Binomial distribution based τ-leap accelerated stochastic simulation

Abhijit Chatterjee and Dionisios G. Vlachos[a)]
Department of Chemical Engineering and Center for Catalytic Science and Technology (CCST), University of Delaware, Newark, Delaware 19716

Markos A. Katsoulakis
Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003

(Received 28 July 2004; accepted 22 October 2004; published online 21 December 2004)

Recently, Gillespie introduced the τ-leap approximate, accelerated stochastic Monte Carlo method for well-mixed reacting systems [J. Chem. Phys. **115**, 1716 (2001)]. In each time increment of that method, one executes a number of reaction events, selected randomly from a Poisson distribution, to enable simulation of long times. Here we introduce a binomial distribution τ-leap algorithm (abbreviated as BD-τ method). This method combines the bounded nature of the binomial distribution variable with the limiting reactant and constrained firing concepts to avoid negative populations encountered in the original τ-leap method of Gillespie for large time increments, and thus conserve mass. Simulations using prototype reaction networks show that the BD-τ method is more accurate than the original method for comparable coarse-graining in time. © *2005 American Institute of Physics.* [DOI: 10.1063/1.1833357]

INTRODUCTION

The foundations of microscopic or exact Monte Carlo (MC) simulation, termed also as stochastic simulation algorithm (SSA), for well-mixed, chemically reacting systems were laid down several years ago by Gillespie.[1,2] Since then, SSA has become one of the most widespread computational tools in chemical sciences. Yet, its microscopic nature, i.e., the execution of one reaction per SSA time increment, has severely limited SSA to relatively short time scales and small and fairly similar size populations. In order to accelerate the exact stochastic simulation, the next reaction method was recently proposed.[3] In addition, several approximate methods that capitalize on separation of time scales have been proposed to accelerate the exact SSA.[4–7] Finally, separation of time scales based on the master equation has also been explored to deal with the inherent stiffness of chemical kinetics.[8,9] A review of the various acceleration MC methods is given in Ref. 10. A major limitation of most acceleration MC methods is that usually the noise is either amplified or reduced substantially (see Ref. 10 for a comparison of various methods).

The τ-leap method was recently introduced by Gillespie[11] for approximate, accelerated MC simulations of chemical kinetics in well-mixed reacting systems. The essence of the τ-leap method is that instead of executing one reaction in every microscopic time interval and changing the participating species by stoichiometric populations, one selects a coarse-time increment, τ, which is usually larger than the microscopic one. In this coarse time increment, one ''fires'' each reaction multiple times and updates the populations after each time step accordingly. The number of times each reaction is fired is selected randomly from a Poisson distribution (see Ref. 11 for details and below). We refer to the original τ-leap method of Gillespie as the Poisson distribution based τ-leap method (the PD-τ method). The midpoint τ-leap[11] and the implicit τ-leap method[12] attempt to improve the accuracy and robustness of the original method.

The PD-τ leap method and its variants partially sacrifice accuracy for greater speed by enabling molecular ''bundles,'' i.e., a large number of firings, sampled from the Poisson distribution to react over coarse (mesoscopic) time intervals. Several examples studied using the PD-τ leap method have evidenced that it is a significant advancement over the exact SSA in terms of computational requirements while providing nearly the correct noise when the time leaps are not as large. This is in contrast to most other acceleration methods mentioned above. Since inclusion and understanding of noise is a main reason for performing a stochastic simulation at the first place, the PD-τ leap method, along with its variations, is the most promising single acceleration technique. It is not then surprising that despite its short life time, the PD-τ leap method has already been employed in various, mainly biological, studies.[13,14]

A problem with the PD-τ leap method is that physically unrealistic negative populations (concentrations in the continuum terminology) may result, arising from the unbounded Poisson random variable and the fact that reaction firings are independent. In fact, this problem occurs with probability 1 given a sufficient long computation. In our experience, this problem is encountered when molecular population sizes are small and/or the time leaps are large. The work of Ref. 14 highlights this problem of the τ-leap method that in their case limited the acceleration of hybrid multiscale simulation of complex reaction networks.

In order to overcome the negative populations resulting in the PD-τ leap method encountered in moderately large values of τ, in this paper we introduce the binomial distribution based τ-leap method (abbreviated as BD-τ method). In

a)Author to whom correspondence should be addressed. Electronic mail: vlachos@che.udel.edu

this new method, aside from choosing the random variable from the bounded binomial distribution, motivated from our recent work on spatially coarse-grained MC simulation,[15–17] the limiting reactant and constrained firings concepts are also invoked to ensure mass conservation. In a mathematical context, mass conservation ensures that a well-defined Markov process can be written for the populations in the system; this step is crucial in obtaining eventually rigorous numerical analysis results for the τ-leap method. Prototype reaction examples are employed to illustrate the performance of the methods. Numerical examples indicate that for comparable coarse-graining in time, the BD-τ method is more accurate than the PD-τ leap method.

