Computing Correlation Times for Parallel Replica Methods

By

Dominic Rufa

A Thesis Submitted to the Faculty of the Department of Chemistry, University of Florida in Partial Fulfillment of the Requirements for the Degree of

Bachelor of Science

Chemistry

University of Florida

February 2018

Abstract

Parallel Replica Dynamics, and the more recent Parallel Trajectory Splicing, are promising time parallelization methods that have been shown to accelerate the dynamics of several classes of atomistic systems characterized by infrequent event transition events. However, perhaps the most notable obstacle inhibiting the extensibility of these methods to arbitrary systems is the dilemma associated with estimating the relaxation time for replica ensembles within metastable states. This report presents three methods to accurately calculate this relaxation, or correlation time (τ_{corr}), for a class of systems. Furthermore, simulations are conducted on both model potentials and Ala_3 in silico to illustrate the advantages and limitations of each method.

Acknowledgements

I would like to thank my advisor, Adrian Roitberg, for his guidance. I would also like to thank Art Voter and Danny Perez at Los Alamos National Laboratory for introducing me to this project. I appreciate their availability and time dedicated to the fruitful discussions concerning the mathematical details and formalizations of the theory herein. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Science Undergraduate Laboratory Internship (SULI) program. This material is also based upon work supported by the University of Florida Undergraduate Scholars Program (USP) of the Center for Undergraduate Research (CUR).

List of Figures and Tables

1. Figure 1: Plot of potential energy and its associated normalized QSD……. 23
2. Figure 2: Log plot of survival probability of a 10,000 replica, single-trajectory ensemble with an associated linear regression plot……...….25
3. Figure 3: Log plot of survival times of a 5,000 vicious walker pair ensemble with an associated linear regression plot…….....26
4. Figure 4: Contour plot of the potential surface defined as $V(x,y) = 2sin(2\pi x)cos(2\pi y) + 2x^2 + 4y^2$. …….28
5. Table 1: Comparison of the eigenvalues for the synchronization-generated six-state substochastic matrix against the eigenvalues found by a two-dimensional Smoluchowski operator discretization method……...29
6. Figure 5: φ/ψ central dihedral angle Ramachandran plot of Ala_3……..31
7. Figure 6: Nodal domains of ensembles A and B of the preliminary training……....32
8. Figure 7: Nodal domains for ensembles A and B for step 3 of the SVM protocol……....33
9. Figure 8: Escape rate plot as a function of iteration step..……...……. 34

Table of Contents

Abstract **2**

Acknowledgements **3**

List of Figures and Tables **4**

Table of Contents **5**

Introduction **6**

Theory **9**

Fokker-Planck Formalism 9

Vicious Walkers 11

Discretization with Langevin Synchronization 13

The Discrete Substochastic Process 13

Langevin Synchronization 13

Markovianity within Substates 15

Particle Currents of Excited States 16

Algorithms **19**

Vicious Walkers 19

Langevin Synchronization and Substate Discretization 20

The Excited State Model 21

Implementation, Results, and Discussions **23**

Vicious Walkers 23

Langevin Synchronization and Substate Discretization 28

The Excited State Model 31

Conclusion **36**

References **37**

Introduction

Accelerated Molecular Dynamics (aMD) methods are used to extend the accessible timescales of atomistic simulations to more effectively compute thermodynamic and kinetic properties of matter, as well as to observe relevant mechanistic functions of complex aggregates. The most successful aMD methods to-date include Temperature Accelerated Molecular Dynamics (TAD) [1], Hyperdynamics [2], and most recently, parallel trajectory methods, which are currently realized by Parallel Replica Dynamics (ParRep) [3] and Parallel Trajectory Splicing (ParSplice) [4]. Parallel trajectory methods have been of particular interest to the aMD community over the past several years for several reasons. First, ParRep and ParSplice have been shown to efficiently boost dynamics for diverse classes of atomistic systems, from the pyrolysis of organic molecules [5] to defect formation processes in metals [6]. Aside from wide applicability, a recent mathematical formalization of the ParRep method suggests that it is even more generalizable than initially recognized [7]. Furthermore, it is anticipated that these methods will become increasingly efficient as the computing power of ever-expanding parallel computing architectures grow.

In effect, parallel trajectory methods accelerate dynamics by surmounting an ubiquitous obstacle in the realm of MD: infrequent event transitions. Many atomistic systems can be described as high-dimensional energy surfaces divided into networks of metastable states within which trajectories execute thermal vibrations as they explore the local environment. Only after sufficient time do these trajectories experience escapes through unstable dividing surfaces into connected states. Essentially, the dynamics of such systems can be recast as Markov State Models (MSMs). From a computational perspective, conducting intrastate dynamics is an expensive task that separates timescales and can inhibit the exploration of essential dynamics. Moreover, conducting MD in these metastable states

yields little-to-no relevant information from a thermodynamic and kinetic standpoint. Parallel trajectory methods circumvent this timescale separation problem by utilizing an ensemble of replicas across connected computing resources to parallelize time, effectively filtering out irrelevant intra-state processes whilst maintaining true, unbiased state-to-state dynamics. This can be summarized into a general two-step process:

1. *Dephasing* stage. Some initial position vector of a trajectory in a pre-defined state is broadcast to each of the N_r computing resources to generate an ensemble of trajectories. Overdamped langevin dynamics is conducted on each replica with a different random number seed to ensure statistical independence. MD must be run until each trajectory has evolved within the state for a correlation time, τ_{corr}, an adjustable, state-dependent parameter. If, at any point in time, a trajectory leaves the state, the replica is terminated, and a Fleming-Viot algorithm is implemented to replicate one of the existing trajectories as a replacement replica, and dephasing for τ_{corr} is re-attempted.
2. *Parallel* stage. Once all trajectories have successfully dephased, the wall clock is advanced by τ_{corr}. Each trajectory initializes independent MD again, and upon the first escape from the state, the wall clock is pushed forward by $N_r t$, where t is the shortest escape time, measured from the end of the *dephasing* stage [3].

The process is repeated with the first escaped trajectory in the newly defined state.

Perhaps the most critical assumption of this method is that the system dynamics approach Markovianity at sufficiently long timescales; however, of key importance for the motivation of this report is that at *any finite time*, the state-to-state dynamics are, in fact, non-Markovian. The instant a trajectory experiences an interstate crossing event, it retains *memory* of its transition path through the

dividing surface. This is a consequence of the fact that, upon an escape from a state, the probability distribution of the trajectory's position is spatially biased toward the coordinate space near the interstate crossing event. As a result, any subsequent interstate transition within a sufficiently short period of time from the previous transition is, de facto, non-Markovian, rendering the state-to-state dynamics biased. Hence, it is necessary to define a kinetic quantity that describes the rate of convergence to Markovianity such that, after a sufficient time, namely τ_{corr}, an escape is nearly Markovian (I.e., an escape is decorrelated from the previous interstate transition path). Probabilistically, this is equivalent to claiming that, after a τ_{corr}, the distribution of trajectories is independent of the location of the most recent interstate transition event. This long-time distribution is analogous to the Boltzmann distribution (I.e., it represents a relaxed distribution that is dependent on the potential energy surface) with the caveat that the distribution at the state boundary is zero, since trajectories that cross the state boundary are no longer defined within the state.

The mathematical underpinnings of this concept have been explored, and τ_{corr} can be defined explicitly in relation to the spectral gap to the solution of the Fokker-Planck Equation for a given state [7]. However, calculating this quantity proves to be quite a complicated task, even for relatively simple, low-dimensional systems.

