Journal of Numerical Mathematics and Stochastics*, 1** (1) : 85-101, 2009 © ***JNM@S
http://www.jnmas.org/jnmas1-9.pdf Euclidean Press, LLC
Online: ISSN 2151-2302

Effective Stochastic Simulation Methods for Chemical Reaction Systems

T. TIAN

Department of Mathematics, University of Glasgow, Glasgow G12 8QW, United Kingdom,
E-mail: t.tian@maths.gla.ac.uk

Abstract. *Recent studies through biological experiments have indicated that noise plays a very important role in determining the dynamic behaviour of biological systems. The advances in stochastic modelling of genetic regulatory networks and cell signalling transduction pathways have proposed chemical reaction systems in which some key reactants have small copy numbers. Over the last years extensive research has been carried out to develop effective τ-leap methods and multiscale simulation methods to reduce the huge computational time of the stochastic simulation algorithm. This paper gives a review of the recent progress in the stochastic simulation methods for chemical reaction systems.*

Key words: Biological System, Discrete Stochastic Simulation, Stochastic Differential Equation, Time Delay.

AMS Subject Classifications : 60H35, 65C05, 92C45

1. Introduction

There is a growing body of experimental evidence which suggests that noise plays a very important role in genetic regulation [22,37]. Experimental studies have demonstrated that noise in cellular processes may result from the small copy number of molecular species, intermittent gene activity and fluctuations of experimental conditions [38,55]. It has been proposed that noise in the form of random fluctuations arises in biological networks in one of two ways: namely as internal (intrinsic) noise or external (extrinsic) noise [50]. The internal noise is mainly derived from the probabilistic events of biochemical reactions in the system due to the small copy numbers of certain key molecular species. External noise mainly refers to the environmental fluctuations or the noise propagation from the upstream biological pathways. Empirical discoveries have stimulated explosive research interests in developing stochastic models for a wide range of biological systems including genetic regulatory networks, cell signalling transduction pathways and metabolic pathways [3, 10 39, 64, 69].

The stochastic simulation algorithm (SSA) represents a discrete modelling approach and an essentially exact procedure for simulating the time evolution of a well-stirred reaction system [25]. However, the large computational time of the SSA is a barrier in the applications of this method to large-scale biological systems. To reduce the computational time of the SSA, Gillespie proposed the Poisson τ-leap method which allows for more reactions to fire in a relatively larger time step [27]. In recent years there have been growing research interests in developing effective simulation methods to address the issues arising from the applications of the Poisson τ-leap methods [28, 49]. These issues include, for example, the stiffness in chemical reaction systems where there are well-separated "fast" and "slow" time scales present, possible negative molecular numbers in stochastic simulation, the simulation accuracy and the selection of the optimal leap size. In addition, sophisticated computer software systems have been developed not only for computational scientists but also for biological researchers.

This paper gives a review of the recent progress in the stochastic simulation methods for chemical reaction systems. After the introduction of the exact simulation methods in Section 2, Section 3 presents various τ-leap methods and discusses the issues arising from the application of the τ-leap methods. The Langevin approach and multiscale simulation methods will be addressed in Section 4. From Section 5, we concentrate on the simulation methods for chemical systems with time delay. We will discuss the delay stochastic simulation algorithm in Section 5, the delay Langevin approach in Section 6 and the delay τ-leap methods in Section 7.

2. Stochastic Simulation Algorithm

It is assumed that a chemical reaction system is a well-stirred mixture at constant temperature in a fixed volume Ω. This mixture consists of N molecular species $\{S_1, \ldots, S_N\}$ that chemically interact through M reaction channels $\{R_1, \ldots, R_M\}$. The dynamic state of this system is denoted as $\mathbf{x}(t) \equiv (x_1(t), \ldots, x_N(t))^\top$, where $x_i(t)$ is the molecular number of species S_i at time t. For each reaction channel R_j ($j = 1, \ldots, M$), a propensity function $a_j(\mathbf{x})$ is defined for a given state $\mathbf{x}(t) = \mathbf{x}$ and the value of $a_j(\mathbf{x})dt$ represents the probability that one reaction R_j will fire somewhere inside Ω in the infinitesimal time interval $[t, t + dt)$. In addition, a state change vector ν_j is defined to characterise the change of molecular numbers due to reaction R_j. The element ν_{ij} of ν_j represents the change in the copy number of species S_i due to reaction R_j. The $N \times M$ matrix ν with elements ν_{ij} is called the stoichiometric matrix.

The SSA is a statistically exact procedure for generating the time and index of the next occurring reaction in accordance with the current values of the propensity functions. There are several forms of this algorithm. The widely used direct method works in the following manner.

Method 1: **The direct method** [25].

1 Calculate the values of propensity functions $a_j(\mathbf{x})$ based on the system state $\mathbf{x}$ at time t and $a_0(\mathbf{x}) = \sum_{j=1}^{M} a_j(\mathbf{x})$.

2 Generate a sample r_1 of the uniformly distributed random variable $\mathbf{U}(0,1)$, and determine the time of the next reaction

$$\mu = \frac{1}{a_0(\mathbf{x})} \ln\left(\frac{1}{r_1}\right). \tag{1}$$

3 Generate another sample r_2 of $\mathbf{U}(0,1)$ to determine the index k of the next reaction occurring in $[t, t+\mu]$,

$$\sum_{j=1}^{k-1} a_j(\mathbf{x}) < r_2\, a_0(\mathbf{x}) \le \sum_{j=1}^{k} a_j(\mathbf{x}). \tag{2}$$

4 Update the state of the system by

$$\mathbf{x}(t+\mu) = \mathbf{x}(t) + \nu_k. \tag{3}$$

5 Go to Step 1 if $t+\mu \le T$, where T is the end time point. Otherwise, the system state at T is $\mathbf{x}(T) = \mathbf{x}(t)$.