THE BD-τ LEAP METHOD ALGORITHM AND ITS MATHEMATICAL RATIONALE

A well-mixed reacting system of N molecular species is considered. The state vector $\mathbf{X}(t)=(X_1(t),\ldots,X_N(t))$ contains the number of molecules (population size) $X_i(t)$ of all species S_i at time t, $1\leqslant i\leqslant N$. Species participate in M chemical reactions denoted by R_j, $j=1,2,\ldots,M$, with a propensity function (or transition probability per unit time) a_j. Here $a_j dt$ is the probability that one reaction R_j will happen in the infinitesimal time interval $[t,t+dt)$ (see discussion in Ref. 11 on propensity functions) and ν_{ij} is the stoichiometric coefficient of species S_i in reaction R_j.

The original PD-τ leap method

A brief outline of the PD-τ leap method is first given. In each time leap τ of the original PD-τ leap method one selects the number k_j of each reaction R_j to be executed (also called ''firings'') from a Poisson distribution

$$P_{\mathrm{PD}}(k_j;a_j\tau)=\frac{e^{-a_j\tau}}{k_j!}(a_j\tau)^{k_j}. \tag{1}$$

In the case of a single reaction, the problem with such a selection step is that k_j may exceed the available population size of one or more chemical species, resulting in negative populations. Furthermore, in the PD-τ leap method the firings of all reactions are independent random numbers. Thus, even though each k_j may not lead to negative populations *per se*, the simultaneous execution of all reactions in a reaction network may do so. In simple words, one needs to constrain k_j to conserve mass at all times. Next we describe an algorithm that is capable of doing that.

The proposed BD-τ leap algorithm

An important conceptual assumption in all τ leap methods is that τ must be sufficiently small so that the change in the state vector $\mathbf{X}(t)$ and consequently in the propensities is negligible. This condition was termed the leap condition in Ref. 11. When the leap condition is approximately satisfied, the reaction firings of reaction R_j can be assumed to be statistically independent from one time step to the next and also of firings of other reactions $R_{j'}$, $j'=1,\ldots,M$, $j'\neq j$ during each time step. Furthermore, the propensity functions would be nearly constant during $[t,t+\tau)$, where $a_j=a_j(t)$ is computed based on $\mathbf{X}(t)$. In practice, of course, one is interested in taking as large changes in population as possible to increase the time steps and reduce the CPU. In doing that, the leap condition is violated, i.e., the propensity functions, which are kept constant during a time step, actually change considerably because of changes in population via the same reaction and other reactions. Violation of the leap condition unavoidably leads to some error and possibly to negative populations.

Negative populations are avoided in the BD-τ leap method by placing an upper bound on the number of S_i molecules, $1\leqslant i\leqslant N$, consumed in the time interval $[t,t+\tau)$. In the BD-τ leap method, molecular bundles for each of the M reactions are allowed to fire sequentially in a given order in $[t,t+\tau)$. The maximum number of firings $k_{\max}^{(j)}$ for reaction R_j is determined by the limiting reactant, i.e., the species that can be consumed completely if the reaction were to go to completion. The elementary chemistry concept of limiting reactant is required for bimolecular and trimolecular reactions only. The identification of the limiting reactant is a crucial step that ensures that the number of firings of each chemical reaction does not lead to negative populations in *a single reaction* for any chemical species participating in this reaction.

Given that out of $k_{\max}^{(j)}$ maximum firings, $a_j\tau$ fire on the average, and the leap condition is satisfied, we will approximate the true microscopic dynamics by assuming that each R_j firing has a probability $p=a_j\tau/k_{\max}^{(j)}$ of happening and a probability $(1-p)$ of failing to occur. It can be mathematically shown that the number of firings k_j of R_j then belongs to the binomial distribution,

$$P_{\mathrm{BD}}(k_j;p,k_{\max}^{(j)})=\frac{k_{\max}^{(j)}!}{k_j!(k_{\max}^{(j)}-k_j)!}p^{k_j}(1-p)^{k_{\max}^{(j)}-k_j} \tag{2}$$

(see for example the binomial distribution generated as sum of coin-tossings in Ref. 18).

Introducing $k_{\max}^{(j)}$ alone is insufficient of ensuring mass conservation of an entire reaction network. Once k_j is determined from Eq. (2), the number of molecules left to react via the remaining reactions has been reduced. To account for this, we introduce the vector $\tilde{\mathbf{X}}$ that tracks the currently available reacting population size during a time leap. At time t (before execution of any reaction), $\tilde{\mathbf{X}}(t)=\mathbf{X}(t)$. After executing the jth reaction, $\tilde{\mathbf{X}}$ is updated by subtracting the number of reactant molecules of all species consumed in R_j. This update also modifies $k_{\max}^{(j+1)}$ for the next reaction. This step ensures that the maximum allowed firings, $k_{\max}^{(j+1)}$, left over in executing the subsequent reaction R_{j+1} would not exceed the actually available populations and ensures mass conservation of an arbitrary complex *reaction network*, which is the main objective of this paper.

