

# Conditional Path Sampling, Drift Relaxation, and Metadynamics

Jacob Price, Panos Stinis

March 30, 2017

## Abstract

This note summarizes the current state of the conditional path sampling and metadynamics project. We include a description of the problem, an incomplete literature review, preliminary results on a toy problem, and short and long term future directions.

## 1 Introduction

The chemical engineering, bioengineering, and nanomaterials communities are interested in the properties of materials consisting of  $O(1000)$  atoms. Examples include proteins, polymers, and carbon nanotubes. Due to the expense and difficulty of experiments, there is high demand for accurate and fast simulations of these materials.

The scale of these materials necessitates the use of molecular dynamics to reliably simulate their motion. Unfortunately, the range of timescales accessible by molecular dynamics simulations does not include the timescales of interest of the system. The size of molecular dynamic integration steps is dictated by atomic motions on the order of femtoseconds. This is far removed from the millisecond or second domain of interest. Thus, tremendous effort has been placed on accelerating molecular dynamics into the domain of biochemical interest.

Most results relate to the computation of equilibrium probability distributions of biochemical systems. One prominent example is metadynamics. Consider a molecular dynamics simulation of, for example, an unfolded protein consisting of  $N$  atoms. Through Langevin dynamics, the system performs a random walk through the  $6N$ -dimensional space of atom positions and velocities. Supposing the unfolded and folded conformations are local potential minima (with the folded conformation perhaps serving as global minimum), this simulation can be expected to explore the folded conformation given a sufficient number of time steps. The waiting time for the transition, however, could be unacceptably long.

Metadynamics accelerates the exploration of conformation space by depositing Gaussian bias periodically throughout the simulation. This induces the

simulation to preferentially explore areas of conformation space it has not yet visited. In this manner, a full description of conformation space can be found. If the bias deposits are scaled according to the amount of bias present prior to a deposition, the potential is expected to converge, and the underlying free energy landscape can be inferred. This is the famed “well-tempered metadynamics” [1].

This technique is among several that lead to the reconstruction of the equilibrium distribution for a reasonable amount of computational effort. Unfortunately, due to the repeated biasing of the underlying potential, the paths traced through conformation space no longer reflect the original dynamics. In fact, the myriad acceleration schemes developed for molecular dynamics almost universally sacrifice the physical dynamics in favor of equilibrium quantities.

However, these dynamical quantities are of great interest to theoreticians and experimentalists alike. Protein folding and polymer conglomeration are inherently dynamic processes. If we better understand these processes, we can apply this understanding to nanomaterial fabrication and learn more about the underlying chemical and biological processes.

In this note, we propose a means of sampling physical paths in the high-dimensional space of molecules consisting of  $O(1000)$  atoms. In Section 2, we overview a literature search of sampling techniques that give rise to dynamical information. In Section 3, we explicitly express the problem as a conditional sampling problem. In Section 4, we apply this methodology to a toy problem. We discuss the advantages and disadvantages, as well as means for improving our results. Finally, in Section 5, we lay out short and long term future directions. Our current position is that we are optimistic we will be able to make a contribution towards the sampling of physical transition paths in this molecular regime.

## 2 Literature Review

There has been some investigation into reconstructing dynamical information from metadynamics [2,3]. Under the assumption that passes through transition states are fast, and thus that no bias will be deposited there, some conclusions can be drawn about probable paths. The result is dynamical information about the rates of transition between metastable states, but not necessarily fully resolved samples of those transition paths, nor their distributions.

The most prominent other method is stochastic, flexible-length shooting, or transition path sampling (TPS). Transition path sampling consists of taking a successful transition, randomly selecting a point during the transition, perturbing its momenta, and then evolving the results forward and backward in time. The acceptance probability is tuned by the degree of perturbation [4–6]. These papers give me the most pause, as they seem to solve the problem we seek to address but in a more general (flexible-path length) manner. I would like to understand and discuss this method further. In [7], Juraszek and Bolhuis use high temperature simulations that achieve the transition, are equilibrated to the proper temperature, then used for transition path sampling. The result

took approximately  $10 \mu\text{s}$  and yielded 3,200 pathways, of which about 100 were independent and uncorrelated. They consist of paths of widely varying length. Modifications of TPS are described in [8].

### 3 Conditional Path Sampling

#### 3.1 Conditional Distribution of Paths

Let  $x \in \mathbb{R}^{6N}$  be the state vector of a  $N$ -particle molecular dynamics system in three dimensional space. Suppose that the system has an associated free energy surface  $U(x)$  in phase space. This may have been inferred from a metadynamics simulation. Define the drift  $f(x) = -\nabla U$ , which acts as the forcing function of the system.

Suppose the system evolves subject to noise in addition to the deterministic dynamics determined by  $f$ . Let  $X_t$  be a random variable corresponding to the system state at time  $t$ . Under Langevin dynamics, the system's evolution is governed by the stochastic differential equation

$$dX_t = f(X_t) dt + \sigma(X_t) dW_t \quad (1)$$

where  $\sigma(X_t)$  is the state-dependent amplitude of noise and  $dW_t$  is standard Brownian motion.

Suppose that at time  $t = 0$ , the state of the system is distributed according to  $g_{X_0}(x_0)$ . Let us fix an end time  $T$  at which the system is distributed according to  $h_{X_T}(x_T)$ . For example, suppose  $g$  corresponds to an ensemble of states in one free energy well (such as “unfolded”) and  $h$  corresponds to an ensemble of states in another free energy well (such as “folded”). We will thus attempt to sample all paths  $\mathbf{x}$  that are unfolded at time  $t = 0$  and folded at time  $t = T$ . The choice of  $g$ ,  $h$ , and  $T$  are left to the modeler and will be discussed at greater length in Section 5.

Consider a discretization of time from  $t = 0$  to  $t = T$ . For simplicity suppose the discretization is in equal time increments (though this is not necessary). For  $M$  increments,  $\Delta t = T/M$  and  $t_i = i\Delta t$  for  $i = 0, \dots, M$ . Let  $\mathbf{X}$  be a random variable corresponding to the path of the system  $\mathbf{X} = (X_0, X_1, \dots, X_{M-1}, X_M = X_T)$ . Then the distribution we seek to sample is:

$$p_{\mathbf{X}}(\mathbf{x}) = g_{X_0}(x_0) \prod_{i=1}^M \Pr(X_i = x_i | X_{i-1} = x_{i-1}) h_{X_T}(x_T) \quad (2)$$

where the conditional probabilities are determined by the SDE (1). We discretize (1) according to the Euler-Maruyama scheme as:

$$X_{t+1} = X_t + f(X_t) \Delta t + \sigma(X_t) \sqrt{\Delta t} \xi_t \quad (3)$$

where  $\xi_t$  is a sample from the standard normal distribution. Then the system position at  $X_{t+1}$  can be thought as drawn from a Gaussian distribution with

mean  $X_t + f(X_t)\Delta t$  and standard deviation  $\sigma(X_t)\sqrt{\Delta t}$ . Therefore

$$\Pr(X_i = x_i | X_{i-1} = x_{i-1}) = \exp \left\{ \sum_{j=1}^6 -\frac{(x_i^j - (x_{i-1}^j + f^j(x_{i-1})\Delta t))^2}{2(\sigma^j(x_{i-1}))^2\Delta t} \right\} \quad (4)$$

where  $x_i^j$  corresponds to the  $j$ th component of the state at time  $i\Delta t$ , and  $f^j$  and  $\sigma^j$  correspond to the  $j$ th component of  $f$  and  $\sigma$  respectively.

