JCTC Journal of Chemical Theory and Computation

Calculation of Point-to-Point Short-Time and Rare Trajectories with Boundary Value Formulation

Dov Bai and Ron Elber*

Department of Computer Science, Upson Hall 4130, Cornell University, Ithaca, New York 14853

Received January 20, 2006

Abstract: Sampling rare, short-time, and reactive trajectories is of considerable interest in molecular simulations. These trajectories, which are also called "activated", hop between stable states separated by energy or entropy barriers. Simulations of activated trajectories with random sampling of initial conditions are inefficient, since most initial conditions lead to trajectories that do not pass the barrier in short times. A boundary value formulation is proposed that selects these rare trajectories, making the sampling of point-to-point reactive trajectories more effective. Earlier boundary value formulations by one of us focused on computations of approximate trajectories. In the proposed method, trajectories are accurate even when we employ a relatively large integration step (by a factor of about 100 compared to initial value methods). The boundary value solutions to short-time reactive trajectories tend to be unique and have significant statistical weights compared to other reactive trajectories of the microcanonical ensemble. Three numerical examples are considered: a transition in the Mueller potential, a conformational change in alanine dipeptide, and an isomerization in a Lennard-Jones cluster.

I. Introduction

Molecular dynamics simulations are powerful tools in the analysis of microscopic phenomena. Kinetic and thermodynamic properties of matter are studied on the computer by solving microscopic equations of motions. While highly successful in many cases, significant limitations remain. One of these limits is of sampling rare events. In numerous cases only a tiny fraction of the trajectories are "productive" after time t (reaching a desired final state). Because of the rarity of these reactive processes the overall rate of an ensemble of these trajectories (or the average time of reaction) can be very long even if the actual transition is rapid. Therefore, sampling of rare transitional events is computationally demanding. This paper focuses on this challenge and suggests a method for efficient sampling of rare, short-time, and reactive paths. We restrict the discussion to Newtonian's trajectories.

We describe the dynamics of the system with the coordinate vector $X(t)$—the (Cartesian) position of the system at a time instant t. We divide the coordinate (conformational) space into three states. The system can be in the reactant state (R), the product state (P), or in a transitional state (Tr) (Figure 1).

We determine the characteristics of trajectories that started in R at time zero and found in P after total time T.

The most straightforward way to compute reactive trajectories is to sample initial conditions $X(0),Q(0)$ from a starting probability density $\rho_0(X,Q)$ (Q is the momentum). The probability density we have in mind is either of the microcanonical ensemble $-(\rho_0(X,Q) = \delta(E(X,Q) - E_0))$ where E is the total energy and E_0 is a predetermined value of the energy) or of the canonical ensemble ($\rho_0(X,Q) = \exp[-\beta E(X,Q)/Z]$ where β is the inverse temperature and Z is the partition function ($Z = \int \exp(-\beta E(X,Q))dXdQ$)). With the initial conditions at hand we integrate the equations of motions up to the specified time T (m is the diagonal mass matrix and U is the potential energy).

$$\frac{dX}{dt} = m^{-1}Q \quad \text{and} \quad \frac{dQ}{dt} = -\frac{dU}{dX} \tag{1}$$

If the system is ergodic, and if it reaches equilibrium, then an ensemble average of a function $f(X,Q)$ - - $\langle f(X,Q)\rangle$ =

* Corresponding author e-mail: ron@cs.cornell.edu.

10.1021/ct060028m CCC: $33.50
Published on Web 03/28/2006

$\int \rho_0(X,Q) f(X,Q) dX dQ$ can be performed as an average over time and a single long trajectory $\langle f(X,Q)\rangle = 1/T \int_0^T f(X(t),Q(t))dt$. Here we do not require ergodicity or equilibrium, which means that averages must be done with respect to the initial distribution $\rho_0(X,Q)$. A trajectory with a final coordinate vector, $X(T) \in P$, is reactive. If the end configuration is not in P, then this trial failed and produced a nonreactive trajectory (nevertheless, the calculations of nonreactive trajectories is not a complete loss, since nonreactive trajectories can be used to estimate the fraction of reactive versus nonreactive paths). It is obvious that if only a tiny fraction of the trajectories initiated according to $\rho_0(X,Q)$ reacts, then a large fraction of straightforward trial trajectories would be a miss. The actual fraction can be so small that straightforward sampling of reactive trajectories may not be practical.

In light of the difficulties in straightforward sampling of reactive trajectories it is desirable to use a less straightforward approach and bias the sampling of the initial conditions to produce trajectories that spend less time in the reactant state and cross the barrier more readily than typical trajectories. This is in the spirit of the "biased sampling" of individual configurations[1] in computational statistical mechanics. Instead of configurations this time we bias the sampling of trajectories. Such biases are used at different levels:

(i) Generation of approximate or average trajectories from R to P to study qualitative reaction mechanisms.[2] Examples of subsets of "approximate" trajectories include minimum energy or free energy paths[3−8] and trajectories with filtered high-frequency modes[9−12] (of course, free energy paths have more uses than qualitative studies of mechanisms).

(ii) Generating exact trajectories from R to P but without accurate knowledge of their statistical weights. These Newtonian trajectories model better kinetic energy and inertial effects compared to option (i). The probability of the reactive trajectories that are sampled this way is assumed to be significant (see also Figure 1). However, because of the lack of exact weight, it is difficult to compute the reactive flux and the rate. Therefore, additional calculations are necessary to compute the rate in this way.

(iii) The third option computes reactive trajectories and their corresponding statistical weights. The weights enable, for example, the calculations of the rate of the reaction. Option (iii) is the most demanding computationally.

Significant progress was made in the direct application of option (iii) for large molecules in the condensed phase.[13,14] Of considerable theoretical interest is the transition path sampling approach.[13] Nevertheless, rate calculations are limited to systems in equilibrium and to reactants and products that are strong attractors. Trajectories are initiated in the transition region, and the equations of motion are integrated in the backward and forward directions to R and P, respectively. With the additional restriction of equilibrium the weights of the "patched" trajectories can be computed. If R and P are only weak attractors, trajectories initiated at Tr are unlikely to terminate at the desired reactants and products. For the last case this approach is not necessarily better than the straightforward method in which the trajectories are initiated simply at R.

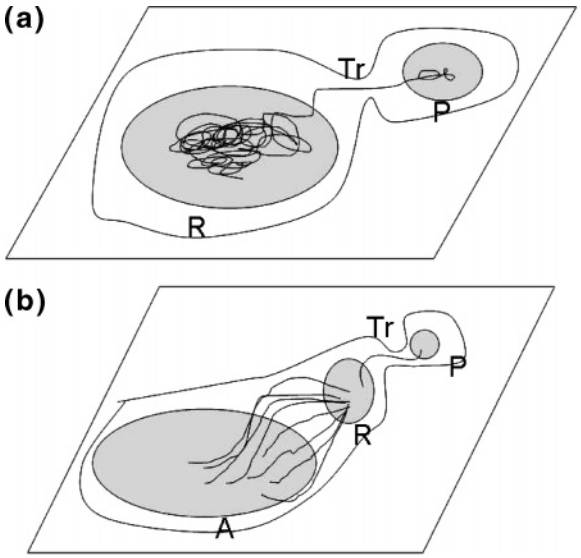

Figure 1. (a) A schematic drawing of an activated trajectory going from R to P. The system is initially trapped in a minimum R and is undergoing a rapid transition to minimum P, making the transition a rare event. Here we are primarily interested in the (short) time in which the trajectory is not in R or in P. (b) A schematic drawing of a rare transition due to geometrical factors. A nonequilibrium distribution is prepared at the state R. This state which is of high energy is relaxed rapidly to P and to A. Since the channel to A is much broader than the channel to P, most trajectories end at A. The transition to P is therefore rare (but fast).