Finally, once all M reactions are allowed to trigger, the populations of all species are updated based on the stoichiometry of the chemical reactions. From an efficient implementation point of view, updating of populations can be done only *locally*, as suggested in Ref. 3, and is well known in the molecular simulation literature,[19] i.e., update only the species participating in a certain reaction.

Summarizing, the BD-τ algorithm consists of the following steps:

(1) Obtain the stoichiometric coefficients ν_{ij}, the initial population size $\mathbf{X}(0)$, and the rate constants c needed in computing the propensities.
(2) Initialize the time, $t := t_0$.
(3) Repeat steps 4–6 until a maximum time, $t_{\max}$, is reached.
(4) Compute the propensities $a_j(\mathbf{X}(t))$ using the population $\mathbf{X}(t)$ and set $\tilde{\mathbf{X}} := \mathbf{X}$.
(5) Select τ (see text below) and update time according to $t := t + \tau$.
(6) For $j = 1$ to M reactions
  (a) Find $k_{\max}^{(j)} = \min_{(i=1,\ldots,N)_{\nu_{ij}<0}}(\mathrm{int}(\tilde{X}_i / |\nu_{ij}|))$, where int( ) is the greatest integer function.
  (b) With p defined as $p = a_j\tau / k_{\max}^{(j)}$, sample k_j from the binomial distribution Eq. (2).
  (c) Set $\tilde{X}_i := \tilde{X}_i + \nu_{ij}k_j$ for $i = 1,\ldots,N$ if $\nu_{ij} < 0$.
(7) (d) Go to step 6a.
(8) Update populations $X_i := X_i + \nu_{ij}k_j$, for $i = 1,\ldots,N$.
(9) Go to step 3.

Steps 1 and 2 initialize the state of the system, and step 4 computes the propensities for the next τ-leap. Step 5 updates the time, while step 6 randomly selects the number of firings for each reaction in the reaction network. These steps occur in every MC simulation.

Note that if $a_j\tau > k_{\max}^{(j)}$, $j = 1,2,\ldots,M$, we set $p = 1$. Thus, all available molecules of the limiting reactant in R_j react. As a result, no more molecules than the available population can ever be consumed. However, in the results presented below, we have never encountered $a_j\tau > k_{\max}^{(j)}$ despite taking large time steps. In case that p approaches 1, the time increment is already too large to provide good results and it should be reduced, i.e., the condition $p = 1$ can be viewed as an upper bound for the time increment of the leap methods. Finally, note that the choice of reaction firings rather than species population and the continuous update of the vector $\tilde{\mathbf{X}}$ after each reaction has been executed are judicious. Specifically, by choosing species population from a Binomial distribution, i.e., carry out the loop (6) over species, instead of chemical reaction firings proposed above, it would be nearly impossible to ensure mass conservation for complex reaction networks where many species are shared between multiple reactions.

Implications of constrained firings in the new algorithm

The departure of the new methodology from the original PD-τ leap method stems from (a) the *binomial distribution*, which provides a bounded random variable, and (b) the fact that the *firings of reactions are coupled or constrained* by the availability of species (through $k_{\max}^{j}$) during the execution of a τ-leap.

As mentioned in Ref. 11, when the time interval τ is microscopic (i.e., of the order of the average time increment of the exact SSA, $1/\Sigma_j a_j$) the time trajectories of species populations predicted by the PD-τ method and the exact SSA match. Furthermore,

$$a_j\tau \pm \sqrt{a_j\tau} \tag{3}$$

firings of reaction R_j are triggered in τ. On the other hand, for the BD-τ method one has from Eq. (2) that

$$a_j\tau \pm (a_j\tau)^{1/2}(1 - a_j\tau/k_{\max}^{(j)})^{1/2} \tag{4}$$

firings are expected in τ, where $a_j\tau$ and $(a_j\tau)^{1/2}(1 - a_j\tau/k_{\max}^{(j)})^{1/2}$ are the mean and standard deviation of the binomial distribution, respectively. When τ is microscopic, $a_j\tau \ll k_{\max}^{j}$, $j = 1,\ldots,M$, the correct zeroth and first moments are also obtained for the BD-τ method. While in the infinite size limit the binomial distribution in Eq. (2) and the Poisson distribution in Eq. (1) yield the same asymptotic behavior, for finite sizes the noise of the BD-τ method is less than that of the BD-τ method. This feature has important ramifications regarding the accuracy of the BD-τ leap method, as numerical examples below demonstrate.

The higher computational cost of generating random variables for the Poisson or binomial distributions in comparison to a uniform distribution (needed in the exact SSA) renders the approximate methods inefficient when τ is microscopic or nearly so (see comparison of CPU in the next section). Thus, in practice one is interested in taking large time leaps where the leap condition is satisfied approximately. At this point one may wonder about the effect of sequential order of reactions in computing $\tilde{\mathbf{X}}$ and $\mathbf{k}_{\max}^{(j)}$. The average number of firings $k_j = a_j\tau$ is unaffected by $k_{\max}^{(j)}$. However, the noise $(a_j\tau)^{1/2}(1 - a_j\tau/k_{\max}^{(j)})^{1/2}$ is affected by $k_{\max}^{(j)}$ and, thus, the sequential (in a deterministic order) execution of reactions could affect the noise of the solution. Numerical comparison between the exact SSA and the approximate methods (see next section) for several reaction networks demonstrates that the accuracy of the solution obtained via both the PD-τ and BD-τ method is lost (due to serious violation of the leap condition) before biased solutions can be noticed. Alternatively, the bias in the noise could potentially be improved by randomly choosing the execution of all M reactions. An example using this approach is also presented below.

