IT Licentiate theses
2008-003

# Numerical simulation of well stirred biochemical reaction networks governed by the master equation

ANDREAS HELLANDER

UPPSALA UNIVERSITY
Department of Information Technology

UPPSALA
UNIVERSITET

# Numerical simulation of well stirred biochemical reaction networks governed by the master equation


BY
ANDREAS HELLANDER


October 2008


DIVISION OF SCIENTIFIC COMPUTING
DEPARTMENT OF INFORMATION TECHNOLOGY
UPPSALA UNIVERSITY
UPPSALA
SWEDEN



Dissertation for the degree of Licentiate of Philosophy in Scientific Computing
at Uppsala University 2008

# Numerical simulation of well stirred biochemical reaction networks governed by the master equation

*Andreas Hellander*
Andreas.Hellander@it.uu.se

*Division of Scientific Computing*
*Department of Information Technology*
*Uppsala University*
*Box 337*
*SE-751 05 Uppsala*
*Sweden*

http://www.it.uu.se/


ISSN 1404-5117
Printed by the Department of Information Technology, Uppsala University, Sweden

**Abstract**

Numerical simulation of stochastic biochemical reaction networks has received much attention in the growing field of computational systems biology. Systems are frequently modeled as a continuous–time discrete space Markov chain, and the governing equation for the probability density of the system is the (chemical) master equation. The direct numerical solution of this equation suffers from an exponential growth in computational time and memory with the number of reacting species in the model. As a consequence, Monte Carlo simulation methods play an important role in the study of stochastic chemical networks. The stochastic simulation algorithm (SSA) due to Gillespie has been available for more than three decades, but due to the multi–scale property of the chemical systems and the slow convergence of Monte Carlo methods, much work is currently being done in order to devise more efficient approximate schemes.

In this thesis we review recent work for the solution of the chemical master equation by direct methods, by exact Monte Carlo methods and by approximate and hybrid methods. We also describe two new and conceptually different numerical methods to reduce the computational time when studying models using SSA. A hybrid method is proposed, which is based on the separation of species into two subsets based on the variance of the copy numbers. This method yields a significant speed-up when the system permits such a splitting of the state space. A different approach is taken in an algorithm that makes use of low-discrepancy sequences and the method of uniformization to reduce variance in the computed density function.

## List of Papers

This thesis is a summary of the following papers. They will be referred to as Paper A, Paper B and Paper C.

A A. Hellander and P. Lötstedt, Hybrid method for the chemical master equation, J. Comput. Phys., 227(1), pp. 100-122, 2007.

B M. Hegland, A. Hellander and P. Lötstedt, Sparse grids and hybrid methods for the chemical master equation, BIT, to appear 2008.

C A. Hellander, Efficient computation of transient solutions of the chemical master equation based on uniformization and quasi–Monte Carlo, J. Chem. Phys.,128, 154109, 2008.

In addition, the following papers are related to the content of the thesis, but not explicitely discussed.

L. Ferm, P. Lötstedt and A. Hellander, A hierarchy of approximations of the master equation scaled by a size parameter, J. Sci. Comp., 34(2), pp. 127–151, 2008.

S. Engblom, L. Ferm, A. Hellander and P. Lötstedt, Simulation of the reaction–diffusion master equation on unstructured meshes, Technical report no. 012, Dept. of Information Technology, Uppsala University, 2008. Available at: http://it.uu.se/research/publications/reports/2008-012/ (Submitted to SIAM J. Sci. Comput.)

## Acknowledgements

I am greatful to my advisor Per Lötstedt for inspiration and for always finding time to quickly give valuable advice. I thank Stefan Engblom, Eddie Wadbro, Emmet Caulfield and the n-dim seminar group for thouroughly reading this manuscript. Input from Markus Hegland also improved the content. I also thank my other colleagues, especially my roommate Sven-Erik Ekström, for many intreresting conversations. Finally, this work would have been much harder without the support from my family and the unceasing love and encouragement from Marie.

This work was supported by the Graduate School in Mathematics and Computing (FMB).

# Contents

**1 Introduction** **1**

**2 Stochastic modeling of chemical reactions in well stirred systems** **2**

**3 Direct solution methods for the CME** **5**

**4 Monte Carlo methods** **6**
4.1 Exact stochastic simulation . . . . . 6
4.2 Approximate methods . . . . . 10
4.3 Hybrid and multiscale methods . . . . . 10

**5 Summary of papers** **11**
5.1 Paper A . . . . . 11
5.2 Paper B . . . . . 12
5.3 Paper C . . . . . 12

**6 Beyond homogeneous models – spatiotemporal dynamics** **13**

# 1 Introduction

The copy number of the macromolecules participating in typical models of e.g. gene regulation is often small. Ribonucleic acid (RNA), the template from which proteins are synthesized, can as an example be present in the range of zero to a few molecules. Transcription factors, proteins directly involved in gene regulation by binding to the DNA might in a typical model vary from ten to a few hundred molecules. This is easily understood, considering the small volume in which the networks operate. The cell volume of *E. Coli* is in the order of $10^{-15}$ l, where a concentration of $1nM$ (nanomolar) roughly corresponds to 1 molecule. In this setting, the much used *reaction rate equations* (RREs), being equations for the time evolution of the concentrations of the chemical species, can not be trusted to capture the dynamics accurately. Treating the participating species as discrete entities offer a better picture of the system state, as we are able to model stochastic effects and study higher moments of the distribution of the copy numbers of the species as well as the correlation between different proteins. Indeed, in much recent work it has been suggested that a stochastic modeling is required in order to explain system behavior, see e.g. [4, 23, 60, 65, 81].

The most frequently occurring mathematical model for chemical reactions on the mesoscale is a continuous–time discrete–space Markov process (CTMC). The state is the copy number of the different chemical species, and the governing equation for the probability density function (PDF) is the chemical master equation (CME). The CME is a consequence of the Markov property, and has a simple interpretation as a balance equation for the probability density. Each species in the model adds one dimension to the problem, and the computational work to obtain a direct numerical solution grows exponentially with the dimension. As a consequence, computation of the PDF by the solution of the CME by a naïve discretization is infeasible for all but low–dimensional problems.

A different approach is to use Monte Carlo (MC) methods to study the time evolution of the jump process. This approach has been used for a long time in many different application areas, and the method goes under different names in different fields. The kinetic MC method was first introduced in [1, 85], and in the case of well stirred chemical reactions Gillespie was able to put the method on a theoretical foundation, introducing the much used Stochastic Simulation Algorithm (SSA) [34]. By replication, statistical quantities such as expected values and correlation coefficients can be computed. In one sense, Monte Carlo simulation is insensitive to the dimensionality of the problem, since the work grows linearly with the number of reaction channels in the model. However, the method suffers from a poor convergence rate, and computation of the PDF can be very time consuming due to the vast number of trajectories that need to be generated. Also, the reaction networks are often stiff due to the presence of different time

scales. This can make the simulation of many interesting models practically impossible, even if the number of different species is modest.