The distribution we wish to sample is given by

$$p_{\mathbf{x}}(\mathbf{x}) = \exp \left\{ \log(g_{X_0}(x_0)) + \log(h_{X_T}(x_T)) - \sum_{j=1}^6 \sum_{i=1}^M \frac{(x_i^j - (x_{i-1}^j + f^j(x_{i-1})\Delta t))^2}{2(\sigma^j(x_{i-1}))^2\Delta t} \right\}. \quad (5)$$

This is a  $6NM$  dimension distribution, and so is exceptionally difficult to sample. We will rely on Monte Carlo sampling methods ideal for sampling high dimensional problems.

### 3.2 Sampling the Distribution

Consider, for example, the hybrid Monte Carlo (HMC) algorithm, which is also known as the Hamiltonian Monte Carlo algorithm [9]. This is a Markov chain Monte Carlo algorithm that proposes samples by devising a Hamiltonian system with  $p_{\mathbf{x}}(\mathbf{x})$  as the stationary distribution. The variables of interest (in this case  $\mathbf{x}$ ) are considered position variables  $\mathbf{q}$ , and the system is augmented with virtual velocity variables  $\mathbf{p}$ . Then a Hamiltonian is defined as

$$H(\mathbf{q}, \mathbf{p}) = -\log(p_{\mathbf{x}}(\mathbf{q})) + \frac{\mathbf{p}^T \mathbf{p}}{2}. \quad (6)$$

The dynamics of a Hamiltonian system are given by

$$\begin{aligned} \frac{d\mathbf{q}}{dt} &= \nabla_{\mathbf{p}} H \\ \frac{d\mathbf{p}}{dt} &= -\nabla_{\mathbf{q}} H \end{aligned}$$

Given a path  $\mathbf{x} = \mathbf{q}_0$ , virtual velocities  $\mathbf{p}_0$  are drawn from a standard normal distribution. The system  $(\mathbf{q}, \mathbf{p})$  is then evolved using a symplectic integrator, such as the velocity Verlet algorithm,  $L$  virtual time steps of step size  $\delta t$  to produce  $(\mathbf{q}', \mathbf{p}')$ . The proposed new path  $\mathbf{q}$  is accepted with probability

$$\Pr(\text{accept}) = \min \left\{ 1, e^{H(\mathbf{q}, \mathbf{p}) - H(\mathbf{q}', \mathbf{p}')} \right\} \quad (7)$$

Because symplectic integrators conserve the Hamiltonian, this acceptance probability is quite high. This methodology allows the Monte Carlo sampling scheme

to take large steps in conformation space with high acceptance probability.  $L$  and  $\delta t$  are parameters that can be used to tune the algorithm.

This sampling method will converge to the correct distribution, however in practice it is insufficient. Namely, samples are highly correlated, and it may take thousands of HMC steps in order to produce an uncorrelated sample. We developed and tested some remedies to this on a toy problem in section 4, and are optimistic that we can apply these alterations to a larger scale computation.

## 4 Preliminary Results

We consider a difficult problem on which to test our methods. The following is inspired largely by [10]. Consider a one-dimensional potential  $U(x) = 10x^2(x^2 - 2)$ . The two wells are quite deep. Suppose the noise level  $\sigma$  is small and independent of the state. We consider a path that begins at time  $t = 0$  with a Gaussian distribution of standard deviation  $s_0$  and mean  $-1$ , and ends at time  $t = 1$  with a Gaussian distribution of standard deviation  $s_T$  and mean  $1$ . That is, we seek to sample paths that make the transition from the left well to the right well some time before  $t = 1$ .

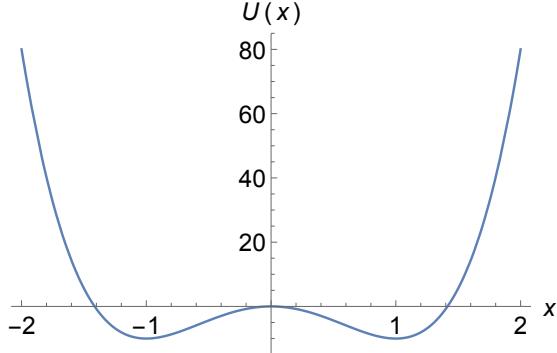


Figure 1: The free energy landscape of our toy problem  $U(x) = 10x^2(x^2 - 2)$ .

The SDE governing these dynamics is

$$dX_t = f(X_t) dt + \frac{1}{2} dW_t \quad (8)$$

where  $f(X_t) = -40x(x^2 - 1)$ . We discretize the path with step size  $\Delta t = 1/M$  and let  $\mathbf{X} = (X_0, X_1, \dots, X_M = X_T)$  be a random path. The distribution of these paths we seek to sample is

$$p_{\mathbf{X}}(\mathbf{x}) = \exp \left\{ -\frac{(x_0 - (-1))^2}{2s_0^2} - \frac{(x_T - 1)^2}{2s_T^2} - \sum_{i=1}^M \frac{(x_i - (x_{i-1} + f(x_{i-1})\Delta t))^2}{2\sigma^2} \right\}. \quad (9)$$

I use the following metaphor to think about operations on these paths. Think about the position of the particle at each time step as a ball sitting on the free energy landscape. Each of these balls would like to fall down into the bottom of whatever well it is in, but the stochastic term serves to jostle each ball continuously. In addition, each ball is connected to its two nearest neighbors by a spring that seeks to pull neighboring points together with spring constant  $k = \frac{1}{\sigma\Delta t}$ . Finally, the two points at the end are *also* connected to their target distributions by springs, but with spring constants  $\frac{1}{s_0^2}$  and  $\frac{1}{s_T^2}$ . Thus, the effectiveness of the model hinges upon the ability of the spring constants connecting points to overcome the steepness of the wells. Initial and final distributions with smaller standard deviations can more effectively be sampled.

By implementing the HMC algorithm, we attempt to sample this distribution. The results look qualitatively correct (Figure 2), but are highly autocorrelated (Figure 3).

There is an additional problem with this method that is not clear from these figures. The system explores the state space in a predictable way. Initially, transitions occur just before the end of time. Slowly, the algorithm sweeps through different transition times, creating the qualitatively reasonable depiction of transitions distributed throughout the interval. Figure 4 depicts a rough heuristic of the time of transition plotted against the sample number. Clearly our exploration is not mixing well.

The problem stems from the stiffness of the problem. For small timesteps, subsequent points in the path cannot be far from one another, as they are drawn from Gaussian distributions with variance  $\sigma\Delta t$ . Consequently, significant changes to the path take a very long time to accrue. One solution is to relax the timestep. For larger  $\Delta t$ , subsequent points can be further from one another, and it is easier to modify the path. Using the ball and spring example, it is as if we have greatly relaxed the spring constant binding neighboring balls, making it easier to move the path around.

The result is shorter autocorrelation times (Figure 6) and qualitatively similar paths (Figure 5). We can see that the path space is explored more efficiently as well (Figure 7). This, however, is simply a bandage. First, the autocorrelations and path explorations are markedly improved, but still quite bad. Second, suppose the desired level of resolution is finer than the resolution with acceptable autocorrelation times. Another solution was needed.

