

Automated estimation of rare event probabilities in biochemical systems

Bernie J. Daigle, Jr.,^{1,a),b)} Min K. Roh,^{1,b),c)} Dan T. Gillespie,^{2,d)} and Linda R. Petzold^{1,e)}¹*Department of Computer Science, University of California Santa Barbara, Santa Barbara, California 93106, USA*²*Dan T Gillespie Consulting, 30504 Cordoba Pl., Castaic, California 91384, USA*

(Received 2 August 2010; accepted 9 November 2010; published online 25 January 2011)

In biochemical systems, the occurrence of a rare event can be accompanied by catastrophic consequences. Precise characterization of these events using Monte Carlo simulation methods is often intractable, as the number of realizations needed to witness even a single rare event can be very large. The weighted stochastic simulation algorithm (wSSA) [J. Chem. Phys. **129**, 165101 (2008)] and its subsequent extension [J. Chem. Phys. **130**, 174103 (2009)] alleviate this difficulty with importance sampling, which effectively biases the system toward the desired rare event. However, extensive computation coupled with substantial insight into a given system is required, as there is currently no automatic approach for choosing wSSA parameters. We present a novel modification of the wSSA—the doubly weighted SSA (dwSSA)—that makes possible a fully automated parameter selection method. Our approach uses the information-theoretic concept of cross entropy to identify parameter values yielding minimum variance rare event probability estimates. We apply the method to four examples: a pure birth process, a birth-death process, an enzymatic futile cycle, and a yeast polarization model. Our results demonstrate that the proposed method (1) enables probability estimation for a class of rare events that cannot be interrogated with the wSSA, and (2) for all examples tested, reduces the number of runs needed to achieve comparable accuracy by multiple orders of magnitude. For a particular rare event in the yeast polarization model, our method transforms a projected simulation time of 600 years to three hours. Furthermore, by incorporating information-theoretic principles, our approach provides a framework for the development of more sophisticated influencing schemes that should further improve estimation accuracy. © 2011 American Institute of Physics. [doi:10.1063/1.3522769]

I. INTRODUCTION

Nature employs a variety of mechanisms to ensure the robustness of biochemical systems. Through principles like feedback and redundancy, many naturally occurring systems exhibit consistent behavior in spite of changing environments.¹ Although biochemical systems are inherently stochastic,² mechanisms conferring robustness often prevent highly variable behavior and the associated catastrophic consequences. These consequences, often manifesting as organismal disease states, are thus inherently rare, yet their accurate characterization is of great interest.³

A common approach for studying stochastic biochemical behavior employs Monte Carlo methods like the stochastic simulation algorithm (SSA).⁴ For particularly rare events (say, $p \leq 10^{-7}$), characterization with the SSA is not feasible, as the number of simulated realizations needed to witness even a single rare event can be very large. To address this limitation, Kuwahara and Mura developed the weighted SSA (wSSA), which combines importance sampling with the SSA to bias the system of interest toward the rare event.⁵ Kuwahara and

Mura demonstrated that by careful selection of reaction biasing parameters, the wSSA enabled rare event probability estimates of equivalent accuracy to the SSA using up to ten orders of magnitude less computation. However, the authors did not provide a systematic approach for selecting favorable biasing parameters, and a wSSA estimate generated using poorly chosen parameter values can be substantially less accurate than one generated with the SSA.

Gillespie *et al.* recently developed an extension to the wSSA that simultaneously calculates rare event probability estimates and associated estimator variances.⁶ This permitted the evaluation of multiple sets of parameter values, followed by selection of the set that confers the lowest variance. They illustrated the effectiveness of their method on small biochemical systems (the largest example comprised six reactions), but their approach still requires substantial system insight when determining which parameter values to test. Without such insight, all parameter value combinations should be tested to ensure an optimal combination, and the computational complexity of this method grows exponentially with the number of reactions in the system. Given these constraints, use of the wSSA for characterizing rare events in real world systems is not tractable.

In this paper, we present a novel modification of the wSSA, called the doubly weighted SSA (dwSSA), that makes possible a principled, fully automated, and efficient method

^{a)}Electronic mail: bdaigle@gmail.com.^{b)}These authors contributed equally to this work.^{c)}Electronic mail: min@cs.ucsb.edu.^{d)}Electronic mail: gillespiedt@mailaps.org.^{e)}Electronic mail: petzold@cs.ucsb.edu.

for reaction biasing parameter selection. Our presentation is structured as follows: Sections II and III present the modified version of the wSSA and the automated parameter selection method, respectively. Section IV provides the detailed algorithms of our method. In Sec. V we apply the dwSSA to four test models of increasing complexity. Finally, in Sec. VI we summarize our contributions and motivate a promising future area of research.

II. MODIFIED WSSA FORMULATION

We begin with a brief review of the wSSA; further details can be found in Kuwahara and Mura⁵ and Gillespie *et al.*⁶ We assume a well-stirred chemical system whose molecular population state at time t is represented by the random process $\mathbf{X}(t)$. The system state can be altered by the firing of M reactions, whose propensities at time t are in the set $\{a_j(\mathbf{X}(t)) : j = 1, \dots, M\}$ with sum $a_0(\mathbf{X}(t))$. The “direct method” implementation of the SSA simulates a reaction trajectory by sequentially generating the time to the next reaction τ and the index of the next reaction j' as exponential and categorical random variables, respectively. We assume that each trajectory is run until some stopping time \mathcal{T} , which is the smaller of the time to reach the rare event and the final simulation time. Thus, the probability of the entire system trajectory $\mathbf{J} \equiv (\tau_1, j'_1, \dots, \tau_{N_T}, j'_{N_T})$ given $\mathbf{X}(0) = \mathbf{x}_0$ is as follows:

$$\begin{aligned} P_{\text{SSA}}(\mathbf{J}) &= \prod_{i=1}^{N_T} \left[a_0(\mathbf{X}(t_i)) e^{-a_0(\mathbf{X}(t_i))\tau_i} d\tau_i \times \frac{a_{j'_i}(\mathbf{X}(t_i))}{a_0(\mathbf{X}(t_i))} \right] \\ &= \prod_{i=1}^{N_T} [a_{j'_i}(\mathbf{X}(t_i)) e^{-a_0(\mathbf{X}(t_i))\tau_i} d\tau_i] \end{aligned} \quad (1)$$

with $t_i \equiv \sum_{j=1}^i \tau_j$ and N_T the total number of reactions that fire in the interval $[0, \mathcal{T}]$.

The wSSA as presented in Ref. 5 biases the selection of reaction indices according to predilection functions $\{b_j(\mathbf{X}(t)) : j = 1, \dots, M\}$, given by

$$b_j(\mathbf{X}(t)) \equiv \gamma_j a_j(\mathbf{X}(t)), \quad b_0(\mathbf{X}(t)) = \sum_{j=1}^M b_j(\mathbf{X}(t)), \quad (2)$$

where each γ_j is a positive constant. The probability of the same reaction trajectory \mathbf{J} under the wSSA is thus given by

$$P_{\text{wSSA}}(\mathbf{J}) = \prod_{i=1}^{N_T} \left[a_0(\mathbf{X}(t_i)) e^{-a_0(\mathbf{X}(t_i))\tau_i} d\tau_i \times \frac{b_{j'_i}(\mathbf{X}(t_i))}{b_0(\mathbf{X}(t_i))} \right]. \quad (3)$$

In order to correct for the above reaction selection bias, the wSSA assigns the following weight to each trajectory, whose product with the probability Eq. (3) equals the probability Eq. (1):

$$W_{\text{wSSA}}(\mathbf{J}) = \prod_{i=1}^{N_T} \frac{a_{j'_i}(\mathbf{X}(t_i))/a_0(\mathbf{X}(t_i))}{b_{j'_i}(\mathbf{X}(t_i))/b_0(\mathbf{X}(t_i))}. \quad (4)$$

We propose a modified version of the wSSA—the “doubly weighted SSA” (dwSSA)—in which both reaction selection and time to the next reaction are perturbed. The general representation of an SSA importance sampling scheme that possesses these two properties can be found in Chap. 11 of Ref. 7. The advantages of the dwSSA over the wSSA are twofold: (1) the dwSSA makes possible the characterization of rare events in some systems that cannot be interrogated with the wSSA, and (2) the dwSSA enables an automated method for properly choosing the predilection function parameters $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_M]$. We illustrate the first advantage with an example in Sec. V, and we present the automated method behind the second advantage in Sec. III.

The dwSSA selects reaction indices in the same way as the wSSA, but it generates the time to the next reaction using a modified exponential distribution with mean $1/b_0(\mathbf{X}(t))$. Thus, the probability of a reaction trajectory \mathbf{J} under the dwSSA take the form:

$$\begin{aligned} P_{\text{dwSSA}}(\mathbf{J}) &= \prod_{i=1}^{N_T} \left[b_0(\mathbf{X}(t_i)) e^{-b_0(\mathbf{X}(t_i))\tau_i} d\tau_i \times \frac{b_{j'_i}(\mathbf{X}(t_i))}{b_0(\mathbf{X}(t_i))} \right] \\ &= \prod_{i=1}^{N_T} [b_{j'_i}(\mathbf{X}(t_i)) e^{-b_0(\mathbf{X}(t_i))\tau_i} d\tau_i]. \end{aligned} \quad (5)$$

The correcting weight, whose product with the probability (5) equals the probability (1), is

$$\begin{aligned} W_{\text{dwSSA}}(\mathbf{J}) &= \prod_{i=1}^{N_T} \left[\frac{a_{j'_i}(\mathbf{X}(t_i)) e^{-a_0(\mathbf{X}(t_i))\tau_i}}{b_{j'_i}(\mathbf{X}(t_i)) e^{-b_0(\mathbf{X}(t_i))\tau_i}} \right] \\ &= \prod_{i=1}^{N_T} [\exp \{(b_0(\mathbf{X}(t_i)) - a_0(\mathbf{X}(t_i))) \tau_i\} \times (\gamma_{j'_i})^{-1}]. \end{aligned} \quad (6)$$

The remaining steps of the dwSSA are identical to the wSSA and are described in algorithm form in Sec. IV. We note that dwSSA trajectories, unlike wSSA trajectories, can be viewed as SSA trajectories of a modified system where the original rate constant of each reaction has been multiplied by the corresponding predilection function parameter γ_j .

