UNIVERSITÄT BIELEFELD

Dissertation

A Hybrid Algorithm for the Simulation of Biochemical Reactions and Diffusion

Dipl. Biochem. Mark Möller

June 2006

Graduate School in Bioinformatics and Genome Research
Technical Faculty

supervised by:
Prof. Dr. Ellen Baake
Dr. Sven Rahmann
Dr. Holger Wagner

Summary

Over the last decades, the amount of data about molecular processes within cells has tremendously increased leading in particular to an increased interest in theoretical investigations of such systems. One basic theoretical approach in this context is to model processes in biological cells as chemical reaction (diffusion-) systems and to study their properties by computer simulations.

One major problem in handling such systems is that they often simultaneously contain substrates with a wide range of possible particle numbers. For example, ribosomes typically exist in small numbers; tRNA-molecules or proteins are represented in intermediate quantities; and some ions, such as potassium or sodium, are typically present in large quantities. However, no conventional algorithm works well for such a wide range of particle numbers: Small particle numbers require stochastic algorithms, whereas intermediate and large particle numbers can only be treated by computationally more efficient, though perhaps less exact modeling.

To address this problem, I developed the COntrollable Approximative STochastic Algorithm (COAST). COAST is a self adjusting algorithm that can be applied to simulate reaction and diffusion systems. It is based on three different levels of modeling: an exact stochastic approach for low particle numbers, an approximative stochastic approach by Gaussian distributions for intermediate numbers, and a description by deterministic kinetics for high particle numbers.

A special characteristic of COAST is that it automatically determines the optimal level of modeling for the reaction channel at each time step. This is done by using criteria, which appropriately depend on one single error control parameter α. One can show that all approximations of COAST lead to errors even smaller than α. Thus, by choosing a suitable value for α, the user can easily find an optimal trade off between accuracy and computational efficiency for an individual simulation system.

It is demonstrated in test simulations that COAST is able to reproduce results of exact stochastic algorithms with small errors. In most cases, the error is much smaller than α. On the other hand, COAST shows a different asymptotic dependence of the runtime on the number of particles N: For n-order reactions, the asymptotic runtime is proportional to N^n for exact algorithms, but proportional to N^{n-1} for COAST. So clearly, COAST provides significant improvements, in particular if N is large and n is small.

Preamble

Diese Arbeit ist meiner Tante Rita und meinem Onkel Fiete gewidmet. Mögen sie beide in Frieden ruhen.

"Ich moechte mich bei allen bedanken, denen ich Dank schulde und auch bei denen, die meinen, dass ich ihnen Dank schulde, auch wenn das jeder Grundlage entbehrt. Ich habe nichts dagegen, allen gegenueber dankbar zu sein, vor allem aber der Schoepfung insgesamt. Ohne sie waere ich moeglicherweise gar nicht hier - oder da." [Nuhr, 2004]

Aber ein paar Personen moechte ich doch hervorheben.
Zunaechst einmal ist da Dr. Holger Wagner. Ohne ihn waere diese Arbeit voellig undenkbar gewesen. Er war die treibende Kraft hinter allem, was ich hier zusammengetragen habe. Darueberhinaus werde ich unsere Gespraeche über Wissenschaft, Politik, Geschichte und Gesellschaft sehr vermissen.

Des weiteren bedanke ich mich bei Prof. Ellen Baake und Dr. Sven Rahmann, die mich adoptiert haben, um diese Arbeit zu einem konstruktiven Abschluss zu bringen.

Ohne die Fuersprache von Dr. Klaus Prank waere ich nicht hier an der Graduate School und daher bin ich auch ihm dankbar verbunden.

Ich danke Leila Taher, Andre Skusa, Dion Whitehead und Bjoern Oleson, die meine nervigen Fragen zur Programmierung in Java bzw. Matlab tapfer haben über sich ergehen lassen.

Julie Stromer hat dankbarerweise geholfen mein "Germish" in eine Form zu bringen, die einem gesunden "Englisch" nahekommt.

Letzendlich gilt mein besonderer Dank dem deutschen Steuerzahler, dem ich meinen Lebensunterhalt verdankt habe.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

COAST	controllable approximate stochastic reaction algorithm
DNA	desoxyribonucleinacid
ESA	exact stochastic method
FCS	fluorescence correlation spectroscopy
FEA	finite element analysis
FPE	Focker-Plack equation
FRAP	flourescence recovery after photobleaching
FRET	fluorescence resonance energy transfer
FRM	first reaction method
GFP	green fluorescent protein
MAPK	mitogen activated protein kinases
MD	molecular dynamics
NLPDE	non linear partial differential equation
NRM	next reaction method
NSM	next subvolume method
ODE	ordinary differential equation
PDE	partial differential equations
PW-DMC	probability-weighted Dynamic Monte Carlo method
RNA	ribonecleinacid
SDE	stochastic differential equation
SSA	stochastic simulation algorithm
tRNA	transfer RNA
VE	volume element
YFP	yellow fluorescent protein

List of Notations

α	error parameter of COAST
γ	friction coefficient
μ	reaction channel
$\sigma_\mu S$	stoichiometric factor of S in reacton channel μ
τ	small time interval
c_μ	stochastic reaction constant of channel μ
D_S	Diffusion constant of Substrate S
E	expectation
$f(i)$	force at i
Fe	iron
Fe_2O_3	iron(III)oxide
k_B	Boltzmann's constant
k_μ	deterministic reaction constant of channel μ
m	mass
N	number of particles
N_S	number of particles of specie S
p	probability
q	probability
$Q_S i$	total transition probability of the substrate S at the lattice-point i
R	resolution parameter of COAST
T	temperature or timepoint
t	time
U	potential
V	variance or volume

1 Introduction

To understand the complexity and dynamics of biological systems, mathematical modeling and computer simulations have become an important area of research in the last decade. Here, I present an algorithm that is especially designed for the simulation of reaction diffusion systems with a wide variety of species present in very different orders of magnitudes of concentration. It is a contribution to the current development of algorithms for systems biology that aims at providing a comprehensive view on chemical processes in general, and cellular processes in particular. In the last years, biological research has focused on the molecular details of the systems under study. Presently, systems biology tries to put these pieces together combining theoretical and experimental approaches.

Computational Modeling Biology was once limited to reductionist approaches, which were very helpful in the past. However, by blinding out connections in those systems, the retained models are incomplete, inaccurate and simply incorrect [Mellman and Misteli, 2003]. To really obtain a comprehensive view on biological systems, we have to be able to process, analyze, and interpret interactions and dynamic events. Computational modeling allows us to explore such events taking account of the components and the pathways established within the cellular systems under investigation.

Modeling has had a long tradition, and remarkable success, in disciplines such as engineering and physics. Physical science, for example, is supported by three pillars: experimental studies, theoretical studies and simulation. Now, the current development in systems biology indicates that simulations will become more and more important in the future of life sciences. The amount of information gained in biological science has developed tremendously over the last years. Biological modeling uses well-established methods such as the finite element method (=finite element analysis, FEA) or numerical techniques to solve ordinary differential equations (ODEs) or partial differential equations (PDEs) describing biological systems [Doyle, 2001]. ODEs and PDEs are commonly used to model biological networks like metabolic or signal transduction networks [Vilar et al., 2003].

The roots of the theory of differential equations go back to the time of the development of infinitesimal calculus by I.Newton (1643-1727) and G.Leibniz (1646-1716) at the end of the 17th century. Since then, famous mathematicians such as J.Bernoulli (1654-1705), who calculated the orbits of the planets, L.Euler (1707-1783), J.L.Lagrange (1736-1813) and C.F.Gauß worked in that field. Differential equations became an important tool in physics (motion, electrical resonant circuits),

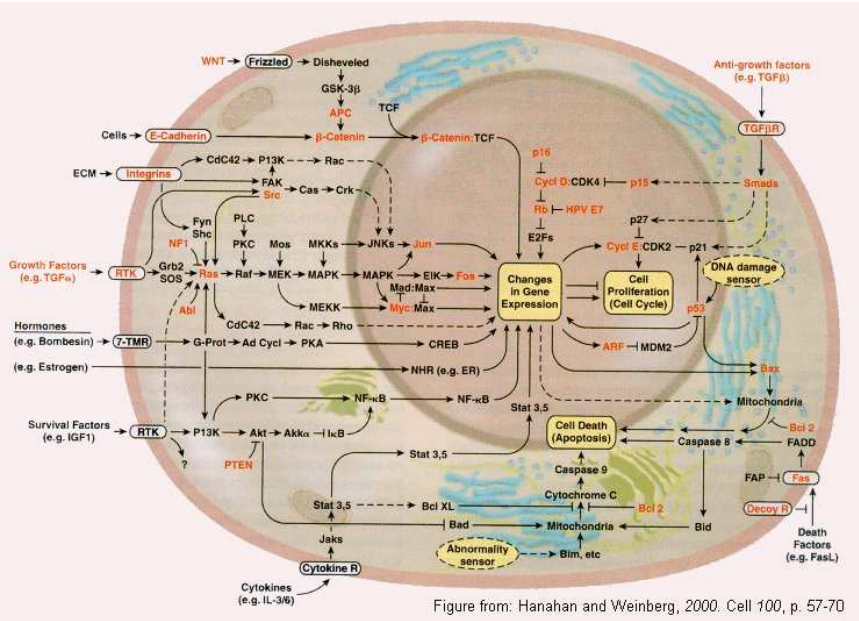

Figure 1.1: *The figure shows the signal transduction network leading to apoptosis presented by Hanahan. It gives an example of how complex cellular processes can be. Reprinted from Cell, Vol.100, Hanahan and Weinberg, The hallmarks of cancer, p57-70, copyright (2000), with permission from Elsevier.*

biology (population dynamics, Lotka-Volterra model of predator-prey relation), chemistry (chemical reactions, carbon dating C14-method) and the financial sector (cycle of growth).

Advantages of Computational Models Some biological concepts have already been discovered by computational modeling. They include bistability [Bhalla and Iyengar, 1999], ultrasensitivity [Ferrell and Machleder, 1998, Ferrell, 1999], and rhythmic behavior [Elowitz, 2000]. Ultrasensitivity defines a response that is more sensitive to ligand concentration as compared to the standard responses as defined by Michaelis-Menten kinetics [Goldbeter and Koshland, 1981]. A classical Michaelis-Menten reaction is described by a hyperbolic reaction velocity curve, while an ultrasensitive reaction is described by a sigmoidal curve. In 1996, Huang and Ferrell analyzed the MAPKinase-signaling-pathway (MAPK: mitogen activated protein kinase) in Xenopus oocytes [Huang and Ferrell, 1996]. They found that one part, the module consisting of the three MAPKinases, worked as a switch, filtering noise and only being activated if the input reached a certain level. This behavior is experimentally observable only if one analyzes a single cell. Another example of computa-

Characteristic	Quantity	Source
proteins	225,000	
ribosomes	15,000	
tRNA-molecules	170,000	
small organic molecules	15,000,000	
ions	25,000,000	
water	25%	[Goodsell, 1993]
no. of genes	4497	[Keseler et al., 2005]
no. of reactions per cell cycle	$10^{14} - 10^{16}$	[Endy and Brent, 2001]

Table 1.1: *The table shows some characteristics of an average E.coli cell. The complexity is about a factor thousand smaller for the smallest cell types (mycoplasms) and about a factor of thousand larger for typical plant and animal cells [Schwehm, 2001].*

tional modeling is the modeling of ion-channels which goes back several years [Levitt, 1999], reaction-diffusion systems simulating transport processes out of the nucleus, or transport processes of proteins between the endoplasmatic reticulum and the Golgi complex [Ladinsky et al., 1999].

These models provide a systematic framework to describe and analyze such complex systems (cf. Table 1.1); this complexity is the result of the number of single nodes within these networks and their interactions. They summarize the current knowledge and hypotheses about missing information. Speaking from a biochemical point of view the nodes are biologically active substrates and the edges are the chemical interactions between them.

Models in general have several advantages. The problem with analyzing complex systems is that the output is far from intuitive; doubling the input does not mean that the output will be doubled [Voit, 2002]. This nonlinearity is caused by synergistic effects, which results in the invalidity of the summation principle of single events. Modeled systems are easier to manipulate than real systems. One clear example for this are multiple knock-out experiments. Another example involves the timescales of biological processes. The time span for molecular movements is within μ-seconds, whereas a human life is approximately 10^9 s (75 years). The former time span is hard to observe and the latter difficult to follow. Simulations allow scientists to capture time and scales together. It should be noted, however, that a model is only an approximation of reality, and all predictions made with those models can only be as good as the model used to make those predictions.

It is not the aim of simulations to replace *in vivo* experiments, but rather to offer important amendments for their planing and analyses. The model can be used to obtain an overview of possible outcomes. Sufficiently detailed and accurate models can serve as a reference for interpreting experimental results and suggesting further

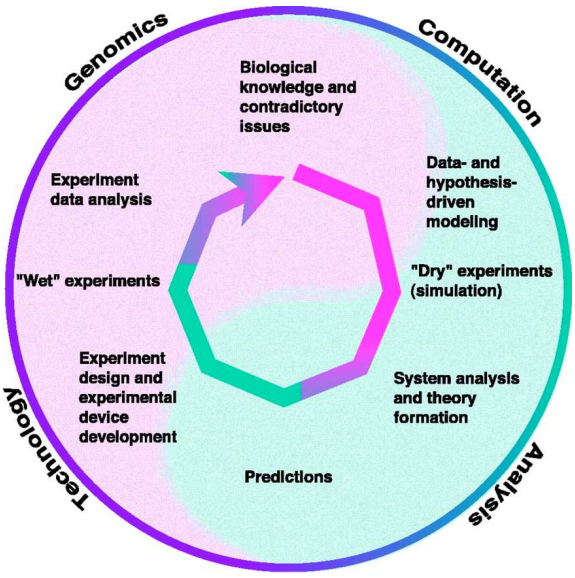

Figure 1.2: *The figure presents the hypothesis-driven research in systems biology demonstrating how mathematical models can be a contribution to research. Reprinted with permission from SCIENCE, Vol.100, Kitano, Systems biology: a brief summary, p1662-1664, March 2002. Copyright (2002) AAAS.*

hypotheses [Takahashi and al., 2002]. Simulations can provide insight into otherwise impossible scenarios and so will be able to save time and money. They are expected to guide wet-lab processes and narrow the experimental search space.

Progress in biochemistry and biology in general has provided science with great detail of cellular processes. Computational biology seeks to understand the principles underlying their dynamic behavior. As Bundschuh et al. [2003] state , there has been considerable effort in the past to model the biochemical network of a whole cell or cellular subsystems. He provides examples for the benefits gained by modeling, namely enhancing our understanding of cell functions, easily observing the designated systems and determining the quantities of interest (measuring them would be only possible by complex experiments). For these reasons, he sees the future of drug development, where the effects of a putative drug on a cell can be immediately tested.

The Problem of Modeling Cellular Processes and Structure The tools needed to establish a working model are provided by mathematics and bioinformatics. For bulk chemical reactions, it is common to find deterministic models resulting from the *mass action law* and formulated by differential equations.
However back in 1930, John Burdon Haldane, the co-founder of population genetics,

expressed that certain critical processes in the cell may be carried out by only a few enzymes [Haldane, 1930]. 16 years later, McIlwain [1946] already repeated this in his nature article as a well known fact. In 1989 P.J.Halling asks, in the title of his publication, "Do the laws of chemistry apply to living cells?" and comes to the conclusion that a cell is a unique chemical system [Halling, 1989]. It is such a small reaction unit, that some species only appear in very small quantities - sometimes only a couple of molecules. In such cases, deterministic models are no longer appropriate. They are misleading and likely to result in incorrect expectations. For situations where chemical species exist in very small quantities, it may be better to use stochastic models.

The first scientists to mention stochastic methods as a tool for modeling chemical reactions were the biophysicist and Nobel prize winner Max Delbrück (1906-1981) and the dutch physicist and (Niels Bohr's first scientific assistant) Hendrik Anton (Hans) Kramers (1894-1952) [Delbrück, 1940, Kramers, 1940].

Delbrück examined enzyme reactions, and Kramers studied Brownian motion in a force field. In the 1950's, Alfred Renyi (1921-1970) was able to show that the Law of Mass Action breaks down for small systems [Renyi, 1954], and K.Singer explained that even small fluctuations can have significant effects on chemical reactions, that consequently can lead to the irreproducibility of experiments. If a system has only species present in low copy numbers, their steady-state fluctuations become significant in comparison to the mean. That is why the system can no longer be described by the deterministic law of mass action.

Many genetic regulatory reactions occur only at low concentrations. However, tiny changes can have a big influence on the whole system, as demonstrated by the phage λ lysis-lysogeny decision circuit [McAdams and Arkin, 1999, Rao et al., 2002]. This stochastic switching has been analyzed using stochastic kinetics and by deterministic models [Srivastava et al., 2002].

To attempt research on a cellular level, new algorithms were required. Gillespie [1977] proposed his *Stochastic Simulation Algorithm (SSA)*, which will be described in the form of the *First Reaction Method* in Section 2.4.2. It is also called the *Exact Stochastic Method (ESA)* [Vereecken et al., 1997]. Since Gillespie's proposal, improvements have been suggested such as the *Next Reaction Method* by Gibson and Bruck [2000] that -as has been recently shown- is not always faster, even though it uses less random numbers, due to its larger computational overhead [Cao et al., 2004b].

So far, several quantitative kinetic tools have been developed to model dynamic systems behavior (e.g.: E-CELL [Tomita et al., 1999], GEPASI [Mendes, 1993], and Virtual Cell [Schaff et al., 1997]). They can all be used for a wide variety of scenarios. If one momentarily ignores usability and implementation, they all include either completely stochastic (a form of the *SSA*) or strictly deterministic algorithms.

As mentioned previously and discussed in 2.3.1, both approaches have their realms of appropriateness in certain environments. Ordinary differential equations (ODE's) have the advantage of being fast and reliable in the macroscopic limit where a large number of molecules is available. However, at low concentrations, they cannot be described by these deterministic methods any more, due to stochastic effects. The presence of stochastic effects in gene expression and signal transduction processes has been shown by both, theoretical and experimental approaches [Levin et al., 1998, McAdams and Arkin, 1997, Ozbudak and al., 2002, Elowitz et al., 2002]. In this case, a stochastic approach such as Gillespie's *Direct Method* for modeling and simulation is biochemically and biophysically more realistic, but computationally limited in view of the high numbers of molecules (particles are traced as individuals simulating their movements and reactions).

Of course, stochastic algorithms are closer to reality, but very time consuming, restricting their use to systems with small and intermediate particle numbers. On the other hand deterministic models are better suited for systems with large particle numbers, but they fail at predictions for systems with intermediate amounts of molecules. Therefore, so called approximative algorithms have been invented that are mostly based on the *Direct Method*, and promise to fill the gap between stochastic and deterministic approaches. They are further described in Section 2.4.2. The key idea is to define a time interval in which the number of reactions is small, so the reaction probabilities can be assumed to be constant. However, those approximative algorithms, which work well for intermediate particle numbers, do not cover the complexity of a cellular system. The idea was conceived to develop hybrid algorithms, that use the afore-mentioned approaches and combine them into one strategy [Kiehl et al., 2004]. The crucial problem is to define the point of transition, i.e., when to switch from a stochastic to an approximate to a deterministic approach.

For example, some algorithms exist that are based on fixed partitions of the system into slow and fast reactions. "Fast" means here that it is likely to have a lot of reactions per time span and "slow" the opposite. With this combination, slow reactions are treated by the *First Reaction Method*, and fast reactions either by deterministic reaction kinetics [Haseltine, 2002, Kiehl et al., 2004, Takahashi et al., 2004], by Langevin equations [Haseltine, 2002], or by random variables distributed according to the probability density functions at a quasi-stationary state [Rao and Arkin, 2003].

The disadvantage is that the user of the algorithm has to partition the system, defining the point of transition. The non-automatic predefined partitioning makes these algorithms unusable for systems with oscillating concentrations, which is likely in most biological systems, because here a reaction changes its condition between "slow" and "fast". This is not the only point of criticism. Another aspect that has not been considered enough is spatial dimension, i.e., the cellular structure. Since automatic adaption does not exist, the ideas described are not appropriate for systems

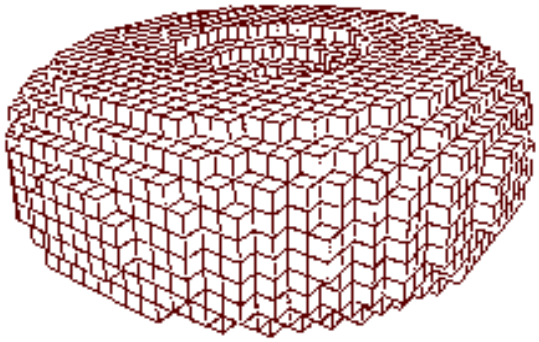

Figure 1.3: *A volume split in volume elements (voxels).*

with non-homogeneously distributed substrates. A cell is not a single reaction entity, it is a complex organization of specialized reaction compartments (i.e., mitochondria, Golgi complex). It has a complex three dimensional structure and whoever tries to simulate this, has to account for that as well.

So, although much data on cellular structure, construction, and constitution has been accumulated, there is still need for efficient algorithms simulating the cell as a reaction-diffusion system. The current implementations lack algorithms for multi-scale particle numbers and a complete representation of three dimensional diffusion processes. Most of the existing simulators divide the cell into a few homogeneous reaction spaces. In those simulators, the information on the whereabouts of molecules within the cell is lost, so that there is no way to reflect the change of gradients any more. Calcium waves are only one example to demonstrate the importance of spatial dimension [Fink et al., 2000].

The New Approach In my approach, the cell is divided into a grid of cubical volume elements (VE) (cf. Figure 1.3). Using this approach, we increase the number of observable spaces and are able to locate species dependently on the grid size.

Another advantage of subvolumes is related to the way reactions are handled. A subvolume is the smallest reaction jar. There are two general ideas of calculating biochemical kinetics. Either the changes of concentrations are predicted based on ordinary and partial differential equations (PDE), or by a stochastic approach using Monte Carlo simulations.

The VE-approach enables me to use a hybrid model that combines the advantages of exact, approximative, and deterministic approaches. With the Controllable Approximative Stochastic reaction-algorithm (COAST) presented herein, I have developed an algorithm fulfilling the needs of state-of-the-art simulators [Möller and

Wagner, 2005].

In recent years the community involved in the simulation of modeling biological systems emphasized the necessity of efficient designs of algorithms [Schwehm, 2001]. Based on the data by Endy and Brent [2001] Schwehm assumed 10^{14} reactions to occur per cell cycle in Escherichia coli and calculated that a stochastic simulation of a whole cell cycle would take about twelve years on a single processor. Therefore he concluded:

"Stochastic whole cell simulation is thus either the realm of massively parallel computing, or it needs new algorithms which can combine deterministic and stochastic simulation techniques."

With COAST, I present such an algorithm based on three different regimes of modeling. The transition points are defined by only a single error parameter α. This value controls the partitioning of the reactions. Three modeling levels are used: an exact way based on Gillespie's *Direct Method* for small particle numbers, for intermediate an approximative method based on Gaussian distributions, and for high particle numbers a deterministic approach.

To prove its reliability and accuracy, I compared COAST to the *First Reaction Method* and the *tau-leap method*, which is used as an example for an approximative algorithm [Gillespie, 1976, 2001]. It turns out that COAST is as accurate as the exact method, but is significantly faster than the exact and the approximative algorithm. It has shown reliable results for simple and complex systems like the Oregonator or Michaelis-Menten kinetics [Gillespie, 1977, Michaelis and Menten, 1913]. The key advantages of COAST are the wide scale of particle numbers covered, the self partitioning of the reactions channels, and the "easy-to-set" error value.

Another advantage of the COAST-algorithm is to model diffusion as well. As already suggested by Stundzia and Lumsden [1996], who considered the transition to neighbor volume elements (voxels) as merely additional reactions using the *First Reaction Method*, reaction algorithms can be adopted for diffusion. Thus far, partial differential equations have been the most efficient way in the three dimensional space next to track single molecules. It depends on the experiment, if the reconstruction of diffusion within a cell is necessary or if it can be neglected. This of course depends on the type of information one wants to gain. Essentially, all cellular processes include some kind of diffusive transport of metabolite- and enzyme-sized solutes [Oelveczky and Verkman, 1998].

It can be shown that diffusion and subcellular compartmentalization influences the signaling chemistry of a cell, which results in different signaling, such as washout of signals, reinforcement of signals, and the conversion of steady responses to transients [Bhalla, 2004]. It is important that one takes the three dimensional structure of a

cell into account. Spacial appearances such as calcium-waves have been modeled and experimentally shown [Fink et al., 2000, Strier et al., 2003]. The diffusion is becoming important since stochastic effects on a cellular scale have a higher impact.

Nowadays, scientist have the tools to study molecular diffusion processes. Biologists are assisted by using the green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* to tag nearly any protein and study their localization, dynamics and interactions in living cells [Lippincott-Schwartz et al., 2001, Tsien, 1998]. Other fluorescent proteins including the yellow fluorescent protein (YFP) are also used. New imaging methods improved the way of observing the GFP fusion proteins, such as fluorescence recovery (FRAP), fluorescence correlation spectroscopy (FCS) and fluorescence resonance energy transfer (FRET). They allow researchers to trace single molecules, measure concentrations of substrates, and analyse their distribution. Lippincott-Schwartz et al. [2001] point out that the technical advances will help scientists to move from a steady-state view to a dynamic model of cellular function. Such data has then been used to establish reliable three dimensional cellular models accurate enough to simulate cellular processes [Oelveczky and Verkman, 1998].

The most important algorithms that can be used to model diffusion are the Random Walk or Wiener Process (cf. Section 2.4.1), and stochastic deterministic equations (Langevin equation). The advantage of the Random Walk is the accurate modeling of all possible interactions that can occur, but the computational expense becomes very high, if the particle numbers rise. The Langevin equations describe the stochastic trajectories of single particles. Furthermore computationally expensive for large particle numbers. Other deterministic diffusion models suppress stochastic fluctuations and are therefore not useful for simulating signaling cellular networks with low particle numbers.

The reaction algorithm COAST can be modified to be a reaction diffusion algorithm, describing both crucial elements of a dynamic cell with one approach. To demonstrate the capabilities of COAST, I applied it to a one dimensional grid and tested speed and accuracy against the random walk as the most elementary way of modeling diffusion.

The accuracy of COAST is better than set by the error parameter α. I have been able to model the diffusion of particles without and in a forcefield. The results show that COAST is able to model the means and variances of the expected distributions accurately.

In this thesis I describe the reaction algorithm COAST and its application to diffusion. Furthermore I tested its accuracy and reliability compared to common algorithms and give a practical reasoning for the points of transition between the modeling regimes. The results indicate that the symbiosis of the diffusion and reaction implementations provide a powerful instrument for simulating cellular processes.

2 State of the Art

In the introduction I have explained the necessity of the development of efficient algorithms to model cellular systems. The main problems are the simulation of reaction processes involving strong changing numbers of particles and therefore the need of using deterministic models but also stochastic models to give a realistic picture of structures with low concentrations. In this chapter I will give an overview on existing methods to model chemical reaction and diffusion processes and discuss some of the existing tools for modeling biological systems.

2.1 Meaning of Chemical Reactions & Diffusion for Biological Systems

2.1.1 Chemical Reactions

Maybe it is true to say that chemistry is not everything, but without it, everything would be nothing. I am sure one will be always able to find exceptions, but when it comes to life and biological systems this statement cannot be more true.

Chemical reactions are the processes that result in the interconversion of chemical substances. The driving force behind chemical reactions is the desire of the reacting species to rearrange themselves into a lower energetic state. This is not limited to the inorganic world. It reaches its highest complexity by using carbon which allows a manifoldly variety of high complex molecules which are the basic components of life in the form we know; therefore the chemistry of carbon is named organic chemistry. Strictly, chemical reactions involve the motion of electrons in the forming and breaking of chemical bonds. However, the general concept of a chemical reaction is also applicable to transformations based on non covalent bondings.