The insufficiency of the macroscopic equations, in combination with the need for fast simulation of increasingly complex biochemical models, has thus created a need for new numerical methods and this has become one main focus area in the field of computational systems biology. Much progress has been made both in the development of direct methods for the CME and in improvements of the Monte Carlo method. The purpose of this thesis is to provide a thourough review of the work that has been done in this area as well as to describe our contributions. In particular, one approach has been to develop hybrid methods, linking the deterministic, macroscale regime to the mesoscale, and in this way reducing the complexity and the stiffness of the problem. We have developed a new such method method which is discussed in Papers A and B. In paper C, we take another approach, reducing the variance in the estimator of the PDF and thus the number of trajectories needed in order to compute an accurate solution.

The remainder of this thesis is organized as follows. In Sect. 2 the mesoscale model is introduced and the CME stated. Here, we also briefly discuss other models of chemical reactions and their relation to the jump process. Direct solution methods are reviewed in Sect. 3 and in Sect. 4.1 and Sect. 4.2 we discuss exact and approximate stochastic simulation. Sect. 4.3 reviews hybrid and multiscale methods and the papers on which this thesis is based are summarized and put into their context in Sect. 5. Finally, we mention a few important directions of future work and look beyond the setting of well stirred systems in Sect. 6.

## 2 Stochastic modeling of chemical reactions in well stirred systems

In this section we consider a system of $N$ chemical species whose copy numbers are the random variables $X_i, i = 1, \dots N$. The state vector is $X = \{X_1, \dots, X_N\}$ and a realization of this variable is denoted $\mathbf{x} = \{x_1, \dots, x_N\}$, which takes values in the integer lattice $\mathbb{Z}_+^N$. The stochastic process $\{X_t\}_{t\geq 0}$ is a collection of such stochastic variables, and a realization of this process will often be referred to as a trajectory or sample path. The conditional probability $Pr(X(t) = \mathbf{x}|X(t_0) = \mathbf{x}_0)$ is denoted $p(\mathbf{x}, t|\mathbf{x}_0, t_0)$.

The chemical species participate in $M$ chemical reactions $R_1, \dots, R_M$, that change the state in discrete jumps. We will denote the state change vector $\mathbf{n}_r$, and it defines how the chemical reaction $r$ changes the state. A reaction is in this notation written as $\mathbf{x} \to \mathbf{x} + \mathbf{n}_r$. A simple example is the association of two proteins $X_1$ and $X_2$ forming a third species $X_3$, $X_1 + X_2 \to X_3$, where $\mathbf{n}_r = [-1, -1, 1]$. To each individual reaction is associated a propensity function, $\omega_r(\mathbf{x})$, which is such that $\omega_i(\mathbf{x})dt$ is interpreted

as the probability that one reaction $R_i$ occurs in the infinitesimal time interval $[t, t + dt)$. For the reaction above this propensity can e.g. take the form $\omega_r(\mathbf{x}) = k_a x_1 x_2$, where $k_a$ is a mesoscopic rate constant. Other functional forms are used when we consider for example enzymatic reactions. The parameters in the models are in some cases known from experiments. The problem of inferring these parameters from given data is a challenging problem, both from a computational and an experimental point of view. However, the methods we discuss in this paper are all concerned with the forward problem, and rates are considered given. s

We regard the molecules as homogeneously distributed in a system volume $\Omega$. This is reasonable when we can assume that the molecules undergo many non–reactive collisions between each occurrence of a reaction. In this setting, the model we consider for the chemical system is a continuous–time discrete–space Markov process. Roughly speaking, given that the process is in state $\mathbf{x}$ at time $t$, the probability to be in a state $\mathbf{x}'$ at a later time is independent of the state at any time $s < t$. The master equation is an equivalent reformulation of the well known Chapman–Kolmogorov equation, but easier to use in practical applications. It can be compactly written as

$$\frac{\partial p(\mathbf{x}, t|\mathbf{x}_0, t_0)}{\partial t} = \sum_{r=1}^{M} [\omega_r(\mathbf{x} - \mathbf{n}_r)p(\mathbf{x} - \mathbf{n}_r, t|\mathbf{x}_0, t_0) - \omega_r(\mathbf{x})p(\mathbf{x}, t|\mathbf{x}_0, t_0)] \tag{1}$$

For a derivation of the master equation and basic properties, see e.g. [83, Chap. V] and [36]. An intuitive interpretation of the master equations is as a balance equation for the density. The first term is the flux of probability into state $\mathbf{x}$ by (feasible) reactions and the second term is the outflow. An exact consequence for the first moment $E[X(t)] = \sum_{\mathbf{x}\in\mathbb{Z}_+^N} \mathbf{x}p(\mathbf{x}, t)$ is the set of equations

$$\frac{d}{dt}E[X_i] = \sum_{r=1}^{M} \mathbf{n}_{r,i}E[\omega_r(X)], \quad i = 1, \ldots, N. \tag{2}$$

For linear propensity functions $\omega_r(\mathbf{x})$ this gives

$$\frac{d}{dt}E[X_i] = \sum_{r=1}^{M} \mathbf{n}_{r,i}\omega_r(E[X]), \quad i = 1, \ldots, N. \tag{3}$$

Introducing the concentration of species $X_i$, $\langle X_i\rangle = E[X_i]/\Omega$ and using (3) we recognize the much used macroscopic reaction rate equations

$$\frac{d}{dt}\langle X_i\rangle = \sum_{r=1}^{M} \mathbf{n}_{r,i}\tilde{\omega_r}(\langle X\rangle), \quad i = 1, \ldots, N, \tag{4}$$

where $\tilde{\omega_r}$ is an appropriate scaling of the mesoscopic rate function $\omega_r$. This set of equations, however, is not an exact consequence of (1) or (2), but provides a good approximation when the copy numbers are large.

The master operator is linear, and the CME can be written compactly

$$\frac{\partial p(\mathbf{x}, t|x_0, t_0)}{\partial t} = \mathcal{M}p(\mathbf{x}, t|x_0, t_0). \tag{5}$$

Though conceptually simple, the work required to solve the CME directly in the straightforward manner is prohibitive for higher dimensions. Even so, the complete picture obtained from the knowledge of the full PDF have motivated many attempts to reduce the computational work and these methods are reviewed in Sect. 3.

A systematic way of approximating the CME is van Kampen's $\Omega$–expansion [83, Chap. X], where $\Omega$ is often interpreted as the system volume. To lowest order, we obtain the macroscopic equation (3). The second order expansion is often termed the *linear–noise approximation*, and it has been used to compute approximate solutions of the CME in a systems biology context by Elf and Ehrenberg in [20]. Engblom suggested a method based on a systematic approximation of the moment equation (2) in [24], and this approach, the RREs and the linear noise approximation are compared in several examples by Ferm et al. [30]. Another approximation of the general master equation that has been used to study biochemical networks is the Fokker–Planck (FP), or forward Kolmogorov, equation. It is a master equation for Itô–diffusions (for an introduction, see e.g. [64, Chap. 7]) and derived by Planck as a second order approximation of the master equation for a general Markov process. For a more complete discussion see e.g. [83, Chap. VIII]. The numerical solution of the FP equation is studied in the thesis by Sjöberg [77].

