Transition pathways in a many-body system: Application to hydrogen-bond breaking in water

Félix S. Csajka and David Chandler

Department of Chemistry, University of California at Berkeley, Berkeley, California 94720

(Received 18 February 1998; accepted 10 April 1998)

We apply a stochastic method introduced by Dellago et al. [J. Chem. Phys. 108, 1964 (1998)] to sample transition paths in high-dimensional systems. The method connects two endpoint regions (for example a reactant and a product region) by a set of space-time paths. This approach is an importance sampling for rare events that does not require prior knowledge of the location of dynamical bottlenecks. Transition paths are generated with a weight corresponding to a chain of Metropolis Monte Carlo steps. We derive Monte Carlo algorithms and apply the technique to the dynamics of hydrogen-bond breaking in liquid water. We obtain averages in a transition path ensemble for the structure and energy along the trajectory. While characterized by a rate constant, hydrogen-bond breaking in water occurs frequently enough to be studied by standard methods. The process therefore provides a useful test of path sampling methods. The comparison between path sampling and standard Monte Carlo demonstrate the feasibility of transition path sampling for a many-body system with a rough potential energy surface. © 1998 American Institute of Physics.

I. INTRODUCTION

This paper demonstrates the feasibility of applying transition path sampling to a dynamical process in a many-body system. The transition path method samples a chain of states, constrained to begin in a reactant region and to end in a product region. As an illustration, we study the breaking of hydrogen bonds (H-bonds) in water at ambient conditions. In addition to demonstrating feasibility, we also illustrate methods for extracting information from the ensembles of transition pathways.

Numerical studies of rare events generally rely on an identification of the transition state. The number of transition states grows exponentially with the dimensionality of the system. As such, an exhaustive search for all transition states is impossible for systems of current interest in chemistry and biology (such as solvation dynamics, conformational transitions in large molecules or transport processes in condensed matter systems). An enumeration of transition states is also irrelevant because there will be many distinct trajectories traversing an energy barrier which is rough on the energy scale of $k_B T$ ($k_B$ is the Boltzmann constant, $T$ is the temperature). These facts indicate a need for stochastic methods which sample rare events, creating an ensemble of transition pathways. Figure 1 illustrates the pertinent issues for a high-dimensional system: there is a qualitative difference between a low-dimensional system with distinct saddle points, and a high-dimensional system with a rugged potential energy barrier.

Traditionally, transition states have been located by searching for saddle points on the potential energy landscape. As Fig. 1 shows, however, this approach is limited to systems with few degrees of freedom. The alternative approach we adopt follows Pratt’s suggestion to use a Markov chain of states between endpoint regions. The chain of states so formed is isomorphic to a polymer in configuration space with endpoint constraints. The specific Markov chain we generate corresponds to the Metropolis Monte Carlo algorithm. Other stochastic or deterministic pathways are possible too. The stochastic nature of dynamical algorithms, or the statistical characterization of initial and final conditions, lead to transition path ensembles that can be sampled by techniques isomorphic to equilibrium statistical mechanics. A similar idea has been used by Zimmer to find the dynamical critical exponent of the two-dimensional Ising model. Others have exploited the chain of states perspective, but in ways limited to low dimensional systems or in ways that only approximate the path ensemble of a concrete dynamical algorithm.

In illustrating the feasibility of true transition path sampling, we examine the process where a tagged pair of water molecules in liquid water evolves from a hydrogen-bonded state to a state with a broken hydrogen bond. In the un-bonded final state, the tagged waters are still nearest neighbors. With this requirement, we focus our attention on the hydrogen bond breaking events within the first solvation shell, and do not consider subsequent diffusion that also affects H-bond kinetics. We look at the patterns of hydrogen-bond breaking, studying the evolution of structural and energy averages along the reaction path. The breaking of a hydrogen bond in water is not strictly a rare event since there is an extensive number of bonds which break and form on a time scale of 1 ps. Their abundance allows a direct comparison between transition path sampling and conventional simu-
lation methods and a consistency check for the transition path method. When the reactive event is truly rare, a comparison is not possible, and some form of transition path sampling provides the only computational way to study the process.

The paper is organized as follows: In Sec. II we discuss the transition path action and Monte Carlo algorithms. Section III gives details on our Monte Carlo Simulations, and Sec. IV presents results on hydrogen-bond breaking in water. We conclude in Sec. V.