Another exact method is the first reaction method which uses M random numbers $\{r_j\}$ at each time step to determine the possible reaction time $\mu_j = -\ln(r_j)/a_j(\mathbf{x})$ of each reaction channel [24]. The reaction firing in the next step R_k is that requiring the smallest reaction time, namely $\mu_k = \min_{1\le j\le M}\{\mu_j\}$. Compared with the direct method, the first reaction method is not effective because $M-1$ random numbers are abandoned at each time step. To improve the efficiency, Gibson and Bruck [23] proposed the next reaction method by recycling the generated random numbers. After the system is updated according to a selected reaction channel R_k, the step size of the selected channel R_k is determined by a newly generated random number. However, the putative step size of the other $M-1$ channels will be updated based on the stepsize of this channel at the previous step and values of the propensity function at these two steps. In addition, a so-called dependency graph was designed to reduce the computing time of propensity functions. Numerical results suggested that the next reaction method is effective for simulating systems with many species and reaction channels.

3. Accelerating Tau-Leap Methods

Since the SSA may be very computationally inefficient, Gillespie proposed the Poisson τ-leap methods in order to improve the efficiency of the SSA while maintaining acceptable losses in accuracy [27]. The idea of the τ-leap methods is to take a larger time step and allow for more reactions to take place in that step. In the Poisson τ-leap method, the number of times that the reaction channel R_j will fire in the time interval $[t, t+\tau]$ is approximated by a Poisson random variable $\mathcal{P}(a_j(\mathbf{x})\tau)$ $(j = 1, \ldots, M)$ with mean $a_j(\mathbf{x})\tau$. Here the leap size τ should satisfy the Leap Condition: a temporal leap by τ will result in a state change λ such that for every reaction channel R_j, the change $|a_j(\mathbf{x}+\lambda) - a_j(\mathbf{x})|$ is "effectively infinitesimal" [27]. This method is given below.

Method 2: **The Poisson τ-leap method** [27].
1 Calculate the values of the propensity functions $a_j(\mathbf{x})$ based on the system state $\mathbf{x}$ at time t.
2 Choose a value for the leap size τ that satisfies the Leap Condition.
3 Generate a sample of the Poisson random variable $\mathcal{P}(a_j(\mathbf{x})\tau)$ for each reaction channel $(j = 1, \ldots, M)$.
4 Perform the update of the system by

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \sum_{j=1}^{M} \nu_j \mathcal{P}(a_j(\mathbf{x})\tau). \tag{4}$$

A major step of the Poisson τ-leap method is to choose an appropriate stepsize that satisfies the Leap condition. Gillespie first proposed a method to determine the leap size τ [27. In this method, the expected change of each propensity function $a_j(\mathbf{x})$ during $[t, t+\tau]$ should be bounded by $\epsilon a_0(\mathbf{x})$ with a given error control parameter ϵ,

$$|a_j(\mathbf{x}+\lambda) - a_j(\mathbf{x})| \leq \epsilon a_0(\mathbf{x}), \tag{5}$$

where λ is the expected net change of the state $\mathbf{x}$ in $[t, t+\tau]$, which is calculated by $\lambda = \tau \sum_{j=1}^{M} \nu_j a_j(\mathbf{x})$. Later, more sophisticated methods have been proposed either to select the optimal leap size or to avoid the possible negative molecular numbers in simulation [30, 17]. For example, by considering the mean and standard deviation of the expected change in the propensity functions, the stepsize is determined by [30],

$$\tau = \min_{j\in[0,M]} \left\{ \frac{\epsilon a_0(\mathbf{x})}{|\mu_j(\mathbf{x})|}, \frac{\epsilon^2 a_0^2(\mathbf{x})}{\sigma_j^2(\mathbf{x})} \right\} \tag{6}$$

where $\mu_j(\mathbf{x}) = \sum_{k=1}^{M} f_{jk}(\mathbf{x}) a_k(\mathbf{x})$, $\sigma_j^2(\mathbf{x}) = \sum_{k=1}^{M} f_{jk}^2(\mathbf{x}) a_k(\mathbf{x})$, and $f_{jk}(\mathbf{x}) = \sum_{i=1}^{N} \frac{\partial a_j(\mathbf{x})}{\partial x_i} \nu_{ik}$. This method is an extension of the method (5) that only considered the mean of the expected change of the propensity functions. It is worthy to note that the leap size is a pre-selected deterministic value and is determined by the error control parameter ϵ. Similar to many other numerical methods, the leap size τ is related to the balance between the computational efficiency and simulation accuracy. In addition, our simulation results suggested that the computing time for selecting the leap size is about a half of the total computing time when method (6) was used in the binomial τ-leap method [65].

Since the τ-leap method uses the initial state $\mathbf{x}$ at time t to approximate the states in the time interval $[t, t+\tau]$, the simulation accuracy can be improved by computing a better approximation to the states in the given time interval. For example, by using an estimation of the system state at the midpoint $t+\tau/2$, the midpoint τ-leap method is given by [27

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \sum_{j=1}^{M} \nu_j P(a_j(\bar{\mathbf{x}})\tau).$$

where $\bar{\mathbf{x}}$ is the predicted system state at the midpiont

$$\bar{\mathbf{x}} = \mathbf{x}(t) + \sum_{j=1}^{M} \nu_j \left[a_j(\mathbf{x}) \frac{\tau}{2} \right].$$

The Poisson τ-leap method, the midpoint τ-leap method, and the implicit τ-leap method [56] are special cases of the Poisson Runge-Kutta methods (PRK) [12]. We also derived the Heun PRK and the R2 PRK method [12]. It was intended to design more accurate methods by choosing the free parameters in the PRK methods. In addition, these leap methods were placed in the context of the numerical methods for the solution of stochastic differential equations

(SDEs) driven by Poisson noise [11]. However, a fundamental issue of the discrete stochastic simulation is the convergence property of the simulation methods. Li addressed this issue by proving that a modified explicit τ-leap method has the mean-square strong convergence order 1/2 and weak convergence order 1 [42]. In addition, the higher order τ-leaping methods have been constructed by adding a random correction to the primitive τ-leaping schemes in each time step [36].