We draw inspiration from the drift relaxation method described in [11]. Besides the strength of the springs, another contributing factor to the stiffness of the problem is the steepness of the wells. For a path that traverses between wells, the forces must balance perfectly between the pull to the original well and the pull of the points ahead in the chain. This leaves little room for modifying paths to find new valid ones, and slows the exploration speed. If the well were less steep, but the spring constants between balls were kept the same, the stiffness of the problem would be reduced, and finding new paths would be easier.

Additionally, if we sample the path distributions for shallower wells, a given

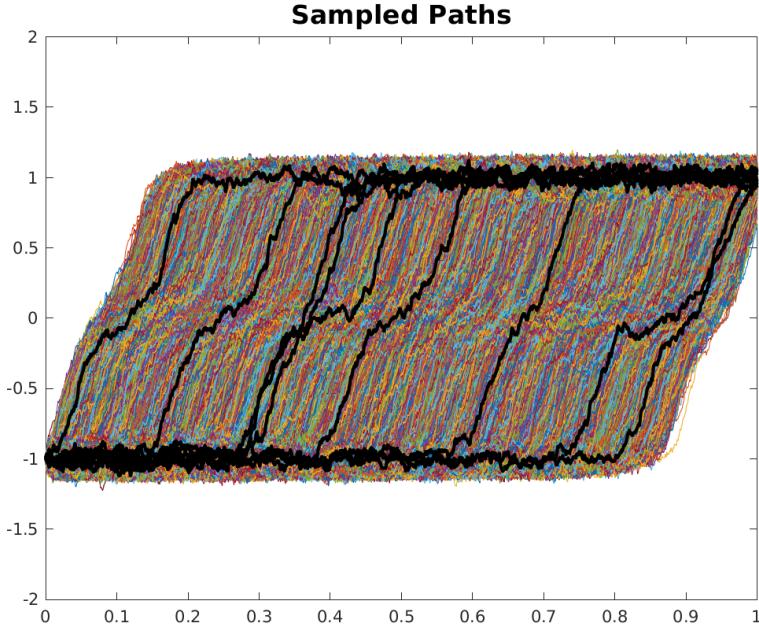


Figure 2: Paths sampled directly from (9) for  $\Delta t = 2^{-10}$ . A selection is highlighted in black for ease of viewing. 100,000 samples were computed, and the first 100 paths were discarded. The initial condition is a Gaussian centered at  $-1$  with standard deviation 0.01. The final condition is a Gaussian centered at  $1$  with standard deviation 0.1. The noise in the SDE is  $\sigma = 0.5$ . The timestep of HMC was selected from a uniform distribution between  $\delta t = 0.002$  and  $\delta = 0.006$  and the end time of each HMC step was 1 (so  $L = 1/\delta t$ ). This is considered the most direct method of sampling.

path that traverses the deep well will not be significantly different from the path that traverses the shallow well. Indeed, it will simply need to sharpen the speed of transition, but otherwise the paths are qualitatively the same. This is the inspiration behind drift relaxation. We sample transitions in shallower wells, which are easier to sample, and use those as initial guesses for sampling the deep well. We can even create a sequence of progressively shallower wells, and allow a sampled path to cascade through them until we reach the original dynamics. Figures 9 and 10 demonstrate how much easier it is to sample the shallower wells.

But again, we are met with a sense of “too little, too late.” The shallower wells are indeed easier to sample (the autocorrelations die faster, the transition time moves around more quickly), but they are still too inefficient. They would

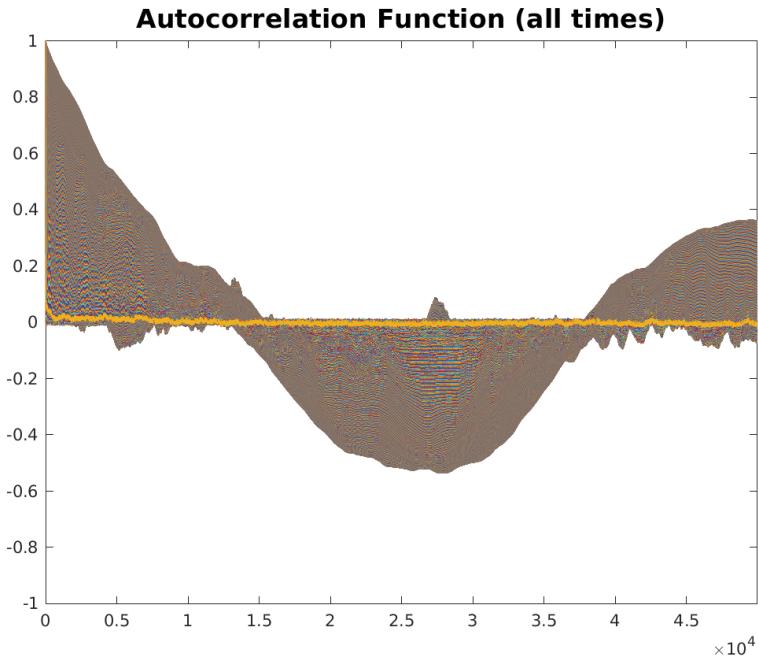


Figure 3: The autocorrelation functions of each position  $n\Delta t$  along the path ( $n = 0, 1, \dots, T/\Delta t$ ) for direct samples of (9). This was computed using a fast Fourier transform approach from the 100,000 sample paths depicted in Figure 2 with  $\Delta t = 2^{-10}$ . Note that the overall autocorrelation of the path as a whole reaches its minimum after around 10,000 samples. That is to say, in order to generate a set of decorrelated samples, one must discard all but every 10,000 paths. Note also the periodic nature of the correlation. These all contribute to making the direct sampling method unfeasible.

need to be orders of magnitude more efficient in order to improve our base sampling method, because sampling must now be done in series in order to generate a single independent path. Furthermore, because the exploration of transitions follows a seemingly deterministic path (i.e. it begins with transitions near the end time and sweeps back and forth), computing a fixed number of samples before sending one to the next step will not necessarily yield independent or decorrelated initial guesses.

The next potential solution was proposed by Jonathan Weare in [10]. He observed, as we did, that relaxing the discretization led to more efficient explorations of the domain. Like us, he noted the qualitative similarity between paths at different discretizations. Finally, he took inspiration from the method of parallel tempering.

