Effective time-stepping for the tau-leaping method for stochastic simulation of well-stirred biochemical networks

by

Mahmuda Binte Mostofa Ruma,

Bachelor of Science, Dhaka University, 2003

A thesis

presented to Ryerson University

in partial fulfilment

of the requirements for the degree of

Master of Science

in the program of

Applied Mathematics

Toronto, Ontario, Canada, 2017

AUTHOR'S DECLARATION FOR ELECTRONIC SUBMISSION OF A THESIS

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research.

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Effective time-stepping for the tau-leaping method for stochastic simulation of well-stirred biochemical networks

Master of Science, 2017

Mahmuda Binte Mostofa Ruma

Applied Mathematics

Ryerson University

Biological processes at the cellular level are noisy. The noise arises due to random molecular collisions, and may be substantial in systems with low molecular counts in some species. This thesis introduces a variable tau-leaping method for the simulation of stochastic discrete mathematical models of well-stirred biochemical systems which is theoretically justified. Numerical tests on several models of biochemical systems of practical interest illustrate the advantages of the adaptive tau-leap method over the existing schemes.

Acknowledgements

At the outset I am pronouncing my thankfulness to the Almighty who gave me the ability to carry out this research work. My gratitude will always be for my supervisor Dr. Silvana Ilie, Associate Professor, Department of Mathematics, Ryerson University, who continuously guided me from all directions. It is my great gratification for having the opportunity to work under her supervision. I would like to express my heart-rending admiration to my supervisor who has encouraged and rightly initiated me to step into the wide area of mathematics and its application in the engineering fields. I am not less grateful and thankful to the efforts, perseverance, sincerity, enormous will-force, clarity, accuracy, completeness, monumental patience, generous co-operation and fellow-feeling she showed for me to venture this research and bring this painstaking task to a successful end.

I would also like to show my deep gratitude to all my teachers in the Department of Mathematics, Ryerson University, for their intelligent and open-minded support. I would also like to thank the staff of the Department of Mathematics, Ryerson University, in providing me all necessary help during the course of my MSc degree. I would like to thank Dr. Dejan Delic, Dr. Katrin Rohlf and Dr. Dzung Minh Ha, the members of my thesis committee for agreeing to

give their valuable time for me.

Last but not the least, I would like to express my heart-felt thanks to

my parents and my husband without whose encouragement

and support this work would not be what it is today.

Finally, I would like to thank my daughter for her co-operation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

In recent years, stochastic modeling and simulations of biochemical networks have been subject of intense research [1, 5, 30, 41]. The current interest in Computational Cell Biology reflects the need for advanced computer tools to understand the complexity of living cells. One of the objectives of Computational Biology is to generate computer simulations for studying biological phenomena, data or patterns.

The deterministic rate laws for modeling chemically reacting systems were employed in biochemistry by Heinrich and Schuster [29] and applied in chemistry by Epstein and Pojman [17]. This deterministic approach to modelling utilizes a simple relation between molecular concentrations of different species and the reaction rates. The law of mass action, introduced by Espenson [18], describes the evolution of the concentrations of each chemical species for all future times.

Robert Brown, the Scottish botanist, first noticed the presence of fluctuations when studying living phenomena at microscopic level (1827). The *macroscopic* model is considered only for chemical systems with large molecular amounts for each species. However, random fluctuations are unavoidable at the molecular level and may have a significant impact when some low molecular numbers exist for some species in the biochemically reacting system. This happens, for example, in

modelling DNA binding sites in gene regulatory networks. McAdams and Arkin [41] showed that the low copy numbers of expressed RNAs are meaningful in the adjustment downstream pathways. Consequently, stochastic modeling is required (Morton-Firth and Bray [42]) as a number of important biological processes have small number of molecules present in the reaction volume.

The stochastic model of well-stired chemical kinetics, known as the Chemical Master Equation, was considered by Van Kampen and Gillespie [23]. The relationship between the stochastic and the deterministic chemical reaction models was discused by Kurtz [35, 36]. Gillespie [25, 26] proposed a Monte Carlo simulation algorithm for generating trajectories in exact agreement with the probability given by the Chemical Master equation.

The Chemical Master Equation has been studied intensely and it was successfully applied for understanding many biochemical proceses in the cell, even when the well- mixed assumption is loosely applicable. Nonetheless, the simulations of stochastic models are computationally much more intense than their deterministic counterparts. Since biochemical systems are quite complex and often entail both fast and slow reactions, Gillespie's algorithm is also quite expensive for these systems. These challenges have increased the interest of the numerical analysis community to design more efficient numerical methods for solving the Chemical Master Equation. A more efficient exact algorithm was proposed by Gibson and Bruck [19]. A number of approximate algorithms have also been developed and more efforts are currently

dedicated to this fast expanding research area [30]. The tau-leaping method was introduced by Gillespie [21] for reducing the computational cost of the stochastic simulation of well-stirred biochemical networks. The tau-leaping strategy requires that some leap condition is satisfied for each time-step. More effective tau-leaping methods were proposed by Cao, Gillespie and Petzold [13], Tian and Burrage [52], Anderson [3] and Chatterjee et al. [16]. The theoretical framework for studying tau-leaping methods, such as consistency and stability, was developed by Rathinam et al. [47], Anderson et. al [4] higher order methods by Li [39] and post-leap checks by Anderson [3].

Leap selection techniques for stochastic discrete models of well-stirred biochemical systems that aim to speed -up the simulation were developed by Cao, Gillespie and Petzold [13, 14]. Moreover, several adaptive strategies were introduced for the pathwise numerical solution [38, 51] and the mean-square numerical solution [37] of stochastic continuous models of well-stirred biochemical networks (known as the Chemical Langevin Equation). A technique to adaptively choose the step size in the spatial tau-leaping method for the Reaction Diffusion Master Equation was proposed in [44].

In this thesis, we describe an improved adaptive time-stepping technique for approximating the exact solution of the Chemical Master Equation using the tau-leaping method. This technique extends a strategy for variable time-stepping in the numerical solution of stochastic differential equations developed by Burrage

and Burrage [10] to efficiently select the steps in the tau-leaping method for the discrete stochastic model of the Chemical Master Equation. Also, the adaptive method for the tau-leap strategy for the Chemical Master Equation presented in this thesis shares similarities with the variable time-stepping technique for the Reaction Diffusion Master Equation introduced in [44]. However, our proposed adaptive scheme dynamically partitions the species into slow and fast subsets.

The numerical solution obtained by tau-leaping must obey the leap condition for some given tolerance [13, 21]. The step is rejected if the leap condition criteria is not satisfied. When the time-step is rejected, a method is applied for guaranteeing that the correct stochastic path is followed. In order to ensure the correct statistics of the sample paths, we apply a technique intoduced by Anderson [3] for the tau-leaping method upon the rejection of steps.

We consider three different types of constrains for adaptive time-stepping. When a future constraint does not exist, we consider the Poisson distribution for generating the amount of reactions in the current time step. But in the presence of a future constraint, we consider conditioning on the previously created steps. There are two cases, one is for stepping before the constraint and another is for stepping after the constraint. Advancing a step before a constraint is performed by generated values with a binomial distribution (see Anderson [3]). We calculate the number of reactions as the sum of new Poisson distributions and the difference between the existing Poisson distributions and the binomial distributions for the step after the

constraint. Our method generates paths with different sequences of time steps; on each path the selection of the leap size is flexible. We implement the leap condition in a strong sense, therefore we expect that the results obtained with our method are very accurate.

The outline of the thesis is given below.

We discuss the background of stochastic models of biochemical networks and stochastic simulation methods of biochemical kinetics in Chapter 2. We introduce an improved adaptive method for the tau-leaping strategy for the Chemical Master Equation in Chapter 3. In Chapter 4, the numerical results on three problems of practical interest illustrate the advantages of the proposed adaptive tau-leap method. In Chapter 5, we summarize our conclusions and discuss directions of future research.

Chapter 2

BACKGROUND

Below, we give an introduction to the stochastic modeling and simulation of discrete stochastic models of well-stirred biochemical kinetics. In this section, four computational algorithms will be introduced, the Gillespie algorithm, the First Reaction Algorithm, the Next Reaction algorithm and the explicit tau-leap algorithm, for the numerical simulation of stochastic homogeneous biochemical systems.

2.1 Stochastic Models of Biochemical Kinetics

We consider a process of N different types of chemical species, $S_1, S_2, ..., S_N$ that take part in M types of chemical reactions, $R_1, R_2, ..., R_M$. The dynamics of the system may be studied by keeping track of the position and the velocity for each molecule under the appropriate laws of physics, considering the impact between the molecules and the result of their interaction. Nonetheless, this Molecular Dynamics approach is very challenging to solve numerically if the total molecular count is large or if integration over a long time interval is required. In case the system is well-stirred or homogeneous, the model may be substantially simplified. By ignoring the spatial distribution, one can evaluate only the molecular count of each species. We also assume that the system is in thermal equilibrium and in a constant volume.

Suppose that at time $t = 0$, the molecular number of every species is known. It is then sufficient to consider the evolution in time of the amount of molecules of each species. In what follows, we are interested in computing the number of molecules of each species as time increases. Consider therefore the state vector of the system, $X(t) = [X_1(t), X_2(t), ..., X_N(t)]^T$, where $X_i(t)$ represents the amount of S_i molecules available at time t.

The state vector $X(t)$ will be changed when one of the M reactions takes place. Since we do not consider spatial information in this case, we wish to evaluate the probability of a reaction taking place, given the current state of the system. For well-stirred systems, each reaction R_j can be characterized by the following:

i) **a propensity function** $a_j(x)$ is defined as: $a_j(x)dt \triangleq$ probability that a single R_j event happens between $[t, t + dt)$, if $X(t) = x$.

ii) **a state change vector** ν_j: given that $X(t) = x$, one reaction R_j changes the state vector to $x + \nu_j$.

Examples:

For any biochemical system:

$$S_1 \xrightarrow{c_j} S_2;$$

the propensity function is $a_j(x)dt = c_j x_1$ and its state change vector is

$$\nu_j = (-1, 1, 0, ..., 0)^T.$$

The entry ν_{ij} represents the change in the population of species S_i caused by a reaction R_j. The propensity functions are computed according to the following

rules justified by the principles of kinetic theory:

For Second Order Reactions: When $m \neq n$, if the reaction is

$$S_m + S_n \xrightarrow{c_j} \text{products},$$

then the propensity is

$$a_j(\mathbf{X}(t)) = c_j X_m(t) X_n(t).$$

Here 'products' represents the summation of molecules of certain species.

Dimerization: For $m = n$, if the reaction is

$$S_m + S_m \xrightarrow{c_j} \text{products},$$

its propensity is then

$$a_j(\mathbf{X}(t)) = c_j \frac{1}{2} X_m(t)(X_m(t) - 1).$$

The propensity of a dimerization, $\frac{1}{2} X_m(t)(X_m(t) - 1)$ is evaluated as the number of ways in which an unordered pair of molecules can be chosen from a total $X_m(t)$ of S_m molecules.

For First Order Reactions: When a reaction of the following type occurs

$$S_m \xrightarrow{c_j} \text{products},$$

its propensity takes the form of

$$a_j(\mathbf{X}(t)) = c_j X_m(t).$$

Consider now the probability of the system to be in a state x at time t. It is of interest to describe the evolution of these probabilities if the initial system state is known.