This report proceeds to explore three potential solutions to this problem with applications of Langevin synchronization, vicious walkers, and a critical analysis of particle flux characteristics of *excited s*tates. Langevin synchronization is a phenomenon by which certain replicas experiencing the same thermal noise sequences undergo exponential convergence to the same trajectory. Several key quantities of interest are calculated with this unique treatment, and it is shown to yield correlation times in the discrete state approximation that constitute faithful resemblances of continuous dynamics. The extensibility of this method is addressed with respect to more complicated systems, and several computational examples are included to illustrate the advantages and limitations. Vicious walkers, on

the other hand, is a thought experiment in classical statistical physics wherein trajectories annihilate upon contact [8]. With this method, all relevant dynamical quantities are generated, and it is shown to yield mathematically exact correlation times for systems characterized by decoupled degrees of freedom. Finally, the *excited* state method is presented mathematically and followed by a brief explanation on the use of support vector machines (SMVs) to converge to a characteristic *excited* QSD, thereby sampling the correlation time. This is complemented by an algorithmic interpretation, which is tested on a benchmark molecule, Ala_3. Together, all of these methods hold potential to solve the correlation time problem in the parallel replica dynamics.

Theory

In this section, parallel trajectory methods are contextualized with the backward Fokker-Planck diffusion equation in order to address the mathematical difficulty in calculating τ_{corr}. This is followed by an explanation of the three proposed methods and their respective applicability to the problem.

Fokker-Planck Formalism

Consider a system in a potential $V(X)$ such that X is a position vector in R^{3N} (I.e., N is the number of atoms) defined on the space Ω; the overdamped langevin equation of motion is given by:

$$dX_t = -\frac{1}{\gamma}\nabla V(X_t)dt + \sqrt{\frac{2}{\beta\gamma}}dW_t \tag{1}$$

where γ is the friction coupling constant, $\beta = 1/k_B T$, $E(dW_t) = 0$, and $E(dW_t \cdot dW_{t'}) = \delta(t - t')$. Define a state Ω_α as a partition of Ω that is bounded by a connected surface $\partial\Omega_\alpha$, wherein the probability density of X at zero time is given by $P(X, t \mid X_0, t_0)$. Since the only trajectories of interest

exist in Ω_α up to time t, Dirichlet boundary conditions are imposed such that $P(X,t) = 0$ at and beyond $\partial\Omega_\alpha$. The probability density evolves according to

$$\partial_t P = LP; \quad L = -\frac{1}{\gamma}\nabla \cdot (\nabla V) + D\nabla^2 \tag{2}$$

where $D = 1/(\beta\gamma)$. A spectral decomposition of L in terms of its eigenvalues and eigenfunctions yields

$$P(X,t) = \sum_{i=1}^{\infty} c_i^0 \rho_i(X) e^{-\lambda_i t}; \quad c_i^0 = \int_\alpha P(X,0)\rho_i(X) d\mu^{-1} \tag{3}$$

where μ is the invariant measure. The eigenvalues are positively-defined such that $0 < \lambda_1 < \lambda_2 < ...$ Hence, each consecutive term in the probability density series decays exponentially faster than the previous. After a sufficient time, namely $\tau_{corr} \gg 1/(\lambda_2 - \lambda_1)$ [7],

$$P(X,t) \simeq c_i^0 \rho_i(X) e^{-\lambda_i t} \tag{4}$$

with an error on the order of $O(e^{-(\lambda_2-\lambda_1)t})$. $\rho_1(X)$ is known as the quasi-stationary distribution, or QSD, since it defines the long-time shape of the probability density of X. In context, the QSD represents the trajectory distribution that must be achieved to optimize the convergence to Markovianity in the long-time limit. Furthermore, Eq. 4 suggests that λ_1 quantifies the equilibrium escape rate from the state; by this logic, literature argues that λ_2 is the slowest *intra-state* transition rate, though the discussion on particle flux characteristics conflicts with this assertion, as will be explained shortly.

In practice, the key quantity of interest is τ_{corr}, or the inverse spectral gap. λ_1 can be measured directly by initializing replicas in Ω_α and employing a rejection method by which trajectories are killed and replicated in the state when they cross $\partial\Omega_\alpha$ (I.e., Fleming-Viot Algorithm) until the distribution of escape times becomes approximately exponential. However, directly measuring λ_2 is a much more involved task.

Vicious Walkers

Consider N replicas occupying the same state in a common coordinate space wherein the trajectories evolve according to Eq. 1 with respect to $V(X) = \sum_{i=1}^{N} v(x_i)$. The probability that all trajectories have evolved within the state without an escape event by time t is governed by a reformulation of Eq. 2, namely

$$\partial_t P(X,t) = D\sum_{i=1}^{N}\frac{\partial^2 P(X,t)}{\partial x_i^2} - \frac{1}{\gamma}\sum_{i=1}^{N}\frac{\partial V(X)}{\partial x_i}\frac{\partial P(X,t)}{\partial x_i}. \tag{5}$$

For a single trajectory in this ensemble, the survival probability, define by $\varrho(x_i,t)$, satisfies

$$\partial_t \varrho(x_i,t) = D\frac{\partial^2 \varrho(x_i,t)}{\partial x_i^2} - \frac{1}{\gamma}\frac{dv(x_i)}{dt}\frac{\partial \varrho(x_i,t)}{\partial x_i} \tag{6}$$

,which is separable in time and space. Thus, assuming the aforementioned Dirichlet boundary conditions, the solution set is the one-dimensional analog to Eq. 3. For N non-interacting trajectories in the same potential, the total probability density, $P(X,t)$, takes the form of a product of N single-trajectory survival probabilities, $P(x_i,t)$. This is a classical analogy to a system of N identical bosons in the regime of the Schrodinger Equation. However, since the annihilation property must be obeyed upon trajectory contact, Dirichlet boundary conditions are imposed such that $P(X,t) = 0$ when $x_i = x_j$. This is satisfied by constructing $P(X,t)$ as an antisymmetric combination of products of $\varrho(x_i,t)$. In the spirit of quantum fermionic systems, a Slater-type determinant can be constructed such that

$$P(X,t) = \sum_{i=1}^{N}\det(Q^{i_1,i_2,\ldots i_N}) \tag{7}$$

where

$$Q_{a,b}^{i_1,i_2,\ldots,i_N} = \frac{c_{i_a}^{0_b}}{\sqrt{N}} \rho_{i_a}(x_b) e^{-\lambda_{i_a} t} \quad . \tag{8}$$

In the semi-infinite time limit, every term in the series of Eq. 7 decays to zero with the exception of the first, which is constructed with the N slowest-decaying single-trajectory eigenfunctions. It follows that $P(X,t) = det(Q^{1,2,\ldots,N})$ where the associated time-dependent term is $e^{-\theta_N t}$ such that [9]

$$\theta_N = \sum_{i=1}^{N} \lambda_i \quad . \tag{9}$$

Since N=2 yields λ_2, two vicious walkers are chosen. Thus, in the long-time limit, the distribution becomes

$$P(X,t) \propto \left[\rho_1(x_2)\rho_2(x_2) - \rho_1(x_2)\rho_2(x_1)\right] e^{-(\lambda_1+\lambda_2)t} \quad . \tag{10}$$

Also, since the paired walker system is fermionic, the spatial eigenfunctions must be antisymmetric about particle exchange. Consequently, there must exist a node that bisects the state wherein $P(X_{node}, t) = 0$ for all t. Equating Eq. 10 to zero gives the trivial solution of $x_1 = x_2$.

Aside from the fact that the solution must be separable to yield mathematically exact results, there is one more conundrum: the slowest-decaying pair of the term in the total survival probability has a rate of $\lambda_1 + \lambda_3$. Considering the spectral gap of a vicious walker pair, the correlation time can be defined as

$$\tau_{corr,2} \gg \frac{1}{\lambda_3 - \lambda_2} \quad . \tag{11}$$

This is problematic with regard to the efficiency of the method; if the difference between the second and third eigenvalues is small with respect to the spectral gap, then converging to the two-particle QSD could take exponentially longer than the actual relevant correlation time of interest. It is agreed that this is an impasse for any method that seeks to achieve sampling of λ_2 by locating the principal node of the solution to the Fokker-Planck Equation.

Discretization with Langevin Synchronization

This section discusses the ability to use Langevin synchronization to detect substates, discretize coordinate space, and in conjunction with the vicious walker formalism, define λ_2.