EXAMPLES COMPARING THE PD-τ AND BD-τ METHODS

The two τ-leap methods are numerically compared for different prototype reaction networks. In all reaction networks the magnitude of τ varies from small, to satisfy the leap condition and numerically validate the BD-τ leap method, to large, where the leap condition is violated in order to observe the effect of coarse time intervals on the solution. In all results below reactions are picked sequentially, as indicated in the loop (step 6) of the algorithm, except for one example where reactions are randomly ordered in each time leap.

The following reaction network of first-order reactions is first considered with rate constants indicated

$$S_1 \underset{1}{\overset{2}{\rightleftarrows}} S_2 \underset{2}{\overset{1}{\rightleftarrows}} S_3 .$$

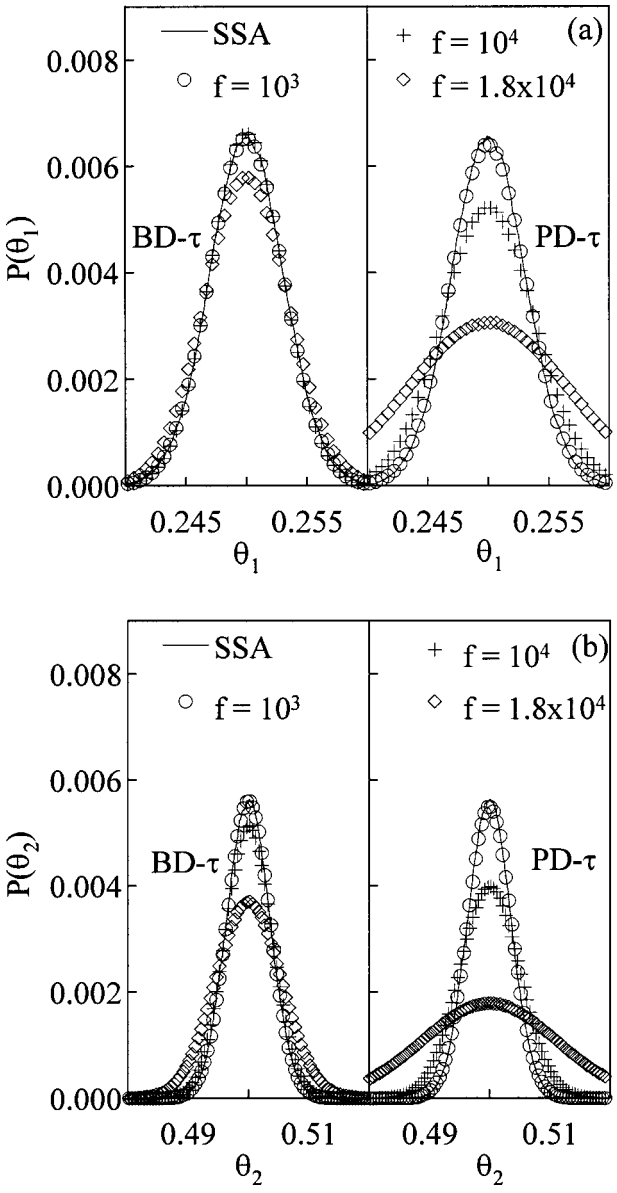

FIG. 1. Equilibrium probability density function (pdf) for species populations S_1 (top) and S_2 (bottom) in the reaction $S_1 \underset{1}{\overset{2}{\rightleftarrows}} S_2 \underset{2}{\overset{1}{\rightleftarrows}} S_3$, using SSA (solid lines) and the BD-τ and PD-τ methods with an initial population of $X_1(0)=20000$, $X_2(0)=X_3(0)=0$. Coarse-graining factors are $f=10^3$ (circles), 10^4 (crosses), and 1.8×10^4 (diamonds). The BD-τ gives more accurate results than the PD-τ method.

Since these are unimolecular reactions, the limiting reactant problem is trivially fulfilled. The choice of this slightly more complicated example compared to a simple isomerization reaction stems from our intention to have a single species, namely S_2, participating in several reactions so we illustrate how the new method overcomes the problem of negative populations via coupled or constrained firings. The time in each MC iteration is updated with time increments

$$\tau = f \Big/ \sum_{j=1}^{M} a_j \tag{5}$$

for both the BD-τ and PD-τ methods. Here $1/\Sigma_{j=1}^{M} a_j$ is the averaged microscopic time increment of the SSA and f is a coarse-graining factor (we choose $f>1$). Besides allowing control of temporal coarse-graining in a simple, transparent way, Eq. (5) ensures nearly the same τ for both BD-τ and PD-τ leap methods and thus enables a direct comparison of accuracy and CPU requirements of the two methods. Obviously, this is not an optimal way of time stepping but is sufficient for method comparison.

