Forward Flux Sampling-type schemes for simulating rare events: Efficiency analysis

Rosalind J. Allen, Daan Frenkel, and Pieter Rein ten Wolde*
FOM Institute for Atomic and Molecular Physics,
Kruislaan 407, 1098 SJ Amsterdam, The Netherlands
(Dated: January 7, 2014)

We analyse the efficiency of several simulation methods which we have recently proposed for calculating rate constants for rare events in stochastic dynamical systems, in or out of equilibrium. We derive analytical expressions for the computational cost of using these methods, and for the statistical error in the final estimate of the rate constant, for a given computational cost. These expressions can be used to determine which method to use for a given problem, to optimize the choice of parameters, and to evaluate the significance of the results obtained. We apply the expressions to the two-dimensional non-equilibrium rare event problem proposed by Maier and Stein. For this problem, our analysis gives accurate quantitative predictions for the computational efficiency of the three methods.

I. INTRODUCTION

Rare events are processes which happen rapidly, yet infrequently. Specialized techniques are required in order to study these events using computer simulation. This is because, in "brute force" simulations, the vast majority of the computational effort is used in simulating the uninteresting waiting periods between events, so that observing enough events for reliable statistical analysis is generally impossible. The quantities of interest from the simulation point of view are generally the rate constant for the rare transitions between the initial and final states and the properties of the Transition Path Ensemble (TPE) - the (correctly weighted) collection of transition trajectories. When computing these quantities, it is very important to know the statistical error in the calculated value, and the likely cost of the computation. In this paper, we derive approximate expressions for these quantities, for three rare event simulation methods which we proposed in a recent publication [1]. These expressions turn out to be surprisingly accurate for simulations of a model rare event problem. Our results allow us to quantify the computational efficiency of the three methods.

The three "FFS-type" simulation methods allow the computation of both the rate constant and the transition paths for rare events in equilibrium or non-equilibrium steady-state systems with stochastic dynamics. In all three methods, a series of interfaces are defined between the initial and final states. The rate constant is given by the flux of trajectories crossing the first interface, multiplied by the probability that these trajectories subsequently reach B. The latter probability is computed by carrying out a series of "trial" runs between successive interfaces; this procedure also generates transition paths, which are chains of connected successful trial runs. The methods differ in the way the trial runs are fired and the transition paths are generated. In the "forward flux sampling" (FFS) method, a collection of points is generated at the first interface and trial runs are used to propagate this collection of points to subsequent interfaces - thus generating many transition paths simultaneously. In the branched growth (BG) method, a single point is generated at the first interface and is used as the starting point for multiple trial runs to the next interface. Each successful trial generates a starting point for multiple trials to the following interface, so that a "branching tree" of transition paths is generated. In the Rosenbluth (RB) method, a single starting point is chosen at the first interface, multiple trial runs are carried out, but only one successful trial is used to propagate the path to the next interface - thus unbranched transition paths are generated. In this method, a re-weighting step is needed to ensure correctly weighted transition paths.

A range of simulation techniques for rare events in soft condensed matter systems are currently available. In Bennett-Chandler-type methods, the rate constant is obtained via a computation of a free energy barrier [2]. In Transition Path Sampling (TPS) [3], transition trajectories (paths) are generated by shooting forwards and backwards in time from already existing paths, and are then sampled using a Monte Carlo procedure. The rate constant is obtained via the computation of a time correlation function. Bennett-Chandler-type methods and TPS are suitable for systems with stochastic or deterministic dynamics, but they require knowledge of the steady state phase space density, which means that the system must be in equilibrium. While the FFS-type methods are only suitable for systems with stochastic dynamics, they do not require the phase space density to be known and can therefore be used for non-equilibrium steady states not satisfying detailed balance. To our knowledge, the only other path sampling method that is suitable for non-equilibrium systems is that proposed recently by Crooks and Chandler [4], which adopts a "TPS"-type methodology, generating new stochastic paths from old paths by changing the random number history.

The origin of the efficiency of the FFS-type methods is that they use a series of interfaces in phase space between the initial and final states to divide up the transition paths into a series of connected "partial paths".

*Electronic address: tenwolde@amolf.nl

These partial paths are generated in a ratchet-like manner - *i.e.* once a particular interface has been reached, the system configuration is stored and is used to initiate trial runs to the next interface. Many other rare event techniques also use a series of interfaces in phase space. In Transition Interface Sampling (TIS) [5] and Partial Path Transition Interface Sampling (PPTIS) [6], interfaces are used to facilitate the generation of transition paths by a TPS-like procedure. In Milestoning [7], trajectories are generated between interfaces assuming a steady-state distribution at each interface, while string methods [8, 9] use a series of planes in phase space to allow a trajectory connecting the initial and final states to relax to the minimum free energy path. The advantages of the FFS-type methods over other transition path and rate constant calculation methods are that no assumptions are made about "loss of memory" during the transition, no *a priori* knowledge is required of the steady state phase space density, and the rate constant is obtained in a simple and straightforward way. We have recently become aware that the BG method bears resemblance to the RESTART method, used for simulating telecommunications networks [10, 11, 12] (this approach was originally introduced by Bayes [13]). The efficiency of that method has also been analysed [11]. A related method, known as Weighted Ensemble Brownian Dynamics, has been applied to protein association reactions [14].

The key aim of a rare event simulation technique is to calculate the rate constant, or in some cases, obtain the TPE, with enhanced efficiency, compared to brute force simulations. However, quantifying the efficiency of a particular simulation method is often difficult. Our aim in this paper is to derive simple but accurate expressions for the computational cost and statistical accuracy of the three FFS-type methods. We define the "efficiency" of the methods to be the inverse of the product of the cost and the variance in the calculated rate constant; our results then allow us to analyse the efficiency of the methods in a systematic way. From a practical point of view, we expect the expressions derived here to be of use to those carrying out simulations in two ways. Firstly, when faced with a rare event problem, one often has a limited amount of computer time available, and specific requirements as to the desired accuracy of the calculated rate constant. Analytical expressions for the cost and statistical accuracy would allow one to estimate, before beginning the calculation, whether the desired accuracy can be obtained within the available time, and thus to make an informed decision as to which, if any, method to use. Secondly, after completing a rate constant calculation, one needs to obtain error bars on the resulting value - this is especially important for rare events, where both experimental and simulation results can be highly inaccurate. In general, error estimation requires the calculation to be repeated several times, which is computationally expensive. However, if analytical expressions were available for the statistical accuracy, in terms of quantities which were already measured during the rate constant calculation, one could obtain the error bars on the predicted rate constant, to within reasonable accuracy, without the need for lengthy additional calculations. In this paper, we derive such analytical expressions.

Approximate expressions are derived for the cost, in simulation steps, and for the variance in the calculated rate constant, for the three FFS-type methods. We initially treat the simple case where all trials fired from one interface have equal probability of succeeding. We then move on to the more realistic case where the probability of reaching the next interface depends on the identity of the starting point. To this end, we include in our calculations the "landscape variance" - the variance in the probability of reaching the next interface, due to the characteristic "landscape" for this particular rare event problem. Our expressions are functions of user-defined parameters, such as the number of trial runs per point at a particular interface, as well as parameters characterizing the rare event problem itself, such as the probability that a trial run succeeds in reaching the next interface.

We analyse the efficiency of the three methods as a function of the parameters, for a "generalized" model system. We find that the optimum efficiency is similar for all three methods, but that the effects of changing the parameter values are very different for the three methods. In particular, the BG method performs well only within a narrow range of parameter values, while the FFS and RB methods are more robust to changes in the parameters. The RB method has consistently lower efficiency, due to its requirement for an acceptance/rejection step - however, RB may be more suitable for applications where analysis of transition paths as well as rates is needed, or where storage of configurations is very expensive.

To test the accuracy of our predictions in the context of a real simulation problem, we then apply the three FFS-type methods to the two-dimensional non-equilibrium rare event problem proposed by Maier and Stein [15, 16, 17]. We measure the computational cost of the methods and the variance in the final value of the rate constant, and we compare these to the cost and variance predicted by the expressions derived earlier. We find that the expressions give remarkably good predictions, both for the cost and the variance. This suggests that the expressions can, indeed, be used to give accurate and easy-to-calculate error estimates for real simulation problems.

In Section II, we briefly describe the three FFS-type methods. Expressions for the computational cost and for the statistical error in the calculated rate constant are derived in Section III. In Section IV, these expressions are shown to be accurate for the two-dimensional non-equilibrium rare event problem proposed by Maier and Stein. Finally, we discuss our conclusions in Section V.

II. BACKGROUND: FFS-TYPE METHODS

The FFS-type methods use the "effective positive flux" expression for the rate constant, which was rigorously derived by Van Erp *et al* [5, 6, 18, 19, 20]. The rare event consists of a transition between two regions of phase space $A(x)$ and $B(x)$, where x denotes the coordinates of the phase space. The transition occurs much faster than the average waiting time in the A state. We assume that a parameter $\lambda(x)$ can be defined, such that $\lambda < \lambda_A$ in A and $\lambda > \lambda_B$ in B. A series of values of λ, $\lambda_0 \dots \lambda_n$, are chosen such that $\lambda_0 \equiv \lambda_A$, $\lambda_n \equiv \lambda_B$ and $\lambda_i < \lambda_{i-1}$. These must constitute a series of non-intersecting surfaces in phase space, such that any transition path leading from A to B passes through each surface in turn. This is illustrated in Figure 1.

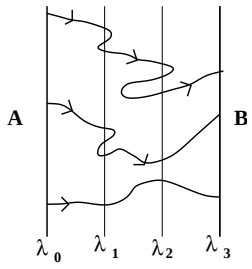

FIG. 1: Schematic illustration of the definition of regions A and B and the interfaces $\lambda_0 \dots \lambda_n$ (Here, $n = 3$). Three transition paths are shown.

The rate constant k_{AB} can be expressed as [20]

$$k_{AB} = \frac{\overline{\Phi}_{A,n}}{\overline{h}_{\mathcal{A}}} = \frac{\overline{\Phi}_{A,0}}{\overline{h}_{\mathcal{A}}} P(\lambda_n|\lambda_0). \qquad (1)$$

In Eq. (1), $h_{\mathcal{A}}$ is a history-dependent function describing whether the system was more recently in A or B: $h_{\mathcal{A}} = 1$ if the system was more recently in A than in B, and $h_{\mathcal{A}} = 0$ otherwise [5, 18, 20]. The over-bar denotes a time average. $\Phi_{A,j}$ is the flux of trajectories with $h_{\mathcal{A}} = 1$ that cross λ_j for the first time - *i.e.* those trajectories that cross λ_j, having been in A more recently than any previous crossings of λ_j. $P(\lambda_j|\lambda_i)$ is the probability that a trajectory that comes from A and crosses λ_i for the first time will subsequently reach λ_j before returning to A: thus $P(\lambda_n|\lambda_0)$ is the probability that a trajectory that leaves A and crosses λ_0 will subsequently reach B before returning to A. Eq.(1) states that the flux of trajectories from A to B can be expressed as the flux leaving A and crossing λ_0, multiplied by the probability that one of these trajectories will subsequently arrive at B rather that returning to A. $P(\lambda_n|\lambda_0)$ can be expressed as the product of the probabilities of reaching each successive interface from the previous one, without returning to A:

$$P(\lambda_n|\lambda_0) = \prod_{i=0}^{n-1} P(\lambda_{i+1}|\lambda_i) \qquad (2)$$

For simplicity of notation, in what follows, we define $P_B \equiv P(\lambda_n|\lambda_0)$, $p_i \equiv P(\lambda_{i+1}|\lambda_i)$, $q_i \equiv 1 - p_i$ and $\Phi \equiv \overline{\Phi}_{A,0}/\overline{h}_{\mathcal{A}}$. We also use the superscript "e" to indicate an estimated value of a particular quantity.

Previously, we described in detail three different approaches - the "forward flux sampling" (FFS), "branched growth" (BG) and "Rosenbluth" (RB) methods - to calculating k_{AB}, based on expressions (1) and (2) [1, 21]. For completeness, we briefly repeat the description here.

A. Forward flux sampling

In FFS, the flux Φ is measured using a free simulation in the basin of attraction of region A. When the system leaves A and crosses λ_0 for the first time (since leaving A), its phase space coordinates are stored and the run is continued. In this way, a collection of N_0 points at λ_0 is generated, after which the simulation run is terminated.

The probabilities p_i are then estimated using a trial run procedure. Beginning with the collection of points at λ_0, a large number M_0 of trials are carried out. For each trial, a point is selected at random from the collection at λ_0. This point is used to initiate a simulation run, which is continued until the system either crosses the next interface λ_1, or re-enters A. If λ_1 is reached, the final point of the run is stored in a new collection. After M_0 trials, p_0 is given by $N_s^{(0)}/M_0$, where $N_s^{(0)}$ is the number of trials which reached λ_1. The probability p_1 is then estimated in the same way: the new collection of points at λ_1 is used to initiate M_1 trial runs to λ_2 (or back to A), generating a new collection of points at λ_2, and so on. Finally, the rate constant is obtained using Eqs (1) and (2).

FFS generates transition paths according to their correct weights in the TPE [1, 21]. In order to analyse these transition paths, one begins with the collection of trial runs which arrive at λ_B from λ_{n-1} and traces back the sequence of connected partial paths which link them to region A. The resulting transition paths are branched - *i.e.* a single point at λ_0 can be the starting point of multiple transition paths.