In this paper, we estimate rare event probabilities of the form $p(\mathbf{x}_0, \mathcal{E}; t)$, defined as the probability that the system starting at time 0 in state \mathbf{x}_0 will first reach any state in the set \mathcal{E} at some time $\leq t$. The Monte Carlo estimator of this quantity using the standard SSA is

$$\hat{p}_{\text{SSA}}(\mathbf{x}_0, \mathcal{E}; t) = \frac{1}{K} \sum_{k=1}^K [I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}}], \quad (7)$$

where \mathbf{J}_k is the k th SSA trajectory simulated over the time interval $[0, \mathcal{T}]$ with initial state \mathbf{x}_0 , K is the total number of trajectories, and $I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}}$ takes a value of 1 if any of the states visited by \mathbf{J}_k (denoted by $S(\mathbf{J}_k)$) are in \mathcal{E} (0 otherwise). The expression in Eq. (7) is equivalent to the number of trajectories that successfully reach the rare event divided by the total number of trajectories.

Similarly, using the weight in Eq. (6), the Monte Carlo estimator for $p(\mathbf{x}_0, \mathcal{E}; t)$ using the dwSSA is

$$\hat{p}_{\text{dwSSA}}(\mathbf{x}_0, \mathcal{E}; t) = \frac{1}{K} \sum_{k=1}^K [I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}} W_{\text{dwSSA}}(\mathbf{J}_k)], \quad (8)$$

where \mathbf{J}_k now represents the k th simulated dwSSA trajectory. Equation (8) is equivalent to the sum of the dwSSA weights for trajectories successfully reaching the rare event divided by the total number of trajectories.

III. AUTOMATIC SELECTION OF DWSSA PARAMETER VALUES

Like the wSSA, the dwSSA requires user-defined predilection function parameters. While the wSSA extension detailed in Ref. 6 allows the user to assess the effects of estimator variance, it does not provide *a priori* guidance for selecting parameter values. However, the dwSSA combined with the information-theoretic concept of cross entropy enables a fully automated, efficient method for selecting low-variance parameter values. In the following subsections we describe the integration of the dwSSA into the general cross-entropy (CE) method of Rubinstein.^{8,9} The first subsection begins by deriving an expression “cross entropy,” that approximates the estimator variance conferred by a given set of dwSSA parameters. Next, we identify an optimization problem whose solution is the set of parameter values that minimizes the cross entropy. Finally, we outline a fully automated algorithm that solves the above problem without requiring any prior system knowledge. The second subsection derives a closed-form solution for the optimal parameter values that requires minimal computational expense.

A. Application of the cross-entropy method to the dwSSA

We saw in Eq. (8) that the dwSSA estimator for $p(\mathbf{x}_0, \mathcal{E}; t) \equiv p$ incorporates the ratio of trajectory probabilities $W_{\text{dwSSA}}(\mathbf{J}) = P_{\text{SSA}}(\mathbf{J})/P_{\text{dwSSA}}(\mathbf{J})$. Assume for the moment that we had complete freedom to weight our system such that $P_{\text{dwSSA}}(\mathbf{J})$ could take the form of $P_{\text{dwSSA}}^*(\mathbf{J})$, defined as:

$$P_{\text{dwSSA}}^*(\mathbf{J}) = \frac{I_{\{S(\mathbf{J}) \cap \mathcal{E}\}} P_{\text{SSA}}(\mathbf{J})}{p}. \quad (9)$$

If we use $P_{\text{dwSSA}}^*(\mathbf{J})$ in place of $P_{\text{dwSSA}}(\mathbf{J})$ to compute $W_{\text{dwSSA}}(\mathbf{J})$, upon substitution into Eq. (8) and some algebraic manipulation we see that $\hat{p}_{\text{dwSSA}}(\mathbf{x}_0, \mathcal{E}; t) = p$. Put another way, our estimator exhibits zero variance and perfect accuracy. However, use of Eq. (9) is obviously impractical, since even if we knew the exact value for p we would have no way of actually producing dwSSA trajectories that satisfied Eq. (9). Instead, suppose we choose $P_{\text{dwSSA}}(\mathbf{J})$ to minimize some measure of distance between itself and $P_{\text{dwSSA}}^*(\mathbf{J})$. Specifically, we minimize the *cross entropy* \mathcal{D} (Kullback–Leibler

divergence), which is defined as:

$$\begin{aligned} \mathcal{D}(P_{\text{dwSSA}}^*, P_{\text{dwSSA}}) & \equiv E_{P^*} \left[\ln \frac{P_{\text{dwSSA}}^*(\mathbf{J})}{P_{\text{dwSSA}}(\mathbf{J})} \right] \\ & = E_{P^*} [\ln P_{\text{dwSSA}}^*(\mathbf{J})] - E_{P^*} [\ln P_{\text{dwSSA}}(\mathbf{J})] \\ & \approx \frac{1}{K} \sum_{k=1}^K [\ln P_{\text{dwSSA}}^*(\mathbf{J}^*_k)] - \frac{1}{K} \sum_{k=1}^K [\ln P_{\text{dwSSA}}(\mathbf{J}^*_k)] \\ & \approx \frac{1}{K} \sum_{k=1}^K [\ln P_{\text{dwSSA}}^*(\mathbf{J}^*_k)] \\ & \quad - \frac{1}{K} \sum_{k=1}^K \left[\ln P_{\text{dwSSA}}(\mathbf{J}_k) \times \frac{P_{\text{dwSSA}}^*(\mathbf{J}_k)}{P_{\text{SSA}}(\mathbf{J}_k)} \right] \\ & \equiv D. \end{aligned} \quad (10)$$

Here, E_{P^*} is the expectation operator with respect to the (impractical) $P_{\text{dwSSA}}^*(\mathbf{J})$ -associated system, and \mathbf{J}^*_k and \mathbf{J}_k represent k th trajectories generated from the $P_{\text{dwSSA}}^*(\mathbf{J})$ -associated and $P_{\text{SSA}}(\mathbf{J})$ -associated systems, respectively. The second to last line in Eq. (10) is a Monte Carlo approximation whose second term depicts the use of importance sampling to transform the reference system from one with trajectory probability $P_{\text{dwSSA}}^*(\mathbf{J})$ to one with probability $P_{\text{SSA}}(\mathbf{J})$.

Substituting Eq. (9) into Eq. (10) and denoting $P_{\text{dwSSA}}(\mathbf{J})$ more precisely as $P_{\text{dwSSA}}(\mathbf{J}; \boldsymbol{\gamma})$, the last two lines of (10) become:

$$\begin{aligned} D(\boldsymbol{\gamma}) & \equiv \frac{1}{K} \sum_{k=1}^K \left[\ln \frac{I_{\{S(\mathbf{J}^*_k) \cap \mathcal{E}\}} P_{\text{SSA}}(\mathbf{J}^*_k)}{p} \right] \\ & \quad - \frac{1}{K} \sum_{k=1}^K \left[\ln P_{\text{dwSSA}}(\mathbf{J}_k; \boldsymbol{\gamma}) \times \frac{I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}}}{p} \right], \end{aligned} \quad (11)$$

where D can be viewed as a function of the dwSSA parameters $\boldsymbol{\gamma}$. Our goal is to minimize Eq. (11) with respect to $\boldsymbol{\gamma}$. This is equivalent to the simpler maximization problem:

$$\max_{\boldsymbol{\gamma}} \left(\sum_{k=1}^K [I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}} \times \ln P_{\text{dwSSA}}(\mathbf{J}_k; \boldsymbol{\gamma})] \right), \quad (12)$$

since the first term in Eq. (11) does not depend on $\boldsymbol{\gamma}$ and p is a constant. In typical applications, the argument in Eq. (12) is a convex function of $\boldsymbol{\gamma}$ (i.e., hill-shaped) and differentiable,⁸ so we can produce Monte Carlo estimates of the dwSSA parameter values that confer minimum cross entropy ($\hat{\boldsymbol{\gamma}}^*$) by taking partial derivatives with respect to each γ_j and setting the resulting expressions to $\mathbf{0}$:

$$\sum_{k=1}^K \left[I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}} \times \nabla_{\boldsymbol{\gamma}} \ln P_{\text{dwSSA}}(\mathbf{J}_k; \hat{\boldsymbol{\gamma}}^*) \right] = \mathbf{0}. \quad (13)$$

The form of $P_{\text{dwSSA}}(\mathbf{J}; \boldsymbol{\gamma})$ given in Eq. (5) enables an analytical solution to Eq. (13) (details below). This is in contrast to the wSSA trajectory probability Eq. (3) which, when substituted into Eq. (13), would require a much more expensive numerical solution. The latter follows from the observation that

the factor $\prod_{i=1}^{N_T} [1/b_0(\mathbf{X}(t_i))]$ in Eq. (3) does not cancel like it does in Eq. (5); consequently, upon taking the logarithm and differentiating Eq. (13), we cannot compute a closed form solution for $\hat{\boldsymbol{\gamma}}^*$. In general, distributions belonging to a natural exponential family (such as the dwSSA trajectory distribution Eq. (5)) lead to closed form expressions for cross-entropy parameter estimates.⁸ In this case, the analytical solution to Eq. (13) provided by the dwSSA results in a computationally efficient, principled method for selecting low-variance parameter values.

