# STOCHASTIC SIMULATION OF COUPLED CHEMICAL REACTIONS USING RECURSIVE METHODS

*Vibha Mane, Mónica F. Bugallo, Petar M. Djurić*

Department of Electrical and Computer Engineering
Stony Brook University, Stony Brook, NY 11794, USA
email: {vmane, monica, djuric} @ece.sunysb.edu

## ABSTRACT

In this paper, we present a new method for stochastic simulation of coupled chemical reactions. In this method we obtain recursive expressions for propagating the first two moments of the probability distributions over time. Its advantage over other simulation methods is that it does not require Monte Carlo simulations, and hence it performs several orders of magnitude faster than existing Monte Carlo methods. Simulation results are presented for some examples of coupled first-order reactions.

***Index Terms***— stochastic simulation, biological systems, biochemical processes.

## 1. INTRODUCTION

We are interested in predicting the time dependent behavior of *genetic networks* such as protein-DNA and DNA-DNA interactions, and *biochemical networks* such as interaction between proteins. There are two popular frameworks for modeling such networks. In the deterministic framework, we have a set of reaction rate equations with the unknowns being the molecular concentrations. In the stochastic framework, we have the chemical master equation with the unknowns being the probability distributions of the number of molecules. The deterministic framework is appropriate for systems with large number of molecules. However, in many inter- and intracellular biochemical reactions, some molecular species occur in very small numbers and therefore the deterministic approach would be inappropriate. Further, if a system operates close to unstable equilibria, stochastic fluctuations can be amplified [1], [9], which again makes the deterministic methods unsuitable. Hence, it has been widely accepted that stochastic approaches can produce more accurate results.

Within the stochastic framework, McQuarrie presented analytical solutions to the chemical master equation for a small number of first-order and second-order reactions using probability generating functions [10]. Laurenzi obtained the analytical solution for the reversible second-order reaction using the Laplace transform [8]. Zhang et al. obtained solutions for a system of first-order reactions using the Laplace transform method [14]. Darvey et al. [3] and Dunstan et al. [4] derived equilibrium distributions for a few second-order reactions. In 1976, Gillespie introduced the stochastic simulation algorithm (SSA), which is a Monte Carlo-based method for obtaining the molecular distributions in complex systems [6]. The SSA is an exact method and its advantage is that it is straightforward to implement. However, this algorithm simulates only one reaction per time step, and therefore it is computationally infeasible if some molecular species have large populations or there are many reactions in the system. Further, it requires the simulation of many realizations of the process to compute the probability distribution of the molecular species.

Several authors proposed accelerated versions of the SSA to reduce its computation time. In these algorithms, multiple firings of the reactions at each time step are allowed, and hence the resulting methods are much more efficient. Gillespie proposed the *Poisson distribution $\tau$-leap method* [7], Tian et al. [12] and Chatterjee et al. [2] presented the *binomial $\tau$-leap methods* and Gibson developed the *next reaction method* [5]. Recently, Pettigrew et al. [11] proposed *multinomial $\tau$-leap methods*, which are an extension of the binomial $\tau$-leap methods with several improvements.

In this paper we present a new method which provides stochastic solutions of complex biochemical systems without requiring the use of Monte Carlo simulations. The biochemical system is a Markovian process, where at each time step the state of the system is determined from the previous time step. Therefore, we obtain recursive expressions for the first two moments of the joint distributions which would update the system. Thus, rather than generating thousands of realizations of the system from which we can construct empirical distributions, we proceed by directly obtaining the moments of the joint distribution. The algorithm is scalable with the sizes of the molecular populations and the number of reactions.

The paper is organized as follows. In the next section we

This work has been supported by the National Science Foundation under Award CCF-0515246 and the Office of Naval Research under Award N00014-06-1-0012.

provide the problem statement. In Section 3 we present the philosophy of the recursive method and then explain how we apply it on a simple example of first-order reversible chemical reaction. In Section 4 we generalize it on a system composed of any number of coupled first-order reactions. Simulation results that show the performance of the method are presented in Section 5, and concluding remarks are made in Section 6.