Every reaction R has a different reaction velocity, quantified by a reaction constant k. The course of a chemical reaction is described by a reaction equation:

$$\text{reactant(s) or educt(s)} \longrightarrow \text{product(s)}$$

All biological processes depend on the formation and breaking of covalent and non covalent bondings. The latter include so called weak bondings which can be specified as electrostatic interactions, hydrogen bonds or van der Waals interactions. These weak interactions are the way of enzymes interact with their substrates, hormones with their receptors and antibodies with antigens.

The cell is the main site of enormous biochemical activity called the metabolism. This is the process of chemical changes which goes on continuously in the living organism. The build-up of tissue, replacement of old tissue, conversion of food to energy, disposal of waste materials and reproduction - all these activities are what we characterize as "life". Life needs chemical reactions.

2.1.2 Diffusion

Diffusion describes the spreading or distribution of a substance because of the thermal movement of their particles. Nowadays we define it as a spontaneous physical process of equilibration along a gradient of concentration, which is degraded during that process.

Diffusion is the most important way of molecular transport within cells, but mainly for small distances. To cover twice the distance a particle needs four times the diffusion time. In this way diffusion is limiting the size of a cell and defines for multicellular organisms the need of other not on diffusion based transport systems (nerve system, bloodstream). The diffusibility of a particle depends on the temperature, its size and its charge. We observe the diffusion of particles as a consequence of a concentration gradient. Corresponding to the second fundamental theorem of thermodynamics, which demands an increase of entropy, one can observe seemingly a directed movement of particles from the area of higher concentration to the area of lower concentration, but it is not a directed movement. The seemingly directed flow is the consequence of the stochastic process occurring here.

Biological processes constantly generate gradients of concentration by producing species in a localized manner, for example the production of proteins at the ribosomes.

2.2 From Systems Biology to Mathematical Modeling

To make biomolecular knowledge useful for medical or technical purposes one needs an integral understanding of cellular systems. Research has been concentrated over the last years on molecular details. Systems biology is the academic field that seeks to integrate different levels of information to understand how biological systems work. It is a "whole-istic" [Chong and Ray, 2002], interdisciplinary approach with methods and concepts of molecular biology, systemiology and informatics to gain a better understanding of cellular processes. It is not concentrated on single genes or proteins, but on the interactions between all components of a system.

H. Kitano [2002] published an excellent overview on *Systems Biology*. According to him it is an examination of structure and dynamics of cellular and organismal function rather than the focus on isolated parts of a cell. Many properties arise at

the systems level only and cannot be derived by looking at details. A cell is an example for a system with a complex microstructure, whose components communicate manifoldly among each other and with the outside world. Voit [2002] supports this by stating, that the challenge dealing with complex systems is a result of synergistic properties, which do not exist in any constituents, but only in their intricate interrelationships .

Knowing the parts of a -for example- gene-regulatory network and their interactions is not enough. We have to understand how changes in one part are affecting the others, how they dynamically interact. Kitano thinks understanding of a system can be gained by insight into four properties: "System structure" (gene interactions, biochemical pathways), "System dynamics" (system behavior over time under changing conditions), "The control method" (mechanisms that control the state of the cell), "The design method" (modify systems to have desired properties).

Biology delivers the data and has the methods to gain them, informatics processes and structures it. Another purpose of informatics is to provide tools to model and visualize. The system sciences provide methods to describe, analyse and abstract the biological systems. Classical examples of systems are the immune system, or the nerve system. The original idea of a system level understanding is not new and goes back to the first half of the last century [Wiener, 1948], but new methods like automated gene sequencing, DNA microarrays, proteome chips and metabolic profiles have provided science with valuable information about the genetic and metabolic responses of organisms to stimuli to make an in silico cell envisionable [Voit, 2002].

Kitano points out, that understanding of the properties of biological systems might have an impact on the future of medicine. Drug discovery through trial and error has been successful throughout the centuries [Voit, 2002]. Then man began to do research on the details of the organism, to disassemble it and its components to optimize the process in finding the fundamental mechanisms of health and disease. This approach is called reductionism and has been useful over the last decades. As described above the knowledge of details is not enough. By knowing more about the interchange of those single parts, pharmaceutical companies would be able to undergo their research much more efficient with less failures and less expense.

There are current ongoing initiatives for *systems biology* and I will only list some examples: Institute for systems biology Seattle, USA, by Leroy Hood; Alliance for cellular signaling, USA, by A. Gilamn; and in Japan the Systems biology group by H.Kitano. [Kitano, 2002]. In Germany the BMBF has financed projects within the scope of the research program "Systeme des Lebens - Systembiologie". The aim is a virtual representation of a cell like a virtual laboratory. This should smooth the way for predictive biology, where comparable research is possible like in a real biological system.

This is not easy. The biological and metabolic systems governing the effects are

dramatically complex. Voit points out, a cell is more than a collection of membranes, organelles and proteins, mixed with some DNA and RNA. Reductionist methods are necessary, but they need to be accompanied by mathematical concepts, which are capable of capturing the essence of complex, integrated systems.

2.3 Modeling

A central role is the mathematical modeling of complex cellular networks. The mathematical models connect the parts of systems biology. The Process of modeling is interactive. That is the model world has to converge towards the reality. All models have to be compared to real data and than be adjusted. In Section 2.5 some approaches toward modeling cellular systems are described.

A first step to get away from a statical biological network (cf. 'Biochemical Pathway', Boehringer®) is to model biological processes by algorithms and to represent their dynamical character with these algorithms. All available data for function, localisation, concentration and interactions are thereby calculatively combined. Models are abstractions that are easier to manipulate than the actual system [Endy and Brent, 2001]. They are typically heuristic and develop alongside the experiments and are inseparable from them. Future development of computational speed will be of critical need to implement high scaled networks. The development of reliable models is crucially dependent on the data the model is based on. Gaps in knowledge about components of the system, interactions and of other parts of related networks can be very harmful. Abstraction is necessary for a model, but the simplification must be done carefully. The impact of molecular crowding is well known [Minton, 2001], but because of its complexity it has not yet been modeled. One criterion of biological complexity is the rich network of interactions among the constituents. These interactions are numerous and have nonlinear characteristics that are difficult to handle with intuition alone [Voit, 2002]. Nonlinearities make complex systems difficult to understand. Only mathematical models are able to help us predict those systems far away from our intuition.

In the 1950s the computer became useful in solving systems of differential equations. However, it took 20 more years until in the 1970s stochastic methods have been developed to model low representations of species (cf. Section 2.3.1 for stochastic and deterministic methods) [Gillespie, 1976]. Those methods have been improved (cf. next section for details) [Gibson and Bruck, 2000] and used in the 90's to describe different systems [Arkin et al., 1998, Bhalla, 2002]. Further improvements can be expected by combining stochastic and deterministic approaches (cf. Section 2.4.2). This dissertation is to be a contribution to the worldwide discussion on that topic. A system can be at some steady state for a certain parameter value. If this value is now raised above a certain threshold, a feedback mechanism can result in

an oscillating system and not a proportional increase of the output. Here are some examples how mathematical models have been used in the past.

J.Tyson [1991] modeled the cell-division-cycle. In his model like in most mathematical models chemical kinetics are represented by involving ordinary and partial differential equations. With such simulators the switch-like behavior of the MAPK-module in signal transduction has been clarified [Ferrell, 1999], which is experimentally only accessible if experiments are done on single cells. Other simulations that have been done were simulations of ion channels [Hodgkin and Huxley, 1952] and the human heart [Noble, 2002]. The work by Hodgkin and Huxley represents one of the highest-points in cellular biophysics and the quantitative model of action potential generation and propagation they developed forms the basis for understanding and modeling the excitable behavior of all neurons; it is the single most successful quantitative model in neuroscience [Hille, 1984]. Smith et al. [2002] were the first to model the transport of molecules into and out of the nucleus. Other examples are calcium waves describing the release of calcium ions from the sarcoplasmic reticulum [Loew and Schaff, 2001].

Several tools are available already to enable biologists to get access to a field that hitherto has been restricted to the design of integrated circuits and chemical processing plants [Doyle, 2001] (cf. Section 2.5).

The main problems of modeling result from simplification and abstraction. The cell is not a well stirred reaction tube. It is highly compartmented with high local concentrations (e.g. mitochondria matrix). Macromolecular crowding has a great impact on diffusion and reaction. Molecules of a certain size are not able to diffuse at all or by a much smaller diffusion coefficient than in experimental buffer. Furthermore endogenous obstacles hinder diffusion. Many reactions occur on two-dimensional membranes or in dimensionally restricted environments.

Another difficulty is the nonlinearity of complex systems. They defy the law of superposition, which means that *Devide and Conquer (Julius Caesar: Divide et Impera)* is not possible. The single parts of linear systems can be analysed independently from each other, but nonlinear systems usually lose essential characteristics when taken apart [Voit, 2000]. Without mathematical analysis it is difficult to predict simple mechanisms like a feedback loop. Does the output increase or decrease or don't change at all? Intuition does not help here. Other systems begin to oscillate under certain conditions and when the parameters are slightly changed it does not respond at all [Kholodenko, 2000].

Modeling gives a flexibility actual lab experiments cannot provide. One is able to model time expensive experiments in a fraction of the otherwise needed time or observe processes in detail which take only microseconds. However, the aim of modeling is to assist traditional laboratory work, to suggest and council, but not to replace it.

2.3.1 Deterministic Versus Stochastic Modeling

Mathematical modeling is a powerful approach for understanding the complexity of biological systems. There have been already successful attempts for simulating complex biological processes like metabolic pathways, gene regulatory networks and cell signaling. The models have not only generated experimentally verifiable hypothesis but have also provided valuable insights into the behavior of complex biological systems [Meng et al., 2004]. Modeling can be divided into three levels:

- macroscopic: dynamics of molecular concentrations, mostly deterministic models (differential equations or S-System [Voit, 2000])
- mesoscopic: dynamics of single molecules, in general without referring to physical forces (mostly stochastic models)
- microscopic: simulation of physical forces between and within molecules, e.g. protein folding, docking, molecular modeling

The most accurate way would be modeling molecular dynamics - modeling on a microscopic level. Therefore one has to track the position and velocity of every molecule in the system. Furthermore every collision has to be observed, if there is a chemical reaction or not. By modeling molecular dynamics we investigate the changes in species populations and their spatial distribution. The main problem with this approach is the computational expense. Although computer technology made a big leap forward during the last decade, such operations are still for supercomputers out of reach referring to complex biological systems in terms of time.

There are several more efficient approaches to model chemical and physical processes like reaction and diffusion. In Section 2.4 some of these models are presented in detail. With slight simplifications the models can be classified in two categories. They can be classified as either stochastic or deterministic, excluding consciously the approximate approaches at this point. Nowadays hybrid models are introduced to close the gap between these two regimes. To simplify the process let us concentrate now on deterministic and stochastic algorithms for chemical reaction and afterwards discuss how the existing ideas can be used to model diffusion.

Deterministic Modeling If we assume the amount of reactive collisions to be low compared to the amount of unsuccessful reactions, a simplification can be made. Particles can now diffuse within a certain area and keep the system in a well-stirred condition. Based on this assumption in chemistry it is quite common to formulate chemical kinetics of a chemical reaction using the Law of Mass Action:

$$2A + B \leftarrow k_2 \,|\, k_1 \rightarrow C \qquad \frac{[A]^2 \cdot [B]}{[C]^2} = K(T,p) = \frac{k_2}{k_1},$$

where K is the reaction constant and a function of the temperature T and the pressure p. This function is derived from:

$$\frac{d[C]}{dt} = k_1 [A]^2 [B] - k_2 [C].$$

Here k_1 and k_2 are the velocity constants for the two single reactions. The velocity constants are proportionality constants equal to the initial rate of a reaction divided by the concentration of the reactant. A,B and C are the substrates and $[A]$,$[B]$ and $[C]$ are the concentrations of the substrates, usually in $\left[\frac{mol}{l}\right]$.

In words the Law of Mass Action says: The product of the concentration of the reaction partners with all concentrations always taken to the power of their stoichiometric factors, equals a constant K which has a numerical value that depends on the temperature and the pressure. K is called the reaction constant. The Law of Mass Action follows if one assumes that the system has reached equilibrium and $\frac{d[C]}{dt} = 0$. Let us emphasize the main problem with determinism directly here. The differential equation assumes that the system is continuously predictable, which is of course not the case for a complex biological system. Furthermore, the differential equation or a system of differential equations works very well for high numbers of particles where fluctuations can be neglected, but often molecules in cell structures are only present in small amounts and show a stochastic behavior. Moreover, the numbers of particles change not continuously, but discrete. The modeling of chemical reactions using deterministic rate laws has proven to be extremely successful in both chemistry and biochemistry [Epstein and Pojman, 1998, Heinrich and Schuster, 1996].

Usually one is interested in the change of metabolic concentrations over time. Therefore a differential equation or a system of differential equations has to be solved. However, exact solutions only exist for very simple systems. Consider the system in the following reaction equation as an example where again A and B mark the substrates and k is the reaction constant:

$$A \xrightarrow{k} B \qquad \Rightarrow \qquad \frac{d[A]}{dt} = -k [A] \qquad \Rightarrow \qquad [A] = [A_0] e^{-kt}$$

More complex systems have to be solved by numerical simulation (e.g., Runge-Kutta Method [Butcher, 1987]). In such a deterministic system of differential equations every substrate and each of its derivates (modified substrate) must have one equation. So the number of reactions is directly depending on the amount of species present in this system. An additional complication can result from differently fast reactions, then the system is called "stiff".

If one applies ordinary differential equations, one makes three implicit assumptions:

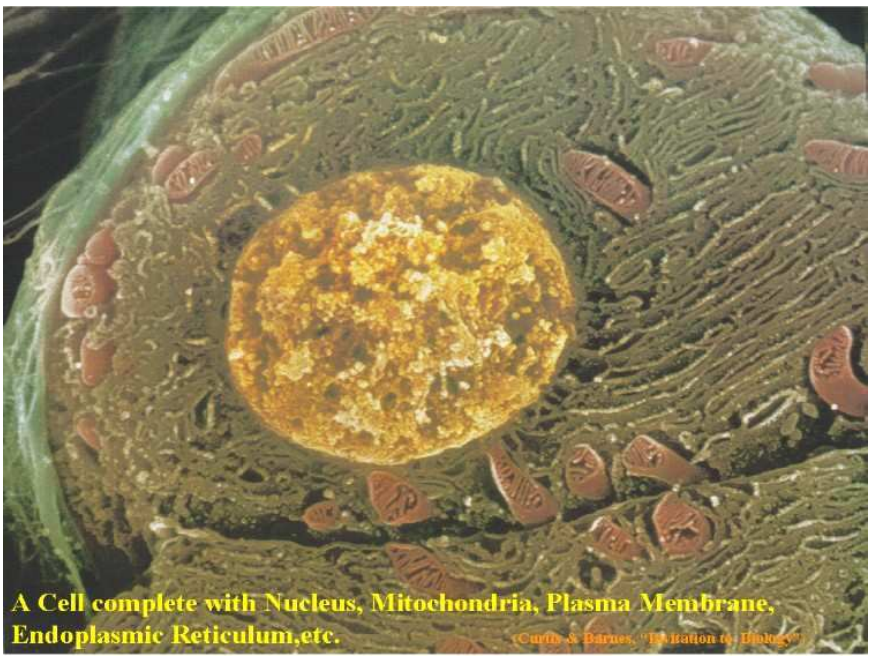

Figure 2.1: *A Cell, a homogenous volume?*
Reprinted from "Invitation to Biology" by H.Curtis & N.S. Barnes, Worth Publishers. Copyright (1994), with permission from W.H. Freeman and Company / Worth Publishers

- a very large reaction volume with high particle numbers present
- an equilibrium of the system
- a homogeneous distribution of all molecules

Let us now have a closer look at those assumptions. The focus of this thesis are algorithms that shall be applied to biological systems. The appropriate reaction volume would be the volume of a single cell.

If we observe the changes of concentrations in the cytosol of a cell, describing them by ODEs is a common way. Therefore one considers chemical reactions to be macroscopic under convective or diffusive stirring, continuous and deterministic [Cox, 1994]. This is evidently a simplification, as it is well understood that chemical reactions involve discrete, random collisions between individual molecules. However, if we only look at small reaction volumes like single vesicles, one cannot speak of a determinism any more. We are used to calculate with concentrations, but on such a low level the changes are molecule by molecule - discrete and not continuous. We reach a level on which a deterministic approach is not useful anymore because it

cannot describe spontaneous changes of the reacting molecule number. The system is now showing a stochastic behavior.

Associated with this is the question about the equilibrium. If an equilibrium is reached in a deterministic system, nothing is changing any more, but because chemical reactions are stochastic events, one cannot neglect fluctuations. Biological systems can leave such apparent stable states. Furthermore, biological systems are quite often close to instable conditions.

The last assumption, the homogeneous distribution, is necessary if one uses ordinary differential equations, because they do not take local resolutions into account. If one wants to do that, one has to use partial differential equations. This is maybe not necessary if the observed system is a lake, but a single cell represents a very complex structured system. The single compartments separated by single or double membranes are specific reaction volumes with optimal reaction conditions for specific reactions (i.e., mitochondria, liposomes, endoplasmatic reticulum, nucleus, Golgi apparatus). To assume everything as one volume is not only far from reality in a structural but also physiological sense, because certain reactions are not able to exist next to each other. Because of the presented limitations, deterministic models are limited to areas in which high numbers of particles occur, like metabolic processes, but they are not suited for signal transduction or gene expression due to the low representation of substrates. Here one has to use mesoscopic models such as stochastic models.

Nevertheless deterministic modeling has led to some interesting results. In all biological systems, it is necessary to increase or decrease activities in response to external and internal signals. The sensitivity of the system to such signals becomes very important. The term ultrasensitivity has now been defined to indicate a case in which the sensitivity is greater than that to be expected from standard hyperbolic (Michaelis-Menten) response [Goldbeter and Koshland, 1984]. In 1996 Huang and Ferrell were solving the rate equations for the MAPKinase-system numerically and predicted the cascade to work as a switch [Huang and Ferrell, 1996] (ultrasensitive). They were able to show the ultra sensitivity of this bistable system in experiments with *Xenopus oocyte*. In 2000 Kholodenko demonstrated how negative feedbacks and ultrasensitivity can lead to oscillations in the mitogen-activated protein kinase cascades [Kholodenko, 2000]. Levchenko et al. [2000] simulated the influence of scaffolding proteins on the MAPK-system.

Another framework, which is worth to be mentioned, was developed over the last 30 years to model complex metabolic pathways and gene regulatory networks: it is called *canonical modeling* [Voit, 1991]. It is based on the Biochemical Systems Theory (BST) [Savageau, 1969]. The variables describing e.g. a signal transduction pathway are metabolites and enzymes. The dynamics of each variable is described by the change of its value over time and this change is governed by the difference of all influxes and effluxes. All fluxes are described by power-law functions which

are justified by the Taylor's expansion based on the calculus of finite differences developed by Brook Taylor (1685-1731). In fact virtually any ordinary differential equation can be written equivalently in canonical form of a Generalized Mass Action (GMA)-, S-, or Lotka-Volterra-system. Equation (2.1) shows the most important type of canonical models, the S-system [Savageau, 1969]:

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} \qquad i = 1, 2, ..., n \tag{2.1}$$

where $X_1, ..., X_n$ stand for dependent variables (dynamic concentrations of internal metabolites), $X_{n+1}, ..., X_{n+m}$ stand for external variables (fixed concentrations of external metabolites), g_{ij}, h_{ij} are kinetic orders, which may be non-integer and non-positive, and α_i, β_i are rate constants. In a nutshell, the functions and variables are represented in logarithmic coordinates. In this coordinate system, the functions are approximated by Taylor series, where only the constant and linear terms are retained.

Stochastic Modeling The occurrence of stochastic phenomena in a variety of physical systems like turbulent fluid flow, is well established. In the recent past attention has shifted to stochasticity, noise and its impact on biological systems [Meng et al., 2004]. On a molecular level random fluctuations are inevitable and get more significant if the number of interacting particles is very low. This is for example the case during transcription where transcription factors interact with DNA binding sites [Ozbudak and al., 2002, Elowitz et al., 2002]. Beyond this McAdams and Arkin [1997] were able to prove that low copy RNA can be significant for the regulation of downstream pathways. Ross and al. [1994] described mRNA being produced in random pulses.

One characteristic of stochastic systems is that identical initial conditions, such as initial concentrations or an initial temperature, can lead to completely different results. One studied example is the lysis/lysogenic switch of the bacteriophage λ infecting E.coli [Arkin et al., 1998]. Due to noise the network may randomly evolve into one of the two bistable states [Hasty, 2000].

The occurring fluctuations in the concentrations or particle numbers can be ascribed to two different effects. That is why one distinguishes between intrinsic and extrinsic noise. Stochastic effects arising due to the inherent nature of biochemical effects are termed as *intrinsic noise* [Meng et al., 2004]. This is for example the case during transcription, while only a few transcription factors and mRNA molecules are interacting with the DNA. Reactions occur here only randomly. On the other hand the subsequent step -the translation- has an *extrinsic* component of noise. The randomly fluctuating factors are the number of ribosomes, the stage of the cell cycle, the mRNA degradation and the cellular environment. They all depend on external environmental conditions.

As we now have seen processes like gene regulation cannot sufficiently be modeled by a deterministic model, observing the system from a macroscopic point of view.

As described above one has to distinguish between intrinsic and extrinsic noise or stochasticity. There are many equivalent formulations of stochastic kinetics. One, the chemical master equation, describes the evolution in time of the probability distribution of system composition. The chemical master equation is a set of linear ordinary differential equations with discrete changes of particle numbers. This set cannot be solved analytically, only numerically. One approach is Gillespies *Stochastic Simulation Algorithm* (cf. Section 2.4.2); another one is the *Chemical Langevin Equation (CLE)* also proposed by Gillespie [2000]. To treat extrinsic stochasticity a stochastic term is introduced into the governing reaction equation, which simulates the fluctuating noise:

$$\frac{d\vec{X}(t)}{dt} = \sum_{j=1}^{M} \vec{\nu}_j \, a_j \left(\vec{X}(t)\right) + \sum_{j=1}^{M} \vec{\nu}_j \sqrt{a_j \left(\vec{X}(t)\right)} \, \Gamma_j(t),$$

where

$$\begin{aligned}
\vec{X} &: \text{stores the number of molecules for all species,} \\
\vec{\nu}_j \equiv (\nu_{1j} ... \nu_{Nj}) &: \text{the change in the number of } S_i \text{ molecules caused} \\
\cong \nu_{ij} &\quad \text{by one } R_j \text{ event,} \\
a_j(\vec{x}) &: \text{a propensity function (given the system in state} \vec{x}\text{), } a_j(\vec{x})dt, \\
&\quad \text{is the probability that one } R_j \text{ event will occur in the next dt .}
\end{aligned}$$

This equation does not refer to diffusion. All other dynamical processes except of reaction are assumed to have come to equilibrium much faster than the composition, so we have the situation of a "well-stirred system". The transitions between different compositions are called propensities. If the noise Γ is Gaussian and white, the probability distribution satisfies a Fokker-Planck equation. Robert Zwanziger [2001] was able to show that Γ is not really a Gaussian distribution, but as an approximation it is sufficient. The discrete stochastic process $\vec{X}(t)$ is now approximated as a continuous stochastic process. The CLE can be invoked, if the reactant population is "sufficiently large".

The stochastic treatment of chemical reactions was initiated by Kramers in 1940 [Turner et al., 2004]. Fundamental is the idea that molecular reactions are essentially random processes; it is not possible to say with complete certainty when the next

reaction will occur within a volume. Turner points out that in macroscopic systems, with large numbers of interacting molecules, the randomness of this behavior averages out so that the overall macroscopic state of the system becomes highly predictable. It is this property of large scale random systems that enables a deterministic approach to be adopted.

2.4 Existing Methods

2.4.1 Methods for Simulating Diffusion

Historical Background In 1827 the English botanist Robert Brown ($*1773, \dagger1858$) observed pollen grains in aqueous solution. He was stunned to see that even after hours of observation they still moved restless. He claimed he was able to reproduce this observation with sulfur, volcanic ash and other fine grained substances, but there are doubts if he really was able to observe it [Deutsch, 1991].

We now refer to stochastic movements of charged or uncharged particles in watery solution as Brownian motion. The first quantitative description of a diffusion process was done by the physiologist A. Fick [1855] ($*1829, \dagger1901$). The relationship he found, known as *Fick's Law of Diffusion*, states that the rate at which the concentration of a substance decreases at any point x in a system is proportional to the curvature of the concentration gradient at that point. The constant of proportionality, $\mathbf{D}$, is the diffusion coefficient or diffusivity in the system [Agutter et al., 2000]. During his PhD thesis in Zuerich in 1905 and in two publications in the "Annalen der Physik" (1905/1906) A. Einstein ($*1879, \dagger1955$) and independently von Smoluchowski ($*1872, \dagger1917$) found an explanation for Fick's law in molecular terms.

Part of the analysis also led to a derivation of Fick's law and to the general inference that the macroscopic diffusion process can be explained by the molecular-kinetic mechanism of Brownian motion in fluid systems where there are concentration gradients. Einstein was able to calculate Avogadro's number, which had so far only been roughly determined. Additionally the theory was seen as an additional proof for the relatively new atomistic theory. The work of Einstein and Smoluchowski further assisted in the development of the theory of stochastic processes. The American N. Wiener ($*1894 - \dagger1964$) used the Einstein-von Smoluchowski equation for the probability distribution of diffusing particles to derive the probability that an individual particle would pass during a stated interval of time between any two points in a defined space in 1923. Today the "Wiener process" is a synonym for Brownian motion. Stochastic theory has been influential in quantum mechanics (e.g. Feynman's path integral method), in mathematics (leading to the discovery of profound connections between functional analysis, differential equations and probability theory), and in

several other fields.

Partial Differential Equations The most conservative way to model diffusion are partial differential equations (PDE's). These can be further divided into two main classes: linear PDE's and non linear PDE's. The usual way of describing diffusion processes would be by using non linear partial differential equations (NLPDE's). The aim here is to model the interactions between the particles of the same species. This results in equations of higher order (non linear). Several techniques are known to integrate them numerically:

- multigrid method
- finite elements method
- "Monte Carlo" method
- spectral theory
- cellular automata
- lattice Boltzmann gas method

If one can assume that the concentration of particles is so low that interactions between particles of the same species can be neglected, linear PDE's are sufficient. There are again two different approaches related to the linear PDE's. The *Smoluchowski-approach* requires the strong friction limit; i.e., the particles do not have an inertia, which results in a Markov-process. For the *Fokker-Planck* differential equation the *strong friction limit* is not used and therefore we do not have a Markov-process, this again results in ones ability to give information about the acceleration of the particles. The Fokker-Planck differential equation describes the time evolution of the particle distribution function.

Partial differential equations are for example used by *Virtual Cell* with the finite volume method [Schaff et al., 1997]. The space is divided into subvolumes and the transfer between the volumes is calculated by PDEs. A smaller grid is producing more accurate results, but with a higher computational cost. The main problem with PDE's is that they are not capable to reflect stochastic effects, but noise is important and it gains on importance the smaller the subvolumes are.

Another way to describe spatial movement of particles are *stochastic differential equations*. They have already been applied in biology with a different focus (population growth [Kiester A.R., 1974], granulocyte movement [Boyarsky et al., 1976] and populations genetics.
The most important form is the *Langevin Equation*:

$$m \cdot a \;=\; F(x(t)) \;-\; \gamma v \cdot m \;+\; \xi(t), \tag{2.2}$$

with F being the interaction force on a single Brownian particle, m the mass of this particle, x the x-position of the particle, γ its friction coefficient, v the velocity and a its acceleration.
$F(x(t))$ describes the interatomic forces and are therefore equivalent to Newton's equation of motion. The second term on the right side in Equation (2.2) represents the frictional force by the solvent. $\xi(t)$ is the random stochastic force due to thermal fluctuations of the solvent. The solvent is not explicitly represented, but its effects on the molecules by the frictional and the stochastic term. If $\xi(t)$ has the mean of zero, the equation is called the *Langevin Equation* (1908). The *Langevin Equation* is an alternative way next to the *Focker Planck Equation* (FPE) to describe Markov Processes and is the calculus of stochastic differential equations (SDE) governing the dynamics of the system.