A number of approaches have been proposed to improve the efficiency and/or accuracy of the Poisson τ-leap methods. The implicit τ-leap method was proposed to address the stiffness in chemical reaction systems [56]. A composite approach was proposed later by combining the explicit and implicit τ-leap methods [18]. The algorithm automatically selects between the explicit and implicit τ-leap methods. In addition, the modified Poisson τ-leap method introduced an additional control parameter in order to avoid possible negative molecular numbers in stochastic simulation [16]. More accurate τ-leap methods have been proposed either by calculating the mean and variance of the propensities during the leap [70] or by adding a random correction to the primitive tau-leaping schemes [36].

Since the samples of a Poisson random variable are unbounded, negative molecular numbers may be obtained if a reactant has small molecular number and a propensity function involving that species has a large value. There are two possible ways of obtaining negative molecular numbers in stochastic simulation [62]. First, the generated sample of reaction number is greater than one of the molecular numbers of the reaction channel. In the second case, a species involves a number of reaction channels and the total reaction number of these channels is greater than the copy number of that species, though the reaction number of each channel is smaller than the copy number of that species.

For tackling the problem of negative numbers, binomial random variables were introduced to avoid the possible negative numbers of the first case by restricting the maximal reaction number [62, 19]. In the binomial τ-leap method, the reaction number of channel R_j is defined by a sample value of the binomial random variable $\mathcal{B}(N_j, a_j(\mathbf{x})\tau/N_j)$ under the condition $0 \le a_j(\mathbf{x})\tau/N_j \le 1$. The maximal possible reaction number N_j has been defined for the widely used three types of elementary reactions. In addition, a sampling technique was designed for sampling the total reaction number of a group of reaction channels in order to avoid the negative numbers of the second case [62]. The binomial τ-leap method is given below.

Method 3: **The binomial τ-leap method** [62].

0 Define the maximal possible reaction number N_j for each reaction channel. If a species involves two or more reaction channels $\{R_{j1}, \dots, R_{jk}\}$, define a maximal total reaction number N_{jk} for these reaction channels.

1 Calculate the values of propensity functions $a_j(\mathbf{x})$ based on the system state $\mathbf{x}$ at time t.

2 Use a method to determine the value of leap size τ. Check the stepsize conditions $0 \le a_j(\mathbf{x})\tau/N_j \le 1$ of the binomial random variables. If necessary, reduce the stepsize τ to satisfy these conditions.

3.1 Generate a sample $\mathcal{B}_j$ of the binomial random variable $\mathcal{B}(N_j, a_j(\mathbf{x})\tau/N_j)$ for reaction channels in which species involve one single reaction.

3.2 When a species involves two or more reaction channels, generate a total reaction number $\mathcal{B}(N_{jk}, \sum_{l=1}^{k} a_{jl}(\mathbf{x})\tau/N_{jk})$ for these reaction channels and then generate the reaction number $\mathcal{B}_j$ for each reaction channel in this group.

4 Perform the updates of the system by

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \sum_{j=1}^{M} \nu_j \mathcal{B}_j. \tag{7}$$

A number of approaches have been proposed to improve the performance of the binomial τ-leap method. A modified binomial τ-leap method was introduced by considering the maximal existing population in a leap [51]. In addition, the binomial distribution was extended to the multinomial distribution in order to estimate the numbers of reactions occurring simultaneously more accurately [52]. Recently, a generalized binomial τ-leap method was proposed to overcome some of the difficulties associated with the binomial τ-leap methods [41]. In addition, the binomial τ-leap method was applied to simulate chemical systems with time delay [41] and spatio-temporal reaction systems [46].

In the τ-leap methods, it was assumed that, during a pre-selected deterministic time step τ, the reaction number of each channel is a sample of a Poisson/binomial random variable. Another major type of leap methods is the k_α-leap method [27]. In an implementation of the k_α-leap method, which is called the R-leap method [4], it was proposed to select a pre-defined number of firings L that may span several reaction channels. Then the leap step τ_L of these L reactions is a sample of the gamma random variable $\Gamma(L, 1/a_0(\mathbf{x}))$. Over the time interval $[t, t+\tau_L]$, the number of firings of each reaction channel K_j (satisfying $\sum_{j=1}^{M} K_j = L$) follows the correlated binomial distributions. A number of techniques have been proposed in the R-leap method to determine the total reaction number L and to sample the firing number K_j of each reaction channel [4]. Another implementation of the k_α-leap method, which is called the K-leap method, was also proposed to achieve the computing efficiency over the exact SSA [15].

4. Chemical Langevin Equation and Multiscale Simulation Methods

The chemical master equation is a time evolution equation associated with the probability of the system at a given state. Instead of simulating the trajectories of a discrete system, the chemical master equation uses the continuous differential equations to accurately describe the stochastic dynamics, given by

$$\frac{\partial}{\partial t} P(\mathbf{x}, t) = \sum_{j=1}^{M} a_j(\mathbf{x} - \nu_j) P(\mathbf{x} - \nu_j, t) - \sum_{j=1}^{M} a_j(\mathbf{x}) P(\mathbf{x}, t), \tag{8}$$

where $P(\mathbf{x}, t)$ is the probability of the system at the state $\mathbf{x}(t) = \mathbf{x}$. In general these parabolic partial differential equations are difficult to solve either analytically or numerically. Therefore, stochastic methods discussed in Sections 2 and 3 have been widely used to simulate the trajectories of reaction systems and statistical properties of the system have been derived from a large number of simulations. However, when the system has a finite number of distinct population vectors or the number of species in the system is not large, the chemical master equation can be solved analytically or numerically. The approximation methods includes, for example, the sparse grid method [34], the finite state projection algorithm [48], and the conjunction of the Krylov method and finite state projection algorithm [44, 45].