## 2. PROBLEM STATEMENT

Consider a biochemical system composed of $N$ molecular species which take part in coupled biochemical reactions. We have knowledge of all the reactions and the associated rate constants. Further, we know the joint distribution of the initial number of molecules of the species in the system. Let the system be represented by the random vector $\underline{X}(t) = \begin{bmatrix} X_1(t) & X_2(t) & \dots & X_N(t) \end{bmatrix}$, where $X_i(t)$ represents the number of molecules of species $S_i$ at time instant $t$. We want to determine the first two moments of the joint distributions of $\underline{X}(t)$.

## 3. A RECURSIVE METHOD

We obtain recursive expressions for the first two moments of the distribution of the species, and therefore we refer to our method as a recursive method (RM). The main idea behind it is to decompose the studied biochemical system into a set of simpler systems, find recursive solutions for the simpler systems and combine them so that they provide a solution for the complete system. We assume that during short time intervals, each of the simple systems evolve independently of the others and that any molecule in the system undergoes at most one reaction. We illustrate the method with an elementary example.

### 3.1. Example - First Order Reversible Reaction

Consider the reversible reaction

$$S_1 \underset{c_{21}}{\overset{c_{12}}{\rightleftharpoons}} S_2 \tag{1}$$

where $S_1$ and $S_2$ denote species, and $c_{12}$ and $c_{21}$ are stochastic rate constants [13]. The reaction simply means that a molecule of species $S_1$ can become a molecule of species $S_2$ and vice versa, a molecule of species $S_2$ can become a molecule of species $S_1$.

The first step of the method is to decompose the system by using elementary reactions. For the addressed system, it is clear that the elementary reactions are

$$S_1 \overset{c_{12}}{\rightarrow} S_2 \tag{2}$$

$$S_2 \overset{c_{21}}{\rightarrow} S_1. \tag{3}$$

We note that the probability that a molecule of $S_1$ converts to a molecule of $S_2$ in some time interval $\Delta t$ is given by

$$p_{12} = 1 - e^{-c_{12}\Delta t}.$$

The expression for the probability that a molecule of $S_2$ becomes a molecule of $S_1$ is analogous.

Suppose next that $X_1(t)$ and $X_2(t)$ are the number of $S_1$ and $S_2$ molecules at time $t$, respectively. Suppose also that $\mu_1(t)$ and $\mu_2(t)$ are the means and $\sigma_1^2(t)$ and $\sigma_2^2(t)$ are the variances of the number of $S_1$ and $S_2$ molecules at time $t$. Here we note that the covariance of $S_1$ and $S_2$ in this example is $-\sigma_1(t)\sigma_2(t)$. We now want to update these means and variances to $\mu_1(t+\Delta t)$, $\mu_2(t+\Delta t)$, $\sigma_1^2(t+\Delta t)$, and $\sigma_2^2(t+\Delta t)$. To that end, we first find the conditional means and variances, and then integrate out the conditioning variables. The conditioning variables are the number of molecular species at time $t$.

In our work we use the analytical results reviewed by McQuarrie in [10]. For example, by summing up the contributions from (2) and (3), we obtain the following expression for the conditional expectations of $X_1(t+\Delta t)$ and $X_1^2(t+\Delta t)$:

$$\begin{aligned} E\left(X_1(t+\Delta t) \,|\, X_1(t), X_2(t)\right) \;=\; & X_1(t)\, e^{-c_{12}\Delta t} \\ & + X_2(t)\, \left(1 - e^{-c_{21}\Delta t}\right) \end{aligned} \tag{4}$$

and

$$\begin{aligned} & E\left(X_1^2(t+\Delta t) \,|\, X_1(t), X_2(t)\right) \;= \\ & X_1(t)\, e^{-c_{12}\Delta t}\, \left(1 - e^{-c_{12}\Delta t}\right) \\ & + X_2(t)\, e^{-c_{21}\Delta t}\, \left(1 - e^{-c_{21}\Delta t}\right) \\ & + E^2\left(X_1(t+\Delta t) \,|\, X_1(t), X_2(t)\right). \end{aligned} \tag{5}$$

