



# New “Tau-Leap” Strategy for Accelerated Stochastic Simulation

Doraiswami Ramkrishna,\* Che-Chi Shu,<sup>†</sup> and Vu Tran

School of Chemical Engineering, Purdue University, West Lafayette, Indiana 47907, United States

## S Supporting Information

**ABSTRACT:** The “Tau-Leap” strategy for stochastic simulations of chemical reaction systems due to Gillespie and co-workers has had considerable impact on various applications. This strategy is reexamined with Chebyshev’s inequality for random variables as it provides a rigorous probabilistic basis for a measured  $\tau$ -leap thus adding significantly to simulation efficiency. It is also shown that existing strategies for simulation times have no probabilistic assurance that they satisfy the  $\tau$ -leap criterion while the use of Chebyshev’s inequality leads to a specified degree of certainty with which the  $\tau$ -leap criterion is satisfied. This reduces the loss of sample paths which do not comply with the  $\tau$ -leap criterion. The performance of the present algorithm is assessed, with respect to one discussed by Cao et al. (*J. Chem. Phys.* **2006**, *124*, 044109), a second pertaining to binomial leap (Tian and Burrage *J. Chem. Phys.* **2004**, *121*, 10356; Chatterjee et al. *J. Chem. Phys.* **2005**, *122*, 024112; Peng et al. *J. Chem. Phys.* **2007**, *126*, 224109), and a third regarding the midpoint Poisson leap (Peng et al., 2007; Gillespie *J. Chem. Phys.* **2001**, *115*, 1716). The performance assessment is made by estimating the error in the histogram measured against that obtained with the so-called stochastic simulation algorithm. It is shown that the current algorithm displays notably less histogram error than its predecessor for a fixed computation time and, conversely, less computation time for a fixed accuracy. This computational advantage is an asset in repetitive calculations essential for modeling stochastic systems. The importance of stochastic simulations is derived from diverse areas of application in physical and biological sciences, process systems, and economics, etc. Computational improvements such as those reported herein are therefore of considerable significance.

## 1. INTRODUCTION

In recent times, stochastic simulations have assumed extraordinary importance in modeling biological and nano systems. The importance of the Monte Carlo simulation methodology arises from its capacity to substitute for the less tractable master equation and mean field approaches. This capacity is derived from its basic simplicity which could, however, be seriously compromised if computational times become forbiddingly large because of the repetitive attribute of the methodology. Consequently, considerable effort has come to pass on finding efficient simulation strategies so that computation of statistical averages of stochastic behavior can become facile. The issue of efficiency arises in negotiating accuracy with sacrifice of rigor or through crafty approximations. In stochastic systems, however, the validity of approximations is uncertain because they invariably involve future values of random variables in the system leading to conservative and hence less efficient heuristics. The remedy must lie in the use of mathematical propositions concerning probabilities with which approximations about random variables are true.

It is the purpose of this work to aid in the quest for efficient stochastic simulation strategies by exploiting an important inequality that could greatly rationalize the process. This is Chebyshev’s inequality that is routinely encountered in the treatment of random variables.<sup>1</sup> The inequality is concerned with the probability with which a random variable deviates from its expectation in terms of its variance.

We are concerned here with a stochastic system of chemical transformations involving several species at a fixed temperature. Temperature variation can be captured in the reaction rate constants which can, in turn, influence the rate at which the

propensity function changes. However, this temperature change needs to be small since one big assumption made throughout the development of  $\tau$ -leap is to assume the propensity function stays constant during each leap. The stochastic behavior arises out of the randomness in reaction rates and is reflected in the numbers of different reaction species at different times. Simulation of this system can be performed with an *exact* algorithm that has come to be known as SSA (stochastic simulation algorithm). The term “exact” is derived from the fact that the simulation is based on rigorously derived distributions of “waiting times” during which the system remains “quiescent” without any reaction event. While the methodology is usually attributed to Gillespie,<sup>2</sup> a basic version of the idea appeared in a paper by Kendall<sup>3</sup> who used the term “quiescent interval” instead of “waiting time”. This method was generalized for the simulation of population balance equations in a doctoral thesis by Shah,<sup>4</sup> subsequently published in the chemical engineering literature<sup>5</sup> independently of Gillespie’s publication<sup>2</sup> in the geophysical literature. Some further insights as to the connection of the foregoing algorithm to the master equation for population balances are given by Ramkrishna.<sup>6</sup>

The virtue of SSA’s exactness is computational accuracy but at the cost of large computation times. SSA also provides the starting point for various improvements pursued by researchers in this area to minimize the cost of computation. The term