2.1.1 Chemical Master Equation (CME)

Let us define $P(x,t)$, to be the probability that the state vector at time t is $\mathbf{X}(t) = x$, if $X(0) = x_0$. We wish to find the probability of being in state x at time $t+dt$ where the step dt is so small that no more than one reaction event happens during $[t, t+dt)$. The system is in state $\mathbf{x}$ at time $t+dt$ if (i) the system was in state $\mathbf{x}$ at time t and there was no reaction during $[t, t+dt)$, or (ii) for $1 \leq j \leq M$ the system state was $\mathbf{x} - \nu_j$ at time t is and one reaction R_j happened during $[t, t+dt)$, therefore after a step dt the system state becomes $\mathbf{x}$. In what follows, we employ a result from the probability theory, namely the law of total probability. If B is the event of interest and the events $H_0, H_1, H_2, .., H_M, H_{M+1}$ obey the following conditions:
(a) exhaustive (one such event occurs) and (b) are disjoint (at most one event occurs), then, according to the law of total probability, we obtain

$$P(B) = \sum_{j=0}^{M} P(B \mid H_j) P(H_j).$$

We denoted the conditional probability of B occurring, if H_j occurred, by $P(B|H_j)$. Let us take B to be the event that the system state is $\mathbf{x}$ at time $t+dt$. We consider H_0 to be the event that the system state at time t was $\mathbf{x}$. Also assume that H_j for $1 \leq j \leq M$ is the event that the system state at time t was $\mathbf{x} - \nu_j$ and that H_{M+1}

is the event that the system state is anything else at time t. We remark that the conditional probability $P(B \mid H_j)$ with $1 \leq j \leq M$ is, in fact, the probability of one reaction R_j happening during $[t, t+dt)$. This probability may be computed using the propensity function of this reaction as

$$P(B \mid H_j) = a_j(\mathbf{x} - \nu_j)dt, \text{ for } 1 \leq j \leq M.$$

With this observation, we can show that $P(B \mid H_0)$ is the probability that no reactive event occurred in $[t, t+dt)$. Thus

$$P(B \mid H_0) = 1 - \sum_{j=1}^{M} a_j(x)dt$$

We also observe that,

$$P(B \mid H_{M+1}) = 0,$$

as H_{M+1} consists of all the possible states that differ by two or more reactions from the given system state x.

Now, from the definition of $P(x,t)$, we have

$$P(x, t+dt) = (1 - \sum_{j=1}^{M} a_j(x)dt)P(x,t) + \sum_{j=1}^{M} a_j(x - \nu_j)dtP(x - \nu_j, t)$$

which may be rewritten in the form

$$\frac{P(x, t+dt) - P(x,t)}{dt} = \sum_{j=1}^{M} (a_j(x - \nu_j)P(x - \nu_j, t) - a_j(x)P(x,t)). \tag{2.1}$$

We take $dt \to 0$ and obtain

$$\frac{dP(x,t)}{dt} = \sum_{j=1}^{M} (a_j(x - \nu_j)P(x - \nu_j, t) - a_j(x)P(x,t)).$$

This is a system of linear ordinary differential equations in $P(x,t)$, known as the Chemical Master Equation (CME) (see Gillespie [23]). The dimension of this system is equal to the number of all possible system states. This depends on the total number of molecules present and the specific form of the reactions. Often, the CME is of very high dimension and therefore it is computationally very challenging.

2.1.2 Chemical Langevin Equation (CLE)

Under certain conditions, the Chemical Master Equation model may be reduced to the Chemical Langevin Equation which is easier to solve numerically than the CME. Following Gillespie [21], we assume that the leap condition is satisfied, that is τ is small enough such that each propensity $a_j(x) \approx$ constant during the time interval $[t, t+\tau)$. Then the number of R_j reactions in $[t, t+\tau)$ may be approximated by a *Poisson random variable* $P_j(a_j(x), \tau)$ with mean and variance $a_j(X(t))\tau$. Then when $X(t) = x$, we obtain:

$$X(t+\tau) = X(t) + \sum_{j=1}^{M} \nu_j P_j(a_j(X(t)), \tau) \tag{2.2}$$

which is known as *the explicit tau-leaping method*. Moreover, if the leap τ is also chosen such that $a_j(X(t))\tau \gg 1$ for all $1 \leq j \leq M$, then we can approximate the Poisson random variable $P_j(a_j(x), \tau)$ with a normal random variable

$$N_j(a_j(x)\tau, a_j(x)\tau) \sim a_j(x)\tau + \sqrt{a_j(x)\tau}\ Z_j$$

with the same mean and variance $a_j(x(t))\tau$. Here Z_j is normally distributed with mean 0 and variance 1.

Thus:

$$P_j(a_j(X(t),\tau) \simeq a_j(X(t))\tau + \sqrt{a_j(X(t))\tau}\ Z_j,$$

Substituting the above in equation (2.2), we obtain

$$X(t+\tau) = X(t) + \tau\sum_{j=1}^{M} \nu_j a_j(X(t)) + \sqrt{\tau}\sum_{j=1}^{M} \nu_j \sqrt{a_j(\mathbf{X}(t))}\ Z_j$$

Recall the following definition of a Wiener process.

Definition: A scalar standard *Wiener process*, over $[0,T]$ is a random variable $W(t)$ that depends continuously on $t \in [0,T]$ and satisfies the following three conditions.

(1) $W(0) = 0$ (with probability 1).

(2) For $0 \leqslant s < t \leqslant T$ the random variable given by the increment $W(t) - W(s)$ is normally distributed with mean zero and variance $t - s$; equivalently, $W(t) - W(s) \sim \sqrt{t-s}N(0,1)$, where $N(0,1)$ denotes a normally distributed random variable with zero mean and unit variance.

(3) For $0 \leqslant s < t < u < v \leqslant T$ the increments $W(t) - W(s)$ and $W(v) - W(u)$ are independent.

If we take the limit $\tau \to dt$ in the previous equation then we derive

$$\mathbf{dX} = \sum_{j=1}^{M} \nu_j a_j(\mathbf{X}(t))dt + \sum_{j=1}^{M} \nu_j \sqrt{a_j(\mathbf{X}(t))}\ dW_j(t) \tag{2.3}$$

Here $W_j(t)$ are the independent scalar Wiener processes, for $1 \leq j \leq M$. The stochastic differential equation (2.3) is called the *Chemical Langevin equation*. The dimension of the Chemical Langevin equation is N, the number of species in the system. The state vector $X(t)$ in the Chemical Langevin Equation is a Markov process continuous in space and in time.

2.1.3 Reaction Rate Equation (RRE)

The deterministic model of well-stirred biochemical kinetics can be obtained when very large numbers of every species exist in the system. Assume that the system is in thermodynamic limit. Thermodynamic limit is achieved when the species densities X_i/Ω stay bounded for all $i = 1, 2, ..., N$, as the populations of species S_i and the model volume Ω approach infinity. In this case, the deterministic part of CLE (2.3) has a similar size as that of the system but the fluctuation part has size similar to the square root of the model size. Consequently, the diffusion term in the Chemical Langevin Equation (2.3) is comparatively much smaller than the drift term (first term) of the CLE. Thus the diffusion term can be ignored, thus we can reduce the CLE to the reaction rate equation model. So, by neglecting the stochastic or fluctuation part from the CLE (2.3), we get the following system of the ordinary differential equations (ODEs):

$$\frac{d\mathbf{X(t)}}{dt} = \sum_{j=1}^{M} \nu_j a_j(\mathbf{X(t)}),$$

which is known as the classical *reaction rate equation (RRE).*

This mathematical model is often given in terms of concentrations rather than in terms of population numbers. The concentrations vector, $\mathbf{y(t)}$, has components $y_i(t) = X_i(t)/(VN_A)$ for $1 \leq i \leq N$ with $N_A = 6.023 \times 10^{23} mol^{-1}$ being Avogadro's constant and V the volume.

The RRE obeys the *law of mass action* that has an empirical rule of thumb. We refer to D. J. Highman [30] for constructing the propensity functions depending on concentrations as below:

Case 1: First order reactions: If the reaction is

$$S_m \overset{k_j}{\rightarrow} \text{ products},$$

the expression of the propensity function becomes

$$a_j(y(t)) = k_j y_m(t).$$

Case 2: Second order reactions: If the reaction is

$$S_m + S_n \overset{k_j}{\rightarrow} \text{ products},$$

with condition $m \neq n$, the propensity function takes the form of

$$a_j(y(t)) = k_j y_m(t) y_n(t).$$

Case 3: Dimerization: for the reactions

$$S_m + S_n \overset{k_j}{\rightarrow} \text{ products},$$

with $m = n$, the propensity function turns into the form of

$$a_j(y(t)) = k_j y_m(t)^2.$$

This form of the propensity functions in terms of concentrations is based on mass action kinetics principles. In conclusion, the reaction rate equation model is valid when large number of molecules of each species are present in the system.

2.2 Simulation Methods of Stochastic Biochemical Kinetics

2.2.1 Exact Methods

Instead of solving directly the Chemical Master Equation (CME) to find the probability to be in each possible state at any time t, one could generate an evolution of the system state, one trajectory at a time. This is the Monte Carlo approach and it is widely used for solving numerically the CME. Gillespie [25] described two exact simulation algorithms for solving the Chemical Master Equation. These methods are known as the Direct Method and the First Reaction Method. Gibson-Bruck [19] gave another such exact strategy, known as the Next Reaction Method. These techniques are describe below.

2.2.1.1 Gillespie Algorithm (or SSA)

The Chemical Master Equation is an accurate model of well-stirred stochastic biochemical kinetics. To derive the stochastic simulation algorithm (SSA), we first define $P_0(\tau|x,t)$ to be the probability, given $X(t) = x$, that no reaction of any type

happened in the time interval $[t, t+\tau)$. When using the definition of the propensity and the laws of probability, we derive

$$P_0(\tau + d\tau|x,t) = P_0(\tau|\ x,t) \times \left[1 - \sum_{j=1}^{M} a_j(x)d\tau\right] \tag{2.4}$$

Indeed,

$$\begin{aligned} P_0(\tau + d\tau|x,t) &= \text{Probability no reaction in } [t, t+\tau) \\ &\quad \times \text{Probability no reaction in } [t+\tau, t+\tau+d\tau) \\ &= P_0(\tau|x,t)(1 - \sum_{j=1}^{M} \text{Prob. of reaction } R_j \text{ in}[t+\tau, t+\tau+d\tau)) \\ &= P_0(\tau|x,t)(1 - \sum_{j=1}^{M} a_j(x)d\tau). \end{aligned}$$

By re-grouping the terms in equation (2.4) and taking to the limit $d\tau \to 0$, we arrive to the following ordinary differential equation

$$\frac{dP_0}{dt}(\tau|x,t) = -a_0(x)P_0(\tau|x,t).$$

The solution to this scalar ordinary differential equation, with the initial condition $P_0(0|x,t) = 1$, is

$$P_0(\tau|\ x,t) = exp(-a_0(x)\tau),$$

where

$$a_0(x) = \sum_{j=1}^{M} a_j(x). \tag{2.5}$$

In what follows, we wish to study $P(\tau, j|x,t)d\tau$, which is the probability that the next reaction is the j-th reaction and this reaction happens during $[t+\tau, t+\tau+d\tau)$,

given that $X(t) = x$.

$$P(\tau, j|x, t) = \text{Prob. no reaction in } [t, t+\tau)$$

$$\times \text{Prob. of reaction } R_j \text{ in } [t+\tau, t+\tau+d\tau)$$

$$= P_0(\tau|\ x, t) a_j(x) d\tau.$$

Thus, we obtain

$$P(\tau, j|x, t) = a_j(x)\ \exp(-a_0(x)\tau). \tag{2.6}$$

We can re-write this as

$$P(\tau, j|x, t) = \frac{a_j(x)}{a_0(x)} \left[a_0(x) e^{-a_0(x)\tau} \right]. \tag{2.7}$$

The expression (2.7) suggests that the time τ to the next occurring reaction is an exponentially distributed random variable with mean $1/a_0(x)$ and the index j of this reaction is the integer random variable with point probability $a_j(x)/a_0(x)$. Recall below some theoretical results (see also Wilkinson [53]).