After integrating the above conditionals over the conditioning variables, we obtain the mean and the variance of the number of $S_1$ molecules

$$\mu_1(t+\Delta t) = \; \mu_1(t)\, e^{-c_{12}\Delta t} + \; \mu_2(t)\, \left(1 - e^{-c_{21}\Delta t}\right) \tag{6}$$

and

$$\begin{aligned} \sigma_1^2(t+\Delta t) = \; & E(X_1^2(t+\Delta t)) - \mu_1^2(t+\Delta t) \\ = \; & \mu_1(t)\, e^{-c_{12}\Delta t}\, \left(1 - e^{-c_{12}\Delta t}\right) \\ & + \mu_2(t)\, e^{-c_{21}\Delta t}\, \left(1 - e^{-c_{21}\Delta t}\right) \\ & + \sigma_1^2(t)\, \left(e^{-c_{12}\Delta t}\right)^2 + \; \sigma_2^2(t)\, \left(1 - e^{-c_{21}\Delta t}\right)^2 \\ & - 2\sigma_1(t)\sigma_2(t)\, e^{-c_{12}\Delta t}\, \left(1 - e^{-c_{21}\Delta t}\right). \end{aligned} \tag{7}$$

Similar expressions can be obtained for the mean and variance of the number of $S_2$ molecules.

## 4. A RECURSIVE METHOD - GENERAL EXPRESSION

Next, we consider a spatially homogeneous system, where the molecules interact through first-order reactions. Let there be $N$ molecular species $\left(S_1, S_2, \dots, S_N\right)$, and as before, let $\underline{X}(t)$ describe the state of the system at time $t$. Let $p_{ji}$ be the transition probability from species $j$ to species $i$. We have

the following expressions for the conditional expectations for $X_i(t+\Delta t)$ and $X_i^2(t+\Delta t)$:

$$E\left(X_i(t+\Delta t)\,|\,X_1(t),\ldots,X_N(t)\right) \;=\; \sum_{j=1}^{N} X_j(t)\;p_{ji} \quad (8)$$

$$E\left(X_i^2(t+\Delta t)\,|\,X_1(t),\ldots,X_N(t)\right) \;=\; \sum_{j=1}^{N} X_j(t)\;p_{ji}\,(1-p_{ji}) + E^2\left(X_i(t+\Delta t)\,|\,X_1(t),\ldots,X_N(t)\right) \quad (9)$$

In the above expressions, the transition probabilities $p_{ji}$ , for the reaction $S_j \overset{c_{ji}}{\rightarrow} S_i$, in the time interval $\Delta t$ are given by

$$p_{ji} = \frac{c_{ji}}{c_j}(1-e^{-c_j\Delta t}), \qquad j \neq i \quad (10a)$$

$$p_{jj} = e^{-c_j\Delta t} \quad (10b)$$

where

$$c_j = \sum_{i=1}^{N} c_{ji}. \quad (11)$$

Again, after integrating the expressions in (8) and (9) over the conditioning variables we obtain the mean and the variance of the number of $S_i$ molecules,

$$\mu_i(t+\Delta t) \;=\; \sum_{j=1}^{N} \mu_j(t)\;p_{ji} \quad (12)$$

$$\sigma_i^2(t+\Delta t) \;=\; \sum_{j=1}^{N} \mu_j(t)\;p_{ji}\,(1-p_{ji}) + \sum_{j=1}^{N} \sigma_j^2(t)\;p_{ji}^2 \;+\; 2\sum_{\substack{j=1\\ j\neq i}}^{N} \rho_{ij}(t)\sigma_i(t)\sigma_j(t)p_{ji}\,p_{ii} \quad (13)$$

where $\rho_{ij}(t)$ is the correlation coefficient of $X_i(t)$ and $X_j(t)$. This coefficient was present in (7) too, and it had a value equal to $-1$. In the next subsection, we describe how we compute the correlation coefficient.

