

NUMERICAL STUDIES OF HIGHER ORDER TAU-LEAPING METHODS

by

Anuj Dhoj Thapa

A thesis

presented to Ryerson University

in partial fulfillment of the
requirements for the degree of

Master of Science

in the program of

Applied Mathematics

Toronto, Ontario, Canada, 2018

©Anuj Dhoj Thapa, 2018

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.

NUMERICAL STUDIES OF HIGHER ORDER TAU-LEAPING METHODS

Master of Science, 2018

Anuj Dhoj Thapa

Applied Mathematics Program

Ryerson University

Abstract

Gillespie's algorithm, also known as the Stochastic Simulation Algorithm (SSA), is an exact simulation method for the Chemical Master Equation model of well-stirred biochemical systems. However, this method is computationally intensive when some fast reactions are present in the system. The tau-leap scheme developed by Gillespie can speed up the stochastic simulation of these biochemically reacting systems with negligible loss in accuracy. A number of tau-leaping methods were proposed, including the explicit tau-leaping and the implicit tau-leaping strategies. Nonetheless, these schemes have low order of accuracy. In this thesis, we investigate tau-leap strategies which achieve high accuracy at reduced computational cost. These strategies are tested on several biochemical systems of practical interest.

Acknowledgements

Initially, I would like to express my wholehearted thanks and courtesy to all the researchers whose journals are cited here and became treasures to enhance my knowledge to complete this thesis. Secondly, I would like to thank my supervisor Prof. Dr. Silvana Ilie; her commendable guidance, professional supervision and elaborative instructions made my research and thesis work come to completion as this document. Thirdly, I would like to thank my family, namely my father (Damber Dhoj Thapa), mother (Angur Thapa), sister (Mamata Thapa), brother (Milan Dhoj Thapa) and wife (Suraksha Khadka), for their constant support, motivation and inspiration through all the ups and downs during the last two years and further. I would like to thank the defense committee chair Dr. Jean-Paul Pascal and members, Dr. Dejan Delic and Dr. Dzung Minh Ha for taking the time to read my thesis and for all their suggestions. In addition to this, I would like to thank Steve Kanellis for his technical support and my course instructor Dr Katrin Rohlf for her guidance. I am always thankful to my friends, Amit Shrestha and Mathbar Singh Raut who keep me always motivated to explore new things.

Contents

Declaration	ii
Abstract	iii
Acknowledgements	iv
List of Figures	vii
1 Introduction	1
2 Background on Mathematical Models of Biochemical Systems and Simulation Tools	6
2.1 Deterministic and Stochastic Models	6
2.2 Description of Stochastic Models	7
2.3 Chemical Master Equation	10
2.4 Tau-Leaping Method:	13
2.5 Chemical Langevin Equation	14
2.6 Reaction Rate Equation	16
2.6.1 Derivation of Reaction Rate Equation:	16
2.7 Stochastic Simulation Algorithm	17
3 Stochastic Methods for Well-stirred Biochemical Systems	25
3.1 The Tau leaping Method for the Chemical Master Equation	25

3.2	Midpoint Tau-leaping Method for Chemical Master Equation	29
3.3	Trapezoidal Tau-Leaping Formula	30
4	Analysis of Tau-Leaping Method	32
4.1	Local Error Analysis	33
4.2	Local Error Formula for the Mean and Covariance of Explicit Tau-leaping Method	37
4.3	Local Error Formula for the Implicit Tau-leaping Method	40
5	Numerical Experiments	43
5.1	Simple Reaction Channel	43
5.2	Decay-dimerization Reaction Channel	50
5.3	Potassium Channel	55
6	Conclusion	67
	Bibliography	74

List of Figures

5.1	Simple reaction channel: Histograms of species S_1 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.05/4$	46
5.2	Simple reaction channel: Histograms of species S_2 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.05/4$	47
5.3	Simple reaction channel: A sample trajectory of number of molecules of species S_1 as function of time.	48
5.4	Simple reaction channel: A sample trajectory of number of molecules of species S_2 as function of time.	48
5.5	Simple reaction channel: Log log plot of the relative global error at time $T = 0.05/4$ as function of step size τ , for all the tau-leaping methods.	49
5.6	Decay dimerization reaction: Histograms of species S_1 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.02/4$	52
5.7	Decay dimerization reaction: Histograms of species S_2 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.02/4$	53
5.8	Decay dimerization reaction: A sample trajectory of number of molecules of species S_1 as function of time.	54

5.9	Decay dimerization reaction: A sample trajectory of number of molecules of species S_2 as function of time.	54
5.10	Decay dimerization reaction: Log log plot of the relative global error at time $T = 0.02/4$ as function of step size τ , for all the tau-leaping methods.	55
5.11	Potassium channel: Histograms of species S_1 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$	58
5.12	Potassium channel: Histograms of species S_2 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$	59
5.13	Potassium channel: Histograms of species S_3 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$	60
5.14	Potassium channel: Histograms of species S_4 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$	61
5.15	Potassium channel: Histograms of species S_5 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$	62
5.16	Potassium channel: A sample trajectory of number of molecules of species S_1 as function of time.	63
5.17	Potassium channel: A sample trajectory of number of molecules of species S_2 as function of time.	63
5.18	Potassium channel: A sample trajectory of number of molecules of species S_3 as function of time.	64
5.19	Potassium channel: A sample trajectory of number of molecules of species S_4 as function of time.	64

5.20 Potassium channel: A sample trajectory of number of molecules of species S_5 as function of time.	65
5.21 Potassium channel: Log log plot of the relative global error at time $T = 5/4$ as function of step size τ , for all the tau-leaping methods. . .	66

Chapter 1

Introduction

Mathematical modeling [29, 1, 2] has been successfully used in several biological disciplines for decades. Among the earliest quantitative biological models studied extensively in the literature is the Lotka-Volterra model of species competition and predator-prey relationships. In cell biology, many studies were dedicated to develop accurate models and advanced simulation tools for examining the structure and dynamics within a cell. Among the models considered were deterministic and stochastic ones. Computational modelling in cell biology [40] has been widely used to make significant progress leading to practical innovations in medicine, drug discovery and other related research. At the same time, the modern software and computational tools [28] contributed to the development and analysis of more realistic and complex models of cellular processes.

In earlier studies, the mathematical modeling approaches of biological systems were employed deterministic such as differential equations, which considered given initial conditions and neglected the noise and randomness within the system, and produces the similar results.

Among the deterministic models, an important class is that of systems of ordinary differential equations (ODE) [23]. For example, ODEs are utilized to model networks of well-stirred biochemical reactions, for which all molecular species have large population

numbers. This ODE model, known as the reaction rate equations (RRE), governs the behavior of the vector of concentrations of all species, as functions of time. Nonetheless, when some molecular species in the cell have low population numbers (such as in genetic networks), then the continuous deterministic model of the RRE fails to accurately describe the system dynamics. Indeed, important species in genetic networks come in low molecular counts (between 1 and a few tens), but these species often significantly influence the system behavior. For such systems, stochastic discrete models are required for capturing the random fluctuations which were observed in experiments [5, 6, 11, 13, 14, 32, 34]. One refined stochastic discrete model of well-stirred biochemical systems is the Chemical Master Equation (CME) [18, 29]. The Chemical Master Equation models the dynamics of the system state. The system state $X(t)$ is a vector with entries the number of molecules of each species at the current time, t , rather than the species concentrations. More precisely, the CME is a system of ordinary differential equations, in which each equation represents the evolution in the probability of the system to be in a particular state. Since often in application the number of all possible system states is very large, so is the dimensions of the CME.

Studying the CME directly is very challenging. A simplifying assumption is that of large population numbers in each species (hundred or more molecules/ species). In this case, the Chemical Master Equation model may be reduced to an approximate time-evolution equation of Langevin type, called the Chemical Langevin Equation (CLE) [20]. In this model, the stochastic process $X(t)$ has entries which are real values, rather than integers, as was the case for Chemical Master Equation. The Chemical Langevin Equation is a system of stochastic differential equations, with one equation for each reactant species in the biochemical system. This model is therefore an easier to simulate numerically and analyze.

Further, under the assumption of the thermodynamic limit, in which the system volume and the species populations tend to infinity, such that the species concentrations

are bounded, the deterministic part dominates the stochastic part of the CLE. Then we can neglect the stochastic terms of the Chemical Langevin Equation and we can reduce the CLE to get a system of ODEs, namely the Reaction Rate Equations (RRE).

Since the Chemical Master Equation is very challenging to solve directly, many studies were dedicated to the development of indirect methods to simulate the solution of this model. Stochastic simulation strategies for the Chemical Master Equations are classified as: exact and approximate methods. The Stochastic Simulation Algorithm (SSA), proposed by Daniel Gillespie in 1976 [16, 17] is an exact such method, in the sense that it generates trajectories of the stochastic process $X(t)$, having a probability in exact agreement with that modeled by the Chemical Master Equation. The SSA is thus a Monte Carlo simulation strategy, Nonetheless, when the biochemical system is subjected to some reactions which are very fast, the exact method is computationally intensive. The next reaction method published in 2000 by Gibson and Bruck [15] is another exact scheme for the CME, which has improved computational cost of simulation compared to SSA. In the next reaction method, the unused reaction times are reused and the algorithm employs a dependency graph for storing the reaction propensities. Only the propensities affected by a reaction fires in the next step are updated. While the next reaction method is more efficient than the SSA, it is still very slow on stiff models of cellular processes. An exact methods for the Chemical Master Equation were quite inefficient on stiff models of biochemical systems, there was a growing demand for faster simulation methods, allowing some minor sacrifice in accuracy in exchange for significant speed up. Examples of faster approximate simulation methods for CME include hybrid methods and tau-leaping schemes. Hybrid methods may combine exact or approximate methods for the CME, with numerical methods for the CLE or the RRE to more efficiently simulate the solution of a model of well-stirred biochemical system [1, 2, 25, 26, 35, 38].

The tau-leaping methods for the CME are based on the following idea: if a larger step-size τ may be chosen, to jump over many reactions of each type and the number

of each reactions may be accurate be approximated, then the system state can be updated accordingly. The explicit tau-leaping method was developed in 2001 by Gillespie [21]. This approximate strategy is applicable when the leap condition is valid, that is the leap size τ is chosen small enough such that no propensity is varied significantly over that step. When the condition is obeyed the number of reactions of the j^{th} type firing in $[t, t+\tau)$ is approximated by a Poisson distribution with mean and variance equal to $a_j(X(t)).\tau$, where a_j is the propensity function of the j^{th} reaction and $X(t)$ is the system state at the beginning of the interval. When many reactions occur over a step τ , the tau-leaping method is more efficient than the exact methods for the CME. But the explicit tau-leaping scheme is inefficient on stiff problems. For such problems Rathinam et al. [37] introduced the implicit tau-leaping scheme.

Both the explicit and implicit tau-leaping methods were shown to have low order of accuracy by Rathinam et al. [37]. More accurate numerical strategies for stochastic discrete biochemical systems are required, to assist researchers in the development and analysis of refined models of cellular processes. Two such higher order of accuracy method are analyzed in this thesis: the midpoint tau-leaping scheme [21] and the trapezoidal tau-leaping strategy [9]. The midpoint tau-leaping scheme is half degree higher order of accuracy than the explicit tau-leaping technique. The trapezoidal tau-leaping is an implicit method, as is the implicit tau-leaping strategy, but of higher order of accuracy. Also the trapezoidal rule is A-stable and does not have the damping of noise effect as the implicit tau-leaping.

Our studies of the two higher order accuracy tau-leaping methods are numerical in nature. We tested these methods on several models of biochemical networks of practical interest and compared the results with those obtained using the explicit and implicit tau-leaping schemes and the exact SSA.

Finally, the outline of thesis is as follows. Chapter 2 presents the backgrounds on stochastic modeling approaches for well-stirred biochemical systems, namely the CME and the CLE, along with the reduced deterministic model of the RRE. Stochastic

simulation methods for the CME and CLE models are given in chapter 3. Chapter 4 gives some theoretical results on the order of accuracy of several existing tau-leaping strategies. Numerical tests of the higher order tau-leaping methods considered in this thesis are included in chapter 5. The summary of our findings and the topics of future research are given in chapter 6.

Chapter 2

Background on Mathematical Models of Biochemical Systems and Simulation Tools

Mathematical models are required to describe the behavior of biochemical systems [12] and effective and accurate computational tools are needed to simulate their numerical solution, to study the biological systems and reveal the factors affecting the system like disease in organs and organisms. In addition to this, the mathematical models also help to predict some data pattern which leads to biological discovery such as treatment of diseases. Mathematical models of biochemical systems are classified into two categories [30, 35]: deterministic and stochastic.

2.1 Deterministic and Stochastic Models

The deterministic model of well-stirred biochemical systems consists of systems of ordinary differential equation (ODE) and with this model the outcomes are predictable and remain the same if the initial conditions are kept constant. Deterministic models are accurate for describing biochemical systems having very large molecular populations,

while stochastic models [39, 41, 43] are also accurate in predicting the dynamics of biochemical systems involving small number of molecules in some species, e.g. in a cell.

2.2 Description of Stochastic Models

In a well-stirred biochemical system, the molecules can be assumed distributed evenly [24, 34, 6, 8, 44]. In other words, a well-mixed system involves randomly distributed reactant species which are no more likely to be found in any sub-part of the system than in another of the same size. However, the natural motion of the molecules can change the concentration of a particular sub-volume of the system [42] and the system becomes heterogeneous and thus, to accomplish the requirements, one should stir the system externally.

Let us consider a well-stirred (well-mixed) biochemical system kept at thermal equilibrium in a constant volume. It consists of N molecular species $\{S_1, S_2, \dots, S_N\}$ undergoing M reaction channels $\{R_1, R_2, \dots, R_M\}$. At time t , the dynamical state

$$X(t) = \begin{pmatrix} X_1(t) \\ X_2(t) \\ \cdot \\ \cdot \\ \cdot \\ X_N(t) \end{pmatrix} \quad (2.2.1)$$

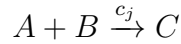
describes the biochemical system; here $X_i(t)$ is the non-negative number of S_i molecules

at time t . Moreover, the state change vector of the reaction R_j is

$$\nu_{ij} = \begin{pmatrix} \nu_{1j} \\ \nu_{2j} \\ \cdot \\ \cdot \\ \cdot \\ \nu_{Nj} \end{pmatrix},$$

where ν_{ij} = the change in the number of X_i molecules produced by one R_j reaction. The array $\{\nu_{ij}\}$ with $1 \leq i \leq N$, $1 \leq j \leq M$ is an $N \times M$ matrix, called the stoichiometric matrix, where N represents the total number of species in a model and M represents the total number of reactions.

To each reaction channel R_j it is associated a propensity function $a_j(x)$ such that $a_j(x)dt$ is defined as the probability that one R_j reaction will happen within the volume V in the next infinitesimal time interval $[t, t + dt)$, given that $X(t) = x$. For instance, if the following reaction happens,



with x_A and x_B being the number of molecules of A and B respectively, the propensity function is given by

$$a_j = c_j x_A x_B$$

where c_j is the reaction rate constant.

The propensity function depends on the molecular amounts of the reacting species, while the reaction rate depends on the concentration of such species. The types of reactions considered in the thesis are outlined below. Since reactions are assumed to be instantaneous events, no more than two molecules may interact at any moment in the system.

First order reaction:



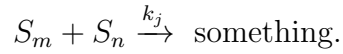
Here, if $x_m(t)$ denotes the number of S_m molecules at time t , then the propensity function is

$$a_j(x) = c_j x_m(t)$$

and the stochastic rate constant c_j depends on the deterministic reaction rate parameters k_j according to:

$$c_j = k_j.$$

Second order reaction:



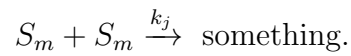
Here when $m \neq n$, if $x_m(t)$ and $x_n(t)$ denote the number of S_m and S_n molecules, respectively, and V is the volume of the system, then the propensity function is

$$a_j(x) = c_j x_m(t) x_n(t).$$

If c_j and k_j are the stochastic and deterministic reaction rate parameters, respectively, and V is the system volume, then

$$c_j = \frac{k_j}{V}.$$

Dimerization reaction:



Here, the propensity function may be written as

$$a_j(x) = \frac{c_j x_m(t)(x_m(t) - 1)}{2},$$

and the following relationship exists between the stochastic and deterministic rate constants:

$$c_j = \frac{2k_j}{V}.$$

2.3 Chemical Master Equation

If we consider a homogeneous biochemical system with fixed volume and temperature [46], at each point in time, the state of the system is $X(t)$ given by (2.2.1). Here $X(t)$ is a Markov process [18, 38, 11], we will show that the stochastic process $X(t)$ is modelled using a system of ordinary differential equations (ODEs), called the Chemical Master Equation (CME) [19, 23].

Let us consider $X(t_0) = x_0$ to be the initial conditions, which are assumed known. To describe the dynamics of the system, we determine

$$P(x, t | x_0, t_0)$$

the probability of the system to be in state x at time t , given that $X(t_0) = x_0$.