Another important group of rare, short-time, and reactive paths is of nonequilibrium trajectories. The distribution $\rho_0(X,Q)$ can be a product of an external perturbation (absorption of a photon, temperature jump, etc.) that prepares the system in an initial nonequilibrium state and then is turned off. Typically, we follow the relaxation of the system to a new state either computationally or experimentally. For this class of problems equilibrium considerations apply only asymptotically at the long time limit. It is necessary to directly initiate trajectories according to the above distribution. If our interest is focused on a small fraction of these nonequilibrium trajectories that end at a known desired product (e.g. Figure 1), then the selection of reactive trajectories is difficult.

Here we propose an alternative approach for the sampling of reactive trajectories that is not based on initial value formulation. Instead we solve a boundary value problem and compute a whole trajectory from reactant to product as a single minimization task. The proposed approach can be classified as an intermediate between methods (ii) and (iii). The computed trajectories are exact, but the precise calculation of the weight can be expensive. In section VI we consider the computations of relative rates.

Studies of activated trajectories with boundary value formulation have a number of advantages compared to initial value solvers. Most importantly, boundary value formulation selects only reactive trajectories that end at the desired products, while initial value solvers are likely to have many misses. Nevertheless, there are also some negative aspects to the use of boundary value formulation. First, the optimiza-

tion of the whole trajectory (many time slices) is demanding in computer memory and expensive in terms of CPU time. These difficulties are to be contrasted with the calculation of a single frame at a time in initial value formulation that requires less memory and CPU per step. For a fixed time step and system size the boundary value formulation is considerably more expensive computationally. In practice boundary value calculations of exact classical trajectories are limited to short-time trajectories for which the cost of the calculation is not high. An estimate of the computational efficiency of boundary value calculations compared to initial value studies is provided in section VII.3

The limitation to short-time trajectories is not a significant restriction for the problem at hand since both of the processes that we are interested in: (a) passage over a large energy barrier, and (b) rapid relaxation to rare products, are quick (if we remove from trajectories of type (a) the "incubation" period, which is the (typically long) time spent at the reactant state prior to crossing the barrier).

Another nontrivial problem with the boundary value formulation is of uniqueness. In contrast to initial value formulation that generates a single unique trajectory (when the initial velocities and coordinates are given), the boundary value formulation (the initial and final coordinates and the total time are fixed) can have more than one solution. In section III.1 we illustrate numerically that the usual boundary formulation, based on Lagrangian mechanics, is ill posed. Numerous (and different) trajectories solve the same boundary value problem. The multiplicity of solutions needs to be addressed to ensure the proper convergence of the calculations to a unique solution and in order to estimate the weight of a particular boundary value. We introduce in section VI an algorithm to compute relative rates in which we sample trajectories, compute averages, and avoid the absolute weight problem.

The multiplicity of boundary value solutions should not come as a surprise. Consider a long time ergodic trajectory that reaches the "statistical-mechanics" limit in which a single trajectory samples extensively the coordinate space. In this case the probability of "hitting" the end point is independent of the initial conditions or the total time of the trajectory. This means that for a sufficiently large number of trials and long times more than one solution for the prespecified end point is likely to be observed.

II. Minimal Time, Point-to-Point Calculation of Classical Trajectories

The problem is formulated as follows: Given the two states R and P and two coordinate sets $X_i \in R$ and $X_f \in P$ of a reactive trajectory, compute a classical mechanics trajectory with a minimal time T and a bias to low energies. More specifically, we seek the stationary path of the classical action,[15] $S = \int_{0,X_i}^{T,X_f} L dt$, for which T is a minimum. The bias to low energies is added to a Monte Carlo procedure that searches for the classical path. The minimal time requirement removes a significant incubation period and, as we demonstrate below, makes the problem better posed.

Boundary value formulations (without the minimal time condition) were proposed by Gillilan and Wilson[16] in the context of a rubber band model of trajectories, by Cho, Doll, and Freeman[17] using Fourier space, and by Olender and Elber[9] in the context of the Onsager-Machlup action.[18] Recently, Passerone et al. studied activated trajectories using another boundary value formulation.[19]

As mentioned in the Introduction the application of the boundary value approach to compute *long time and exact* dynamics is difficult because of the following: (a) The optimization of an accurate long trajectory (including incubation time) is demanding in computer memory and CPU time compared to the calculation of a single frame at a time in initial value formulation. (b) We illustrate below that the solution for a long time boundary value problem is not unique. The above two problems are addressed by the minimal time condition that selects trajectories that are (i) nearly unique and (ii) more direct and therefore easier to compute. Hence besides focusing on the most interesting component of the transitions, these trajectories are more accessible to boundary value calculations.

For computational purposes it is convenient to write a discrete version of the classical action. The action becomes a function of the set of intermediate coordinates

$$S = \sum_{i=1}^{N} L_i \Delta t = \sum_{i=1}^{N} \left(\frac{m}{2} \left(\frac{X_i - X_{i-1}}{\Delta t} \right)^2 - U(X_i) \right) \Delta t \qquad (2)$$

The discrete coordinate sets provide a "basis-set" for the trajectory. X_i is the vector of (Cartesian) coordinates of time slice i. The condition that the action is stationary provides a set of equations for the intermediate coordinates, while X_0 and X_N are kept fixed.

$$\frac{1}{\Delta t} \frac{\partial S}{\partial X_i} = m \frac{X_{i+1} + X_{i-1} - 2X_i}{\Delta t^2} + \frac{\mathrm{d}U(X_i)}{\mathrm{d}X_i} = 0 \; i = 1,...,N-1 \qquad (3)$$

These equations are solved iteratively as follows:

1. Use a guess to interpolate between the two minima. In the examples below a straight line interpolation was used. Pick an initial time for the trajectory that should be small based on some knowledge of the system properties (e.g. a transition time in peptides is less than a picosecond). Use a small number of time slices (e.g. 20−40) to describe the trajectory.

2. Solve the second-order system of Verlet's equations (eq 3) with fixed boundaries. From (3) the residual vectors for the internal points $i = 1,...,N-1$ are $r_i = X_{i+1} + X_{i-1} - 2X_i + m^{-1}\Delta t^2 \, \partial U/\partial X_i$. The goal of the solution is to reduce all r_i to 0. This is accomplished by point-by-point iterations either by simulated annealing (SA) that minimize the sum of $r_i^t r_i$ or by Kaczmarz iterations[20] (the simple Gauss-Seidel iterations[21] do not converge). In Kaczmarz iterations, in contrast to Gauss-Seidel's, changes are made not only to a single point (e.g. point i) but also to all its neighboring points as well. The neighboring points are all those points appearing in the equation for the ith point. The coefficients of the neighboring points in the ith point equation determine the weights of the changes of these points. While iterating on the ith equation, the positions X_i, X_{i-1}, and X_{i+1} are updated so that r_i vanishes. A global minimizer (like simulated

