



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Adaptive Importance Sampling Monte Carlo Simulation of Rare Transition Events

Maurice de Koning, Wei Cai, Babak Sadigh,
Tomas Opperstrup, Malvin H. Kalos, Vasily V.
Bulatov

September 2, 2004

Journal of Chemical Physics

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Adaptive Importance Sampling Monte Carlo Simulation of Rare Transition Events

Maurice de Koning¹, Wei Cai², Babak Sadigh², Tomas Oppelstrup², Malvin H. Kalos² and Vasily V. Bulatov²

¹*Instituto de Física, Universidade de São Paulo,
CP 66318, 05315-970 São Paulo - SP, Brazil and*

²*Lawrence Livermore National Laboratory, P.O. Box 808, L-371, Livermore, CA 94550*

We present an adaptive importance sampling method for quantifying the statistics of rare-event processes in atomistic simulations. The approach is based on an explicit evaluation of the probability that a sequence of states (or path) initiating in a state A leads to a reactive transition event to final state B . The importance sampling method seeks to bias the sampling of system trajectories such that those that contribute significantly, i.e. those that characterize reactive transitions, are generated more frequently. This is accomplished by means of an importance function, which modifies the transition probabilities among the microstates that comprise a path. For each problem there exists an optimal importance function, which biases that path sampling in such a manner that each path initiating in A leads to a successful event. The fact that the optimal function obeys a variational principle, then leads to an adaptive method in which a trial function form containing a set of adjustable parameters is chosen. The parameters are then adjusted so as to bring the trial function as close as possible to the optimal importance function. We demonstrate the method in two model problems.

PACS numbers:

I. INTRODUCTION

The purpose of predictive modeling and simulation at the atomistic level is to characterize and quantify the atomistic unit mechanisms that control the macroscopic behavior of complex systems. This objective is common to many fields of research, including chemistry, physics, biology, and materials science, where a fundamental understanding of the elementary processes requires detailed insight into the atomic-scale configurations and rearrangements. The atomistic modeling techniques of molecular dynamics (MD) and Monte Carlo (MC) provide powerful tools in this context¹, allowing a detailed observation of atomic-scale structures during controlled computational “experiments” that are difficult or impossible to realize in a laboratory.

In general, a meaningful atomistic simulation should satisfy three conditions: (i) the description of the interatomic interactions should be sufficiently accurate, (ii) the number of particles considered in the simulation should be statistically significant, and (iii) the simulation should cover a sufficiently long time interval to cover the processes of interest. While the development of modern electronic-structure theory² and the evolution of massively parallel computing resources³ have significantly boosted the capabilities of atomistic simulations with regard to the first two conditions, the remaining barrier to predictive atomistic modeling concerns the limited time scale accessible to MD and MC simulations.

This key limitation stems from the fact that MD and MC naturally operate on the time scale of typical atomic vibrations in the system. For instance, in solid-phase systems this scale is defined by a characteristic phonon frequency, typically of the order of 10^{13} Hz, restricting the accessible simulation time to the order of nanoseconds. On the other hand, many interesting phenomena,

such as relaxation in glasses and protein folding, occur on time scales of milliseconds or even longer. The origin of this time-scale disparity can be traced to topographical features of the potential-energy surface, which typically contains several deep energy basins surrounded by barriers many times higher than the thermal-energy scale. Given that the rate of inter-basin transitions typically decreases exponentially with increasing barrier height, such transitions represent rare events on the time scale of atomic/molecular motion. Consequently, attempts to simulate such transitions using conventional MD or MC methods are hopelessly inefficient since virtually all CPU cycles are spent on the “irrelevant” equilibrium motion within a basin.

As a result, significant attention has been given to the development of special simulation tools that enable the study of rare transition events. A large portion of this effort has been devoted to methods within the framework of transition state theory (TST),⁴ ranging from a variety of techniques designed to locate saddle point configurations in the potential-energy landscape,⁵⁻⁸ to accelerated dynamics methods.^{9,10} While TST-based approaches are effective in systems with only a few particles, they are inefficient for handling transitions that involve collective motion of large numbers of atoms and multiple reactive mechanisms. To handle such complex problems effectively one should, instead of focusing on saddle points, adopt an approach based on system trajectories. In this context, a successful approach would be to modify the standard MD and MC techniques in such a manner that the probability of sampling a successful transition event is enhanced while spending less CPU cycles on the “irrelevant” equilibrium trajectories. A good *importance sampling* scheme of this kind should satisfy two conditions: (i) the relative probabilities of different transition trajectories must remain unaltered, and (ii), the abso-

lute probability of sampling a successful transition event should be enhanced by a known amount. These two conditions should be met to ensure that different transition mechanisms are sampled with correct relative probabilities, and allow a straightforward calculation of the corresponding transition rates.

The development of the transition path sampling (TPS) method by D. Chandler and coworkers^{11–14} was a fundamental step in this direction. TPS samples the subset of transition paths through a random-walk procedure that generates a subsequent transition event by slightly modifying the previous one. This approach satisfies the first condition and preserves the relative importance of different transition mechanisms. On the other hand, TPS does not satisfy the second condition; the ratio of the probabilities of successful to unsuccessful events is altered by an unknown amount. As a result, the calculation of transition rates requires a significant amount of computational effort.¹⁵

In this paper we discuss a new path-based importance sampling (IS) strategy for simulating rare-event processes at the atomistic level.¹⁶ The approach is based on the importance sampling MC formalism¹⁷ and seeks to enhance the probability of sampling successful transition events by using an *importance function*. By selecting it appropriately, one focuses predominantly on the successful transition events, while keeping track of the quantitative changes in their absolute probabilities. In this manner, the rare-event problem is reformulated into an optimization problem for the best-possible importance function. Once this function is known, the problem is solved completely: only successful transition events are generated while all others are neglected. Moreover, the formulation satisfies *both* conditions mentioned earlier. The relative probabilities among different successful paths remain unaltered and the absolute sampling probabilities are enhanced by a *known* amount, facilitating the computation of transition rates.

The main challenge of the method is the identification of an appropriate importance function, especially for problems involving large numbers of degrees of freedom. Yet, we find close resemblance between our search for the optimal importance function and the variational optimization of the ground state many-electron wave function in variational Quantum Monte Carlo (QMC) simulations.¹⁸ The latter involves the optimization of a trial wavefunction characterized by a set of parameters $\{\alpha\}$ with respect to an objective function, usually minimizing the energy or its variance, and has been successfully applied to large systems.^{18,19} Here, we show that an identical procedure can be used to find a suitable importance function for a rare-event problem. In this case, a trial importance function containing a set of parameters is optimized with respect to the variance in the statistical weights of the trajectories generated in the IS-MC simulation. To demonstrate the IS-MC methodology and associated numerical algorithms we apply the technique to two model problems.

The remainder of the paper has been organized as follows. In section II we discuss the theory behind the IS-MC path sampling formalism. Section III describes the details of the applications, including the used optimization algorithms and discusses the obtained results. We conclude the discussion in IV.

II. METHODOLOGY

A. The rare-event problem and Monte Carlo quadrature

Consider a classical N -particle system in which the interactions are described in terms of a potential-energy function $E(\mathbf{R})$, where the $3N$ -dimensional vector $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ specifies the microscopic configuration of the system. Let us assume that $E(\mathbf{R})$ is known to possess the metastable states A and B , as shown schematically in Fig. 1, and we are interested in computing the rate constant for transitions from region A to B .

The starting point of our approach is the fact that we can unambiguously slice a reactive trajectory, such as the one shown in Fig. 1, into a sequence of *failure* paths followed by a single *successful* path. Here, we define a failed path as a sequence of microstates that initiates in region A , exits it at some instant, but returns to it *before* reaching B . In contrast, a successful segment is defined as a sequence of states that initiates in A and succeeds in reaching B before returning to state A . The reactive trajectory shown in Fig. 1, for instance, consists of 3 failed paths, namely the sequences of states $0 \rightarrow 1$, $1 \rightarrow 2$, and $2 \rightarrow 3$, and the successful path $3 \rightarrow 4$. In this view, a transition event is considered rare if the expectation value of the number of failed paths observed before detecting a successful one is very large. In other words, a transition event is rare when the *success probability* p_s of sampling a successful path from some initial condition in A is very small.