In between the discrete jump Markov process and the continuous macroscopic description, chemical reaction networks are frequently modeled as a continuous stochastic process. The chemical Langevin equation is a stochastic differential equation (SDE)

$$dX_t^{(i)} = \sum_{r=1}^{M} \mathbf{n}_{r,i}\omega_r(X_t)dt + \sum_{r=1}^{M} \mathbf{n}_{r,i}\sqrt{\omega_r(X_t)}dB_t, \quad i = 1, \dots N, \tag{6}$$

where $B_t$ is standard Brownian motion, and its master equation is the FP equation. Kurtz has shown that in the thermodynamic limit $\Omega \to \infty$ the difference between the jump process and the process $X_t$ in (6) is proportional to $\log(\Omega)$ for polynomial $\omega_r$ [32, pp. 254–255]. This means that for finite times the difference between the corresponding concentration processes $(X_t/\Omega)$ dissapears asymptotically as $\log(\Omega)/\Omega$, giving some validity to (6) and thus to the FPE. The relevance of arguments based on the thermodynamic limit can be questioned for the study of biochemical systems

in a finite volume. Gillespie instead advocates the validity of eq. (6) in the presence of a "domain of macroscopically infinitesimal time intervals" [37].

The methods considered in the following all build on the formalism described above. Many other mathematical frameworks have been used to study genetic networks. For a recent review, see [16].

# 3 Direct solution methods for the CME

The direct, naïve solution of the CME is conceptually simple. The master operator is linear and time–independent, and given a truncation of the state space (which is a subset of the lattice of non-negative integers) the PDF at time $t$ is given by $p(\mathbf{x}, t) = e^{At}p(\mathbf{x}, 0)$. In a few dimensions and with a small state space, this is a feasible approach. However, it is unlikely that these systems will be of much interest to the modeler. The possibility of obtaining an accurate solution and avoiding the $\mathcal{O}(N^{-1/2})$ convergence of Monte Carlo methods has led to the development of methods that reduce the state space in different ways. Using the Fokker-Planck approximation, and thus going from a discrete to a continuous state space, Sjöberg et al. use the finite volume method to obtain solutions at lower cost than for stochastic simulation methods [22, 79]. Further gains are made using space–time adaptivity by Ferm et al. [31] and problems in up to five dimensions are solved efficiently by [77]. The FPE approximation may not be sufficiently accurate for some of the species in the model if they are present in very few copies, and Sjöberg devises a hybrid CME-FPE method [78] to deal with this.

Many methods based on state space aggregation have appeared. Ferm and Lötstedt propose a method that aggregates states and derive a master equation for the average in the larger cells, thus reducing the state space for the CME. They further use adaptivity to efficiently obtain solutions in low dimensions [29]. A discrete spectral method is developed by Engblom [25] where he obtains adaptivity implicitly in the choice of basis. A similar method is proposed by Deuflhard et al. in [17]. The sparse grids method has been adapted to solve the CME by Hegland et al. and they solve a problem in 10 dimensions in [42]. They further rely on the efficient computation of the action of the matrix exponential on vectors in order to evolve the PDF in time [75]. Huber used sparse grids to solve a problem in 100 dimension for a special type of cascade in [46].

Munsky and Khammash propose to use the finite state projection method (FSP) [62], where states that are more unlikely to be reached in a finite time are grouped together, thus reducing the state space considerably. Peles et al. combine the FSP method with approximations based on singular perturbation theory to further improve on the method by taking the time scale separation into account [66]. These methods use an explicit computation of

the matrix exponential, while Burrage et al. uses Krylov methods [6, 75]. This approach are improved further by MacNamara et al. [57] and MacNamara gives a thourough overview of the Krylov FSP method in his thesis [58]. A different approach is taken by Zhang et al. [86] where they propose to use external uniformization to solve the CME efficiently for models with few species. Sidje et al. also use this in combination with inexact matrix–vector products in [76].

# 4 Monte Carlo methods

Even though much progress has been made in the area of direct solution methods the most common way to study systems in higher dimensions is to use Monte Carlo methods. In Sect. 4.1 we review the Gillspie algorithm and recent work to improve the efficiency of exact MC simulation. Approximate simulation methods are discussed in Sect. 4.2 and hybrid and multi–scale methods are reviewed in Sect. 4.3.

An important aspect of the methods described is their implementation. The Systems Biology Markup Language (SMBL), a standard format for sharing models, has been developed by Hucka et al. [47]. Many different software packages supporting this standard have been developed and are publically available. A recent review covering approximate methods, hybrid methods and some available software is provided by Li et al. in [51].

## 4.1 Exact stochastic simulation

The Markov property implies that the time between events is exponentially distributed with parameter equal to the sum of the rates of all individual events. Discrete event simulation of a Markovian system basically amounts to sampling a time for the next event from this distribution and selecting the reaction that occurs at that time. These are the basic steps of the widely used kinetic Monte Carlo (KMC) method. Young and Elcock [85] and independently Lebowitz, Bortz and Kalos [1] introduced the method, known as the KBL algorithm.

Gillespie made kinetic Monte Carlo popular in the chemical and biochemical communities, naming the algorithm the Stochastic Simulation Algorithm (SSA) in his seminal papers [34, 35]. In Gillespie's original paper, he describes two equivalent formulations of the method, the first reaction method (FRM) and the direct method (DM). The DM is more efficient than the FRM, and is outlined in Algorithm 1. We denote the propensity of an event $\omega_r(\mathbf{x})$, where $r$ is an index of the particular event, $r = 1, \dots R$ and $\mathbf{x}$ is the current state of the system. Figure 4.1 shows results from a simulation of a model of the heat shock response (HSR) in *E.Coli*, taken from [12]. The model consist of 28 chemical species participating in 61 reactions. To the

**Algorithm 1** The direct method (DM)/KBL algorithm

Initialize: Compute all propensities $\omega_r(\mathbf{x}), r = 1, \ldots, M$, $t = 0$.
**while** $t < T$ **do**
  Compute the sum of the propensities, $\lambda = \sum_{r=1}^{M} \mu_r(\mathbf{x}_j)$.
  Draw to random numbers $u_1$ and $u_2$ from the uniform distribution $U(0, 1)$.
  Sample the next reaction time (by inversion), $\tau = -\frac{\log(u_1)}{\lambda}$.
  Sample the next reaction event (by inversion), i.e find $r$ such that $\sum_{i=1}^{r-1} \omega_i(\mathbf{x}) < \lambda u_2 < \sum_{i=1}^{r} \omega_i(\mathbf{x})$
  Update the state vector, $\mathbf{x} = \mathbf{x} - \mathbf{n_r}$.
  $t = t + \tau$
**end while**

left we see the path of a single trajectory and to the right the estimated marginal PDF of one of the species based on $10^3$ trajectories.

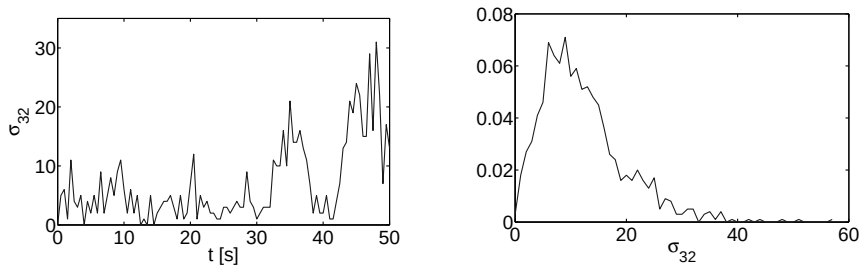



Fig. 1: The hsr model simulated to a final time 50s. The figure shows one trajectory of the species $\sigma_{32}$ (left) and its approximated density (right). As can be seen, the resolution of the PDF is poor, and much more trajectories would be needed to compute a good approximation using the MC method. This however, is time consuming due to the stiffness of the model, see the text and Fig 2 for more details.

