Efficient dynamic importance sampling of rare events in one dimension

Daniel M. Zuckerman[1,*] and Thomas B. Woolf[1,2,†]

[1]*Department of Physiology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*
[2]*Department of Biophysics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205*
(Received 25 May 2000; published 27 December 2000)

Exploiting stochastic path-integral theory, we obtain *by simulation* substantial gains in efficiency for the computation of reaction rates in one-dimensional, bistable, overdamped stochastic systems. Using a well-defined measure of efficiency, we compare implementations of ''dynamic importance sampling'' (DIMS) methods to unbiased simulation. The best DIMS algorithms are shown to increase efficiency by factors of approximately 20 for a $5k_BT$ barrier height and 300 for $9k_BT$, compared to unbiased simulation. The gains result from close emulation of natural (unbiased), instantonlike crossing events with artificially decreased waiting times between events that are corrected for in rate calculations. The artificial crossing events are generated using the closed-form solution to the most probable crossing event described by the Onsager-Machlup action. While the best biasing methods require the second derivative of the potential (resulting from the ''Jacobian'' term in the action, which is discussed at length), algorithms employing solely the first derivative do nearly as well. We discuss the importance of one-dimensional models to larger systems, and suggest extensions to higher-dimensional systems.

I. INTRODUCTION

The rapid computation of the transition rate by numerical simulation for an overdamped stochastic particle confined to a one-dimensional double-well potential (Fig. 1) is a deceptively simple problem, and of crucial importance to making progress in multidimensional systems. While straightforward unbiased simulation ultimately yields the rate to any desired precision (e.g., [1]), it requires the simulator to endure many times the waiting period between the rare transition events [Fig. 2(a)]. Such unbiased simulation, moreover, is utterly impracticable for larger systems—particularly biomolecules possessing thousands of atoms, for which simulating a waiting time of 1 ms could require tens or hundreds of centuries of computer time. The well-known analytic methods for rate computations in simple systems [2] are also insufficient for high-dimensional, rough energy landscapes because many barriers and metastable states of unknown geometry intercede along multiple unknown pathways between the two states of interest (see, e.g., [3,4]).

The inadequacy of both analytic methods and straightforward simulation for large biochemical systems points directly to the need for importance sampling and related methods [5–18]. Importance sampling techniques focus computational effort on transition events, typically generating an ensemble of transition trajectories from which to estimate rates and/or paths. Yet despite the early successes and formal appeal of these methods, a quantitatively successful computational tool for large protein systems with 10^4 atoms has not been achieved. We believe the apparently trivial bistable one-dimensional system must be completely understood *in a simulation context* before substantial progress can be expected for importance sampling in larger systems.

The present paper is explicitly computational, or ''simulational:'' the sole objective is to develop efficient simulation methods for one-dimensional systems which are directly or indirectly applicable to large systems. We concern ourselves only with the ''dynamic importance sampling'' (DIMS) methods developed by Woolf [8] and Zuckerman and Woolf [14], which generate ensembles of fully independent transition trajectories. We hope and believe our results will be of practical use in other multidimensional methods, but such applications are beyond the scope of this work. Rather, the ''bottom-line'' questions we attempt to answer are as follows: (i) *Within the DIMS framework, what are the most efficient methods for rate computation*? (ii) *Using a well-defined measure, how efficient are these methods compared to unbiased simulation*? We will also discuss the maximum efficiency possible using DIMS and related methods in one-dimensional (1D) problems, as well as attempting to extrapolate to larger systems.

Achieving efficiency for a 1D problem using dynamic importance sampling, as we will show, requires the appreciation of a variety of theoretical results—particularly relating to stochastic path integrals. The review by Mortensen [19]
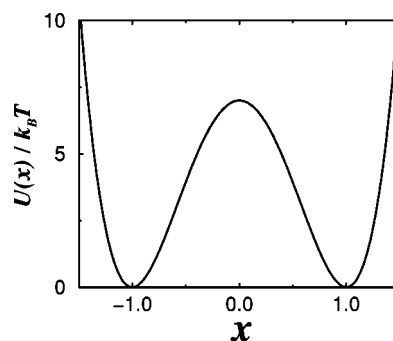

FIG. 1. The symmetric bistable potential studied in the text, Eq. (5.1), shown for a barrier height of $E_b = 7k_BT$ and length parameter $l = 1$, the unit of the x axis.

*Email address: dmz@groucho.med.jhmi.edu
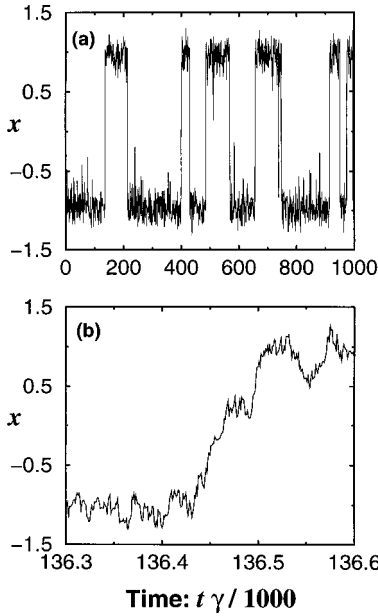†Email address: woolf@groucho.med.jhmi.edu

FIG. 2. Time scales and crossing events. An unbiased trajectory exhibits two vastly different time scales: for a substantial barrier height, the waiting time between events greatly exceeds the time for a single crossing event. (a) The top plot shows an unbiased trajectory of an overdamped particle in the double well of Eq. (5.1), with a barrier height of $E_b = 7k_BT$. (b) The bottom plot isolates a single crossing event from the trajectory in (a) and highlights the rapidity (steepness) of the crossing, including the ascent. Note that the unit of x in both plots is the length scale l; see Eq. (5.1).

and Gardiner's book [20] give introductions to stochastic integrals—of which Ito's and Strantonovich's are the most basic types. Historically, Onsager and Machlup first formulated a stochastic path integral to describe overdamped Brownian dynamics [21]. Since then, uniqueness and other formal aspects of the Onsager-Machlup formulation have been discussed by a variety of workers (e.g., [22–28]). Indeed, formal properties of stochastic functional integrals also play an important role in the present investigation. These integrals have been used, for example, to address the question of the most likely crossing event [see Figs. 2(b) and 3] [29–39]. Related functional integral approaches directly address the Smoluchowski (overdamped Fokker-Planck) equation for the evolution of the entire probability distribution [22,40,29,41,42].

