On the efficiency of biased sampling of the multiple state path ensemble

Rogal, J.; Bolhuis, P.G.

Link to publication

On the efficiency of biased sampling of the multiple state path ensemble

Jutta Rogal[1,a)] and Peter G. Bolhuis[2]

[1]*Interdisciplinary Centre for Advanced Materials Simulation, Ruhr-Universität Bochum, Stiepeler Str. 129, 44801 Bochum, Germany*
[2]*Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*

Developed for complex systems undergoing rare events involving many (meta)stable states, the multiple state transition path sampling aims to sample from an extended path ensemble including all possible trajectories between any pair of (meta)stable states. The key issue for an efficient sampling of the path space in this extended ensemble is sufficient switching between different types of trajectories. When some transitions are much more likely than others the collective sampling of the different path types can become difficult. Here we introduce a Wang–Landau based biasing approach to improve the sampling. We find that the biasing of the multiple state path ensemble does not influence the switching behavior, but does improve the sampling and thus the quality of the individual path ensembles. © *2010 American Institute of Physics.* [doi:10.1063/1.3449144]

I. INTRODUCTION

Many physical processes in complex systems involve transitions between stable regions in phase space that occur on a time scale much longer than the molecular time scale due to the presence of large free energy barriers separating these stable regions. Examples include nucleation, protein conformational change, and chemical reactions. A sampling of such rare events employing regular all-atom molecular dynamics simulations with a time step of a few femtoseconds becomes quickly unpractical, as the system tends to spend most of the time in a stable state and only occasionally undergoes a transition to another stable state. Yet, these rare events dominate the long time behavior of the system. To address this rare event problem various computational techniques have been developed. Several approaches use biasing potentials along reaction coordinates as e.g., umbrella sampling,[1] blue moon sampling,[2] local elevation,[3] flooding,[4] hyperdynamics,[5] and metadynamics.[6] These methods have been quite successful but all require knowledge of a reaction coordinate along which the bias is being applied. A poor choice of such a reaction coordinate can lead to poor sampling of the transition. Transition path sampling (TPS)[7] provides a possibility to explore transitions between stable states in phase space without prior knowledge of a reaction coordinate, only a clear definition of the stable state regions is required. TPS has been successfully applied to a variety of problems, examples include crystal nucleation,[8] the folding of small proteins[9] (an overview of applications in biological systems is given in Ref. 10), and the transformation of nanoparticles under pressure.[11]

To obtain reliable results from TPS it is important that the path space is sampled sufficiently within a simulation. There are two main difficulties that can occur while sampling the path space: one is the problem of multiple reaction channels and the other one of intermediate (meta)stable states. Multiple reaction channels pose a problem since the TPS simulation can get stuck in one of the channels. The results can thus depend on the initial path and important mechanism that might even exhibit a lower barrier can be missed. A solution to this problem has been proposed by applying a replica exchange scheme to trajectories belonging to different interfaces within the transition interface sampling (TIS).[12–14] Several intermediate states are problematic since in the original formulation of TPS only trajectories connecting two stable states are included in the path ensemble. Trajectories ending up in a different state are rejected and the sampling can become very inefficient or even impossible due to a very low acceptance. We have addressed this problem by developing a multiple state TPS (MSTPS) approach.[15] This approach views the rare event process not as a two-state process, but as a Markovian state model (MSM) with multiple states.[16–18] Transitions between each pair of states are possible, and the system is assumed to lose its memory in between jumps. The MSTPS is effectively a method to sample all transitions in such a MSM. Within the approach all trajectories connecting any of the stable states in the system are included in the path ensemble. The advantage of the multiple state path sampling is that only one simulation is needed to sample the entire path ensemble instead of one simulation for each transition between each pair of stable states. The acceptance ratio is higher and the decorrelation of subsequent pathways is faster resulting in a more efficient sampling of the path space. The higher acceptance is due to the fact that all pathways connecting any two states in the system contribute to the ensemble. The faster decorrelation results from the switching between different types of pathways. The switching between different types of pathways is essential to efficiently sample the multiple state path ensemble. If the switching goes to zero the original two-state sampling is recovered. The MSTPS should be applied to systems where the switching actually poses a problem to the

a)Electronic mail: jutta.rogal@rub.de.

two-state sampling, i.e., systems with a low acceptance ratio due to trajectories that do not return to one of the two stable states. Examples of such processes include protein folding via intermediate states and cluster reorganization.

The trajectories in the multiple state ensemble are sampled with their correct weights, i.e., for transitions out of the same stable state the ratio of pathways is equal to the ratio of the transition probabilities. This can be a problem if the transition probabilities in the investigated system differ significantly. If, e.g., the transition probability for a $i \rightarrow j$ transition is a factor of 1000 larger than $i \rightarrow k$ then only one out of 1000 trajectories is a $i \rightarrow k$ trajectory. To address this problem and achieve an equal sampling of all types of pathways in the multiple state path ensemble we apply in this work a biasing scheme based on the Wang–Landau sampling approach.[19,20] We compare the efficiency of the Wang–Landau scheme with a simple biasing function and with the unbiased path sampling simulations.

The paper is organized as follows. In Sec. II a brief review of the MSTPS is given and the acceptance rule is extended to incorporate the corresponding biasing functions. In Sec. III the Wang–Landau biasing and the simple biasing scheme are applied to a model system. The efficiency is discussed with respect to the switching between different path types and the quality of the individual path ensembles. The results are summarized in Sec. IV.

II. THEORETICAL BACKGROUND

We briefly review the basic ideas of the multiple state path sampling (a detailed discussion can be found in Ref. 15) and introduce how the Wang–Landau biasing scheme and a simple biasing scheme can be applied to the sampling of the path ensemble.

A. Multiple state transition path sampling

In the MSTPS the original formulation of TPS[7] and TIS[12] is extended to include not only two distinct stable states A and B in the sampling but several stable states.[15] Here we focus on the multiple state version of TIS. We represent a dynamical trajectory as a discretized path $\mathbf{x}(L) = \{x_0, x_1, \ldots, x_L\}$ with a total length $L\Delta t$, where Δt is the time step between consecutive time slices $x \equiv \{r^N, p^N\}$ containing all N coordinates r and momenta p of the particles. The path ensemble consists of all dynamical trajectories connecting any two stable states in the system

$$\mathcal{P}_{\text{MSTIS}} = \sum_{i,j} \mathcal{P}_{ij}[\mathbf{x}(L)], \tag{1}$$

where

$$\mathcal{P}_{ij}[\mathbf{x}(L)] \equiv Z^{-1} \widetilde{h}_{ij}[\mathbf{x}(L)] \mathcal{P}[\mathbf{x}(L)] \tag{2}$$

is the path ensemble containing all trajectories connecting states i and j that cross the hypersurface (also called interface) defined by an order parameter λ as $\{x : \lambda(x) = \lambda_{mi}\}$ with $\lambda_{mi} \in \mathbb{R}$. The index m refers to the "outer" interface in the MSTIS framework,[15] and is from here on considered a fixed parameter for each state i. $\mathcal{P}[\mathbf{x}(L)]$ is the dynamical path probability, Z is a normalization factor defined as the integral

over all possible paths of all lengths, and $\widetilde{h}_{ij}[\mathbf{x}(L)]$ is a product of indicator functions

$$\widetilde{h}_{ij}[\mathbf{x}(L)] = h_i(x_0) h_j(x_L) \hat{h}_i^m[\mathbf{x}(L)] \prod_k \bar{h}_k[\mathbf{x}(L)]. \tag{3}$$