B. The branched growth method

In the BG method, which was inspired by techniques for polymer sampling [2, 22, 23], branched transition paths are generated one by one, rather than simultaneously, as in FFS. The generation of each path begins with a single point at λ_0, obtained using a simulation in the basin of attraction of A, as in the FFS method. This point is used to initiate k_0 trial runs, which are continued until they either reach λ_1, or return to A. Each of the $N_s^{(0)}$ end points at λ_1 becomes a starting point for k_1 trial runs to λ_2 or back to A. Each of the $N_s^{(1)}$ successful trial runs to λ_2 initiates k_2 trials to λ_3, and so on until λ_n

is reached. An estimate P_B^e of P_B is obtained as the total number of branches that eventually reach λ_n, divided by the total possible number: $P_B^e = N_s^{(n-1)} / \prod_{i=0}^{n-1} k_i$. If, at any interface, no trials were successful, $P_B^e = 0$. To generate the next branching path, we obtain a new starting point at λ_0 from the simulation in the basin of attraction of A. After many branching paths have been generated, an average is taken over the P_B^e values of all the paths. The flux Φ is meanwhile obtained from the simulation run in region A. The branched transition paths that are generated in the BG method are correctly weighted members of the TPE [1]. We note that the BG method bears resemblance to methods developed for telecommunication networks [10, 11, 12] and to a method used for protein association [14].

C. The Rosenbluth method

The RB path sampling method is related to the Rosenbluth scheme for sampling polymer configurations [2, 24, 25]. The RB method generates unbranched transition paths, one at a time. An initial point at λ_0 is obtained using a simulation in the A basin, which is continued until the trajectory crosses λ_0 for the first time, as in the FFS and BG methods. This point is used to initiate k_0 trials, which are continued until they either reach λ_1 or return to A. If $N_s^{(0)} > 0$ of these trials reach λ_1, one successful trial is selected at random and its end point at λ_1 is used to initiate k_1 trials to λ_2 or back to A. Once again a successful trial is chosen at random and the process is repeated until either no trials are successful or λ_n is reached. The generation of the next path then begins with a new point at λ_0, obtained using the simulation run in the A basin.

The Rosenbluth method as outlined above does not, however, generate paths according to their correct weights in the TPE: for correct sampling, paths must be re-weighted by a "Rosenbluth factor". The Rosenbluth factor for a partial path up to interface i is given by:

$$W_i = \prod_{j=0}^{i-1} N_s^{(j)} \tag{3}$$

Note that the re-weighting factor W_i depends on the number of successful trials obtained at all the previous interfaces, while generating the path up to λ_i. The correct re-weighting can be achieved using a Metropolis-type acceptance/rejection scheme [2], in which a newly generated path is either accepted or rejected based on a comparison of its Rosenbluth factor with that of a previously generated path. Ensemble averages of any quantity of interest are then taken over all accepted paths. Here, the quantity which we wish to calculate is the probability p_i that a trial run fired from λ_i will reach λ_{i+1}, for each interface i. When we fire k_i trial runs from λ_i, we obtain an estimate for p_i: $p_i^e \equiv N_s^{(i)}/k_i$. We require the correctly weighted ensemble average for p_i^e at each interface i; we note, however, that the same procedure could also be used to calculate the ensemble average of any other property of the ensemble of paths from λ_0 to λ_i.

From a practical point of view, each interface has associated with it *two* values of W_i and p_i^e. The first set of values: $W_i^{(\mathrm{n})}$ and $p_i^{e(\mathrm{n})}$, are associated with the transition path that is currently being generated (the "new" path). $W_i^{(\mathrm{n})}$ depends on the number of successful trials generated in creating this transition path as far as λ_i, and $p_i^{e(\mathrm{n})} \equiv N_s^{(i)}/k_i$ depends on the number of successful trials fired from the point at λ_i to λ_{i+1}. The other set of values, $W_i^{(\mathrm{o})}$ and $p_i^{e(\mathrm{o})}$, are the "old" values for this interface. These values correspond to the last "acceptance" event at this interface.

The recipe for obtaining k_{AB} within the RB method is as follows. Transition paths are generated as described above. When the path generation procedure reaches λ_i, we calculate the Rosenbluth factor $W_i^{(\mathrm{n})}$ (using Eq.(3)) and we fire k_i trial runs to obtain $p_i^{e(\mathrm{n})} \equiv N_s^{(i)}/k_i$. We then calculate the ratio $W_i^{(\mathrm{n})}/W_i^{(\mathrm{o})}$ and draw a random number $0 < s < 1$. If $s < W_i^{(\mathrm{n})}/W_i^{(\mathrm{o})}$, an acceptance event takes place. In this case, the previous values of $W_i^{(\mathrm{o})}$ and $p_i^{e(\mathrm{o})}$ are replaced by the newly obtained values $W_i^{(\mathrm{n})}$ and $p_i^{e(\mathrm{n})}$. If, however, $s > W_i^{(\mathrm{n})}/W_i^{(\mathrm{o})}$, a rejection occurs and $W_i^{(\mathrm{o})}$ and $p_i^{e(\mathrm{o})}$ remain unchanged for this interface. Regardless of the outcome of the acceptance/rejection step, the accumulator for the probability p_i^e is incremented by the current value of $p_i^{e(\mathrm{o})}$ - this may be either a newly generated value (if an acceptance just occurred) or an old value that may have been already added to the accumulator several times (if several rejections have happened in a row). To proceed to the next interface, a successful trial run is chosen out of those that have been newly generated, and its end point at λ_{i+1} is used as the starting point for k_{i+1} trial runs to λ_{i+2}. A corresponding acceptance/rejection step is then carried out at λ_{i+1}. We note that the "old" values $W_i^{(\mathrm{o})}$ and $p_i^{e(\mathrm{o})}$ for different interfaces need not correspond to the same transition path. After many complete transition paths have been generated, k_{AB} is obtained using Eq.(1), where an estimate of the flux Φ is calculated from the simulation run in region A. A "pseudo-code" corresponding to the above procedure is given in our previous publication [1], together with a description of an alternative, "Waste Recycling" [26] re-weighting scheme. In this paper, however, we shall consider only the Metropolis acceptance/rejection approach.

III. COMPUTATIONAL EFFICIENCY

In this section, we derive approximate expressions for the computational efficiency of the three methods. Following Mooij and Frenkel [27], we use the following defi-

nition for the efficiency, $\mathcal{E}$:

$$\mathcal{E} = \frac{1}{\mathcal{C}\mathcal{V}} \tag{4}$$

In Eq. (4), $\mathcal{C}$ represents the computational cost, which we define to be the average number of simulation steps, per initial point at λ_0. The statistical error in the estimated value k^e_{AB} of the rate constant is represented by $\mathcal{V}$. Denoting the mean (expectation value) of variable u by $E[u]$ and its variance by $V[u]$, we define $\mathcal{V}$ to be the variance $V[k^e_{AB}]$, per initial point at λ_0, divided by the square of the expectation value $E[k^e_{AB}]$:

$$\mathcal{V} = \frac{N_0 V[k^e_{AB}]}{(E[k^e_{AB}])^2} = N_0 \frac{V[k^e_{AB}]}{k^2_{AB}} \tag{5}$$

where N_0 is the number of starting points at λ_0 used in obtaining the estimate k^e_{AB}. The expectation value of k^e_{AB} is, of course, the true rate constant: $E[k^e_{AB}] = k_{AB}$. The error bar for k^e_{AB} is given by $k_{AB}\sqrt{\mathcal{V}/N_0}$.

A. Computational Cost

We define the computational cost $\mathcal{C}$ of a particular method to be the average number of simulation steps required by that method, per starting point at λ_0. In making this definition, we ignore any other contributions to the CPU time, such as memory storage. To estimate the value of $\mathcal{C}$, we consider a generic system that makes a rare transition between states A and B. A parameter λ and interfaces $\lambda_0 \ldots \lambda_n$ are chosen as in Section II.

There are two contributions to the cost $\mathcal{C}$. The first is the average cost R, in simulation steps, of generating one starting point at λ_0. This is related to the flux Φ from the A region to λ_0 by $R = 1/(\Phi dt)$, where dt is the simulation timestep.

The second contribution to $\mathcal{C}$ is the cost of the trial run procedure. We first consider the cost C_i of firing one trial run from interface λ_i. The run is continued until it reaches either the next interface λ_{i+1} (with probability p_i), or the boundary λ_A of region A (with probability q_i). We make the assumption that the average length (in simulation steps) of a trajectory from interface λ_i to another interface λ_j is linearly proportional to $|\lambda_j - \lambda_i|$, with proportionality constant S. C_i is then given by:

$$C_i = S\left[p_i(\lambda_{i+1} - \lambda_i) + q_i(\lambda_i - \lambda_A)\right] \tag{6}$$

The basis for the assumption of linearity in Eq.(6) is that we suppose that the system undergoes one-dimensional diffusion along the λ coordinate in the presence of a "drift force" of fixed magnitude. For an equilibrium system, the origin of the drift force is the free energy barrier. Farkas and Fülöp have presented analytical solutions [28] for the mean time to capture for a particle undergoing one-dimensional diffusion with constant drift force, in the presence of two absorbing boundaries. In Appendix A, we show how these results lead to Eq.(6). Eq.(6) is shown to be valid for the two-dimensional Maier-Stein problem in Section IV A (Figure 7).

Expressions for the cost

Given Eq.(6), we can compute the average cost $\mathcal{C}$ per starting point at λ_0 of the three methods.

In FFS, we make M_i trial runs from interface i and, providing at least one of these is successful, we proceed to the next interface $i+1$. In practice, M_i is expected to be large enough that at least one trial run reaches λ_{i+1}. In this case, the expected cost per starting point at λ_0 is:

$$\mathcal{C}^{\rm ffs} = R + \frac{1}{N_0}\sum_{i=0}^{n-1} M_i C_i \tag{7}$$

Defining k_i such that $k_i = M_i/N_0$, Eq.(7) can be rewritten as:

$$\mathcal{C}^{\rm ffs} = R + \sum_{i=0}^{n-1} k_i C_i \tag{8}$$

If, however, M_i is small, we must take account of the possibility that none of the trial runs from λ_i reach λ_{i+1}. In this case, the FFS procedure is terminated at interface i and the cost is accordingly reduced. Since the probability of reaching interface $i > 0$ is $\prod_{j=0}^{i-1}\left(1 - q_j^{M_j}\right)$ (this is the probability that at least one trial is successful at all interfaces $j < i$), Eq.(8) is replaced by:

$$\mathcal{C}^{\rm ffs} = R + k_0 C_0 + \sum_{i=1}^{n-1}\left[k_i C_i \prod_{j=0}^{i-1}\left(1 - q_j^{N_0 k_j}\right)\right] \tag{9}$$

Although the cost is reduced by failing to reach later interfaces, this of course results in a less accurate prediction of the rate constant, since the terminated FFS calculation makes no contribution to the estimate of p_i for later interfaces. This will be reflected in our expression for the statistical error in Section III B.

We now turn to the BG method. Here, we generate a "branching tree" of paths, with N_i points at interface i originating from a single point at λ_0. We fire k_i trial runs for each of these N_i points. The average value of N_i is:

$$\overline{N_i} = \prod_{j=0}^{i-1} p_j k_j \qquad (i > 0) \tag{10}$$

Of course $\overline{N_0} = 1$. The average cost per starting point at λ_0 is therefore:

$$\begin{aligned} \mathcal{C}^{\rm bg} &= R + \sum_{i=0}^{n-1} k_i C_i \overline{N_i} \\ &= R + k_0 C_0 + \sum_{i=1}^{n-1}\left[k_i C_i \prod_{j=0}^{i-1} p_j k_j\right] \end{aligned} \tag{11}$$

Finally, we come to the RB method. In this algorithm, we generate unbranched paths by firing k_i trials from interface i, choosing one successful trial at random and proceeding to interface $i+1$. If no trial runs are successful, we start again with a new point at λ_0. The probability of reaching interface $i>0$ is $\prod_{j=0}^{i-1}\left(1-q_j^{k_j}\right)$. The cost of the RB method, per starting point at λ_0, is therefore:

$$\mathcal{C}^{\rm rb} = R + k_0 C_0 + \sum_{i=1}^{n-1}\left[k_i C_i \prod_{j=0}^{i-1}\left(1-q_j^{k_j}\right)\right] \qquad (12)$$

Once again, the "price" of failing to reach later interfaces will be paid in the form of an increased variance in the calculated rate constant. The effect of the Metropolis acceptance/rejection step in the RB method appears only in the variance in k^e_{AB} (Section III B), and not in the cost.

Illustration

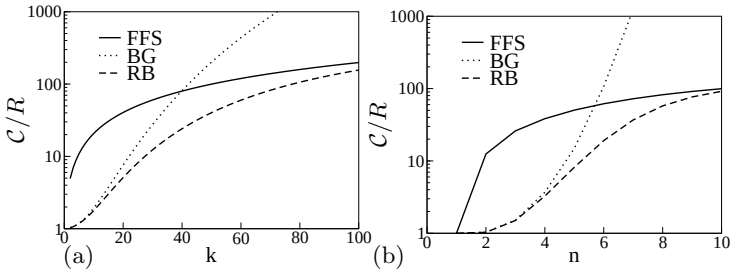

FIG. 2: Cost $\mathcal{C}/R$, for evenly spaced interfaces, $p_i = p$, $k_i = k$, $R = S$, $N_0 = 1000$ and $P_B = 10^{-8}$. (a): $\mathcal{C}/R$ as a function of k, for $n = 5$. (b): $\mathcal{C}/R$ as a function of n, for $k = 25$.

