform the initial state (q,p) to the final state (q^*,p^*).

- Step 1: q is unchanged, p is drawn from the Gaussian distribution. This happens for a single step.
- Step 2: Leapfrog (or another integration scheme) is used to update the positions of (q). Momentum is negated at the end. This step may repeat many times. The final state after the leapfrog step(s) is the proposed move (q^*,p^*).

The proposal, (q^*,p^*), is accepted with probability

$$\min\left\{1, \exp\left[-H(q^*,p^*) + H(q,p)\right]\right\} = \min\left\{1, \exp\left[-U(q^*) + U(q) - K(p^*) + K(p)\right]\right\} \tag{2.42}$$

If the proposal is accepted, the new positions are updated $(q,p) = (q^*,p^*)$. If rejected, the new configuration is reset to the configuration before the 2 steps $(q,p) = (q,p)$. Looking at the sampling from the joint probability distribution, step 1 is responsible for all of the movement of the probability density of (p,q). The second step is responsible for a large movement in the probability of q but $H(q,p)$ is nearly constant without step 1.

2.5.3 Properties of the HMC method

A crucial property of the HMC algorithm is one that affects the probability distribution in the first step in section 2.5.2. The choice of random momenta, importantly drawn to correspond to the Boltzmann distribution, will not necessarily leave the probability invariant, but all other steps will. The numerical integration of the Hamiltonian flow is symplectic and thus conserves phase space. The rejections in the Metropolis-Hastings step is needed to preserve detailed balance. These two together ensure that the algorithm samples the target probability distribution.

The question still remains: Why does the sampling improve on the original Brownian dynamics? As stated previously, introducing a Hamiltonian flow between the Brownian dynamics and the Metropolis test serves to dramatically increase the movement in phase space. These coordinate changes have a magnitude comparable to the standard deviation of the motion in the most highly constrained direction. That is, the changes are constrained to be approximately the square root of the smallest eigenvalue of the covariance matrix. Configurations which are not highly constrained benefit the most from the ballistic Hamiltonian flow[32].

The choice of the number of steps for step 2 in section 2.5.2 is optimally determined by examining how long it takes to sample in the least constrained direction. This will lead to the most highly independent states for the given configuration in phase space. A number of steps needed the approximately equal to the square of the ratio of the most highly constrained to least constrained eigenvalues. This leads to an advantage over the simple random walk that is proportional to this ratio.

Appendix 2

2.A Properties of Hamiltonian Dynamics

Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p} \qquad \dot{p} = -\frac{\partial H}{\partial q} \tag{2.43}$$

$$H(q,p) = Uq + K(p) \qquad K(p) = p^T M^{-1} p/2 \tag{2.44}$$

Analog of this is uses the vector $z = (q,p)$. Hamiltons equations are then given by

$$\frac{dz}{dt} = J\nabla H(z) \qquad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$$

Invariance under time reversal

Here the $\rightarrow$ symbol is the time reversal.

$$t \rightarrow -t' \qquad q \rightarrow q' \qquad p = m\frac{dq}{dt} \rightarrow -m\frac{dq'}{dt'} = -p'$$

$$H = H(q,p) = V(q) + \frac{p^2}{2m} \rightarrow V(q') + \frac{(p')^2}{2m} = H(q',p') = H' \tag{2.45}$$

Now look at Hamilton's equations under this transformation

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} \rightarrow -\frac{dq'}{dt'} = \frac{\partial H'}{\partial p'} \tag{2.46}$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} \rightarrow \frac{dp'}{dt'} = -\frac{\partial H'}{\partial q'} \tag{2.47}$$

These equations are invariant under time reversal. This should not be too unfamiliar, as it is rooted into our basic intuition of mechanics. A ball rolling up a hill can be time and momentum reversed and it is transformed to a system with identical energy but rolling down the hill.

Conservation

$$\frac{dH}{dt} = \sum \frac{\partial H}{\partial q}\frac{dq}{dt} + \frac{\partial H}{\partial p}\frac{dp}{dt} = \sum \frac{\partial H}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial H}{\partial p}\frac{\partial H}{\partial q} = 0 \tag{2.48}$$

An important note is that H is only conserved with the MHMC algorithm accepts all proposals, which will be approximately true for very small step size.

Volume Preservation

Liouville's Theorem. Divergenceless vector field preserves volume.

$$\nabla \cdot \dot{z} = \frac{\partial}{\partial q}\frac{dq}{dt} + \frac{\partial}{\partial p}\frac{dp}{dt} = \frac{\partial}{\partial q}\frac{\partial H}{\partial p} + \frac{\partial}{\partial p}\frac{\partial H}{\partial q} = 0 \tag{2.49}$$

Symplecticness

This condition defines the symplectic property,

$$B_S^T J^{-1} B_S = J^{-1} \tag{2.50}$$

Chapter 3

Path Sampling using Onsager-Machlup

In this chapter I turn my attention to an alternative approach to the sampling problem. In the previous chapter, I reviewed some of the conventional procedures in which to sample the Brownian SDE, along with a brief analysis of the methods. However, there are disadvantages associated with using the forward methods, which I will discuss below.

In the previous chapter, the discussion was solely about correctly generating the forward trajectory from which physical properties can be inferred. In these trajectories, the only information known *a priori* was the current position. This makes any forward facing trajectory act like taking a shot in the dark, where you somewhat randomly push particles around and then wait for rare and interesting events to occur.

The forward methods become especially unfavorable when the temperature is small compared to the energy barriers present in the system, and the waiting times for the observation of a rare event can become extremely large. Starting the trajectory in a specified energy basin, one finds the ending configuration will be approximately distributed according to the posterior probability distribution only if one waits a long enough time. The only option available is to wait for the simulation to spontaneously perform the action of interest. Waiting for the system to hop over a barrier is a frustrating experience. One of the aims of the work presented in this thesis is to mitigate some of these disadvantages.

There are two ideas which combine to overcome the frustration of the long waiting times. First, I will consider an alternative to the standard forward trajectory, and consider paths, where the entire motion is defined for a predetermined amount of time. I will impose an extra constraint where the paths are constrained to have predefined starting and ending configurations, thus forming a double-ended path. This double-endedness rectifies the difficulties with waiting times of transition to a specific energy basin or configuration, as the ending state is defined by the algorithm itself. I will then introduce the concept of the Onsager-Machlup probability measure[16], which is used to determine the probability of a predefined trajectory in time. This probability will be used in later sections as a probability distribution which can be sampled.

The question for the subsequent chapters will become: How would one use the HMC method to sample these double-ended paths in a way which remains consistent with thermodynamics.

3.1 Double-Ended Paths

The sampling methods presented for the remainder of this work will utilize an object referred to as a *path*. A path is distinctly different from what I have referred to as a trajectory. A trajectory is a forward looking sequence of positions, where the past and current positions are defined, but the future times remain unknown. On the other hand, a path is sequence of positions which are defined for all time steps between the initial time, t=0, in the final time, t=T. These paths do not have a single set of initial positions as the starting point like a forward trajectory does. Rather, the initial state is a full path, where all positions between times t=0 and t=T are defined.

The path object is fairly strictly defined in this work. There is a position vector of length N which is declared for each degree of freedom. Every path in a given simulation has a fixed time step, Δt, between each of the N positions. Using this time step and the number of total positions along the path defines the length of the path, T

$$T = N\,\Delta t \tag{3.1}$$

The path is fully defined by the time step Δt, the path time length T, and the total positions along the path N.

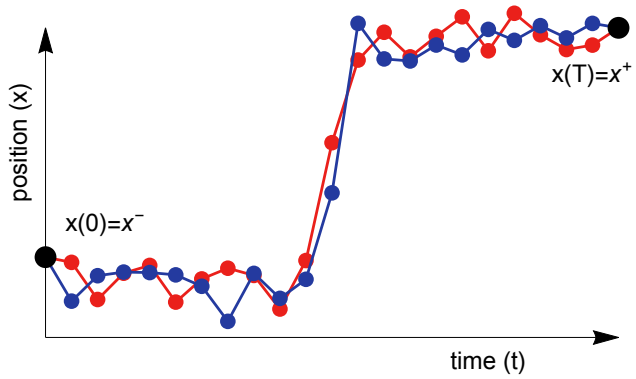


Figure 3.1: A cartoon example of two paths making a transition from an initial position (x^-) to a final position (x^+). All positions along both paths are defined at each time interval, separated by the time step along the path, Δt. The initial path, shown in red, is used as a starting point in the generation of the proposed path, shown in blue, which is separated by the time step between paths, $\Delta\tau$.

You may be wondering how introducing this new path object leads to any improvements in sampling physical systems. The previous chapter outlines how to proceed with sampling by creating trajectories and these newly introduced path objects act like trajectories of some fixed length and time. The paths are constrained at the initial time (analogously to forward methods), and are unconstrained at the ending point. I now impose additional constraint on the path, where the final position is also pinned to some predetermined point. These boundary conditions will be given the label x^- for the initial path position and x^+ for the final path position.

Imposing the constraint that the path starts and ends at specified positions requires each path to contain a transition from one energy basin to another. This constraint only affects the boundaries, and thus the center of the path is free to explore the free energy landscape between the two basins. This pinning alleviates the long waiting times needed to sample rare events.

The objective of the new sampling methods which are introduced in the subsequent chapters is to generate a sequence of paths, all of which are consistent with the thermodynamics of the system. The artificial time step between subsequent paths is referred to as $\Delta\tau$ and is different from the time step along the path (Δt). $\Delta\tau$ will turn out to be a tuning parameter which can be adjusted to obtain efficient sampling, while still maintaining an appropriate level of error. An illustration of two paths is shown in Figure 3.1.

3.2 The Onsager-Machlup Functional

The approach for writing the probability of a given path was first derived by Onsager and Machlup in 1953[16]. In this article, the authors find "the solution to the problem of finding the probability of any path." The idea has foundations in idea of the linearity of irreversible processes that won the first author the Nobel Prize in Chemistry, but adds a stochastic fluctuation in the form of a random force. Under the assumption of a Markovian process, they use the conditional PDF for the Ornstein-Uhlenbeck process to put formula for the path probability "into a particularly interesting form". Indeed, this work is at the foundation of path sampling methods which are widely used in many fields, and the result has been stated as being "incapable of improvement either in form or in their mode of derivation."[33]

The central idea for the use of this result is that, rather than blindly generating succession states, one can leverage the underlying noise history of a system to calculate the probability of the pathway. This procedure removes the restraint of only being able to sample trajectories which evolve forward in time, instead being able to calculate the probability of any path which is defined for a given time period.

The Onsager Machlup article wrote down the probability of a sequence of randomly fluctuating positions in a formula analogous to the probability $\mathbb{P}_{OM}$

$$\mathbb{P}_{\mathbb{OM}} = \prod_i \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\xi_i^2\right\} \tag{3.2}$$

Using the definition of the Brownian SDE, and its discretized form

$$dx = F(x)dt + \sqrt{2\epsilon}dW \tag{3.3}$$

$$\Delta x_i = F(x_i)\Delta t + \sqrt{2\epsilon\Delta t}\,\xi_i \tag{3.4}$$

Using the discrete Brownian SDE, one can simply solve for the square of the random Gaussian variate. Putting this this expression into $\mathbb{P}_{OM}$ yields an expression for the probability which can

be solved in terms of the path positions themselves.

$$\mathbb{P}_{OM} \propto \prod_i \exp\left\{-\frac{\Delta t}{2\epsilon}\left(\frac{1}{2}\left(\frac{\Delta x_i}{\Delta t} - F(x_i)\right)^2\right)\right\} \tag{3.5}$$

$$-\ln \mathbb{P}_{OM} \propto \frac{\Delta t}{2\epsilon}\sum_i \frac{1}{2}\left(\frac{\Delta x_i}{\Delta t} - F(x_i)\right)^2 \tag{3.6}$$

The statement of the path of probability is entirely defined for a given sequence of fluctuations.

Chapter 4

Path-Space Hybrid Monte-Carlo

The methods presented in Chapter 2 were aimed at creating trajectories which evolve forward in time. The previous chapter motivated an alternative approach to the forward trajectory, where sampling is performed on a double-ended path which is defined for all times between $t = 0$ and $t = T$, as well as introduced a method of calculating the probability of such a path using the Onsager-Machlup functional.

The focus of this chapter is to develop a computational framework which can be used to sample these double ended paths in the continuous-time limit. In this limit, the Onsager-Machlup function can be used to create an effective *thermodynamic action* of the system, which uses the continuous-time *path potential*. The standard way to treat this new limit is to use Ito calculus and the Girsanov theorem[18] to create an alternative path probability[17][20], $\mathbb{P}_{IG}$, which is used as the measure from which the samples are drawn.

I will explain how to go about sampling this measure by using the HMC algorithm, which in continuous-time is called the Path-Space Hybrid Monte Carlo method[13] (PSHMC). This method leverages the joint probability distribution to untangle the kinetic term in the Hamiltonian from the paths to form auxiliary velocities, which inherit the role of the random Gaussian variates from the forward methods. The details of the algorithm will then be reviewed, along with the detailed calculation of the energy drift which is necessary for the Metropolis-Hastings correction. I end the chapter with a discussion of many of the numerical considerations as well as proposing a framework for implementing the algorithm on a computer.

4.1 Thermodynamic Action and the Path Potential $G(x)$

This section will lay the groundwork for the continuous-time probability measure from which the path sampling method will be based upon. The discussion begins with the Euler-Maruyama form of the Brownian SDE from equation 2.3

$$\frac{\Delta x_i}{\Delta t} = F(x_i) + \sqrt{\frac{2\epsilon}{\Delta t}}\xi_i \tag{4.1}$$

Solving for the Gaussian random variates squared yields

$$\xi_i^2 = \frac{\Delta t}{2\epsilon}\left(\frac{\Delta x_i}{\Delta t} - F_i\right)^2 \tag{4.2}$$

The Onsager-Machlup probability is then written down for the Brownian SDE, which was seen before in equation 3.6.

$$\mathbb{P}_{OM} \propto \prod_i \exp\left(-\frac{\xi^2}{2}\right) = \exp\left[\frac{\Delta t}{2\epsilon}\sum_i \frac{1}{2}\left(\frac{\Delta x_i}{\Delta t} - F_i\right)^2\right] \tag{4.3}$$

Lets now take a closer look at the argument of this sum, which has been called the Onsager-Machlup function[21]

$$\mathscr{L}_i = \frac{1}{2}\left(\frac{\Delta x_i}{\Delta t} - F_i\right)^2 \tag{4.4}$$

$$= \frac{1}{2}\left(\frac{\Delta x_i}{\Delta t}\right)^2 + \frac{1}{2}\left|F_i\right|^2 - \frac{\Delta x_i}{\Delta t}F_i \tag{4.5}$$

In order to cast the equation in a continuous (differential) form, the first and the last term in this sum must be carefully considered. The $\Delta x/\Delta t(dx/dt)$ term in this sum diverges in the continuous-time limit due to the fractal nature of the Brownian path. The first term will be handled by using a change in measure, to be discussed later. The last term in equation 4.5 can be approximated by employing the quadratic variation rule along the path, $(\Delta x_i)^2 \approx 2\epsilon\Delta t$ (see Appendix 4.A), in the

following way

$$\frac{\Delta x_i}{\Delta t}F_i = \frac{1}{2}\left(\frac{\Delta x_i}{\Delta t}(F_{i+1} - F_i)\right) - \frac{1}{2}\left(\frac{\Delta x_i}{\Delta t}(F_{i+1} + F_i)\right) \tag{4.6}$$

$$= \frac{1}{2}\left(\frac{(\Delta x_i)^2}{\Delta t}\frac{F_{i+1} - F_i}{\Delta x_i}\right) - \left(\frac{\Delta x_i}{\Delta t}\frac{F_{i+1} + F_i}{2}\right) \tag{4.7}$$

$$\approx -\epsilon\frac{dF}{dx} - \bar{F}\cdot\frac{dx}{dt} \tag{4.8}$$

Recognise that the last term is the force multiplied by the velocity, which is simply the power, or when integrated, the difference in the potentials at the boundary in the continuous-time limit. The new probability measure can then be written (informally) as a continuous-time process

$$-\ln\mathbb{P}_{\text{Informal}} = \frac{1}{2\epsilon}\int_0^T dt\left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}\big|F(x)\big|^2 - \frac{1}{2}\frac{(dx)^2}{dt}\frac{dF}{dx} + \bar{F}\frac{dx}{dt}\right] \tag{4.9}$$

$$= \frac{1}{2\epsilon}\left(V(x^+) - V(x^-) + \int_0^T dt\left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + \frac{1}{2}\big|F(x)\big|^2 - \epsilon\nabla^2 V(x)\right]\right) \tag{4.10}$$

$$= \frac{1}{2\epsilon}\left(V(x^+) - V(x^-) + \int_0^T dt\left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + G(x)\right]\right) \tag{4.11}$$

$G(x)$ is the so-called continuous-time *path potential*, which is defined as

$$G(x) = \frac{1}{2}\big|\nabla V(x)\big|^2 - \epsilon\nabla^2 V(x). \tag{4.12}$$

The second term in G is sometimes referred to as the Jacobian correction. For an explanation see [34].

The mathematical concern with this integral expression (Equation 4.11) is that it is formally divergent, due to the fractal nature of the Brownian path. Using the Radon-Nikodym derivative to express the change in measure relative to the free Brownian case leaves the relative change in measure as

$$-\ln\mathbb{P}_{IG} = -\ln\frac{d\mathbb{P}_{\text{Informal}}}{d\mathbb{Q}_p} = \frac{1}{2\,\epsilon}\left(V(x^+) - V(x^-) + \int_0^T dt\; G(x_t)\right) \tag{4.13}$$

This expression is known as the *Ito-Girsanov path probability*, and is this measure which is to be sampled with the PSHMC algorithm.

4.2 Path Space Hamiltonian

To write down the HMC algorithm in the continuous-time space, one first needs to define an effective Hamiltonian for the system. The informal probability distribution is used to define the effective path space Hamiltonian analogous to the joint distribution defined in Equations 2.40 and 2.41. The use of the label $\Lambda \to H/(2\epsilon)$ serves as a reminder this is not a Hamiltonian in the regular sense, it is a quantity which serves as an effective Hamiltonian. The informal probability distribution in Equation 4.11 is then written as

$$\mathbb{P}_{\text{Informal}} = \frac{1}{\mathbb{Z}} \exp(-\tilde{\Lambda}(x,p)) \tag{4.14}$$

Again, this is an informal equation because of the fractal nature of the path; the paths are almost nowhere differentiable.

As the PSHMC method is only interested in relative changes in probability, all constant terms will be dropped from this expression for $\tilde{\Lambda}$. The partition function (which can be excluded from the entire calculation because of the Metropolis-Hastings adjustment) and the evaluation of the potential energy change between the boundary conditions will be absorbed into this proportionality.

$$\tilde{\Lambda} \propto -\ln \mathbb{P}_{\text{Informal}} - \frac{1}{2\epsilon}\left(V(x^+) - V(x^-)\right) \tag{4.15}$$

$$\propto \frac{1}{2\epsilon}\left(\int_0^T \frac{1}{2}\left(\frac{dx}{dt}\right)^2 dt + \int_0^T G(x)dt\right) \tag{4.16}$$

The full effective Hamiltonian then includes an extra set of variable which form the random variates, adding a stochastic term into this distribution. The auxiliary variables (that is to say, the velocities) are generated to be consistent with the temperature. This addition defines Λ, which will used for sampling with the PSHMC method. An integration by parts is then applied to the first integral in Equation 4.16 to yield Λ as

$$\Lambda(x,p) = \frac{1}{2\epsilon}\left(\frac{1}{2}\int_0^T dt\big(p \cdot \boldsymbol{M}^{-1} \cdot p\big) - \frac{1}{2}\int_0^T dt\big(x \cdot \frac{d^2}{dt^2} \cdot x\big) + \int_0^T G(x)dt\right) \tag{4.17}$$

At this point it is convenient to introduce another operator $\boldsymbol{L}$, the second derivative with respect

to time

$$\boldsymbol{L} = \frac{d^2}{dt^2} \tag{4.18}$$

One of the large advantages of the path sampling is that the mass matrix can be chosen to be $-\boldsymbol{L}$ (a positive definite operator), as was shown in the work of Beskos *et.al.* [13]. This transformation allows the high frequency modes to be treated exactly.

$$\Lambda(x,p) = \frac{1}{2\epsilon}\left(-\frac{1}{2}\int_0^T dt\big(p\cdot\boldsymbol{L}^{-1}\cdot p\big) - \frac{1}{2}\int_0^T dt\big(x\cdot\boldsymbol{L}\cdot x\big) + \int_0^T G(x)dt\right) \tag{4.19}$$

There are a number of key points which I would like to address:

1. The momenta chosen in Equation 4.17 obey a known distribution (Gaussian). These auxiliary variables (that is to say, the velocities) are generated to be consistent with the temperature.

2. The choice of the Mass matrix $\boldsymbol{M} \to -\boldsymbol{L}$ means that the high frequencies modes are treated exactly[13]. This can be understood by realizing that the path potential $G \to 0$ for the high frequency oscillations of the path; the path is dominated by the noise of the stochastic process in this high frequency regime. For this case of the free particle, the Hamiltonian flow (and thus the high frequency modes) is treated exactly with a particular splitting to be discussed in 4.3. Thus, an analytic solution is obtained for this integration step, which avoids the problem of dimensionality as $\Delta t \to 0$.

3. The Hamiltonian flow is used to generate large movements in phase space. This flow is integrated over a large interval to facilitate efficient sampling, thereby de-correlating adjacent paths in the Markov chain. The algorithm uses a leap-frog integration over τ, using a time step, $\eta = \sqrt{2\,\Delta\tau}$, and a large number, N_{MD}, of steps. Note if $N_{MD} = 1$, then this method reduces to the Stochastic Partial Differential Equation of Stuart *et. al.*[35].

4. Errors are controlled by using a Metropolis-Hastings acceptance criterion based on the total change in the energy over multiple integration steps ($\Delta\Lambda$). The crucial point here is, without resorting to the subtraction of large numbers, that the effective-energy change can be determined in path space.

4.3 HMC in Path Space

In the previous two sections I have introduced continuous-time thermodynamic action and, using the potential G, have defined an effective path space Hamiltonian which can be used to sample paths. In this section, I will derive a framework of Beskos *et. al.*[13], which will be used to sample such paths. Reiterating the most important results from these sections, the effective path-space Hamiltonian, Λ, and the path potential, G, are

$$\Lambda(x,p) = \frac{1}{2\epsilon}\int_0^T dt\left(-\frac{1}{2}\big(p\cdot\boldsymbol{L}^{-1}\cdot p\big) - \frac{1}{2}\big(x\cdot\boldsymbol{L}\cdot x\big) + G(x)\right) \tag{4.20}$$

$$G(x) = \frac{1}{2}\big|\nabla V(x)\big|^2 - \epsilon\nabla^2 V(x). \tag{4.21}$$

where $\boldsymbol{L} = d^2/dt^2$ is the second derivative operator, $x = x(t)$, and $p = p(t)$.

