Importance sampling of rare transition events in Markov processes

Wei Cai, Malvin H. Kalos, Maurice de Koning, and Vasily V. Bulatov
Lawrence Livermore National Laboratory, University of California, Livermore, California 94550
(Received 9 May 2002; published 17 October 2002)

We present an importance sampling technique for enhancing the efficiency of sampling rare transition events in Markov processes. Our approach is based on the design of an importance function by which the absolute probability of sampling a successful transition event is significantly enhanced, while preserving the relative probabilities among different successful transition paths. The method features an iterative stochastic algorithm for determining the optimal importance function. Given that the probability of sampling a successful transition event is enhanced by a known amount, transition rates can be readily computed. The method is illustrated in one- and two-dimensional systems.

DOI: 10.1103/PhysRevE.66.046703 PACS number(s): 02.70.−c, 02.50.−r, 05.10.−a, 45.10.−b

I. INTRODUCTION

In many situations of interest, the evolution of a Markov process is characterized by a series of rare transitions. In this case, the direct application of Monte Carlo simulation techniques becomes very inefficient, since the vast majority of the CPU cycles are spent on local moves, which, most of the time, do not contribute to the overall evolution of the system under consideration.

An illustrative example of this problem is the kinetic Monte Carlo (kMC) simulation of dislocation motion by the kink-pair mechanism [1–3]. In this approach, a perfect dislocation is modeled as a straight line positioned on a two-dimensional lattice, as shown in Fig. 1(a). Under the influence of an externally applied stress, the dislocation line lying along the horizontal direction moves in the vertical direction by means of the creation of kink pairs, which are rectangular-shaped excitations that move a part of the dislocation line to the next lattice position in the vertical direction, as shown in Fig. 1(b). In this manner, the dislocation motion occurs by the nucleation of embryonic kink pairs of unit width, followed by their expansion along the dislocation line, as shown in Figs. 1(c)–1(e). A direct kMC simulation of this process would involve the generation of a stochastic sequence of the three elementary processes in the model: nucleation of embryonic kink pairs, their expansion along the dislocation line, and possible kink annihilation events. The efficiency of such a direct kMC approach, however, is extremely low given that embryonic kink pairs are very unstable and have a large tendency to recombine among themselves. Only after a kink pair has reached a certain critical width i_c, does it become stable against recombination and able to contribute to the overall dislocation motion. Accordingly, a direct kMC simulation will spend most of the CPU cycles on series of local moves that involve a nucleation event immediately followed by a recombination event, whereas the important critical-width kink pairs are only rarely sampled.

This particular problem has been solved [1–3] by modifying the event catalog such that only the nucleation of critical-width kink pairs is considered, while the explicit evolution of embryonic kink pairs leading to such critical-width states is removed from consideration. While successful, this specific approach provides a solution only for the particular problem of kMC simulation of dislocation motion. Yet, the problem of efficient kMC simulations of Markov processes controlled by rare events is much more general. For instance, standard displacement MC simulation of a system whose evolution is governed by activated processes suffers from the same kind of problem. In this case, the local MC moves sample mostly ''unimportant'' states near the deep energy basins of the potential-energy landscape, while the important states near the energy barriers are visited only very occasionally. Given their importance in general, it would be of considerable interest to develop a general approach to the efficient simulation of rare transition events in Markov processes.

In this paper we present a general *importance sampling* [4] framework for this purpose. The foundation of the method is based on the fact that the low efficiency of direct MC simulations is a result of the extremely low probabilities of generating sequences of states, or *paths*, that lead to successful transition events. Within this context, the importance sampling approach seeks to modify the direct MC sampling algorithm in such a manner that the absolute probability of the successful sequences is enhanced, while also preserving their relative probabilities. Furthermore, the importance sampling method enhances the absolute probability of generating a successful path by a *known* amount, such that the evaluation of transition rates is straightforward. The scheme involves the design of a suitable *importance function* that controls the manner in which the sampling of local MC moves is

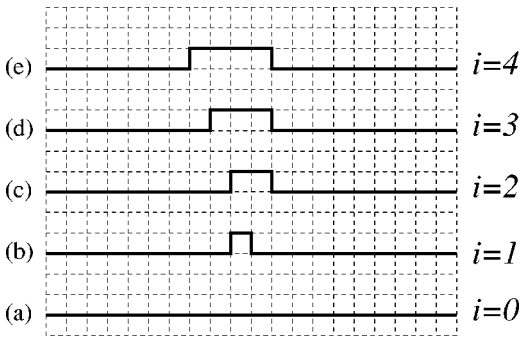

FIG. 1. Dislocation motion by kink-pair nucleation and growth.

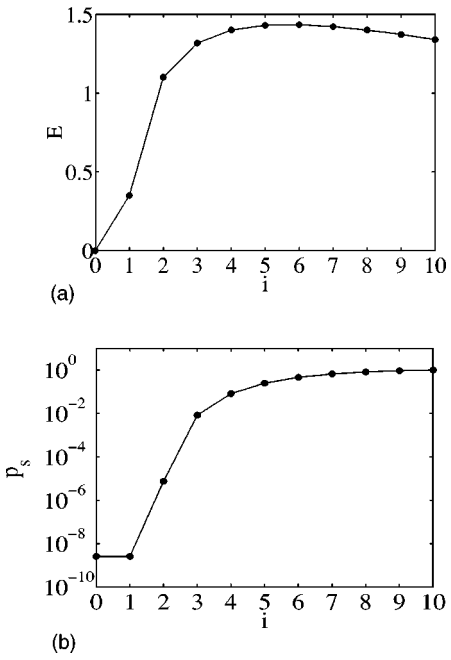

FIG. 2. (a) A typical energy landscape of kink-pair nucleation growth-recombination process under external stress represented by a one-dimensional random walk. Random walks reaching kink-pair width $N=10$ are counted as successful. (b) Survival probability function $p_s(i)$ obtained by direct diagonalization of the transition matrix $\tilde{K}_{ij}=K(i\rightarrow j)$ (for $j\neq 0$, $\tilde{K}_{i0}=0$).

modified and determines by how much the absolute probability of generating a successful path is changed. We show that there exists an optimal importance function for which this probability becomes equal to unity and develop an iterative algorithm for finding a numerical approximation to this function.