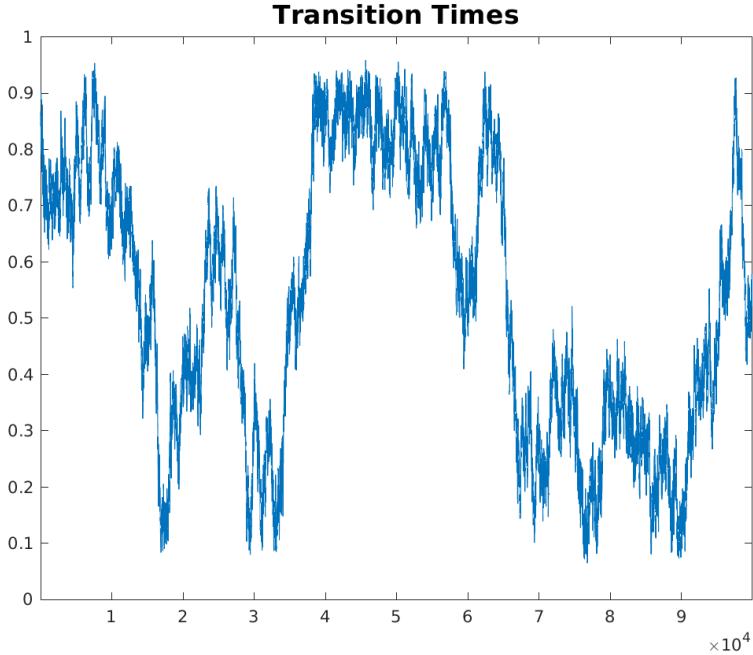


Figure 4: The transition times of the 100,000 samples taken directly from (9). The  $x$ -axis is the sample number and the  $y$ -axis is the time (between zero and one) that the transition between the  $X = -1$  well and the  $X = 1$  well occurred. This was computed heuristically as the proportion of points on each path that are below 0. Note the scale of the axis and the slow speed at which the accepted paths explore potential transition times.

Namely, consider running  $n$  parallel sampling algorithms like the one above, but with different discretizations (the  $n$ th having the desired resolution). The most convenient is to, at each level  $l$ , let the path be discretized into  $2^l$  steps. Thus, each level higher has twice as many steps as the previous.

After each algorithm samples a path, randomly select  $l$  as one of the  $n - 1$  underresolved samples and attempt to exchange this path with the one above it using a Metropolis accept-reject step. The interested reader may see [10] for tremendous detail, but here we will give a brief description of this accept-reject scheme.

Any Metropolis scheme has the following procedure. Begin with an initial state. In this case, this is the ordered pair of paths  $(\mathbf{x}_l, \mathbf{x}_{l+1})$ . We are seeking to conduct Metropolis steps that do not modify the underlying distribution we are trying to sample. Because the path sampling algorithms run in parallel, they are independent of one another. Thus, the distribution we are trying to sample

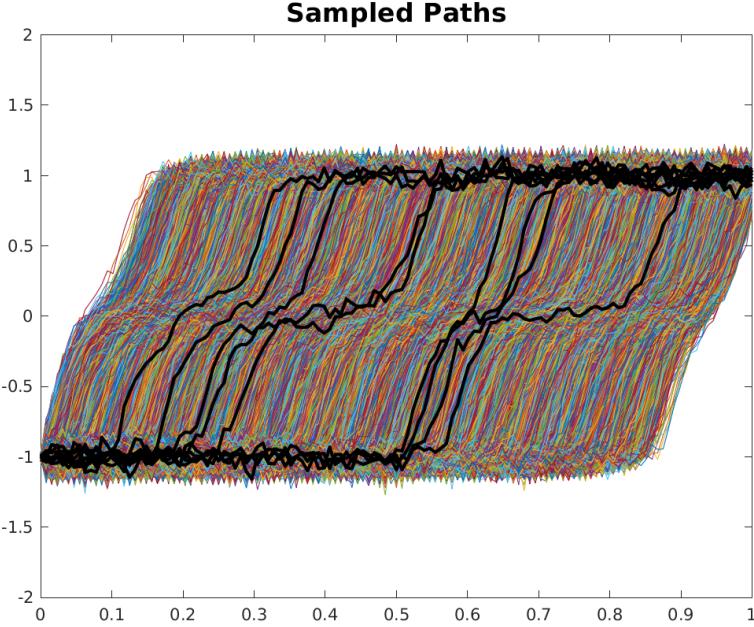


Figure 5: Paths sampled directly from (9) for  $\Delta t = 2^{-7}$ . A selection is highlighted in black for ease of viewing. 100,000 samples were computed, and the first 100 paths were discarded. The initial condition is a Gaussian centered at  $-1$  with standard deviation 0.01. The final condition is a Gaussian centered at  $1$  with standard deviation 0.1. The noise in the SDE is  $\sigma = 0.5$ . The timestep of HMC was selected from a uniform distribution between  $\delta t = 0.002$  and  $\delta = 0.006$  and the end time of each HMC step was 1 (so  $L = 1/\delta t$ ). Compare these results to Figure 2. The paths appear qualitatively the same.

for this ordered pair is simply  $p_{\mathbf{x}_l}(\mathbf{x}_l)p_{\mathbf{x}_{l+1}}(\mathbf{x}_{l+1})$ .

We then define a distribution from which to sample new proposals  $(\mathbf{x}'_l, \mathbf{x}'_{l+1})$ , conditioned upon the current state. Call this conditional distribution  $g((\mathbf{x}'_l, \mathbf{x}'_{l+1}) | (\mathbf{x}_l, \mathbf{x}_{l+1}))$ . This should describe the path exchange we desire. The new level  $l$  path  $\mathbf{x}'_l$  would consist of every other entry in  $\mathbf{x}_{l+1}$  with probability one, and the new level  $l+1$  path would consist of  $\mathbf{x}_l$  for every other entry, with the ones in between drawn from the conditional distributions:

$$p_{X_i|X_{i-1}, X_{i+1}}(x_i|x_{i-1}, x_{i+1}) = \exp \left\{ -\frac{(x_i - (x_{i-1} + f(x_{i-1})\Delta t))^2}{2\sigma^2 \Delta t_{l+1}} - \frac{(x_{i+1} - (x_i + f(x_i)\Delta t))^2}{2\sigma^2 \Delta t_{l+1}} \right\} \quad (10)$$

The presence of the  $f(x_i)$  makes this distribution very hard to sample from.

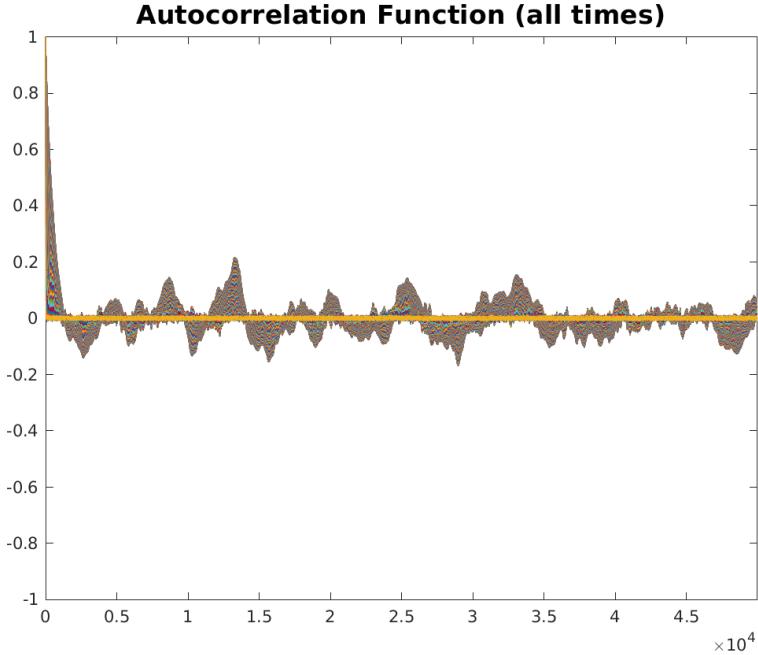


Figure 6: The autocorrelation functions of each position  $n\Delta t$  along the path ( $n = 0, 1, \dots, T/\Delta t$ ) for direct samples of (9) with  $\Delta t = 2^{-7}$ . This was computed using a fast Fourier transform approach from the 100,000 sample paths depicted in Figure 5. Compare this against Figure 3, and observe that the autocorrelation reaches its minimum after only about 2,000 samples instead of 10,000. Furthermore, though the periodic behavior is still present, it leads to less pronounced autocorrelations. These results, however, still leave much to be desired.