Proposition 1: If $\tau_j \sim Exp(a_j)$ where $j = 1, 2, ..., m$ are independent exponential random variables, then

$$\tau_0 \equiv \min_j \tau_j \sim Exp(a_0),$$

$$\text{with } a_0 = \sum_{j=1}^{m} a_j.$$

Proof:

$$P(\tau_0 > x) = P(\min_j \tau_j > x)$$

$$= P([\tau_1 > x] \cap P([\tau_2 > x]) \cap ..... \cap P([\tau_m > x])$$

$= P(\tau_1 > x).P(\tau_2 > x).....P(\tau_m > x)$

$= e^{-\tau_1 x}.e^{-\tau_2 x}.....e^{-\tau_m x}$

$= e^{-x\sum_{j=1}^{m}\tau_j}$

$= e^{-a_0 x}.$

Hence, $P(\tau_0 \leq x) = 1 - e^{-\tau_0 x}$ and then also $\tau_0 \sim Exp(X_0)$.

Lemma: If $U \sim Exp(a)$ and $V \sim Exp(b)$ are independent exponential random variables, then

$$P(U < V) = \frac{a}{a+b}.$$

Proof:

$P(U < V) = \int_0^\infty P(U < V \mid V = v)f(v)dv$

$= \int_0^\infty P(U < v)f(v)dv$

$= \int_0^\infty (1 - e^{-av})be^{-bv}dv$

$= \dfrac{a}{a+b}$

Proposition 2: If $\tau_j \sim Exp(a_j)$ where $i = 1, 2, ..., m$ are independent exponential random variables and if j is the smallest index of the τ_j, then j is a discrete random variable with probability mass function

$$\omega_j = a_j/a_0,\ j = 1, 2, ..., n,\ \text{where}\ \ a_0 = \sum_{j=1}^{m} a_j$$

Proof:

$\omega_i = P(\tau_i < \min\limits_{j\neq i}\tau_j)$

$= P(\tau_i < Y)$, where $Y = \min\limits_{j\neq i}\tau_j$

$= \dfrac{a_i}{a_i + a_{-i}}$(from the lemma)

$$= \frac{a_i}{a_0}$$

By Gillespie's inversion method from the Monte Carlo theory, we can generate random samples of the two joint density functions of τ and j as follows: take two random numbers r_1 and r_2 uniformly distributed in the unit-interval and τ and j can be select according to

$$\tau = \frac{1}{a_0(x)} \ln(\frac{1}{r_1}) \tag{2.8}$$

and

$$j = \text{the lowest integer satisfying} \sum_{j'=1}^{j} a_{j'}(x) > r_2 a_0(x). \tag{2.9}$$

Now if $X(t) = x$, the state vector is upgraded to $X(t+\tau) = x + \nu_j$ to show that one reaction R_j fired, given that $X(t) = x_0$. This process is repeated until the solution is advanced to the final time T_{final}. The steps for the implementation of Gillespie's Direct Method [1976] can be summarized as follows:

Gillespie Algorithm

While $t < T_{final}$ do

Calculate $a_j(x)$ for $1 \leq j \leq M$ and calculate $a_0(x) = \sum_{j=1}^{M} a_j(x)$.

Generate $r_1, r_2 \backsim U(0,1)$

Compute $\tau = \frac{1}{a_0(x)} \ln(\frac{1}{r_1})$ and

$j =$ the lowest integer satisfying $\sum_{j'=1}^{j} a_{j'}(x) > r_2 a_0(x)$,

Update $t = t + \tau$, $X(t + \tau) = x + \nu_j$

end while

The SSA is exact for the CME, since it is generates a distribution in exact agreement with the distribution obtained when solving directly the CME. But SSA has a disadvantage, for the strategy of simulating every reaction event, one at a time, is too time consuming when applied to many real systems.

2.2.1.2 The First Reaction Method

The *First Reaction Method* [25, 26] produces a possible reaction time τ_j for the reaction R_j in accordance to the following expression:

$$\tau_j = (1/a_j(x)) \ln(1/r_j), \tag{2.10}$$

where $r_1, r_2, ..., r_M$ are uniform random numbers in $(0, 1)$. Then the next firing reaction is that which happens first, i.e,

$$\tau = \min_{1 \leqslant j \leqslant M} \tau_j \tag{2.11}$$

The index of this reaction is,

$$j = \text{index of reaction for the smallest } \tau_j \tag{2.12}$$

The steps of the First Reaction Method [25, 26] are as follows:

First Reaction Algorithm

Set M reactions, N species , at time $t = t_0$ initial state $X(t_0) = x_0$

and final time T_{final}.

While $t < T_{final}$ do

Calculate $a_j(x)$ for $1 \leq j \leq M$ and $a_0(x) = \sum_{j=1}^{M} a_j(x)$,

Generate $r_1, r_2, .., r_M$ uniform random numbers in $(0, 1)$.

Compute $\tau_j = \dfrac{1}{a_j(x)} \ln(1/r_j)$

calculate τ and j by using the formulas:

$\tau = \min_{1 \leqslant j \leqslant M} \tau_j$ and

$j =$ index of reaction for the smallest τ_j

update $t = t + \tau$, $X(t + \tau) = x + \nu_j$

end while

The simulation using First Reaction Method (FRM) is more time-consuming than that of the Direct Method (DM) because the FRM utilizes M random numbers for each step while the DM only computes two random numbers per step. By applying a few changes to the First Reaction Method, it can be turned into a more efficient algorithm, namely the Next Reaction Method.

2.2.1.3 Next Reaction Method

The Next Reaction Method is also known as the Gibson-Bruck algorithm [19]. It is more efficient than the First Reaction Method or the Direct Method. It utilizes an ordered binary tree P to find the next reaction and its possible time, and a Dependency Graph G to recalculate the propensities. The updating is done only to the propensities effectively changed after the firing of the chosen event. The Gibson-Brock [19] algorithm is given below:

Next Reaction Algorithm

Choose $t = t_0$ for $X(t_0) = t_0$ for $1 \leq j \leq M$

While $t < T_{final}$ do

Calculate $a_j(x)$ and calculate $a_0(x) = \sum_{j=1}^{M} a_j(x)$.

Use $\tau = \min_{1 \leqslant j \leqslant M} \tau_j$ and $j =$ index of reaction for the smallest τ_j.

Also store the value of (j, τ_j) into P.

Update $t = t + \tau$, $X(t + \tau) = x + \nu_j$

if $(j, j') \in$ G

Update $a_j^{'}$

Consider the case (i) for $j' \neq j$ thus $\tau_{j'} = (a_{j'}, old/a_{j'}, new)(\tau_{j'} - t) + t$.

else (ii) if $j' = j$, then generates τ_j by using $\tau_j = (1/a_j(x)) \ln(1/r_j)$

update $\tau_{j'}$ in P by setting $t = t + \tau$.

end while

Several things can be noted about this algorithm. This method advances the time from the current to the next time an event occurs, that is 'relative' times to 'absolute' times. This algorithm keeps track of the propensities affected by each reaction in an efficient way.

2.2.2 Approximate Methods

Any exact strategy computes a sequence of all reactions that happen in the system, thus becoming computationally quite intense on applications with some fast reactions. Another approach, which speeds-up the simulation, is to step with a predefined time-step over many reactions. In this case, the numerical solution is required to satisfy some accuracy criteria. We describe some of these approximate methods below.

2.2.2.1 Tau-leaping method

Gillespie introduced in 2001 [21] the tau-leap method for speeding-up the stochastic simulation of well-stirred biochemical system, by firing many reactions of the same type in a given time step τ. This method is applicable when a leap condition is obeyed. The *Leap Condition* requires that:

$\tau > 0$ is small enough such that each $a_j(x)$ function remains almost constant in the interval $[t, t+\tau)$. Then the number of R_j firings over the time interval $[t, t+\tau)$ can be estimated by a *Poisson random variable*, denoted by $P_j(a_j(x), \tau)$, with mean and variance $a_j(x)\tau$. Indeed,

$$X(t+\tau) = x + \sum_{j=1}^{M} \nu_j P_j \left(\int_t^{t+\tau} (a_j(s))ds \right)$$

when $X(t) = x$

Note that

$$\int_t^{t+\tau} (a_j(s))ds \backsimeq a_j(t)\tau$$

if the leap condition is satisfied. So, instead of updating the time from each reaction to the next, the technique advances the system with the largest suitable τ for the tau-leaping condition, and draws the number of firings for every reaction R_j from a Poisson random variable $P_j(a_j(x), \tau)$. Then the system is updated according to the following rule:

$$X(t+\tau) = x + \sum_{j=1}^{M} \nu_j P_j(a_j(x), \tau). \tag{2.13}$$

This approximation is the *τ-leaping method*.

In order to implement the tau-leaping technique effectively, a method is needed to efficiently estimate the largest value of τ satisfying the Leap Condition. The Leap Condition proposed by Gillespie [21] is satisfied in a weak sense if the expected change in each propensity function $a_j(x)$ during the leap is bounded by $\epsilon a_0(x)$, where ϵ is an error controlling parameter with $0 < \epsilon \ll 1$. In their paper, Gillespie and Petzold [22] proposed that the largest estimation of τ satisfying the above requirement can be considered as follows.

Firstly, calculate the auxiliary quantities $M^2 + 2M$

$$f_{jj'}(x) \equiv \sum_{i=1}^{N} \frac{\delta a_j(x)}{\delta x(i)} \nu_{ij'},\ j, j' = 1, .., M, \tag{2.14}$$

$$\mu_j(x) \equiv \sum_{j'=1}^{M} f_{ij'}(x) a_{j'}(x),\ j = 1, ..., M, \tag{2.15}$$

$$\sigma_j^2(x) \equiv \sum_{j'=1}^{M} f_{jj'}^2(x) a_{j'}(x),\ j = 1, ..., M, \tag{2.16}$$

then calculate τ as:

$$\tau = \min_{j \in [1,M]} \{ \frac{\epsilon a_0(x)}{\mid \mu_j(x) \mid} \frac{(\epsilon a_0(x))^2}{\sigma_j^2(x)} \}. \tag{2.17}$$

Here, $\mu_j(x)\tau$ estimates the expected mean change in $a_j(x)$ during a time interval of length τ, $\sqrt{\sigma_j^2(x)\tau}$ estimates the expected standard deviation of $a_j(x)$, and equation (2.17) is bounded by $\epsilon a_0(x)$ for all $1 \leq j \leq M$. Recently, an improved tau selection formula was proposed by Cao et al [13], which is:

$$\tau = \min_{i \in I_{rs}} \{ \frac{\max\{\epsilon x_i / g_i, 1\}}{\mid \mu_j(x) \mid}, \frac{\max\{\epsilon x_i / g_i, 1\}^2}{\sigma_j^2(x)} \} \tag{2.18}$$

where μ_i and σ_i^2 are defined by

$$\mu_j(x) \triangleq \sum_{j \in J_{ncr}} \nu_{ij} a_j(x), \text{ for all } i \in I_{rs} \tag{2.19}$$

$$\sigma_j^2(x) \equiv \sum_{j=1}^{M} \nu_{ij}^2 a_j(x), \text{for all } i \in I_{rs} \tag{2.20}$$

and g_i is the order of the highest reaction of species S_i and $\epsilon_i = \dfrac{\epsilon}{g_i}$ where ϵ is the user-specified tolerance. Also I_{rs} represents the set of all reactant species indeces and J_{ncr} is the set of all non-critical reactions indeces. A reaction is called critical if it is within n_c firings of eliminating one of its reactants ($n_c < 10$). Otherwise, a reaction is called non-critical. We note that the above tau-selection strategies, while computationally effective, are not very accurate. Their drawback is that the leap condition is satisfied only in weak sense (in mean and in variance).