The paper is organized as follows. In Sec. II the importance sampling framework is developed in the context of the one-dimensional kink pair nucleation problem mentioned earlier. Section III describes results obtained in further applications of the proposed methodology to relatively simple one- and two-dimensional problems. Section IV concludes the discussion.

II. METHODOLOGY

A. Kink-pair nucleation and growth problem: One-dimensional random walk

Consider the kink-pair nucleation and growth problem from the preceding section. The configuration of a kink pair is modeled in terms of its width *i*, measured in units of lattice spacing along the dislocation line. Accordingly, the various states of a growing kink are enumerated by $i=0,1,2,3,\ldots$, where $i=0$ corresponds to a perfectly straight dislocation line with no kinks present, $i=1$ corresponds to an embryonic kink pair of unit width, $i=2$ is a kink pair of width 2, and so on (see Fig. 1). A typical energy landscape $E(i)$ under the influence of external stress is shown in Fig. 2(a). Throughout the paper energies are measured in arbitrary units. For narrow kink pairs the attractive interaction, which tends to restore the perfectly straight dislocation line, is dominant. After reaching a certain critical width, in this case around $i_c=6$, the kink pair becomes stable against annihilation and its expansion under the influence of external stress becomes energetically favorable.

A direct kMC simulation of the process of embryonic kink-pair formation and migration involves a series of one-dimensional random walkers that initiate from the perfect dislocation configuration, state $i=0$. The walkers continue until they either return to state $i=0$, in which case the perfectly straight dislocation line is restored, or reach a certain width $i=N$, with $N\geqslant i_c$, in which case the kink pair has overcome the attractive forces. The transition probabilities $K(i\rightarrow j)$ for $i=1,\cdots,N-1$ are

$$K(i\rightarrow i-1)=\frac{r_-(i)}{r_-(i)+r_+(i)},$$

$$K(i\rightarrow i+1)=\frac{r_+(i)}{r_-(i)+r_+(i)},$$

$$K(i\rightarrow j)=0, \quad j\neq i+1,i-1, \tag{1}$$

with

$$r_-(i)\equiv\exp\left[-\frac{E(i-1)-E(i)}{2k_BT}\right],$$

$$r_+(i)\equiv\exp\left[-\frac{E(i+1)-E(i)}{2k_BT}\right], \tag{2}$$

and $K(0\rightarrow 1)=1$ for the initial state. Within this definition, the walker jumps either to the right or to the left at a given step, but cannot remain in its current state [5].

A direct simulation based on these transition probabilities is extremely inefficient. This can be seen from Fig. 2(b), which shows the survival probability $p_s(i)$ that a walker which initiates in a state i reaches the state $i=10$ before annihilating at a temperature $k_BT=0.069$ (computed in Sec. III). Since $p_s(0)\sim 10^{-9}$, on average, only one out of every 10^9 random walkers that start from the nucleation of an embryonic kink pair will lead to a relevant event of the formation of a kink pair with a width $i=10$.

One solution to this problem is to change the event catalog in such a manner that the nucleation of embryonic kink pairs is removed from the description and is replaced by the formation of kink pairs with a width $i=N$. In order to guarantee that the statistical properties remain unaffected, one determines the effective nucleation rate of such wider kink pairs, which can be done in a variety of ways [3]. In the present paper we adopt a different approach to the problem. Instead of modifying the space of accessible states, that is, removing narrow kink pairs from description, we develop a methodology that enhances the efficiency of the simulation by modifying the transition probabilities in Eqs. (1) and (2), that is, we leave the system and its kinetics unchanged but alter the Monte Carlo evolution. The purpose of this modification is to bias the generation of random walks in such a way that the absolute probability of sampling a successful path (i.e. one that reaches $i=N$) is increased, while keeping

the relative probability of different successful random walks unaltered. In addition, the modification of the transition probabilities is accomplished in a controlled manner, such that the absolute probability is enhanced by a known amount and an effective nucleation rate of the kink pairs of width $i=N$ can be readily computed.

B. Importance sampling of the paths

Let $\Gamma_{0\alpha}=(0,i_1,i_2,\ldots,i_{L-1},\alpha)$ be the sequence of states visited along a random walk, or *path*, of length L that initiates in state $i=0$ and ends in state α, with either $\alpha=0$ (failure) or $\alpha=N$ (success). In the specific case of kink-pair nucleation and growth, we are particularly interested in the probability of generating a successful path since this probability determines the effective rate of kink-pair formation. In addition, the distribution function for the length L of the successful paths is useful because it allows one to compute the time for a successful embryonic kink pair to reach the critical width. Let $f(L)$ be the statistical distribution of the length L of *all* paths. Considering the two possible outcomes of a random-walk simulation, $f(L)$ can be written as

$$f(L)=P_s f_s(L)+(1-P_s)f_f(L), \tag{3}$$

where $P_s\equiv p_s(0)$ is the probability of generating a successful path starting from $i=0$, $f_s(L)$ is the length distribution function of successful paths, and $f_f(L)$ is that for failed paths. Since we are interested only in the first term on the right-hand side of Eq. (3), a direct sampling based on the transition probabilities (1) and (2) is hopelessly inefficient, given that the success probability $P_s\sim 10^{-9}$.