For the purposes of illustration, let us consider a hypothetical rare event problem for which $\lambda_0 = \lambda_A = 0$ and $\lambda_n = \lambda_B = 1$. We suppose that the interfaces are evenly spaced in λ, have equal values of p_i, and that the firing parameter k_i is the same at each interface: *i.e.* $\lambda_i = i/n$, $p_i = P_B^{1/n}$ (from Eq.(2)) and $k_i = k$. We also suppose that $R = S$ and $N_0 = 1000$. The resulting values of the cost $\mathcal{C}$, obtained from Eqs (9), (11) and (12), are plotted in Figure 2a and b as functions of k and n. In the regime of small k or small n (implying small p), the BG and RB methods converge, while the cost of the FFS method is higher. This is because, for BG and RB, the probability of reaching later interfaces is low and the cost is dominated by the trial runs fired from early interfaces. The FFS procedure is less likely to be terminated at early interfaces (note the factor of $1-q_i^{N_0 k_i}$ in Eq.(9) as opposed to $1-q_i^{k_i}$ in Eq. (12)), and is therefore more expensive, per initial point at λ_0. In the regime of large k or large n (implying large p), a different scenario emerges. Here, the BG method becomes by far the most expensive, with a cost that increases dramatically with increasing k or n. This effect is due to the rapidly increasing number of branches per starting point at λ_0. In this regime, the FFS and RB methods converge to the same cost, since Eqs (9) and (12) become equivalent when $1-q^k \approx 1-q^{N_0 k} \approx 1$.

B. Statistical Error

We now turn to the relative variance $\mathcal{V}$ in the estimated value k^e_{AB} of the rate constant, per starting point at λ_0. k^e_{AB} is the product of the estimated flux through λ_0, multiplied by the estimated probability of subsequently reaching B: $k^e_{AB} = \Phi^e P^e_B$ (Eq.(1)).

In this paper, we shall ignore the error in Φ^e. Φ^e is obtained by carrying out a simulation run in the basin of attraction of A and measuring the average number of simulation steps between successive crossings of λ_0 (coming directly from A). As long as λ_0 is positioned close enough to the A region, the simulation run in A can be made long enough to estimate Φ with high accuracy, with a computational cost that is minimal compared to the cost of estimating P_B. We therefore obtain:

$$\mathcal{V} \equiv N_0 \frac{V[k^e_{AB}]}{(E[k^e_{AB}])^2} \approx N_0 \frac{\Phi^2 V[P^e_B]}{(\Phi E[P^e_B])^2} = N_0 \frac{V[P^e_B]}{P_B^2} \qquad (13)$$

In Eq.(13), we have used the general relation [29]

$$V[ax] = a^2 V[x] \qquad (14)$$

where a is a constant.

In what follows, we shall make the important assumption that the numbers $N_s^{(i)}$ of successful trial runs at different interfaces i are *uncorrelated* - *i.e.* that if, during the generation of a transition path, one is particularly successful or unsuccessful at interface i, this will have no effect on the chances of success at interface $i+1$. In reality, of course, there will be correlation between interfaces, especially if the interfaces are closely spaced or the system dynamics have a large degree of "memory". We expect this assumption to be the major limiting factor in the applicability of our results to real systems; however, as we shall see in Section IV, the results are surprisingly accurate for the two-dimensional Maier-Stein problem. We expect that the expressions derived here could be modified to include the effects of correlations between interfaces; for highly correlated systems this may prove necessary.

Expressions for the variance

The basis of our analysis is the fact that on firing k_i trial runs from interface i, the number of successful trials $N_s^{(i)}$ is binomially distributed [29], with mean

$$E[N_s^{(i)}] = k_i p_i \qquad (15)$$

and variance

$$V[N_s^{(i)}] = k_i p_i q_i \tag{16}$$

For now, we assume that all trial runs fired from interface λ_i have equal probability p_i of reaching λ_{i+1}. This assumption will later be relaxed. We shall need to express the variance in P_B^e in terms of the variance $V[p_i^e]$ in the estimated values of p_i at each interface. To do this, we recall that $P_B^e = \prod_{i=0}^{n-1} p_i^e$ (Eq.(2)), and we make use of the following relation [29]:

$$V[f(x,y,\dots)] = \left(\frac{\partial f}{\partial x}\right)^2 V[x] + \left(\frac{\partial f}{\partial y}\right)^2 V[y] + \dots \tag{17}$$

where $f(x,y,\dots)$ is a function of multiple uncorrelated variables $x, y, \dots$ and the partial derivatives are evaluated with all variables at their mean values. By "uncorrelated variables" we mean that the covariance $\mathrm{Cov}[u,v] = 0$ for all pairs of variables u and v. Identifying $x, y \dots$ with $p_i^e, p_{i+1}^e \dots$ and taking $f(p_0^e \dots p_{n-1}^e) = \prod_{i=0}^{n-1} p_i^e$, we find that $\partial f/\partial p_i^e = [\prod_{j=0}^{n-1} p_j^e]/p_i = P_B^e/p_i^e$, so that

$$V[P_B^e] = \sum_{i=0}^{n-1} E\left[\frac{P_B^e}{p_i^e}\right]^2 V[p_i^e] \approx P_B^2 \sum_{i=1}^{n} \frac{V[p_i^e]}{p_i^2} \tag{18}$$

We now use the above results to calculate $\mathcal{V}$ for the FFS method. In this method, we begin with a collection of N_0 points at λ_0. For each interface, p_i^e is obtained by firing $M_i \equiv N_0 k_i$ trial runs: $p_i^e = N_s^{(i)}/M_i$, where $N_s^{(i)}$ is the number of trials which reach λ_{i+1}. Using Eq.(14), $V[p_i^e] = V[N_s^{(i)}]/M_i^2$. Using Eq.(16), we find that $V[N_s^{(i)}] = M_i p_i q_i$. Noting also that $E[p_i^e] = p_i$ and using Eq.(18), we obtain

$$V^{\mathrm{ffs}}[P_B^e] = P_B^2 \sum_{i=0}^{n-1} \frac{q_i}{p_i M_i} = \frac{P_B^2}{N_0} \sum_{i=0}^{n-1} \frac{q_i}{p_i k_i} \tag{19}$$

and from Eq.(13)

$$\mathcal{V}^{\mathrm{ffs}} = \sum_{i=0}^{n-1} \frac{q_i}{p_i k_i} \tag{20}$$

As for the cost calculation, we have assumed that M_i is large enough that there is always at least one trial run which reaches the next interface. If this is not the case, we must also take account of the possibility that interfaces $i > 0$ may not be reached. The probability of reaching interface $i > 0$ is $\prod_{j=0}^{i-1}\left(1 - q_j^{M_j}\right)$, so that

$$V[p_i^e] = \frac{p_i q_i \left(1 - q_i^{M_i}\right)}{M_i \prod_{j=0}^{i}\left(1 - q_j^{M_j}\right)} \tag{21}$$

Eq.(21) is written in this form so that for $i = 0$, we recover $V[p_0^e] = p_i q_i / M_i$. Eqs (19) and (20) must then be replaced by:

$$V^{\mathrm{ffs}}[P_B^e] = P_B^2 \left[\sum_{i=0}^{n-1} \frac{q_i \left(1 - q_i^{M_i}\right)}{p_i M_i \prod_{j=0}^{i}\left(1 - q_j^{M_j}\right)}\right] \tag{22}$$

and

$$\mathcal{V}^{\mathrm{ffs}} = \sum_{i=0}^{n-1} \frac{q_i}{p_i k_i} \left[\frac{1 - q_i^{N_0 k_i}}{\prod_{j=0}^{i}\left(1 - q_j^{N_0 k_j}\right)}\right] \tag{23}$$

We now turn to the BG method. Here, we begin with a single point at λ_0. From this point, we generate a branching "tree" of paths connecting A to B. The value of P_B is estimated by

$$P_B^e = \frac{N_s^{(n-1)}}{\prod_{i=0}^{n-1} k_i} \tag{24}$$

where $N_s^{(n-1)}$ is the total number of trials reaching $\lambda_n \equiv \lambda_B$. We denote the number of points in the branching tree at interface i by N_i. For a given number N_{n-1} of points at λ_{n-1}, the total number of trials fired is $N_{n-1}k_{n-1}$ and the variance in $N_s^{(n-1)}$ is $V[N_s^{(n-1)}|N_{n-1}] = N_{n-1}k_{n-1}p_{n-1}q_{n-1}$ (using Eq.(16)). However, the situation is complicated by the fact that N_{n-1} itself varies; in fact, N_{n-1} is simply the number of successful trial runs reaching λ_{n-1} from λ_{n-2}, and in general:

$$N_i = N_s^{(i-1)} \qquad [i > 0] \tag{25}$$

At this point, we need to calculate the variance in a quantity Y which is conditional upon the value of another quantity X. Here, and several times in the rest of the paper, we will use the general relation

$$V[Y] = E\left[V[Y|X]\right] + V\left[E[Y|X]\right] \tag{26}$$

where the mean and variance on the r.h.s. of Eq.(26) are taken over the distribution of values of X. Since $E[N_s^{(n-1)}|N_{n-1}] = N_{n-1}k_{n-1}p_{n-1}$,

$$\begin{aligned} V\left[E\left[N_s^{(n-1)}|N_{n-1}\right]\right] &= k_{n-1}^2 p_{n-1}^2 V\left[N_{n-1}\right] \\ &= k_{n-1}^2 p_{n-1}^2 V\left[N_s^{(n-2)}\right] \end{aligned} \tag{27}$$

(using Eqs (14) and (25)). We also know that

$$\begin{aligned} E\left[V\left[N_s^{(n-1)}|N_{n-1}\right]\right] &= k_{n-1}p_{n-1}q_{n-1}E\left[N_{n-1}\right] \\ &= k_{n-1}p_{n-1}q_{n-1}\prod_{i=0}^{n-2} k_i p_i \end{aligned} \tag{28}$$

so that

$$V[N_s^{(n-1)}] = q_{n-1}\prod_{i=0}^{n-1} k_i p_i + k_{n-1}^2 p_{n-1}^2 V\left[N_s^{(n-2)}\right] \tag{29}$$

Using the same arguments, we can generalize Eq.(29) to

$$V[N_s^{(i)}] \;= q_i \prod_{j=0}^{i} k_j p_j + k_i^2 p_i^2 V[N_s^{(i-1)}] \quad [i>0] \qquad (30)$$
$$q_i k_i p_i \qquad [i=0]$$

Using Eq.(30), we can solve Eq.(29) recursively, to obtain $V[N_s^{(n-1)}]$. Using Eqs.(24) and (14), we then arrive at the variance in the estimated value of P_B:

$$V^{\rm bg}[P_B^e] = \frac{P_B^2}{N_0} \sum_{i=0}^{n-1} \frac{q_i}{\prod_{j=0}^{i} p_j k_j} \qquad (31)$$

where we have divided by N_0 to account for the fact that P_B^e is calculated by averaging results over N_0 starting points at λ_0. We then obtain from Eq.(13):

$$\mathcal{V}^{\rm bg} = \sum_{i=0}^{n-1} \frac{q_i}{\prod_{j=0}^{i} p_j k_j} \qquad (32)$$

Finally, let us derive the equivalent expression for the RB method. Here, we again use Eq.(18). If we ignore for the moment the effect of the acceptance rejection step, we can use Eqs.(16) and (14) to obtain an expression for the variance in p_i^e:

$$V^{\rm rb}[p_i^e] = \frac{p_i q_i}{N_0 k_i} \frac{\left(1-q_i^{k_i}\right)}{\prod_{j=0}^{i}\left(1-q_j^{k_j}\right)} \qquad (33)$$

where we have taken account of the fact that the probability of reaching interface $i>0$ is $\prod_{j=0}^{i-1}\left(1-q_j^{k_j}\right)$, and that the p_i^e value is averaged over N_0 separate path generations. Eq.(33) is very similar to the FFS result, Eq.(21).