The success probability p_s is the most fundamental quantity in this formulation of the rare-event problem because if it is known, the computation of the transition rate becomes straightforward. Specifically, the forward transition rate $k_{A \rightarrow B}$ is given by

$$k_{A \rightarrow B} = \frac{1}{\tau_w}, \quad (1)$$

where τ_w is the average waiting time before a transition from A to B occurs. This time can be easily computed when realizing that a reactive trajectory, on average, is expected to consist of $N_f = 1/p_s$ failed paths followed by a single successful one. Accordingly, the average waiting time becomes

$$\tau_w = \frac{\langle \tau \rangle_f}{p_s} + \langle \tau \rangle_s, \quad (2)$$

where $\langle \tau \rangle_f$ and $\langle \tau \rangle_s$ represent the average duration of failed and successful paths, respectively. The first term,

which represents the overwhelmingly dominant part of the waiting time, can be estimated from the simulation of an ensemble of failed paths. This, of course, can be done very accurately and efficiently given that the probability of sampling such failed paths is essentially equal to 1. The second term is significantly more challenging, however, given that p_s is small and a direct simulation will effectively not sample any successful paths. But this term is usually negligibly small compared to the first one and, in most cases, can be safely ignored.

How do we go about the calculation of p_s ? For this purpose, we consider the statistics of the ensemble of all possible successful and failed paths. Let us start by considering the ensemble of all possible paths (successful, failure, or neither²⁰) of a fixed length τ , specified by the sequences of microstates $\mathbf{R}(\tau) = (\mathbf{R}_0, \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_\tau)$. Let $P[\mathbf{R}(\tau)]$ be the corresponding distribution function, which in a typical MC random-walk simulation can be written in the form

$$P[\mathbf{R}(\tau)] = \rho(\mathbf{R}_0) \prod_{i=0}^{\tau-1} K(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1}), \quad (3)$$

where $\rho(\mathbf{R}_0)$ denotes the equilibrium distribution (e.g. canonical ensemble) of initial microstates \mathbf{R}_0 and the $K(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1})$ is a set of transition probabilities. $P[\mathbf{R}(\tau)]$ is properly normalized, i.e.

$$\int D\mathbf{R}(\tau) P[\mathbf{R}(\tau)] = 1, \quad (4)$$

where the notation¹³ $\int D\mathbf{R}(\tau)$ indicates a summation over all possible paths of length τ . Next, we wish to restrict this ensemble to the subset of successful and failed paths according to the definition adopted above. The probability distribution function of the restricted fail/success ensemble of paths with a length τ can then be written as

$$P_{\text{fs}}[\mathbf{R}(\tau)] = \frac{P[\mathbf{R}(\tau)](f_{\text{F}}[\mathbf{R}(\tau)] + f_{\text{S}}[\mathbf{R}(\tau)])}{Z_{\text{fs}}(\tau)}, \quad (5)$$

where $f_{\text{F}}[\mathbf{R}(\tau)]$ and $f_{\text{S}}[\mathbf{R}(\tau)]$ are characteristic path functions that indicate whether the path is a failure or a success:

$$f_{\text{F(S)}}[\mathbf{R}(\tau)] = \begin{cases} 1, & \text{if path is a failure (success)} \\ 0, & \text{otherwise} \end{cases}, \quad (6)$$

and the denominator is the partition function of the fail/success ensemble of paths with length τ ,

$$Z_{\text{fs}}(\tau) = \int D\mathbf{R}(\tau) P[\mathbf{R}(\tau)] (f_{\text{F}}[\mathbf{R}(\tau)] + f_{\text{S}}[\mathbf{R}(\tau)]). \quad (7)$$

To compute the success probability p_s , however, we need to remove the constraint of considering only paths with a fixed length and look at the ensemble of all possible fail/success paths of *any* length τ . The probability

of sampling the path $\mathbf{R}(\tau)$ from the *full* fail/success ensemble then becomes

$$P_{\text{FS}}[\mathbf{R}(\tau)] = \frac{P[\mathbf{R}(\tau)](f_{\text{F}}[\mathbf{R}(\tau)] + f_{\text{S}}[\mathbf{R}(\tau)])}{Z_{\text{FS}}}, \quad (8)$$

where the denominator

$$Z_{\text{FS}} = \int_0^\infty d\tau Z_{\text{fs}}(\tau) \quad (9)$$

is the partition function of the full fail/success ensemble. The success probability p_s is then given by the ensemble average of the characteristic path function f_{S} over the fail/success ensemble, i.e.

$$p_s = \int_0^\infty d\tau \int D\mathbf{R}(\tau) P_{\text{FS}}[\mathbf{R}(\tau)] f_{\text{S}}[\mathbf{R}(\tau)]. \quad (10)$$

In similar fashion, the failure probability p_f is given by

$$p_f = \int_0^\infty d\tau \int D\mathbf{R}(\tau) P_{\text{FS}}[\mathbf{R}(\tau)] f_{\text{F}}[\mathbf{R}(\tau)]. \quad (11)$$

so that P_{FS} is properly normalized, i.e.

$$p_s + p_f = 1. \quad (12)$$

In principle, the formulation of p_s in terms of the integral in Eq. (10) is ideally suited for standard MC quadrature¹⁷. In this method we sample N trajectories $\mathbf{R}^i(\tau)$ from P_{FS} and estimate the integral as the simple "hit-or-miss" arithmetic mean of the function $f_{\text{S}}[\mathbf{R}^i(\tau)]$,

$$p_s = \langle f_{\text{S}} \rangle_{P_{\text{FS}}} \approx f_{\text{S}}^N = \frac{1}{N} \sum_{i=1}^N f_{\text{S}}[\mathbf{R}^i(\tau)]. \quad (13)$$

Unfortunately, a direct application of this approach is hopelessly ineffective since f_{S} is nonzero only for a very small fraction of the fail/success ensemble. As a result, the procedure leads to very poor statistics, which is reflected by a large variance in the estimator f_{S}^N . More specifically, the quality measure¹⁷

$$\frac{\text{var}(f_{\text{S}}^N)}{\langle f_{\text{S}} \rangle_{P_{\text{FS}}}^2} = \frac{1}{N} \left(\frac{1 - p_s}{p_s} \right),$$

shows that the proportionality factor of the typical $1/N$ behavior for the variance of N -sample estimators¹⁷ is extremely large here since $p_s \approx 0$.

To resolve this problem we resort to a strategy that is frequently used to reduce the variance in MC quadrature calculations, namely *importance sampling*¹⁷. For this purpose, we rewrite Eq. (10) in the form

$$\begin{aligned} p_s &= \int_0^\infty d\tau \int D\mathbf{R}(\tau) \left[\frac{P_{\text{FS}}[\mathbf{R}(\tau)] f_{\text{S}}[\mathbf{R}(\tau)]}{\tilde{P}_{\text{FS}}[\mathbf{R}(\tau)]} \right] \tilde{P}_{\text{FS}}[\mathbf{R}(\tau)] \\ &= \int_0^\infty d\tau \int D\mathbf{R}(\tau) \tilde{f}_{\text{S}}[\mathbf{R}(\tau)] \tilde{P}_{\text{FS}}[\mathbf{R}(\tau)], \end{aligned} \quad (14)$$

where we have introduced a new path probability density function $\tilde{P}_{\text{FS}}[\mathbf{R}(\tau)]$ that satisfies the conditions

$$\tilde{P}_{\text{FS}}[\mathbf{R}(\tau)] \geq 0, \int_0^\infty d\tau \int D\mathbf{R}(\tau) \tilde{P}_{\text{FS}}[\mathbf{R}(\tau)] = 1. \quad (15)$$

Instead of generating paths according to P_{FS} and computing the path average of the function $f_S(\mathbf{R}(\tau))$, we now sample from the alternative distribution function \tilde{P}_{FS} and compute the trajectory average of the function $\tilde{f}_S[\mathbf{R}(\tau)]$. In this manner, p_s is estimated as

$$p_s = \langle \tilde{f}_S \rangle_{\tilde{P}_{\text{FS}}} = \tilde{f}_S^N \approx \frac{1}{N} \sum_{i=1}^N \tilde{f}_S[\mathbf{R}^i(\tau)]. \quad (16)$$

How do we choose \tilde{P}_{FS} ? The best-possible choice is the one for which the variance of the new estimator in Eq. (16) is minimized. Formally, it can be shown¹⁷ that the optimal path distribution function is given by

$$\tilde{P}_A^{\text{opt}}[\mathbf{R}(\tau)] = \frac{P_{\text{FS}}[\mathbf{R}(\tau)] f_S[\mathbf{R}(\tau)]}{p_s}, \quad (17)$$

for which this variance is zero and only successful paths are generated. As we will see below, the optimal path distribution function can, in principle, be determined explicitly. Unfortunately, this is practically feasible only for relatively simple problems involving few degrees of freedom.