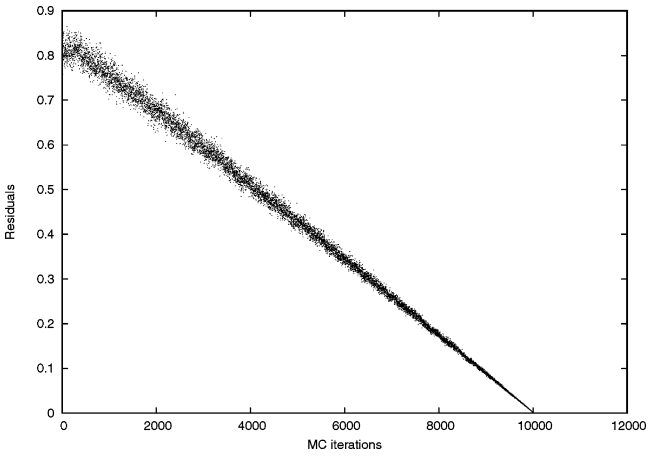

Figure 2. The convergence of the target function *R*2, which is the sum of the square of the residuals as a function of the simulated annealing step. Note the initial increase in the value of the target function before the calculations settle into approximate linear decay of the residuals. The example is from a calculation for the Lennard-Jones cluster (see section V for a more detailed discussion).

annealing) is necessary since elimination of local minima is required. The observation that the residuals sometimes increase before decreasing suggests that the system was trapped in a local minimum. A global minimizer also enables the search for multiple solutions. Convergence is assumed when the norm of residuals $||r|| = 1/N \sum_{i=1}^{N-1} |r_i|$ falls below a cutoff value (e.g. for the Mueller potential the cutoff is 10^{-9}) and the energy is conserved up to a few percents. A typical convergence plot for a simulated annealing run is shown in Figure 2.

3. Gradually decrease the total time T and use the spatial coordinates obtained with the larger T of the previous time as an initial approximation for the smaller T. Solve the resulting system as described in step 2.

4. When T becomes small enough in step 3, either convergence fails because of a sudden and significant change in the shape of the trajectory or some other significant changes occur in the system, such as an abrupt rise in the energy. The process stops at that T, and the solution of the previous total time is kept. We locate a *local* minimum of time in the neighborhood of the initial guess (of the minimal time) and in the neighborhood of a given energy. We do not search for a global minimum (which is a straight line and of high energy). A search bias toward lower energies can be added in the annealing process by a penalty function E_{penalty} which, at the time slice i, is equal to 0 if the energy E_i is below a given value E and to $\lambda(E_i - E)$ when $E_i > E$. λ is a positive constant that determines the relative weight of the penalty. In the examples presented in this work such a bias was added in the calculations of alanine dipeptide trajectories and in the trajectory calculations of the Lennard-Jones cluster.

III. Numerical Experiments on the Mueller Potential

To examine a number of alternative solutions that satisfy the same boundary value conditions, we perform an extensive sampling of trajectories using initial value formulation on the Mueller potential[22]

$$U(x,y) = \sum_{i=1}^{4} A_i \exp[a_i(x - x_i)^2 + b_i(x - x_i)(y - y_i) + c_i(y - y_i)^2]$$

$$A = (-200,-100,-170,15) \quad a = (-1,-1,-6.5,0.7)$$
$$b = (0,0,11,0.6)$$
$$c = (-10,-10,-6.5,0.7) \quad x = (1,0,-0.5,-1)$$
$$y = (0,0.5,1.5,1) \quad (4)$$

We choose the Mueller potential since it is simple and accessible to detailed analysis. It was used extensively in the past to study reaction paths, and many of its properties are well understood. Below we describe both initial and boundary value calculations of trajectories on the Mueller potential.

III.1. Initial Value Calculations To Solve the Boundary Value Problem. The Verlet integrator[23] is used with a time step of 0.0001. The mass was set to 1. The starting coordinates were at the lowest energy minimum (−0.5,1.5), the initial velocities were sampled with a random direction, and the total energy was fixed at −40.5 which is just above the barrier height. If the coordinate vector at any time slice during the trajectory was found within a distance of 0.01 from another minimum (0.5,0) we consider the computed trajectory a solution of the boundary value problem, and the time T of the trajectory was recorded (the end point can be determined only approximately within the neighborhood of the second minimum). We have repeated the calculation using two radii to check the convergence of our calculations and obtained similar results. Below we show a histogram plot (Figure 3) of the probability of finding a trajectory with the prespecified end points as a function of the total time of the trajectory.

The most striking feature of this plot is that trajectories obtained by the boundary value formulation are far from unique (at least in the numerical sense). The trajectories so computed are even "overdetermined" since the energy (in addition to total time and end points) was also fixed. In practice, depending on the calculation history, different trajectories may be obtained. The abrupt decay of the probability at time ≈ 400 is a reflection of the limited sample that we used rather than a property of the true distribution.

Since the directions of the velocities were sampled at random, it is possible that some of the sampled trajectories considered in Figure 3 are very similar. A question of interest is how different are computed trajectories that made it from reactant to product at the same total time and energy.

In Figure 4 we measure the similarity of 32 alternative trajectories with the same end points, energy, and total time T. The coordinates are sampled at three time slices: $\tau = T/4, T/2, 3T/4$ of each trajectory. At each time slice the variance $\sigma^2(\tau) = \langle (X_i - \langle X_i \rangle)^t (X_i - \langle X_i \rangle) \rangle$ of the coordinates X_i, $i = 1,...,32$, is computed from the trajectories. The shortest time trajectories converge to a very small variance (about zero), while at somewhat longer times a rapid increase in the variance (suggesting distinct trajectories) is evident.

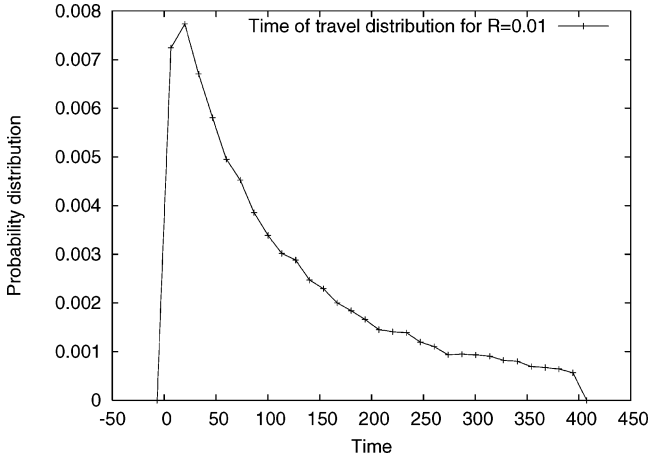

Figure 3. The distribution of the total time of the trajectories that are satisfying spatial boundary condition: The trajectories start at the lowest energy minimum (−0.5,1.5) and arrive at the second lowest minimum at (0.5,0). The total energy was fixed at −40.5, and the direction of the velocity was sampled at random. The drop of the probability density near 400 is due to limited statistics. The radius at the product was taken to be 0.01.

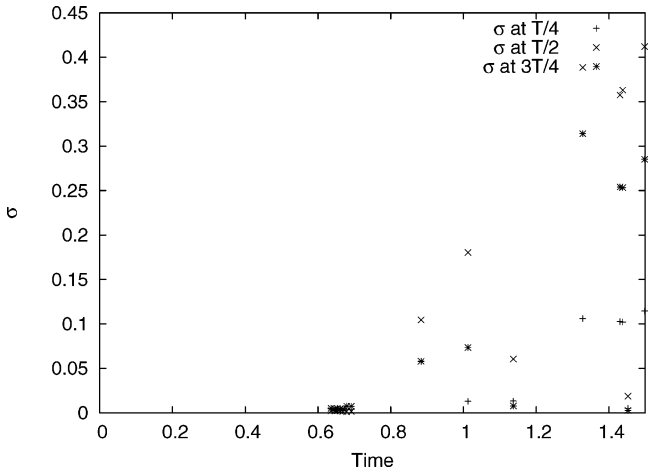

Figure 4. The variance of the alternative trajectories that solve the same boundary condition. As in Figure 3 the trajectories started from the same energy minima, and the total energy was the same. In contrast to Figure 3 we consider trajectories with the same total time and measured the variance of these trajectories as described in the text. It is evident that the variance is small for the shortest time trajectories, while the variation between the trajectories increases rapidly as a function of the trajectory time.