In principle, we can solve Eq. (13) for $\hat{\boldsymbol{\gamma}}^*$ by simulating K unweighted system trajectories with the SSA. However, a practical difficulty arises: since we are trying to estimate the probability of a *rare* event, the vast majority of the $I_{\{S(\mathbf{J}_k) \cap \mathcal{E}\}}$ will be zero. Fortunately, an extension of the CE method circumvents this difficulty. Just as the dwSSA uses importance sampling (IS) to bias system trajectories toward a rare event of interest, a *multilevel* version of the CE method invokes IS (here, in the form of the dwSSA) to bias trajectories for the optimization problem in Eq. (12) toward the rare event.⁸ Given some reference parameter vector $\boldsymbol{\gamma}^{(0)}$, which we assume biases the system towards the rare event, we can rewrite Eq. (12) as the asymptotically ($K \rightarrow \infty$) equivalent:

$$\max_{\boldsymbol{\gamma}} \left(\sum_{k=1}^K [I_{\{S(\mathbf{J}_k^{(0)}) \cap \mathcal{E}\}} \times W_{\text{dwSSA}}(\mathbf{J}_k^{(0)}; \boldsymbol{\gamma}^{(0)}) \times \ln P_{\text{dwSSA}}(\mathbf{J}_k^{(0)}; \boldsymbol{\gamma})] \right), \quad (14)$$

where $\mathbf{J}_k^{(0)}$ is the k th trajectory generated using the dwSSA parameterized with $\boldsymbol{\gamma}^{(0)}$, and $W_{\text{dwSSA}}(\mathbf{J})$ in (6) is written more precisely as a function of $\boldsymbol{\gamma}^{(0)}$. Similarly, Eq. (13) can be rewritten as the asymptotically equivalent:

$$\sum_{k=1}^K [I_{\{S(\mathbf{J}_k^{(0)}) \cap \mathcal{E}\}} \times W_{\text{dwSSA}}(\mathbf{J}_k^{(0)}; \boldsymbol{\gamma}^{(0)}) \times \nabla_{\boldsymbol{\gamma}} \ln P_{\text{dwSSA}}(\mathbf{J}_k^{(0)}; \hat{\boldsymbol{\gamma}}^*)] = \mathbf{0}. \quad (15)$$

The challenge now becomes how to choose $\boldsymbol{\gamma}^{(0)}$ correctly. The multilevel CE method obviates this difficulty by defining a series of intermediate “less rare” events and sequentially

biasing the system towards these events until the final rare event is reached.

We begin by simulating K trajectories of the system in the interval $[0, T]$ using the dwSSA with all parameters set to 1 (\equiv SSA). We record the top $\lceil \rho K \rceil$ trajectories (where ρ is typically $\sim 10^{-2}$) that evolve farthest in the direction of the set \mathcal{E} , and we label those states reached by the $\lceil \rho K \rceil$ recorded trajectories that are closest to \mathcal{E} (one per trajectory) as \mathcal{E}_0 . The set \mathcal{E}_0 represents a “less rare” event [since $\hat{P}_{\text{dwSSA}}(\mathbf{x}_0, \mathcal{E}_0; t) \geq \rho$] which can be used to generate an intermediate set of dwSSA parameters that bias the system in the direction of the original rare event. Specifically, if we replace \mathcal{E} in Eq. (13) with \mathcal{E}_0 , we can solve for the corresponding optimal dwSSA parameters $\hat{\boldsymbol{\gamma}}^{(0)}$ (details presented in the following subsection). Since at least $\lceil \rho K \rceil$ trajectories will have reached \mathcal{E}_0 during the course of the simulation, this solution should be relatively robust.

If we then simulate K trajectories $\mathbf{J}_1^{(0)}, \dots, \mathbf{J}_K^{(0)}$ of the system using the dwSSA parameterized with $\hat{\boldsymbol{\gamma}}^{(0)}$, we can define a set of states \mathcal{E}_1 that is analogous to \mathcal{E}_0 but closer to \mathcal{E} . We then solve a version of Eq. (15) where $\boldsymbol{\gamma}^{(0)}$ has been replaced with $\hat{\boldsymbol{\gamma}}^{(0)}$, and \mathcal{E} and $\hat{\boldsymbol{\gamma}}^*$ have been replaced with \mathcal{E}_1 and $\hat{\boldsymbol{\gamma}}^{(1)}$, respectively. As a result, we will have identified a second set of intermediate dwSSA parameters $\hat{\boldsymbol{\gamma}}^{(1)}$ that bias the system farther in the direction of the original rare event.

The above procedure can then be repeated n times, until the intermediate set of states \mathcal{E}_n is contained in \mathcal{E} . At this point, we can substitute the last set of generated trajectories $\mathbf{J}_1^{(n-1)}, \dots, \mathbf{J}_K^{(n-1)}$ and corresponding parameter estimates $\hat{\boldsymbol{\gamma}}^{(n-1)}$ into an otherwise unaltered version of Eq. (15) to solve for $\hat{\boldsymbol{\gamma}}^*$. In so doing, we will have computed a robust estimate of the optimal dwSSA parameters through a series of intermediate steps which gradually bias the system toward the original rare event. We provide a detailed algorithmic description of the above method in Sec. IV below.

B. Closed-form solution for low-variance dwSSA parameter values

We now present a derivation of the analytical solution to Eq. (15) [and hence Eq. (13)]. Upon substituting Eq. (5) into Eq. (15) and suppressing detail in the first two factors inside the summation, we obtain:

$$\begin{aligned} \mathbf{0} &= \sum_{k=1}^K \left[I_k \times W_k \times \nabla_{\boldsymbol{\gamma}} \ln \left(\prod_{i=1}^{N_{\tau_k}} [b_{j_{ki}}(\mathbf{X}_k(t_{ki})) e^{-b_0(\mathbf{X}_k(t_{ki}))\tau_{ki}} d\tau_i] \right) \right] \\ &= \sum_{k=1}^K \left[I_k \times W_k \times \nabla_{\boldsymbol{\gamma}} \ln \left(\prod_{i=1}^{N_{\tau_k}} \left[\hat{\boldsymbol{\gamma}}_{j_{ki}}^* a_{j_{ki}}(\mathbf{X}_k(t_{ki})) \exp \left\{ -\tau_{ki} \sum_{j=1}^M [\hat{\boldsymbol{\gamma}}_j^* a_j(\mathbf{X}_k(t_{ki}))] \right\} d\tau_i \right] \right) \right]. \end{aligned} \quad (16)$$

Upon taking the logarithm, collecting terms not depending on $\hat{\gamma}^*$ in C_{ki} , and simplifying, we get:

$$\begin{aligned} \mathbf{0} &= \sum_{k=1}^K \left[I_k \times W_k \times \nabla_{\boldsymbol{\gamma}} \left(\sum_{i=1}^{N_{T_k}} \left[\ln(\hat{\gamma}_{j_{ki}}^*) - \tau_{ki} \sum_{j=1}^M [\hat{\gamma}_j^* a_j(\mathbf{X}_k(t_{ki}))] + C_{ki} \right] \right) \right] \\ &= \sum_{k=1}^K \left[I_k \times W_k \times \nabla_{\boldsymbol{\gamma}} \left(\sum_{j=1}^M [n_{kj} \ln(\hat{\gamma}_j^*)] - \sum_{i=1}^{N_{T_k}} \left[\tau_{ki} \sum_{j=1}^M [\hat{\gamma}_j^* a_j(\mathbf{X}_k(t_{ki}))] + C_{ki} \right] \right) \right] \\ &= \sum_{k=1}^K \left[I_k \times W_k \times \nabla_{\boldsymbol{\gamma}} \left(\sum_{j=1}^M [n_{kj} \ln(\hat{\gamma}_j^*)] \right) - I_k \times W_k \times \nabla_{\boldsymbol{\gamma}} \left(\sum_{i=1}^{N_{T_k}} \left[\tau_{ki} \sum_{j=1}^M [\hat{\gamma}_j^* a_j(\mathbf{X}_k(t_{ki}))] \right] \right) \right]. \end{aligned} \quad (17)$$

where n_{kj} is the total number of times reaction j fires in the k th trajectory. After differentiation, we obtain a scalar version of Eq. (17) for each reaction j :

$$0 = \sum_{k=1}^K \left(I_k \times W_k \times \frac{n_{kj}}{\hat{\gamma}_j^*} - I_k \times W_k \times \sum_{i=1}^{N_{T_k}} [a_j(\mathbf{X}_k(t_{ki})) \tau_{ki}] \right), \quad (18)$$

which leads to the following detailed closed-form expression for each reaction's optimal parameter estimate:

$$\hat{\gamma}_j^{(n)} = \frac{\sum_k' (W_{\text{dwSSA}}(\mathbf{J}_k^{(n-1)}; \hat{\boldsymbol{\gamma}}^{(n-1)}) \times n_{kj})}{\sum_k' (W_{\text{dwSSA}}(\mathbf{J}_k^{(n-1)}; \hat{\boldsymbol{\gamma}}^{(n-1)}) \times \sum_{i=1}^{N_{T_k}} [a_j(\mathbf{X}_k^{(n-1)}(t_{ki})) \tau_{ki}]}. \quad (19)$$

For clarity, rare event indicators $I_{\{\cdot\}}$ in Eq. (19) have been replaced by summations \sum_k' , where k iterates only over trajectories reaching the rare event. We note that Eq. (19) represents one of M uncoupled equations from the *final* step of the multilevel algorithm discussed above, where $\hat{\boldsymbol{\gamma}}^{(n)} \equiv \hat{\boldsymbol{\gamma}}^*$. In practice, we solve each of these equations at each step of the CE method until the final parameter estimates are obtained; in this way, each step's computation of estimates relies upon the previous step's values. For an intuitive explanation of why Eq. (19) works, we note that the numerator represents a weighted sum of the total number of times reaction j fires across the successful trajectories, while the denominator is a weighted sum of the *expected* total number of times reaction j will fire across those same trajectories. Reactions that are needed to fire more often than their average behavior to reach the rare event will thus acquire a $\hat{\gamma}_j^*$ greater than 1, while reactions needed to fire less often than average will acquire a $\hat{\gamma}_j^*$ less than 1.

IV. ALGORITHMS

Algorithm 1 (modeled after the wSSA in Ref. 6) implements the dwSSA, which modifies Kuwahara and Mura's wSSA by using $b_0(\mathbf{X}(t))$ to generate both reaction indices and times to the next reaction. It returns an estimate of the rare event probability $p(\mathbf{x}_0, \mathcal{E}; t)$ for a given set of dwSSA parameters $\boldsymbol{\gamma}$.

Algorithm 1. The dwSSA.