The probability of generating a path $\Gamma_{0\alpha}$ starting in state $i=0$ and ending in state $i=\alpha$ is given by the product of the respective transition probabilities $K(i\to j)$,

$$\mathcal{P}(\Gamma_{0\alpha})=K(0\to i_1)K(i_1\to i_2)\cdots K(i_{L-1}\to\alpha). \tag{4}$$

The success probability P_s and associated length distribution $f_s(L)$ are then defined as

$$P_s=\sum_{\Gamma_{0N}}\mathcal{P}(\Gamma_{0\alpha}) \tag{5}$$

and

$$f_s(L)=\frac{1}{P_s}\sum_{\Gamma_{0N}}\mathcal{P}(\Gamma_{0\alpha})\,\delta_{\hat{L}(\Gamma_{0\alpha}),L}, \tag{6}$$

where the summation is over all successful paths Γ_{0N} of any length, $\hat{L}(\Gamma_{0\alpha})$ is the length of path $\Gamma_{0\alpha}$, and $\delta_{\hat{L},L}$ is the Kronecker delta. The probabilities are normalized as follows:

$$\sum_{\Gamma_{00}}\mathcal{P}(\Gamma_{0\alpha})+\sum_{\Gamma_{0N}}\mathcal{P}(\Gamma_{0\alpha})=1. \tag{7}$$

Now, for each state i let us define an *importance function* $I(i)$ and use it to modify the transition probabilities $K(i\to j)$ according to

$$K'(i\to j)=K(i\to j)\cdot\frac{I(j)}{I(i)}\quad(\text{for } j\neq 0),$$

$$K'(i\to 0)=0, \tag{8}$$

for $i=0,\ldots,N-1$. Since only the ratio $I(j)/I(i)$ is relevant, we may fix the value of the importance function in state $i=N$ at an arbitrary constant. For convenience we set $I(N)=1$. Using the modified transition probabilities, the probability Eq. (4) for generating a successful path Γ_{0N} becomes [4]

$$\mathcal{P}'(\Gamma_{0N})=K'(0\to i_1)K'(i_1\to i_2)\cdots K'(i_{L-1}\to N)$$

$$=\mathcal{P}(\Gamma_{0N})\frac{I(N)}{I(0)}=\mathcal{P}(\Gamma_{0N})/I(0), \tag{9}$$

in which all fractions $I(j)/I(i)$ from Eq. (8) have canceled except for $I(N)/I(0)$. Since the probability of generating *any* successful path is modified by the same factor $I(N)/I(0)$, the relative ratio of the probabilities of two successful paths is not altered. The modified success probability P_s' and length distribution function $f_s'(L)$ become

$$P_s'=\sum_{\Gamma_{0N}}\mathcal{P}'(\Gamma_{0\alpha})=P_s\frac{I(N)}{I(0)}, \tag{10}$$

and

$$f_s'(L)=\frac{1}{P_s'}\sum_{\Gamma_{0N}}\mathcal{P}'(\Gamma_{0\alpha})\,\delta_{\hat{L},L}=f_s(L). \tag{11}$$

Equations (9)–(11) show that a simulation based on the modified transition probabilities $K'(i\to j)$ allows one to modify the absolute probability of sampling a successful path by a factor $I(N)/I(0)$, while leaving the relative probabilities of different successful paths and the corresponding length distribution $f_s'(L)=f_s(L)$ unaltered. In this way, by selecting a suitable importance function $I(i)$, one can significantly enhance the simulation efficiency.

The implementation of a random-walk simulation based on the modified transition probabilities (8), however, is somewhat more involved than that for the original system. The reason is related to the normalization of the transition probabilities. While the original transition probabilities in Eq. (1) are properly normalized at each site i,

$$n(i)=\sum_j K(i\to j)=1, \tag{12}$$

the sums

$$n'(i)=\sum_j K'(i\to j)=\sum_{j\neq 0}K(i\to j)\frac{I(j)}{I(i)} \tag{13}$$

may no longer satisfy this condition. In order to handle such non-normalized transition probabilities, we adopt a procedure in which we assign variable *statistical weights* w' to each path and allow for the possibility that a path is declared

unsuccessful and terminated in *any* state i. In contrast, in a simulation based on the original transition probabilities Eq. (1) each path has unit weight and is allowed to terminate only when it is currently in state $i=1$ or $N-1$.

An algorithm for simulating random walkers based on the modified transition probabilities in Eq. (8) is as follows.

Function. Random walk generation according to transition probabilities $K'(i\to j)$.

Input. Modified transition probabilities $K'(i\to j)$.

Output. Success measure s ($s=0$ if the path is a failure, $s=w'$ if the path is successful, where w' is the statistical weight of the path).

(1) Initialize $i:=0$, $w':=1$.

(2) Compute normalization, $n'(i)=\Sigma_j K'(i\to j)$.

(3) Update weight, $w':=w'n'(i)$.

(4) If $w'\geqslant 1$ go to 6.

(5) Draw a random number $\xi\in U[0,1]$. If $\xi>w'$ then $s:=0$ and exit. Otherwise reset $w':=1$.

(6) Select the next state j according to the normalized transition probabilities $K'(i\to j)/n'(i)$. Update $i:=j$.

(7) If $i=0$ then $s:=0$ and exit. If $i=N$ then $s:=w'$ and exit. Otherwise go to 2.

The algorithm initiates the random walk in state $i=0$ and assigns unit weight to it, $w'=1$. In the next step, the normalization factor $n'(i)$ in Eq. (13) is computed for the current state, after which the statistical weight w' for the path is updated by multiplying it by the current normalization factor. Only when the current state is properly normalized, i.e., $n'(i)=1$, will the weight of the path remain unaltered. Otherwise, it will either decrease if $n'(i)<1$, or increase in case $n'(i)>1$. Based on the current value of the weight w', the sampling for the next step may then proceed in two different ways. If $w'\geqslant 1$, the sampling proceeds by selecting either state $i-1$ or $i+1$ according to the normalized transition probabilities $K'(i\to i-1)/n'(i)$ and $K'(i\to i+1)/n'(i)$. In case $w'<1$, however, we allow for the possibility that the walker is terminated at state i with a probability $1-w'$. If this occurs, the path is ended and declared unsuccessful. Otherwise, its weight is reset to unity and the selection of the next state proceeds in the way described above. As output, the algorithm provides the success measure s of the path. It is equal to zero if the path is unsuccessful, while $s=w'$ in case the walker reaches the state $i=N$.