Figure 5a shows three trajectories satisfying the same boundary conditions that are clearly different. Figure 5b shows three trajectories at the short time limit that (within our numerical accuracy) correspond to essentially the same path.

III.2. Boundary Value Calculations. Here we used our proposed boundary value algorithm (section II). Trajectories were computed between the two energy minima of the Mueller potential that were studied with initial value formulation (see the previous section). Since the time step used for the discretization of the functional (eq 2) is significantly larger than the time step of the initial value

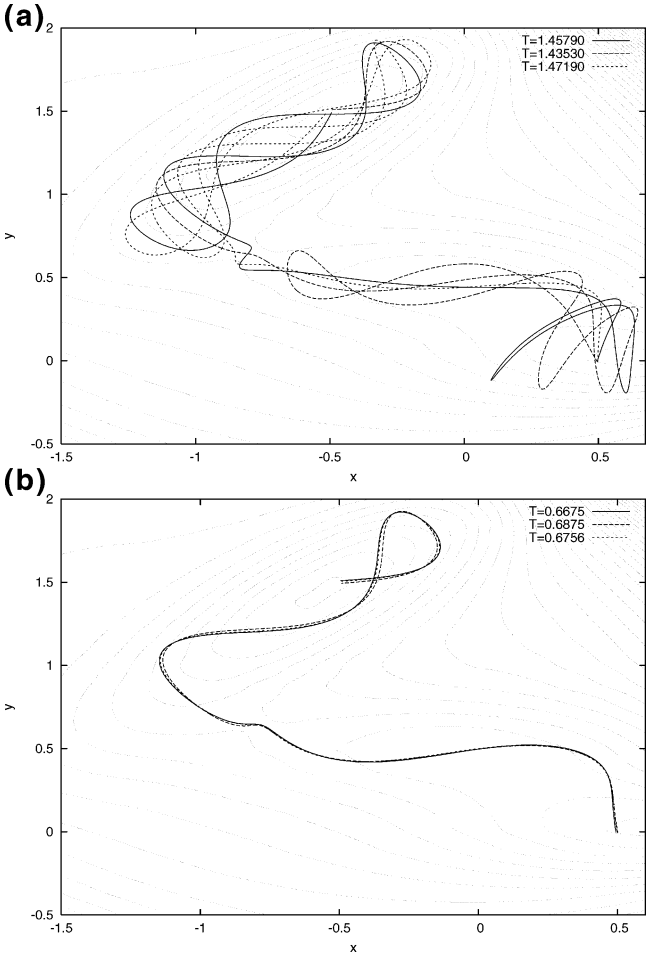

Figure 5. Sample of trajectories that satisfy the same boundary conditions as described in the text and the legends of Figures 3 and 4. (a) Three solutions of the same boundary value problem that are clearly different. (b) Three solutions of the same boundary value problem that seem to converge to the same solution.

formulation (0.01 versus 0.0001), it is desirable to use a refinement procedure. The refinement should convert the coarse grained boundary value trajectories to trajectories with the same time step as the initial value formulation, enabling a meaningful comparison of algorithm accuracies. We note however that for numerous other applications (like the calculation of the relative rate in section VI) the refinement is not necessary.

The proposed refinement is using iterations; i.e., given a coarse grained trajectory with time step ΔT, every iteration generates a new finer trajectory with half the time step ($\Delta T/2$). The boundaries of the finer trajectory are fixed to the same values as the coarse trajectory. Intermediate points are added using linear interpolation between points of the coarser trajectory as an initial approximation and followed by several Newton−Raphson steps that minimize the residuals of the equations of motion (eq 3). At each of the Newton−Raphson steps the unknown corrections δx_i and δy_i to the coordinates are calculated by solving simultaneously for all internal points $i = 1,N - 1$ the system of equations consisting of $\delta x_{i+1} + \delta x_{i-1} - (2 - \Delta t^2\ \partial^2 U/\partial x_i^2)\delta x_i + \Delta t^2\ \partial^2 U/\partial x_i \partial y_i\ \delta y_i = -r_{x_i}$ and their equivalent in the y direction.

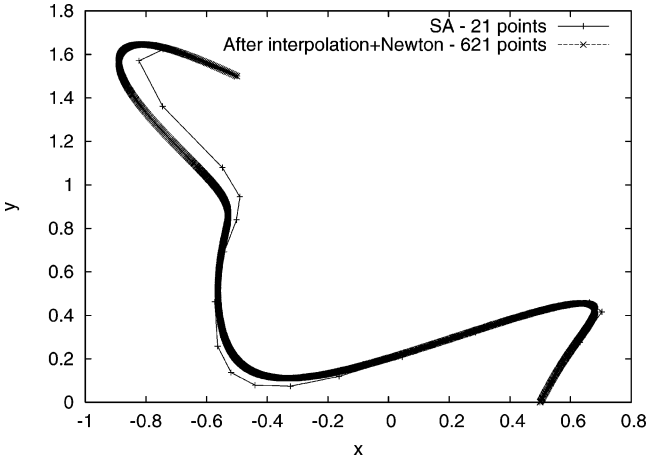

Figure 6. Coarse grained and highly refined boundary value trajectories on the Mueller potential. The refined trajectory is essentially identical to the short-time solution of the Verlet algorithm. The coarse grained trajectory (using a time step larger by a factor of 30) is quite accurate.

In Figure 6 we show a coarse grained and a highly refined trajectory. The coarse grained trajectory, which includes 21 points, is a minimal time trajectory obtained by the procedure described in section II. The initial guess was a straight line between the two end points. The initial total time was 0.5, and Kaczmarz iterations[20] were used until the residuals dropped to below 10^{-9}. The total time was gradually decreased until convergence stopped at total time 0.375. The coarse grained trajectory is the one obtained by the end of this process. The refined trajectory is very close to the exact Verlet trajectory. Even the coarse grained trajectory with a time step 30 times larger is quite accurate. This example demonstrates the benefit of working with short-time trajectories. The particle does not spend much time in oscillatory motion near the minima and focuses instead on executing the more interesting type of motion (the transition). We argue below that these trajectories have significant statistical weight in the microcanonical ensemble. Therefore they are worthy of investigation, even when the overall time scale is slow. Of course it is easier to compute exact trajectories if their total time is short, and if they are less curved and more direct.