Path-integral formulations lead naturally to the notion of an average path, which is critical to the present discussion—even in one dimension. While the geometric pathway for barrier crossing is trivial in a one-dimensional potential like that of Fig. 1, the "speed" (average displacement) at every position constitutes another dimension of the average path—and a critical one to the present discussion. Such 1D average paths (average speeds and distributions) were considered by Dykman and co-workers [43] and by Luchinsky and McClintock [44], both theoretically and experimentally. Zuckerman and Woolf [14] considered average paths in higher dimensions, computationally, while the problem of finding optimal reaction paths in multidimensional systems has a long history (e.g., [45–48,6,37]).
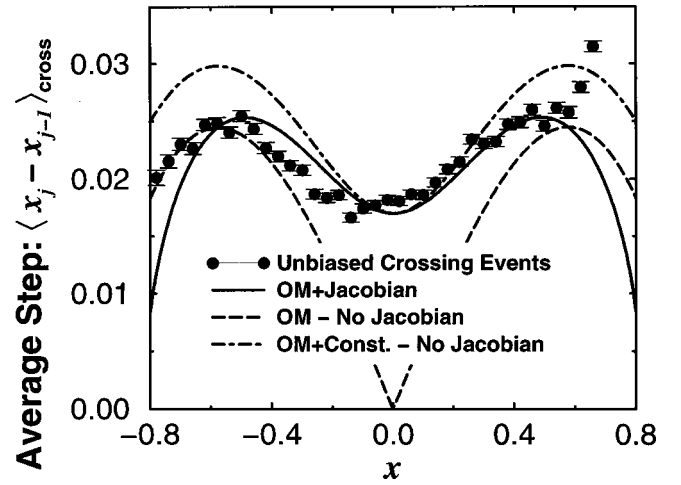

FIG. 3. Average simulation step sizes $(x_j - x_{j-1})$ from many unbiased crossing events generated by Eq. (2.3). A typical crossing event is shown in Fig. 2(b). Solid circles are the data averaged from *unbiased crossing events only* for the potential of Eq. (5.1), with a barrier height $E_b = 7k_BT$. The solid line is the most probable path of Eq. (3.6) based on the Onsager-Machlup action with the stochastic correction (Jacobian) term; the dashed lines depict the uncorrected Onsager-Machlup extremal path, Eq. (3.7) with $C=0$; and, the dot-dash lines show the uncorrected Onsager-Machlup path with a nonvanishing constant—i.e., Eq. (3.7) with $C = -2k_BTU''(0)/(m\gamma)^2$. The statistical error in the simulation data may be gauged by the noise in the data. The data for $|x| > 0.6$ are affected "artifactually" by the procedure of extracting crossing events from long trajectories; see the text, however. The data of this figure only were generated using the parameters $\Delta t = 10^{-4}\gamma^{-1}$, $\gamma^{-1} \equiv 1$, $m = 10.98$, and $k_BT = 249.462$, largely following Refs. [8], [14]; all lengths are given in units of the length scale l of Eq. (5.1).

The paper is organized as follows. We first review the basic formalism for overdamped Langevin dynamics and importance sampling in Sec. II. Section III introduces pertinent path-integral results based on the Onsager-Machlup action, and reviews the so-called "Jacobian" term in some detail, as well as presenting new numerical data. Section IV discusses a method for producing an efficient biased ensemble of trajectories, while Sec. V gives the actual results from the biasing technique. Section VI addresses the relevance of our results to multidimensional problems. Finally, conclusions and a summary of the results are given in Sec. VII.

II. OVERDAMPED LANGEVIN DYNAMICS AND IMPORTANCE SAMPLING

This section briefly reviews the fundamental stochastic dynamics equations and the dynamic importance-sampling formalism for rate computations. We consider solely stochastic dynamics governed by the overdamped Langevin equation in the presence of Gaussian white noise. In the notation of [14], such "Brownian dynamics" are described by

$$dx/dt = f/m\gamma + R(t), \qquad (2.1)$$

where x is the configurational coordinate, t is time, $f(x) = -\nabla U(x)$ is the force, m is the particle's "mass," γ is the

friction constant, and the noise R is assumed Gaussian with zero mean and variance given according to

$$\langle R(t)R(t')\rangle = (2k_BT/m\gamma)\delta(t-t'), \qquad (2.2)$$

where k_B is Boltzmann's constant and T the temperature. Equation (2.1) is presumed to be simulated according to the ''Ito-like'' (Euler) discretization [49,50]

$$x_{j+1} = x_j + (f_j/m\gamma)\Delta t + \Delta x_R, \qquad (2.3)$$

where the subscripts j indicate quantities evaluated at time $j\Delta t \equiv t_j$, so that $f_j = f(x_j)$, and Δx_R is chosen from a Gaussian distribution of zero mean and variance

$$\sigma^2 = 2\Delta t k_B T/m\gamma. \qquad (2.4)$$

The discretization (2.3) is considered to be Ito-like because the force—assumed to be constant over the interval Δt—is evaluated at the beginning of the interval; a loosely termed ''Stratonovich-like'' approach would instead consider, formally [29,24,20],

$$f_j \rightarrow (f_j + f_{j+1})/2. \qquad (2.5)$$

We note that discretizations more sophisticated than the Euler scheme (2.3) are well known—i.e., Runge-Kutta schemes of various orders [51–53]. However, in explicit tests on *rate calculations only* for our simple system (Fig. 1), the lowest-order Runge-Kutta (Heun) method permitted only a marginally larger time step—without losing accuracy—which did not justify the additional computational expense (data not shown). Larger time steps tend to produce *systematic* errors, in addition to the easier-to-diagnose statistical error, regardless of the algorithm [53]. Rate calculations, moreover, may be especially sensitive. We also note that methods of higher-order than the Heun algorithm are not of interest here, because computing higher-order derivatives is extremely costly for the larger bimolecular systems toward which the present research is ultimately oriented. Thus, given our focus on rate computations and our interest in large systems, there seems little reason to employ anything more complicated than the Euler procedure (2.3).

It is useful to consider, in parallel to the dynamical algorithm, the equivalent probabilistic description of the dynamics (2.3). In particular, the explicit single-step transition probability density described above is given by

$$T_{\Delta t}(x_{j+1}|x_j) = (2\pi\sigma^2)^{-1/2} \exp\{-[(x_{j+1}-x_j)$$
$$-(f_j/m\gamma)\Delta t]^2/2\sigma^2\}. \qquad (2.6)$$

Thus the probability density for a whole trajectory $\zeta = (x_0, x_1, x_2, \ldots, x_n)$ is given by the product of the single-step densities:

$$\tilde{Q}(\zeta) = \prod_{j=0}^{n-1} T_{\Delta t}(x_{j+1}|x_j). \qquad (2.7)$$

Importance sampling is effected [8,14] by performing *biased* simulations which follow a prescription different from