For simplicity, we denote this probability by $P(x, t)$. We wish to find $P(x, t + dt)$ for an infinitesimally small time dt and any system state x . In other words, here the system jumps from $X(t)$ to $X(t + dt) = x$. We have two options to be in the latter state: either the system was already in state x at time t and no reaction took place during $[t, t + dt)$ or, for some $1 \leq j \leq M$ the system was in state $(x - \nu_j)$ at time t and the j^{th} reaction happened during $[t, t + dt)$ resulting in the system being in state x at $t + dt$. In order to account for all such events, we have to use a result from the probability theory, called *the law of total probability*.

Let us assume A to be the event of interest and suppose that the events $H_0, H_1, H_2, \dots, H_M, H_{M+1}$ are such that:

- a) no more than one can happen.
- b) at least one event must occur.

Using the law of total probability, we obtain,

$$P(A) = \sum_{j=0}^{M+1} P(A|H_j)P(H_j). \quad (2.3.1)$$

where $P(A|H_j)$ is being the probability of getting A given that H_j happens. Here A is the event which describes the system in the state x at time $t + dt$. Let us assume that H_0 is the event that the system is in state x at time t , and H_j for $1 \leq j \leq M$ to describe the event that the system is in state $x - \nu_j$ at time t . Assume that H_{M+1} is the event that the system is in any other state at time t . Note that, for

$$1 \leq j \leq M,$$

$P(A|H_j)$ represents the probability of the reaction R_j happening during $[t, t + dt)$. Now, using the definition of propensity function, we get

$$P(A|H_j) = a_j(x - \nu_j)dt \quad (2.3.2)$$

for $1 \leq j \leq M$. We also remark that, $P(A|H_0)$ represents the probability that no reaction happens during $[t, t + dt)$. The probability of no reaction happening during $[t, t + dt)$ is 1 minus the probability of any event happening in $[t, t + dt)$, hence

$$P(A|H_0) = 1 - \sum_{j=1}^M a_j(x)dt \quad (2.3.3)$$

We note that,

$$P(A|H_{M+1}) = 0 \quad (2.3.4)$$

since H_{M+1} refers to all the states that are two or more reactions away from x and during $[t, t+dt)$ we assumed that at most one reaction can occur. Now, from equations (2.3.1)- (2.3.4), we get

$$P(x, t + dt) = (1 - \sum_{j=1}^M a_j(x)dt)P(x, t) + \sum_{j=1}^M a_j(x - \nu_j)dtP(x - \nu_j, t) \quad (2.3.5)$$

After simplifying the above equation, we obtain,

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

As $dt \rightarrow 0$, the above equation becomes an ordinary differential equation:

$$\frac{dP(x, t)}{dt} = \sum_{j=1}^M \left[a_j(x - \nu_j)P(x - \nu_j, t) - a_j(x)P(x, t) \right]. \quad (2.3.6)$$

Equation (2.3.6) is known as the Chemical Master Equation [19]. Recall that the expectation is defined as of a function $f(\cdot)$ of the stochastic process X evaluated at t .

$$\mathbb{E}(f(X(t))) \equiv \sum_x f(x)P(x, t) \quad (2.3.7)$$

Applying the expectation to equation (2.3.6) and using equation (2.3.7), we derive

$$\frac{d\mathbb{E}(X_i(t))}{dt} = \sum_{j=1}^M v_{ij}\mathbb{E}(a_j(X(t))),$$

for $j = 1, \dots, M$.

Suppose that all the reactions are at most of first order, then the propensity

functions will be *linear functions* of the state variables or constants and thus

$$\mathbb{E}\left(a_j(X(t))\right) = a_j\left(\mathbb{E}(X(t))\right).$$

If there are some bi-molecular reactions, we may obtain

$$\mathbb{E}\left(a_j(X(t))\right) \neq a_j\left(\mathbb{E}(X(t))\right).$$

If all the reactions are of at most first order, then we derive

$$\frac{d}{dt}\mathbb{E}(X(t)) = \sum_{j=1}^M \nu_j a_j(\mathbb{E}(X)) \quad (2.3.8)$$

If $Y(t) = \mathbb{E}(X(t))$, then equation (2.3.8) is the Reaction Rate Equation (RRE) in the average trajectory, $Y(t)$. The reaction rate equation is a set of N coupled first-order ODEs. This model was traditionally employed for studying chemical kinetics, where molecular populations of all species are in large molecules.

2.4 Tau-Leaping Method:

In 2001, Gillespie proposed the tau-leaping method [21]. This method may significantly reduce the computational cost of Gillespie algorithm without loss of accuracy. In this strategy the step size τ is previously chosen. (with applying Leap Condition). Moreover, it is an approximate method.

The tau-leap scheme may be applied when the following leap condition is obeyed:

$\tau > 0$ is small enough such that $a_j(X(s)) \simeq a_j(X(t))$ for any $1 \leq j \leq M$ and any $t \leq s \leq t + \tau$. The number of events that happens in $[t, t + \tau)$ is $P_j(\int_t^{t+\tau} a_j(X(s)ds)) \simeq P_j(a_j(X(t)).\tau)$, where P_j is a Poisson distribution. We assumed that the leap condition applies. Provided that $X(t) = x$, the number of reaction R_j that fire during the interval $[t, t + \tau]$ has a Poisson distribution with parameter $a_j(X(t))\tau$. Using the leap

condition, we can write the basic tau-leaping method as

$$X(t + \tau) = x + \sum_{j=1}^M P_j(a_j(x)\tau)\nu_j \quad (2.4.1)$$

if $X(t) = x$.

2.5 Chemical Langevin Equation

Under certain assumptions, the biochemical systems may be described using a simplified stochastic model [20, 25, 26]. These assumptions are:

- (1). The step size τ is small enough such that each propensity function $a_j(X(t))$ changes insignificantly over $[t, t + \tau)$.
- (2). τ is large enough such that the product $a_j(X(t)).\tau$ is much larger than 1 for any $1 \leq j \leq M$.

Derivation of the Chemical Langevin Equation (CLE): Since $a_j(X(t)).\tau \gg 1$, then we can approximate the Poisson random variable $P_j(a_j(X(t)).\tau)$ with a normal random variable $N_j(a_j(X(t))\tau, a_j(X(t))\tau)$ with the same mean and variance, namely $a_j(X(t))\tau$, in the tau-leaping method. Thus

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

and substituting in (2.4.1) we obtain

$$X(t + \tau) = X(t) + \tau \sum_{j=1}^M \nu_j(a_j(X(t))) + \sqrt{\tau} \sum_{j=1}^M \nu_j \sqrt{a_j(X(t))} Z_j \quad (2.5.1)$$

where Z_j are independent normal distributions with mean 0 and variance 1. Here, the entries of the state vector are real numbers, since the amount of molecular species

are computed using real-valued normal random variables rather than integer valued Poisson random variables.

Equation (2.5.1) represents the Euler-Maruyama method for a stochastic differential equation (SDE).

Taking, $\tau \rightarrow dt$ an infinitesimal step, in the limit, we obtain, the following stochastic differential equation,

$$dX(t) = \sum_{j=1}^M \nu_j(a_j(X(t)))dt + \sum_{j=1}^M \nu_j \sqrt{a_j(X(t))} dW_j(t) \quad (2.5.2)$$

where $W_j(t)$ are independent scalar Wiener processes for $1 \leq j \leq M$. The stochastic differential equation (2.5.2) is known as the Chemical Langevin Equation(CLE) [20]. This model is used when conditions (1) and (2) above simultaneously apply. These two conditions can be applied if all the molecular populations for all species are sufficiently large.

We remind the reader that a Weiner process is a stochastic process $W : [0, 1]$ such that

1. $W(0) = 0$ with probability one.
2. for any $0 \leq s \leq t$, $W(t) - W(s)$ is normally distributed with mean 0 and variance $(t - s)$.
3. for any $0 \leq u \leq v \leq s \leq t$, $W(t) - W(s)$ and $W(v) - W(u)$ are independent random variables.

Note that the Chemical Langevin Equation is an N-dimensional SDE and it may be derived by reducing the CME in the regime of large population numbers.

2.6 Reaction Rate Equation

The law of mass action governs the reaction rates in biochemistry, in other words it is the fundamental empirical law and it states that, for a reaction in a well-mixed and free medium, the reaction rate is proportional to the concentrations of the individual reactants involved. This law leads to the deterministic model of biochemical system known as the Reaction Rate Equations (RRE).

Thermodynamic limit: This is defined as the state at which the molecular amount of each species and the volume V of the system approach infinity such that $\frac{X_i}{V}$ is bounded for any $1 \leq i \leq N$. In this limit, the size of the stochastic term in equation (2.5.2) is much smaller compared to the size of the deterministic term, thus it may be ignored. Therefore, the CLE (2.5.2) may be simplified to the reaction rate equation:

$$\frac{dX(t)}{dt} = \sum_{j=1}^M \nu_j(a_j(X(t))). \quad (2.6.1)$$

Equation (2.6.1) is valid when the molecular amounts of all species are very large.

2.6.1 Derivation of Reaction Rate Equation:

Indeed, if the molecular population is very large, the stochastic part of Chemical Langevin Equation (CLE) can be neglected and the new equation is an ordinary differential equation, called the reaction rate equation [27]. In order to get the reaction

rate equation, let us take the expectation in equation (2.5.2), and we get

$$\begin{aligned}
\mathbb{E}(dX(t)) &= \mathbb{E}\left(\sum_{j=1}^M \nu_j(a_j(X(t)))dt + \mathbb{E}\left[\sum_{j=1}^M \nu_j \sqrt{a_j(X(t))}dW_j(t)\right]\right) \\
\mathbb{E}(dX(t)) &= \sum_{j=1}^M \mathbb{E}(\nu_j(a_j(X(t))))dt + \sum_{j=1}^M \nu_j \mathbb{E}\left[\sqrt{a_j(X(t))}dW_j(t)\right] \\
\mathbb{E}(dX(t)) &= \sum_{j=1}^M \nu_j \mathbb{E}((a_j(X(t))))dt
\end{aligned}$$

Dividing both sides by dt , we get,

$$\frac{d}{dt}\mathbb{E}(X) = \sum_{j=1}^M \nu_j \mathbb{E}((a_j(X(t)))).$$

For a first order reaction with a propensity function of the form $a_j(x) = c_j X_m$ or $a_j(x) = c_j$ we have, then $\mathbb{E}(a_j(x)) = a_j(\mathbb{E}(x))$. However, for a second order reaction with a propensity function either $a_j(x) = c_j x_m x_n$ for $m \neq n$ or $a_j(x) = c_j \frac{x_m(x_m-1)}{2}$ for $m = n$, in general $\mathbb{E}(a_j(x)) \neq a_j(\mathbb{E}(x))$. Consequently, for biochemical systems with reactions of order zero or one, the behavior of the average trajectory of the CME satisfies the RRE:

$$\frac{dX(t)}{dt} = \sum_{j=1}^M \nu_j(a_j(X(t))) \tag{2.6.2}$$

In conclusion, equation (2.6.2) representing the RRE, can be derived from Chemical Langevin Equation or from the Chemical Master Equation when all the reactions are first or zero order.

2.7 Stochastic Simulation Algorithm

The Chemical Master Equation is a system of ordinary differential equations of very large dimension. It has one equation in $P(x, t)$ for each possible state of the system, x . Due to its very large dimension, it can not be solved directly, in general, by analytical or numerical methods [27]. However, there exist Monte Carlo simulation methods for

generating trajectories in exact agreement with the solution of the Chemical Master Equation. The Stochastic Simulation Algorithm (SSA) is one such method and it was proposed by Gillespie[16, 17].

The SSA is based on the idea of generating trajectories of $X(t)$ obeying a certain probability $P(\tau, j|x, t)$. Here, $P(\tau, j|x, t)$, is the probability that, provided that $X(t) = x$, the next reaction will be fired in the infinitesimal time interval $[t + \tau, t + \tau + d\tau)$ where $d\tau$ is an infinitesimal time and this reaction will be R_j . We so obtain later that, in fact, $P(\tau, j|x, t)$ is the joint probability density function of two random variables which are as follows: τ = time to the new reaction and j = index of the new reaction.

Assume that $P_0(\tau|x, t)$ is the probability of no reaction firing in the time interval $[t, t + \tau)$, given that $X(t) = x$. From the laws of probability, we can write,

$P(\tau, j|x, t)$ = probability of no reaction in $[t, t + \tau) \times$ probability of reaction R_j in $[t + \tau, t + \tau + d\tau)$. Thus

$$P(\tau, j|x, t)d\tau = P_0(\tau|x, t)a_j(x)d\tau. \quad (2.7.1)$$

Now we compute the conditional probability $P_0(\tau|x, t)$. The probability $P_0(\tau|x, t)$ equals the probability of no reaction during $[t, t + \tau + d\tau)$, or, equivalently, the probability of no reaction during $[t, t + \tau)$ and the probability of no reaction over $[t + \tau, t + \tau + d\tau)$. However, the probability of no reaction over the infinitesimal time interval $[t + \tau, t + \tau + d\tau)$ is 1 minus sum of probabilities of each reaction R_j occurring in $[t + \tau, t + \tau + d\tau)$, $1 \leq j \leq M$, which is

$$1 - \sum_{j=1}^M a_j(x)d\tau$$

To summarize,

$$P_0(\tau + d\tau|x, t) = P_0(\tau|x, t) \left(1 - \sum_{j'=1}^M a_{j'}(x)d\tau \right)$$

After doing some algebraic simplifications on the latter equation with the limit $d\tau \rightarrow 0$, we obtain an ordinary differential equations of the form,

$$\frac{dP(\tau|x, t)}{d\tau} = - \sum_{j'=1}^M a'_{j'}(x)$$

with the initial condition $P(0|x, t) = 1$. Solving this equation leads to :

$$P_0(\tau|x, t) = e^{(-a_0(x)\tau)},$$

where

$$a_0(x) = \sum_{j'=1}^M a'_{j'}(x).$$

Substituting this into equation (2.7.1) ,we get

$$P(\tau, j|x, t) = a_j(x)e^{(-a_0(x)\tau)}$$

We can re-write this equation as

$$P(\tau, j|x, t) = \frac{a_j(x)}{a_0(x)} \cdot \left[a_0(x)e^{(-a_0(x)\tau)} \right] \quad (2.7.2)$$

Equation (2.7.2) is the mathematical foundation for the SSA procedure [12]. Indeed, (2.7.2) says that the joint density function of time to the next reaction, τ , and the index of the next reaction, j can be expressed as the product of the τ density function, which is $a_0(x)\exp(-a_0(x)\tau)$, and the j -density function which is $a_j(x)/a_0(x)$. Consequently, we can generate random samples from the above density functions by using the inversion method of the Monte-Carlo theory [15]. If r_1 and r_2 are random numbers from the uniform distribution in the unit time interval, then τ may be computed as

$$\tau = \frac{1}{a_0(x)} \ln \frac{1}{r_1} \quad (2.7.3)$$

while j may be calculated as the smallest integer $1 \leq j \leq M$ such that

$$\sum_{j'=1}^{j-1} a_{j'}(x) \leq r_2 a_0(x) < \sum_{j'=1}^j a_{j'}(x) \quad (2.7.4)$$

To justify the computation of the index j to the next reaction in the SSA, we give below some theoretical results. These results explain the choice in equation (2.7.4).

Proposition 2.7.1. [45] *If $X_i = \text{Exp}(a_i)$, $i = 1, 2, \dots, M$ are independent random variables distributed exponentially with parameters a_i , then the minimum of them is exponentially distributed with parameter a_0 as follows,*

$$X_0 \equiv \min_i \{X_i\} \sim \text{Exp}(a_0), \text{ where } a_0 = \sum_{i=1}^M a_i.$$

Proof. Note that for an exponential distribution X with parameters a , $X \sim \text{Exp}(a)$ we have $P(X > \tau) = e^{-(a\tau)}$ for τ real scalar.