### 4.1. Computing the correlation coefficient

Given any two molecules $S_i$ and $S_j$, we say that they are *linked* if $p_{ij}$ and/or $p_{ji}$ is different from zero. We define the sets $C_i$ and $C_j$ as sets with elements that represent molecules that are linked to $S_i$ and $S_j$, respectively, but excluding $S_i$ and $S_j$. If both $C_i$ and $C_j$ are empty sets, as in the reaction $S_1 \rightleftharpoons S_2$, then $\rho_{ij}(t) = -1$. As described by Zhang et al. [14], if a system starts with a single source initially[1], the joint distribution at a later time is multinomial, and if a system starts with multiple sources, the joint distributions are convolutions of multinomials. Here we use this fact to obtain expressions for $\rho_{ij}(t)$ with one source. The method can be generalized for more than one source and the expressions for such scenarios will be presented elsewhere.

With a single source, the marginal distribution $p(x_i(t), x_j(t))$ of any two species is trinomial. We used the moment generating functions to obtain the following expression for the correlation coefficient:

$$\rho_{ij}(t) = -\sqrt{\frac{\pi_i(t)\pi_j(t)}{(1-\pi_i(t))(1-\pi_j(t))}} \quad (14)$$

where $\pi_i(t)$ and $\pi_j(t)$ are the probabilities of being in states $i$ and $j$ at time $t$, respectively. These probabilities are approximately given by

$$\pi_i(t) = \frac{X_i(t)}{X_T(t)} \quad (15)$$

where $X_T(t)$ is the total molecular population, i.e.,

$$X_T(t) = X_T(0) = \sum_{j=1}^{N} X_j(0). \quad (16)$$

Note that (14) is valid for the case of a single source. If there are two or more sources, $\rho_{ij}(t)$ has different forms.

## 5. SIMULATION RESULTS

In this section we present results obtained by the proposed method, and we compare it with the SSA method. Consider the reaction

$$S_1 \underset{c_{21}}{\overset{c_{12}}{\rightleftharpoons}} S_2 \underset{c_{32}}{\overset{c_{23}}{\rightleftharpoons}} S_3. \quad (17)$$

The parameters were $X_1(t=0) = 100$, $X_2(t=0) = 0$, $X_3(t=0) = 10000$, $c_{12} = 0.02\text{s}^{-1}$, $c_{21} = 0.01\text{s}^{-1}$, $c_{23} = 0.04\text{s}^{-1}$, $c_{32} = 0.01\text{s}^{-1}$, and $\Delta t = 0.1$s. We applied the SSA method and simulated $4000$ realizations, which were used for constructing empirical distributions. Fig. 1 shows the results of the obtained distributions for the three species at $t = 10$s. The two methods are in very good agreement.

We also compared the computing times of SSA and our method. The simulation times were compared on a 2.8 GHz Pentium machine and we found that our method provides results below 0.01s regardless of the sizes of the populations, whereas SSA needed several minutes. The advantage of our method in computing time will be even more impressive if the number of molecules and/or the number of reactions in the system increases. This advantage can also be observed when the method is compared to the faster $\tau$-leap class of methods [2], [11], [12].

[1] All the species populations have zero initial values except for one of them which represents a source.