**Special Issue:** Ganapati D. Yadav Festschrift

**Received:** July 22, 2014

**Revised:** September 16, 2014

**Accepted:** September 22, 2014

**Published:** September 22, 2014

“tau-leap”, coined by Gillespie and co-workers, aptly describes the underlying strategy of “leaping” over relatively less significant events (as reflected by minimal changes in the so-called “propensities” or transition rates of individual reactions). For all of the methods discussed in this work, each leap is checked after each calculation. Anderson’s method<sup>7</sup> is a precheck method and hence produces a very different mechanism of filtering leaps. For any given selection of  $\varepsilon$ , postcheck methods inevitably encounter some ineffective leap steps at which the leap condition is violated. Depending upon the number of violations and the extent of each violation, the accuracy of the simulation will be affected. Our method introduces a way to promote the likelihood of acquiring more efficient leap steps by appropriate increase of  $\delta$ . Anderson’s method introduces a procedure to do precheck before performing a calculation of leaps. Only steps where the criteria are satisfied will continue with the calculation of the corresponding leaps. Moreover, only the first step of picking  $\tau$  will follow the formula of Cao et al.<sup>8</sup> The subsequent steps of evaluating  $\tau$  strictly follow their procedure. For each new calculated  $\tau$ , it is then evaluated with the leap condition. Depending upon the value of  $\tau$ , different scaling factors can be applied to adjust  $\tau$ . If a leap for  $\varepsilon = \varepsilon'$  is rejected,  $\tau$  will be decreased by multiplying with some  $p < 1$ . If a leap is accepted for  $\varepsilon = \varepsilon'$  but would fail if  $\varepsilon = 3\varepsilon'/4$ , it will be reduced by multiplying it by some  $p^*$  that satisfies  $p < p^* < 1$ . If a leap is accepted for both cases,  $\tau$  can be enlarged by raising it to the power of  $q$  such that  $q$  is between 0 and 1. Thus, the algorithm for precheck will only accept those  $\tau$  that satisfy the leap condition. However, in stochastic simulations, good results can still be accomplished, as long as the number of unsatisfied leaps is low. Every time a leap is rejected, a new calculation will need to be made; therefore the question is a balance between accuracy and speed. Anderson’s approach focuses on checking the leap condition, whereas all other postcheck methods focus on the issue of how to select a  $\tau$  value. Therefore, the method cannot be included in our comparison of postcheck methods with the approach of this work. For that reason, this work will exclusively focus on comparison among postcheck methods. Negative population is another problem associated with the usage of Poisson random number generation. In the later version of the method of Cao et al.<sup>8</sup> the author introduced ways of capturing and treating critical reactions in which the negative population might occur; Peng et al. implemented a binomial leap strategy<sup>9–11</sup> to resolve the problem. In addition to the preceding treatments, we implement a novel strategy using the Chebyshev inequality to improve the simulation process. In this work, we show how the Chebyshev inequality can be used to craft this temporal leap by specifying the probability with which it influences the deviations in the stochastic numbers of chemical species from their expected values. In so doing, following Cao et al.,<sup>8</sup> we retain the Taylor series expansion of the propensities about the state of the system specified at instant  $t$  truncated beyond the quadratic terms. A new algorithm is derived simply using different probabilistic criteria for approximations.

## 2. CHEBYSHEV INEQUALITY

We recall that for any random variable  $X$  with finite expectation  $EX$  and variance  $VX$  the following inequality<sup>1</sup> is satisfied for an arbitrary  $\varepsilon > 0$ .

$$\Pr\{|X - EX| \geq \varepsilon\} \leq \frac{VX}{\varepsilon^2} \quad (1)$$

For our purposes, the preceding inequality may be restated as

$$\Pr\{|X - EX| < \varepsilon\} > 1 - \frac{VX}{\varepsilon^2} \quad (2)$$

Clearly inequality 2 stipulates the minimum of the probability with which the random variable  $X$  deviates from  $EX$  by less than  $\varepsilon$ .

## 3. DEVELOPMENT OF THE NEW “TAU-LEAP” STRATEGY

As observed earlier, we adhere to the stochastic system formulated by Gillespie and Petzold.<sup>8</sup> Thus, there are  $N$  chemical species in numbers represented by the random vector  $\mathbf{X}$ , involved in  $M$  transformations (reactions) with the vector  $\nu_j \equiv \{\nu_{ij}, j = 1, 2, \dots, N\}$  denoting the change in the numbers of species due to each  $j$ th reaction event. Further, the reaction propensity vector, conditional on  $\mathbf{X} = \mathbf{x}$ , is represented by  $\{a_j(\mathbf{x}); j = 1, 2, \dots, M\}$ . If at time  $t$ ,  $\mathbf{X}(t) = \mathbf{x}$ , then after time  $\tau$  (i.e.,  $\tau > 0$ ), we have<sup>1</sup>

$$\mathbf{X}(t + \tau) - \mathbf{x} \equiv \Lambda(\tau; \mathbf{x}) = \sum_{j=1}^M K_j(\tau; \mathbf{x}) \nu_j \quad (3)$$

in which  $K_j(\tau; \mathbf{x})$  represents the number of  $j$ th reaction events during the interval from  $t$  to  $t + \tau$ . For the formulation of  $\tau$ -leap,  $K_j(\tau; \mathbf{x})$  is assumed to be a Poisson process  $\mathcal{P}(a_j(\mathbf{x}); \tau)$  with mean and variance given by  $a_j(\mathbf{x}) \tau$ , notwithstanding the change in  $\mathbf{X}$  during this time interval. Thus, (3) is replaced by

$$\Lambda(\tau; \mathbf{x}) = \sum_{j=1}^M \mathcal{P}(a_j(\mathbf{x}); \tau) \nu_j \quad (4)$$