The Gillespie algorithm is of fundamental importance in computational systems biology, and efforts have been made to improve its efficiency. In the DM, the selection of the next reaction channel to execute is made by a linear search, giving a complexity $\mathcal{O}(R)$. More efficient formulations have been obtained by improving on this step of the algorithm. The first of these was made in the application area of epitaxial crystal growth. Maksym divided the set of reactions into subsets [59], giving a complexity of $\mathcal{O}(R^{1/2})$. Blue et al. extended this approach by subsequent division of subsets [5], where a K-level method results in a search time proportional to $R^{1/K}$. Taking K to the limit, they achieved the complexity $\mathcal{O}(\log R)$ using a binary tree structure. An overview of these methods and some other improvements is presented by Schulze in [73, 74]. More known to the systems biology

community is the Next Reaction Method (NRM) by Gibson and Bruck [33]. In this work they dealt with two additional optimizations. By switching to absolute time, they were able to reduce the number of random numbers needed in each step from two to one. They also introduced the use of a dependency graph, minimizing the number of propensities that need to be recalculated for every timestep. For every reaction channel, the time until that reaction occurs is computed and maintained in an indexed priority queue (efficiently implemented as a binary heap). The selection of the next event is done in constant time, while the update of the propensities is done in logarithmic time. The cost for maintaining the priority queue is relatively high and the method has its main merits for systems with many reaction channels and where relatively few propensities change with each reaction. Such systems arise when diffusion is added to the models and the reaction–diffusion master equation is simulated. For such models, a variant of the NRM called the Next Subvolume Method (NSM) has been developed by Elf et al. [21] and can be thought of as a clever combination of the ideas of NRM and Maksym's method.

Biochemical network models are often stiff in that some reactions may be much faster than the others. This is often due to very large reaction constants or that species that participate in these reactions are present in large copy numbers. We are frequently interested in the species with low copy numbers, for which the effects of fluctuations are most pronounced. The interesting dynamics of these variables are often given on a timescale determined by the slow reactions. However, for a stiff system the SSA will spend a large fraction of the CPU time sampling the fast events, in which the slow species may or may not participate. This is the motivation for the approximate and hybrid methods discussed later, but also for some of the exact methods. Cao, Li and Petzold suggest that an optimization of the reaction ordering, placing the most frequently occurring events first in the reaction list, combined with a dependency graph, outperforms the NRM even for moderately large systems. Their optimized direct method (ODM) [12] relies on time scale separation to keep the average search depth small. In the ODM, the reaction ordering is determined prior to the simulation by a presimulation for a shorter time, or with few trajectories. If the copy numbers of the species change much from the initial values, this optimal order may change drastically during the course of simulation. In the sorting direct method (SDM) by McCollum et al. [61], the optimal ordering is maintained during the simulation by reordering the reaction list. This paper also contains a detailed overview of the difference in the implementation of DM, NRM, ODM and SDM. Li and Petzold illustrate the efficiency of the logarithmic method in [52].

The HSR model is an example of a very stiff system and has been used to illustrate the need for multi scale methods in e.g. [7]. Fig. 2 shows the number of executions of each reaction when the system is simulated with a

single trajectory to the final time $T_f = 500s$. Clearly, a few reactions are much faster than the others, taking up almost all computing time. Table. 4.1 illustrates the difference in performance of the DM, ODM and NRM by comparing the time needed to generate an ensemble of $10^3$ trajectories on a MacBook Core Duo 2.0 GHz laptop with 2GB ram. The ordering of the reactions in the DM run corresponds to Fig. 2.

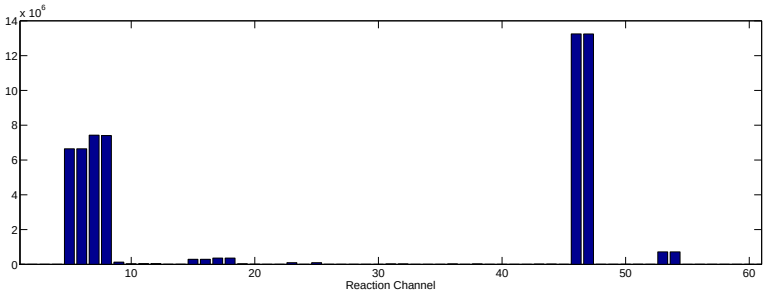



Fig. 2: The number of occurrences of each event when the system was simulated with a single trajectory to $T_f = 500s$. We clearly see the stiffness of the model in that six out of 61 reactions are responsible for almost all events.

| Method | DM | ODM | NRM |
|---|---|---|---|
| CPU time [s] | 3734 | 1978 | 4036 |

Table 1: A comparison of the CPU time needed to simulate an ensemble of $10^3$ trajectories to a final time $T_f = 50s$. The different methods were implemented in the $C$ programming language wrapped in Matlab mex–files, and all use sparse state update. As can be seen, optimizing the reaction ordering makes simulation much faster. Also, the overhead associated with maintaining the heap makes NRM less efficient for this stiff but comparatively small model.

A different approach is taken by Hellander, where variance reduction is obtained by uniformization and quasi–Monte Carlo [43]. There the aim is to reduce the number of trajectories needed to compute an approximation of the PDF at the price of a higher cost per trajectory. Another recent effort to reduce the number of simulations is found in [55].

While parallelization of a single trajectory is hard, it is very beneficial to generate ensembles of trajectories in parallel. Li has implemented SSA on clusters [80] and Li et al. on the graphics processing unit [53, 54]. Field programmable gate arrays have also been used to simulate biochemical reaction networks in [71, 84].

## 4.2 Approximate methods

As seen in the previous section, exact stochastic simulation has been improved substantially since the introduction of the method. Even so, as for the direct methods, there is an inherent limitation in the amount of speed-up obtainable. A different approach to make simulation of larger, stiff systems possible is to introduce some approximation in order to reduce the stiffness or the number of species in the stochastic model. Clearly, such methods could be made much faster than the exact methods. However, in this category of methods the Markov chain is no longer simulated exactly, so the validity of the approximations become a major issue. An important and rapidly evolving approximate MC method is the so called $\tau$–leap method. It was introduced by Gillespie in 2001 [38] and rely on the existence of a macroscopic time where the propensities change slowly, cf. [37]. The basic idea is to let several (fast) reactions fire in the same time interval. The number of events in that interval is approximately Poisson distributed, and can thus be conveniently sampled. The early versions of the algorithm had the problem that negative populations can arise, and several methods have been proposed to deal with this. One way of avoiding negative populations is to abandon the Poisson approximation, as in the binomial leaping method developed independently by Tian et al. [82] and Chatterjee et al. [14] or based on the multinomial distribution by Pettigrew et al. in [67]. Cao et al. instead improve on the Poisson methods in several papers, see e.g. [8, 10, 11]. Anderson takes a different approach, introducing post–leap checks in [2].