Eq. (2.3). Naturally, trajectories generated from a biased simulation are distributed not according to $\tilde{Q}$, but according to some other function, $D(\zeta)$. This alternative distribution can be used to compute probabilities and, hence, transition rates. One first requires the conditional probability $P_B(t|0;x_0)$ to be in the final state (B) at time t having started at x_0 at $t=0$, which can be written in two equivalent ways for nonvanishing D,

$$P_B(t_n|0;x_0) \simeq \int d\zeta\, \tilde{Q}(\zeta)h_B(x_n) = \int D(\zeta)d\zeta\, \frac{\tilde{Q}(\zeta)}{D(\zeta)}h_B(x_n), \qquad (2.8)$$

where $d\zeta = \prod_{j=1}^{n} dx_j$, and $h_B(x)$ is an indicator function which is unity for x in state B and zero otherwise. The approximation here is simply the neglect of the discretization error. In a biased simulation, the integral is estimated by

$$P_B(t_n|0;x_0) \simeq \frac{1}{M} \sum_{\zeta \in S_D} \frac{\tilde{Q}(\zeta)}{D(\zeta)}h_B(x_n), \qquad (2.9)$$

where S_D is a sampling ensemble of M trajectories chosen according to D, and where normalization, $\int d\zeta\, D(\zeta) = 1$, is required. Note that the function D must be known for the simulation to be performed, just as Eqs. (2.6) and (2.7) are the distributions from which unbiased steps and trajectories are sampled. Ultimately, rates are estimated by computing the slope of the linear regime in a probability vs time plot, as discussed in Sec. IV.

III. PATH INTEGRALS AND MOST PROBABLE CROSSING EVENTS

Achieving efficiency for a 1D problem using dynamic importance sampling requires a variety of theoretical results for stochastic path integrals and their ''instantons.'' These approaches have been used, for example, to address the question of the most probable crossing event [see Figs. 2(b) and 3], which can be derived in a straightforward way from the continuum Onsager-Machlup action used in the path-integral formulation [29,31,33–39]. In this section we discuss the Onsager-Machlup action and its optimization at somewhat greater length than is required for the computational aims of this paper. This is because, beyond the theoretical interest in the action, and especially in the so-called ''Jacobian'' term, some older literature deserves a fresh look and detailed discussion.

The path-integral story begins with Onsager and Machlup (OM) [21] and their well-known action for weighting continuum representations of overdamped Brownian trajectories. In that formulation, a *continuous* trajectory—presumably intended to represent a smoothed stochastic trajectory—is weighted according to

$$\text{probability} \propto \exp\left[-\frac{S}{(k_BT/m\gamma)}\right] \qquad (3.1)$$

where

$$S = \tfrac{1}{4} \int_{t_i}^{t_f} dt \, \mathcal{L}(t), \tag{3.2}$$

and

$$\mathcal{L}_{\mathrm{OM}}(t;x,\dot{x}) = \left[\dot{x} - \frac{f(x)}{m\gamma} \right]^2, \tag{3.3}$$

and where $\dot{x} = dx/dt$ and $f(x) = -dU/dx$. In their original paper [21], Onsager and Machlup applied the Euler-Lagrange equation for functional minimization,

$$\frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{x}} = 0. \tag{3.4}$$

to the case of a single harmonic well.

Subsequent contributions amended the original Onsager-Machlup formulation of the action, Eqs. (3.2) and (3.3). First, Stratonovich realized in 1962 [22] that if one starts from the product of discrete-step probability densities (2.7) and determines the continuum limit in a rigorous fashion, an additional term arises in the integral representation of the action (3.2), so that instead of Eq. (3.3), the effective Lagrangian takes the Onsager-Machlup-Jacobian (OMJ) form, namely

$$\mathcal{L}_{\mathrm{OMJ}} = \mathcal{L}_{\mathrm{OM}} + \frac{2k_B T}{(m\gamma)^2} U''(x). \tag{3.5}$$

This result was confirmed by Bach *et al.* [25]. In other words, the original description by Onsager and Machlup was incomplete for simulations performed according to the Ito-like algorithm (2.3). Note that this new term is proportional to the curvature of the potential and is dominant at a barrier top, for any nonzero noise amplitude ($T>0$).

In the mid-1970s Graham [40,29] rederived the term as a Jacobian resulting from changing variables from the fluctuation coordinate to the configuration coordinate—assuming a Stratonovich-like discretization (2.5). Thereafter, apparently, the term has been viewed as a Jacobian, although the propriety of that name in a practical *simulation* context is questionable since, to the authors' knowledge, only the Ito discretization (2.3) is usable precisely, and that results in a constant Jacobian. More appropriate terminology might be the "stochastic correction term."

Both Stratonovich and Graham indicated, importantly, that the inclusion of the stochastic-correction/"Jacobian" term led to a path-integral representation of the probability distribution consistent with the overdamped Fokker-Planck (Smoluchowski) equation.

Graham also derived the most probable crossing event associated with the corrected action using the Euler-Lagrange equation (3.4), namely [29],

$$[\dot{x}_c^{\mathrm{OMJ}}]^2 = \left[\frac{f(x)}{m\gamma} \right]^2 - \frac{2k_B T}{(m\gamma)^2} U''(x) + C, \tag{3.6}$$

where C is a constant of integration. This result is called an "instanton" by field theorists (e.g., [54]) because it de-scribes a rapid transition, as illustrated in Fig. 2. To appreciate this, note that by taking either the positive or negative square root for $\dot{x}_c^{\mathrm{OMJ}}$ in Eq. (3.6), the velocity of ascent is seen to be the same as for descent. Yet this feature was apparently not appreciated until later; see below. It will, however, prove central to our goal of obtaining efficiency in simulations. Below, we also address the computational cost associated with computing a second derivative of the potential.

It is worth noting that although the extremal path, formally, results when $C<0$ in Eq. (3.6) [from minimizing Eq. (3.2) with Eq. (3.5)], one cannot necessarily take the constant to be negative or even to vanish. Observe that the left-hand side of Eq. (3.6) must be positive. Thus, in regions of relatively large positive curvature—such as near a well bottom—the second term of the right-hand side could exceed the first in magnitude, requiring $C>0$. While this observation appears to be immaterial to our investigation of the simple bistable well (see Fig. 3), one could imagine the positivity of the constant having physical consequences in a more complicated potential with metastable intermediate states.