Hamilton's Equations

Applying Hamiltons equations to this effective Hamiltonian yields the following two equations of motion

$$\frac{\partial x}{\partial \tau} = \frac{\partial \Lambda}{\partial p} = -\boldsymbol{L}^{-1}\cdot p = \dot{x} \tag{4.22}$$

$$\frac{\partial p}{\partial \tau} = -\frac{\partial \Lambda}{\partial x} = \boldsymbol{L}\cdot x - \vec{\nabla}G \tag{4.23}$$

Note that the time scale τ has absorbed all constant prefactors of the effective Hamiltonian in Equation 4.19. Applying the $\boldsymbol{L}$ operator from the left to Equation 4.22 gives

$$p = -\boldsymbol{L}\cdot\dot{x} \tag{4.24}$$

Replacing p from Equation 4.24 into equation 4.23 and commuting the differentials yields

$$-\boldsymbol{L}\cdot\frac{\partial^2 x}{\partial\tau^2} = \boldsymbol{L}\cdot x - \vec{\nabla}G \tag{4.25}$$

Finally, application of the $\boldsymbol{L}^{-1}$ operator from the left gives the final second order coupled differential equation of motion.

$$\frac{\partial^2 x}{\partial \tau^2} = -x + \boldsymbol{L}^{-1} \cdot \vec{\nabla} G \tag{4.26}$$

Integrating Forward in Path Time τ

These coupled first order equations are numerically integrated using a particular splitting[13] of the Liouville time evolution operator[36]

$$v = \frac{\partial x}{\partial \tau} \tag{4.27}$$

$$\frac{\partial v}{\partial \tau} = -x + \boldsymbol{L}^{-1} \cdot \vec{\nabla} G \tag{4.28}$$

First Half Step

$$w_0 = v_0 + \frac{\eta}{2} \boldsymbol{L}^{-1} \cdot \vec{\nabla} G \tag{4.29}$$

Full Step

$$\frac{\partial x}{\partial \tau} = \; v \tag{4.30}$$

$$\frac{\partial v}{\partial \tau} = -x \tag{4.31}$$

Exact integration over a time η gives

$$\begin{pmatrix} x_1 \\ w_1 \end{pmatrix} = \begin{pmatrix} \cos\eta & \sin\eta \\ -\sin\eta & \cos\eta \end{pmatrix} \begin{pmatrix} x_0 \\ w_0 \end{pmatrix} \tag{4.32}$$

Crank-Nicholson[37] applied to these coupled differential equations gives

$$\frac{x_1 - x_0}{\eta} = \frac{w_0 + w_1}{2} \qquad \text{and} \qquad \frac{w_1 - w_0}{\eta} = -\frac{x_0 + x_1}{2} \tag{4.33}$$

whose solution is

$$\begin{pmatrix} x_1 \\ w_1 \end{pmatrix} = \begin{pmatrix} \frac{4-\eta^2}{4+\eta^2} & \frac{4\eta}{4+\eta^2} \\ -\frac{4\eta}{4+\eta^2} & \frac{4-\eta^2}{4+\eta^2} \end{pmatrix} \begin{pmatrix} x_0 \\ w_0 \end{pmatrix} \tag{4.34}$$

Thus, $\cos\theta = \cos\eta$ or $\cos\theta = \frac{4-\eta^2}{4+\eta^2}$ and correspondingly $\sin\theta = \sin\eta$ or $\sin\theta = \frac{4\ \eta}{4+\eta^2}$. The full step then corresponds to mixing Brownian Bridges.

Second Half Step

$$v_1 = w_0 + \frac{\eta}{2}\ \boldsymbol{L}^{-1} \cdot \vec{\nabla} G(x_1) \tag{4.35}$$

Combining the Integrations

The above can now be combined to form the full step generated by the leapfrog integration. To perform this calculation, the initial path (x_0), which has the correct thermal fluctuation is required as input. In order to preserve the boundary conditions, the auxiliary velocities (v_0) are required to have zero velocity at the boundaries, and thus forms a Brownian Bridge. The next bridge path is given by

$$x_1 = \sin\theta\ \frac{\eta}{2}\ \boldsymbol{L}^{-1} \cdot \vec{\nabla} G(x_0) + \sin\theta\ v_0 + \cos\theta\ x_0 \tag{4.36}$$

In the above equation, take note of two points. The first is that x_0 is a Brownian path and v_0 is a Brownian Bridge. Each has the appropriate thermal fluctuation corresponding to the temperature (via the quadratic variation sum rule). This calculation is therefore simply a mixture of Brownian Bridges in the limit when the potential vanishes. This mixture (x_1) will have the correct quadratic variation if both x_0 and v_0 do, as $\sin^2\theta + \cos^2\theta = 1$ from the rotation in the full step. The second important note is that the boundary conditions require $\frac{\eta}{2}\sin\theta + \cos\theta = 1$. The Crank-Nicholson prescription for the value of $\sin\theta$ and $\cos\theta$ satisfy the boundary conditions, while the exact rotation only satisfies the condition when the paths begin and end at zero.

Generating the Hamiltonian Flow

Leveraging the reversibility of the Hamiltonian flow (see Appendix 2.A) one creates an equation which expresses the path x_0 in terms of the path x_1 and the path velocities v_1. In this equation,

the indices are switched and the momentum (velocity) is reversed as is required by Hamiltonian dynamics.

$$x_0 = -\sin\theta\ \frac{\eta}{2}\ \boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1) - \sin\theta\ v_1 + \cos\theta\ x_1 \tag{4.37}$$

Performing the leapfrog integration on the path x_1 generates the second path in the sequence, x_2. The useful part of this equation is that it may be rewritten in a form which does not involve any velocities.

$$x_2 = \sin\theta\ \frac{\eta}{2}\ \boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1) + \sin\theta\ v_1 + \cos\theta\ x_1 = (x_1 - x_0) + (2\cos\theta - 1)\, x_1 \tag{4.38}$$

This equation is then repeated many times to perform the required number of Molecular Dynamics (N_{MD}) steps for simulation parameters, the choice of which will be discussed in Section 4.5.3.

4.4 Calculating the Integration Errors

The Metropolis-Hastings algorithm is employed in order to be sure that the posterior probability is conserved under the above integration. This requires the calculation of the total drift in Λ between any two paths. In order to make the computation efficient, it is important that this calculation avoid requiring knowledge of any of the auxiliary velocities $(v_0, v_1, ...)$. The manipulations below show how to simplify this drift into only a function of the positional paths.

In order to shorten the equations, $\delta\Lambda$ will be separated into two parts, the first involving only the position and velocities and the second only involving the path potential G.

$$\delta\Lambda = \delta\Lambda^{(1)} + \delta\Lambda^{(2)} \tag{4.39}$$

The first part of the error in Λ is

$$\delta\Lambda^{(1)} = -\frac{1}{2\,\epsilon}\int_0^T dt\ \left(\frac{1}{2}x_1\cdot\boldsymbol{L}\cdot x_1 + \frac{1}{2}v_1\cdot\boldsymbol{L}\cdot v_1\right) + \frac{1}{2\,\epsilon}\int_0^T dt\ \left(\frac{1}{2}x_0\cdot\boldsymbol{L}\cdot x_0 + \frac{1}{2}v_0\cdot\boldsymbol{L}\cdot v_0\right) \tag{4.40}$$

By using the definitions of w_0 and w_1 from the two half steps in the leapfrog integration (Equations

4.29 and 4.29)

$$v_0 = w_0 - \frac{\eta}{2}\,\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_0) \qquad\qquad v_1 = w_1 + \frac{\eta}{2}\,\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1)$$

the drift can be rewritten as

$$\begin{aligned}\delta\Lambda^{(1)} =& -\frac{1}{2\,\epsilon}\int_0^T dt\,\left(\frac{1}{2}x_1\cdot\boldsymbol{L}\cdot x_1+\frac{1}{2}w_1\cdot\boldsymbol{L}\cdot w_1\right)\\ &+\frac{1}{2\,\epsilon}\int_0^T dt\,\left(\frac{1}{2}x_0\cdot\boldsymbol{L}\cdot x_0+\frac{1}{2}w_0\cdot\boldsymbol{L}\cdot w_0\right)\\ &-\frac{1}{2\,\epsilon}\int_0^T dt\,\left(\frac{\eta}{2}\,w_1\cdot\vec{\nabla}G(x_0)\,+\,\frac{\eta}{2}\,w_2\cdot\vec{\nabla}G(x_1)\right)\\ &+\frac{1}{2\,\epsilon}\int_0^T dt\,\left(\frac{\eta}{2}\vec{\nabla}G(x_0)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_0)\,-\,\frac{\eta}{2}\vec{\nabla}G(x_1)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1)\right)\end{aligned} \tag{4.41}$$

The first 2 terms vanish since (x_0, w_0) and (x_1, w_1) are related by an unitary transformation (a rotation).

$$\begin{aligned}\delta\Lambda^{(1)} =& -\frac{1}{2\,\epsilon}\int_0^T dt\,\left(\frac{\eta}{2}\,w_1\cdot\vec{\nabla}G(x_0)\,+\,\frac{\eta}{2}\,w_2\cdot\vec{\nabla}G(x_1)\right)\\ &+\frac{1}{2\,\epsilon}\int_0^T dt\,\left(\frac{\eta}{2}\vec{\nabla}G(x_0)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_0)\,-\,\frac{\eta}{2}\vec{\nabla}G(x_1)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1)\right)\end{aligned} \tag{4.42}$$

In order to remove the dependence of calculation of the velocities at the half step the following relations are employed

$$w_0 = -x_0\cot\theta\,+\,x_1\,\csc\theta \qquad\qquad w_1 = -x_0\csc\theta\,+\,x_1\,\cot\theta \tag{4.43}$$

This yields the final expression for the drift in $\Lambda^{(1)}$

$$\begin{aligned}\delta\Lambda^{(1)} =& \frac{\eta}{4\,\epsilon}\int_0^T dt\,\left((x_0\cot\theta\,-\,x_1\,\csc\theta)\cdot\vec{\nabla}G(x_0)\right)\\ &+\frac{\eta}{4\,\epsilon}\int_0^T dt\,\left((x_0\csc\theta\,-\,x_1\,\cot\theta)\cdot\vec{\nabla}G(x_1)\right)\\ &+\frac{\eta}{4\,\epsilon}\int_0^T dt\,\left(\vec{\nabla}G(x_0)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_0)\,-\,\vec{\nabla}G(x_1)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1)\right)\end{aligned} \tag{4.44}$$

For the Crank-Nicholson integrator, $\cot\theta = \frac{4-\eta^2}{4\ \eta}$ and correspondingly $\csc\theta = \frac{4+\eta^2}{4\ \eta}$. It is important to note that the computational effort required to calculate this drift term is significantly reduced

by performing the above simplification. Each of the $\cot\theta$ and $\csc\theta$ terms can simply be calculated, (each are a constant for the simulation) and then the multiplication of paths are all that is necessary to calculate the drift.

The second contribution to the drift can simply be calculated along the path in its current form

$$\delta\Lambda^{(2)} = -\frac{1}{2\,\epsilon}\int_0^T dt\,\Big(G(x_1) - G(x_0)\Big) \tag{4.45}$$

Overall Drift in Λ Between Paths x_0 and x_1

The final form of the drift equation looks rather complicated but calculation is straightforward.

$$\begin{aligned}\delta\Lambda =& \frac{\eta}{4\,\epsilon}\int_0^T dt\,\Big((x_0\cot\theta\;-\;x_1\;\csc\theta)\cdot\vec{\nabla}G(x_0)\Big)\\ &+\frac{\eta}{4\,\epsilon}\int_0^T dt\,\Big((x_0\csc\theta\;-\;x_1\;\cot\theta)\cdot\vec{\nabla}G(x_1)\Big)\\ &+\frac{\eta}{4\,\epsilon}\int_0^T dt\,\Big(\vec{\nabla}G(x_0)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_0)\;-\;\vec{\nabla}G(x_1)\cdot\boldsymbol{L}^{-1}\cdot\vec{\nabla}G(x_1)\Big)\\ &-\;\frac{1}{2\,\epsilon}\int_0^T dt\,\Big(G(x_1)-G(x_0)\Big)\end{aligned} \tag{4.46}$$

4.5 Numerical Considerations

4.5.1 Crank Nicholson

The PSHMC algorithm uses Hamilton's equations to define two first order Partial Differential Equations (PDE). In order to guarantee the convergence of these PDEs, I use the Crank-Nicholson method[37]. This method uses the finite difference method for the time derivatives and the midpoint method for the spacial derivatives. The advantage of using Crank-Nicholson is that the method is unconditionally stable for any choice of constants in the PDE. The other advantage of using the Crank-Nicholson method appears when applying boundary conditions (BCs) to a solution of a PDE as was discussed directly following Equation 4.36

4.5.2 Implementation of the Algorithm

It is now instructive to give an overview of a possible way to design algorithm in order to efficiently sample paths using this method. This section will not cover the actual code which was

used, but does require a language with pointers to operate efficiently. All notation in this section will follow coding paradigms of the C programming language.

By far the most expensive computational piece of the algorithm is a calculation of the forces (F), the path potential (G) and its derivatives (∇G). An appropriate data structure would therefore have the following vectors all of $N = T/\Delta t$ length.

Program 1: Data structure used in the implementation of the PSHMC algorithm in `dim` dimensional space. A structure of this type is used for each particle in the system.

```
position[dim]  \\ stores positions along the path (vector quantity)
V              \\ potential along the path
G              \\ path potential along the path
gradG[dim]     \\ gradient of the path potential (vector quantity)
LinverseG[dim] \\ Linverse of the path potential (vector quantity)
```

The calculation of each of these arrays constitutes the majority of the computational time. Note that the algorithm used to compute `LinverseG` is reviewed in Appendix 4.D.

Fortunately, the expense of calculating this structure is counterbalanced by the perfectly parallel nature of the force calculations (all but the $\boldsymbol{L}^{-1}\nabla G$ vector, which is reviewed in Appendix 4.D). Each of the time steps along the path (Δt) are completely independent from their neighbors when calculating these potentials and fields. The vast majority of the calculation can therefore be parallelized on distributed memory by splitting the path into N pieces and distributing them across N nodes.

There are a total of three complete paths required to implement the PSHMC algorithm, each with the above structure for every particle, which will be called `pathOld`, `pathCurrent`, and `pathNew`. The first step of the algorithm reads the positions (from disk or from memory) to `pathOld.position` and proceeds with calculation of the rest of the path potential calculations to fill the structure in Program 1. The second step then uses a Brownian Bridge (v_0) to calculate the updated path positions (x_1) in Equation 4.36 which are stored in `pathCurrent.position` followed by filling the `pathCurrent` structure. This step is finished with a calculation of the energy drift between paths x_1 and x_0 in the effective path Hamiltonian, as stated in Equation 4.46.

At this point the Stochastic part of the algorithm is complete, and the Hamiltonian flow is calculated a large number of times. The path positions (x_2) are calculated according to Equation 4.38, stored to `pathNew.positions` and the structure is filled. Again, the energy drift is calculated

according to 4.46.

The important point to note at this point is the `pathOld` structure is not required for future calculation. In order to avoid copying large amounts of memory, the structure pointers are rotated (`pathCurrent` $\rightarrow$ `pathOld`, `pathNew` $\rightarrow$ `pathCurrent`, `pathOld` $\rightarrow$ `pathNew`). The memory space can now be used to store the next step in the Hamiltonian Flow.

4.5.3 Tuning the PSHMC Method

The HMC algorithm may also be ergodic which means the algorithm will explore all of the available state space (given enough time). This will fail to hold if $N_{MD} * \eta \approx 2\pi$ as the positions will return to approximately the same configurations. This occurs because of the exact rotation of Equation 4.32.

This exact integration helps determine the parameters for optimal sampling rate. In order to obtain the maximal movement away from the initial state the number of steps of Hamiltonian dynamics (N_{MD}), and the algorithmic step size ($\Delta\tau = \eta^2/2$) must be tuned. $\Delta\tau$ is chosen to give the largest appropriate acceptance rate to the MHMC step, and is tuned to give an acceptance rate between 50% and 90%. This time step, $\eta = \sqrt{2\,\Delta\tau}$, multiplied by the number of Hamiltonian dynamics steps, N_{MD}, should be on the order of π. This choice reflects the fact that the positions and auxiliary variables resemble a harmonic oscillator which has a period of 2π and thus a large move corresponds to an integration over approximately half of the period. However, one should not use a fixed N_{MD} steps, as this may enhance unphysical resonances[32].

The step size Δt is chosen so the noise will dominate over the potential term ($\langle G \rangle$). We have chosen the noise to be 100 times larger than the potential term.

$$\Delta t = -\frac{100\langle G \rangle}{\varepsilon} \tag{4.47}$$

where the average $\langle G \rangle$ is over the Boltzmann distribution. This calculation is shown in detail in Appendix 4.A

4.5.4 Consistency Check

A consistency check lies in the relationship between the equilibrium averages of the components of G. This will provide some information about the (time) length of our paths, i.e. whether our choice of T is appropriate.

$$\langle |\nabla V(x)^2| \rangle = \epsilon \, \langle \nabla^2 V(x) \rangle = -2\langle G \rangle. \tag{4.48}$$

The details of this calculation are shown in Appendix 4.C

Appendix 4

4.A Quadratic Variation and Path Length

The correct thermal noise along the path is governed by the quadratic variation. From Brownian dynamics

$$\frac{\Delta x}{\Delta t} = F + \sqrt{\frac{2\epsilon}{\Delta t}}\,\xi_i \tag{4.49}$$

On average, the force (F) is zero and the expectation of $\xi_i^2 = 1$.

$$\sum_i \left(\frac{\Delta x}{\Delta t}\right)^2 = \sum_i \frac{2\epsilon}{\Delta t}\,\xi_i^2 \tag{4.50}$$

$$\sum_i (x_i - x_{i-1})^2 = 2\,\epsilon\,T \tag{4.51}$$

I use this relationship to calculate an estimate for the length of the path, T. The underlying dynamics should be dominated by the random fluctuations and not the path potential. Separating the path integral into the fluctuation part and the path potential part, and taking $\tilde{G} = \max(G)$ along the path yields

$$I_1 = \int dt \sum_{i,\alpha} \frac{1}{2}\left(\frac{dx}{dt}\right)^2 = \frac{1}{2}\frac{N_D N_p}{\Delta t}\,2\,\epsilon\,T \tag{4.52}$$

$$I_2 \approx \int dt\ \tilde{G} = \tilde{G}T \tag{4.53}$$

where N_D is spacial dimensions and N_p is the number of particles. Now for good statistics I impose

$I_1 = 100 I_2$, giving an estimate for an appropriate upper limit for Δt

$$\Delta t = \frac{N_D N_p \epsilon}{100 \tilde{G}} \tag{4.54}$$

4.B Heuristic Derivation of the Path Potential G

It is instructive to show how the path potential can be derived using the Stratonovich form of stochastic calculus. The form of the path potential, G, is a result of using the Onsager-Machlup functional. The discrete Brownian equation of motion is given as

$$\frac{\Delta x}{\Delta t} = F + \sqrt{\frac{2\epsilon}{\Delta t}}\, \xi_i \tag{4.55}$$

The Stratonovich form of stochastic calculus requires the evaluation at a half step both forward and backward and average the two

$$\text{Forward:} \qquad \frac{\epsilon}{2} \sum \xi_{i,\rightarrow}^2 = \sum_i \frac{\Delta t}{4} \left[\left(\frac{\Delta x}{\Delta t} \right)^2 + |F_i|^2 - 2F_i \frac{x_{i+1} - x_i}{\Delta t} \right] \tag{4.56}$$

$$\text{Backward:} \qquad \frac{\epsilon}{2} \sum \xi_{i,\leftarrow}^2 = \sum_i \frac{\Delta t}{4} \left[\left(\frac{\Delta x}{\Delta t} \right)^2 + |F_{i+1}|^2 - 2F_{i+1} \frac{x_i - x_{i+1}}{\Delta t} \right] \tag{4.57}$$

$$\text{Average:} \qquad \frac{\epsilon}{4} \sum \left(\xi_{i,\rightarrow}^2 + \xi_{i,\leftarrow}^2 \right) = \sum_i \frac{\Delta t}{4} \left[\left(\frac{\Delta x}{\Delta t} \right)^2 + F^2 + F' \frac{\Delta x^2}{\Delta t} \right] \tag{4.58}$$

With the substitutions: $F' = -\nabla^2 V$ and $\frac{\Delta x^2}{\Delta t} = 2\epsilon$ the path action is found with the path potential, G,

$$\frac{\epsilon}{2} \sum \left(\xi_{i,\rightarrow}^2 + \xi_{i,\leftarrow}^2 \right) = \sum \Delta t \left[\frac{1}{2} \left(\frac{\partial x}{\partial t} \right)^2 + \underbrace{\frac{1}{2} |F|^2 - \epsilon \nabla^2 V}_{G} \right] \tag{4.59}$$

4.C Consistency Check

The equilibrium Average of G can be calculated for a system in i dimensions with α particles

$$G = \sum_{i,\alpha} \left\{ \frac{1}{2} \left(\frac{\partial V}{\partial x_{i\alpha}} \right)^2 - \epsilon \left(\frac{\partial^2 V}{\partial x_{i\alpha}^2} \right) \right\} \tag{4.60}$$

The expected value is the calculated using the Boltzmann distribution as the probability measure

$$\left\langle \left(\frac{\partial V}{\partial x_{i\alpha}} \right)^2 \right\rangle = \frac{1}{Z} \int dx^{dN} \left(\frac{\partial V}{\partial x_{i\alpha}} \right)^2 \exp\left(-\frac{V}{\epsilon} \right) \tag{4.61}$$

Using integration by parts

$$u = -\exp\left(-\frac{V}{\epsilon} \right) \qquad \text{then} \qquad du = \frac{dx}{\epsilon} \frac{\partial V}{\partial x_{i\alpha}} \exp\left(-\frac{V}{\epsilon} \right).$$

$$v = \frac{\partial V}{\partial x_{i\alpha}} \qquad \text{then} \qquad dv = \frac{\partial^2 V}{\partial x_{i\alpha}^2}\, dx.$$

$$\left\langle \left(\frac{\partial V}{\partial x_{i\alpha}} \right)^2 \right\rangle = \frac{\epsilon}{Z} \int v\, dt = -\frac{\epsilon}{Z} \int u\, dv = \epsilon \left\langle \frac{\partial^2 V}{\partial x_{i\alpha}^2} \right\rangle \tag{4.62}$$

These expectations then give a value for the equilibrium value of the path potential G as

$$\langle\, G \,\rangle = -\frac{1}{2} \sum_{i,\alpha} \left\langle \left(\frac{\partial V}{\partial x_{i\alpha}} \right)^2 \right\rangle = -\frac{\epsilon}{2} \sum_{i,\alpha} \left\langle \left(\frac{\partial^2 V}{\partial x_{i\alpha}^2} \right) \right\rangle \tag{4.63}$$

This can be thought of as an analog of the Virial theorem in classical dynamics.

4.D Numerical Algorithm for L^{-1}

I present an algorithm to perform Gaussian elimination on a statically defined tridiagonal matrix and a vector, $\vec{x}$. The operator $\boldsymbol{L}$ performs the following operation:

$$\boldsymbol{L} = \frac{d^2}{dt^2} \tag{4.64}$$

This operator can be described by a tridiagonal matrix with -2 on the main diagonal and 1 on

the off diagonals (boundary conditions excluded). We want to perform Gaussian elimination on this matrix with $\vec{x}$ transformed to a new vector $\vec{g}$ that will be used to form the solution to the Gaussian elimination. $\vec{x}$ and $\vec{g}$ are of length $N+2$, and are uniformly spaced in time, with interval dt, and possess boundary conditions $x_0 = g_0 = x^-$ and $x_{N+1} = g_{N+1} = x^+$.

$$\begin{pmatrix} -2 & 1 & 0 & 0 & 0 & 0 & x_1 \\ 1 & -2 & 1 & 0 & 0 & 0 & x_2 \\ 0 & 1 & -2 & \ddots & 0 & 0 & x_3 \\ 0 & 0 & \ddots & \ddots & 1 & 0 & x_4 \\ 0 & 0 & 0 & 1 & -2 & 1 & \vdots \\ 0 & 0 & 0 & 0 & 1 & -2 & x_N \end{pmatrix} \Rightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & g_1\,dt^3 - x^- \\ 0 & 1 & 0 & 0 & 0 & 0 & g_2\,dt^2 \\ 0 & 0 & 1 & 0 & 0 & 0 & g_3\,dt^2 \\ 0 & 0 & 0 & \ddots & 0 & 0 & g_4\,dt^2 \\ 0 & 0 & 0 & 0 & 1 & 0 & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 & g_N\,dt^2 - x^+ \end{pmatrix} \tag{4.65}$$

The matrix of interest is of size (N x N) and the vector contains N elements (boundaries are dropped). This problem is easily solved using simple Gaussian elimination and back substitution but the algorithm is hard to parallelize efficiently as it is an iterative routine (calculation of the next step relies upon knowing the value of the current step).

Before looking at the discrete version, it is useful to look at the analytic, continuum, version.

