

A cutoff phenomenon in accelerated stochastic simulations of chemical kinetics via flow averaging (FLAVOR-SSA)

Basil Bayati,¹ Houman Owhadi,^{2,a)} and Petros Koumoutsakos^{1,b)}

¹Computational Science and Engineering Laboratory, ETH Zurich, CH-8092, Switzerland

²Applied and Computational Mathematics, California Institute of Technology, Pasadena, California 91125, USA

(Received 20 August 2010; accepted 30 October 2010; published online 30 December 2010)

We present a simple algorithm for the simulation of stiff, discrete-space, continuous-time Markov processes. The algorithm is based on the concept of *flow averaging* for the integration of stiff ordinary and stochastic differential equations and ultimately leads to a straightforward variation of the well-known *stochastic simulation algorithm* (SSA). The speedup that can be achieved by the present algorithm [flow averaging integrator SSA (FLAVOR-SSA)] over the classical SSA comes naturally at the expense of its accuracy. The error of the proposed method exhibits a cutoff phenomenon as a function of its speed-up, allowing for optimal tuning. Two numerical examples from chemical kinetics are provided to illustrate the efficiency of the method. © 2010 American Institute of Physics. [doi:10.1063/1.3518419]

I. INTRODUCTION

Continuous-time Markov processes are ubiquitous in the modeling of the dynamics of complex systems.^{1,2} An important example is stochastic chemical kinetics which describes the time evolution of chemically reacting systems by taking into account the fact that molecules come in whole numbers and exhibit randomness in their dynamical behavior.^{3,4}

The transition probabilities of such processes obey the *Chapman–Kolmogorov* equation, which in turn is equivalent to the so-called *master equation*.^{5,6} The number of variables in this master equation is large for all but the simplest systems, so analytical or direct numerical integration methods are oftentimes impractical. Alternatively, Monte Carlo samples of the stochastic process can be numerically generated, via *stochastic simulation algorithms* (SSAs),^{7,8} so that the only error introduced is the sampling error.

SSA, however, is a rather inefficient method when the system of chemical reactions is stiff, involving a large number of disparate temporal scales either due to disparate reaction rates or due to large variations in the molecular populations. In recent years, several algorithms have been proposed to simulate such stiff processes. Some of the methods that have been developed rely upon the assumption that the fast processes are in so-called *quasi-equilibrium*.^{9,10} R-leaping,¹¹ an accelerated stochastic simulation algorithm, efficiently handles systems with disparate rates by sorting the reaction events and by acting first on the reactions with large propensities. Peles *et al.*¹² have formulated an efficient algorithm for the direct solution of the master equation with multiple time-scales. In a separate line of work, Weinan *et al.*^{13,14} have used drift averaging techniques¹⁵ to accelerate stochastic simulations, while Gear and Kevrekidis¹⁶ and Kevrekidis *et al.*¹⁷ have employed projective integration methods for

stiff differential equations. The method presented herein relies upon a flow averaging principle, *flow averaging integrators* (called FLAVORS), that has been introduced for the integration of stiff ordinary and stochastic differential equations.¹⁸ The extension of FLAVORS to stochastic systems results in a simple and non-intrusive algorithm for the effective simulation of a stiff set of reactions. The objective of this article is to provide a description of a novel method, based on the concept of flow averaging, for the simulation of stiff discrete-space, continuous-time Markov processes. It shall be shown that ultimately the algorithm amounts to a simple rescaling of the fast propensities at every other time-step. Like the original FLAVOR algorithm, the algorithm presented here (FLAVOR-SSA) is non-intrusive, in that it can be coupled to any stochastic simulation algorithm including accelerated algorithms.³

II. STOCHASTIC SIMULATION ALGORITHMS

SSAs (Refs. 8 and 19) are a class of Monte Carlo methods for the simulation of discrete-space, continuous-time Markov processes. The discrete-state of the system is denoted by $\mathbf{x} \in \mathbb{N}^N$, where N is the number of species in the system. SSA deals with a realization of the time-dependent stochastic processes, namely a trajectory $\mathbf{X}(t) \in \mathbb{N}^N$. The process is simulated over time by the following update scheme:

0. Initialize the time $t = t_0$ and the system's state $\mathbf{X} = \mathbf{X}_0$.
1. With the system in state \mathbf{x} at time t , evaluate all the reaction rates (propensities) $a_j(\mathbf{X})$ and their sum $a_0(\mathbf{X}) := \sum_{j=1}^M a_j(\mathbf{X})$.
2. Generate values for τ and j where τ is an exponential random variable with parameter $a_0(\mathbf{X})$ and j is a discrete random variable on $\{1, \dots, N\}$ with $\mathbb{P}[j = k] = a_k(\mathbf{X})/a_0(\mathbf{X})$.