II. ACTION AND ALGORITHMS FOR TRANSITION PATHS

A. The Metropolis path action

A path in space-time is given by an ordered sequence of \( L+1 \) copies of phase space, \( \{ r_0 \rightarrow r_1 \rightarrow \cdots \rightarrow r_L \} \). \( r_0 \) denotes a point in configuration space, and \( \tau = 0, \ldots, L \) labels successive steps on the path. The transition path action therefore depends on \( D \times (L+1) \) coordinates. The simplest choice is then a product of Markovian transition probabilities\(^1\)

\[
\exp(-S[r_0, r_1, \ldots, r_L]) = e^{-\beta E(r_0)} h_A(r_0) \left( \prod_{\tau=0}^{L-1} p(r_\tau \rightarrow r_{\tau+1}) \right) h_B(r_L). \tag{1}\]

Here, \( \beta = 1/k_B T \) and \( E(r_\tau) \) is the potential energy of configuration \( r_\tau \). \( h_A(r_0) \) and \( h_B(r_L) \) are the initial and final boundary constraints, respectively, forcing the path to start in region \( A \) (the reactant region) and evolve to region \( B \) (the product region) in \( L \) steps.

In this paper, we restrict ourselves to the transition probability \( p(r \rightarrow r') \) for a Markov process generated by the Metropolis Monte Carlo algorithm, which does not depend on the momenta. In that case,

\[
p(r \rightarrow r') = \omega(r \rightarrow r') + \delta(r \rightarrow r') Q(r),
\]

\[
\omega(r \rightarrow r') = \eta(r, r') \min\{1, \exp[-\beta E(r') + \beta E(r)]\},
\]

\[
Q(r) = 1 - \int dr'' \omega(r \rightarrow r'').
\]  \tag{2}

\( \omega(r \rightarrow r') \) gives the transition probability for an accepted new element \( r' \) in the Markov chain; \( Q(r) \) is the probability of rejecting a trial move at \( r \). The delta distribution appears because there is a finite rejection probability in the probability density \( p(r \rightarrow r') \). \( \eta(r, r') \) is a symmetric function which decays rapidly for growing \( |r - r'| \), e.g., a Gaussian or a characteristic function with support in a small interval \( \Delta r \) around \( r \) (the standard choice in Metropolis Monte Carlo). The \( \min \)-function returns the smaller of its arguments.

The transition probability (2) corresponds to the well-known Metropolis Monte Carlo rule with generating probability \( \eta(r, r') \) and acceptance probability \( \min\{1, \exp[-\beta E(r') + \beta E(r)]\}.\) We note that \( p(r \rightarrow r') \) is normalized by construction and satisfies detailed balance.

B. Configurational bias sampling

The configurational bias Monte Carlo (CBMC) algorithm samples equilibrium conformations of a polymer by regrowing the entire chain in a biased fashion.\(^{21-23}\) Since a chain generated at random is very likely to overlap with itself and/or with its neighbors, each segment is regrown with a bias proportional to its Boltzmann weight. There is a correction for the biased growth, known as the “Rosenbluth weight,” \( W \). The weight enters the acceptance probability in the CBMC algorithm. It is defined as the ratio \( W_a = P_a/e^{-S_a} \), where \( P_a \) is the total CBMC generating probability of path \( a \), and \( S_a \) is path \( a \)'s action. It follows from the detailed balance condition that the acceptance probability for a trial \( a \rightarrow b \) is \( P_{ab} e^{-S_b} = \min\{1, W_a/W_b\} \).

To enhance the chance that the transition path satisfies the final endpoint constraint \( h_B(r_L) \), we use a guiding field \( \phi_x(r_x) \) to bias the trial moves.\(^1,24\) Guided CBMC algorithms use the identity

\[
\prod_{\tau=0}^{L-1} p(r_\tau \rightarrow r_{\tau+1}) = \prod_{\tau=0}^{L-1} \left\{ p(r_\tau \rightarrow r_{\tau+1}) \times \exp[\phi_x(r) - \phi_{x+1}(r_{\tau+1})]\right\}. \tag{3}\]

This identity holds if the guiding field \( \phi_0 = 0 \) and \( \phi_L = 0 \). Using such a field, the generating probability is

\[
P_{\text{gen}}(r_{\tau-1} \rightarrow r_\tau) = \bar{n}(r_{\tau-1}, r_\tau) \hat{\omega}(r-1), \tag{4}\]

where

\[
\bar{n}(r_{\tau-1}, r_\tau) = \left[ \min\{1, \exp[-\beta E(r_\tau) + \beta E(r_{\tau-1})]\} + \delta(r_{\tau-1} - r_\tau) Q^{\text{est}}(r_{\tau-1}) \right] \exp[\phi_x(r)] - \phi_{x+1}(r_{\tau+1}) \tag{5}\]

with

\[
Q^{\text{est}}(r) = 1 - \frac{1}{k} \sum_{i=1}^{k} \min\{1, \exp[-\beta E(r_i^\tau) + \beta E(r)]\} \tag{6}\]