The Metropolis acceptance/rejection step (described in Section II) increases the variance in p_i^e. On reaching interface i, we fire k_i trials and obtain an estimate $p_i^{e,(\rm n)} = N_s^{(i)}/k_i$. We either accept or reject this estimate. If we reject, $p_i^{e,(\rm n)}$ makes no contribution to the average value of p_i^e - instead, the previously accepted estimate, $p_i^{e,(\rm o)}$, is added to the average, even though $p_i^{e,\rm o}$ was already added to the average in the previous acceptance/rejection step. If, instead, we accept $p_i^{e,(\rm n)}$, it makes a contribution to p_i^e, and, if the subsequent estimates happen to be rejected, it may repeat this contribution multiple times. The final estimate, p_i^e, is therefore an average over all the values of $N_s^{(i)}/k_i$ that were generated, weighted by the number of times Q that each of these values contributed to p_i^e:

$$p_i^e = \frac{\sum_{l=1}^{N_g} Q_l \left[N_s^{(i)}/k_i\right]_l}{N_g^{(i)}} \qquad (34)$$

where the sum is over all generated $N_s^{(i)}/k_i$ values and $N_g^{(i)}$ is the total number of these. In fact,

$$N_g^{(i)} = N_0 \frac{\prod_{j=0}^{i}\left(1-q_j^{k_j}\right)}{\left(1-q_i^{k_i}\right)} \qquad (35)$$

since the number of times we fire trials from λ_i is simply the number of times we begin a path generation from λ_0 and succeed in reaching λ_i. Using Eq.(14), the variance p_i^e is then

$$V[p_i^e] = \frac{\sum_{l=1}^{N_g^{(i)}} Q_l^2 V\left[N_s^{(i)}/k_i\right]_l}{(N_g^{(i)})^2} = \frac{V\left[N_s^{(i)}\right]}{k_i^2 (N_g^{(i)})^2} \sum_{l=1}^{N_g^{(i)}} Q_l^2 \qquad (36)$$

(assuming that the distributions of the stochastic variables Q_l and $[N_s^{(i)}]_l$ are uncorrelated). Eq.(36) is equivalent to:

$$V[p_i^e] = \frac{V\left[N_s^{(i)}\right]}{k_i^2 N_g^{(i)}} \sum_{Q=0}^{\infty} Q^2 P(Q) = \frac{p_i q_i}{k_i N_g^{(i)}} \sum_{Q=0}^{\infty} Q^2 P(Q) \qquad (37)$$

In order to find the distribution $P(Q)$, we define a new variable θ_i. θ_i is the probability that we accept a newly generated estimate $p_i^{e,(\rm n)} = N_s^{(i)}/k_i$. $P(Q)$ is then:

$$P(Q) = \quad (1-\theta_i) \qquad Q=0 \qquad (38)$$
$$P(Q) = \; \theta_i^2 (1-\theta_i)^{Q-1} \qquad Q>0$$

Eq.(38) can be understood as follows: $Q=0$ corresponds to a $p_i^{e,(\rm n)}$ value that is generated but is immediately rejected and therefore contributes zero times to the average. This occurs with probability $1-\theta_i$. $Q>0$ corresponds to a $p_i^{e,(\rm n)}$ value that is generated and accepted (with probability θ_i) - the next $Q-1$ values that are generated are rejected (with probability $(1-\theta_i)^{Q-1}$), then finally a new value is generated which is accepted (with probability θ_i), so that the original value ceases to contribute to the average. The distribution (38) has the property that [30]

$$\sum_{Q=0}^{\infty} Q^2 P(Q) = \frac{2-\theta_i}{\theta_i} \qquad (39)$$

so that Eq.(37) for the variance in p_i^e per point at λ_i becomes

$$V[p_i^e] = \frac{p_i q_i}{k_i N_g^{(i)}} \left[\frac{2-\theta_i}{\theta_i}\right] \qquad (40)$$

Using Eq.(35), we obtain:

$$V^{\rm rb}[p_i^e] = \frac{p_i q_i}{N_0 k_i} \left[\frac{(2-\theta_i)}{\theta_i}\right] \frac{\left(1-q_i^{k_i}\right)}{\prod_{j=0}^{i}\left(1-q_j^{k_j}\right)} \qquad (41)$$

Comparing to Eq.(33), we see that the effect of the acceptance/rejection step is to multiply $V[p_i^e]$ by a factor $(2-\theta_i)/\theta_i$. Using Eq.(18), the relative variance in P_B^e is:

$$\frac{V^{\rm rb}[P_B^e]}{P_B^2} = \frac{1}{N_0} \sum_{i=0}^{n-1} \frac{q_i}{p_i k_i} \frac{(2-\theta_i)}{\theta_i} \frac{\left(1-q_i^{k_i}\right)}{\prod_{j=0}^{i}\left(1-q_j^{k_j}\right)} \qquad (42)$$

so that using Eq.(13),

$$\mathcal{V}^{\rm rb} = \sum_{i=0}^{n-1} \frac{q_i}{p_i k_i} \frac{(2-\theta_i)}{\theta_i} \frac{\left(1-q_i^{k_i}\right)}{\prod_{j=0}^{i}\left(1-q_j^{k_j}\right)} \qquad (43)$$

We show in Appendix B that the acceptance probability θ_i for $i > 0$ [note that $\theta_0 = 1$] can be approximated as:

$$\theta_i = \frac{1}{2} - \frac{\sqrt{\pi}}{4}\left[2{\rm Erf}\left(\frac{\sigma_i}{2}\right) - 1\right] \qquad (i > 0) \qquad (44)$$

where ${\rm Erf}(x)$ is the error function: ${\rm Erf}(x) = (2/\sqrt{\pi})\int_0^x e^{-t^2} dt$, and σ_i is given by:

$$\sigma_i^2 = \sum_{j=0}^{i-1}\left[\frac{(1-q_j^{k_j})q_j}{k_j p_j} - q_j^{k_j}\right] \qquad (45)$$

Eqs (44) and (45) can be substituted into Eq.(43) to give a complete expression for the relative variance in the estimated rate constant for the RB method.

Illustration

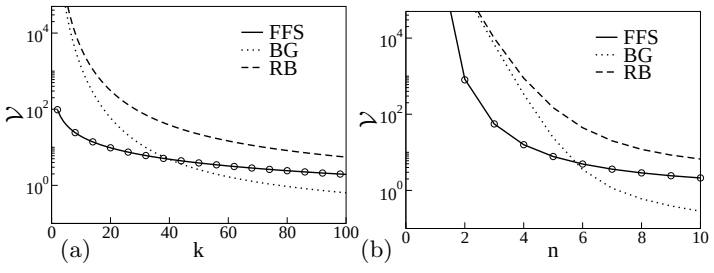

FIG. 3: Relative variance $\mathcal{V}$, for $p_i = p$, $k_i = k$ and $P_B = 10^{-8}$. The circles show the function $\sum_{i=0}^{n-1} q_i/(p_i k_i)$. (a): $\mathcal{V}$ as a function of k, for $n = 5$. (b): $\mathcal{V}$ as a function of n, for $k = 25$.

Returning to the hypothetical rare event problem with evenly spaced interfaces introduced above, Figure 3 shows $\mathcal{V}$ as a function of k (for $n = 5$) and of n (for $k = 25$), for $p_i = p = P_B^{1/n}$, $k_i = k$, $N_0 = 1000$ and $P_B = 10^{-8}$. The circles show the limiting form $\sum_{i=0}^{n-1} q_i/(p_i k_i)$, which is in good agreement with the FFS results, since $1 - q^{N_0 k} \approx 1$. For small k or small n (small p), the RB and BG results tend to converge, since the probability of reaching later interfaces is small and the results are dominated by the early interfaces. In this regime, the FFS method gives the smallest variance, since the chance of terminating the trial run procedure at early interfaces is lower than for the other methods.

It is interesting to compare expressions (23), (32) and (43). All three expressions are of the form

$$\mathcal{V} = \sum_{i=0}^{n-1} \frac{q_i}{p_i k_i X_i} \qquad (46)$$

However, X_i takes different forms for the three methods:

$$X_i^{\rm ffs} = \frac{\prod_{j=0}^{i}(1-q_j^{N_0 k_j})}{(1-q_i^{N_0 k_i})} \qquad (47)$$

$$X_i^{\rm bg} = \prod_{j=0}^{i} p_j k_j \qquad (48)$$

and

$$X_i^{\rm rb} = \frac{\theta_i}{(2-\theta_i)} \frac{\prod_{j=0}^{i-1}(1-q_j^{k_j})}{(1-q_i^{k_i})} \qquad (49)$$

We note that $X_i^{\rm ffs} > X_i^{\rm rb}$, so that $\mathcal{V}^{\rm ffs}$ is always less than $\mathcal{V}^{\rm rb}$, even for $\theta_i = 1$. Both $X_i^{\rm ffs}$ and $X_i^{\rm rb}$ are always less than unity: $\mathcal{V}^{\rm ffs}$ approaches the limiting form $\sum_{i=0}^{n-1} q_i/(p_i k_i)$ from above as k_i increases (in fact in Fig. 3a it takes this form for all k) and $\mathcal{V}^{\rm rb}$ approaches $\sum_{i=0}^{n-1}(2-\theta_i)q_i/(p_i k_i \theta_i)$. For the BG method, however, $X_i^{\rm bg}$ can increase indefinitely as k_i increases, so that this method produces the smallest variance for large k_i, as in Figure 3a. However, comparing with Figure 2, we see that this is also the regime in which the BG method becomes very expensive.

Landscape Variance

So far in our analysis, we have assumed that all the points at interface λ_i have to same p_i value - *i.e.* that on firing a trial run to λ_{i+1} we have the same probability of success, no matter which point at λ_i we start from. In reality, this is not the case; we expect there to be a distribution of p_i values among the points at each interface λ_i. We call the variance of this distribution the "landscape variance" U_i at interface i, and we expect it to make a contribution to the variance in P_B^e. We now extend our analysis to include the potentially important effect of the landscape variance.

Let us suppose that each point j at λ_i has an associated probability $p_i^{(j)}$ that a trial run fired from that point will reach λ_{i+1}. The distribution of $p_i^{(j)}$ values encountered during the rate constant calculation has mean $E[p_i^{(j)}] = p_i$ and variance $V[p_i^{(j)}] \equiv U_i$. Of course, the values of U_i depend on the number and placement of the interfaces.

In Appendix C, we re-derive expressions for the relative variance in the estimated rate constant, taking into account the landscape variance. The final results are:

$$\mathcal{V}^{\rm ffs} = \sum_{i=0}^{n-1} \left\{ \left[\frac{q_i}{p_i k_i} + \frac{U_i N_0}{p_i^2 N_i}\left(1 - \frac{1}{N_0 k_i}\right)\right] \times \frac{\left(1-q_i^{N_0 k_i}\right)}{\prod_{j=0}^{i}\left(1-q_j^{N_0 k_j}\right)} \right\} \qquad (50)$$

where $N_i = N_0 k_{i-1} p_{i-1}$ for $i > 0$ and $N_i = N_0$ for $i = 0$.

$$\mathcal{V}^{\rm bg} = \sum_{i=0}^{n-1} \left[\frac{k_i q_i p_i + U_i \left(k_i^2 - k_i\right)}{k_i p_i \prod_{j=0}^{i} p_j k_j} \right] \tag{51}$$

and

$$\mathcal{V}^{\rm rb} = \sum_{i=0}^{n-1} \left\{ \left[\frac{q_i}{p_i k_i} + \frac{U_i}{p_i^2} \left(1 - \frac{1}{k_i}\right) \right] \times \left[\frac{(2-\theta_i)}{\theta_i} \right] \frac{\left(1 - q_i^{k_i}\right)}{\prod_{j=0}^{i} \left(1 - q_j^{k_j}\right)} \right\} \tag{52}$$

Comparing Eqs (50), (51) and (52) to their equivalent forms without landscape variance, (23), (32) and (43), we see that for each interface the "binomial" terms of the form $p_i q_i / k_i$ are now supplemented by additional terms describing the landscape variance. In the limit of very large k_i, the relative variance no longer tends to zero. Instead, as $k_i \to \infty$ (for all i), the FFS and BG expressions (50) and (51) tend to the constant value U_0/p_0^2, while the RB expression (52) tends to $\sum_{i=0}^{n-1} U_i/p_i^2$. While the "binomial" contribution to the variance can be reduced by firing many trial runs per point, the "landscape" contribution can only be reduced by sampling many points. In the FFS and BG methods, branching paths are generated. For very large k_i, each point at λ_0 generates many points at subsequent interfaces, so that only U_0 remains in Eqs (50) and (51) as $k_i \to \infty$. In the RB method, however, paths are not branched, so that each point at λ_0 corresponds to one (or less than one) point at each subsequent interface. In this case, as $k_i \to \infty$, all the U_i values continue to contribute to $\mathcal{V}$.

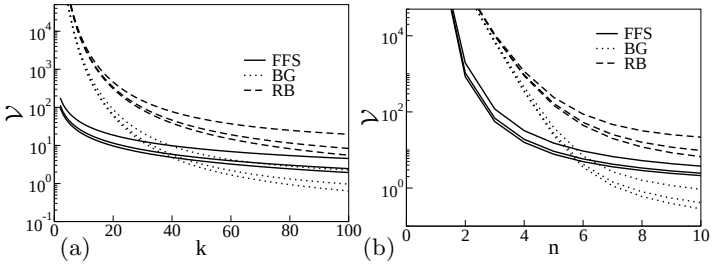

FIG. 4: Relative variance $\mathcal{V}$ in k^e_{AB}, as predicted by Eqs (50), (51) and (52), for the model problem of Figs 2 and 3, with $P_B = 10^{-8}$ and $U_i = U$. The upper curves in each group correspond to $U = 5p^2/n$, the middle curves to $U = p^2/n$ and the lower curves to $U = 0$. (a): $\mathcal{V}$ as a function of k, keeping $n = 5$. (b): $\mathcal{V}$ as a function of n, keeping $k = 25$.

In Figure 4, we revisit the simple model problem of Figs 2 and 3, adding in the effects of landscape variance. We take U_i to be the same for all interfaces: $U_i = U$. We choose, somewhat arbitrarily, $U = p^2/n$ or $U = 5p^2/n$. These turn out to be quite realistic values for the Maier-Stein system discussed in Section IV. Figure 4 shows the relative variance $\mathcal{V}$ (as in Figure 3), calculated with $U = 5p^2/n$ (upper curves), $U = p^2/n$ (middle curves) and $U = 0$ (lower curves). Although the landscape variance does not change the general trend that $\mathcal{V}$ decreases as k or N increases, it does have the qualitative effect that $\mathcal{V}$ no longer tends to zero (as discussed above). Depending on the value of U, the quantitative effects of the landscape contribution can be very significant, especially as k or N becomes large.

C. Efficiency

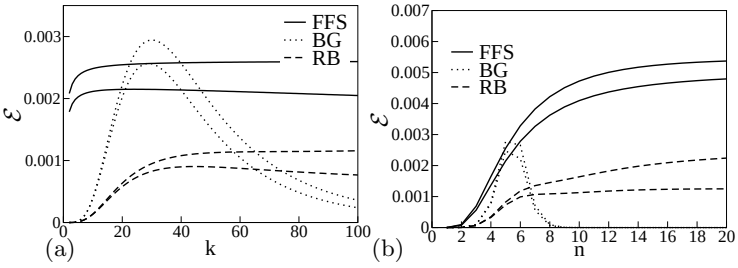

FIG. 5: Efficiency $\mathcal{E}$, calculated using Eq.(4), for the simple model of Figs 2, 3 and 4. For each method, results are plotted with $U = p^2/n$ (lower curves) and $U = 0$ (upper curves). (a): $\mathcal{E}$ as a function of k for $n = 5$. (b): $\mathcal{E}$ as a function of n for $k = 25$.