Following this algorithm, the probability P'_s of sampling a successful path and the corresponding length distribution $f'_s(L)$ are estimated as averages over the success measures $s(\Gamma^k_{0\alpha})$ along a series of M random walkers $\Gamma^k_{0\alpha}$,

$$\overline{P'_s}=\frac{1}{M}\sum_{k=1}^{M} s(\Gamma^k_{0\alpha}) \tag{14}$$

and

$$\overline{f'_s(L)}=\frac{\frac{1}{M}\sum_{k=1}^{M} s(\Gamma^{(k)}_{0N})\,\delta_{L_k,L}}{\overline{P'_s}}, \tag{15}$$

where L_k is the length of path k. Using Eqs. (10) and (11), these results then give estimators for the success probability P_s in the original system

$$P_s\approx\overline{P'_s}\frac{I(0)}{I(N)}, \tag{16}$$

and the length distribution function, $f_s(L)\approx\overline{f'_s(L)}$.

C. Optimizing the importance function

According to Eq. (10) the probability P'_s for sampling a successful random walk depends only on the ratio $I(0)/I(N)$ of the initial and final values of the importance function. Therefore, *any* function $I(i)$ that satisfies the given boundary conditions for this ratio gives the *same* expectation values for the success probability P'_s and length distribution function $f'_s(L)$. However, since the purpose of the simulation is the estimation of P'_s and $f'_s(L)$, we must rely on their calculation by means of the averages, Eqs. (14) and (15). While the expectation values are determined only by the value $I(0)/I(N)$, the *variances* depend on the function $I(i)$ at the intermediate points. In this sense there exists an *optimal* importance function $I_{\rm opt}(i)$, for which these variances are minimized.

The variances are closely related to the degree of normalization of the modified transition probabilities $K'(i\to j)$. Unless the transition probabilities are normalized at every state, the sampling procedure will generate successful paths with weights larger than or equal to unity and other paths that are prematurely terminated. This will give rise to finite variances in Eqs. (14) and (15). For instance, the variance in the estimator $\overline{P'_s}$ is given by

$$\begin{aligned}\sigma\frac{2}{\overline{P'_s}}&=\frac{1}{M}\sum_{k=1}^{M}[s(\Gamma^k_{0\alpha})-P'_s]^2\\&=\frac{1}{M}\sum_{k=1}^{M_s}[w'(\Gamma^k_{0N})-P'_s]^2+\frac{1}{M}\sum_{k=1}^{M_{\rm f}}(P'_s)^2,\end{aligned} \tag{17}$$

where M_s and M_f are the numbers of successful and failed random walks, respectively. Equation (17) shows that the minimum variance is achieved if $P'_s=1$, $M_f=0$ and the weights of all successful paths are equal to unity. In this situation, both the first and the second terms on the right-hand side of Eq. (17) vanish and $\sigma(2/P'_s)=0$.

In this manner, the optimized importance function $I_{\rm opt}(i)$ is the one for which the $K'(i\to j)$ are normalized at all states $i=0,\ldots,N-1$,

$$\sum_j K'(i\to j)=\sum_{j\neq 0} K(i\to j)\frac{I_{\rm opt}(j)}{I_{\rm opt}(i)}=1, \tag{18}$$

with $I_{\rm opt}(N)=1$.

In order to determine $I_{\rm opt}(i)$, it is useful to rewrite Eq. (18) in the form

$$I_{\mathrm{opt}}(i)=\sum_{j\neq 0} K(i\rightarrow j)I_{\mathrm{opt}}(j), \tag{19}$$

for $i=0,1,\ldots,N-1$ and $I_{\mathrm{opt}}(N)=1$. It follows that the function $I_{\mathrm{opt}}(i)$ is the right eigenvector of the transition probability matrix $\tilde{K}_{ij}=K(i\rightarrow j)$ (for $j\neq 0$ and $\tilde{K}_{i0}=0$) with unit eigenvalue. The physical interpretation of this function becomes clear after recognizing that the survival probability function $p_s(i)$, defined as the probability that a random walk initiating in state i reaches state N before annihilating in the original system [see Sec. II A and Fig. 2(b)], satisfies the same recursive relation,

$$p_s(i)=\sum_{j\neq 0} K(i\rightarrow j)\, p_s(j), \tag{20}$$

where $p_s(0)=P_s$ and $p_s(N)=1$. Accordingly, the value of the optimal importance function $I_{\mathrm{opt}}(i)$ at state i is equal to the survival probability function $p_s(i)$ in the system described by the original transition probabilities.

The success probability function $p_s(i)$ can be obtained by direct diagonalization of the transition probability matrix $\tilde{K}_{ij}$. This approach has been adopted in several earlier methods for increasing simulation efficiency [6,7]. Given that $I_{\mathrm{opt}}(i)$ is the eigenvector corresponding to the largest eigenvalue of $\tilde{K}_{ij}$ [8], the optimal importance sampling function can be obtained by repetitive multiplication by the transition probability matrix. Setting initial values $I^1(i)=1$ for all states i and updating the importance function values according to the iterative scheme

$$I^{k+1}(i)=\sum_{j\neq 0} K(i\rightarrow j)I^k(j) \quad \text{for } i=0,1,\ldots,N-1,$$

$$I^k(N)=1 \quad \text{for all } k, \tag{21}$$

progressively suppresses all eigenvectors with eigenvalues smaller than unity, retaining only the desired eigenvector with eigenvalue 1.

While this scheme can be easily implemented for systems characterized by a limited number of states i, it will no longer be practical for high-dimensional systems involving large numbers of degrees of freedom. In this case even the mere storage of the importance function $I(i)$ for all accessible states i may pose an enormous challenge. In this situation one should focus on finding an importance function that is close to the optimal one for the ''relevant'' states in the problem. This can be achieved by combining the iterative procedure (21) with a series of random walk simulations. In this approach, one starts with a series of simulations based on the original transition probabilities and records a list of the states i that are visited during a sequence of M random walks. Next, one updates the importance function $I(i)$ and transition probabilities $K'(i\rightarrow j)$ [by Eqs. (21) and (8)] *only* for the states i in this list, leaving all other values unaltered. Using the updated transition probabilities, one carries out another series of random-walk simulations and records a new list of visited states, for which the importance function and transition probabilities are updated. This iterative procedure is repeated several times, until the success probability P_s' reaches a reasonable level, preferably close to unity, and the importance function becomes sufficiently close to the optimal importance function for the relevant states in the problem. In terms of the theory of rare events [9], the relevant states may be interpreted as those most closely involved in the transition of interest.