denoting \( Q(r) \) as estimated with a finite number, \( k \), of configurations. These configurations are chosen at random from the distribution \( \eta(r, r^\tau) \). The function \( \hat{\omega}(r_{\tau-1}) \) is the partial Rosenbluth weight obtained by summing
The generating probability is $P_n(r_{\tau-1}, r_{\tau}^{(i)})$ over those same, $k$, configurations, $r_{\tau}^{(i)}$, $i = 1, 2, \ldots , k$. The full Rosenbluth weight $\tilde{W}$ is then
\[
\tilde{W} = e^{-\beta E(r_0)} \prod_{\tau=0}^{L-1} \tilde{w}(r_{\tau}).
\] (7)

Finally, the overall acceptance probability is $P_{\text{acc}}(a \rightarrow b) = \min[1, \tilde{W}_b / \tilde{W}_a] h_A(r_0) h_B(r'_L)$. Obviously, all paths begin in region $A$ and are accepted only if they reach region $B$.

The choice for the guiding field strongly affects the rate of convergence of the simulation. In this work, we use a field which pulls the transition path towards the product region; see Eq. (20) below.

C. Reptation and shooting algorithms\(^\text{12}\)

The reptation algorithm removes time slices from one endpoint and adds them to the other end of the path. The shooting algorithm grows a segment of the path by starting at intermediate time slice $T$ and attempting to reach an endpoint region. The algorithms are especially useful for treating paths associated with deterministic dynamics.\(^\text{12(b)}\)

Both algorithms generate a chain of trial moves with a generating probability identical to the transition probability, which simplifies matters considerably. For a chain of forwards steps, the total generating probability is identical to the product of transition probabilities: any trial chain which reaches region $B$ is accepted. For both algorithms, it is necessary to define a time-reversed transition probability $\tilde{p}(r_{\tau+1} \rightarrow r_{\tau})$ in order to perform "backwards shots" towards the region $A$ or add time slices in region $A$ for reptation moves. For Metropolis paths, the operation of time reversal is trivial:
\[
\tilde{p}(r_{\tau+1} \rightarrow r_{\tau}) = \omega(r_{\tau+1} \rightarrow r_{\tau}) + \delta(r_{\tau} - r_{\tau+1})Q(r_{\tau}).
\] (8)

The acceptance probability for a chain of backwards moves is $P_{\text{acc}}(a \rightarrow b) = \min[1, W_b / W_a]$ with $W_a = e^{-S}/P_{\text{gen}}(a)$. The generating probability is $P_{\text{gen}} = \prod_{\tau=0}^{L-1} \tilde{p}(r_{\tau+1} \rightarrow r_{\tau})$. The ratio $e^{-S}/P_{\text{gen}}$ may be written as a product of factors of the form
\[
\frac{\omega(r_{\tau+1} \rightarrow r_{\tau})}{\omega(r_{\tau+1} \rightarrow r_{\tau})} = \frac{\eta(r_{\tau}, r_{\tau+1})}{\eta(r_{\tau}, r_{\tau+1})} \min[1, e^{-\beta \Delta E}] = e^{-\beta \Delta E},
\] (9)

where $\Delta E = E(r_{\tau+1}) - E(r_{\tau})$. Equation (9) also holds in case of a rejection, where $\Delta E = 0$, and it is assumed that $\eta(r, r')$ is symmetric. Then the ratio $e^{-S}/P_{\text{gen}}$ becomes
\[
e^{-S}/P_{\text{gen}} = \prod_{\tau=0}^{L-1} e^{-\beta [E(r_{\tau+1}) - E(r_{\tau})]} = e^{-\beta E(r_0)} = e^{-\beta E(r_T)},
\] (10)

but this ratio is identical for the old and the trial path, since both paths share the same initial time slice $r_T$. This result is a consequence of the fact that $p(r_{\tau} \rightarrow r_{\tau+1})$ satisfies detailed balance.

Summarizing, the acceptance rules for the shooting algorithm of Metropolis paths are
\[
P_{\text{acc}}(r_0, \ldots , r_T, r_{T+1}, \ldots , r_L) = h_A(r_0), \quad \text{shooting towards } A;
\] (11)
\[
P_{\text{acc}}(r_0, \ldots , r_T, r_{T-1}, r_T, \ldots , r_L) = h_B(r_L'), \quad \text{shooting towards } B;
\] (12)

In Eq. (11), the trial changes the $T$ final time steps $\{r_{T+1}, \ldots , r_L\}$; in Eq. (12), the first $T$ time steps are changed, i.e., $\{r_0, \ldots , r_T\}$.