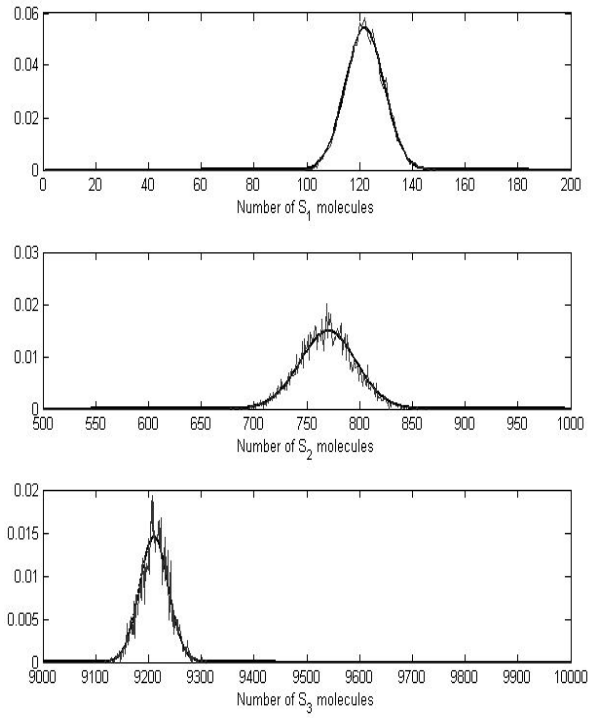


**Fig. 1**. Comparison of results between the RM (solid line) and the SSA (noisy plot) methods, for the reaction in equation (17). The plot shows the distribution of the number of $S_1$, $S_2$ and $S_3$ molecules at $t = 10$s. The used parameters are given in Section 5.

## 6. CONCLUSION

All of the existing stochastic approaches to simulation of complex chemical systems are based on Monte Carlo methods. The advantage of the Monte Carlo methods is that SSA is an exact method and this and some of the accelerated versions of the SSA are straightforward to implement. However, the computation times for these methods can be prohibitively long for large systems. In this paper we propose a recursive mechanism for computing the first two moments of the joint distributions. Simulation results were presented for a few first-order systems. We made also comparisons with existing methods and found that the proposed method yields accurate results with significant savings in computer time. The goal is to extend the recursive method to second order systems, and develop a general recipe which would facilitate its implementation.

## 7. REFERENCES

[1] J. M. Bower and H. Bolouri, Eds., *Computational Modeling of Genetic and Biochemical Networks*, MIT Press, 2004.

[2] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, "Binomial distribution based $\tau$-leap accelerated stochastic simulation," *The Journal of Chemical Physics*, vol. 122, pp. 024112, 2005.

[3] I. G. Darvey, B. W. Ninham, and P. J. Staff, "Stochastic models for second order chemical reaction kinetics," *The Journal of Chemical Physics*, vol. 45, pp. 2145–2155, 1966.

[4] F. D. J. Dunstan and J. F. Reynolds, "Normal approximations for distributions arising in the stochastic approach to chemical reaction kinetics," *Journal of Applied Probability*, vol. 18, no. 1, pp. 263–267, 1981.

[5] M. A. Gibson and J. Bruck, "Efficient exact stochastic simulation of chemical systems with many species and many channels," *The Journal of Physical Chemistry*, vol. 104, pp. 1876–1889, 2000.

[6] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *Journal of Computational Physics*, vol. 22, pp. 403–434, 1976.

[7] D. T. Gillespie, "Approximate accelerated stochastic simulation of chemically reacting systems," *The Journal of Chemical Physics*, vol. 115, pp. 1716–1733, 2001.

[8] I. J. Laurenzi, "An analytical solution of the stochastic master equation for reversible biomolecular reaction kinetics," *The Journal of Chemical Physics*, vol. 113, pp. 3315–3322, 2000.

[9] H. H. McAdams and A. Arkin, "Stochastic mechanisms in gene expression," *Proceedings of the National Academy of Sciences*, vol. 94, pp. 814–819, 1997.

[10] D. A. McQuarrie, "Stochastic approach to chemical kinetics," *Journal of Applied Probability*, vol. 4, pp. 413–478, 1967.

[11] M. F. Pettigrew and H. Resat, "Multinomial tau-leaping method for stochastic kinetic simulations," *The Journal of Chemical Physics*, vol. 126, pp. 084101, 2007.

[12] T. Tian and K. Burrage, "Binomial leap methods for simulating stochastic chemical kinetics," *The Journal of Chemical Physics*, vol. 121, pp. 10356–10364, 2004.

[13] D. J. Wilkinson, *Stochastic Modeling for System Biology*, Chapman & Hall/CRC, 2006.

[14] X. Zhang, K. De Cock, M. F. Bugallo, and P. M. Djurić, "A general method for the computation of probabilities in systems of first order chemical reactions," *The Journal of Chemical Physics*, vol. 122, no. 10, pp. 104101 (8), 2005.