## 4.3 Hybrid and multiscale methods

Quite a few methods have emerged, combining the meso and macroscales in order to reduce stiffness. A general strategy here is to mix the deterministic description using reaction–rate like or Langevin equations for subsets of the reactions, the chemical species or both. It has been illustrated in many cases that the speedup obtained using this strategy can be substantial. However, as for the approximate stochastic methods, they rely on the validity of different approximations. Haseltine and Rawlings combine deterministic or Langevin equations for fast reactions with SSA for slow channels [40]. Saliz and Kaznessis devise a hybrid method in which they partition the reactions and evolve the process defined by the fast reactions with the chemical Langevin equation [70]. Puchalka and Kierzek combine the NRM with $\tau$–leaping [68] in their "maximal time step method". Lötstedt and Ferm propose a dimension reduction strategy [56] and Hellander and Lötstedt use this to partition the species based on their variance in [44]. E and Vanden–Eijnden suggest an inner SSA for the fast scale, and an outer for the slow scale in their nested SSA (nSSA) [18, 19]. Erban et al. use the equation-free framework to handle the different time scales [27]. Samant and Vlachos

propose another multiscale MC algorithm in [72].

Many methods based on the quasi–steady state and partial equilibrium assumptions have emerged. Rao and Arkin improve the efficiency of the SSA by using Michaelis–Menten kinetics in their QSSA [69]. A similar approach is taken in [7, 9] where Cao et al. use the stochastic partial–equilibrium assumption of a "virtual fast system". Cao and Petzold further combine this approach with $\tau$–leaping in [13]. Goutsias expands on ideas from [40, 69] in [39].

# 5 Summary of papers

In this section, we give a brief summary of the papers on which this thesis is based. In these three papers, two different new numerical methods are described. Paper A and Paper B deal with a hybrid method based on dimension reduction of the CME [56] and falls in the category of methods in Sect. 4.3. Paper C on the other hand, accounts for a new exact Monte Carlo method where we achieve variance reduction and an improved convergence rate by the use of low–discrepancy sequences and the method of uniformization combined with SSA.

## 5.1 Paper A

In this paper we devise a new hybrid method based on the separation of the set of reacting species $\mathbf{X} = \{X_1, X_2, ..., X_S\}$ into one subset with smaller relative variance, $\mathbf{Y} = \{Y_i\}_{i=1}^n$ and one whose components need stochastic treatment, $\mathbf{X}' = \{X_i'\}_{i=1}^m$, $m+n = S$. The continuous species in the set $\mathbf{Y}$ are modeled as normally distributed random variables with a small variance $\sigma^2$, pairwise independent and independent of $\mathbf{X}'$. The discrete species in $\mathbf{X}'$ are simulated using SSA to approximate the marginal probability distribution, and the expected values $\mathbf{y}$ of the species in the set $\mathbf{Y}$ are given by the system of ordinary differential equations

$$\frac{d\mathbf{y}}{dt} = \sum_{r=1}^{R} \sum_{x \in \mathbb{Z}_+^m} \mathbf{n}_{r,i} \omega(\mathbf{x}, \mathbf{y}) p_{\mathbf{X}'}(\mathbf{x}, t), i = 1, \ldots, n, \tag{7}$$

where $\omega(\mathbf{x}, \mathbf{y})$ are propensity functions and $R$ is the number of different chemical reactions.

If the sum in (7) is evaluated directly, the exponential growth is not avoided. Instead, it is approximated using a quasi–Monte Carlo method. This obviously introduces another source of error, but makes a fast solution of the ODEs possible. The improvement of simulation time comes from two sources: dimension reduction, giving a smaller state space and a smaller number of reaction events in the SSA simulations, and reduction of stiffness when fast reactions involving only species in $\mathbf{Y}$ are handled by the

deterministic equations (7). The resulting ODE system is likely to be stiff. This however, is efficiently handled using an implicit, adaptive time stepping scheme. The time step controller takes both the error in the PDF and the error in the computation of the right hand side sum into account.

In this setting, we report improvements of execution time for a reduced model of a signalling cascade by a factor of 10 compared to SSA for the completely stochastic system.

### 5.2 Paper B

The purpose with this paper is to further illustrate the potential of the hybrid method from Paper A, as well as to compare and contrast the method to a direct method. We apply the hybrid method described in Paper A on a very stiff model of a mitogen activated protein kinase (MAPK) cascade with 22 species [45]. The summation of the right hand side is done differently, since we do not aim to compute the density at each time step, and we report a speedup compared to the ODM by more than two orders of magnitude when we simulate the system with a few stochastic species. The paper is coauthored with Markus Hegland, who solves the CME for a similar model in fewer dimensions with the sparse grids method.

### 5.3 Paper C

In Papers A and B, the purpose of the dimension reduction is to reduce the average time per trajectory, and thus compute the solution of the CME faster. In Paper C, on the other hand, the aim is to reduce the variance in the estimate of the PDF. This is achieved by applying a quasi–Monte Carlo method for the simulation of discrete–time Markov Chains [50]. In order to apply the method, the CTMC is transformed into a DTMC subordinate to a Poisson process using the method of uniformization or Jensen's method [48]. The probability that the system is in state $\mathbf{x}_i$ at time $T$ is written

$$p(\mathbf{x}_i, T) = \sum_{k=0}^{\infty} Pr(N_T = k) p_k(\mathbf{x}_i, k). \tag{8}$$

where $N_T \in Po(\lambda_{max} T)$ is a Poisson random variable and $p_k(\mathbf{x}_i, k)$ is the probability of the discrete time chain resulting from the uniformization being in state $\mathbf{x}_i$ at step $k$. $p_k(\mathbf{x}, k)$ is approximated from realizations generated with SSA modified to incorporate the QMC method. The variance reduction comes from the introduction of correlation between trajectories, while each individual trajectory is still a correct realization of the process. The only approximation made is when the sum in (8) is truncated. This error can be made arbitrary small and the method thus falls in the category of exact MC methods in Sect. 4.1.

In this way, we reduce the number of replications needed to arrive at an accurate approximation of the PDF. It is demonstrated in one example with four species and one with eight species that this can be achieved if the systems are non–stiff and of moderate dimensionality. Even if this is useful in its own right, we anticipate that the method's main merits will be in the combination with other hybrid and multi scale that reduce stiffness. In particular, the modified systems obtained by the partitioning of species in Paper A will be both less stiff and of lower dimensionality and they will thus be well suited for simulation by this method.

# 6 Beyond homogeneous models – spatiotemporal dynamics

In this section we introduce some of our most recent work and outline some directions of future work. As has been outlined in this thesis, much effort has been invested into the development of more efficient numerical methods for the study of biochemical networks. Undoubtedly, much work remains to be done, as models become larger and more complex. However, there are many processes in the living cell that can not be explained in a well–stirred context. Examples include signaling pathways where the spatial organization plays an important role for the function of the system, e.g. MAPK kinase cascades, where translocation of ERK to the nucleus leads to the activation of many transcription factors. The macroscopic equation in most use is here the (non–linear) reaction–diffusion equation

$$\frac{\partial c}{\partial t} = \Delta c + R(c). \tag{9}$$

Being based on the same approximation as the RREs, it suffers from the same limitation: when the number of molecules become small, the approximation is not valid. It is possible to formulate a master equation that includes diffusion, the reaction-diffusion master equation (RDME), where the geometry is discretized with small computational subvolumes. The diffusion is then treated as jumps from a cell to adjacent cells. This approach has been used by e.g. Elf et al. to study the effect of stochasticity on a few different systems [21, 28]. They have also developed the software mesoRD [41], based on the next subvolume method (NSM) [21].