4.D.1 Algebraic Solution

Consider the following equation:

$$\frac{d^2x}{dt^2} = g(t) \tag{4.66}$$

with $x(0) = x^-$ and $x(T) = x^+$. Using the operator $\boldsymbol{L}$ defined above, we can write the matrix equivalent as

$$\boldsymbol{L} \cdot x = g \qquad \text{or} \qquad x = \boldsymbol{L}^{-1} \cdot g \tag{4.67}$$

where $\boldsymbol{L}$ is the finite representation of the second derivative operator, $g(t)$ is a known function (at least on a uniform grid) and $x(t)$ is to be determined. The boundary conditions for x are known (see above). This differential equation is separable and can be solved by integration. Furthermore, the algorithm can be simplified by changing the order of integration of the double integral.

Consider the functions:

$$I_0(t) = \int_0^t g(w)\,dw \qquad\qquad I_1(t) = \int_0^t w\,g(w)\,dw \tag{4.68}$$

The general form of $\vec{x}$ is given by the following integral (α and β are determined by BCs)

$$\begin{aligned} x(t) &= \int_0^t dv \int_0^v g(w)dw + \alpha + \beta t && (4.69)\\ &= \int_0^t dw \int_w^v g(w)dv + \alpha + \beta t && (4.70)\\ &= \int_0^t (w-t)g(w)dw + \alpha + \beta t && (4.71)\\ &= I_1(t) - tI_0(t) + \alpha + \beta t && (4.72) \end{aligned}$$

Now matching boundary conditions we have

$$x(0) = x^- = I_1(0) + \alpha \quad \Rightarrow \quad \alpha = x^- \tag{4.73}$$

$$x(T) = x^+ = I_1(T) - TI_0(T) + \alpha + \beta T \quad \Rightarrow \quad \beta = I_0(T) + \frac{1}{T}\left[x^+ - x^- - I_1(T)\right] \tag{4.74}$$

Giving the final $x(t)$ as

$$x(t) = x^- - I_1(t) + \frac{t}{T}\Big(I_1(T) \;+\; x^+ \;-\; x^-\Big) + t\,\Big(I_0(t) - I_0(T)\Big) \tag{4.75}$$

Then we can verify that the solution explicitly

$$x'(t) = -t\,g(t) + \Big(I_1(T) \;+\; x^+ \;-\; x^-\Big)\frac{1}{T} + t\,g(t) + \Big(I_0(t) - I_0(T)\Big) \tag{4.76}$$

$$\boldsymbol{L}\cdot x = x''(t) = g(t) \tag{4.77}$$

In addition, we see that the boundary conditions are satisfied.

Program 2: Example code written in the *Mathematica* programming language which shows the analytic and discrete algorithm for calculating L^{-1}.

```
(*----------- analytic algorithm ------------*)
(* Mathematica code to check analytic solution*)
I0[u_]:=Integrate[g[w],{w,0,u}]
I1[u_]:=Integrate[w g[w],{w,0,u}]
Linvg[u_]:=xm-I1[u]+(I1[U]+xp-xm) (u/U)+(I0[u]-I0[U]) u
Linvg=Collect[Linvg[u],{1/U,u}]
xm==Linvg/.u->0//Expand//FullSimplify
xp==Linvg/.u->U//Expand//FullSimplify
g[u]==D[Linvg,{u,2}]//Expand

(*----------- discrete algorithm ------------*)
(* test the implementation of the above method on a finite grid *)

(* Total number of beads *)
Numbeads=8;
(*size of vectors and matrices; exludes BCs*)
Num=Numbeads-2;
(* generate a general vector g with the step size squared included *)
g=dt^2*Table[ToExpression["g"<>ToString[i]],{i,1,Num}];
(* initialize vectors for i0 and i1 *)
i0=Table[0,{i,1,Num}];
i1=Table[0,{i,1,Num}];
(*set initial conditions *)
i0[[1]]=i1[[1]]=g[[1]];
(* calculate i0 and i1 up to step u for every u *)
Do[
  i0[[n]]=i0[[n-1]]+g[[n]];
  i1[[n]]=i1[[n-1]]+n g[[n]];
,{n,2,Num}]
(*evaluation of the BCs*)
lasti=i0[[Num]]-(xp-xm+i1[[Num]])/(Num+1);
(*form solution vector*)
x=Table[xm+n (i0[[n]]-lasti)-i1[[n]],{n,1,Num}]//Expand//Simplify;
(* add the BCs to the beginning and end of the solution vector *)
AppendTo[x,xp];PrependTo[x,xm];
(* print the solution *)
x//Expand//MatrixForm

(*Check that L x=g*)

Table[x[[i+1]]-2 x[[i]]+x[[i-1]]==g[[i-1]],{i,2,Numbeads-1}]//Expand//Simplify
```

4.D.2 Parallelization of L^{-1} on Distributed Memory

N_{nodes} = Total number of compute nodes. u^- is the beginning bead on the node. u^+ is the ending bead on the node.

The integrals I_0 and I_1 can be calculated on a distributed memory system with a bit of care. The integral J_j^α needs to be calculated on each node (named α) with a subset of sequential beads.

$$J_j^\alpha(u) = \int_{u^-}^{u} w^j g(w) dw \tag{4.78}$$

This result is then broadcast to all of the other nodes and a barrier is placed on the computation until the broadcast is complete. Upon completion of the broadcast each node can calculate the integral I_j^α and `lasti`

$$I_j^\alpha(u) = \sum_{\beta<\alpha} J_j^\beta(u^+) + J_j^\alpha(u) \tag{4.79}$$

$$\texttt{lasti} = \frac{\sum_{\alpha=1}^{N_{nodes}} J_0^\alpha(u^+) - (x^+ - x^- + \sum_{\alpha=1}^{N_{nodes}} J_1^\alpha(u^+)}{N_{beads} - 1} \tag{4.80}$$

The solution on the α^{th} node is then

```
x[n_] = Table[xm+n(I0[[n]] - lasti)-I_i[[n]],{n,1,Nbead-2}]
```

The only part of the calculation that must be shared to all nodes is the end point of the integration of J_j^α for each node. The endpoints can be broadcast across all nodes and then `lasti` can be calculated. There are two separate sums that must be computed from the shared Js. All nodes must calculate the sum of J_0 and J_1 from all other nodes. This will give you $I_0(U)$ and $I_1(U)$ that is needed for the calculation of `lasti`.

Chapter 5

The Lennard-Jones System

The previous chapters, I have introduced the concept of a path and developed a possible way to sample double-ended paths for a continuous-time process. In this chapter, I present results from using this algorithm to sample paths in systems which possess an interesting conformational transition. These systems are purely theoretical in nature and were chosen to test the path sampling method, rather than to gain new physical insight of specific physical properties of the systems themselves.

A similar Hybrid Monte Carlo (HMC) algorithm has been implemented in a study of polymers[38] where H was the actual Hamiltonian and where the sums in H are finite (corresponding to the number of atoms in the polymer). Similarly, one can sample paths from the corresponding measure: $\mathbb{P}_{path} \propto \exp\left[-(H/2\epsilon)\right]$.

The system I will consider is a cluster of particles in three dimensions governed by the Lennard-Jones (LJ) potential. This is a historically well studied problem[39], but the relatively large number of degrees of freedom obscures all easily accessible physical properties of the system from being analyzed. In order to better understand the dynamics produced by the continuous-time HMC path sampling method, a simpler system is required. In a later chapter, I will present the study of a single particle fluctuating at thermal equilibrium in an externally imposed potential to answer some of the questions raised by this LJ study.

The interested reader might wonder why the discussion these physical systems proceeds by starting with the much more complicated (many degrees of freedom) Leonard Jones clusters, and then returns to the much simpler external potential in one dimension. The explanation of this

choice is somewhat historical, but will also be used to gain insight into the complex nature of sampling algorithms. Often, the dynamics of the systems studied is so complex that the underlying properties of the method are washed out by the complexity of the system. This is the case when applying the Path Space Hybrid Monte-Carlo (PSHMC) method to the Lennard Jones system as will be shown in this chapter.

5.1 Clusters of Lennard-Jones Particles

The original motivation for this work was to develop a novel sampling method which can be used to understand the protein folding problem. These systems exhibit collective motions of extremely large numbers of atoms, have relatively small transition rates, and exceedingly complex free energy landscapes. Sampling with a diffusive algorithm like the forward SDE or deterministic molecular dynamics is inefficient at sampling these complex systems. It is thus very beneficial to use any known constraints to guide the sampling. In many complex chemical and biological systems, it is possible to experimentally determine the initial system state (the amino acid sequence for the protein folding problem) and the question becomes how this special initial state transforms to a structure with a higher level function. In many cases, the final state of the folded protein is also long lived and can be experimentally measured. This provides two boundary conditions which, when applied to the continuous-time HMC method from chapter 4, can be used to force the transition to occur in some predefined time period, increasing the efficiency of identifying the interesting dynamics.

With this motivation in mind, one requires a problem that has a somewhat rigid set of conditions as a test for the proposed continuous-time HMC sampling method. First of all, in order for the results of this analysis to be meaningful to the physics community, the chosen system must be well studied and documented in the literature. In order to test code base, I required the system reside in a three dimensional space, yet it must also be simple enough for use as a testing system. Finally, as stated previously, the system must have predefined beginning and ending configurational states.

The system which I have chosen which satisfies all these requirements uses the well-known Lennard-Jones (LJ) potential as a self interaction potential between a cluster of particles. This

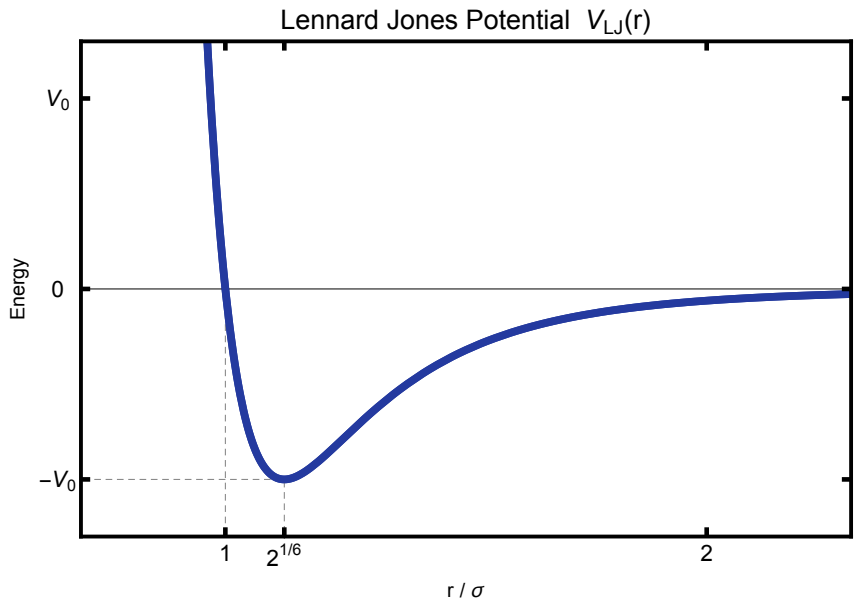


Figure 5.1: The Lennard-Jones potential in the radial direction. The $1/r^{12}$ term in this potential produces a very large repulsive barrier for positions $r < \sigma$, which leads to LJ particles having an extremely hard core. At long distances, the potential tends to zero for any choice of parameters σ and V_0

system is described with the following pairwise potential

$$U(r) = \sum_{i<j}^{N} V_{LJ}(r_{ij}) \tag{5.1}$$

where each degree of freedom follows the radial force defined from the potential

$$V_{LJ}(r_{ij}) = 4V_0 \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{5.2}$$

The ground state configurations of these LJ clusters were studied by Hoare and Pal [39] who found the optimal configurations for a given number of particles, N, in the cluster. For a small number of particles, these configurations are easily understood ($N = 3 \rightarrow$ equilateral triangle, $N = 4 \rightarrow$ tetrahedron) but for larger clusters, the ground state configurations become much more complicated. These previous studies have identified a set of particle numbers which exhibit very efficient packing geometry, which have been given the title of a *magic number* of particles. For these magic cluster numbers, the difference in average energy per particle between a cluster with a magic number particles and a cluster with one less (or more) particle is significantly larger than

the same difference for non-magic cluster numbers. This is a somewhat analogous statement to the magic numbers having a minima of chemical potential. A medium-size cluster with a magic number is N=13, and the geometry for the ground state is icosahedral (see Figure 5.2). For N=14, the ground state geometry is the same icosahedral structure with an added article residing on the surface. I will subsequently call these clusters LJ_{13} and LJ_{14}, respectively.

5.2 Transformation at Zero Temperature

I investigate a previously discovered low lying mode[19] in which the surface atom penetrates the cluster and pushes an internal atom out to the surface. The LJ_{13} cluster has previously been studied by Barry et al [19] using molecular dynamics techniques. This study discovered an interesting mode of transformation where single atom becomes excluded from the icosahedral structure to reside on the surface and a vacancy is concurrently created on the opposite side of the cluster. This collective motion occurs when the motion of 3 central atoms, linked together to form a *chain* of three particles (N_{in}) form the center of the cluster, coherently push through to the other side, leaving a vacancy and a surface atom on opposite ends. The surrounding shell of particles (N_{out}) move only a small amount between the initial (icosahedral) structure and the final excited configuration. Note the extra atom is at the bottom of the cluster initially, moves into the cluster and pushes the chain up through the cluster. This collective motion is shown graphically, where I have removed the thermal fluctuations for clarity, in figure 5.3. I will use these starting and ending configurations as boundary

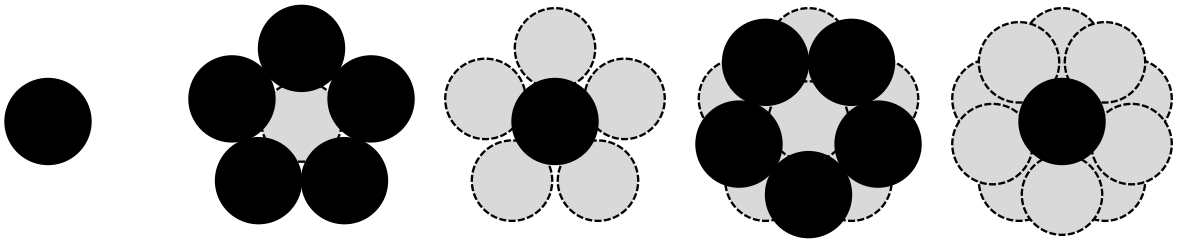


Figure 5.2: Construction of the ground state configuration of 13 Lennard Jones particles (LJ_{13}. The construction proceeds on successive levels starting with the leftmost panel of a single particle which is surrounded by five close-packed LJ particles in a plane. This structure then repeats itself, with the fourth level being a copy of the second level rotated by $2\pi/5$, and is finally topped off the single particle on the fifth level. The radial distance between the particles is determined by the parameters (l and ϵ) in the Lennard Jones potential (equation 5.1) such that the particles reside the minimum of $V(r)$.

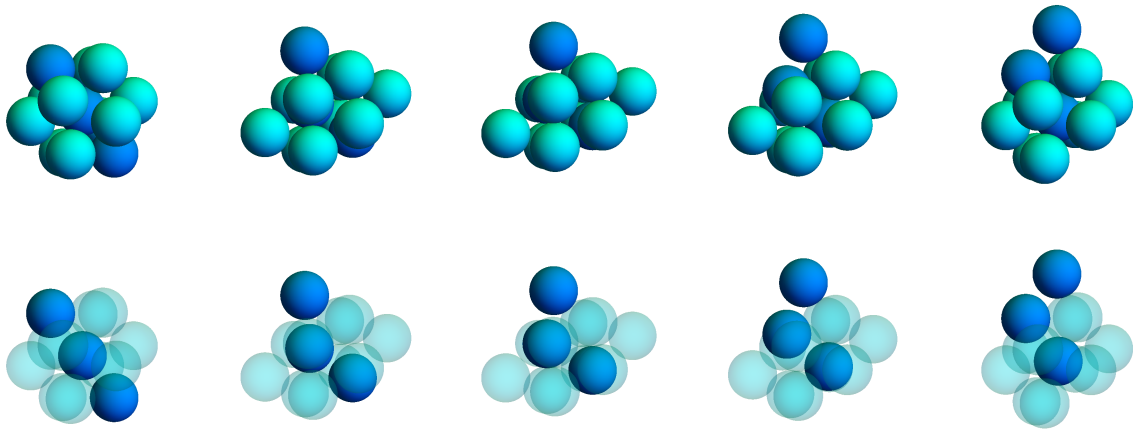


Figure 5.3: 3-Dimensional representation of the zero-temperature LJ_{13} cluster transformation. Five distinct states are observed in this transformation, from the global minimum energy configuration (left panel) to the excited minima with a surface excitation (right panel). The *outer shell* particles are shown in cyan and the *inner chain* particles are shown in blue. The bottom figures show the identical transformation with the outer shell particles transparent to better illustrate the chain movement. An animation of this motion is included in the supplementary material (Animation 5.1).

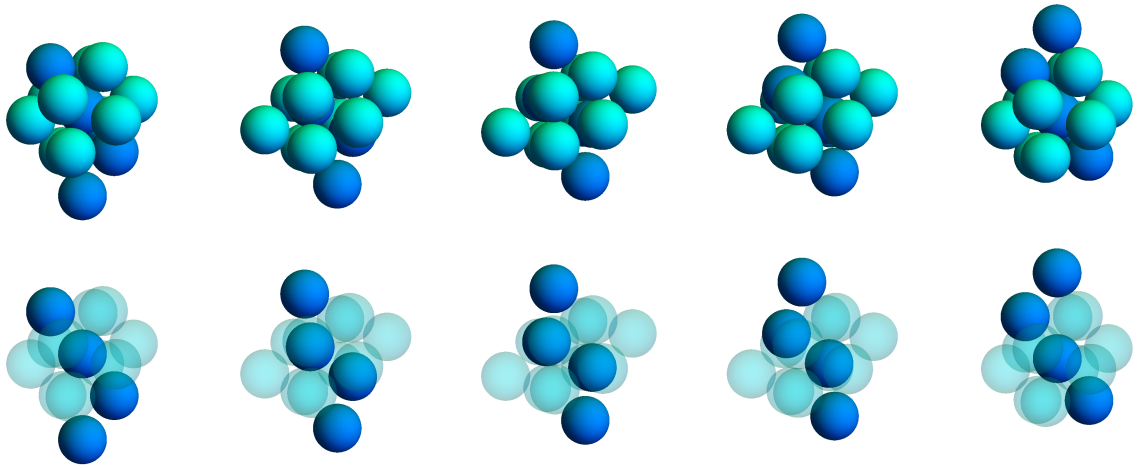


Figure 5.4: 3-Dimensional representation of the zero-temperature LJ_{14} cluster transformation. This transformation is similar to the transformation shown in Figure 5.3, with five states present, but is a very symmetric motion which starts and ends with identical (albiet rotated) configurations. The transformation proceeds with the bottom chain particle excluded from the LJ_{13} structure and ends with the top chain particle being excluded. An animation of this motion is included in the supplementary material (Animation 5.2).

conditions in the path sampling algorightm.

The zero-temperature path is obtained by minimizing the Large Deviation result, but this is

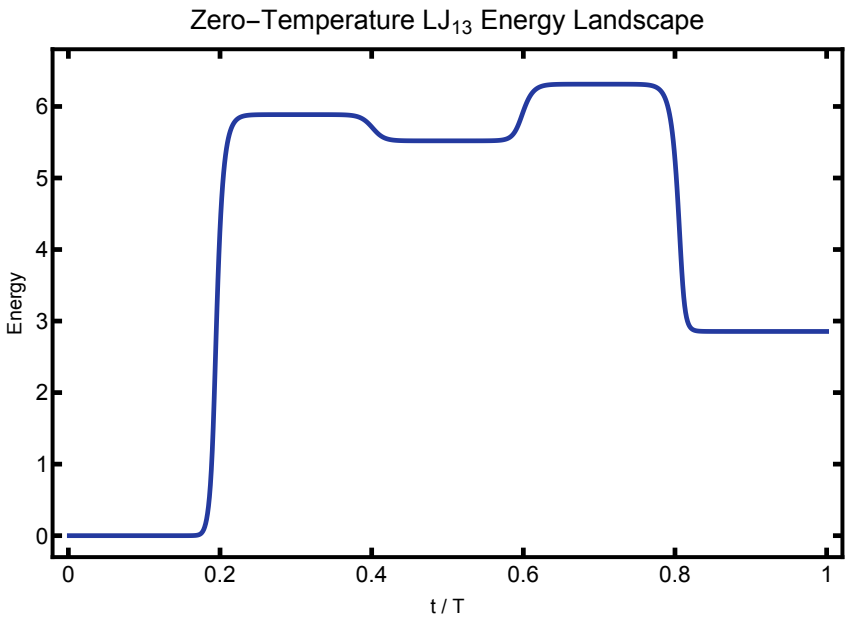


Figure 5.5: Energy of the zero-temperature transition pathway of the LJ_{13} cluster from the ground state structure (icosahedron) at $t = 0$ to the configuration with a vacancy and surface particle at $t = T$. The path shows five distinct configurations, each of which exist for a roughly equal length of time, at each of the critical points in this energy landscape. Note that the initial and final boundary configurations are of different energy.

unphysical. It is the thermal energy that propels the conformation change, but at zero temperature such energy is nonexistent. It is important to note for these zero-temperature configurations, all critical points of the conformational change have approximately equal path length (time). This can be understood realizing is equally hard to fall off of a peak in the potential as it is to fall up from the trough in the absence of all thermal fluctuation*.

5.2.1 Energy Landscape of the LJ clusters

The potential energy of the sequence of configurations along the path time, t, is the most obvious way to characterize the state of the cluster. This calculation can be performed for each configuration by calculating the pairwise potential given in equation 5.1. This energy along the path will then be used as a physical measure of the state of that transition. Both σ and V_0 have been set to unity and the ground state (icosahedral) configurational energy set to zero, for ease of calculation.

Both LJ_{13} and LJ_{14} exhibit 5 distinct intermediate stages that these critical points for the zero-

*In fact, it is *infinitely hard* in either case.

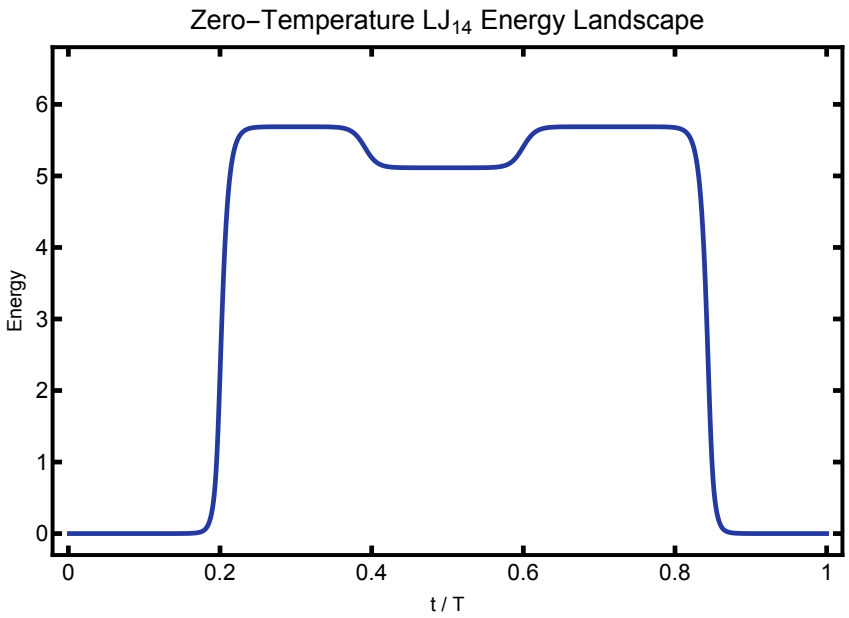


Figure 5.6: Energy of the zero-temperature transition pathway of the LJ_{14} cluster. The initial and final state are identical geometries, as are the two intermediate states, and are thus degenerate. Analogously to the LJ_{13} cluster, this path shows five distinct configurations at all critical points which exist for a roughly equal length of time.

temperature systems (refer to Figures 5.5 and 5.6). LJ_{13} has no energy degeneracy for any of these zero-temperature states as the transformation lacks symmetry between the forward and backward motion. LJ_{14} has two doubly degenerate energy states which arise from the highly symmetric transition and, importantly for the following analysis, the boundary conditions themselves are degenerate. This degeneracy will permit a further insight into what information can be gained from the continuous-time HMC transition rate. If the transition rate is accessible via the method, for this symmetric system, the transition should occur at half of the path length, as long as the path is long enough.