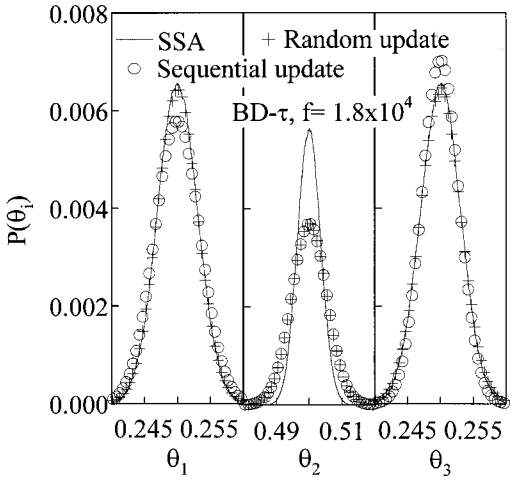

FIG. 2. Comparison of equilibrium probability distribution functions (pdf) for all species in the reaction $S_1 \underset{1}{\overset{2}{\rightleftarrows}} S_2 \underset{2}{\overset{1}{\rightleftarrows}} S_3$ using the BD-τ method with sequential (circles) and random (crosses) execution of reactions. The coarse-graining factor is $f=1.8\times10^4$. For this example the noise is slightly better when reactions are executed in random order.

Initially only species S_1 is assumed to be present, with $X_1(0)=20\,000$ molecules, $X_2(0)=X_3(0)=0$. The equilibrium probability density function (pdf) for species S_1 and S_2 using the exact SSA and the BD-τ and PD-τ methods is plotted in Fig. 1 for different values of f. Note that in all figures the normalized population, $\theta_i=\langle X_i\rangle/\Sigma_{j=1}^{N} X_j(0)$, is graphed. For small f (such as $f=10^3$) the equilibrium pdfs of all methods overlap. For larger values of f (e.g., 10^4 and 1.8×10^4) the mean population size is correct. The noise of the BD-τ is fairly close to the exact SSA, whereas that of the PD-τ methods is substantially overestimated. For even larger values of f, negative populations result in the PD-τ method, whereas the BD-τ method does not give negative populations.

Simulations for this first reaction system indicate that the new BD-τ leap method is potentially more accurate than the PD-τ leap method for similar acceleration while preventing negative populations. The latter aspect permits use of larger time steps, and thus further acceleration, that cannot be realized with the PD-τ method. The difference in noise between the BD-τ and PD-τ methods observed in Fig. 1 can be rationalized in terms of the mean and the standard deviation of the probability distributions of Eqs. (1) and (2). Specifically, according to (4), the noise in the BD-τ leap method is lower than that of the PD-τ method because of the term $(1-a_j\tau/k_{\max}^{(j)})^{1/2}$. In the limit of large populations, $X_i\to\infty$, this term is equal to 1, and the two τ-leap methods give identical solutions.

Figure 2 compares the pdf for all three species for the above example obtained by sequentially sampling the six