The Discrete Substochastic Process

Eq. 2 can be recast as a substochastic Markov matrix equation in the limit that the number of states that defines Ω_α, say q, approaches infinity in the continuous limit. However, using a finite discretization scheme that preserves the slowest intrastate transitions can be shown to maintain a faithful representation of its continuous-case counterpart. Ideally, this would entail locating and cataloging all of the potential minima in Ω_α, partitioning the state with transition boundaries between these minima, and finding the first-order transition rates k_i between these predefined substates. At this point, the row substochastic matrix Q can be computed such that the entry $Q_{i,j} = k_{i\to j}$ where $i \neq j$. $Q_{i,i} = -\sum_j k_{i\to j}$. Finally, the diagonalization of Q yields the QSD and the spectral gap.

However, using traditional minimization techniques to locate and catalog all potential minima in Ω_α is likely a prohibitively expensive process that carries with it the prospect of recovering unnecessarily detailed features on the coordinate space. An alternative method is to employ Langevin synchronization, which, in principle, recovers significant harmonic substates whilst overlooking the finer features of the energy surface.

Langevin Synchronization

Rewriting Eq. 1 in the more workable form of a single degree of freedom yields

$$\frac{dx_i}{dt} = -\frac{1}{\gamma}\frac{\partial V(X)}{\partial x_i} + F_i(t) \quad . \tag{12}$$

Consider the difference in position of the same degree of freedom for two particles (I.e., $\Delta x = x_i - x_j$). In the harmonic approximation that $V(x_{i,j}) = \frac{a}{2}(x_{i,j} - x_0)^2$ and using the same thermal noise sequence for both trajectories (I.e., $F_i(t) = F_j(t)$), one finds

$$\frac{dx_i}{dt} - \frac{dx_j}{dt} = -\frac{a}{\gamma}(x_i - x_j) \Rightarrow \frac{d\Delta x}{dt} = -\frac{a}{\gamma}\Delta x \Rightarrow \Delta x = c_0 e^{-\frac{at}{\gamma}} \quad . \tag{13}$$

In the frequency regime, $a \equiv m\omega^2$; hence, the synchronization rate is proportional to the quare of the angular frequency of the harmonic well.

Of course, the harmonic well is a highly-idealized approximation to the actual potential energy surfaces of complex systems. In reality, potential wells do not have constant curvature, which results in a position-dependent frequency, namely

$$\omega(x) = \sqrt{\frac{1}{2m}\frac{d^2V}{dx^2}} \tag{14}$$

Nevertheless, this detail allows synchronization, albeit not in the first order.

The true hindrance to synchronization of trajectories is two-fold. First, if a trajectory experiences thermal excitement into a region of convex character, then its local frequency is imaginary, resulting in local divergence at short timescales. This divergence, however, becomes more long-term when a trajectory moves through a dividing surface into a different basin entirely, though this can be avoided by running dynamics at low temperatures.

In the regime of uncoupled harmonic wells in higher dimensions, the synchronization is clearly independent among each degree of freedom for all time. However, when coupling terms among harmonic (or approximately harmonic) degrees of freedom are present, an interesting phenomenon emerges. Initially, the faster synchronizing mode dominates, but the coupling term will eventually disrupt convergence in favor of the slower mode. Thus, there must exist a crossover at

some time between the faster-to-slower mode that is dependent upon the relative strengths of the coupling term, which is, in turn, temperature-dependent. This phenomenon was studied extensively on the Voter potential in [9].

Markovianity within Substates

In order to maintain true dynamics in the Markov limit, one need not only find τ_{corr} for the entire metastable state, but also τ_{corr} for each substate. Granted, correlations within sub-states exist on a much shorter time scale relative to those of the entire state, and thus, can often be pre-set as a conservative parameter that doesn't significantly impact the efficiency of the ParRep method. However, it is advantageous to know the conditions of and limits to the approximation.

τ_{corr} is computed in the decoupled infinite harmonic well approximation. Given the harmonic potential $V(X) = \sum_{i=1}^{N} a_i x_i^2/2,$ the backward Fokker-Planck Equation reads

$$\partial_t p(x_i,t) = D\frac{\partial^2 p(x_i,t)}{\partial x_i^2} - \frac{a_i x_i}{\gamma}\frac{\partial p(x_i,t)}{\partial x_i} \quad . \tag{15}$$

This equation takes nearly the same form as the harmonic potential Schrodinger Equation. In this spirit, Eq. 15 transforms with $p(x_i,t) = e^{a_i x_i^2/4D}\,\psi(x_i,t)$ to give

$$\partial_t \psi(x_i,t) = D\frac{\partial^2 \psi(x_i,t)}{\partial x_i^2} + \left[\frac{a_i}{2\gamma} - \frac{(a_i x_i)^2}{4D}\right] p(x_i,t) \tag{16}$$

The solutions have the form $\psi(x_i,t) = e^{-\lambda_i t}\,u(x_i),$ where $u(x_i)$ satisfies the eigenequation

$$\left[D\frac{\partial^2}{\partial x_i^2} + \frac{a_i}{2\gamma} - \frac{(a_i x_i)^2}{4D}\right] u(x_i) = -\lambda_i u(x_i) \tag{17}$$

The eigenfunctions of the equation are known:

$$u_n(x_i) = H_n\left(x_i\sqrt{\frac{a_i}{2D}}\right) e^{-a_i x_i^2/4D} \tag{18}$$

$H_n(\cdot)$ are Hermite polynomials and the corresponding eigenvalues are $\lambda_{n,i} = na_i/\gamma$.

The one-dimensional infinite-well case gives spectral gaps of a_i/γ. The result can be easily generalized to N harmonic wells with different curvatures using separability. This is equivalent to

$$P(X,t) = P(x_1,t) \oplus P(x_2,t) \oplus \ldots \oplus P(x_N,t) \tag{19}$$

This requires the first eigenvalue of the N-dimensional eigenequation to be zero since all $\lambda_{i,0} = 0$ for all i degrees of freedom. The first *excited* state in the interchanges $\lambda_{i,0}$ with $\lambda_{i,1}$ for the term corresponding to the smallest spectral gap. By the fact that, for a given oscillator, each eigenvalue is equidistant from an adjacent eigenvalue, the spectral gap is proportional to a_i for a constant γ. The smallest a_i corresponds to the slowest mode of oscillation, which defines the minimum spectral gap. Hence, in order to estimate τ_{corr} for a given sub-state, one need only measure the slowest mode of synchronization, which is equal to the reciprocal of the long-time trajectory convergence rate. [10] shows that this estimation is, in fact, a conservative approximation of the true substate τ_{corr}, and that in the event the transition state is symmetric about some oscillator, the spectral gap is unchanged for the aforementioned calculation.

Particle Currents of *Excited States*

Often, discretizing a state into approximately-harmonic substates and yielding results that are in agreement with the true, continuous-case diffusion equation is impossible. This arises when there are large convex portions of potential, or when there exist coupled oscillators with curvatures that are orders of magnitude apart. However, by far the most common reason for the failure of substate discretization is much more intrinsic to the assumptions used in the scheme. The parallel replica dynamics are most effective when there are hierarchies of barriers such that one barrier dominated over the rest by at least an order of magnitude. Defining such a barrier is not necessarily a difficult task, but

the existence of substantial heterogeneous distributions of barrier heights within substates is rather problematic, especially since it is difficult to choose γ parameters that will capture all relevant substates through synchronization.

Consequently, it may be necessary to abandon the discretization method in its entirety and rely upon diffusion in continuous space. The first step herein is to reformulate Eq. 2 as a particle current, where

$$\partial P = -\nabla \cdot J; \quad J \equiv -\gamma^{-1}(\nabla V)P - D\nabla V \tag{20}$$

The objective of this novel method is straightforward, though the mathematical formalization is arduous. The QSD of the state has no nodes since it corresponds to the smallest possible eigenvalue; however, since the Fokker-Planck Equation is a particular case of the Sturm-Liouville eigenproblem, solutions must not only be mutually orthogonal (under the invariant weight), but the number of nodes increases such that the i^{th} eigenfunction has $i-1$ nodes. Hence, the eigenfunction corresponding to λ_2 must have a single node that splits the state into two nodal domains (one positively- and one negatively-defined in probability density). Immediately, it should be obvious that the *excited* states are nonphysical, since probability cannot be negative. However, it is perhaps advantageous to think of the positively- and negatively- defined domains as probability densities of two flavors of mutually annihilating particles (i.e. vicious walkers).