- 1: $m_K \leftarrow 0$
- 2: **for** $k = 1$ to K **do**
- 3: $t \leftarrow 0$, $\mathbf{x} \leftarrow \mathbf{x}_0$, $w \leftarrow 1$
- 4: evaluate all $a_j(\mathbf{x})$ and $b_j(\mathbf{x})$; calculate $a_0(\mathbf{x})$ and $b_0(\mathbf{x})$
- 5: **while** $t \leq t_f$ **do**
- 6: **if** $\mathbf{x} \in \mathcal{E}$ **then**
- 7: $m_K \leftarrow m_K + w$
- 8: break out of the while loop
- 9: **end if**
- 10: generate two unit-interval uniform random numbers r_1 and r_2
- 11: $\tau \leftarrow b_0^{-1}(\mathbf{x}) \ln(1/r_1)$
- 12: $j \leftarrow$ smallest integer satisfying $\sum_{i=1}^j b_i(\mathbf{x}) \geq r_2 b_0(\mathbf{x})$
- 13: $w \leftarrow w \times (\gamma_j)^{-1} \times \exp\{(b_0(\mathbf{x}) - a_0(\mathbf{x}))\tau\}$
- 14: $t \leftarrow t + \tau$, $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{v}_j$
- 15: update all $a_j(\mathbf{x})$ and $b_j(\mathbf{x})$; recalculate $a_0(\mathbf{x})$ and $b_0(\mathbf{x})$
- 16: **end while**
- 17: **end for**
- 18: **return** $\hat{p}_{\text{dwSSA}}(\mathbf{x}_0, \mathcal{E}; t) = m_K/K$

In the above, \mathbf{v}_j and t_f represent the state change vector for reaction j and the simulation end time, respectively. The computational complexity of Algorithm 1 is identical to that of the wSSA with given biasing parameters.

Algorithm 2 implements the multilevel cross-entropy method for optimal dwSSA parameter estimation. It returns the vector of optimal parameter estimates $\hat{\boldsymbol{\gamma}}^*$.

Algorithm 2. Optimal dwSSA parameter estimation by multilevel cross-entropy method.

- 1: $\boldsymbol{\gamma} \leftarrow [1 \quad 1 \quad \dots \quad 1]$, $i \leftarrow -1$
- 2: **repeat**
- 3: $i \leftarrow i + 1$
- 4: run **Algorithm 1**; mark the $\lceil \rho K \rceil$ trajectories evolving farthest in the direction of \mathcal{E}
- 5: $\mathcal{E}_i \leftarrow$ at most $\lceil \rho K \rceil$ states closest to \mathcal{E} reached by the marked trajectories (one per trajectory)
- 6: $\boldsymbol{\gamma} \leftarrow$ result of (19) evaluated using \mathcal{E}_i and trajectories from step 4
- 7: **until** $\mathcal{E}_i \subseteq \mathcal{E}$
- 8: **return** $\hat{\boldsymbol{\gamma}}^* = \boldsymbol{\gamma}$

As written, the computational complexity of Algorithm 2 is roughly $n \times$ the complexity of the dwSSA, where n is the number of iterations needed for $\mathcal{E}_n \subseteq \mathcal{E}$. However, when taken together, steps 4–6 of Algorithm 2 require the storage of K independent dwSSA trajectories. For large K ($\geq 10^7$), this requirement becomes prohibitive. To circumvent this difficulty, we typically run step 4 twice for all but the final iteration of the loop—once to identify \mathcal{E}_i in step 5, and a second time immediately afterwards (using the same random number seed as for the first) to compute the current optimal parameter estimates $\boldsymbol{\gamma}$. Using this modified approach, we only have to store the $\lceil \rho K \rceil$ states closest to \mathcal{E} that are reached by the marked trajectories. The practical complexity of Algorithm 2 is thus roughly $(2n - 1) \times$ the complexity of the dwSSA.

Overall, the computation of a rare event probability estimate $\hat{p}_{\text{dwSSA}}(\mathbf{x}_0, \mathcal{E}; t)$ requires one run of Algorithm 2 to produce $\hat{\boldsymbol{\gamma}}^*$ followed by one run of Algorithm 1 using those parameters as input. This leads to a total complexity of $2n \times$ the complexity of the dwSSA. As we discuss in Sec. V, all the examples we tested required $n \leq 4$. We note that in practice, the number K of realizations used in Algorithm 2 is several orders of magnitude smaller than what is typically used in the final run of Algorithm 1. Consequently, the time required to estimate parameters using the multilevel cross-entropy method is a small fraction of the total simulation time.

V. EXAMPLES

We now illustrate dwSSA performance on the following four examples: a pure birth process, a birth-death process, an enzymatic futile cycle, and a yeast polarization model. We estimate rare event probabilities for each example by first running four independent realizations of Algorithm 2 with $K = 10^5$. We compute the mean of the resulting parameter estimates to arrive at a consensus $\hat{\boldsymbol{\gamma}}^*$. We then run four independent realizations of Algorithm 1 with varying K , yielding distributions of estimates for the rare event probability. Finally, using four independent ensembles of $K = 10^7$ trajectories each (4×10^7 total independent simulations), we compute the mean probability estimate as well as the estimate uncertainty (using methods described in Gillespie *et al.*⁶) to produce a 68% confidence interval. When possible, we also analyze the examples using the wSSA parameterized with the

optimal values given in Ref. 6 to compare the accuracy of the two methods.

A. Pure birth process

The only algorithmic difference between the original wSSA and the dwSSA is that the latter biases the time to the next reaction τ as well as reaction selection [see Eqs. (5) and (6)]. If the probability of an event is small because its occurrence requires certain reactions to fire considerably more/less than their average number of firings, an importance sampling approach that biases reaction selection but not τ (i.e., the wSSA) may not be sufficient. Although the wSSA indirectly influences times to the next reaction through altered reaction selection, the direct biasing of τ (i.e., using the dwSSA) can increase/decrease the average number of reactions fired during a simulation to a much greater degree. The simplest example illustrating this point is a pure birth process (homogeneous Poisson process): a single reaction model where only the time to the next reaction can be weighted ($(a_j/a_0)/(b_j/b_0) = 1$ since $a_0 = a_j$ and $b_0 = b_j$). The model is specified as follows:



with $\mathbf{x}_0 = [0]$. For this and the following examples, we simplify our definition of a rare event by limiting the states of interest \mathcal{E} to those governed by only a single species S . Specifically, we define a threshold species count θ^S above/below which the event occurs, rewriting $p(\mathbf{x}_0, \mathcal{E}; t)$ as $p(\mathbf{x}_0, \theta^S; t)$. For the pure birth process, we compute estimates of $p([0], 75; 50)$ —the probability that the population of S reaches 75 before time 50. The mean population of S at $t = 50$ is 35, which is equivalent to an average of 35 reactions occurring during a single simulation. Thus, to reach the rare event the system must produce more than twice the average number of S molecules.

Table I summarizes the results of running Algorithm 2 with $\rho = 0.01$ on the pure birth process. We see that the algorithm has converged to the original rare event threshold by $n = 3$ steps, yielding an optimal parameter estimate $\hat{\boldsymbol{\gamma}}^* = [2.194]$. We substituted this value into Algorithm 1 and ran it independently four times each for $K \in \{10^4, 10^5, 10^6, 10^7, 10^8\}$; Fig. 1 displays the results. Using the properties of the homogeneous Poisson process, we computed the exact value of $p([0], 75; 50) (= 2.981 \times 10^{-9})$, which we display as a green line in Fig. 1. While the dwSSA estimates quickly converge to the true probability, we did not observe a single rare event occurrence from wSSA simulations of equivalent K . Because the wSSA is identical to the SSA applied to a one reaction system, both the SSA and the wSSA are inefficient in simulating the above system. Finally, we computed the mean probability estimate and uncertainty for the dwSSA with four independent runs of $K = 10^7$, yielding an estimate that is identical to the exact probability:

$$\hat{p}_{\text{dwSSA}}([0], 75; 50) = 2.9808 \times 10^{-9} \pm 0.0011 \times 10^{-9}. \quad (20)$$

Using the formula described in Ref. 6, we would expect an SSA (equivalently, a wSSA) estimate of similar accuracy to

TABLE I. Results of the multilevel cross-entropy algorithm applied to the pure birth process. The first column denotes the iteration number, the second column labels which of four independent realizations are displayed, the third column specifies the intermediate rare event threshold, and the fourth column presents the intermediate optimal parameter estimate. By the third step, at least $\lceil \rho K \rceil$ of the dwSSA simulated trajectories have reached the original rare event threshold.

Step (i)	Trial No.	θ_i^S	$\hat{\gamma}^{(i)}$
1	1	49	[1.480]
	2	49	[1.483]
	3	49	[1.484]
	4	49	[1.482]
2	1	69	[2.029]
	2	69	[2.026]
	3	69	[2.028]
3	4	69	[2.025]
	1	75	[2.194]
	2	75	[2.194]
	3	75	[2.194]
	4	75	[2.194]

Eq. (20) to require over 10^{15} trajectories, which corresponds to a dwSSA computational gain of $> 10^7$.

Although the pure birth process is a somewhat extreme illustration of the advantage of the dwSSA over the wSSA, we note that the same advantage will exist for more complex models in which the average number of reactions occurring

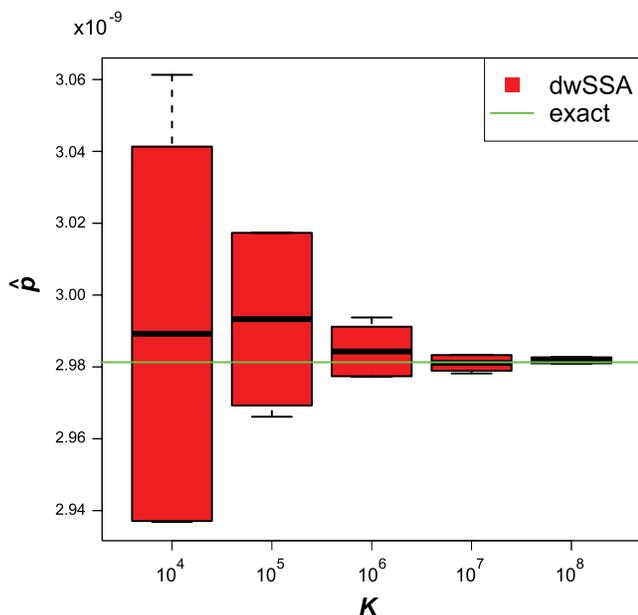


FIG. 1. Convergence plot of rare event probability estimate (\hat{p}) vs simulation ensemble size (K) for the pure birth process. Each boxplot displays (moving outwardly) the mean, ± 1 standard deviation, and minimum and maximum of four independent dwSSA ensembles for a given K . We parameterized the dwSSA with $\hat{\gamma}^* = [2.194]$, determined by calculating the mean of four independent realizations of Algorithm 2 ($K = 10^5$). The horizontal green line indicates the analytically determined rare event probability (p). As K increases, the dwSSA estimates converge to the true probability. Results for the wSSA are not shown, as it is algorithmically identical to the unweighted SSA for this model and does not result in the observation of any rare event occurrences.