When the molecular numbers x_i $(i = 1, \ldots, N)$ in a chemical reaction system are quite large, the value of $a_j(\mathbf{x})\tau$ may be large for an appropriately selected step size τ. In this case, the Poisson random variable $\mathcal{P}(a_j(\mathbf{x})\tau)$ can be approximated by a normal random variable with the same mean and variance, given by [26],

$$\mathcal{P}(a_j(\mathbf{x})\tau) \approx N(a_j(\mathbf{x})\tau, a_j(\mathbf{x})\tau) = a_j(\mathbf{x})\tau + \sqrt{a_j(\mathbf{x})\tau}\ N(0,1). \tag{9}$$

Then the Poisson τ-leap method (4) can be approximated by the following formula with normal random variables, given by

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \sum_{j=1}^{M} \nu_j a_j(\mathbf{x})\tau + \sum_{j=1}^{M} \nu_j \sqrt{a_j(\mathbf{x})\tau}\ N_j(0,1), \tag{10}$$

where the M normal variables $N_j(0,1)$ are statistically independent. The above scheme is the explicit Euler method [9] for solving the chemical Langevin equation

$$d\mathbf{x}(t) = \sum_{j=1}^{M} \nu_j a_j(\mathbf{x})dt + \sum_{j=1}^{M} \nu_j \sqrt{a_j(\mathbf{x})}\ dW_j(t), \tag{11}$$

where $W_j(t)$ is the Wiener process. If the molecular numbers in the system are very large, the value of $a_j(\mathbf{x})\tau$ may be still very large for a given stepsize τ. In this case, compared with the drift term $a_j(\mathbf{x})\tau$, the diffusion terms $\sqrt{a_j(\mathbf{x})\tau}\,N_j(0,1)$ in (10) are neglectable. Finally, we obtained the explicit Euler method for solving the chemical rate equation

$$\frac{d\mathbf{x}(t)}{dt} = \sum_{j=1}^{M} \nu_j a_j(\mathbf{x}). \tag{12}$$

The chemical Langevin equation (11) links three types of important modelling regimes, namely the discrete stochastic models simulated by the SSA or τ-leap method, continuous stochastic models in terms of SDEs, and continuous deterministic models of ordinary differential equations (ODEs). It also provides an approach to describe internal noise in the continuous SDE regime [63]. The chemical Langevin equation was used to develop the quasi-steady-state assumption which is the theoretical basis of the multiscale simulation methods [54, 32].

The multiscale methods are promising methods for simulating chemical reaction systems. Based on the molecular numbers and values of propensity functions, chemical reactions are partitioned into different reaction subsets at different time steps, and different simulation methods are employed to simulate different subsets of chemical reactions. For example, Burrage et al [13] proposed an adaptive approach to divide a reaction system into slow, intermediate and fast reaction subsets and used the SSA, the Poisson τ-leap method and a SDE solver to simulate the reactions in different subsets. Different partitioning techniques and different combinations of simulation methods have led to a number of effective methods and software systems. The combinations of simulation methods include, for example, the SSA and the Poisson τ-leap method [53], the SSA and a SDE solver [58, 59], the SSA and an ODE solver [35, 68], the SSA, the Poisson τ-leap method and a SDE solver [13], together with the SSA, the Poisson τ-leap method, a SDE and an ODE solvers [31]. In addition, a number of well-developed computer software systems have been published on the internet [40, 43, 67, 21]. These methods include a number of sophisticated approaches to address the stiffness in chemical reaction systems. In addition, we may use a robust method such as the SSA to simulate reactions that may generate negative molecular numbers and use a more efficient method such as a τ-leap method or a SDE solver to simulate reactions with large molecular numbers. In this way we may not only improve the computing efficiency but also solve the problem of the possible negative molecular numbers in stochastic simulations.

5. Delay Stochastic Simulation Algorithm

The SSA assumes that the next reaction will fire in the next reaction time interval $[t, t+\mu]$ with small values of μ. For systems including both fast and slow reactions, however, this assumption may not be valid if the slow reactions take a much longer time than the fast reactions. The large reaction time of slow reactions should be realized by time delay if we hope to put both fast and slow reactions in a system consistently and to study the impact of slow reactions on the system dynamics [47]. Recently, the delay stochastic simulation algorithm (DSSA) was designed to simulate chemical reaction systems with time delays [8, 5]. These methods have been used to validate stochastic models for biological systems with slow reactions [57,60]. However, compared with the significant progress in designing simulation methods for biological systems without time delay [28, 49], the development of effective methods for simulating chemical reaction systems with time delay is still in the very early stage of development.

It is assumed that a well-stirred chemical reaction system contains N molecular species $\{S_1, \dots, S_N\}$ with number $x_i(t)$ of the species S_i at time t. Among the M reaction channels, the first M_1 reactions $\{R_1, \dots, R_{M_1}\}$ are assumed to have time delay $\{T_1, \dots, T_{M_1}\}$, respectively, and the last $M - M_1$ reactions have no time delay. Reactions with delay are classified into consuming reactions and nonconsuming reactions [5]. When a nonconsuming reaction fires, the copy numbers of all the reactants do not change. Similar to the simulation methods for reactions without delay, both nonconsuming and consuming reactions are within a single framework by defining different state change vectors.

It is assumed that a delayed reaction with delay T_j initiates in an infinitesimal time interval $[t+\Delta, t+\Delta+d\Delta)$ with the next reaction time Δ and finishes at $[t+\Delta+T_j]$. Thus the firing of a delayed reaction takes two steps, namely the initiation step and update step. Consequently, two state change vectors (v_j and u_j) are defined to represent the initiation process and update process for each reaction channel. The key difference between the SSA and the DSSA is that, after the time step Δ of the next reaction is determined in the DSSA, we need to check if there is any scheduled update of delayed reactions in the time interval $[t, t+\Delta]$. The DSSA is given in detail as follows.

Method 4: **The delay stochastic simulation algorithm** [5].

1 Initialize. Determine the delay time and state change vectors for each delay reaction.

2 Calculate the values of propensity functions $a_j(\mathbf{x})$ of both delayed and nondelayed reactions based on the system state $\mathbf{x}$ at time t and $a_0(\mathbf{x}) = \sum_{j=1}^{M} a_j(\mathbf{x})$.

3 Use a sample r_1 of $\mathbf{U}(0,1)$ to determine the time of the next reaction

$$\Delta = \frac{1}{a_0(\mathbf{x})} \ln\left(\frac{1}{r_1}\right). \tag{13}$$

4 Check if there is any update of delayed reaction that is scheduled to finish in $[t, t+\Delta]$.

4.1 If the delayed reaction R_j has the earliest updating time and is scheduled to be updated at $t+\mu$ $(0 < \mu < \Delta)$, discard the next reaction time Δ, update the delayed reaction by using

$$\mathbf{x}(t+\mu) = \mathbf{x}(t) + u_j, \tag{14}$$

and then go to Step 2.