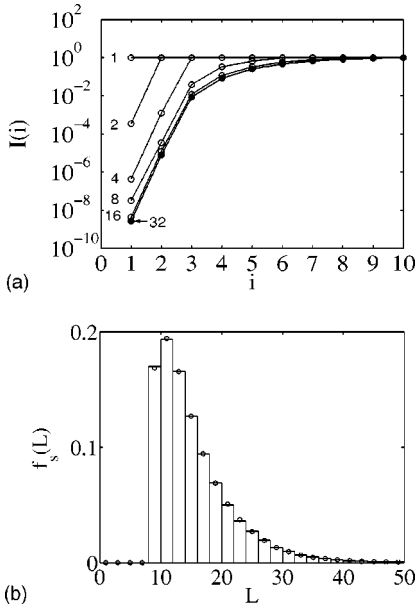

FIG. 3. (a) Convergence of importance function $I(i)$ towards the optimal solution $p_s(i)$. The numbers in the figure give the number of iterations for each curve. $I(0)$ is always equal to $I(1)$ and is not plotted. (b) Length distribution $f_s(L)$ of successful paths on the one-dimensional random walk by importance sampling of 10^5 *successful* paths, using the optimal importance function from the exact solution (bars) and an importance function improved iteratively ($\bigcirc$).

Having determined a suitable importance function using either of the two procedures described above, one proceeds by measuring the relevant statistical estimators from a series of random-walk simulations based on the modified transition probabilities. Using Eqs. (10) and (11), the results can then be translated into the desired estimators for the properties of the original system.

III. RESULTS

A. One-dimensional random walk

As a first case study we apply the importance sampling method to one-dimensional random walks on the energy profile of Fig. 2(a) at a temperature $k_BT=0.069$. We define the success state at $i=N=10$. Direct evaluation of the eigenvector of the transition probability matrix $K(i\rightarrow j)$ with unit eigenvalue gives the exact optimal importance function $I_{\mathrm{opt}}(i)$, as shown by the filled circles in Fig. 3(a). Given that $I_{\mathrm{opt}}(i)$ is equal to $p_s(i)$, the survival probability of walkers initiating in state $i=0$ is of the order of only 10^{-9}, meaning that a direct simulation based on the original transition probabilities will be ineffective. Notice that $p_s(0)=p_s(1)$, since $K(0\rightarrow 1)=1$.

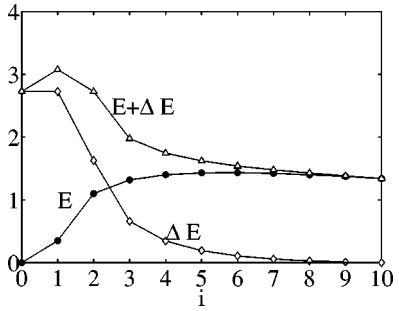

FIG. 4. The bias potential $\Delta E(i) = -2k_B T \ln p_s(i)$ ($\diamond$) corresponding to the optimal importance function. The original potential energy $E(i)$ ($\bullet$) and the resulting modified potential $E(i) + \Delta E(i)$ (Δ) are also shown.

The open circles in Fig. 3(a) show the results of the iterative algorithm used to obtain an estimate for the optimal importance function. The initial condition corresponds to the case where the transition probabilities are not altered, i.e. $I^1(i) = 1$ for all states i. Using the corresponding transition probabilities, a series of 200 random-walk simulations is carried out and a list of the visited states i is compiled. Next, the importance function is updated only for the states in this list, and a new series of 200 random-walk simulations based on the updated estimate for the importance function is conducted. Although after the third iteration the importance function has been adjusted only for the states $i = 0$ and $i = 1$, the survival probability has already increased by a factor $\sim 10^6$. Subsequent iterations further refine the importance function, adjusting it on an increasing number of states. After 32 iterations, the importance function has essentially converged and differences with the exact solution are no longer visible on the scale of the figure. In addition, the length histogram $f_s(L)$ of successful paths as measured from a series of 10^5 random walks using the importance function obtained after 32 iterations is essentially identical to the exact one. This is shown in Fig. 3(b), where the circles denote the data compiled using the iterative importance function and the bars represent the results based on the exact importance function.

It may be useful to interpret the optimal importance function in terms of associated effective energy bias on the original random-walk system. According to Eqs. (1) and (2), the transition probabilities $K(i \to j)$ are given in terms of the energies $E(i)$ of the different kink-pair configurations i. Similarly, the modified transition probabilities $K'(i \to j)$ may be associated with a modified energy landscape $E'(i)$,

$$E'(i) = E(i) + \Delta E(i), \tag{22}$$

where the bias potential $\Delta E(i)$ is related to the importance function $I(i)$ according to

$$\Delta E(i) = -2k_B T \ln I(i). \tag{23}$$

This association becomes exact when the modified transition probability $K'(i \to j)$ is normalized everywhere, i.e., when the importance function is optimal.

Figure 4 shows $E'(i)$ and $\Delta E(i)$ for the optimal importance function $I_{\mathrm{opt}}(i)$, as well as the original potential-energy landscape. $\Delta E(i)$ approximately inverts the potential landscape near $i = 0$, transforming the attractive energy minimum in the original landscape into a repulsive maximum. While a walker in the original system is likely to remain confined to the energy basin of state $i = 0$, the walkers that belong to the successful path ensemble rapidly escape to the destination state $i = N$ as if they were repelled from the origin $i = 0$.

This observation suggests a simple approach to construct a suitable importance function. In principle, one could use a bias function that inverts the original energy landscape. The inversion results in a left-right reflection of the transition probability, i.e., $K'(i \to i-1) = K(i \to i+1)$ and $K'(i \to i+1) = K(i \to i-1)$. This transformation preserves the normalization at each state and, for the one-dimensional random-walk problem leads to a success probability P_s' close to unity. For problems in higher dimensions, however, this simple inversion approach is generally not appropriate, since the optimal bias potential-energy will show approximate inversion of the energy function only along certain directions in the potential energy landscape, i.e. those involved in the reaction coordinates of the transition event under consideration (see the following section).