Other features of the most probable crossing have also proved of great interest. Dykman and Krivoglaz made important observations in their study of transitions in nonequilibrium systems using an uncorrected Onsager-Machlup functional integral, in 1979 [30]. In a series of papers begun in 1989 and aimed primarily at systems with colored noise, Bray, McKane, and co-workers [31,33–36] also provided some insights pertinent to the case of white noise *without* the correction term. (While the latter group included the Jacobian term formally, it was omitted from their analysis of the low-temperature limit.) Both groups derived essentially a special case of Graham's result (3.6)

$$[\dot{x}_c^{\mathrm{OM}}]^2 = \left[\frac{f(x)}{m\gamma} \right]^2 + C, \tag{3.7}$$

where C is again a constant of integration which vanishes for the most probable case. It also is equivalent to Graham's result at inflection points of the potential. Bray, McKane, and co-workers noted the crucial fact mentioned above: for systems with detailed balance (i.e., time-invariant potentials), the ascent described by $\dot{x}_c^{\mathrm{OM}}$ is equally rapid as, and symmetric with, the descent [see Fig. 2(b)]. This point is also implicit in the work of Dykman and Krivoglaz. Furthermore, both groups recognized that paradoxically the external trajectory ($C=0$) *never occurs* [30,35] because an infinite amount of time is required to reach (or descend from) a parabolic barrier top. This may be seen by computing the barrier crossing time,

$$t_b = \int_{t_i}^{t_f} dt = \int_{x_i}^{x_f} \frac{dx}{\dot{x}_c}. \tag{3.8}$$

Such apparently formal observations belie the fact that optimal paths have proven useful in a variety of recent studies. Dykman *et al.* determined weighted *distributions* of paths (including the optimal) leading to large, nontransitioning fluctuations in a time-oscillatory potential [55]. Olender

and Elber [37] discussed the Onsager-Machlup extremal barrier crossing, and employed a discrete version in developing an algorithm for steepest-descent path-finding between known initial and final states. Bier *et al.* [38] discussed extremal crossings in the context of rate calculations for fluctuating barriers. Elber and Shalloway suggested a temperature dependence for the integration constant of Eq. (3.7) to overcome the divergent-crossing time difficulty, and successfully distinguished optimal paths for different regimes of an effective temperature parameter [39].

Both in order to confirm the analytic results reviewed above and to gather data that could usefully inform our efforts to gain efficiency in rate computations, we studied numerically generated crossing events. We performed very long unbiased simulations of a particle moving according to Eq. (2.3) in a one-dimensional bistable well (Fig. 1) and ''snipped out'' a large number of crossing events like the one shown in Fig. 2(b). Binning the observed step sizes of *only* the crossing events according to the x position, Fig. 3 shows that the average step sizes $\langle x_j - x_{j-1} \rangle_{\text{cross}}$ closely follow the prediction (3.6) which includes the stochastic correction term. The two theoretical predictions lacking the correction term—namely Eq. (3.7) with two values for C—appear far less adequate by contrast. We note that the binned distributions appear to be highly Gaussian so that the average and most probable values essentially coincide.

A comment on the data for $|x| \gtrsim 0.6$ in Fig. 3 is in order. The substantial turn up of the data for $x > 0.6$ is certainly an artifact of isolating crossing events: by definition, the final step or steps are right moving. Thus, if one defines crossing events to end when the value 0.7 is exceeded, rather than 0.8 as in the figure, the turn up occurs for correspondingly smaller x values. Interestingly, while the same argument applies to the left side of the figure, the data are not similarly affected. Indeed, the asymmetry between the left and right edges of the plot is striking. While, at present, we can offer no convincing explanation, it is worth noting that the ''OM+Jacobian'' description becomes less than credible as x increases beyond the right inflection point of the potential, $x \simeq 0.58$. That description, Eq. (3.6) with $C=0$, suggests counterintuitively that a particle which just completed a transition should fall more slowly than the drift velocity (effectively given by ''OM–No Jacobian'' for $x>0$ in Fig. 3). Yet it is not clear to the present authors precisely why the applicability of the corrected OM theory should be limited to the region between the inflection points; intuitively, of course, the inflection points roughly mark the boundaries between the stable states and the barrier-top transition rate.

IV. A HIGHLY EFFICIENT BIAS METHOD FOR DIMS CALCULATIONS

The primary goal of the present work is to improve and quantify the level of efficiency in one-dimensional rate calculations. To that end we now introduce and evaluate a biasing method, which is a variation on those discussed in our earlier work [8,14]. The method combines two essential ingredients: biased crossing events that emulate the most prob-

able path, together with a threshold at which the biasing is triggered.

A. Motivation

Computing reaction rates by simulation requires the generation of an ensemble of appropriately weighted trajectories (or a correspondingly long trajectory) exhibiting many barrier-crossing events. While many sampling methods are capable of generating a suitable ensemble of trajectories in a long period of time, it is no easy task to rapidly generate a truly *important* ensemble—containing highly probable crossing events—even for an apparently trivial potential like a symmetric, one-dimensional double well. The reason is not hard to see. In an unbiased simulation (2.3), each step is chosen from a Gaussian distribution centered at the deterministic step in the direction of the force. For *crossing events* which climb over a barrier, a trajectory necessarily takes steps in a direction *opposing* the force. As the center of the Gaussian distribution is always downhill from the present step, the typical step in the *ascent* of a crossing event is in the ''tail'' of the Gaussian distribution (more or less so depending on the parameters entering into the width, σ). As the full probability for a crossing event is then a product of steps in the tail of the distribution, it seems clear that only small fluctuations about the most probable crossing event will occur. Large fluctuations away from the most probable path will be exponentially damped out, as they require steps yet further out in the tail of the distribution.

Our interest in the most probable path for a crossing event, then, is hardly surprising: to generate an ''important'' (albeit biased) ensemble of trajectories, one must hew to the most probable path. This we shall do explicitly below, using the results quoted in Sec. III.

But once one knows the most probable path for a single crossing event, the DIMS method still requires the generation of such events at appropriate intervals. That is, there is an ideal distribution of waiting times between events in a biased simulation [cf. Fig. 2(a) for the unbiased case]. To see the reason why a distribution of wait times (first passage times) is required, consider Fig. 4, which demonstrates the rate calculation for biased *and* unbiased simulation. One first calculates the probability to arrive in the final state (the right well in Fig. 1), having started in the initial state (the left well), as a function of time. One then determines the slope of the linear regime. While an unbiased simulation of sufficient length will automatically generate data with roughly the same precision for all times shown along the horizontal axis in Fig. 4, a biased DIMS simulation must be designed to do so. As discussed in [14], one essentially has to evaluate a separate integral for each time point, so part of the goal is to spread the information gathered evenly over the necessary times. This is the motivation behind the ''thresholding'' detailed below.