5.2.2 Reaction Coordinate for the LJ clusters

The complex movements of the transition of these LJ clusters in positional space is often not very illuminating. The movements of the particles are dominated by the seemingly random fluctuations while the interesting parts of the path remain hidden from view. For this reason it is convenient to define a single quantity that maps configuration along the path to the state of progression of the transition. This reaction coordinate, labeled ρ, varies from ρ^- at the start of the path to ρ^+ at the end of the path. Values between ρ^- and ρ^+ then provide a measure of the evolution of the

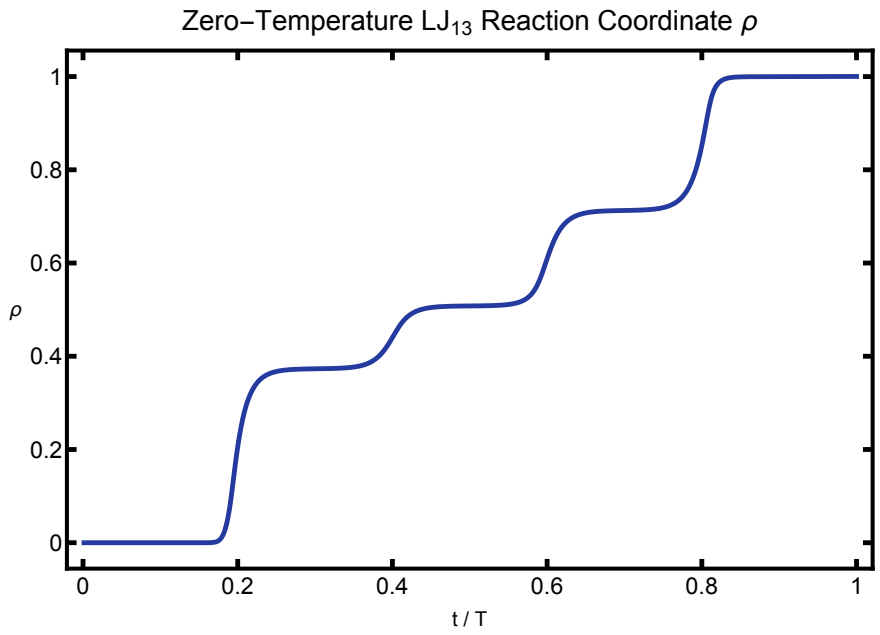


Figure 5.7: The reaction coordinate of the zero-temperature LJ_{13} transition pathway which is used to quantify the progression of the transition between the two boundary configurations. Larger values of ρ correspond to a larger progression through the transition pathway. There is a corresponding plateau for each of the critical points seen in the energy landscape in Figure 5.5.

transition. This reaction coordinate is defined to be

$$\rho = \left(\vec{R}_{in} - \vec{R}_{out}\right) \cdot \hat{n} \tag{5.3}$$

where $\vec{R}_{in} - \vec{R}_{out}$ is the difference in the center of mass of the outside shell of LJ particles (10 atoms, excluding the chain particles) to the center of mass of the chain (4 particles for LJ_{14} and 3 for LJ_{13}). The unit vector $\hat{n}$ is the eigenvector corresponding to the minimum eigenvalue of the moment of inertia tensor of the chain of particles. This vector roughly points in the direction of the movement of the chain as it traverses through the shell. Figures 5.7 and 5.8 illustrate the behavior of ρ for the zero-temperature LJ_{13} and LJ_{14} clusters respectively. The coordinate shows distinct plateaus for each of the critical points of the zero-temperature configuration.

5.3 Path Evolution of the Thermally Excited LJ Clusters

Let us now take a step back and review what was achieved in the previous sections. In the previous section, I have developed the framework for a cluster of Lennard-Jones particles which

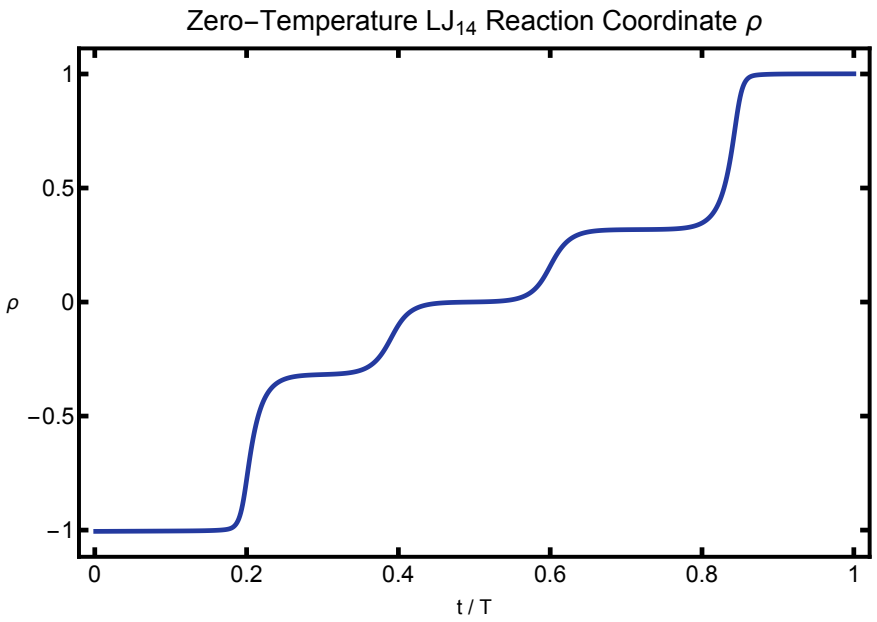


Figure 5.8: The reaction coordinate of the zero-temperature LJ_{14} transition pathway which uniquely identifies the cluster configurations even for states which are degenerate in energy. Larger values of ρ correspond to a larger progression through the transition pathway. Note the value of ρ is symmetric and varies between -1 and $+1$ for this cluster.

forms a three-dimensional system in the absence of thermal excitation. One of the manifestations of the nonphysical nature of the zero-temperature configurations is that the cluster spends the same order of time at each of the critical points, even at a saddle point. At finite temperature, as the path passes through the largest energy barrier, a physical path should quickly make the transition to the next valley; passing quickly through the saddle point. This is the standard instanton picture. The zero-temperature configurations are clearly nonphysical in the sense that the transition instanton is broken. The question now becomes how to proceed in performing the path sampling when the system has reached thermal equilibrium with its surrounding heat bath?

I've used the method described in chapter 4 to generate a sequence of paths, using the thermalized zero-temperature path above as a starting point. I have taken the smooth zero-temperature configuration and added noise to form a path which exhibits thermal fluctuations at a temperature of $\epsilon = 0.13$. The number of degrees of freedom is equal to special dimension ($d = 3$) multiplied by the number of particles in the cluster ($N = 13$ for LJ_{13} and $N = 14$ for LJ_{14}). The total time of each of the paths is $T = 10$ and the time step along the path between each configuration is $\Delta t = 0.0001$, which specifies each path to have $100,001$ configurations. The number of molecular dynamics steps per MHMC loop was chosen to be $N_{MD} = 3000$, in accordance with the guidelines

in section 4.5.3.

Using this physically naive starting path serves as a relevant real-world test of the dynamics of the PSHMC method. This path is very easily generated, but obviously non-physical because of the large amount of time spent in the high energy transition state. For more complex systems (folding of a protein), these physically naive guesses are often all that is available, so it is required that the sampling handle this input path correctly. If the sampling method is robust, the sequence of paths generated will evolve to the point where the instanton has healed. The convergence of the nonphysical starting path to a physical path is of critical importance because computer experimenters, in general, study problems which have unknown free energy barriers. This means choosing a realistic starting path is quite difficult. By starting the sampling with this naive guess, I have attempted to include such a difficulty.

As can be seen in Figures 5.9 and 5.10, adding noise to the smooth path completely washes out two of the three intermediate states from either of the zero-temperature paths. The result is a path which is described correctly by the Onsager-Machlup functional in Equation 3.6, but spends the vast majority of its time sitting on top of the high-energy barrier.

A closer examination of the reaction coordinate (equation 5.3) reveals more information about this intermediate high-energy state than can be inferred from the energy landscape alone. For the case of LJ_{13}, Figure 5.11 clearly shows the preferred configuration of this thermalized path is a combination of the second and third plateau from the zero-temperature path (Figure 5.8). The highest energy barrier of the zero-temperature reaction coordinate ($\rho \approx 0.7$) is only achieved once the cluster begins falling into the ending energy basin. This state is entropically stabilized. The reaction coordinate for the LJ_{14} cluster is similarly illuminating (Figure 5.12), but displays distinctly different behavior. The highly symmetric nature of the LJ_{14} structure causes the cluster to oscillate between any of the intermediate high energy states ($-0.3 < \rho < 0.3$).

As the calculation proceeds, large chunks of the path which reside on top of the high energy barrier (parts of the path where $0.3 < \rho < 0.8$ for LJ_{13} and $-0.4 < \rho < 0.4$ for LJ_{14}) are preferentially moved into the low-energy basins. This movement is not monatonic, where chumks of the path only fall into the basin. Rather, the chunks will often fall into the basin and other times stumble back onto the barrier. It is only the motion over many paths which exhibit the trend to the path spending most of its time in the low energy states.

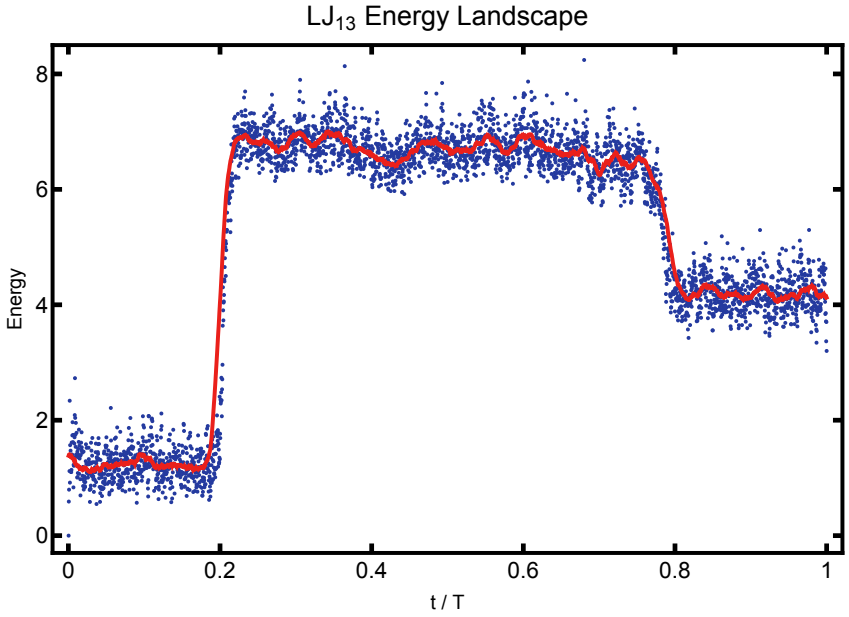


Figure 5.9: The energy landscape of the starting path used in the PSHMC sampling method for the LJ_{13} cluster. This initial path is simply a thermalized version of the zero-temperature configuration and displays the unphysical broken instanton behavior. The rapidly fluctuating energy along the path (shown in blue dots) describes a cluster at a configurational temperature of $\epsilon = 0.13$. The red line is a moving average (70 points) of the path energies and serves to guide the eye toward the average energy along the path.

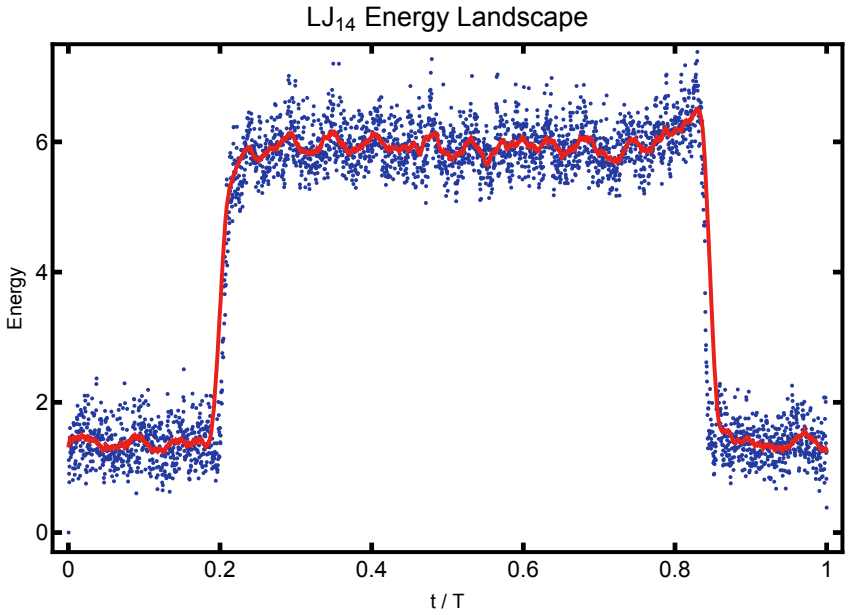


Figure 5.10: The energy landscape of the starting path used in the PSHMC sampling method for the LJ_{14} cluster which shows the unphysical broken instanton behavior. The configurational of temperature of this cluster is ϵ. The blue dots show the energy along the path and the red line the moving average (70 points) that serves as a guide to the eye.

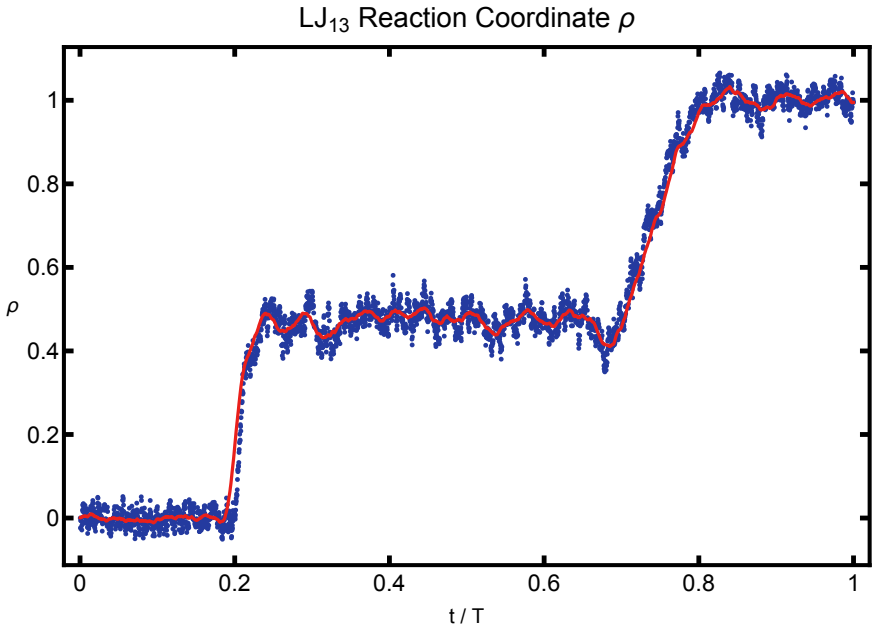


Figure 5.11: The reaction coordinate for the LJ_{13} initial path. Central (high energy) portion of the path fluctuates about $0.4 < \rho < 0.5$, which excludes the highest energy plateau of the zero-temperature pathway ($t \approx 0.7$ in Figure 5.7), which is highly disfavored at this configurational temperature.

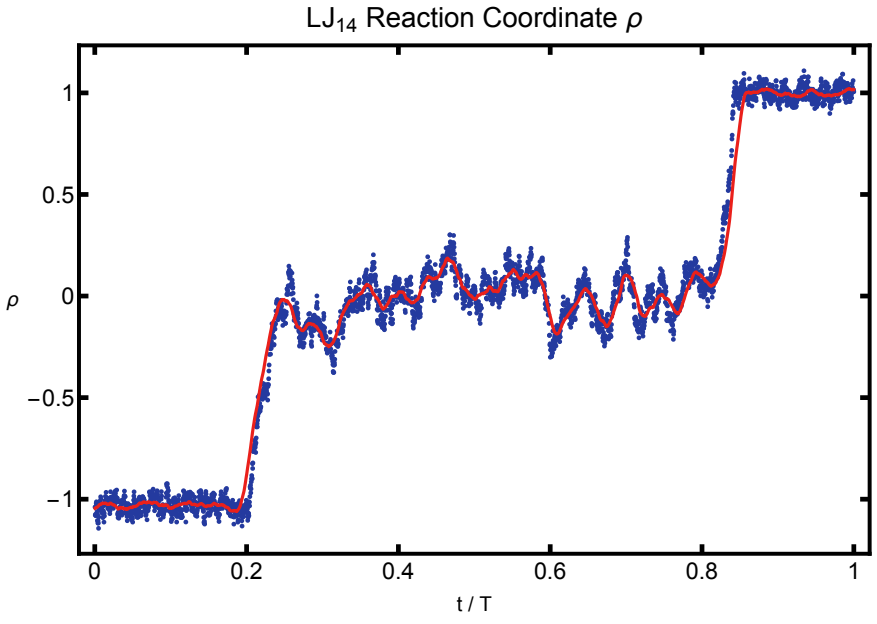


Figure 5.12: The reaction coordinate for the LJ_{14} initial path. Central portion of the path fluctuates between all three high-energy configurations ($-0.3 < \rho < 0.3$) of the zero-temperature path. At this configurational temperature the chain of particles is free to move through the cluster on top of the barrier.

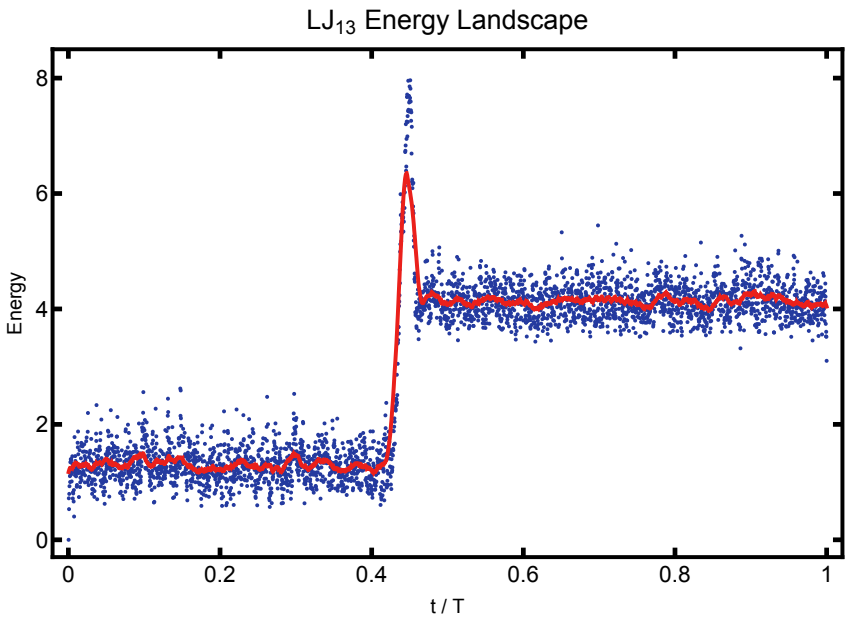


Figure 5.13: The ending path of the LJ_{13} cluster after sampling with the PSHMC method. The transition occurs in a very short amount of time, and spends almost no time on top of the high-energy barrier. This is the standard instanton picture. An animation of the evolution (in sampling time τ) of the energy landscape and the motion of the cluster for this final path is included in the supplementary materials (Animations 5.3 and 5.5).

After generating many paths using the continuous-time HMC method, the initially unphysical transition paths converge to what resembles a physical transition for each cluster (see Figures 5.13 and 5.15). The grid spacing of the path is also sufficiently small as to resolve the actual dynamics of the transition, as can be seen in Figures 5.14 and 5.16. These figures show that the path does not look strictly like an instanton with this fine of a grid spacing. The reaction coordinate for the ending paths, shown in Figures 5.17 and 5.18, reveal that the transition is a very fluid motion. Once the cluster decides to make the transition over the energy barrier, it does not linger in the high energy states, and quickly falls into the opposite barrier.

5.3.1 Conclusions from sampling the LJ clusters

There are many conclusions to be drawn from this analysis. First, the ending path created by this sampling method healed the broken instanton, revealing a path between the two boundaries which resembles a physical transition. This path is defined on a fine enough grid that the transition itself is resolved and can be analyzed.

Performing the same gradient descent analysis on the ending path for the LJ_{13} cluster to find

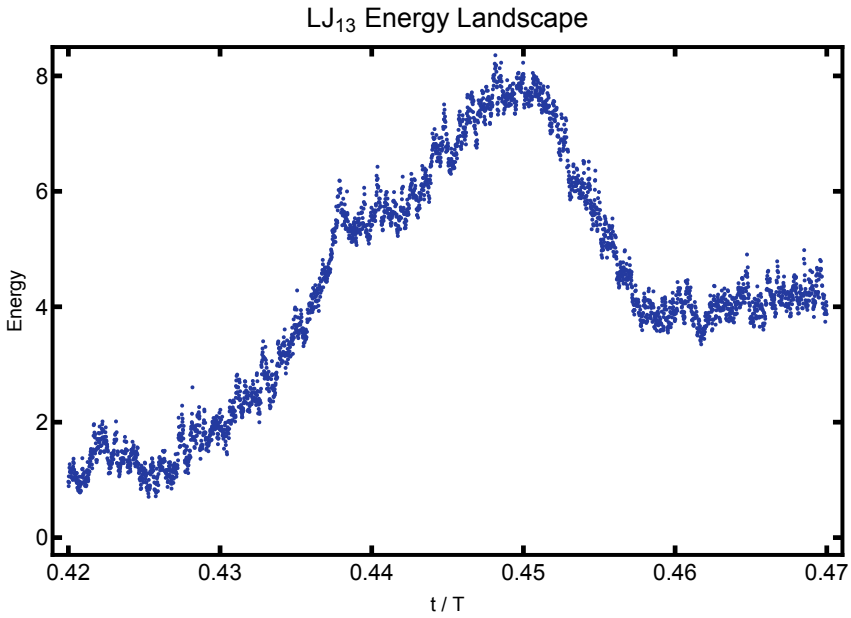


Figure 5.14: Closer view showing the details of the transition region of the LJ_{13} ending path. This figure illustrates the full structure of the transition region for the parameters defined in section 5.3.

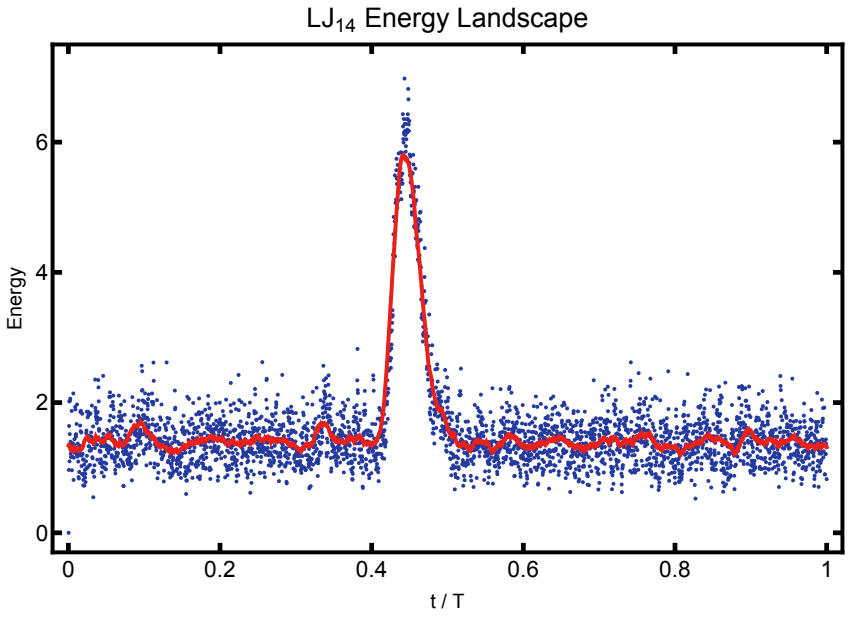


Figure 5.15: Energy landscape of the ending path of the LJ_{14} cluster. The instanton is now healed and the transition time length is small. Note that the transition does not occur at $t = 0.5T$ which suggests that the transition rates remain inaccessible with this path sampling method. An animation of the evolution (in sampling time τ) of the energy landscape and the motion of the cluster for this final path is included in the supplementary materials (Animations 5.4 and 5.6).