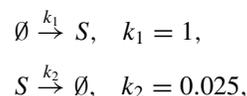
TABLE II. Results of the multilevel cross-entropy algorithm applied to the birth–death process. The column identities match those of Table I. In this example, only two steps were required for $\lceil \rho K \rceil$ of the dwSSA trajectories to reach the original rare event threshold.

Step (i)	Trial No.	θ_i^S	$\hat{\gamma}^{(i)}$
1	1	61	[1.255 0.805]
	2	61	[1.260 0.800]
	3	61	[1.256 0.800]
	4	61	[1.252 0.801]
2	1	80	[1.452 0.693]
	2	80	[1.458 0.685]
	3	80	[1.452 0.685]
	4	80	[1.454 0.679]

by time t is far larger/smaller than the number needed to reach the rare event.

B. Birth–death process

Our second example is a birth–death process. This system consists of two reactions and thus requires the estimation of two biasing parameters. The model description is as follows:



with $\mathbf{x}_0 = [40]$. We note that this model is identical to the single species production–degradation model in Refs. 5 and 6, with the unchanging S_1 removed for simplicity. In the above description, the kinetic constants and initial conditions are set such that the system is in stochastic equilibrium. The rare event probability we examine is $p(\mathbf{x}_0, \theta^S; t) \equiv p([40], 80; 100)$. In Gillespie *et al.*,⁶ the authors estimated the optimal parameters for the wSSA as $\hat{\gamma}_{\text{wSSA}}^* = [1.30 \ 0.769]$, where the second parameter is simply the reciprocal of the first. By adopting this reciprocal constraint originally introduced by Kuwahara and Mura,⁵ Gillespie *et al.* reduced the parameter space to a single parameter; nevertheless, their selection algorithm required a minimum of seven parameter evaluations, each consisting of 4×10^7 runs of the wSSA (2.8×10^8 runs total). Were the parameter space not reduced, $> 10^{16}$ runs would have been required to estimate optimal parameters. In contrast, our multilevel cross-entropy approach coupled with the dwSSA did not require a parameter space reduction, and we recovered optimal parameter estimates in $4 \times (2n - 1) \times 10^5 = 1.2 \times 10^6$ runs of the dwSSA ($n = 2$ steps in this example). Table II summarizes the results of running Algorithm 2 with $\rho = 0.01$, which we averaged to obtain $\hat{\gamma}^* = [1.454 \ 0.686]$.

The wSSA and dwSSA optimal parameter estimates for the birth–death process are not identical. The discrepancy is due to the added τ weighting employed by the dwSSA; however, we note that the reciprocal relationship of the two parameters is roughly preserved. To test the sensitivity of Algorithm 2 to the proportion of trajectories required to cross the rare event threshold, we ran an additional four independent realizations with $\rho = 0.1$. The results were very similar

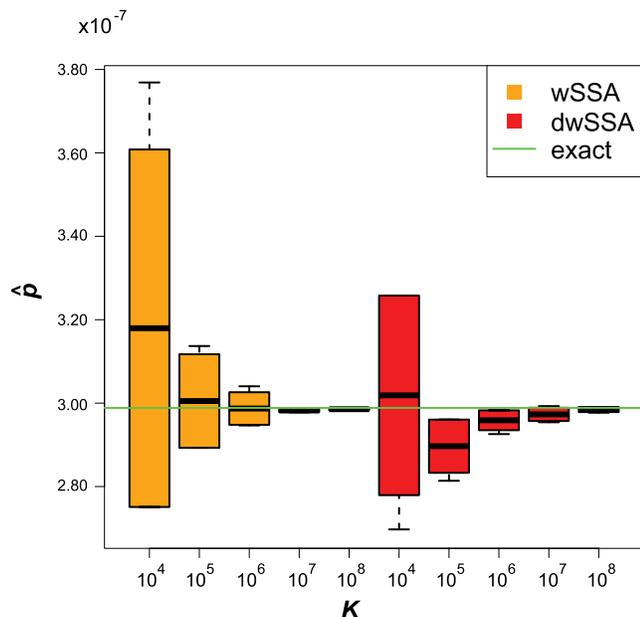


FIG. 2. Convergence plot of rare event probability estimate (\hat{p}) vs simulation ensemble size (K) for the birth–death process. Boxplots are constructed as in Fig. 1, summarizing results of four independent wSSA or dwSSA simulation ensembles for each value of K . We parameterized the dwSSA with the mean of four realizations of Algorithm 2, yielding $\hat{\boldsymbol{\gamma}}^* = [1.454 \ 0.686]$. The wSSA was parameterized with the optimal values discovered in Ref. 6: $\hat{\boldsymbol{\gamma}}_{wSSA}^* = [1.30 \ 0.769]$. As before, the green line denotes the exact rare event probability. With increasing K , both the wSSA and dwSSA estimates converge to the true probability.

($\hat{\boldsymbol{\gamma}}^* = [1.452 \ 0.692]$), suggesting that the cross-entropy approach is not particularly sensitive to ρ .

Figure 2 displays the results of running four independent realizations of both the wSSA and dwSSA on the birth–death process with varying K . Both methods’ estimates for $p([40], 80; 100)$ converge to the true probability ($= 2.986 \times 10^{-7}$), obtained using the system generator matrix (details in Ref. 5) as the simulation ensemble size increases. We computed the mean probability estimate and uncertainty for the dwSSA with four independent ensembles of $K = 10^7$, yielding:

$$\hat{p}_{dwSSA}([40], 80; 100) = 2.971 \times 10^{-7} \pm 0.007 \times 10^{-7}. \quad (21)$$

Our multilevel cross-entropy approach to estimating optimal dwSSA parameters relies on a close correspondence between cross entropy and estimator variance. To evaluate this correspondence, we performed a sensitivity analysis of dwSSA estimator variance with respect to $\boldsymbol{\gamma}$; Fig. 3 displays the results. Across the range of parameter values tested (we simulated $K = 10^8$ dwSSA trajectories for each parameter combination), the variance ranged from 1.06×10^{-11} (enclosed by the red rectangle) to 5.34×10^{-7} . The optimal parameter combination returned by Algorithm 2 achieved a variance of 3.32×10^{-11} (enclosed by the green rectangle). The similarity of the minimum variance overall and the variance associated with Algorithm 2 suggests that the assumption made by the cross-entropy method—i.e., that minimum cross entropy closely approximates minimum variance—is justified. Thus, we expect optimal parameter estimates de-

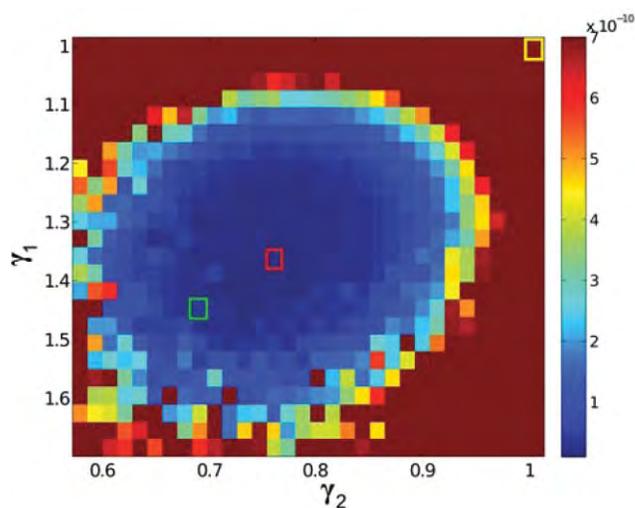
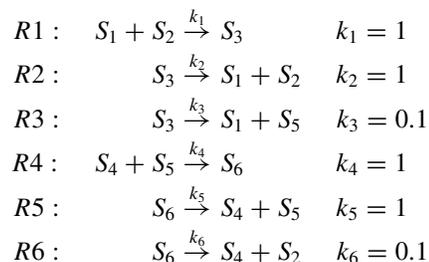


FIG. 3. Sensitivity of rare event probability estimator variance to dwSSA parameter values γ_1 and γ_2 for the birth–death process. Each rectangle displays the dwSSA estimator variance when run with $K = 10^8$ using the corresponding values of γ_1 and γ_2 . Pseudocolor represents variance magnitude, with dark red denoting the highest variance and dark blue the lowest (best performance). Parameter combinations conferring variance $\geq 7 \times 10^{-10}$ were colored with the darkest red shade; the maximum observed was 5.34×10^{-7} . Variance of the unweighted system (\equiv SSA) is depicted by the yellow rectangle. The green rectangle outline depicts the optimal parameter combination identified using Algorithm 2, while the red rectangle outline corresponds to the minimum variance observed for all combinations tested. The discrepancy between the two is likely due to the imperfect correspondence between minimum variance and minimum cross entropy.

rived from our proposed method to be those that effectively minimize estimator variance.

C. Enzymatic futile cycle

Next we consider an enzymatic futile cycle, which appeared in Kuwahara and Mura⁵ and was later revisited by Gillespie *et al.*⁶ The system is characterized by the following set of six reactions:



with $\mathbf{x}_0 = [1 \ 50 \ 0 \ 1 \ 50 \ 0]$.

This mechanism was described by Samoilov *et al.* and is widely used in such diverse regulatory processes as membrane transport and GTPase cycles.¹⁰ As with the birth–death process, the above model is in stochastic equilibrium. The rare event probability of interest is $p(\mathbf{x}_0, \theta^{S_5}; t) \equiv p([1 \ 50 \ 0 \ 1 \ 50 \ 0], 25; 100)$. Gillespie *et al.*⁶ estimated the optimal wSSA parameters as $\hat{\boldsymbol{\gamma}}_{wSSA}^* = [1 \ 1 \ 0.35 \ 1 \ 1 \ 2.857]$, where the sixth parameter is the reciprocal of the third and the remaining parameters are fixed at 1. These constraints were introduced by Kuwahara and

TABLE III. Results of the multilevel cross-entropy algorithm applied to the futile cycle model. The column identities match those of Tables I–II. In this example, two steps were required for $[\rho K]$ of the dwSSA trajectories to reach the original rare event threshold.