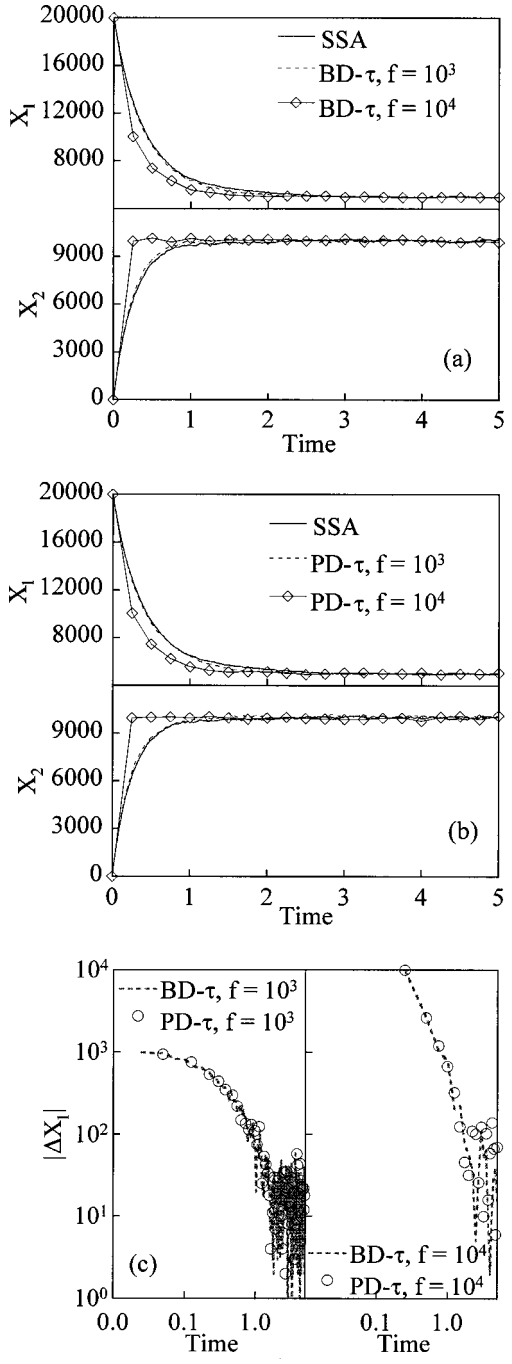

FIG. 3. Single trajectories using (a) BD-τ and (b) PD-τ methods. Deviations of the BD-τ and PD-τ transient solutions from those of SSA occur only for large time increments τ. (c) Corresponding changes in population X_1 per time leap versus time. Other parameters are those of Fig. 1.

elementary reactions in a deterministic manner ($j=1,...,M=6$) and by randomly ordering the reactions at each time leap. It is clear that while the means of the pdf are not affected by how the reactions are picked, the noise of species S_1 and S_3 improves whereas that of S_2 remains relatively unaffected, when random selection of reactions is implemented. This change in noise stems from the constraint firings and the fact that the noise depends on $k_{\max}^{(j)}$ for relatively large time leaps. Based on limited examples, it appears that random sampling may perform better. However, further work is needed to fully understand this issue.

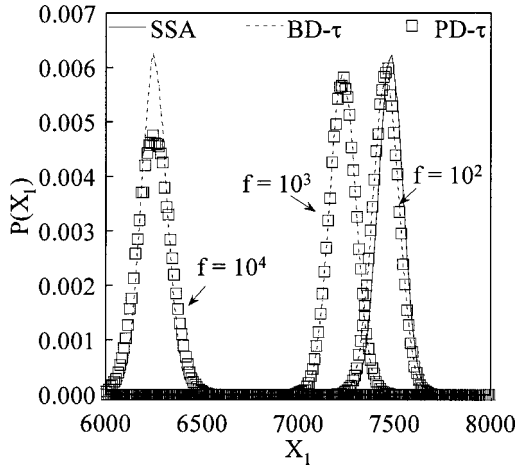

FIG. 4. Probability density functions (pdf) for species population S_1 at $t=0.75$ time units. SSA is denoted by solid line, and the BD-τ and PD-τ results are denoted by dotted lines and squares, respectively. Deviations in pdf are observed in both BD-τ and PD-τ for large coarse-graining factors f. Other parameters are those of Fig. 1.

Figures 3(a) and 3(b) show single trajectories obtained with the two τ-leap methods and the exact SSA, and Fig. 3(c) shows the corresponding change in the population size X_1 per MC iteration as a function of time. Note that the molecular bundles of the two τ-leap methods follow each other relatively closely, so CPU comparison (discussed below) is meaningful. For relatively small values of f, the transient, single trajectory solutions of the approximate methods agree well with the solution of the exact SSA [Figs. 3(a) and 3(b)]. On the other hand, for $f=10^4$ too many firings occur, and deviations are visible at short times. These errors at short times for large values of f persist upon ensemble average over 10^4 trajectories as shown in Fig. 4 for a chosen time, i.e., these are not an artifact of examining the single trajectories displayed in Fig. 3. While at short times the accuracy can be improved by taking lower values of f (see pdfs in Fig.

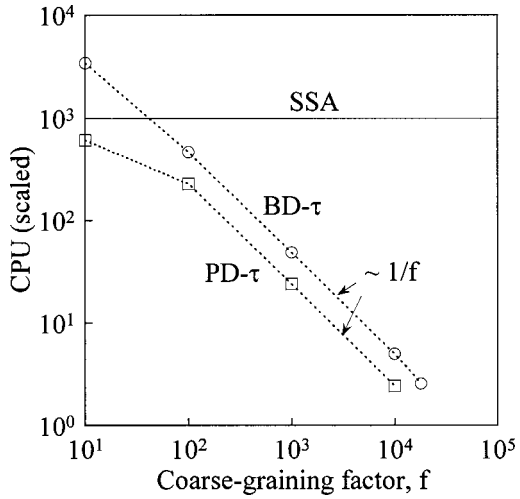

FIG. 5. CPU comparison between various methods for parameters of Fig. 1. The maximum error in noise is set as 40%. The BD-τ method is twice as slow as the PD-τ method for the same value of f but larger values of f are possible for the former method. Significant acceleration compared to the exact SSA is found for large τ-leaping. Other parameters are those of Fig. 1.