B. The algorithm: most probable crossings above a threshold

Specifically, trajectories are generated according to the following algorithm. Each trajectory is started from the minimum of the left well, $x = -1$, of the potential (5.1) shown in
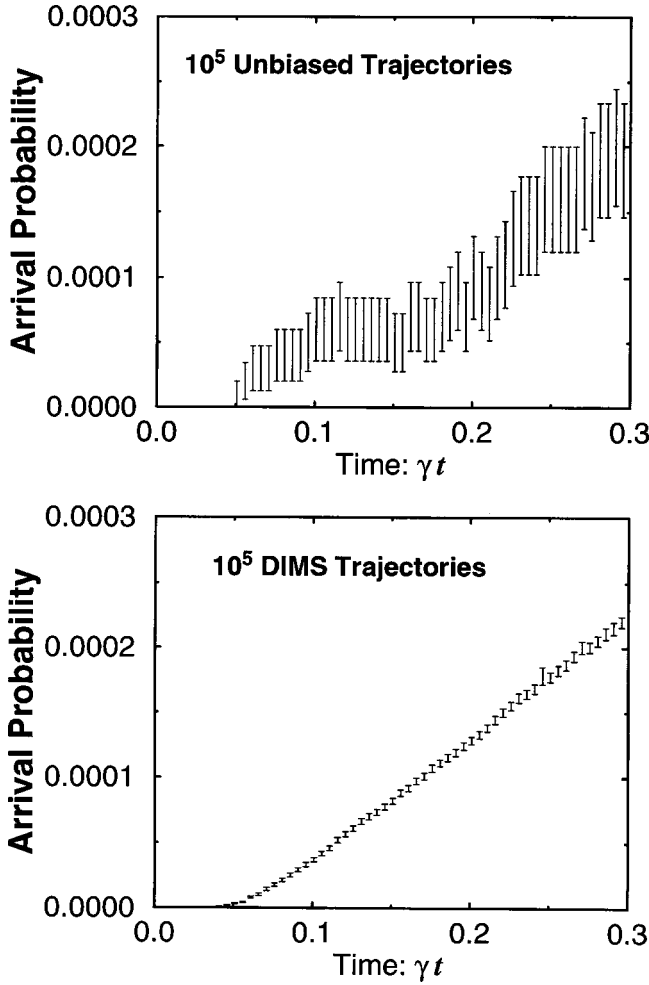
FIG. 4. Evolution of the end-state arrival probability for rate calculations. The probability to arrive in the right well of the potential Eq. (5.1) with $E_b = 9k_BT$, depicted in Fig. 1, is plotted against time, based on trajectories initiated at the bottom of the left well ($x = -1$) at time $t = 0$. Both unbiased (top) and DIMS (bottom) results are shown, with the latter using Eqs. (4.1) and (3.6). The rate k is computed as a fit to the slope of the linear regime. The DIMS computation (bottom) shows a dramatic improvement in efficiency, which is quantified in Fig. 5.

Fig. 1, and run for a fixed total amount of time exceeding the transient time, t_b. Predefining a threshold value, $-1 < x_h < 0$, of the coordinate x, we perform unbiased simulation while $x < x_h$ according to Eq. (2.3). If and when the threshold value is exceeded ($x > x_h$), we select steps according to

$$x_{i+1} = x_i + \dot{x}_c(x_i)\Delta t + \Delta x_R, \qquad (4.1)$$

with $\dot{x}_c$ given by either the Jacobian-augmented result (3.6) or that without (3.7), and with Δx_R chosen from the same Gaussian distribution as in the unbiased case. In other words, instead of using the deterministic step $f_i \Delta t / m\gamma$ as in Eq. (2.3), we use the most probable step *in a crossing event* to emulate unbiased crossing events. The use of a threshold away from the well minimum ensures a sufficiently broad distribution of waiting times between the artificial crossing events, which in turn permits the acquisition of data for the range of times necessary to compute the rate as in Fig. 4. We note, however, that it is no more difficult to run a single long trajectory and compute correlation functions to determine rates (e.g., [1]).

If a Heun scheme [51–53] were used for the unbiased dynamics, the biased dynamics just given (4.1) could be readily modified in the same way the Euler scheme itself (2.3) is modified in the Heun approach. Given that the underlying continuum action should be the same for the two integration methods, a Heun modification of Eq. (4.1)—or even the unmodified version—should give similar results to those found here with the Euler method.

C. Curvature-adjusted sampling width

Yet another refinement for the one-dimensional biasing techniques is possible, motivated by the time-dependent width of the Gaussian noise σ_j, which is required for the provably optimal computation of the probability *density* at a single point on a curvature-free potential surface [50]. Note that the biasing methods just described always used a constant width, $\sigma_j = \sigma$ given in Eq. (2.4). In a spirit similar to Wagner's approach described in [50], one can ask the question: ''What is the optimal sampling scheme to travel between two fixed points, with one intermediate step, on a surface with arbitrary curvature?'' Answering this question is not difficult and suggests that the local curvature influences the optimal width.

Following Wagner (see [50]), one needs to observe first that the optimal *sampling density* at time $t = \Delta t$ is exactly the distribution of *unbiased* two-step paths which begin at x_0 at $t = 0$ and end at x_2 at $t = 2\Delta t$. This constitutes the perfect *sampling* distribution because it is precisely that (tiny) subset of unbiased trajectories which end at the predefined value of interest, and which are distributed naturally. Using the single-step Gaussian distributions (2.6), the two-step distribution is

$$T_{\Delta t}(x_2|x_1)T_{\Delta t}(x_1|x_0) = \frac{1}{2\pi\sigma^2}\exp\left\{-\frac{[(x_2-x_1)-(f_1/m\gamma)]^2 + [(x_1-x_0)-(f_0/m\gamma)]^2}{2\sigma^2}\right\}, \qquad (4.2)$$

where σ still represents the unbiased value and f_i again gives the force at x_i. By rearranging the terms, completing the square, and approximating $f_i \simeq f_0 - (x_i - x_0)U''(x_0)$, one finds

$$T_{\Delta t}(x_2|x_1)T_{\Delta t}(x_1|x_0) \simeq c \times \exp\left\{-\frac{\left[x_1 - \frac{x_0+x_2}{2}\left(\frac{1-\alpha}{1-\alpha+\alpha^2/2}\right)\right]^2}{2[\sigma/\sqrt{2(1-\alpha+\alpha^2/2)}]^2}\right\}, \qquad (4.3)$$

where c is a constant independent of x_1 in this approximation, and the dimensionless curvature is $\alpha \equiv U''(x_0)\Delta t/m\gamma$.

The result (4.3) for the optimal distribution of x_1 values illustrates several interesting points. First, the distribution is independent of the force to first order in α. Second, while the correction to the expected mean of the distribution, $(x_0+x_2)/2$, is second order in α, that for the width σ is first order—noting that the factor $\sqrt{2}$ is expected from the work of Wagner and actually approaches unity for a large number of time steps [50]. The optimal sampling width in fact *decreases* at the barrier top, where $\alpha<0$, in the small Δt limit ($\alpha \gg \alpha^2$). Thus, the distribution (4.3) motivates a further bias for sampling, namely use of the width

$$\sigma_{\text{samp}} = \sigma/\sqrt{1-\alpha+\alpha^2/2} \qquad (4.4)$$

for sampling from Gaussian distributions near the point x, where α is to be computed using $U''(x)$. We investigate this refinement in the next section.