Now, consider

$$\begin{aligned} P(X_0 > \tau) &= P\left(\min_{i=1, \dots, M} \{X_i\} > \tau\right) \\ &= P\left(\left[X_1 > \tau\right] \cap \left[X_2 > \tau\right] \cap \dots \cap \left[X_M > \tau\right]\right) \\ &= \prod_{i=1}^M P\left(X_i > \tau\right) \\ &= \prod_{i=1}^M e^{-(a_i \tau)} \\ &= e^{-\sum_{i=1}^M a_i \tau} \\ &= e^{-(a_0 \tau)}. \end{aligned}$$

As a consequence, we derived that,

$$P(X_0 \leq \tau) = 1 - e^{-(a_0\tau)}$$

and hence $X_0 \sim \text{Exp}(a_0)$, is an exponential distribution with parameter $a_0 = \sum_{i=1}^M a_i$. □

Lemma 2.7.1. [45] *Suppose that $X \sim \text{Exp}(a)$ and $Y \sim \text{Exp}(b)$ are independent random variables exponentially distributed with parameters a and b . Then the following is true,*

$$P(X < Y) = \frac{a}{a + b}.$$

Proof. We can write the following derivation

$$\begin{aligned} P(X < Y) &= \int_0^\infty P(X < Y | Y = y) f(y) dy \\ &= \int_0^\infty P(X < y) f(y) dy \\ &= \int_0^\infty (1 - e^{-ay}) b e^{-by} dy \end{aligned}$$

Therefore,

$$P(X < Y) = \frac{a}{a + b}.$$

□

Proposition 2.7.2. [45] *Assume $X_j \sim \text{Exp}(a_j)$, $j = 1, 2, \dots, M$ are independent random variables exponentially distributed each with parameter a_i , and let i be the index of the smallest of the X_j . Then i is a discrete random variable with probability*

mass function

$$\pi_j = \frac{a_j}{a_0}, \text{ where } a_0 = \sum_{j=1}^M a_j.$$

Proof. Note that,

$$\begin{aligned} \pi_i &= P\left(X_i < \min_{j \neq i} \{X_j\}\right) \\ &= P(X_i < Y) \end{aligned}$$

Here we denoted $Y = \min_{j \neq i} \{X_j\}$,

$$\begin{aligned} \text{so that } Y &\sim \text{Exp}(a_{-i}) \text{ and } a_{-i} = \sum_{j \neq i} a_j \\ \pi_i &= \frac{a_i}{a_i + a_{-i}} \end{aligned}$$

Then finally using lemma (2.7.1),

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

□

This leads to the SSA [17] as follows:

- 1 Take $t = t_0$ and $x = x_0$ as initial time and state of the system respectively.
- 2 Calculate $a_j(x)$ and $a_0(x)$ at time t and state x .
- 3 Find values for τ and j using equations(2.7.3) and (2.7.4)
- 4 Update the system by replacing $t \leftarrow t + \tau$ and $x \leftarrow x + \nu_j$.
- 5 Save (x, t) as desired and repeat it by going to step 2 or, otherwise, end the simulation.

We note that the SSA is slow for some biochemical systems involving some fast reactions but easy to implement. The reason for its inefficiency on many systems

arising in applications is due to the factor $1/a_0(x)$ in equation (2.7.3). This factor becomes very small if the population of the reactant species is large. Another cause of the high computational cost of the SSA is due to searching for the index j of the first reaction and updating all the propensities functions after each time step, in case the biochemical system has many reactions (M large).

Gibson and Bruck [15], in 2000, proposed the Next Reaction Method which is an improvement in terms of efficiency over the Gillespie algorithm [17]. The Next Reaction Method reduces the computational cost of computing exact trajectories for the CME by updating only the parts of the system changed by a reaction and by storing them in a special structure. At each time-step, Gillespie's First Reaction Method [16, 17] consists of generating all the possible time at which each reaction R_k fires, namely τ_k for each R_k , and then choosing j such that $\tau_j = \min_{k=1,\dots,M} \tau_k$. The system is thus updated as $x \leftarrow x + \lambda_j$ and $t \leftarrow \tau_j$. The three time consuming steps (which occur during every iteration of Gillespie's First Reaction Method) are proportional to the number of reactions R_j :

1. updating of all reactions associated with propensity functions a_j
2. generating a possible time, τ_k for the next reaction R_k and
3. identifying the smallest possible time, $\tau_j = \min_{k=1,\dots,M} \tau_k$

The Next Reaction Method [15] discards each of these steps and uses new ideas which are as follows:

1. Store not only a_l but also τ_l ,
2. Recalculate a_l with τ_l only if they are changed by the reaction that occurred (use a dependency graph, which allows to update the minimum number of propensities a_l)
3. Re-use τ_l wherever appropriate.

Further, the algorithm of the Next Reaction Method proceeds as follows:

- 1 Initialize the initial number of molecules $X = X_0$ at time $t = 0$.
- 2 Find the propensity function, $a_j(X)$ for each reaction R_j .
- 3 Generate M independent, uniform(0,1) random numbers r_j .
- 4 Find τ_j , $\tau_j = \frac{1}{a_j} \ln \frac{1}{r_j}$, for each reaction R_j , $j = 1, \dots, M$.
- 5 Compute $t = \min_j \{\tau_j\}$ and suppose τ_k be the time where the minimum is realized.
- 6 The molecular amount of each species is updated according to reaction R_k .
- 7 Update all the propensity functions affected by the occurrence of reaction R_k , let them be a'_j .
- 8 Set all $\tau_j = \frac{a_j}{a'_j} (\tau_j - t) + t$ for each $j \neq k$.
- 9 If r is uniform random number in (0,1), then set $\tau_k = \frac{1}{a'_k} \ln \frac{1}{r} + t$.
- 10 Set $a_j = a'_j$ for each j .
- 11 Go to step 5 or else end the simulation.

Chapter 3

Stochastic Methods for Well-stirred Biochemical Systems

3.1 The Tau leaping Method for the Chemical Master Equation

In 2001, Gillespie proposed the tau-leap method [21] to achieve significant speed-up of the computation without significant loss in accuracy. Gillespie's algorithm is slow when the system is subject to very fast reactions, as it simulates all reactions in the system, one at a time. By contrast, the tau-leaping method [3, 10] collects details of how many reactions of each type occur in a certain time interval. If the time interval is large enough for many reactions to happen, that leads to noticeable gain in the computational time. On the other hand, in order to maintain accuracy, an appropriate length of the time interval, τ , has to be chosen, which should be small enough so that the change in each propensity function is negligible. This is called the leap condition.

The leap condition: It requires that the size of τ is chosen such that the relative changes induced in the propensity functions during τ are small. In other words, for

the current state x , it requires τ to be small enough that the change in the state during $[t, t + \tau)$ will be so small that no propensity function will suffer an appreciable change in its value. Tau-leaping proceeds with a pre-selected time τ , during which many reactions fire. If τ is chosen sufficiently small to satisfy the leap condition and $X(t) = x$ (which makes propensity function almost constant), then each reaction R_j happens nearly $P_j(a_j(X(t))\tau)$ times in $[t, t + \tau)$, where $P_j(a_j(X(t))\tau)$ is a Poisson distribution of mean and variance $a_j(X(t))\tau$. The exact formula is

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

but due to the leap condition,

$$P_j \left(\int_t^{t+\tau} a_j(X(s)) ds \right) \simeq P_j(a_j(X(t)), \tau).$$

Therefore, we can write the following approximate formula

$$X(t + \tau) = x + \sum_{j=1}^M v_j P_j(a_j(X(t)), \tau). \quad (3.1.1)$$

For this algorithm M Poisson random numbers need to be generated for every leap. Formula (3.1.1) is called the *tau-leaping method*. It is more efficient than SSA when many reactions are fired during a leap. If $\tau \rightarrow 0$, the tau-leaping schemes becomes the SSA. Then it becomes very inefficient as most random numbers in equation (3.1.1) will be zero.

The tau-leaping method is an approximate Monte Carlo strategy for simulating the solution of the CME. The scheme (3.1.1) is known also as the *explicit tau-leaping method* and it is appropriate for simulating stochastic models of non-stiff biochemical systems. Stiffness arises in biochemical systems when both fast and slow reactions are present, the fastest dynamics being stable. Stiffness presents challenges for solving

deterministic models of biochemical systems and these challenges are more significant when stochastic models are solved. In the deterministic setting, implicit strategies are more efficient for solving numerically stiff problems than the explicit technique. This is also true when dealing with stiff stochastic models of biochemical systems. For these, the so-called *implicit tau-leaping method* was proposed Rathinam et al.[36]. In fact, in the implicit tau-leaping scheme only the deterministic term is implicit, the stochastic term of mean zero is in an explicit form. The implicit tau-leaping scheme outperforms in terms of efficiency the explicit tau-leaping strategy and the SSA for stiff biochemical systems.

Explicit tau-leaping method : As discussed before, if τ satisfies the leap condition then the explicit tau-leap method is given by,

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

where $X(t) = x$. Here $P_j(a_j(X(t)), \tau)$ are independent Poisson random variables with means and variances $(a_j(X(t))\tau)$.

For stability, the explicit tau-leaping method requires a small step size when fast reactions are present.

Algorithm of Explicit tau-leaping method :

- 1 Select τ satisfying the leap condition and an initial state of $x \leftarrow x_0$ at time $t \leftarrow t_0$.
- 2 Generate the number of firings k_j of channel R_j in $[t, t + \tau)$ as $k_j = P(a_j(x)\tau)$ with $(j = 1, \dots, M)$.
- 3 Update $t \leftarrow t + \tau$ and $x \leftarrow x + \sum_{j=1}^M k_j \nu_j$.

4 Record (x, t) . Then return to step (1) or else end the simulation.

Here, if the generated value of k_j are noticeably large, the explicit tau-leaping method will be faster than the SSA.

Implicit tau-leaping method : Recall that when solving numerically an ordinary differential equation (ODE) of the form

$$\frac{dX}{dt} = f(X, t) \quad (3.1.2)$$

We can apply explicit Euler method [22]

$$X(t + \Delta t) = X(t) + f(X(t), t)\Delta t.$$

When the ODE (3.1.2) is stiff, the implicit Euler method is preferred, as it is unconditionally stable. This method may be written as

$$X(t + \Delta t) = X(t) + f(X(t + \Delta t), t + \Delta t)\Delta t.$$

To compute $X(t + \Delta t)$ an implicit equation needs to be solved. To solve the implicit equation in $X(t + \Delta t)$, we can use the Newton iteration method or something similar may be applied to the tau-leaping method. The implicit tau-leaping scheme can be written as

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

where $X(t) = x$.

Here, the mean of the Poisson random variable $P_j(a_j(X(t)), \tau)$ is subtracted and changed to be the value at time $(t + \tau)$. The explicit tau-leaping method may be

written in the form:

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

where $\sum_{j=1}^M \nu_j a_j X(t) \tau$ is the deterministic part and $\sum_{j=1}^M \nu_j [P_j(a_j(X(t)), \tau) - a_j(X(t)) \tau]$ is the stochastic part of mean zero.

Now, we can make the method implicit in the deterministic part as below:

$$X(t + \tau) = X(t) + \sum_{j=1}^M \nu_j [P_j(a_j X(t), \tau) - a_j(X(t)) \tau] + \sum_{j=1}^M \nu_j a_j(X(t + \tau)) \tau. \quad (3.1.3)$$

Equation (3.1.3) can be solved numerically to estimate $X(t + \tau)$ by using Newton's method with the initial guess $X(t)$. The implicit tau-leaping scheme is noticeably faster than explicit tau-leaping method on stiff systems as it allows larger step sizes. The former method is unconditionally stable where the latter is conditionally stable.

3.2 Midpoint Tau-leaping Method for Chemical Master Equation

In 2000, Gillespie introduced the mid-point tau-leaping method [21] to get higher accuracy of the numerical simulation.

When solving numerically an ordinary differential equation of the form

$$\frac{dX}{dt} = f(X(t))$$

by the explicit or implicit Euler's method, then the accuracy of the numerical solution is $O(\Delta t)$. In order to use this technique along with the tau-leaping method, we first create the increment,

$$\lambda' = \sum_{j=1}^M (a_j(x) \tau) \nu_j,$$

on the interval $[t, t + \tau)$, where $X(t) = x$. The midpoint is

$$x' = x + \left[\frac{\lambda'}{2}\right]$$

. Using this predicted midpoint, the state at the end of the interval is

$$X(t + \tau) = x' + \sum_{j=1}^M P_j(a_j(x'), \tau) \nu_j$$

This formula is known as the midpoint tau-leaping method [21].

Midpoint Tau-leaping Algorithm :

- 1 Initiate the system state $x \leftarrow x_0$ at $t \leftarrow t_0$ and set rate parameters.
- 2 Select the τ which satisfies the leap condition.
- 3 Compute the expected state change

$$\lambda' = \tau \sum_j a_j(x) \nu_j$$

during $[t, t + \tau]$.

- 4 Using $x' = x + \left[\frac{\lambda'}{2}\right]$, generate for each $j = 1 \dots M$ a sample value k_j of the Poisson random variable $P_j(a_j(x')\tau)$.
- 5 Update $t \leftarrow t + \tau$ and $x \leftarrow x + \sum_{j=1}^M k_j \nu_j$.
- 6 Record (x, t) . If $t < T$ return to step [2] or else end the simulation.

3.3 Trapezoidal Tau-Leaping Formula

In 1998, Ascher and Petzold schemed the trapezoidal rule to be A-stable [7]. In this section we discuss the extension of the trapezoidal method for ODEs to the tau-leaping

scheme by Cao and Petzold [9]. Though it is of first order for the variance, it is of second order for the mean. Unlike the implicit tau-leaping scheme it is free from the damping effect.

The trapezoidal rule has a higher convergence order than the explicit and implicit Euler method for ODEs. Let us consider an ordinary differential equation of the form

$$\frac{dX}{dt} = f(X, t).$$

Then, the trapezoidal method to solve the above equation is given by,

$$X(t + \Delta t) = X(t) + \frac{1}{2}\Delta t(f(X(t)) + f(X(t + \Delta t)))$$

The trapezoidal tau-leaping formula [9] is

$$X^{(tr)}(t + \tau) = x + \sum_{j=1}^M \nu_j [P(a_j(x), \tau) - \frac{\tau}{2}a_j(x) + \frac{\tau}{2}a_j(X^{tr}(t + \tau))], \quad (3.3.1)$$

where $X(t) = x$. We note that (3.3.1) is an implicit strategy in $X(t + \tau)$. To solve it, we may apply the iterative Newton's method with the initial guess taken to be $X(t)$.

Trapezoidal Tau-leaping Algorithm:

- 1 Initialize $t \leftarrow t_0$ and the system state $x \leftarrow x_0$.
- 2 Calculate the propensity function, $a_j(x)$ for $j = 1, 2, 3, \dots, M$.
- 3 Generate M independent random numbers, k_j from $P(a_j(x)\tau)$ with $j = 1, 2, 3, \dots, M$.
- 4 Solve (3.3.1) for $X^{(tr)}$ by Newton's Method.
- 5 Update the system state and time by step (4) and $t + \tau$, respectively.
- 6 If $t < T$ return to step (2) or else end the simulation.

Chapter 4

Analysis of Tau-Leaping Method

Tau-leaping method uses a piece-wise Poisson approximation with propensity function assumption $a_j(x) + o(\tau)$ as $\tau \rightarrow 0$ and for such a small value of τ , the behavior of both explicit and implicit-tau leaping is obvious as suggested by Stochastic Simulation Algorithm. Here, a quantitative estimate of error analysis as a function of stepsize τ is done to develop idea of automatic stepsize selection and to gather information for future study of higher-order accurate tau-leaping methods.

Let us suppose $X^{(e)} = x$ is the state at time t and for time step $t + \tau$ where $\tau > 0$ is the state can be written as

$$X^{(e)}(t + \tau) = x + \sum_{j=1}^M \nu_j K_j^{(e)}(x, \tau)$$

Here, superscript "e" stands for explicit tau-leaping method. Where $K_j^{(e)}(x, \tau) = P_j(a_j(x), \tau)$ for $j = 1, \dots, M$ are independent Poisson random variables with mean and variance $a_j(x)\tau$.

Suppose that $X^{(i)}(t) = x$ is the state at time t , the state at time $t + \tau$ using the

implicit tau (unrounded version) method is given by

$$X^{(i)}(t + \tau) = x + \sum_{j=1}^M \nu_j a_j(X^{(i)}(t + \tau))\tau + \sum_{j=1}^M \nu_j (P_j(a_j(x)\tau) - a_j(x)\tau)$$

where superscript "i" is for implicit-tau leaping method.

Now we are going to see the local error using the local expansion of Taylor series in τ . Below we follow the analysis by Rathinam et al.[37].