B. Two-dimensional random walk

As a second application we use the importance sampling method to investigate the statistics of random walks on a two-dimensional square lattice described by the energy function

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

defined on the domain $x,y \in [-1.5,1.5]$ and with lattice spacings $\Delta x = \Delta y = 0.1$. Figures 5(a) and 5(b) show three-dimensional and contour plot representations of the energy landscape on the specified domain, respectively. The energy function has two energy minima at the positions N_0: $(x = -1.1, y = 0)$ and N_1: $(x = 1.1, y = 0)$ with energies $V(N_0) = V(N_1) = -0.0812$, respectively. The two energy basins are separated by two saddle points at S_A: $(x = 0, y = 1)$ and S_B: $(x = 0, y = -1)$ with energies $V(S_A) = 1.02$ and $V(S_B) = 0.98$.

In this problem, the transition probability matrix for a given state i has four nonzero entries corresponding to the states immediately below and above, and right and left of state i. We now consider the simulation of random walks initiating in state N_0 and ending in state N_1. The exact optimal importance function $I_{\mathrm{opt}}(i,j)$, determined by directly computing the eigenvector with unit eigenvalue for $k_B T = 0.043$, is shown in Fig. 6. A direct simulation will be ineffective since the probability of generating a successful random walk starting in N_0 using the original transition probabilities is only $\sim 10^{-12}$.

Figure 7 shows the results of the importance function obtained using the iterative procedure described in the previous two sections. The result shown in the figure was obtained after a series of 1000 iterations, each of which involved a

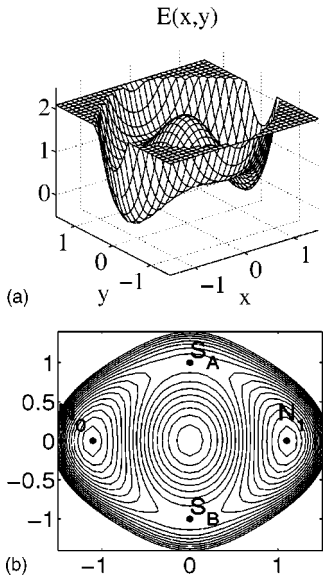

FIG. 5. (a) The energy landscape $E(x,y)$ for two-dimensional random walks. For illustrative purposes, energies higher than 2 are truncated in the plot, but not in the actual calculation. (b) Contour of $E(x,y)$.

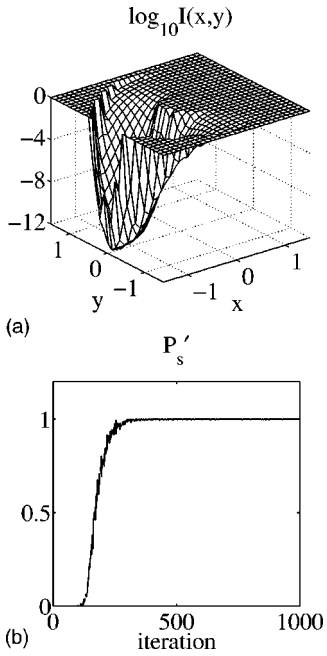

FIG. 7. (a) Importance function for the 2D random walk optimized by the stochastic algorithm after 1000 iterations. (b) Convergence of the success probability P_s' to 1 with the number of iterations.

series of 200 random walks generated using the current importance function modified after each iteration. Comparing this result to the exact optimal importance function shown in Fig. 6, it is evident that there are significant differences between the two, in particular, near the edges of the domain. This is due to the fact that the iterative procedure updates the values of the importance function only for those states that are visited during the simulation. The regions near the edges of the domain are only very rarely visited, even if one were to use the exact optimal importance function for the sampling, so that the convergence of the importance function for these states is relatively slow. Fortunately, however, the iterative procedure is required only to provide a good approximation to the optimal importance function for those states that are important for the transition under study. This is reflected in Fig. 7(b), which shows the success probability P_s' as a function of iteration step. While there are clear differences between the exact optimal importance function shown and the one obtained after 1000 iterations, the probability for generating a successful random walk becomes essentially equal to unity after 500 iterations. This indicates that, although quite different in irrelevant regions, the iterative approximation is very close to the exact importance sampling for those states that are effectively involved in the transitions between N_0 and N_1.

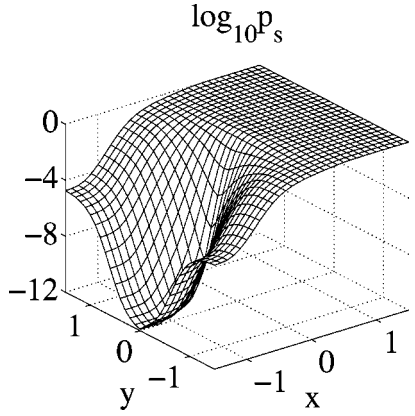

FIG. 6. Survival probability function $p_s(i)$, i.e., the optimal importance function for the 2D random-walk system.

Using both the iterative function (after 1000 iterations) and the exact solution, we now construct the length distribution function for the paths connecting N_0 and N_1. Given that the energy landscape contains two distinct saddle-point regions, we can identify two distinct transition mechanisms depending on whether the path passes through the saddle region around S_A or around S_B. In this manner we may write the total length distribution function $f_s(L)$ as

$$f_s(L)=\frac{P'_{s,A}f'_{s,A}(L)+P'_{s,\,B}f'_{s,B}(L)}{P'_s}, \tag{24}$$

where $P'_{s,A}$ ($P'_{s,B}$) is the probability that a successful path traverses through the region of saddle point S_A (S_B). Figure 8 shows the two contributions on the right-hand side of Eq. (24) as obtained from simulations of 10^4 successful paths. The full (dashed) curves show the contribution due to the paths passing near saddle point S_A (S_B), as obtained using the exact importance function. Since the energy of saddle point S_A is higher than that of S_B, its contribution to the total length distribution function is smaller. The filled (empty) circles are the data obtained using the iterative approximation for the importance function. As expected, the agreement between the two sets of data is excellent. In this example, the