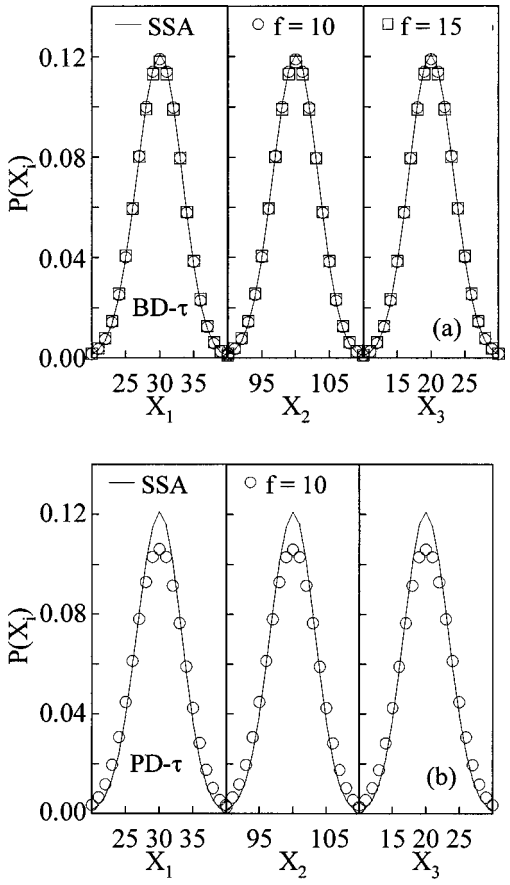

FIG. 6. Equilibrium probability density function (pdf) for species populations in the reaction $S_1+S_2 \underset{150}{\overset{1}{\rightleftarrows}} S_3$, using SSA (solid lines) and (a) the BD-τ and (b) PD-τ method with an initial population of $X_1(0)=50$, $X_2(0)=120$, and $X_3(0)=0$. Coarse-graining factors are $f=10$ (circles) and 15 (squares). Negative populations are encountered using the PD-τ method for $f>13$ (not shown).

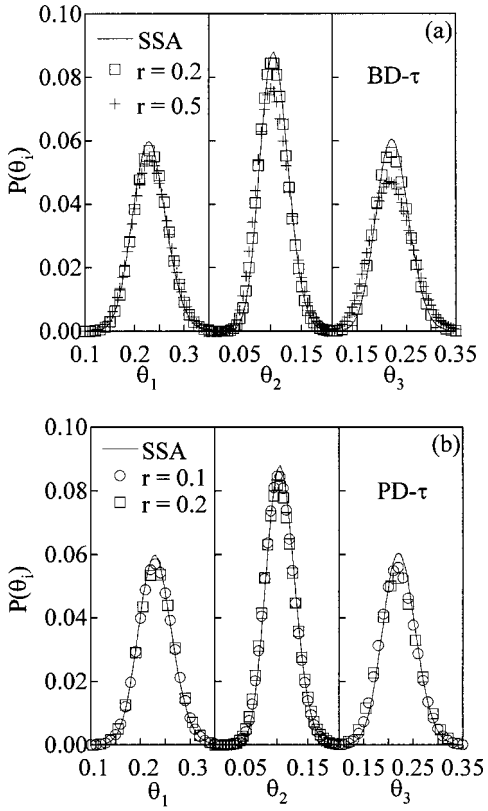

FIG. 7. Equilibrium probability density function (pdf) for species populations in the complex reaction network using SSA (solid lines) and the BD-τ and PD-τ methods with an initial population of $X_1(0)=100$, $X_2(0)=75$, $X_3(0)=25$, and $X_4(0)=0$. Coarse-graining factors are $r=0.1$ (circles), 0.2 (squares), and 0.5 (crosses). The results of the BD-τ method for $r=0.1$ are indistinguishable from SSA (not shown for clarity). Negative populations are encountered using the PD-τ method for $r>0.2$ (not shown), whereas the BD-τ method still gives reasonable results for $r=0.5$.

3 for various values of f), comparison of Figs. 2 and 3 clearly indicates that adaptive f or τ-leap strategies are highly desirable. Some such strategies have already been proposed[11,20] and an alternative approach is proposed in the conclusions.

Figure 5 compares the CPU time of the BD-τ and PD-τ methods for advancing the same real time at equilibrium with a maximum allowable error in noise of 40%. The CPU decreases inversely proportional with increasing f, and significant savings occur upon τ-leaping compared to the exact SSA. We have found that the higher cost of the BD-τ compared to the PD-τ is associated with the specific implementation of random number generation, that here is done following Ref. 21. This comparison may change if different algorithms of random number generation are followed. Finally, larger time steps can be realized with the BD-τ leap method, since mass is always conserved and the method appears to be more accurate. Larger time increments may allow for lower computational cost when the BD-τ leap method is used.

In addition to the above example, the reversible recombination reaction

$$S_1+S_2 \underset{150}{\overset{1}{\rightleftarrows}} S_3$$

with an initial population of $X_1(0)=50$, $X_2(0)=120$, and $X_3(0)=0$ was also studied. This example tests the species constraint arising from the presence of a limiting reactant. The results, shown in Fig. 6, are qualitatively similar with the previous example.