One particularly interesting result of this approach is that it would explain the location of the node. Clearly, the location of the node of ρ_2 is time independent, and this can be explained rather plainly with Eq. 20. The particle current, J, is continuous across the node since P is defined as zero on the node and the gradient of V is continuous. This means that the local particle flux through every infinitesimal surface, dS, on the nodal hyperplane is equal and opposite. This unique (assuming no degeneracy) steady-state solution would also be realized in the long time limit of two contiguous populations of mutually annihilating particles A and B in the reaction scheme of $A + B \rightarrow \varnothing$.

Aside from the fact that both nodal domains of the second eigenfunction have the same local flux along the nodal domain, the populations must also have the same nodal escape rate λ_2. This stipulation also requires that each nodal domain experiences the same total flux out of the state, which is intuitive, though the mathematical formalization is rather dense. It is explained below in 'full' detail.

Given the distribution

$$P(X) = c_2^0 \rho_2(X) e^{-\lambda_2 t} \tag{21}$$

such that c_2^0 normalizes $|\rho_2(X)|$, Eq. 20 can be integrated over the coordinate space of the state such that

$$\int_{\Omega_\alpha} \frac{d}{dt} P dV = \frac{d}{dt} \int_{\Omega_\alpha} P dV = -\int_{\partial\Omega_\alpha} J \cdot dS \tag{22}$$

The Divergence Theorem requires that one computes the total flux through the state boundary, not through the node, itself. Considering the last term in Eq. 22,

$$\int_{\partial\Omega_\alpha} J \cdot dS = \int_{\partial\Omega_{\alpha+,state}} J \cdot dS + \int_{\partial\Omega_{\alpha-,state}} J \cdot dS \tag{23}$$

where $\partial\Omega_{\alpha+/-,state}$ is the surface that defines the state in the nodal domain + (likewise, for -).

On the other hand, if attention is restricted to domain + (-), then

$$\int_{\Omega_{\alpha\pm}} \frac{d}{dt} P dV = \frac{d}{dt} \int_{\Omega_{\alpha\pm}} P dV = -\int_{\partial_{\alpha\pm}} J \cdot dS \tag{24}$$

In a single nodal domain ($\alpha + / -$), the rightmost side of Eq. 24 is rewritten as

$$-\int_{\partial\Omega_{\alpha\pm}} J \cdot dS = -\int_{\partial\Omega_{\alpha\pm,node}} J \cdot dS = -\int_{\partial\Omega_{\alpha\pm,state}} J \cdot dS \tag{25}$$

Since the escape rate λ_2 characterizes the entire state and each nodal domain, then the leftmost side of Eq. 22 is equal to the leftmost side of Eq. 24. Therefore,

$$\int_{\partial\Omega_{\alpha\pm,node}} J \cdot dS + \int_{\partial\Omega_{\alpha\pm,state}} J \cdot dS = \int_{\partial\Omega_{\alpha+,state}} J \cdot dS + \int_{\partial\Omega_{\alpha-,state}} J \cdot dS \tag{26}$$

The two terms closest to the '=' are identically the same if $\alpha+$ is chosen, so

$$\int_{\partial\Omega_{\alpha\pm,node}} J \cdot dS = \int_{\partial\Omega_{\alpha-,state}} J \cdot dS \tag{27}$$

Hence, the total flux from the nodal boundary of domain + is equal to the total flux leaving domain - from the state boundary. Also, since the flux is locally equal and opposite on the node, the flux from the node boundary of domain - is equal to the total flux leaving domain +. Therefore, the total fluxes out of each nodal domain are equal, so in order to preserve Eq. 24, each domain must contain the same magnitude of total probability. These features will be used in the next section to rationalize algorithms for sampling λ_2 directly.

Algorithms

This section realizes the physical underpinnings of the vicious walkers model, the Langevin synchronization discretization scheme, and the *excited states* treatment with algorithms to sample correlation times of states.

Vicious Walkers

The vicious walkers method is straightforward. It proceeds as such:

1. Initialize two walkers in a state with random phase space coordinates.
2. Execute independent overdamped Langevin dynamics on each replica.
3. If Δx of the two trajectories ever changes sign, record the wall clock time from initialization and kill the replicas; zero the wall clock

4. Likewise, if either of the replicas escapes the predefined state boundaries, record the wall clock time, kill replicas, and zero the wall clock

After a 'sufficient' number of trials, the escape time distribution can be plotted, and the tail of the distribution is exponential in $\lambda_1 + \lambda_2$. A maximum likelihood method can be implemented on the data points collected to yield the rate.

Langevin Synchronization and Substate Discretization

The following procedure is proposed to calculate correlation times for metastable states for discretization methods coupled with Langevin synchronization signaling:

1. Distribute and propagate N independent master trajectories across a defined state using a Fleming-Viot rejection method
2. After a time τ_m, halt dynamics, initialize probe trajectories (one for each master) with the same initial coordinates and nose sequence as the corresponding master trajectory for a time τ_s, which is on the order of a synchronization time. It should be noted that the noise sequence should be initialized from the time $\tau_m - \tau_s$.
3. Compare the tail of the master and the probe trajectories. If they overlap within some predetermined tolerance, then run synchronization on the pair until the slowest convergence mode can be quantified. Record the convergence rates and catalog the sub-states.
4. Repeat the previous steps until all of the substates have been cataloged (I.e., no new substates are discovered within a certain number of loops). If ever a master trajectory experiences a synchronization within a previous cataloged state, skip step 2 and proceed to another repetition of step 1.
5. Distribute replicas among the substates and run independent dynamics for τ_{corr} corresponding to each substate. Use the probe trajectory method outlined in step 3 to flag for escapes before

decorrelation is completed. Each escaped trajectory re-attempts decorrelation with the aforementioned Fleming-Viot procedure.

6. After decorrelation, run independent dynamics to sample the intra-state transition rates.
7. Construct the sub-stochastic transition rate matrix, diagonalize, and compute the spectral gap for the entire state.

The primary consideration to be made before synchronizing trajectories is that the replicas are sufficiently distributed across the state such that all synchronizing substates are accessible in a sufficiently short time. This is ensured as long as the state is partitioned with the highest barrier, and assumption that is made a priori in ParRep.

The signaling method is also capable of returning false positive and negative synchronizations. A false positive, which is defined as signaling a transition that did not occur, could arise when the probe trajectory escapes a substate before synchronizing to the master, so the tails of the trajectories differ. A false negative, which is simply a failure to detect a transition, will occur if the probe synchronizes quickly to the master and follows its path out of the substate [9].

These issues can be resolved in some regard by employing a sufficiently large ensemble of replicas and observing the average characteristics of master and probe trajectories. As a matter of fact, if all the master trajectories are also given the same noise sequence, then true synchronizations in the state and escapes will become obvious without even using probe trajectories, through the dimensionality of the state may prohibit the use of enough trajectories to observe this.

The *Excited State* Model

The following algorithm describes the method for using an SVM with an initial node ansatz to converge to ρ_2, and thus, sample λ_2.