Instead, we sample simply from a Gaussian distribution centered at the midpoint of  $X_{i-1}$  and  $X_{i+1}$  with variance  $\sigma\Delta t_{l+1}$ . Thus, we have constructed our proposed new paths  $(\mathbf{x}'_l, \mathbf{x}'_{l+1})$ , which have exchanged entries. The acceptance probability of this exchange is then defined as:

$$P(\text{accept}) = \min \left\{ 1, \frac{p_l(\mathbf{x}'_l)p_{l+1}(\mathbf{x}'_{l+1})g((\mathbf{x}_l, \mathbf{x}_{l+1}) | (\mathbf{x}'_l, \mathbf{x}'_{l+1}))}{p_l(\mathbf{x}_l)p_{l+1}(\mathbf{x}_{l+1})g((\mathbf{x}'_l, \mathbf{x}'_{l+1}) | (\mathbf{x}_l, \mathbf{x}_{l+1}))} \right\}. \quad (11)$$

With probability  $P(\text{accept})$  we change  $(\mathbf{x}_l, \mathbf{x}_{l+1}) \rightarrow (\mathbf{x}'_l, \mathbf{x}'_{l+1})$ , essentially exchanging paths between levels of discretization. Otherwise, we leave the paths as they are.

The results of this parallel tempering-like method are tremendous. The paths look qualitatively similar to the direct sampling method (Figure 11), but

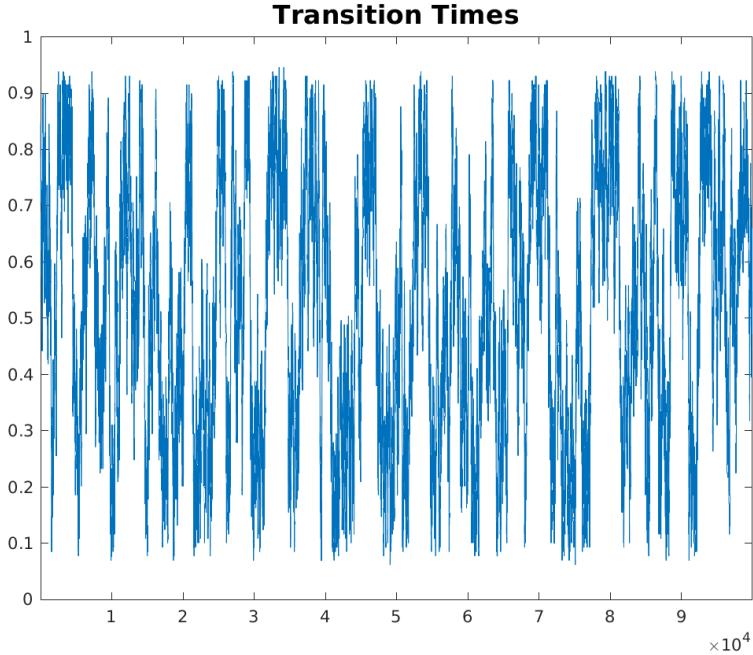


Figure 7: The transition times of the 100,000 samples taken directly from (9) for  $\Delta t = 2^{-7}$ . Compare against Figure 4. Note the increased speed with which the algorithm explores potential transition times. This contributes to the attenuated autocorrelation time.

the autocorrelations (Figure 12) and exploration of different transition times (Figure 13) demonstrate the hugely improved efficiency. Furthermore, the path exchanges succeed with an acceptance rate of about 0.5 for high levels, but lower for low levels.

We applied a similar idea to path exchanging algorithms with parallel simulations of differing well depth. The Metropolis acceptance probability is simpler in this case, as the paths are already the same length and no additional sampling is required. The results were similarly promising (Figures 14-16). This is the current state of the conditional path sampling project. We have many ideas for next steps, and are eager to explore them.

## 5 Next Steps

### 5.1 Investigation of Robustness

There are numerous algorithms included in this note, and we have only just begun to explore them. We should explore parameter space to learn more about optimizing these methods. For example,

- What is the best way to quantify decorrelation times so that we can compare different methods?
- Are there better choices of  $\delta t$  and  $L$  from HMC that would facilitate path exploration?
- What other Monte Carlo algorithms exist for high-dimensional spaces? Might they be better suited than HMC for this specific problem?
- Does adding additional levels of lower resolution improve the time-sampling algorithm?
  - On a related note, what is the best discretization for stable low-resolution samples? Our current implicit scheme gives nonsense results.
- Is there an optimal well-depth ratio for the free energy relaxation methods?
- Can we somehow couple the time exchange and energy relaxation methods to more easily produce independent samples?
  - Perhaps this will provide the means to incorporate the serial drift relaxation as originally proposed by Panos.
  - In each shallow well, time exchange algorithms will accelerate the mixing of potential paths, which can then be used as seeds for deeper wells (with time exchange algorithms).
  - We can construct independent samples perhaps faster than the decorrelation time.
- How do these many methods perform against different free energy potentials?
- Is there any practical or theoretical gain from these ideas over those of transition path sampling?

### 5.2 Theoretical Investigations

At present, we are relying largely on physical intuition to interpret the sampled paths. It is difficult to validate the method and identify if bugs are present when we do not know what the paths themselves ought to look like. Further investigation into theoretical benchmarks we can use for validation is warranted.

For example, if one can determine the distribution of first passage times between wells, we can compare this to histograms computed from the data used in Figures 4, 7 13, and 16.

One can also compute the relative probability of paths using  $p_X(\mathbf{x})$  in order to determine whether paths are being computed roughly in proportion with their probabilities in the long run.

Many of these validation methods will require longer simulations and more data. We may need to move to a better language than Matlab and employ better data curating. Furthermore, parallelizing the simulations and using more powerful computers will also speed the production of data for comparison.

### 5.3 Generalization

The generalization of this method to multiple dimensions is not difficult. We simply need to compute gradients instead of derivatives. This will bring us closer to a formulation that is applicable to protein folding experiments.

Another step towards application would be to generalize the initial and final distributions. I believe the initial and final distributions will be functions of the particle positions only through the reaction coordinates we discussed. This is actually not a problem, as we can rewrite our path sampling scheme with an additional chain rule step to connect the two levels.

Once these two steps are done, we can potentially apply the problem to a simple protein folding scenario, and through this try to poke and prod for other problem areas we can address afterwards.

### 5.4 Metadynamics and Conditional Path Sampling

We have not yet forgotten the original proposal: to use the metadynamics-generated potentials as intermediates in the proposed well relaxation scheme. This may still be a fruitful area of inquiry, as the potential relaxation method using metadynamics intermediaries coupled with time exchanges may be the most efficient path sampling method we can propose. Perhaps better even than using scaled free energies as our intermediaries. Once the above extensions have been completed, it would be relatively simple to implement this and begin testing.