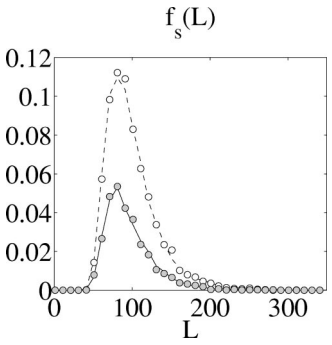

FIG. 8. Length distribution $f_s(L)$ of the successful paths for the two-dimensional random walk computed by importance sampling of 10^4 *successful* paths. The contributions from paths crossing saddle region S_A are plotted with the solid line and filled circles, while those associated with path S_B are plotted with the dashed line and open circles, using the optimal (Fig. 6) and close to optimal (Fig. 7) importance functions, respectively.

computer time spent on optimizing the importance function and on generating successful transition path samples is comparable. In comparison, direct sampling without importance function would require 10^{12} more computer time to generate the same ensemble of successful paths, given its low success probability P_s.

Finally, we consider the properties of the modified and bias potential energy landscapes corresponding to the optimal importance function. For this purpose we consider their behavior along two distinct paths connecting N_0 and N_1, as shown in Fig. 9(a). The path marked *A* is representative of important sequences of states, connecting both minima through saddle point S_A. In contrast, path *B* represents a relatively unimportant path, linking N_0 and N_1 through a straight line in the configurational space. Figures 9(b) and 9(c) show the modified and bias functions $E'(i)$ and $\Delta E(i)$, as well as the original energy landscape along the respective paths. Along path *A*, the bias function effectively ''inverts'' the energy landscape in the neighborhood of minimum N_0, allowing random walkers on the modified landscape $E'(i)$ to diffuse to state N_1 without encountering any saddle points. Along path *B*, however, the optimal bias function only shifts the energy level of state N_0 with respect to that of N_1, but does not remove the energy barrier. This indicates that the approximate inversion property observed for the one-dimensional (1D) random-walk problem holds only for directions in the potential-energy landscape involving a reaction coordinate of the transition of interest.

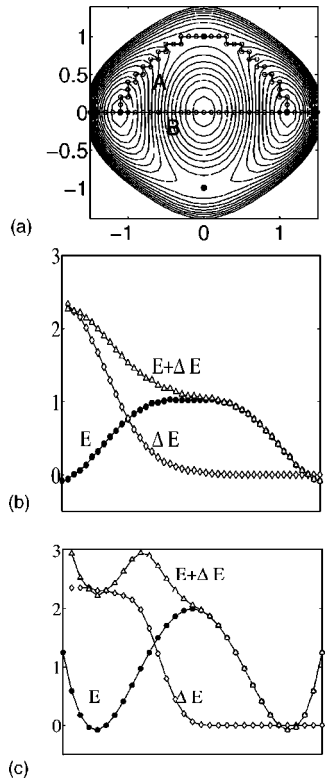

FIG. 9. (a) Paths *A* and *B* across the energy landscape. (b) Optimal bias potential ΔE along path *A* (c) ΔE along path *B*.

C. Two-dimensional random walk using one-dimensional importance function

The optimal importance function I_{opt} is defined on the domain of all states accessible to the system under consideration. In the previous two applications the number of accessible states is relatively small so that the successive trial importance functions can be stored for all microscopic states. The resulting iterative procedure eventually converges to a near optimal solution. It is evident, however, that this task becomes increasingly difficult for high-dimensional systems involving large numbers of degrees of freedom. In such cases one may no longer be able to record or adjust the value of a trial importance function on all accessible states. In this situation one may proceed by using an importance function that, instead of being tabulated for all accessible states, is defined in terms of a reduced set of variables. Although such an importance function will generally not converge to the exact optimal solution, it still may significantly increase the success probability while controlling the variance, provided the reduced variable space is selected appropriately. In particular, an appropriate trial function would be defined on those degrees of freedom that are actively involved in the transition of interest, i.e. the reaction coordinates. Unfortunately, the reaction coordinates are usually not known *a priori*. For these reasons, developing a robust solution that applies to general high-dimensional systems still remains a great challenge. In this section, however, we demonstrate that the above mentioned approach can be practical, i.e. one can indeed use a lower-dimensional importance function to boost sampling efficiency of a higher-dimensional system.

As an illustration, let us modify the 2D random-walk problem of the preceding section by redefining success as the walker reaches states with energies higher than a preselected value E_c. In this problem, we could use a trial importance function defined in terms of the state energy E itself. Motivated by the behavior of optimal importance functions in the previous sections, we can use a trial importance function of the form

$$I(i) = \exp\left(-\frac{\Delta E(i)}{2k_B T} \right), \tag{25}$$

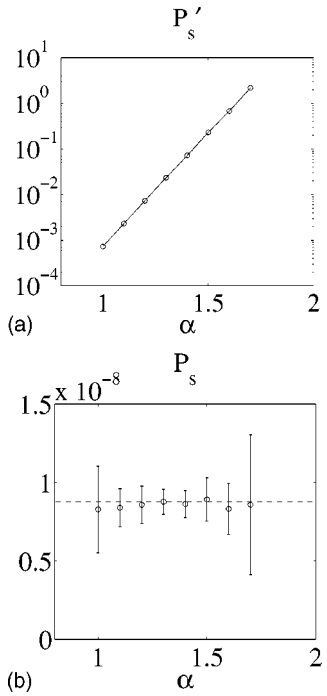

FIG. 10. (a) The estimated success probabilities $\overline{P_s'}$ for the importance-sampled random walks with different choices of α. The bias potential is $\alpha(E_c-E)$. (b) The estimated success probability $\overline{P_s}=\overline{P_s'}I(0)$ for the original system. The averages from 20 independent measures are plotted, each drawn from 10^4 random walkers. The dashed line shows exact solution (8.772×10^{-9}), which is the optimal importance function at state N_0. The error bar shows the estimated standard deviation, which is minimized for $\alpha=1.3$.