The indicator functions define the stable state regions and ensure that the path starts in a state i, $h_i(x_0)$, ends in a state j, $h_j(x_L)$, does not enter any other stable state k in between i and j, $\bar{h}_k[\mathbf{x}(L)]$, and crosses the hypersurface λ_{mi}, $\hat{h}_i^m[\mathbf{x}(L)]$. The Metropolis acceptance rule that ensures a proper sampling of the path ensemble in Eq. (1) with flexible path length is given by

$$P_{\text{acc}}[\mathbf{x}^{(\text{o})} \rightarrow \mathbf{x}^{(\text{n})}] = \sum_{ij} \widetilde{h}_{ij}[\mathbf{x}^{(\text{n})}(L^{(\text{n})})] \min\left[1, \frac{L^{(\text{o})}}{L^{(\text{n})}} \frac{\rho(x_{\text{sp}}^{(\text{n})})}{\rho(x_{\text{sp}'}^{(\text{o})})}\right], \tag{4}$$

where $\rho(x_{\text{sp}}^{(\text{n})})$ and $\rho(x_{\text{sp}'}^{(\text{o})})$ are the steady state distributions at the shooting points of the old (o) and new (n) path, respectively. Following the ideas of TIS the rate constant for a transition from a state i to a state j within the path ensemble can be expressed as

$$k_{ij} = \langle \phi_{mi} \rangle P_i(\lambda_{0j} | \lambda_{mi}), \tag{5}$$

where $\langle \phi_{mi} \rangle$ is the flux through the interface λ_{mi}. $P_i(\lambda_{0j} | \lambda_{mi})$ is the probability that a trajectory coming from state i and crossing the corresponding interface λ_{mi} reaches state j (i.e., crosses λ_{0j}) before returning to i. The crossing probability $P_i(\lambda_{0j} | \lambda_{mi})$ can directly be obtained from the path ensemble in Eq. (1) and is given by the number of pathways n_{ij} that start in i, cross λ_{mi}, and end in j divided by the number of all pathways starting in i and crossing λ_{mi},

$$P_i(\lambda_{0j} | \lambda_{mi}) = \frac{\int \mathcal{D}\mathbf{x}(L) \mathcal{P}_{ij}[\mathbf{x}(L)]}{\int \mathcal{D}\mathbf{x}(L) \sum_j \mathcal{P}_{ij}[\mathbf{x}(L)]} \approx \frac{n_{ij}}{\sum_j n_{ij}}. \tag{6}$$

To obtain the rate constants for all transitions between the different stable states in the system the flux through the λ_{mi} interface has to be calculated only once for each state. All crossing probabilities are calculated simultaneously within a single MSTPS/TIS simulation. Compared to the original two-state sampling only one simulation including all transitions has to be performed instead of one simulation for each transition. For trajectories originating in the same stable state i the ratio of the rate constants is given by the ratio of the corresponding crossing probabilities

$$\frac{k_{ij}}{k_{ik}} = \frac{P_i(\lambda_{0j} | \lambda_{mi})}{P_i(\lambda_{0k} | \lambda_{mi})}. \tag{7}$$

As can be seen from Eq. (7) an efficient sampling of the entire multiple state path ensemble will be difficult if there are large differences in the rate constants. If, e.g., the rate constant for a transition $i \rightarrow j$ is a factor of 10^4 larger than for the $i \rightarrow k$ transition, also the corresponding crossing probabilities have this ratio. Since the trajectories are sampled with their correct weights determined by the crossing probabilities, cf. Eq. (6), only 1 out of 10^4 is a $i \rightarrow k$ pathway. To address this problem we apply a biasing function to obtain an equal sampling of all types of pathways.

B. Wang–Landau biasing of the MS-TPE

The Wang–Landau sampling scheme[19,20] is a Monte Carlo (MC) algorithm that estimates the density of states $g(E)$ accurately via a random walk that produces a flat histogram in energy space. The random walk is performed with a probability proportional to the reciprocal of the density of states $1/g(E)$. Since the density of states is unknown at the beginning it is set initially to 1 for all energies. During the sampling the acceptance of a new energy level is given by

$$P_{\text{acc}}[E^{(o)} \rightarrow E^{(n)}] = \min\left[1, \frac{g(E^{(o)})}{g(E^{(n)})}\right]. \quad (8)$$

Each time an energy level is visited the corresponding density of states is updated by multiplication with a modification factor $f > 1.0$ and a histogram $H(E)$ is recorded. The sampling is continued until the histogram is *flat*, usually until for all E the histogram $H(E)$ is $x\%$ of the average histogram $\langle H(E) \rangle$, where x depends on the size and complexity of the system and the desired accuracy. For a flat histogram the density of states converges to the true value with an accuracy proportional to $\ln(f)$. In a next step, the modification factor is reduced, e.g., $f_{i+1} = \sqrt{f_i}$ (or any other rule that decreases f monotonically toward 1.0), the histogram is reset, whereas the density of states is retained and the sampling is continued until the new histogram is again flat. The sampling is converged if the modification factor is smaller than some defined value, e.g., $f_{\text{final}} = \exp(10^{-8})$.

The basic idea of the Wang–Landau algorithm can also be applied to the sampling of different types of pathways in the multiple state path ensemble. The type of a pathway determines which two states are connected, i.e., $i \rightarrow j$, $i \rightarrow k$ trajectories, etc. The fraction of different pathways then corresponds to the density of states, i.e., a *density of path types*. As shown in Eq. (7) the path fraction is related to the crossing probability which is needed to calculate the rate constants. The Wang–Landau acceptance criterion is included by extending the acceptance rule, Eq. (4), with respect to the path-type density g,

$$P_{\text{acc}}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = \min\left[1, \frac{L^{(o)}}{L^{(n)}} \frac{\rho(x_{\text{sp}}^{(n)})}{\rho(x_{\text{sp}'}^{(o)})}\right]$$
$$\times \sum_{ijkl} \widetilde{h}_{ij}[\mathbf{x}^{(n)}(L^{(n)})]\widetilde{h}_{kl}[\mathbf{x}^{(o)}(L^{(o)})]\min\left[1, \frac{g_{kl}}{g_{ij}}\right], \quad (9)$$

where g_{ij} is the density of trajectories connecting state $i \rightarrow j$. The path-type density g_{ij} and the corresponding histogram H_{ij} are updated according to the Wang–Landau scheme. Note that the fourfold sum over states picks out the corresponding path-type densities for the old and new paths. There is always one combination of k and l for which $\widetilde{h}_{kl} = 1$ since we assume that the old path connects a pair of stable states. If the sampling is converged the relative path-type densities are equal to the relative number of path n,

$$\frac{n_{ij}}{n_{kl}} = \frac{g_{ij}}{g_{kl}}. \quad (10)$$

Due to numerical reasons usually the logarithm of the density of path type is recorded. The crossing probability is given by

$$P_i(\lambda_{0j}|\lambda_{mi}) \approx \frac{n_{ij}}{\Sigma_k n_{ik}} = \frac{1}{\Sigma_k \dfrac{n_{ik}}{n_{ij}}} = \frac{1}{\Sigma_k \dfrac{g_{ik}}{g_{ij}}}$$
$$= \frac{1}{\Sigma_k \exp[\ln(g_{ik}) - \ln(g_{ij})]}. \quad (11)$$

The fraction of each path type with respect to all sampled pathways is correspondingly

$$\frac{n_{ij}}{n_{\text{tot}}} = \frac{1}{\Sigma_k \Sigma_l \exp[\ln(g_{kl}) - \ln(g_{ij})]}. \quad (12)$$

Using the Wang–Landau scheme to sample the multiple state path ensemble we can directly obtain the crossing probabilities from the path-type density and achieve an equal sampling of all path types due to the requirement of a flat histogram H_{ij}.