The FPE is a deterministic partial differential equation, which can be solved either by numerical or analytical methods:

$$\frac{\partial p}{\partial t} = \left[-\frac{\partial}{\partial v} v + \frac{\partial}{\partial v} \left(\gamma v - \frac{F(x)}{m} + \frac{\gamma k_B T}{m} \frac{\partial^2}{\partial v^2} \right) \right] p,$$

where p is the transition probability, m the mass of a particle, γ the friction coefficient, T the temperature, k_B the Boltzmann constant, v the velocity of the particle and $F(x)$ the external force field.

Its original purpose has been to describe microscopic processes in the presence of random forces (noise). Three coupled *Langevin equations* are needed to describe the motion of a single particle in three dimensions. To solve SDE's there are in principal two possible ways: one is to model single trajectories by using for example Gillespie's SSA (cf. Section 2.4.2) or integrating the SDE's to find the solving probability distribution. If the frictional and the random forces are zero, the *Langevin Equation* reduces to Newton's equation of motion, which is the mathematical simple description of *molecular dynamics* (MD). All forces affecting every molecule in a calculated and this results in the computational cost of MD simulation to increase linearly with the number of interacting atoms. MD's are despite of being most accurate not feasible for whole cell simulations. So far it was only used for small numbers of particles and little volumes [Baynes and Trout, 2004, Friedel and Shea, 2004].

While the FPE is a statistic approach to calculate the change of the probability density over time, the *Langevin Equation* on the other hand was originally describing the temporal change of the velocity of particles due to a stochastic force.

Brownian dynamics Another approach are Brownian dynamics which are realized in the tool *MCell* [Stiles and al., 1998]. This is a stochastic approach where single molecules are followed. Their movement is according to the *Langevin equation*, which includes random forces. These random forces reflect the interactions between substrates and solvent molecules. With this method crowded environments can be simulated, but it is causing high computational effort.

Cellular Automata Using a different approach, Weimar [1997] describes two classes of *cellular automata* for reaction diffusion systems. The first type realizes diffusion by a random walk of particles on a lattice (reactive lattice gas automata) and the second one, moving average cellular automata, is based on a local average. The molecules are replaced by idealized particles. These cellular automata evolve on a square lattice on which particles propagate in two dimensions, with nearest neighbor interactions only.

Direct Method Nowadays exact stochastic approaches have become the norm in biochemical simulations. However, it was not until the 90's of last century when researchers were thinking of modeling molecular movement in the cell. Stundzia and Lumsden [1996] extended Gillespie's *Direct Method* (cf. Section 2.4.2) to diffusion by treating the diffusion from one subvolume to an adjacent volume as an additional reaction step. The time step is calculated stochastically by a probability function, which is determined by the intrinsic reaction kinetics and diffusion dynamics.

Mesoscopic Approach In the same year Ander et al. [2004] published *Smart -Cell*, a framework to simulate cellular reaction diffusion processes. It uses a mesoscopic reaction model to simulate diffusion and localisation of particles. In contrast to *MCell* and *Smoldyn* [Lipkow et al., 2005], *SmartCell* (cf. Section 2.5 for details) does not treat diffusion as a random walk, where all molecules are simulated individually. The algorithm is very similar to the *Next Subvolume Method* by Hattne and Elf, which was independently developed and published in 2005 [Hattne et al., 2005]. *SmartCell* is based also on the idea to separate the volume into subvolume elements. Similar to the approach by Stundzia and Lumbsden they treat the diffusion as an additional stochastic reaction, but using the *Next Reaction Method* by Gibson and Bruck to model diffusion as a single translocation of a molecule between adjacent voxels. Like in many other simulators, the particles do not have a volume here, so excluded volume effects cannot be modeled. For each event a probability is calculated and then a timespan τ. The event with the lowest τ is executed and the probabilities are recalculated.

Method	abbr.	Space	Scale	Stochastic	Excluded
molecular dynamics	MD	Particle	Micro	-	+
Brownian dynamics	BD	Particle	Micro	+	+
Smoldyn	-	Particle	Micro	+	-
cellular automata	CA	Discrete	Micro	+	+
Spatial Gillespie	-	Discrete	Meso	+	-
partial differential equations	PDE	Mesh	Macro	-	-
Gillespie	-	-	Meso	+	-
ordinary differential equations	ODE	-	Macro	-	-

Table 2.1: *Summary of existing approaches to model diffusion processes (freely adapted from [Takahashi et al., 2005])*
MD: molecular dynamics; BD: Brownian dynamics; CA: Cellular automata; PDE: partial differential equations; ODE: ordinary differential equations

Next Subvolume Method Hattne and Elf introduced the *Next Subvolume Method* one year after *SmartCell* [Hattne et al., 2005]. The algorithm is based on the reaction diffusion master equation (RDME,[Baras and Mansour, 1996]). The space is divided into subvolumes, which must be chosen small enough to ensure homogeneity and the RDME is applied to every voxel. The Diffusion is now a first order elementary reaction between the subvolumes. The RDME is complex and with analytical solutions hard to come by. Thats why so far only 1D systems where modeled [Baras and Mansour, 1996, Górecki et al., 1999]. The *Next Subvolume Method* is the implementation of the RDME to more dimensions. With an increasing number of voxels Gillespies *Direct Method* is not feasible any more, because the computational effort rises linear with the amount of voxels. The *Next Subvolume Method* is using the *Direct Method* by Gillespie for sampling the time for a next reaction or diffusion event and the *Next Reaction Method* by Gibson and Bruck to decide in which subvolume the next event occurs. They claim the computational time of the algorithm increases only logarithmically, than linear with the amount of subvolumes.

The approaches by Stundzia and Lumbsden, by Ander et al. and by Hattne and Elf are three similar diffusion algorithms, that is why they are also called as "spatial Gillespie" approaches [Takahashi et al., 2005].

2.4.2 Reaction Algorithms

Stochastic Simulation Algorithm - SSA Gillespie [1976] presented a stochastic reaction algorithm based on Newtonian physics and thermodynamics. Furthermore he described two possible implementations of his algorithm, namely the *Direct Method* and the *First Reaction Method*. His model assumes a system of N chemical species

$(S_1, ..., S_N)$ that interact through M reaction mechanisms (or channels) $(R_1, ..., R_M)$ in a specified Volume V at a constant temperature T. The *Grand Probability Function* $P(\vec{X}; t)$ gives the probability that there will be present in V at time t, X_i of species S_i, where $\vec{X} \equiv (X_1, X_2, ..., X_N)$ is a vector of molecular species populations [Turner et al., 2004]. The knowledge of this function provides a complete understanding of the probability distribution of all possible states at all times.

If the system is well stirred or the amount of reactive collisions is much smaller than the number of nonreactive collisions, each reaction R_μ can be described by the *propensity function* which is also known as the *Fundamental Hypothesis* of the stochastic formulation of chemical kinetics:

$$a_\mu \, dt \equiv h_\mu \, c_\mu \, dt. \tag{2.3}$$

The propensity function in Equation (2.3) gives the probability a_μ of reaction μ occurring in the time interval $[t, t + dt]$. μ is an index $(1 \leq \mu \leq M)$. h_μ denotes the number of possible combinations of reactant molecules involved in reaction μ. The Table 2.2 shows some examples.

reaction		$c_\mu =$	h_μ
monomolecular reaction	$S_i \rightarrow \mathrm{P}$	k_μ	$\lvert S_i \rvert$
bimolecular reaction	$S_i + S_j \rightarrow \mathrm{P}$	$\frac{k_\mu}{V}$	$\lvert S_i \rvert \cdot \lvert S_j \rvert$
bimolecular reaction	$2S_i \rightarrow \mathrm{P}$	$\frac{2\,k_\mu}{V}$	$\frac{1}{2} \lvert S_i \rvert \cdot (\lvert S_j \rvert - 1) = \binom{\lvert S_i \rvert}{2}$

Table 2.2: Conversion from kinetic to stochastic reaction constants

k_μ is the macroscopic velocity constant of a chemical reaction. To measure it, one only needs macroscopic properties of the chemical system, mainly the concentrations of the participating species. However, c_μ is the mesoscopic velocity constant, which is different from k_μ, but can be calculated from k_μ by knowing the volume of the observed system and the kind of chemical reaction taking place (cf. Table 2.2).

If one considers an infinitesimal small time interval $(t, t + dt)$, in which either one or zero reactions occur, there are only $M + 1$ possible ways to lead to the state $\vec{X}$ at time $t + dt$. So one can formulate:

$$\begin{aligned} P(\vec{X}, t + dt) &= P(\vec{X}, t) \cdot P(\text{no state change over dt}) \\ &+ \sum_{\mu=1}^{M} P(\vec{X} - \vec{v}_\mu, t) \cdot P(\text{state change to } \vec{X} \text{ over dt}), \end{aligned}$$

where $\vec{\mu}$ is a stoichiometric vector defining the result of reaction μ on the state vector $\vec{X}(\vec{X} \rightarrow \vec{X} + \vec{v}_\mu \text{after reaction } \mu)$ and further

$$\begin{aligned} P(\text{no state change over dt}) &= 1 - \sum_{\mu=1}^{M} a_\mu(\vec{X})\, dt \text{ and} \\ P(\text{state change to } \vec{X} \text{ over dt}) &= \sum_{\mu=1}^{M} P(\vec{X} - \vec{v}_\mu, t) a_\mu(\vec{X} - \vec{v}_\mu)\, dt. \end{aligned}$$

By using

$$\lim_{dt \to 0} \frac{P(\vec{X}, t+dt) - P(\vec{X}, t)}{dt} = \frac{\partial P(\vec{X}, t)}{\partial t},$$

one obtains the *Chemical Master Equation* that describes the stochastic dynamics of the system:

$$\frac{\partial P(\vec{X}, t)}{\partial t} = \sum_{\mu=1}^{M} a_\mu(\vec{X} - \vec{v}_\mu)\, P(\vec{X} - \vec{\mu}_\mu, t) \; - a_\mu(\vec{X}) P(\vec{X}, t).$$

To simulate now a system of chemical reactions, one has to be able to give information about two things:

- when is the next reaction going to occur
- which reaction will it be

Gillespie [1977] introduced a probability distribution to describe the system governed by the master equation. $P(\tau, \mu)dt$ is the probability for the next reaction to occur in the interval [t+τ,t+τ+dτ] and is of type μ. $P(\tau, \mu)dt$ is a two dimensional density function with the continuous variable $\tau (0 \leq \tau < \infty)$, which gives information of the point in time, and the discrete variable μ ($\mu = 1, ..., M$), which states which reaction is occurring. The probability for the next reaction can now be formulated as following

$$P(\tau, \mu)d\tau = P_0(\tau) h_\mu c_\mu d\tau,$$

where $P_0(\tau)$ is the probability for no reaction within [t,t+τ]. This is multiplied by $h_\mu c_\mu d\tau$, the probability that in the upcoming interval [t+τ,t+τ+dτ] the reaction μ takes place. To calculate $P_0(\tau)$ one can divide the interval [t,t+τ] in K partial intervals $\epsilon = \frac{\tau}{K}$ of equal size. The probability for no reaction in the first interval [t,t+ϵ] is now

$$\prod_{\mu=1}^{M} 1 - h_\mu c_\mu \epsilon = 1 - \sum_{\mu=1}^{M} h_\mu c_\mu \epsilon.$$

1.	initialisation	t=0, set initial particle numbers
2.	calculate	$a_i = h_i c_i$for all i =1,...,M
3.	τ_i	generate for all reactions the corresponding τ_i according to an exponential distribution using the *First Reaction Method* $\rightarrow \tau_i = \frac{1}{a_i} \ln \frac{1}{r}$
4.	reaction	execute the reaction with the lowest τ_i and adjust particle numbers
5.	time step	t=t+τ_i
6.	loop	go to step 2

Table 2.3: Procedure of the SSA using the *First Reaction Method*

This is also true for all partial intervals K:

$$P_0(\tau) = \left(1 - \sum_{\mu=1}^{M} h_\mu c_\mu \epsilon\right)^K = \left(1 - \sum_{\mu=1}^{M} h_\mu c_\mu \frac{\tau}{K}\right)^K.$$

If one now forms the limit for $K \rightarrow \infty$, one obtains:

$$P_0(\tau) = e^{-\sum_{\mu=1}^{M} h_\mu c_\mu \tau}.$$

This leads to the wanted density function $\mathrm{P}(\tau, \mu)$

$$P(\tau, \mu) = h_\mu c_\mu e^{-\sum_{\mu=1}^{M} h_\mu c_\mu \tau}.$$

So $P_\nu(\tau) = e^{-a_\nu \tau} a_\nu d\tau \, (with \, a_\nu = h_\nu c_\nu)$ is the probability that now the reaction ν is happening in the interval $[t+\tau, t+\tau+d\tau]$ and before that nothing. The reaction R_ν for which the probability $P_\nu(\tau)$ is the highest, is the next to occur. The *First Reaction Method* is now calculating the next occurring reaction, updating the particles numbers and starting with the next iteration. The time τ_i of the single reactions are calculated by the inverse function of $P_\nu(\tau)$:

$$P_\nu^{-1}(\tau) = \tau_\nu = \frac{1}{a_\nu} \ln \frac{1}{r},$$

where r is a uniformly distributed random number.
This is resulting in a linear time complexity.

The *First Reaction Method* works fine, but it is very time consuming. All a_i-values are recomputed in step two, although the value has not changed for some reactions. The *First Reaction Method* and the *Direct Method* differ in the way how the random pair (τ, μ) is calculated from the joint probability density function $P(\tau, \mu)$. Schwehm [2001] describes the difference as following:

For each reaction in the *Direct Method* a probability is computed by multiplying the rate constant of each reaction with the concentration of its substrates. Then a

	deterministic modeling	stochastic modeling
	rates	probabilities
	concentrations	particle numbers
	One set of starting values → one solution	One set of starting values → different solutions
⊕	fast good for metabolic processes	more realistic: pays respect to small volumes, heterogeneity, instabilities low particle numbers → fluctuations good for signal-transduction and gene expression
⊖	demands high particle numbers	computationally expensive

Table 2.4: Characteristics of deterministic and stochastic models

random number is used to perform a roulette-wheel selection according to the relative probabilities of all reactions, and a second random number determines the execution time used for this reaction. The *Direct Method* used two random numbers for each reaction selection.

The *First Reaction Method* computes, as described above, for each reaction (using one random number for each reaction) a tentative execution time. Then the reaction with the smallest execution time is selected. This method uses one random number for each reaction and iteration. This leads to the following conclusion: The *First Reaction Method* requires as many reaction numbers as there are reactions, the *Direct Method* on the other hand demands only two numbers. If the number of reactions exceeds two, the *Direct Method* is more efficient. The *First Reaction Method* has the advantage of being easier to implement.

In 2000 Gibson & Bruck have introduced an improvement, the *Next Reaction Method* [Gibson and Bruck, 2000]. They introduced a dependency graph, to capture the relations between single reactions and made an update only for those variables which have really changed. To achieve this, only used random numbers are newly generated the others are reused by transferring the a_i-values to the changed time scale. By this the complexity is reduced from linear to logarithmic. However, Cao et al. [2004a] have just recently published a comparison between the *Direct Method* and the *Next Reaction Method* and claimed the *Direct Method* to be for all but a very specialized class of problems much more efficient than the *Next Reaction Method*.

Approximative Algorithms and Hybrid Methods At this point it has to be clarified that there is no strict nomenclature referring to approximative and hybrid algorithms. For this reason I would like to make my own definition to simplify the further discussion. Let us define approximative algorithms as the closing gap between stochastic and deterministic approaches. An approximative algorithm does not predict a single reaction stochastically, but several events at once. However, in contrast to determinism the results can still differ if one compares single experiments, this is due to the fact, that random numbers are still used. Hybrid methods use now stochastic and/or approximative and/or deterministic algorithms together, partitioning the reactions by specific rules into one of the categories and calculate the events per time. COAST is a hybrid modeling tool using a stochastic, approximative and deterministic algorithm.

τ-leap method One approximative algorithm is the *τ-leap method* [Gillespie, 2001, Gillespie and Petzold, 2003], which fits the regime of intermediate particle numbers quite well. The key idea of this method is to determine time-intervals of length τ (so called τ-leaps), in which the number of reactions is so small that the propensity functions (reaction probabilities) are assumed to be approximately constant. By doing so, all reaction probabilities are formulated in terms of Poisson-distributions, and the length of these τ-leaps is computed dependently on an error control parameter. Note that the assumption of approximative constant reaction probabilities allows for the successive computation of reaction numbers in the different reaction channels. It also facilitates the description of reactions of higher order (e.g. $2\,X \rightarrow P$ or $X + Y \rightarrow P$) as independent probability experiments with identical distributions.

Beside of this there are also some negative aspects. In the *τ-leap method* all reaction probabilities are formulated in terms of Poisson-distributions, which are binomial-distributions limited by definition to infinitely large particle numbers and infinitely small reaction probabilities [Giri, 1974, p. 65]. Hence, the usage of Poisson-distributions does not fit the description of reaction channels with small particle numbers. Additionally the usage of Poisson distributions can lead to negative particle numbers, this led to the development of versions of the *τ-leap method* based on the binomial distribution [Tian and Burrage, 2004, Chatterjee et al., 2005]: the so-called binomial leap methods. The number of steps necessary for evaluating Poisson and binomial-distributions is asymptotically, proportional to the number of particles (or equivalently: from the expectation). In contrast the costs for the evaluation of a Gaussian-distribution or of the deterministic reaction kinetics are constant, so the latter models must be advantageous for large particle numbers.

Thus, all the algorithms mentioned so far are well adapted to a certain range of particle numbers, but not for the entire range from low to high numbers. Consequently, algorithms have been developed, which use different levels of modeling for

the different particle numbers. For example there exist some algorithms, that are based on fixed partitions of the system into slow and fast reaction channels. With this combination, slow reactions are treated by the *First Reaction Method*, and fast reactions are treated either by: deterministic reaction kinetics [Haseltine, 2002, Kiehl et al., 2004, Takahashi et al., 2004]; by Langevin equations [Haseltine, 2002]; or by random variables distributed according to the probability density functions at quasi stationary state [Rao and Arkin, 2003]. However, these partitioning methods require direct intervention of the modeler to partition the system into reaction sets covering different time and concentration ranges. Thus these algorithms cannot be appropriate for the simulation of systems with strongly fluctuating particle numbers (e.g. the Oregonator which will be discussed in Section 4.1.2). Furthermore, fixed partitioning is not suitable for systems with heterogeneously distributed substrates, which is especially problematic if applied to reaction-diffusion models.

Maximum Reaction Time Method Another modeling approach is the *Maximum Reaction Time Method* [Puchalka and Kierzek, 2004]. It describes slow reactions by the *Next Reaction Method* of Gibson and Bruck, and fast reactions by the *τ-leap method*. The partitioning into slow and fast reaction channels is performed automatically in each time step by criteria depending on two error control parameters. A third error control parameter is the value of the maximum time step.

The automatic partitioning makes the *Maximum Reaction Time Method* approach very interesting. However, there remain some problems. For example, it is very difficult to define appropriate values for the error control parameters. To show this, let us consider the influence of the error control parameter r, which defines a threshold value for the treatment by the *τ-leap method*. In a system with M reaction channels, the *τ-leap method* is only applied to a reaction channel μ if

$$r < f_\mu := \frac{a_\mu}{\sum_{\nu:=1}^{M} a_\nu}\,, \qquad (\mu \in \{1,..,M\}) \tag{2.4}$$

where a_μ is the propensity function of the reaction channel μ. Hence, for constant value a_μ, f_μ gets smaller if the number of reaction channels gets larger (The most simple case: If all a_ν have the same value, then $f_\mu = 1/M$). Thus, for constant a_μ and r, it is more likely that the reaction channel R_μ is treated by the *τ-leap method* if it is embedded in a small system than in a large system, which does not make too much sense. Thus, r depends on the number M of reaction channels.

Furthermore, let us consider the system $A+B \rightleftharpoons C+D$ and let us assume that all particle numbers $\#_A$, $\#_B$, $\#_C$, and $\#_D$ are proportional to a scaling factor z. The system can be split into the two single reactions $A+B \to C+D$ and $C+D \to A+B$ with corresponding values $f_1 := \frac{a_1}{a_1+a_2}$ and $f_2 := \frac{a_2}{a_1+a_2}$. Consequently, the reaction rates $a_1 := k_1 \#_A \#_B$ and $a_2 := k_2 \#_C \#_D$, are proportional to z^2, but $f_1 := \frac{a_1}{a_1+a_2}$ and $f_2 := \frac{a_2}{a_1+a_2}$ (cf. Equation (2.4)) are independent of z. Hence, it can happen

that $f_1 < r$ for all z so that $A + B \rightarrow C + D$ is always treated by the *next reaction method*. However, for large z, the propensity a_1 can reach arbitrarily large values.

These two simple examples show that the error control parameter r depends on at least two quantities: the number of reaction channels, and the number of particles, where the latter can fluctuate strongly during a simulation. Analogous considerations are applicable for other error control parameters. Thus, the search for optimal values of error control parameters is quite a difficult task which should, in my opinion, not be left to the user.

A further problem may be that all 'slow' reaction channels are only evaluated by the *Next Reaction Method*, which evaluates each reaction channel for time intervals corresponding to the mean time between two reactions. However, it can happen that the mean time between two reactions in a 'slow' reaction channel is of the same order of magnitude as the mean time between two reactions of a fast reaction channel. In this case, the fastest slow reaction dictates the size of the time steps; so that, on average, only a few fast reactions occur in each time step. Hence, the gain obtained from the *τ-leap method* is minimized.

probability-weighted Dynamic Monte Carlo method Another approximative approach is the *probability-weighted Dynamic Monte Carlo method* (PW-DMC) ,published by Resat et al. [2001]. In this method, reactions with large probabilities are allowed to occur in "bundles", which means that a single Monte Carlo step corresponds not only to a single reaction, but to several reactions in the same channel. As a consequence, the reactions follow a completely different statistics in each time step than in the original model. Suppose, for example, that there are two reaction channels with similar reaction probabilities. Then, each PW-DMC time step – maybe given by hundreds or thousands of reactions in one channel, but no reaction in the other – represents a very unlike event in the real reaction system. As a consequence, PW-DMC particularly leads to larger fluctuations than an exact algorithm. Resat et al. [2001] argue that this statistical error significantly cancels out if many simulation trajectories are averaged, which is true for stable dynamical systems, but not necessarily true for unstable dynamical systems. Furthermore, it does not make much sense to reduce the description of stochastic model to the mean of many trajectories. Such an average can more efficiently be computed by deterministic reaction kinetics. Instead of this, one has to try to also reflect the fluctuations correctly.

Date of reception	Author	Content	Reference
Jan.1976	Gillespie	*Direct Method,First Reaction Method*	Gillespie,1976
Feb.1977	Gillespie	*Direct Method, First Reaction Method*	Gillespie,1977
Sep.1999	Gillespie	The chemical Langevin equation	Gillespie,2000
Oct.1999	Gibson & Bruck	*next reaction method*	Gibson,2000
Dec.2000	Gillespie	*Poisson τ-leap Method,estimated midpoint method*	Gillespie,2001
Jun.2002	Haseltine & Rawlings	partition system: stochastic & deterministic	Haseltine,2002
Aug.2002	Rao & Arkin	quasi-steady state assumption to Gillespie algorithm	Rao,2003
Dec.2002	Kiehl	hybrid simulation:*direct method* & deterministic	Kiehl,2004
Jan.2003	Rathinam &Cao & Petzold &Gillespie	*implicit τ-leaping method*	Rathinam,2003
Apr.2003	Takahashi& Tomita	hybrid simulation: *Next Reaction Method* & deterministic	Takahashi,2004
May 2003	Gillespie & Petzold	improved leap-size selection	Gillespie,2003
Oct.2003	Burrage & Tian	A Multi-scaled Approach: SSA, *τ-leap* & Euler-method	Burrage, 2004
Aug.2003	Puchalka & Kierzek	*Maximal Timestep Method*: *next reaction method* & *τ-leap-method*	Puchalka,2004
Mar.2004	Cao & Petzold	*Optimized Direct Method*	Cao,2004a
Apr.2004	Tian& Burrage	*Binomial τ-leap Method*	Tian,2004
Jul.2004	Cao & Gillespie &Petzold	a theoretical improvements	Cao,2005
Dec.2004*	Cao & Petzold & Gillespie	Stability of leaping methods	Cao,2004
Jan.2005	Chatterjee	another *Binomial τ-leap Method*	Chatterjee,2005

Table 2.5: *This tabel gives a chronological overview of important publications on reaction algorithms for biological modeling, sorted by date of reception. *:date of availability*

2.5 Existing Implementations

The increasing availability of data and the complexity of cellular systems have motivated several programmers to provide integrative support to biology, after all: the overall goal of computational cell biology is to enable cell biologists to build and exercise predictive models of cellular processes. There are several tools for the simulation of dynamic biochemical systems available using the described reaction and diffusion algorithms. They are mainly freeware and can be downloaded from the authors websides. The Table 2.6 gives an overview on some of the described tools. The list is not exhaustive. It shows some important products sorted by the year of its first release. Most of the named simulators are still under maintenance, so there might have been some improvements, that are not included in this thesis. The focus here will be the implementation of algorithms on reaction and diffusion. Further details on the software can be obtained directly from the distributors.

The simulators can be in general classified by two characteristics: if they can model spatial aspects and by the main strategy to simulate (stochastic or deterministic). *GEPASI* is for example a non-spatial, ODE based tool for complex biochemical reaction pathways and therefore deterministic simulator; spatial information is not included (cf. Appendix C.1). Non-spatial deterministic simulators are typically ODE-solvers applied to mass action equations. The simulated space is just one entity and stochastic effects are not simulated. This of course makes *GEPASI* fast but sets its limitations.

Like *GEPASI NEURON* is also a deterministic simulator and designed to simulate electrophysiological behavior of single neurons using ODEs (cf. Appendix C.1).

Using ODE-solver is not without any problems. If systems includes both very fast and very slow dynamics, that is, some reactions are much faster than the others, the system is called stiff [Haavisto, 2004]. Stiff systems are hard to simulate since the fast dynamics require for short step size and the slow dynamics increase the total simulation time interval. Using a small stepsize, the simulation of the whole process becomes very slow. Consequently, some numerical algorithms are developed especially for the simulation of this kind of systems.

An example representing the fraction of stochastic spatial simulators is *MCell* (cf. Appendix C.1). The authors describe it as a "general Monte Carlo simulator of cellular physiology". *MCell* captures stochastic fluctuations seen with small numbers of particles and models diffusion by simulating Brownian random walk.