While the homogeneous problem often presents considerable computational difficulties, spatiotemporal simulation is even worse. The dimensionality of the problem becomes very high, even for a very small number of computational cells. Solving the RDME for the density by direct methods is not likely to be possible for realistic problems and even an approximation by replication of MC simulations is probably going to be too time consuming for most problems. Instead, insight into the dynamic behavior of the system

must be gained by studying a small number of trajectories. The simulation will still have the same issue of stiffness due to different time scales of the reactions, but will in addition require the generation of very large number of diffusion events. The diffusion of the different species may in turn vary in rate due to e.g. different molecular size or localization in the cell. An important challenge for the future in the field of computational systems biology is therefore to develop more efficient methods for simulation of spatially dependent models. A step in this direction has been taken by Engblom et al. [26], where NSM is extended to simulations on an unstructured mesh by making connections to the finite element method. In addition to making complex geometries easy to handle, we also suggest a hybrid method to speed up simulation when the number of some of the species is large by treating diffusion of those species macroscopically. We are currently working on software that implements this method using COMSOL Multiphysics 3.4 [15] to handle meshes and pre and postprocessing, and we hope to make it available to the public in the near future. Fig. 6 shows a stochastic simulation of the MinD–MinE system from [28] simulated with our code. The results are in good agreement with the simulations on a structured mesh with mesoRD in [28].

Even though the diffusion makes the time scale separation less apparent than in the homogeneous case, we anticipate that some of the ideas from the approximate and hybrid methods reviewed in the previous sections can be adapted to this setting. In particular, we have made a first attempt to apply tau–leaping to a simple model problem, and are currently continuing to develop the hybrid method proposed in [26].

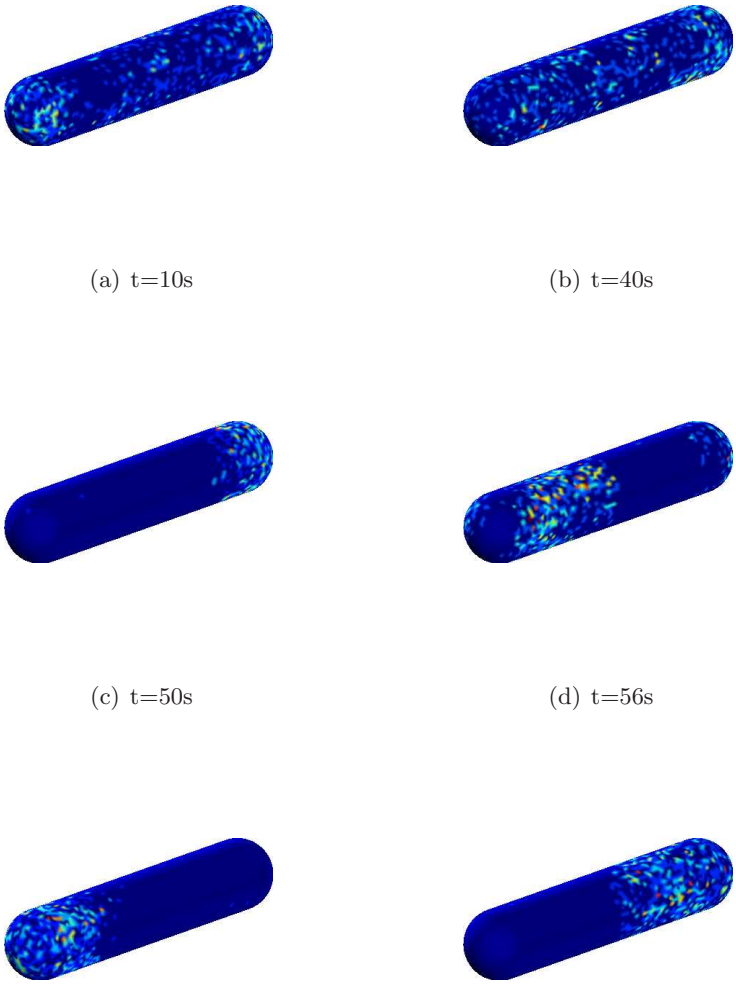
We also believe that it is important to consider how more complicated mechanisms of cellular transport should be modeled and simulated. As an example, it has been suggested that simple diffusive transport may not be a sufficient means of signal propagation in MAPK pathways since slow diffusion and high phospatase activity may attenuate the signal too much. It is suggested by Kholodenko that microtubuli dependent endocytic trafficking may be important in these type of networks [49]. If it is important to consider these mechanisms in a stochastic spatiotemporal model, and how to model it in a computationally tractable manner is an open question. The method developed by us in [26] can be extended to include convection terms, but this may be an oversimplification. Explicit treatment of microtubuli would be prohibitively expensive since they are too small to be resolved by a discretization, but it may be possible to devise efficient multiscale methods that take active transport into account and we believe that it is worthwhile to investigate this further.

The environment in a cell is crowded and the size of most of the chemical species in typical models is relatively large. It is therefore natural to question the validity of normal diffusion, especially when we consider diffusion in membranes [3, 63]. Many models of complex phenomena inside the cell will

inevitably involve 2-D like diffusion on membrane surfaces and 3-D diffusion in the cytosol or inside organelles. Maybe we need to allow for different diffusion models in different parts of the cell, and how to introduce this into the mesoscopic model in a correct and feasible way is also an interesting direction of future work.

Apart from this last section, this thesis has been focused on numerical methods for solving the chemical master equation in high dimension or, when the systems consist of too many species to yet permit a complete solution, more efficient simulation methods to study realizations of the process. While the dimensionality of the problem is a severe limiting factor in the homogeneous models, it will not directly be the increase of chemical species in spatiotemporal models that will be limiting when we turn to more complicated models of the cell. Instead, it will be the inevitable multiscale and multiphysics nature of any detailed description of intracellular phenomena that will pose the largest challenge.

Future methods will need to handle biochemical reactions and different types of transport mechanisms happening on different time scales, and in the same time combine different descriptions of processes, going from a stochastic microscopic description of some, a mesoscopic model of others all the way up to the deterministic macroscopic models when appropriate. Also, different descriptions of the same process may be appropriate in different regions in space and time, so an efficient method should ideally adapt both the model and the necessary resolution to the local requirements. This will be a fundamental and exciting challenge for the methods development branch of systems biology in the future.



(a) t=10s (b) t=40s

(c) t=50s (d) t=56s

(e) t=65s (f) t=70s

Fig. 3: Simulation of a model of MinD oscillations in *E. Coli* on an unstructured mesh with the method developed in [26]. At $t = 0s$, no MinD is present on the membrane. After an initial period when the number of molecules are increasing on the whole membrane in (a)–(b), the membrane bound MinD oscillates from pole to pole in (c)–(f).

## References


[1] J. L. A.B. Bortz, M.H. Kalos. A new algorithm for Monte Carlo simulation of ising spin systems. *J. Comput. Phys.*, 17(1):10–18, January 1975.

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

[3] D. S. Banks and C. Fradin. Anomalous diffusion of proteins due to molecular crowding. *Biophys. J.*, 89(5):2960–2971, 2005.