V. RESULTS: COMPARISON OF EFFICIENCY

We demonstrate the capability of the DIMS algorithms of Sec. IV by quantifying their efficiency for rate computations in the simple bistable potential shown in Fig. 1,

$$U(x) = E_b[(x/l)^2 - 1]^2, \qquad (5.1)$$

where E_b is the barrier height and l is the length scale of the problem. The central result is that one must account for the most probable crossing to gain maximum speed-up in rate computation as compared with unbiased simulation. Not surprisingly, the efficiency increases with barrier height. Yet even for the relatively low barrier height of $E_b=5k_BT$, we achieve roughly a 20-times efficiency improvement—i.e., DIMS rate calculations are approximately 20 times as fast for a given level of precision. That factor increases to 300 for a $9k_BT$ barrier. While such gains will not be readily extendible to multidimensional systems, it is important to understand and demonstrate the ingredients necessary for optimal performance.

Figure 5 shows our results for the potential (5.1) for two different barrier heights, $E_b/k_BT=5$ and 9. The biasing methods accounting for the most probable crossing are significantly superior for the larger barrier. The ''DIMS-Wagner'' algorithm—which takes no account of the most probable path—refers to the technique described in Ref. [14]. It performs only modestly well for the $9k_BT$ barrier, and its efficiency is very sensitive to the fixed simulation length. All the other DIMS procedures employed the algorithm of Sec. IV, Eq. (4.1), above a threshold $x_h = -0.7$, with trajectories initiated at $x = -1$. The presence or absence of the ''Jacobian'' (stochastic correction term) reflects whether Eq. (3.6) or (3.7) was used to complete Eq. (4.1), and ''CURV'' indicates that the curvature-modified width (4.4) was used in place of the unbiased width (2.4). Rates, k, were calculated using the method noted in Sec. IV A and Fig. 4.

Efficiency is computed by estimating the relative simulation lengths needed to obtain a desired degree of precision, given by the variance of the rate estimates,

$$\sigma_k^2 = \frac{1}{n}\sum_{i=1}^{n} (k_i - \langle k\rangle)^2, \qquad (5.2)$$

where $n=20$ is the number of simulations performed for each data point, k_i is the rate computed for the ith simulation, and $\langle k\rangle$ is the mean of the n rate estimates. The horizontal arrow spanning the DIMS and unbiased results for $E_b=9k_BT$ at a precision of $\sigma_k = 10^{-4}$, for example, exceeds two decades—indicating that the DIMS computations are more than 100 times faster.

The effectiveness of the DIMS formulation *excluding* the Jacobian—i.e., based on Eqs. (4.1) and (3.7) with $C=0$—deserves further comment. While the Jacobian-augmented DIMS simulation is slightly superior for $E_b = 9k_BT$, the insensitivity to including the correction term is surprising given the sharp contrast demonstrated in Fig. 3. The lesson appears to be that, at least for the parameters studied, the motion in the immediate neighborhood of the barrier top (where the correction term, proportional to the curvature, has the greatest effect) is less important to the anatomy of a crossing event than the (rapid) climb and descent. In the long term, the success of the uncorrected approach could facilitate the extension of DIMS to multidimensional systems, since that approach does not require the computation of second derivatives of the potential. Future work may show this to save a substantial amount of computer time.

We note that our efficiency estimates have excluded the ''overhead'' cost of implementing the DIMS method. This cost depends on the optimization of one's code, and as it happens, our code is suboptimal for *un*biased simulation, so that there is no overhead at all. There are, however, inherent overhead costs in DIMS that cannot be optimized away. While the dynamics employing the Jacobian-augmented most probable path (3.6) requires the computation of a second derivative at every step, our results show that the simpler form (3.7) is nearly as good and requires only the force. Calculating the force, it should be remembered, is not an overhead cost because this must be done in unbiased simulation (2.3) anyway. The only notable cost inherent in the DIMS method, then, is computing the error associated with biased computation—in order to correct for it as discussed in Refs. [8] and [14]. This correction entails computing the ratio of two Gaussian terms (or, equivalently, the difference of two logarithms) at every step. Compared with the fixed cost of unbiased simulation—computing the force and generating a high quality pseudorandom number at every step—and its inherent inefficiency with long waiting times, the DIMS costs are far overshadowed by the efficiency gains which here exceed one and two orders of magnitude.

For completeness, we give a number of further details, which apply to both the unbiased and DIMS results. A simulation consisted of N trajectories (see Fig. 5) initiated at $x = -1$ at time $t=0$. The time steps were $\Delta t = 0.003\gamma^{-1}$ for $E_b=5k_BT$ (having set $\gamma \equiv 1$) and $\Delta t = 0.001\gamma^{-1}$ for E_b
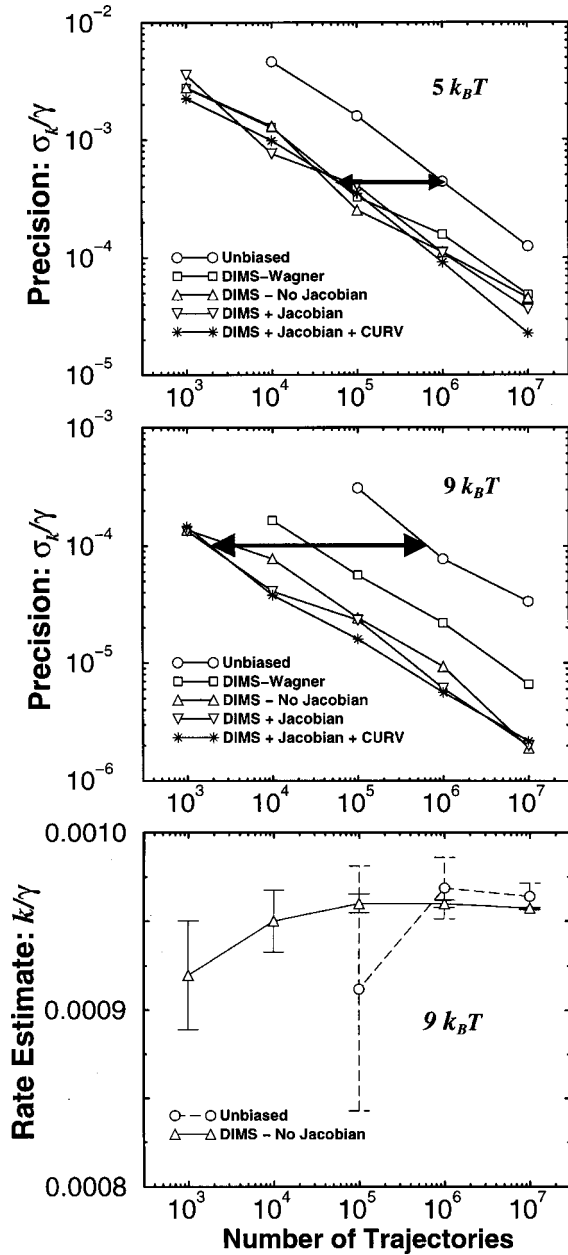
FIG. 5. Efficiency in rate computations using the DIMS approach. The standard deviation σ_k of a set of 20 rate estimates computed by either unbiased or DIMS simulation is plotted against the length of each simulation (top two plots). The spans of the horizontal arrows measure the efficiency by indicating the differences in simulation length required to obtain a given variance (i.e., precision). For example, when the barrier height is $E_b = 9k_BT$, the best DIMS approaches are roughly 300 times faster in obtaining a precision given by $\sigma_k = 10^{-4}$. The ''DIMS-Wagner'' algorithm is from our earlier work [14], while the other DIMS simulations use the algorithm of Eq. (4.1) and the surrounding text. For the latter, the most probable step is chosen according to either the Jacobian-augmented formulation (3.6) or that without (3.7), as indicated. The label ''CURV'' indicates that the Gaussian sampling width was modified according to Eq. (4.4). The bottom plot shows rate estimates k for the $9k_BT$ barrier height. Both DIMS [Eqs. (4.1) and (3.6)] and unbiased estimates converge toward a common result with increasing simulation length. The error bars indicate the standard error of the mean, and underestimate the 95% confidence interval.