^{a)}Electronic mail: owhadi@caltech.edu.

^{b)}Electronic mail: petros@ethz.ch.

3. Execute the next reaction by replacing $t \rightarrow t + \tau$ and $\mathbf{X} \rightarrow \mathbf{X} + \mathbf{v}_j$ where \mathbf{v}_j is the stoichiometric vector that denotes the change induced by reaction j .
4. Record (\mathbf{X}, t) . Return to step 1, or end.

The random variable $\mathbf{X}(t)$ is a sample from the probability distribution $\mathbf{P}(\mathbf{X}_0, t)$ (the solution of the master equation) of the chemical species at time t . Hence, the empirical distribution associated with m independent samples of $\mathbf{X}(t)$ approximates $\mathbf{P}(\mathbf{X}_0, t)$ with an accuracy of the order of $m^{-(1/2)}$. In the classical formulation of the SSA, the reaction rates a_k are defined by

$$a_k(\mathbf{X}) := c_k \Omega \prod_{j=1}^N \left\{ \frac{X_j(X_j - 1) \dots (X_j - r_j + 1)}{\Omega^{r_j}} \right\}, \quad (1)$$

where c_k is the reaction rate, Ω is the system volume, X_j is the number of molecules of species j , and r_j (a constant) is the number of reactants of the species j . It should be noted that although Eq. (1) holds for arbitrary order reactions, reactions only have a physical meaning for order 0, 1, and 2.

Write $a_k(t) := a_k(\mathbf{X}(t))$. Under the assumption that there is a clear separation of fast and slow processes, the total propensity can be decomposed as follows:

$$a_0(t) = \frac{1}{\epsilon} \sum_i \tilde{a}_i^{(\text{fast})}(t) + \sum_j a_j^{(\text{slow})}(t), \quad (2)$$

where $a_i^{(\text{fast})}(t) = (1/\epsilon) \tilde{a}_i^{(\text{fast})}(t)$ and $\epsilon \ll 1$ is the stiffness parameter that is either specified by the problem (a fast reaction rate) or emerges during the evolution of the system (disparities in the propensities). It should be noted that this dimensionless parameter has been introduced only for the clarity of the presentation and the analysis of the method and in practice its specific value is not strictly needed. The fast processes evolve on a time-scale of $\mathcal{O}(\epsilon)$ and the slow processes are on a time-scale of $\mathcal{O}(1)$. The inefficiency of the SSA is incident to the fact that $a_0 \approx \mathcal{O}(\tilde{A}^{(\text{fast})} \epsilon^{-1})$, where $\tilde{A}^{(\text{fast})} = \sum_i \tilde{a}_i^{(\text{fast})}(t)$ has units [1/time] and is $\mathcal{O}(1)$ and consequently $\mathbb{E}[\tau] \approx \mathcal{O}(\epsilon)$.

Assume that the propensities are in ascending order and indexed by k . Then, ϵ can be determined by

$$\epsilon = \min_k \left\{ \frac{a_{k-1}}{a_k} \right\}. \quad (3)$$

Partitioning of the propensities can be performed by indexing the fast propensities by $i \geq k'$ and the slow propensities $j < k'$, where k' is the arg min (argument of the minimum) of Eq. (3). We note that this partitioning can also be performed hierarchically for systems with multiple time-scales, although in this article such systems are not investigated. Moreover Eq. (3) assumes that the time-scales are clearly separated. For example, if $a_k = 2a_{k-1}$ for $k = 2, \dots, K$, then although there is a large disparity between a_1 and a_K , there is not a clear separation between fast and slow processes.

Letting $\Phi^{(1/\epsilon)}$ denote one iteration of the SSA, the numerical method can be written as

$$(\mathbf{X}_n, t_n) = (\Phi^{(1/\epsilon)})^n (\mathbf{X}_0, t_0), \quad (4)$$

which is equivalent to performing n iterations of SSA.