An exceptional simulator is *StochSim* (cf. Appendix C.1). It is using a very own reaction algorithm. This stochastic simulator was developed by Carl Firth (formerly known as Morton-Firth) back in 1998 as a biochemical simulator - simulating complex

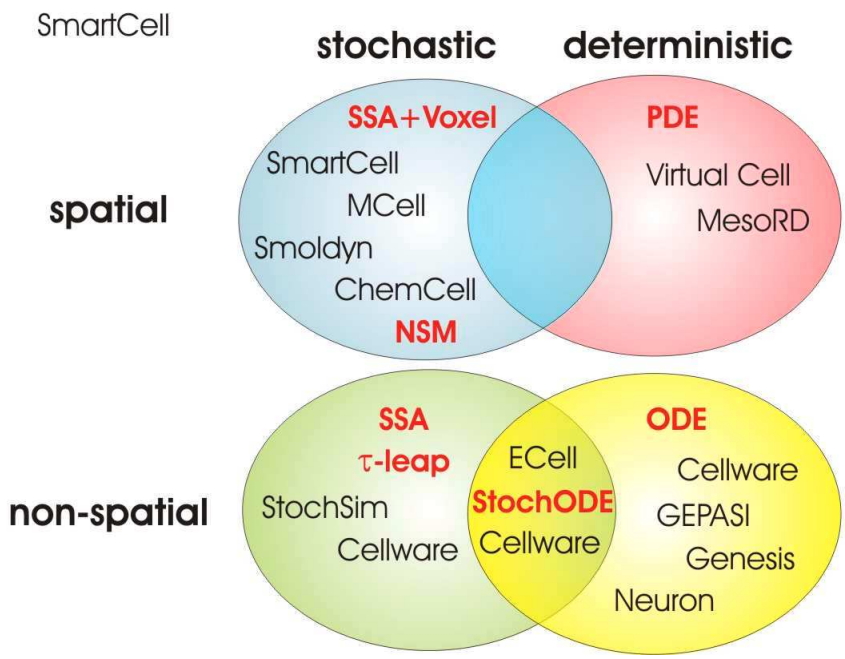

Figure 2.2: *Characterisation of existing cellular simulators; red: algorithms, black: implementations*

stochastic signaling pathways in bacterial chemotaxis. Single molecules are treated as single objects or intracellular automata. *StochSim* is capable of handling multistate molecules. For small numbers of reactions and single state molecules it is slower than SSA, but in other cases it is much faster and more accurate. Gillespie's algorithm cannot identify molecules as individual elements, their states, positions and velocities within the reaction volume cannot be followed over time and multistate molecules cannot be represented. At each time step, two molecules are picked and a random number generator is used to decide, if a reaction occurs or not using a lookup table of probabilities of all possible reactions. Since version 1.2 *StochSim* can model in two dimensions with squares forming the tessellation. Since v1.4 also triangles and hexagons can be used but there are no representations of cellular compartments. Speed gained by look up tables for reactions.

Cellware is a relatively new tool first released in 2004. It uses several reaction algorithms. One own development is *StochODE*, which is solving ODE's plus an external noise term; therefore *StochODE* is a solver for SDE's [Dhar et al., 2004]. Others used are NRM, SSA, tau-leap and several ODE-solver. Diffusion is not modeled, although simple compartments are represented. *Cellware* can only use one algorithm at a time. The much older *E-Cell* (first appearance in 1996) is using a hybrid approach. Parts

of the reactions are modeled using Gillespie's SSA while others use ODE's. Speed and accuracy are combined to model the stochastic behavior of -for example- gene expression. Like *Cellware E-Cell* is not able to model diffusion.

Some simulators especially the newer ones model diffusion by either random walk or partial differential equations and paying respect to spatial aspect of the cell. This has been encouraged by confocal and two-photon excited fluorescence microscopy, that permit investigators to study the structure and dynamics of living cells with submicrometer threedimensional spatial resolution and with time resolutions as fast as milliseconds [Slepchenko et al., 2002].

With *Virtual Cell* (VCell) Schaff Schaff et al. [1997] introduced a simulation tool, that uses the *finite element method* (FE) to solve reaction diffusion PDE's if a spatial resolution is demanded; otherwise ODE's are taken (cf. Appendix C.1). In the FE-approach the volume is divided in subvolumes and for each volume one assumes well-mixed conditions. Differential equations , which describe mass action kinetics are used to compute fluxes between and reaction rates within each voxel. The problem is that with realistic cellular structure, the grid has to be very fine or irregular in shape. In the first case, the finer the grid, the higher the computational cost, in the later the grid itself becomes a computational problem. The less voxels are taken, the worse the assumption becomes that a voxel represents a homogeneous space. *Virtual Cell* represents a typical deterministic simulator, which can pay respect to spatial aspects, but is not able to reflect the influence of stochastic events/noise. By downsizing the finite subvolumes the effects of noise are amplified, because the molecule numbers in each subvolume are getting smaller than when they were taken as whole (Bhalla [2004]).

In September 2004 Andrews and Lipkow introduced *Smoldyn* [Lipkow et al., 2005]. The name is derivated from "Smoluchowski dynamics". This tool is designed to model chemical reactions networks especially to look at the effects of cellular architecture and molecular crowding on signal transduction pathways. Each molecule is treated as a single point (centers of mass), so there is no volume and no inertia. The molecules diffuse freely in the test volume. All particles have a given binding radius. If two molecules get close enough, so the distance is smaller than the binding radius, a reaction occurs. It has to be emphasized that the binding radius and the sum of the molecule radii are not the same. Because of the fact that most reactions occur at a slower rate because of a reaction activation energy, the sum of the molecular radii is replaced by a smaller binding radius. For reversible reactions Andrews defined a debinding radius, which is totally artificial, but helped to prevent two molecules from immediately recombine after just being split. Steric interactions between particles that cannot react, are ignored. The leap length of a particle is derived from Fick's law $s_B = \sqrt{2\,D\,t}$. A problem is the calculation for bimolecular reactions. If two par-

ticles A and B were moving, the question is, if their distance during the last δt has ever been smaller than the reaction radius. Because of the computational complexity of answering that question for multi particle systems. Andrews et al. simplified the concept by only looking at the final positions of all particles and ckecking if any distances fall below a binding radius. The accuracy now depends on the setting of δt.

Two simulators presented in 2004 are *SmartCell* and *MesoRD* (cf. Appendix C.1). They are both using the *Next Subvolume Method (NSM)* by Elf and Ehrenberg [2004] to model diffusion. SmartCell was developed to simulate diffusion-reaction frameworks in a whole cell-context [Ander et al., 2004]. Because of the fact that the distribution of entities can be crucial for certain processes, SmartCell is using the idea of deviding the modeling space into subvolumes and was at first using the NRM of Gibson and Bruck to model diffusion and reaction but recently changed to the *Next Subvolume Method*. This makes *SmartCell* a spatial stochastic simulator. Within the single volume elements the particles are assumed to be equidistributed, so the stochastic algorithm can be used. *SmartCell* does not simulate excluded volume effects because the simulated particles have the volume 0. MesoRD was using the NSM from the beginning [Hattne et al., 2005]. The NSM scales logarithmically with the number of subvolumes, the NRM by Gibson and Bruck also, but memory requirements and operations per second are higher. Gillespies SSA on the other hand scales linearly and is therefore much slower.

What to expect In the last two subsection I have presented several approaches to model cellular systems. In Section 2.4.2 the most important reaction algorithms were described. By applying them to a spatial grid as demonstrated by some implementations in Section 2.5 they can be applied to simulate diffusion as well.

The main problems of the existing algorithms are their limitations. No algorithm alone is capable of performing efficient and accurate simulations. If they are accurate like the *First Reaction Method* they lack of speed and if they are fast like a deterministic approach they do not reproduce stochastic fluctuations anymore. And not only that, but τ*-leap methods* using Poisson distributions are also based on the wrong assumption when applied to small particle numbers. The hybrid methods are the logical consequence, but the solutions so far do not cover the whole spectrum of occurring particle numbers. Either they only use an exact approach and an approximative method like the *maximum reaction time method* or they totally blind out intermediate particle numbers like the hybrid method by Takahashi et al. [2004].

What is needed is an algorithm covering small, intermediate and high particle numbers and simulating them as accurate as necessary and as fast as possible. The limited partitioning is not the only problem the hybrid methods so far have. The partitioning is not very intuitive. In the approaches by Haseltine [2002], Kiehl et al.

Simulator	year of 1st release	reaction algorithms	diffusion algorithms	subvolumes	compartments	3D	area of application	reference
E-Cell	1996	SSA, ODE, NRM	∅	∅	√	∅	gene regulatory networks, signal transduction, metabolic networks	Tomita,1999
VCell	1997	ODE's and /or PDEs	PDE's with finite volume method	√	√	√	distribution and dynamics of intracellular biochemical processes	Schaff,1997
MCell	1997	bimol.react.: collision unimol. react.: similar to SSA	3D Random Walk	∅	√	√	neurotransmission, signal transduction, trasmembrane flux simulations in 3D	Stiles,1998
Smoldyn	2003	bimol.react.: collision unimol. react.: similar to SSA	3D Random Walk	∅	√	√	signal transduction, general biochemical simulator	Andrews,2004
Cellware	2004	ODE, *Direct Method*, NRM, StochODE, explicit τ-leap	∅	∅	√	∅	signal transduction, gene regul. networks metabolic networks	Dhar,2005
SmartCell	2004	NSM	NSM	√	√	√	diff. & react.networks	Ander,2004
MesoRD	2004	stoch. mode: NSM deter. mode: PDE	stoch. mode: NSM deter. mode: PDE	√	√	√	stoch. & determ. sim. of chemical reactions and diffusion in 3D	Hattne,2005

Table 2.6: *A Selection of present cellular simulators with respect to how they model reaction and diffusion, how they pay respect to spatial aspects and what is their main area of application.*

[2004] and Takahashi et al. [2004] the user has to divide the reaction channels into the different modeling classes. This is very inconvenient and furthermore inappropriate if the system is oscillating or at least one species would have to be reclassified as 'slow' or 'fast'. A complex system like the *Oregonator* (cf. Section 4.1.2) is a good example for this.

So far only the *Maximum Reaction Time Method* and the *probability weighted Dynamic Monte Carlo method* provide an automatic shifting between a limited amount of modeling levels (cf. Table 3.3). However, they use more than one error parameter and they are not very intuitive.

Rao et al. [2002] made an important and for this thesis crucial statement regarding the existing reaction algorithms in the journal "Nature" : "Although a few new strategies have been proposed to increase the efficiency of the Gillespie algorithm (tau-leap and NRM), there are currently no satisfactory approaches simulating processes concurrently across multiple scales of time, space and concentration. An alternative approach is to separate timescales explicitly and reduce the model by singular perturbations. Yet another approach is to construct hybrid models involving continuous and discrete representations. Both these approaches require direct intervention by the modeler - a cumbersome and sometimes impossible task. The long-term goal is to develop algorithms that do this both automatically and adaptively."

With COAST I am confident to present in the following chapter an algorithm that fulfills this demand. This algorithm covers exact stochastic, approximative and deterministic cases. However, at the same time its accuracy is only defined by one single parameter α.

3 COAST for Reaction and Diffusion

So far I have introduced the problem and main goal of this thesis in the first chapter. In the last chapter I gave an overview on existing methods to model reaction and diffusion and their advantages and disadvantages for modeling cellular processes. At the end I presented some existing simulators in this scientific area.
In the following chapter I will introduce the *Controllable Approximative Stochastic reaction-algorithm* (COAST). COAST is a hybrid algorithm using three levels of modeling and is controlled by one single error parameter α. This chapter explains step by step the algorithm for reaction problems and its application to diffusion scenarios.

3.1 Concept of COAST

Gillespie's approach (cf. Section 2.4.2) answers two important questions:

- Which reaction will occur next?
- When is the reaction going to occur?

With COAST, the questions have slightly changed:

- Which reaction will define the next time step?
- How long is this time step?

COAST uses some ideas of the *maximum reaction time method* - particularly the automatic partitioning of the reaction channels into classes with different levels of modeling (cf. Section 2.4.2). COAST allows for all reaction channels to perform several reactions within a single time step. Within this given time step, the different reaction channels are evaluated successively using three different levels of modeling:

- an exact stochastic level based on Gillespie's *First Reaction Method* for small numbers of particles,
- an approximative stochastic modeling by Gaussian-distributions for intermediate particle numbers,
- and the deterministic reaction kinetics for large numbers of particles.

Therefore, the partitioning into three levels of modeling is done automatically in every time step.

In contrary to the *First Reaction Method*, the stochastic method used in COAST allows for more than one reaction to take place within a given time step.

As previously mentioned, the subdivision of the reaction channels into the three different modeling levels also depends on a single error control parameter α. This control parameter α is chosen so that the error of COAST is always smaller than $(\alpha \cdot 100)\,\%$ of the value of an exact algorithm. In Section 3.10 I will give some firther information on the different errors that are estimated by α. Furthermore, in practice, I show that the error in simulations is usually much smaller than the upper bound given by this parameter. So, an α-value of 0.05 would mean that one allows an error of 5% in all calculations.

Thus, the algorithm can be controlled by the choice of $\alpha \in [0, 1]$. This makes it easy to find an optimal trade off between accuracy and performance for a given simulation system.

In the next section I present the mathematical background supporting COAST. Section 3.3 describes the single steps of the COAST-algorithm.

3.2 Derivation of the Fundamentals

In contrast to other existing hybrid algorithms, α offers a precise method to determine when to switch from one modeling level to another. The usual way to apply the *First Reaction Method* in order to calculate which is the next reaction and when it is going to occur is by evaluating binomial distributions. This is computationally expensive for more than one occurring reaction since several random numbers have to be chosen. By using Gaussian distributions one can compute random numbers with less computational effort. It is a well known property of binomial distributions to converge toward a Gaussian distribution if the size of the set increases.

For the algorithm, two essential problems must be solved: Firstly, one has to determine time spans in which the particle numbers and, thus, the reaction probabilities are nearly constant. Secondly, one needs - at least for intermediate and large particle numbers - methods which allow to compute the number of reactions efficiently without too large errors.

3.2.1 Methods

In this paragraph I will derive the necessary transition criteria for the three applied regimes of COAST. The criteria result in two requirements formulated in Equation

(3.17) and Equation (3.18) (and Equations (3.19) and (3.20) respectively).

Exact stochastic model: If the particle numbers are low the reactions are calculated by a modified *First Reaction Method*, where we allow more than one reaction until the reaction probabilities change by more than $\alpha \cdot 100\,\%$. This first regime is called Σ.
Approximative stochastic model: Since the reaction probabilities (propensities) are nearly constant, the number of reactions during such a time step can approximatively be described by binomial distributions and, thus, for sufficiently large particle numbers by discrete Gaussian-distributions. This defines the second regime Γ. A critical question is defining the point of transition between Σ (nearly binomial-distributed) and Γ.

I will now explain when it is appropriate to switch from a binomial distribution to a Gaussian distribution with an error of α.

Let $P_B(k; N, p)$ be the probability for k events given by a binomial distribution with parameters N and p

$$P_B(k; N, p) = \left(\begin{array}{c} N \\ k \end{array} \right) p^k \, (1-p)^{N-k} \, . \tag{3.1}$$

The expectation is $E := N \cdot p$ and the variance is $V := N \cdot p \cdot (1-p)$. In terms of a reaction system, $P_B(k; N, p)$ would be the probability for k reactions occurring with originally N particles in the system and the probability p for a single reaction to occur.
Further, let X be a standard normal variable and Z be a probability variable

$$Z := \text{Round}\left(\sqrt{V}\, X + E \right) \tag{3.2}$$

with the Round-procedure

$$\text{Round}\,(x) := \left\{ \begin{array}{cl} [\,x\,]+1\,, & \text{if } x - [\,x\,] \geq 1/2\,, \\ [\,x\,]\,, & \text{if } x - [\,x\,] < 1/2\,. \end{array} \right. \qquad ([\,x\,] := \max\{n \in \mathbb{Z}, n \leq x\,\}) \tag{3.3}$$

Let $P_G(k, N)$ be the probability of a "discrete Gaussian distribution" for $Z = k$ events with the same expectation E and variance V.
Equation (3.2) implies that the probability for $Z = k$ is

$$P_G(k, N) = \tfrac{1}{\sqrt{2\,\pi}} \int_{\frac{k-E-1/2}{\sqrt{V}}}^{\frac{k-E+1/2}{\sqrt{V}}} e^{-x^2/2}\, dx\,, \tag{3.4}$$

so that it must be shown that $P_B(k; N, p)$ and $P_G(k, N)$ are approximatively identical for large N, where, for a fixed value of the error control parameter α, $P_G(k, N)$ is a valid approximation for $P_B(k; N, p)$ for all $N > N_0(\alpha) \in \mathbb{N}$ if the supremum norm

$$\sup\Big\{|P_B(k;N,p) - P_G(k,N)|; 0 \le k \le N\Big\} < \alpha \qquad \text{for all } N \ge N_0(\alpha)\,. \tag{3.5}$$

This is a relatively simple approximation, other approaches like the DeMoivre Laplace limit theorem [Feller, 1970, 182pp] will possibly give better approximations with a positive impact on the algorithms performance.
The aim of this section is the derivation of an appropriate value $N_0(\alpha)$. To this aim, we will firstly prove three Lemmas. The value $N_0(\alpha)$ itself is the content of the theorem at the end of the section.

Lemma 1 *: If $|k-E| > \sqrt{\frac{V}{\alpha}}$, then $|P_B(k;N,p) - P_G(k,N)| < \alpha$ is fulfilled for all k with $|k-E| > \sqrt{\frac{V}{\alpha}}$.*
Proof: In accordance with Tchebycheff's inequality [deFinetti, 1974, p.172f.]

$$P_B\left(|Y-E| > \sqrt{\tfrac{V}{\alpha}}, N\right) < \alpha\,, \qquad P_G\left(|Z-E| > \sqrt{\tfrac{V}{\alpha}}, N\right) < \alpha\,. \tag{3.6}$$

Since all $P_B > 0$ and $P_G > 0$, $|P_B(k;N,p) - P_G(k,N)| < \alpha$ is, thus, always fulfilled for all k with $|k-E| > \sqrt{\frac{V}{\alpha}}$. □

In what follows, we will determine $N_B(\alpha)$ and $N_G(\alpha)$, so that for all $|k-E| < \sqrt{\frac{V}{\alpha}}$

$$D_B(k,N) \;:=\; |P_B(k;N,p) - \pi(k,N)\,| < \frac{\alpha}{2}\,, \qquad \forall N \ge N_B(\alpha)\,, \tag{3.7}$$

$$D_G(k,N) \;:=\; |P_G(k,N) - \pi(k,N)\,| < \frac{\alpha}{2}\,, \qquad \forall N \ge N_G(\alpha)\,, \tag{3.8}$$

where

$$\pi(k,N) := \frac{\exp\left(-\frac{(k-N\,p)^2}{2\,N\,p\,(1-p)}\right)}{\sqrt{2\,\pi\,N\,p\,(1-p)}}$$

Criterion (3.5) is then fulfilled for all $N > N_0(\alpha) := \max\{N_B(\alpha), N_G(\alpha)\}$.

Lemma 2 *: Let $|k-E| < \sqrt{\frac{V}{\alpha}}$. Then $D_B(k,N) < \frac{\alpha}{2}$ for all $N > N_B(\alpha)$ with*

$$N_B(\alpha) := \frac{1}{3\,\alpha\,p\,(1-p)}\,.$$

Proof: We replace the binomial coefficient in Equation (3.1) by the extension to Sterling's formula of Buchner [1951] . Therefore we define a function $\zeta(N,k)$ in the following way:

$$\zeta(N,k) \;=\; \ln\left(\frac{\binom{N}{k}}{Q(N,k)}\right) \quad \Rightarrow \quad \binom{N}{k} \;=\; e^{\zeta(N,k)}\, Q(N,k) \tag{3.9}$$

where

$$Q(N,k) \;=\; \sqrt{\frac{N}{2\,\pi\,k\,(N-k)}}\left(\frac{N}{k}\right)^{k}\left(\frac{N}{n-k}\right)^{N-k}. \tag{3.10}$$

Notice that k will be replaced by $\kappa \cdot N$.
This leads to

$$\begin{aligned}
P_B(k;N,p) &= e^{\zeta(N,k)}\sqrt{\frac{N}{2\pi k(N-k)}}\left(\frac{N}{N\kappa}\right)^{\kappa n}\left(\frac{N}{N(1-\kappa)}\right)^{N(1-\kappa)} p^{\kappa\ N}\,(1-p)^{N(1-\kappa)} \\
&= \frac{e^{\zeta(N,k)+\frac{1}{2}\ln\frac{1}{\kappa(1-\kappa)}+(\kappa N)\ln\frac{1}{\kappa}+N(1-\kappa)\ln\frac{1}{1-\kappa}+(\kappa N)\ln p+N(1-\kappa)\ln 1-p}}{\sqrt{2\pi N}} \\
&= \frac{e^{\zeta(N,k)-\frac{1}{2}\ln\frac{\kappa(1-\kappa)\ p(1-p)}{p(1-p)}-\kappa N(\ln\kappa-\ln p)-N(1-\kappa)(\ln(1-\kappa)-\ln(1-p))}}{\sqrt{2\pi N}} \\
&= \frac{e^{\zeta(N,k)-\frac{1}{2}\ln\frac{\kappa(1-\kappa)}{p(1-p)}+\ln\left(p(1-p)^{-\frac{1}{2}}\right)-\kappa N\ln\frac{\kappa}{p}-N(1-\kappa)\ln\frac{1-\kappa}{1-p}}}{\sqrt{2\pi N}} \\
P_B(k;N,p) &= \frac{\exp\left(\zeta(N,k)-\frac{1}{2}\ln(\frac{\kappa\,(1-\kappa)}{p\,(1-p)})-N\,(\kappa\,\ln(\frac{\kappa}{p})+(1-\kappa)\,\ln(\frac{1-\kappa}{1-p}))\right)}{\sqrt{2\,\pi\,N\,p\,(1-p)}}
\end{aligned} \tag{3.11}$$

To continue we perform a Taylor-expansion of the exponent in Equation (3.11). κ later on is replaced by $\frac{k}{N}$. Furthermore we neglect all terms of the order N^{-1}, which has a consequence the disappearance of the ζ-function. For the single parts of the exponent one obtains the following derivations:

$$\begin{aligned}
f(x) &= -\frac{1}{2}\ln\frac{\kappa(1-\kappa)}{p(1-p)} \\
f'(x) &= \frac{-1+2\kappa}{2\kappa(1-\kappa)} \\
f''(x) &= \frac{2\kappa^2-2\kappa+1}{2(\kappa-\kappa^2)^2} \\
f'''(x) &= \frac{3\kappa-3\kappa^2+2\kappa^3-1}{(\kappa-\kappa^2)^3}
\end{aligned} \tag{3.12}$$

$$\begin{aligned} f(x) &= -n\left(\kappa \ln \frac{\kappa}{p} + (1-\kappa)\ln\frac{1-\kappa}{1-p}\right) \\ f'(x) &= n\left(-\ln\frac{\kappa}{p} + \ln\frac{1-\kappa}{1-p}\right) \\ f''(x) &= \frac{-n}{\kappa(1-\kappa)} \\ f'''(x) &= \frac{n(1-2\kappa)}{\kappa^2(1-\kappa)^2} \end{aligned} \tag{3.13}$$

Since $(\kappa - p)$ is of the order $N^{-1/2}$, a Taylor-expansion of the exponent in Equation (3.11) results in:

$$\begin{aligned} P_B(k;N,p) &\approx \frac{\exp\left(\frac{-1+2p}{2p(1-p)}(\kappa-p) + \frac{1}{2}\frac{-N}{p(1-p)}(\kappa-p)^2 + \frac{1}{6}\frac{N(1-2p)}{p^2(1-p)^2}(\kappa-p)^3\right)}{\sqrt{2\,\pi\,V}} \\ &\approx \frac{\exp\left(\frac{-1+2p}{2p(1-p)}(\frac{k}{N}-p) + \frac{1}{2}\frac{-N}{p(1-p)}(\frac{k}{N}-p)^2 + \frac{1}{6}\frac{N(1-2p)}{p^2(1-p)^2}(\frac{k}{N}-p)^3\right)}{\sqrt{2\,\pi\,V}} \\ &\approx \frac{\exp\left(-\frac{1-2p}{2Np(1-p)}(k-Np) - \frac{1}{2}\frac{N}{pN^2(1-p)}(k-Np)^2 + \frac{1}{6}\frac{N(1-2p)}{N^3p^2(1-p)^2}(k-Np)^3\right)}{\sqrt{2\,\pi\,V}} \\ &\approx \frac{\exp\left(-\frac{(k-N\,p)^2}{2\,V} + \frac{1-2\,p}{6\,V^2}(k-N\,p)^3 - \frac{1-2\,p}{2\,V}(k-N\,p)\right)}{\sqrt{2\,\pi\,V}} \end{aligned} \tag{3.14}$$

By using $1-e^{-x} \approx x$ $(e^{-x} \approx 1-x)$ for small x , Equation (3.14) can be reformulated to:

$$P_B(k;N,p) \approx g(k) := \frac{e^{-\frac{(k-Np)^2}{2V}}}{\sqrt{2\pi V}}\left(1 + \frac{1-2p}{6V^2}(k-Np)^3 - \frac{1-2p}{2V}(k-Np)\right).$$

The upper bound of $D_B(k,N) :\approx |g(k) - \pi(k,N)|$ in the relevant interval $[E - \sqrt{\frac{V}{\alpha}}, E + \sqrt{\frac{V}{\alpha}}]$ is given by a local maximum (see Figure 3.1). Now we set $x = \frac{k-Np}{\sqrt{V}}$ to simplify calculations. This results in

$$D_B(x) \approx \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi V}}\left(\frac{1-2p}{6}\frac{x^3}{V} - \frac{1-2p}{2}\frac{x}{\sqrt{V}}\right)$$

and the first deviation

$$D'_B(x) \approx \frac{e^{-\frac{x^2}{2}}(1-2p)}{2\sqrt{2\pi}V}\left(-\frac{x^4}{3} + 2x^2 - 1\right).$$

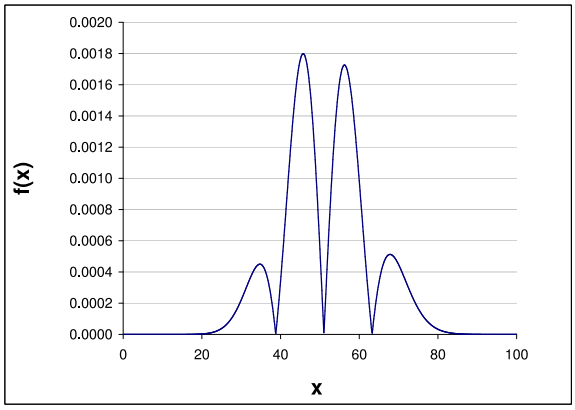

Figure 3.1: *The graph of the function* $f(x) := |P_B(x; N, p) - \pi(x, N)|$ *for* $p = \alpha = 0.02$ *and* $V = 50$ *(*$\Rightarrow N = 2551$*), where for* P_B *the approximative expression (3.14) is used. As implied by (3.6) , the interval relevant for the estimation of* $D_B(x, N)$ *is given by* $[E - \sqrt{\frac{V}{\alpha}}, E + \sqrt{\frac{V}{\alpha}}] = [1.02, 101.02]$.