Furthermore, metadynamics can give us information about the starting and ending distributions  $g$  and  $h$  as well as the transition time  $T$ . If the transition time is itself subject to a known distribution, we can sample that distribution and use that to compute paths. By sampling the  $T$  distribution before each independent path is generated, we can create paths with correctly distributed lengths.

## References

- [1] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical review letters*, 100(2):020603, 2008.
- [2] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physical review letters*, 111(23):230602, 2013.
- [3] Matteo Salvalaglio, Pratyush Tiwary, and Michele Parrinello. Assessing the reliability of the dynamics reconstructed from metadynamics. *Journal of chemical theory and computation*, 10(4):1420–1425, 2014.
- [4] Phillip L Geissler, Christoph Dellago, and David Chandler. Kinetic pathways of ion pair dissociation in water. *The Journal of Physical Chemistry B*, 103(18):3706–3710, 1999.
- [5] Peter G Bolhuis, David Chandler, Christoph Dellago, and Phillip L Geissler. Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annual review of physical chemistry*, 53(1):291–318, 2002.
- [6] Christoph Dellago, Peter G Bolhuis, Félix S Csajka, and David Chandler. Transition path sampling and the calculation of rate constants. *The Journal of chemical physics*, 108(5):1964–1977, 1998.
- [7] J Juraszek and PG Bolhuis. Sampling the multiple folding mechanisms of trp-cage in explicit solvent. *Proceedings of the National Academy of Sciences*, 103(43):15859–15864, 2006.
- [8] Ryan Gotchy Mullen, Joan-Emma Shea, and Baron Peters. Easy transition path sampling methods: Flexible-length aimless shooting and permutation shooting. *Journal of chemical theory and computation*, 11(6):2421–2428, 2015.
- [9] Radford M Neal et al. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, 2:113–162, 2011.
- [10] Jonathan Weare. Parallel marginalization monte carlo with applications to conditional path sampling. *arXiv preprint arXiv:0709.1721*, 2007.
- [11] Panos Stinis. Conditional path sampling for stochastic differential equations through drift relaxation. *Communications in Applied Mathematics and Computational Science*, 6(1):63–78, 2011.

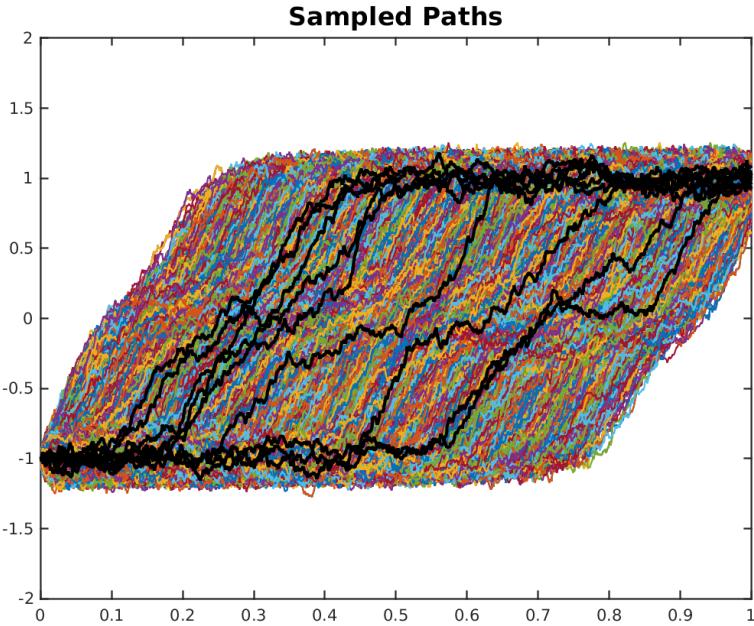


Figure 8: Paths sampled directly from (9) for  $\Delta t = 2^{-10}$ , but with the free energy scaled by  $1/2$  (yielding shallower wells and a less stiff problem). A selection is highlighted in black for ease of viewing. 100,000 samples were computed, and the first 100 paths were discarded. The initial condition is a Gaussian centered at  $-1$  with standard deviation 0.01. The final condition is a Gaussian centered at  $1$  with standard deviation 0.1. The noise in the SDE is  $\sigma = 0.5$ . The timestep of HMC was selected from a uniform distribution between  $\delta t = 0.002$  and  $\delta t = 0.006$  and the end time of each HMC step was 1 (so  $L = 1/\delta t$ ). Compare these results to Figure 2. The paths appear qualitatively the same, but less steep.

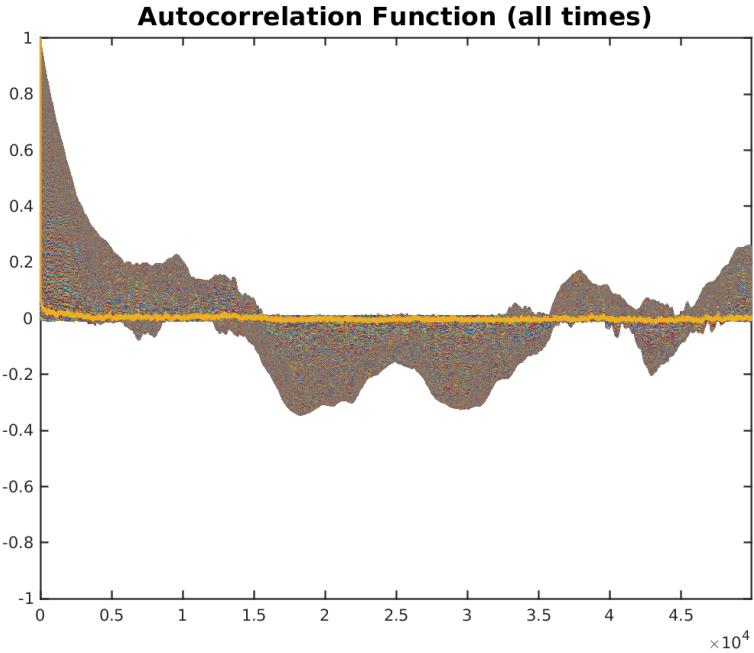


Figure 9: The autocorrelation functions of each position  $n\Delta t$  along the path ( $n = 0, 1, \dots, T/\Delta t$ ) for direct samples of (9) with  $\Delta t = 2^{-10}$ , but with the free energy scaled by  $1/2$ . This was computed using a fast Fourier transform approach from the 100,000 sample paths depicted in Figure 8. Compare this against Figures 3 and 6, and observe that the autocorrelation reaches its minimum after only about 5,000 samples instead of 10,000. Furthermore, though the periodic behavior is still present, it leads to slightly less pronounced autocorrelations. These results, however, still leave much to be desired, especially in comparison to the more promising results from lower resolution paths of the original dynamics.