III. FLAVORIZED-SSA

Tao *et al.*¹⁸ presented a class of numerical methods, called FLAVORS, for the integration of stiff ordinary and stochastic differential equations. The method is based on averaging the instantaneous flow of a dynamical system with hidden fast and slow variables. The advantage of the method is that the computational cost is determined by the slow processes instead of the fast ones. In order to briefly outline the FLAVOR method we consider the stiff ordinary differential equation

$$\dot{\mathbf{u}} = \mathbf{G}(\mathbf{u}) + \frac{1}{\epsilon} \mathbf{F}(\mathbf{u}). \quad (5)$$

Assume that there exists a possibly unknown and nonlinear diffeomorphism η such that $(x, y) = \eta(\mathbf{u})$ is the solution of the following stiff set of ordinary differential equations:

$$\dot{x} = g(x, y), \quad (6)$$

$$\dot{y} = \frac{1}{\epsilon} f(x, y). \quad (7)$$

Let $\phi^{(1/\epsilon)}$ be the flow of a first order accurate legacy integrator for Eq. (5), i.e., $\bar{\mathbf{u}}_{t+h} = (\phi_h^{(1/\epsilon)})(\bar{\mathbf{u}}_t)$, where $\bar{\mathbf{u}}$ is a numerical approximation of the solution of Eq. (5). The FLAVOR method works by averaging the flow of Eq. (5) [instead of Eqs. (6) and (7)] by splitting and resynchronization. By splitting we refer to a composition $(\phi_{\delta-\tau}^{(0)} \circ \phi_{\tau}^{(1/\epsilon)})$, where $\phi_{\delta-\tau}^{(0)}$ denotes the flow with the stiffness parameter $1/\epsilon = 0$ and $\delta \gg \tau$ are the time-steps. By resynchronization we refer to the distinct time-steps δ and τ , the effects of which are to advance the internal clock of fast variables by τ every time-step of length δ . The averaging applied M times means that $(\phi_{\delta-\tau}^{(0)} \circ \phi_{\tau}^{(1/\epsilon)})^M \approx (\phi_{\tau}^{(1/\epsilon)})^L$, where $L\tau = M\delta$. The variable M denotes the number of samples used to average the flow. The approximate scheme is therefore $\bar{\mathbf{u}}_t = (\phi_{\delta-\tau}^{(0)} \circ \phi_{\tau}^{(1/\epsilon)})^M(\bar{\mathbf{u}}_0)$, where $M\delta = t$. It is shown in Ref. 18 that the proposed flow averaging is accurate (in a strong sense with respect to (possibly hidden) slow variables and in the sense of measures with respect to (possibly hidden) fast variables) provided that fast variables are locally ergodic and $(\tau/\epsilon)^2 \ll \delta \ll \tau/\epsilon$. The condition $\delta \ll \tau/\epsilon$ ensures that the slow dynamic has been averaged with respect to that of the fast variables. The condition $(\tau/\epsilon)^2 \ll \delta$ ensures that the error caused by fast variables on slow ones (when η is nonlinear) remains small. After optimizing on τ the accuracy of the proposed method is, in the worst case scenario, $\delta^{1/3}$ while its cost is $1/\delta$ (and in particular, both are independent from ϵ). The FLAVOR method is not completely equivalent to simulating Eq. (5) with a larger ϵ' : indeed, the accuracy of such simulation (with a time-step proportional to ϵ') would be $e^{CT/\epsilon'} \epsilon'$ (the constant $e^{CT/\epsilon'}$ typically results from Gronwall's lemma and the fact the total simulation of Eq. (6) has to be rescaled by $1/\epsilon'$ (the error is multiplicative at a rate $1 + C/\epsilon'$)) which, in particular, cannot be made arbitrarily small with respect to ϵ' .