(i) Explicit Tau-leaping:

This method gives an explicit formulation to update the system state X at time $t + \tau$, given that $X(t) = x$. The tau-leaping algorithm for (2.2) was proposed by Gillespie [21] and can be briefed as follows:

Explicit Tau-leaping Algorithm

Initialize $X(t_0) = x_0$ at time t_0

while $t < t_{final}$

Generate $a_j(x)$, the propensity function for $j = 1, 2, ..., M$ and also step size τ that satisfies the leap condition according to the leap selection strategy.

Draw samples $(p_j)_{j=1}^{M}$ from independent Poisson variables $P_j(a_j(X(t)))$ for all $1 \leq j \leq M$.

Set $X(t+\tau) = x + \sum_{j=1}^{M} \nu_j P_j(a_j(x), \tau)$ and set $t = t + \tau$.

end while

Stiffness arises when well-separated fast and slow scales are present in the system, with the fast dynamics being stable. This is often encountered in models of biochemical systems arising in applications, some reactions being fast and others being slow. The explicit tau-leaping method exhibits instability for stiff systems when large step sizes are employed. The method is essentially an extension of the explicit Euler method for ODEs to discrete stochastic systems, and as such it is conditionally stable.

(ii) Implicit Tau-leaping: In numerous applications, issues of stiffness may arise. Rathinam et al. [47], investigated the effective simulation of stiff models of stochastic discrete biochemical systems. Note that an implicit equation may be written in the deterministic term $a_j(X(t+\tau))\tau$, while the stochastic term $(P_j(a_j(x), \tau) - a_j(x)\tau)$, that has zero mean, is computed at the current time t. The

implicit τ- leaping formula can be derived from the explicit method (2.2) as follows:

$$X(t+\tau) = x + \sum_{j=1}^{M} \nu_j \left[\tau a_j(X(t+\tau)) + P_j(a_j(x), \tau) - \tau a_j(x)\right] \tag{2.21}$$

This equation is solved by, for example, Newton's method for estimating $X(t+\tau)$. This is similar to the deterministic case. The implicit tau-leaping method was proposed for overcoming the limitation on the step-size due to stiffness. The implicit tau-leaping algorithm shares similarities with the explicit tau-leaping method, except that the system state update is (2.21) instead of (2.13). Also, an extra implicit solver is applied, e.g. Newton's method.

2.2.2.2 Hybrid Methods

Hybrid models and techniques aim to speed-up the simulation by representing the species with large molecular amounts by more efficient models (e.g. CLE or RRE), while species with low population numbers are modeled with the CME. These tools are combining the continuous and deterministic, or continuous and stochastic modelling, as well as the discrete and stochastic methodologies to study the behaviour of homogeneous biochemical systems [2, 8, 28, 46]. This approach may be useful for stochastic biochemical networks with large amounts of molecules, still such methods may neglect some stochastic variations which occur when a few molecules of certain species are available. By contrast, stochastic simulation techniques represent these random fluctuations accurately.

These strategies pick a division criterion that permits to arrange the reactions into

fast, moderate and slow subsystems. The dynamics of the fast subsystem is expected to advance independently of the moderate/slow subsystems in a given step of the latter ones. However, the dynamics of the moderate subsystem is in general considered dependent of the fast system. This variation in approach is due to the fact that the moderate subsystem can not develop independently of the first one as the molecules of species that react in moderate reactions are, usually, species whose amounts are changed by fast reactions. In addition, fast reactions occur more often than moderate reactions and the adjustment induced by the event of the moderate reactions may be reduced, compared to the change induced by the event of the fast reactions. Also, the synchronization of the time-stepping for the various simulation subsystems is needed (e.g. sum of molecules and concentrations).

Some hybrid algorithms, due to Neogi [43], Alfonsi et. al. [2], Salis [48], are used in a stochastic algorithm, considering time-varying propensities, for partitioning of the fast reaction subsystem and the slow subsystem simulation. Also, a number of these methods [28, 45, 31] are used as a combination of stochastic simulation and numerical integration for ODEs/SDEs without considering time-varying propensities in the simulation of the slow subsystem. Puchalka et.al. [46] proposed a hybrid strategy based on a combination of the Next Reaction Method and the tau-leaping scheme.

It may happen that reactions in the fast subsystem can evolve in a way, such that to recompute them dynamically one must embed them into the moderate sub-system,

and viceversa. Hybrid strategies have a dynamical partitioning in reactions or species. Characterization of these methods is based on which set of techniques are used (SSA,ODE's, tau - leaping, SDE), regardless of whether it utilizes dynamic/automatic or user characterized partitioning, or it is considering the time dependent or constant propensities [45]. Pahle [45] gave a comprehensive survey of the state-of-the-art hybrid techniques for stochastic biochemical networks.

2.2.2.3 Euler-Maruyama Method for CLE

A differential equation of the form:

$$dX(t) = f(X(t))dt + g(X(t))dW(t) \tag{2.22}$$

is known as a stochastic differential equation, shortly SDE, where $W(t)$ is an independent Wiener process. The initial condition is $X(0) = X_0$. By considering X_0 =constant and if $g \equiv 0$, we can obtain an ordinary differential equation, namely

$$\frac{dX(t)}{dt} = f(X(t)), \text{ for } t = 0,$$
$$\text{with } X(t_0) = X_0.$$

Now, we present a numerical method for SDE. When integrating on a time interval $[0, T]$, we choose a time step $\triangle t = T/L$ where L is a positive integer and set, $t_j = j \triangle t$, $j = 0, 1, 2, , ..., L$. Denote by $X_j \simeq X(t_j)$. The *Euler-Maruyama method* for the SDE (2.22) can be formulated as:

$$X_j = X_{j-1} + f(X_{j-1}) \triangle t + g(X_{j-1})(W(\tau_j) - W(\tau_{j-1}))$$

where $j = 1, 2, ..., L$.

The strong convergence of the Euler-Maruyama method requires a bound for the expected value of the difference between the exact solution $X(t_j)$ and the numerical approximation X_j, $E|X_j - X(t_j)|$, where E represents the *expected value*. The convergence of a numerical method for solving (2.22) is said to be of *strong order of convergence* γ if:

$$E|X_j - X(t_j)| \leq C \triangle t^{\gamma}$$

$$\text{for any } t_j = j \triangle t,\ 0 \leq t_j \leq T.$$

Here C is a constant independent of the step $\triangle t$, for $\triangle t$ sufficiently small.

The Euler-Maruyama method has strong order of convergence for $\gamma = \frac{1}{2}$ if f and g satisfy appropriate conditions [33]. Euler-Maruyama for CLE becomes

$$X(t+\tau) = X(t) + \sum_{j=1}^{M} \nu_j a_j(X(t)) \triangle t + \sum_{j=1}^{M} \nu_j \sqrt{a_j(X(t)) \triangle t} \triangle W_j(t),$$

where $\triangle W_j(t) = W_j(t+\tau) - W_j(t)$.

2.2.2.4 Stiff/non-stiff solvers for RRE with non-negative option for population numbers

Stiffness is exhibited for a system of ODEs, which has both fast and slow dynamics. The fast dynamics is such that the trajectory approaches the stable manifold. After a short transient, the *slow modes* determine the dynamics of the system.

The evolution of such a system displays a rapid change for a short time interval, known as the transient (of time-scale given by the fast modes). After the transient,

the system is advanced slowly, according to the time scales of the slow modes. On the transient, the model is non-stiff, while after it the model becomes stiff. Explicit methods use time steps that are similar to the fastest time scale. The explicit methods advance the solution from one time to the next by approximating the slope of the solution curve at or near the beginning of the time interval. Larger time steps for explicit methods lead to numerical instability. This is a significant disadvantage of explicit methods when applied to stiff models.

On the other hand, an implicit method does not approximate the slope of the trajectory near the beginning of the interval of a time step. Instead, it gives more weight to the slope at the unknown point at the end of the current time step. This tends to avoid the above-described instability, but at the cost of having to solve a nonlinear system of equations for the future point, at each time step. The gain in speed-up of implicit over explicit methods on stiff problems may be significant. Among the siff solvers in MATLAB that can be used to solve stiff RREs, we mention, 'ode15s', 'ode23s' or 'ode23tb'. For non-stiff RRE, one can use 'ode45', 'ode25' or 'ode113'. The option 'NonNegative' will prevent negative population numbers in each molecular species.

Chapter 3

ADAPTIVE METHOD FOR TAU-LEAPING STRATEGY FOR CME

We propose below an adaptive method for the tau-leaping strategy to solve numerically the Chemical Master Equation. When the numerical solution does not satisfy the leap condition, the step is rejected. The method requires that the leap condition is satisfied on every step, on every trajectory, therefore we predict that it gives a very accurate solution. To ensure that the rejection does not result in a biased numerical solution, we condition the random variables for the step chosen after rejection on the random variables of the step that was rejected. The theoretical framework on which this method is based was provided by Anderson [3] in the context of post-leap checks.

For the tau-leaping strategy, a step is rejected when the Leap Condition is not fulfilled or at the point when some population numbers become negative. Rejections should be done such that the approximation is not biased. When the solution violates the accuracy criteria, the step is rejected, but the samples of the Poisson random variables already created are stored for future use. When a smaller step is attempted, the new random quantities for this step should be conditioned on the samples of the Poisson variables already created for the larger step τ, to guarantee that the numerical solution is computed on the correct trajectory is followed.

For this conditioning, we applied a method proposed by Anderson [3] for post-leap checks.

For predicting a step τ, one generates the number of reactions R_j fired between $[t, t+\tau]$ by sampling from a Poisson random variable $P_j(a_j(x(t)), \tau)$. We denote this random number by $p_{\tau,j}$. If the error criteria corresponding to the predicted step τ is not satisfied, then τ is rejected. However, the number of reactions R_j, $p_{\tau,j}$ is recorded for future use. A new step, τ', is tried such that $0 < \tau' < \tau$. Given the future constraint, the number of R_j firings in $[t, t+\tau']$ denoted by q_j depends on how many reactions happened between $[t, t+\tau]$. In fact, the number of times R_j happens in $[t, t+\tau']$ should be *conditioned* on the number of times R_j occurs in $[t, t+\tau]$. According to Anderson [3], the correct way of conditioning is to choose q_j, a sample from a binomial distribution $B_j(p_{\tau,j}, \tau'/\tau)$. If there are $p_{\tau,j}$ reactions R_j in $[t, t+\tau]$ and only q_j in $[t, t+\tau']$, then clearly they are $p_{\tau,j} - q_j$ such reactions in $[t+\tau', t+\tau]$. This strategy ensures that the right sample path is maintained (see also Anderson (3)).