IV. Molecular Example I: Transitional Trajectories in Alanine Dipeptide

We have repeated the boundary value computation also for the considerably more complex system of a conformational transition in alanine dipeptide. We consider the conformational transition from a helix to an extended chain. The coarse grained description of the trajectory includes 20 points. The time step was 0.00875 ps, making the total trajectory time 0.175 ps (which is the minimal time required for the transition as we confirmed using the procedure described in section II). This minimal time was also observed by extensive sampling of initial value solutions and recording the time of arrival to the product. The initial guess for the path was a straight-line interpolation, and a simulated annealing algorithm was used to minimize the norm of the residuals, defined identically to the Mueller problem. The starting residual norm was 10^5, and the cooling continued until the residuals dropped to about 10^{-5} where changes in space coordinates for the Metropolis algorithm were about 10^{-7} Å. In Figure 7 we show boundary value trajectories for several total times. All these trajectories are similar to the minimal time trajectory found by extensive sampling of initial value trajectories. The sampling of initial value trajectories was performed in a similar spirit to the calculations described for the Mueller potential. The alanine dipeptide system is considerably more complex than the Mueller potential, and it is possible to find new converged trajectories while making the total time shorter and shorter. However, as we decrease the total time, we obtain an abrupt rise in the total energy, shifting to a different class of trajectories. We terminate the process as described in section II when such a sharp change in the energy is observed.

V. Molecular Example II: An Isomerization Process in a Lennard-Jones Cluster

A more complex molecular example is the isomerization of a Lennard-Jones cluster. The potential energy of the system is given by $U = 4\sum_{j>k}[[1/q_{jk})^{12} - (1/q_{jk})^6]$ where q_{jk} is the distance between any two particles. The masses were set to one. A 38 atom LJ cluster has two minima with nearly the same lowest energy: the first is at the MacKay icosahedra and the other at a face-centered-cubic truncated octahedron[24] (Figure 8). In the following calculations the two boundaries are set to minima in the basins of the two distinct conformations.

To provide a rough estimate of the minimal time of the trajectory the period of oscillation of a Lennard-Jones term is computed by harmonic analysis as ~0.85. The expectation is that the minimal transition time will correlate with the vibrational time. The total time of the transition in the calculation was therefore set to $T = 2.0$. The number of slices was 33, and the uniform time step was set accordingly to $\Delta T = T/32$. The time was decreased in the optimization following section II until it was impossible to conserve the energy. The energy conservation was used as a stopping criterion.

The solution of the boundary value problem was done by Monte Carlo simulated annealing, minimizing the functional $R2 = \sum_{i=1}^{N-1} r_i^t \cdot r_i$. The term on the right-hand side is an inner product of the residual vector at each nonboundary time slice i. The slices are numbered between 0 and N. The initial approximation of the coordinates at nonboundary slices is obtained by linear interpolation from the boundaries.

The simulated annealing starts at a temperature $T_{\text{init}} = 5.0\text{e-}4$. The cooling protocol is linear, decreasing the temperature by a constant ΔT at each cooling step. The number of temperatures was $N_T = 1000$ and $\Delta T = T/N_T = 5.0\text{e-}7$. The process ends when the temperature reaches its final value, ΔT. To avoid trapping in a local minima the values of all space coordinates were discretized with a small mesh size $H_{\text{SA}} = 1.0\text{e-}4$. The initial coordinates are rounded to the nearest value which is an integral multiple of H_{SA}. The changes to coordinates in the Metropolis algorithm are always in integral multiples of H_{SA}. For each temperature, 200 Monte Carlo sweeps were done. A single sweep consists of a Metropolis step for each coordinate in the internal time

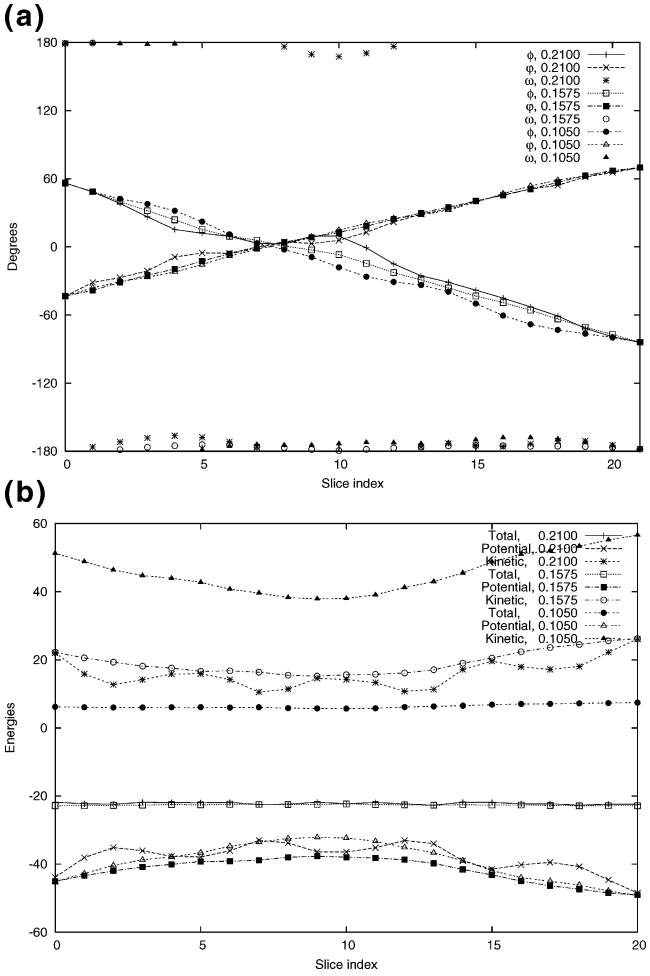

Figure 7. (a) A short-time trajectory for a conformational transition in alanine dipeptide. The calculations include all the 36 degrees of freedom of this molecular system. However, the projection onto the two soft degrees of freedom (the torsion angles φ and ψ) and one stiff degree of freedom (ω) is shown. A sequence of three times is shown to demonstrate our convergence criterion (shown in part b). (b) The energetic of the conformational transition in alanine dipeptide for three different trajectories of short time. Note the significant jump in the total energy and the kinetic energy when the time of trajectory is shortened from 0.1575 to 0.1050. We use this as an indicator that the minimal time trajectory (of low energy) converged at 0.1575.

slices. The value of the functional $R2$ was monitored to ensure that it decreases linearly with the temperature (Figure 2). At the initial temperature the maximum step size in the Metropolis algorithm was 400 H_{SA}, and the attempted changes to coordinates in the Metropolis algorithm were distributed uniformly between $-400H_{SA}$ and $400H_{SA}$. At later temperatures, the maximum allowable step size was decreased by a factor of 0.8 if the acceptance rate of the Metropolis algorithm fell below 0.4 in the previous temperature. The residuals decrease in the annealing process from a starting value of about 0.8 to a final value 0.002.

Figure 9 displays the total, potential, and kinetic energies at convergence as a function of time. The maximum of the potential energy is near -150 which is consistent with low kinetic energy when passing transition states. The total energy is nearly conserved, and the small fluctuations are likely to be due to the discretization.

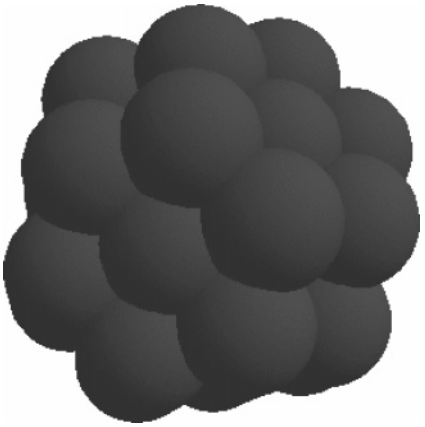

Figure 8. The initial structure of the conformational transition in the Lennard-Jones cluster (MacKay icosahedra). For the complete transition check http://www.cs.cornell.edu/ron/movies/lj_img.gif.