Having calculated the computational cost and the statistical accuracy of the three methods, we are now in a position to assess their overall computational efficiency, as defined by Eq.(4). Figure 5 shows the efficiency of the three methods as a function of k (Fig. 5a) and of n (Fig. 5b), for the simple model case of Figs 2, 3 and 4. Note the altered scale on the n axis in comparison to Figures 2 and 3. For each method, the upper curve shows the results without the landscape contribution to the variance ($U = 0$) and the lower curve includes a landscape contribution of $U = p^2/n$.

Firstly, we note that the optimum values of $\mathcal{E}$ are of the same order of magnitude for all three methods, although $\mathcal{E}$ is consistently lower for RB, due to the acceptance/rejection step. However, the dependence of the efficiency on the parameter values k and n is very different for the three methods. For the BG method, the efficiency shows a pronounced peak, both as a function of k and of n. Although for an optimum choice of parameters, this method can be the most efficient, its performance is highly sensitive to the choice of parameters, decreasing sharply for non-optimal values of k or n. The FFS and RB methods are much less parameter-sensitive - in fact, as long as k or n is not too small, the choice of parameters appears not to be at all critical for these methods. In general, Fig.5 seems to indicate that the method of choice is FFS, since this method is highly robust to changes in the parameters, is the most efficient method at small k or

n, and remains efficient as k and n become large. However, this interpretation must be treated with care, since several important factors are not included in the analysis leading to Fig.5. Firstly, our analysis does not include the effects of correlations between interfaces. This has the effect that neither the FFS or RB methods shows a maximum in efficiency as a function of n in Fig.5b. In our simple model, one can always gain more information by sampling at more closely spaced interfaces - however, in reality, correlations between interfaces are likely to make very closely spaced interfaces computationally inefficient. Another important factor to be considered is the fact that both the FFS and BG methods generate branched transition paths. In FFS, in fact, an effect analogous to "genetic drift" means that if the number of points in the collections at the interfaces is small enough to be of the order of the number of interfaces, then all the paths that finally reach B can be expected to originate from a small number of initial points at λ_0. If there is "memory loss" - *i.e.* no correlations between interfaces, this may be unimportant. However, if the history of the paths is important, then the RB method may be the method of choice, since this generates independent, unbranched paths. Furthermore, the RB method requires much less storage of system configurations than FFS (for which a whole collection of points must be stored in memory at each interface) - for some systems, this may be a significant factor in the computational cost.

Figure 5 also shows the effects of landscape variance on the efficiency of the three methods. Including landscape variance always decreases the efficiency, but produces rather few qualitative effects for this simple model problem. It is interesting to note, however, that in Figure 5a both the FFS and RB methods show a maximum in efficiency as a function of k only when the landscape contribution is included. When the landscape contribution is not considered, the equations predict that arbitrarily high accuracy can be obtained by firing an infinitely large number of trials from a single point. In this example, we took the landscape variance to be the same for all interfaces: $U_i = U$. However, one can easily imagine that for some systems, there is much greater variation among transition paths when they are close to the A basin, while for others, paths tend to diverge as they approach B. In the former case, we can expect the RB and BG methods to have an advantage relative to FFS, because in these methods, relatively more points are sampled at early interfaces (since the probability of failing to complete a transition path is higher). Conversely, if the landscape variance is very large close to the B basin, the BG method may be advantageous, since it samples many points at later interfaces due to its branching tree of paths.

IV. THE MAIER-STEIN SYSTEM

In this section, we test the expressions derived in Section III for a real rare event simulation problem. As our test case, we simulate the two-dimensional non-equilibrium rare event problem proposed by Maier and Stein [15, 16, 17]. This system has been extensively studied both theoretically and experimentally [15, 16, 17, 31, 32] and was also used by Crooks and Chandler [4] as a test case for their non-equilibrium rare event method. We hope that the conclusions obtained for this system will also prove to be applicable to more computationally intensive rare event problems.

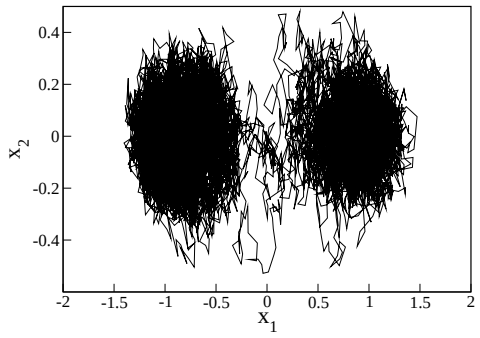

FIG. 6: Typical trajectory for a brute-force simulation of the Maier-Stein system, with $\alpha = 6.67$, $\mu = 2$ and $\epsilon = 0.1$.

The Maier-Stein system consists of a single particle moving with over-damped Langevin dynamics in a two-dimensional force field. The position vector (x_1, x_2) of the particle satisfies the stochastic differential equation:

$$\dot{x_i} = f_i(x) + \xi_i(t) \tag{53}$$

where the force field $\mathbf{f} = (f_1, f_2)$ is given by:

$$\mathbf{f} = \left(x_1 - x_1^3 - \alpha x_1 x_2^2\,,\, -\mu x_2(1 + x_1^2)\right) \tag{54}$$

and the stochastic force $\xi = (\xi_1, \xi_2)$ satisfies:

$$\langle \xi_i(t) \rangle = 0 \;;\; \langle \xi_i(t+\tau)\xi_j(t) \rangle = \epsilon\, \delta(t-\tau)\delta_{ij} \tag{55}$$

This system is bistable, with stable points at $(\pm 1, 0)$ and a saddle point at $(0, 0)$. If $\alpha \neq \mu$, the force field $\mathbf{f}$ cannot be expressed as the gradient of a potential. In this case, the system is intrinsically out of equilibrium and does not satisfy detailed balance. The parameter ϵ controls the magnitude of the stochastic force acting on the particle. For $\epsilon > 0$, the system makes stochastic transitions between the two stable states, at a rate which decreases as ϵ decreases. Figure 6 shows a typical trajectory generated by a brute-force simulation. Here, and in the rest of this Section, we use $\alpha = 6.67$, $\mu = 2.0$ (following Crooks and Chandler [4]) and $\epsilon = 0.1$. Eq.(53) is integrated numerically with timestep $\delta t = 0.02$ [33]. For our calculations using the FFS-type methods, we define $\lambda(\mathbf{x}) = x_1$, $\lambda_A \equiv \lambda_0 = -0.7$ and $\lambda_B \equiv \lambda_n = 0.7$.

A. Measuring the parameters

In order to test the expressions of Section III, we must measure the cost parameters R and S, the probability

P_B of reaching B and, for a given set of n interfaces, the probabilities $\{p_i\}$ and the landscape variance values $\{U_i\}$. For most of our calculations, we used $n = 7$, and the interfaces were positioned as listed in Table I. For the results of Figs 8b, 9b and 11b, where n was varied, we kept the interfaces evenly spaced between $\lambda_0 = -0.7$ and $\lambda_n = 0.7$. R, the cost of generating an initial point at λ_0, was measured using a simulation in region A to be $R = 590 \pm 50$ steps. In these calculations, points at λ_0 were collected upon every 10th crossing of λ_0 from A. To measure S (the proportionality constant in Eq.(6)), we carried out an FFS run, measuring the average length (in simulation steps) of successful and unsuccessful trials from each interface. The results are shown in Figure 7. Here, the filled circles show the average length, in simulation steps, of successful trials from interface λ_i (plotted on the x axis) to $\lambda_{i+1} = \lambda_i + 0.2$. Since $|\lambda_i - \lambda_j| = 0.2$ for all these trials, Eq.(6) predicts that all the filled circles should have show the same average trial length. The open circles show the average length of unsuccessful trials, which begin at λ_i and end at $\lambda_A = -0.7$, so that $|\lambda_i - \lambda_j| = \lambda_i + 0.7$: Eq.(6) predicts that all the open circles should lie on a straight line. Combining all the data, we obtain an average value of $S = 131$ steps. This value is used to plot the solid lines in Figure 7. The very good agreement that is observed between the solid lines and the circles implies that the drift-diffusion approximation, Eq.(6), is reasonable for this problem. The most significant deviation occurs for the successful trial runs between $\lambda = -0.7$ and $\lambda = -0.5$; these are unexpectedly short, perhaps because the "drift force" is weaker in this region.

Interface	λ_i	p_i	U_i
0	-0.7	0.1144 ± 0.0001	0.00350 ± 0.00003
1	-0.5	0.2651 ± 0.0002	0.00368 ± 0.00008
2	-0.3	0.3834 ± 0.0002	0.0031 ± 0.0003
3	-0.1	0.5633 ± 0.0003	0.0021 ± 0.0002
4	0.1	0.7702 ± 0.0003	0.0008 ± 0.0001
5	0.3	0.9152 ± 0.0002	0.0003 ± 0.0001
6	0.5	0.9747 ± 0.0002	0.00005 ± 0.00002

TABLE I: Positions of the interfaces and measured values of $\{p_i\}$ and $\{U_i\}$ for the Maier-Stein problem.

Using FFS, we obtained $P_B = [4.501 \pm 0.007] \times 10^{-3}$. The values of $\{p_i\}$ were also measured (using FFS) and are given in Table I. The landscape variance $\{U_i\}$ was measured using the procedure described in Appendix D: after generating a correctly weighted collection of points at interface λ_i (for example using FFS), one fires k_i trials from each point j and records the number of successes, $N_s^{(i)}|j$. One then calculates the variance among points $V[N_s^{(i)}]$. The intrinsic variance is given by

$$U_i = \frac{V[N_s^{(i)}]/k_i - p_i q_i}{k_i - 1} \tag{56}$$

Table I shows that for this problem U_i/p_i^2 is rather small

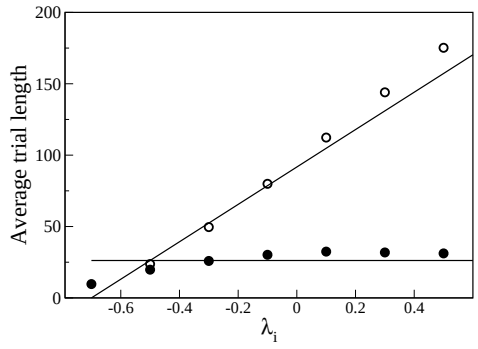

FIG. 7: Costs of trial runs between interfaces, for the Maier-Stein system. The average length, in simulation steps, of "successful" trials (to λ_{i+1}) are shown as filled circles. For these trials, $\lambda_j = \lambda_i + 0.2$ and $|\lambda_i - \lambda_j| = 0.2$. The average length of "unsuccessful" trials (to $\lambda_A = -0.7$) are shown as open circles. For these trials, $|\lambda_i - \lambda_j| = \lambda_i + 0.7$. The solid lines show the linear approximation, Eq.(6), with $S = 131$.

(a maximum of 0.27 for interface 0), indicating that the landscape variance is unlikely to have important effects in this case. However, this may not be the case for more complex systems in higher dimensions.

B. Testing the expressions

We now measure directly the cost, in simulation steps, the error in the calculated rate constant, and thus the efficiency of the three methods, for the Maier-Stein problem, and compare our simulation results to the predictions of Section III. For each method, simulations were carried out in a series of blocks. For FFS, a block consists of a complete FFS calculation with N_0 starting points. For the RB and BG methods, a block consists of N_0 starting points at λ_0. Each block produces a result P_B^e for the probability of reaching B. To find $V[P_B^e]$, we calculate the variance between blocks:

$$V[P_B^e] = \overline{(P_B^e)^2} - (\overline{P_B^e})^2 \tag{57}$$

where the over-line denotes an average over the blocks. The cost $\mathcal{C}$ per starting point at λ_0 is the average number of simulation steps per block, divided by N_0.

Figure 8 shows a comparison between the simulation values of $\mathcal{C}$ and the theoretical predictions (Eqs (9), (11) and (12)), for the three methods, as a functions of k (Fig.8a) and of n (Fig.8b). In these calculations, the same value of k was used for all interfaces: $k_i = k$ for all i. To obtain the data in Fig.8b, we used interfaces which were evenly spaced in λ and a fixed value $k = 3$. We observe remarkably good agreement between the predicted and observed values for the cost, verifying that at least for this problem, Eqs (9), (11) and (12) are very accurate.

The predicted and measured values of $\mathcal{V}$ are shown in Figure 9, for all three methods. Agreement is again ex-

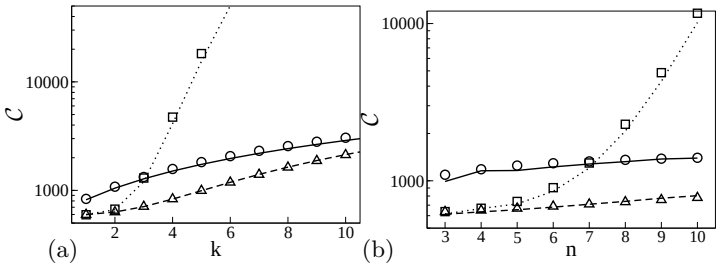

FIG. 8: Predicted and measured values of $\mathcal{C}$, for the Maier-Stein problem as described in Section IV. The lines show the theoretical predictions for the FFS (solid line), BG (dotted line) and RB (dashed line) methods. The symbols show the simulation results. Circles: FFS method, squares: BG method, triangles: RB method (with Metropolis acceptance/rejection). Simulation results were obtained with 400 blocks of $N_0 = 1000$ starting points for FFS and 2000 starting points per block for BG and RB. (a): $\mathcal{C}$ as a function of k, for $n = 7$. (b): $\mathcal{C}$ as a function of n, for $k = 3$, for evenly spaced interfaces.