Instead, the strategy we will adopt is to search for a \tilde{P}_{FS} that is “similar” to the optimal path distribution function in the sense that it leads to an acceptable low-variance estimator for p_s . In practice, this implies that we attempt to construct a probability distribution function \tilde{P}_{FS} for which the events that contribute significantly to the integral in Eq. (10), i.e. those that represent successful transition events, occur more frequently. The variance serves as the guiding principle in this process, representing an unambiguous quantitative quality measure. Conceptually, this approach is similar to the variational quantum MC approach^{18,19} in which the many-electron wave function of a system of interest is obtained by minimizing the variance in the energy for a given trial wave function. This variance is zero if and only if the trial wavefunction corresponds to the exact ground state. In the path importance sampling case the variance of interest is zero if and only if the optimal path distribution function $\tilde{P}_A^{\text{opt}}[\mathbf{R}(\tau)]$ is used.

B. Optimized importance sampling of Markovian transition paths

Let us now consider the procedure for constructing a suitable path probability distribution function \tilde{P}_{FS} . For this purpose, we now restrict the discussion to Markovian path simulations, which are based on a set of transition probabilities $K(\mathbf{R}_i \rightarrow \mathbf{R}_j)$ that describe the statistics of

transitions from microstate \mathbf{R}_i to \mathbf{R}_j and are properly normalized, i.e

$$n(\mathbf{R}_i) \equiv \sum_j K(\mathbf{R}_i \rightarrow \mathbf{R}_j) = 1. \quad (18)$$

For a system in contact with a heat reservoir at a constant temperature, for instance, the matrix $K(\mathbf{R}_i \rightarrow \mathbf{R}_j)$ may be constructed according to the Metropolis algorithm^{17,21}. The probability P_{FS} of sampling a given sequence of L states from the fail/success ensemble is then given by

$$P_{\text{FS}}[\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_L] = \frac{\rho(\mathbf{R}_0)}{Z_{\text{FS}}} \prod_{i=0}^{L-1} K(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1}), \quad (19)$$

with $\rho(\mathbf{R}_0)$ the canonical equilibrium distribution of initial microstates \mathbf{R}_0 in region A .

The goal of constructing \tilde{P}_{FS} is to reduce the probability of generating failed paths. To this end we apply two transformations to the original transition probabilities. The first is given by

$$K'(\mathbf{R}_i \rightarrow \mathbf{R}_j) = \begin{cases} 0, & \text{if } i \notin A \wedge j \in A, \\ K(\mathbf{R}_i \rightarrow \mathbf{R}_j), & \text{otherwise,} \end{cases} \quad (20)$$

Its purpose is to prohibit a trajectory from re-entering region A once it has exited. At first sight one might think this would eliminate failed paths altogether. But obviously this is not the case, given that transformation (20) introduces a set of states that no longer satisfy the normalization condition Eq. (18). More specifically, for all those states \mathbf{R}_i outside of region A which have nonzero matrix elements $K(\mathbf{R}_i \rightarrow \mathbf{R}_j)$ for states \mathbf{R}_j inside region A , we have

$$n'(\mathbf{R}_i) = \sum_j K'(\mathbf{R}_i \rightarrow \mathbf{R}_j) = \sum_{\mathbf{R}_j \notin A} K(\mathbf{R}_i \rightarrow \mathbf{R}_j) < 1. \quad (21)$$

How do we handle such undernormalized transition probabilities in a MC path simulation? In order to preserve the interpretation of K' as being a matrix describing transition probabilities, a renormalization is required. An appropriate way of doing so is to interpret undernormalization as the possibility for a path to be declared a failure, even when the current (undernormalized) state \mathbf{R}_i is outside of region A . The probability for this to occur is determined by the *degree of undernormalization*, defined as

$$m'(\mathbf{R}_i) = 1 - n'(\mathbf{R}_i) \quad (22)$$

In this manner, when the current state \mathbf{R}_i is undernormalized, the sampling procedure involves two stages. The first step determines whether the path is declared a premature failure and terminated, or if it is allowed to continue. The second step is carried out only if the latter

is the case, selecting the next microstate in the path. The first step involves the sampling of a random number ξ between 0 and 1, and if $\xi < m'(\mathbf{R}_i)$ the path is terminated. Otherwise, the next microstate in the path is sampled according to the nonzero elements of the transition probability matrix K' , but which have now been *renormalized*, i.e. $K'(\mathbf{R}_i \rightarrow \mathbf{R}_j)/n'(\mathbf{R}_i)$.

It is not difficult to see that transformation (20) does *not* affect the probabilities of sampling successful or failed segments. This is due to the fact that it causes only the states immediately adjacent to region A to be undernormalized and that their degree of undernormalization is precisely exactly equal to the sum of the transition probabilities to states \mathbf{R}_j inside A for the original transition probability matrix $K(\mathbf{R}_i \rightarrow \mathbf{R}_j)$. In other words, when in a state adjacent to region A , the probability for a path to fail due to undernormalization of the matrix K' is precisely equal to the probability of re-entering region A when using the original transition probability matrix K . Similarly, for a given successful path $\Gamma = (\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_L)$, the path probability can be written as

$$\begin{aligned} P'_{FS} &= \frac{\rho(\mathbf{R}_0)}{Z_{FS}} \prod_{i=0}^{L-1} \frac{(1 - m'(\mathbf{R}_i)) K'(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1})}{n'(\mathbf{R}_i)} \\ &= \frac{\rho(\mathbf{R}_0)}{Z_{FS}} \prod_{i=0}^{L-1} \frac{n'(\mathbf{R}_i) K'(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1})}{n'(\mathbf{R}_i)} \\ &= P_{FS}[\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_L], \end{aligned} \quad (23)$$

where the factors $(1 - m'(\mathbf{R}_i))$ represent the probability that a path is continued. Therefore, the sole difference between path simulations based on K and K' is that the *mechanism* for path failure has changed. Instead of re-entry into region A , the only mechanism for path failure while using the transition matrix K' is undernormalization.

The purpose of the second transformation now is to reduce the degree of undernormalization of K' . This is accomplished by means of an *importance function* $I(\mathbf{R})$, defined on the space of microstates accessible to the system, and which operates on the elements of matrix K' according to

$$\tilde{K}(\mathbf{R}_i \rightarrow \mathbf{R}_j) = K'(\mathbf{R}_i \rightarrow \mathbf{R}_j) \frac{I(\mathbf{R}_j)}{I(\mathbf{R}_i)}. \quad (24)$$

The best-possible alternative path distribution function \tilde{P}_A^{opt} referred to in the previous section is now obtained by using the *optimal importance function*, $I^{\text{opt}}(\mathbf{R})$, which completely eliminates the undernormalization of matrix K' and produces a new transition probability matrix that is fully normalized, i.e.

$$\begin{aligned} \tilde{n}(\mathbf{R}_i) &= \sum_j \tilde{K}(\mathbf{R}_i \rightarrow \mathbf{R}_j) \\ &= \sum_j K'(\mathbf{R}_i \rightarrow \mathbf{R}_j) \frac{I^{\text{opt}}(\mathbf{R}_j)}{I^{\text{opt}}(\mathbf{R}_i)} = 1 \end{aligned} \quad (25)$$

A path simulation based on matrix \tilde{K}_{opt} will therefore sample *only* successful paths, allowing an exact (i.e. zero-variance) determination of the success probability p_s .

Rewriting (25) in the form

$$\sum_j K'(\mathbf{R}_i \rightarrow \mathbf{R}_j) I^{\text{opt}}(\mathbf{R}_j) = I^{\text{opt}}(\mathbf{R}_i), \quad (26)$$

shows that the optimal importance function I^{opt} is the right-eigenvector with unit eigenvalue of the transition probability matrix K' obtained after transformation (20).