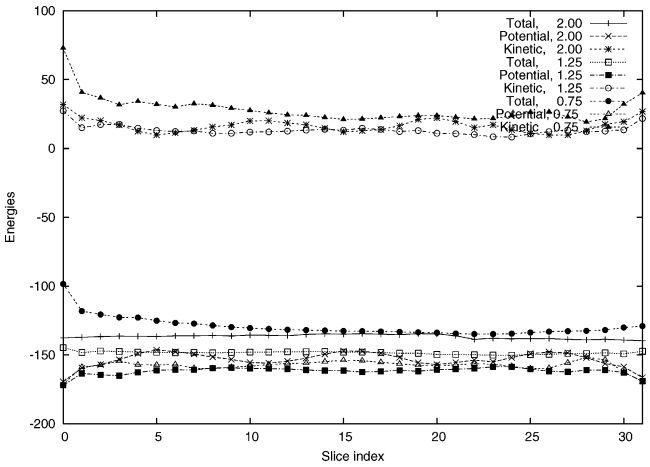

Figure 9. The energetic of a transition in 38 atoms of a Lennard-Jones cluster. Note that the total energy is clearly a constant for a total time of 2.00 but is fluctuating significantly for shorter times.

VI. Calculations of Relative Rates

Here we show how the short-time trajectories can be used in the calculation of rates and relative rates at short times. The restriction for short times is of particular interest for downhill but rare events and for activated processes (Figure 1).

In initial value formulation the reaction probability from R to P between time T and $T + dT$ is given by

$$p(T)dT = \left[\int_{\substack{X\in R \\ X(T)\in P}} Y_{RP}(X,Q,X(T),Q(T))\rho_0(X,Q)dXdQdX(T)\right]dT$$

where $Y_{RP}(X,Q,X(T),Q(T))$ is the probability density (in phase space and time) that a trajectory that starts in the phase space point (X,Q) in R will make a transition to P at time T. We used the shortcut (X,Q) to denote $(X(0),Q(0))$. For the (deterministic) Newton's equations of motion, $Y_{RP}(X,Q,X$-

$(T),Q(T))$ is either one or zero. Estimating the above integral there are two functions that require accurate and efficient sampling - - Y_{RP} and ρ_0. In initial value formulation we sample ρ_0 exactly, but the calculation of Y_{RP} can be difficult. If the process is rare, Y_{RP} will be zero most of the time. In the boundary value formulation Y_{RP} is always one, but the sampling is not done according to ρ_0 which complicated the calculation of the weight of the trajectories.

In the point-to-point approach we do not have the initial momentum as input. We sample configurations at both the reactants and the products sites $X(0) \in R$ and $X(T) \in P$ and compute trajectories of fixed time length T between any pair of sample points. The initial momentum is an output that we can compute by a finite difference formula from the calculated trajectory (i.e. $Q(0) \cong m(X_1 - X_0)/\Delta t$), once it is available.

Here we show how the relative rates of two different systems can be estimated, relying on a set of reactive trajectories only. Calculation of relative rates of similar systems is widely used in computational chemistry and biophysics. For example, we may consider changes of rates following the replacement of a hydrogen by a deuterium atom or a single residue mutation of a reacting protein. The time itself can also be a perturbation (e.g. by changing the size of Δt keeping the number of time slices fixed).

A concrete case is of a reaction at constant temperature. In the corresponding canonical ensemble, we have (E is the total energy)

$$\rho_0(X,Q) = w/Z$$

$$w = \exp(-\beta U(X)) \cdot \exp(-\beta Q^2/2m) \equiv \exp(-\beta E)$$

$$Z = \int \exp(-\beta E) dX dQ$$

Consider two systems with reactants and products R,P and $\bar{R},\bar{P}$, respectively, the ratio of the reaction probability densities is

$$\frac{\bar{p}(T)}{p(T)} = \frac{Z}{\bar{Z}} \cdot \frac{\int_{X,Q\in\bar{R}} \bar{w}(X,Q)\bar{Y}_{RP}(X,X(T))dXdQdX(T)dQ(0)}{\int_{X,Q\in R} \bar{w}(X,Q)Y_{RP}(X,X(T))dXdQdX(T)dQ(0)}$$

The ratio of the partition functions $(Z/\bar{Z})$ can be computed with standard methods in equilibrium statistical mechanics such as the umbrella sampling, free energy perturbation, dynamic integration, and more (for a recent discussion on alternative ways to compute free energy see ref 25).

$$\frac{\bar{Z}}{Z} = \frac{\int_{X\in R} \exp(-\beta(\bar{E} - E)) \exp(-\beta E) dX}{\int_{X\in R} \exp(-\beta E) dX} \equiv \langle \exp(-\beta(\bar{E} - E)) \rangle_{\exp(-\beta E)}$$

The above integration is restricted to the reactant state R.

Assume that a trajectory is found by the point-to-point approach. The weight of the initial structure is $\exp(-\beta\bar{E})$. This trajectory is perturbed to a new trajectory with a new energy, E. Since the *total energy* is not expected to vary violently from a trajectory to a trajectory, the average below should not be too difficult to perform.

The second ratio of integrals is called RR for "Relative Rates".

$$\mathrm{RR} = \frac{\int \exp(-\beta(E - \bar{E})) \frac{Y_{RP}(X,X(T))}{\bar{Y}_{RP}(X,X(T))} \exp(-\beta\bar{E}) \bar{Y}_{RP}(X(T) - X_P) dX dX(T)}{\int \bar{Y}_{RP}(X,X(T)) \exp(-\beta\bar{E}) dX_R dX_P}$$

$$= \left\langle \exp(-\beta(E - \bar{E})) \frac{Y_{RP}(X,X(T))}{\bar{Y}_{RP}(X,X(T))} \right\rangle_{\bar{Y}_{RP}(X,X(T))\exp(-\beta\bar{E})}$$

The existence of a corresponding point-to-point trajectory with a slightly different Hamiltonian is tested first (in our experience we were always able to find a nearby perturbed trajectory with a different energy), and the average of the energy differences is performed explicitly. This is an important difference from the initial value formulation in which the ratio of the Ys can be zero most of the time.

The algorithm to compute the RR is therefore as follows:

(i) Sample start and end configurations in R and P.

(ii) Compute a trajectory according to the "sampling" Hamiltonian using total fixed time T that is short but not minimal with the point-to-point approach,

(iii) Compute a trajectory for the perturbed Hamiltonian starting from the trajectory in (ii) as an initial guess. If convergence fails add zero to the average and go to (v). It is also possible to compute the change in energy by perturbation theory with no need to reoptimize the reference trajectory.

(iv) Estimate the difference in energy of the two trajectories and add to the RR.

(v) Check for termination conditions (convergence of the average or maximum number of allowed trajectories, etc.). Return to (i) if not terminating.

An example of a calculation of relative rates on the Mueller energy surface is sketched below. The original Mueller potential was used for sampling, and the perturbation was linear in the coordinate y, i.e., the perturbed potential is $U + 5y$, where U is the Mueller potential of eq 4. The reactant and product were defined with a radius of 0.2 around the location of the corresponding minima. The reactant was the minimum at the upper left corner of the Mueller potential. The total time was set to 0.38, slightly longer than the minimal time of 0.375 observed earlier. We first computed initial value trajectories with the Verlet algorithm (the time step was 0.001) and estimate the absolute rate for the two Hamiltonians (the absolute reaction probability of the unperturbed Hamiltonian was 0.0012). We then use a coarse grained trajectory with 38 points and estimated the perturbation. The rate difference was 10% in both the boundary and initial value approaches.