For the reptation algorithm, the acceptance rules for Metropolis paths are
\[
P_{\text{acc}}(r_T, \ldots , r_L', r_{L-T}, \ldots , r_0) = h_A(r_0) h_B(r_L'), \quad \text{reptating } T \text{ beads from } A \text{ to } B,
\] (13)
\[
P_{\text{acc}}(r_0^{\prime}, \ldots , r_L'^{-1}, r_0, \ldots , r_T) = h_A(r_0) h_B(r_T), \quad \text{reptating } T \text{ beads from } B \text{ to } A.
\] (14)

In Eq. (13), the trial moves consists of deleting beads $\{r_0, \ldots , r_T\}$ and adding $T$ beads after $r_L$. In Eq. (14), the trial move deletes the beads $\{r_T, \ldots , r_L\}$ and adds them before $r_0$. Note that the time labels have to be adjusted after each accepted trial.

D. Metropolis paths for many-body systems

We construct the Markov chain of a many-particle system by defining the transition path as a product of Metropolis Monte Carlo single-particle moves. Neighboring elements of the Markov chain differ by the position of a single particle (accepted MC step), or they are identical (rejected MC step). The action becomes a product over $L$ passes: within a pass, we visit each degree of freedom once:
\[
\exp(-S[r_0, r_1, \ldots , r_L])
\] 
\[= e^{-\beta E(r_0)} h_A(r_0)
\] 
\[\times \prod_{\tau=0}^{L-1} \prod_{m=1}^{N} p(r_{p(m), \tau} \rightarrow r_{p(m), \tau+1}) h_B(r_L).
\] (15)

$r_{n, \tau}$ labels the $n$th particle at pass $\tau$. $N$ is the number of particles, and $P(m)$ makes a random permutation of the list $\{1, 2, \ldots , N\}$ such that each element occurs once, i.e., the lists $\{1, 2, \ldots , N\}$ and $\{P(1), P(2), \ldots , P(N)\}$ contain the same $N$ elements.

The action (15) is suited to the study of solvation problems. We add boundary constraints to the solute, but evolve the solvent particles without constraints. Since the path action without constraints is just the probability for generating a path by direct Monte Carlo, we sample the constrained degrees of freedom (the solute) by the guided CBMC method and the unconstrained solvent particles by direct Monte Carlo. We regrow the path globally and accept or reject a trial path according to its Rosenbluth weight. There are no special memory requirements for this scheme. One only needs to save the initial configuration $r_0$ and the Rosenbluth weight from the previous path.
III. SIMULATIONS

A. Implementation of Transition Path Sampling

All simulations were done in the NVT-ensemble with $N=64$ water molecules at the density $\rho=1.000$ g/cm$^3$, $T=298$ K and $N=64$ water molecules. This number of water molecules is sufficient for our purpose, since we show below that hydrogen bond breaking involves only local changes in the structure. We assume water molecules have rigid intramolecular structure, and use the single point charge (SPC) intermolecular potential.\textsuperscript{25} The model gives a reasonable description of hydrogen bonding despite its simplicity.\textsuperscript{26} We use periodic boundary conditions and treat the Coulombic long-range interactions with Ewald sums.\textsuperscript{27} Long-range interactions with Ewald sums.\textsuperscript{27} The intermolecular potential.\textsuperscript{25} The model gives a reasonable description of hydrogen bonding despite its simplicity.\textsuperscript{26} We use periodic boundary conditions and treat the Coulombic long-range interactions with Ewald sums.\textsuperscript{27}

A step time corresponds to 64 single-particle moves, such that each degree of freedom participates in exactly one trial move. As noted after Eq. (15), the order with which the degree of freedom is considered within a time step is random.\textsuperscript{28} The Monte Carlo moves are performed as follows: a water molecule is translated by drawing a trial displacement $\Delta R$ from a Gaussian distribution, $\propto \exp(-\Delta R^2/2\sigma_R^2)$ for each Cartesian coordinate. For the rotations, we randomly choose an orientation for the rotation axis from the unit sphere; then we rotate the water by an amount $\Delta \Phi$ drawn from a Gaussian distribution, $\propto \exp(-\Delta \Phi^2/2\sigma_\Phi^2)$. A trial move always consists of a rotation and a translation. The variances were chosen by running a standard MC simulation with trial moves drawn from a Gaussian distribution, and requiring that the acceptance probability is about 35%. This procedure gives the variances $\sigma_R=0.09$ Å and $\sigma_\Phi=12.6^\circ$.

Transition paths are generated with three different methods: (1) standard Metropolis Monte Carlo with boundary constraints $h_A(r_0)$ and $h_B(r_L)$ (paths which do not satisfy these constraints are discarded); (2) configurational bias Monte Carlo with a guiding field; (3) a combination of shooting and reptation algorithms.