4.1 Local Error Analysis

Here we derive recursive integral forms for the solution of Chemical Master Equations with local Taylor expansions of those integrals. This is the base to find the errors generated by the tau-leaping methods [4]. Let us suppose the multi-index $k = (k_1, \dots, k_l)$, where $k_j \in \{1, \dots, M\}$, represents a sequence of reaction events R_{k_j} firing in this order. Again, suppose $|k| = l$ is the number of reactions and $p(k; x, \tau)$ represents the probability that the sequence of reactions that happened in the interval $(t, t + \tau]$ is exactly k given that $X(t) = x$. Then

$$p(\phi; x, \tau) = e^{-a_0(x)\tau}$$

where ϕ means no reactions occurred. Let us consider the increment in the moments be $X(t + \tau) - X(t)$ given that $X(t) = x$. The conditional expectation may be written as

$$\mathbb{E}(X(t + \tau) - X(t) | X(t) = x) = \sum_{l=1}^{\infty} \sum_{k, |k|=l} p(k; x, \tau) \left(\sum_{\alpha=1}^l \nu_{k_\alpha} \right) \quad (4.1.1)$$

while the order r conditional moment is

$$\mathbb{E}((X(t + \tau) - X(t))^r | X(t) = x) = \sum_{l=1}^{\infty} \sum_{k, |k|=l} p(k; x, \tau) \left(\sum_{\alpha=1}^l \nu_{k_\alpha} \right)^r \quad (4.1.2)$$

For any state x and multi-index k , $p(k; x, \tau)$ is an analytic function of τ . The components of $\mathbb{E}((X(t + \tau) - X(t))^r | X(t) = x)$ all analytic function of τ on a neighbourhood of O . In order to compute the above moments up to terms including of order 2 in τ , by assuming the moments exist, we have to add terms with $l = 1$ and $l = 2$ in equations (4.1.1) and (4.1.2) and then derive

$$\begin{aligned} \mathbb{E}(X(t + \tau) - X(t) | X(t) = x) &= \sum_{j=1}^M \nu_j p((j); x, \tau) \\ &+ \sum_{j_1=1}^M \sum_{j_2=1}^M (\nu_{j_1} + \nu_{j_2}) p((j_1, j_2); x, \tau) + O(\tau^3) \end{aligned}$$

It can be shown that

$$p((k_1, k_2, \dots, k_l); x, \tau) = \int_0^\tau p((k_1, k_2, \dots, k_{l-1}); x, s) \times a_{k_l}(x + \nu_{j_1} + \dots + \nu_{j_{l-1}}) e^{-a_0(x + \nu_{j_1} + \dots + \nu_{j_l})(\tau - s)} ds$$

By induction, we can show that $p(k; x, \tau) = O(\tau^{|k|})$ in the limit as $\tau \rightarrow 0$. We have,

$$p((j); x, \tau) = \int_0^\tau e^{-a_0(x)s} a_j(x) e^{-a_0(x + \nu_j)(\tau - s)} ds \quad (4.1.3)$$

and also

$$p((j_1, j_2); x, \tau) = \int_0^\tau p((j_1); x, s) \times a_{j_2}(x + \nu_{j_1}) e^{-a_0(x + \nu_{j_1} + \nu_{j_2})(\tau - s)} ds \quad (4.1.4)$$

for any $j, j_1, j_2 \in \{1, 2, \dots, M\}$. Then using Taylor series expansion for the equations (4.1.3) and (4.1.4), we get

$$p((j); x, \tau) = a_j \tau - \frac{1}{2} \tau^2 \sum_{j_1=1}^M a_j(x) \{a_{j_1}(x + \nu_j) + a_{j_1}(x)\} + O(\tau^3)$$

and

$$p((j_1, j_2); x, \tau) = \frac{1}{2} \tau^2 a_{j_1}(x) a_{j_2}(x + \nu_{j_1}) + O(\tau^3).$$

Also, we obtain for the r^{th} order conditional moment

$$\begin{aligned}\mathbb{E}((X(t+\tau) - X(t))^r | X(t) = x) &= \sum_{j=1}^M \nu_j^r p((j); x, \tau) \\ &+ \sum_{j_1=1}^M \sum_{j_2=1}^M (\nu_{j_1} + \nu_{j_2})^r p((j_1, j_2); x, \tau) + O(\tau^3)\end{aligned}$$

Using the Taylor expansion and simplifying for the mean, we get

$$\begin{aligned}\mathbb{E}(X(t+\tau) - X(t) | X(t) = x) &= \tau \sum_{j=1}^M \nu_j a_j(x) \\ &+ \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} a_{j_2}(x) \{a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)\} \\ &+ O(\tau^3)\end{aligned}\tag{4.1.5}$$

Similarly, for the order r moment, we derive

$$\begin{aligned}\mathbb{E}((X(t+\tau) - X(t))^r | X(t) = x) &= \tau \sum_{j=1}^M \nu_j^r a_j(x) \\ &- \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^r a_{j_1}(x) a_{j_2}(x) \\ &- \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^r a_{j_1}(x) a_{j_2}(x + \nu_{j_1}) \\ &+ \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M (\nu_{j_1} + \nu_{j_2})^r a_{j_1}(x) a_{j_2}(x + \nu_{j_1}) \\ &+ O(\tau^3)\end{aligned}$$

Using the expression, for $p((j); x, \tau)$ and $p((j_1, j_2); x, \tau)$ we can compute the conditional

co-variance $\text{Cov}(X(t + \tau)|X(t) = x)$ as

$$\begin{aligned}
\text{Cov}(X(t + \tau)|X(t) = x) &= \text{Cov}(X(t + \tau) - X(t)|X(t) = x) \\
&= \mathbb{E}((X(t + \tau) - X(t))^2|X(t) = x) - (\mathbb{E}(X(t + \tau) - X(t)|X(t) = x))^2 \\
&= \tau \sum_{j=1}^M \nu_j^2 a_j(x) \\
&+ \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^2 a_{j_2}(x) \{a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)\} \\
&+ \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \nu_{j_2} a_{j_1}(x) \{a_{j_2}(x + \nu_{j_1}) - a_{j_2}(x)\} \\
&+ \frac{1}{2} \tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \nu_{j_2} a_{j_2}(x) \{a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)\} \\
&+ O(\tau^3)
\end{aligned} \tag{4.1.6}$$

We will find in what follows that the explicit tau-leaping method is weakly consistent to first order in τ .

Theorem 4.1.1. *[37] Order-1 weak consistency of the explicit tau-leaping method*

Let x be any possible system state and $r \geq 1$ be any integer, then there exists a constant $M_r > 0$ and $\delta_r > 0$ such that

$$\begin{aligned}
&\| \mathbb{E}(X^{(e)}(t + \tau)) - \mathbb{E}(X^{(e)}(t))^r | X^{(e)}(t) = x) - \mathbb{E}((X(t + \tau) - X(t))^r | X(t) = x) \| \\
&\qquad \qquad \qquad < M_r \tau^2
\end{aligned}$$

for any $0 \leq \tau \leq \delta_r$.

Proof. Note that, given $X^{(e)}(t) = x$, we have for the explicit tau-leap solution

$$\begin{aligned}
\mathbb{E}\left((X^{(e)}(t+\tau) - X^{(e)}(t))^r | X^{(e)}(t) = x\right) &= \\
&= \mathbb{E}\left(\sum_{j=1}^M \nu_j P_j(a_j(x)\tau)\right)^r \\
&= \tau \sum_{j=1}^M \nu_j^r a_j(x) + O(\tau^2). \quad (4.1.7)
\end{aligned}$$

but for the exact solution we have

$$\begin{aligned}
\mathbb{E}\left((X(t+\tau) - X(t))^r | X(t) = x\right) &= \\
&= \tau \sum_{j=1}^M \nu_j^r a_j(x) + O(\tau^2). \quad (4.1.8)
\end{aligned}$$

The theorem is proved by combining (4.1.7) and (4.1.8). \square

Since, we know that the explicit method is weakly consistent to first order, now we need to see that the mean and covariance of local error is $O(\tau^2)$.

4.2 Local Error Formula for the Mean and Covariance of Explicit Tau-leaping Method

If we use the equation (4.1.5) and the fact that

$$\mathbb{E}(X^{(e)}(t+\tau) - X^{(et)}(t) | X^{(e)}(t) = x) = \tau \sum_{j=1}^M \nu_j a_j(x)$$

then, we get the local error in the mean for the explicit tau-leaping method [37] as

$$\begin{aligned}\mathbb{E}(X^{(e)}(t + \tau) - X(t + \tau) | X^{(e)}(t) = X(t) = x) &= \\ &- \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} a_{j_2}(x) \{a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)\} \\ &+ O(\tau^3)\end{aligned}$$

Since the Poisson random numbers $P_j(a_j(x), \tau)$ are independent, we have,

$$\text{Cov}(X^{(e)}(t + \tau) | X^{(e)}(t) = x) = \tau \sum_{j=1}^M \nu_j^2 a_j(x). \quad (4.2.1)$$

Now using the equations (4.1.6) and (4.2.1), we will get the local error in covariance for explicit tau-leaping method as below (see also [37])

$$\begin{aligned}\text{Cov}(X^{(e)}(t + \tau) | X^{(e)}(t) = x) &= \\ &- \text{Cov}(X(t + \tau) | X(t) = x) = \\ &- \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1}^2 a_{j_2}(x) \{a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)\} \\ &- \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \nu_{j_2} a_{j_1}(x) \{a_{j_2}(x + \nu_{j_1}) - a_{j_2}(x)\} \\ &- \frac{1}{2}\tau^2 \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \nu_{j_2} a_{j_2}(x) \{a_{j_1}(x + \nu_{j_2}) - a_{j_1}(x)\} \\ &+ O(\tau^3).\end{aligned}$$

Similar results may be obtained for the implicit tau-leaping method.

Theorem 4.2.1. [37] *Consistency of rounded implicit tau-leaping method*

If bounded Poisson random numbers are used then for any polynomial $g : \mathbb{R}^N \rightarrow \mathbb{R}$

and any possible system state $x \in \mathbb{Z}_+^N$ there exist $C > 0$ and $\delta > 0$ such that

$$|E(g(X^{(ir)}(t + \tau)) - g(X(t + \tau)) | X^{(ir)}(t) = X(t) = x) | < C\tau^2$$

for all $\tau \in [0, \delta]$.

Here, 'ir' denotes rounded implicit of the tau-leaping method.

The proof uses the observation that $X^{(0)}(t + \tau) = X^{(e)}(t + \tau)$ with probability 1, provided that $X^{(0)}(t) = X^{(e)}(t) = x$, for any $\tau \in [0, \delta]$.

Theorem 4.2.2. [37] *Consistency of unrounded implicit tau-leaping method*

If bounded Poisson random numbers are used, then for any polynomial $g : \mathbb{R}^N \rightarrow \mathbb{R}$ and any possible system state $x \in \mathbb{Z}_+^N$ there exist $C > 0$ and $\delta > 0$ such that

$$|E(g(X^{(i)}(t + \tau)) - g(X(t + \tau)) | X^{(i)}(t) = X(t) = x) | < C\tau^2$$

for all $\tau \in [0, \delta]$.

Let us find the formula for the local error of the implicit tau-leaping method using Taylor expansion in τ .

4.3 Local Error Formula for the Implicit Tau-leaping Method

Using Taylor's formula and writing $X^{(i)}(t + \tau) = X'$ and the intermediate state X'_e where X' is a deterministic function of X'_e we have:

$$\begin{aligned} X' &= X'_e + \tau \sum_{j=1}^M \nu_j \{a_j(X'_e) - a_j(x)\} \\ &+ \frac{\tau^2}{2} \sum_{j_1=1}^M \sum_{j_2=1}^M \nu_{j_1} \frac{\partial a_{j_1}}{\partial x}(X'_e) \nu_{j_2} \{a_{j_2}(X'_e) - a_{j_2}(x)\} \\ &+ O(\tau^3) \end{aligned} \tag{4.3.1}$$

Assuming that bounded Poisson numbers are used, for any function $g : \mathbb{R}^N \rightarrow \mathbb{R}$ that is Lipschitz continuous on any bounded domain of \mathbb{R}^N and any possible system state x there exist $K > 0$ and $\delta > 0$ such that

$$|E(g(X^{(e)}(t + \tau) - g(x) | X^{(e)}(t) = x)| < K\tau$$

for all $\tau \in [0, \delta]$.

This shows that double summation component in equation (4.3.1) can be replaced by $O(\tau^3)$. Hence, the equation given below can be used to compare the local error in implicit tau-leaping method to that of explicit tau-leaping scheme:

$$X^{(i)}(t + \tau) = X^{(e)}(t + \tau) + \tau \sum_{j=1}^M \nu_j \{a_j(X^{(e)}(t + \tau)) - a_j(x)\} + O(\tau^3)$$

Here, we assume $X^{(i)}(t) = X^{(e)} = x$. In order to calculate the r^{th} moment of $X^{(i)}(t + \tau)$, we need to raise the above equation to the power r and take expectation on both sides.

Below we present some theorems of weak convergence of the explicit and implicit tau-leaping methods.

Theorem 4.3.1. [37] *Order-1 weak convergence of explicit tau-leaping method*

Assuming the propensity functions are linear and x is a possible state, if the explicit tau-leaping method is applied with $X^{(e)}(t_0) = x$ and the time steps $\tau_1, \tau_2, \dots, \tau_n$ are such that $\sum_{j=1}^n \tau_j = T$, $\tau = \max_{j=1, \dots, n} \{\tau_j\}$, then for any positive integer r , there exist constants $C > 0$ and $\delta > 0$ such that

$$\|\mathbb{E}((X^{(e)}(t_0 + T))^r) - \mathbb{E}((X(t_0 + T))^r)\| < C\tau$$

for all $\tau \in (0, \delta)$.

The proof of this theorem, which is beyond the scope of this thesis, is based on the standard result in numerical analysis stating that consistency and zero-stability imply convergence.

Theorem 4.3.2. [37] *Order-1 weak convergence of implicit tau-leaping method*

Assuming the propensity functions are linear and x is a possible state, if the implicit tau-leaping method is applied with $X^{(i)}(t_0) = x$ and the time steps $\tau_1, \tau_2, \dots, \tau_n$ are such that $\sum_{j=1}^n \tau_j = T$, $\tau = \max_{j=1, \dots, n} \{\tau_j\}$, then for any positive integer r , there exist constants $C > 0$ and $\delta > 0$ such that

$$\|\mathbb{E}((X^{(i)}(t_0 + T))^r) - \mathbb{E}((X(t_0 + T))^r)\| < C\tau$$

for all $\tau \in (0, \delta)$.

Under the assumption that all propensity functions are locally Lipschitz, Li[31] showed that the explicit tau-leaping method has strong order of convergence 1/2 in

L^2 and weak order of convergence 1. In this case, if the tau-leaping method is applied on a mesh $0 = t_0 < t_1 < \dots < t_n = T$ of the interval $[0, T]$, then there exists $C > 0$ such that, for $\tau = \max_{0 \leq l \leq n-1} (t_{l+1} - t_l)$ we obtain

$$\sup_{l \leq n} \mathbb{E} \left(|X^{(e)}(t_l) - X(t_l)|^2 \right) \leq C\tau$$

and

$$|\mathbb{E}(g(X^{(e)}(t_n))) - \mathbb{E}(g(X(t_n)))| \leq C\tau.$$

This shows that the explicit tau-leaping scheme has strong order of convergence $1/2$ and weak order of convergence 1. This is similar with the results for the Euler-Maruyama strategy for SDE.

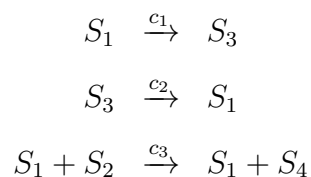
Chapter 5

Numerical Experiments

In this chapter , we study the behavior of the higher order tau-leaping methods, the midpoint and trapezoidal tau-leaping strategies, compared to the lower order tau-leaping schemes and the exact stochastic simulation algorithm. We test these methods on three models of biochemical systems of practical interest. On each model, we simulate 10,000 trajectories for each method and compare the resulting histograms with the histograms obtained using 10,000 trajectories generated with the exact stochastic simulation algorithm of Gillespie.

5.1 Simple Reaction Channel

Consider first the simple reaction model from having four species and three reaction channels [36]:



The total number of S_1 and S_3 molecules is constant (denoted by x_T). If we do not consider the by-product S_4 , then this model can be expressed as system with two variables, x_1 and x_2 , where x_1 and x_2 are the molecular counts of S_1 and S_2 respectively.

The initial conditions and the reaction rate parameters are

$$\begin{aligned}c_1 &= 10^3 \\c_2 &= 10^3 \\c_3 &= 0.0005 \\x_1(0) &= 1000 \\x_2(0) &= 100 \\x_T &= 2000.\end{aligned}$$

The propensity functions for the reactions above are as follows:

$$\begin{aligned}a_1 &= c_1 x_1 \\a_2 &= c_2 (x_T - x_1) \\a_3 &= c_3 x_1 x_2\end{aligned}$$

The system is studied on the time interval $[0, \frac{0.05}{4}]$.

The state change vectors for the reactions R_1, R_2 and R_3 , respectively are

$$v_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

We simulated 10000 trajectories for each of the midpoint, trapezoidal, implicit and explicit tau-leaping methods and Gillespie's algorithm and plotted the histograms for these strategies at $T = 0.05/4$ for each of the species. Figure 5.1 shows the histograms for the species S_1 , while figure 5.2 presents the histograms for the species S_2 .

Sample time-evaluation trajectories for the species S_1 and S_2 are presented in figures 5.3 and 5.4, respectively.

From Figures 5.1 and 5.2, we see that the midpoint and trapezoidal tau-leaping schemes are slightly more accurate for the same size of the leap τ than the implicit and explicit tau-leaping techniques.

In Figure 5.5, we showed the graphs of the relative global errors of each of the tau-leaping methods at time $T = 0.05/4$, for the following sequence of step sizes $\tau = 3.125 \times 10^{-7}, 1.563 \times 10^{-7}$ and 7.81×10^{-8} . The loglog plots are presented. We can clearly see that the errors in the higher order methods are smaller than those for the lower order methods.