VII. Discussion

VII.1 Why Is It Possible To Use a Larger Time Step in Boundary Value Formulation? Equation 3, which is a discrete version of the Newton's equations of motion, is similar in accuracy to initial value algorithms. Rearranging eq 3 we can recover an initial value algorithm:

$$X_{i+1} = 2X_i - X_{i-1} - m^{-1}\Delta t^2 \frac{dU(X_i)}{dX} \quad (5)$$

If the coordinates at X_0 and X_1 are given, we can use the above formula to propagate the solution to its final configuration at X_N. A solution of the discrete boundary value problem is also a solution of the discrete initial value problem. This relationship is particularly intriguing in light of the observation on stability and accuracy made in the present study. We stated that the boundary value formulation is accurate and considerably more stable numerically compared to initial value formulation. Given that a solution of the boundary value formulation is identical to a solution obtained with initial value formulation, as discussed above, the statement seems false.

Therefore we have performed the reverse numerical experiment in which we start with initial value description. We sample different initial values (X_0 and X_1) and use eq 5 to generate trajectories of the same characteristics as in the boundary value formulation. If we employ the large time step that was used in the boundary value formulation, most of the initial value trajectories lose stability and "explode" in accord with the usual observations. Nevertheless, it is possible to find rare initial value trajectories with large time steps that maintain their stability (and accuracy). These are the trajectories that are picked by the boundary value formulation. Since the end structure in the boundary value formulation is sound (by construction) all the "exploding" unstable trajectories are eliminated at the start. Only the stable trajectories remain. This explains the observed stability of the approach proposed here. The enhanced stability of the boundary value formulation also supports the Stochastic Difference Equation approach[10−12] which focuses on generating approximate long-time (or length) trajectories that qualitatively describe reaction mechanisms. The present approach is centered on an accurate calculation of short-time trajectories.

VII.2. Are Minimal-Time Activated Trajectories Significant? We have invested considerable effort to show that the minimal time trajectories are nearly unique and as such can be computed accurately and efficiently with boundary value formulation. However, typical calculations of physical observations (such as calculations of rate) rely on an ensemble of trajectories and not on a unique path. Therefore the significance of minimal time trajectories and their contribution to the ensemble of reactive paths require examination.

A first impression suggests that these trajectories are not significant. Since all initial values (coordinates and momenta) in the microcanonical ensemble have the same statistical weight, a single trajectory cannot be representative (in this section we consider trajectories with fixed energy and time).

A more subtle argument focuses on the transitional component of the trajectories. Consider reactive trajectories of total time T. We will divide this long-time trajectory into two components: (i) an incubation part of length $(T - \tau)$ in which the system spends some time either in R or in P and (ii) a transient component (length τ) in which the system is in Tr. Since both R and P are attractive states we expect that $\tau \ll T$. A transitional trajectory is initiated at the interface between R and Tr, called RTr, and is terminated at the interface between Tr and P called TrP. A transitional trajectory spends all its time (τ) in Tr without "touching" RTr nor TrP at any time after initiation or before termination. A sample of such boundary value transitional trajectories makes it possible to compute, $P(\tau)$, the probability of observing a transitional trajectory of time−length τ. The details of the distribution will depend on the underlining energy surface, and we cannot be more specific. However we expect (due to the assumed transient nature of the trajectory) that $P(\tau)$ is a rapidly decreasing function of τ. Moreover, at very short times $P(\tau)$ must go to zero. Therefore it has at least one maximum.

Experiments rarely measure properties of transient times and trajectories. More often the overall time t (including the incubation time of component (i) of the trajectory) and rate are available. We therefore need to extend our argument to trajectories of total length t and transient time τ (communicated to us by Cristopher Dellago).

We now consider a discrete (numerical) representation of the trajectory. The numerical trajectory of total fixed time is described by a sequence of coordinates separated by a time step of Δt: $X_1,X_2,...,X_n,X_{n+1},...,X_m,X_{m+1},...,X_k$. Similarly to the continuous trajectory the discrete path is divided into three parts, $X_1,...,X_n$—a set of configurations that reside in the reactants, R, $X_{n+1},...,X_m$—a set of configurations that reside in the transition state, Tr, and $X_{m+1},...,X_k$—a set of configurations in the product, P. The transitional trajectory that we have in mind is the middle piece (from X_{n+1} to X_m, or to make the transition complete perhaps from X_n to X_{m+1}, we ignore this small difference). Assuming that we use eq 5, then with the initial pair of coordinates $-X_1$ and X_2, we can generate the whole trajectory from X_1 to X_k. Note however that we are not interested in the whole trajectory. We are only interested in the transitional component from X_{n+1} to X_m. The statistical weight of this trajectory is the number of initial conditions in the reactant space that results in the same *transitional* trajectory. Similarly to the reptation mechanism of polymer transport we now remove the first coordinate set and create a new coordinate set at the end of this trajectory (creating a trajectory from X_2 to X_{k+1}, X_{k+1} is uniquely defined since X_k and X_{k-1} are known). This is a trajectory of the same total length and exactly the same transitional trajectory (for $n > 1$). However, it has different initial conditions in the reactant state (they are now X_2 and X_3). We can repeat this process and move the first coordinate to the end of the trajectory, keeping the same transitional trajectory, n times. Since we consider a trajectory of total fixed length of time, a larger n means a smaller $m-n$. The shorter is the transitional trajectory the larger is n, and more initial conditions provide the same transitional trajectory.

Typically all initial conditions will have the same weight if sampled from the microcanonical ensemble; however, the above argument, restricted to an ensemble of trajectories with fixed energy and total time, suggests that shorter transitional trajectories have larger statistical weights that skew the original $P(\tau)$ to short times. The degree of skewing will depend on the particular energy surface.

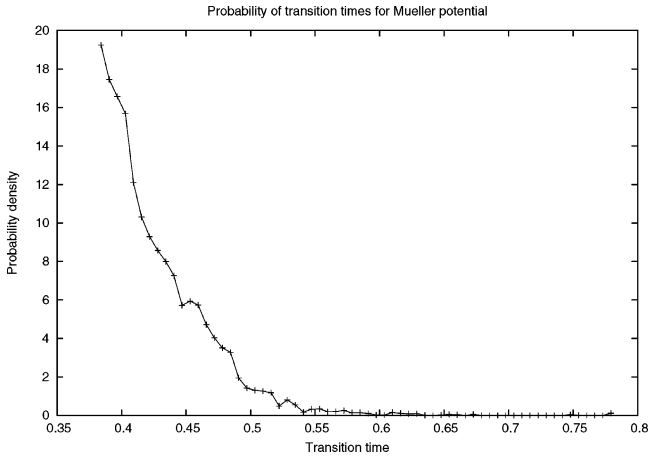

Figure 10. Distribution of transition times from two states in the Mueller potential. The reactant state is a circle with a radius 0.35 centered at (−0.5,1.5), and the product is another circle with a radius 0.3 centered at (0.5,0). Transitional trajectories were sampled by the Verlet algorithm for a total energy of −38 and a total time of 50. The time in which the trajectory is not in the reactant or the product is the transition time. Note that in the canonical ensemble (that prefers low energies and slower transition times) the peak shifts to the right.

To illustrate the above argument we computed the distribution of the transitional times for the Mueller potential. The probability density of transitional times for the Mueller potential is highly peaked at the shortest time (shown in Figure 10).

Note that for more diffusive transitions, and trajectories that cross slowly and/or many times the barrier, the argument above is not valid.