1. Define some preliminary node with an ansatz. This is often achieved by considering $V(X)$, collecting the largest subset of separable (or approximately separable, in some local coordinate space) terms, propagating the Fokker-Planck operator on each of these numerically, and placing the node corresponding to the separable term with the smallest spectral gap.
2. Initialize N_r independent replicas in each domain and propagate with independent dynamics until local QSDs are achieved (i.e. the tail of the distributions of each domain is exponential).
3. Perform the iterative process:
    a. Run dynamics for t_{dyn}.
    b. Train both ensembles with an SVM. It is advised to use a Gaussian activation function, since this allows for the largest basis set projection on to coordinate space. The training features are dependent on the system of interest, though the most conservative route is to use all atomic internal coordinates.
    c. Conduct the Fleming-Viot subroutine to re-initialize all replicas that have escaped according to the new state definitions.
    d. Run dynamics on each ensemble for time t_{rlx} to relax replicas in newly-defined domains whilst simultaneously conducting the Fleming-Viot subroutine on escaped replicas.
    e. Repeat

It should be noted that the γ and C parameters in the Gaussian activation function should be optimized to consistently return contiguous domains. Any other result is non-physical, unless some dimensions are periodic. In the interest of brevity, a discussion of SVMs and the double minimization equation has been omitted. These details are secondary to the objective of this report.

Implementation, Results, and Discussions

This section reports on the implementation of the three proposed methods. The vicious walkers model was applied to a one-dimensional potential, whilst the Langevin discretization and *excited* state methods were applied to a coupled two-dimensional problem. The success of the *excited* state model warranted an application to Ala_3 with promising results for future applications. These results are followed by a critical assessment of the advantages and limitations of the methods.

Vicious Walkers

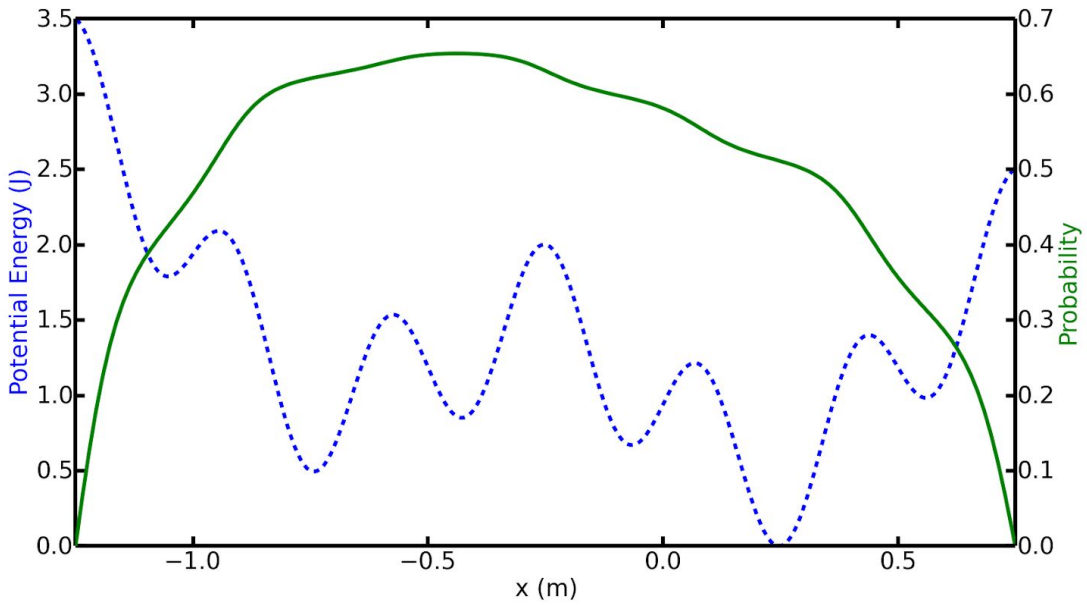

Fig. 1: Plot of potential energy (dotted blue) and its associated normalized QSD (solid green). The one-dimensional state is defined on the range (-1.25 m, 0.75 m) and the potential therein has the form $V(x) = sin(2\pi x)cos(4\pi x) + x^2$.

Fig. 1 illustrates the potential energy function and the corresponding QSD. It should be noted that two local maxima in the region around zero were chosen as the state boundaries in order to stay true to the ParRep method of choosing a state boundary such that the spectral gap is maximized. The calculation of the QSD was achieved by diagonalizing the matrix-equivalent of a discretized Smoluchowski operator; generating a stochastic transition rate matrix with a rate from state n to state n+1 is defined as [10]

$$k_{n,n+1} = \frac{D}{d^2} e^{-\beta(V(x_{n+1})-V(x_n))/2} \tag{28}$$

where d is the distance between x_n and x_{n+1}. The diagonal entries are defined as the negative sums of the rates in their respective rows. D and β were taken as unity. 1000 equidistant points were used to generate the matrix. The substochastic matrix was extracted as a sub-block of the stochastic matrix with the endpoints defined as the absorbing states. Diagonalization yielded λ_1 and λ_2 values of $1.165\ s^{-1}$ and $4.687\ s^{-1}$, respectively.

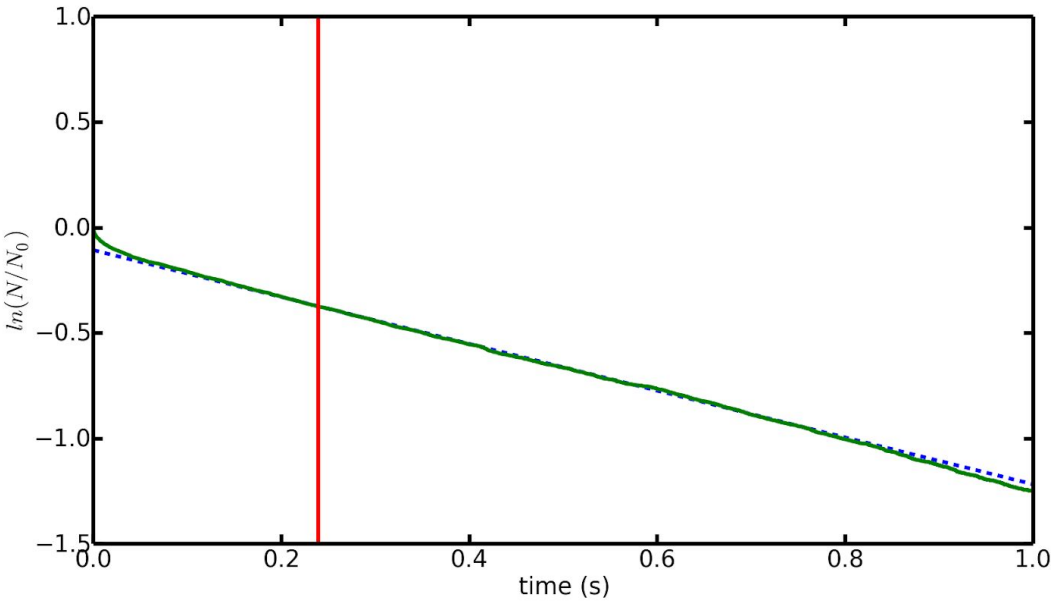

Fig. 2: Log plot of survival probability of a 10,000 replica, single-trajectory ensemble (solid green) with an associated linear regression plot (dotted blue). The linear regression plot was fit to data between t=0.4 s and t=0.9 s, as this interval maximized the R^2 value of the regression fit. The fit has a slope of -1.164 ± 1.285 x 10^{-3} s^{-1} and a y-intercept of -0.07630 ± 9.204 x 10^{-4} s^{-1}. The error of λ_1 compared to the theoretical value of 1.164 s^{-1} is -0.08584 percent. The red line denotes the correlation time of 0.2389 s.

Fig. 2 illustrates the escape-time distribution of 10,000 replicas initialized randomly within the state, and whose trajectories were integrated according to Eq. 1 (D and β were taken as unity). The potential gradient scheme used a dx value of 1 x 10^{-6} m to compute forces.