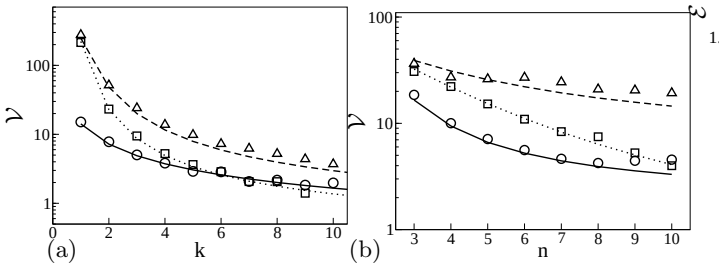

FIG. 9: Predicted and measured values of $\mathcal{V}$, for the Maier-Stein problem. The lines show the theoretical predictions for the FFS (solid line), BG (dotted line) and RB (dashed line) methods. The symbols show the simulation results. Circles: FFS method, squares: BG method, triangles: RB method (with Metropolis acceptance/rejection). Simulation results were obtained with 400 blocks of $N_0 = 1000$ starting points for FFS and 2000 starting points per block for BG and RB. Interfaces were evenly spaced between $\lambda_A = -0.7$ and $\lambda_B = 0.7$ (a): $\mathcal{V}$ as a function of k, for $n = 7$. (b): $\mathcal{V}$ as a function of n, for $k = 3$. In (b), the landscape contribution is not included in the theoretical calculation.

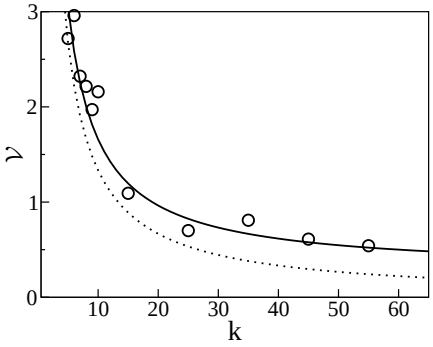

FIG. 10: Predicted and measured values of $\mathcal{V}$, for the Maier-Stein problem, for the FFS method. Solid line: Eq.(50) (with landscape variance), dotted line: Eq.(23) (no landscape variance), circles: simulation results.

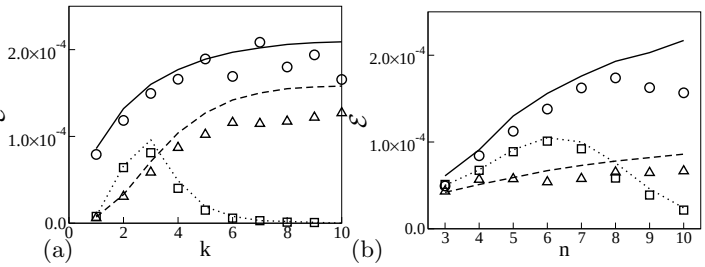

FIG. 11: Predicted and measured efficiency $\mathcal{E}$, for the Maier-Stein system. The lines show the theoretical predictions for the FFS (solid line), BG (dotted line) and RB (dashed line) methods. The symbols show the simulation results. Circles: FFS method, squares: BG method, triangles: RB method (with Metropolis acceptance/rejection). Simulation results were obtained with 400 blocks. For FFS, each block had $N_0 = 1000$ starting points and for BG and RB each blocks had 2000 starting points. Interfaces were evenly spaced. (a): $\mathcal{E}$ vs k for $n = 7$. (b): $\mathcal{E}$ vs n for $k = 3$.

cellent, showing that the approximations of Section III B are justified, at least for this problem. The landscape contribution to $\mathcal{V}$ is included in Figure 9 for panel (a) but not for (b). In Figure 10, we show the effect of neglecting this contribution (note the altered scales on both axes). Although the landscape contribution is small for this problem, it becomes significant for large k as the "binomial" contribution decreases.

The efficiency $\mathcal{E}$ is plotted in Figure 11. Excellent agreement is obtained between simulation and theory. It is also interesting to note that the trends in $\mathcal{E}$ as a function of k are qualitatively very similar to those obtained for the model problem of Fig. 5. The BG method shows high efficiency only within a relatively narrow range of parameter values, while the FFS and RB methods are much more robust to changes in the parameters. The RB method is consistently less efficient than FFS, due to the acceptance/rejection step. As the number of interfaces n becomes large, we would expect the correlations between interfaces (which are not included in our analysis) to have a greater effect, and the theoretical predictions to become less accurate. This effect is observed to a certain extent: the efficiency of FFS, for example, decreases relative to the predicted value as n increases. However, this is not a dramatic effect, and in fact, even on increasing n further, as far as 100 interfaces, we find a decrease of only a few percent in the efficiency of FFS. It seems therefore, that for FFS at least, one can use any number n of interfaces, as long as n is not too small or so very large that memory requirements become the limiting factor.

The remarkable agreement between the theoretical predictions and the simulation results shown in Figures 8, 9 and 11 perhaps reflects the simplicity of the Maier-Stein

problem. The main assumption for the calculation of $\mathcal{V}$ - that the sampling of p_i at different interfaces is uncorrelated - seems to be well justified in this case. We would expect our theoretical predictions to be less accurate for more complex problems, perhaps with strong correlations between interfaces. In fact, on investigating the two examples presented in our previous paper [1] - the flipping of a genetic switch and the translocation of a polymer through a pore - we find that the quantitative estimates of both the cost and variance can differ by a factor of about 10 from the theoretical predictions. Even with this caveat, however, we believe that the expressions of Section III will prove to be of practical use for a wide range of rare event simulation problems.

V. DISCUSSION

In this paper, we have derived simple analytical expressions for the computational cost of the three FFS-type rare event simulation methods and the statistical accuracy of the resulting estimate of the rate constant. The expressions were found to be in remarkably good agreement with simulation results for the two-dimensional non-equilibrium rare event problem proposed by Maier and Stein [15, 16, 17].

Our analysis allows us to draw some general conclusions about the relative merits of the three FFS-type methods. Firstly, the optimum efficiencies of the methods are all of the same order of magnitude, at least for the simple test problem studied here. However, the methods show very different sensitivities to the choice of parameters. The Branched Growth method in particular is highly sensitive, performing well only for a narrow range of parameter values. Within this range, however, it performs well in comparison to the other methods. The FFS method is the most robust to changes in the parameters, performing consistently well, even for parameter values where the other methods are very inefficient. The Rosenbluth method is lower in efficiency than the others, as a consequence of the Metropolis acceptance/rejection step which is required in order to obtain paths with the correct weights in the Transition Path Ensemble.

These observations provide a very useful guide for choosing a rate constant calculation method. In general, unless one has a very good idea of the optimum parameters, the BG method carries a risk of being low in efficiency. Of course, strategies could be envisaged to overcome this problem - for example, one could imagine terminating a certain percentage of the branches to avoid the high cost of sampling later interfaces. The analysis used here could easily be extended to predict the likely success of such approaches. The RB method appears from this analysis to be of relatively low efficiency. However, that is not to say that one should not use the Rosenbluth method. On the contrary, this is the only method which generates unbranched paths, making it highly suitable for situations where one wishes to analyse the paths, in order to study the transition mechanism. The RB and BG methods also require much less storage of system configurations than FFS (for which all N_i points at interface i must be stored in memory), making them potentially suitable for large systems. As a general conclusion, however, the results of this paper show that the FFS method is highly robust to parameter changes and is probably the method of choice for calculations of the rate constant where effects such as the storage of many configurations in memory are not important.

These results could also suggest possible strategies for choosing the parameters for the three methods. One approach would be to use the analytical expressions derived here in an optimization scheme for finding $\{k_i\}$, $\{\lambda_i\}$ and n. This is likely to be useful for the BG method, but may be less essential for the FFS and RB methods, where the choice of parameters is much less critical.

We expect that the predictions of the cost and statistical error derived here will be useful not only for parameter optimization, but also for assessing, before beginning a calculation, which method to use and, indeed, whether to proceed at all. Some preliminary calculation would be needed in order to obtain rough estimates for R,S, P_B, $\{p_i\}$ and (if required) $\{U_i\}$. These preliminary calculations are expected to be much cheaper than a full simulation. While the expressions for the cost and variance will be less accurate if only rough estimates for the parameters are available, we expect the results to be nevertheless accurate enough to be of use.

Furthermore, the expressions for $\mathcal{V}$ can be used, after a rate constant calculation has been completed, to obtain error bars on the calculated value of k_{AB}. In this case, the values of P_B and $\{p_i\}$ are known. The intrinsic variances $\{U_i\}$ can also be easily obtained during the rate constant calculation, as explained in Appendix D. These values can be substituted into the expressions to obtain a reliable estimate of the statistical error in the resulting rate constant.

In this work, we provide a way to compare the efficiency of the three FFS-type methods. It would also be very useful to compare their efficiency to that of other methods, such as the method of Crooks and Chandler [4] for non-equilibrium rare event problems, or TPS [3] or Transition Interface Sampling (TIS) [5, 20] for equilibrium problems. We have carried out preliminary calculations using the Crooks-Chandler method for the Maier-Stein system. We find that the value of the rate constant is in agreement with that of the FFS-type methods, but that the FFS-type methods are much more efficient. However, a thorough comparison would require a detailed investigation, optimizing the parameter choices of all the methods. We therefore leave this to a future study.

In conclusion, we have presented expressions for the computational cost and statistical accuracy of three recently introduced rare event simulation methods. We believe that the expressions presented here will be valuable in using these methods to compute rate constants and in evaluating the results of such computations.

Acknowledgments

The authors thank Axel Arnold for his careful reading of the manuscript. This work is part of the research program of the "Stichting voor Fundamenteel Onderzoek der Materie (FOM)", which is financially supported by the "Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO)". R.J.A. was funded by the European Union Marie Curie program.

APPENDIX A: COST OF TRIAL RUNS

In order to estimate the cost of a trial run, we assume that the system undergoes one-dimensional diffusion along the λ coordinate, with a constant drift velocity (the origin of which is a force due to the "free energy barrier"). The problem is then equivalent to that of a particle which undergoes diffusion with drift along the x axis, after being released between two absorbing boundaries. We are interested in the mean time $\tau_{\leftarrow}$ or $\tau_{\rightarrow}$ that the particle takes to be captured at the left or right boundary, *given* that it is eventually captured at that particular boundary. Farkas and Fülöp have studied the problem of one dimensional diffusion with drift [28]. They give analytical expressions for the probabilities $n_{\leftarrow}$ and $n_{\rightarrow}$ that the particle is absorbed at the left and right boundaries, respectively, and the rates of absorption, $j_{\leftarrow}$ and $j_{\rightarrow}$ at the left and right boundaries. The mean first passage time τ is the average time before the particle is absorbed at one of the boundaries:

$$\tau = \int_0^\infty t\,[j_{\leftarrow} + j_{\rightarrow}]\,dt \tag{A1}$$

To compute $\tau_{\leftarrow}$ and $\tau_{\rightarrow}$, we require integrals similar to Eq.(A1), but including only events where the particle reaches the desired boundary. The integrals must also be normalized by the probability of reaching that boundary:

$$\tau_{\leftarrow} = \frac{\int_0^\infty t j_{\leftarrow} dt}{n_{\leftarrow}} \quad ; \quad \tau_{\rightarrow} = \frac{\int_0^\infty t j_{\rightarrow} dt}{n_{\rightarrow}} \tag{A2}$$

Carrying out the integrals (A2) using the expressions of Farkas and Fülöp for $j_{\leftarrow}$, $j_{\rightarrow}$, $n_{\leftarrow}$ and $n_{\rightarrow}$ (Eqs (3-5) of their paper [28]), we arrive at:

$$\begin{aligned}\tau_{\leftarrow} &= \frac{L}{v}\left[\coth\left(\frac{Lv}{2D}\right) - (1-\alpha)\coth\left(\frac{(1-\alpha)Lv}{2D}\right)\right] \\ \tau_{\rightarrow} &= \frac{L}{v}\left[\coth\left(\frac{Lv}{2D}\right) - \alpha\coth\left(\frac{\alpha Lv}{2D}\right)\right]\end{aligned} \tag{A3}$$

where v is the drift velocity, D is the diffusion constant, the absorbing boundaries are at $x = 0$ and $x = L$ and the particle is released at $x = \alpha L$ at time t. In the limit that the drift velocity is large, $\cosh[Lv/(2D)] \to 1$ and $\tau_{\leftarrow}$ and $\tau_{\rightarrow}$ reduce to:

$$\tau_{\leftarrow} = \frac{\alpha L}{v} \quad ; \quad \tau_{\rightarrow} = \frac{(1-\alpha)L}{v} \tag{A4}$$

In this case, the average time for a particle to be captured at a specified boundary is linearly proportional to the distance between the starting point of the particle and that boundary, and the proportionality constant is the same for particles moving against or with the drift velocity. It is therefore appropriate to approximate the cost of a trial run between λ_i and λ_j by $S|\lambda_j - \lambda_i|$, as in Eq.(6).