4.2 If there is no scheduled update of delayed reaction in $[t, t+\Delta]$, generate a sample r_2 of $\mathbf{U}(0,1)$ to determine the index k of the next reaction occurring in $[t, t+\Delta]$,

$$\sum_{j=1}^{k-1} a_j(\mathbf{x}) < r_2 a_0(\mathbf{x}) \leq \sum_{j=1}^{k} a_j(\mathbf{x}). \tag{15}$$

5.1 If a nondelayed reaction is selected, update the state of the system by

$$\mathbf{x}(t+\Delta) = \mathbf{x}(t) + \nu_k. \tag{16}$$

5.2 If a delayed reaction is selected, use (16) to update the initiation process and set a record to finish the reaction at $t+\Delta+T_k$.

The DSSA, which is also called the rejection method, abandons the random number and next reaction time if a scheduled update of delay reaction is within the next time step. To improve the efficiency, Cai proposed a direct method in which no random number was discarded [14]. The difference between the DSSA and the direct method is the method to determine the stepsize, namely Step 3 in Method 4. The direct method is based on the simulation methods for reaction systems with time dependent propensity functions [1]. When the propensities are functions of time t, the stepsize Δ of the next reaction has the following distribution function

$$1 - \exp\left(-\sum_{k=1}^{M} \int_t^{t+\Delta} a_k(\mathbf{x}(s)\,,\ s)\, ds \right) \equiv 1 - \exp\left(-\int_t^{t+\Delta} a_0(\mathbf{x}(s)\,,\ s)\, ds \right). \tag{17}$$

Using a sample r of $\mathbf{U}(0,1)$, the stepsize Δ is determined by solving

$$\int_t^{t+\Delta} a_0(\mathbf{x}(s)\,,\ s)\, ds = \ln\left(\frac{1}{r}\right). \tag{18}$$

Note that the above formula is that in the SSA (1) if the propensity functions are constant within $[t, t+\Delta]$.

The propensity functions are piecewise constant if they are the functions of the system state only. Assuming that the system state varies at time points $t < t_1 < \ldots < t_k < \tau$, the integral in (18) is

$$I = (t_1 - t)a_0(\mathbf{x}(t)) + (t_2 - t_1)a_0(\mathbf{x}(t_1)) + \ldots + (\Delta - t_k)a_0(\mathbf{x}(t_k)). \tag{19}$$

If the scheduled updating time points of delayed reactions are $t_1, t_2, \ldots, t_k$, the direct method finds the first scheduled update at t_j in order that

$$(t_1 - t)a_0(\mathbf{x}(t)) + (t_2 - t_1)a_0(\mathbf{x}(t_1)) + \ldots + (t_{j+1} - t_j)a_0(\mathbf{x}(t_j)) > \ln\left(\frac{1}{r}\right). \tag{20}$$

The direct method updates the scheduled delay reactions at $t_1, \ldots, t_{j-1}$ and then the next reaction will fire at

$$t^* = t_j + \frac{\ln(1/r) - (t_1 - t)a_0(\mathbf{x}(t)) - \ldots - (t_j - t_{j-1})a_0(\mathbf{x}(t_{j-1}))}{a_0(\mathbf{x}(t_j))}.$$

If the first update time t_1 is large or there is no scheduled update, the next reaction will fire at $t^* = t + \ln(1/r)/a_0(\mathbf{x}(t))$, which is the method in the SSA.

By combining the DSSA and the next reaction method for simulating systems without delays, Anderson proposed a modified next reaction method for systems with time dependent propensities and time delay [1]. In the next reaction method, each reaction has a putative reaction time and the reaction firing in the next step is that requiring the smallest reaction time. In the modified next reaction method, the updating time of delay reactions is also regarded as putative reaction time. An update of delayed reaction is selected if this updating time is the smallest putative reaction time. Therefore the approach for selecting the next reaction in the modified next reaction method is consistent with that in the rejection method. Then the updated system state is used to calculate the putative reaction time of all reaction channels.

6. Delay Chemical Master Equation

The chemical master equation (8) can be used to write down a system of ODEs that describes the deterministic behaviour of the mean associated with the SSA or an SDE that represents the intrinsic noise in continuous form. In a similar manner the DSSA described in Section 5 has corresponding representations as delay differential equations or stochastic delay differential equations (SDDEs). In the DSSA, the state vector $\mathbf{x}(t)$ is a non-negative N-dimensional jump stochastic process but is not a Markov process any more due to the time delay. For reaction systems with time delay, the chemical master equation should be based on the current system state at time t for reactions without time delay and the system state at $t - T_j$ for the reaction R_j with time delay T_j. The delay chemical master equation is given by [66] as

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{x},t) = &-\sum_{j=1}^{M_1}\sum_{\mathbf{x}_i \in I(\mathbf{x})} a_j(\mathbf{x}_i)P(\mathbf{x},t;\mathbf{x}_i,t-T_j) + \sum_{j=1}^{M_1}\sum_{\mathbf{x}_i \in I(\mathbf{x})} a_j(\mathbf{x}_i)P(\mathbf{x}-\nu_j,t;\mathbf{x}_i,t-T_j) \\ &- \sum_{j=M_1+1}^{M} a_j(\mathbf{x})P(\mathbf{x},t) + \sum_{j=M_1+1}^{M} a_j(\mathbf{x}-\nu_j)P(\mathbf{x}-\nu_j,t),\end{aligned} \tag{21}$$

where $P(\mathbf{x},t;\mathbf{x}_i,t-T_j)$ is the probability that the system is both in the state $\mathbf{x}$ at t and in the state $\mathbf{x}_i$ at $t - T_j$, and $I(\mathbf{x})$ is the set of all possible system states. When all molecular numbers are very large and fluctuations are not important, we can get the delay reaction rate equation from the chemical master equation (21), given by

$$\frac{d\mathbf{x}(t)}{dt} = \sum_{j=1}^{M_1} \nu_j a_j(\mathbf{x}(t-T_j)) + \sum_{j=M_1+1}^{M} \nu_j a_j(\mathbf{x}(t)). \tag{22}$$

Note, if there are no delays, the delay reaction rate equation (22) is an ODE describing the standard chemical kinetics rate equations (12).