As mentioned above, the value of p_s can be directly inferred from the function I^{opt} . To see this, let us consider in which way the probability of sampling a successful path using matrix \tilde{K}_{opt} is altered relative to a simulation based on the original set of transition probabilities K . To this end, we compare the respective sampling probabilities P_{FS} and \tilde{P}_A^{opt} of a given successful sequence of states $\Gamma = (\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_L)$. The former is given by Eq. (19), whereas we have

$$\begin{aligned} \tilde{P}_{FS}^{\text{opt}}(\Gamma) &= \frac{\rho(\mathbf{R}_0)}{Z_{FS}} \prod_{i=0}^{L-1} \tilde{K}(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1}) \\ &= \frac{\rho(\mathbf{R}_0)}{Z_{FS}} \prod_{i=0}^{L-1} \left[K'(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1}) \frac{I^{\text{opt}}(\mathbf{R}_{i+1})}{I^{\text{opt}}(\mathbf{R}_i)} \right] \\ &= \left[\frac{\rho(\mathbf{R}_0)}{Z_{FS}} \prod_{i=0}^{L-1} K(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1}) \right] \frac{I^{\text{opt}}(\mathbf{R}_L)}{I^{\text{opt}}(\mathbf{R}_0)} \\ &= P_{FS}(\Gamma) \frac{I^{\text{opt}}(\mathbf{R}_L)}{I^{\text{opt}}(\mathbf{R}_0)}. \end{aligned} \quad (27)$$

for the alternative one. Note that in the third line of Eq. (27) we have substituted transition probability matrix K' by K since a path, once exiting region A , will never revisit it. Moreover, since all states are properly normalized, the MC sampling can be carried out without the need for adopting renormalization measures such as those in Eq. (23), and each path will represent a successful transition event. Eq. (27) shows that the probability of sampling a successful path is enhanced by a factor that depends only on the values of the optimal importance function I^{opt} in the initial and final microstates of the path. In addition, given that the optimal importance function can be shown to be constant for all the microstates in regions A and B , with values I_A^{opt} and I_B^{opt} , respectively, the sampling probability for *all* possible successful paths will be enhanced *uniformly* by the factor $I_B^{\text{opt}}/I_A^{\text{opt}}$. Eq. (14) then reveals the relation between the optimal importance function and the success probability p_s ,

$$\begin{aligned} p_s &= \int_0^\infty d\tau \int D\mathbf{R}(\tau) \left[\frac{P_{FS}[\mathbf{R}(\tau)] f_S[\mathbf{R}(\tau)]}{\tilde{P}_{FS}^{\text{opt}}[\mathbf{R}(\tau)]} \right] \tilde{P}_{FS}^{\text{opt}}[\mathbf{R}(\tau)] \\ &= \frac{I_A^{\text{opt}}}{I_B^{\text{opt}}} \int_0^\infty d\tau \int D\mathbf{R}(\tau) \tilde{P}_{FS}^{\text{opt}}[\mathbf{R}(\tau)] = \frac{I_A^{\text{opt}}}{I_B^{\text{opt}}}. \end{aligned} \quad (28)$$

C. Variational optimization of the importance function

In principle, the recipe for identifying the optimal importance function is given by Eqs. (20), (24) and (26). Unfortunately, an exact solution to the eigenequation (26) is feasible only in very simple cases. In realistic situations the numbers of degrees of freedom and accessible microstates in the system are usually too large to render an explicit solution possible. This situation resembles the problem of solving the Schrödinger equation for a many-body problem in quantum mechanics. This similarity naturally leads to the development of an approximate scheme for the identification of an appropriate importance function based on a standard approach utilized in quantum mechanics: the variational method.²²

In this scheme we select a trial function form for the importance function involving a set of free parameters. The parameters are then adjusted so as to bring the trial function as close as possible to the optimal importance function. The degree of normalization of the transition probability matrix \tilde{K} associated with the chosen trial function serves as the quality measure in this process. Specifically, since the optimal importance function is the one for which all microstates are normalized, the variational optimization process attempts to adjust the parameters such that the normalization factors of the microstates become as close to unity as possible. As in the variational method in quantum mechanics, the quality of the resulting importance function is only as good as the selected trial function form.

1. MC algorithm

To examine the practical aspects of the optimization process we consider the case in which we have selected a trial importance function of the form $I = I(\mathbf{R}, \{\alpha\})$ which involves a set of adjustable parameters $\{\alpha\}$. After constructing the normal transition probability matrix K of the system, we apply the transformations (20) and (24) to obtain the modified transition probability matrix $\tilde{K}(\{\alpha\})$, which now depends explicitly on the values of the adjustable parameters. The optimization of the parameters then proceeds by sampling paths using a MC simulation based on $\tilde{K}(\{\alpha\})$ and adjusting them such that Eq. (26) is satisfied as closely as possible.

However, as in the case of the truncated matrix K' , the sampling procedure is somewhat more elaborate due to the appearance of unnormalized states. In particular, in addition to the undernormalized states we have seen earlier, the matrix $\tilde{K}(\{\alpha\})$ will generally also contain *overnormalized* states \mathbf{R}_i for which

$$\tilde{n}(\mathbf{R}_i; \{\alpha\}) = \sum_{\mathbf{R}_j} \tilde{K}(\mathbf{R}_i \rightarrow \mathbf{R}_j; \{\alpha\}) > 1. \quad (29)$$

so that renormalization is required to preserve the interpretation of \tilde{K} as being a matrix describing a set of

transition probabilities.

This is accomplished by attributing *statistical weights* w to each path. The value of w varies along the sequence of microstates of the path and its current value determines how the sampling of the next state is carried out. At the initial microstate \mathbf{R}_0 of a path $\Gamma = (\mathbf{R}_0, \mathbf{R}_1, \dots)$, the weight is equal to unity, $w(\mathbf{R}_0; \Gamma) = 1$. The subsequent values of w are then found by multiplying the previous value by the current normalization factor. Accordingly, for the second microstate \mathbf{R}_1 in the path we have

$$w(\mathbf{R}_1; \Gamma) = w(\mathbf{R}_0; \Gamma) \tilde{n}(\mathbf{R}_1; \{\alpha\}), \quad (30)$$

and so forth for subsequent states. As in the case of K' , the path sampling algorithm based on $\tilde{K}(\{\alpha\})$ now involves two steps. The first consists of a check of the value of the path weight in the current state \mathbf{R}_i .

If $w(\mathbf{R}_i; \Gamma) < 1$ we allow for the possibility of the path to be declared a failure with a probability $\tilde{m}(\mathbf{R}_i; \Gamma) = 1 - w(\mathbf{R}_i; \Gamma)$. Drawing a uniform random number ξ between 0 and 1, the path is then terminated if $\xi < \tilde{m}(\mathbf{R}_i; \Gamma)$, otherwise it is allowed to continue. The second step, only in case of path continuation, then involves resetting the path weight to unity, i.e. $w(\mathbf{R}_i; \Gamma) = 1$, followed by selecting the next state according to the renormalized transition probability matrix elements of $\tilde{K}(\mathbf{R}_i \rightarrow \mathbf{R}_j; \{\alpha\})/\tilde{n}(\mathbf{R}_i)$.

If $w(\mathbf{R}_i; \Gamma) \geq 1$, however, we immediately proceed to the second step by selecting the next state according to the renormalized transition probability matrix, however *without* resetting the path weight $w(\mathbf{R}_i; \Gamma)$ to unity.

In this manner, for any given successful path $\Gamma = (\mathbf{R}_0, \mathbf{R}_1, \dots, \mathbf{R}_L)$, the final statistical weight is greater or equal to unity, $w(\Gamma) = w(\mathbf{R}_L; \Gamma) \geq 1$. More specifically, the final path weight $w(\Gamma)$ is given by

$$w(\Gamma) = \prod_{j=k+1}^L \tilde{n}(\mathbf{R}_j), \quad (31)$$

where \mathbf{R}_k is the last state along the path at which the weight was found to be smaller than unity. The corresponding sampling probability for the path can be shown to be

$$\begin{aligned} \tilde{P}_{\text{FS}}(\Gamma; \{\alpha\}) &= \frac{\rho(\mathbf{R}_0)}{Z_{\text{FS}}} \prod_{i=0}^{L-1} \frac{\tilde{K}(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1}; \{\alpha\})}{w(\Gamma)} \\ &= \left[\frac{\rho(\mathbf{R}_0)}{Z_{\text{FS}}} \prod_{i=0}^{L-1} \frac{K(\mathbf{R}_i \rightarrow \mathbf{R}_{i+1})}{w(\Gamma)} \right] \frac{I(\mathbf{R}_L; \{\alpha\})}{I(\mathbf{R}_0; \{\alpha\})} \\ &= \left(\frac{P_{\text{FS}}(\Gamma)}{w(\Gamma)} \right) \frac{I(\mathbf{R}_L; \{\alpha\})}{I(\mathbf{R}_0; \{\alpha\})}. \end{aligned} \quad (32)$$

2. Optimization criteria

Using the above simulation algorithm, we now need to optimize the parameter set $\{\alpha\}$. This process can be

guided by a variety of optimization criteria. One of them is based on the fact that the transition probability matrix \tilde{K} is optimized if, and only if, all states are properly normalized. In this view, if we define the quantity W as the simple product of all normalization factors in a path, disregarding whether it is successful or a failure, i.e.