The “leap condition” of Gillespie and Petzold<sup>1</sup> is given by

$$|a_j(\mathbf{x} + \Lambda(\tau; \mathbf{x})) - a_j(\mathbf{x})| \leq \varepsilon a_o(\mathbf{x}), \quad a_o(\mathbf{x}) \equiv \sum_{j=1}^N a_j(\mathbf{x}) \quad (5)$$

where the symbol  $\varepsilon$  is to be distinguished from that appearing in the Chebyshev inequality (1) or (2). Other replacements considered by these authors for  $a_o(\mathbf{x})$  do not affect the treatment of this work. Defining

$$f_{jj'}(\mathbf{x}) \equiv \sum_{i=1}^N \frac{\partial a_j(\mathbf{x})}{\partial x_i} \nu_{ij'}, \quad j, j' = 1, 2, \dots, M \quad (6)$$

$$\mu_j(\mathbf{x}) \equiv \sum_{j'=1}^M f_{jj'}(\mathbf{x}) a_{j'}(\mathbf{x}), \quad \sigma_j^2(\mathbf{x}) \equiv \sum_{j'=1}^M f_{jj'}^2(\mathbf{x}) a_{j'}(\mathbf{x}), \quad j = 1, 2, \dots, M \quad (7)$$

The change in propensity during the time interval between  $t$  and  $t + \tau$  is given by

$$a_j(\mathbf{x} + \Lambda(\tau; \mathbf{x})) - a_j(\mathbf{x}) \equiv \Delta a_j(\tau; \mathbf{x}) \approx \sum_{j'=1}^M f_{jj'} \mathcal{P}(a_{j'}(\mathbf{x}); \tau) \quad (8)$$

The random variable  $\Delta a_j(\tau; \mathbf{x})$ , in view of the approximation in (8), has expectation  $E\Delta a_j(\tau; \mathbf{x}) \equiv \mu_j(\mathbf{x})\tau$  and variance

$V\Delta a_j(\tau; \mathbf{x}) = \sigma_j^2(\mathbf{x})\tau$ . Gillespie and Petzold choose the interval  $\tau$  in accord with

$$\tau = \min_{j=1,2,\dots,M} \{\tau_j\}, \quad \tau_j \equiv \min \left( \frac{a_0 \varepsilon}{|\mu_j|}, \frac{a_0^2 \varepsilon^2}{\sigma_j^2} \right) \quad (9)$$

We now consider the application of the Chebyshev inequality to the random variable  $\Delta a_j(\tau; \mathbf{x})$  toward determining a suitable criterion that will ensure the leap condition (5). We observe first that the triangular inequality implies that

$$|\Delta a_j(\tau; \mathbf{x})| \leq |\Delta a_j(\tau; \mathbf{x}) - \mu_j(\mathbf{x})\tau| + |\mu_j(\mathbf{x})\tau| \quad (10)$$

By imposing that

$$|\Delta a_j(\tau; \mathbf{x}) - \mu_j(\mathbf{x})\tau| + |\mu_j(\mathbf{x})\tau| \leq \varepsilon a_0(\mathbf{x}) \quad (11)$$

inequality (10) implies the leap condition (5). From (11), it follows that

$$|\Delta a_j(\tau; \mathbf{x}) - \mu_j(\mathbf{x})\tau| \leq \varepsilon a_0(\mathbf{x}) - |\mu_j(\mathbf{x})\tau|$$

The preceding inequality does require that the right-hand side be positive so that

$$\tau < \frac{\varepsilon a_0(\mathbf{x})}{|\mu_j(\mathbf{x})|} \quad (12)$$

to the right of which is a potential candidate for  $\tau$  as prescribed by Cao et al.<sup>8</sup>

Invoking the Chebyshev inequality (2) by replacing the  $\varepsilon$  which appears there by  $(\varepsilon a_0(\mathbf{x}) - |\mu_j(\mathbf{x})\tau)$ , we obtain

$$\begin{aligned} & \Pr\{|\Delta a_j(\tau; \mathbf{x}) - \mu_j(\mathbf{x})\tau| < a_0(\mathbf{x})\varepsilon - |\mu_j(\mathbf{x})\tau|\} \\ & > 1 - \frac{\sigma_j^2(\mathbf{x})\tau}{[a_0(\mathbf{x})\varepsilon - |\mu_j(\mathbf{x})\tau|]^2} \end{aligned} \quad (13)$$

Since the right-hand side of (13) must be nonnegative, it is readily found that an upper bound for  $\tau$  is  $a_0^2 \varepsilon^2 / \sigma_j^2$ , thus recovering the second candidate for  $\tau$ -leap due to Gillespie and Petzold. However, we produce a more effective choice of  $\tau$ -leap in this development.

If we let the right-hand side of (13) be appropriately close to unity, a choice becomes available for  $\tau$ . Thus, setting its value equal to  $\delta$  (e.g.,  $\delta = 0.95$ ), we have (dropping henceforth the argument  $\mathbf{x}$  for notational simplicity)

$$\tau^2 - 2 \left[ \frac{a_0 \varepsilon}{|\mu_j|} + \frac{1}{2(1-\delta)} \left\{ \frac{\sigma_j}{|\mu_j|} \right\}^2 \right] \tau + \left[ \frac{a_0 \varepsilon}{|\mu_j|} \right]^2 = 0 \quad (14)$$

which obtains the following solution for  $\tau_j$ , the subscript referring to its being specific to the  $j$ th reaction.

$$\begin{aligned} \tau_j = & \frac{a_0 \varepsilon}{|\mu_j|} + \frac{1}{2(1-\delta)} \left\{ \frac{\sigma_j}{|\mu_j|} \right\}^2 \\ & \pm \sqrt{\left[ \frac{a_0 \varepsilon}{|\mu_j|} + \frac{1}{2(1-\delta)} \left\{ \frac{\sigma_j}{|\mu_j|} \right\}^2 \right]^2 - \left[ \frac{a_0 \varepsilon}{|\mu_j|} \right]^2} \end{aligned} \quad (15)$$