VII.3. Efficiency of the Calculation. The boundary value calculations of this manuscript are reasonably efficient. The basic unit for cost evaluation is the cost of energy calculation which is roughly proportional to the number of particles in the system, N. The cost of a single action evaluation is proportional to $L \times N$ where L is the number of time slices. The number of relaxation steps depends on the system but (in practice) is no worse than another factor of $L \times N$. Nevertheless, this factor is perhaps the major source of uncertainty in the formula provided below, and in some cases it can be larger. Accepting this estimate, the total cost is proportional to $(L \times N)^2$. This is similar to earlier boundary value calculations[9−12] in much larger systems, producing approximate trajectories.

For comparison, the computational cost for a single initial value trajectory is proportional to $\bar{L} \times N$, where $\bar{L}$ is the number of time slices we require in an initial value solver. We denote the probability of an initial value trajectory to be reactive by p. The computational cost for a reactive trajectory is therefore proportional to $p^{-1}(\bar{L} \times N)$. The ratio of costs of the two approaches is

$$\text{ratio of costs} = a\frac{(L \times N)^2}{p^{-1}(\bar{L} \times N)} = a\frac{L^2 \times N}{p^{-1} \times \bar{L}} \qquad (6)$$

where a is a numerical constant which is estimated to be of order of 10. In the most complex example of this manuscript (the Lennard-Jones cluster) L is 33 and $\bar{L}$ is about 100 times larger. The number of degrees of freedom, N, is 38×3. We can estimate what is the reaction probability in which the cost of computing a reactive trajectory using an initial value or the boundary value approach is roughly the same (ratio of costs equal one). $1 = 33^2 \times 38 \times 3 \times 10/p^{-1} \times 3300\ p \approx 0.001$, which is a process that is only marginally activated (i.e. not so rare). The above formula also suggests that as the system becomes larger the reaction probability better be smaller to make the boundary value calculations of the type described here more effective. Indeed the reaction probabilities of complex systems at short times are significantly smaller than 10^{-3}.

Parallelization does not change the above conclusions in an obvious way. Let the number of processors be n. Since the time slices can be distributed between the processors with limited communication overhead we can replace one of the L by L/n in eq 5 (the required number of minimization steps is reduced from the original cost of $L \times N$, but the extent is not clear). Initial value trajectories can be parallelized in a trivial way (launching independent trajectories on separate processors), and we therefore replace p^{-1} by $(np)^{-1}$. The cost ratio for n processors is therefore unchanged or perhaps slightly favors the boundary value formulation.

Finally we note that the calculation of the perturbed trajectory for the study of relative rate takes little effort within the boundary value formulation. In initial value formulation it takes the same effort as the generation of the unperturbed trajectory.

Acknowledgment. This research is supported by NIH Grant GM59796 to R.E. We thank Eric vanden Eijnden and David Shalloway for useful discussions. The initial and final coordinate sets of the Lennard-Jones cluster were kindly provided by David Wales.

Note Added after ASAP Publication. This article was released ASAP on March 28, 2006, with the incorrect Received Date. The correct version was posted on April 12, 2006.

References

(1) Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte-Carlo free energy estimation - umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.

(2) Elber, R. a chapter 2, Reaction path studies of biomolecules. In *Recent developments in theoretical studies of proteins*; World Scientific: Singapore, 1996.

(3) Ulitsky, A.; Elber, R. A new technique to calculate steepest descent paths in flexible polyatomic systems. *J. Chem. Phys.* **1990**, *92*, 1510−1511.

(4) Jonsson, H.; Mills, G.; Jacobson, K. W. Nudge Elastic Band Method for Finding Minimum Energy Paths of Transitions. In *Classical and Quantum Dynamics in Condensed Phase Simulations;* Berne, B. J.. Ciccotti, G.. Coker, D. F., Eds.; World Scientific: Singapore, 1998; p 385.

(5) Weinan, E.; Weiqing, R.; Vanden-Eijnden R. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*, Art. No. 52301, 1−4.

(6) Ren, W.; Vanden-Eijnden, E.; Maragakis, P.; Weinen, E. Transition pathways in complex systems: Applications of the finite-temperature string method to the alanine dipeptide. *J. Chem. Phys.* **2005**, *123*, Art No. 134109, 1−12.

(7) Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the density of states by history-dependent metadynamics. *Phys. Rev. Lett.* **2004**, *92*, Art No. 170601, 1−4.

(8) Noe, F.; Ille, F.; Smith, J. C.; Fischer, S. Automated computation of low energy pathways for complex rearrangements in proteins: Application to the conformational switch of ras p21. *Proteins: Struct., Funct., Bioinformatics* **2005**, *59*, 534−544.

(9) Olender, R.; Elber, R. Calculation of classical trajectories with a very large time step: Formalism and numerical examples. *J. Chem. Phys.* **1996**, *105*, 9299−9315.

(10) Elber, R.; Ghosh, A.; Cardenas, A. Long time dynamics of complex systems. *Acc. Chem. Res.* **2002**, *35*, 396−403.

(11) Cardenas, A. E.; Elber, R. The kinetics of cytochrome C folding: Atomisitically detailed simulation. *Proteins: Struct., Funct., Genet.* **2003**, *51*, 254−257, 2003.

(12) Cardenas, A. E.; Elber, R. Atomically detailed simulations of helix formation with the stochastic difference equation. *Biophys. J.* **2003**, *85*, 2919−2939.

(13) Dellago, C.; Bolhuis, P. G.; Geissler, P. L. Transition Path Sampling. *Adv. Chem. Phys.* **2002**, *123*, 1−78.

(14) Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S. How well can simulation predict protein folding kinetics and thermodynamics. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 43−69. Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78*, 3908−3911.

(15) Landau, L. D.; Lifshitz, E. M. *Mechanics*; Butterworth-Heinenann: Oxford, 1993; Chapter 1.

(16) Gillilan, R. E.; Wilson, K. R. Shadowing, rare events and rubber bands−a variational Verlet algorithm for molecular dynamics. *J. Chem. Phys.* **1992**, *97*, 1757−1772.

(17) Cho, A. E.; Doll, J. D.; Freeman, D. L. The construction of double ended classical trajectories. *Chem. Phys. Lett.* **1994**, *229*, 218−224.

(18) Machlup, S.; Onsager, L. Fluctuations and irreversible processes 2. Systems with kinetic energy. *Phys. Rev.* **1953**, *91*, 1512−1515.

(19) Passerone, D.; Parrinello, M. Action-derived molecular dynamics in the study of rare events. *Phys. Rev. Lett.* **2001**, *87*, art. No. 108302, 1−4.

(20) Kaczmarz, S. Approximate solution of systems of linear equations. *Bull. Acad. Polonaise Sci. Lett. A* **1937**, *5*, 355−377.

(21) Golub, G. H.; Van Loan, C. F. *Matrix Computations;* John Hopkins University Press: 1996.

(22) Mueller, K. Reaction paths on multidimensional energy hypersurfaces. *Angew. Chem.* **1980**, *19*, 1−13.

(23) Verlet, L. Computer experiments on classical fluids. I. Thermodynamical properties of Lennard Jones molecules. *Phys. Rev.* **1967**, *159*, 98−103.

(24) Doye, J. P. K.; Wales, J. D. Thermodynamics of global optimization. *Phys. Rev. Lett.* **1998**, *80*, 1357−1360.

(25) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. Assessing the efficiency of free energy calculation methods. *J. Chem. Phys.* **2004**, *120*, 3563−3578.

CT060028M