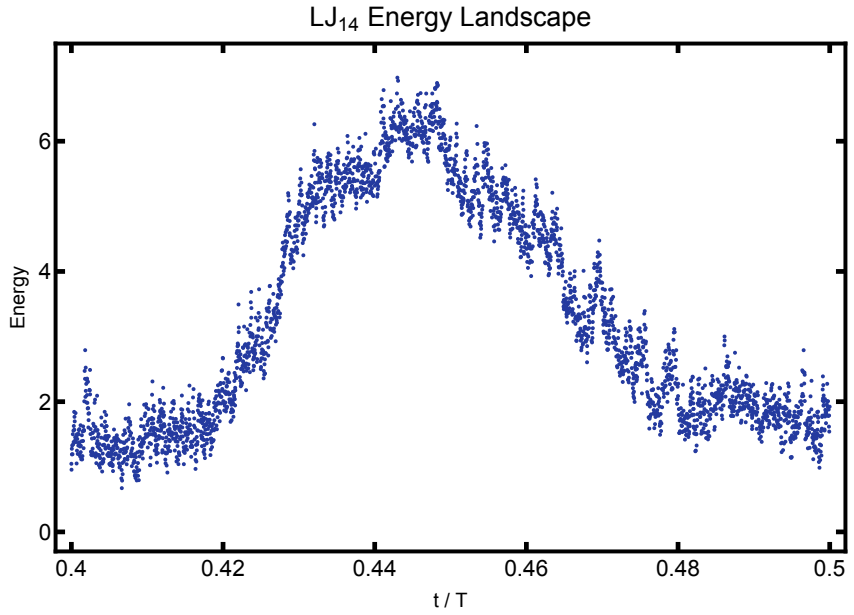


Figure 5.16: Closer view showing the details of the transition region of the LJ_{14} ending path. This figure illustrates the full structure of the transition region for the parameters defined in section 5.3.

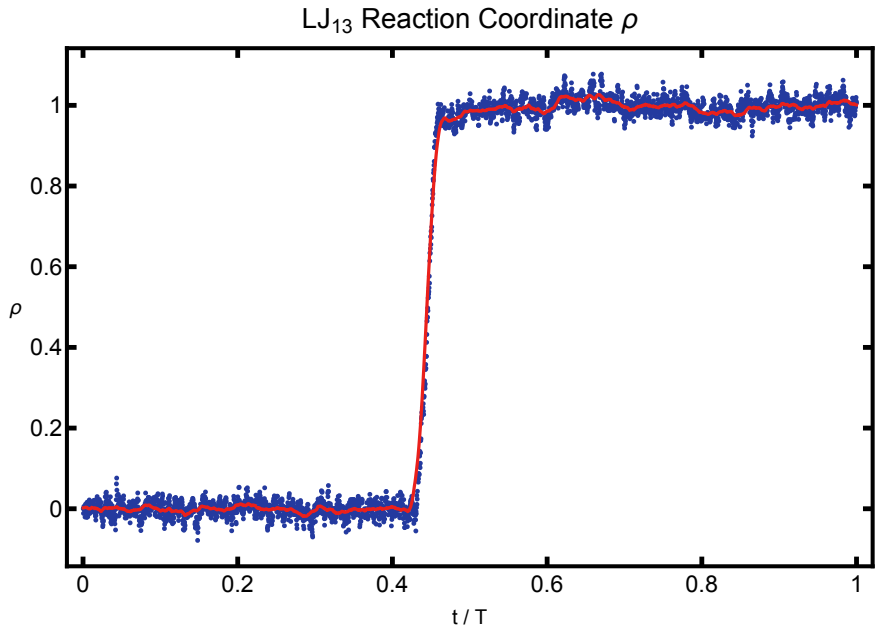


Figure 5.17: The reaction coordinate for the ending LJ_{13} path. This final configuration shows the healed instanton picture, which is the correct thermodynamic picture for the transition. The cluster only spends appreciable time in the basins, at $\rho \approx 0$ and $\rho \approx 1$, and spends as little time as is possible in the higher energy transition states. The evolution of the reaction coordinate as a function of sampling index is shown in the supplementary material (Animation 5.7).

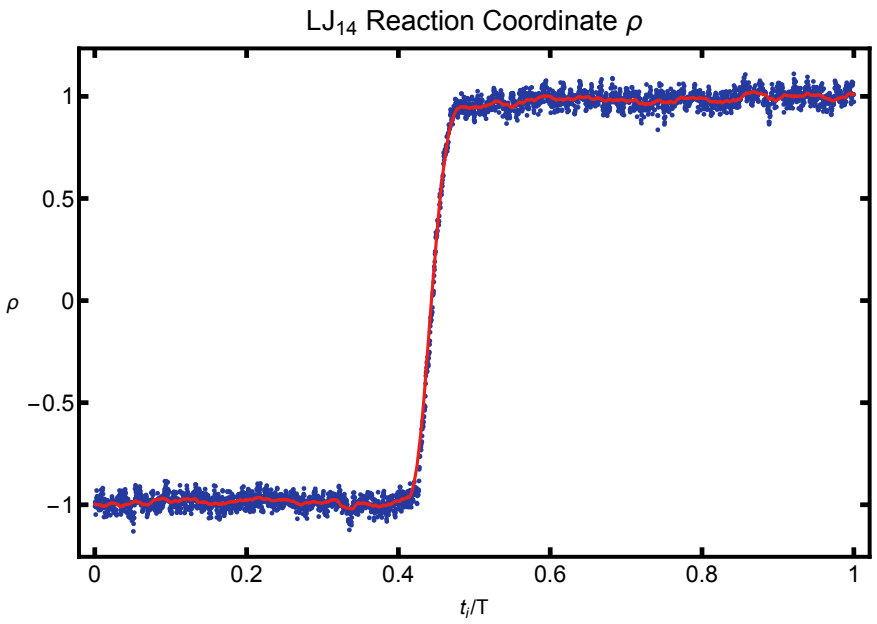


Figure 5.18: The reaction coordinate for the ending LJ_{14} path. This final configuration shows the healed instanton picture. The transition time for the LJ_{13} cluster in Figure 5.17 is shorter than this particlar LJ_{14} path. Similarly to the LJ_{13} case, the cluster spends very little time in the high energy transition region for this ending path. The evolution of the reaction coordinate as a function of sampling index is shown in the supplementary material (Animation 5.8).

the quenched path reveals very different dynamics than the results shown in section 5.1. The energy landscape, shown in figure 5.19, exhibits only two high-energy plateaus, with very small energy changes occurring on each of these these plateaus. The reaction coordinate, shown in figure 5.20, reveals a much more complicated structure to these transitions. A more thorough analysis of this motion reveals that the difference between these two paths stems from the fluctuations of the cluster breaking the symmetric breathing of the outside shell particles, which forms a fracture in the hexagonal ring.* While this analysis yields interesting dynamics for the Lennard Jones clusters, it only serves to complicate the understanding of this particular method of path sampling.

While using the Lennard Jones clusters as a testing problem to analyze the continuous-time path space HMC method is interesting physically, the complexity of the problem also has many drawbacks. The motion of the cluster becomes very complex as the path evolves with each particle seemingly fluctuating at random for the vast majority of time for each path. Understanding and explaining this complicated motion is unreasonably difficult when this is only a test case and no

*The breaking of this symmetry is obvious in hindsight. Any fluctuations of the ring particles off of the local equilibrium will break this perfectly symmetric breathing mode.

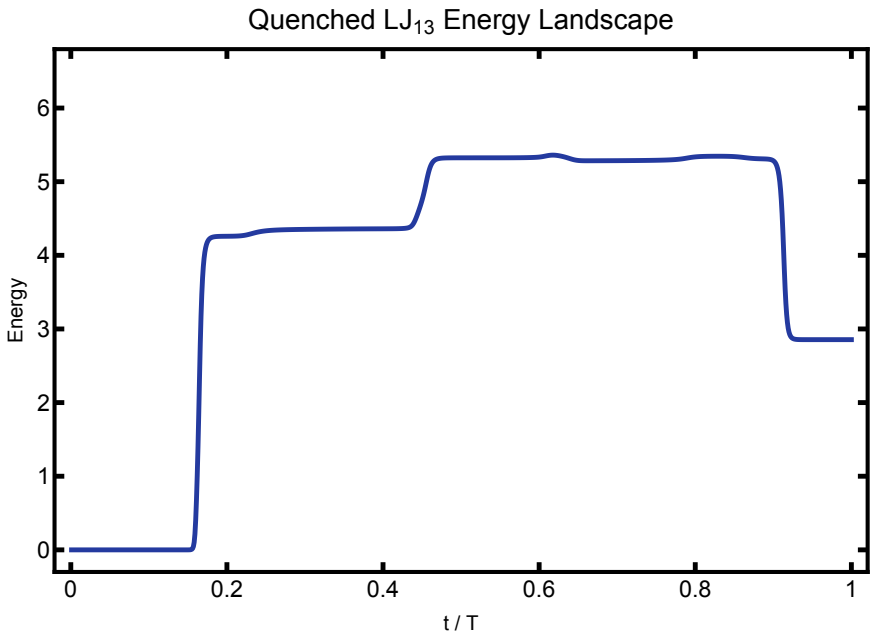


Figure 5.19: Energy landscape of the path resulting from quenching the fluctuations from of the ending LJ_{13} path. The high energy states between $0.2 < t < 0.4$ and $0.6 < t < 0.8$ shown in figure 5.13 have been eliminated, in favor of lower energy configurations which break the symmetric breathing mode. This was not seen in the original zero temperature configuration because the symmetric breathing mode moves through a unstable critical path. The motion of the cluster after performing the quench of thermal fluctuations is shown in the supplementary material (Animation 5.9).

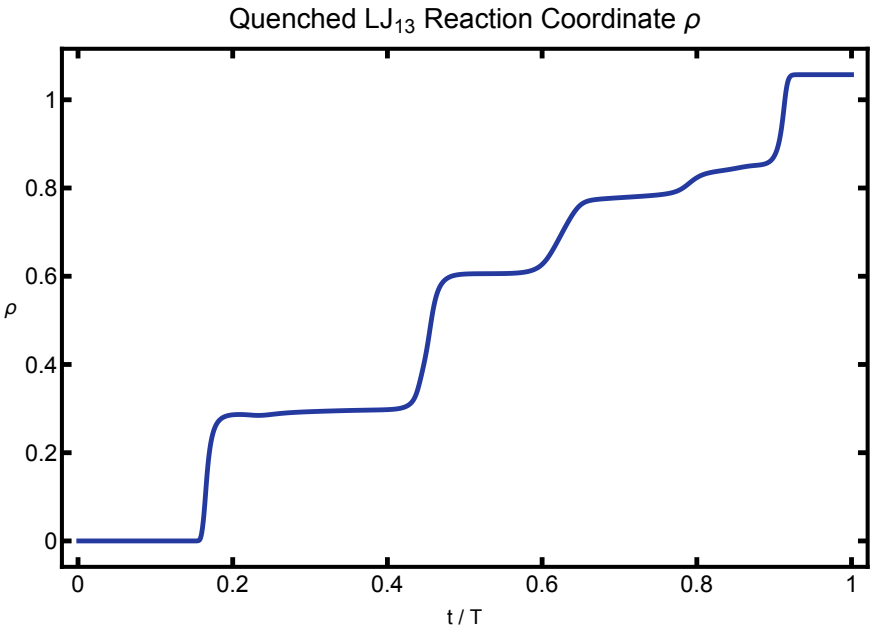


Figure 5.20: Reaction coordinate of the path resulting from quenching the fluctuations from of the ending LJ_{13} path. The transition evolves continuously from the ground state configuration to the excited state which is seen by the monotonically increasing nature of ρ. There are 4 distinct intermediate states for this particular quenching.

new physical insight will be gained. These complications originate for a few distinct reasons. The cluster being defined in three dimensions means that it is free to rotate and translate. Subtracting these rotational/translational degrees of freedom from the overall motion of the cluster is a difficult problem which, when solved, still will not increase the understanding of this proposed sampling method. Furthermore, the hard-core in the Leonard Jones potential is exceedingly stiff which, in turn, requires the step size in algorithmic time (τ), to be dramatically decreased. It is then difficult to gather useful statistics as calculation speed slows.

It is important to note that I have displayed only a single realization of an infinite number of possible transition pathways. To find the most probable pathway, one would need to calculate the density of paths with the specified boundary conditions. This is an exceedingly difficult problem for any structure with a complex free energy landscape and/or many degrees of freedom. Remember that each path is only a single realization of an infinite number of possible paths. The time at which transition occurs for any of these ending paths remains unconstrained, which implies that recovering any sort of density is inaccessible. The complexity of this test problem is obviously getting in the way of a deeper understanding the behavior of the sampling method.

Appendix 5

5.A The Path Potential G for the LJ clusters

Preliminaries

We need to perform a calculation of a few potentials in order to implement the HMC algorithm. We require the Energy, G and the gradient of G where

$$V = \frac{1}{2}\sum_{i\neq j} 4\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}}\right) \tag{5.4}$$

$$G = \frac{1}{2}|F|^2 - \epsilon\nabla^2 V \tag{5.5}$$

The potential is based on the Lennard-Jones (12-6) potential

$$V_2(r_{ij}) = 4\left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}}\right) \tag{5.6}$$

so that the full potential can be written as

$$V = \frac{1}{2}\sum_{i\neq j} V_2(r_{ij}) \tag{5.7}$$

Now for some notation.

All greek indices (α, γ) denote cartesian coordinates (ie: 1,2,3) and roman indices denote particle indices (ie: 1 to 13 for LJ_{13}). The Euclidean distance and its derivative with respect to a cartesian direction is

$$r_{ij} = \left[\sum_{\alpha} \left(x_{i\alpha} - x_{j\alpha} \right)^2 \right]^{1/2} \tag{5.8}$$

$$\frac{\partial r_{ij}}{\partial x_{k\gamma}} = \frac{\left(x_{i\gamma} - x_{j\gamma} \right) \delta_{ki}}{r_{ij}} + \frac{\left(x_{j\gamma} - x_{i\gamma} \right) \delta_{kj}}{r_{ij}} \tag{5.9}$$

where δ is the Kronecker delta

Energy

The first and easiest potential to calculate is the total potential energy stored by the system

$$V = \frac{1}{2} \sum_{i \neq j} 4 \left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}} \right) = 4 \sum_{i<j} \left(\frac{1}{r_{ij}^{12}} - \frac{1}{r_{ij}^{6}} \right) \tag{5.10}$$

Force

The calculation of G is more involved than the energy. Instead of deriving the entire result in one step I will show separate calculations of the Force and then the Laplacian in the next section. The force on particle k in the direction γ is $F_{k\gamma} = \vec{F}_k$

$$\begin{aligned} F_{k\gamma} &= -\frac{\partial V(r)}{\partial x_{k\gamma}} = -\sum_{i} \frac{\partial V}{\partial r_{ik}} \frac{\partial r_{ik}}{\partial x_{k\gamma}} \\ &= -\sum_{i} \left(x_{k\gamma} - x_{i\gamma} \right) \frac{V_2'(r_{ik})}{r_{ik}} \\ &= \sum_{i} \left(x_{k\gamma} - x_{i\gamma} \right) W(r_{ik}) \end{aligned} \tag{5.11}$$

where $W(r_{ij}) = \frac{-V_2'(r)}{r}$

Laplacian of V

The Laplacian of the potential is given as $\vec{\nabla} \cdot (\vec{\nabla} V)$

$$\begin{aligned}\mathcal{L} = \sum_{\gamma} \frac{\partial^2 V(r)}{\partial x_{k\gamma}^2} &= \sum_{\gamma} \frac{\partial}{\partial x_{k\gamma}} \left(-F_{k\gamma}\right) \\ &= \sum_{\gamma} -\frac{\partial}{\partial x_{k\gamma}} \left[\sum_{i} \left(x_{k\gamma} - x_{i\gamma}\right) W(r_{ik})\right] \\ &= \sum_{\gamma} \left[-\sum_{i} W(r_{ik}) - \sum_{i} (x_{k\gamma} - x_{i\gamma})^2 \frac{W'(r_{ik})}{r_{ik}}\right] \\ &= -\sum_{i} (3W(r_{ik}) - r_{ik} W'(r_{ik}))\end{aligned} \tag{5.12}$$

The above quantity is a scalar. Now we introduce $\mathcal{L}$ (not the Lagrangian) as the Laplacian summed over all particles

$$\mathcal{L} = \sum_{k\gamma} \frac{\partial^2 V}{\partial x_{k\gamma}^2} = \sum_{ik} \mathcal{Z}(r_{ik}) \tag{5.13}$$

$$\mathcal{Z}(r) = -3W(r) - rW'(r) \tag{5.14}$$

Gradient of G

The gradient of G is much more involved than the previous derivations. Lets start by seperating G into two parts: G_1 and G_2

$$G = G_1 + G_2 \tag{5.15}$$

$$G_1 = \frac{1}{2} \sum_{i\alpha} F_{i\alpha} F_{i\alpha} \qquad \text{and} \qquad G_2 = -\epsilon \mathcal{L} \tag{5.16}$$

The gradient of the force squared is

$$\frac{\partial G_1}{\partial x_{k\gamma}} = \sum_{i\alpha} F_{i\alpha} \frac{\partial F_{i\alpha}}{\partial x_{k\gamma}} \tag{5.17}$$

$$= \sum_{i\alpha} F_{i\alpha} \left(\frac{\partial}{\partial x_{k\gamma}} \left(\sum_{j} \left(x_{i\alpha} - x_{j\alpha}\right) W(r_{ij})\right)\right) \tag{5.18}$$

$$= \sum_{i\alpha} F_{i\alpha} \left[\sum_{j} W(r_{ij}) \frac{\partial}{\partial x_{k\gamma}} (x_{i\alpha} - x_{j\alpha}) + \sum_{j} (x_{i\alpha} - x_{j\alpha}) \frac{\partial}{\partial x_{k\gamma}} W(r_{ij})\right] \tag{5.19}$$

Chapter 6

Results in One Dimension

In this chapter, I present convincing numerical evidence that sampling the Ito-Girsanov form of the OM functional produces unphysical paths. The cause of this unphysical nature will later be shown to be due to a fundamental flaw in the construction of the continuous-time measure.

To frame the objective of this chapter, I remind the reader of the goal of this work. I am interested in understanding and developing improved sampling methods, rather than forming a deeper understanding Lennard-Jones cluster dynamics. The complexity of the LJ clusters obscures how well the sampling procedure actually probes the underlying physics and dynamics of the system. Having spent much time trying to understand the complexities of these larger systems, I will divert my focus to a simpler problem.

To clearly understand the results of this analysis, I will study a single particle which resides in an externally imposed one-dimensional potential. Obviously, this is a much simpler system than the one used for the LJ clusters, and thus it is easier to explore in a detailed fashion. Paths in these systems should be consistent the Boltzmann distribution at finite temperature. In one dimension, visualization can be used to understand if an ensemble of paths are physical in this sense. In a one-dimensional system, I do not need to consider the complication of the translational and rotational degrees of fredom revealed in the study of the 3-dimensional LJ systems. These potentials can also be tailored to probe specific physical properties of the path, for example understanding entropic forces as seen in the work of Pinski and Stuart[40]. The generation of the initial path can also be performed using a forward method from Chapter 2, which absolves the algorithm from having to heal the broken instanton which would be present when using the thermalized zero-temperature

path. This saves a large amount of computational time as the calculation does not need to anneal the system before statistics can be gathered.

At the conclusion of this chapter, I hope to convince the reader that there are serious problems with the Ito-Girsanov continuous-time formulation of the PSHMC sampling method. This chapter will only present the evidence for this claim. In the subsequent chapter, I will employ the Metropolis algorithm to identify what errors are made using this method.

6.1 Externally Imposed One-Dimensional Potential

For this study, the construction of the externally imposed potential in which the particle resides is of utmost importance. To study rare transition events, the potential is required to have, at minimum, two energy basins which are spatially separated by an energy barrier, which is large compared to the available thermal energy. The paths of interest will start in one of these basins, transition over the energy barrier, and end in the other basin. Possibly the simplest potential of this form is a symmetric potential which possesses to generate energy basins at $x = -1$ and $x = +1$ and a barrier of unit energy at $x = 0$, and is defined by the polynomial

$$V(x) = \left(x^2 - 1\right)^2 \tag{6.1}$$

Each of the basins in this potential are quadratic near the minimum of the potential and thus the leapfrog integration should lead to correct sampling. There been previous studies which have looked at potentials with symmetric energy basins using past sampling procedures similar to the one outlined here[35]. I am interested in finding paths which, on average, are compatible with the Boltzmann distribution. The posterior probability distribution is extremely difficult to find to high precision. To leave an appreciable amount of path in each well, the choice of basins which are degenerate in energy is advantageous, as any disparity in energy in the basins will drive the vast majority of the path to the lower energy well.

The requirement that the energy basins be degenerate in energy allows the unequal entropic contributions of the basins to be probed. The path sampling methods discussed here are designed to probe the free energy landscape, which includes both entropic and energy effects. To probe

entropic effects in a symmetric potential one would have to resolve the tails of the path distribution, as the probability of being in one basin is the same as being in the other. To explore this issue, I use a potential which is degenerate but has a one narrow basin and one wide basin, which will give rise to different entropic basins. The wide well is entropically preferred at any temperature $\epsilon > 0$. For paths to be consistent with the Boltzmann distribution, the particle should be observed preferentially to the in the wide well over the narrow well. The analysis can now be performed by looking at the basin probability, rather than requiring resolution of the tails of the posterior probability distribution.

The potential I have chosen to perform this study on is an asymmetric potential which is given by the following polynomial expression

$$V(x) = \frac{(8-5x)^8(2+5x)^2}{2^{26}} \tag{6.2}$$

which has two degenerate energy basins separated by a barrier of height unity. The narrow well has a quadratic shape near the minima, similar to the symmetric potential in Equation 6.1, while the wide well is almost nearly flat near the minima. This potential, which is illustrated in Figure 6.1, will be referred to as the *Thin-Broad* potential and will be the focus of the following numerical results.

6.1.1 Simulation at Finite Temperature

The generation of a starting path for the one dimensional problem is significantly simplified compared to the Lennard-Jones case. For the LJ clusters, the starting paths were by adding noise to the zero-temperature paths which contained a broken instanton. This feature required significant computational time to thermalize. For the 1D problems of interest here, I will avoid using a path with a broken instanton by simply selecting a specific part of a forward trajectory. Carefully constructing the path will still provide a non-Boltzmann starting path which serves as a test for the method to heal to a Boltzmann-like state. In particular, in the starting path, the particle spends about 50% of the time in the wide well, rather than the 90% required by Boltzmann. The path will have a total of 3 crossings of the energy barrier, starting in the thin well and ending in the wide well.