The positive solution is in conflict with (12) so that we must seek only the smaller root with the negative sign which is indeed less than  $a_0(\mathbf{x})\varepsilon/|\mu_j(\mathbf{x})|$ . Thus, we have

$$\begin{aligned} \tau_j = & \frac{a_0 \varepsilon}{|\mu_j|} + \frac{1}{2(1-\delta)} \left\{ \frac{\sigma_j}{|\mu_j|} \right\}^2 \\ & - \sqrt{\left[ \frac{a_0 \varepsilon}{|\mu_j|} + \frac{1}{2(1-\delta)} \left\{ \frac{\sigma_j}{|\mu_j|} \right\}^2 \right]^2 - \left[ \frac{a_0 \varepsilon}{|\mu_j|} \right]^2} \end{aligned} \quad (16)$$

which is of course consistent with (12). We conclude from this analysis that the required  $\tau$  is given by

$$\tau = \min_{j=1,2,\dots,M} \{\tau_j\} \quad (17)$$

which assures the required leap condition (5) with a probability of at least  $\delta$ .

It is next of interest to compare the development here with the Gillespie–Petzold solution (hereafter referred to as GPS). Because (12) represents a common requirement, we only need consider the other alternative  $\varepsilon^2 a_0^2 / \sigma_j^2$  for  $\tau_j$  with that arising from expression 16. In passing, the Cauchy–Schwarz inequality (which states that the absolute value of the inner product between any two vectors is at most equal to the product of their norms) can be used on (7) to show that

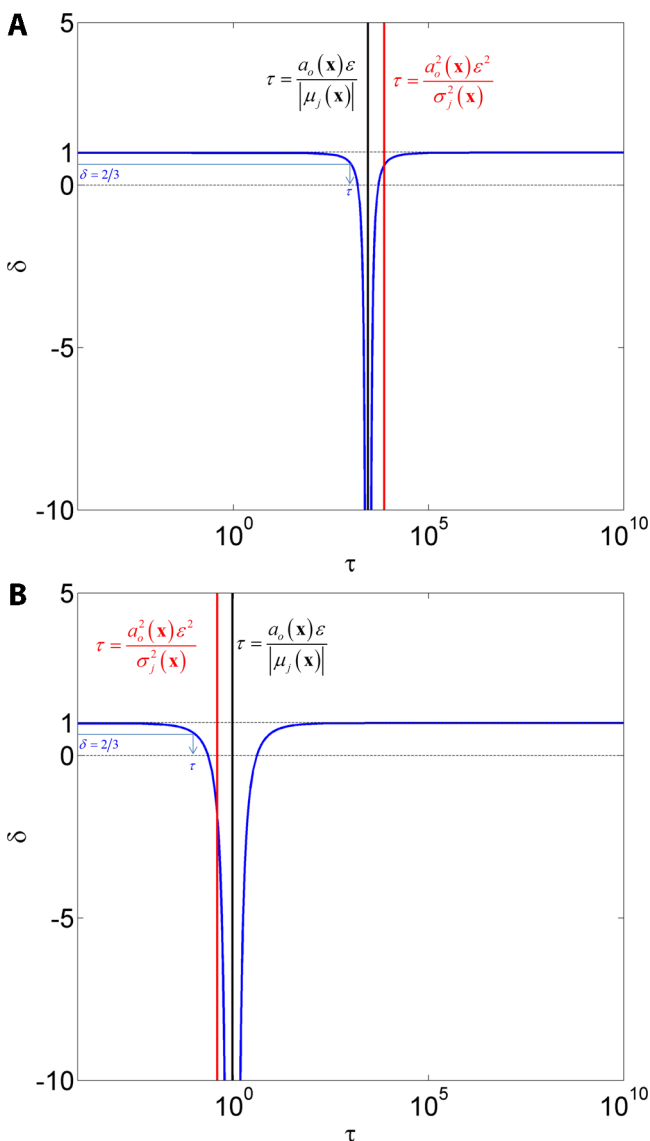
$$|\mu_j|^2 \leq \sigma_j^2 a_0 \quad (18)$$

which serves to further calibrate the plot of  $\delta$  versus  $\tau$ . We return to compare our prescription for  $\tau_j$  to that by GPS. Suppose a specific value of  $\varepsilon$  is chosen. If we begin with choosing