C. Simple bias

As an alternative we can also bias the MSTPS by assigning a fixed weight w_{ij} to each path type. These weights are likewise included in the acceptance criterion

$$P_{\text{acc}}[\mathbf{x}^{(o)} \rightarrow \mathbf{x}^{(n)}] = \min\left[1, \frac{L^{(o)}}{L^{(n)}} \frac{\rho(x_{\text{sp}}^{(n)})}{\rho(x_{\text{sp}'}^{(o)})}\right]$$
$$\times \sum_{ijkl} \widetilde{h}_{ij}[\mathbf{x}^{(n)}(L^{(n)})]\widetilde{h}_{kl}[\mathbf{x}^{(o)}(L^{(o)})]\min\left[1, \frac{w_{kl}}{w_{ij}}\right]. \quad (13)$$

The unbiased number of trajectories for a certain path type is given by

$$n_{ij} = n_{ij}^{\text{bias}} w_{ij}, \quad (14)$$

where n_{ij}^{bias} is the number of trajectories in the biased sampling. Using fixed biasing weights it is not essential to use the logarithm of the weights but in most cases it is numerically more stable. The ratio of two path types is expressed as

$$\frac{n_{ij}}{n_{kl}} = \frac{n_{ij}^{\text{bias}}}{n_{kl}^{\text{bias}}} \exp[\ln(w_{ij}) - \ln(w_{kl})]. \quad (15)$$

The corresponding crossing probabilities and path-type fraction are calculated using Eqs. (11) and (12). If the biasing weights are set to the converged path-type densities from the Wang–Landau sampling then all types of trajectories are sampled equally, i.e., $n_{ij}^{\text{bias}}/n_{kl}^{\text{bias}} = 1.0$, due to the flat histogram criterion in the Wang–Landau scheme and Eq. (15) is equal to Eq. (10).

D. Correlation function

An important aspect in evaluating the quality of a simulation is the correlation within the path ensemble. Within the Metropolis scheme for sampling the trajectory space a new

trial pathway is created on the basis of an old one and therefore subsequent pathways are correlated up to a certain degree. The strength of the correlation depends on the perturbation made at the shooting point and the dynamics of the system. The larger the changes at the shooting point the more the subsequent pathway differs from the old one. But this also influences the acceptance. The larger the perturbation at the shooting point the less likely the new trajectory is a valid pathway and the old trajectory is counted again in the ensemble. If the dynamics are deterministic the displacement at the shooting point can be tuned to obtain the fastest decorrelation. It was shown in Ref. 7 that the decorrelation is fastest if the overall acceptance is around 0.4.

In the multiple state path sampling the decorrelation is enhanced due to the switching between different path types. A $i \rightarrow j$ pathway that has been created by a shooting move from a $k \rightarrow l$ trajectory is already quite different. To quantify this behavior a correlation function is defined,

$$c(n) = \frac{\frac{1}{N-n}\Sigma_{i=1}^{N-n}(G_i - \langle G \rangle)(G_{(i+n)} - \langle G \rangle)}{\frac{1}{N}\Sigma_{i=1}^{N}(G_i - \langle G \rangle)^2}, \quad (16)$$

where N is the total number of sampling steps, G is the quantity of interest, e.g., path type or path length, and $\langle G \rangle$ is the average value of G, i.e., $\langle G \rangle = [1/N]\Sigma_i G_i$. A fast decay of the correlation function indicates a good sampling of the trajectory space.

III. RESULTS

We have implemented the biasing of the multiple state path ensemble using both the Wang–Landau sampling scheme and a simple bias. We apply the biased sampling to a model system using deterministic dynamics.

A. Model system

The model consists of a two-dimensional (2D) triatomic molecule embedded in a Weeks–Chandler–Andersen (WCA) fluid[21] as described in Ref. 15. The interactions of the N particles in the system are determined by the WCA-potential

$$V_{WCA}(r) = \begin{cases} 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \epsilon & \text{if } r \leq r_{WCA} \equiv 2^{1/6}\sigma \\ 0 & \text{if } r > r_{WCA}, \end{cases} \quad (17)$$

where r is the distance between two particles, ϵ is the interaction strength, and σ is the interaction radius. In addition the particles in the trimer interact via a double well potential

$$V_{dw}(r) = h\left[1 - \frac{(r - r_{WCA} - w)^2}{w^2}\right]^2, \quad (18)$$

where h is the barrier height and w is the barrier width. Only particles 1–2 and 2–3 are coupled via the double well potential, i.e., there are four minima in the free energy surface as a function of the interparticle distances r_{12} and r_{23}. Within the minima the interparticle distance can be either in a short (sh)
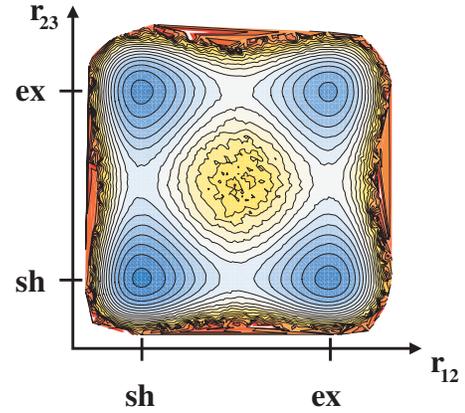

FIG. 1. Free energy surface of the 2D triatomic molecule as a function of the interparticle distances r_{12} and r_{23}. The interparticle distance can be either in a short (sh) or extended (ex) conformation resulting in four stable states.

$r = r_{WCA}$, or an extended (ex) $r = r_{WCA} + 2w$, conformation and the four stable states are correspondingly labeled shsh, exsh, shex, and exex, cf. Fig. 1. Transitions can occur between all states in the system changing either one or both interparticle distances. Including trajectories that return to their initial stable state after crossing the λ_{mi} interface there are 16 different types of pathways that are sampled simultaneously in the simulations.

The results are presented in reduced units where σ is the unit of length, ϵ of energy, m of mass, and $\tau \equiv (\sigma^2 m / \epsilon)^{1/2}$ of time. All simulations are performed in a microcanonical ensemble with a total energy of $E = 20\epsilon (\sim T = 0.171\epsilon/k_B)$. The total number of particles is $N = 100$ with a particle density of $\rho = 0.6\sigma^{-3}$ and periodic boundary conditions are applied. The barrier height is $h = 1.0\epsilon$ and the width $w = 0.25\sigma$. The velocity Verlet algorithm is used to integrate the equation of motion with a time step of $\Delta t = 0.001\tau$. New pathways are created by a shooting move with a maximum displacement of the momentum at the shooting point of δp_{max}. Two sets of simulations are performed one with $\delta p_{max} = 5.0(\epsilon m)^{1/2}$ and one with $\delta p_{max} = 0.5(\epsilon m)^{1/2}$. For the unbiased simulations the overall acceptance ratios are ~ 0.48 and ~ 0.57, respectively. The acceptance ratios of the biased simulations are discussed in the corresponding sections.