Analogously to the FLAVOR algorithm, the FLAVOR ized-SSA method is defined as the composition of two SSA steps:

$$(\mathbf{X}_n, t_n) = (\Phi^{(\xi/\epsilon)} \circ \Phi^{(1/\epsilon)})(\mathbf{X}_{n-1}, t_{n-1}), \quad (8)$$

where $\xi \in [0, 1]$. The SSA step $\Phi^{(\xi/\epsilon)}$ advances the simulation clock by a time interval of size δ , which is an exponential random variable with parameter

$$\hat{a}_0(t, \xi) := \frac{\xi}{\epsilon} \sum_i \tilde{a}_i^{(\text{fast})}(t) + \sum_j a_j^{(\text{slow})}(t). \quad (9)$$

It also selects a fast reaction i with probability $(\xi/\epsilon)(\tilde{a}_i^{(\text{fast})}(t)/a_0(t))$ or slow reaction j with probability $a_j^{(\text{slow})}(t)/a_0(t)$. This is simply equivalent to identifying fast reactions and multiplying fast propensities by ξ at every other step. If $\xi = 1$, the algorithm reduces to the standard formulation of the SSA. The only difference between the *FLAVOR*ized-SSA and the SSA is that the fast (stiff) propensities are rescaled by ξ at every other iteration. Intuitively, this amounts to slowing down the fast processes to be on the same time-scale as the slow processes, thus permitting a larger time-step to be taken. For $\xi/\epsilon \gg 1$, fast variables have the time to converge in law towards their equilibrium distribution before any significant change on slow variables. Conversely, if ξ/ϵ is of $\mathcal{O}(1)$, then fast variables do not have the time to converge in law towards their equilibrium distribution before significant changes on the slow variables. Hence, speed-up and error are both increasing as a function of $1/\xi$. However, although speed-up is proportional to $1/\xi$, we will show that the error exhibits a sharp transition with respect to the value of ξ , i.e., a cutoff phenomenon.

IV. CUTOFF PHENOMENON

Stiff systems, and in particular fast processes, force stochastic simulation algorithms to take prohibitively small time-steps as discussed above. These fast processes are on several occasions not as important as the slow processes in determining the overall dynamics of the system. Hence, if one were to artificially slow down the fast reactions by a factor of $0 \leq \xi < 1$ then the stiffness would be alleviated and the problem of small time-steps would be mended. Naturally, an error would be introduced since one is not simulating the original system, but rather a rescaled version. It will be shown below that the error introduced by this artificial slowing down of fast processes is not linear with respect to ξ , but rather changes significantly when ξ approaches ϵ . In other words, when $\xi > \epsilon$ virtually no error is made, yet when $\xi < \epsilon$ a large error is made. This sharp transition of the error around a certain value, in this case $\xi \approx \epsilon$, can be likened to a so-called cutoff phenomenon as studied by Diaconis.²⁰ Diaconis proved that fewer than six riffle shuffles of a deck of playing cards is not sufficient to bring the deck to its stationary distribution (sc. a deck with a random distribution of cards), yet—rather surprisingly—seven shuffles suffice. In the case of chemical kinetics, the fast processes (i.e., reactions) can be artificially slowed down without substantially modifying the dynamical properties of the system, the result of which is a decrease in the computational time.

Diaconis²⁰ showed that with fewer than six riffle shuffles, a deck of playing cards remains strongly correlated to its initial configuration whereas seven shuffles suffice to bring it close to the the uniform distribution on the permutation group.

This sharp transition in mixing properties of Markov chains has been called cutoff phenomenon. Sharp convergence of the present numerical method with respect to ξ is investigated next. Since the value of ξ dictates both the error and speed-up of the simulation, we inquire into its optimal value. The proof of convergence of *FLAVOR*-SSA with error bounds is similar to those given in Ref. 18 and will not be reproduced here (we also refer to theorem 2.1 of Ref. 15).

The value of ξ controls the distance, in distribution, between fast variables and their local (with frozen slow variables) invariant distribution. In presence of a spectral gap λ , this distance decays like $e^{-\xi\lambda t/\epsilon}$ (see, e.g., Ref. 21). This is why the cutoff is observed when the value of ξ is slightly larger than ϵ .

V. NUMERICAL EXAMPLE 1: EMERGENCE OF TIME-SCALES

The following chemical system, which was investigated in Ref. 13, exhibits a disparity in time-scales at $t \approx 1$:



This system does not have an explicit stiffness parameter ϵ , but rather the stiffness arises owing to the evolution of the reaction set. Moreover, this chemical system is indicative of real-world problems in which the value of ϵ is not known *a priori*. The reversible reaction shown in Eq. (12) at time $t \approx 1$ has propensity values of roughly 10^3 times larger than the propensities of the other reactions. The value of ϵ varies in the range of 10^{-3} and 10^{-4} when $t > 1$. The corresponding propensities and stoichiometric vectors are

$$a_1^{(\text{slow})} = X_1, \quad \mathbf{v}_1 = (-1, +1, 0, 0)^T,$$

$$a_2^{(\text{slow})} = X_2, \quad \mathbf{v}_2 = (+1, -1, 0, 0)^T,$$

$$a_3^{(\text{slow})} = X_1, \quad \mathbf{v}_3 = (-1, 0, +1, 0)^T,$$

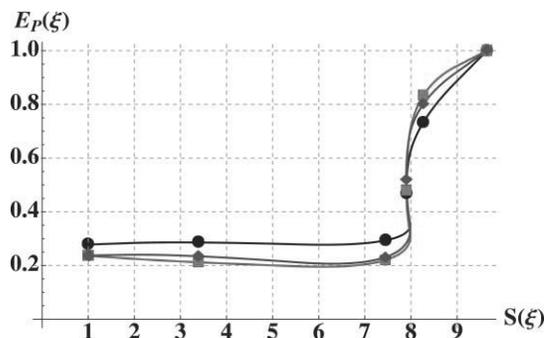


FIG. 1. Optimal value of ξ : error of the probability density function, $E_P(\xi)$, versus speed-up, $S(\xi)$. I^∞ : \circ ; I^1 : \square ; I^2 : \diamond . $\xi = 10^{-j}$, where $j = 0, 1, \dots, 5$ and where $\xi = 1$ is the leftmost point and $\xi = 10^{-5}$ is the rightmost point. The maximum error has been rescaled to 1 and the error at $\xi = 1$ is the sampling error. The markers denote simulated points and the lines are second-order interpolations. The optimal value is $\xi = 10^{-2}$ where $S(\xi) \approx 7$ and the error is negligible.

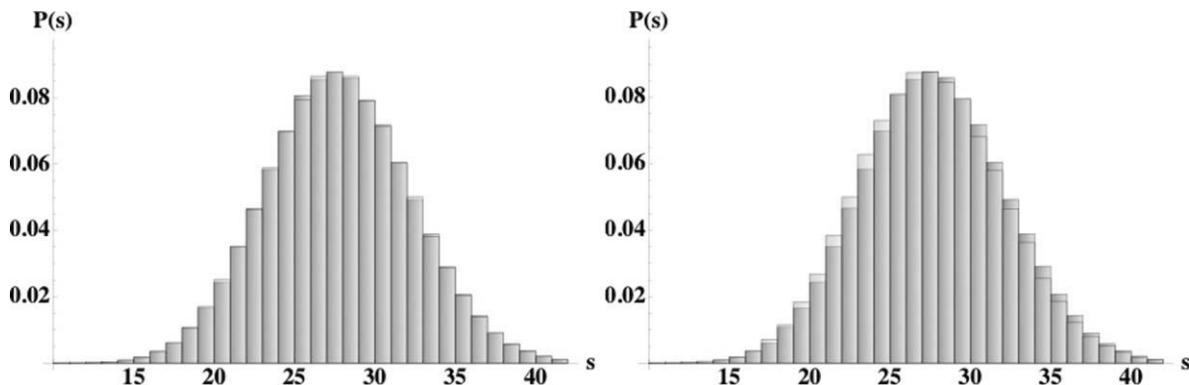


FIG. 2. Comparison of solutions: FLAVORized-SSA [light gray (orange)], SSA [dark gray (black)]. $\mathbf{P}^{\xi=10^{-2}}(t = 4)$ (left) and $\mathbf{P}^{\xi=10^{-5}}(t = 4)$ (right) for the species S_1 . Note that the large error for $\mathbf{P}^{\xi=10^{-5}}(t = 4)$ results from a shift in the distribution.

$$\begin{aligned}
 a_4^{(\text{slow})} &= X_3, & \mathbf{v}_4 &= (+1, 0, -1, 0)^T, \\
 a_5^{(\text{fast})} &= 2X_2(X_2 - 1)X_3, & \mathbf{v}_5 &= (0, -1, -2, +3)^T, \\
 a_6^{(\text{fast})} &= 2X_4(X_4 - 1)(X_4 - 2), \\
 \mathbf{v}_6 &= (0, +1, +2, -3)^T.
 \end{aligned}$$

Simulations were performed with $\mathbf{X}(t) = (X_1(t), X_2(t), X_3(t), X_4(t))^T$ and $\mathbf{X}(0) = (100, 3, 3, 3)^T$ in $t \in [0, 4]$. The FLAVORized-SSA was turned on once there was a disparity in the system, namely at time $t = 1$. A total of $N = 2 \times 10^5$ samples were performed.