Step (i)	Trial No.	$\theta_j^{S_5}$	$\hat{\gamma}^{(i)}$
1	1	38	[0.999 1.000 0.492 0.996 0.999 1.914]
	2	38	[0.998 1.000 0.502 0.995 0.995 1.932]
	3	38	[0.997 1.001 0.487 1.001 1.001 1.920]
	4	38	[1.001 0.994 0.505 1.007 1.000 1.930]
2	1	25	[0.998 1.001 0.321 1.004 0.995 3.007]
	2	25	[1.003 1.004 0.321 1.004 0.993 3.004]
	3	25	[0.997 1.002 0.320 1.005 0.992 3.012]
	4	25	[1.001 1.005 0.317 1.002 0.994 3.009]

Mura,⁵ and like the birth–death process, they reduce the effective parameter space to a single parameter. In light of this reduction, Gillespie’s *et al.* parameter selection algorithm required a minimum of seven parameter evaluations, each consisting of 4×10^5 wSSA runs (2.8×10^6 total). However, the above parameter space simplification requires considerable insight into the behavior of the enzymatic futile cycle. Specifically, we note that the obvious choice of reactions to perturb via the wSSA would be $R3$, $R4$, and $R5$, since they directly modify the population of S_5 . In contrast, Kuwahara and Mura (and Gillespie *et al.*) chose to perturb only $R3$ and $R6$. Results from numerical experiments we have conducted suggest that perturbing any reactions other than $R3$ and $R6$ returns a rare event probability estimate with considerably lower accuracy (not shown). Without this insight, a naïve application of Gillespie’s *et al.* parameter selection algorithm would have required $> 10^{38}$ runs.

In contrast, our approach required no prior insight and we recovered optimal parameter estimates in 1.2×10^6 runs of the dwSSA ($n = 2$ steps). Table III summarizes the results of running Algorithm 2 with $\rho = 0.01$, which we averaged to obtain $\hat{\gamma}^* = [1.000 \ 1.003 \ 0.320 \ 1.003 \ 0.993 \ 3.008]$. We note that the third parameter of $\hat{\gamma}^*$ is approximately the reciprocal of the sixth parameter and the remaining parameters are very close to 1. Thus, our multilevel cross-entropy approach recovered the optimal parameter constraints automatically.

Figure 4 displays the results of four independent runs of the wSSA and dwSSA (using their respective optimal parameter estimates) on the futile cycle with varying K . Both methods converge to the true probability ($= 1.738 \times 10^{-7}$), obtained using the system generator matrix.⁵ Computation of the mean probability estimate and uncertainty for the dwSSA with four independent ensembles of $K = 10^7$ gave a result that is identical to the true probability:

$$\hat{p}_{\text{dwSSA}}([1 \ 50 \ 0 \ 1 \ 50 \ 0], 25; 100) = 1.7381 \times 10^{-7} \pm 0.0004 \times 10^{-7}. \quad (22)$$

D. Yeast polarization

For our final example, we modified a model of the pheromone-induced G-protein cycle in *Saccharomyces cerevisiae* so that it does not start in nor reach stochastic equilibrium within a 20s simulation time. The original model is described in Drawert *et al.*¹¹ Our modified system consists of

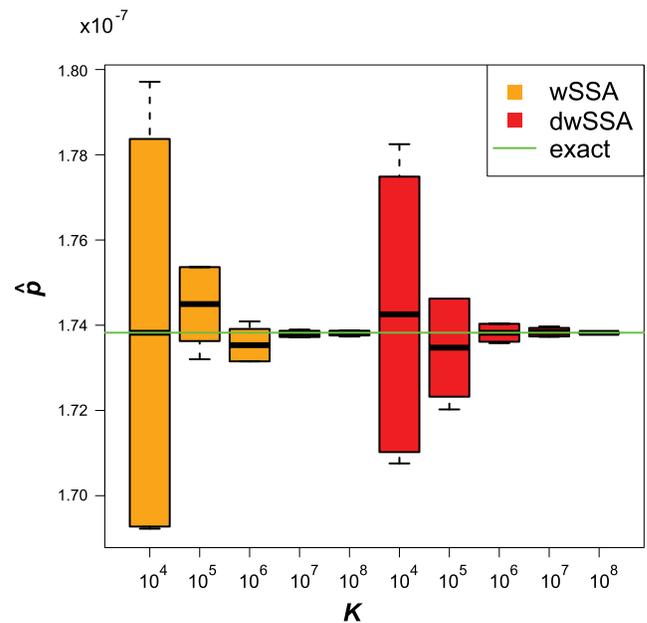
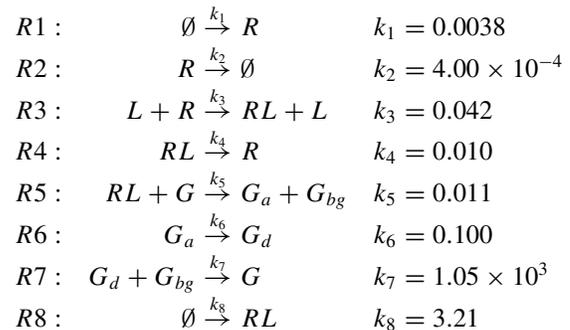


FIG. 4. Convergence plot of rare event probability estimate (\hat{p}) vs simulation ensemble size (K) for the futile cycle model. Boxplots are constructed as in Figs. 1 and 2, summarizing results of four independent wSSA or dwSSA simulation ensembles for each value of K . As before, we parameterized the dwSSA with the mean of four realizations of Algorithm 2, yielding $\gamma = [1.000 \ 1.003 \ 0.320 \ 1.003 \ 0.993 \ 3.008]$. The wSSA was parameterized with the optimal values discovered in Ref. 6: $\gamma = [1 \ 1 \ 0.350 \ 1 \ 1 \ 2.857]$. The green line denotes the exact rare event probability. With increasing K , both the wSSA and dwSSA estimates converge to the true probability.

seven species $\mathbf{x} = [R \ L \ RL \ G \ G_a \ G_{bg} \ G_d]$ and is characterized by the following eight reactions:



with $\mathbf{x}_0 = [50 \ 2 \ 0 \ 50 \ 0 \ 0 \ 0]$. We examine the event probability $p(\mathbf{x}_0, \theta^{G_{bg}}; t) \equiv p([50 \ 2 \ 0 \ 50 \ 0 \ 0 \ 0], 50; 20)$. This event has not been previously characterized, so we began by simulating the system with the unweighted SSA. Using four ensembles of $K = 10^7$, we computed the mean event probability estimate $\hat{p}_{\text{SSA}} = 1.125 \times 10^{-6}$. The event in question is not exceptionally rare; however, the mean uncertainty associated with the estimate leads to a large 68% confidence interval ($[0.9573, 1.293] \times 10^{-6}$) and illustrates the high intrinsic stochasticity of the system. To reduce the uncertainty of the estimate, we ran four independent realizations of Algorithm 2 ($K = 10^5$) with $\rho = 0.01$. These runs were exceedingly slow to converge ($n > 10$) due to the high system stochasticity. To speed up convergence, we re-ran Algorithm 2 using $\rho = 0.005$. Table IV displays the results, in which all runs converged with $n = 3$.

TABLE IV. Results of the multilevel cross-entropy algorithm applied to the yeast polarization model. The column identities match those of Tables I–III. In this example, three steps were required for $[\rho K]$ of the dwSSA trajectories to reach the original rare event threshold.

Step (i)	Trial No.	$\theta_i^{G_{bg}}$	$\hat{\gamma}^{(i)}$
1	1	45	[1.223 0.872 1.066 0.924 1.057 0.764 1.007 1.182]
	2	45	[1.218 0.988 1.060 0.917 1.064 0.760 0.998 1.188]
	3	45	[1.069 0.900 1.056 0.906 1.064 0.756 0.998 1.179]
	4	45	[0.923 0.963 1.077 0.917 1.064 0.755 1.005 1.178]
2	1	49	[0.954 0.533 1.071 0.933 1.099 0.660 1.010 1.272]
	2	49	[1.372 0.865 1.090 0.943 1.128 0.671 0.977 1.258]
	3	49	[1.600 0.824 1.065 0.817 1.093 0.671 1.012 1.250]
	4	49	[0.783 0.754 1.091 0.944 1.114 0.650 1.002 1.234]
3	1	50	[0.444 0.225 1.073 2.101 1.115 0.657 1.000 1.069]
	2	50	[0.583 0.875 1.069 0.950 1.132 0.629 1.043 1.286]
	3	50	[0.378 0.342 1.065 0.909 1.070 0.623 0.933 1.280]
	4	50	[1.951 3.037 1.191 0.991 1.128 0.675 1.065 1.207]

In contrast to the previous three examples, several of the optimal parameter estimates showed high variability across the independent realizations (see $\hat{\gamma}_1^*$, $\hat{\gamma}_2^*$, $\hat{\gamma}_4^*$). To better characterize this variability, we ran an additional 100 independent realizations of Algorithm 2. Figure 5 shows the results, in which we see that $\hat{\gamma}_1^*$ and $\hat{\gamma}_2^*$ display extremely high variability, ranging from values near 0 to >4 . The remaining six parameters exhibit relatively consistent estimates, with $\{\hat{\gamma}_3^*, \hat{\gamma}_5^*, \hat{\gamma}_8^*\}$ each > 1 , $\hat{\gamma}_6^* < 1$, and $\{\hat{\gamma}_4^*, \hat{\gamma}_7^*\} \sim 1$. It is worth noting that the two parameters whose values are consistently farthest from 1 (γ_6 and γ_8) correspond to two reactions which do not directly influence the species in the rare event description (G_{bg}). Thus, our method discovers a weighting strategy

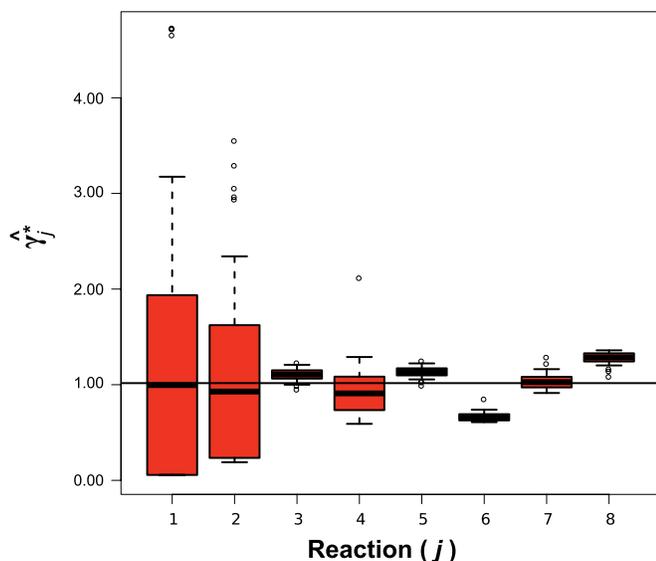


FIG. 5. Variability of optimal parameter estimates for the yeast polarization model. Boxplots summarize parameter estimates for the eight model reactions from 104 independent realizations of Algorithm 2 ($K = 10^5$). Unlike Figs. 1, 2, and 4, box whiskers extend to the mean ± 2 standard deviations. More extreme values are displayed as individual points. Parameters $\gamma_3 - \gamma_8$ exhibit relatively consistent estimates, whereas γ_1 and γ_2 vary widely across the different realizations.

that is not obvious, but is nonetheless critical for delivering a low variance probability estimate.