B. Wang–Landau sampling of the multiple state transition path ensemble

Within the multiple state path sampling trajectories of different path types are sampled according to their probability distribution. For the above described model system the path fractions [cf. Eq. (12)] as obtained from an unbiased simulation are (errors in the last digit are indicated by the subscript)

$\rightarrow$	shsh	exsh	shex	exex
shsh	0.267$_7$	0.053$_2$	0.052$_2$	0.00010$_2$
exsh	0.053$_2$	0.168$_6$	0.00024$_4$	0.025$_2$
shex	0.052$_2$	0.00025$_3$	0.165$_6$	0.025$_2$
exex	0.00012$_2$	0.025$_2$	0.025$_2$	0.089$_5$

If the sampling is sufficient the number of $i \rightarrow j$ pathways has to be the same as $j \rightarrow i$ pathways due to the detailed balance
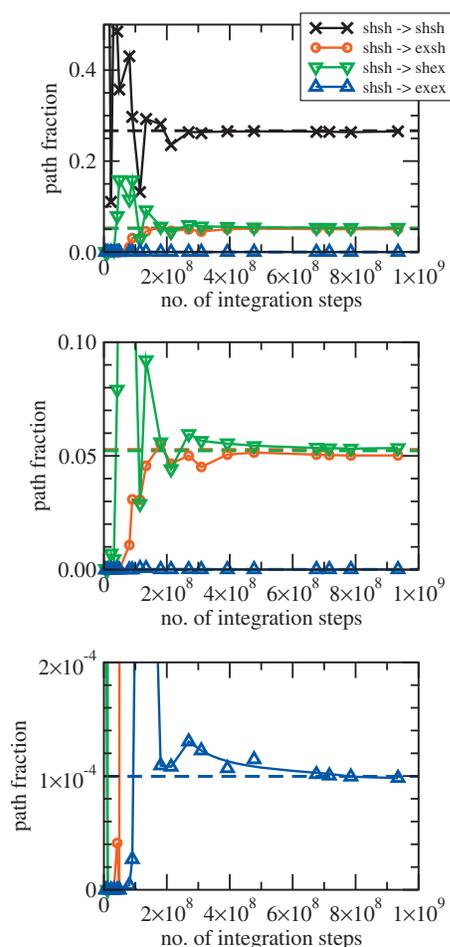
FIG. 2. Path fraction calculated from the path-type density in the Wang–Landau scheme. Results are shown for transitions out of the shsh state, black crosses mark shsh→shsh (likely), red circles shsh→exsh (medium likely), green down-triangles shsh→shex (medium likely), and blue up-triangles shsh→exex (unlikely) trajectories. In the top graph all four trajectories are shown, the middle and bottom graphs zoom in on the y-axis. The converged results from the unbiased multiple state path sampling are marked by dashed lines. The path fraction obtained within the Wang–Landau scheme converges to the value of the unbiased simulation.

criterion and the resulting matrix containing the path fractions is symmetric. The values on the diagonal denote pathways that return to their initial state and depend on the position of the λ_{mi} interfaces. The ratio of the remaining path fractions within one column, respectively row, depends on the actual transition probabilities [cf. Eq. (7)] and is thus inherent to the system, i.e., a shsh→exsh trajectory will be sampled more often than a shsh→exex trajectory by a factor of $0.053/0.0001 \approx 530$. Applying the Wang–Landau scheme all pathways will be sampled equally and the corresponding path fraction can be extracted from the path-type density in Eq. (12). In Fig. 2 the results are shown for the four different path types initiating in the shsh state. The path fraction is plotted versus the number of Verlet integration steps. The top graph shows all four path types, the middle and bottom graphs zoom in on the shsh→exsh, shsh→shex, and shsh →exex pathways, respectively. Within the Wang–Landau sampling the modification factor f is reduced 19 times from $\ln(f_0)=1$ to $\ln(f_{final})=2^{-19}=1.91\times10^{-6}$ and the number of integration steps and path-type density is recorded at each

step. A maximum deviation of 20% from the average histogram $H(\mathbf{x})$ (flat histogram) was allowed before reducing the modification factor. The results for the path fraction of the converged unbiased sampling are indicated by the dashed lines. Comparing the results from the Wang–Landau sampling and the unbiased simulations we find a very good agreement. With decreasing modification factor f the Wang–Landau path-type densities converge nicely toward the reference values of the path fraction. The same behavior is also found for the 12 remaining types of pathways in the system. The total number of integration steps is quite large, though, $\sim 9.3\times10^8$, sampling a total number of $\sim 2.7\times10^6$ pathways until the Wang–Landau sampling is converged. For $\delta p_{max} = 5.0(\epsilon m)^{1/2}$ the overall acceptance ratio is ~ 0.3, which is somewhat lower than in the unbiased simulation due to the additional condition in the acceptance rule, Eq. (9). 35% of the rejections are due to the Wang–Landau density criterion.