Based on the numerical methods for solving the delay differential equations (22), we can use the general modelling approach [64] to develop effective stochastic simulation methods. Similar to the concept of Poisson τ-leap methods to speed up the computational performance of the SSA, we can adopt the same approach for the DSSA and this gives rise to the delay Poisson τ-leap method, given by [66]

$$\mathbf{x}(t_n + h) = \mathbf{x}(t_n) + \sum_{j=1}^{M_1} v_j P(a_j(\mathbf{x}(t_n - T_j))h) + \sum_{j=M_1+1}^{M} v_j P(a_j(\mathbf{x}(t_n))h). \tag{23}$$

If the mean of the Poisson random variables in (23) is very large, the Poisson random variable $P(a_j(\mathbf{x}(t))h)$ can be approximated by the Gaussian random variable $N(a_j(\mathbf{x}(t))h, a_j(\mathbf{x}(t))h)$ [26]. Then the discrete stochastic processes $\mathbf{x}(t)$ in (23) can be approximated by the continuous processes with Gaussian random variables, which is the explicit Euler scheme for solving the SDDEs that describe the evolution of reaction systems with time delay

$$d\mathbf{x} = \textstyle\sum_{j=1}^{M_1} v_j a_j(\mathbf{x}(t - T_j))dt + \sum_{j=M_1+1}^{M} v_j a_j(\mathbf{x}(t))dt$$

$$+ \sum_{j=1}^{M_1} v_j \sqrt{a_j(\mathbf{x}(t - T_j))}\ dW_j(t) + \sum_{j=M_1+1}^{M} v_j \sqrt{a_j(\mathbf{x}(t))}\ dW_j(t). \tag{24}$$

In the case that there are no delays this corresponds to the chemical Langevin equation (11).

7. Binomial τ-DSSA Method

Because of the issue of the possible negative molecular numbers in stochastic simulation, the proposed Poisson delay τ-leap method (23) has not been implemented for simulating reaction systems with time delay. Recently the binomial τ-DSSA method was proposed to improve the efficiency of the DSSA [41]. Based on the binomial τ-leap method [62] and the modified binomial τ-leap method [51], a generalized binomial τ-leap method was first proposed to overcome some of the difficulties associated with the binomial τ-leap methods [41. The binomial τ-DSSA method was then developed from the generalized binomial τ-leap method to simulate reaction systems with time delay. The additional step in the binomial τ-DSSA method is to determine the exact firing time of each delayed reaction. Once the reaction number $\mathcal{B}_j$ of a delay reaction firing in $[t, t+\tau]$ is determined, we need to determine the $\mathcal{B}_j$ firing time points in the time interval $[t, t+\tau]$. It was proposed that the $\mathcal{B}_j$ delay reactions initiate at $t + r_j\tau$ $(j = 1, \ldots, \mathcal{B}_j)$, where r_j is a sample of $\mathbf{U}(0, 1)$. Thus the scheduled updating time points of this delayed reaction with delay T_j are $t + r_j\tau + T_j$. In addition, the final step for updating the system should include the delayed reactions whose products are scheduled to be manifested within the leap step. Thus the update of the system is given by

$$\mathbf{x}(t+\tau) = \mathbf{x}(t) + \sum_{j=1}^{M_0} u_j \mathcal{B}_j^{(1)} + \sum_{j=M_0+1}^{M_1} u_j \mathcal{B}_j^{(2)} + \sum_{j=M_1+1}^{M} v_j \mathcal{B}_j^{(3)},$$

which represents the update of consuming delayed reactions, nonconsuming delayed reactions and nondelayed reactions [41].

The binomial τ-DSSA method maintains a queue structure of the delayed reactions for update. At each leap step, the queue is checked to determine whether the product of the delayed reactions ought to be manifested. Simultaneously, delay reactions initiated at that step are added to the queue and the time points of manifestation are determined by random numbers generated from $\mathbf{U}(0,1)$. To minimize the computational effect, Bayati et al [7] proposed the D-leap method which was aimed at reducing the time to generate and store the firing time of delayed reactions. Instead of generating and recording the updating time points of delayed reaction R_j in a leap step $[t, t+\tau]$, the D-leap method records the number of delayed reactions $\mathcal{B}_j$, the first possible execution time of these delayed reactions $t+\tau+T_j$ and the leap size τ which is called span. When the earlier possible updating time is within the interval of the current time step, the number of reactions that will be executed in the current time step will be determined by the overlap of the span of these $\mathcal{B}_j$ reactions with the current leap step [7].

7. Conclusion

This paper gave an overview for the recent progress in the development of effective stochastic methods for simulating chemical reaction systems. The τ-leap methods and multiscale simulation methods are two major types of efficient methods that have been extensively studied since the development of the Poisson τ-leap methods. In addition, a number of simulation methods have been designed to simulate chemical reaction systems with time delay. However, compared with the significant progress in designing simulation methods for biological systems without time delay, the development of simulation methods for chemical reaction systems with time delay is still at the very early stage of development. This will be one of the interesting topics for future research. In addition, more work is needed to design adaptive simulation methods that can be used to simulate a wide range of chemical reaction systems. Further, there are a number of theoretical issues that are related to the convergence property of simulation methods. The progress in this area may lead to simulation methods with higher order accuracy. Finally, the huge computational time of the spatio-temporal stochastic simulation is a significant challenge in stochastic simulation. Although substantial progress has been achieved in this area [6, 10, 2, 33, 29, 20, 61], the spatio-temporal stochastic simulation is also one of the interesting topics for future research.

References

[1] D. F. Anderson, A modified next reaction method for simulating chemical systems with time dependent propensities and delays, *Journal of Chemical Physics* **127**, (2007), 214107.

[2] S. S. Andrews, and D. Bray, Stochastic simulation of chemical reactions with spatial resolution and single molecule detail, *Physical Biology* **1**, (2004), 137151.