$$\Delta E(i)=\alpha[E_c-E(i)], \tag{26}$$

where α is a adjustable parameter for optimization. A choice $\alpha\approx2$ appears particularly appropriate considering Figs. 4 and 9. Figure 10(a) shows the measured success probability of the importance sampled random walks $\overline{P_s'(\alpha)}$. The results for a range of α are plotted, with critical energy $E_c=0.9$, which is slightly lower than the energy of both saddle points. Notice that for difference choices of α, $\overline{P_s'(\alpha)}$ changes over several orders of magnitude [the single data point for $\overline{P_s'(\alpha)}$ that is above 1.0 indicates that the average weights of the successful paths in this case are more than 1.0]. On the other hand, the estimated success probability for the original system $\overline{P_s(\alpha)}=\overline{P_s'(\alpha)}\cdot I(N_0)$ remains nearly constant for this range of α [Fig. 10(b)]. The horizontal dashed line represents the exact success probability as determined from the optimal importance function. Each data point and its associated standard deviation $\sigma_{\overline{P_s}}$ were obtained from 20 independent sets of 10^4 random-walk simulations based on the respective importance function. The standard deviations of these estimates are plotted as error bars, showing a minimum around $\alpha\approx1.3$. Even though for $\alpha=1.3$ the success probability P_s' is only about 0.01, rather far from optimal, the importance sampling procedure based on this bias energy function Eq. (26) with $\alpha=1.3$ still gives a very significant efficiency boost compared to a direct simulation based on the original energy landscape, with $P_s=8.772\times10^{-9}$. The weighted average (with weight being the inverse of the standard deviation) of these estimates for P_s is $\overline{P_s'}=8.529\times10^{-9}$, in good agreement with the exact solution.

IV. DISCUSSION

The importance sampling approach discussed here shows interesting similarities with the transition path sampling (TPS) approach developed by Chandler and co-workers [10,11]. The TPS method samples an ensemble of successful transition paths through a Metropolis algorithm in which a current successful trajectory is slightly distorted to generate a new trial path. As in our importance sampling scheme, this algorithm samples predominantly successful transition paths while preserving the relative probabilities of different successful events, and has shown to be very useful for the study of rare-event phenomena in complex systems. The major difference between the TPS scheme and the current importance sampling method is that the latter is based on the use of an explicit importance function. In this manner, one not only preserves the relative distribution of successful transition paths, but also retains quantitative information about the amount by which the sampling of successful transition paths has been enhanced. This information is not available in the TPS approach, where one focuses mainly on the successful transition events but the ratio of relative probabilities of successful and unsuccessful paths is lost. As a consequence, TPS does not directly yield the absolute transition rates but requires a significant additional effort to compute them using other methods, e.g., umbrella sampling. Given that this ratio is always known in the present importance sampling framework, it is free of this limitation and, in principle, allows a direct estimation of absolute transition rates.

Although it has shown to be very effective in the demonstrated one-dimensional and two-dimensional problems, the extension of the importance sampling scheme to more realistic problems involving large numbers of degrees of freedom remains a challenge. In the context of the present method, the main hurdle is the determination of a suitable importance function. In this sense, the problem of sampling rare transition events has been transformed into a problem of optimizing the importance function. This transformation appears promising, given that the optimization of the many-body importance function shows striking similarities to quantum Monte Carlo methods [12], in which the Schrödinger's equation is solved by optimizing the trial many-body wave functions.

V. SUMMARY

We have developed an importance sampling method for effectively sampling rare transition events in Markov processes. The approach is based on the use of an importance function that modifies the transition probabilities such that the probability of generating successful transition paths is enhanced, while preserving the relative probabilities of different successful events. In addition, the importance sampling method enhances the absolute probability of generating

a successful path by a *known* amount, which facilitates the evaluation of absolute transition rates. It is shown that an optimal importance function I_{opt} exists and can be approximated using an iterative stochastic algorithm. Importance sampling based on the optimal importance function leads to maximum efficiency, for which each simulated sequence of states is successful and the variances in the estimators for statistical quantities are minimized.

The extension of this method to problems involving large numbers of degrees of freedom remains a challenge. However, the general approach appears promising, given that similar problems of optimization in high dimensions have been successfully addressed in other areas of computational sciences, e.g., quantum Monte Carlo.

ACKNOWLEDGMENTS

This work was performed under the auspices of U.S. Department of Energy by 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. M.K. and V.V.B. also acknowledge support from the Office of Basic Energy Sciences, U.S. Department of Energy.

[1] W. Cai, V. V. Bulatov, and S. Yip, J. Comput.-Aided Mater. Des. **6**, 175 (1999).

[2] W. Cai, V. V. Bulatov, J. F. Justo, A. S. Argon, and S. Yip, Phys. Rev. Lett. **84**, 3346 (2000).

[3] C. S. Deo and D. J. Srolovitz, Model. Simul. Mater. Sci. Eng. **10**, 581 (2002).

[4] M. H. Kalos and P. A. Whitlock, *Monte Carlo Methods* (Wiley, New York, 1986).

[5] One can also use a Metropolis algorithm, in which case the walker does have a finite probability to stay at its current state. The jump probability will then be $K(i\rightarrow i-1)=1/2\ \min[r_{-}^{2}(i),1]$, $K(i\rightarrow i+1)=1/2\ \min[r_{+}^{2}(i),1]$, $K(i\rightarrow i)=1-K(i\rightarrow i-1)-K(i\rightarrow i+1)$.

[6] D. T. Gillespie, *Markov Processes: An Introduction For Physical Scientists* (Academic Press, San Diego, 1992).

[7] M. A. Novotny, Phys. Rev. Lett. **74**, 1 (1995).

[8] It can be shown that the largest eigenvalue (in absolute values) of $\tilde{K}_{ij}$ is 1.

[9] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, New York, 1987).

[10] D. Chandler, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. J. Berne, G. Ciccotti and D. F. Coker (World Scientific, Singapore, 1998), pp. 51–66.

[11] C. Dellago, P. G. Bolhuis, and D. Chandler, J. Chem. Phys. **108**, 9236 (1998).

[12] M. P. Nightingale and C. J. Umrigar, *Quantum Monte Carlo Methods in Physis and Chemistry*, Proceedings of NATO Advanced Study Institute (Kluwer, Dordrecht, 1999).