To assess the efficiency of the Wang–Landau biasing of the multiple state path sampling we compare the average value of the path fraction and the corresponding error to the results of the unbiased simulation. This comparison is not entirely straightforward since in the unbiased simulation the error can be calculated as the standard deviation over block averages of a certain block length. In the Wang–Landau sampling we obtain an estimate of the error by performing ten independent calculations. The average of the path fraction is taken over these ten simulations for each value of the modification factor f, i.e., a new *block* is defined for each value of f. Since the number of pathways that are necessary to obtain a flat histogram varies for a certain f also an average number of integration steps is calculated for each block. In Fig. 3 the results are shown for three transitions out of the shsh state. The graphs on the left side, (a-I)–(c-I), show the average path fraction as a function of the total number of integration steps obtained using the Wang–Landau biasing of the multiple state path sampling and the unbiased sampling, respectively. The graphs on the right side, (a-II)–(c-II), show the corresponding relative errors given by the standard deviation over block averages. Again, both simulations converge to the same value of the path fraction, i.e., the Wang–Landau biasing does converge to the correct results. Regarding the relative error, however, the Wang–Landau biasing appears to be less efficient than the unbiased simulation. Similar results are found for all 16 types of pathways in the model system. One problem in the Wang–Landau biasing is the convergence with respect to the modification factor f. If the modification factor is reduced the computational cost to achieve a flat histogram should in general increase, which is not always the case in our simulations. Also, the number of sampled pathways for a certain value of f differs quite a bit for the ten different runs. This is due to a rather slow decorrelation between subsequent pathways leading to a different convergence behavior of the histogram in the various runs and for various modification factors f. To obtain more accurate results the tolerance of the flat histogram would have to be decreased but this would, in turn, make the simulations even more costly, although it might also significantly reduce the relative errors. Another reason why this particular Wang–Landau scheme is not efficient might be that the path-type
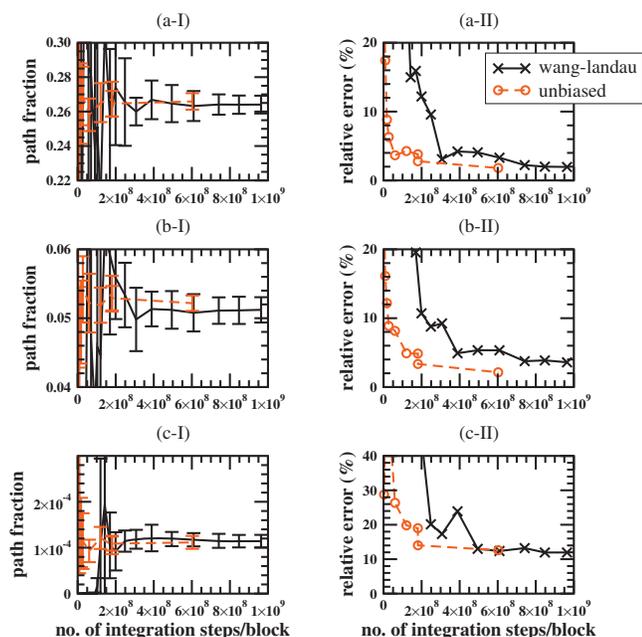
FIG. 3. Path fraction calculated from the path-type density in the Wang–Landau scheme (black, solid line with crosses) compared to the results from an unbiased simulation (red, dashed line with circles). Graphs (a-I)–(c-I) show the path fraction and (a-II)–(c-II) show the relative error as a function of the number of integration steps. Results are presented for three different path types: (a) shsh→shsh (likely), (b) shsh→exsh (medium likely), and (c) shsh→exex (unlikely). The unbiased simulation is more efficient in determining the path fraction than the Wang–Landau biasing scheme for all three cases.
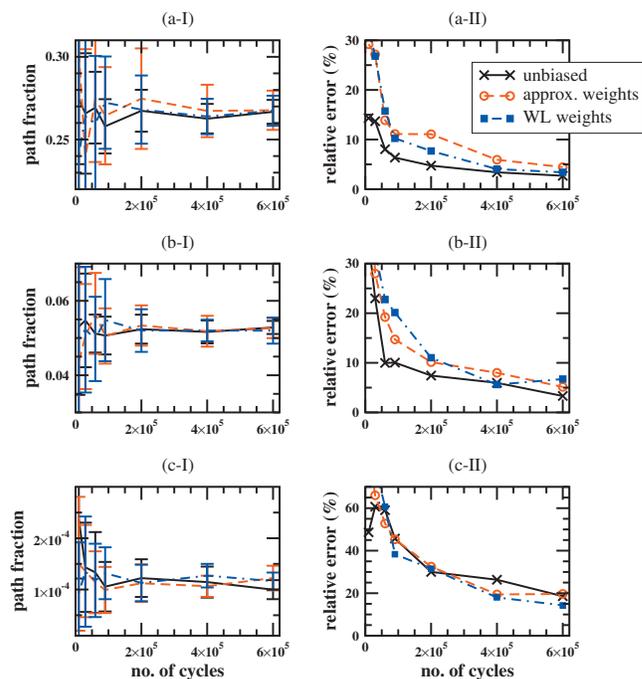

FIG. 4. Path fraction calculated using two sets of weights (approximate weights of 1, 10 100—red, dashed line with circle; weights from Wang–Landau path-type density—blue, dashed-dotted line with squares) and no biasing (black, solid line with crosses). Graphs (a-I)–(c-I) show the path fraction and (a-II)–(c-II) show the relative error as a function of the number of simulation cycles per block. Results are presented for three different path types: (a) shsh→shsh (likely), (b) shsh→exsh (medium likely), and (c) shsh→exex (unlikely). No significant difference in determining the path fraction is found for the biased and unbiased simulations. The results are comparable for all path types.

density is discrete. Usually, the energy density sampled using Wang–Landau MC is a continuous function, whereas in our model system the path-type density only has 16 entries, corresponding to the 16 different pathways.

To obtain a better understanding on how the biasing influences the multiple state path sampling we perform simulations applying a simple biasing function.

C. Simple bias

Applying the Wang–Landau sampling scheme to the multiple state path sampling leads to correct results but does not appear to be efficient. In a next step we apply a simple biasing function to the acceptance rule, cf. Eq. (13), to investigate the influence of the biasing function on the switching between different path types and the quality of the individual path ensembles.

Looking at the path fraction matrix and, e.g., Fig. 2, it can be seen that the trajectories in our model system can be divided into three groups:

- very unlikely (shsh→exex, exsh→shex, shex→exsh, exex→shsh),
- medium likely (shsh→exsh, shsh→shex, exsh→shsh, exsh→exex, shex→shsh, shex→exex, exex→exsh, exex→shex), and
- very likely (shsh→shsh, exsh→exsh, shex→shex, exex→exex).

In an unbiased simulation likely trajectories will be

sampled more often than unlikely ones. The biasing of the ensemble was done with two different sets of weights. One set has approximate weights of 1 (unlikely), 10 (medium likely), and 100 (likely) for the different types of pathways. In the second set the weights are set to the path-type densities obtained in the Wang–Landau simulation. Results for three transitions originating in the shsh state are shown in Fig. 4 and are representative for all other path types. The graphs on the left side, (a-I)–(c-I), show the path fraction as a function of the number of simulation cycles per block, i.e., number of MC moves/block, averaged over 20 blocks. The graphs on the right side, (a-II)–(c-II), show the corresponding relative errors. The simulations applying a biased sampling converge to the same value as the unbiased simulation with increasing block size. Also the relative errors behave the same with increasing block size and are of similar magnitude for the three different simulations. Again it appears that biasing the multiple state path ensemble with either approximate weights or Wang–Landau weights gives the correct path fractions, but does not improve the convergence.

We can investigate the effect of the biasing function for the different types of pathways by looking at the correlation function, cf. Eq. (16). The correlation function is evaluated with respect to the different path types, i.e., $G_i = h^{pt}[\mathbf{x}_i]$, where $h^{pt}[\mathbf{x}_i]$ is an indicator function that is equal to 1 if a certain path is of the desired type and 0 otherwise. The correlation function for the path type exex→shsh (unlikely transition) is shown in Fig. 5(a). In the unbiased simulation
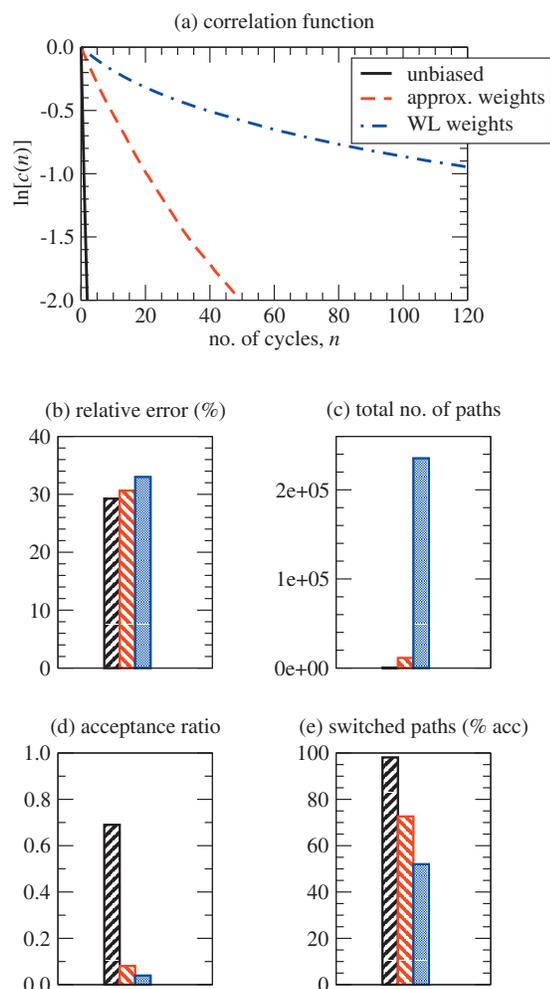
FIG. 5. (a) Correlation function $c(n)$ with respect to the path type for unbiased (black, solid line), biased using approximate weights (red, dashed line) and using Wang–Landau weights (blue, dashed-dotted line) MSTPS simulations; (b) relative error, (c) total number of trajectories, (d) acceptance ratio, and (e) percentage of switched trajectories for a simulation over 20 blocks with 2×10^5 simulation cycles. Left bars mark the unbiased simulation, center bar the approximate weights, and right bars the Wang–Landau weights. Results are shown for the exex→shsh transition (unlikely). The decorrelation is fastest in the unbiased simulation, but the size of the path ensemble (total number of paths) is much smaller leading to a similar relative error in all three cases. The overall acceptance drops in the biased simulations due to the additional factor in the acceptance rule.