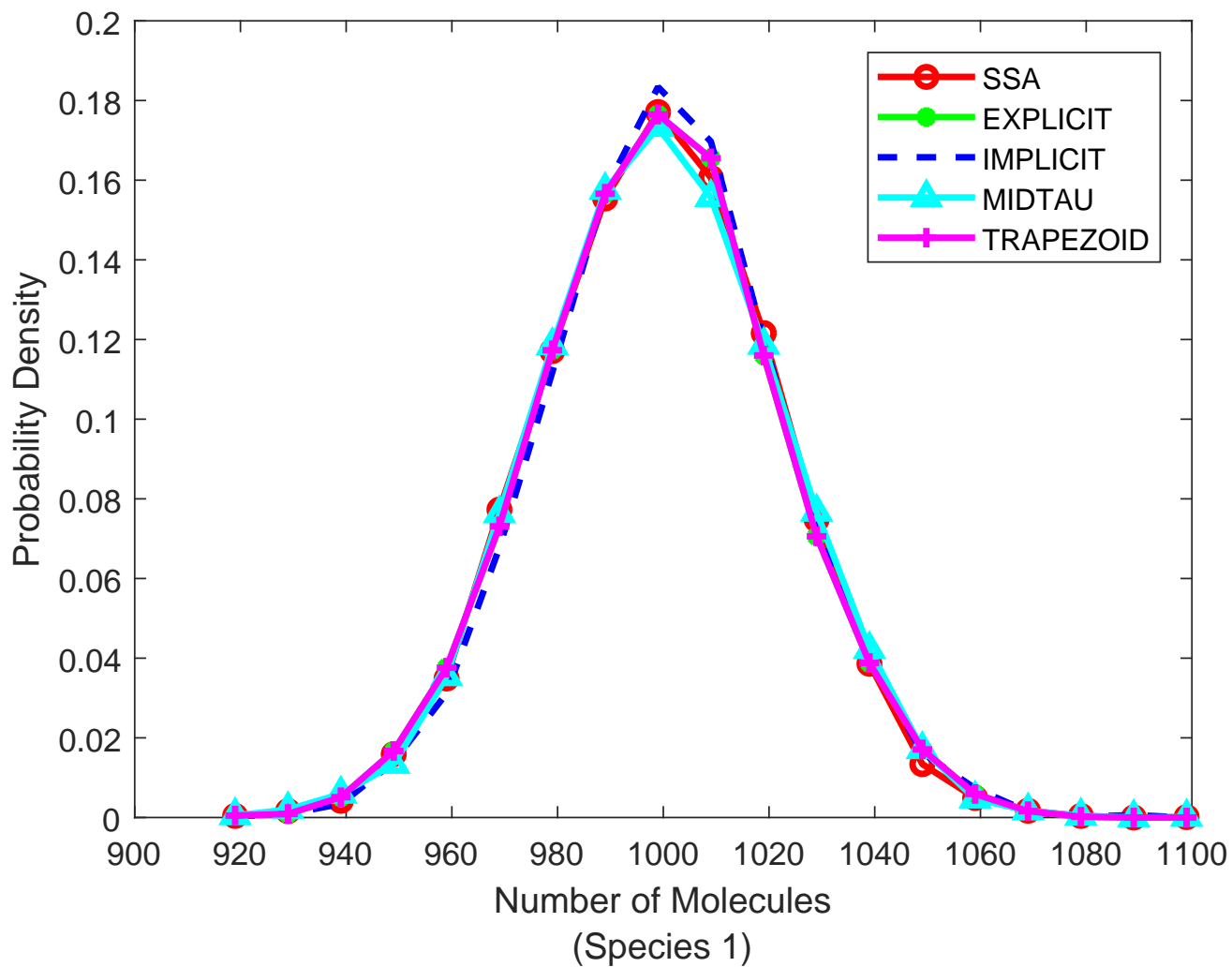


Figure 5.1: Simple reaction channel: Histograms of species S_1 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.05/4$.

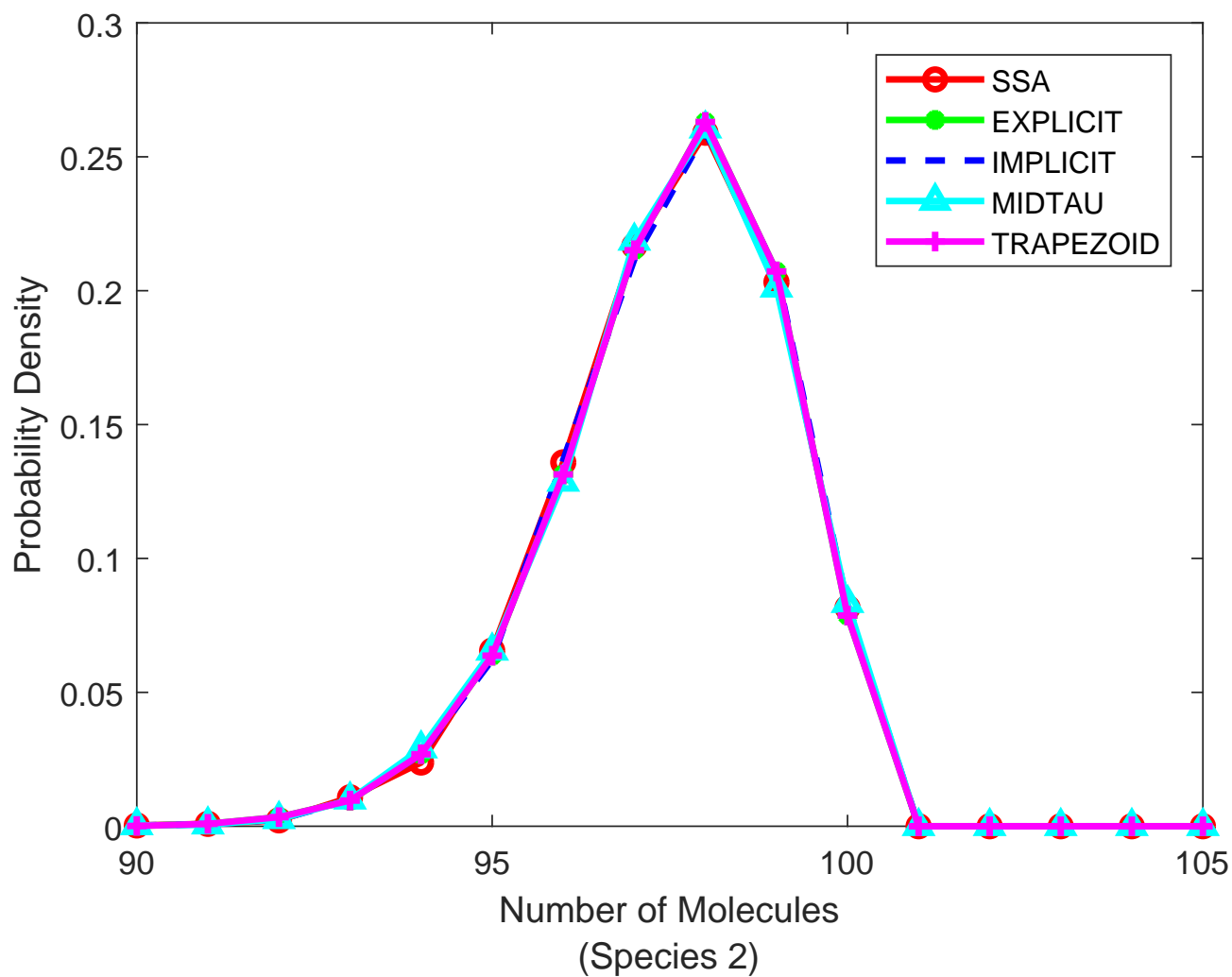


Figure 5.2: Simple reaction channel: Histograms of species S_2 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.05/4$.

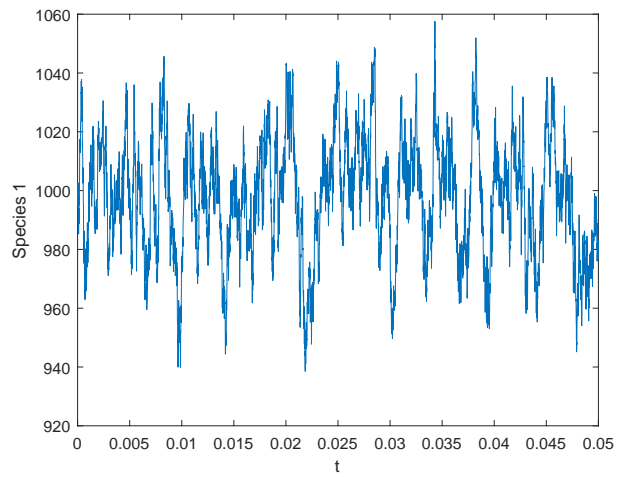


Figure 5.3: Simple reaction channel: A sample trajectory of number of molecules of species S_1 as function of time.

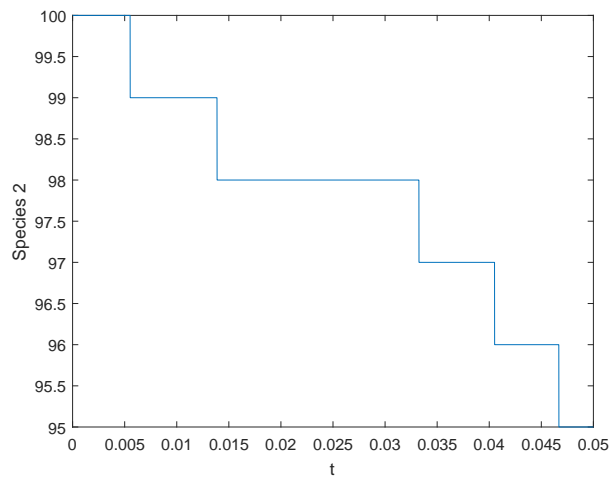


Figure 5.4: Simple reaction channel: A sample trajectory of number of molecules of species S_2 as function of time.

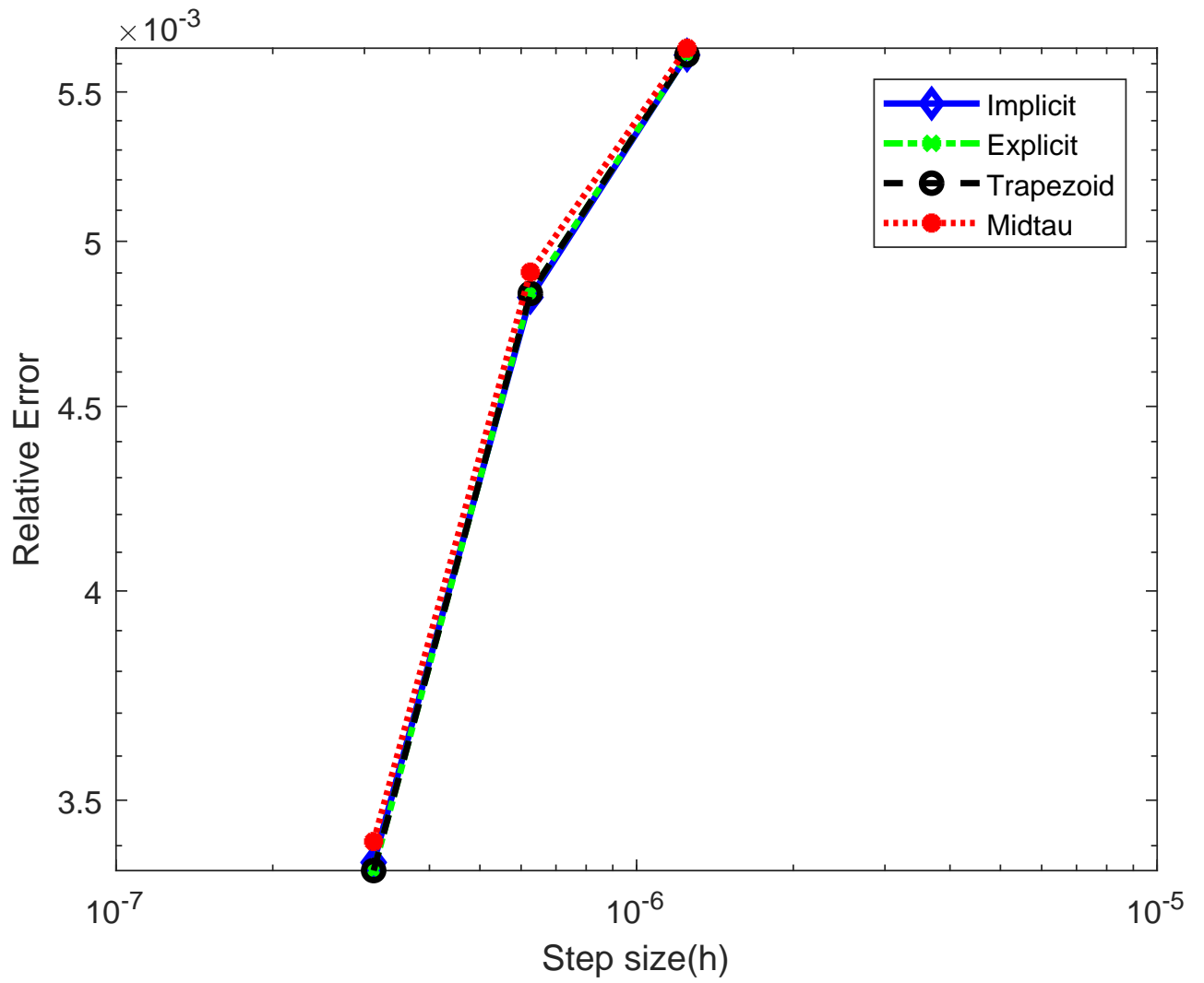
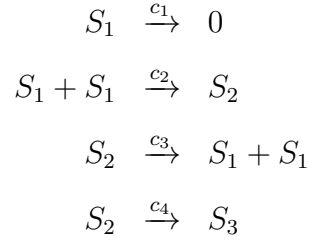


Figure 5.5: Simple reaction channel: Log log plot of the relative global error at time $T = 0.05/4$ as function of step size τ , for all the tau-leaping methods.

5.2 Decay-dimerization Reaction Channel

The second model considered in this thesis is the Decay-dimerization reaction system [36]. The system involves three species and four reactions:



The model is subjected to the following initial conditions:

$$\begin{aligned} x_1(0) &= 400 \\ x_2(0) &= 798 \\ x_3(0) &= 0. \end{aligned}$$

It has the following parameter values $c_1 = 1, c_2 = 10, c_3 = 1000$ and $c_4 = 0.1$.

To the reaction channels above correspond the following propensities

$$\begin{aligned} a_1 &= x_1 \\ a_2 &= 5x_1(x_1 - 1) \\ a_3 &= 1000x_2 \\ a_4 &= 0.1x_2 \end{aligned}$$

This model is integrated on the time interval $[0, \frac{0.2}{4}]$.

The state change vectors are

$$v_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} -2 \\ 1 \end{bmatrix}, v_3 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, v_4 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

As with the previous system, the decay-dimerization model is simulated using 10,000 trajectories using the midpoint, trapezoidal, implicit and explicit tau-leaping methods with the same value of the step-size τ and with 10,000 trajectories generated with the SSA. The histograms at $T = 0.02/4$ obtained using each of these methods for the species S_1 and S_2 are given in figure 5.6 and figure 5.7, respectively. It is worth noting that this model is stiff, having both fast and slow reactions.

The trapezoidal tau-leaping method is the most accurate on this model. The midpoint tau-leaping scheme is slightly less accurate on this stiff problem.

The evaluations in time of the molecular counts for species S_1 and S_2 on a sample SSA are plotted in Figures 5.8 and 5.9, respectively. In addition, the loglog plot of the relative error of each tau-leaping method at $T = 0.02/4$ compared to the reference SSA solution, as a function of the step size τ is shown in Figure 5.10. This Figure confirms that higher order tau-leaping methods are more accurate than lower order ones.