To estimate the time interval spanned by the path simulations, we calculate the mean square displacement for both translations and rotations. For translations, we use the Einstein relation $\langle |\mathbf{R}(\tau) - \mathbf{R}(0)|^2 \rangle = 6D\tau$, where $\mathbf{R}(\tau)$ is the oxygen position of molecule $i$ at time step $\tau$. Using the self-diffusion constant for SPC water, $D=0.36$ Å$^2$/ps,\textsuperscript{28} we find a time step of $0.6 \pm 0.2$ fs with $\sigma_R=12.6^\circ$ and $\sigma_\Phi=0.09$ Å.

For rotations, we study rotational diffusion using $C_i(\tau) = \langle P_i | \mathbf{e}_i(\tau) \cdot \mathbf{e}_i(0) \rangle$, where $P_i(x)$ is the order $l$ Legendre polynomial of $x$, and $\mathbf{e}_i(\tau)$ is the unit vector along the dipole moment of the water molecule $i$ at time step $\tau$. For $l=1$, we find that $C_i(\tau)$ decays roughly linearly due to the stochastic sampling of the rotations.\textsuperscript{30} This is in contrast to Newtonian molecular dynamics simulations, where one sees an initial period of libration ($\approx 20$ fs) followed by a roughly exponential decay.\textsuperscript{31,32} Using the dipole relaxation time $\tau_r \approx 3$ ps for SPC water,\textsuperscript{29} one finds that a time step in the path corresponds to $1.0 \pm 0.3$ fs (same $\sigma_R$ and $\sigma_\Phi$ as above).

B. Averages in the transition path ensemble

We collect averages for the observable $O$ weighted by the path action:

\begin{align}
\langle O(\tau) \rangle_{AB} &= \frac{1}{Z} \int \left( \prod_{\tau' = 0}^{L} dr_{\tau'} \right) O(r_{\tau}) \nonumber \\
& \quad \times \exp(-S[r_0, \ldots, r_\tau, \ldots, r_L]);
\end{align}

\begin{align}
Z &= \int \left( \prod_{\tau' = 0}^{L} dr_{\tau'} \right) \exp(-S[r_0, \ldots, r_L]).
\end{align}

Note, the definition of the action, Eq. (15), includes the boundary constraints on the transition path. The path average $\langle O(\tau) \rangle_{AB}$ is a time profile of observable $O$ while the system evolves from $A$ to $B$. Since we study an ensemble of paths, we must address the fact that different realizations of the trajectories, even if rather similar in shape, can reach the transition state region at different time steps. To take meaningful averages, we must match up the paths. Matching can be done with a time correlation function of the form

\begin{equation}
O(\tau) = \sum_{\tau' = -\tau}^{L-\tau} \langle \hat{h}(\tau')O(\tau' + \tau) \rangle_{AB}.
\end{equation}

Here, a blip-function $\hat{h}(\tau)$ marks the instant the system passes through a dividing surface that lies somewhere between the boundaries $A$ and $B$:

\begin{equation}
\hat{h}(\tau) = \begin{cases} 
1, & \text{if } r_x \text{ is on the product side} \\
1, & \text{of the dividing surface for all } \tau' > \tau; \\
0, & \text{product side or if } \tau < 0 \text{ or } \tau > L.
\end{cases}
\end{equation}

The blip-function $\hat{h}(\tau)$ is therefore $+1$ if a path crosses from the reactant to the product side of the surface and zero for noncrossing paths. The time correlation of $\hat{h}(\tau)$ with $O(\tau')$ is most useful if the paths crossing the dividing surface do so without many recrossings. With such a surface, the dispersion of matched paths can be used to assess whether or not there is one dominant reaction channel in the transition dynamics.

C. Boundary constraints for hydrogen bond breaking

We applied boundary constraints to the first and last time step to sample hydrogen bond breaking in liquid water. We took an equilibrium configuration and arbitrarily put a “tag” on two hydrogen bonded waters. As with any dynamical process going from one metastable state to another, we identified appropriate boundary constraints from equilibrium statistics. In particular, during an equilibrium simulation of water, we collected histograms of the hydrogen bond angles and of the pair energies between waters in the first solvation shell (\(R_D \leq 3.5\) Å). Both the hydrogen bond energy (the pair energy of a hydrogen bonded dimer) and the hydrogen bond angle provide reasonable boundary constraints. In this work, we used a set of geometric criteria, setting the hydrogen bond angle $\alpha$ (see Fig. 2) to less than 10 degrees initially and to more than 40 degrees finally. The matching blip function was chosen after comparing the number of recrossings for different choices of $\alpha$: 