(black, solid line) the decorrelation is immediate, i.e., the same type of pathway is hardly sampled in succession. If the sampling is biased towards the unlikely transition the correlation increases and the same type of pathway is sampled more often in a row. The effect is largest for the weights extracted from the Wang–Landau path-type density. The relative error of the path fraction, Fig. 5(b), for 20 blocks with 2×10^5 cycles each is not significantly influenced by the biasing. The convergence, and therefore the error of a simulation, is a tradeoff between the total number of pathways and the decorrelation within the path ensemble. As shown in Fig. 5(c) the total number of pathways increases dramatically upon biasing whereas the decorrelation decreases, Fig. 5(a). In the unbiased simulation the path ensemble for the unlikely transition is rather small, only 458 trajectories, but the acceptance ratio is high, ~0.7, cf. Fig. 5(d), and as shown in Fig.

5(e) 98% of the accepted pathways of this type have been created from a different type of pathway, i.e., there was a switching between different types of pathways, leading to the extremely fast decorrelation. In the biased simulation using the Wang–Landau weights the total number of trajectories is much larger, 235 434, but the overall acceptance has decreased to ~0.04 and only 52% of the accepted pathways switched from a different path type.

The results are comparable for all unlikely transitions. For medium likely transitions the effect of the biasing on the decorrelation and ensemble size is less pronounced and for likely transitions the trend is even slightly reversed. Here, there are less trajectories in the path ensemble of the biased simulation, but the decorrelation is somewhat faster. Again the relative errors are comparable.

1. Switching in the biased path ensemble

As discussed above, to obtain a good sampling of the entire multiple state ensemble the switching between different types of trajectories is essential. The MSTPS scheme will only be efficient if different types of trajectories are created by the shooting moves. If the transition probability for two transitions $i \rightarrow j$ and $i \rightarrow k$ differs significantly, e.g., $k_{ij}/k_{ik} = 1000$, the trajectories will be sampled accordingly and the switching will be low. Again, for $k_{ij}/k_{ik}=1000$ only 1 out of 1000 trajectories will be a $i \rightarrow k$ pathway and therefore there can only be one switching event every 1000 pathways, which is schematically shown in the top left graph (a.1) in Fig. 6. Applying a biasing function the number of sampled trajectories will be equal for both transitions, i.e., $n_{ij}=n_{ik}=500$, and several switches between the two types of trajectories would in principle be possible, as depicted in the middle left graph (a.2) of Fig. 6. The increased switching would then also enhance the sampling of the path ensemble. From the discussion above, though, and especially from analyzing the correlation function plotted in Fig. 5(a), it becomes clear that the biasing function changes the switching behavior to the situation depicted in the bottom left graph (a.3) of Fig. 6. The number of rare and frequent trajectories in the ensembles is now equal but the switching is still low. This is due to the fact that in the applied biasing scheme the number of unlikely trajectories is increased by a rejection of switching events from an unlikely to a likely pathway, i.e., by modifying the acceptance rule. This is also shown in the two right graphs, (b.1) and (b.2), in Fig. 6 for the trimer system. The switching between the likely and medium likely path types is good for both the biased and unbiased simulations whereas switching to the unlikely path type is poor. The bias does not increase the switching between likely/medium and unlikely trajectories but increases the sampling of unlikely pathways. To effectively increase the switching between different path types it would be necessary to change the generation probability of a certain path type. This would involve a modification of the shooting move. Changing the generation probability of a certain path type is not straightforward. It requires prior knowledge about the system and will greatly depend on the investigated problem. In this case a successive two-state path sampling might be more appropriate.
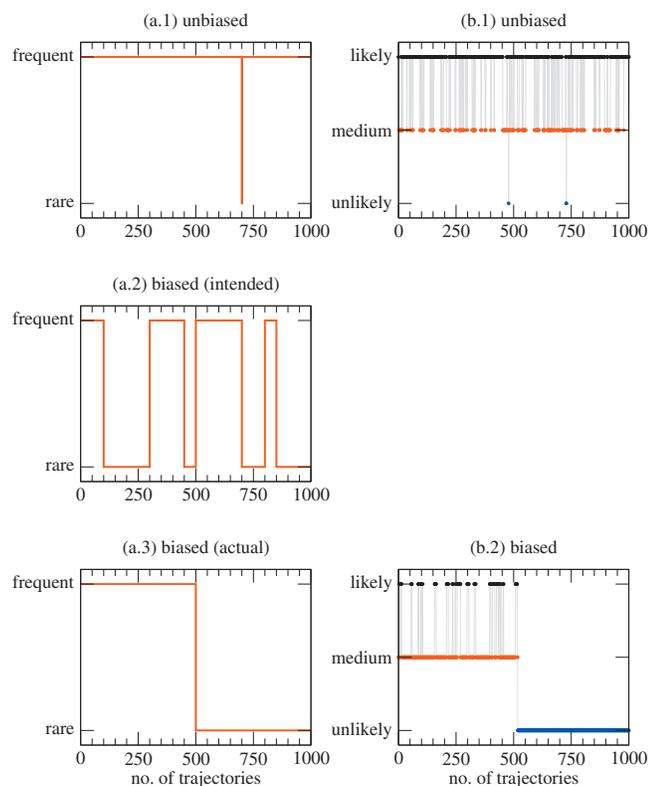
FIG. 6. (a) Schematic figure of the switching behavior between different types of pathways for a system where the ratio of the transition probabilities of two transitions is 1:1000. Graph (a.1) exemplifies the unbiased simulation, only 1 out of 1000 trajectories is a rare trajectory. In the middle graph (a.2) both types of pathways are sampled equally and a frequent switching occurs. This is the behavior we intended to induce by applying a biasing function. The bottom graph (a.3) displays the switching behavior as obtained by applying the biasing function. Once the simulation switches from a frequent to a rare trajectory the biasing function in the acceptance rule keeps the sampling on the unlikely pathway. (b) Graphs (b.1) and (b.2) show the switching between likely, medium, and unlikely trajectories for 1000 pathways of the trimer system, graph (b.2) reflects the behavior schematically shown in figure (a.3).