We hypothesize that the reason for the high variability of the first two parameter estimates is the lack of sensitivity of \hat{p}_{dwSSA} to their values. If true, this hypothesis suggests that we could assign any combination of values within $[0, 4]$ to γ_1 and γ_2 with little effect on the resulting probability estimate. We performed this experiment by simulating the dwSSA ($K = 10^7$) four times each for 16 total combinations of values for the two parameters. We also performed an identical experiment perturbing γ_6 and γ_8 , whose original parameter estimates were very consistent. Figure 6 displays the results. As expected, perturbing γ_1 and γ_2 had almost no effect on the value of \hat{p}_{dwSSA} , whereas any perturbation of γ_6 and γ_8 away from their optimal estimates consistently had a negative effect on the precision of \hat{p}_{dwSSA} . These observations suggest that our multilevel cross-entropy approach coupled with the dwSSA can identify optimal parameter estimates and simultaneously provide insight into the sensitivity of the rare event probability estimate to each parameter.

Using the optimal parameter estimates from Table IV, we computed the mean for each parameter to obtain $\hat{\gamma}^* = [0.839 \ 1.120 \ 1.099 \ 1.238 \ 1.111 \ 0.646 \ 1.010 \ 1.211]$. We used these estimates to run four independent realizations of the dwSSA for varying K , yielding the results shown in Fig. 7. As before, the values of \hat{p}_{dwSSA} converge with increasing K , although we do not have an analytical form for the exact probability in this example. Figure 7 does not display results from the wSSA, as the method outlined by Gillespie *et al.*⁶ would certainly be intractable without prior system insight for an eight reaction model. Finally, we computed the mean probability estimate and uncertainty for the dwSSA with four independent ensembles of $K = 10^7$, yielding:

$$\begin{aligned} \hat{p}_{dwSSA}([50 \ 2 \ 0 \ 50 \ 0 \ 0 \ 0], 50; 20) \\ = 1.13 \times 10^{-6} \pm 0.03 \times 10^{-6}. \end{aligned} \quad (23)$$

We note that the uncertainty associated with the dwSSA estimate Eq. (23) is over five times smaller than the uncertainty

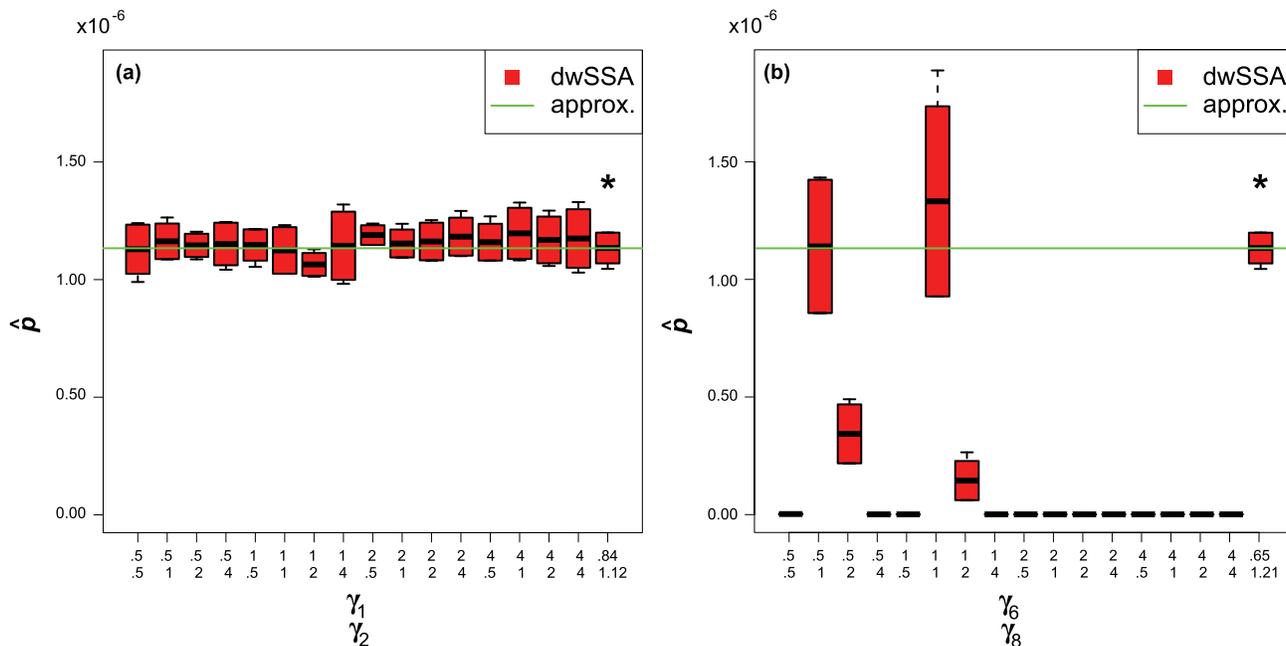


FIG. 6. (a) Sensitivity of rare event probability estimate to dwSSA parameter values γ_1 and γ_2 for the yeast polarization model. We tested 16 combinations of γ_1 and γ_2 values spanning the ranges observed in Fig. 5 along with the optimal combination discussed in Fig. 7 (boxplot marked with *). Boxplots summarize results of four independent dwSSA ensembles ($K = 10^7$) using the indicated values of γ_1 and γ_2 along with the optimal values of $\gamma_3 - \gamma_8$ detailed in Fig. 7. The green line denotes the estimate of p achieved with 10^8 realizations of the dwSSA. Varying γ_1 and γ_2 has little effect on the resulting values of \hat{p}_{dwSSA} , presumably due to the insensitivity of the rare event to these two parameters. (b) A similar plot in which we modify values of γ_6 and γ_8 in the same manner as in (a). This time, varying of parameters leads to a profound increase in \hat{p}_{dwSSA} variability, with most parameter combinations yielding zero observations of the rare event. These results provide an explanation for the differences in parameter estimate precision seen in Fig. 5: The multilevel cross-entropy method provides optimal parameter estimates with a precision commensurate to the sensitivity of \hat{p}_{dwSSA} to their values.

of the original SSA estimate. Using the formula described in Ref. 6, we would expect an SSA estimate of similar accuracy to Eq. (23) to require over 10^9 trajectories, which corresponds to a dwSSA computational gain of > 25 . In terms of running time, on a desktop computer with a single 3 GHz processor, use of the dwSSA reduces an SSA run of ~ 14 days to ~ 13 h.

To apply our method to a rare event probability whose estimation is substantially beyond the capabilities of the SSA or wSSA, we modified the original problem as follows: $p(\mathbf{x}_0, \theta^{G_{bg}}; t) \equiv p([50 \ 2 \ 0 \ 50 \ 0 \ 0 \ 0], 40; 5)$. As before, we ran four independent realizations of Algorithm 2 ($K = 10^5$) with $\rho = 0.01$. Table V displays the results, where all runs converged with $n = 4$. Using these optimal parameter esti-

TABLE V. Results of the multilevel cross-entropy algorithm applied to the modified yeast polarization model. The column identities match those of Tables I-IV. In this example, four steps were required for $\lceil \rho K \rceil$ of the dwSSA trajectories to reach the original rare event threshold.

Step (i)	Trial No.	$\theta_i^{G_{bg}}$	$\hat{\gamma}^{(i)}$
1	1	24	[1.022 0.885 1.298 0.779 1.263 0.562 1.010 1.353]
	2	24	[1.158 0.904 1.285 0.744 1.255 0.565 0.979 1.365]
	3	24	[1.063 1.013 1.286 0.836 1.258 0.568 0.984 1.361]
	4	24	[1.083 0.988 1.297 0.798 1.257 0.560 1.025 1.354]
2	1	32	[0.675 0.983 1.567 0.657 1.450 0.384 0.955 1.676]
	2	32	[0.436 0.692 1.560 0.666 1.445 0.400 1.021 1.673]
	3	32	[1.130 0.454 1.529 0.653 1.431 0.394 0.993 1.693]
	4	32	[1.409 1.467 1.530 0.594 1.427 0.427 1.024 1.700]
3	1	37	[0.298 0.640 1.677 0.292 1.591 0.337 0.897 1.877]
	2	37	[0.874 1.129 1.772 0.694 1.589 0.305 0.862 1.928]
	3	37	[1.140 0.666 1.722 0.648 1.601 0.303 1.016 1.930]
	4	37	[1.308 0.661 1.665 0.427 1.537 0.260 1.083 1.948]
4	1	40	[0.059 0.299 1.636 0.824 1.767 0.230 0.908 1.928]
	2	40	[1.123 4.309 1.774 0.528 1.630 0.254 1.154 2.129]
	3	40	[1.154 1.364 1.704 0.326 1.629 0.245 0.757 2.097]
	4	40	[0.746 0.846 1.775 0.571 1.703 0.259 1.083 2.110]