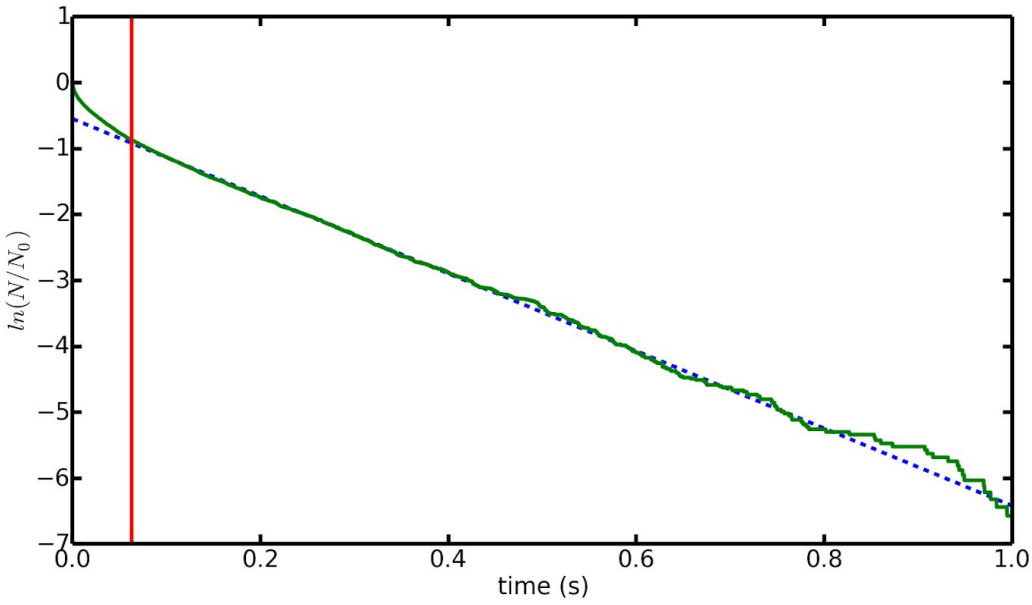

Fig. 3: Log plot of survival times of a 5,000 vicious walker pair ensemble (solid green) with an associated linear regression plot (dotted blue). The linear regression plot was fit to data between t=0.2 s and t=0.7 s, as this interval maximized the R^2 value of the regression fit. The fit has a slope of -5.871 ± 5.448 x 10^{-3} s^{-1} and a y-intercept of -0.5483 ± 2.374 x 10^{-3}. The error of $\lambda_1 + \lambda_2$ as compared to the theoretical value is 0.3247 percent. The red line annotates the second correlation time at 0.06272 s.

Fig. 3 reports on the numerical experiment. It is fortunate that the second correlation time is shorter than the first, as this allowed for a faster convergence to the two-particle QSD.

The measured correlation time of 0.2822 s was found to agree with the accepted value of 0.2839 s with a percent error of -0.5824. A more critical examination of the method suggests that the source of error in the measured correlation time is twofold. First, the error is bounded by the timestep size, which was set as 1 x 10^{-4} s. Specifically, the rejection criterial are checked between every

timestep to determine escape events for each trajectory in the ensemble; if at least one rejection criterion is met at step i, then the escape event presumably occurred somewhere between δt_i and δt_{i-1}, the error of which, is the size of the associated timestep. Often, however, this error is many orders of magnitude smaller than the correlation time of the state. Secondly, the experimental procedure requires that a sufficient number of replicas are present at any given time step in order to minimize noise in the survival probability. If this is not satisfied, then the linear regression will be prone to statistically significant errors in the least squares fitting procedure. This is especially common if the long-time escape rate is high, as the population of the ensemble will approach zero rapidly; the effect of this can be seen in the tails of the distributions of Figs. 2 and 3. Lastly, and perhaps the most critical limitation, is the fact that the Hamiltonian must be separable. This prohibits applications to atomistic problems, which are characterized by all-atom radial interactions; however, it is this separability restriction that invites a discussion of discretization in the following section.

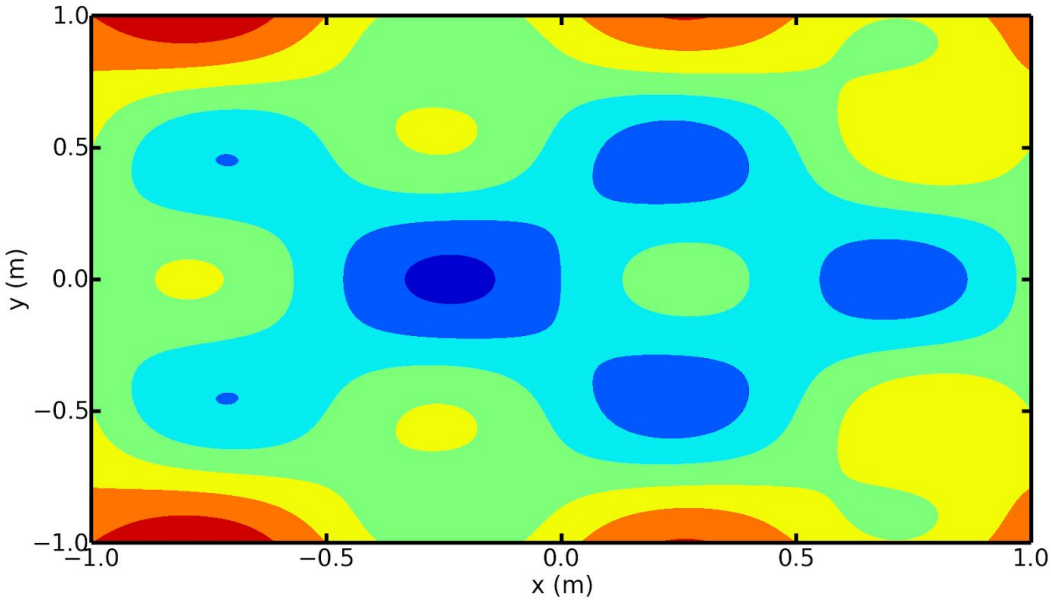

Fig. 4: Contour plot of the potential surface defined as $V(x,y) = 2sin(2\pi x)cos(2\pi y) + 2x^2 + 4y^2$. The blue and red indicate low- and high-potential regions, respectively.

Fig. 4 illustrates the two-dimensional coupled potential that was used to test the Langevin Synchronization substate discretization method. An ensemble of 100 replicas was implemented to probe the state with the langevin integrator defined in the previous section. Initial coordinates were chosen as uniformly distributed random variables in x and y within the confines of the state definition of x: (-1,1), y: (-1,1). τ_m was set to 100 ms for master trajectories. Subsequently, probe trajectories were run from the same initial coordinates as their respective masters for $\tau_s = 25$ ms with the same noise sequence as the last 25 ms of the master trajectories. To register synchronizations, and hence, catalog substates with their respective convergence rates, Δx and Δy of the probe and master

trajectories were computed, linearized, and fit with regression. If the fits both yielded $R^2 > 0.98$, then a synchronization was registered, the convergence rate was computed from the slope of the regression, and the state was cataloged.

Likely due to the rather fast intrastate convergence rates and the locally harmonic regions, all 6 substates were detected within the first iteration of the Langevin synchronization subroutine. 62 replicas registered synchronizations, 21 escaped the state entirely, and the remaining 17 replicas registered transitions to different states within τ_s. Upon further examination of each master and probe trajectory pair, no false positives or negatives were registered.

Table 1. Comparison of the eigenvalues for the synchronization-generated six-state substochastic matrix against the eigenvalues found by a two-dimensional Smoluchowski operator discretization method.

λ_i	**Synchronization** [s^{-1}]	**Smoluchowski Op.** [s^{-1}]
1	3.216	3.018
2	8.307	7.660
3	10.741	8.102
4	17.063	12.344
5	44.851	31.262
6	47.285	33.104

Afterward, intrastate correlation times were computed from the synchronization rates in the harmonic approximation. The states with minima at (-0.71 m, 0.45 m), (0.23 m, 0.43 m), (-0.24 m, 0 m), (0.7 m, 0 m), (-0.71 m, -0.45 m), and (0.23 m, -0.43 m)were found to have substate correlation times of 24.2, 23.1, 25.4, 23.2, 24.3, and 23.6 ms, respectively. The 100 replicas were divided among

these states, initialized at the final master trajectory coordinates, and carried out independent dynamics. Whenever a trajectory escaped, the time was recorded only if it had evolved for its respective substate correlation time. The Fleming-Viot subroutine was implemented to maintain efficient sampling. This process was run for 200 ms, at which point, the time distribution was fit to compute intrastate transition rates, the substochastic matrix was constructed, and then diagonalized to yield the left column in Table 1.