$= 9k_BT$. These were determined to be close to the maximum values for which the rate estimates did not change as Δt increased in unbiased simulation. As discussed above, the rate is computed as the slope of the linear regime in a plot of the arrival probability (to be in the right well, $x > 0$) as a function of time; see Fig. 4. The slopes (rates, k_i) were computed from a least-squares fit to 10 data points, the t values of which were held fixed for a given barrier height. The particle mass m and the thermal energy k_BT were both set to 1. Note that Fig. 3 uses the parameters given in the caption.

A. What is the optimal efficiency?

''The system is so simple. How does one know whether a 300-time improvement in efficiency is impressive?'' So a skeptic might wonder, and the attempt to answer seems a worthwhile exercise.

The basic point is that computing a rate by simulation involves the simultaneous calculation of a series of difficult integrals of the form (2.8) discussed in Sec. II. Each data point in Fig. 4 is an estimate for such an integral.

We can try to estimate the minimum number of trajectories needed for the rate computation by multiplying together estimates for the following: (i) the number of trajectories needed to estimate the probability *density* to be at a single location x in state B at a given time, $P(t_j; x \in B | 0; A)$; (ii) the number of discrete locations x in state B required to estimate the probability for the whole state; and (iii) the number of independent time points needed to estimate a slope in the linear regime. Regarding (i), only in the case of a constant force can $P(t_j; x \in B | 0; A)$ be computed exactly [50]—equivalently, with a single trajectory. One might expect that at least ten trajectories would be required for any real surface, setting (i). In a similar manner for (ii), at least ten points should be required to characterize a state (which, in principle, is known only numerically). The number of independent time points is a slightly more complex issue since some (through not all) trajectories from a given time point may also be used to estimate another. Conservatively, then, we use the estimate three for (iii). Our estimate for the minimum number of trajectories required to calculate a rate is thus 300. We believe our DIMS results of Fig. 5 compare favorably with this heuristic—and conservative—theoretical minimum.

VI. POSSIBLE EXTENSIONS TO MULTIDIMENSIONAL PROBLEMS

While the results presented here for a one-dimensional potential seem a far cry from a high-dimensional biomolecular system, we believe they teach important lessons for the large system. The simplest point is that a biased trajectory must closely mimic the natural barrier-crossing dynamics if efficiency in rate computations is to be obtained. Indeed, we have found that a poor bias can be worse than no bias at all. It is not enough to know—as one does automatically in one dimension—even the optimal geometric channel for a transition: the size of the steps along that geometric path are critical, as we have shown.

Although the problem of finding channels is extremely

challenging in itself, let us ask ''How can one compute the rate for a large system assuming the channels are known?'' A natural choice would be to start with an initial trajectory within each channel, and then attempt either to optimize it [56,46–48,6,37,57,58] or to generate an ensemble of trajectories from it [7,15]. Given the importance of closely following the optimal course, it seems natural to use an optimization or sampling scheme which builds in the knowledge of the most probable path (3.7) and its multidimensional analog. The risks, otherwise, could be great: since a high-dimensional path will be very rough, it is easy to imagine a multistep segment of a trajectory becoming trapped in a region of the potential surface with far too few or too many steps to be even close to optimal. One idea for overcoming this difficulty would be to use a scheme capable of removing and inserting time steps, in order to search for an appropriate distribution of steps along a predefined geometric path. We intend to pursue further investigations along these lines.

Returning to the issue of finding multidimensional channels in the first place, we note that the DIMS method is ideally suited to attack this problem since it generates an ensemble of completely *independent* trajectories. Indeed, we have already developed an algorithm which has proved capable of efficiently finding distinct, important channels in a multidimensional system [59]. We have named the idea the ''soft-ratcheting algorithm,'' and we note that its efficiency is thus far limited to finding channels, rather than determining rates. The essence of the technique is simple: generate an unbiased step and accept it with a *probability* (hence the ''softness''), depending on how far the trajectory has progressed toward the target state. To complete the calculation in the DIMS formulation, one then estimates the overall acceptance probability—which is an inexpensive calculation in a large system. Note that the soft-ratcheting algorithm requires no second derivatives of the potential.

Finally, a time-scale problem could prove serious, even though we do not expect it to be nearly the handicap it is for molecular dynamics. In particular, a fundamental limitation of applying the DIMS method (or a related approach [5–15]) to multidimensional problems is the barrier-crossing time t_b. In practical terms, t_b shows up as the transient time prior to the linear regime in a plot used for rate evaluation (Fig. 4). Probability cannot accrue, after all, until crossing events have occurred. In a large system, t_b is the limiting time scale for applying a method like DIMS to computing rates. Since a reasonable number, say N, of crossing events will be needed to estimate the rate, one would have to simulate for a time exceeding Nt_b. The authors are unaware of any estimates of t_b for biomolecular systems, but we note that—for the DIMS method to potentially yield a rate estimate—t_b would have to

be less than a nanosecond for a large, explicitly solvated protein. Recent work on large time steps [60,57,58,15] holds promise, however, for attacking the time-scale problem.

VII. SUMMARY AND CONCLUSIONS

We have demonstrated substantial increases in efficiency for simulation-based calculations of transition rates in bistable potentials with modest barrier heights, $5k_BT$ and $9k_BT$, using biasing methods which are extensions of the dynamic importance sampling (DIMS) formulation [8,14]. Computations were sped up by a factor of approximately 300 for the $9k_BT$ barrier, and the primary results (Fig. 5) suggest the speed-up will increase significantly for larger barriers. The critical ingredient in our efficiency was close emulation of probable crossing events suggested by the Onsager-Machlup action, Eqs. (3.2) and (3.3).

The simple one-dimensional problem has been addressed from a variety of theoretical and numerical perspectives in an effort to pave the way for more difficult problems. In Sec. III we examined the stochastic correction—or ''Jacobian''—term (3.5) in the Onsager-Machlup action from theoretical and numerical perspectives. There, we also discussed the desired distribution of waiting times between artificial crossing events, as well as the effect (and utility) of the curvature of the potential. After presenting the explicit results for efficiency levels in rate computations, we discussed the optimal efficiency one could hope to attain in principle.