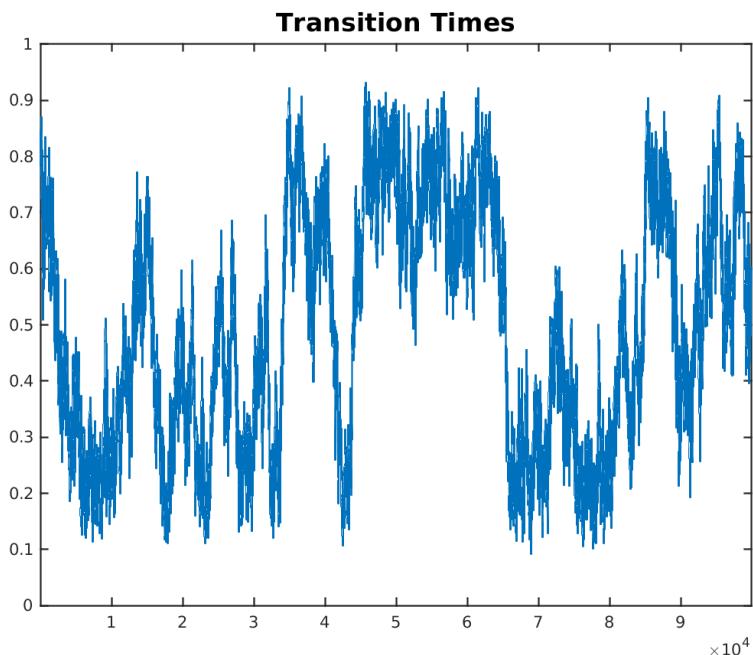


Figure 10: The transition times of the 100,000 samples taken directly from (9) for  $\Delta t = 2^{-10}$  but with the free energy scaled by 1/2. Compare against Figures 4 and 7. The algorithm explores new transition times more quickly than the original algorithm, but more slowly than the lower resolution sampling method.

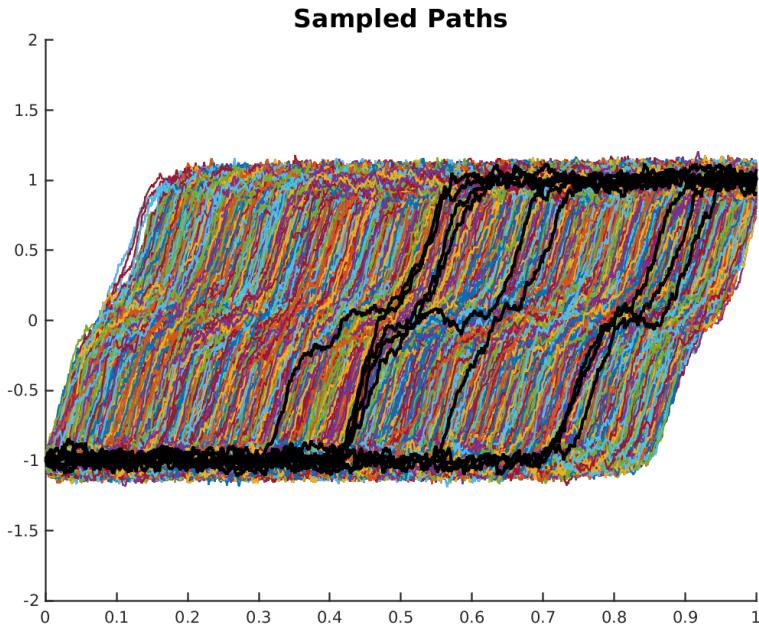


Figure 11: Paths sampled using the time exchange algorithm (a selection is highlighted in black for ease of viewing). Pictured are only paths with  $\Delta t = 2^{-10}$ . Parallel simulations with  $\Delta t = 2^{-7}, 2^{-8}$ , and  $2^{-9}$  were also conducted, with exchanges between pairs. 10,000 samples were computed, and the first 100 paths were discarded. The initial condition is a Gaussian centered at  $-1$  with standard deviation 0.01. The final condition is a Gaussian centered at  $1$  with standard deviation 0.1. The noise is  $\sigma = 0.5$ . The timestep of HMC was selected from a uniform distribution between  $\delta t = 0.002$  and  $\delta = 0.006$  and the end time of each HMC step was 1 (so  $L = 1/\delta t$ ). Compare this to Figure 2 to see that the paths are qualitatively the same.

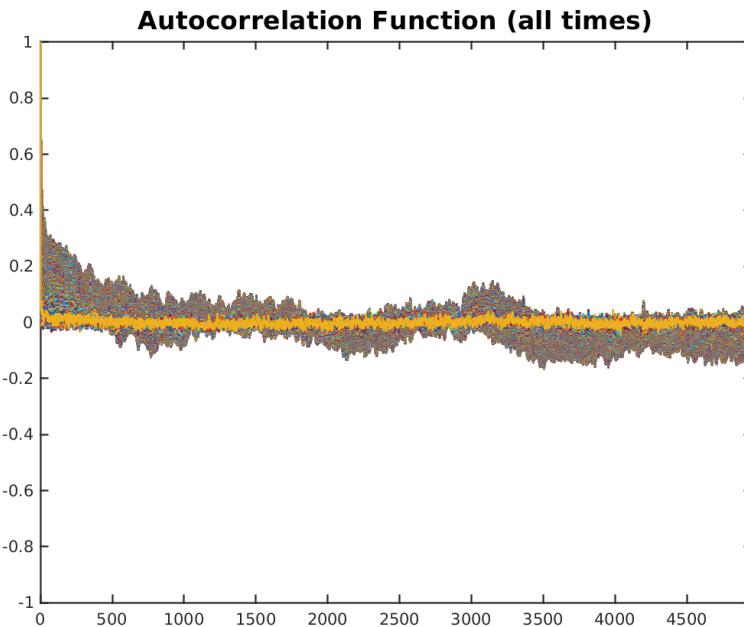


Figure 12: The autocorrelation functions of each position  $n\Delta t$  along the path ( $n = 0, 1, \dots, T/\Delta t$ ) for samples from the time-exchange algorithm with  $\Delta t = 2^{-10}$  and auxiliary simulations with  $\Delta t = 2^{-9}, 2^{-8}$ , and  $2^{-7}$ . This was computed using a fast Fourier transform approach from the 10,000 sample paths depicted in Figure 11. Compare this against Figures 3 and 6, and note the different scale on the  $x$ -axis. Samples are now decorrelated after only about 500 samples. Furthermore, the periodicity is essentially eliminated. These results are far more promising.

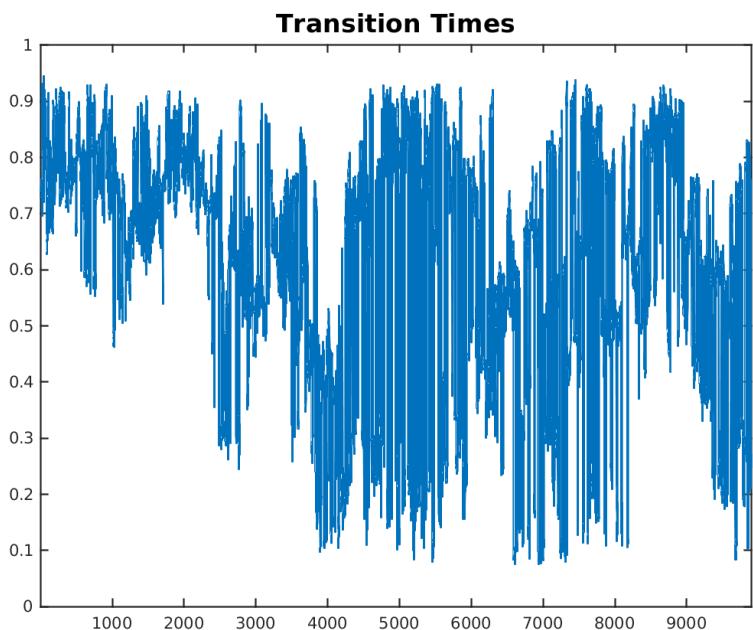


Figure 13: The transition times of the 10,000 samples from the time-exchange algorithm for  $\Delta t = 2^{-10}$  and auxiliary simulations with  $\Delta t = 2^{-9}, 2^{-8}$ , and  $2^{-7}$ . Compare against Figures 4 and 7. The algorithm explores new transition times much more quickly than the original algorithm, and even rivals the exploration speed of the lower resolution model.