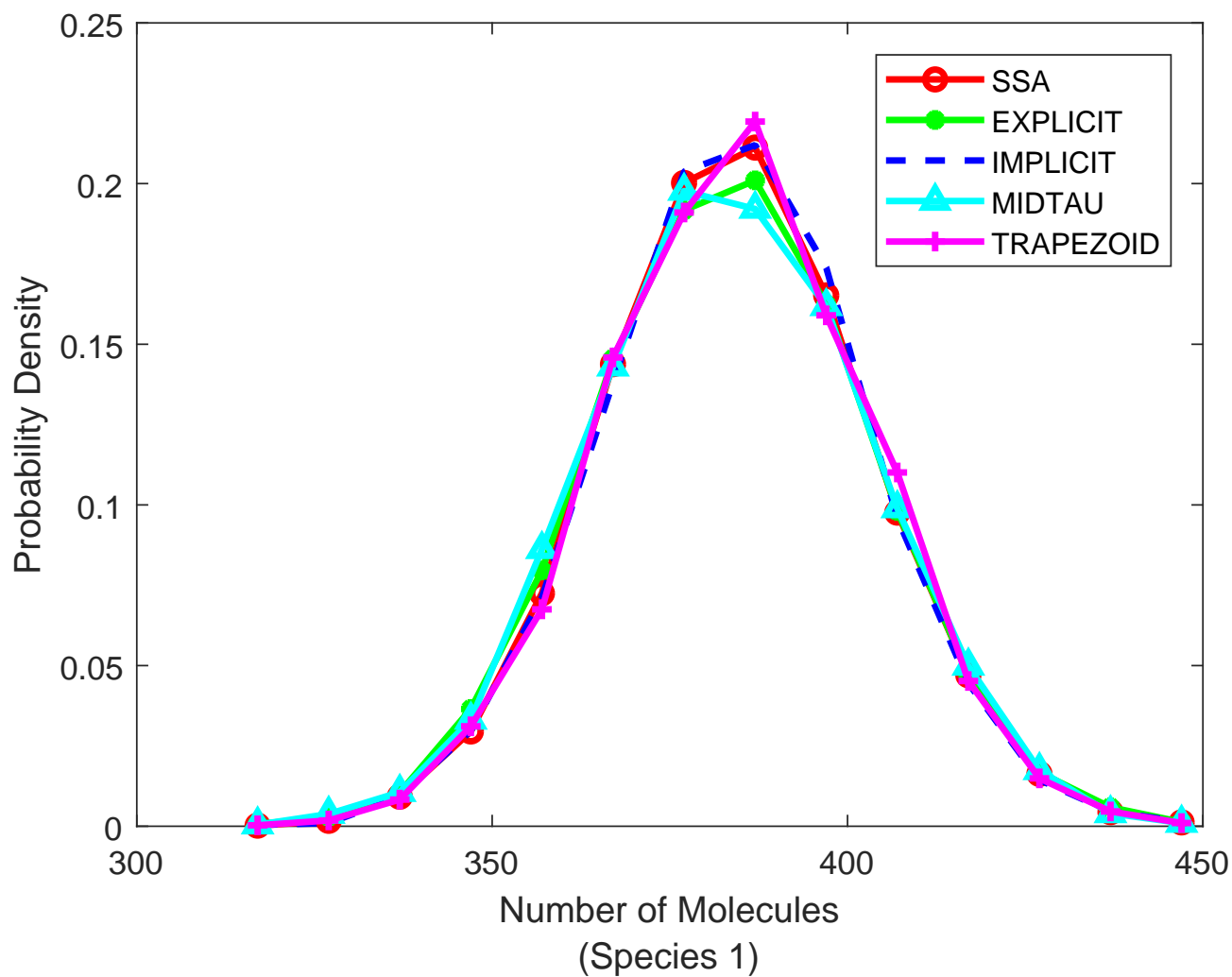


Figure 5.6: Decay dimerization reaction: Histograms of species S_1 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.02/4$.

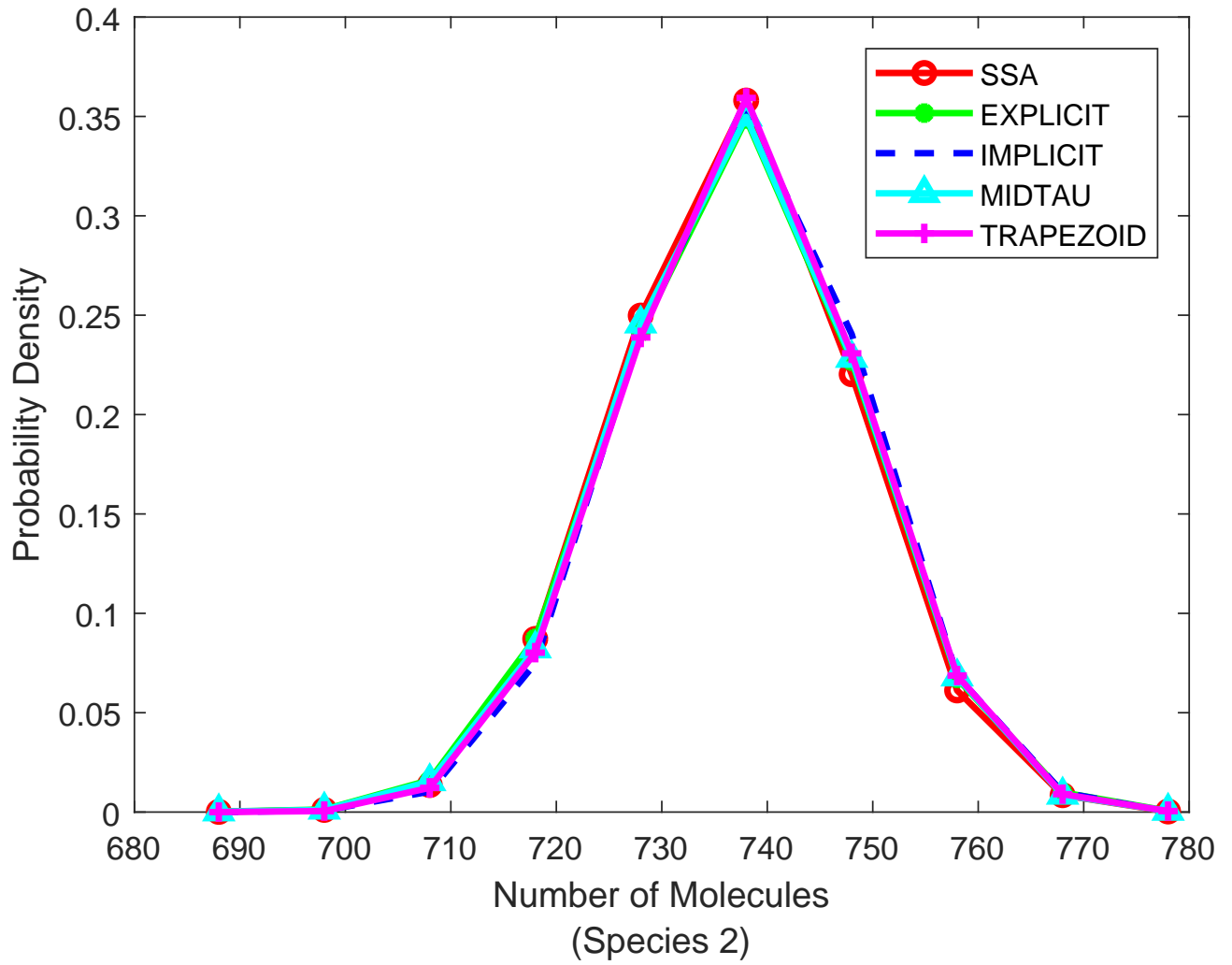


Figure 5.7: Decay dimerization reaction: Histograms of species S_2 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 0.02/4$.

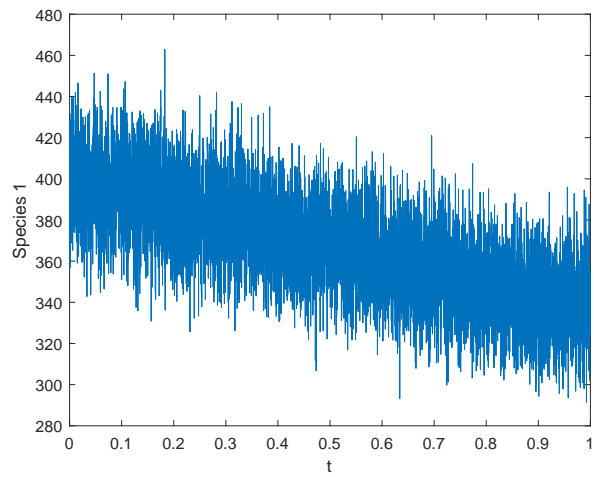


Figure 5.8: Decay dimerization reaction: A sample trajectory of number of molecules of species S_1 as function of time.

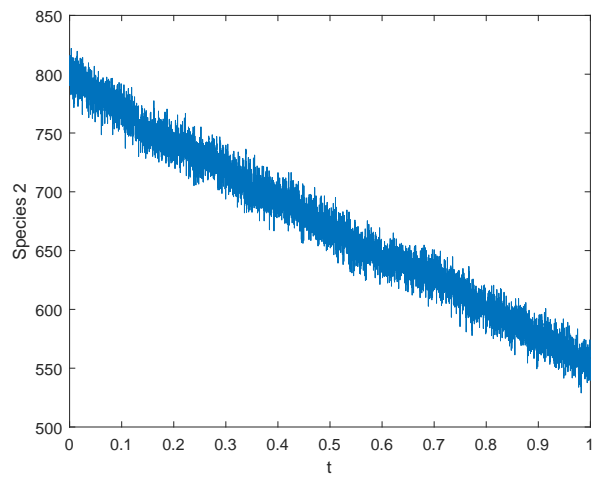


Figure 5.9: Decay dimerization reaction: A sample trajectory of number of molecules of species S_2 as function of time.

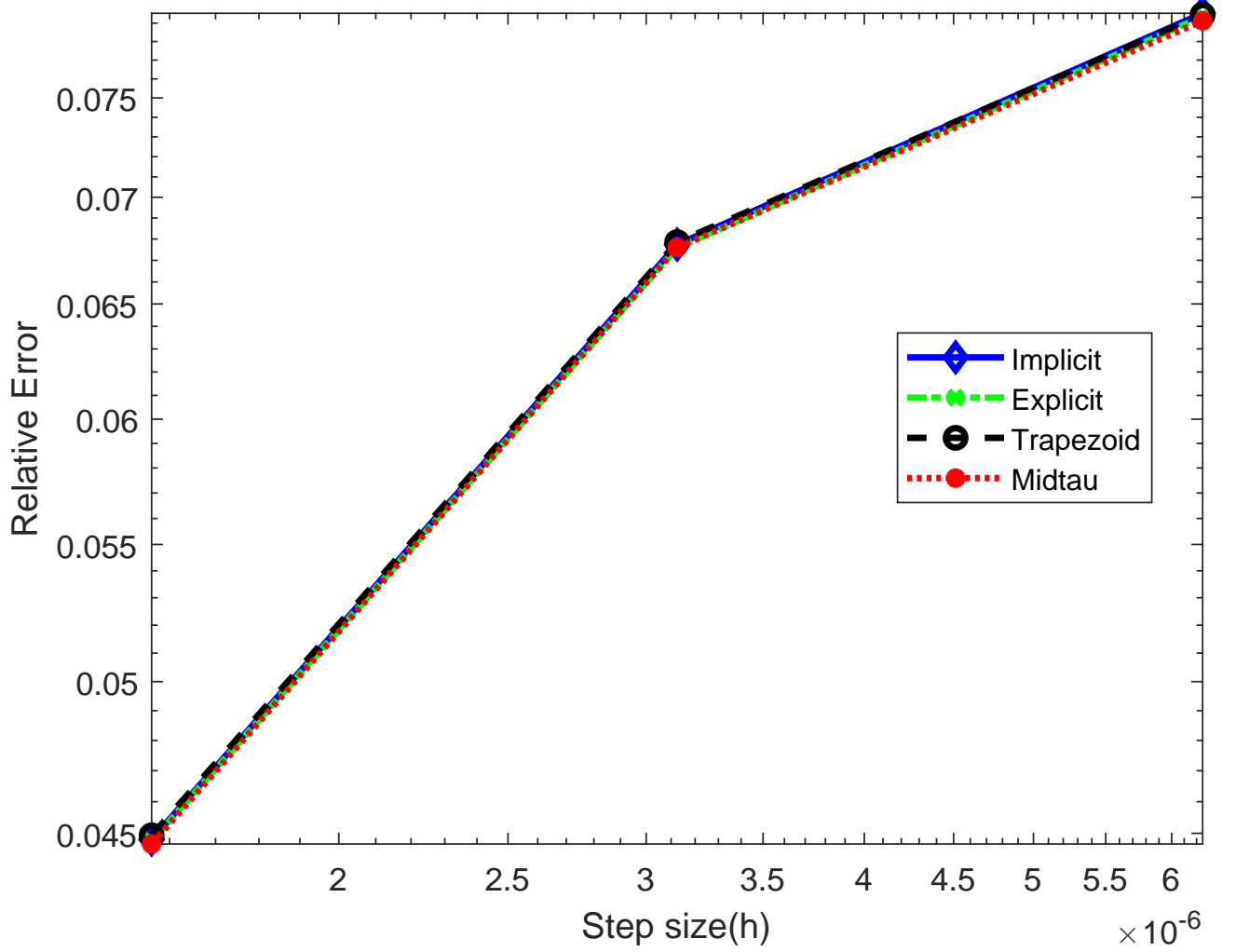
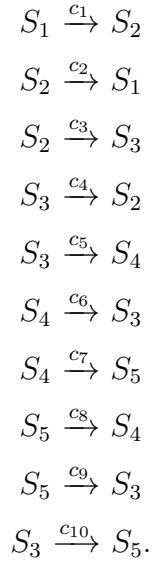


Figure 5.10: Decay dimerization reaction: Log log plot of the relative global error at time $T = 0.02/4$ as function of step size τ , for all the tau-leaping methods.

5.3 Potassium Channel

The final model we analyze is the Potassium channel (A K^+ channel) system. It consists of three closed states (S_1, S_2, S_3), one open state (S_4) and one activation state (S_5), there are $N = 5$ chemical species (S_1, S_2, S_3, S_4, S_5) reacting through 10 reactions as mentioned below [33],



The propensity functions of the reactions above are, respectively,

$$\begin{aligned}
a_1(x) &= c_1 x_1 \\
a_2(x) &= c_2 x_2 \\
a_3(x) &= c_3 x_2 \\
a_4(x) &= c_4 x_3 \\
a_5(x) &= c_5 x_3 \\
a_6(x) &= c_6 x_4 \\
a_7(x) &= c_7 x_4 \\
a_8(x) &= c_8 x_5 \\
a_9(x) &= c_9 x_5 \\
a_{10}(x) &= c_{10} x_3.
\end{aligned}$$

The molecular amount of the S_i species is denoted by x_i . The stoichiometric matrix of the Potassium channel model may be written as

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

. The initial conditions are set to $X(0) = [100, 50, 100, 50, 100]^T$,
while the values of the reaction rate parameters are given by

$$c(1) = 0.1$$

$$c(2) = 0.1$$

$$c(3) = 0.1$$

$$c(4) = 0.1$$

$$c(5) = 0.1$$

$$c(6) = 0.1$$

$$c(7) = 0.1$$

$$c(8) = 0.1$$

$$c(9) = 0.1$$

$$c(10) = 0.1$$

.

This relatively large size system is integrated on the time interval $[0, 5]$.

A sample size of 10,000 trajectories are simulated with the tau-leaping methods considered in this thesis, using the same leap size τ and the results are compared to those generated employing 10,000 trajectories of the exact SSA. The histograms at $T = 5$ of each of the species S_1, S_2, S_3, S_4 and S_5 are given in Figures 5.11, 5.12, 5.13,

5.14 and 5.15., respectively. All tau-leaping methods are accurate on this model with the step-size used, the midpoint and trapezoidal tau-leaping techniques are slightly more accurate than the lower order implicit and explicit tau-leaping schemes, as expected. As with the previous models, for this larger model the higher order tau-leaping algorithms have higher order of accuracy as tested numerically.

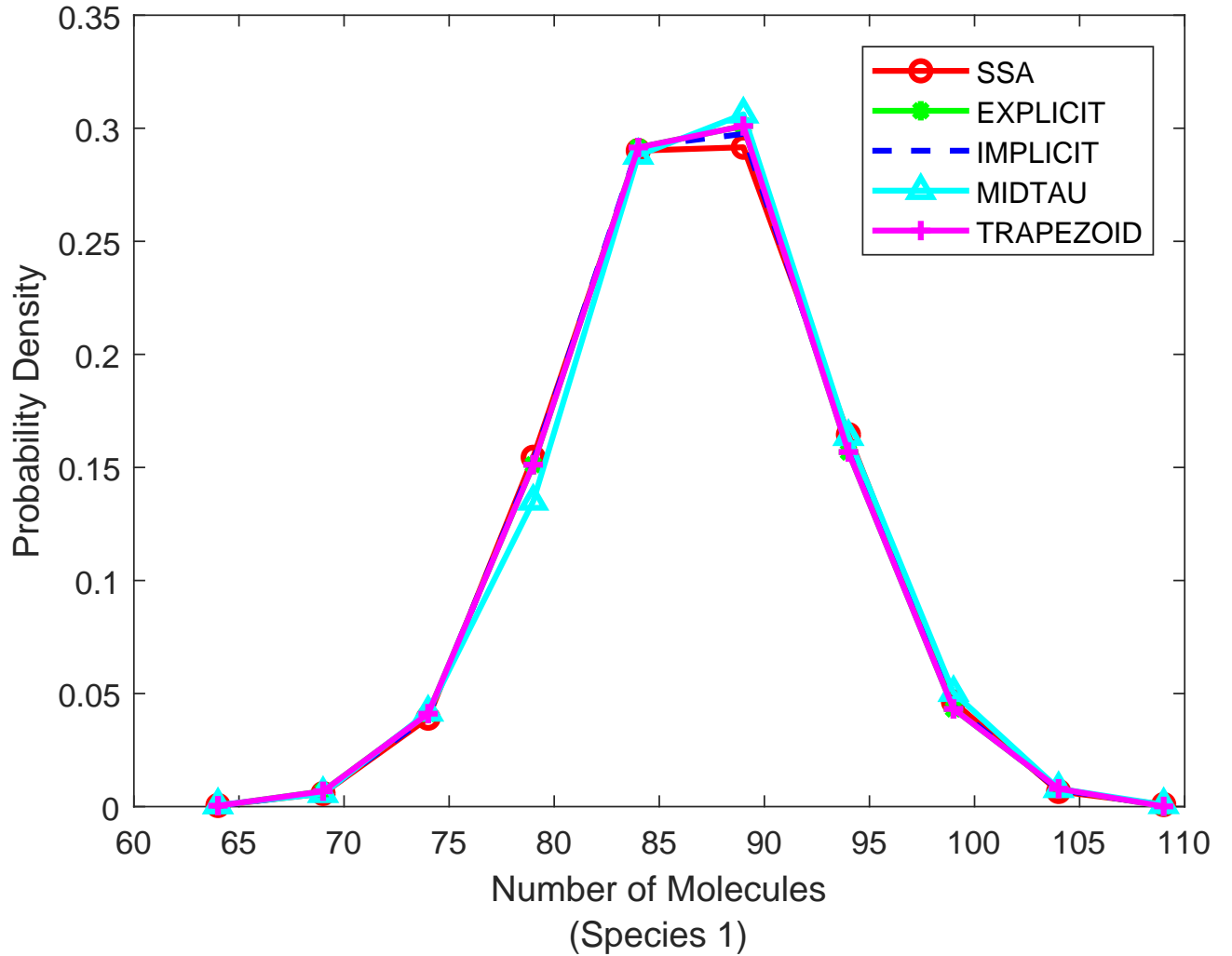


Figure 5.11: Potassium channel: Histograms of species S_1 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$.