$$W = \prod_{i=0}^{L-1} \tilde{n}(\mathbf{R}_i; \{\alpha\}), \quad (33)$$

the optimal importance function $I^{\text{opt}}(\mathbf{R})$ is obtained only if $W = 1$ for all paths. In this context, the optimization of the set of parameters $\{\alpha\}$ may proceed by minimizing the quality measure

$$Q_1(\{\alpha\}) \equiv (\ln W)^2. \quad (34)$$

Since Q_1 will be rigorously zero for all paths only for $I^{\text{opt}}(\mathbf{R})$, the optimized set of parameters $\{\alpha\}_{\text{opt}}$ for the selected function form can then be obtained by minimizing the path average

$$\langle Q_1(\{\alpha\}_{\text{opt}}) \rangle = \min_{\{\alpha\}} [Q_1(\{\alpha\})]. \quad (35)$$

A second criterion is based on the variance of the alternative estimator for p_s for the modified path probability Eq. (32). Using Eqs. (14), (16) and (32) it follows that, for a sample of N paths generated using $\tilde{P}_{\text{FS}}(\Gamma; \{\alpha\})$, p_s is estimated as the average

$$\begin{aligned} p_s &= \langle \tilde{f}_S \rangle_{\tilde{P}_{\text{FS}}} \approx \frac{1}{N} \sum_{i=1}^N w(\Gamma_i) f_S[\Gamma_i] \left(\frac{I(\mathbf{R}_{0(\Gamma_i)})}{I(\mathbf{R}_{L(\Gamma_i)})} \right) \\ &= \tilde{p}_s \frac{1}{N_s} \sum_{i=1}^{N_s} w(\Gamma_i) \left(\frac{I(\mathbf{R}_{0(\Gamma_i)})}{I(\mathbf{R}_{L(\Gamma_i)})} \right) \end{aligned} \quad (36)$$

where $\mathbf{R}_{0(\Gamma_i)}$ and $\mathbf{R}_{L(\Gamma_i)}$ are the initial and final states of the paths and $\tilde{p}_s = N_s/N$ with N_s the number of successful paths.

An effective evaluation of the success probability p_s through the estimator Eq. (36) requires that its variance be as small as possible. In this context, an alternative optimization criterion involves the minimization of the variance in the weights of *successful* paths, i.e.

$$Q_2(\{\alpha\}) \equiv \langle w^2 \rangle_s - \langle w \rangle_s^2. \quad (37)$$

3. Optimization algorithms

Once an optimization criterion has been selected, we need a minimization algorithm for the evolution of the parameter set in the optimization process. For this purpose a number of numerical approaches is available. Similar to variational quantum MC,^{18,19} the minimization can be carried out using simulated annealing and genetic algorithms for global optimization purposes as well as steepest-descent or conjugate-gradient algorithms for

further local refinement in the space of adjustable parameters. In the present work we rely mostly on a combination of a genetic algorithm and local steepest-descent minimization.

III. APPLICATIONS

To illustrate the practical operation of the presented importance sampling framework we consider its application to two simple model problems, a one-dimensional potential well and a two-dimensional system with two stable states separated by two distinct barriers.

A. One-dimensional potential well

First we consider the following simple one-dimensional system, described by the potential-energy function

$$V(x) = x^2 \exp(-0.5x^2), x \in [-2, 5] \quad (38)$$

shown in Fig. 2. For convenience we discretize the domain into the set of states $x_i = -2 + (i-1)\Delta x$, with $\Delta x = 0.1$ and $i = 1, 2, \dots, 71$. The system contains a stable minimum at $x = 0$ and we are interested in transitions that cross the barrier and reach the state $x = 5$ at a temperature $k_B T = 0.2585$ using the Metropolis MC algorithm. For this purpose, we define the fail and success regions A and B according to the individual states $x_A = 0$ and $x_B = 5$, respectively. The conventional Metropolis transition probability matrix for this system is given as

$$K(x_i \rightarrow x_j) = \begin{cases} \frac{1}{2} \min[1, \exp(-\Delta V_{ij}/k_B T)], j = i \pm 1 \\ 1 - \sum_{j=i \pm 1} K(x_i \rightarrow x_j), j = i \end{cases} \quad (39)$$

with $\Delta V_{ij} = V(x_j) - V(x_i)$. Proper boundary conditions are adopted at the limits $x = -2$ and $x = 5$ of the domain, allowing trial steps to the left or right only.

The first step in the importance sampling scheme is to implement the transformation defined in Eq. (20), eliminating transitions that take the system from a state outside A to one within A , having

$$K'(x_i \rightarrow x_j) = \begin{cases} 0, \text{ if } x_i \neq x_A \wedge x_j = x_A \\ K(x_i \rightarrow x_j), \text{ otherwise} \end{cases} \quad (40)$$

Based on Eq. (26), and given the simplicity of the problem we can explicitly determine the optimal importance function I^{opt} by identifying the right-eigenvector with eigenvalue 1 of matrix K' . The resulting eigenvector, normalized such that its value in the success state $x_B = 5$ equals 1, is plotted in Fig. 3. The plot shows that the value of the importance function in the fail

region is very close to zero, indicating that the probability of sampling a successful transition event, as defined in Eq. (28), is extremely low. More specifically, Fig. 4, which shows $I^{\text{opt}}(x)$ on a logarithmic scale, reveals that the probability of sampling a successful transition is $p_s = I^{\text{opt}}(x_A)/I^{\text{opt}}(x_B) = 5.843 \cdot 10^{-6}$ for the system and temperature under consideration. Note that the plot only covers the states $x \geq 0$ since $I^{\text{opt}}(x) = 0$ for all states $x < 0$, implying that any path starting in such a state fails. This, of course, is a consequence of the low dimensionality of the problem, forcing the path to pass through the fail state $x = x_A$ in order to reach x_B .

Let us now re-examine the problem using the variational method outlined above and compare the results to the ones obtained from the exact matrix diagonalization method. The first element in the application of the variational method involves the choice of a function form for the importance function. As shown in Fig. 4, it should be able to describe a variation of several orders of magnitude between the regions of failure and success. A simple functional form that allows such variations is the following exponential of a Gaussian

$$I(x; A, \alpha, x_0) = \exp \left\{ A \exp \left[-\alpha(x - x_0)^2 \right] \right\}, \quad (41)$$

in which the three parameters A , α and x_0 describe the height, width and center position of the Gaussian, respectively.

Using this trial functional form we utilize a genetic optimization algorithm²³ for the minimization of the quality factor Q_1 in Eq. (34). Each iteration in the algorithm involves a generation consisting of a population of 30 distinct parameter sets. Initially, the sets are generated randomly, with $A \in [-20, 0]$, $\alpha \in [0, 2]$ and $x_0 \in [-1, 1]$. For each parameter set we generate a series of $N = 500$ paths using the respective modified matrices \tilde{K} and measure the corresponding average of the quality measure Q_1 ,

$$\langle Q_1(A, \alpha, x_0) \rangle = \frac{1}{N} \sum_i Q_1(\Gamma_i; A, \alpha, x_0), \quad (42)$$

where the summations runs over *all* generated paths Γ_i , both successful and failed. In addition we also compute the path averages of the *derivatives* of Q_1 with respect to the 3 parameters, $\langle \partial Q_1 / \partial A \rangle$, $\langle \partial Q_1 / \partial \alpha \rangle$ and $\langle \partial Q_1 / \partial x_0 \rangle$, which represent generalized forces and can be used in a local steepest-descent minimization scheme. Based on the estimators obtained from the 500 trajectories for each parameter set, the members of the population are then ranked according to increasing value of Q_1 . Next, a new generation of parameter sets is generated by first carrying out one steepest-descent step for all 30 members using the measured generalized forces. Next, we add 6 more members to the population; three correspond to a random perturbation of the 3 fittest parameter sets (i.e., those with the lowest Q_1 -values), and the other 3 are chosen completely randomly. Using the new population of 36 members, 500 MC trajectories are generated to determine the new Q_1 -values. The members are again ranked

according to increasing Q_1 , after which the worst 6 parameter sets are discarded. This procedure is repeated until convergence, which is reached after the genetic evolution produces a generation in which all members are essentially identical.

Fig. 5 shows the evolution of the genetic optimization algorithm along a cycle of 17 generations. The plot shows the values of the parameters A and α for the 30 population members for three different generations. Panel a) shows the initial population generated randomly on the intervals specified above. Panel b) shows the population after the 14-th iteration, showing a clear clustering. The process may be considered converged after generation 17, for which all 30 members of the population are essentially the same, as shown in panel c). The parameter values for the best set are $A = -12.19599$, $\alpha = 0.191672$, and $x_0 = -0.398851$, respectively, which gives the importance function shown as the line in Fig. 4 and $\langle Q_1(A, \alpha, x_0) \rangle \approx 1.39$, which implies an average path normalization factor of $W \approx 4.01$.