[4] N. Barkai and S. Leibler. Circadian clocks limited my noise. *Nature*, 403:267–268, 2000.

[5] J. L. Blue, I. Beichl, and F. Sullivan. Faster Monte Carlo simulations. *Phys. Rev. E*, 51(2):R867–R868, Feb 1995.

[6] K. Burrage, M. Hegland, S. MacNamara, and R. B. Sidje. A krylov–based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems. In A. Langville and W. Stewardt, editors, *150th Markov Anniversary Meeting*, Boson Books, pages 21–38. Raleigh, NC, 2006.

[7] Y. Cao, D. Gillespie, and L. Petzold. Multiscale stochastic simulation algorithm with partial equilibrium assumption for chemically reacting systems. *J. Comput. Phys.*, 206:395–411, 2005.

[8] Y. Cao, D. T. Gillespie, and L. R. Petzold. Avoiding negative populations in explicit poisson tau-leaping. *J. Chem. Phys.*, 123(5):054104, 2005.

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

[10] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124(4):044109, 2006.

[11] Y. Cao, D. T. Gillespie, and L. R. Petzold. Adaptive explicit-implicit tau-leaping method with automatic tau selection. *J. Chem. Phys.*, 126(22):224101, 2007.

[12] Y. Cao, H. Li, and L. Petzold. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys*, 121(9):4059–4067, 2004.

[13] Y. Cao and L. Petzold. Slow scale tau-leaping method. *Comput. Meth. Appl. Mech. Engin.*, 2008 (to appear).

[14] A. Chatterjee, D. G. Vlachos, and M. A. Katsoulakis. Binomial distribution based tau-leap accelerated stochastic simulation. *J. Chem. Phys.*, 122(2):024112, 2005.

[15] COMSOL, Inc. http://www.comsol.com.

[16] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.*, 9(1):67–103, 2002.

[17] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow. Adaptive discrete Galerkin methods applied to the chemical master equation. Technical Report 04, Konrad–Zuse–Zentrum für Informationstechnik Berlin, 2007.

[18] W. E, D. Liu, and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J. Chem. Phys.*, 123, 194107, 2005.

[19] W. E and E. Vanden-Eijnden. Nested stochastic simulation algorithms for chemical kinetic systems with multiple time scales. *J. Comput. Phys.*, 221(1):158–180, 2007.

[20] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Res.*, 13(11):2475–2484, 2003.

[21] J. Elf and M. Ehrenberg. Spontaneous separation of bi–stable biochemical systems into spatial domains of opposite phases. *Syst. Biol.*, 1(2), 2004.

[22] J. Elf, P. Lötstedt, and P. Sjöberg. Problems of high dimension in molecular biology. In *Proc. 19th GAMM Seminar Leipzig*, pages 21–30, 2003.

[23] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[24] S. Engblom. Computing the moments of high dimensional solutions of the master equation. *Appl. Math. Comput.*, 180(2):498–515, 2006.

[25] S. Engblom. A discrete spectral method for the chemical master equation. Technical Report 005, Dept. Information Technology, Uppsala University, 2008.

[26] S. Engblom, L. Ferm, A. Hellander, and P. Lötstedt. Simulation of stochastic reaction–diffusion processes on unstructured meshes. Technical Report 012, Dept. Information Technology, Uppsala University, 2008.

[27] R. Erban, I. G. Kevrekidis, D. Adalsteinsson, and T. C. Elston. Gene regulatory networks: A coarse-grained, equation-free approach to multiscale computation. *J. Chem. Phys.*, 124(8):084106, 2006.

[28] D. Fange and J. Elf. Noise-induced Min phenotypes in *E. coli*. *PLOSB*, 2(6):0637–0647, 2006.

[29] L. Ferm and P. Lötstedt. Adaptive solution of the master equation in low dimensions. *Applied Numerical Mathematics*, In Press, Corrected Proof.

[30] L. Ferm, P. Lötstedt, and A. Hellander. A hierarchy of approximations of the master equation scaled by a size parameter. *J. Sci. Comp.*, 34(2):127–151, 2008.

[31] L. Ferm, P. Lötstedt, and P. Sjöberg. Conservative solution of the Fokker–Planck equation for stochastic chemical reactions. *BIT*, 46(0):61–83, 2006.

[32] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer-Verlag, 3rd edition, 2004.

[33] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 104:1876–1889, 2000.

[34] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reacting systems. *J. Comput. Phys.*, 22:403–434, 1976.

[35] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[36] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A*, 188:404–425, Sept. 1992.

[37] D. T. Gillespie. The chemical Langevin equation. *J. Chem. Phys.*, 113(1):297–306, 2000.

[38] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.*, 115(4):1716–1733, 2001.

[39] J. Goutsias. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys.*, 122(18):184102, 2005.

[40] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117(15):6959–6969, 2002.

[41] J. Hattne, D. Fange, and J. Elf. Stochastic reaction–diffusion simulation with MesoRD. *Bioinformatics*, 21(12):2923–2924, 2005.

[42] M. Hegland, C. Burden, L. Santoso, S. MacNamara, and H. Booth. A solver for the stochastic master equation applied to gene regulatory networks. *J. Comput. Appl. Math.*, 205(2):708–724, 2007.

[43] A. Hellander. Efficient computation of transient solutions of the chemical master equation based on uniformization and quasi-Monte Carlo. *J. Chem. Phys.*, 128(15):154109, 2008.

[44] A. Hellander and P. Lötstedt. Hybrid method for the chemical master equation. *J. Comput. Phys.*, 227(1):127–151, 2008.

[45] C.-Y. F. Huang and J. James E Ferrel. Ultrasensitivity in the mitogen–activated protein kinase cascade. *Proc. Nat. Acad. Sci. USA*, 93:10078–10083, 1996.

[46] L. Huber. Mathematical modeling, numerical approximation and computational exploration of gene regulatory processes in the lambda phage. Master's thesis, Swiss Federal Institute of Technology, Institute of Computational Science, Zurich, 2006.

[47] M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, , the rest of the SBML Forum:, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J.-H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner, and J. Wang. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4):524–531, 2003.

[48] A. Jensen. Markoff chains as an aid in the study of Markoff processes. *Skand. Aktuarietidskr.*, 36:87–91, 1953.

[49] B. N. Kholodenko. MAP kinase cascade signaling and endocytic trafficking: a marriage of convenience? *Trends Cell Biol.*, 12(4):173–177, 2002.

[50] C. Lécot and B. Tuffin. Quasi–Monte Carlo methods for estimating transient measures of discrete time Markov chains. In H. Neiderriter, editor, *Monte Carlo and Quasi Monte Carlo Methods 2002*. Springer-Verlag, 2002.

[51] H. Li, Y. Cao, L. Petzold, and D. Gillespie. Algorithms and software for stochastic simulation of biochemical reacting systems. *Biotechnology Progress*, 24(1):56–61, 2008.

[52] H. Li and L. Petzold. Logarithmic direct method for discrete stochastic simulation of chemically reacting systems. Technical report, Department of Computer Science, University of California, Santa Barbara, http://www.engineering.ucsb.edu/∼cse/Files/ldm0513.pdf, 2006.