Theorem:

If $Y(t)$ is a Poisson process with intensity a, and $0 \leq s < \mu < t$, then, the increment $Y(\mu) - Y(s)$ conditioned on $Y(s)$ and $Y(t)$, has a binomial distribution, $B(Y(t) - Y(0), \frac{\mu - s}{t - s})$.

Proof: We may assume, without loss of generality, that $s = 0$ and $Y(0) = 0$. Consider $Y(t) = p$ and $0 < \mu < t$.

From the definition of the conditional probability, we get

$$P(Y(\mu)) = j \mid Y(t) = p) = \frac{P(Y(\mu) = j) \cap (Y(t) = p))}{P(Y(t) = p)} \qquad (i)$$

The fraction is equal to

$$\frac{P(Y(\mu) = j) \cap (Y(t) = p))}{(P(Y(t) = p))} = \frac{P(Y(t) - Y(\mu) = p - j)P(Y(\mu) = j)}{P(Y(t) = p)} \qquad (ii)$$

From the properties of a Poisson distribution, we can write that

$$\frac{P(Y(t) - Y(\mu) = p - j)P(Y(\mu) = j)}{P(Y(t) = p)} = \frac{e^{-a(t-\mu)}(a(t-\mu))^{p-j})(a\mu)^j e^{-a\mu} p!}{j! e^{-at}(at)^p (p-j)!} \qquad (iii)$$

From (i), (ii) and (iii), we get, after simplifying the expression

$$P(Y(\mu) = j \mid Y(t) = p) = \binom{p}{j}(1 - \frac{\mu}{t})^{p-j}(\frac{\mu}{t})^j$$

i. Leap Condition

We remark that numerically imposing the leap condition in terms of molecular population numbers is preferred to imposing the condition in the terms of propensity functions (see Cao et al [13]). The former gives a more accurate numerical solution. The version of the leap condition imposed on the molecular amount is

$$|X_i(t + \tau) - X_i(t)| \leq max\{\epsilon_i X_i(t), 1\} \qquad (3.1)$$

with $i \in I_{rs}$ where I_{rs} is the set of indices of all reactant species (so $i \in I_{rs}$ if and only if at least one propensity function depends on X_i). The equation (3.1) requires that the relative change in every species X_i is less than some small parameter ϵ_i, or the absolute change in the population number is by at least one molecule. The value of ϵ_i depends on the user specific tolerance according to:

$$\epsilon_i = \frac{\epsilon}{g_i}, \qquad (3.2)$$

for each $i \in I_{rs}$. Here the tolerance ϵ is bounded by $0 < \epsilon < 1$.

The function $g_i = g_i(x)$ depends on the highest order of reaction in which the species S_i occurs in a reaction.

(i) If **highest order of reaction** is 1, we have to choose $g_i = 1$.

(ii) If **highest order of reaction** is 2, take $g_i = 2$, but in the event that any second-order reaction requires two S_i molecules, take

$$g_i = \left(2 + \frac{1}{x_i - 1}\right).$$

(iii) If **highest order of reaction** is 3, choose $g_i = 3$, but in the event that some third-order reaction requires two S_i molecules, take

$$g_i = \frac{3}{2}\left(2 + \frac{1}{x_i - 1}\right).$$

The exception is if some third-order reaction requires three S_i molecules, then choose

$$g_i = \left(3 + \frac{1}{x_i - 1} + \frac{2}{x_i - 2}\right).$$

According to Cao et. al. [13], the leap condition (3.1) produces a more accurate numerical solution than the version of the leap condition based on propensities:

$$| a_j(X(t+\tau)) - a_j(X) | \leq \max\{\epsilon a_j(x), c_j\},\ 1 \leq j \leq M.$$

ii. Variable time-stepping strategy for the tau-leaping method

Before we propose a reliable variable time-stepping strategy, we first discuss our error criteria. The error criteria we use requires that the leap condition is satisfied. Thus, for each step, each molecular population X_i must obey the condition (3.1).

Denote by $X_i(t_n)$ the approximation of the number of S_i molecules at time t_n. Let us split the set of reactants in two distinct subsets: X_{low} is the subset of the species in low molecular amounts and X_{large} is the subset of the species in large molecular amounts, for the interval $[t_n, t_{n+1}]$, with

$$t_{n+1} = t_n + \tau_n.$$

For the user-specific tolerance ϵ, the species S_i belong to the subset of *low* number species when

$$\max\{\epsilon_i X_i(t_n), 1\} = 1 \tag{3.3}$$

Otherwise the species S_i belongs to the *large* subset, that is

$$\max\{\epsilon_i X_i(t_n), 1\} = \epsilon_i X_i(t_n). \tag{3.4}$$

Note that, for any $X_i(t_n)$ with a large population the leap condition (3.1) means that, the relative change number of molecules is such that

$$r_i(t_n) = \frac{|X_i(t_{n+1}) - X_i(t_n)|}{\epsilon_i |X_i(t_n)|} \leq 1 \tag{3.5}$$

The leap condition (3.1) applied to any $X_i(t_n)$ with a low population implies that the absolute change in its number of molecules is bounded by:

$$A_i(t_n) = |X_i(t_{n+1}) - X_i(t_n)| \leq 1. \tag{3.6}$$

We can now define the error for the species with large population numbers as:

$$r_{large}(X^n, X^{n+1}, \tau^n) = \max_{\{i:\ X_i(t_n)\ large\}} \{r_i(t_n)\} \tag{3.7}$$

while for the species with low population the error criteria is defined as:

$$A_{low}(X^n, X^{n+1}, \tau^n) = \max_{\{i:\ X_i(t_n)\ low\}} \{A_i(t_n)\} \tag{3.8}$$

over the time interval $[t_n, t_{n+1}]$.

Let us now discuss step-size adaptivity for the tau-leaping strategy.

The idea behind the variable time-stepping technique we wish to propose extends the standard technique for predicting the future step in the numerical ODE solving [27], [49] and numerical SDE solving [10, 9]. For ODEs and SDEs, the future step τ^{n+1} is predicted as:

$$\tau^{n+1} = \tau^n \left(\frac{\theta}{e(X^n, \tau^n)} \right) \tag{3.9}$$

where $e(X^n, \tau^n)$ is the previous error measurement and τ^n is the previous step when the scaled error is required to obey:

$$e(X^n, \tau^n) \leq 1,$$

The parameter of θ in (3.9) is a safety factor. It is introduced to minimize the occurrences of step rejections and it is restricted by $0 < \theta < 1$. To avoid the number of step rejections the step should not be too large or too small compared to the previous step. Thus, instead of (3.9) the following strategy to predict the future step is used in implementation [10]

$$\tau^{n+1} = \tau^n \min \left(\alpha, \max \left(\beta, \left(\frac{\theta}{e(X^n, \tau^n)} \right) \right) \right) \tag{3.10}$$

where the maximal step increment factor $\alpha > 1$ and the minimal step decrease factor $\beta < 1$ are chosen to decrease the chances to reject a step.

In the case of stochastic discrete models of biochemical kinetics simulated by the tau-leaping method, the scaled relative change in amount,

$$\frac{|X_i(t_{n+1}) - X_i(t_n)|}{\epsilon_i |X_i(t_n)|},$$

for the large population species should satisfy:

$$r_{large}(X^n, X^{n+1}, \tau^n) \leq 1. \tag{3.11}$$

Applying a predictor similar to (3.10)

$$\tau_1^{n+1} = \tau^n \; min\left(\alpha, \; max\left(\beta, \left(\frac{\theta}{r_{large}(X^n, X^{n+1}, \tau^n)}\right)\right)\right)$$

Since the scaled absolute change in quantity for the small population species $|X_i(t_{n+1}) - X_i(t_n)|$ should be such that:

$$A_{low}(X^n, X^{n+1}, \tau^n) \leq 1 \tag{3.12}$$

then, similarly, the next step for these species may be chosen as:

$$\tau_2^{n+1} = \tau^n \; min\left(\alpha, \; max\left(\beta, \left(\frac{\theta}{A_{low}(X^n, X^{n+1}, \tau^n)}\right)\right)\right)$$

We require that the error criteria (3.11) and (3.12) are satisfied, so we can choose

$$\tau^{n+1} = min(\tau_1^{n+1}, \tau_2^{n+1}). \tag{3.13}$$

We have the following cases in our variable step size method.

Case 1: No future condition When no future step was created on the current path, then the number of reactions of type R_j on the time interval $[t, t + \tau']$ are

computed by sampling a Poisson distribution $P_j(a_j(X(t)), \tau')$ where $X(t) = x$. If we denote this sample by $p_{\tau',j}$, then the tau-leaping method applied on the current step is:

$$X(t + \tau') = X(t) + \sum_{j=1}^{M} \nu_j p_{\tau',j}.$$

The error criteria (3.11) and (3.12) will be then verified and, if the step is accepted, the solution is advanced to $t+\tau'$. Otherwise a step in the future was created $(t+\tau')$, along with the samples of Poisson distributions, $p_{\tau',j}$, which will be used for future conditioning.

Case 2: Future condition and step before the condition Assume that a step was created in the future (on the time interval $[t, t+\tau]$) and the corresponding leap, τ, was rejected as either (3.11) or (3.12) was violated. Associated to the future step, the samples $p_{\tau,j}$ were sampled from Poisson random variables.

A new, smaller step τ' is then chosen $(0 < \tau' < \tau)$. The number of reactions of type R_j on the smaller interval $[t, t + \tau')$, q_j, are computed by conditioning on the number of reactions of the same type in $[t, t + \tau]$, namely $p_{\tau,j}$. According to the Theorem above we obtain that q_j is a sample from $B_j(p_{\tau,j}, \tau'/\tau)$, where B_j are binomial random variables.

The tau-leaping strategy applied to the time interval $[t, t + \tau']$ may be written as:

$$X(t + \tau') = X(t) + \sum_{j=1}^{M} q_j \nu_j \tag{3.14}$$

where $X(t) = x$. If the solution (3.14) satisfies (3.11) and (3.12), then the step τ' is accepted and the numerical solution is advanced at time $t + \tau'$.

Case 3: Future condition and step after the condition If on the time interval $[t, t + \tau]$ the step τ was rejected (and $p_{\tau,j}$ were created and stored) and a step $0 < \tau' < \tau$ was accepted (and q_j were created), then on the current time interval $[t + \tau', t + \tau)$ there are

$$p_{\tau,j} - q_j$$

reactions of type R_j.

A new step is predicted using formula (3.13). Denote this step by $\tau'' > \tau$. The number of reactions of type R_j between $[t + \tau', t + \tau'']$ equal to the number of R_j firings in $[t+\tau', t+\tau]$ to which we add the number of R_j occurrences in $[t+\tau, t+\tau'']$. Since there is no future condition in $[t + \tau, t + \tau'']$, the latter is r_j, a sample from a Poisson random variable $P_j(a_j(X), (\tau'' - \tau))$.

The total number of firings of the j-th reaction channels in $[t + \tau', t + \tau'']$ is

$$(p_{\tau,j} - q_j) + r_j.$$

Thus the tau-leaping estimation of the system state at $X(t + \tau'')$ is

$$X(t + \tau'') = X(t) + \sum_{j=1}^{M} [(p_{\tau,j} - q_j) + r_j]\nu_j. \tag{3.15}$$

If the new step is accepted, then the solution is advanced to $t + \tau''$ and no condition exists in the future, for the current path.