\begin{align}
\langle O(\tau) \rangle_{AB} &= \frac{1}{Z} \int \left( \prod_{\tau' = 0}^{L} dr_{\tau'} \right) O(r_{\tau}) \nonumber \\
& \quad \times \exp(-S[r_0, \ldots, r_\tau, \ldots, r_L]);
\end{align}

\begin{align}
Z &= \int \left( \prod_{\tau' = 0}^{L} dr_{\tau'} \right) \exp(-S[r_0, \ldots, r_L]).
\end{align}
The time derivative of the conditional probability $P_{A \rightarrow B}$ is

$$h_A(r_0) = \begin{cases} 1, & \text{if } R_{0,0} < 3.5 \text{ Å}, \alpha < 10 \text{ degrees;} \\ 0, & \text{otherwise.} \end{cases}$$

$$h(r_s) = \begin{cases} 1, & \text{if } \alpha > 25 \text{ degrees;} \\ 0, & \text{otherwise.} \end{cases}$$

$$h_B(r_L) = \begin{cases} 1, & \text{if } R_{0,0} < 3.5 \text{Å}, \alpha > 40 \text{ degrees;} \\ 0, & \text{otherwise.} \end{cases}$$

The initial trajectory has $L+1$ copies of the system which are identical except for the H-donor molecule, which is rotated in small increments to satisfy the final time step constraint.

The water transition paths we have studied are 25 time steps long. This number was chosen after studying the probability of reaching region $B$ from region $A$ in $\tau$ steps, $P_{A \rightarrow B}(\tau) = \langle h_A(0)h_B(\tau)\rangle/\langle h_A \rangle$ (Fig. 3, upper part). One finds a linear increase with $\tau$ after a transient time. The reactive flux, $k(t) = \langle h_A(0)h_B(\tau)\rangle/\langle h_A \rangle$ (Fig. 3, bottom part), is the time derivative of the conditional probability $P_{A \rightarrow B}(\tau)$. After $P_{A \rightarrow B}(\tau)$ enters the linear regime, $k(t)$ is reasonably constant. The transition probability for $\tau = 25$ is $(7.7 \pm 0.5) \times 10^{-3}$, where its slope (or rate constant) is $(5 \pm 1) \times 10^{-4}$ per time step. This slope is consistent with the known rate constant for H-bond breaking $k_{A \rightarrow B} = 0.7 \text{ ps}^{-1}$ (Ref. 20) if a time step along the path is $0.7 \pm 0.2$ fs.

D. Path sampling methods

1. Direct Monte Carlo runs

A straightforward Monte Carlo simulation has been used to check the consistency of the transition path simulations. We have used the same Gaussian variances to generate trial moves. In addition, we labeled the water pairs in the system which satisfy the initial endpoint constraint $h_A(0)$. After 25 Monte Carlo passes, we screened the labeled pairs for broken hydrogen bonds, coinciding with the constraint $h_B(L)$, with $L = 24$. Water pairs which satisfy both endpoint constraints are included into the path averaging. Path averages are evaluated with the time-correlation approach, Eq. (16).

In a standard Monte Carlo run with $(L+1) \times 10^3$ passes we found 1837 H-bond breaking events. The efficiency of the path sampling algorithms was estimated via the number of statistically independent transition paths per unit time. For stretches of trajectory $L = 24$ passes long, there was an average of 0.18 successful paths. At any given point on the trajectory, there were typically 24 H-bonds satisfying the constraint $h_A(0)$. In other words, there were 0.0077 successful paths per H-bond.

2. Configurational bias runs

Here, the tagged water pair was sampled by the guided CBMC algorithm, while the untagged waters were sampled by standard MC. During the construction of a transition path, we made 25 trial moves on both the tagged dimer and the untagged waters.

The validity of the CBMC algorithm depends on an accurate estimation of the rejection factor $Q^{\text{est}}(r)$. We have studied the convergence of $Q^{\text{est}}(r)$ with increasing upper summation limit $k$, as shown in Fig. 4. The sum in Eq. (3) can be conveniently sampled by Monte Carlo integration. One generates trials from a Gaussian distribution and samples the min-function only. This approach is equivalent to von Neumann rejection. The rate of convergence depends on the variance for the trial translations, $\sigma_T$, and rotations, $\sigma_R$; for the set of values $\sigma_T = 0.09 \text{ Å}$, $\sigma_R = 12.6^\circ$, we find a summation limit $k \approx 200$ ensures reasonable accuracy.
This accuracy is borne out in the study of structural averages, which do not change significantly for $k$ larger than that value.