[3] A. Arkin, J. Ross, and H. H. McAdams, Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected Escherichia coli cells, *Genetics* **149**, (1998), 1633-1648.

[4] A. Auger, P. Chatelain, and P. Koumoutsakos, R-leaping: accelerating the stochastic simulation algorithm by reaction leaps, *Journal of Chemical Physics* **125**, (2006), 084103.

[5] M. Barrio, K. Burrage, A. Leier, and T. Tian, Oscillatory regulation of Hes1: discrete stochastic delay modelling and simulation, *PLoS Computational Biology* **2**, (2006), e117.

[6] T. M. Jr Bartol, B. R. Land, E. E. Salpeter, and M. M. Salpeter, Monte Carlo simulation of miniature endplate current generation in the vertebrate neuromuscular junction, *Biophysical Journal* **59**, (1991), 290-307.

[7] B. Bayati, P. Chatelain, and P Koumoutsakos, D-leaping: accelerating stochastic simulation algorithms for reactions with delays, *Journal of Computational Physics* **228**, (2009), 5908-5916.

[8] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Delay-induced stochastic oscillations in gene regulation, *Proceedings of the National Academy of Sciences USA* **102**, (2005), 14593-14598.

[9] K. Burrage, P. M. Burrage, and T. Tian, Numerical methods for strong solutions of stochastic differential equations: an overview, *Proceedings of the Royal Society of London A* **460**, (2004), 373-402.

[10] K. Burrage, J. Hancock, A. Leier, D. V. Jr Nicolau, Modelling and simulation techniques for membrane biology, *Briefing in Bioinformatics* **8**, (2007), 234-244.

[11] K. Burrage, S. MacNamara, and T. Tian, Accelerated Leap Methods for Simulating Discrete Stochastic Chemical Kinetics, *Lecture Notes in Control and Information Sciences* **341**, (2006), 359-366.

[12] K. Burrage, and T. Tian, Poisson Runge-Kutta methods for chemical reaction systems, In: *Advances in Scientific Computing and Applications*, Y. Y. Lu et al. eds. (2004), 82-96, Science Press, Beijing.

[13] K. Burrage, T. Tian, and P. Burrage, A multi-scaled approach for simulating chemical reaction systems, *Progress in Biophysics and Molecular Biology* **85**, (2004), 217-234.

[14] X. Cai, Exact stochastic simulation of coupled chemical reactions with delays, *Journal of Chemical Physics* **126**, (2007), 124108.

[15] X. Cai, and Z. Xu, K-leap method for accelerating stochastic simulation of coupled chemical reactions, *Journal of Chemical Physics* **126**, (2007), 074102.

[16] Y. Cao, D. T. Gillespie, and L. R. Petzold, Avoiding negative populations in explicit Poisson tau-leaping, *Journal of Chemical Physics* **123**, (2005), 054104.

[17] Y. Cao, D. T. Gillespie, and L. R. Petzold, Efficient step size selection for the tau-leaping

simulation method, *Journal of Chemical Physics* **124**, (2006), 044109.

[18] Y. Cao, D. T. Gillespie, and L. R. Petzold, Adaptive explicit-implicit tau-leaping method with automatic tau selection, *Journal of Chemical Physics* **126**, (2007), 224101.

[19] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis, Binomial distribution based τ-leap accelerated stochastic simulation, *Journal of Chemical Physics* **122**, (2005), 024112.

[20] A. Chatterjee, and D. G. Vlachos, Multiscale spatial Monte Carlo simulations: multigriding, computational singular perturbation, and hierarchical stochastic closures, *Journal of Chemical Physics* **124**, (2006), 64110.

[21] P. K. Dhar, T. C. Meng, S. Somani, L. Ye, K. Sakharkar, A. Krishnan, A. Ridwan, S. H. K. Wah, M. Chitre, and Z. Hao, Grid cellware: the first grid-enabled tool for modelling and simulating the cellular processes, *Bioinformatics* **21**, (2005), 1284-1287.

[22] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Stochastic gene expression in a single cell, *Science* **297**, (2002), 1183-1186.

[23] M. A. Gibson, and J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, *Journal of Physical Chemistry A* **104**, (2000), 1876-1889.

[24] D. T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *Journal of Computational Physics* **22**, (1976), 403-434.

[25] D. T. Gillespie, Exact stochastic simulation of coupled chemical reactions, *Journal of Chemical Physics* **81**, (1977), 2340-2361.

[26] D. T. Gillespie, The chemical Langevin equation, *Journal of Chemical Physics* **113**, (2000), 297-306.

[27] D. T. Gillespie, Approximate accelerated stochastic simulation of chemically reacting systems, *Journal of Chemical Physics* **115**, (2001), 1716-1733.

[28] D. T. Gillespie, Stochastic simulation of chemical kinetics, *Annual Review of Physical Chemistry* **58**, (2007), 35-55.

[29] D. T. Gillespie, The multinomial simulation algorithm for discrete stochastic simulation of reaction-diffusion systems, *Journal of Chemical Physics* **130**, (2009), 094104.

[30] D. T. Gillespie, and Petzold LR. Improved leap-size selection for accelerated stochastic simulation, *Journal of Chemical Physics* **119**, (2003), 8229-8234.

[31] L. A. Harris, and P. Clancy, A partitioned leaping approach for multiscale modeling of chemical reaction dynamics, *Journal of Chemical Physics* **125**, (2006), 144107.

[32] E. L. Haseltine, and J. B. Rawlings, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics, *Journal of Chemical Physics* **117**, (2002), 6959-6969.

[33] J. Hattne, D. Fange, and J. Elf, Stochastic reaction-diffusion simulation with MesoRD, *Bioinformatics* **21**, (2005), 2923-2924.

[34] M. Heglanda, C. Burdenc, L. Santosob, S. MacNamaraa, and H. Boothmaltese, A solver for the stochastic master equation applied to gene regulatory networks, *Journal of Computational and Applied Mathematics* **205**, (2007), 708-724.

[35] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer, COPASI a complex pathway simulator, *Bioinformatics* **22**, (2006) 3067-3074.

[36] Y. Hu, and T. Li, Highly accurate tau-leaping methods with random corrections, *Journal of Chemical Physics* **130**, (2009), 124109.