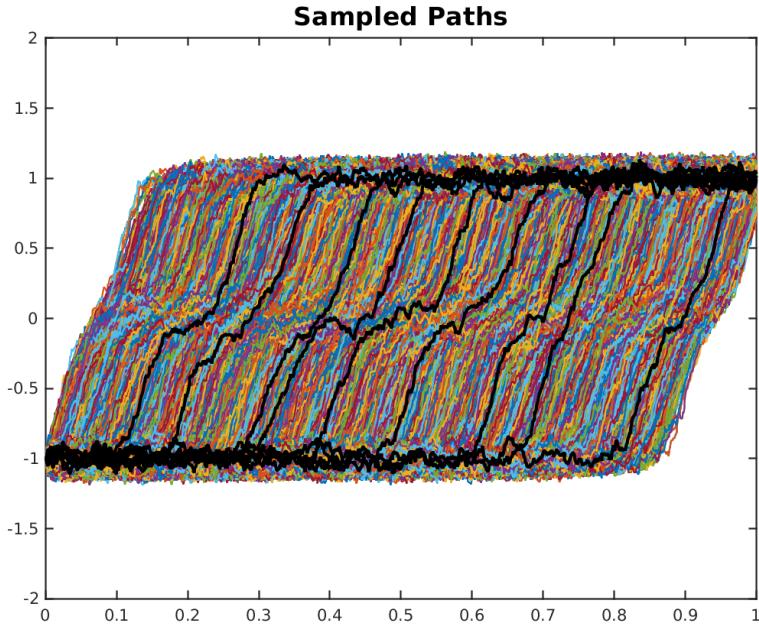


Figure 14: Paths sampled using the relaxed free energy exchange algorithm (a selection is highlighted in black for ease of viewing). Pictured are only paths of the original dynamics with  $\Delta t = 2^{-10}$ . Parallel simulations with  $\Delta t = 2^{-10}$  with the free energy scaled by 0.75, 0.8, 0.85, 0.9 and 0.95 were also conducted, with exchanges between neighbors. 100,000 samples were computed, and the first 100 paths were discarded. The initial condition is a Gaussian centered at  $-1$  with standard deviation 0.01. The final condition is a Gaussian centered at  $1$  with standard deviation 0.1. The noise is  $\sigma = 0.5$ . The timestep of HMC was selected from a uniform distribution between  $\delta t = 0.002$  and  $\delta = 0.006$  and the end time of each HMC step was 1 (so  $L = 1/\delta t$ ). Compare this to Figure 2 to see that the paths appear the same qualitatively.

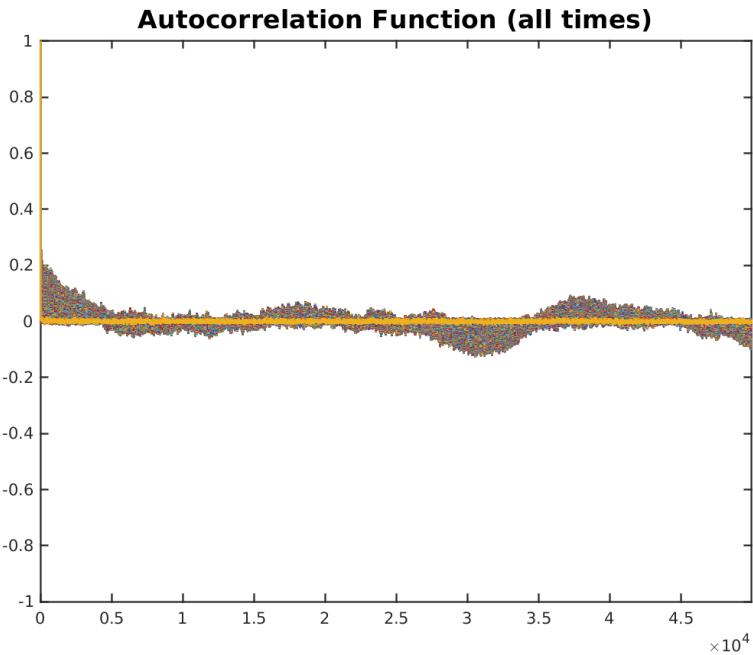


Figure 15: The autocorrelation functions of each position  $n\Delta t$  along the path ( $n = 0, 1, \dots, T/\Delta t$ ) for samples from the free energy exchange algorithm with  $\Delta t = 2^{-10}$  and auxiliary simulations with the free energy scaled by 0.75, 0.8, 0.85, 0.9 and 0.95. This was computed using a fast Fourier transform approach from the 100,000 sample paths depicted in Figure 14. Compare this against Figures 3, 6, and 12. Samples are now decorrelated after only about 5,000 samples and have much lower autocorrelations beyond that. Furthermore, the periodicity is essentially eliminated. These results are promising, but not as impressive as the time exchange autocorrelations.

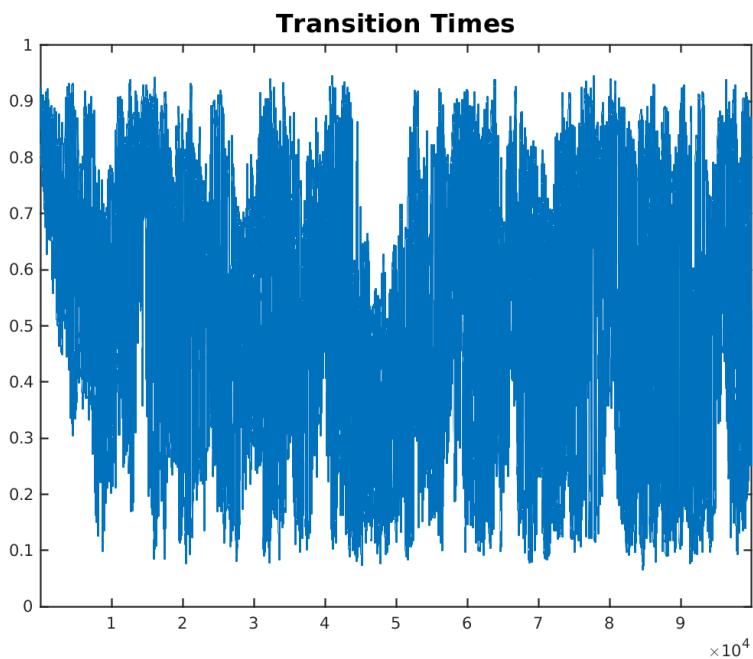


Figure 16: The transition times of the 100,000 samples from the free energy exchange algorithm for  $\Delta t = 2^{-10}$  and auxiliary simulations with the free energy scaled by 0.75, 0.8, 0.85, 0.9 and 0.95. Compare against Figures 4, 7, and 13. The algorithm explores new transition times much more quickly than either the original algorithm or the time exchange algorithm.