$$\varepsilon > \frac{\sigma_j^2}{a_0 |\mu_j|} \quad (19)$$

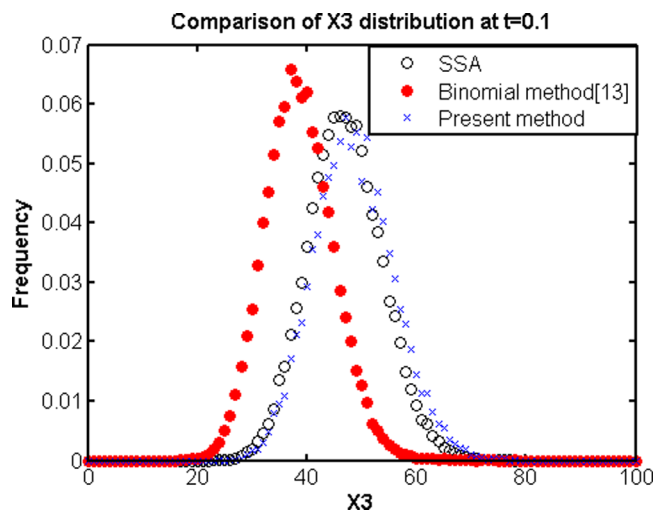
which implies that  $a_0 \varepsilon / |\mu_j| < \varepsilon^2 a_0^2 / \sigma_j^2$ , so that from GPS  $\tau_j = a_0 \varepsilon / |\mu_j|$ . For this  $\varepsilon$ , Figure 1a shows the layout of  $\delta$  and  $\tau$  and thus the extent to which they may be negotiated for an efficient simulation as guided by Chebyshev's inequality. In particular, it would appear that  $\tau_j$  should in fact be safely less than  $a_0 \varepsilon / |\mu_j|$  (instead of being equal to it as in GPS) so that  $\delta$  has a suitably positive value. Of course the resulting  $\tau$ -leap would be lowered, but sample paths generated from the process would more reliably satisfy the  $\tau$ -leap condition (5) which is indicative of higher simulation efficiency. It is noteworthy that the GPS choice is not backed by the probabilistic assurance of satisfying the  $\tau$ -leap condition (5). If, on the other hand,  $\varepsilon < \sigma_j^2 / a_0 |\mu_j|$ , the GPS prescription would call for  $\tau_j = a_0^2 \varepsilon^2 / \sigma_j^2 < \sigma_j^2 / |\mu_j|^2$ ; furthermore, it can be shown (see Appendix in the Supporting Information) that  $\tau_j > \tau_0$ , where  $\tau_0$  is obtained from (16) by setting  $\delta = 0$ ; thus the  $\delta - \tau$  scenario would change to that shown in Figure 1b. Thus, regardless of the choice of  $\tau_j$  in (9), Figure 2 locates it in the domain of negative  $\delta$  thus denying any probabilistic assurance that the  $\tau$ -leap condition 5 would be satisfied. It does not preclude, however, the generation of meaningful sample paths for the process.

As long as  $\tau$  is chosen to satisfy (12), there is no explicit restriction on the choice of  $\varepsilon$ , although its smallness would indeed govern the quality of the  $\tau$ -leap criterion. Throughout the simulation, both the method of Cao et al. as well as the current method encounter a number of defective steps at which the leap condition fails. As the number of these wasteful steps increases, the accuracy will be negatively influenced (i.e., it will require a higher number of sample paths to accomplish a fixed



**Figure 1.** Plots showing how  $\tau_j$  (indicated as  $\tau$ ) may be chosen for various values of  $\delta$ , the probability with which the  $\tau$ -leap condition is satisfied for the case  $\varepsilon > \sigma_j^2/a_0|\mu_j|$ . (a, top) The GPS solution for  $\tau_j$  appears at the asymptote! (b, bottom) The GPS obtains  $\tau_j$  in the region for which  $\delta < 0$ .

accuracy). Cao et al. proposed a procedure to generate leaps but never proved mathematically how it can be satisfied. Hence, some generated leaps might violate the leap condition badly and affect strongly the overall quality of the simulation. One way to avoid this problem and therefore to optimize the process is to increase the probability of getting a good selection of leaps. Two different ways which can accomplish this goal are as follows. Clearly, smaller values of  $\varepsilon$  will be progressively conservative and hence less effective in reducing computation time. On the other hand, if we fix  $\varepsilon$ , the use of the Chebyshev inequality may be regarded as a way to reduce the number of wasteful simulations that violate the  $\tau$ -leap criterion with the stipulated  $\varepsilon$ . From (17), it is transparent that, for fixed  $\varepsilon$ , as  $\delta \rightarrow 1$ ,  $\tau \rightarrow 0$ , which would be undesirable. Higher values of  $\tau$  can be chosen by negotiating the value of  $\delta$  to be suitably less than 1. While it may at first seem that the lower leaps in our approach may, not surprisingly, lead to more accurate solutions, it must be understood that the methodology lies in its efficiency in that



**Figure 2.** Comparison of the accuracy for the histogram between the two methods: binomial method<sup>13</sup> and the present method for  $\varepsilon = 0.02$ .

a solution of given accuracy is obtained with a smaller number of simulations. The proposed method thrives in the facility to manipulate both the parameters  $\varepsilon$  and  $\delta$  to promote efficiency. The computational demonstration to follow would of course confirm the foregoing observations.