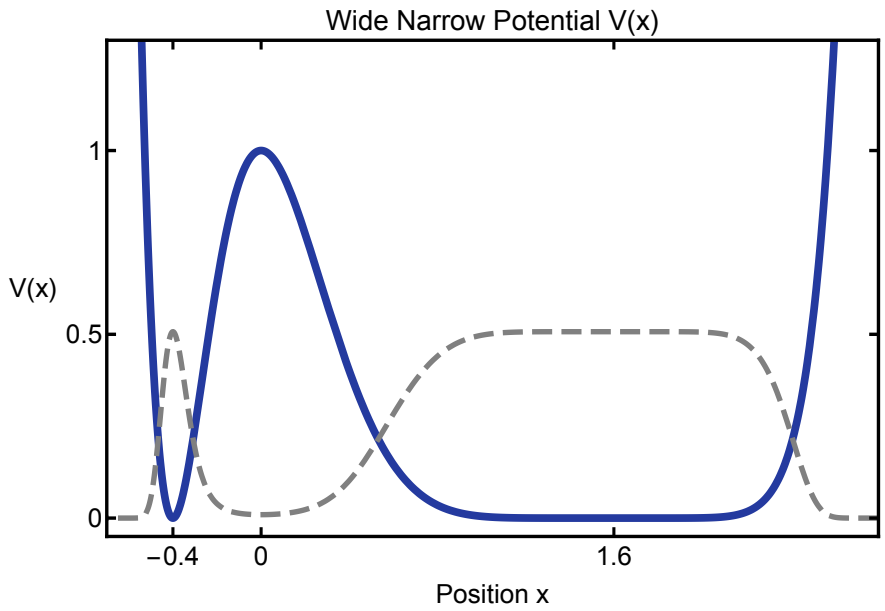


Figure 6.1: The thin broad potential (Equation 6.2) which is the system study for the chapter. The blue line shows the potential in which the particle resides, starting in the left degenerate minima ($x = 0.4$) and making a transition(s) over the barrier to the right minima ($x = 1.6$). The dashed grey line is the Boltzmann probability distribution.

The details of the procedure I have used is as follows. Using a forward method from Chapter 2, I generate a trajectory and select a portion which spends a large time in one of the basins, makes a transition over the energy barrier, and then spends a large time in the opposite basin. Next, I cut a section of this path which is approximately centered about the transition. Make this new section of the trajectory have a length of 1/3 of the final required path length time and modify the end points of this path to have the boundary conditions of choice (first position is at x^- and final position is at x^+). To create the final path make a sequence of three of these paths end to end, with the middle path time reversed. See Figure 6.2 for an illustration of this initial path. This forms a path which has a positional distribution of approximately 60% in one well and 40% in the other, and has 3 separate and equally spaced transitions.

For the asymmetric Thin-Broad potential shown in Equation 6.2, this path is obviously non-Boltzmann. The particle spends only 60% of the time in the broad well when the true distribution should approximately 90% of time spent at $\epsilon = 0.25$. A robust sampling algorithm is required to heal this poor initial guess.

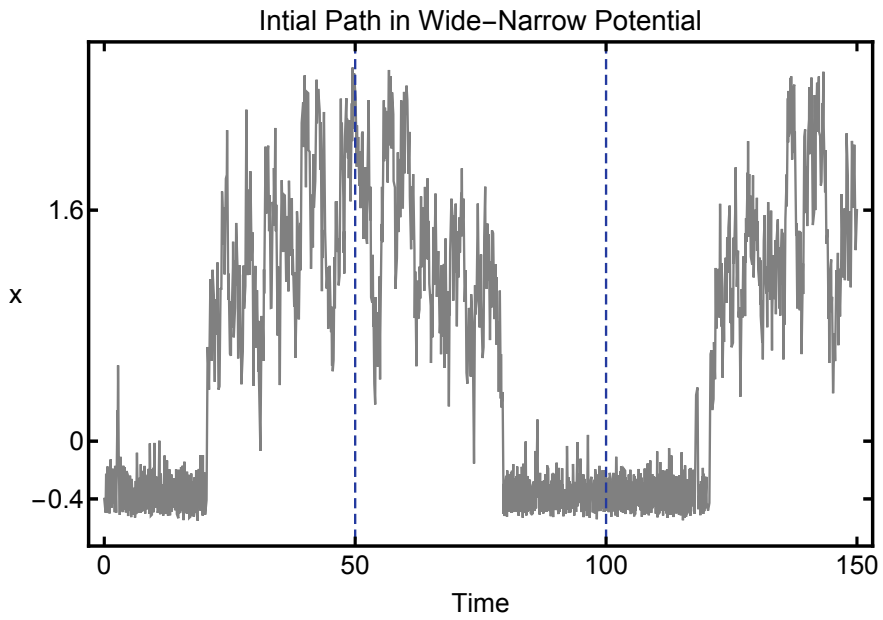


Figure 6.2: Path used as the initial guess created by combining three trajectories, which are visually separated at that dashed lines. The boundary conditions for the past are ($x^- = 0.4$) and($x^+ = 1.6$). The boundary conditions require at minimum one transition to occur along the path, and in this case there are three transitions.

6.1.2 Final Paths

I have used the initial path shown in Figure 6.2 and applied the PSHMC method as well as a novel method (novelHMC) to be introduced in the next chapter, to generate a sequence of paths. The time step along the path, Δt, chosen in accordance with the guidelines in Appendix 4.A, is sufficiently small as to preserve a high level of accuracy in the simulation. This time step size is very small for the sampling in the flat regions of the potential. It becomes necessary to have this small of a step size only when sampling the very steep outer walls of the potential, especially for the thin well which has a much steeper potential*.

The parameters of these paths are: temperature $\epsilon = 0.25$ with a path time length of $T = 150$. The step size along the path is $\Delta t = 0.005$, which sets the number of time steps along the path to be $N_B = T/\Delta t + 1 = 30,001$. The time step between paths, τ, is chosen to give an acceptance rate for the MHMC step greater than 0.50 and will differ for each of the methods. For the PSHMC simulation $\Delta\tau = 10^{-7}$ yields an acceptance rate of 60% and for NovelHMC $\Delta\tau = 10^{-6}$ yields an acceptance rate of 91%. This small step size therefore only allows small changes between two

*For example, the minimum of the final path is $x \approx -0.55$ which corresponds to $G \approx 250$ which leads to the step size chosen. See Appendix 4.A for details.

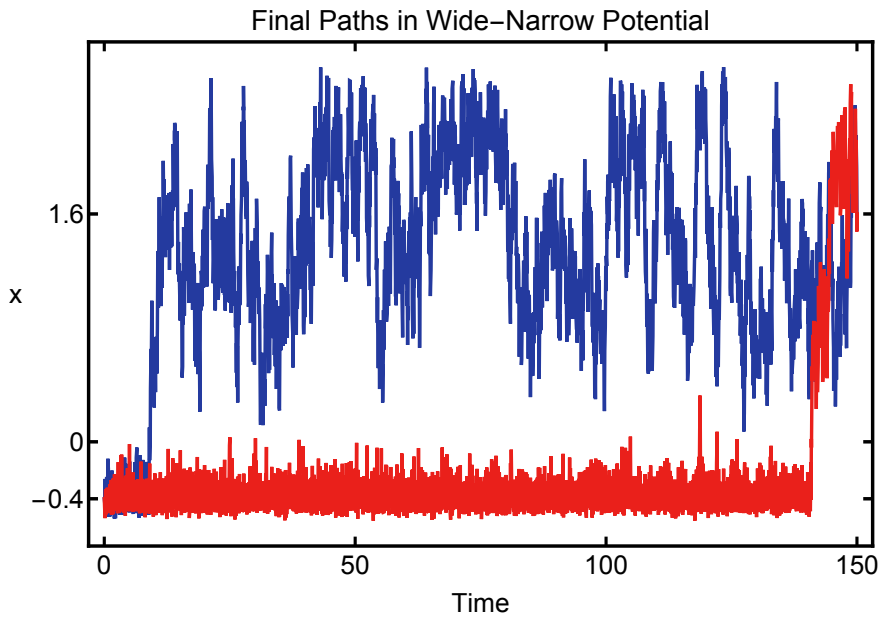


Figure 6.3: The ending paths of two HMC path sampling methods after a sequence of paths have been generated. The red path was generated using PSHMC and the blue path via a novel HMC path sampling method to be introduced in the following chapter. Note that the PSHMC spends the majority of his time in the entropically disfavored, thinwell.

subsequent paths, which means that a large sequence of paths must be generated by each method in order to infer the long-time behaviour. The number of Molecular Dynamics (MD) steps per MHMC adjustment for PSHMC is $N_{MD} = 1000$ and for NovelHMC is $N_{MD} = 500$. This number of MD steps leads to efficient sampling for the methods (see Section 4.5.3). Note that the larger time step between paths ($\Delta\tau$) corresponds to increased movement of the path, which decreases the computational effort required to find the steady state behavior.

The ending paths for both methods are shown in Figure 6.3. There is a major problem with the PSHMC result. It is clearly seen that the PSHMC method drives the path into the thin well, and as will be seen, reaches the steady state in this unphysical configuration. The NovelHMC method appears to sample the potential correctly (see Chapter 7 for details).

The underlying Boltzmann distribution provides a metric to gauge the physicality of the path ensemble. Minimizing the free energy of the system means that the majority of the positions along each path should be loacted in the broad well, as there are more states to occupy. This potential was designed specifically to probe such entropic factors. The calculation of time spent in each well is performed very easily using the Heaviside function Θ to define $B(s)$ to be the fraction of the

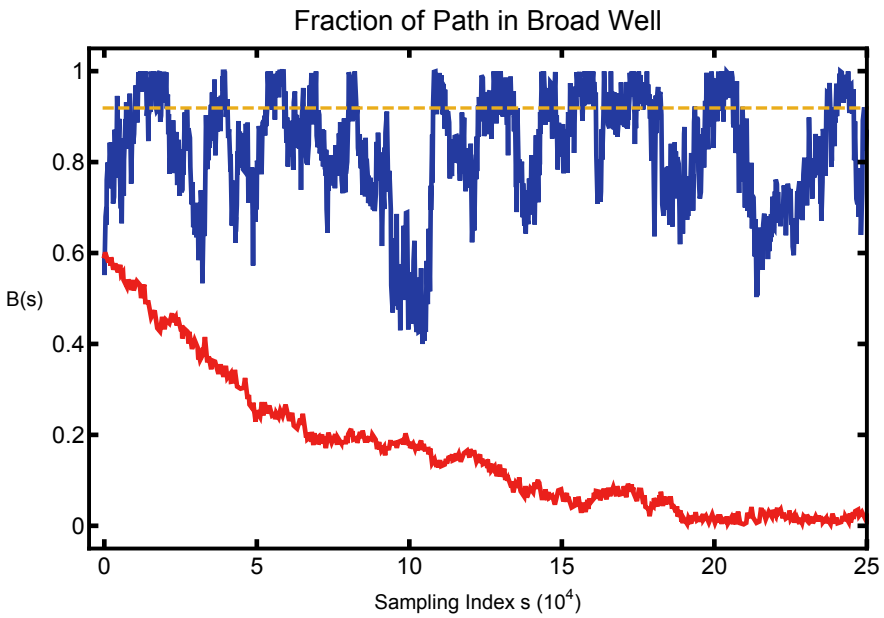


Figure 6.4: The broad well fraction (Equation 6.3) shown paths are generated. The expected Boltzmann state is shown as the dashed orange line. The red, lower curve corresponds to the PSHMC sequence and shows the method driving the system to a non-Boltzmann state. The blue curve displays B(s) for the novel HMC method oscillating near the expected Boltzmann value.

path which is contained in the broad well, namely

$$B(s) = \frac{1}{T}\int_0^T dt\ \Theta(x_t^{(s)}) \approx \frac{1}{N}\sum_i \Theta(x_i^{(s)}) \tag{6.3}$$

where the sampling indexes denoted as s, and the corresponding path is $x_i^{(s)}$. If the sequence of paths generated using a sampling method disagrees with the Boltzmann prediction for this metric, then the paths generated with the method are clearly not physical.

$B(s)$ is shown as sampling proceeds in Figure 6.4. The PSHMC method quickly drives the sampling to the incorrect well according to the Boltzmann distribution. The percent of time spent in this wide well steadily decreases from approximately 60% to 5%. Once the path for PSHMC falls into the narrow well, there is no recovery to the Boltzmann state. These paths are most assuredly not physical. The NovelHMC method behaves in a much different, and in many respects, more sporadic manner, where the fraction of the path in the broad well fluctuates between $0.5 < B(s) < 1.0$. This behaviour is compatible with physical intuition as the paths should have

large fluctuations but be roughly centered on the expected Boltzmann result.* Note that these results have not converged to the target distribution, as the MHMC method does for forward integration, as the paths still have an error which is related to the step size Δt, and the length of the path, T. Clearly, the ensemble of paths generated using novelHMC is much better behaved than the paths generated with PSHMC. †

6.2 Equivalence of Path Probability

I would now like to consider a computer experiment which creates paths using the following very simple algorithm. First, generate a set of random Gaussian variates $\{\xi\}$, of length N, where $T = N * \delta t$ is much larger than the first passage time as given in Figure 2.1. Next, generate a trajectory using the leapfrog (Euler-Maruyama) integrator

$$x_{i+1} = x_i + F(x_i)\Delta t + \sqrt{2\epsilon\Delta t}\xi_i \tag{6.4}$$

where x_0 is set to be in the minimum of the wide well ($x_0 = 1.6$) and record the endpoint of this trajectory. Performing this integration gives a single trajectory of length T, and end point $x_{end} = x(T)$. Up to this point the analysis is exactly equivalent to the forward leapfrog method, and the OM path probability is simply

$$\mathbb{P}_{OM} = \prod_i \frac{1}{\sqrt{2\pi}} \exp(-\frac{\xi_i^2}{2}) \tag{6.5}$$

Now, rather than generating an entire new set of random Gaussian variates, simply scramble the set into a completely different ordering. It is important to emphasize that this analysis uses *exactly* the same set of random variates, and thus the probability of all such paths will be equal. Using the scrambled set $\{\xi\}$, generate a new path using the leapfrog integrator and record the endpoint of the trajectory, $x_{end} = x(T)$.

This basic experiment has been performed on the globally Lipschitz force defined from the potential in Equation 6.2. The length of trajectory was taken to be $T = 1000$, according to a

*The paths should lie exactly on the Boltzmann result only in the limit of the path time $T \to \infty$.

† See the attached animations to see a sample of how the paths evolve for each of these methods.

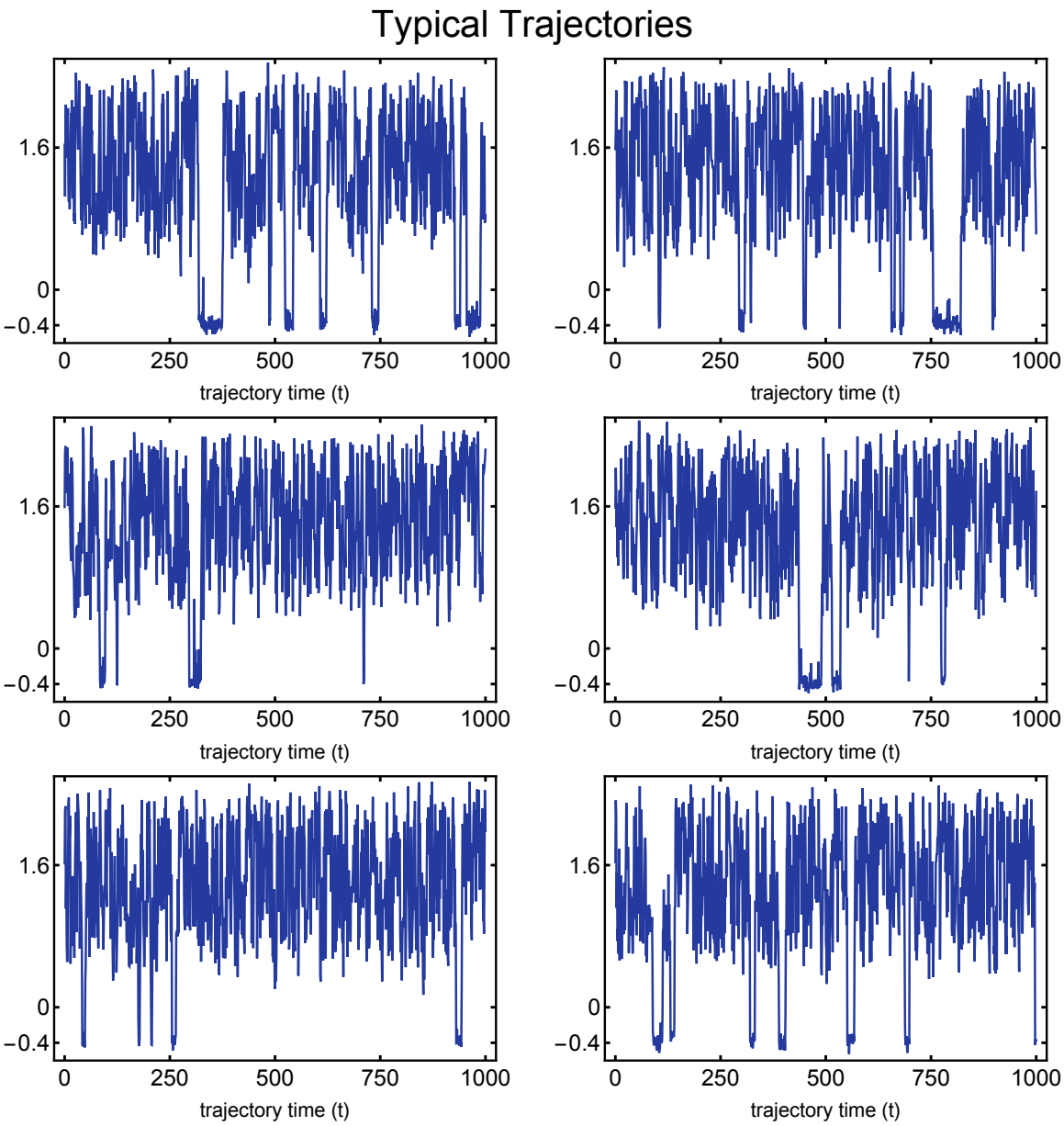


Figure 6.5: Examples of 6 paths generated using the leapfrog integration method. Each of the paths in this figure use an identical set of random Gaussian fluctuations and the same starting point $x_- = x(0)$, the only difference is the ordering of the random numbers. Every one of the paths in the figure have *equivalent* OM probabilities, and there are $N! = (2 * 10^6)! \approx 10^{11733474}$ total paths in this set. Note how much variation in path structure is obtained from only this single set of random numbers. Surely almost every imaginable trajectory is contained in this set of random numbers alone.

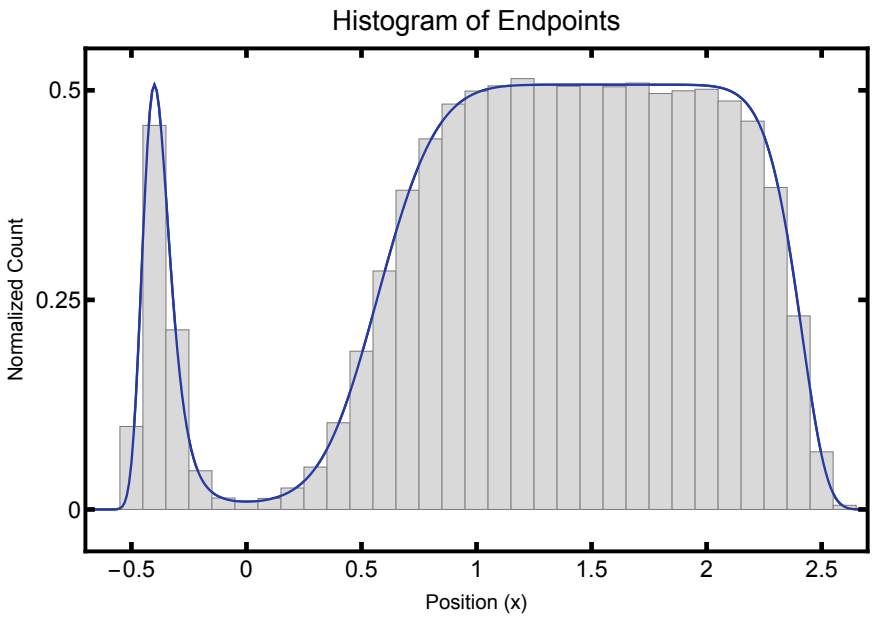


Figure 6.6: Distribution of end points of the set of trajectories with equal OM probability. Here the endpoints of 472,640 different trajectories are shown. The blue line is the Boltzmann distribution for the configurational temperature $\epsilon = 0.25$

similar analysis performed in Chapter 1 (see Figure 2.1). To sample as close as possible to the actual Boltzmann distribution (remember there is no MHMC test performed in this experiment), I have chosen a very small time step of $\Delta t = 0.0005$, giving a total number of steps, $N = 2 * 10^6$. A sample of 6 of the possible paths is shown in Figure 6.5, all of which are extremely different realizations of the trajectory. One could ask the question: What path is the most probable path? The question has no answer, as all of these paths are equally probable.

This procedure will give a set of a maximum of $N!$ different trajectories (this number surpasses the distinction of astronomically large for this choice of N), all of which have *equivalent* probabilities under the OM probability. I have sampled a set of 472,640 scrambled paths in this experiment, and found that the distribution of endpoints created by the calculation agree with Boltzmann up to an error which is due to the finite step size (discussed in Figure 2.3 for a similar potential). The normalized distribution of endpoints are shown in Figure 6.6.

Now recall the regularized probability measure for the continuous-time method

$$\frac{d\mathbb{P}}{d\mathbb{Q}} = \exp\left(-\frac{1}{2\epsilon}\left(V(x^+) - V(x^-) + \int_0^T dt G(x_t)\right)\right)$$

This path probability has an explicit dependence on the specific path, as can be seen in the last term

of the above equation. The looming question is now determining why some specific paths, which minimize $G(x_t)$, maximize the probability. This could help in the understanding of the unphysical results obtained when sampling the continuous-time form of the OM functional.

6.3 Questions raised from this analysis

The results shown in this chapter motivate two questions which are fundamentally important to path sampling methods.

- Why does the PSHMC fail so catastrophically when sampling the Ito-Girsanov form of the OM probability?

The analysis of the evolution of paths generated with the PSHMC method shows the continuous-time OM functional exhibits unphysical evolution. Identifying what exactly has gone wrong is very difficult when examining the problem using simple Brownian Bridges. * All of the errors for these Brownian processes are simply ignored when integrating the path forward in time.

- If all paths have (almost) equivalent probabilities, how can one find a 'Most Probable Path'?

In previous works, the path potential G was used as an indicator that one could optimize paths. This formed a notion of a Most Probable Path (MPP)[21]. The analysis of the equivalence of path probability in Section 6.2 raises serious questions into the validity of the idea of MPPs. Furthermore, long paths will all have almost equivalent OM probabilities. Why should some of the paths in this set be most probable? Path probability equivalence would suggest that the path density should be the object of interest, rather than a single path with maximal probability.

*It turns out that sampling with only the Brownian Bridge fails to sample the full measure in Equation 4.13. This point will be discussed further in Chapter 7.

Chapter 7

Finite Method

Clearly the evidence from Chapter 6 is compelling enough to ask serious questions about the path probability distribution derived using the Ito-Girsanov change of measure ($\mathbb{P}_{IG}$). The more difficult question is what methods are available that are understood well enough to definitively answer such questions. Evidently, the understanding of Brownian dynamics has led us astray, and therefore should not be used as the foundation to probe these tough questions.

In this chapter, I will present the framework of a novel view using the Metropolis algorithm which is analogous to the PSHMC method of Chapter 4, but not based on a continuous-time formalism. This method is a powerful tool that is proven to generate paths which are consistent with the Boltzmann probability distribution, and can be employed to examine errors made in the integration. Using the Metropolis algorithm as a lens, I will illuminate a fundamental misunderstanding of this continuous time probability measure ($\mathbb{P}_{IG}$).

7.0.1 Preliminaries

The notation used in this chapter is adapted from the PSHMC method of Chapter 4. The time step along the path is Δt and between paths is $\Delta\tau$, each of which are related to the leapfrog time steps in the following manner

$$h = \sqrt{2\,\Delta t} \qquad\qquad \eta = \sqrt{2\,\Delta\tau} \tag{7.1}$$

It is very useful to introduce the average and difference between path positions with the following notation

$$\bar{x}_n = \frac{x_{n+1} + x_n}{2} \qquad\qquad \Delta x_n = x_{n+1} - x_n$$

Note that $\bar{x}_n$ and Δx_n depend on the $n+1$ index, which will become important when performing the derivatives later in this chapter.