[37] D. A. Hume, Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression, *Blood* **96**, (2000), 2323-2328.

[38] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Stochasticity in gene expression: from theories to phenotypes, *Nature Reviews Genetics* **6**, (2005), 451-464.

[39] S. Kar, W. T. Baumann, M. R. Paul, and J. J. Tyson, Exploring the roles of noise in the eukaryotic cell cycle, *Proceedings of the National Academy of Sciences USA* **106**, (2009), 6471-6476.

[40] A. M. Kierzek, STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm, *Bioinformatics* **18**, (2002), 470-481.

[41] A. Leier, T. T. Marquez-Lago, and K. Burrage, Generalized binomial tau-leap method for biochemical kinetics incorporating both delay and intrinsic noise, *Journal of Chemical Physics* **128**, (2008), 205107.

[42] T. Li, Analysis of explicit tau-leap schemes for simulating chemical reaction systems, *Multiscale Modeling and Simulation* **6**, (2007), 417-436.

[43] L. Lok, and R. Brent, Automatic generation of cellular reaction networks with Moleculizer 1.0, *Nature Biotechnology* **23**, (2005), 131136.

[44] S. MacNamara, A. M. Bersani, K. Burrage, and R. B. Sidje, Stochastic chemical kinetics and the total quasi-steady-state assumption: Application to the stochastic simulation algorithm and chemical master equation, *Journal of Chemical Physics* **129**, (2008), 095105.

[45] S. MacNamara, K. Burrage, and R. B. Sidje, Multiscale modeling of chemical kinetics via the master equation, *Multiscale Modeling and Simulation* **6**, (2008), 11461168.

[46] T. T. Marquez-Lago, and K. Burrage, Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics, *Journal of Chemical Physics* **127**, (2007), 104101.

[47] N. A. M. Monk, Oscillatory expression of Hes1, p53, and NF-kappaB driven by transcriptional time delay, *Current Biology* **33**, (2003), 1409-1413.

[48] B. Munsky, and M. Khammash, The finite state projection algorithm for the solution of the chemical master equation, *Journal of Chemical Physics* **124**, (2006), 044104.

[49] J. Pahle, Biochemical simulations: stochastic, approximate stochastic and hybrid approaches, *Briefing in Bioinformatics* **10**, (2009), 53-64.

[50] J. Paulsson, Summing up the noise in gene networks, *Nature* **427**, (2004), 415-418.

[51] X. Peng, W. Zhou, and Y. Wang, Efficient binomial leap method for simulating chemical kinetics, *Journal of Chemical Physics 126*, (2007), 224109.

[52] M. F. Pettigrew, and H. Resat, Multinomial tau-leaping method for stochastic kinetic simulations. *Journal of Chemical Physics* **126**, (2007), 084101.

[53] J. Puchalka, and A. M. Kierzek, Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks, *Biophysical Journal* **86**, (2004), 1357-1372.

[54] C. V. Rao, and A. P. Arkin, Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm, *Journal of Chemical Physics* **118**, (2003), 4999-5010.

[55] C. V. Rao, D. W. Wolf, and A. P. Arkin, Control, exploitation and tolerance of intracellular noise, *Nature* **420**, (2002), 231-237.

[56] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, Stiffness in stochastic chemically reacting systems: the implicit tau-leaping method, *Journal of Chemical Physics* **119**, (2003), 12784-12794.

[57] M. R. Roussel, and R. Zhu, Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression, *Physical Biology* **3**, (2006), 274-284.

[58] H. Salis, and Y. N. Kaznessis, Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions, *Journal of Chemical Physics* **122**, (2005), 054103.

[59] H. Salis, and Y. N. Kaznessis, An equation-free probabilistic steady-state approximation: dynamic application to the stochastic simulation of biochemical reaction networks, *Journal of Chemical Physics* **123**, (2005), 214106.

[60] R. Schlicht, and G. Winkler, A delay stochastic process with applications in molecular biology, *Journal of Mathematical Biology* **57**, (2008), 613-648.

[61] B. M. Slepchenko, J. C. Schaff, J. H. Carson, and L. M. Loew, Computational cell biology: spatiotemporal simulation of cellular events, *Annual Review of Biophysics and Biomolecular Structure* **31**, (2002), 423-441.

[62] T. Tian, and K. Burrage, Binomial leap methods for simulating chemical kinetics, *Journal of Chemical Physics* **121**, (2004), 10356 -10364.

[63] T. Tian, and K. Burrage, Bistability and switching in the lysis/lysogeny genetic regulatory network of Bacteriophage lambda, *Journal of Theoretical Biology* **227**, (2004), 229-237.

[64] T. Tian, and K. Burrage, Stochastic models for regulatory networks of the genetic toggle switch, *Proceedings of the National Academy of Sciences USA* **103**, (2006), 8372-8377.

[65] T. Tian, and K. Burrage, An efficient stepsize selection procedure for discrete simulation of biochemical reaction system, *ANZIAM Journal* **48**, (2006) C1022-C1040.

[66] T. Tian, K. Burrage, P. M. Burrage, and M. Carletti, Stochastic delay differential equations for genetic regulatory networks, *Journal of Computational and Applied Mathematics* **205**, (2007), 696-707.

[67] M. Tomita, K. Hashimoto, K. Takahashi, T. S. Shimizu, Y. Matsuzaki, F. Miyoshi, K. Saito, S. Tanida, K. Yugi, J. C Venter, and C. A. Hutchison III, E-CELL: software environment for whole-cell simulation, *Bioinformatics* **15**, (1999), 72-84.

[68] K. Vasudeva, and U. S. Bhalla, Adaptive stochastic-deterministic chemical kinetic simulations, *Bioinformatics* **20**, (2004), 78-84.

[69] D. J. Wilkinson, Stochastic modelling for quantitative description of heterogeneous biological systems, *Nature Reviews Genetics* **10**, (2009), 122-133.

[70] Z. Xu, and X. Cai, Unbiased τ-leap methods for stochastic simulation of chemically reacting systems, *Journal of Chemical Physics* **128**, (2008), 154112.