Even though the importance function resulting from the variational optimization procedure is quite different from the optimal one, a MC simulation based on it already shows good efficiency. Running a batch of $1 \cdot 10^6$ path simulations, the modified success probability is measured to be $\tilde{p}_s = 0.211 \pm 0.001$, so that approximately 1 in every 5 attempts leads to a successful transition event, with an the average success path weight $w_s = 3.61 \pm 0.07$ and a variance $Q_2 = \langle w^2 \rangle_s - \langle w \rangle_s^2 \approx 100$. The success probability p_s in the original system is then estimated according to Eq. (36), which, with $I(\mathbf{R}_0(\Gamma_i))/I(\mathbf{R}_L(\Gamma_i)) = 7.6253 \cdot 10^{-6}$ for the optimized parameter set, gives $p_s = (5.8 \pm 0.1) \cdot 10^{-6}$, in good agreement with the exact result. The simulation also provides an estimator for the average successful path length,

$$\langle \tau \rangle_s = \frac{1}{N_s} \frac{1}{\langle w \rangle_s} \sum_{i=1}^{N_s} w(\Gamma_i) L(\Gamma_i), \quad (43)$$

giving $\langle \tau \rangle_s = (3.79 \pm 0.01) \cdot 10^2$ MC steps.

Once the success probability p_s and average successful path length are known, the computation of the forward transition rate constant k_{AB} becomes straightforward, requiring only the average length of failed paths. The latter is easily found by simulating a series of trajectories according to the unbiased transition probability matrix K . From a set of 10^3 paths we obtain $\langle \tau \rangle_f = (9.78 \pm 0.01)$. Using Eq. (1) and (2) we estimate the average waiting time and rate as $\tau_w = (1.68 \pm 0.03) \cdot 10^6$ MC steps and $k_{AB} = (5.9 \pm 0.1) \cdot 10^{-7}$ per MC step. These results are in good agreement with the estimates obtained from the 10^3 direct path MC simulations in which the average waiting time and the corresponding transition rate were found to be $\tau_{w,\text{direct}} = (1.71 \pm 0.06) \cdot 10^6$ MC steps and $k_{AB} = (5.8 \pm 0.2) \cdot 10^{-7}$ per MC step.

B. Two-dimensional case

As a second application, we consider rare transition events in the two-dimensional potential-energy surface

$$V(x, y) = \frac{1}{6} \{4[1 - x^2 - y^2]^2 + 2[x^2 - 2]^2 + [(x + y)^2 - 1]^2 + [(x - y)^2 - 1]^2 - 2\} + 0.02y, \quad (44)$$

which is shown in the contour plot of Fig. 6. The model contains two metastable states A and B at $(x, y)_A = (-1.1, 0)$ and $(x, y)_B = (1.1, 0)$ with energies $V(A) = V(B) = -0.0812$, separated by two distinct barriers S_1 and S_2 at $(x, y)_{S_1} = (0, 1)$ and $(x, y)_{S_2} = (0, -1)$ with energies $V(S_1) = 1.02$ and $V(S_2) = 0.98$, respectively.

We discretize the space of accessible states in terms of a two-dimensional square grid with spacing $\Delta x = \Delta y = 0.1$ on the domain $x, y \in [-1.5, 1.5]$. As in the one-dimensional problem, we define the fail and success regions as the individual states A and B , respectively. The MC dynamics of the system is governed by the Metropolis algorithm in which only the four first-neighbor states are directly accessible. Considering the system at a temperature $k_B T = 0.08617$, the explicit diagonalization of the truncated transition matrix K' yields the optimal importance function $I^{\text{opt}}(x, y)$, shown in Fig. 7, and the associated success probability $p_s = 6.154 \cdot 10^7$.

We now re-examine the problem using the variational optimization approach. Once again we adopt a simple exponential of a Gaussian as our trial function form for the importance function, writing

$$I(x, y) = \exp \{ A \exp [-\alpha(x - x_0)^2 - \beta(y - y_0)^2] \}, \quad (45)$$

which involves the set of 5 adjustable parameters A , α , β , x_0 and y_0 .

As in the one-dimensional problem, we utilize a genetic algorithm²³ and start by optimizing the parameter set with respect to quality measure Q_1 . The algorithm is essentially the same as the one used in the one-dimensional problem. In each generation we considered a population of 40 parameter sets. Initially, the 40 sets are generated randomly, with $A \in [-50, 0]$, α and $\beta \in [0, 4]$, x_0 and $y_0 \in [-2, 2]$. For each parameter set, a series of 10^3 MC path simulations was carried out to measure the average value of Q_1 over both failed and successful paths. The members of the population are then ranked according to increasing value of Q_1 after which a new generation is produced by carrying out one steepest-descent step for all 40 members using the measured generalized forces. As before, we add 10 more members to the population; three correspond to a random perturbation of the 5 parameter sets with the lowest Q_1 -value, and the remaining 5 are chosen completely randomly. Using the new population of 50 members, 10^3 MC trajectories are generated to determine the new Q_1 -values. The members are again ranked according to increasing Q_1 , after which the worst 10 parameter sets are discarded.

Convergence requires around 20 generations, after which the 40 population members in a generation are essentially the same, reaching an average value of $Q_1 \approx 0.84$ or an average path normalization factor $W \approx 2.3$, and a modified success probability of $\tilde{p}_s \approx 0.53$ at generation 20. The associated fittest parameter set is given by $A = -22.497462$, $\alpha = 2.449118$, $\beta = 0.550239$, $x_0 = -1.143219$, and $y_0 = -0.804378$.

From the insight gained in the one-dimensional problem, one might expect that the low Q_1 -value and high success probability \tilde{p}_s , are indicative of the parameter set's suitability for an accurate calculation of p_s . Analysis of the obtained importance function according to quality measure Q_2 , however, shows that is not the case here. In contrast to the 1-dimensional problem, the parameter set optimized with respect to Q_1 leads to an extremely large variance in the path weights w of successful paths. Specifically, a simulation based on 10^6 trajectories reveals that the average weight of a successful path is $\langle w \rangle_s \approx 2 \times 10^3$, while its variance is more than 6 orders of magnitude larger, at $\langle w^2 \rangle_s - \langle w \rangle_s^2 \approx 4 \times 10^9$. As a result, the importance function optimized with respect to Q_1 , despite its elevated success probability \tilde{p}_s , is unreliable for an accurate estimation of p_s .

Therefore, we repeat the above optimization process, but now searching for parameter sets that lead to low values of quality measure Q_2 . To this end we employ the same genetic algorithm used above, but now ranking the population members according to Eq. (37). Convergence is again reached after approximately 20 generations, producing the parameter set $A = -14.646743$, $\alpha = 1.882146$, $\beta = 0.845562$, $x_0 = -1.089187$, and $y_0 = 0.621533$. With this result the average successful path weight becomes $\langle w \rangle_s \approx 1.22$ with a variance $\langle w^2 \rangle_s - \langle w \rangle_s^2 \approx 5.6$ and an average success probability $\tilde{p}_s \approx 1.21 \cdot 10^{-2}$. Running a batch of 10^6 path simulations using this parameter set, we estimate the success probability in the original system to be $p_s = (3.8 \pm 0.2) \cdot 10^{-7}$.

While the statistical quality of this estimate is good, its value is significantly lower than the exact result obtained by explicit matrix diagonalization. The reason for this becomes clear after analyzing the successful trajectories generated by the optimized importance function. It shows that only one of the two reactive mechanisms is sampled, with all reactive paths passing through barrier S_2 . Accordingly, the obtained estimator for p_s corresponds only to the probability $p_s(S_2)$ of observing a reactive event associated with mechanism S_2 and thus underestimates the total success probability.

The failure of the above importance function to sample both mechanisms is due to the fact that the position of its minimum is closer to saddle point S_2 . In this view, an improved importance function may be obtained by forcing its minimum to coincide with the position of the fail state A . For this purpose we repeat the genetic optimization algorithm, but allowing only the parameters A , α and β to vary, while $x_0 = x_A$ and $y_0 = y_A$ are held fixed. Convergence is reached after 30 gener-

ations, resulting in the parameter set $A = -10.285377$, $\alpha = 3.601954$, $\beta = 1.554864$, with $x_0 = -1.1$ and $y_0 = 0$, with an average successful path weight $\langle w \rangle_s \approx 1.0061$ and variance $\langle w^2 \rangle_s - \langle w \rangle_s^2 \approx 3.66 \cdot 10^{-3}$ at a success probability $\tilde{p}_s \approx 1.8 \cdot 10^{-2}$. The drastic reduction of the variance in the weight of successful paths is indicative of the improvement of the obtained importance function compared to the previous one. This is reflected in the fact that the importance function now samples both transition mechanisms, as is illustrated by the distribution of y -values of the reactive paths as they cross $x = 0$ shown in Fig. (8). As a result, the estimated success probability $p_s = (6.18 \pm 0.04) \cdot 10^{-7}$, as measured from a series of 10^6 path simulations, is now in excellent agreement with the exact result $p_s = 6.15 \cdot 10^{-7}$ obtained by explicit matrix diagonalization. The relative contributions due to both mechanisms, as obtained from the distribution function in Fig. (8), are $p_s(S_1) = (2.49 \pm 0.02) \cdot 10^{-7}$ and $p_s(S_2) = p_s - p_s(S_1) = (3.69 \pm 0.02) \cdot 10^{-7}$, respectively, the latter of which is in good agreement with the result obtained using the previous importance function.