Note that the split in the subsets of low population species and of large population species is done dynamically in our algorithm.

Below we present our new adaptive tau-leaping strategy for simulating well-stirred biochemical networks.

Algorithm	
Set tolerance ϵ, the control factor θ, the minimal factor α, the maximal factor β, the initial step τ_0 and the number of trajectories.	**Step 0: Initializing**
For every trajectory do 2-6.	**Step 1: Loop for trajectories**
Set $X \leftarrow x_0$ for $t \leftarrow 0$, the elementary step $\tau \leftarrow \tau_0$, $\tau' \leftarrow \frac{\tau}{2}$	**Step 2: Initializing a trajectory**
While $(t < T_{final})$ do 4-6.	**Step 3: loop for one trajectory**
Set X_{low} and X_{large}	**Step 4: Low and large subsets**
Repeat 6-7 until a step is accepted.	**Step 5: Steps**
If constraint: **(A)** Sample Poisson distributions: $p_{\tau',j} \leftarrow P_j(a_j(X), \tau')$ (B) Find X' using (2.13). **(C)** c_1**:** If $(e(X, X', \tau') \leq 1)$, then accept step. Update, $t \leftarrow t + \tau'$ and $X \leftarrow X'$. c_2**:** Else, reject step. Future constraint exists and is recorded. $\tau \leftarrow \tau'$, $p_{\tau,j} \leftarrow p_{\tau',j}$. **(D)** Update τ' using (3.13).	**Step 6: No future condition**.

Algorithm

If ($\tau' \leq \tau$), then

(A) Sample binomial distribution:

$q_j \leftarrow B_j(p_{\tau,j}, \tau'/\tau)$

(B) Find X' using(3.14).

(C) c_1**:** If ($e(X, X', \tau') \leq 1$), then accept step.

Update $t \leftarrow t + \tau'$ and $X \leftarrow X'$.

(c_2**:**) Else, reject step.

(D) Update τ' using (3.13). **Step 7 (7.1): Step before constrain**.

Else ($\tau' > \tau$):

(A) Sample Poisson distribution and set:

$p_{\tau',j} \leftarrow P_j(a_j(X), \tau' - \tau) + p_{\tau,j} - b_j$

(B) Find X' using (3.15)

(C) c_1**:** If ($e(X, X', \tau') \leq 1$), then accept step.

Update, $t \leftarrow t + \tau'$ and $X \leftarrow X'$.

Future constraint removed.

c_2**:** Else, reject step.

Store the new future constraint. $\tau \leftarrow \tau'$,

$p_{\tau,j} \leftarrow p_{\tau',j}$.

(D) Update τ' using (3.13). **Step 7 (7.2): Step after constraint**.

The high accuracy of the proposed variable time-stepping method is due to the application of the exact leap condition (3.1) for each accepted step. The leap condition is not applied approximately, as in the existing strategies in the literature for the numerical solution of the CME.

We wish to emphasize that our algorithm employs a dynamic partitioning of the subset of low and large molecular species. More precisely, we do not assume that the subset of low and large biochemical species are fixed during the integration. This flexibility, while increasing slightly the computational cost of the simulation, makes our method applicable to a wide variety of well-stirred biochemical networks.

Chapter 4

NUMERICAL RESULTS

In this section, we illustrate the advantages of the proposed adaptive scheme for the tau-leaping method compared to the exact stochastic simulation algorithm of Gillespie. The numerical tests are performed on a numbers of biochemical networks arising in practice. In each case, the accuracy of the adaptive scheme is studied, more precisely we compare the numerical results generated with our method, with those obtained with Gillespie's algorithms.

Simulations were done for all three models for 10,000 trajectories. We chose $\alpha = 2$, $\beta = 0.5$, the given tolerance $\epsilon = 0.5$ and the value of the safety parameter $\theta = 0.8$. Also, we calculated the efficiency gain of the variable step-size tau-leap technique by the ratio of the CPU-time of the SSA and that of our model,

speed-up = CPU (SSA)/ CPU (adaptive tau-leaping).

4.1 Simple Reaction Model

In order to study the accuracy and efficiency of the proposed strategy, we initially apply this technique on a simple reaction model [15]. The model consists of three species S_1 ,S_2 , and S_3 and three reaction channels. We choose the parameter values $c_1 = 1, c_2 = 1000$ and $c_3 = 0.0005$ with initial conditions $X_1(0) = 5 \times 10^3, X_2(0) =$

Case	Reaction	Propensity
1	$S_1 + S_2 \to S_3$	$c_1 X_1 X_2$
2	$S_3 \to S_1 + S_2$	$c_2 X_3$
3	$S_1 + S_4 \to S_5 + S_1$	$c_3 X_4 X_1$

Table 4.1: Simple Reaction Model

$5 \times 10^3, X_3(0) = 10^2$ and $X_4(0) = 10^2$. The integration is performed on the time interval [0,0.01]. We remark that the problem is stiff, since the propensities can be partitioned into the fast and slow groups. Numerical results are given based on simulation of 10,000 sample paths.

The efficiency gain achieved by our variable time-stepping technique over the SSA, for the increased levels of stiffness for the biochemical system in Table 4.1, are shown in Table 4.2.

-	$\xi = 1$	$\xi = 50$	$\xi = 100$
Speed-up of adaptive tau over SSA	4.99	9.80	11.82

Table 4.2: Simple reaction model: speed-up of the new tau-leaping time-stepping over the SSA, for the model reaction rate parameters $[\xi c_1, \xi c_2, c_3]$.

This shows that our adaptive tau-leaping method between 5 to 12 times faster than the SSA, for the sets of parameters considered.

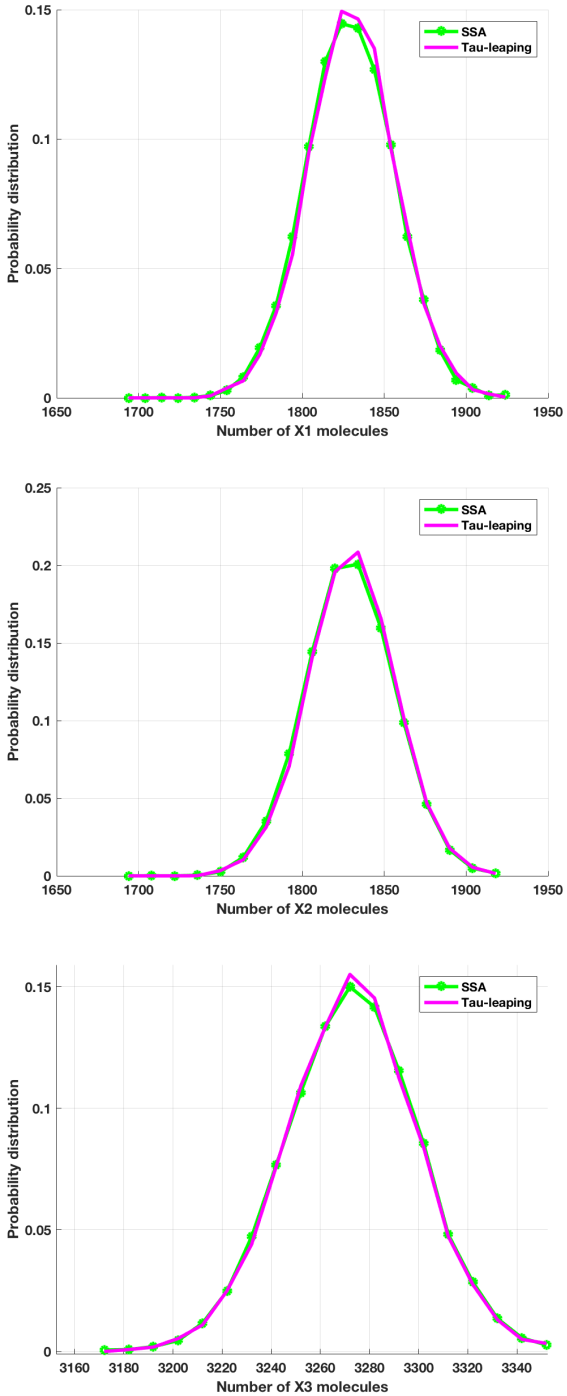

Figure 4.1: Simple reaction model: Comparison of the histograms of species X_1, X_2 and X_3 computed at T=0.01 using the SSA and the adaptive tau-leaping scheme.

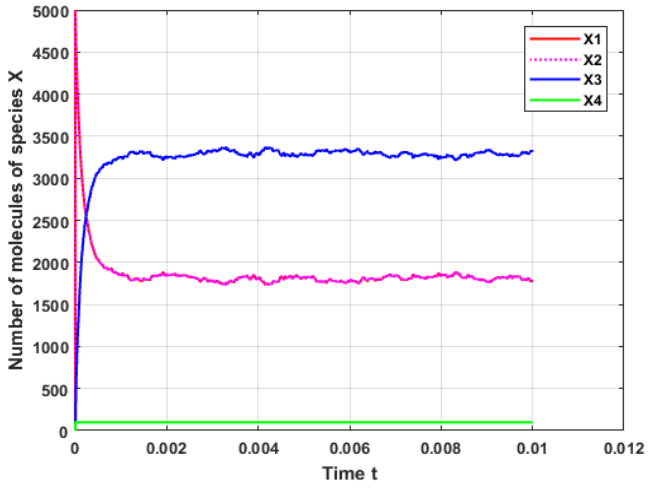

Figure 4.2: The evolution in time t of all species in the simple reaction model, for one sample path.

Figure 4.1 displays the histogram for the species S_1, S_2 and S_3 obtained using the adaptive tau-leaping methods and the exact method SSA, respectively, for time $t = 0.01$.

From Figure 4.1, it is clear that our proposed adaptive method has very similar accuracy compared to the exact stochastic simulation algorithm (or Gillespie's Direct Method), thus our variable time-stepping tau-leaping strategy is very accurate, while being more efficient than the SSA.

4.2 Cycle Model

Consider now a system consisting of five species S_1, S_2, S_3, S_4 and S_5 and five reactions [38]. The biochemically reacting system, representing a cycle model is presented in Table 4.2.

Case	Reaction	Propensity
1	$S_1 \rightarrow S_2$	$c_1 X_1$
2	$S_2 \rightarrow S_3$	$c_2 X_2$
3	$S_3 \rightarrow S_1$	$c_3 X_3$
4	$S_1 + S_4 \rightarrow S_5$	$c_4 X_1 X_4$
5	$S_5 \rightarrow S_1 + S_4$	$c_5 X_5$

Table 4.3: Cycle Model

This system has reaction rate constants as follows,

$c_1 = 1.5 \times 10^3, c_2 = 5 \times 10^3, c_3 = 10^3, c_4 = 1.66 \times 10^-4$ and $c_5 = 8 \times 10^-2$, while the initial conditions are $X_1(0) = 1000, X_2(0) = 800, X_3(0) = 400, X_4(0) = 40$ and $X_5(0) = 50$.

The state-change vectors of the system reactions are the corresponding columns

appearing in the system's stoichiometric matrix below:

$$\nu = \begin{bmatrix} -1 & 0 & 1 & -1 & 1 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix},$$

The cycle model is integrated on the time interval $[0, 0.01]$.