A parametric plot of the error of the probability density function versus the speed-up compared to SSA is shown in Fig. 1. The errors are defined as

$$l^p = \|\mathbf{P}^{\xi} - \mathbf{P}\|_p = \left(\sum_k |P_k^{\xi} - P_k|^p \right)^{1/p}, \quad (13)$$

for $p = 1, 2$, where the vector \mathbf{P} denotes the discrete probability density function of an exact stochastic simulation and \mathbf{P}^{ξ} is the density function of the FLAVORized algorithm with parameter ξ . The special case of the l^{∞} norm is defined as $l^{\infty} = \|\mathbf{P}^{\xi} - \mathbf{P}\|_{\infty} = \max_k |P_k^{\xi} - P_k|$. The l^{∞} norm was also used since the l^1 and l^2 norms are not sensitive to an out-

lier, sc. a large error at a single point. The l^1 and l^2 norms are both considered since Diaconis²⁰ considered the so-called *total variation distance* (similar the l^1 norm) and the use of the spectral gap used to show the cutoff phenomenon in this work relies upon the l^2 norm.²¹ Here we show that—up to a normalization—they are all in agreement with respect to the cutoff phenomenon.

It should be noted that the speed-up is relatively modest, since SSA is performed in the interval $0 \leq t < 1$. The value of ξ was varied in the range $\xi = 10^{-j}$, where $j = 0, 1, \dots, 5$. The optimal value for ξ in terms of the error and speed-up is $\xi = 10^{-2}$ for this particular system. This confirms the cutoff analysis above that showed that the error was minimized once $\xi > \epsilon$. Additionally, a comparison of the probability density functions for $\xi = 10^{-2}$ and $\xi = 10^{-5}$ is shown in Fig. 2. The larger error for $\xi = 10^{-5}$ is a result of the shift of the distribution.

VI. NUMERICAL EXAMPLE 2: MINIMAL STIFF CHEMICAL SYSTEM

The following simple stiff chemical system is considered (as in Ref. 13):

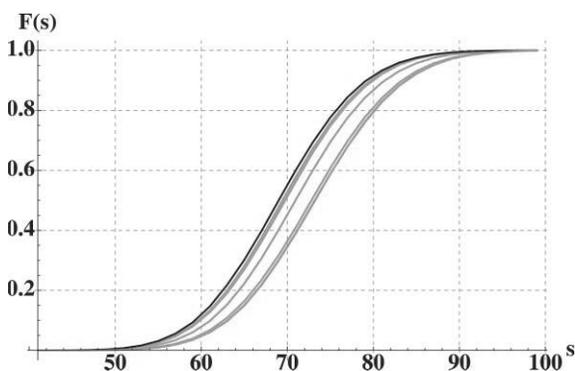


FIG. 3. Mind the shift: shift from left to right of the cumulative distribution functions of $\mathbf{P}^{\xi}(t = 1)$ for the species S_2 as the value of ξ becomes smaller. Light gray (orange) lines denote FLAVORized-SSA, dark gray (black) SSA. $\xi = 10^{-j}$, where $j = 1, \dots, 8$. The shift becomes prominent when $\xi = \epsilon = 10^{-4}$ (solitary orange line) and is a result of the rescaling of time for fast variables.

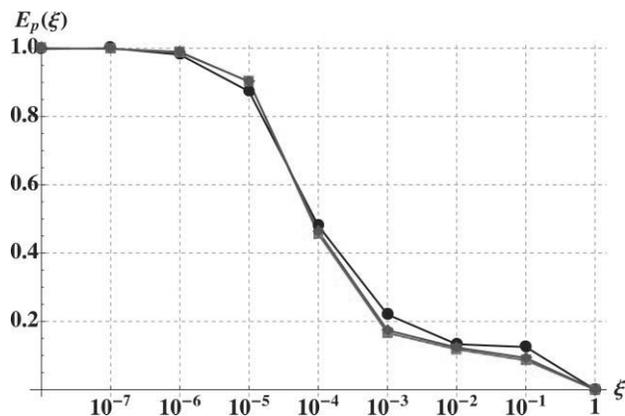


FIG. 4. Error analysis: $\|\mathbf{P}^{\xi}(t = 1) - \mathbf{P}(t = 1)\|$, where l^{∞} : \circ ; l^1 : \square ; l^2 : \diamond ; $\xi = 10^{-j}$, $j = 0, 1, \dots, 8$. The errors have been rescaled to be in the range $[0, 1]$ and the x-axis is logscale.