For the guiding field, we choose the function

$$\exp[-\phi(x)] = \begin{cases} 1, & \text{if } \tau = 0, L \text{ or } \alpha \geq \alpha_{\text{max}}; \\ 2\exp\left[(\alpha - \alpha_{\text{max}})/10\right], & \text{otherwise}; \end{cases}$$

where $\alpha_{\text{max}} = 40^\circ$ was chosen. This choice gives a reasonable acceptance rate for the H-bond breaking process.

To run a CBMC transition path simulation, we first constructed an initial transition path by keeping all the water molecules fixed except the tagged H-donor water, which we rotated in small increments to satisfy the final endpoint constraint. After every CBMC trial, we sampled the initial configuration by standard MC (3200 trial steps) to avoid statistical correlations between trial paths. The acceptance rate for CBMC trial paths was 12%. Thus, the algorithms generated statistically independent paths after roughly 10 CBMC trials.

After the equilibration, we performed a production run, saving independent CBMC trial paths. A total of 1052 independent paths were assembled from roughly 30 000 CBMC trials.

3. Runs with the shooting and reptation algorithms

We repeated the path sampling simulations by combining the shooting and reptation algorithms. The initial path for the equilibrium run was taken from the CBMC equilibration run. During the production runs we chose a time step at random and used it as a starting point for a shooting or a reptation move. Reptation moves which violate the first endpoint constraint in Eqs. (13) and (14) were immediately discarded.

We generated 1500 paths with the combined shooting and reptation scheme. A path was saved after 50 shooting trials and 50 reptation trials. The acceptance probability depends strongly on the location of the time slice from which the trial is initiated; this fact is illustrated in Fig. 5 for shooting trials (left part) and reptation trials (right part). 1500 paths were produced with a correlation length of 3 to 4 paths. In other words, about 400 independent paths are generated from 15 000 shooting and reptation Monte Carlo steps.

4. Comparison

Our estimates of efficiency for the two path sampling algorithms give similar numbers. In CBMC, the computation of $Q(r)$ is expensive. On the other hand, the bias enhances the likelihood of finding a successful path. To decorrelate the paths, one samples the initial time slice, which is cheaper than performing path moves. Long transition paths and large energy barriers demand a careful design of the CBMC guiding field function. It is probably easiest to use the shooting and reptation algorithms to sample long transition paths.

IV. RESULTS

This section discusses the simulation results for hydrogen bond breaking pathways in water, obtained by transition path sampling. All the averages are collected using the time-correlation method discussed in Eq. (16). For example, the evolution of the hydrogen bond angle in Fig. 6 is

$$\tilde{\alpha}(\tau) = \sum_{\tau'} - (\tilde{h}(\tau')\alpha(\tau' + \tau))_{\text{ab}}.$$  

As a consequence of the time-correlation average, time labels can run from $-24$ to $+24$. 

FIG. 5. Acceptance probability for trial moves with the shooting algorithm (left side) and the reptation algorithm (right side).

FIG. 6. Hydrogen bond angle $\alpha$ in the path ensemble: direct Monte Carlo simulation (dotted line), CBMC transition path simulation (solid line) and a combination of shooting and reptation sampling (dashed line). The path averages are aligned with a marker, $\tilde{h}(0)$, as explained in the text.

FIG. 7. Oxygen-oxygen distance (upper plots) and the oxygen-H-bond hydrogen (lower plots) in the path ensemble. Direct MC sampling (dotted lines), CBMC path sampling (solid lines) and shooting and reptation sampling (dashed lines).
and the number of data points is largest around time step 0 where the system crosses the dividing surface.

The average evolution of the hydrogen bond angle, the oxygen-oxygen separation, and of oxygen-donor hydrogen separation are shown in Figs. 6 and 7. The meanings of these coordinates are illustrated in Fig. 2. The averaging of their evolution is carried out using three different methods for sampling the transition path ensemble. The three methods give essentially the same results, as they should if each method is correct.

The figures show that the hydrogen bond angle increases monotonically on the time scale of the simulation, which is about 50 fs. This behavior is consistent with the experimental time scale for angular motion, which is roughly 60 fs for a librational period.\(^3_4\) In that same time stretch, the average O-O separation increases monotonically from 3.0 to 3.2 Å and the average O-H separation grows from 2.0 to 2.6 Å in accordance with the angular motion. The most rapid pace of the averaged evolution occurs between \(t = 5_0\) and \(t = 5_1\). This behavior is a consequence of the alignment, Eq. \(\sim 17\), which requires all paths to cross \(\alpha = 25^\circ\) in that interval.