Finally, a more complex, nonlinear, stiff reaction network is considered

$$2S_1 \underset{1}{\overset{0.02}{\rightleftarrows}} S_2$$

$$S_1+S_2 \underset{1}{\overset{1}{\rightleftarrows}} 2S_3$$

$$S_3+S_4 \underset{1}{\overset{0.05}{\rightleftarrows}} S_1$$

$$S_2 \underset{1}{\overset{1}{\rightleftarrows}} S_4$$

with initial conditions $X_1(0)=100$, $X_2(0)=75$, $X_3(0)=25$ and $X_4(0)=0$. The equilibrium population sizes are rela-

tively small, namely $\langle X_1\rangle = 46.2$, $\langle X_2\rangle = 21.3$, $\langle X_3\rangle = 44.6$, and $\langle X_4\rangle = 20.9$ molecules. In this specific example, we have explored an alternative method of choosing τ that is based on small changes in the instantaneous conversion of all reactants, namely using $\tau = \min_i(rX_i/\Sigma_{j=1}^{M}|\nu_{ij}|a_j)$, s.t. $0<r<1$, $\nu_{ij}<0$. The coarse-graining scaled factor r captures the magnitude of propensities. Note that this selection is simply postulated (it has been found to work well for all reaction networks tested herein) and, as noted in Refs. 11 and 20, efficient adaptive criteria should further be researched.

Figure 7 shows the equilibrium pdf of selected species for different values of the coarse-graining factor r. Qualitatively, the same conclusions are reached as for the simpler reaction networks discussed above, with the differences between the two τ-leap methods being less profound for small values of r. For moderate values of r, the PD-τ method leads again to negative concentrations during the course of a computation.

CONCLUSIONS

Motivated by the recent work of Gillespie[11] for time acceleration of Monte Carlo (MC) methods in well-mixed systems, a binomial distribution based τ-leap (BD-τ) method was introduced as an approximate, accelerated MC technique that can reach longer time scales of well-mixed systems. This method combines the bounded nature of the binomial distribution variable with the limiting reactant and constrained firing concepts to avoid negative populations encountered in the τ-leap, Poisson distribution based (PD-τ) method of Ref. 11, and thus conserve mass. Furthermore, mass conservation ensured by the BD-τ leap method allows one to write accelerated dynamics as a well-defined population Markov process. It would be desirable to derive accelerated, τ-leap type of methods, from the microscopic MC via coarse-graining in time of the underlying generator in the spirit of systematic derivations of spatial coarse-graining of microscopic MC.[16] The latter point is important in order to mathematically derive theoretical error estimates, using ideas from information loss theory (for such an example of error derivation in spatial MC, see Ref. 17), which would allow dynamic error control in time leaping analogous to well-established techniques employed in ordinary differential equations. These points will be explored in future work.

Numerical examples indicate that the BD-τ method is more accurate than the PD-τ method for comparable time increments. Several extensions of the BD-τ method are possible for future work. As an example, in analogy to the implicit τ-leap method based on the Poisson distribution of Ref. 12, an implicit BD-τ scheme is possible. As another extension discussed in Ref. 10, combination of the BD-τ method with spatially coarse-grained MC methods,[15–17] is also entirely possible.

ACKNOWLEDGMENTS

This work was partially supported by the NSF through Grant Nos. ITR-0219211, DMS-0100872, and CTS-0312117.

[1] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[2] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[3] M. A. Gibson and J. Bruck, J. Phys. Chem. A **104**, 1876 (2000).
[4] J. He, H. Zhang, J. Chen, and Y. Yang, Macromolecules **30**, 8010 (1997).
[5] H. Resat, H. S. Wiley, and D. A. Dixon, J. Phys. Chem. B **105**, 11026 (2001).
[6] D. G. Vlachos, Chem. Eng. Sci. **53(1)**, 157 (1998).
[7] M. A. Snyder, A. Chatterjee, and D. G. Vlachos, Comput. Chem. Eng. (in press).
[8] C. V. Rao and A. P. Arkin, J. Chem. Phys. **118**, 4999 (2003).
[9] E. L. Haseltine and J. B. Rawlings, J. Chem. Phys. **117**, 6959 (2002).
[10] D. G. Vlachos, Adv. Chem. Eng. (to be published).
[11] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[12] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, J. Chem. Phys. **119**, 12784 (2003).
[13] J. Pucha and A. M. Kierzek, Biophys. J. **86**, 1357 (2004).
[14] K. Burrage, T. H. Tian, and P. Burrage, Prog. Biophys. Mol. Biol. **85(2–3)**, 217 (2004).
[15] M. Katsoulakis, A. J. Majda, and D. G. Vlachos, Proc. Natl. Acad. Sci. U.S.A. **100(3)**, 782 (2003).
[16] M. A. Katsoulakis, A. J. Majda, and D. G. Vlachos, J. Comput. Phys. **186**, 250 (2003).
[17] M. A. Katsoulakis and D. G. Vlachos, J. Chem. Phys. **119**, 9412 (2003).
[18] A. N. Shiryaev, *Probability* (Springer Verlag, Berlin, 1989).
[19] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford Science, Oxford, 1989).
[20] D. T. Gillespie and L. R. Petzold, J. Chem. Phys. **119**, 8229 (2003).
[21] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge, 1986).