The difference between the last two optimization results is indicative of difficulties with the optimization procedure. While the first attempt could, in principle, have converged to the last importance function, it encountered a worse local minimum of Q_2 , leading to the sampling of only one of the available reactive mechanisms. This demonstrates that an indication of good statistical quality alone is not always sufficient to guarantee an accurate estimate for p_s . On the other hand, this problem is not a particular flaw of the present importance sampling approach. Rather, it is a problem of global optimization in general, in which one can never guarantee that an obtained result actually corresponds to the global minimum of the problem at hand. Moreover, other rare-event techniques are also sensitive to similar issues. The TPS method, for instance, may suffer ergodicity problems when the Markov chain of transition paths becomes dependent on the initial transition path in systems where two or more competing saddle points are separated by a high barrier. In the context of the present approach, this issue calls for the application of robust global optimization strategies including more elaborate genetic algorithms than the one utilized here and simulated annealing techniques. In this light, the importance sampling approach will benefit from the continuing efforts in the development of effective global optimization techniques²³.

IV. SUMMARY

The importance sampling method presented in this work represents a novel approach to the rare-event problem within the framework of the statistical mechanics of Markovian system trajectories. The starting point of our approach is the fact that one can unambiguously slice a reactive trajectory from state A to B into a sequence of

failed paths followed by a single successful path. The former is defined as a sequence of microstates that initiates in region A , exits it at some instant, but returns to it *before* reaching B , while a successful path is defined as a sequence of states that initiates in A and succeeds in reaching B before returning to state A . In this formulation, the probability p_s of sampling a successful path from some initial condition in A is the most fundamental a transition event because if it is known the calculation of the transition rate becomes straightforward.

The idea of the importance sampling method now is to compute the success probability using a “hit-or-miss” Monte Carlo quadrature approach, in which one samples a series of trajectories from the fail/success ensemble and counts the fraction of successful paths. However, if no special measures are adopted this approach is doomed to fail since the probability of generating successful paths is typically very low, leading to very poor statistics through a large variance.

The strategy of the method then is to bias the sampling of system paths so as to favor the generation of successful paths. This is accomplished by means of an importance function, which affects the transition probabilities used to sample the sequence of microstates that comprise a path. The advantage of using an importance function is that it contains all quantitative information needed to determine the amount by which the sampling probability of a given path has been biased. Formally, there exists an optimal importance function for which the bias is optimized and unsuccessful paths are suppressed entirely, generating only reactive transition events. In this manner, the rare-event problem has been transformed into an optimization problem, that of identifying the best-possible importance function. In practice, however, the identification of the optimal importance function is feasible only in simple problems involving few degrees of freedom. On the other hand, the optimal importance function obeys a variational principle, which, similar to the variational method for determining a many-body ground state wave function in quantum mechanics, provides a systematic approach toward finding a suitable approximation. The scheme is based on the choice of a trial function form containing a set of adjustable parameters for the importance function. The parameters are then adjusted so as to bring the chosen trial function as close as possible to the optimal importance function.

In the two model applications, the variational approach has demonstrated to be effective in generating good approximations to the optimal importance function, allowing an accurate and efficient evaluation of the success probabilities and the associated transition rates. In principle, the importance sampling scheme should be widely applicable, providing a generic framework for quantifying the statistics of rare-event processes in atomistic simulations. The two main challenges now are to deduce guidelines for the construction of suitable trial function forms for rare-event processes involving multiple degrees of freedom and to evaluate different minimization strate-

gies for the optimization of the corresponding sets of adjustable parameters. However, in light of the conceptual similarity with the variational MC method in quantum mechanics, which has been successfully applied to complex many-electron systems, this task may be faced with confidence.

Acknowledgments

M.K. gratefully acknowledges financial support from the Brazilian agency Fundação de Amparo à Pesquisa

do Estado de São Paulo under contract no. 03/01243-2. This work was performed under the auspices of the U. S. Department of Energy by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48. W.C. is supported by the University Relationship Program at LLNL.

-
- ¹ D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, San Diego, 1996).
 - ² S. Goedecker, *Rev. Mod. Phys.* **71**, 1085 (1999).
 - ³ F. F. Abraham (2002).
 - ⁴ P. Hänggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990).
 - ⁵ L. R. Pratt, *J. Chem. Phys.* **85**, 5045 (1986).
 - ⁶ R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987).
 - ⁷ M. Nastar, V. V. Bulatov, and S. Yip, *Phys. Rev. B* **53**, 13521 (1996).
 - ⁸ H. Jónsson, *Annu. Rev. Phys. Chem.* **51**, 623 (2000).
 - ⁹ A. F. Voter, *Phys. Rev. Lett.* **78**, 3908 (1997).
 - ¹⁰ A. F. Voter, *J. Chem. Phys.* **106**, 4665 (1997).
 - ¹¹ P. Bolhuis, C. Dellago, and D. Chandler, *Faraday Discussion* **110**, 421 (1998).
 - ¹² C. Dellago, P. G. Bolhuis, and D. Chandler, *J. Chem. Phys.* **110**, 6617 (1999).
 - ¹³ C. Dellago, P. G. Bolhuis, and P. L. Geissler, *Adv. Chem. Phys.* **123**, 1 (2002).
 - ¹⁴ P. G. Bolhuis, C. Dellago, P. L. Geissler, and D. Chandler, *Annu. Rev. Phys. Chem.* **53**, 291 (2002).
 - ¹⁵ J. Ramirez and M. Laso, *J. Chem. Phys.* **115**, 7285 (2001).
 - ¹⁶ W. Cai, M. H. Kalos, M. de Koning, and V. V. Bulatov, *Phys. Rev. E* **66**, 046703 (2002).
 - ¹⁷ M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986).
 - ¹⁸ W. M. C. Foulkes, L. Mitás, R. J. Needs, and G. Rajagopal, *Rev. Mod. Phys.* **73**, 33 (2001).
 - ¹⁹ P. R. C. Kent, R. J. Needs, and G. Rajagopal, *Phys. Rev. B* **59**, 12344 (1999).
 - ²⁰ A path is considered neither a failure nor a success if it does not satisfy any of the two definitions. An example of such a path is a sequence of microstates that initiates outside of region *A*.
 - ²¹ M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).
 - ²² A. Messiah, *Quantum Mechanics* (Dover, New York, 1999).
 - ²³ A. K. Hartmann and H. Rieger, *Optimization Algorithms*

in Physics (Wiley-VCH, Berlin, 2001).

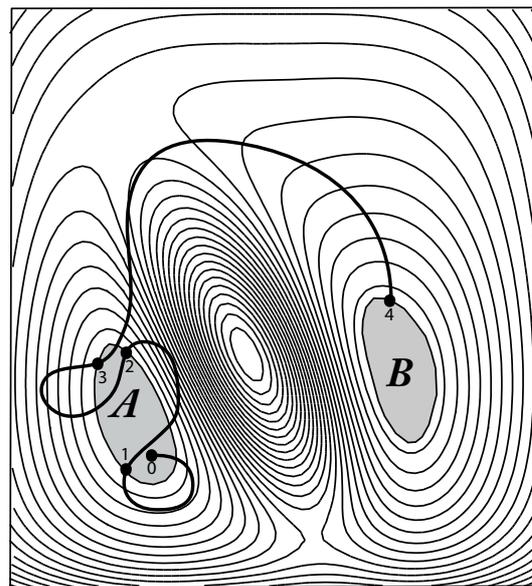


FIG. 1: Schematic representation of a reactive trajectory in an energy landscape $E(\mathbf{R})$ with the (meta)stable states *A* and *B*. Such a trajectory can be unambiguously sliced into a sequence of *failure* paths followed by a single *successful* path. A failed path is defined as a sequence of microstates that initiates in region *A*, exits it at some instant, but returns to it *before* reaching *B*. In contrast, a successful segment is defined as a sequence of states that initiates in *A* and succeeds in reaching *B* before returning to state *A*. The shown reactive trajectory consists of 3 failed paths, namely the sequences of states $0 \rightarrow 1$, $1 \rightarrow 2$, and $2 \rightarrow 3$, and the successful path $3 \rightarrow 4$.