This results in $x_0 \approx -\sqrt{3-\sqrt{6}}$. Hence, one can estimate

$$\begin{aligned} \max\{D_B(k,N)\} &\approx D_B(-\sqrt{3-\sqrt{6}}, N) \\ &\approx \frac{(1-2p)\sqrt{3-\sqrt{6}}\sqrt{6}\, e^{\frac{-3+\sqrt{6}}{2}}}{\sqrt{2\pi}V6} \\ &< \frac{1}{6V}, \end{aligned}$$

so that $D_B(k, N) < \frac{\alpha}{2}$ (cf. Equation (3.7)) is fulfilled for

$$N > \frac{1}{3\,p\,(1-p)\,\alpha}\,. \qquad \square$$

Lemma 3 *: Let* $|k - E| < \sqrt{\frac{V}{\alpha}}$*. Then* $D_G(k, n) < \frac{\alpha}{2}$ *for all*

$$N := \frac{1}{9\,\alpha^{2/3}\,p\,(1-p)}\,.$$

Proof: $P_G(k, N)$ (cf. Equation (3.4)) can be rewritten by a Taylor-expansion of

$f(x) = \exp(-x^2/2)$ at $x_0 = \frac{k-E}{\sqrt{V}}$ in the form:

$$\begin{aligned}
P_G(k,N) &= \frac{1}{\sqrt{2\,\pi}} \int_{\frac{k-E-1/2}{\sqrt{V}}}^{\frac{k-E+1/2}{\sqrt{V}}} \left(f(x_0) + f'(x_0)\,(x-x_0) + f''(x_0)\frac{(x-x_0)^2}{2} \right) dx \\
&= \frac{1}{\sqrt{2\,\pi}} \int_{\frac{k-E-1/2}{\sqrt{V}}}^{\frac{k-E+1/2}{\sqrt{V}}} e^{-\frac{(k-E)^2}{2V}} - 2\left(\frac{k-E}{\sqrt{V}}\right) e^{-\frac{(k-E)^2}{2V}} \quad \left(x - \frac{k-E}{\sqrt{V}}\right) \\
&\quad + f''(x_0)\,(x-x_0)^2\,dx \\
&= \frac{1}{\sqrt{2\,\pi}} \int_{\frac{k-E-1/2}{\sqrt{V}}}^{\frac{k-E+1/2}{\sqrt{V}}} e^{-\frac{(k-E)^2}{2V}} - 2\left(\frac{k-E}{\sqrt{V}}\right) e^{-\frac{(k-E)^2}{2V}} x + 2\left(x - \frac{(k-E)^2}{V}\right) e^{-\frac{(k-E)^2}{2V}} \\
&\quad + f''(x_0)\,(x-x_0)^2\,dx \\
&= \pi(k,N) + \frac{1}{\sqrt{8\,\pi}} \int_{\frac{k-E-1/2}{\sqrt{V}}}^{\frac{k-E+1/2}{\sqrt{V}}} f''(x_0)\,(x-x_0)^2\,dx\,,
\end{aligned}$$

By neglecting terms of the order $\mathcal{O}(n^{-1})$, one thus obtains for Equation (3.8)

$$\begin{aligned}
D_G(k,n) &= \left| \frac{1}{\sqrt{8\,\pi}} \int_{\frac{k-E-\frac{1}{2}}{\sqrt{V}}}^{\frac{k-E+\frac{1}{2}}{\sqrt{V}}} f''(x_0)\,(x-x_0)^2\,dx \right| \\
&\leq \frac{\max\{|f''(x_0)|\}}{\sqrt{8\pi}} \int_{x_0-\frac{1}{2\sqrt{V}}}^{x_0+\frac{1}{2\sqrt{V}}} (x-x_0)^2\,dx
\end{aligned}$$

now we use $k = x_0\sqrt{V} + E$

$$\begin{aligned}
&\leq \frac{\max\{|f''(x_0)|\}}{\sqrt{8\pi}} \left| \frac{x^3}{3} - x^2 x_0 + x x_0^2 \right|_{x_0-\frac{1}{2\sqrt{V}}}^{x_0+\frac{1}{2\sqrt{V}}} \\
&\leq \frac{\max\{|f''(x_0)|\}}{\sqrt{8\pi}} \frac{1}{12\,V^{3/2}} \\
&\leq \frac{1}{\sqrt{8\,\pi}\,12\,V^{3/2}}\,. \qquad (3.15)
\end{aligned}$$

Notice that $|f''(x_0)|$ has three extrema (at $x_1 = -\sqrt{3}, x_2 = 0$ and $x_3 = \sqrt{3}$) and $\max\{|f''(x_0)|\}$ is at position $x = 0$ with $|f''(0) = 1|$. Hence, $D_G(k,N) < \frac{\alpha}{2}$ is fulfilled for all k if

$$\frac{\alpha}{2} > \frac{1}{\sqrt{8\pi}12V^{\frac{3}{2}}}.$$

Because $V = N\,p\,(1-p)$ we obtain

$$\frac{\alpha}{2} > \frac{1}{\sqrt{8\pi}12(N^{\frac{3}{2}}p^{\frac{3}{2}}(1-p)^{\frac{3}{2}})}.$$

By solving for N we receive

$$\begin{aligned} N^{\frac{3}{2}} &> \frac{1}{\sqrt{8\pi}6\alpha p^{\frac{3}{2}}(1-p)^{\frac{3}{2}}} \\ &> \frac{1}{(288\,\pi)^{1/3}\,\alpha^{2/3}\,p\,(1-p)}. \end{aligned}$$

By approximating the existing quotient the resulting inequality is

$$N > \frac{1}{9\,\alpha^{2/3}\,p\,(1-p)} \qquad \square \tag{3.16}$$

Theorem 1 *: sup*$\Big\{|P_B(k;N,p) - P_G(k,N)|; 0 \le k \le N\Big\} < \alpha$ *for all* $N > \frac{1}{3\,\alpha\,p\,(1-p)}$.
Proof: Due to Lemma (1), the estimations can be restricted to $|k-E| < \sqrt{\frac{V}{\alpha}}$. On the other hand, Lemma (2) and Lemma (3) lead to the result, that $\sup\Big\{|P_B(k;N,p) - P_G(k,N)|; 0 \le k \le N\Big\} < \alpha$ (cf. Equation (3.5)) for all $|k-E| < \sqrt{\frac{V}{\alpha}}$ if

$$N \ge N_0(\alpha) := \max\{N_B(\alpha), N_G(\alpha)\} = \frac{1}{3\,\alpha\,p\,(1-p)}\,. \qquad \square$$

So we can conclude that the binomial distribution $P_B(k;N,p)$ and the Gaussian distribution $P_G(k,N)$ are referred to be the same with respect to the error α for all

$$N \ge \frac{1}{3\,\alpha\,p\,(1-p)}\,. \tag{3.17}$$

Deterministic reaction kinetics: Furthermore, one can define the transition point between the approximative and the deterministic regime by applying similar considerations. It is a well know fact that for large particle numbers the statistic fluctuations can be neglected and we reach the regime of determinism [Ethier and Kurtz, 2005].

In the following text I will consider three possible ways to calculate the transition between the regimes Γ and Δ. The first one introduced is comparably crude, but based on the well known Tschebyscheff's inequality [deFinetti, 1974, p.172f.]. The calculated criterion (3.18) was used in the simulations of COAST. It is possible to estimate an earlier point of transition using more accurate approaches. These improvements could be used to improve the runtime results of COAST. The second approach uses the quantiles of the normal distribution and the third approximates the Gaussian distribution by another e-function.

To unify the three approaches one has to standardize the distributions. A random variable Y is given by the Gaussian distribution $Y \approx N\left(\mu, \frac{\sigma^2}{N}\right)$ where $\mu = p$ and

$\sigma^2 = p(1-p)$. Its distribution is transformed to a the standard normal distribution $\frac{(\overline{x}-\mu)\cdot\sqrt{N}}{\sigma}$ with $\overline{x} := \frac{1}{N}\sum y_i$.

Now we demand that the probability for a certain number of reactions being further away from the expected value than a given distance ϵ is $P(|Z| \geq \epsilon) \leq \alpha$.
First let us consider Tschebyscheff's inequality.

$$P(|Z| \geq \epsilon) \leq \frac{V}{N \cdot \epsilon^2}.$$

This results in $\alpha = \frac{V}{N \cdot \epsilon^2}$. With this we can give a definition for ϵ.

$$\epsilon = \sqrt{\frac{V}{N \cdot \alpha}} = \frac{\sigma}{\sqrt{\alpha \cdot N}}$$

The second step is to demand the deviation ϵ to be very small compared to the expected value μ.

$$\frac{\sigma}{\sqrt{\alpha \cdot N}} << \mu$$

This inequality has to be quantified to be useful. Because we only want one single error parameter we use α again to simplify the estimation:

$$\frac{\sigma}{\sqrt{\alpha \cdot N}} \quad < \quad \alpha \cdot \mu.$$

Squaring both sides results in

$$\frac{\sigma^2}{\alpha \cdot N} \quad < \quad \alpha^2 \cdot p^2.$$

Using $\sigma^2 = p(1-p)$ we can conclude

$$\frac{(1-p)}{\alpha^3 \cdot p} \quad < \quad N_{Tschebyscheff}. \tag{3.18}$$

With Equation (3.18) we have an estimation when to apply the deterministic instead of the Gaussian distribution due to the fact that the expected value used is the one given by the deterministic dynamics.

However, Tschebyscheff's inequality is relatively coarse and there are better estimations possible. The second approach is based on the quantile-function. We first define an ϵ, so that the probability for a value to be further away than ϵ from the expected value is less or equal the error α. $\Phi(\epsilon)$ is defined as the integral $\Phi(\epsilon) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\epsilon} e^{-\frac{t^2}{2}}\, dt$. This concludes to

$$2\,(1 - \Phi(\epsilon)) \quad = \quad \alpha.$$

So we can obtain

$$1 - \frac{\alpha}{2} \quad = \quad \Phi(\epsilon).$$

By applying the inverse function Φ^{-1}, we receive the following expression for ϵ

$$\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \quad = \quad \epsilon.$$

Again one has to demand ϵ to be much smaller than the expected value; i.e. we allow a relative error of $\alpha \cdot 100\%$ and therefor ϵ is set as $\alpha \cdot \mu$. It has to be emphasized that due to the standardization one now has to rescale ϵ with $\frac{\sigma}{\sqrt{N}}$

$$\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{N}} \quad < \quad \alpha \cdot \mu.$$

By using the fact $\mu = p$ the inequality changes to

$$\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \cdot \frac{\sigma}{\sqrt{N}} \quad < \quad \alpha \cdot p.$$

Finally we solve the inequality for N:

$$\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)^2 \frac{(1-p)}{\alpha^2 \cdot p} \quad < \quad N_{Quantiles}. \tag{3.19}$$

Using (3.19) is the most accurate way to calculate the point of transition. However handling the tabulated values for the Φ^{-1}-function can be complicated. Therefore it might be better to find an approximative solution which is our third approach to this problem.

$$P(|Z| \geq \epsilon) \quad = \quad 2\,(1 - \Phi(\epsilon))$$

Now we can apply the definition for Φ and obtain

$$\begin{aligned} &= \quad 2 \cdot \int_{\epsilon}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\, dx \\ &= \quad \frac{2}{\sqrt{2\pi}} \cdot \int_{\epsilon}^{\infty} e^{-\frac{x^2}{2}}\, dx. \end{aligned}$$

The factor in front of the integral is smaller than 1. The problem is the integral, it cannot be solved analytically. However, we can replace it by another larger integral of which we know the antiderivative [Wasserman, 2006, p.8].

$$\begin{aligned} P(|Z| \geq \epsilon) \quad &\leq \quad \int_{\epsilon}^{\infty} x\, e^{-\frac{x^2}{2}}\, dx \\ &\leq \quad \left[-e^{-\frac{x^2}{2}}\right]_{\epsilon}^{\infty} \\ &\leq \quad e^{-\frac{\epsilon^2}{2}}. \end{aligned}$$

This results in the following criterion for ϵ, because the described probability is supposed to be α

$$\epsilon = \sqrt{\ln \frac{1}{\alpha^2}}.$$

Like in the two other approaches we set $\epsilon = \alpha\,\mu$ and undo the standardization, which leads to

$$\sqrt{\ln \frac{1}{\alpha^2}} \cdot \frac{\sigma}{\sqrt{N}} \;<\; \alpha \cdot \mu.$$

Squaring both sides of the inequality results in

$$\sqrt{\ln \frac{1}{\alpha^2}} \cdot \frac{\sigma}{\sqrt{N}} \;<\; \alpha \cdot \mu.$$

By solving the inequality for N we obtain finally

$$\ln\left(\frac{1}{\alpha^2}\right) \cdot \frac{1-p}{p\,\alpha^2} \;<\; N_{Approximation}. \tag{3.20}$$

Figure 3.2 compares the three approaches. Using the quantile-function or the approximation with the modified e-function results in an earlier switch between the Γ and Δ-regime, demanding less particles to be present with a given reaction probability. The improvement is depending on the given α-value. For low α-values it cuts the needed particle number by more than 90%.

With Equation (3.17) and Equation (3.18), I have defined the borders between the three regimes. Equation (3.17) marks the transition between Σ and Γ, and Equation (3.18) marks the transition between Γ and Δ.

Illustration of the findings: Let us consider the derived criteria for $\alpha = 0.05$ (respectively $\alpha = 0.01$) in more detail. For small reaction probabilities p (Note: $p \leq \alpha$), $1 \approx 1 - p$. Thus, one obtains from Equation (3.17) and Equation (3.18) the following estimations of the mean number of reactions $N\,p$:

$$N\,p \approx N\,p\,(1-p) \geq \tfrac{1}{3\,\alpha} \text{ and } N\,p \approx N\,p\,(1-p) \geq \tfrac{1}{\alpha^3}\,.$$

Hence, criteria (3.17) is fulfilled if the mean reaction number is larger than 7 (for $\alpha = 0.01 : 34$). This is also the upper bound for the amount of random numbers per reaction channel necessary in a time-interval, since for larger reaction numbers (on average), Equation (3.2) can be used, for which only one random number is required. Analogously, one can see, that deterministic reaction kinetics can be used if the expectation of the reaction number is larger than 8000 (For $\alpha = 0.01 : 10^6$). Then, obviously, no random number is necessary. Of course, the actual number of particles depends on the reaction probability p.

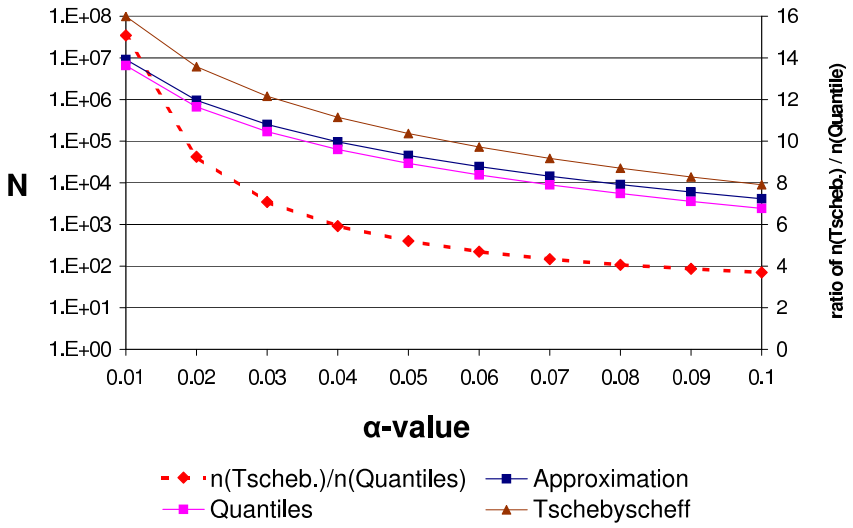

Figure 3.2: *This figure demonstrates the three approaches presented to model the transition between the* Γ *and the* Δ*-regime give by the three Equations (3.18), (3.19) and (3.20). The probability* p *was assumed to be always equal to* α*. The number of particles needed is presented with respect to* α*. The most accurate approach is the one using the quantile-functions, closely followed by an approximation using another e-function. These two methods allow up to only one 1/15th of the original amount of particles for the algorithm to switch from* Γ *to* Δ*.*

3.2.2 Length of the Time Steps

One of the important properties of the COAST algorithm is the assumption of nearly constant reactions probabilities, which is again defined by the error parameter α. We have to clarify how many reactions are allowed to occur without a change of the probabilities by more than α. This paragraph will solve this problem. Equation (3.24) defines the critical number of reactions per reaction channel. It is possible with Equation (3.22) to calculate the timespan in which these reactions are going to occur depending on the type of reaction.

Let us consider a single reaction channel $A+B \to P$ with a stochastic reaction constant c and particle numbers $N_A \leq N_B$, where P is an undefined product. According to the *First Reaction Method*, $l \leq N_A$ reactions have occurred after the time

$$\delta t := \sum_{i=0}^{l-1} \frac{-\ln(r_i)}{c\,(N_A - i)\,(N_B - i)}\,,$$

where r_i are independent random variables uniformly distributed in $[0, 1]$. The mean till l reactions have occurred is given by

$$\langle \delta t \rangle (l) \;:=\; \sum_{i=0}^{l-1} \left(\frac{-1}{c\,(N_A - i)\,(N_B - i)} \int_0^1 \ln(x)\,dx \right).$$

The integral can be solved to -1 and by factoring out N_A twice in the nominator we obtain

$$\langle \delta t \rangle (l) \;=\; \frac{1}{c\,N_A^2} \sum_{i=0}^{l-1} \frac{1}{(1 - \frac{i}{N_A})\,(\frac{N_B}{N_A} - \frac{i}{N_A})}\,.$$

Now $\frac{i}{N_A}$ is replaced by x. The occurring sum can be interpreted as a Riemann sum for the corresponding integral. With an estimated error of $\mathcal{O}(N_A^{-1})$ the equation changes to

$$\langle \delta t \rangle (l) \;=\; \frac{1}{c\,N_A} \left(\int_0^{\frac{l}{N_A}} \frac{1}{(1-x)\,(\frac{N_B}{N_A} - x)}\,dx + \mathcal{O}(N_A^{-1}) \right).$$

Thus, by eliminating terms of the order $\mathcal{O}(N_A^{-2})$, one obtains [Gradshteyn and Ryshik, 1980, p.68 (2.172)]:

$$\langle \delta t \rangle (l) = \begin{cases} \frac{1}{c\,(N_B - N_A)} \ln\left(\frac{N_B - l}{N_A - l}\,\frac{N_A}{N_B}\right), & \text{if } N_A \neq N_B\,, \\ \frac{1}{c\,N_A}\,\frac{l}{N_A - l}\,, & \text{if } N_A = N_B\,. \end{cases} \tag{3.21}$$

This result can also be obtained by an deterministic approach. As derived in the Appendix A.5, $A(t)$ the concentration of the species A in the reaction $A + B \xrightarrow{c} P$ after a timespan t can be described by

$$A(t) = \frac{(N_A N_B - N_A^2) \cdot e^{-kt(N_B - N_A)}}{N_B - N_A \cdot e^{-ct(N_B - N_A)}},$$

where N_A and N_B mark the starting concentrations of A and B.
Because we are looking for the timespan τ for l expected reactions, we have to calculate l first:

$$\begin{aligned} l &= N_A - A(t) \\ \Rightarrow l &= N_A - \frac{(N_A N_B - N_A^2) \cdot e^{-c\tau(N_B - N_A)}}{N_B - N_A \cdot e^{-c\tau(N_B - N_A)}} \\ \Rightarrow l &= \frac{N_A\,N_B\,(1 - e^{-c\tau(N_B - N_A)})}{N_B - N_A\,e^{-c\tau(N_B - N_A)}} \end{aligned}$$

This can be reformulated to

$$\tau(l) = \frac{1}{c(N_B - N_A)} \ln\left(\frac{N_B - l}{N_A - l}\,\frac{N_A}{N_B}\right),$$

which is the same we have obtained in Equation (3.21). The deterministic and the stochastic way led both to the same result.

The deterministic uses usually concentrations for a single substrates, but this is not an obstacle, because if a constant volume is used through the simulations all concentrations are equivalent to specific particle numbers. Furthermore one has to notify that the constant used here is the stochastic reaction constant.

For all type of reactions it is possible to calculate an expectancy for the time span τ until l reactions have occurred, either by a stochastic or a deterministic approach. The deterministic way has the advantage of being much easier to calculate and the simple relation of time and reactions is enough for the purpose of this thesis.

In more detail, for first and second order reactions the time span τ for l reactions is

$$\tau(l) = \begin{cases} \frac{1}{c} \ln\left(\frac{N_A}{N_A - l}\right), & \text{for } A \to P, \\ \frac{1}{c\,(N_B - N_A)} \ln\left(\frac{N_B - l}{N_A - l}\,\frac{N_A}{N_B}\right), & \text{for } A + B \to P \ (N_A \neq N_B), \\ \frac{1}{c\,N_A}\,\frac{l}{N_A - l}, & \text{for } A + B \to P \ (N_A = N_B), \\ \frac{2}{c_\mu\,N_A}\,\frac{l}{N_A - 2\,l}, & \text{for } 2\,A \to P, \end{cases} \tag{3.22}$$

which is equivalent to the occurring number of reactions

$$l(\tau) = \begin{cases} N_A\,(1 - e^{-c\,\tau}), & \text{for } A \to P, \\ \frac{N_B\,N_A\,(1 - e^{-(N_B - N_A)\,c\,\tau})}{N_B - N_A\,e^{-(N_B - N_A)\,c\,\tau}}, & \text{for } A + B \to P (N_A \neq N_B), \\ \frac{N_A^2\,c\,\tau}{1 + N_A\,c\,\tau}, & \text{for } A + B \to P, (N_A = N_B), \\ \frac{N_A^2\,c\,\tau}{2 + 2\,N_A\,c\,\tau}, & \text{for } 2\,A \to P. \end{cases} \tag{3.23}$$

Analogous results can be derived for higher order functions.

I will now show that all reaction probabilities are considered constant up to $\alpha \cdot 100\%$, if for all reaction channels $\mathcal{R}_\mu$ with $\sigma_\mu(A)\,A + \sigma_\mu(B)\,B \to \sigma_\mu(P)\,P$ the number of reactions is smaller than

$$l_\mu := \min\left\{\frac{\alpha\,N_S}{2\,\varrho(S)\,\sigma_\mu(S)}; S \in \{A, B, P\}\right\} \tag{3.24}$$

where $\varrho(S)$ is the number of reaction channels in which S occurs and $\sigma_\mu(S)$ is the stoichiometric factor of S in the reaction channel μ.

I.e., a criterion will be derived for how many reaction steps can be allowed without changing any reaction probability in a relevant fashion. To this aim, let us consider a small variation ϵ of the particle numbers N_A and N_B in a (second order) reaction

channel $A + B \longrightarrow P$. In this case, the expected number of reactions in a time interval of length τ is (cf. Equation (3.23)):

$$l(N_A, N_B) = \frac{N_B\, N_A\, (1 - e^{-(N_B - N_A)\, c\, \delta t})}{N_B\, - N_A\, e^{-(N_B - N_A)\, c\, \delta t}}\,. \tag{3.25}$$

This leads in a zeroth order Taylor expansion to ($N_B > N_A$)

$$l(N_A, N_B) \approx \frac{N_B \cdot N_A}{N_B\, - \, N_A}. \tag{3.26}$$

We define the reaction probabilities (propensities) for $N_A + \epsilon$ and $N_B + \epsilon$ are approximately the same as for N_A and N_B if

$$|\, l(N_A + \epsilon, N_B + \epsilon) - l(N_A, N_B)\,| < \alpha\, l(N_A, N_B)\,.$$

By using approximation (3.26), one obtains

$$|\epsilon\, N_B + \epsilon\, N_A\,| < \alpha\, N_A\, N_B\,. \tag{3.27}$$

For $A \rightarrow P$ and $2\,A \rightarrow P$, one obtains analogously (cf. Equation (3.22)):

$$|\epsilon| < \alpha\, N_A \quad \text{and} \quad |\epsilon| < \frac{\alpha\, N_A}{2} \tag{3.28}$$

respectively, where for the latter estimation, one has to assume that $1 + (N_A + \epsilon)\, c\tau \approx 1 + N_A\, c\tau$.

Let us assume that substrate S occurs in $\varrho(S)$ reaction channels. Then, inequalities (3.27) and (3.28) are valid if the number of reactions l_μ in each reaction channel $\mathcal{R}_\mu$ with $\sigma_\mu(A)\, A + \sigma_\mu(B)\, B \rightarrow \sigma_\mu(P)\, P$ fulfills

$$l_\mu \leq \min\left\{ \tfrac{\alpha\, N_S}{2\, \varrho(S)\, \sigma_\mu(S)};\; S \in \{A, B, P\} \right\}\,.$$

In this case, chemical reactions can, in a first approximation, be considered as independent, identically distributed events, so that the reaction probabilities can be approximated by binomial- or (for large particle numbers) discrete Gaussian-distributions.

3.3 The Reaction Algorithm

COAST follows the scheme in Figure 3.3 and a detailed list of all steps is presented in Table 3.1. After initialization, the length τ of a time interval is estimated, where reaction probabilities are expected to be nearly constant. This is the case if the expected number of reactions is smaller than l_μ, as defined in Equation (3.24). This is done in the subroutine “Next evaluation time”.

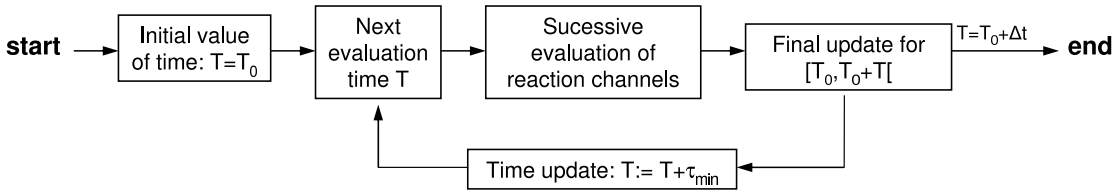

Figure 3.3: *Schematic representation of* COAST. *The scheme shows the determination of the number of reactions at a lattice point i in the time interval* $[t, t + \Delta t[$.

These nearly constant reaction probabilities allow one to consider higher order reactions as nearly independent processes. Furthermore, one can also compute the number of reactions in the different reaction channels successively, since the mutual influences of the reactions can be considered small. Note that the nearly constant particle numbers imply that the exact order of the evaluations of the reaction channels is not essential for the outcome.

Accordingly, the number of reactions in the different reaction channels during this interval of length τ is computed successively by application of the subroutine "Evaluation of reaction channels". Finally, an update of particle numbers is performed, partly in "Evaluation of reaction channels", partly in "Final update".

This procedure is repeated until time $T_0 + \Delta t = t_{stop}$ is reached. For pure reaction systems, $T_0 + \Delta t$ represents the end of the simulation. However, in Section 5.3, I will discuss how to extend this reaction-algorithm to a reaction-diffusion algorithm, where reactions and thermal motions in the same time interval are determined successively. In this context, $[T_0, T_0 + \Delta t[$ represents only a short simulation step. In what follows, I will consider in more detail the three most important steps in the algorithm: "Next evaluation time", "Evaluation of reaction channels" and "Final update".

3.3.1 Next Evaluation Time T

Initially, a value for the error parameter α must be chosen; a lower α results in increasing accuracy, but at the expense of increasing computational cost. Then, the critical number of reactions l_μ (cf. Equation (3.24)) is computed for each reaction channel $\mathcal{R}_\mu$ with an additional simplification:

In a time step $[T, T + \tau[$ no particle can react twice[1].

Accordingly, the criterion in Equation (3.24) is restricted to the educts of the reactions, simplifying the computation without leading to unreliable results. This

[1]Note that the probability for a single reaction of a particle in a time interval is smaller than α. Hence, the probability for two or more reactions of a particle in a time interval is smaller than α^2.

I.	**Preparation Phase**	
	α set by user	
	∘ t = 0	
	∘ calculate c_μ out of all k_μ	
II.	**Main loop**	until T=t_{stop}
	a) Next evaluation time	
	∘ calculate $\lambda(S)$	(max. no. of allowed reacting particles)
	∘ calculate l_μ	(max. no. of allowed reactions per channel)
	∘ calculate τ_μ	(time till l_μ reactions occur)
	∘ sort channels by τ	(lowest τ first)
	∘ T = T + τ_{min}	
	b) Evaluation of reaction channels	
	∘ loop over all reaction channels μ	
	• calculate p_μ	(reaction probabilities)
	• divide in Σ, Γ, Δ	
	• calculate κ_μ	(no. of occurring reactions)
	• update educts	
	c) Final Update	
	∘ update products	

Table 3.1: *The single steps of the reaction algorithm*
The process is split into two phases. During the preparation phase α is set and the kinetic constants are transformed into stochastic reaction constants. In the main loop the three routines "Next evaluation time", "Evaluation of reaction channels" and "Final update" are executed until the time stop t_{stop} is reached.