APPENDIX B: ACCEPTANCE PROBABILITY FOR THE RB METHOD

This section is concerned with the Metropolis acceptance/rejection step in the Rosenbluth method. We derive the approximate expression (44) for the probability θ_i that a newly generated estimate $p_i^{e(\mathrm{n})} = N_s^{(i)}/k_i$ for the probability p_i is accepted. Upon reaching interface i, we calculate the Rosenbluth factor $W_i^{(\mathrm{n})} = \prod_{j=0}^{i-1} N_s^{(j)}$ corresponding to the newly generated path leading to interface i. We compare this to the Rosenbluth factor $W_i^{(\mathrm{o})}$ corresponding to the previous path to have been accepted at interface i. Acceptance occurs if the ratio $Z_i \equiv W_i^{(\mathrm{n})}/W_i^{(\mathrm{o})}$ is greater than a random number $0 < s < 1$. If we know the distribution function $P(Z_i)$, the acceptance probability is given by:

$$\theta_i \;=\; \int_0^1 ds \int_s^\infty dZ_i \; P(Z_i) \tag{B1}$$

We would therefore like to calculate $P(Z_i) \equiv P(W_i^{(\mathrm{n})}/W_i^{(\mathrm{o})})$. To obtain this, we require the distribution functions for both $W_i^{(\mathrm{n})}$ and $W_i^{(\mathrm{o})}$. We begin with $W_i^{(\mathrm{n})}$, which we can write as

$$\log[W_i^{(\mathrm{n})}] = \sum_{j=0}^{i-1} \log[N_s^{(j)}] \tag{B2}$$

We assume that the $\log[N_s^{(j)}]$ for each interface j are independent variables (*i.e.* that the sampling at different interfaces is uncorrelated). Since we are adding many independent variables, we apply the Central Limit Theorem [29] to Eq.(B2). In the limit of a large number of interfaces, the distribution of $y_i^{(\mathrm{n})} = \log[W_i^{(\mathrm{n})}]$, is:

$$p(y_i^{(\mathrm{n})}) = \frac{1}{\sigma_i\sqrt{2\pi}} \exp\left[-\frac{(y_i^{(\mathrm{n})} - \mu_i)^2}{2\sigma_i^2}\right] \tag{B3}$$

where

$$\mu_i = \sum_{j=0}^{i-1} E[\log N_s^{(j)}] \tag{B4}$$

and

$$\sigma_i^2 = \sum_{j=0}^{i-1} V[\log N_s^{(j)}] \tag{B5}$$

The expectation value $E[\log N_s^{(j)}]$ can be found approximately by performing a Taylor expansion of $\log N_s^{(j)}$ about $E[N_s^{(j)}]$, to give:

$$\log N_s^{(j)} \approx \log E[N_s^{(j)}] + \frac{\left(N_s^{(j)} - E[N_s^{(j)}]\right)}{E[N_s^{(j)}]} - \frac{1}{2}\frac{\left(N_s^{(j)} - E[N_s^{(j)}]\right)^2}{E[N_s^{(j)}]^2} \quad \text{(B6)}$$

taking the expectation value of Eq.(B6), we obtain:

$$E[\log N_s^{(j)}] \approx \log E[N_s^{(j)}] - \frac{V[N_s^{(j)}]}{2E[N_s^{(j)}]^2} \quad \text{(B7)}$$

Using the variance relation (17), we find that

$$V[\log N_s^{(j)}] \approx \frac{1}{E[N_s^{(j)}]^2} V[N_s^{(j)}] \quad \text{(B8)}$$

We now need to know $E[N_s^{(j)}]$ and $V[N_s^{(j)}]$. On firing k_i trials from interface i, we know that the number of successes follows a binomial distribution. However, the variable $N_s^{(j)}$ in Eqs (B13) and (B14) refers to the number of successes at interface j, given that we know the path subsequently reached interface $i > j$. We therefore know that $N_s^{(j)} > 0$, so that

$$p(N_s^{(j)}) = \frac{1}{(1-q_j^{k_j})} \frac{k_j!}{(k_j - N_s^{(j)})!(N_s^{(j)})!} p_j^{N_s^{(j)}} q_j^{k_j - N_s^{(j)}} \quad \text{(B9)}$$

so that

$$E(N_s^{(j)}) = \frac{k_j p_j}{(1-q_j^{k_j})} \quad \text{(B10)}$$

$$E({N_s^{(j)}}^2) = \frac{\left[k_j p_j q_j + k_j^2 p_j^2\right]}{(1-q_j^{k_j})} \quad \text{(B11)}$$

and

$$V[N_s^{(j)}] = \frac{\left[(1-q_j^{k_j}) k_j p_j q_j - k_j^2 p_j^2 q_j^{k_j}\right]}{(1-q_j^{k_j})^2} \quad \text{(B12)}$$

Substituting (B10) and (B12) into (B7) and (B8), we obtain:

$$E[\log N_s^{(j)}] \approx \log\left[\frac{k_j p_j}{1-q_j^{k_j}}\right] - \frac{1}{2}\left[\frac{(1-q_j^{k_j})q_j}{k_j p_j} - q_j^{k_j}\right] \quad \text{(B13)}$$

and

$$V[\log N_s^{(j)}] \approx \frac{q_j(1-q_j^{k_j})}{k_j p_j} - q_j^{k_j} \quad \text{(B14)}$$

Substituting (B13) and (B14) in turn into (B4) and (B5) leads to

$$\mu_i = \sum_{j=0}^{i-1} \log\left[\frac{k_j p_j}{(1-q_j^{k_j})}\right] - \frac{1}{2}\left[\frac{(1-q_j^{k_j})q_j}{k_j p_j} - q_j^{k_j}\right] \quad \text{(B15)}$$

and

$$\sigma_i^2 = \sum_{j=0}^{i-1} \frac{q_j(1-q_j^{k_j})}{k_j p_j} - q_j^{k_j} \quad \text{(B16)}$$

Finally, the distribution function $f(W_i)$ for the Rosenbluth factor of the newly generated path can be found by making the change of variables $W_i = \exp[y_i^{(\mathrm{n})}]$ in Eq.(B3), to give:

$$f(W_i) = \frac{1}{\sigma_i\sqrt{2\pi}}\left[\frac{1}{W_i}\right]\exp\left[-\frac{(\log[W_i] - \mu_i)^2}{2\sigma_i^2}\right] \quad \text{(B17)}$$

We now turn to the distribution function $g(W_i)$ for the Rosenbluth factor $W_i^{(\mathrm{o})}$ of the previous path to have been accepted at interface i. $W_i^{(\mathrm{o})}$ does not follow the same distribution as $W_i^{(\mathrm{n})}$, because the "previous" path has survived at least one round of acceptance/rejection. We know that the acceptance/rejection procedure re-weights paths by a factor proportional to the Rosenbluth factor (see Section II C), so if we assume that $W_i^{(\mathrm{o})}$ has been "fully" re-weighted (note that this is an approximation), we can say that

$$g(W_i) \approx \frac{W_i f(W_i}{\int_0^\infty W' f(W') dW'} \quad \text{(B18)}$$

The denominator of Eq.(B18) ensures that $g(W_i)$ is properly normalized. Substituting (B18) into (B17), we find that:

$$g(W_i) = \frac{1}{I}\frac{1}{\sqrt{2\pi}\sigma_i}\exp\left[-\frac{(\log[W_i] - \mu_i)^2}{2\sigma_i^2}\right] \quad \text{(B19)}$$

where

$$I = \int_0^\infty W_i f(W_i) dW_i = \exp\left[\mu_i + \frac{\sigma_i^2}{2}\right] \quad \text{(B20)}$$

Armed with Eqs (B17) and (B19), we can now find the distribution function $P(Z_i)$ for the ratio $Z_i \equiv W_i^{(\mathrm{n})}/W_i^{(\mathrm{o})}$. This is given by:

$$P(Z_i) = \int_0^\infty \int_0^\infty dW_i dW_i' g(W_i)\, f(W_i')\, \delta\left(\frac{W_i'}{W_i} - Z_i\right) \quad \text{(B21)}$$

Changing the variable of the second integral to $Z_i' = W_i'/W_i$, we obtain

$$P(Z_i) = \int_0^\infty \int_0^\infty dW_i dZ_i'\; W_i g(W_i)\, f(Z_i' W_i)\, \delta\left(Z_i' - Z_i\right) = \int_0^\infty dW_i\; W_i\, g(W_i)\, f(Z_i W_i) \quad \text{(B22)}$$

Substituting (B17) and (B19) into (B22), we obtain:

$$P(Z_i) = \frac{1}{2\pi\sigma_i^2 I Z_i} \times \int_0^\infty dW_i \ \exp\left[-\frac{(\log[W_i]-\mu_i)^2+(\log[Z_i W_i]-\mu_i)^2}{2\sigma_i^2}\right] \quad \text{(B23)}$$

This integral can be carried out analytically [30], to give:

$$P(Z_i) = \frac{\exp\left[-\frac{\sigma_i^2}{4}\right]}{2\sigma_i Z_i\sqrt{\pi}} \exp\left[-\frac{(\log Z_i)^2}{4\sigma_i^2} - \frac{\log Z_i}{2}\right] \quad \text{(B24)}$$

We are now finally in a position to calculate the acceptance probability θ_i, using Eq.(B1). Substituting Eq.(B24) into (B1) and integrating over Z_i, we obtain [34]:

$$\begin{aligned}\theta_i &= \frac{1}{2}\int_0^1 ds \ \left[1 - \frac{\sqrt{\pi}}{2}\mathrm{Erf}\left[\frac{\sigma_i}{2} + \frac{\log s}{2\sigma_i}\right]\right] \quad \text{(B25)}\\ &= \frac{1}{2} - \frac{\sqrt{\pi}}{4}\left[2\mathrm{Erf}\left(\frac{\sigma_i}{2}\right) - 1\right]\end{aligned}$$

where $\mathrm{Erf}(x)$ is the error function: $\mathrm{Erf}(x) = (2/\sqrt{\pi})\int_0^x e^{-t^2}dt$.

Although Eq.(B25) is a simple and convenient expression for the acceptance probability θ_i, its derivation required several approximations. We have therefore tested the validity of Eq.(B25). We first carried out a "simulated simulation", in which we defined a series of $N = 15$ interfaces, each with the same value of $p_i = p = 10^{-6/15}$, and "simulated" the Rosenbluth calculation, each time drawing a random number to determine the outcome of a given "trial run", for a given number of trial runs $k_i = k$, taken to be the same for all interfaces. We measured the acceptance probabilities at each interface after 2×10^6 Rosenbluth "path generations", and compared these to Eq.(B25). The results are shown in Figure 12a, for $k = 2$, $k = 5$ and $k = 8$. The agreement with the "simulation" is very reasonable. To compare with real simulation results, we also measured the acceptance probabilities θ_i for the RB simulations of the Maier-Stein system described in Section IV. The results are compared with the predictions of Eq.(B25) in Figure 12b. Again, quite good agreement is obtained.

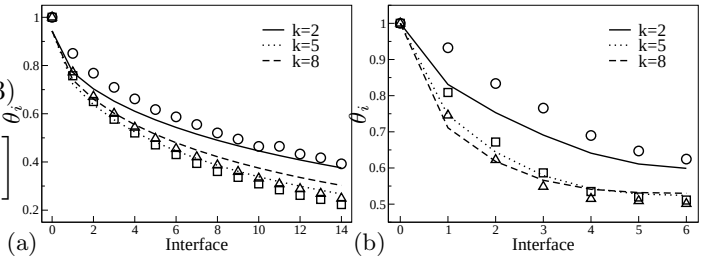

FIG. 12: (a): "Simulated" and predicted acceptance probabilities θ_i for interfaces $0 \le i \le 14$, for the "simulated simulation" described in the text, for $k = 2, 5, 8$. (b): Simulated and predicted values of θ_i for $0 \le i \le 6$ for the Maier-Stein problem of Section IV, for $k = 2, 5, 8$. In both plots, solid lines represent predicted values for $k = 2$, dotted lines, $k = 5$ and dashed lines, $k = 8$. Symbols represent simulation results: circles: $k = 2$, squares: $k = 5$ and triangles: $k = 8$.

APPENDIX C: THE EFFECTS OF LANDSCAPE VARIANCE

In this section, we include the effects of the "landscape variance" in our expressions for the relative variance $\mathcal{V}$ of P_B^e. The result will be that expressions (23), (32) and (43) are transformed into (50), (51) and (52). As described in Section III, we suppose that point j at interface λ_i has probability $p_i^{(j)}$ that a trial run fired from it will reach λ_{i+1}, rather than λ_A. The variance in the $p_i^{(j)}$ values for points at λ_i (sampled according to their expected occurrence in the trial run firing procedure) is the "landscape variance", U_i.

If we choose a particular point j, fire k_i trial runs and measure the number of successes $N_s^{(i)}$, we expect to obtain a mean value $E[N_s^{(i)}|j] = k_i p_i^{(j)}$, and a variance $V[N_s^{(i)}|j] = k_i p_i^{(j)} q_i^{(j)}$ (in analogy with Eqs (15) and (16)). We now average over many points j at interface λ_i, using the general variance relation (26):

$$\begin{aligned}V[N_s^{(i)}] &= E\left[V\left[N_s^{(i)}|j\right]\right] + V\left[E\left[N_s^{(i)}|j\right]\right] \quad \text{(C1)}\\ &= E\left[k_i p_i^{(j)} q_i^{(j)}\right] + V\left[k_i p_i^{(j)}\right]\end{aligned}$$

where the mean and the variance are taken over the distribution of points j. Since $E[p_i^{(j)} q_i^{(j)}] = E[p_i^{(j)} - (p_i^{(j)})^2] = E[p_i^{(j)}] - E[(p_i^{(j)})^2]$ and $U_i = E[(p_i^{(j)})^2] - (E[p_i^{(j)}])^2$, we can deduce that $E\left[k_i p_i^{(j)} q_i^{(j)}\right] = k_i(p_i - p_i^2 - U_i) = k_i(p_i q_i - U_i)$. Using Eq.(14), we have $V[k_i p_i^{(j)}] = k_i^2 V[p_i^{(j)}] = k_i^2 U_i$, so that

$$V[N_s^{(i)}] = k_i p_i q_i + U_i k_i^2\left[1 - \frac{1}{k_i}\right] \quad \text{(C2)}$$

This first term on the r.h.s. of Eq.(C2) corresponds to Eq.(16): the binomial contribution arising from the limited number of trial runs per point. The second term is an extra contribution, due to the landscape variance.