Figure 4.3 gives the histograms at time $T = 0.01$ of the species S_1, S_2, S_3 and obtained with the SSA and the improved variable tau-leaping method. This figure shows that the proposed variable time-stepping strategy has excellent accuracy, when compared to the exact SSA, for this model. The histograms for species S_4 and S_5, which have similar accuracy, were omitted. The speed-up of the adaptive tau-leaping scheme over the SSA is

$$\text{speed-up} = 8.61.$$

In conclusion, our method is almost an order of magnitude faster than the SSA, while maintaining similar accuracy.

Figure 4.4 shows a sample path with the evolution of the state vector X as a function of time.

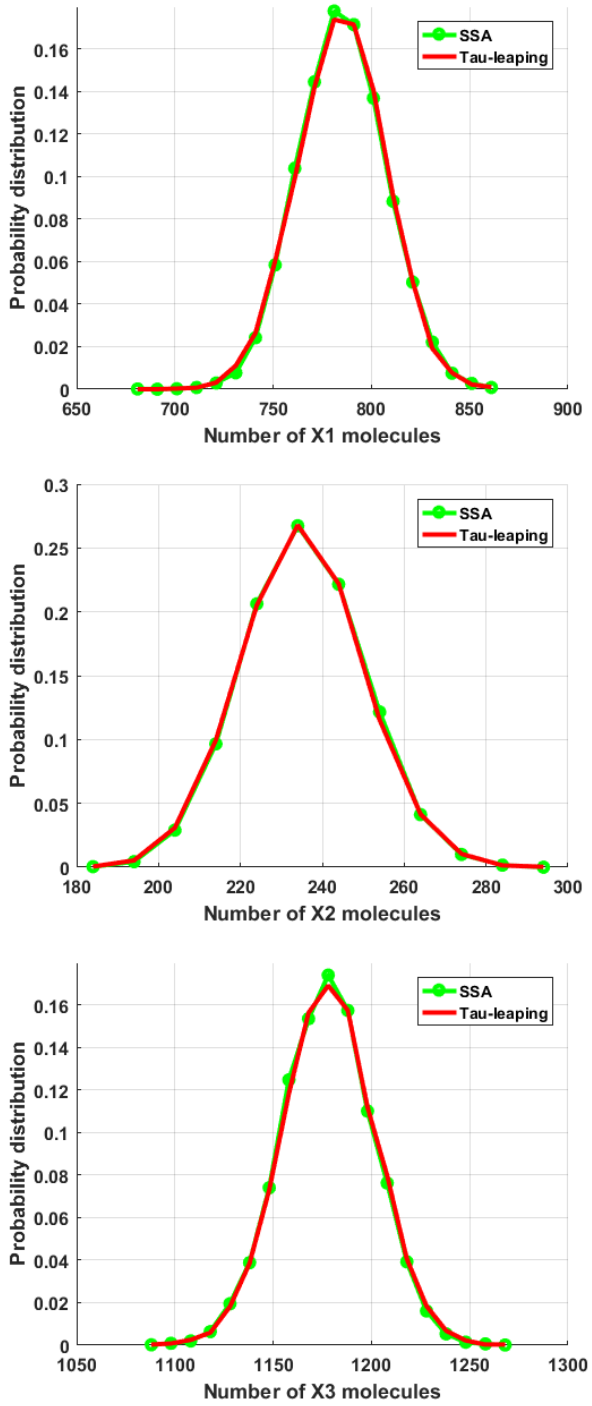

Figure 4.3: Cycle model: The histograms of species X_1, X_2 and X_3 at time T=0.01 computed with the SSA and the adaptive tau-leaping scheme.

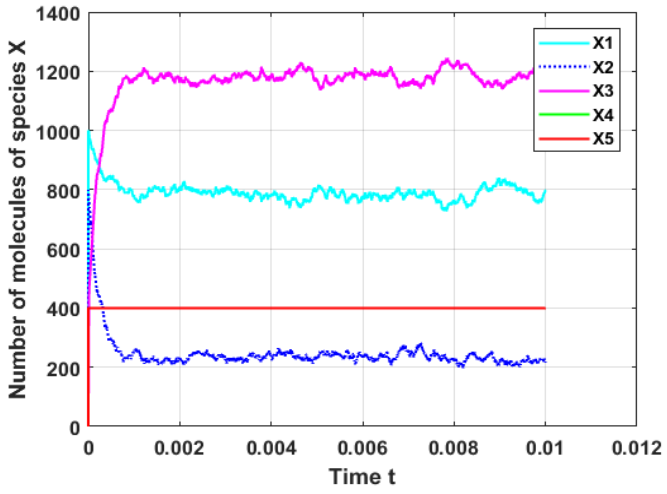

Figure 4.4: The evolution in time t of all species in the Cycle model, for one sample path.

4.3 Complex Model

Finally, we study, a complex model representing a genetic network, [40] . This model is composed of twelve reactions and ten species ($A, B, S_A, S_B, S_AB, S_AB_2, S_BA, S_BA_2, P_A$ and P_B respectively). The initial values of the molecular amounts are $A(0) = 800, B(0) = 800, S_A(0) = 500, S_B(0) = 500, S_AB(0) = 400, S_AB_2(0) = 500, S_BA(0) = 400, S_BA_2(0) = 500, P_A(0) = 0$ and $P_B(0) = 0$. The simulation is performed on the time interval [0, 0.01]. The reactions, their propensities and the reaction rate constants are shown in Table 4.3.

Case	Reaction	Propensity	Reaction rates
1	$S_A \to S_A + A$	$c_1 S_A$	$c_1 = 0.16$
2	$S_B \to S_B + B$	$c_2 S_B$	$c_2 = 0.16$
3	$S_A + B \to S_A B$	$c_3(S_A)B$	$c_3 = 40$
4	$S_A B \to S_A + B$	$c_4(S_A)B$	$c_4 = 2000$
5	$S_A B + B \to S_A B_2$	$c_5(S_A B)B$	$c_5 = 2.5$
6	$S_A B_2 \to S_A B + B$	$c_6(S_A)B_2$	$c_6 = 1600$
7	$A \to P_A$	$c_7 A$	$c_7 = 0.1$
8	$S_B + A \to S_B A$	$c_8(S_B)A$	$c_8 = 2$
9	$S_B A \to S_B + A$	$c_9(S_B)A$	$c_9 = 2000$
10	$S_B A + A \to S_B A_2$	$c_{10}(S_B A)A$	$c_{10} = 2.5$
11	$S_B A_2 \to S_B A + A$	$c_{11}(S_B)A_2$	$c_{11} = 1600$
12	$B \to P_B$	$c_{12} B$	$c_{12} = 0.1$

Table 4.4: Complex Model

We see in Figure 4.5 that for the complex reaction network, the proposed adaptive tau-leaping technique gives very good accuracy. Indeed, we remark that the histograms for species X_2, X_5 and X_7 generated with our method and with the (exact) SSA match very well (The histograms for the other species, of similar accuracy, were omitted for brevity). Moreover, our adaptive tau-leaping technique is more efficient than the SSA.

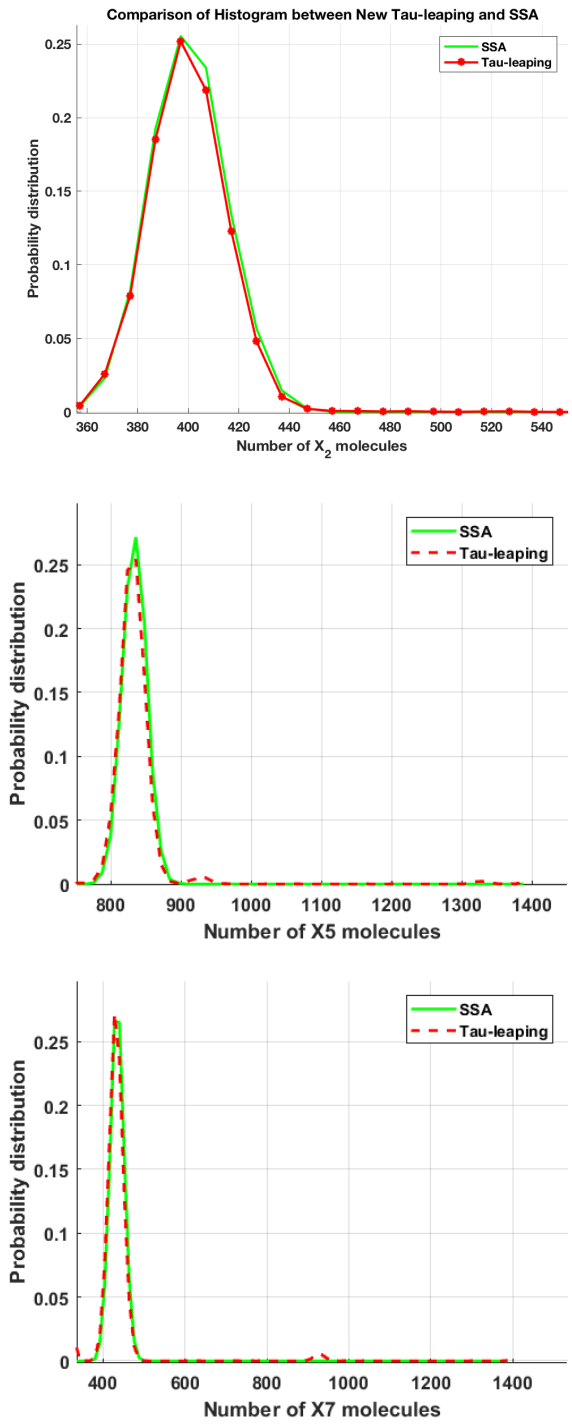

Figure 4.5: Complex model: the histograms of species X_2, X_5 and X_7 at time T=0.01 computed with the SSA and the adaptive tau-leaping scheme.

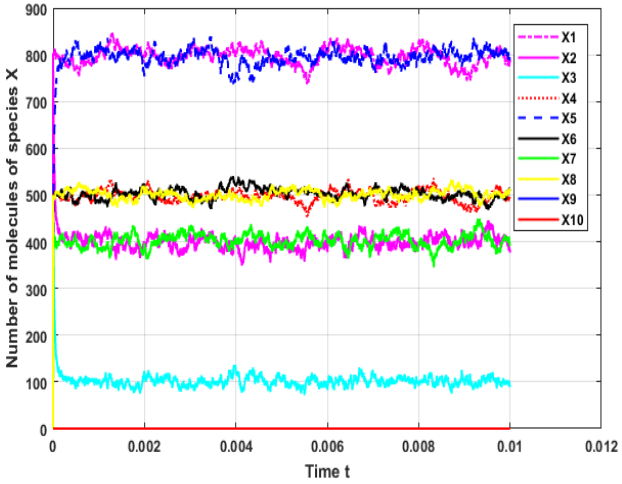

Figure 4.6: The evolution in time t of all species in the Complex model, for one sample path.

Figure 4.6 presents the dependence on time of all molecular species. Numerical simulations with the given set of parameters show that the efficiency gain is

$$\text{speed-up} = 3.53,$$

thus our method is almost 4 times more efficient than the SSA, while being very accurate. The diagram of this biochemical reaction network is represented in Figure 4.7, for reference.

We see that the adaptive time-stepping strategy for the tau-leaping scheme scales well with the dimension of the system. In our future work we shall consider how to reduce further the computational complexity of the adaptive algorithm depending on the dimension of the system.