In commenting on the extension of the DIMS method to large, high-dimensional systems in Sec. VI, we noted that the problem may be conceptually broken up into two parts: finding the geometric channels and then sampling trajectories within those channels. The DIMS method is ideally suited for the first step, channel finding, and we described an explicit algorithm effective in that capacity. Our hope is that the results of the present paper will be useful in constructing techniques for the second stage, single-channel trajectory sampling.

ACKNOWLEDGMENTS

[1] R. W. Pastor, in *The Molecular Dynamics of Liquid Crystals*, edited by G. R. Luckhurst and C. A. Veracini (Kluwer Academic, Dordrecht, 1994), pp. 85–138.

[2] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).

[3] D. Chandler, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, edited by B. Berne, G. Ciccotti, and D. Coker (World Scientific, Singapore, 1998), pp. 3–23.

[4] D. Chandler, in *Classical and Quantum Dynamics in Condensed Phase Simulations* Ref. [3], pp. 51–66.

[5] L. R. Pratt, J. Chem. Phys. **85**, 5045 (1986).

[6] E. M. Sevick, A. T. Bell, and D. N. Theodorou, J. Chem. Phys. **98**, 3196 (1993).

[7] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).

[8] T. B. Woolf, Chem. Phys. Lett. **289**, 433 (1998).

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

[10] F. S. Csajka and D. Chandler, J. Chem. Phys. **109**, 1125 (1998).

[11] P. G. Bolhuis, C. Dellago, and D. Chandler, Faraday Discuss. **110**, 421 (1998).

[12] C. Dellago, P. G. Bolhuis, and D. Chandler, J. Chem. Phys. **110**, 6617 (1998).

[13] O. Mazonka, C. Jarzyński, and J. Blocki, Nucl. Phys. A **641**, 335 (1998); **650**, 499(E) (1999).

[14] D. M. Zuckerman and T. B. Woolf, J. Chem. Phys. **111**, 9475 (1999).

[15] P. Eastman, N. Grønbech-Jensen, and S. Doniach (unpublished).

[16] S. Huo and J. Straub, J. Chem. Phys. **107**, 5000 (1997).

[17] M. J. Ruiz-Montero, D. Frenkel, and J. J. Brey, Mol. Phys. **90**, 925 (1997).

[18] A. Ulitsky and D. Shalloway, J. Chem. Phys. **109**, 1670 (1998).

[19] R. E. Mortensen, J. Stat. Phys. **1**, 271 (1969).

[20] C. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 1985), Chap. 4.

[21] L. Onsager and S. Machlup, Phys. Rev. **91**, 1505 (1953).

[22] R. L. Stratonovich, *Selected Translations in Mathematical Statistics and Probability* (American Mathematical Society, Providence, RI, 1972), Vol. 10, p. 273 (translation of 1962 work in Russian).

[23] H. Haken, Z. Phys. B **24**, 321 (1976).

[24] R. Graham, Z. Phys. B **26**, 281 (1976).

[25] A. Bach, D. Dürr, and B. Stawicki, Z. Phys. B **26**, 191 (1977).

[26] C. Wissel, Z. Phys. B **35**, 185 (1979).

[27] B. Caroli, C. Caroli, and B. Roulet, J. Stat. Phys. **26**, 83 (1981).

[28] M. Dykman, M. Millonas, and V. Smelyanskiy, Phys. Lett. A **195**, 53 (1994).

[29] R. Graham, in *Fluctuations, Instabilities, and Phase Transitions*, edited by T. Riste (Plenum, New York, 1975), pp. 215–280, note slips in Eq. (2.49).

[30] M. Dykman and M. Krivoglaz, Zh. Eksp. Teor. Fiz. **77**, 60 (1979) [Sov. Phys. JETP **50**, 30 (1979)].

[31] A. J. Bray and A. J. McKane, Phys. Rev. Lett. **62**, 493 (1989).

[32] M. Dykman, Phys. Rev. A **42**, 2020 (1990).

[33] A. J. McKane, H. C. Luckock, and A. J. Bray, Phys. Rev. A **41**, 644 (1990).

[34] A. J. Bray, A. J. McKane, and T. J. Newman, Phys. Rev. A **41**, 657 (1990).

[35] H. C. Luckock and A. J. McKane, Phys. Rev. A **42**, 1982 (1990).

[36] A. J. McKane, J. Phys. A **26**, 5629 (1993).

[37] R. Olender and R. Elber, J. Mol. Struct.: THEOCHEM **398–399**, 63 (1997).

[38] M. Bier, I. Derényi, M. Kostur, and R. D. Astumian, Phys. Rev. E **59**, 6422 (1999).

[39] R. Elber and D. Shalloway, J. Chem. Phys. **112**, 5539 (2000).

[40] R. Graham, in *Quantum Statistics in Optics and Solid-State Physics*, edited by G. Höhler, Springer Tracts in Modern Physics (Springer-Verlag, Berlin, 1973), Vol. 66, pp. 1–97.

[41] U. Weiss, Phys. Rev. A **25**, 2444 (1982).

[42] V. Baibuz, V. Zitserman, and A. Drozdov, Physica A **127**, 173 (1984).

[43] M. I. Dykman *et al.*, Phys. Rev. Lett. **68**, 2718 (1992).

[44] D. G. Luchinsky and P. V. E. McClintock, Nature (London) **389**, 463 (1997).

[45] K. Müller and L. D. Brown, Theor. Chim. Acta **53**, 75 (1979).

[46] R. Czerminski and R. Elber, J. Chem. Phys. **92**, 5580 (1990).

[47] C. Choi and R. Elber, J. Chem. Phys. **94**, 751 (1991).

[48] S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).

[49] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).

[50] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, 1992).

[51] W. Rümelin, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **19**, 604 (1982).

[52] G. Batrouni *et al.*, Phys. Rev. D **32**, 2736 (1985).

[53] A. Brańka and D. Heyes, Phys. Rev. E **58**, 2611 (1998).

[54] S. Coleman, in *The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum, New York, 1979), pp. 805–942.

[55] M. Dykman, D. Luchinsky, P. McClintock, and V. Smelyanskiy, Phys. Rev. Lett. **77**, 5229 (1996).

[56] M. Berkowitz, J. D. Morgan, J. A. McCammon, and S. H. Northrup, J. Chem. Phys. **79**, 5563 (1983).

[57] R. Olender and R. Elber, J. Chem. Phys. **105**, 9299 (1996).

[58] R. Elber, J. Meller, and R. Olender, J. Phys. Chem. B **103**, 899 (1999).

[59] D. M. Zuckerman and T. B. Woolf (unpublished).

[60] R. E. Gillilan and K. R. Wilson, J. Chem. Phys. **97**, 1757 (1992).