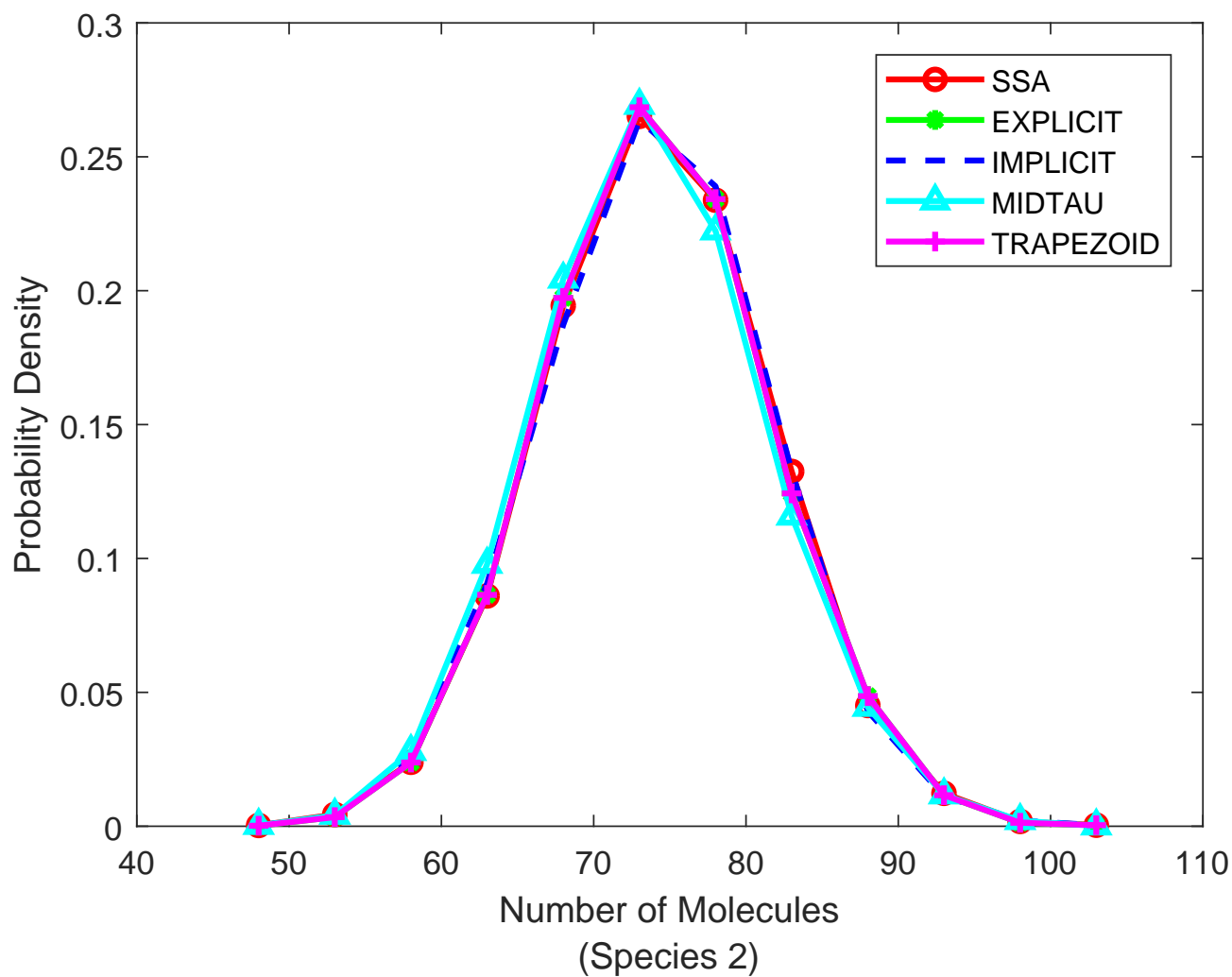


Figure 5.12: Potassium channel: Histograms of species S_2 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$.

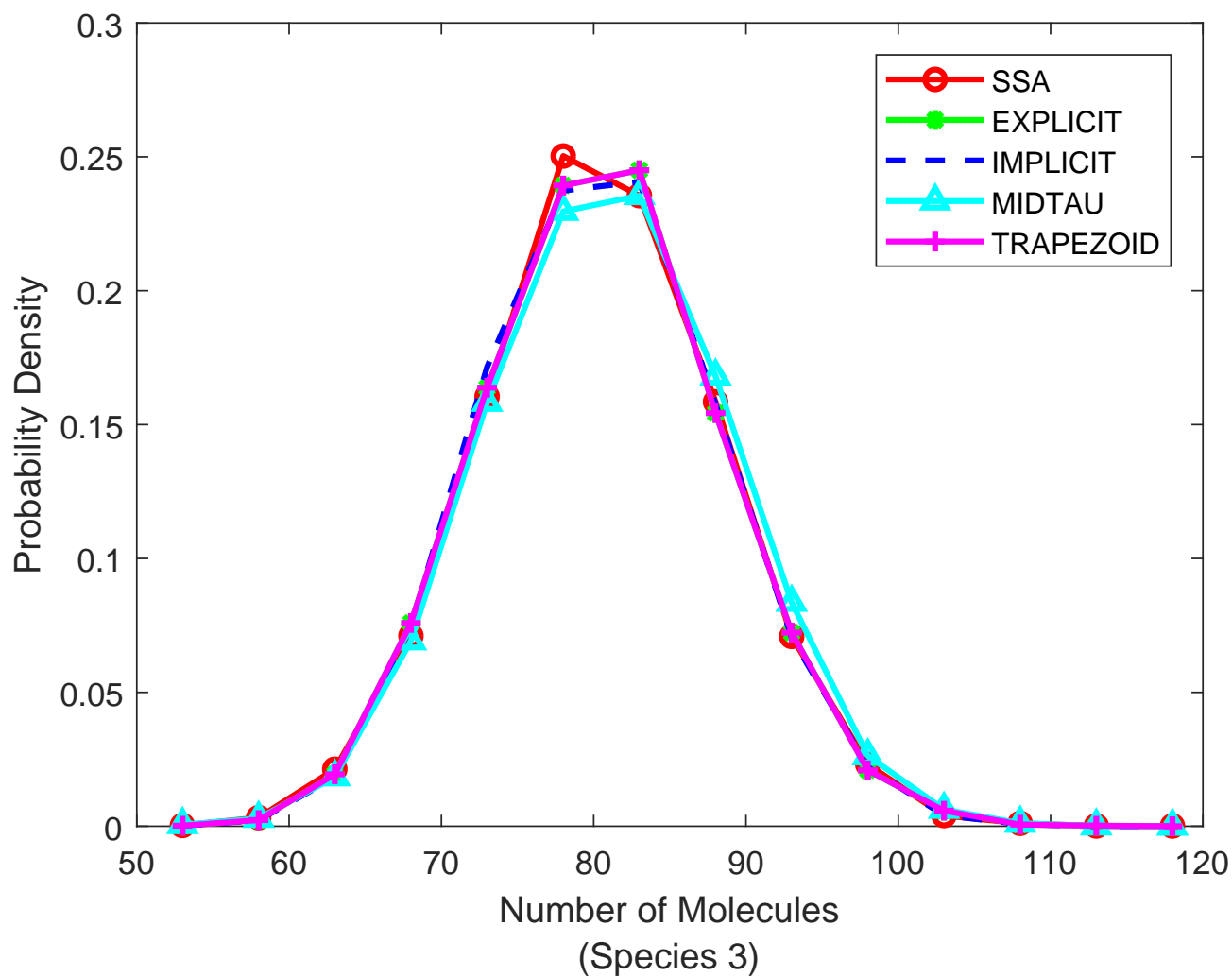


Figure 5.13: Potassium channel: Histograms of species S_3 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$.

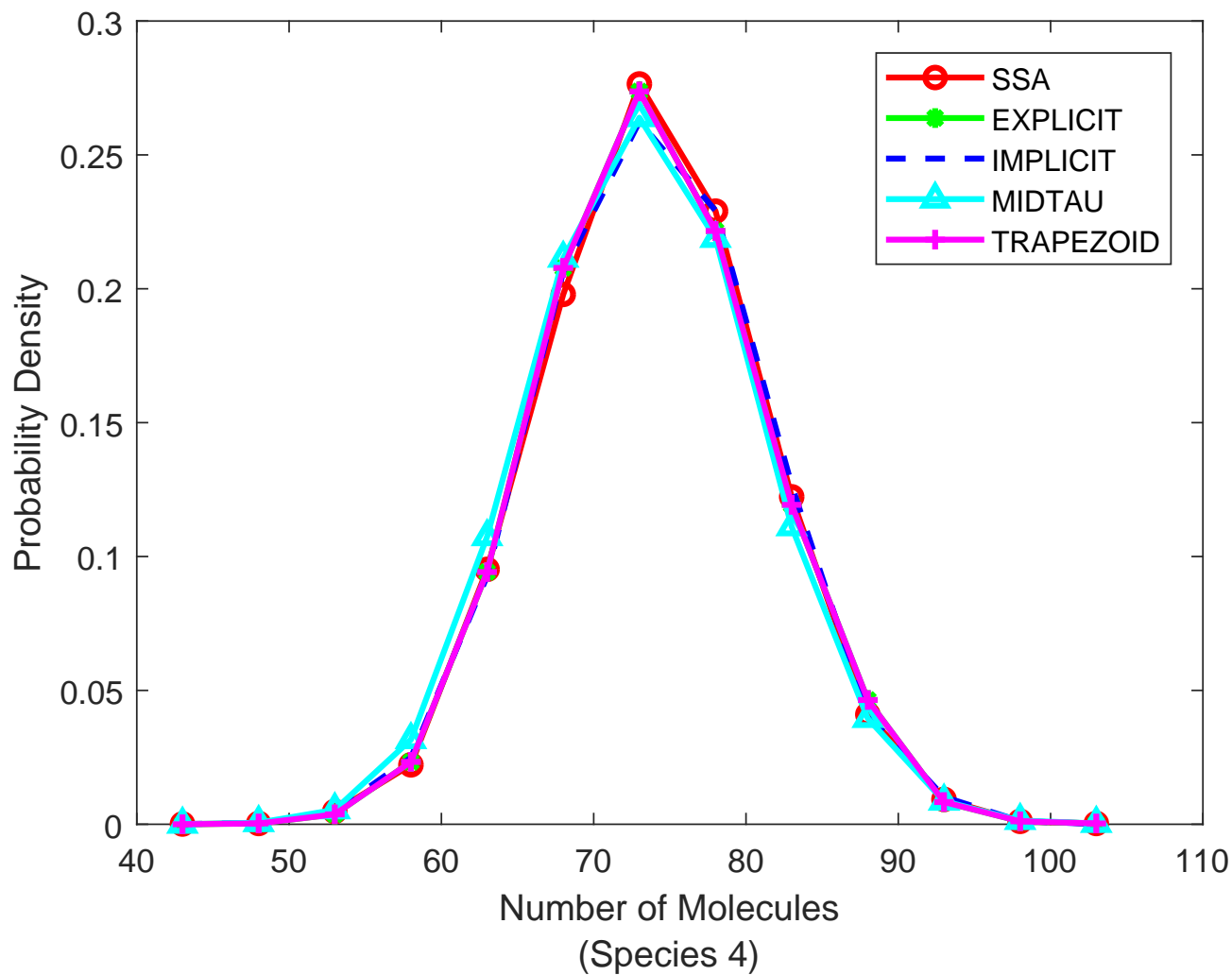


Figure 5.14: Potassium channel: Histograms of species S_4 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$.

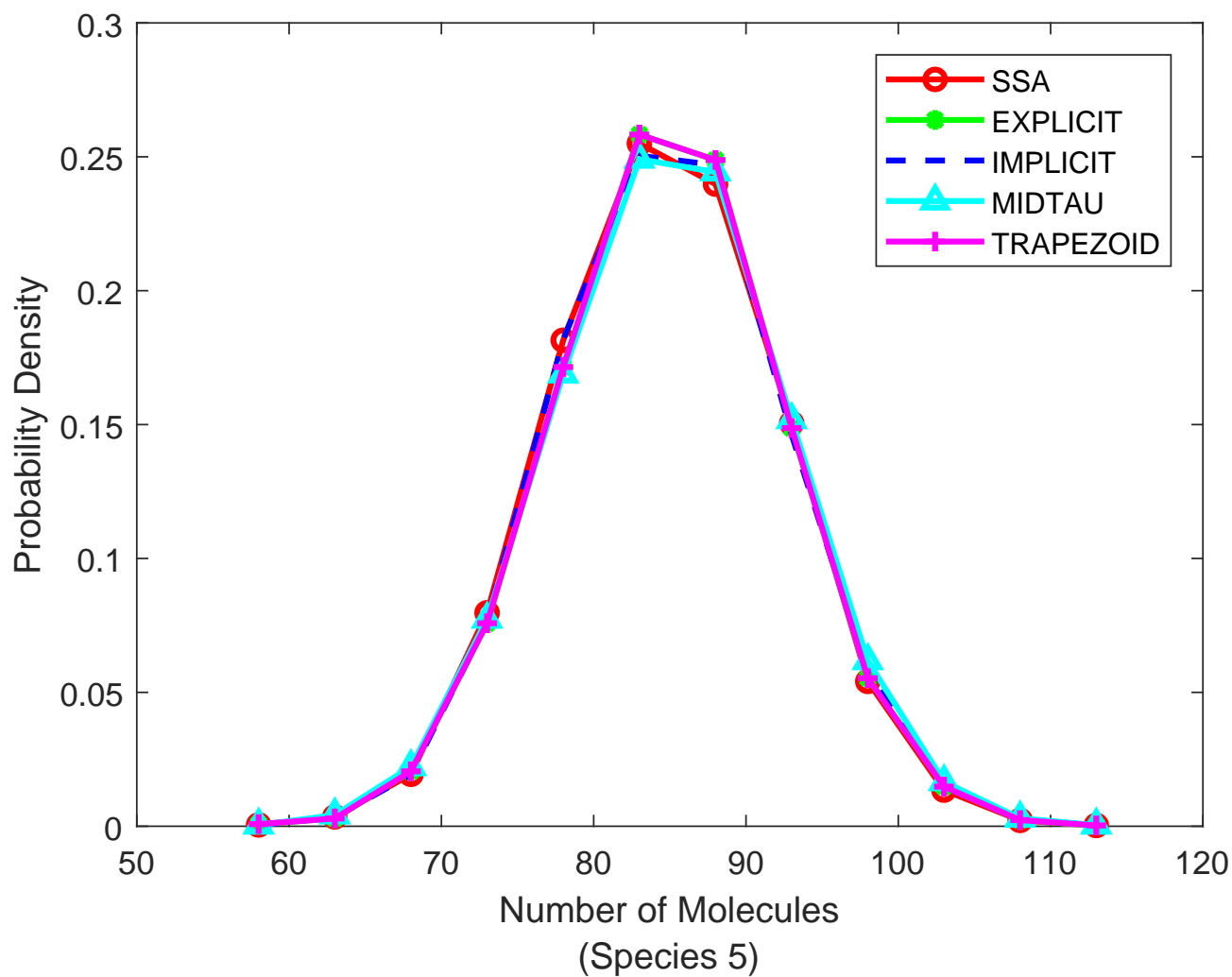


Figure 5.15: Potassium channel: Histograms of species S_5 computed using the SSA, the explicit, implicit, midpoint and trapezoidal tau-leaping methods at time $T = 5$.