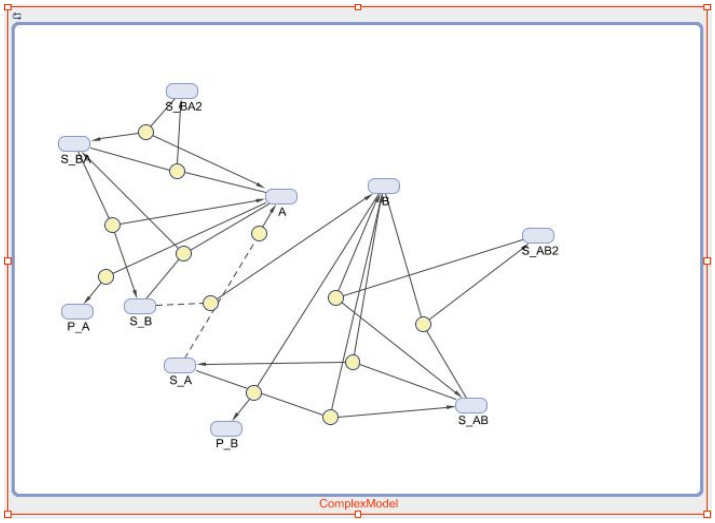

Figure 4.7: The diagram of the reaction network of the complex model

Chapter 5

CONCLUSION

One of the important research areas in Computational Biology focuses on stochastic modelling and simulation of complex networks of biochemical reactions. There are a wide range of interesting problems concerning the development of effective and reliable simulation tools for these stochastic models as well as for formulating the theoretical framework for studying these tools. We discussed some of the key stochastic models of well-stirred biochemical models as well as described the state-of the art simulation strategies for them. In particular, we presented the widely used stochastic model of well-stirred biochemical networks, the Chemical Master Equation.

In this thesis, we propose an improved adaptive time-stepping scheme for the tau-leaping strategy for approximating the solution of the Chemical Master Equation. The tau-leaping method has been effectively utilized to simulate numerous reaction systems emerging in applications. Our technique guarantees that the leap condition is satisfied over each step, on each trajectory, making it a very accurate strategy. Being an extension of variable time-stepping methodologies based on the integral controllers used for solving SDEs or ODEs, such methods are shown to be very efficient computationally.

Our method may sometimes prompt step rejections, when the accuracy criteria is not satisfied, still, the strategy guarantees that the statistics of the numerical solution is not biased. The variable tau-leaping strategy was shown to be more efficient than the stochastic simulation algorithm which is often employed for solving the Chemical Master Equation model for models which are moderately stiff. Our technique is quite flexible in chosing the step size in the tau-leaping method, being particularly effective on systems with multiple scales in time. It uses a dynamic splitting of the set of all reacting species into low and large molecular species. Our future work will focus on efficient and reliable strategies to adapt the time-step for higher order tau-leaping methods.

REFERENCES

[1] D. Adalsteinsson, D. McMillen, T.C. Elston, 2004, Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks, *BMC Bioinformatics* 5(24).

[2] A. Alfonsi et al., 2005, Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems, *ESAIM Proc.* 14, 1-13.

[3] D.F. Anderson, 2008, Incorporating postleap checks in tau-leaping, *J. Chem. Phys.*, 128(5), 054103.

[4] D.F. Anderson, A. Ganguly, T.G. Kurtz, 2011, Error analysis of tau-leap simulation methods, *The Annals of Applied Probability*, 21(6), 2226-2262.

[5] A.P. Arkin, J. Ross, H.H. McAdams, 1998, Stochastic kinetic analysis of developmental pathway bifurcation in phage-infected Escherichia coli cells, *Genetics* 149, 1633.

[6] K. Ball, T.G. Kurtz, L. Popovic, G. Rempala, 2006, Asymptotic analysis of multiscale approximations to reaction networks, *Ann. Appl. Prob.* 16, 1925–1961.

[7] M. Barrio, K. Burrage, A. Leier, T. Tian, 2006, Oscillatory regulation of Hes1: discrete stochastic delay modelling and simulation, *PLoS Comput. Biol.* 2(9), 1017–1030.

[8] M. Bentele, R. Eils, 2004, General stochastic hybrid method for the simulation of chemical reaction processes in cells. In, Computational Methods in Systems Biology *CMSB, LNCS, Heidelberg Springer*, 248-251.

[9] P.M. Burrage et al., 2004, Adaptive stepsize based on control theory for stochastic differential equations, *J. Comput. Appl. Math.* 170, 317–336.

[10] K. Burrage, P. M Burrage, 2002, A variable stepsize implementation for stochastic differential equations, *SIAM J. Sci. comput.* 24(3), 848-864.

[11] K. Burrage, P.M. Burrage, and T. Tian, 2004, Numerical methods for strong solutions of stochastic differential equations: an overview, *Proc. R. Soc. Lond. A* 460, 373–402.

[12] Y. Cao, D.T. Gillespie, L. Petzold, 2005, The slow-scale stochastic simulation algorithm, *J. Chem. Phys.* 122, 014116.

[13] Y. Cao, D.T. Gillespie, L. Petzold, 2006, Efficient step-size selection for the tau-leaping method, *J. Chem. Phys.* 124, 044109.

[14] Y. Cao, D.T. Gillespie, L.Petzold, 2007, Adaptive explicit-implicit for the tau-leaping with automatic tau-selection, *J. Chem. Phys.* 126, 224101.

[15] Y. Cao, L. Petzold, 2008, Slow scale tau-leaping method, , *Comput MethodsAppl Mech Eng.*; 197(43-44), 3472–3479.

[16] A. Chatterjee, D.G. Vlachos, M.A. Katsoulakis, 2005, Binomial distribution based tau-leap accelerated stochastic simulation, *J. Chem. Phys.,* 122, 024112.

[17] I.R. Epstein, J.A. Pojman, 1998, An introduction to nonlinear chemical dynamics: oscillations, waves, patterns and chaos. *Oxford University Press, Oxford.*

[18] J.H. Espenson, 1995, Chemical Kinetics and Reaction Mechanisms, *McGraw-Hill*, Singapore.

[19] M.A. Gibson, J. Bruck, 2000, Exact stochastic simulation of chemical systems with many species and many channels, *J. Chem. Phys.* 105, 1876-1889.

[20] D.T. Gillespie, 2000, The chemical langevin equations, *J. Phys. Chem.* 113, 297–306.

[21] D.T. Gillespie, 2001, Approximate accelerated stochastic simulation of chemically reacting systems, *J. Chem. Phys.* 115, 1716–1733.

[22] D.T. Gillespie, L. Petzold, 2006, Numerical simulation for biochemical kinetics, *Systems Modelling in Cellular Biology*, Eds. Z. Szallasi, J. Stelling, V. Periwal, MIT Press, 331–354.

[23] D.T. Gillespie, 1992 (a), A rigorous derivation of the chemical master equation, *physica A,* 188, 402-425.

[24] D.T. Gillespie, 1992(b), Markov Processes: An Introduction for Physical Scientists. *Academic Press.*

[25] D.T. Gillespie, 1977, Exact stochastic simulation of coupled chemical reactions, *J. Phys. Chem.* 81, 2340-2361.

[26] D.T. Gillespie, 1976, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, *Journal of computational physics*, 22(4), 403-434.

[27] E. Hairer, S.P. Nørsett, and G. Wanner, 2009, Solving Ordinary Differential Equations I, 2nd ed. *Springer, Berlin.*

[28] EL. Haseltine, JB. Rawlings, 2002, Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics, *J Chem Phys*, 117(15), 6959-6969.

[29] R. Heinrich, S. Schuster, 1996, The Regulation of Cellular Systems, *Chapman and Hall*, New York.

[30] D. J. Higham, 2008, Modeling and simulating chemical reactions, *SIAM Rev.* 50(2), 347–368.

[31] S. Hoops, S. Sahle, R. Gauges, et al., 2006, COPASI – a Complex Pathway Simulator, *Bioinformatics,* 22(24), 3067-3074.

[32] M. Kerker, 1974, Brownian movements and molecular reality prior to 1900. *J. Chem. Educ.* 51, 764-768.

[33] P.E. Kloeden, E. Platen, 1992, Numerical Solution of Stochastic Differential Equations. Berlin etc., *Springer-Verlag*, 632.

[34] W. Koh, K.T. Blackwell, 2012, Improved spatial direct method with gradient-based diffusion to retain full diffusive fluctuations, *J. Chem. Phys.*, 137, 154111.

[35] T.J Kurtz, 1978, Strong approximation theorems for density dependent Markov chains, *Stoch. Proc. Appl.*, 6(3), 223-240.

[36] T.J Kurtz, 1972, The relationship between stochastic and deterministic models for chemical reactions, *J. Chem. Phys.* 57(7), 2976-2978.

[37] S. Ilie, A. Teslya, 2012, An adaptive stepsize method for the chemical langevin equation, *J. Chem. Phys.*, 136, 184101.

[38] S. Ilie, 2012, Variable time-stepping in the pathways numerical solution of the chemical langevin equation, *J. Chem. Phys.,* 137, 234110.

[39] T. Li, 2007, Analysis of explicit tau-leaping schemes for simulating chemically reacting systems, *SIAM Multi. Model. Simul.* 6, 417-436.

[40] T. Marquez-Lago, K. Burrage, 2007, Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics, *J. Chem. Phys.* 127, 10410.

[41] H.H. McAdams, A. Arkin, 1997, Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 94, 814-819.

[42] C.J. Morton-Firth, D. Bray, 1998, Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* 192, 117-128.

[43] NA. Neogi, 2004, Dynamic partitioning of large discrete event biological systems for hybrid simulation and analysis. In, Hybrid Systems: Computation and Control, *HSCC, LNCS 2993, Heidelberg Springer*, 463-476.

[44] Jill M.A Padgett, Silvana Ilie, 2016, An adaptive tau-leaping method for stochastic simulations of reaction-diffusion systems, *AIP Advances*, 6(3), 035217.

[45] J. Pahle, 2009, Biochemical simulations: stochastic, approximate stochastic and hybrid approaches, *Brief Bioinform*, 10, 53–64.

[46] J. Puchalka, AM. Kierzek, 2004, Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks, *Biophys J*, 86(3), 1357-1372.

[47] M. Rathinam, L. Petzold, Y. Cao, D.T. Gillespie, 2005, Consistency and stability of tau-leaping schemes for chemical reaction systems, *SIAM Multi. Model. Simul.* 4,867-895.

[48] H. Salis, V. Sotiropoulus, Y. N. Kaznessis, 2006, Multiscale Hy3S: Hybrid stochastic simulation for supercomputers, *BMC Bioinform*, 7, 93.

[49] G. Söderlind, 2002, Automatic Control and Adaptive Time-Stepping, *Numer. Algorithms*, 31, 281–310.

[50] V. Sotiropoulos, M.N. Contou-Carrere, P. Daoutidis, Y.N. Kaznessis, 2009, Model reduction of multiscale chemical langevin equations: a numerical case study, *IEEE/ACM Trans. Comput. Biol.Bioinf.* 6, 470.

[51] V. Sotiropoulos, Y.N Kaznessis, 2008, An adaptive time step scheme for a system of stochastic differential equations with multiple multiplicative noise: Chemical Langevin equation, a proof of concept, *J. Chem. Phys.,* 128, 014103.

[52] T. Tian, K. Burrage, 2004, Bionomial leap methods for simulating stochastic chemical kinetics, *J. Chem. Phys.*, 121(1), 10356-10364.

[53] D. J. Wilkinson, 2006, Stochastic Modelling for systems biology, *Mathematical and Computational Biology series*, Chapman and Hall/CRC, ISBN 1-58488-540-8.