[53] H. Li and L. Petzold. Efficient parallellization of stochastic simulation algorithm for chemically reacting systems on the graphics processing unit. Technical report, Dept. Computer Science, University of California, Santa Barbara, 2008.

[54] H. Li and L. R. Petzold. Stochastic simulation of biochemical systems on the graphics processing unit. Technical report, Dept. of Computer Science, University of California, Santa Barbara, 2008.

[55] A. Lipshtat. An "all possible steps" approach to the accelerated use of Gillespie's algorithm. *J. Chem. Phys.*, 126, 184103, 2007.

[56] P. Lötstedt and L. Ferm. Dimensional reduction of the Fokker–Planck equation for stochastic chemical reactions. *Multiscale Model. Simul.*, 5(2):593–614, 2006.

[57] S. MacNamara, K. Burrage, and R. B. Sidje. Multiscale modeling of chemical kinetics via the master equation. *Multiscale Model. Simul.*, 6(4):1146–1168, 2008.

[58] S. F. MacNamara. *Krylov and Finite State Projection Methods for Simulating Stochastic Biochemical Kinetics via the Chemical Master Equation*. PhD thesis, The University of Queensland, Australia, 2008.

[59] P. A. Maksym. Fast Monte Carlo simulation of MBE growth. *Semiconductor Science and Technology*, 3(6):594–596, 1988.

[60] H. H. McAdams and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, 15(2):65–69, 1999.

[61] J. M. McCollum, G. D. Peterson, C. D. Cox, M. L. Simpson, and N. F. Samtova. The sorting direct method for stochastic simulation of biochemical systems with variying reaction execution behaviour. *Comput. Biol. Chem.*, 30:39–49, 2005.

[62] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4):044104, 2006.

[63] J. Nicolau, Dan V., J. F. Hancock, and K. Burrage. Sources of anomalous diffusion on cell membranes: A monte carlo study. *Biophys. J.*, 92(6):1975–1987, 2007.

[64] B. Øksendahl. *Stochastic Differential Equations.* Springer–Verlag, Berlin Heidelberg, 6th edition, 2005.

[65] J. Paulsson, O. G. Berg, and M. Ehrenberg. Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation. *Proc. Nat. Acad. Sci. USA*, 97(13):7148–7153, 2000.

[66] S. Peles, B. Munsky, and M. Khammash. Reduction and solution of the chemical master equation using time scale separation and finite state projection. *J. Chem. Phys.*, 125(20):204104, 2006.

[67] M. F. Pettigrew and H. Resat. Multinomial tau-leaping method for stochastic kinetic simulations. *J. Chem. Phys.*, 126(8):084101, 2007.

[68] J. Puchalka and A. M. Kierzek. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophys. J.*, 86(3):1357–1372, 2004.

[69] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J. Chem. Phys.*, 118(11):4999–5010, 2003.

[70] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.*, 122, 054103, 2005.

[71] L. Salwinski and D. Eisenberg. *In silico* simulation of biological network dynamics. *Nature Biotechnology*, 22(8):1017–1019, 2004.

[72] A. Samant and D. G. Vlachos. Overcoming stiffness in stochastic simulation stemming from partial equilibrium: A multiscale Monte Carlo algorithm. *J. Chem. Phys.*, 123(14):144114, 2005.

[73] T. P. Schulze. Kinetic Monte Carlo simulations with minimal searching. *Phys. Rev. E*, 65(3):036704, Feb 2002.

[74] T. P. Schulze. Efficient kinetic Monte Carlo simulation. *J. Comput. Phys.*, 227(4):2455–2462, 2008.

[75] R. B. Sidje. Expokit: a software package for computing matrix exponentials. *ACM Trans. Math. Softw.*, 24(1):130–156, 1998.

[76] R. B. Sidje, K. Burrage, and S. MacNamara. Inexact uniformization method for computing transient distributions of markov chains. *SIAM J. Sci. Comput.*, 29(6):2562–2580, 2007.

[77] P. Sjöberg. *Numerical Methods for Stochastic Modeling of Genes and Proteins*. PhD thesis, Uppsala University, 2007.

[78] P. Sjöberg. Partial approximation of the master equation by the Fokker–Planck equation. *Applied Parallel Computing. State of the Art in Scientific Computing, Lecture Notes in Scientific Computing, Springe Verlag*, pages 637–646, 2007.

[79] P. Sjöberg, P. Lötstedt, and J. Elf. Fokker-Planck approximation of the master equation in molecular biology,. *Comput. Vis. Sci. (to appear)*, 2008.

[80] StochKit Team. *User's Guide for StochKit*. http://www.engr.ucsb.edu/∼cse.

[81] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Nat. Acad. Sci. USA*, 98:8614–8619, 2001.

[82] T. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *J. Chem. Phys.*, 121(21):10356–10364, 2004.

[83] N. G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, third edition, 2007.

[84] M. Yoshiini, Y. Iwaoka, Y. Nishikawa, T. Kojima, Y. Osana, Y. Shibata, N. Iwanaga, H. Yamada, H. Kitano, A. Funahashi, N. Hiroi, and H. Amano. FPGA implementation of a data-driven stochastic biochemical simulator with the next reaction method. *Field Programmable Logic and Applications, 2007. FPL 2007. International Conference on*, pages 254–259, Aug. 2007.

[85] W. Young and E. Elcock. Monte Carlo studies of vacancy migration in binary ordered alloys. *Proc. Phys. Soc.*, 89:735–747, 1966.

[86] J. Zhang, L. T. Watson, and Y. Cao. A modified uniformization method for the solution of the chemical master equation. Technical Report TR-07-31, Department of Computer Science, Virginia Tech, http://eprints.cs.vt.edu/archive/00000992/01/IETSB07.pdf, 2008.

**Recent licentiate theses from the Department of Information Technology**

**2008-002** Ioana Rodhe: *Query Authentication and Data Confidentiality in Wireless Sensor Networks*

**2008-001** Mattias Wiggberg: *Unwinding Processes in Computer Science Student Projects*

**2007-006** Björn Halvarsson: *Interaction Analysis and Control of Bioreactors for Nitrogen Removal*

**2007-005** Mahen Jayawardena: *Parallel Algorithms and Implementations for Genetic Analysis of Quantitative Traits*

**2007-004** Olof Rensfelt: *Tools and Methods for Evaluation of Overlay Networks*

**2007-003** Thabotharan Kathiravelu: *Towards Content Distribution in Opportunistic Networks*

**2007-002** Jonas Boustedt: *Students Working with a Large Software System: Experiences and Understandings*

**2007-001** Manivasakan Sabesan: *Querying Mediated Web Services*

**2006-012** Stefan Blomkvist: *User-Centred Design and Agile Development of IT Systems*

**2006-011** Åsa Cajander: *Values and Perspectives Affecting IT Systems Development and Usability Work*

**2006-010** Henrik Johansson: *Performance Characterization and Evaluation of Parallel PDE Solvers*

**2006-009** Eddie Wadbro: *Topology Optimization for Acoustic Wave Propagation Problems*

**2006-008** Agnes Rensfelt: *Nonparametric Identification of Viscoelastic Materials*

**2006-007** Stefan Engblom: *Numerical Methods for the Chemical Master Equation*

UPPSALA
UNIVERSITET

Department of Information Technology, Uppsala University, Sweden