The development of this work is readily adapted to the new  $\tau$ -leap method presented by Cao et al.,<sup>8</sup> who relate the relative change in propensity to that of the stochastic state variables involved. For example, if  $a_i(\mathbf{x})$  involves a first order reaction in the  $i$ th species (alone), we have

$$\frac{\Delta a_j}{a_j} = \frac{\Delta x_i}{x_i} \quad (20)$$

For the foregoing case, the  $\tau$ -leap condition in the  $j$ th propensity translates to  $|\Delta X_i| \leq \varepsilon x_i$  in state variable domain. For more general cases, Cao et al.<sup>8</sup> obtain the  $\tau$ -leap condition in the state variables as

$$|\Delta X_i| \leq \max \left\{ \frac{\varepsilon x_i}{g_i}, 1 \right\} \quad (21)$$

where  $g_i (\geq 1)$  represents the highest order of reaction with respect to species  $i$ . Note that (21) implies a change of at least one molecule; for an explanation of this and other aspects of this algorithm, the reader is referred to Cao et al.'s report.<sup>8</sup>

The application of our methodology leads to the following expression for  $\tau$  which will satisfy the leap condition for  $\mathbf{X}$  with probability of at least  $\delta$ .

$$\tau = \min_{i=1,2,\dots,N} \left\{ \frac{\alpha_i}{|\hat{\mu}_i|} + \frac{1}{2(1-\delta)} \left\{ \frac{\hat{\sigma}_i}{|\hat{\mu}_i|} \right\}^2 - \sqrt{\left[ \frac{\alpha_i}{|\hat{\mu}_i|} + \frac{1}{2(1-\delta)} \left\{ \frac{\hat{\sigma}_i}{|\hat{\mu}_i|} \right\}^2 \right]^2 - \left[ \frac{\alpha_i}{|\hat{\mu}_i|} \right]^2} \right\} \quad (22)$$

where  $\alpha_i \equiv \max\{\varepsilon x_i/g_i, 1\}$ ,  $\hat{\mu}_i(\mathbf{x}) \equiv \sum \nu_{ij} a_j(\mathbf{x})$ , and  $\hat{\sigma}_i^2(\mathbf{x}) \equiv \sum \nu_{ij}^2 a_j(\mathbf{x})$ .

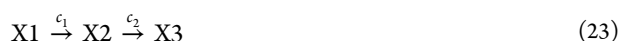
The efficacy of the proposed algorithm is shown by comparing it with that of Cao et al.<sup>8</sup>

The stochastic algorithm is usually employed to obtain the average behavior of the system for a chosen time interval by averaging several sample path simulations using the algorithm. The SSA<sup>9,12</sup> serves as a benchmark in evaluating the accuracy of any algorithm by comparing the histogram obtained by sample path averaging with that of SSA. In this regard, we use the same metric as that used by Cao and Petzold.<sup>13</sup> The comparison between the algorithms presented in this work with others has been made in two different ways. Thus, we compare the accuracies obtained for a fixed computation time, which is more focused on accuracy than computation time. Alternatively, where accuracy is prescribed, computation times can be compared for any two algorithms.

We employ two different examples for our demonstration. They are (i) the Schlogl model noted for its bistability and (ii) a linear consecutive reaction system. The algorithms used for comparison are those due to Cao's method,<sup>8</sup> Gillespie's midpoint Poisson  $\tau$ -leap method,<sup>11,12</sup> and the binomial leap method<sup>9,10</sup> subsequently improved by Peng et al.<sup>11</sup>

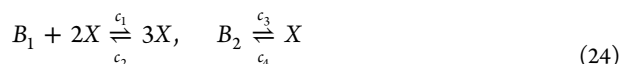
#### 4. EXAMPLES

**Example 1.** We consider the consecutive linear reaction system as in (23).



We apply our  $\tau$  selecting method with formula 22 for the generation of  $\tau$ . With rate constants  $c_1 = 1$  and  $c_2 = 1$  and initial conditions  $X1 = 10^4$ ,  $X2 = 1$  and  $X3 = 0$ , we calculate  $X3$  until time is equal to 0.1.

**Example 2.** Schlogl's chemical reaction model is shown in (24). This model is noted for its bistability.  $B_1$  and  $B_2$  are constants with their particle numbers as  $N_1$  and  $N_2$ , respectively.



The values of parameters, adapted from Cao et al.,<sup>8</sup> are  $c_1 = 3 \times 10^{-7}$ ,  $c_2 = 10^{-4}$ ,  $c_3 = 10^{-3}$ ,  $c_4 = 3.5$ ,  $N_1 = 1 \times 10^5$ , and  $N_2 = 2 \times 10^5$ . The initial condition of  $X$  is 250, and we simulate the system up to  $t = 4$ .

#### 5. DISCUSSION OF RESULTS

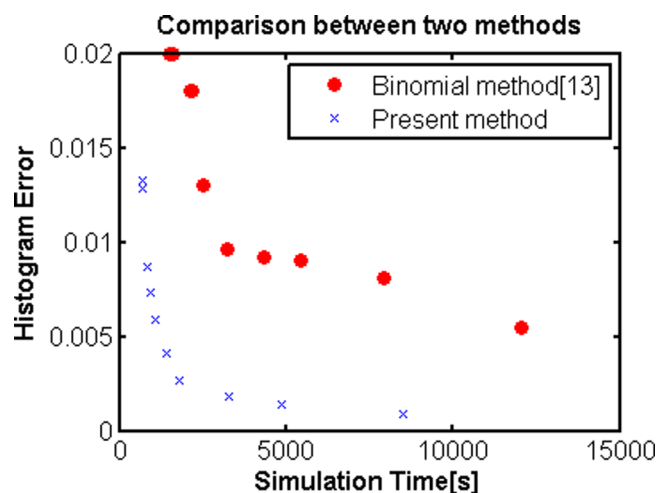
Figure 2 shows the probability density for  $X3$  at  $t = 0.1$  for SSA, the present method and the binomial method of Peng et al.,<sup>11</sup> using  $\varepsilon = 0.02$ . The higher and close proximity of the distribution by the present method to the SSA is clearly evident. Table 1 shows the number of sample paths used for averaging is only 10,000 for an accuracy notably higher than that from the algorithm under comparison. Although the computational time for each sample path for the present

**Table 1. Comparison of Results Generated from Two Different Methods: Binomial Method<sup>13</sup> and the Present Method for the Consecutive Linear Reaction<sup>a</sup>**

	binomial method	present method
no. of trajectories	30,000	10,000
$\varepsilon$	0.02	0.02
$\delta$		0.667
histogram error	0.0090	0.0018
total simulation time (s)	$5.417 \times 10^3$	$3.269 \times 10^3$ (40% faster)

<sup>a</sup>This table shows all of the parameters and results shown in Figure 2.