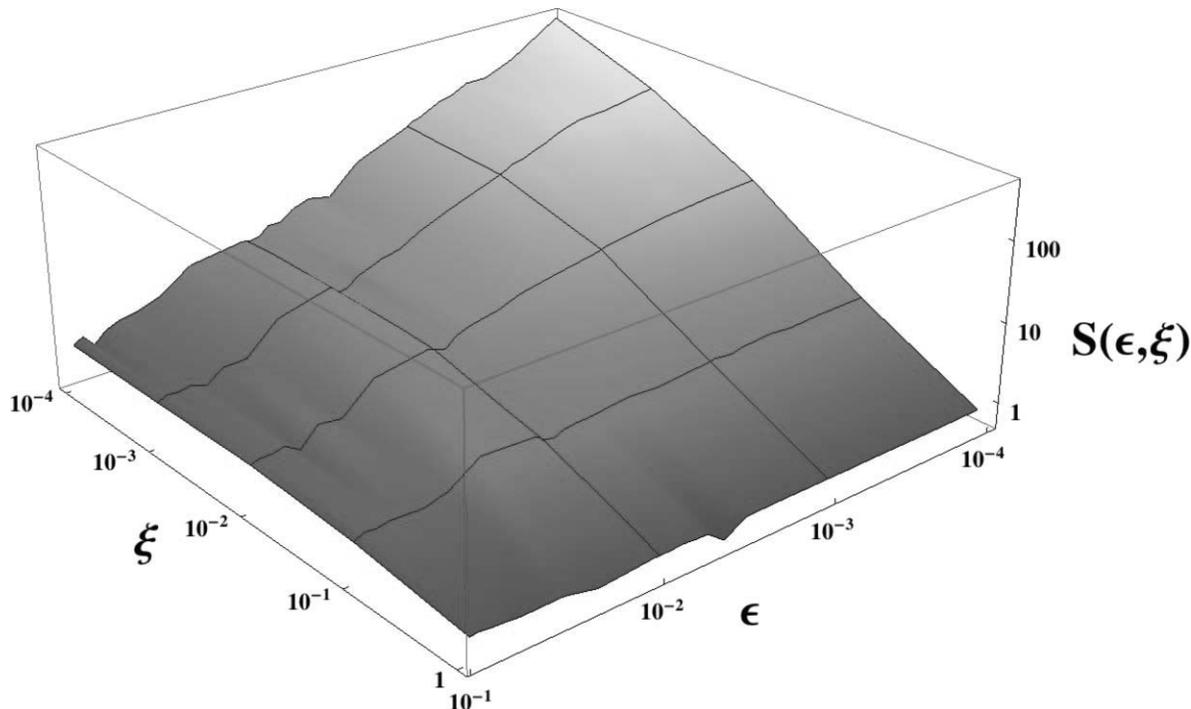


FIG. 5. Computational efficiency: speed-up compared to SSA, $S(\epsilon, \xi)$, with respect to the scaling parameter, ξ , and the stiffness parameter, ϵ . All axes are logscale.

where $\epsilon \ll 1$. The propensities and stoichiometric vectors for this system are:

$$a_1^{(\text{fast})} = \frac{1}{\epsilon} X_1, \quad \nu_1 = (-1, +1)^T, \quad (15)$$

$$a_2^{(\text{fast})} = \frac{1}{\epsilon} X_2, \quad \nu_2 = (+1, -1)^T, \quad (16)$$

$$a_3^{(\text{slow})} = X_2, \quad \nu_3 = (0, -1)^T, \quad (17)$$

$$a_4^{(\text{slow})} = \Omega, \quad \nu_4 = (0, +1)^T. \quad (18)$$

For simplicity, the species S_3 is assumed to be constant. Simulations were performed in $t \in [0, 1]$, where $\mathbf{X}(t) = (X_1(t), X_2(t))^T$, $X_i(t)$ denotes the number of molecules or particles of species S_i at time t , $\mathbf{X}(0) = (50, 50)^T$, $\Omega = 100$, $N = 2 \times 10^5$, and $\epsilon = 10^{-4}$. The value of ξ was changed in order to study the convergence of the simulation. Figure 3 shows the cumulative distributions of $\mathbf{P}^\xi(t=1)$ for the species S_2 . A shift in distributions is noticeable when the value of ξ becomes smaller than ϵ . The error analysis in Fig. 4 shows a clear cutoff around $\xi = \epsilon = 10^{-4}$. The errors have been rescaled to be in the range $[0, 1]$.