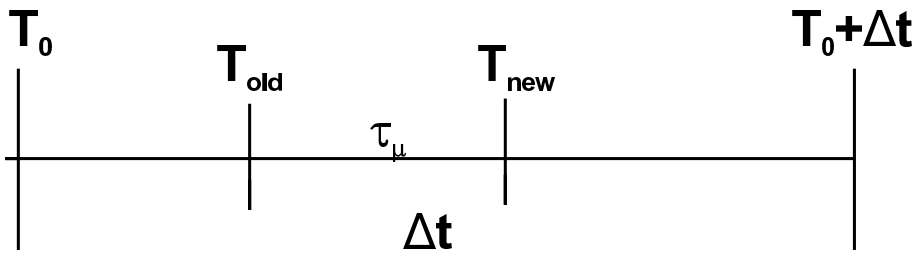

Figure 3.4: *Illustration for the different time symbols used in the algorithm.*

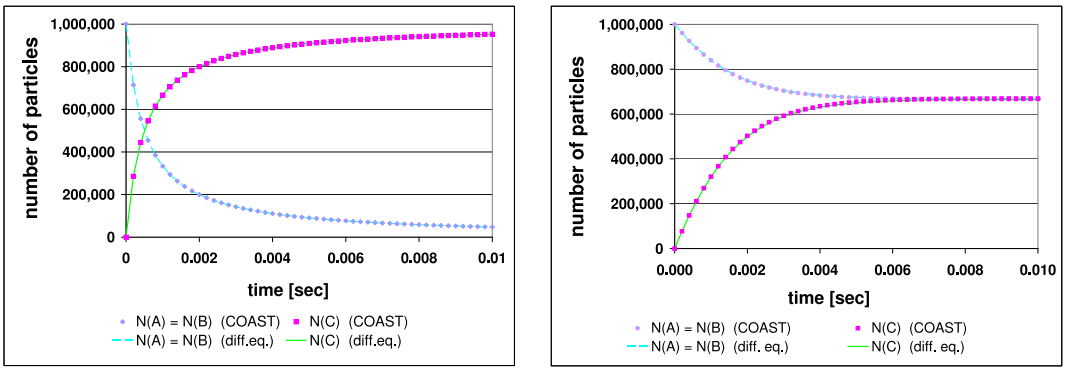

Figure 3.5: *Time-evolution of the particle numbers N_A and N_C in both systems for* COAST*-simulations ($\alpha = 0.05$) and for the deterministic reaction kinetics. The left (right) diagram shows the behavior for $A+B \leftrightharpoons C$ ($A+B \leftrightharpoons 2\,C$). $k_{left} = 0.002$ $k_{right} = 0.0002$. In the beginning,* COAST *applies the deterministic reaction kinetics to $A + B \to 2\,C$, but the first reaction method to the back reaction.*

assumption can be eliminated by applying Equation (3.24) to both educts and products.

The most stringent test for this simplifying assumption is the investigation of the time-evolution of a system with a very fast and a slow reaction channel, so that the fast channel is treated by the deterministic reaction kinetics, and the slow channel by the *First Reaction Method*. Such systems are shown in Figure 3.5, where the reaction systems $A+B \leftrightharpoons C$ and $A+B \leftrightharpoons 2C$ are considered. The initial conditions $N_A = N_B = 10^6$, $N_C = 0$ where chosen such that, in the beginning, $A + B \to 2C$ is treated by the deterministic reaction kinetics, but the back reaction by the *First Reaction Method*. Clearly, the mean value of the COAST-simulations coincide with the values of the ODE-solutions, so that one cannot observe a (relevant) error due to the assumption "no particle reacts twice".

Subsequently, for each reaction channel $\mathcal{R}_\mu$, the time $T_{old} + \tau_\mu$ is determined at which l_μ reactions are expected (cf. Equation (3.22)). The next evaluation time T_{new} is given either by the minimum $T_{old} + \tau_\mu$ or by $T_0 + \Delta t$, where $T_0 + \Delta t$ is either the end of the whole simulation or, in reaction-diffusion models, the end of a time step[2]. In more detail, the module "Next evaluation time" is composed of the following three steps:

Step 1: For each substrate A, compute the maximal number of particles per species, which is allowed to react such that the propensity is not changing by more than α (derivation in Equation (3.22))

$$\lambda(A) := \mathsf{max}\left\{1,\, \tfrac{\alpha\, N_A}{2\,\varrho(A)}\right\}$$

where $\varrho(A)$ is the number of reaction channels in which A occurs as a reactant.

Step 2: For each reaction channel $\mathcal{R}_\mu$ with $\sigma_\mu(A)\, A + \sigma_\mu(B)\, B \rightarrow \sigma_\mu(C)\, C + \sigma_\mu(D)\, D$, compute the maximal number of allowed reactions

$$l_\mu := \mathsf{min}\left\{\tfrac{\lambda(A)}{\sigma_\mu(A)}, \tfrac{\lambda(B)}{\sigma_\mu(B)}\right\}. \tag{3.29}$$

and $\tau_\mu(l_\mu$ given by Equation (3.22)).

Step 3: Determine

$$T_{new} := \mathsf{min}\{T_{old} + \mathsf{min}\{\,\tau_\mu\,\}, T_0 + \Delta t\}\,.$$

So T_{new} is either the sum of the lowest τ of all channels and the old T, or the timespan till the end of $T_0 + \Delta t$.

3.3.2 Evaluation of Reaction Channels

The successive evaluation of the reaction numbers starts at the reaction channel with minimum τ_μ and ends at the reaction channel with maximum τ_μ[3]. Accordingly, the first step is the ordering of the reaction channels $\mathcal{R}_\mu$ according to the τ_μ's. In the second step, one determines to which of the model classes Σ, Γ and Δ each reaction channel belongs, where Σ represents the *First Reaction Method* of Gillespie [1976, 1977], Γ a Gaussian-distribution (cf. Equation (3.2)), and Δ the deterministic reaction kinetics. Correspondingly, this classification is performed by the criteria in

[2] In reaction -diffusion models one often computes reactions and thermal motions in the same time step successively [Hebert, 1992, Möller and Wagner, 2005]

[3] Since the reaction probabilities are nearly constant during a time step, the exact succession of the evaluation steps do not have a strong influence on the outcomes

Equation (3.17) and Equation (3.18), where the probabilities p_μ are given by the expectations in Equations (3.23) of the *First Reaction Method* divided by the (smaller) particle number.

Between the evaluation of two reaction channels, a first update of the number of particles is performed. This first update is restricted to a reduction of particle numbers corresponding to the consumption of educts. The second update due to the production of particles in reactions will be performed in the "Final update" at the end of each time step. Note that this splitting of updates is in accordance with the assumption that no particle reacts twice in $[T, T+\tau[$.

Step 1: If there are m reaction channels $\mathcal{R}_\mu$, determine the sequence $\mathsf{a}(\nu_1, ..., \nu_m)$ $(\nu_i \in \{1, .., m\})$, so that for all $i < j$: $\mathcal{R}_{\nu_i} \neq \mathcal{R}_{\nu_j}$ and $\tau_{\nu_i} \leq \tau_{\nu_j}$. (sorting the channels with lowets τ first)

Step 2: For $i := 1$ to m do:

(a) Compute

$$a_{\nu_i} := \begin{cases} N_A\,, & \text{for } A \to ?, 2\,A \to ?\,, \\ \min\{N_A, N_B\}\,, & \text{for } A + B \to ?\,, \end{cases}$$

and the reaction probabilities $p_{\nu_i} := \frac{l_{\nu_i}(\tau)}{a_{\nu_i}}$, where $l_{\nu_i}(\tau)$ is given by Equation (3.23).

(b) Perform the classification

$$\mathcal{R}_{\nu_i} \in \begin{cases} \Delta\,, & \text{if } a_{\nu_i} > \frac{1}{\alpha^3\, p_{\nu_i}\,(1-p_{\nu_i})}\,, \\ \Gamma\,, & \text{if } \frac{1}{\alpha^3\, p_{\nu_i}\,(1-p_{\nu_i})} \geq a_{\nu_i} > \frac{1}{3\,\alpha\, p_{\nu_i}\,(1-p_{\nu_i})}\,, \\ \Sigma\,, & \text{if } \frac{1}{3\,\alpha\, p_{\nu_i}\,(1-p_{\nu_i})} \geq a_{\nu_i}\,. \end{cases}$$

(c) Compute the number of reactions κ_{ν_i} in $[T, T+\tau[$ by

If $\mathcal{R}_{\nu_i} \in \Sigma$:

$$\kappa_{\nu_i} := \min\left\{ m \in \mathbb{N}_0, t_0 - \sum_{j=0}^{m} \frac{\ln(r_{\nu_i}^{(j)})}{Q_{\nu_i}(j)} > T \right\}\,,$$

where $r_{\nu_i}^{(j)}$ are random variables equidistributed in $[0, 1]$ and where

$$Q_{\nu_i}(j) := \begin{cases} c_{\nu_i}\,(N_A - j)\,, & \text{for } A \to C + D, \\ c_{\nu_i}\,(N_A - j)\,(N_B - j) & \text{for } A + B \to C + D, \\ \frac{c_{\nu_i}}{2}\,(N_A - 2\,j)\,(N_A - 2\,j - 1)\,, & \text{for } 2\,A \to C + D\,. \end{cases}$$

If $\mathcal{R}_{\nu_i} \in \Gamma$:

$$\kappa_{\nu_i} = \min\left\{\frac{N_A}{\sigma_{\nu_i}(A)}, \frac{N_B}{\sigma_{\nu_i}(B)}, \max\{0, n_{\nu_i}\}\right\}$$

with Equation (3.3):

$$n_{\nu_i} := \text{Round}\left(\sqrt{N_A\, p_{\nu_i}\,(1-p_{\nu_i})}\, X + N_A\, p_{\nu_i}\right)$$

where the normally distributed random variable X can efficiently be computed by the Box-Muller algorithm [Box and Muller, 1958].

If $\mathcal{R}_{\nu_i} \in \Delta$:

$$\kappa_{\nu_i} := \text{Round}\,(N_A\, p_{\nu_i})\ .$$

(d) Update of educts: If $\mathcal{R}_{\nu_i}$ is given by $\sigma_{\nu_i}(A)\,A + \sigma_{\nu_i}(B)\,B \to \sigma_{\nu_i}(C)\,C + \sigma_{\nu_i}(D)\,D$, then $N_A = N_A - \sigma_{\nu_i}(A)\,\kappa_{\nu_i}$ and $N_B = N_B - \sigma_{\nu_i}(B)\,\kappa_{\nu_i}$.

3.3.3 Final Update

In the final update, the particle numbers are increased according to the number of reactions. Thus, the final update can be described in the following fashion:

Update of products: For all reaction channels $\mathcal{R}_\mu$ with $\sigma_\mu(A)\,A + \sigma_\mu(B)\,B \to \sigma_\mu(C)\,C + \sigma_\mu(D)\,D$ do: $N_C := N_C + \sigma_\mu(C)\,\kappa_\mu$, $N_D := N_D + \sigma_\mu(D)\,\kappa_\mu$,

3.4 Extending COAST to Diffusion

3.4.1 Problems and Approaches

The models for the description of thermic motions of particles is composed of two classes (cf. Figure 3.2): The stochastic description of the trajectories of single particles and diffusion models reflects the time-evolution of the probability distribution of such a particle. Correspondingly, the first class of models is able to reflect stochastic effects due to small particle numbers, whereas their simulations are computationally very expensive for large particle numbers. On the other hand, diffusion models are computationally very efficient, but their deterministic time-evolution suppresses stochastic fluctuations, so that they are only suitable for large particle numbers. Consequently, both kinds of models are not suitable to represent cellular networks, since they often contain substrates with a wide range of possible particle numbers [Goodsell, 1991, Endy and Brent, 2001].

Algorithm/Model	**Reference**	**Modeling of**	**kind of model**
Molecular dynamics	Baynes 2004, Friedel 2004	single particles	deterministic
Langevin-equation	Stiles 1998	single particles	stochastic
Smoldyn	Lipkow 2005	single particles	stochastic
(Spatial) Gillespie	Takahashi 2004	single particles	stochastic
Gibson-Bruck	Hattne 2005, Stundzia 1996	single particles	stochastic
Diffusion model (PDE)	Evans 1999	distributions of particles	deterministic

Table 3.2: *An overview of the algorithms for (reaction-) diffusion models. Note that molecular dynamics (MD) requires a description of* all particles *in a system, whereas all other models allow a consideration of subsystems.*

However, the thermal motion of particles can be interpreted as a kind of "reaction": one considers molecules of the substrate with different positions as different substrates and, thus, the transitions from one lattice point to another as a reaction channel. Accordingly, reaction-diffusion algorithms can be considered published [Elf et al., 2003, Stundzia and Lumsden, 1996] treating not only reactions, but also the diffusive motions by exact stochastic reaction-algorithms [Gillespie, 1977, Gibson and Bruck, 2000]. For large particle numbers, these methods lead to high computational costs. Consequently, they can only be efficiently applied to systems with small or intermediate particle numbers.

		small	intermediate	large	subdivision
τ-leap	Gillespie 1977	Poisson	Poisson	Poisson	
binomial leap	Chatterjee 2005 Tian 2004	binomial	binomial	binomial	
hybrid methods	Takahashi 2004	NRM	NRM	deterministic	by user, fixed
	Kiehl 2004	NRM	NRM	deterministic	by user, fixed
	Cao 2005 Rao 2003	FRM	FRM	quasi steady state	by user, fixed
	Haseltine 2002	FRM	FRM	Langevin-equation	by user, fixed
maximum time step	Puchalka 2004	NRM	Poisson	Poisson	automatic in each step
PW-DMC	Resat 2004	Monte Carlo with single reactions	Monte Carlo with bundles	Monte Carlo with bundles	automatic in each step
COAST		FRM	Gauss	deterministic	automatic in each step

Table 3.3: *Characterization of the recently published reaction-algorithms: FRM denotes the First Reaction Method or Direct Method of Gillespie Gillespie [1977], NRM the Next Reaction Method of Gibson and Bruck Gibson and Bruck [2000]. "bundle" means several reactions of the same type.*

Some recently published reaction-algorithms (cf. Figure 3.3) try to solve the dilemma between the exactness of modeling and computational costs by using different levels of modeling for the different ranges of particle numbers. I will now describe how COAST , as a multi-level algorithm, can be applied to diffusion processes and by keeping its original function extending it to a reaction-diffusion-algorithm.

3.4.2 Outline

Here I will discuss the adoption of the COAST to the needs of linear diffusion models. Thereby, *linear diffusion model* means that the diffusion rates of each described substrate is independent from the concentrations of all of these explicitly described substances. This is a reasonable approximation if the interactions between these substrates are small compared to the interactions with other substrates. Thus, linear diffusion models may not be suitable for all biological systems [Agutter et al., 1995], but are always appropriate if the concentrations of the explicitly described substrates are low enough.

On the other hand, linear diffusion models allow the subdivision of the diffusion model into (approximatively) independent subunits: the thermal motions of different substrates can be treated independently, and the transitions from different lattice points can, for small time steps, also be considered as approximatively independent events. Last but not least, the transitions from the same lattice site into different

directions can also be treated independently; provided that one uses appropriately constrained probabilities.

This allows to decompose the dynamics into (nearly) independent processes help to simplify the algorithm enormously. Additionally, since linear diffusion models correspond to first order reactions, they can work with constant time steps, which additionally allows for a simplification of the algorithm.

I emphasize here the concept of error control of COAST to linear diffusion models, which means that the errors due to the discretization of the spatial coordinates are estimated dependently from two error control parameters, namely the parameter α mentioned before and a parameter R corresponding to the spatial resolution of the diffusion model.

In the following section, the diffusion model and the corresponding random walk used by COAST are introduced. The content of Section 3.6 is the estimation of the errors due to the necessary discretization of time and space dependently from error parameters.

3.5 The Discrete Diffusion Model

In this paragraph I will describe how to get from the continuous diffusion model to a discrete diffusion model. This approach allows us to approximate the continuous diffusion by a discrete approach and gain with Equation (3.35) an quantitative expression for the transition probability between two adjacent volume elements. For the discussion of the diffusion-model on which COAST is based, namely the Smoluchowski-equation, let us consider the case of a one-dimensional motion of a single substrate A with a friction coefficient γ and an external force $f_A(x)$.

The motion of a particle A in a time span δt is given by the Langevin-equation in the strong friction limit (i.e. $m\ddot{x} \to 0$):

$$x(t+\delta t) - x(t) = \int_t^{t+\delta t} \frac{f_A(x(s))}{\gamma_A}\, ds + \sqrt{2\, D_A\, \delta t}\, W\,,$$

where k_B is the Boltzmann's constant, T is the absolute temperature, and W is a normally distributed random number with density

$$\sigma(W) := \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{W^2}{2}}\,. \tag{3.30}$$

D_A is the diffusion coefficient of the substrate A and is related to the Boltzmann-constant k_B, the temperature T and the friction coefficient of substrate A by

$$D_A := \frac{k_B\, T}{\gamma_A}.$$

By Ito-integration [Oksendahl, 1985, p. 20 ff.] of Equation (3.30), one obtains the diffusion-equation, which describes the time-evolution of the corresponding probability density function $\varrho(x,t)$:

$$\frac{\partial}{\partial t}\,\varrho(x,t) = \frac{-1}{\gamma_A}\frac{\partial}{\partial x}\,(\,f_A(x)\,\varrho(x,t)\,) + D_A\,\frac{\partial^2}{\partial x^2}\,\varrho(x,t)\,, \tag{3.31}$$

namely the Smoluchovski-equation.
Now let us consider, the discrete Smoluchovski-equation based on the lattice Λ and open boundary conditions:

$$\Lambda := \Big\{(i\,\Delta x\,, j\,\Delta t\,)\,|\,-n \le i \le n\,; j \in \mathbb{N}_0\Big\}\,,$$

and

$$\varrho(-n,t) = \varrho(n,t) = 0 \quad \forall t \in \mathbb{N}_0\,.$$

we obtain from subtracting or adding respectively the two Taylor-expansions of g around the point b

$$\begin{aligned} g(b+\Delta b) &= g(b) + \Delta b\,g'(b) + \tfrac{\Delta b^2}{2}\,g''(b) + \tfrac{\Delta b^3}{6}\,g'''(b) + \\ &\quad + \tfrac{\Delta b^4}{24}\,g''''(b) + \mathcal{O}(\Delta b^5)\,, \\ g(b-\Delta b) &= g(b) - \Delta b\,g'(b) + \tfrac{\Delta b^2}{2}\,g''(b) - \tfrac{\Delta b^3}{6}\,g'''(b) \\ &\quad + \tfrac{\Delta b^4}{24}\,g''''(b) - \mathcal{O}(\Delta b^5) \end{aligned}$$

the approximations

$$\begin{aligned} g'(b) &= \tfrac{g(b+\Delta b)-g(b-\Delta b)}{2\,\Delta b} + \mathcal{O}((\Delta b)^2), \\ g''(b) &= \tfrac{g(b+\Delta b)-2\,g(b)+g(b-\Delta b)}{(\Delta b)^2} + \mathcal{O}((\Delta b)^2). \end{aligned} \tag{3.32}$$

For $\frac{\partial}{\partial t}\varrho(x,t)$ we need a slightly different approach. To keep it a Markov-process, we approximate this expression by another Taylor-polynomial around the point t:

$$\begin{aligned} \varrho(t+\Delta t) &= \varrho(t) + \varrho'(t)\cdot((t+\Delta t) - t) \\ \varrho'(t) &= \frac{\varrho(t+\Delta t) - \varrho(t)}{\Delta t}. \end{aligned} \tag{3.33}$$

Inserting Equation (3.32) and Equation (3.33) into Equation (3.31) and substituting b by x, leads to:

$$\begin{aligned}\varrho(i,t+\Delta t) &= \varrho(i,t)\,(1-\Delta t\left(\frac{D_A}{\Delta x^2}-\frac{f(i-1)}{\gamma_A\,2\Delta x}\right)\\ &\quad -\Delta t\left(\frac{D_A}{\Delta x^2}+\frac{f(i+1)}{\gamma_A\,2\Delta x}\right))\\ &\quad +\varrho(i-1,t)\,\Delta t\left(\frac{D_A}{\Delta x^2}+\frac{f(i)}{\gamma_A\,2\Delta x}\right)\\ &\quad +\varrho(i+1,t)\,\Delta t\left(\frac{D_A}{\Delta x^2}-\frac{f(i)}{\gamma_A\,2\Delta x}\right)\end{aligned} \tag{3.34}$$

where the first term on the right side describes the particles staying at lattice point i between t and $t+1$. The second term describes the particles moving from $i-1$ to i and the third term the particles moving from $i+1$ to i. The factors connected to the density function ϱ are the transition probabilities. Therefore I define:

$$q(i+\nu|i;\Delta t) := \Delta t\,\left(\tfrac{D_A}{(\Delta x)^2}+\nu\,\tfrac{f_A(i+\nu)}{2\,\gamma_A\,\Delta x}\right) \tag{3.35}$$

as the probability for the transition $i \to i+\nu$ to obtain the discrete diffusion-model:

$$\begin{aligned}\varrho(i,t+\Delta t) &= \Big(1-q(i+1|i;\Delta t)-q(i-1|i;\Delta t)\Big)\,\varrho(i,t)\\ &+ \quad q(i|i+1;\Delta t)\,\varrho(i+1,t)+q(i|i-1;\Delta t)\,\varrho(i-1,t)\,.\end{aligned} \tag{3.36}$$

An analogous derivation of the diffusion process in reversed order can be found in standard stochastic literature (e.g. Feller [1970, 354pp]).

3.6 The Values of the Discretization Parameters Δx and Δt

In this section, appropriate choices for the discretization parameters Δx and Δt are presented. To this aim, I will firstly set up four conditions, which will result in definitions for Δx and Δt. In doing so, we will always consider the case of a single substrate A. At the end of this section, the derived findings will be summarized and the extension to systems with many substrates will be discussed.

3.6.1 First Condition: Approximation of Continuous Distributions

By approximation of a continuous distribution by a discrete distribution we gain a criterion for Δx.

In a diffusion model like the Smoluchowski-equation, the particles are described by continuous distributions, so that the number of particles in the interval $[x-\frac{\Delta x}{2}, x+\frac{\Delta x}{2}[$ is given by the integral over a density function ϱ

$$\begin{aligned} P_{cont} &:= \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \varrho(y)\,dy \\ &\quad \text{(2nd grade Taylor-polynomial of } \varrho(y)) \\ &= \int_{x-\frac{\Delta x}{2}}^{x+\frac{\Delta x}{2}} \varrho(x) + \varrho'(x)\,(y-x) + \tfrac{(y-x)^2}{2}\,\varrho''(x) + \mathcal{O}((\Delta x)^3)\,dy \\ &= \varrho(x)\,\Delta x + \varrho'(x)\;\; 0 \; + \tfrac{2(\Delta x)^3}{24}\,\tfrac{\varrho''(x)}{2} + \mathcal{O}((\Delta x)^5) \\ &= \varrho(x)\,\Delta x + \tfrac{(\Delta x)^3}{24}\varrho''(x) + \mathcal{O}((\Delta x)^5)\,. \end{aligned} \tag{3.37}$$

On the other hand, the simulations are based on a discrete distribution assuming that the particles are homogeneously distributed within a voxel. Thus, the number of particles in the interval $[x-\frac{\Delta x}{2}, x+\frac{\Delta x}{2}[$ is given in a discrete model by

$$P_{disc} \approx \varrho(x)\,\Delta x\,. \tag{3.38}$$

Accordingly, the condition $|P_{cont} - P_{disc}| < \alpha\, P_{cont}$ can, as a first approximation, be written as

$$\begin{aligned} \left|\varrho(x)\,\Delta x + \tfrac{(\Delta x)^3}{24}\varrho''(x) \; - \varrho(x)\,\Delta x\right| &< \alpha\,\left|\varrho(x)\,\Delta x + \tfrac{(\Delta x)^3}{24}\varrho''(x)\right| \\ &\qquad \text{with } \alpha\,\tfrac{(\Delta x)^3}{24}\varrho''(x) \approx 0 \\ \Leftrightarrow \qquad \left|\tfrac{(\Delta x)^3}{24}\varrho''(x)\right| &< \alpha\,|\varrho(x)\,\Delta x| \\ \Leftrightarrow \qquad \Delta x &< \sqrt{24\,\alpha\, min\left|\tfrac{\varrho}{\varrho''}\right|}. \end{aligned} \tag{3.39}$$

For a reformulation of this inequality, an assumption about the exact form of ϱ is necessary. Such an assumption is naturally problematic since ϱ usually depends on time. On the other hand, in most cases ϱ will be nearly a Gaussian distribution - for example at the local minima of the potential. Hence, we define ϱ as a probability density function of a Gaussian-distribution with standard deviation R

$$\varrho(x) := \tfrac{1}{\sqrt{2\,\pi}\,R}\,e^{-\frac{x^2}{2\,R^2}}\,, \tag{3.40}$$

which results in

$$\Delta x < \sqrt{24\,\alpha\,\left|\tfrac{R^4}{x^2-R^2}\right|}. \tag{3.41}$$

Due to Tschebyscheff's inequality, x is smaller than $\frac{R}{\sqrt{\alpha}}$ with probability $1-\alpha$. With similar considerations like the one justified in Section 3.2.1 it is possible to derive

a different value for x, which would result in a larger value for Δx and possibly a better performance in total for the implementation. However, the general ideas are the same and therefore I limit the discussion only to the value for x derived from Tschebyscheffs inequality. By neglecting events with probability smaller than α, one obtains the estimation

$$\Delta x \;<\; \sqrt{24\,\alpha\,\left|\tfrac{R^4}{x^2-R^2}\right|}.$$

Now we substitute x by $\frac{R}{\sqrt{\alpha}}$ to obtain

$$\begin{aligned}
\Delta x &< \sqrt{24\,\alpha\,\left|\frac{R^4}{\left(\frac{R}{\sqrt{\alpha}}\right)^2-R^2}\right|} \\
&< \sqrt{24\,\alpha\,\left|\frac{R^4}{\frac{R^2}{\alpha}-R^2}\right|} \\
&< \sqrt{24\,\alpha\,\left|\frac{R^2}{\frac{1}{\alpha}-1}\right|} \\
&< \sqrt{24\,\alpha\,\left|\frac{R^2\cdot\alpha}{1-\alpha}\right|}.
\end{aligned}$$

For small values of α one can estimate $1-\alpha \approx 1$ which results in

$$\Delta x \;<\; \sqrt{24}\,\alpha\,R, \tag{3.42}$$

so that the standard deviation R of the Gaussian-distribution can be used as parameter describing the spatial resolution of the system: Distributions with standard deviations smaller than R can show additional errors.

3.6.2 Second Condition: Approximation of Moments

The discrete diffusion model shown in Equation (3.36) is very different from the Langevin-equation (cf. Equation (3.5)). Particularly, it is less similar to the Langevin-equation than the random walk:

$$x(t+\tilde{\tau}) - x(t) \;= \tilde{\tau}\,\tfrac{f_A(x(t))}{\gamma_A} + \sqrt{2\,D_A\,\tilde{\tau}}\,W\,. \tag{3.43}$$

In this equation γ_A is the friction coefficient of substrate A and $f_A(x(t))$ is the force on A as a function of the location x and the time t. W is a normally distributed random number. One has to impose the requirement that both dynamical models result in nearly the same distribution of particles, where these distributions will be characterized by the expectation and variance.