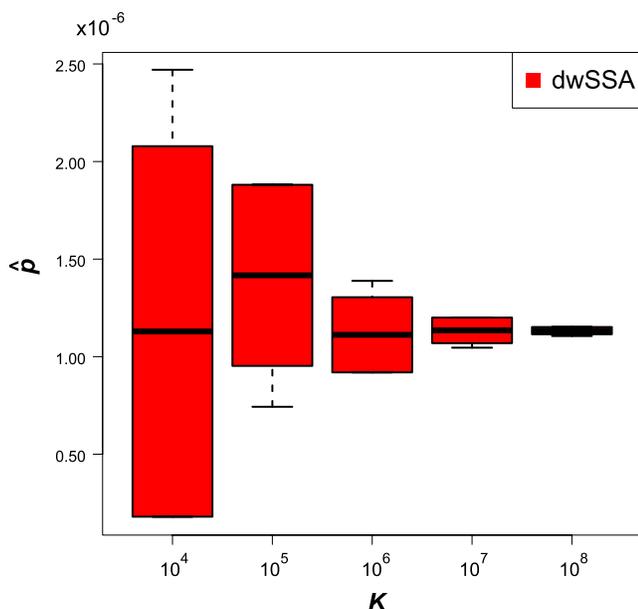


FIG. 7. Convergence plot of rare event probability estimate (\hat{p}) vs simulation ensemble size (K) for the yeast polarization model. Boxplots are constructed as in Figs. 1, 2, and 4, summarizing results of four independent dwSSA simulation ensembles for each value of K . We parameterized the dwSSA with the mean of four realizations of Algorithm 2, yielding $\hat{\gamma}^* = [0.839 \ 1.120 \ 1.099 \ 1.238 \ 1.111 \ 0.646 \ 1.010 \ 1.211]$. As K increases, the dwSSA estimates approach $\hat{p}_{\text{dwSSA}} = 1.131 \times 10^{-6}$; the true rare event probability is unknown. Results for the wSSA are not shown, as the computational cost for determining its optimal parameter values is prohibitive.

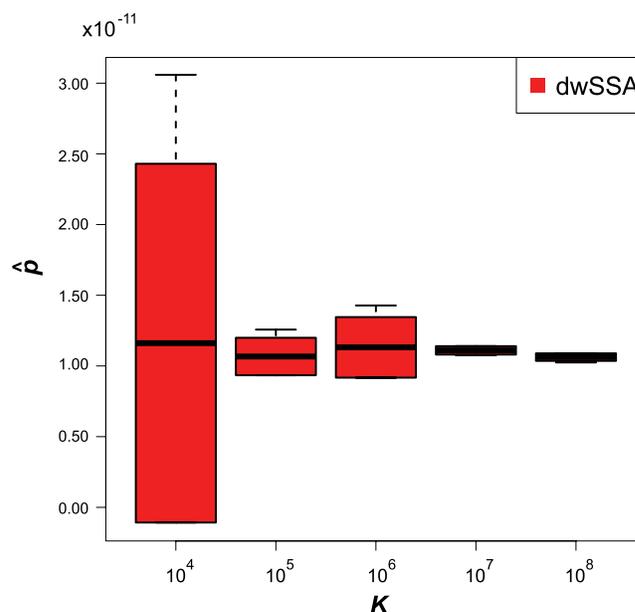


FIG. 8. Convergence plot of rare event probability estimate (\hat{p}) vs simulation ensemble size (K) for the modified yeast polarization model. Boxplots are constructed as in Figs. 1, 2, 4, and 7, summarizing results of four independent dwSSA simulation ensembles for each value of K . We parameterized the dwSSA with the mean of four realizations of Algorithm 2, yielding $\hat{\gamma}^* = [0.771 \ 1.705 \ 1.722 \ 0.562 \ 1.682 \ 0.247 \ 0.975 \ 2.066]$. As K increases, the dwSSA estimates approach $\hat{p}_{\text{dwSSA}} = 1.059 \times 10^{-11}$; the true rare event probability is unknown. Results for the wSSA are not shown, as the computational cost for determining its optimal parameter values is prohibitive.

mates, we computed the mean for each parameter to obtain $\hat{\gamma}^* = [0.771 \ 1.705 \ 1.722 \ 0.562 \ 1.682 \ 0.247 \ 0.975 \ 2.066]$. When compared to the original optimal parameter estimates, aside from parameters whose optimal values tightly spanned 1 (γ_4 and γ_7), the optimal parameter values for the modified system exhibit deviations from 1 that are identical in direction but larger in magnitude. As we were attempting to use our multilevel cross-entropy approach to estimate a smaller rare event probability (i.e., one whose estimation would require greater system biasing), this trend was expected.

We used the above mean optimal parameter estimates to run four independent dwSSA realizations for varying K , yielding the results shown in Fig. 8. Again, the values of \hat{p}_{dwSSA} converge with increasing K . Upon computing the mean probability and uncertainty for the dwSSA with four independent ensembles of $K = 10^7$, we obtain:

$$\begin{aligned} \hat{p}_{\text{dwSSA}}([50 \ 2 \ 0 \ 50 \ 0 \ 0 \ 0], 40; 5) \\ = 1.11 \times 10^{-11} \pm 0.04 \times 10^{-11}. \end{aligned} \quad (24)$$

Although use of the SSA to analyze such a rare event is infeasible, we can use the same technique as above to calculate the computational gain conferred by using the cross-entropy approach coupled with the dwSSA. The result suggests that an SSA estimate of similar accuracy to (24) would require over 10^{13} trajectories, corresponding to a dwSSA computational gain of $> 1.7 \times 10^6$. For this modified example, use of the dwSSA reduces a projected SSA run time of ~ 600 yr to ~ 3 h.

VI. CONCLUSIONS

This paper describes two main research contributions. First, it presents a novel modification of the wSSA—the dwSSA—that weights both reaction selection and time to the next reaction. Second, it shows how an information-theoretic technique, the cross-entropy method, can be used together with the dwSSA to provide an automated mechanism for learning low variance reaction biasing parameters. Importantly, the mathematical properties of the dwSSA combined with the cross-entropy method enable an analytical form for optimal parameter estimates which would not be possible with the wSSA. This attribute of the dwSSA attaches substantial value to its novelty, as the practical power of the cross-entropy method is wholly unavailable to users of the wSSA. The multilevel cross-entropy method requires a single user-defined parameter, ρ , which determines the proportion of trajectories in which the rare event must occur before computing parameter estimates. We have found that results are not very sensitive to the choice of ρ , and in some cases a smaller value for ρ can accelerate convergence of the algorithm.

To demonstrate the performance of our method, we tested it on four different biochemical systems ranging in size from one to eight reactions. In each example, the multilevel cross-entropy method coupled with the dwSSA provided optimal parameter estimates at a fraction of the simulation cost required by the wSSA. These parameter estimates, when used

in the dwSSA, delivered rare event probability estimates of equivalent accuracy to existing methods. Each example tested provided unique insight into the properties of our proposed method. Results from the pure birth process illustrated the usefulness of the dwSSA on a class of rare events in which the number of reaction firings needed to satisfy the rare event deviate considerably from the average. For this class of problems, the wSSA is totally ineffective, as it is effectively identical to the SSA. Analysis of the birth-death process demonstrated that parameter estimates minimizing cross entropy closely correspond to parameters that minimize estimator variance. Previous results using the wSSA to study the enzymatic futile cycle showed that considerable insight was required to properly choose biasing parameters. In contrast, the multilevel cross-entropy method coupled with the dwSSA automatically selects low variance parameters that preserve constraints present in wSSA optimal parameters. Finally, successful characterization of the yeast polarization example demonstrated the utility of our method applied to a realistic biochemical system, where the try-and-test procedure required by the wSSA⁶ would be computationally infeasible. In addition, results from this example illustrated how the cross-entropy method can provide insight into the sensitivity of the rare event probability estimate to each biasing parameter.

As researchers continue to model larger and more comprehensive systems, methods requiring exhaustive parameter searches to estimate rare event probabilities quickly become inadequate. Given that the wSSA is the current state of the art for importance sampling with the SSA, this represents a major limitation. In response to this limitation, the contributions made in this work provide an automated approach whose complexity scales linearly with system size, enabling efficient estimation of rare event probabilities for large systems that could not previously be interrogated. Furthermore, by incorporating information-theoretic principles, our approach provides a framework for the development of more sophisticated

influencing schemes that should further improve estimation accuracy. Future work will focus on this task.

ACKNOWLEDGMENTS

The authors acknowledge the following financial support: B.J.D.J. was supported by Army Grant No. W911NF-09-D0001. M.R. and L.R.P. were supported by Grant No. R01EB007511 from the National Institute of Biomedical Imaging and Bioengineering, DOE Grant No. DEFG02-04ER25621, and the Institute for Collaborative Biotechnologies through Grant No. DFR3A-8-447850-23002 from the U.S. Army Research Office. D.T.G. was supported by the California Institute of Technology through Consulting Agreement No. 102-1080890 pursuant to Grant No. R01GM078992 from the National Institute of General Medical Sciences and through Contract No. 82-1083250 pursuant to Grant No. R01EB007511 from the National Institute of Biomedical Imaging and Bioengineering, and also from the University of California at Santa Barbara under Consulting Agreement No. 054281A20 pursuant to funding from the National Institutes of Health.

¹H. Kitano, *Science* **295**, 1662 (2002).

²M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, *Nat. Rev. Genet.* **6**, 451 (2005).

³M. Csete and J. Doyle, *Trends Biotechnol.* **22**, 446 (2004).

⁴D. Gillespie *J. Phys. Chem.* **81**, 2340 (1977).

⁵H. Kuwahara and I. Mura, *J. Chem. Phys.* **129**, 165101 (2008).

⁶D. T. Gillespie, M. Roh, and L. R. Petzold, *J. Chem. Phys.* **130**, 174103 (2009).

⁷G. Rubino and B. Tuffin, *Rare Event Simulation Using Monte Carlo Methods* (Wiley, Chichester, UK, 2009).

⁸R. Y. Rubinstein and D. P. Kroese, *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning* (Springer, New York, 2004).

⁹R. Y. Rubinstein, *Eur. J. Operational Res.* **99**, 89 (1997).

¹⁰M. Samoilov, S. Plyasunov, and A. P. Arkin, *Proc. Natl. Acad. Sci. USA* **102**, 2310 (2005).

¹¹B. Drawert, M. J. Lawson, L. Petzold, and M. Khammash, *J. Chem. Phys.* **132**, 074101 (2010).