Because this is a finite time method, the second time derivative operator, $\boldsymbol{L}$, from the PSHMC method is the 2^{nd} order finite difference operation:

$$L\, x_n = \frac{x_{n+1} - 2x_n + x_{n-1}}{\Delta t} \tag{7.2}$$

7.1 Integrating Forward

In order to extend this analysis for use with multiple integration methods, I will perform the analysis with a *weighted Force*. This force will take one of 4 forms here

- Leapfrog: $F_w(x_n, x_{n+1}) = F(x_n)$
- Mid Point: $F_w(x_n, x_{n+1}) = F_w(x_{n+1}, x_n) = F\left(\frac{x_{n+1}+x_n}{2}\right) = F(\bar{x_n})$
- Trapezoid: $F_w(x_n, x_{n+1}) = F_w(x_{n+1}, x_n) = \frac{F(x_n)+F(x_{n+1})}{2}$
- Simpson's: $F_w(x_n, x_{n+1}) = F_w(x_{n+1}, x_n) = \left(F(x_n) + 4F(\bar{x}_n) + F(x_{n+1})\right)/6$

Leapfrog is the only weighted Force shown here which is not symmetric about the forward and backward direction. The symmetric Forces will benefit from a significant simplification to the energy drift (δe_n), which will be discussed in an Section 7.1.2. For this particular analysis, I will focus on the Mid-Point weighted Force, but the full derivation using the other 3 methods are provided in Appendix 7.A, along with an explanation of the advantages and disadvantages associated with the each choice of weighted Force.

7.1.1 Forward Evolution

The method begins with the prescription for the forward and backward integration in time

$$x_{n+1} = x_n + v_n^i h + \frac{1}{2}h^2 F_w(x_n, x_{n+1}) \tag{7.3}$$

$$x_n = x_{n+1} - v_n^f h + \frac{1}{2}h^2 F_w(x_{n+1}, x_n) \tag{7.4}$$

Note that these equations require two velocities at every index, n. Solving for each of the velocities yields the following set of equations

$$v_n^f = \frac{1}{h}(x_{n+1} - x_n + \frac{h^2}{2}F_w(x_{n+1}, x_n)) \tag{7.5}$$

$$v_n^i = \frac{1}{h}(x_{n+1} - x_n - \frac{h^2}{2}F_w(x_n, x_{n+1})) \tag{7.6}$$

7.1.2 Energy Drift Between Steps

The drift in energy, δe_n, is simply the potential energy plus the kinetic energy, which is easily written in terms of the path variables and the weighted force

$$\delta e_n = \Delta PE + \Delta KE \tag{7.7}$$

$$= U(x_{n+1}) - U(x_n) + \frac{1}{2}\left(v_n^f\right)^2 - \left(v_n^i\right)^2 \tag{7.8}$$

$$= U(x_{n+1}) - U(x_n) + \frac{1}{2}\left(v_n^f + v_n^i\right)\left(v_n^f - v_n^i\right) \tag{7.9}$$

The energy drift is due to the approximate numerical integration necessitated by using discrete time steps. A manipulation can now be employed with Equations 7.5 and 7.6, which makes the calculation of the energy drift much easier. Adding and subtracting the initial and final velocities

$$v_n^f - v_n^i = \frac{h}{2}(F_w(x_{n+1}, x_n) + F_w(x_n, x_{n+1})) \tag{7.10}$$

$$v_n^f + v_n^i = \frac{h}{2}(x_{n+1} - x_n + \frac{h^2}{4}(F_w(x_{n+1}, x_n) - F_w(x_n, x_{n+1}))) \tag{7.11}$$

which can be directly inserted into the equation for the drift in energy.

$$\delta e_n = U(x_{n+1}) - U(x_n) + \frac{1}{2}\left(x_{n+1} - x_n + \frac{h^2}{4}(F_w(x_{n+1}, x_n) - F_w(x_n, x_{n+1}))\right)(F_w(x_{n+1}, x_n) + F_w(x_n, x_{n+1})) \tag{7.12}$$

$$= U(x_{n+1}) - U(x_n) + \Delta x_n \left(\frac{F_w(x_{n+1}, x_n) + F_w(x_n, x_{n+1})}{2}\right) + \frac{\Delta t}{4}\left(F_w(x_{n+1}, x_n)^2 - F_w(x_n, x_{n+1})^2\right) \tag{7.13}$$

Referring back to Section 7.1, the last term will vanish for symmetric (that is, Mid-Point, Trapezoid and Simpson's) weighted Forces. This dramatically simplifies the equation for the energy drift for the following analysis.

7.2 Probability Between Steps Along a Path

The goal of this analysis is to derive an effective Hamiltonian which resembles Λ from Equation 4.19. The OM statement holds in relation to the probability as a function of the canonical variables $\mathbb{P}_{OM}(q, p) = \mathbb{P}_{OM}(x_n, v_n)$, which is used in a forward trajectory. Here I am interested in deriving an analogous probability between two points on a path, namely $\mathbb{P}_{OM}(x_n, x_{n+1})$. By applying a change of variables, the OM statement can also be used on this probability*

$$\mathbb{P}_n = \mathbb{P}(x_n, x_{n+1}) = J_n \mathbb{P}(x_n, v_n) = J_n \exp\left(-\frac{\xi_n^2}{2}\right) \tag{7.14}$$

where the Jacobian transformation between a variable v_n and x_{n+1} is

$$J_n = h\left|\frac{\partial v_n^i}{\partial x_{n+1}}\right| \tag{7.15}$$

The scaling of the Gaussian random variates, ξ_n, with the configurational temperature, ϵ, is given by the Maxwell-Boltzmann distribution with unit mass, namely

$$v_n^i = \sqrt{\epsilon}\,\xi_n \tag{7.16}$$

*This statement is only loosely true as the sum is necessary for paths to have the correct probability. This term will become the full OM statement when summing the paths to form the effective Hamiltonian in the next section

The probability can now be written in terms of the velocities along the path

$$\mathbb{P}_n = J_n \exp\left(-\frac{\xi_n^2}{2}\right) = J_n \exp\left(-\frac{(v_n^i)^2}{2\epsilon}\right) \tag{7.17}$$

Performing some algebraic manipulations and substituting the path positions for v_n^i from Equation 7.6 yields an expression which resembles the original OM probability

$$-\ln \mathbb{P}_n = \frac{(v_n^i)^2}{2\epsilon} - \ln J_0 \tag{7.18}$$

$$= \frac{\Delta t}{2\epsilon}\left(\frac{1}{2}\left|\frac{x_{n+1}-x_n}{\Delta t} - F_w(x_n, x_{n+1})\right|^2\right) - \ln J_0 \tag{7.19}$$

In keeping with the aim of this analysis, I will now simply expand this term. The next step is to add and subtract the energy error along the path δe. This will allow a close examination of exactly what errors have been made in the ensuing analysis.

$$\begin{aligned}-\ln \mathbb{P}_n = \frac{\Delta t}{2\epsilon}\Bigg[&\frac{1}{2}\left(\left(\frac{x_{n+1}-x_n}{\Delta t}\right) - 2F_w(x_n, x_{n+1})(x_{n+1}-x_n) + F_w(x_n, x_{n+1})^2\right) - \frac{\delta e}{\Delta t} \\ &+ \frac{U(x_{n+1}) - U(x_n)}{\Delta t} + \left(\frac{x_{n+1}-x_n}{\Delta t}\right)^2 (F_w(x_{n+1}, x_n) + F_w(x_n, x_{n+1})) \\ &+ \frac{1}{4}(F_w(x_{n+1}, x_n)^2 - F_w(x_n, x_{n+1})^2)\Bigg] - \log J_0 \end{aligned} \tag{7.20}$$

$$\begin{aligned}= \frac{\Delta t}{2\epsilon}\Bigg[&\frac{U(x_{n+1}) - U(x_n)}{\Delta t} - \frac{\delta e}{\Delta t} + \frac{1}{2}\left|\frac{x_{n+1}-x_n}{\Delta t}\right|^2 + \frac{1}{4}|F_w(x_{n+1}, x_n)|^2 + \frac{1}{4}|F_w(x_n, x_{n+1})|^2 \\ &+ \frac{1}{2}\left(\frac{x_{n+1}-x_n}{\Delta t}\right)(F_w(x_{n+1}, x_n) - F_w(x_n, x_{n+1}))\Bigg] - \log J_0 \end{aligned} \tag{7.21}$$

This formula is a fairly complicated but has the form that is appropriate to perform the splitting of the symplectic integrator[36]. Note that the last term in the square brackets will vanish for all methods introduced previously other than leapfrog.

The quantity, $\mathbb{P}_n$, is the probability of the path segment between positions x_n and x_{n+1}. Creating the full effective Hamiltonian requires taking the product over all path positions.

$$\prod_n \mathbb{P}_n \propto \prod_n \exp(-\widetilde{H}_{eff}) = \exp \sum_n -(\widetilde{H}_{eff}) \tag{7.22}$$

Note that the potential terms telescope in this expression. Because of the fixed boundary conditions they only contribute to the normalization constant. Here, I am only concerned with relative changes in the probability where the normalization constant cancels, analogously to the derivation in the PSHMC method (Equation 4.15).

$$\widetilde{H}_{eff} = \frac{U(x_{n+1}) - U(x_n)}{2\varepsilon} - \log \mathbb{P} \tag{7.23}$$

To form the final effective Hamiltonian, simply add in the auxiliary variables with unit mass;

$$H_{eff} = \widetilde{H}_e ff + \sum_{n=1}^{N} \frac{1}{2} |p_n|^2 \tag{7.24}$$

7.3 Non-Zero Time-Step Effective Hamiltonian

The full form of the effective Hamiltonian is now

$$H_{eff} = \sum_{n=1}^{N} \frac{1}{2}|p_n|^2 + \frac{1}{2}\sum_{n=1}^{N} \left| \frac{x_{n+1} - x_n}{\Delta t} \right|^2 + \sum_{n=1}^{N} \Phi_n \tag{7.25}$$

This has been written specifically to resemble the PSHMC effective Hamiltonian in Equation 4.19, shown here for reference

$$\Lambda(x,p) = \frac{1}{2\epsilon}\left(-\frac{1}{2}\int_0^T dt\big(p \cdot \boldsymbol{L}^{-1} \cdot p\big) - \frac{1}{2}\int_0^T dt\big(x \cdot \boldsymbol{L} \cdot x\big) + \int_0^T G(x) dt \right)$$

The path potential, and its spacial derivative, for this novel HMC (analogous to G) is given as

$$\begin{aligned} \Phi_n =& \frac{1}{4}|F_w(x_{n+1}, x_n)|^2 + \frac{1}{4}|F_w(x_n, x_{n+1})|^2 \\ &+ \frac{\Delta x_n}{2\Delta t}(F_w(x_{n+1}, x_n) - F_w(x_n, x_{n+1})) - \frac{2\varepsilon}{\Delta t}\log J_0 - \frac{\delta e_n}{\Delta t} \end{aligned} \tag{7.26}$$

$$\Phi_n = \Psi_n - \frac{\delta e_n}{\Delta t} \tag{7.27}$$

$$\phi_n = \sum_{n=1}^{N} \frac{\partial \Phi_n}{\partial x_n} = \frac{\partial}{\partial x_n}\left(\Psi_n + \Psi_{n-1} - \frac{\delta e_n}{\Delta t} - \frac{\delta e_{n-1}}{\Delta t} \right) \tag{7.28}$$

7.3.1 Mid-Point

I will use the weighted force corresponding to evaluation at the Mid-Point to derive the path potentials shown above. The force is defined to be

$$F_w(x_n, x_{n+1}) = F_w(x_{n+1}, x_n) = F\left(\frac{x_{n+1} + x_n}{2}\right) = F(\bar{x}) \tag{7.29}$$

The Jacobian transformation is not unity, as it is in the leapfrog integration, and must be included explicitly in the expression for the effective Hamiltonian.

$$J_n = h\frac{\partial v_n^i}{\partial x_{n+1}} = 1 - \frac{\Delta t}{2}F'(\bar{x}) \tag{7.30}$$

Importantly, this integration is symplectic, which is shown by evaluating the following (details of how to interpret this equation are given in Appendix 7.B).

$$\mathbb{J}_n = \left|\frac{\partial v_n^i}{\partial x_{n+1}}\left(\frac{\partial v_n^f}{\partial x_n}\right)^{-1}\right| = 1 \tag{7.31}$$

The energy drift is very simple for this symmetric weighted force

$$\delta e = U(x_{n+1}) - U(x_n) + (x_{n+1} - x_n)F(\bar{x}) \tag{7.32}$$

Finally, the calculation of the components of the analog to the path potential are shown for completeness. These quantities do not require further manipulation for the derivation of the algorithm, but are necessary to perform the path sampling on a computer.

$$\Psi_n = \frac{1}{2}F(\bar{x})^2 - \frac{2\varepsilon}{\Delta t}\log\left(1 - \frac{\Delta t}{2}F'(\bar{x})\right) \tag{7.33}$$

$$\frac{\partial \Psi_n}{\partial x_n} = \frac{1}{2}\left(F'(\bar{x}_n)F(\bar{x}_n) + \frac{\varepsilon}{J_n}F''(\bar{x}_n)\right) \tag{7.34}$$

$$\frac{\partial \Psi_{n-1}}{\partial x_n} = \frac{1}{2}\left(F'(\bar{x}_{n-1})F(\bar{x}_{n-1}) + \frac{\varepsilon}{J_{n-1}}F''(\bar{x}_{n-1})\right) \tag{7.35}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_n}{\Delta t}\right) = -\frac{1}{\Delta t}\left(F(x_n) - F(\bar{x}_n) + \frac{1}{2}\Delta x_n F'(\bar{x}_n)\right) \tag{7.36}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_{n-1}}{\Delta t}\right) = -\frac{1}{\Delta t}\left(-F(x_n) + F(\bar{x}_{n-1}) + \frac{1}{2}\Delta x_{n-1}F'(\bar{x}_{n-1})\right) \tag{7.37}$$

7.4 Forward Propagation of Paths

Generating updated paths will follow an analogous procedure as was shown in Section 4.3. The integration starts with Hamilton's equations

$$\frac{dx_n}{d\tau} = \frac{\partial H_{eff}}{\partial p_n} = p_n \tag{7.38}$$

$$\frac{dp_n}{d\tau} = -\frac{\partial H_{eff}}{\partial x_n} = Lx_n - \phi_n \tag{7.39}$$

Integrating Forward in Path Time τ

Using the same splitting used for the PSHMC method[13], I will find an integration scheme to generate these path updates. The velocities are chosen initially from the random Gaussian variates

$$p_{n,}^{(0)} = \sqrt{\frac{2\,\epsilon}{\Delta t}}\;\; \xi_n \tag{7.40}$$

First Half Step

$$\pi_n^{(0)} = p_n^{(0)} - \frac{\eta}{2}\phi_n^{(0)} \tag{7.41}$$

Full Step

Using Crank-Nicolson[37] on the full step, I arrive at

$$\pi_n^{(1)} - \pi_n^{(0)} = \frac{\eta}{2}\left(Lx_n^{(1)} + Lx_n^{(0)}\right) \tag{7.42}$$

$$x_n^{(1)} - x_n^{(0)} \;=\; \frac{\eta}{2}\left(\pi_n^{(1)} + \pi_n^{(0)}\right) \tag{7.43}$$

By using Crank-Nicolson at this stage, the quadratic variation (fluctuations of the path) is preserved [35], since the Crank-Nicholson method is accurate to second order in Δt.

Second Half Step

$$p_n^{(1)} = \pi_n^{(1)} - \frac{\eta}{2}\phi_n^{(1)} \tag{7.44}$$

Combining the Integrations

The Stochastic Partial Differential Equation (SPDE) is generated by eliminating $\pi_n^{(1)}$ from the system of Equations 7.42 and 7.43. Using Equation 7.41 to eliminate $\pi_n^{(0)}$ and simplify to get the SPDE given below

$$\left(\mathbb{I} - \frac{\Delta\tau}{2}L\right) x_n^{(1)} = \left(\mathbb{I} + \frac{\Delta\tau}{2}L\right) x_n^{(0)} \; - \; \Delta\tau\,\phi_n^{(0)} \; + \; \eta\, p_n^{(0)} \tag{7.45}$$

Generating the Hamiltonian Flow

The reversibility of the Hamiltonian flow allows the backward equation to be written (remembering to negate the momenta) as

$$\left(\mathbb{I} - \frac{\Delta\tau}{2}L\right) x_n^{(0)} = \left(\mathbb{I} + \frac{\Delta\tau}{2}L\right) x_n^{(1)} \; - \; \Delta\tau\,\phi_n^{(1)} \; - \; \eta\, p_n^{(1)} \tag{7.46}$$

Propagating the SPDE forward in time a second step generates the following equation

$$\left(\mathbb{I} - \frac{\Delta\tau}{2}L\right) x_n^{(2)} = \left(\mathbb{I} + \frac{\Delta\tau}{2}L\right) x_n^{(1)} \; - \; \Delta\tau\,\phi_n^{(1)} \; + \; \eta\, p_n^{(1)} \tag{7.47}$$

Note that $p_n^{(1)}$ is an unknown and is not defined by the random Gaussian variates as $p_n^{(0)}$ is. Using Equations 7.45, 7.45 and 7.45 to eliminate both of the momenta paths ($p_n^{(1)}$ and $p_n^{(0)}$) creates an equation to generate updated paths using 2 previous paths.

$$\left(\mathbb{I} - \frac{\Delta\tau}{2}L\right) x_n^{(2)} = \left(\mathbb{I} + \frac{\Delta\tau}{2}L\right)\left(2\,x_n^{(1)}\right) - \left(\mathbb{I} - \frac{\Delta\tau}{2}L\right)\left(x_n^{(0)}\right) \; - \; 2\,\Delta\tau\,\phi_n^{(1)} \tag{7.48}$$

It is now possible to calculate $x_n^{(2)}$ from $x_n^{(1)}$ and $x_n^{(0)}$ with this equation Note that this equation

can be rewritten to be computationally efficient in the following way

$$\left(\mathbb{I} - \frac{\Delta\tau}{2}\,L\right) x_n^{(2)} \;=\; \mathbb{I}\left(2\,x_n^{(1)} - x_n^{(0)}\right) + \left(\frac{\Delta\tau}{2}\,L\right)\left(2\,x_n^{(1)} + x_n^{(0)}\right) \;-\; 2\,\Delta\tau\,\phi_n^{(1)} \tag{7.49}$$

7.5 Calculating the Integration Errors

The MHMC adjustment will be performed between paths. The energy drift between two paths $x^{(1)}$ and $x^{(0)}$ is simply the difference in the effective Hamiltonian

$$\Delta E_{eff} = H_{eff}^{(1)} - H_{eff}^{(0)} \tag{7.50}$$

Rewrite two of the above equations 7.44 and 7.41 and form, ΔE_{eff}, the error in the effective energy:

$$p_n^{(1)} - p_n^{(0)} = \frac{\eta}{2}\left(Lx_n^{(1)} - \phi_n^{(1)} + Lx_n^{(0)} - \phi_n^{(0)}\right) \tag{7.51}$$

$$p_n^{(1)} + p_n^{(0)} = \frac{2}{\eta}\left(x_n^{(1)} - \frac{\eta^2}{4}\phi_n^{(1)} - x_n^{(0)} + \frac{\eta^2}{4}\phi_n^{(0)}\right) \tag{7.52}$$

Using these equations it is possible to eliminate need for explicitly calculating the momenta. The resulting energy drift (error) is now

$$\begin{aligned}\Delta E_{eff} = \Phi^{(1)} - \Phi^{(0)} &+ \frac{1}{2}\sum_{n=1}^{N_t}\left|\frac{x_{n+1}^{(1)} - x_n^{(1)}}{\Delta t}\right|^2 - \frac{1}{2}\sum_{n=1}^{N_t}\left|\frac{x_{n+1}^{(0)} - x_n^{(0)}}{\Delta t}\right|^2 \\ &+ \frac{1}{2}\sum_{n=1}^{N_t}\left(Lx_n^{(1)} - \phi_n^{(1)} + Lx_n^{(0)} - \phi_n^{(0)}\right)\cdot\left(x_n^{(1)} - x_n^{(0)} - \frac{\Delta\tau}{2}\left(\phi_n^{(1)} - \phi_n^{(0)}\right)\right)\end{aligned} \tag{7.53}$$

Recognize cancellation of terms in the expression for ΔE_{eff}

$$\begin{aligned}\Delta E_{eff} = \Phi^{(1)} - \Phi^{(0)} &- \frac{1}{2}\sum_{n=1}^{N_t}\left(\phi_n^{(1)} + \phi_n^{(0)}\right)\cdot\left(x_n^{(1)} - x_n^{(0)}\right) \\ &- \frac{1}{2}\,\frac{\Delta\tau}{2}\sum_{n=1}^{N_t}\left(Lx_n^{(1)} - \phi_n^{(1)} + Lx_n^{(0)} - \phi_n^{(0)}\right)\cdot\left(\phi_n^{(1)} - \phi_n^{(0)}\right)\end{aligned} \tag{7.54}$$

7.6 Numerical Considerations

I have used the algorithm introduced in this chapter to sample paths of a particle in the one dimensional well shown in Equation 6.2. The ensemble of paths is shown to sample paths which are in much better agreement with the equilibrium Boltzmann value (refer to Figure 6.4). The discrepancy from the exact equilibrium value is an unavoidable side effect of sampling paths, and can only be handled correctly by using a vanishingly small time step (Δt) in the path sampling algorithm, which was an oversight in the continuous time version of the algorithm.

In the Brownian formulation, all discussion of the energy error is absent. Now it is clear that the accuracy of the integration method plays an important role when the time-step (in the original SDE) is not zero. The midpoint integrator is more accurate than leapfrog. When interpreting paths, the complete sequence, $\{x_n\}$, is given. This eliminates the numerical iteration usually encountered when using the midpoint method. Furthermore, the midpoint method can be interpreted as a Stratonovich approach.

7.6.1 Advantages of Using a Non-zero Time Step

The main reason for using a method which is based on the Metropolis algorithm is that the paths are ensured to be consistent with Boltzmann statistics. The inconsistencies with the underlying Boltzmann distribution present in the PSHMC results is what drove the development of this new perspective to path sampling. Along the way, new insights are gained.

One of the more important insights comes from the full understanding of the method. The steps derived in this new method are all easily analyzed using standard numerical techniques. The complexities of the "infinities" in the continuous time PSHMC path probability are avoided. This allows the full form of the effective path potential, Φ, to be written without requiring additional assumptions, such as the reliance on the quadratic variation sum rule holding along the path.

7.6.2 Connection to the $\Delta t \to 0$ limit

The connection with the PSHMC method can be readily seen by examining the differences between the path potential, G, and its analog from the Metropolis method, Φ, shown in Equation 7.26. Referring to the formulas in Appendix 7.A.1 reveals an effective path potential Φ for leapfrog

to be

$$\Phi_n = \frac{1}{4}F(x_{n+1})^2 + \frac{1}{4}F(x_n)^2 + \frac{1}{2}\left(\frac{x_{n+1}-x_n}{\Delta t}\right)(F(x_{n+1}) - F(x_n)) - \frac{\delta e_n}{\Delta t} \tag{7.55}$$

In order to recover the original path potential, G, from the continuous time limit, two important assumptions are required. In the limit when δe_n is zero (an assumption which will be discussed later), and when the quadratic variation sum is valid along the path ($\Delta x_n \approx 2\epsilon\,\Delta t$), the middle term becomes

$$\frac{1}{2}\left(\frac{\Delta x_n}{\Delta t}\right)(F(x_{n+1}) - F(x_n)) = \frac{1}{2}\frac{(\Delta x_n)^2}{\Delta t}\frac{(F(x_{n+1}) - F(x_n))}{\Delta x_n} \approx -\epsilon\nabla^2 V(x) \tag{7.56}$$

which is equivalent to the path potential, G, given in Equation 4.12. This approximation is only valid when the noise, embodied by Δx^2, is not correlated with the particle position. The consequence of this assumption, made in the derivation of the Ito-Girsanov path probability measure, is a source of the unphysical nature of the Ito-Girsanov measure.