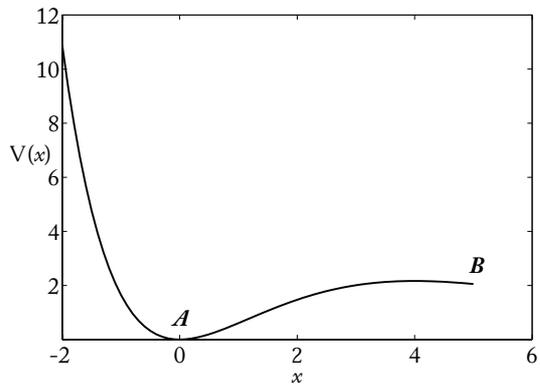


FIG. 2: Potential-energy surface of Eq. (38). The fail and success regions *A* and *B* are defined as $x_A = 0$ and $x_B = 5$, respectively.

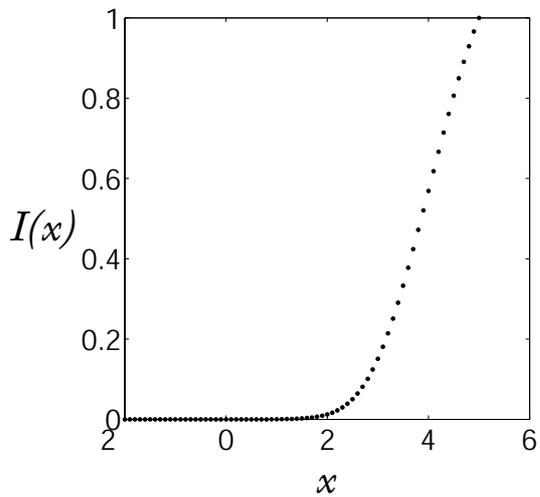


FIG. 3: Optimal importance function for the one-dimensional escape problem at a temperature $k_B T = 0.2585$.

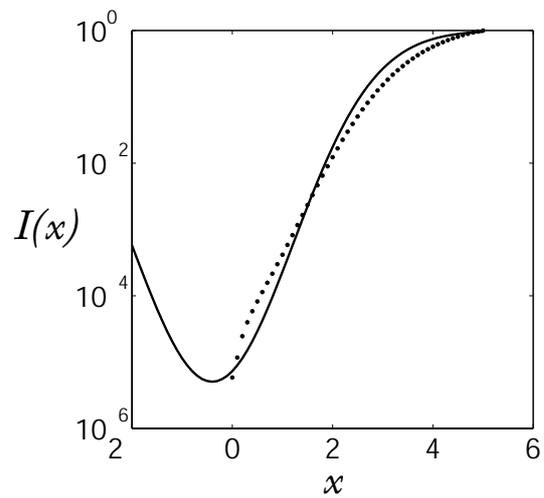


FIG. 4: Optimal importance function (circles) and result of genetic minimization algorithm (line) using trial function form Eq. (41) for the one-dimensional escape problem at a temperature $k_B T = 0.2585$. Values of parameters are $A = -12.19599$, $\alpha = 0.191672$, and $x_0 = -0.398851$. For the optimal importance function values are shown only for states $x \geq 0$ because $I^{\text{opt}}(x) = 0$ for all states $x < 0$.

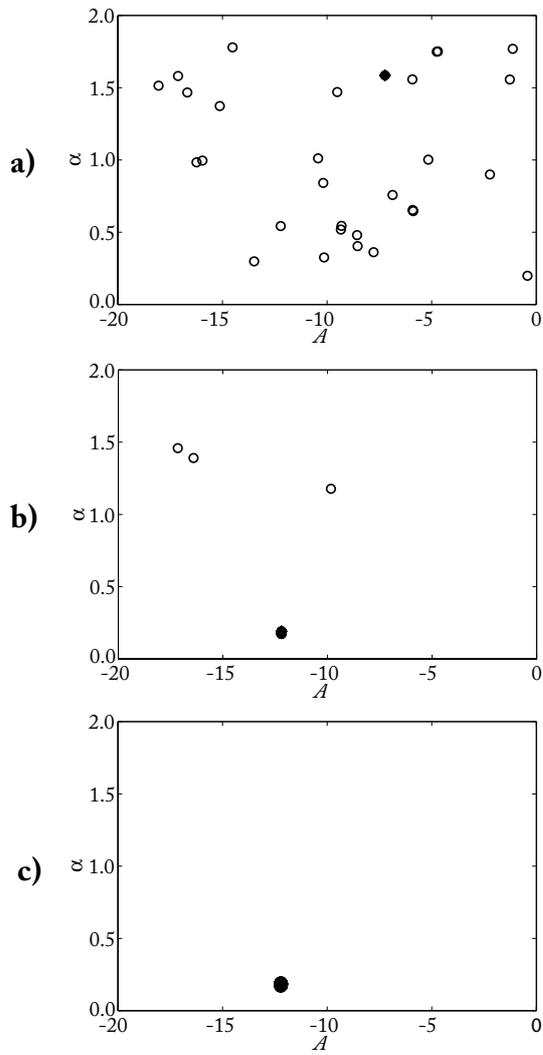


FIG. 5: Evolution of genetic algorithm in the optimization of the parameters A and α in the trial function form Eq. (41). Panel a) shows values of the 30 parameter sets before the first iteration. Panel b) shows distribution of population members in the 14-th generation. Panel c) shows result after 17 generations.

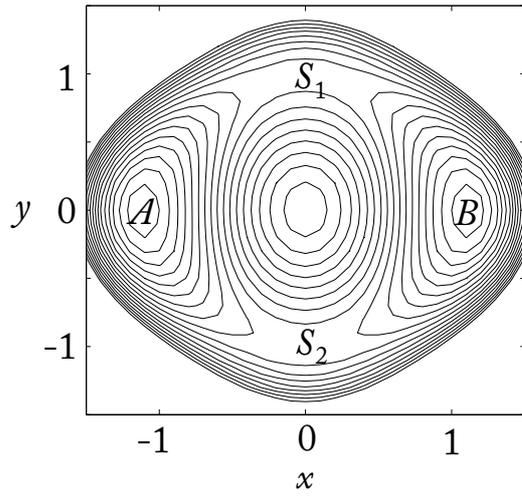


FIG. 6: Potential-energy contours of 2-dimensional system described by Eq. (44).

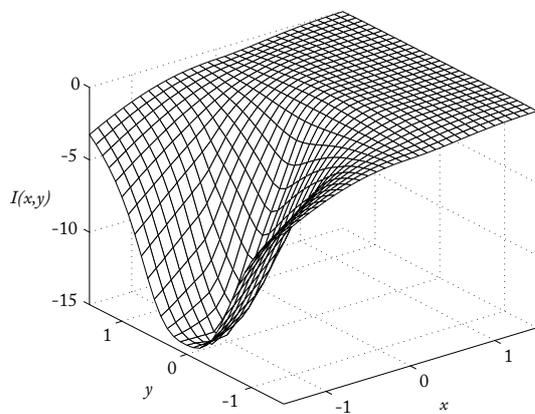


FIG. 7: The natural logarithm of the optimal importance function $I^{\text{opt}}(x, y)$.

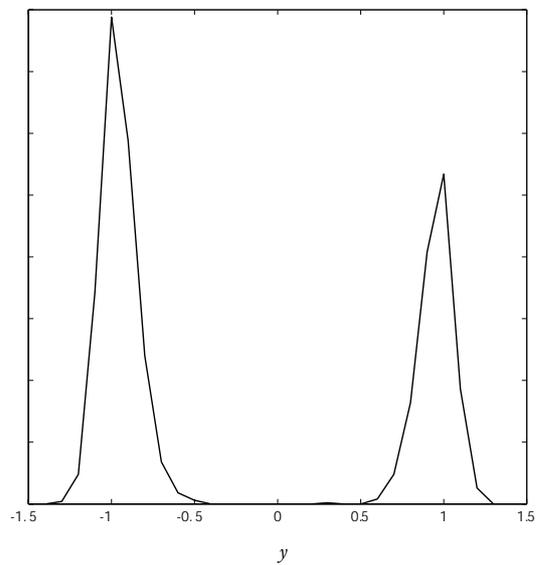


FIG. 8: Distribution of y -values of the reactive trajectories at $x = 0$ as sampled with the optimized importance function with fixed x_0 and y_0 .