The here applied biasing scheme does not increase the switching between different path types and does therefore not improve the convergence with respect to the path fraction which is needed for the calculation of the crossing probabilities. Concerning the path fraction the unbiased multiple state path sampling is as efficient as the biased sampling. This is different from a biased sampling in conformational space where the sampling of high energy configuration is low due to a rejection of these configurations in the Metropolis scheme, whereas they are still generated with a similar probability as the low energy configurations. Applying the shooting moves in the biased MSTPS, on the other hand, already generates the different types of pathways with a different probability. This is due to the fact that the generation probability depends not only on the displacement at the shooting point, δp_{max}, but also on the underlying dynamics. The biasing does therefore not increase the acceptance of new, unlikely pathways but increases the *rejection* of new likely trajectories resulting in the behavior depicted in the bottom graphs in Fig. 6. Nevertheless the sampling of unlikely trajectories does increase and the individual path ensembles of the corresponding path types contain much more trajectories

in the biased simulations. Quantities derived from these ensembles for each path type therefore have a higher accuracy, as discussed in Sec. III C 2.

2. Quality of individual path ensembles

The quality of the individual path ensembles was investigated by evaluating the average path length, the average potential and kinetic energy of the entire system and of the trimer only, as well as the maximum potential and kinetic energy along the trajectories of the entire system and the trimer. All quantities are recorded for each path type separately and block averages and errors are calculated over 20 blocks of various block lengths. Simulations are performed applying no bias, the approximate weights, and the Wang–Landau weights to the multiple state path sampling. In addition two-state TPS simulations are performed for each path type which would effectively correspond to a weight of 1 for the sampled path type and an infinite weight for all other path types.

The relative error of the average path lengths for different individual ensembles is shown in Fig. 7. Figure 7(a) shows the results for the ensemble of shsh→shsh (likely) trajectories, Fig. 7(b) of shsh→shex (medium likely), and Fig. 7(c) of shsh→exex (unlikely) trajectories. The relative error is smaller for the likely and medium likely pathways indicating that the sampling of unlikely trajectories is more difficult in general. For the likely and medium likely pathways [Figs. 7(a) and 7(b)] the multiple state path sampling is much more efficient than the two-state sampling, which is due to the faster decorrelation induced by the switching between different types of pathways. The difference between the unbiased and biased multiple state path sampling is small. In the unbiased simulation the path ensemble for the likely transitions is large and decorrelation is good. The biasing function in the acceptance rule reduces the number of likely transitions but does not have a significant influence on the correlation, i.e., the biasing shifts the number of sampled trajectories from the likely to the unlikely transitions, effectively avoiding an oversampling of likely trajectories.

For the unlikely transition, Fig. 7(c), the situation is slightly different. The most efficient sampling is achieved in the multiple state path sampling applying Wang–Landau weights in the biasing function. The two-state sampling of unlikely trajectories appears to be more efficient than applying no bias or approximate weights to the multiple state path sampling. Again this can be understood by analyzing the correlation function and number of pathways in the ensemble, as shown in Fig. 8. The correlation function is evaluated with respect to the path length, i.e., $G_i \equiv L(\mathbf{x}_i^{pt})$, where $L(\mathbf{x}_i^{pt})$ is the length of path $\mathbf{x}_i$ of type pt. In Fig. 8 results are shown for the shsh→exex transition. Similar to the path fraction the decorrelation with respect to the average path length in the ensemble of an unlikely transition is fastest in the unbiased multiple state path sampling [Fig. 8(a)]. The correlation within the ensemble increases with increasing bias (approximate weights, Wang–Landau weights and two-state sampling), which is due to less switching, respectively for the two-state sampling no switching, between different types of trajectories. The relative error, Fig. 8(b), calculated
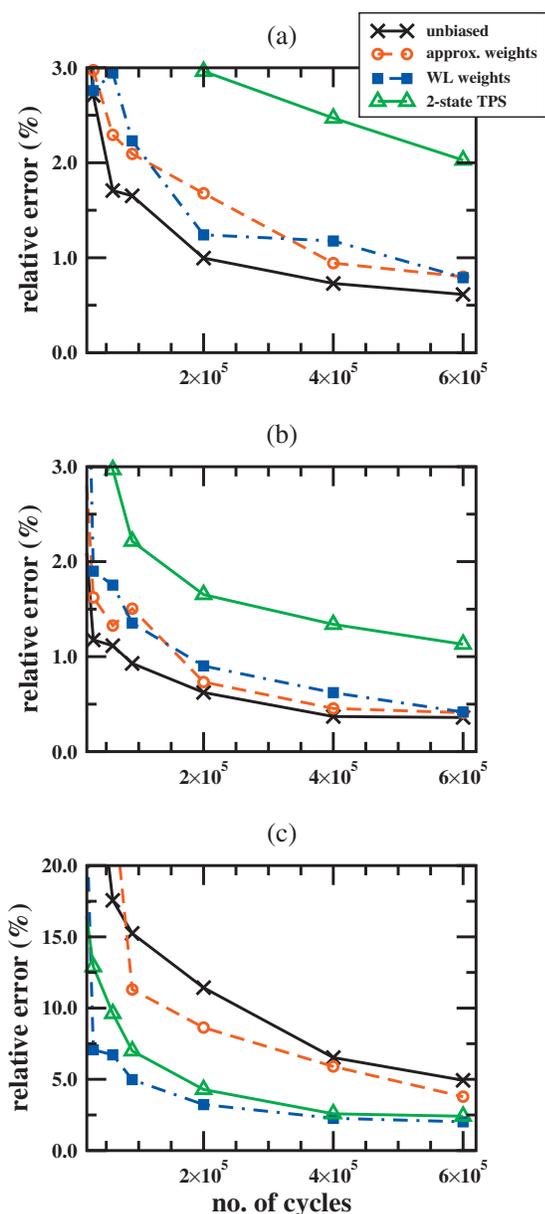
FIG. 7. Relative error of the average path length vs the number of simulation cycles/block for three different transitions out of the shsh state: (a) shsh→shsh (likely), (b) shsh→exsh (medium likely), and (c) shsh→exex (unlikely) with $\delta p_{max} = 0.5(\epsilon m)^{1/2}$. Results are shown for MSTPS simulations with no bias (black, solid line with crosses), biased applying approximate weights (red, dashed line with circles) and Wang–Landau weights (blue, dashed-dotted line with square), and two-state TPS simulations (green, solid line with up-triangles). Errors are calculated as standard deviation over block averages with 20 blocks of various length. For likely and medium likely transitions the multiple state path sampling is always more efficient than the two-state sampling. Biasing the multiple state path ensemble is important for a better sampling of unlikely transitions.
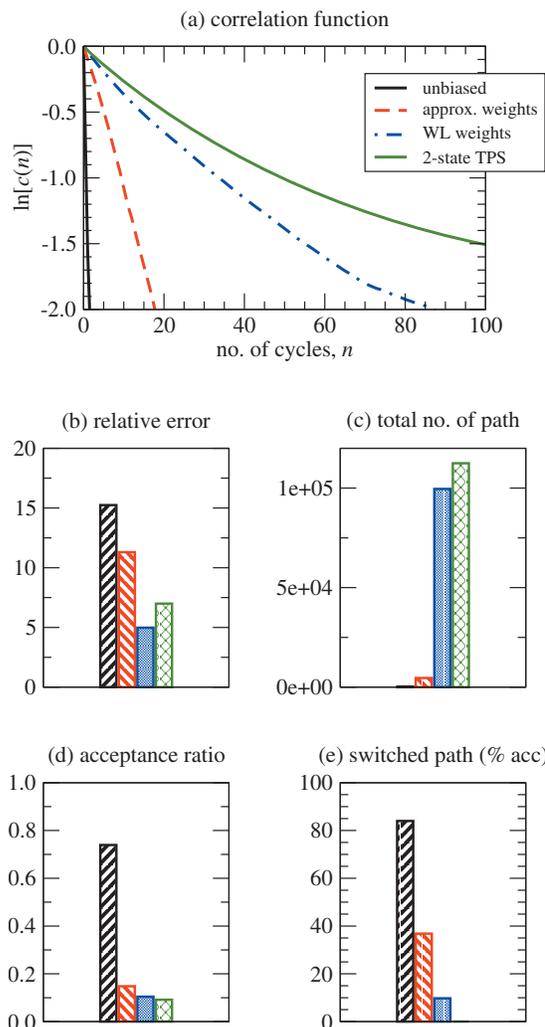