We now repeat the derivations of Section III B, simply replacing Eq.(16) by Eq.(C2). We begin with the RB method, for which Eq.(41) becomes

$$\begin{aligned}V^{\mathrm{rb}}[p_i^e] &= \left[\frac{1}{N_0}\right]\left[\frac{p_i q_i}{k_i} + U_i\left(1 - \frac{1}{k_i}\right)\right] \quad \text{(C3)}\\ &\times \left[\frac{(2-\theta_i)}{\theta_i}\right]\frac{\left(1-q_i^{k_i}\right)}{\prod_{j=0}^{i}\left(1-q_j^{k_j}\right)}\end{aligned}$$

and Eq.(43) is replaced by Eq.(52):

$$\mathcal{V}^{\rm rb} = \sum_{i=0}^{n-1} \quad \left\{ \left[\frac{q_i}{p_i k_i} + \frac{U_i}{p_i^2} \left(1 - \frac{1}{k_i} \right) \right] \times \left[\frac{(2-\theta_i)}{\theta_i} \right] \frac{\left(1 - q_i^{k_i} \right)}{\prod_{j=0}^{i} \left(1 - q_j^{k_j} \right)} \right\}$$

For the BG method, Eq.(30) is replaced by:

$$\begin{aligned} V[N_s^{(i)}] &= \left[k_i p_i q_i + U_i \left(k_i^2 - k_i \right) \right] \prod_{j=0}^{i-1} k_j p_j & \text{(C4)} \\ &\quad + k_i^2 p_i^2 V[N_s^{(i-1)}] & (i > 0) \\ &= k_i p_i q_i + U_i \left(k_i^2 - k_i \right) & (i = 0) \end{aligned}$$

and Eq.(32) becomes Eq.(51):

$$\mathcal{V}^{\rm bg} = \sum_{i=0}^{n-1} \left[\frac{k_i q_i p_i + U_i \left(k_i^2 - k_i \right)}{k_i p_i \prod_{j=0}^{i} p_j k_j} \right]$$

For the FFS method, the situation is slightly more complicated, because the number of trials fired from point j at interface i is not fixed. We make M_i trials from the N_i points at λ_i, each time selecting a starting point at random (so that the probability a particular point is chosen is $1/N_i$). Since we no longer assume that all points at interface i are identical, we must now take account of the distribution of the number of times m_j that point j is selected. This is in fact a multinomial distribution [29, 35], which has average $E[m_j] = M_i/N_i$ and variance $V[m_j] = M_i \left[1/N_i(1 - 1/N_i) \right]$. Let us now do a "thought experiment" in which we first decide how many trial will be fired from each point - *i.e.* we fix the set of values $\{m_j\}$ (of course, $\sum_j m_j = M_i$). We then fire these trials and measure the total number N_s^{tot} which reach λ_{i+1}. The expectation value for N_s^{tot} is

$$E[N_s^{tot}|\{m_j\}] = \sum_j m_j p_i^j = M_i p_i \qquad \text{(C5)}$$

and the variance is found using Eq.(C2), with k_i replaced by m_j, multiplying by m_j^2 and summing over all j:

$$V[N_s^{tot}|\{m_j\}] = \sum_j \left[m_j p_i q_i + U_i \left[m_j^2 - m_j \right] \right] \qquad \text{(C6)}$$

We now imagine that we average the results over many sets of values $\{m_j\}$. Using the general relation (26), we obtain:

$$\begin{aligned} V[N_s^{tot}] &= V\left[E[N_s^{tot}|\{m_j\}] \right] + E\left[V[N_s^{tot}|\{m_j\}] \right] & \text{(C7)} \\ &= V\left[M_i p_i \right] + E\left[M_i p_i q_i + U_i \sum_j m_j^2 - U_i M_i \right] \\ &= M_i p_i q_i + U_i \left[N_i E[m_j^2] - M_i \right] \end{aligned}$$

Here, the variance and expectation are with respect to the distribution of $\{m_j\}$ values. The last line follows from the fact that $V\left[M_i p_i \right] = 0$ as both M_i and p_i are constants with respect to changes in $\{m_j\}$. Since $V[m_j] = M_i \left[1/N_i(1 - 1/N_i) \right] = E[m_j^2] - E[m_j]^2$, we find that $E[m_j^2] = (M_i/N_i)(1 - 1/N_i) + M_i^2/N_i^2$. Substituting this into Eq.(C7), we obtain:

$$V[N_s^{tot}] = M_i p_i q_i + \frac{U_i}{N_i} \left[M_i^2 - M_i \right] \qquad \text{(C8)}$$

Since $p_i^e = N_s^{tot}/M_i$, we must divide Eq.(C8) by M_i^2 to obtain $V[p_i^e]^{\rm ffs}$:

$$V[p_i^e]^{\rm ffs} = \frac{p_i q_i}{M_i} + \frac{U_i}{N_i} \left(1 - \frac{1}{M_i} \right) \qquad \text{(C9)}$$

This leads to:

$$\mathcal{V}^{\rm ffs} = N_0 \sum_{i=0}^{n-1} \quad \left\{ \left[\frac{q_i}{p_i M_i} + \frac{U_i}{p_i^2 N_i} \left(1 - \frac{1}{M_i} \right) \right] \times \frac{\left(1 - q_i^{M_i} \right)}{\prod_{j=0}^{i} \left(1 - q_j^{M_j} \right)} \right\} \qquad \text{(C10)}$$

where $N_i = M_{i-1} p_{i-1}$ for $i > 0$ and $N_i = N_0$ for $i = 0$. Rewriting in terms of $k_i \equiv M_i/N_0$, we obtain Eq.(50):

$$\mathcal{V}^{\rm ffs} = \sum_{i=0}^{n-1} \quad \left\{ \left[\frac{q_i}{p_i k_i} + \frac{U_i N_0}{p_i^2 N_i} \left(1 - \frac{1}{N_0 k_i} \right) \right] \times \frac{\left(1 - q_i^{N_0 k_i} \right)}{\prod_{j=0}^{i} \left(1 - q_j^{N_0 k_j} \right)} \right\}$$

APPENDIX D: MEASURING THE INTRINSIC VARIANCE

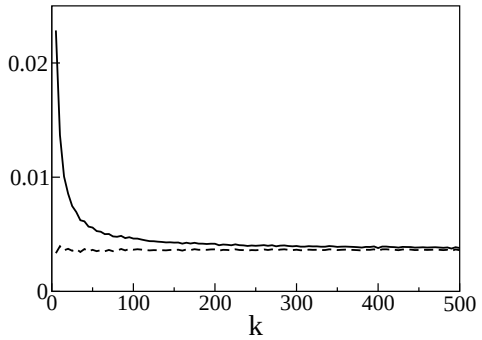

FIG. 13: $V[N_s^{(0)}]/k^2$ (solid line) and $(1/(k-1)) \left[V[N_s^{(0)}]/k - p_0 q_0 \right]$ (dashed line), as functions of $k = M_0/N_0$, calculated using FFS as described in Section D, for the Maier-Stein problem of Section IV with 10000 points at the first interface $\lambda_0 = -0.7$.

In this section, we describe a simple and computationally cheap procedure for measuring the landscape variance parameters U_i. Given a correctly weighted collection of N_i points at interface λ_i (obtained, for example, using FFS), we could fire an extremely large number k of trial runs from each point and measure the variance among points in the values of $N_s^{(i,j)}$ - where $N_s^{(i,j)}$ denotes the number of successful trials from point j:

$$U_i = V[p_i] = \frac{V[N_s^{(i)}]}{k^2} = \frac{1}{k^2}\left\{\sum_{j=1}^{N_i}\frac{{N_s^{(i,j)}}^2}{N_i} - \left[\sum_{j=1}^{N_i}\frac{N_s^{(i,j)}}{N_i}\right]^2\right\} \quad \text{(D1)}$$

This is likely to be an expensive procedure. Fortunately, however, it is not necessary to fire a very large number of trial runs from each point. Instead, we make use of expression (C2), which can be written as

$$U_i = \frac{kV[p_i^e] - p_i q_i}{k-1} = \frac{1}{(k-1)}\left[\frac{V[N_s^{(i)}]}{k} - p_i q_i\right] \quad \text{(D2)}$$

where the expression now holds for any value of k. In the limit that $k \to \infty$, Eq.(D2) reduces to (D1). As a practical procedure, therefore, we generate a collection of N_i points at interface i (using, for example, FFS), and fire k trials from each point - k does not have to be a large number. For each point j, we record the number of successful trials $N_s^{(i,j)}$. The variance $V[N_s^{(i)}]$ of these values is inserted into Eq.(D2) to give a value for U_i. Figure 13 shows the results of this procedure for the Maier-Stein problem of Section IV. For the first interface (λ_0 ($k = 10$), U_i was calculated using Eq.(D2), using k trials for each of 10000 points collected at λ_0. The solid line is the measured value of $V[N_s^{(i)}]/k^2$, while the dashed line is the value of U_i obtained from Eq.(D2). The two lines converge, of course, for large values of k. Figure 13 shows that accurate results for U_i can be obtained using Eq.(D2), using only a small number of trial runs per point.

[1] R. J. Allen, D. Frenkel, and P. R. ten Wolde, J. Chem. Phys. **124**, 024102 (2006).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation. From Algorithms to Applications* (Academic Press, Boston, 2002), 2nd ed.
[3] C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).
[4] G. E. Crooks and D. Chandler, Phys. Rev. E **64**, 026109 1 (2001).
[5] T. S. van Erp, D. Moroni, and P. G. Bolhuis, J. Chem. Phys. **118**, 7762 (2003).
[6] D. Moroni, P. G. Bolhuis, and T. S. van Erp, J. Chem. Phys. **120**, 4055 (2004).
[7] A. K. Faradjian and R. Elber, J. Chem. Phys. **120**, 10880 (2004).
[8] W. E, W. Ren, and E. Vanden-Eijnden, Phys. Rev. B **66**, 052301 (2002).
[9] W. E, W. Ren, and E. Vanden-Eijnden, J. Phys. Chem. B **109**, 6688 (2005).
[10] M. Villén-Altamirano and J. Villén-Altamirano, in *Proceedings of the 13th International Telegraphic Congress* (1991), pp. 71–76.
[11] M. Villén-Altamirano, A. Martínez-Marrón, J. L. Gamo, and F. Fernández-Cuesta, in *Proceedings of the 14th International Telegraphic Congress* (1994), pp. 797–810.
[12] M. Villén-Altamirano and J. Villén-Altamirano, European Transactions on Telecommunications **13**, 373 (2002).
[13] A. J. Bayes, Australian Computer Journal **2**, 180 (1970).
[14] G. A. Huber and S. Kim, Biophys. J. **70**, 97 (1996).
[15] R. S. Maier and D. L. Stein, Phys. Rev. E **48**, 931 (1993).
[16] R. S. Maier and D. L. Stein, J. Stat. Phys. **83**, 291 (1996).
[17] D. G. Luchinsky, R. S. Maier, R. Mannella, P. V. E. McClintock, and D. L. Stein, Phys. Rev. Lett. **82**, 1806 (1999).
[18] T. S. van Erp, Ph.D. thesis, Universiteit van Amsterdam (2003).
[19] D. Moroni, Ph.D. thesis, Universiteit van Amsterdam (2005).
[20] T. S. van Erp and P. G. Bolhuis, J. Comp. Phys. **205**, 157 (2005).
[21] R. J. Allen, P. B. Warren, and P. R. ten Wolde, Phys. Rev. Lett. **94**, 018104 (2005).
[22] P. Grassberger, Phys. Rev. E **56**, 3682 (1997).
[23] D. Aldous and U. Vazirani, Information and Computation **117**, 181 (1995).
[24] M. N. Rosenbluth and A. W. Rosenbluth, J. Chem. Phys. **23**, 356 (1955).
[25] D. Frenkel, G. C. A. M. Mooij, and B. Smit, J. Phys. Condens. Matter **3**, 3053 (1991).
[26] D. Frenkel, Proc. Natl. Acad. Sci. USA **101**, 17571 (2004).
[27] G. C. A. M. Mooij and D. Frenkel, Mol. Sim. **17**, 41 (1996).
[28] Z. Farkas and T. Fülöp, J. Phys. A: Math. Gen. **34**, 3191 (2001).
[29] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics and Engineering* (Cambridge University Press, Cambridge, 2000).
[30] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, USA, 2000), sixth ed.
[31] D. G. Luchinsky and P. V. E. McClintock, Nature **389**, 463 (1997).
[32] D. G. Luchinsky, R. S. Maier, R. Mannella, P. V. E. McClintock, and D. L. Stein, Phys. Rev. Lett. **79**, 3109 (1997).
[33] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
[34] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1970).
[35] Http://mathworld.wolfram.com/MultinomialDistribution.html.