The expectation and variance of the random walk (cf. Equation (3.43)) are

$$E_{rw} = \tilde{\tau}\,\frac{f_A(x)}{\gamma_A}\,, \quad V_{rw} = 2\,D_A\,\tilde{\tau}\,. \tag{3.44}$$

On the other hand, the expectancy of the discrete diffusion model (cf. Equation (3.36)) is

$$E_{ddm} \;=\; \Delta x\,(q(x+\Delta x)\;-\;q(x-\Delta x)).$$

By using the definitions for the transition probabilities given by Equation (3.35) and setting $\Delta t = \tau$, the equation changes to

$$\begin{aligned} E_{ddm} &= \Delta x\,\left(\frac{\tau\,D}{\Delta x^2} + \frac{\tau\,f(x+\Delta x)}{2\,\gamma\,\Delta x} - \left(\frac{\tau\,D}{\Delta x^2} - \frac{\tau\,f(x-\Delta x)}{2\,\gamma\,\Delta x}\right)\right) \\ &= \frac{\tau}{2\,\gamma}\,(f(x+\Delta x)\;+\;f(x-\Delta x))\,. \end{aligned}$$

Replacing the functions $f(x+\Delta x)$ and $f(x-\Delta x)$ by the corresponding Taylor-polynomials we obtain for the expectancy of the direct diffusion model

$$E_{ddm} \;=\; \frac{\tau\,f(x)}{\gamma} + \tilde{\tau}\,\frac{f_A''(x)\,(\Delta x)^2}{2\,\gamma_A} + \mathcal{O}(\Delta x^4). \tag{3.45}$$

The variance is described as the sum over the three jump options (left, right and stay) by

$$\begin{aligned} V_{ddm} &= \Delta x^2\;\; q(i+1|i)\;+\;\Delta x^2\;\; q(i-1|i)\;+\;0^2\;\; q(i|i) \\ &= \Delta x^2\;\; \big(q(i+1|i)\;+\;q(i-1|i)\big). \end{aligned}$$

We replace again the transition probabilities by their definitions given by Equation (3.35) and set $\Delta t = \tilde{\tau}$ to obtain

$$= \;\; \tilde{\tau}\,(2\,D_A + \frac{(f_A(x+\Delta x) - f_A(x-\Delta x))\,\Delta x}{2\,\gamma_A}) - \left(\tilde{\tau}\,\frac{f_A(x+\Delta x) + f_A(x-\Delta x)}{2\,\gamma_A}\right)^2.$$

Finally by applying Taylor-polynomials of the involved functions the variance can be defined as

$$V_{ddm} \;=\; \tilde{\tau}\,(2\,D_A + \frac{f_A'(x)\,\Delta x^2}{\gamma_A}) - \left(\tilde{\tau}\,\frac{f_A(x)}{\gamma_A}\right)^2 + \mathcal{O}(\Delta x^4)\,. \tag{3.46}$$

We therefore consider V_{ddm} as nearly identical to V_{rw} if for a value $\alpha \in [0,1]$ the following conditions hold:

$$2\,\alpha\,D_A > \frac{f'_A(x)\,\Delta x^2}{\gamma_A} \;\Leftrightarrow\; \Delta x < \sqrt{\frac{2\,\alpha\,k_B\,T}{|f'|}} \quad \left(\text{with } D_A = \frac{k_B T}{\gamma}\right),$$

and

$$2\,\alpha\,D_A\,\tilde{\tau} > \left(\frac{f_A(x(t))\tilde{\tau}}{\gamma_A}\right)^2 \;\Leftrightarrow\; \tilde{\tau} < \frac{2\,\alpha\,k_B\,T\,\gamma_A}{|f_A|^2}\,. \tag{3.47}$$

Furthermore we consider E_{ddm} as nearly identical to E_{rw} if for a value $\alpha \in [0,1]$ the following condition holds:

$$\begin{aligned} E_{ddm} - E_{rm} &< \alpha\,E_{rm} \\ E_{ddm} &< (\alpha+1)\,E_{rm} \\ 1 + \frac{f''_A(x)\Delta x^2}{2} &< \alpha+1 \\ \Delta x &< \sqrt{\frac{2 f_A \alpha}{f''_A}} \end{aligned}$$

Concluding, we define that the random walk and the discrete diffusion model lead to nearly the same distributions, if

$$|E_{ddm} - E_{rm}| < \alpha\,|E_{rm}| \;\text{ and }\; |V_{ddm} - V_{rm}| < \alpha\,V_{rm}\,. \tag{3.48}$$

By neglecting terms of the order Δx^4, these conditions are fulfilled if

$$\Delta x < \min\left\{\sqrt{\frac{2\,f_A\,\alpha}{f''_A}}\,,\,\sqrt{\frac{2\,\alpha\,k_B\,T}{f'_A}}\right\}, \tag{3.49}$$

$$\tilde{\tau} < \min\left\{\frac{2\,\alpha\,k_B\,T\,\gamma_A}{|f_A|^2}\right\}, \tag{3.50}$$

T is the temperature and k_B the Boltzmann's-constant.

3.6.3 Third Condition: Positive Probabilities

In the discrete diffusion model (3.36) the transition probabilities q between adjacent volume elements or lattice points are described as (cf. Equation (3.35)):

$$q(i+\nu|i;\Delta t) := \Delta t\,\left(\frac{D_A}{(\Delta x)^2} + \nu\,\frac{f_A(i+\nu)}{2\,\gamma_A\,\Delta x}\right). \tag{3.51}$$

To guarantee positive transition probabilities one has to demand:

$$0 < q(i+\nu|i;\Delta t) \;\Leftrightarrow\; \Delta x < \frac{2\,k_B\,T}{f_{max}(A)} \quad \left(D_A = \frac{k_B\,T}{\gamma_A}\right). \tag{3.52}$$

The variables are named like in the sections above. $f_{max}(A)$ is the maximal force on A within the observed spatial interval. This Equation (3.51) will be used as an additional criterion for the distance between two lattice points Δx.

3.6.4 Fourth Condition: Small Changes of Particle Numbers

In each time step of length $\tilde{\tau}_A$, the probability of a transition from any lattice point i must be smaller than α, which implies

$$q(i+1|i) + q(i-1|i) = \tilde{\tau}_A \left(\tfrac{2\,D_A}{(\Delta x)^2} + \tfrac{f'_{max}(A)}{\gamma_A} \right) \leq \alpha\,. \tag{3.53}$$

By inserting Equation (3.49), Equation (3.53) can be rewritten as

$$\tilde{\tau}_A < \tfrac{\alpha\,(\Delta x)^2}{2\,D_A\,(1+\alpha)}\,. \tag{3.54}$$

3.7 Summary of Formulas for Δx and τ

To derive an appropriate lattice distance Δx, one can use Equation (3.49), (3.52) and (3.42). On the other hand, it is also desirable that the length of the interval $[a, b]$, in which the system is simulated, is a natural multiple of the lattice distance Δx. Hence, we define

$$\Delta x := \frac{b-a}{\left[\frac{b-a}{\delta}\right] + 1}\,, \qquad ([\,x\,] := \max\{n \in \mathbb{Z} \,|\, n \leq x\,\})\,, \tag{3.55}$$

where

$$\delta := \min\left\{ \sqrt{\tfrac{2\,f_A}{f''_A}}\,, \sqrt{\tfrac{2\,\alpha\,k_B\,T}{f'_A}}\,,\ \tfrac{2\,k_B\,T}{f_A},\ \sqrt{24}\,\alpha\,R \,|\, x \in [a,b]\,,\ A \in \mathbb{S} \right\} \tag{3.56}$$

is the generalization of Equations (3.49), (3.52) and (3.42) to systems with many substrates.

Starting from this value for Δx, one can compute for each substrate A the length Δt of a time step in the following fashion (cf. Equations (3.50) and (3.54))

$$\Delta t := \min\left\{ \tfrac{2\,\alpha\,k_B\,T\,\gamma_A}{\max\{f_A^2\}}, \tfrac{\alpha\,(\Delta x)^2}{2\,D_A\,(1+\alpha)} | x \in [a,b],\ A \in \mathbb{S} \right\}. \tag{3.57}$$

3.8 Calculation of Transitions

As described in Section 3.2 one basic idea of COAST is to subdivide the system into independent subprocesses: First, the diffusion of different substrates are independent

processes. Second, the transitions of the same substrates from different lattice points i and j are independent processes. By using the probabilities $p := q(i+1|i)$(for $i \to i+1$) and $p := \frac{q(i-1|i)}{1-q(i+1|i)}$ (for $i \to i-1$), the transition numbers $i \to i+1$ and $i \to i-1$) can be computed successively without additional errors.

Starting from these probabilities, I will present in this section three methods to compute the number of transitions in one of the directions $i \to i \pm 1$. The choice of the method depends on the number of particles N_i at a lattice point i.

Exact stochastic model (Σ): For small numbers of particles the transitions from lattice point i to $i \pm 1$ can be computed by successive evaluation of binomial-distributions of the form (N_i:=number of particles at point i)

$$P_B(\kappa, N_i) = \left(\begin{array}{c} N_i \\ \kappa \end{array} \right) p^\kappa \, (1-p)^{N_i - \kappa} \, , \tag{3.58}$$

where one has to use suitable conditioned probabilities for the second transition (cf. Figure 3.8).

Approximative stochastic model (Γ): For sufficiently large N_i, Equation (3.58) can be approximated by the distribution of the random variable

$$Z := \text{Round}\left(p\, N_i + X \sqrt{N_i\, p\, (1-p)} \right) \, , \tag{3.59}$$

where X is a normally distributed probability variable and where Round is given by Equation (3.3).
In more detail:
If $P_G(\kappa, N_i)$ is the probability for $Z = \kappa$, then $\sup\{|P_B(\kappa, N_i) - P_G(\kappa, N_i)|\} < \alpha$ (cf. Equation (3.5) and (1) for details), for

$$N_i \geq \tfrac{1}{3\,\alpha\, p\,(1-p)} \, . \tag{3.60}$$

Partial differential equation (Δ): It is described in Equation (3.17), that the deviations from the expectation E are, with probability $1-\alpha$, smaller than $\alpha\, E$ when

$$N_i \geq \tfrac{1-p}{\alpha^3\, p} \, . \tag{3.61}$$

In this case, the deterministic description can be applied:

$$\kappa := \text{Round}\,(N_i\, p) \tag{3.62}$$

In COAST, Equations (3.59) and (3.62) will be used for efficient computations of the transition numbers in the case of intermediate and large particle numbers.

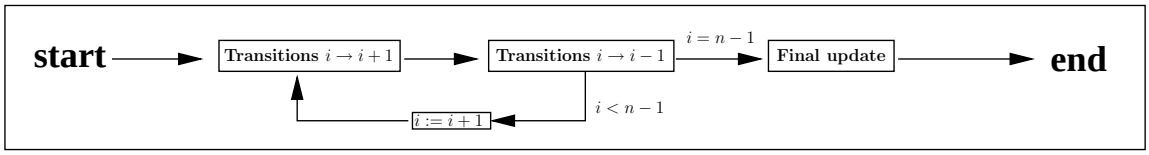

Figure 3.6: *Schematic representation of* COAST. *The scheme illustrates the computation of the transitions for a substrate S during $[t, t+\Delta t[$. In doing so, one starts at lattice point $-n+1$ and ends at $n-1$.*

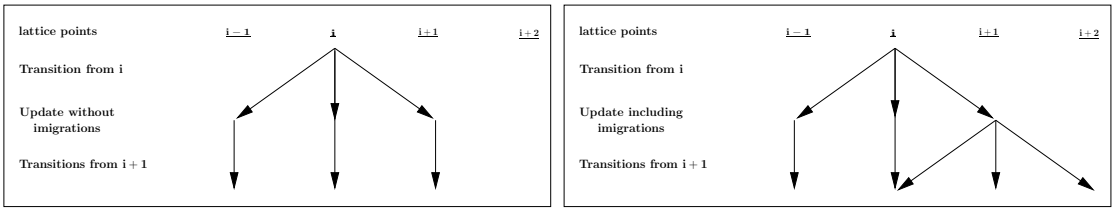

Figure 3.7: *Comparison between an immediate update including immigrations (left) and a consideration of immigrations in a final update after computing all transitions (right). It is assumed that at time t there is a single particle at lattice point i. The successive computation of the transitions from the different lattice leading in the left scenario to articifical, asymmetric transition scheme.*

3.9 The Algorithm

3.9.1 Overview

Assume that space and time coordinates have been discretized by using the parameters Δx (cf. Equation (3.55)) and Δt (cf. Equation (3.57)). Furthermore, suppose that the discretization of the space coordinate x has led to $2n+1$ lattice points $i \in \{-n, ..., n\}$, where $\varrho(\pm n) = 0$ reflects open boundary conditions. Then, the application of COAST to diffusion follows the scheme shown in Figure 3.6: For each substrate A, the computation of the thermal motions in a time interval $[t, t+\Delta t[$ one computes successively the transitions from each lattice point $i \in \{-n+1, ..., n-1\}$, where first of all one always computes the number of transitions in the positive direction $i \to i+1$ and then the transitions in negative direction (cf. Section 3.8). The number of transitions are computed in the following fashion:

Firstly, the subroutine "Transitions" (cf. Section 3.9.2) is used to compute the number of transitions from i to $i+1$. Then the same subroutine is used to determine the transitions from i to $i-1$. Subroutine "Transitions" also includes an update restricted to a reduction of particle numbers due to emigrations. The other part of the update, namely the increase of particle numbers due to immigrations, is shifted to

the subroutine "Final update", performed after the computation of all transitions in the time interval $[t, t+\Delta t[$. Note that this split of the particle update is necessary (cf. Figure 3.7): A complete update immediately after the computation of the transitions from a lattice point i would lead to the artifact that, in a time interval $[t_0, t_0 + \Delta t[$, a particle can jump from lattice point i to all lattice points $i + j$ with $j > 0$ (in the force free case: with probability q^j), but to no lattice point $i - k$ with $k > 1$. The update of the particle numbers at a specific lattice point after one direction, for example $\gamma = 1$, has been processed, is necessary. Otherwise the possible amount of transitions calculated for $\gamma = -1$ may be larger than what would be left after the first transition. This could result in negative particle numbers.

In what follows, the two subroutines "Transitions" and "Final update" are presented in more detail.

3.9.2 Subroutines

Transitions In a first step, one has to define the transition probabilities. Assume we calculate firstly the transition $i \to i+1$, then the transition probability $q_i^{(+1)}$ can be used. However, for the subsequently computed number of transitions $i \to i-1$, one must not use $q_i^{(-1)}$, but the conditioned probability that there was no transition $i \to i+1$ (cf. Figure 3.8).

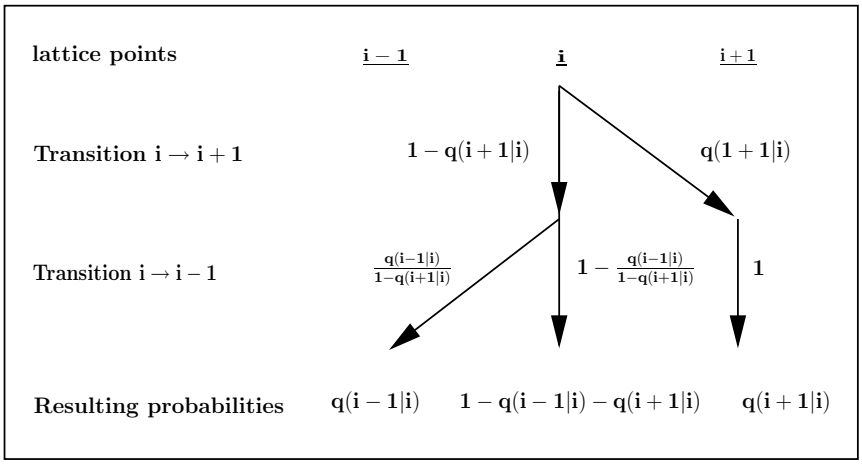

Figure 3.8: *Successive computation of the transitions from lattice point i in the different directions. The numbers at the edges of the graph are the probabilities used in the two steps. The resulting probabilities, given by the products of the probabilities in both steps, are in agreement with the correct transition probabilities.*

Secondly, the number of transitions $\kappa(i + \gamma|i)$ from i to $i + \gamma$ is computed, where three modeling levels are used:

If the criterion in Equation (3.60) is not fulfilled, two binomial distributions (cf. Equation (3.58)) are evaluated, which is defined as the Σ-regime.

If Equation (3.60) is valid, but Equation (3.61) is not, then $\kappa(i+\nu|i)$ is computed by evaluating two Gaussian distributions (cf. Equation (3.59)), where one has to take care that neither the number of particles nor the number of transitions become negative. This is the Γ-regime.

Finally, if Equation (3.61) is valid, the deterministic description (cf. Equation (3.62)) is used, which is named the Δ-regime.

In the last step, an update of the particle numbers is performed, which is restricted to the reduction of the particle number N_i due to emigrations.

Step 1: Defining the probabilities:

$$p := \begin{cases} q(i+1|i)\,, & \text{if } \gamma = +1\,, \\ \frac{q(i-1|i)}{1-q(i+1|i)}\,, & \text{if } \gamma = -1\,, \end{cases}$$

Step 2: Compute the number of transitions $\kappa(i+\gamma|i)$:
Σ : If $N_i \leq (3\,\alpha\,p\,(1-p))^{-1}$:

$$\kappa(i+\gamma|i) := \max\left\{ m \in \mathbb{N}_0 \,|\, \sum_{k=0}^{m} P(k, N_i, p) < r \right\}\,,$$

where r is a random number equidistributed in $[0,1]$ and P_B follows Equation (3.58) so that

$$\begin{aligned} P(0, N_i, p) &:= (\,1-p\,)^{N_i} \\ P(l+1, N_i, p) &:= \tfrac{(N_i-l)\,p}{(l+1)\,(1-p)}\,P(l, N_i, p) \quad (l \geq 0)\,. \end{aligned}$$

Γ : If $(3\,\alpha\,p\,(1-p))^{-1} < N_i \leq (\alpha^3\,p\,(1-p))^{-1}$:

$$\kappa(i+\nu|i) := \begin{cases} 0\,, & \text{if } X < -\sqrt{\frac{N_i\,p}{1-p}}\,, \\ N_i\,, & \text{if } X > \sqrt{\frac{N_i\,(1-p)}{p}}\,, \\ \text{Round}\left(\sqrt{N_i\,p\,(1-p)}\,X + N_i\,p\right), & \text{otherwise}\,. \end{cases}$$

with normally distributed random variable X.

Δ :If $(\alpha^3\,p\,(1-p)^{-1} < N_i$:

$$\kappa(i+\gamma|i) = \text{Round}\,(\,N_i\,p\,)\,.$$

Step 3: Update due to emigrations

$$N_i := N_i - \kappa(i + \gamma|i)$$

.

Final update due to immigrations After computing all $\kappa(i+\nu|i)$, a final update of the particle numbers is performed reflecting the additional particles due to immigrations.

Loop over all lattice points $(-n < i < n)$

$$N_i := N_i + \kappa(i|i+1) + \kappa(i|i-1)\,.$$

3.10 The Error Parameter α

After the development of COAST in the past paragraphs there are maybe some accentuations necessary regarding the error parameter α. Since it was the intention to present an algorithm depending on as least different parameters as necessary. The decision was made that all occurring approximations during the derivation of necessary formulas for COAST had at the end to be adapted by only one single value, which then was defined by α.

I am fully aware of the fact that α is approximating relatively different errors. In the first paragraph about the "Exact Stochastic Model" in Section 3.2.1 α is meant as the maximal deviation allowed for the change of the reaction probabilities for every reaction channel.
In the following paragraph "Approximative Stochastic Model" α describes the supremum norm of the two distributions.

Then in "Deterministic Reaction Kinetics" α is used in two ways. First of all α is set as the error probability for the Tschebyscheff inequality and afterwards a second error parameter is introduced which quantifies the expression of the expectancy μ being much larger $\sqrt{\frac{V}{N \cdot \alpha}}$ (α describes the width of an interval). To simplify the resulting formula (cf. Equation (3.18))the second error is assumed to be of the same value as α.

To adopt *COAST* to diffusion processes *alpha* had to gain an additional meaning in Section 3.6.4. Here the sum of the transition probabilities with respect to a single lattice point or (equivalently volume element) has to be smaller than α.

4 Test Simulations

In chapter one I have introduced the problem of simulating reaction and diffusion processes in cellular structures. Chapter two gave an overview on existing methods covering solution strategies for this question. In the last chapter I presented my own hybrid approach, COAST, the *Controllable Approximative Stochastic reaction-algorithm*. I have explained how differently this algorithm is working depending on the reaction probabilities. It uses three levels of modeling, an exact stochastic method, an approximative method based on Gaussian distributions and a deterministic method. The switching between the three levels is controlled by one error control parameter.

Furthermore, I have explained how the basic ideas of COAST can be applied to diffusion problems and presented the mathematical background for diffusion in one dimension.

In the upcoming chapter I demonstrate the ability of COAST to cope with basic reaction problems as well as multi scale scenarios like the *Oregonator* and the *Circadian Clock*. COAST gives very reliable results even better than demanded by the error parameter α. Especially due to its second modeling level COAST outruns many of the existing implementations based on exact methods or binomial distributions.

The application of COAST to linear diffusion is also tested in this chapter. I have chosen sceneries with and without an external forcefield and COAST reflects very well the results predicted by random walk simulations, but with a much better runtime behavior.

4.1 Test Simulations Using COAST

In this section, an assessment of COAST will be performed. To this aim, I will compare COAST and the *First Reaction Method* by comparing simulation results and computational costs of the two approaches for different reaction systems [Gillespie, 1976, 1977].

Particularly, I will consider the influence of the error control parameter α on the outcomes of the simulations and on the computing time.

4.1.1 Basic Systems

To begin with, I consider the two elementary chemical reaction-systems given by

$$A + B \overset{k_1}{\rightleftharpoons} C \qquad \text{and} \qquad A + B \overset{k_2}{\rightleftharpoons} 2\,C\,, \tag{4.1}$$

where, in both systems, the forward- and the backward-reactions have the same deterministic reaction rates k_1 or k_2 respectively. Both reaction rates are linked to the stochastic reaction constants c_μ used in COAST and the FRM via [Gillespie, 1976, 1977]:

$$c_\mu := \begin{cases} \frac{k_\mu}{V}\,, & \text{for } X \to P, X + Y \to P\,, \\ \frac{k_\mu}{V} \cdot 2\,, & \text{for } 2\,X \to P, \end{cases} \tag{4.2}$$

where V is the volume of the reaction system and P an arbitrary product. In all simulations performed here, $k_1 = k_2 = 0.2\frac{1}{s}$, $V := 1$, and $N_C(0) = 0$.
To compare the computational cost, both systems were simulated by the FRM, by the *τ-leap method* [Dhar et al., 2005], and by COAST for different initial values $N_A(0) = N_B(0)$. In doing so, $\alpha = 0.05$ was set for A+B $\overset{0.2}{\rightleftharpoons}$ C and $\alpha = 0.03$ for A+B $\overset{0.2}{\rightleftharpoons}$2C.

The run time of these simulations are monitored, the results of which are summarized in Figure 4.1. Since it is not my intention to discuss the effects of different implementations, but rather the effects of different algorithms, I do not present absolute run times. Instead, I have defined the run time of the simulation for each algorithm with $N = 100$ as 1. Furthermore, to illustrate the effects of the different modeling levels in COAST on the run time, the relative frequencies of the usage of model class Γ are also shown in the same figure.

As can be seen from Figure 4.1, all algorithms were noticeably fast for small $N_A(0)$. However, the run time behavior of the FRM and τ-leap was qualitatively different from the run time behavior of COAST, when modeling levels Γ and Δ are predominantly used in COAST. To illustrate these different behaviors, I performed least mean square (lms)-fits of the measured run times in the range of particle numbers dominated by Γ and Δ. For A+B $\rightleftharpoons$ C, the leading terms of these fitted functions were proportional to $N_A^{1.01}$ for FRM, proportional to $N_A^{0.98}$ for the *τ-leap method*, but proportional to $N_A^{0.4}$ for COAST. Similarly, the fit curves for A+B $\rightleftharpoons$ 2 C were proportional to $N_A^{1.99}$ for FRM, to $N_A^{1.97}$ for τ-leap, but proportional to $N_A^{0.96}$ for COAST. The reasons for these different behaviors will be discussed in Section 5.1.

For A+B $\rightleftharpoons$ C, the fraction of the stochastic model Γ is decreasing for large particle numbers, which reflects the increasing usage of the deterministic modeling level Δ. As one can see, the increasing usage of Δ does not lead to a strong reduction of computational costs when compared to the costs of stochastic model Γ.

In order to investigate whether COAST is able to reproduce the results of the FRM, I simulated both reaction systems by FRM and by COAST with different

values of α, and initial values $N_A = N_B = 10000$ ($k_1 = k_2 = 0.2\frac{1}{s}$; $V = 1$; $N_C(0) = 0$). The simulation times were $t = 0.5s$ for $A + B \leftrightharpoons C$ and $t = 1s$ for $A + B \leftrightharpoons 2\,C$. Since both the FRM and COAST are stochastic algorithms, one cannot compare a single COAST-run with a single FRM-run. Therefore one must compare collections of identical simulations. Accordingly, I repeated all simulations 1000 times and stored N_A at the end of each of these runs, which is, due to the conservation laws

$$N_A(t) - N_B(t) = \text{const} \tag{4.3}$$

and

$$\begin{aligned} N_A(t) + N_B(t) + 2\,N_C(t) &= \text{const}\,, \quad \text{for } A + B \leftrightharpoons C\,, \\ N_A(t) + N_B(t) + N_C(t) &= \text{const}\,, \quad \text{for } A + B \leftrightharpoons 2\,C\,, \end{aligned} \tag{4.4}$$

sufficient to also characterize N_B and N_C.

The description by deterministic reaction kinetics leads to the ODEs:

$$\begin{aligned} A + B \rightleftharpoons C:\ \dot{N}_A &= -k\,N_A^2 + k\,(10000 - N_A)\,, \\ A + B \rightleftharpoons 2\,C:\ \dot{N}_A &= -k\,N_A^2 + k\,(20000 - N_A)^2\,, \end{aligned} \tag{4.5}$$

where the conservation laws in Equation (4.3) and Equation (4.4) as well as the initial condition $N_A(0) = N_B(0) = 10000$, $N_C(0) = 0$ are used. The equilibrium states of these models, which are defined by $\dot{N}_A = 0$, are given by

$$\begin{aligned} A + B \rightleftharpoons C:&\quad N_A = 99.5\,, \\ A + B \rightleftharpoons 2\,C:&\quad N_A = 6666.7\,. \end{aligned} \tag{4.6}$$

The derivations of the equilibria can be found in the Appendix A. Since the outcomes of the 1000 runs with identical algorithms are given by independent, identically distributed random variables, the collections of outcomes are always approximations of Gaussian-distributions [Feller, 1970, p. 182 f.] completely defined by their expectations and their variances.

In Figure 4.2, the time-evolution of N_A is given for both systems, where the outcomes of COAST ($\alpha = 0.05$) are compared with the results of the deterministic reaction kinetics. Obviously, one cannot observe systematic deviations between the results of COAST and the values of deterministic reaction kinetics.

Figure 4.3 shows the mean value and the standard deviation of N_A at the end of the simulations. These values are shown for COAST-simulations as a function of α. These values are compared with the outcomes of FRM-simulations and the equilibrium values of the deterministic reaction kinetics (cf. Equation (4.6)), these are $N_A = 99.25$ ($A + B \rightleftharpoons C$) and $N_A = 6666$ ($A + B \rightleftharpoons 2\,C$). Thus, the graphs f

and g, given the values of the deterministic reaction kinetics times $(1 \pm \alpha)$, can be used to illustrate whether or not the algorithm is as good as stated.

For both reaction systems, the mean values of FRM are in agreement with the results of the deterministic reaction kinetics. Furthermore, the deviations between the outcome of COAST and the results of FRM are much smaller and therefore even better than the promised $\alpha \cdot 100\,\%$.