The value of ϵ was then varied in order to determine the speed-up of the numerical method compared to the SSA. Figure 5 shows the speed-up with respect to ξ and ϵ . Note that to ensure accuracy, the value of ξ should be larger than ϵ . It is therefore possible to obtain a speed-up of more than $10\times$ if ϵ is very small. As expected, the speed-up is largest when $\xi \rightarrow 0$ and $\epsilon \rightarrow 0$.

VII. CONCLUSIONS AND FUTURE WORK

A straightforward numerical method (FLAVOR-SSA) was presented for the simulation of stiff Markov processes. The method can be easily coupled with any existing stochastic simulations algorithms. The error associated with the proposed method has been shown to be characterized by a cut-off phenomenon allowing for the determination of an optimal trade-off between error and speed-up. Future work will involve a thorough comparison of the present method with various other methods that have been developed, e.g., Refs. 10 and 12–14. Moreover, multiple time-scales and adaptive calculations of ϵ will be investigated. In the same vein, multiple ξ variables can be introduced to scale propensities on different time-scales.

ACKNOWLEDGMENTS

B.B. and P.K. wish to thank Eric Mjolsness at the University of California, Irvine for many helpful discussions regarding the present work. Part of the work of H.O. has been supported by NSF Grant CMMI-092600. The authors also thank the anonymous reviewer, whose comments and suggestions greatly improved the intelligibility of this article.

¹J. Honerkamp, *Stochastic Dynamical Systems* (Wiley-VCH, Weinheim, 1994).

²S. Chandrasekhar, *Rev. Mod. Phys.* **15**, 1 (1943).

³D. T. Gillespie, *Annu. Rev. Phys. Chem.* **58**, 35 (2007).

⁴D. T. Gillespie, *Physica A: Stat. and Theor. Phys.* **188**, 404 (1992).

⁵N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (North Holland, Amsterdam, 2007).

⁶C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed. (Springer-Verlag, Berlin, 2009).

- ⁷A. B. Bortz, M. H. Kalos, and J. L. Lebowitz, *J. Comput. Phys.* **17**, 10 (1975).
- ⁸D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).
- ⁹Y. Cao, D. T. Gillespie, and L. R. Petzold, *J. Chem. Phys.* **122**, 014116 (2005).
- ¹⁰Y. Cao, D. Gillespie, and L. Petzold, *J. Comput. Phys.* **206**, 395 (2005).
- ¹¹A. Auger, P. Chatelain, and P. Koumoutsakos, *J. Chem. Phys.* **125**, 084103 (2006).
- ¹²S. Peles, B. Munsky, and M. Khammash, *J. Chem. Phys.* **125**, 204104 (2006).
- ¹³E. Weinan, D. Liu, and E. Vanden-Eijnden, *J. Chem. Phys.* **123**, 194107 (2005).
- ¹⁴E. Weinan, D. Liu, and E. Vanden-Eijnden, *J. Comput. Phys.* **221**, 158 (2007).
- ¹⁵A. V. Skorokhod, *Asymptotic Methods in the Theory of Stochastic Differential Equations*. Translations of Mathematical Monographs, Vol. 78 (American Mathematical Society, Providence, RI, 1989). (Translated from the Russian by H. H. McFaden.)
- ¹⁶C. W. Gear and I. G. Kevrekidis, *SIAM J. Sci. Comput. (USA)* **24**, 1091 (2002).
- ¹⁷I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos, *Commun. Math. Sci.* **1**, 715 (2003).
- ¹⁸M. Tao, H. Owhadi, and J. E. Marsden, *SIAM Multiscale Modeling Simulation* **8**, 1269 (2010).
- ¹⁹D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- ²⁰P. Diaconis, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1659 (1996).
- ²¹D. A. Levin, Y. Peres, and E. L. Wilmer, *Markov Chains and Mixing Times* (American Mathematical Society, Providence, RI, 2009).