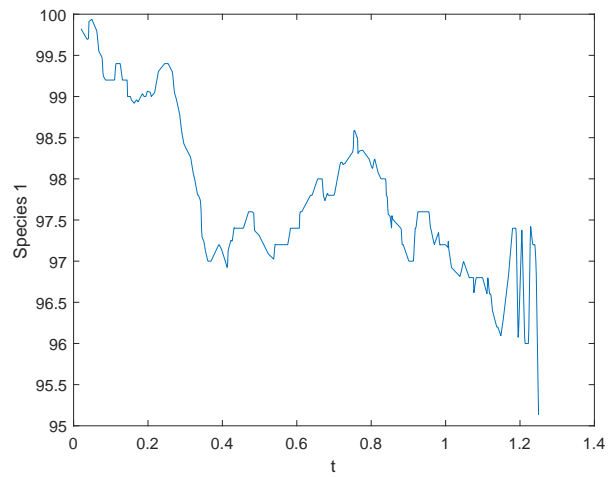


Figure 5.16: Potassium channel: A sample trajectory of number of molecules of species S_1 as function of time.

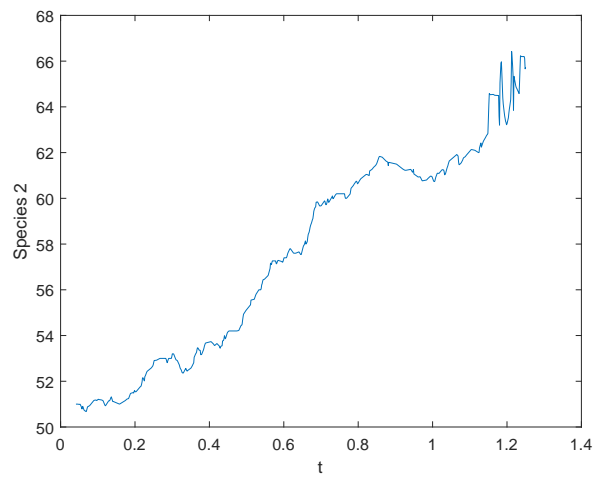


Figure 5.17: Potassium channel: A sample trajectory of number of molecules of species S_2 as function of time.

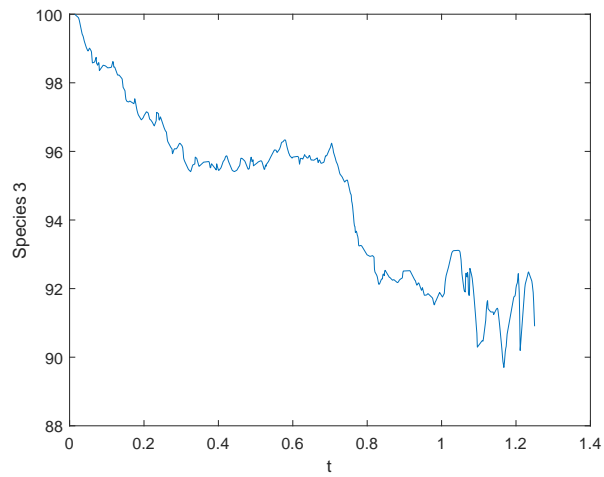


Figure 5.18: Potassium channel: A sample trajectory of number of molecules of species S_3 as function of time.

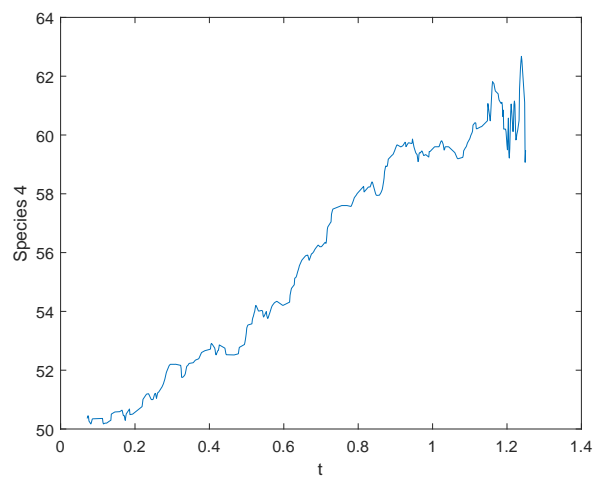


Figure 5.19: Potassium channel: A sample trajectory of number of molecules of species S_4 as function of time.

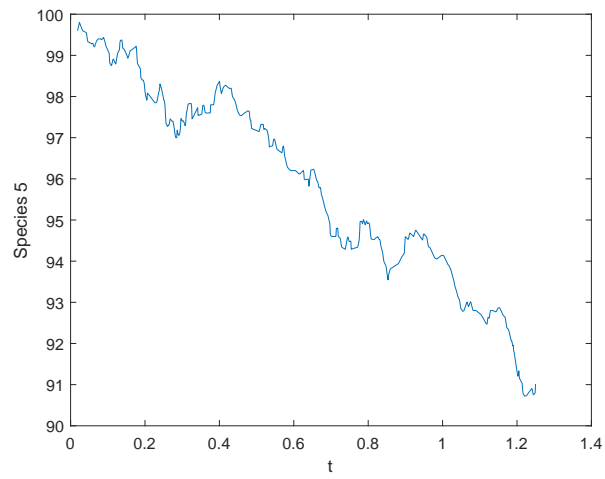


Figure 5.20: Potassium channel: A sample trajectory of number of molecules of species S_5 as function of time.

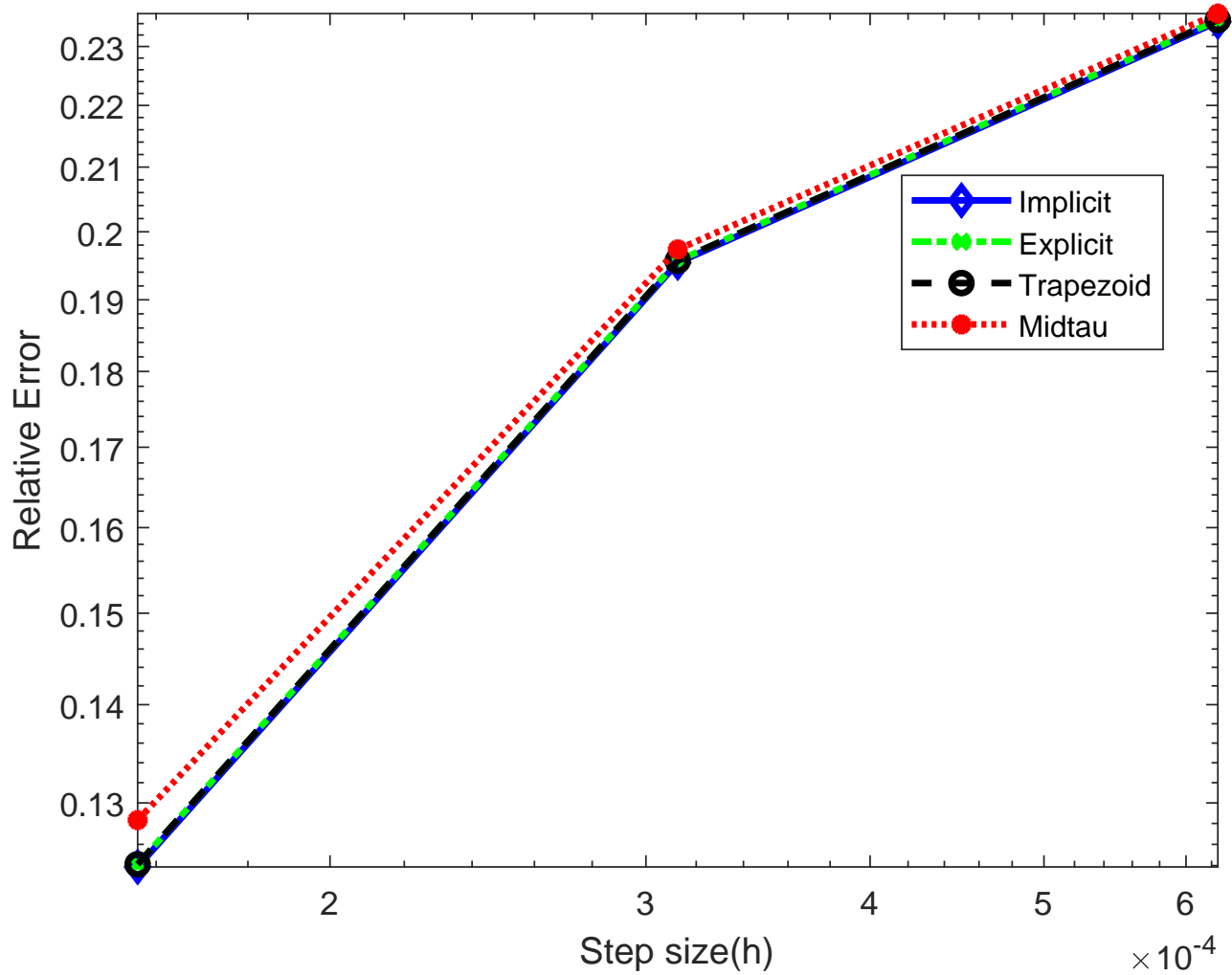


Figure 5.21: Potassium channel: Log log plot of the relative global error at time $T = 5/4$ as function of step size τ , for all the tau-leaping methods.

Chapter 6

Conclusion

Mathematical modeling and simulation of biochemical systems are important research areas in systems biology. In recent years, mathematical modeling and computer simulations have been recognized as some of the key tools for studying biochemical process taking place in a single cell. With the advance of technology, more accurate biological data become available. Those data may be used for building more realistic models of critical processes at the cellular trend. In the case of processes involving some biochemical species in low molecular counts (such as DNA and RNA), stochastic models are necessary for an accurate description of the system dynamics. Stochastic models are more complex than the deterministic ones and more challenging to solve numerically.

In this thesis, we study a stochastic discrete model of well-stirred biochemical systems known as the Chemical Master Equation. Gillespie proposed an exact Monte Carlo method for the Chemical Master Equation, the Stochastic Simulation Algorithm or the SSA. Nonetheless, the SSA is very expensive on many realistic biochemical systems arising in applications. To deal with this difficulty, Gillespie introduced the approximate tau-leaping method, which steps over many reactions during one simulation step. While the explicit tau-leaping method is applicable to non stiff models of biochemical reactions, the implicit tau-leaping scheme is more efficient

for stiff model. The explicit and implicit tau-leaping techniques have low order of accuracy. This work focuses on the study of higher order tau-leaping methods, namely the midpoint and trapezoidal tau-leaping schemes. We study these two higher order methods and compare their performances with that of the low order methods, the explicit and implicit tau-leaping schemes and the exact, but expensive SSA.

Our numerical tests on several systems encountered in applications showed that the higher order leaping strategies give more accurate numerical results than the low order leaping schemes and are therefore preferred when high accuracy of the simulation is required. This is important, for example, when studying biochemical systems which are sensitive with respect to some of their parameters.

In the future, we plan to study efficient time-stepping strategies for higher order tau-leaping methods.

Bibliography

- [1] David Adalsteinsson, David McMillen, and Timothy C Elston. Biochemical network stochastic simulator (bionets): software for stochastic modeling of biochemical networks. *BMC Bioinformatics*, 5(1):24, 2004.
- [2] Aurélien Alfonsi, Eric Cances, Gabriel Turinici, Barbara Di Ventura, and Wilhelm Huisinga. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. In *ESAIM: proceedings*, volume 14, pages 1–13. EDP Sciences, 2005.
- [3] David F Anderson. Incorporating postleap checks in tau-leaping. *The Journal of Chemical Physics*, 128(5):054103, 2008.
- [4] David F Anderson, Arnab Ganguly, Thomas G Kurtz, et al. Error analysis of tau-leap simulation methods. *The Annals of Applied Probability*, 21(6):2226–2262, 2011.
- [5] Steven S Andrews, Tuan Dinh, and Adam P Arkin. Stochastic models of biological processes. In *Encyclopedia of Complexity and Systems Science*, pages 8730–8749. Springer, 2009.
- [6] Adam Arkin, John Ross, and Harley H McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ -infected escherichia coli cells. *Genetics*, 149(4):1633–1648, 1998.

- [7] Uri M Ascher and Linda R Petzold. *Computer methods for ordinary differential equations and differential-algebraic equations*, volume 61. Siam, 1998.
- [8] Karen Ball, Thomas G Kurtz, Lea Popovic, Greg Rempala, et al. Asymptotic analysis of multiscale approximations to reaction networks. *The Annals of Applied Probability*, 16(4):1925–1961, 2006.
- [9] Yang Cao and Linda Petzold. Trapezoidal tau-leaping formula for the stochastic simulation of biochemical systems. *Proceedings of Foundations of Systems Biology in Engineering (FOSBE 2005)*, pages 149–152, 2005.
- [10] Abhijit Chatterjee, Dionisios G Vlachos, and Markos A Katsoulakis. Binomial distribution based τ -leap accelerated stochastic simulation. *The Journal of Chemical Physics*, 122(2):024112, 2005.
- [11] Edmund J Crampin and Santiago Schnell. New approaches to modelling and analysis of biochemical reactions, pathways and networks, 2004.
- [12] Hana El Samad, Mustafa Khammash, Linda Petzold, and Dan Gillespie. Stochastic modelling of gene regulatory networks. *International Journal of Robust and Nonlinear Control*, 15(15):691–711, 2005.
- [13] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.
- [14] Nina Fedoroff and Walter Fontana. Small numbers of big molecules. *Science*, 297(5584):1129–1131, 2002.
- [15] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

- [16] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.
- [17] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
- [18] Daniel T Gillespie. *Markov processes: an introduction for physical scientists*. Elsevier, 1991.
- [19] Daniel T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.
- [20] Daniel T Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.
- [21] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.
- [22] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.
- [23] Daniel T Gillespie and L Petzold. Numerical simulation for biochemical kinetics. *Systems Modelling in Cellular Biology*, pages 331–354, 2006.
- [24] Didier Gonze and Adama Ouattara. Stochastic simulations. 2007.
- [25] Eric L Haseltine and James B Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *The Journal of chemical physics*, 117(15):6959–6969, 2002.
- [26] Eric L Haseltine and James B Rawlings. On the origins of approximations for stochastic chemical kinetics. *The Journal of Chemical Physics*, 123(16):164115, 2005.

- [27] Desmond J Higham. Modeling and simulating chemical reactions. *SIAM review*, 50(2):347–368, 2008.
- [28] Tobias Jahnke. An adaptive wavelet method for the chemical master equation. *SIAM Journal on Scientific Computing*, 31(6):4373–4394, 2010.
- [29] Hiroaki Kitano. Computational systems biology. *Nature*, 420(6912):206, 2002.
- [30] Thomas G Kurtz. The relationship between stochastic and deterministic models for chemical reactions. *The Journal of Chemical Physics*, 57(7):2976–2978, 1972.
- [31] Tiejun Li. Analysis of explicit tau-leaping schemes for simulating chemically reacting systems. *Multiscale Modeling & Simulation*, 6(2):417–436, 2007.
- [32] Harley H McAdams and Adam Arkin. Itsa noisy business! genetic regulation at the nanomolar scale. *Trends in Genetics*, 15(2):65–69, 1999.
- [33] Bence Mélykúti, Kevin Burrage, and Konstantinos C Zygalakis. Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical langevin equation. *The Journal of Chemical Physics*, 132(16):164109, 2010.
- [34] Johan Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415, 2004.
- [35] Jacek Puchalka and Andrzej M Kierzek. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophysical Journal*, 86(3):1357–1372, 2004.
- [36] Muruhan Rathinam, Linda R Petzold, Yang Cao, and Daniel T Gillespie. Stiffness in stochastic chemically reacting systems: The implicit tau-leaping method. *The Journal of Chemical Physics*, 119(24):12784–12794, 2003.

- [37] Muruhan Rathinam, Linda R Petzold, Yang Cao, and Daniel T Gillespie. Consistency and stability of tau-leaping schemes for chemical reaction systems. *Multiscale Modeling & Simulation*, 4(3):867–895, 2005.
- [38] Asawari Samant and Dionisios G Vlachos. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: a multiscale monte carlo algorithm. *The Journal of Chemical Physics*, 123(14):144114, 2005.
- [39] Michael Samoilov, Sergey Plyasunov, and Adam P Arkin. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7):2310–2315, 2005.
- [40] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.
- [41] Kouichi Takahashi, Kazunari Kaizu, Bin Hu, and Masaru Tomita. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics*, 20(4):538–546, 2004.
- [42] Mukund Thattai and Alexander Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.
- [43] Karan Vasudeva and Upinder S Bhalla. Adaptive stochastic-deterministic chemical kinetic simulations. *Bioinformatics*, 20(1):78–84, 2004.
- [44] José MG Vilar, Hao Yuan Kueh, Naama Barkai, and Stanislas Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.

- [45] Darren J Wilkinson. *Stochastic modelling for systems biology*. Chapman and Hall/CRC, 2006.
- [46] Verena Wolf, Rushil Goel, Maria Mateescu, and Thomas A Henzinger. Solving the chemical master equation using sliding windows. *BMC Systems Biology*, 4(1):42, 2010.