The right column of Table 1 was generated by constructing a 100 by 100 point grid and computing the transition rates between adjacent points in accordance with Eq. 28. The first eigenvalue gave an error of 6.6 percent and the second gave an error of 8.4 percent relative to the Smoluchowski discretization method. Clearly, this error increases with increasing λ_i as expected, since the method cannot sufficiently sample the finer features of the potential energy surface in the discrete regime. In total, the synchronization and Smoluchowski discretization methods gave τ_{corr} values of 215 and 196 ms, respectively, for a percent error of 8.8.

It is manifest that the synchronization method yields fairly accurate results, at least in the regime of coupled, nearly harmonic substates of low-dimensional space. It is also worthy of noting that the total time for the ensemble to sufficiently approximate substate correlation times, considering τ_s, τ_m , the rate sampling time, and the substate correlation times, amounted to nearly twice the calculated total correlation time. This time is likely an overestimate since the substate correlation times are extremely conservative, and that only 100 replicas were employed to sample transition rates. This shows promise for the extensibility of the method if one is able to bound the substate correlations more tightly, or if one employs more replicas in the sampling stage to generate a sufficiently-sized ensemble of transitions.

Consistent with the aforementioned argument, the error in the spectral gap would have likely been reduced had the definition of the state been expanded to include more substates. This is another promising prospect for applications to real systems with large number of synchronizable substates.

The *Excited* State Model

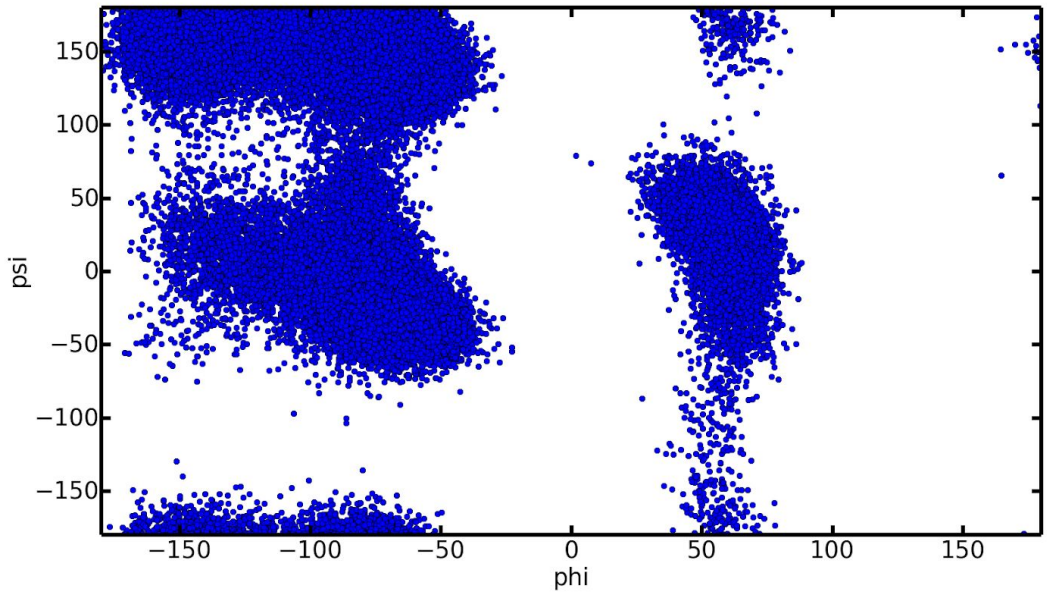

Fig. 5: φ/ψ central dihedral angle Ramachandran plot of Ala_3.

Fig. 5 illustrates the Phi/Psi dihedral angle Ramachandran plot generated from a 500 nm simulation of Ala_3 using the AMBERff14-SB protein force field [11] with a Langevin thermostat, a temperature of 300K, a collision frequency of 50 ps^{-1} , a timestep of 4 fs (with Hydrogen Mass Repartitioning) and an implicit Born water solvent. There appeared to be a rather high barrier at the φ $=0^o$ vertical line, so a state of interest was defined at $-180^0 < \varphi < 0^o$, $-180 < \psi < 180$ for the central dihedral (residue 2). For the purposes of simplicity, the initial node was chosen at $\psi = 0^0$.

In the spirit of the algorithm, 500 independent replicas were initialized randomly in each nodal domain, and each trajectory evolved for a time of 50 ps to converge to local QSDs. Afterward, t_{dyn} was set to 1ps, t_{rlx} was set to 4ps (a conservatively long relaxation time), and escapes were checked every 1 ps. The total iteration time was set to 50 ps (I.e., 10 steps). After each step, a 200 ps escape rate-sampling time was conducted to monitor convergence. The SVM parameters were set with a radial basis function, a γ of 1 x 10^{-3}, and a C correction factor of 1,000. The features trained were exclusively the central dihedral angles of residue 2.

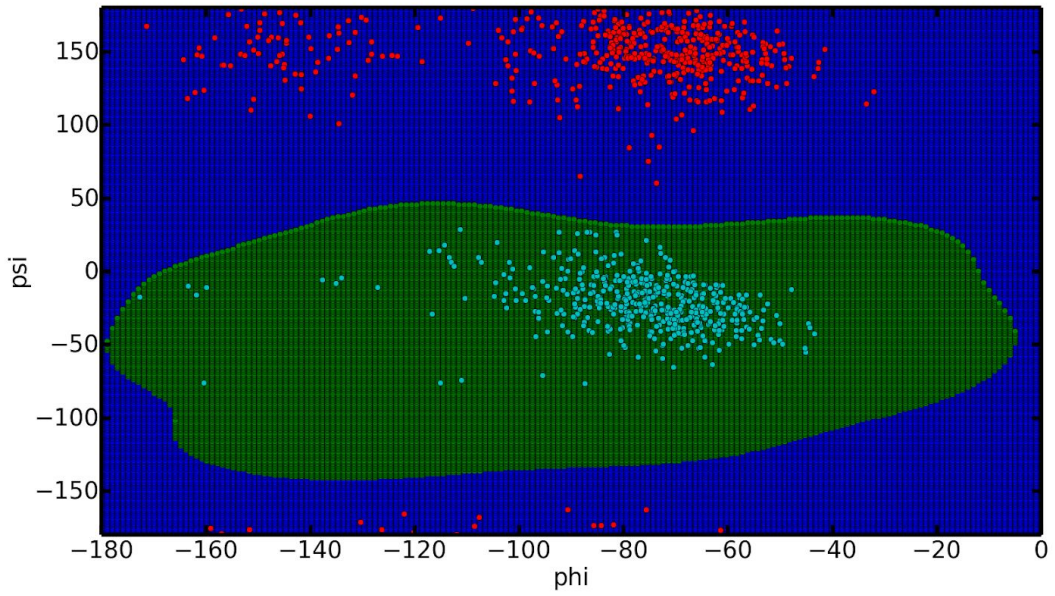

Fig. 6: Nodal domains of ensembles A and B of the preliminary training. Blue represents the domain of the red ensemble A, whereas green represents the domain of the teal ensemble B.

Fig. 6 shows the shape of the initial node based on the ansatz. It is clear that the result is unphysical since the green domain does not experience escapes from the state. As the number of iterations of the protocol increased, it became apparent that the nodal domain was quickly achieving

some equilibrium position in the state space. The similarities of the domains can be seen when comparing the nodes between Figs. 6 and 7, shown below.