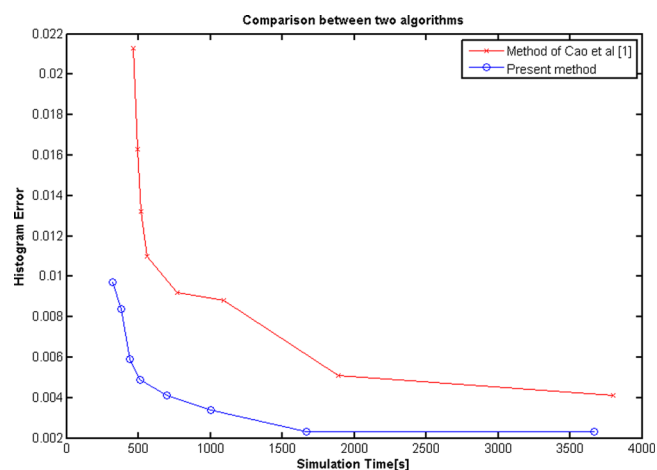
method is higher, a more accurate solution is obtained with 40% higher speed. Figure 3 shows the accuracy of the



**Figure 3.** Comparison of histogram error corresponding to different simulation times for the two methods being used to model the consecutive linear reaction system. The binomial method of Peng et al. is shown in red whereas ours is shown in blue.

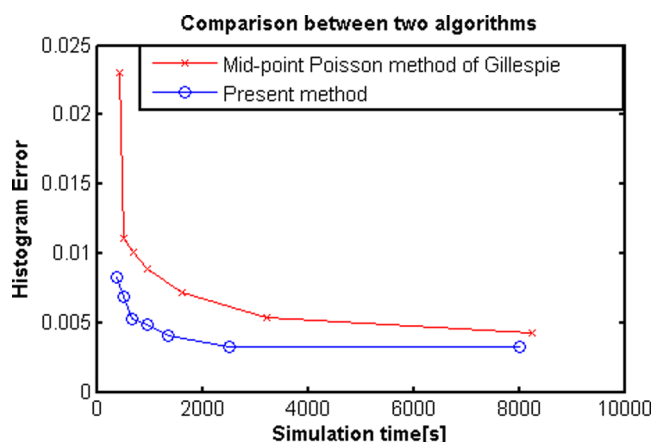
algorithms in comparison as a function of simulation time at different values of  $\varepsilon$ . Histogram distance errors are measured by  $10^6$  samples and  $10^5$  samples generated from the SSA method and the two  $\tau$ -leap methods, respectively, at different values of  $\varepsilon$ . For the same simulation time, the accuracy of the histogram is notably higher for the present method.

Figure 4 shows comparison of the histogram error as a function of simulation time for the present method against that



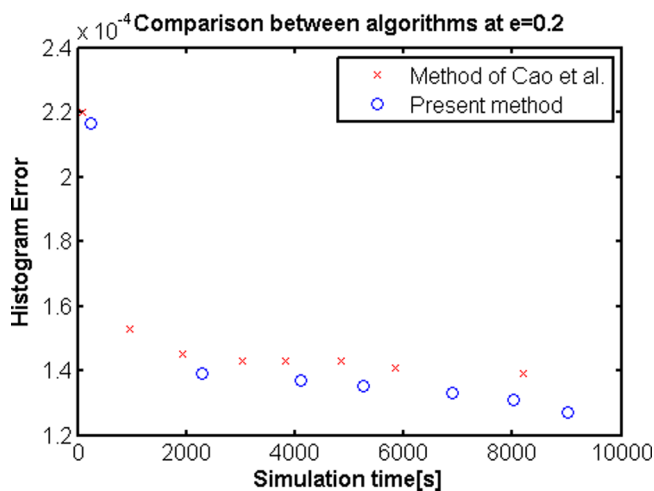
**Figure 4.** Comparison of histogram error with respect to different simulation times of two different methods: the method of Cao et al.<sup>8</sup> in red and ours in blue. This plot shows the accuracy level of the two methods for Schlogl's system at different  $\varepsilon$ .

by Cao et al.<sup>8</sup> for example 2. As in example 1, the notably higher accuracy of the current method is evident for this example also. Figure 5 compares the algorithm of Gillespie's midpoint Poisson method with the present method for the same system. In both Figures 4 and 5, the number of SSA samples is  $10^6$ . The comparison was made based on a fixed number of simulations for those methods at different values of  $\varepsilon$ . The simulation time for the former is considerably higher



**Figure 5.** Comparison of histogram error with respect to different simulation times of two different methods: Gillespie's midpoint Poisson method in red and ours in blue. This plot shows the accuracy level of the two methods for Schlogl's system at different  $\epsilon$ .