Let us now look at one of the above assumptions required of the error along the path Δe_n. The path potential is only valid in its full form when this error $\Delta e_n \to 0$. Using the leapfrog force, Equation 7.61 states

$$\frac{\delta e_n}{\Delta t} = \frac{U(x_{n+1}) - U(x_n)}{\Delta t} + \frac{(x_{n+1}-x_n)}{\Delta t}\left(\frac{F(x_{n+1}) + F(x_n)}{2}\right) + \frac{1}{4}\left(F(x_{n+1})^2 - F(x_n)^2\right) \tag{7.57}$$

Note that this error being non-zero also means that the integration violates detailed balance along the path. This term enters directly into the effective Hamiltonian and has an order 1 error in the last term and therefore in the path. Using the midpoint force eliminates this order one error.

7.6.3 Interpretation of δe_n

There are two errors present in this calculation, one along the path (δe_n), and the other between paths (ΔE_{eff}), which have very different interpretations in the algorithm. The error between paths (ΔE_{eff}) is used in the MHMC test, which suppresses paths which are inconsistent with the Boltzmann distribution. The error along the path (δe_n), on the other had, is not modified by any Metropolis-Hastings adjustment, as there are no rejections along the path, only between complete path proposals.

Without incorporating rejections along the path, minimizing δe_n will directly improve the quality of paths. The specific choice of weighted force determines the magnitude of this error. The role of δe is overlooked in the Brownian view of this process for non-zero time steps. Therefore, it is omitted in the implementation of the PSHMC algorithm described in Chapter 4.

Appendix 7

7.A Alternate Weighted Forces

7.A.1 LeapFrog

The leapfrog integration is central to this discussion as it is analogous to the Euler-Maruyama which was the basis for the derivation of the Ito-Girsanov probability. The local integration error of this method is $O(\Delta t^2)$ and it is a symplectic integration method (see Appendix 7.B). It is imporant to note that, unlike the other weighted forces discussed here, the error along the path per time period $\delta e/\Delta t$, which is a term in the effective Hamiltonian, has an order one error. Forces which are symmetric about the position, that is where $F(x_{n+1}, x_n) = F(x_n, x_{n+1})$, avoid this error.

$$F_w(x_n, x_{n+1}) = F(x_n) \tag{7.58}$$

$$J_n = h\frac{\partial v_n^i}{\partial x_{n+1}} = 1 \tag{7.59}$$

$$\mathbb{J}_n = -1 \tag{7.60}$$

$$\delta e_n = U(x_{n+1}) - U(x_n) + (x_{n+1} - x_n)\left(\frac{F(x_{n+1}) + F(x_n)}{2}\right) + \frac{\Delta t}{4}\left(F(x_{n+1})^2 - F(x_n)^2\right) \tag{7.61}$$

$$\Psi_n = \frac{1}{4}F(x_{n+1})^2 + \frac{1}{4}F(x_n)^2 + \frac{1}{2}\left(\frac{x_{n+1} - x_n}{\Delta t}\right)(F(x_{n+1}) - F(x_n)) \tag{7.62}$$

$$\frac{\partial \Psi_n}{\partial x_n} = \frac{1}{2}F'(x_n)F(x_n) - \frac{1}{2\Delta t}(F(x_{n+1}) - F(x_n)) - \frac{\Delta x_n}{2\Delta t}F'(x_n) \tag{7.63}$$

$$\frac{\partial \Psi_{n-1}}{\partial x_n} = \frac{1}{2}F'(x_n)F(x_n) + \frac{1}{2\Delta t}(F(x_n) - F(x_{n-1})) + \frac{\Delta x_{n-1}}{2\Delta t}F'(x_n) \tag{7.64}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_n}{\Delta t}\right) = -\frac{1}{\Delta t}\left(F(x_n) - \left(\frac{F(x_{n+1}) + F(x_n)}{2}\right) + \frac{\Delta x_n}{2}F'(x_n) + \frac{\Delta t}{2}F'(x_n)F(x_n)\right) \tag{7.65}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_{n-1}}{\Delta t}\right) = -\frac{1}{\Delta t}\left(-F(x_n) - \left(\frac{F(x_n)+F(x_{n-1})}{2}\right) + \frac{\Delta x_{n-1}}{2}F'(x_n) + \frac{\Delta t}{2}F'(x_n)F(x_n)\right) \tag{7.66}$$

7.A.2 Trapezoid

The trapezoidal force, where the average force is evaluated over the interval, has order ($O(\Delta t^3)$) local integration error. This is the same order as the midpoint method which was thuroughly investigated in Chapter 6. However, the mothod is not symplectic, as can be seen in Equation 7.6.3 below. Symplectic integration is not necessary for this path sampling method as the integration lengths are not large enough to require their use, but is favored if there is no loss in the order of integration error. Thus, the Mid-Point method is preferred.

$$F_w(x_n, x_{n+1}) = \frac{F(x_n)+F(x_{n+1})}{2} \tag{7.67}$$

$$J_n = h\frac{\partial v_n^i}{\partial x_{n+1}} = 1 - \frac{\Delta t}{2}F'(x_{n+1}) \tag{7.68}$$

$$\mathbb{J}_n = \frac{1-\Delta t\, F'(x_{n+1})/2}{1-\Delta t\, F'(x_n)/2} \tag{7.69}$$

$$\delta e = U(x_{n+1}) - U(x_n) + \Delta x_n\left(\frac{F(x_n)+F(x_{n+1})}{2}\right) \tag{7.70}$$

$$\Psi_n = \frac{1}{2}\left(F_n + F_{n+1}\right)^2 - \frac{2\varepsilon}{\Delta t}\log(1-\Delta t\, F'(x_{n+1})/2) \tag{7.71}$$

$$\frac{\partial \Psi_n}{\partial x_n} = \frac{F(x_n)+F(x_{n+1})}{4} \tag{7.72}$$

$$\frac{\partial \Psi_{n-1}}{\partial x_n} = \frac{F(x_{n-1})+F(x_n)}{4} + \frac{\epsilon\, F''(x_n)}{1-\Delta t F'(x_n)/2} \tag{7.73}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_n}{\Delta t}\right) = -\frac{1}{\Delta t}\left(\frac{F(x_n)-F(x_{n+1})}{2} + \Delta x_n\frac{F'(x_n)}{2}\right) \tag{7.74}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_{n-1}}{\Delta t}\right) = -\frac{1}{\Delta t}\left(\frac{F(x_{n-1})-F(x_n)}{2} + \Delta x_{n-1}\frac{F'(x_n)}{2}\right) \tag{7.75}$$

7.A.3 Simpsons

Simpsons method is a higher order method than the other weighted forces considered here, with local error $O(\Delta t^4)$. However, it is not a syplectic method and, in order to perform a more rigorous

analysis, was not studied in great detail here. This force could lead to an improvement in path fidelity over the other methods.

$$F_w(x_n, x_{n+1}) = \left(F(x_n) + 4F(\bar{x}_n) + F(x_{n+1})\right)/6 \tag{7.76}$$

$$J_n = h\frac{\partial v_n^i}{\partial x_{n+1}} = 1 - \frac{h^2}{12}\left(F'(x_{n+1}) + 2F'(\bar{x}_n)\right) \tag{7.77}$$

$$\mathbb{J}_n = \frac{1 - \frac{\Delta t}{6}\left(F'(x_{n+1}) + 2F'(\bar{x}_n)\right)}{1 - \frac{\Delta t}{6}\left(F'(x_n) + 2F'(\bar{x}_n)\right)} \tag{7.78}$$

$$\delta e = U(x_{n+1}) - U(x_n) + \frac{\Delta x_n}{6}\left(F(x_n) + 4F(\bar{x}_n) + F(x_{n+1})\right) \tag{7.79}$$

$$\Psi_n = \frac{1}{2}\left(\frac{F(x_n) + 4F(\bar{x}_n) + F(x_{n+1})}{6}\right)^2 - \frac{2\epsilon}{\Delta t}\ln\left(1 - \frac{\Delta t}{6}\left(F'(x_{n+1}) + 2F'(\bar{x}_n)\right)\right) \tag{7.80}$$

$$\frac{\partial \Psi_n}{\partial x_n} = \frac{1}{36}[F(x_n) + 4F(\bar{x}_n) + F(x_{n+1})][F'(x_n) + 2F'(\bar{x}_n)] - \frac{\epsilon F''(\bar{x}_n)}{3J_n} \tag{7.81}$$

$$\frac{\partial \Psi_{n-1}}{\partial x_n} = \frac{1}{36}[F(x_{n-1}) + 4F(\bar{x}_{n-1}) + F(x_n)][F'(x_n) + 2F'(\bar{x}_{n-1})] + \frac{\epsilon(F''(x_n) + F''(\bar{x}_{n-1}))}{3J_{n-1}} \tag{7.82}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_n}{\Delta t}\right) = -\frac{1}{\Delta t}\left(F(x_n) - \frac{1}{6}(F(x_n) + F(\bar{x}_n) + F(x_{n+1})) + \frac{\Delta x_n}{6}(F'(x_n) + 2F'(\bar{x}_n))\right) \tag{7.83}$$

$$-\frac{\partial}{\partial x_n}\left(\frac{\delta e_{n-1}}{\Delta t}\right) = -\frac{1}{\Delta t}\left(-F(x_n) + \frac{1}{6}(F(x_{n-1}) + F(\bar{x}_{n-1}) + F(x_n)) + \frac{\Delta x_{n-1}}{6}(F'(x_n) + 2F'(\bar{x}_{n-1}))\right) \tag{7.84}$$

7.B Symplectic Integration

It is important to note that the various weighted Forces in this chapter are not all symplectic integration schemes. Integrators which are symplectic conserve the probability of paths in the following way

$$\mathbb{P}(x_n, v_n^i) = \mathbb{P}(x_{n+1}, v_n^f) \tag{7.85}$$

It is difficult to perform this operation as the final velocities are not defined in terms of other velocities. For this reason, I will show a useful trick of how to perform this change of variables.

As an aside, the choice of a symplectic integration scheme is a smaller issue for this path

sampling method than it is for large-time integrations of a Hamiltonian system. For this path sampling algorithm, the integration time is of much shorter length than a full scale Hamiltonian simulation. Furthermore, the Metropolis-Hastings adjustment will handle the errors made with a non-symplectic integration scheme. For the results shown above, I have only used the symplectic Mid-Point weighted force.

Forward Change of Variables

Begin with the standard change of variables from $v_n^{(i)}$ to x_{n+1}

$$\mathbb{P}(x_n, v_n^i) = J_n^{\rightarrow}\, \mathbb{P}(x_n, x_{n+1}) \tag{7.86}$$

where

$$J_n^{\rightarrow} = \left| h \frac{\partial v_n^i}{\partial x_{n+1}} \right| \tag{7.87}$$

Backward Change of Variables

Similarly, changing variables from $v_n^{(f)}$ to x_n

$$\mathbb{P}(x_{n+1}, v_n^f) = J_n^{\leftarrow}\, \mathbb{P}(x_{n+1}, x_n) \tag{7.88}$$

where

$$J_n^{\leftarrow} = \left| h \frac{\partial v_n^f}{\partial x_n} \right| \tag{7.89}$$

Inverting this operation is done by inverting the operator

$$\mathbb{P}(x_{n+1}, x_n) = (J_n^{\leftarrow})^{-1}\, \mathbb{P}(x_n, v_n^f) \tag{7.90}$$

Conserving Probability Through Full Step

Combining these two change of variables

$$\mathbb{P}(x_n, v_n^i) = \mathbb{J}_n\, \mathbb{P}(x_{n+1}, v_n^f) \tag{7.91}$$

where

$$\mathbb{J}_n = J_n^{\rightarrow} \ (J_n^{\leftarrow})^{-1} = \left| \frac{\partial v_n^i}{\partial x_{n+1}} \left(\frac{\partial v_n^f}{\partial x_n} \right)^{-1} \right| \tag{7.92}$$

Therefore, the integration is symplectic if $\mathbb{J}_n = 1$, (leapfrog and the midpoint weighted forces introduced in this chapter).

Chapter 8

Concluding Remarks

The original purpose of this work was to develop a path sampling method based on a continuous-time formalism to study rare events in physical systems. The unphysical results of the PSHMC algorithm required a change of focus. Using a novel path sampling method based on the Metropolis algorithm, I have uncovered interesting and enlightening insights into path sampling with the Onsager-Machlup functional.

8.1 Ito-Girsanov Path Probability

The Ito-Girsanov expression for the probability has been studied for almost 40 years[17]. This expression was used to define a path probability

$$-\ln \mathbb{P}_{IG} = -\ln \frac{d\mathbb{P}_{\text{Informal}}}{d\mathbb{Q}_p} = \frac{1}{2\,\epsilon}\left(V(x^+) - V(x^-) + \int_0^T dt\; G(x_t) \right)$$

as discussed in Equation 4.13. This work shows that the Ito-Girsanov expression has been misidentified as a path probability distribution. The paths generated using this probability measure possess an unphysical quality, which is shown in this thesis in two ways.

This unphysical nature was uncovered by direct calculation using the PSHMC algorithm in Section 6.1.2. The ensemble of paths, all of which traverse an energy barrier in a simple 1D potential, generated using this probability distribution are demonstrably unphysical.

In section 6.2, I displayed the results of a simple calculation was performed which displays one

of the manifestations of the original OM statements, that the probability of every path of the same length have an equivalent path probability. The original OM statement states that the log of the path is proportional to the square of Gaussian variates, or equivalently, the random fluctuations along the path. This noise history is imposed on the system by the temperature bath and is not dependent on the configuration of the system. The path probability given from the Ito-Girsanov theory implies just the opposite, that some paths are more probable than others. This preference for some paths leads to a correlation between the positions and the noise, which is a fundamental violation of the theory of statistical fluctuations. The PSHMC method is simply optimizing the Ito-Girsanov measure, which leads to paths where correlation is rampant, which is at the core of the unphysical paths seen in Section 6.1.2.

This misidentification can also be understood with the following observation. In the derivation of the Ito-Girsanov probability, an infinity was regularized out of the expression using the fluctuations of a Brownian bridge ($\mathbb{Q}$). This means that the Brownian bridge is a fundamental object in $\mathbb{P}_{IG}$. Now consider the Ornstein-Uhlenbeck (OU) process[41], where the exact solution to the sampling is known. The OU process is a description of a Brownian motion of a particle who experiences a linear restoring force (quadratic potential). The OU process has a frequency (ν) spectrum which decays with a $1/\nu$ tail, but is finite at low frequencies. The spectrum of the free process (Brownian bridges for the case of double ended paths) still possess this $1/\nu$ tail but diverges at low frequencies. In schemes like the Random Walk Metropolis (RWM)[42, 43], mixing of Brownian bridges yields yet another Brownian bridge, and thus will never yield a spectrum consistent with the OU spectrum. This argument then motivates the conclusion that the Ito-Girsanov change of measure is only an indicator of differences between the motion of a free particle and the true system dynamics. The measure, $\mathbb{P}_{IG}$ is not related to the relative probability of paths.

8.2 Probing the Underlying Measure

The question of why this has not been seen in previous studies remains. There have been many independent studies which use this path probability measure[44, 45, 46], surely someone will have seen the catastrophic nature of the paths produced. This question is at the heart of this research, and the answer is very subtle. Using many sampling methods, such as Random

Walk Metropolis (RWM), the random variates are drawn from the Gaussian distribution and the sampling is performed directly with them. These methods only probe a subspace of solutions admitted from the underlying measure. Recall that the change of measure used to derive the Ito-Grisanov probability distribution is relative to the Brownian bridge. The RWM algorithm leads to sensible results for short paths as the frequency spectrum is a fair approximation in the finite case. For longer paths, numerical realizations will have an increasingly incorrect low frequency spectrum. A consequence is that the number of Metropolis rejections dramatically increases. This is one point that has been overlooked in the literature. In the PSHMC method, energy shifts between the various modes during the deterministic integration step, driven by the effective Hamiltonian. It is in this way that correlations get introduced.

8.3 The Lens of the Metropolis Algorithm

In this thesis, I develop a novel path space framework to sample double-ended paths which are thermodynamically consistent with the underlying (Boltzmann) distribution. The important contribution of this method is the errors are clearly understood. This contrasts to the hidden integration errors which are present in the Brownian SDE. Furthermore, by carefully choosing an integration procedure which minimizes the numerical error, alternative OM-like functionals can be written which yield more accurate description of the path distribution.

8.4 Future Work

As previously discussed, the largest impact of this work for the continual development of path sampling methods is the thorough analysis of the Ito-Girsanov probability. The identification that this probability produces unphysical paths was previously undiscovered and any previous works using the probability[1, 22, 23, 24, 25] could succumb to the same faults found using the PSHMC method. These works need to be reexamined to see how my findings impact them.

Further work is also required to streamline the novel path space HMC method introduced in Chapter 7. The current algorithm used to implement this method requires the Hessian of the potential, which becomes extremely complicated and unwieldy for larger systems.

Bibliography

[1] Carsten Hartmann, Ralf Banisch, Marco Sarich, Tomasz Badowski, and Christof Schütte. Characterization of rare events in molecular dynamics. *Entropy*, 16(1):350–376, December 2013.

[2] Mike P. Allen and Dominic J. Tildesley. *Computer simulation of liquids*. Oxford university press, 1989.

[3] Benedict Leimkuhler and Sebastian Reich. *Simulating Hamiltonian Dynamics*. Cambridge University Press, Cambridge, 2005.

[4] Peter G. Bolhuis, David Chandler, Christoph Dellago, and Phillip L. Geissler. TRANSITION PATH SAMPLING: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.*, 53(1):291–318, 2002.

[5] Christoph Dellago, Peter Bolhuis, and Phillip L. Geissler. Transition path sampling. *Adv. Chem. Phys.*, 123:1–78, 2002.

[6] Titus S. van Erp, Daniele Moroni, and Peter G. Bolhuis. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.*, 118(17):7762–7774, May 2003.

[7] Rosalind J. Allen, Daan Frenkel, and Pieter Rein ten Wolde. Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *J. Chem. Phys.*, 124(19):194111, May 2006.

[8] Rosalind J. Allen, Chantal Valeriani, and Pieter Rein ten Wolde. Forward flux sampling for rare event simulations. *J. Phys.: Condens. Matter*, 21(46):463102, November 2009.

[9] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. String method for the study of rare events. *Phys. Rev. B*, 66(5):052301, August 2002.

[10] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Finite temperature string method for the study of rare events†. *J. Phys. Chem. B*, 109(14):6688–6693, April 2005.

[11] Anton K. Faradjian and Ron Elber. Computing time scales from reaction coordinates by milestoning. *J. Chem. Phys.*, 120(23):10880–10889, June 2004.

[12] Anthony M. A. West, Ron Elber, and David Shalloway. Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. *J. Chem. Phys.*, 126(14):145104, April 2007.

[13] A. Beskos, F. J. Pinski, J. M. Sanz-Serna, and A. M. Stuart. Hybrid monte carlo on hilbert spaces. *Stochastic Processes and their Applications*, 121(10):2201–2230, October 2011.

[14] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21(6):1087–1092, June 1953.

[15] Simon Duane, A. D. Kennedy, Brian J. Pendleton, and Duncan Roweth. Hybrid monte carlo. *Physics Letters B*, 195(2):216–222, September 1987.

[16] Lars Onsager and S. Machlup. Fluctuations and irreversible processes. *Phys. Rev.*, 91(6):1505–1512, 1953.

[17] Robert Graham. Macroscopic theory of fluctuations and instabilities in optics and hydrodynamics. In Tormod Riste, editor, *Fluctuations, Instabilities, and Phase Transitions*, number 11 in NATO Advanced Study Institutes Series, pages 215–279. Springer US, 1975.

[18] Bernt Øksendal. *Stochastic Differential Equations*. Universitext. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[19] R.S. Berry, T.L. Beck, H.L. Davis, and J. Jellinek. SOLID-LIQUID PHASE BEHAVIOR IN MICROCLUSTERS. *Adv. Chem. Phys. Prigogine Rice Eds Wiley N. Y.*, 70:75–138, 1988.

[20] Robert Graham. Path integral formulation of general diffusion processes. *Z Physik B*, 26(3):281–290, September 1977.

[21] Detlef Dürr and Alexander Bach. The onsager-machlup function as lagrangian for the most probable path of a diffusion process. *Commun.Math. Phys.*, 60(2):153–170, June 1978.

[22] Thomas Speck, Andreas Engel, and Udo Seifert. The large deviation function for entropy production: the optimal trajectory and the role of fluctuations. *J. Stat. Mech.*, 2012(12):P12001, December 2012.

[23] Weinan E, Weiqing Ren, and Eric Vanden-Eijnden. Minimum action method for the study of rare events. *Comm. Pure Appl. Math.*, 57(5):637–656, May 2004.

[24] P. Faccioli, M. Sega, F. Pederiva, and H. Orland. Dominant pathways in protein folding. *Phys. Rev. Lett.*, 97(10):108101, September 2006.

[25] Hiroshi Fujisaki, Motoyuki Shiga, and Akinori Kidera. Onsager–machlup action-based path sampling and its combination with replica exchange for diffusive and multiple pathways. *J. Chem. Phys.*, 132(13):134101, April 2010.

[26] Robert Brown. On the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. *Edinb. New Philos. J.*, 5:358–371, 1828.

[27] Albert Einstein. Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Ann. Phys.*, 4, 1905.

[28] Crispin W. Gardiner and others. *Handbook of stochastic methods*, volume 4. Springer Berlin, 1985.

[29] Gisiro Maruyama. Continuous markov processes and stochastic equations. *Rend. Circ. Mat. Palermo*, 4(1):48–90, January 1955.

[30] Peter Eris Kloeden, Eckhard Platen, and Henri Schurz. *Numerical solution of SDE through computer experiments*. Springer Science & Business Media, 2012.

[31] Peter J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82(4):711–732, December 1995.

[32] R. Neal. MCMC using hamiltonian dynamics. *Handb. Markov Chain Monte Carlo*, pages 113–162, 2011.

[33] H.P. McKean. Onsager and machlup's work on fluctuations in statistical physics. In *The Collected Works of Lars Onsager: With Commentary, 196 Lars Onsager and P.C. Hemmer*, pages 769–771. World Scientific, Singapore, 1998.

[34] Alan J. McKane. Stochastic processes. In *Encyclopedia of Complexity and Systems Science*, pages 8766–8783. Springer, 2009.

[35] Andrew M. Stuart, Jochen Voss, and Petter Wilberg. Conditional path sampling of SDEs and the langevin MCMC method. *Commun. Math. Sci.*, 2(4):685–697, December 2004.

[36] Michael P. Allen. Introduction to molecular dynamics simulation. *Comput. Soft Matter Synth. Polym. Proteins*, 23:1–28, 2004.

[37] J. Crank and P. Nicolson. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Math. Proc. Camb. Philos. Soc.*, 43(01):50–67, 1947.

[38] A. Irbäck. Hybrid monte carlo simulation of polymer chains. *J. Chem. Phys.*, 101(2):1661–1667, July 1994.

[39] M. R. Hoare and P. Pal. Statics and stability of small cluster nuclei. *Nature*, 230(9):5–8, March 1971.

[40] F. J. Pinski and A. M. Stuart. Transition paths in molecules at finite temperature. *J. Chem. Phys.*, 132(18):184104, May 2010.

[41] G. E. Uhlenbeck and L. S. Ornstein. On the theory of the brownian motion. *Phys. Rev.*, 36(5):823–841, September 1930.

[42] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer New York, New York, NY, 2004.

[43] Jonathan C. Mattingly, Natesh S. Pillai, and Andrew M. Stuart. Diffusion limits of the random walk metropolis algorithm in high dimensions. *Ann. Appl. Probab.*, 22(3):881–930, June 2012.

[44] Katharine L. C. Hunt and John Ross. Path integral solutions of stochastic equations for nonlinear irreversible processes: The uniqueness of the thermodynamic lagrangian. *J. Chem. Phys.*, 75(2):976–984, July 1981.

[45] Daniel M. Zuckerman and Thomas B. Woolf. Efficient dynamic importance sampling of rare events in one dimension. *Phys. Rev. E*, 63(1):016702, December 2000.

[46] Artur B. Adib. Stochastic actions for diffusive dynamics: reweighting, sampling, and minimization†. *J. Phys. Chem. B*, 112(19):5910–5916, May 2008.