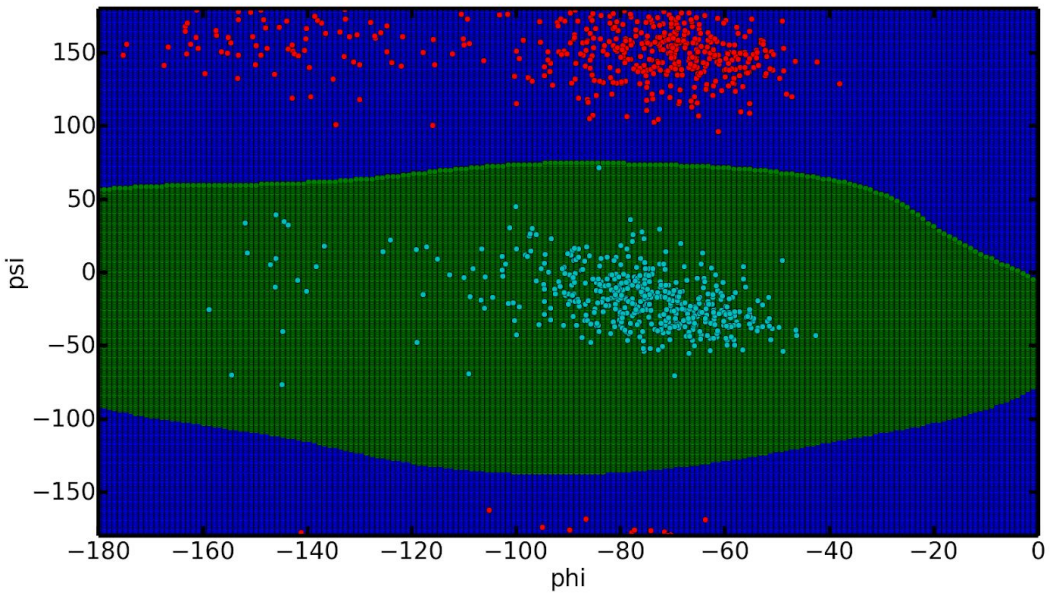

Fig. 7: Nodal domains for ensembles A and B for step 3 of the SVM protocol. The labels are consistent with those in Fig. 6

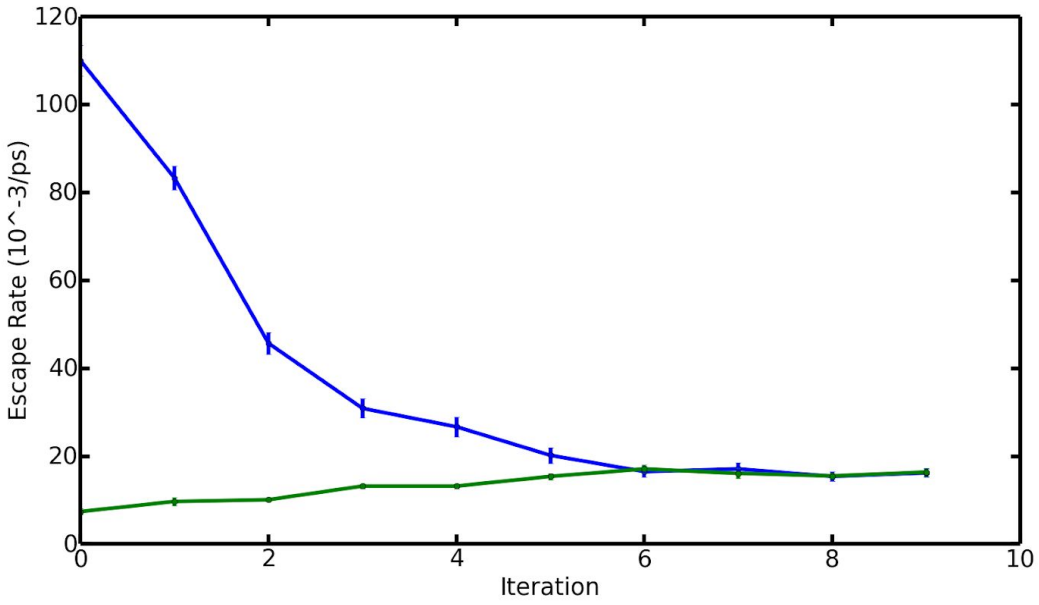

Fig. 8: Escape rate plot as a function of iteration step. Blue and green represent the long-time escape rates of ensembles A and B, respectively. The last iteration yielded an average escape rate of 19.27 x 10^{-3} ps^{-1}.

Fig. 8 clearly shows the converging trend of the escape rates. Whilst the initial node ansatz yielded a large gap in the escape rate (nearly 100 x 10^{-3} ps^{-1}), the final iteration yielded a gap of only 0.21 x 10^{-3} ps^{-1}. The escape rates were calculated from a maximum likelihood calculation in the exponentially distributed escape times for each post-training 200 ps sampling run.

In total, it is clear that the method is effective in its ability to converge to a singular escape rate, and considering the near impossibility of degeneracy in this system, it is likely that the final rate is nearly λ_2, though verifying this with a discretization operator is impossible (there are 42 atoms in the system).

This model clearly has advantages over the previous two. Aside from the fact that the *excited* state procedure is applicable to systems with non-separable Hamiltonians, it also has a distinct

advantage over the discretization method in that it does not require a catalog of all possible transition rates, which could become unwieldy if the friction coefficient is chosen too large or the state space has too many substates. Furthermore, the *excited* state method captures any and all relevant features on the potential energy surface (assuming there are a sufficient number of replicas) since the state is continuously defined.

As far as disadvantages are concerned, the current method is prone to nodal stagnation if there is a sufficiently high barrier crossing that impedes flux. Similarly, if the number of replicas is not large enough, the node will likely have a low signal-to-noise ratio. This becomes particularly problematic when the coordinate space is extremely high-dimensional.

Conclusion

This paper provides a brief description of three methods that are potentially capable of solving the correlation time problem in the parallel replica dynamics regime, which essentially revolves around the dilemma of sampling the second eigenvalue of the solution to the diffusion equation with zero boundary conditions. The vicious walkers method is mathematically exact and relies upon a fermionic treatment of trajectories that yields the sum of the first two eigenvalues in the long-time limit, though it requires a separable solution. On the other hand, the Langevin synchronization/discretization method is a rather accurate numerical approximation to the solution since it involves the reformulation of the diffusion equation with a more coarse-grained approach, though it potentially falls short in retrieving relevant detailed quantities. Lastly, the *excited* state method uses a rigorous mathematical formalism that succeeds in that it maintains the continuous state-space nature of the diffusion equation whilst converging to perhaps the exact second eigenvalue; however, it likely falls short in efficiency and fails with degenerate systems. Overall, the mathematical formalisms and numerical demonstrations of each method effectively elucidate the advantages and limitations.

References

1. M. Sorensen and A. Voter, J. Chem. Phys. **112** , 9599 (2000).
2. A. Voter, Phys. Rev. Lett. **78**, 3908 (1997).
3. A. Voter, Phys Rev. B **57,** 985 (1998).
4. D. Perez, E. Cubuk, A. Waterland, E. Kaxiras, and A. Voter, J. Chem. Theory Comput. **12**, 18 (2016)
5. O. Kum, B. Dickinson, S. Stuart, B. Uberuaga, and A. Voter, J. Chem. Phys. **121**, 9808 (2004)
6. D. Warner and W. Curtin, S. Qn. Nat. Mater. **6**, 177
7. C. L. Bris, T. Lelievre, M. Luskin, and D. Perez, Monte Carlo Method Appl. **18**, 119 (2012)
8. A. Bray and K. Winkler, J. Phys. A **37,** 5493 (2004).
9. B. Uberuaga, M. Anghel, and A. Voter, J. Chem. Phys. **120,** 6363 (2004)
10. C. Lu, A. Voter, and D. Perez, J. Chem. Phys. **140**, 4052 (2014).
11. J.A. Maier, C. Martinez, K. Kasvajhala, L. Wickstrom, K.E. Hauser, and C. Simmerling, J. Chem. Theory Comput. **11,** 3696 (2015).