than that for the latter. When higher accuracy is needed, the computational time needed for Gillespie's method is substantially longer. Another verification for our claim is shown in Figure 6. In this figure,  $\epsilon$  is fixed to be 0.2, and results generated



**Figure 6.** Comparison of histogram error with respect to different simulation times of two different methods: Gillespie's midpoint Poisson method in red and ours in blue. This plot shows the accuracy level of the two methods for Schlogl's system at different numbers of samples at  $\epsilon = 0.2$ .

using two different methods at different numbers of samples were then compared. Similarly, we observed that our present algorithm improves both speed and accuracy. Although the computation time for each sample path is notably larger for the present work because of smaller time intervals, the higher efficiency of the resulting sample path allows accurate results with a significantly smaller number of sample paths as established in Figures 3, 4, and 5. Indeed the foregoing results bear testimony to notable improvements with our algorithm arising from the use of Chebyshev's inequality.

## 6. CONCLUSION

The algorithm presented in this work has attributes of efficiency earned from being able to account for the likelihood with which approximations for  $\tau$ -leap are satisfied by simulated sample

paths. The choice of this probability is a rational guideline to Monte Carlo simulations.

## ■ ASSOCIATED CONTENT

### Supporting Information

Text outlining the detailed proof of how picking  $\tau$ , according to (9) would lead to  $\delta < 0$  and a part which shows  $\tau_0 < \tau_j$ . This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [ramkrish@purdue.edu](mailto:ramkrish@purdue.edu).

### Present Address

<sup>†</sup>School of Chemical Engineering and Biotechnology, National Taipei University of Technology, Taipei 10608, Taiwan.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This paper is dedicated to Professor G. D. Yadav in recognition of his significant contributions to chemical engineering. This work was supported by National Institutes of Health Grant GM081888 (to Wei-Shou Hu).

## ■ REFERENCES

- (1) Feller, W. *An Introduction to Probability Theory and Its Applications*; John Wiley and Sons: New York, 1968; p 71.
- (2) Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361 Note: [Online] <http://pubs.acs.org/doi/abs/10.1021/j100540a008> (accessed Sep. 4, 2014).
- (3) Kendall, D. G. Random fluctuations in the age-distribution of a population whose development is controlled by the simple "birth-and-death" process. *J. R. Stat. Soc. Ser. B (Method.)* **1950**, *12*, 116–119 Note: [Online] <http://www.jstor.org/stable/2983985> (accessed Sep. 4, 2014).
- (4) Shah, B. Ph.D. Dissertation, Indian Institute of Technology in Kanpur, 1974.
- (5) Shah, B. H.; Ramkrishna, D.; Borwanker, J. D. Simulation of particulate systems using the concept of the interval of quiescence. *AIChE J.* **1977**, *23*, 897–904 Note: [Online] <http://onlinelibrary.wiley.com/doi/10.1002/aic.690230617/abstract> (accessed Sep. 4, 2014).
- (6) Ramkrishna, D. Analysis of population balance—IV: The precise connection between Monte Carlo simulation and population balances. *Chem. Eng. Sci.* **1981**, *36*, 1203–1209 Note: [Online] <http://www.sciencedirect.com/science/article/pii/0009250981850683> (accessed Sep. 4, 2014).
- (7) Anderson, D. F. Incorporating postleap checks in tau-leaping. *J. Chem. Phys.* **2008**, *128*, No. 054103, Note: [Online] <http://scitation.aip.org/content/aip/journal/jcp/128/5/10.1063/1.2819665> (accessed Sep. 4, 2014).
- (8) Cao, Y.; Gillespie, D. T.; Petzold, L. R. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.* **2006**, *124*, No. 044109, Note: [Online] <http://scitation.aip.org/content/aip/journal/jcp/124/4/10.1063/1.2159468> (accessed Sep. 4, 2014).
- (9) Tian, T.; Burrage, K. Binomial leap methods for simulating stochastic chemical kinetics. *J. Chem. Phys.* **2004**, *121*, 10356–10364 Note: [Online] <http://scitation.aip.org/content/aip/journal/jcp/121/21/10.1063/1.1810475> (accessed Sep. 4, 2014).
- (10) Chatterjee, A.; Vlachos, D. G.; Katsoulakis, M. A. Binomial distribution based  $\tau$ -leap accelerated stochastic simulation. *J. Chem. Phys.* **2005**, *122*, No. 024112, Note: [Online] <http://scitation.aip.org/content/aip/journal/jcp/122/2/10.1063/1.1833357> (accessed Sep. 4, 2014).

(11) Peng, X.; Zhou, W.; Wang, Y. Efficient binomial leap method for simulating chemical kinetics. *J. Chem. Phys.* **2007**, *126*, No. 224109, Note: [Online] <http://scitation.aip.org/content/aip/journal/jcp/126/22/10.1063/1.2741252> (accessed Sep. 4, 2014).

(12) Gillespie, D. T. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **2001**, *115*, No. 1716, Note: [Online] <http://scitation.aip.org/content/aip/journal/jcp/115/4/10.1063/1.1378322> (accessed Sep. 4, 2014).

(13) Cao, Y.; Petzold, L. Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems. *J. Comput. Phys.* **2006**, *212*, 6–24 Note: [Online] <http://www.sciencedirect.com/science/article/pii/S0021999105003074> (accessed Sep. 4, 2014).