The hydrogen bond angle increases, and the hydrogen bond duly breaks. Indeed, the average evolution of the potential energy between the initially hydrogen bonded pair, \(U_{\text{HB}}\), rises from a value near \(-15\) kJ/mol at \(\tau = -20\) to roughly \(-5\) kJ/mol at \(\tau = 20\). This rise is accompanied by a fall in solvation energy, as shown in Fig. 8. In particular, we graph the path averaged behavior of

\[ U^{(i)} = \sum_{j \neq i} U(j,i), \]

the potential energy of water molecule \(i\) (\(i\) refers to either the donor, D or acceptor, A) less that between the initially hydrogen bonded pair. In its arranged evolution, graphed in Fig. 9, the average \(U^{(A)} + U^{(D)}\) decreases by approximately 10 kJ/mol. The statistical uncertainties in these averages is between 1 to 2.5 kJ/mol. The smallest uncertainties are near \(\tau = 0\), the time slice with the greatest number of data points.

Hydrogen bond populations around the tagged waters are studied in Fig. 10. Again, comparison with standard sampling demonstrates the success of the path sampling algorithms. The equilibrium coordination number for SPC water at ambient conditions is about 3.5. The hydrogen bonds to the tagged waters show a drop from almost 3.8 to 2.9 for the acceptor and from 3.3 to 2.7 for the donor. The difference in coordination partners mirrors the evolution of the potential energy differences discussed above. The asymmetry in the evolution of the donor and the acceptor is due to our asymmetric criteria for initial and final states. The H-acceptor prefers a high initial coordination number and loses roughly one H-bond, whereas the H-donor has about the equilibrium number of hydrogen bonds initially. In about 40% of the trajectories, the H-bond donor immediately finds a new H-bonding partner. This observation agrees with Stillinger’s ‘‘switching of allegiance’’\(^3_5\) description of dynamics in liquid water.

To understand the structural role of the neighboring waters around the tagged pair, it is useful to represent the local water distribution in an averaged way. In particular, to assess the dynamical changes in the local structure during H-bond
breaking, we use time-dependent cylindrical distribution functions for oxygens of water molecules bonded to the tagged acceptor-donor pair. This function is determined at a given time slice $\tau$, by counting waters that satisfy the geometrical H-bonding rule ($\alpha_{HB}<30^\circ, R_{OO}<3.5$ Å) with one or the other of the two tagged waters. A cylinder axis is defined by the oxygens of the tagged waters; the midpoint between donor and acceptor is the origin of the cylindrical coordinates. We project out the component along the cylinder axis ($z$) and the distance from the axis ($r$) for waters within the first solvation shell of either the donor or the acceptor. Normalization then yields the distribution, $g(z,r;\tau)$. This distribution is considered in Figs. 11 and 12.

The figures show that the switching of allegiance is associated with a distortion of the local H-bond network. Nearest neighbors move from most favored tetrahedral arrangements to populate those configurations at the middle and ends of the pictured cylinders. After H-bond breaking, relaxation of the distorted network will lead to diffusion.\(^{20}\)

V. CONCLUSION

This paper shows that it is indeed practical to obtain meaningful results from transition path sampling performed on many-body systems. Specifically, we have shown that implementation with one or more algorithms leads to statistically reliable results. Further, the results can be interpreted through time correlation function analysis.

The transition path method finds trajectories between two given boundaries automatically. No prior assumptions about transition mechanisms or transition states is required. There is also no restriction to a single reactant and product configuration or a single transition pathway. The method allows arbitrary regions of configuration space to act as reactant and product basins, and from these regions, an ensemble of pathways emerges.

In this paper, we have focused our attention on the action functional consistent with Metropolis paths. The algorithms we have described can be applied to any problem where an energy function is available and where the dynamics is stochastic on the time scale of interest, including systems where the definition of a force is not straightforward, such as lattice Hamiltonians and quantum systems. The extension to larger system size is merely an issue of computer memory. Other classes of dynamics can be considered with closely related techniques. For example, transition path sampling can be performed in principle for systems obeying Langevin equations, either in the high-friction limit or with an inertial term,\(^{1}\) as well as deterministic dynamics.\(^{12}\)

ACKNOWLEDGMENTS

We are grateful to Peter Bolhuis and Christoph Dellago for sharing with us their ideas about path sampling, in general, and the shooting algorithm, in particular. We have also benefitted from conversations with Phillip Geissler, Ka Lum, Alenka Luzar, Jordi Marti and Lawrence Pratt. This work was supported in its initial stages by the National Science Foundation, and in its final stages by the Department of Energy through the Chemical Sciences Division of the Lawrence Berkeley National Laboratory.

In the conventional Metropolis algorithm, one usually picks a particle at random, which gives one trial per particle on average. The difference between this procedure and the one we employ herein is negligible in the thermodynamic limit.