FIG. 8. (a) Correlation function $c(n)$ with respect to the average path length for unbiased (black, solid line), biased using approximate weights (red, dashed line) and using Wang–Landau weights (blue, dashed-dotted line), MSTPS simulations, and two-state TPS simulations (green, solid line); (b) relative error, (c) total number of trajectories, (d) acceptance ratio, and (e) percentage of switched trajectories for a simulation over 20 blocks with 9 $\times 10^4$ simulation cycles. Left bars mark unbiased simulation, center-left bars the approximate weights, center-right bars the Wang–Landau weights, and right bars the two-state sampling. Results are shown for the shsh→exex transition (unlikely) with $\delta p_{max} = 0.5(\epsilon m)^{1/2}$. The decorrelation is fastest in the unbiased simulation, but the size of the path ensemble (total number of paths) is much smaller. The relative error is smallest for the biased multiple state path ensemble applying Wang–Landau weights indicating the best balance between size of the path ensemble and decorrelation.

over 20 blocks of 9×10^4 cycles is largest for the unbiased MSTPS, which is the result of the small ensemble size. In the unbiased simulation there are only 211 trajectories, whereas the ensemble in the two-state TPS simulation contains 112 500 pathways, Fig. 8(c). It is indeed remarkable that such a small ensemble of trajectories in the unbiased MSTPS simulation already gives such a small relative error. This is mainly due to the high switching between the different path types, as shown in Fig. 8(e). 84% of all accepted shsh

→exex pathways in the unbiased simulation are created from a different path type. The most efficient sampling is the MSTPS applying Wang–Landau weights in the biasing function. The path ensemble is rather large and the decorrelation is still reasonable due to a certain amount of switching, whereas in the two-state sampling there is no switching by definition.

The results for the other investigated quantities, the average potential and kinetic energy of the entire system and of the trimer only as well as the maximum potential and kinetic energy along the trajectories of the entire system and the trimer, are comparable to the results for the path length. Also, the findings for the other path types are similar to the results in Figs. 7 and 8. For all likely and medium likely trajectories

the unbiased multiple state path sampling is most efficient. For the rare transitions applying a bias is most efficient if the weights correspond to the true weights. For the quality of the individual ensembles biasing the MSTPS is advantageous for the unlikely transitions since an oversampling of the likely transitions is avoided.

IV. CONCLUSIONS

Within the MSTPS scheme transitions between several stable states within a system can be sampled simultaneously. To obtain an efficient sampling of all transitions a switching between different types of pathways is essential. A problem occurs if the transition probabilities between different states differ significantly. Since the trajectories are sampled with the corresponding probability distribution unlikely transitions will be sampled less than likely transitions. To address this problem we have applied a biasing function to the acceptance rule in the multiple state path sampling in a Wang–Landau and simple biasing scheme. We find that the biasing function does improve the overall quality of the path ensembles for the different path types. However, the convergence with respect to the path fraction, which is needed to determine the crossing probabilities, is not enhanced. This is due to the fact that the biasing function adjusts the number of trajectories of each path type but does not increase the switching between different path types. Unlike configurations in a *regular* MC sampling different path types are generated with different probabilities and the biasing leads to an increase in the rejection of likely trajectories instead of an increase in the acceptance of unlikely ones. To change the switching behavior the generation probability of new trajectories would have to be changed. This is, however, not trivial since the generation probability of a certain path type depends not only on the displacement at the shooting point δp_{max}, but also on the underlying dynamics. Nevertheless, the biasing of the MSTPS does improve the quality of the individual ensembles since an oversampling of likely trajectories is avoided and the ensemble of unlikely trajectories is increased. The MSTPS is not applicable if there is no switching between different path types. In such a case a successive two-state TPS is preferred. The employed sampling algorithm does therefore depend on the investigated problem. In applications to a realistic system, such as a solvated protein conformational change, one cannot expect to obtain the number of pathways that we sample in this work. Nevertheless, on the order of 10^4 pathways are feasible, resulting in a reasonable sampling of path space. We emphasize that the multiple state path sampling is always more efficient if there is some switching between different transitions. Biasing the multiple state transition path ensemble does in general provide an equally good sampling and furthermore improves the quality of the individual path ensembles of unlikely transitions.

[1] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).
[2] E. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).
[3] T. Huber, A. Torda, and W. van Gunsteren, J. Comput.-Aided Mol. Des. **8**, 695 (1994).
[4] H. Grubmüller, Phys. Rev. E **52**, 2893 (1995).
[5] A. F. Voter, J. Chem. Phys. **106**, 4665 (1997).
[6] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).
[7] C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).
[8] D. Moroni, P. R. ten Wolde, and P. G. Bolhuis, Phys. Rev. Lett. **94**, 235703 (2005).
[9] J. Juraszek and P. G. Bolhuis, Proc. Natl. Acad. Sci. U.S.A. **103**, 15859 (2006).
[10] C. Dellago and P. G. Bolhuis, in *Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations*, Topics in Current Chemistry Vol. 268, edited by M. Reiher (Springer, Berlin, Heidelberg, 2007), p. 291.
[11] M. Grünwald, C. Dellago, and P. L. Geissler, J. Chem. Phys. **127**, 154718 (2007).
[12] T. S. van Erp and P. G. Bolhuis, J. Comput. Phys. **205**, 157 (2005).
[13] T. S. van Erp, Phys. Rev. Lett. **98**, 268301 (2007).
[14] P. G. Bolhuis, J. Chem. Phys. **129**, 114108 (2008).
[15] J. Rogal and P. G. Bolhuis, J. Chem. Phys. **129**, 224107 (2008).
[16] S. Park and V. S. Pande, J. Chem. Phys. **124**, 054118 (2006).
[17] F. Noé, I. Horenko, C. Schütte, and J. C. Smith, J. Chem. Phys. **126**, 155102 (2007).
[18] J. D. Chodera, N. Singhal, V. S. Pande, K. A. Dill, and W. C. Swope, J. Chem. Phys. **126**, 155101 (2007).
[19] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[20] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[21] J. D. Weeks, D. Chandler, and H. C. Andersen, J. Chem. Phys. **54**, 5237 (1971).