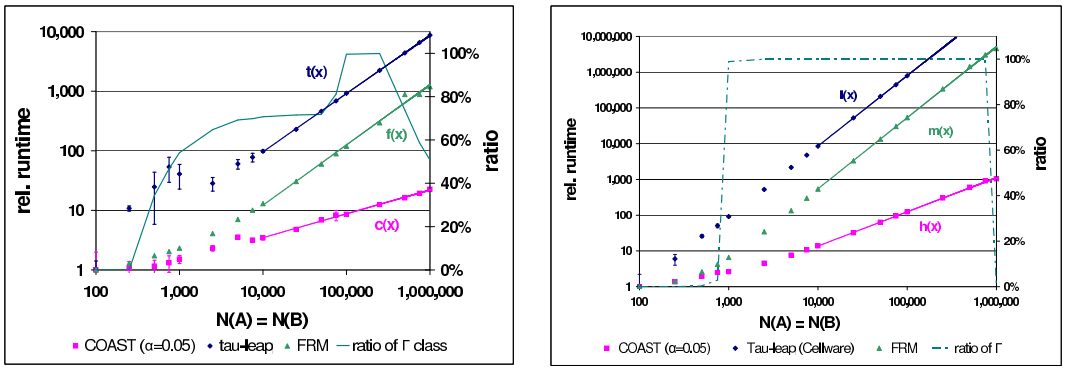

Figure 4.1: *Run time behavior of* COAST , FRM *and* τ-leap method *for* $A+B \rightleftharpoons C$ *(left) (simulation time t=100s) and* $A+B \overset{0.2}{\rightleftharpoons} 2C$ *(right) (simulation time t=0.3s), where* $\alpha = 0.05$ *was used for* COAST. *In all cases, the run times of the simulations with* N_A*=100 were defined as 1. Additionally, the amount of reaction channels evaluated in* COAST *by* Γ *and* Δ *are shown. The following functions were determined by least mean square fit to the run times of the different algorithms led to:* $t \propto N_A^{0.98}$ *(*τ*-leap),* $f \propto N_A^{1.01}$ *(*FRM*),* $c \propto N_A^{0.40}$ *(*COAST*);* $l \propto N_A^{1.97}$ *(*τ*-leap),* $m \propto N_A^{1.99}$ *(*FRM*),* $h \propto N_A^{0.96}$ *(*COAST*).*

These results bring up the obvious question of how useful the Δ-regime is within COAST. This question is not easily answered. Without any doubt in theory there is an advantage by using no random number to using one like in the case of the modeling level of Γ. I performed an experiment to reveal the influence of the Δ-regime on the performance of a simulation (cf. Figure 4.4). To test the speedup, I had to generate a system in which the educts are kept constant, so their changing does not have an influence on the reaction. The model was found to be a reaction of the kind:

$$\overline{A} + \overline{B} \rightarrow C$$

with the parameters k=1, V=1, δt=0.001s, t=0.018s, α=0.05

The values presented are the mean of ten single runs. One can observe that at an initial concentration of more than 310,000 particles of A and B, the criterion in Equation (3.18) is fulfilled and for A=B=320,000 particles 99,99% of all reactions

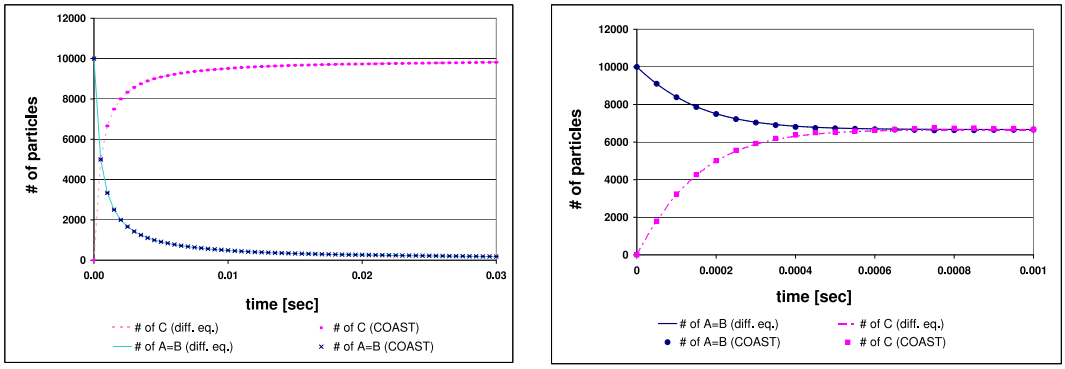

Figure 4.2: *Time-evolution of the particle numbers N_A in both systems for* COAST*-simulations ($\alpha = 0.05$) and for the deterministic reaction kinetics. The left (right) diagram shows the behavior for $A + B \leftrightharpoons C$ ($A + B \leftrightharpoons 2\,C$).*

are calculated by the Δ-regime. The graph is rising constantly because the time intervals decrease with increasing particle numbers, and therefore the computational effort (cf. Equation (3.22)). The benefit is a reduction of the run time by 10% and a better linear run time bahavior for higher particle numbers.

However, it is difficult to say how likely the activation of the Δ-regime is in a specific case. This has to be tested individually.

4.1.2 The Oregonator

In addition to the very simple system described in Section 4.1.1, I also investigated the behavior of COAST when simulating a more complex reaction system, namely the Oregonator. In this system, different substrates have radically different particle numbers, and the particle number of a given substrate is subject to strong fluctuations over time. Before presenting my simulation results, I will give some background information on this interesting topic.

Historical Background A chemical system, in which during the reaction the concentration of at least one specie periodically fluctuates, is called an oscillating system. The most famous oscillating chemical reaction is the Belousov-Zhabotinsky (BZ) reaction [Belousov, 1958]. It involves the oxidation of an organic acid by acidified bromate in the presence of a metal ion catalyst (often cerium ion). The BZ reaction is a classical example of instability and self organization in non equilibrium systems. Oscillating reactions can also be found in biological systems (e.g. oscillations at cell membranes, stimulus transition, oscillating enzyme reactions and circadian rhythm [Goldbeter, 1996]).

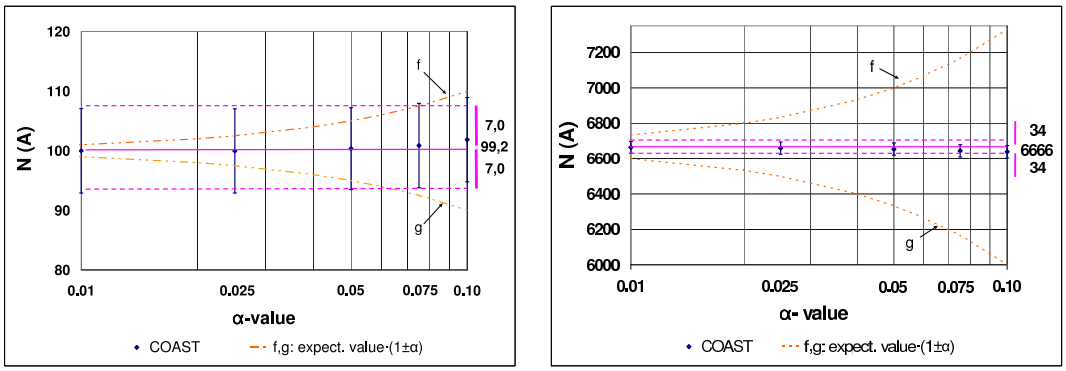

Figure 4.3: *Mean value and standard deviation of N_A for simulations of $A+B \overset{0.2}{\rightleftharpoons} C$ (left) and $A+B \overset{0.2}{\rightleftharpoons} 2C$ (right) in dependence from α. The error bars of the* FRM*-simulations are shown by the solid and the dashed lines. The corresponding values are given at the right hand side of the diagrams. The mean values of the* FRM*-simulations (99.25/6666) are in nearly perfect agreement with the equilibrium values of the ODEs (cf. Equation (4.6)). f and g are given by the values of the ODE times $(1 \pm \alpha)$.*

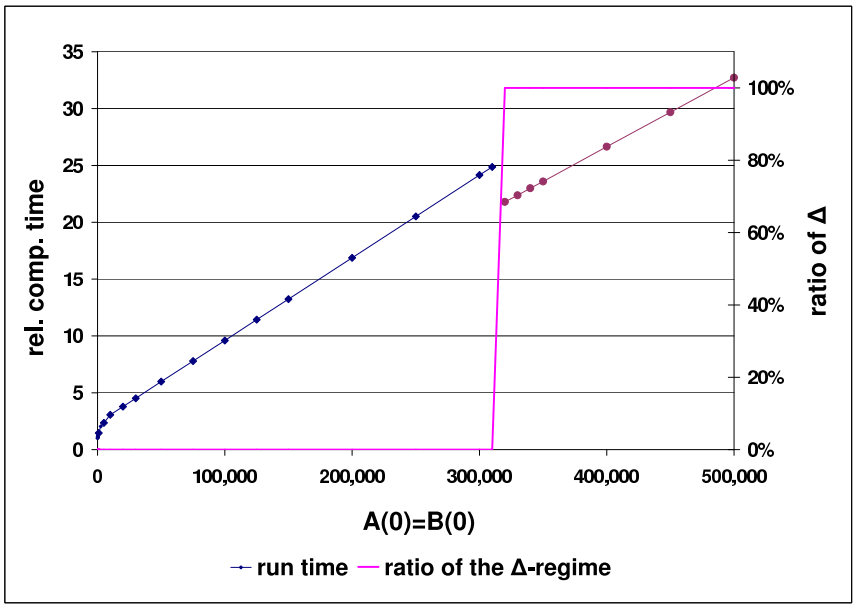

Figure 4.4: *The diagram shows the behavior for $A + B \rightarrow C$ with k=1, V=1, δt=0.001s, t=0.018s, α=0.05*

In August 1825, J.F.W.Herschel (∗1792, †1871) examined the passivity of iron in nitric acid. He discovered that the occurring reaction activity is oscillating between pure Fe and passive Fe_2O_3. But it lasted until 1828 when G.T.Fechner was the first to publish about an oscillating chemical reaction (with silver nitrate treated iron in sulfuric acid)[Fechner, 1828]. Other observations of temporal oscillating reactions followed [Schönbein, 1842, Joule, 1844].

F.F. Runge [1850], the father of paper chromatography, was the first to describe the spontaneous formation of spatial structures in chemical systems. Although similar discoveries were also made, scientists did not believe in chemical oscillations. The conversion of the educts should continue until thermodynamical equilibrium is reached. Oscillations were seen as a contradiction to the second law of thermodynamics. This demands that a closed system (system without energy and mass transfer) aspires towards an equilibrium.

In 1958, the Russian chemist B. Belousov discovered a homogenous oscillating reaction: He tried to oxidate citric acid in sulfuric acid with potassic bromate and a cerium(IV)-salt. He observed the rhythmic appearance of the yellow cerium(IV)-ion [Belousov, 1958, Tyson, 1976]. A. Zhabotinsky repeated his work in 1961, and improved the chemical compositions [Zhabotinsky, 1964]. It took until 1967 that western world became familiar with the results that have been made in the former Soviet Union.

The Belgian scientist I. Prigogine realized that classical thermodynamics only apply for closed systems, which are next to their equilibrium. All open systems (i.e. systems having energy and mass transfer with the surrounding) are in a state of non-equilibrium. Systems like the human body maintain their identity by means of energy flow from a variety of separate sources. Prigogine was able to demonstrate that these systems operate far from the realms of equilibrium and therefore could exhibit strange and unexpected behavior patterns (in full coherence with the second law of thermodynamics). Prigogine gave such systems the name *dissipative systems* [Glansdorff and Prigogine, 1971], because the ability to do work as a consequence of the increase entropy is being lost (dissipated) as the process unfolds.
He and his coworkers suggested a mathematical model of a chemical non-living dissipative system consisting of four single reactions. This model is known as the *Brüsselator*. His work was later recognized with a Nobel price in 1977, leading to full acceptance of oscillating reactions.

The *Brüsselator* has one major problem, it includes a trimolecular reaction, which can be regarded a quite unlikely. In 1972, R.J. Field, E. Körös and R.M. Noyes developed a mechanism for the BZ-reaction consisting of 18 single reactions with 21 different molecules [Noyes et al., 1972]. The Field-Koros-Noyes model can be broken down in 5 essential reactions: the *Oregonator* (named after their patron institution, the University of Oregon) and will be described in the following section.

The FKN-Model There are certain demands that have to be fulfilled so that oscillating reactions are likely to occur in a chemical system:

- the chemical system has to be far away from thermodynamic equilibrium (this is necessary to have chemical reactions at all)
- the chemical system must be an open system (so energy transfer and multiplication of entropy are possible)
- there have to be at least two meta-stable states in the system
- the chemical system must contain a feedback loop (with different impacts on the two states)

As with in all chemical reactions, the educts are consumed while the concentration of the products increase. If the concentration of the educts is too low, the reaction stops. In principle, all oscillating chemical systems are capable of developing spatial structures, because even small random gradients of concentration can be amplified. Only open systems allow undamped oscillations. Table 4.1 shows the five reactions of the FKN-model.

The FKN-model includes one auto-catalytic step with bromous acid ($HBrO_2$) as an auto catalytic intermediate product. Reactions one and two describe a negative feedback loop in which $HBrO_2$ is captured by bromide (auto inhibition). The auto-catalytic increase of bromous acid $HBrO_2$ is slowed down by the disproportion in reaction four. In reaction five bromide is reproduced and the catalyst is reduced under the influence of the organic compounds malonic acid (MA) and bromomalonic acid (BrMA). The oscillations occur because the system is changing between two conditions. In the reduced condition, with high bromide concentration, the catalyst is mainly present as cerium(III) and malonic acid is brominated. The bromide is reduced by the reaction with bromate. If the Br^- -concentration (bromide), is below a critical concentration the auto-catalytic reaction begins and Ce(III) is oxidized to Ce(IV). The system switches to the oxidized condition, which is characterized by high concentrations of $HBrO^-$ and Ce(IV) and by oxidation and bromination of the organic compounds.

In other words: reactions one and two consume bromide ions. If the amount of bromide-ions becomes too low, reaction two is no longer the dominant channel for reaction of $HBrO_2$, and reaction three takes over. In this auto-catalytic reaction, $HBrO_2$ is produced at a rate that depends on the $HBrO_2$ concentration. The growth of $HBrO_2$ is limited by reaction four, which accelerates as the $HBrO_2$ concentration increases. Reaction four has another important effect: it regenerates the reactant bromate. Reaction five regenerates now Ce^{3+} and Br^-. This last reaction is only important when the level of Ce(IV) is high enough. There is a delay between the

(1)	$Br^- + BrO_3{}^- + 2H^+$	$\rightarrow$	$HBrO_2 + HOBr$	$\overline{A} + B$	$\rightarrow$	C
(2)	$Br^- + HBrO_2 + H^+$	$\rightarrow$	$2HOBr$	$C + B$	$\rightarrow$	D
(3)	$2Ce^{3+} + BrO_3{}^- + HBrO_2 + 3H^+$	$\rightarrow$	$2Ce^{4+} + 2HBrO_2 + H_2O$	$\overline{E} + C$	$\rightarrow$	$2C + F$
(4)	$2HBrO_2$	$\rightarrow$	$HOBr + BrO_3{}^- + H^+$	$2C$	$\rightarrow$	G
(5)	$Ce^{4+} + org.comp.$	$\rightarrow$	$fBr^- + Ce^{3+} + org.comp.$	$\overline{H} + F$	$\rightarrow$	B

Table 4.1: *The five chemical reactions describing the Field-Körös-Noyes model of the oregonator.*
$\overline{A} = BrO_3{}^-(bromate)$; $\overline{H} = all\,oxidiz.org.species$; $D = HOBr\,(hypobromousacid)$; $C = HBrO_2(bromousacid)$; $B = Br^-(bromide)$; $F = Ce^{4+}(cerium - 4)$; $f \approx 1$; $\overline{E}, G = simplifications$

reactions which consume bromide and Ce^{3+}, and those which regenerate these reactants. As a result, the system cycles from high values of Ce^{3+} and Br^-, and back again. The oscillations can be nicely illustrated if the oscillating specie is colorfull. In this experiment the color of the Ce^{3+}-ion is magenta and the one of the Ce^{4+}-ion is blue.

Figure 4.5 presents such a system where instead of cer ferroin (cf. Appendix B.1) is used as redox indicator where iron is changing between two states. A ferroin solution is colloquial for a 1,10-phenanthroline ferrous sulfate solution ($(C_{12}H_8N_2)_3FeSO_4$). It is used as a redox indicator, because of its reversible color change from the red hexammineiron(II) complex (reduced form) to the blue hexammineiron(III) complex (oxidized form). The oxidized form is called ferriin. The reactions and composition of this experiment can be found in the appendix.

Experimentally, the H^+-concentration is held constant by a buffer system. Furthermore, the oscillations are observed when the bromate is in large excess, such that its concentration is approximately constant.

Chemical oscillators only appear to be contradictory to the second law of thermodynamics. But the changes one observes is only a small part of all reactions that are occurring. The important reaction is the oxidation of malonic acid by bromate. Their concentrations are constantly decreasing without oscillations, just by two reactions taking it in turns.

The experimental setup of the Oregonator The setup is identical to the implementation of Gillespie [1977], which he used for testing his SSA. It should be noted that there are tiny differences between the original FKN-model [Field and Noyes, 1974] and the setup of Gillespie, but these differences do not effect the outcome of the experiment. With this knowledge, the presented equations are used without any adjustments. This simple model consists of the following five reaction channels (values given are the deterministic rate constants):

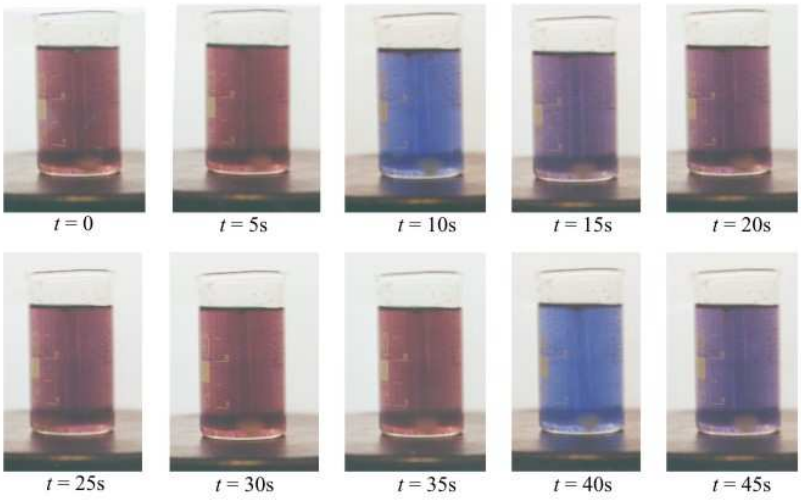

Figure 4.5: *This picture illustrates the temporal changes of the Belousov-Zhabotinsky reaction under the presence of ferroin as a redox indicator.*

$$\overline{A} + B \xrightarrow{0.004} C\,, \qquad C + B \xrightarrow{0.1} D\,, \qquad \overline{E} + C \xrightarrow{0.104} 2C + F\,,$$
$$2\,C \xrightarrow{0.008} G\,, \qquad \overline{H} + F \xrightarrow{0.013} B\,, \tag{4.7}$$

where $\overline{S}$ denotes that N_S is considered constant in time. This means the system is open for that species. Additionally, volume parameter V=1 and the following initial conditions are chosen ($x \in \mathbb{R}^+$):

$$N_A(0) = 500 \cdot x\,, \quad N_B(0) = 1000 \cdot x\,, \quad N_C(0) = 500 \cdot x\,, \quad N_D(0) = 0\,,$$
$$N_E(0) = 1000 \cdot x\,, \quad N_F(0) = 2000 \cdot x\,, \quad N_G(0) = 0\,, \quad N_H(0) = 2000 \cdot x\,. \tag{4.8}$$

This system was simulated for a time span $t = 1$ with $x = 5$ by FRM, and by COAST with $\alpha = 0.05$. In Figure 4.7, the time-behavior of N_F in the FRM- and COAST-simulation is monitored.

It is worth noting that the initial conditions shown in Equation (4.8) are the equilibrium state of the ordinary differential equation, so that an application of deterministic reaction kinetics results in time-constant particle numbers. Conversely, the oscillating particle numbers shown in Figure 4.7 are only due to the application of stochastic dynamics. Hence, it should come as no surprise that, N_F initially exhibits very different behavior in the two simulations. This behavior depends on the exact

fluctuations from the equilibrium state: For the FRM, N_F immediately decreases, whereas in the case of COAST, N_F increases to a small local maximum.

After this starting time, N_F oscillates in both simulations with nearly the same amplitude and nearly the same period. For the determination of the amplitudes and periods, I performed a simulation of three seconds of the Oregonator with a time resolution of 0.00005 s by both algorithms, using the same parameters as mentioned above. As a result, I obtained for FRM an amplitude of 42587 ± 471 and a periodic time of 0.1405 ± 0.0016, and for COAST an amplitude of 42355 ± 864 and a periodic time of 0.1405 ± 0.0016. A numerical solution of the ODEs from deterministic reaction kinetics led to an amplitude 42040 and a periodic time of 0.1405.

The determination of the amplitudes requires the computation of local extrema of the particle numbers, which is a non trivial task in stochastic systems. These difficulties are the reason for the different values of the variances of the amplitudes.

In order to demonstrate that the Oregonator is very suitable to test an algorithm, let us consider the time scales of the different reaction channels in this system.

For the FRM, the time of the next reaction in an arbitrary reaction channel μ is given by

$$\delta_\mu := \tfrac{-\ln(r)}{Q_\mu}\,, \tag{4.9}$$

where r is a random variable equidistributed in $[0,1]$, and where Q_μ is the propensity (cf. Equation (3.3.2)). Thus, the mean time until the next reaction is given by

$$\langle\, \delta_\mu \,\rangle = \int_0^1 \tfrac{-\ln(r)}{Q_\mu}\, dr = \tfrac{1}{Q_\mu}\,, \tag{4.10}$$

such that Q_μ^{-1} is an appropriate quantity to characterize the time scale of a reaction channel. Figure 4.6 shows Q_μ^{-1} dependently of time for three reaction channels, these are $B + C \to D$, $C + E \to 2\,C + F$, and $2\,C \to G$, where $x = 5$ was used again. For $2\,C \to G$, the expectation of the time step length Q_μ^{-1} has values between $10^{-1.9}$s and $10^{-6.2}$s, for $C + E \to 2\,C + F$ and $B + C \to D$, Q_μ^{-1} has values between $10^{-4.8}$s and $10^{-6.8}$s or $10^{-5.3}$s and $10^{-6.5}$s, respectively. The Q_μ^{-1} of the two other reaction channels are always between the values for $2\,C \to G$ and $C + E \to 2\,C + F$.

Thus, the Oregonator is not only a multiple time scale-system, but the time scales are subject to strong fluctuations. Consequently, the Oregonator is suitable to test both the ability of an algorithm to treat reactions with different time scales and the ability to adapt itself to rapidly changing conditions.

To allow a comparison between FRM and COAST, I also show in Figure 4.6 the time step length τ_μ of COAST for the three reaction channels $B + C \to D$, $C + E \to 2\,C + F$, and $2\,C \to G$. The most obvious differences between the time-behavior of the τ_μ and the Q_μ^{-1} is that τ_μ has larger values and a smoother behavior, where the larger values of the τ_μ's imply that COAST works faster than the FRM.

Since all reaction channels of the Oregonator are second-order reactions, all Q_μ^{-1} are proportional to x^{-2} (cf. Section (3.3.2)), where x is the scaling factor introduced in Equation (4.8). On the other hand, since l_μ (cf. Equation (3.29)) (the expected number of reactions of the channel μ) is in a first approximation proportional to the number of particles, the τ_μ -the timespan until all reactions l have occurred in channel μ- of second-order reactions (cf. Equation (3.22)) are proportional to x^{-1}. Consequently, one can expect that the computational cost is proportional to x^2 for FRM, but proportional to x for COAST.

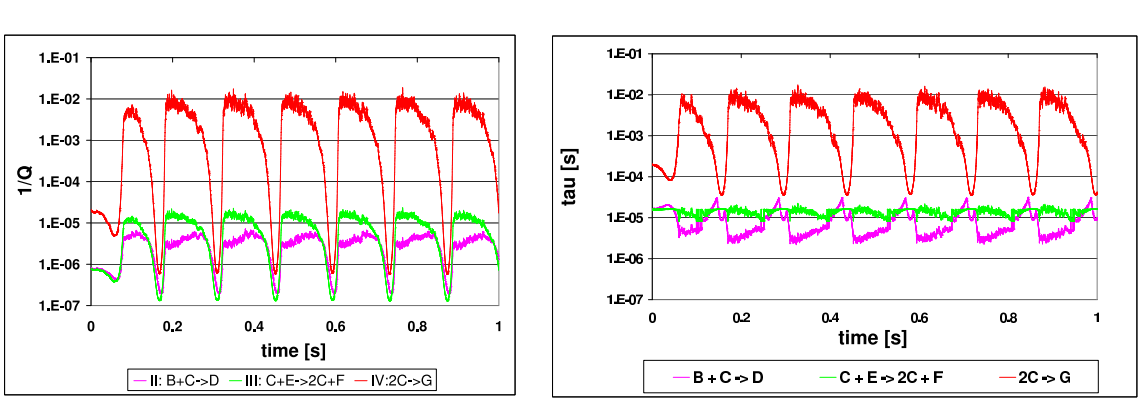

Figure 4.6: *Characterisation of the time-scales of reaction channels in the Oregonator with scaling factor $x = 5$ (cf. Equation (4.8)). Q_μ^{-1} (left diagram) is the expectation of the time span till the next reaction in the channel with the FRM. τ_μ (right diagram) is the possible length of a time step computed by* COAST *for a reaction channel.*

To test this hypothesis, I compared the run time of both methods. Therefore I performed again FRM- and COAST-simulations ($\alpha = 0.05$) with $t = 1s$ for different values of the factor x (cf. Equation (4.8)) and measured the run time of each of these simulations. The results are shown in Figure 4.8, where the portion of the reaction channels evaluated by Σ and Γ is also presented.

To characterize the asymptotic dependence of the run times on the particle number, least mean square fits were performed on the run times in the range with more than 80 % evaluations by the Γ-regime in COAST.

Again, one can see that the ratios between the run times of COAST and FRM decreases with larger numbers of Γ evaluations. Furthermore, as can be seen from the fitted functions, the asymptotic run time behavior of COAST is proportional to x, but proportional to $x^{1.9}$ for the FRM, which is consistent with the hypothesis about the run time behaviors of these systems derived from the analyses of the Q_μ^{-1} and τ_μ.

Figure 4.9 provides more insight information on this experiment by presenting more α-values. The total run time of this experiment was slightly changed to $t = 0.5s$. As

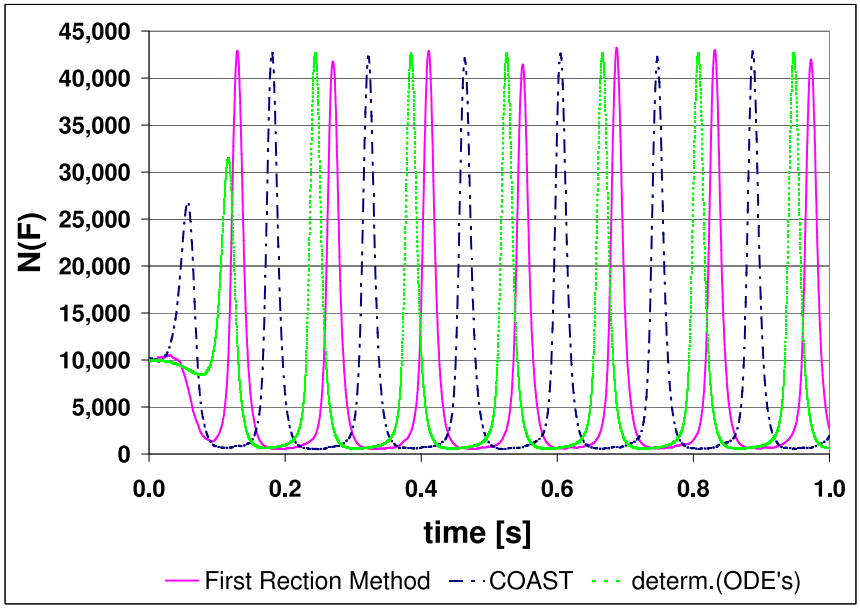

Figure 4.7: *Comparison of the time-evolution of N_F in the* COAST- *and FRM-simulations of the Oregonator. For* COAST, $\alpha = 0.05$ *was chosen. The initial values of the particle numbers were given by Equation (4.8) with* $x = 5$.

one can observe, the total run time for the COAST-experiments depends very much on the set α-value. The higher the α-value, the earlier the algorithm will switch from the Σ-regime to the Γ-regime, which processes the reaction much faster due to the fact, that it is using less random number operations. For an α-value of 5%, the Γ-regime is used very early and for an α-value equal to 1% relatively late as can be seen by observing the solid lines. Another interesting fact is, that the performance of COAST for an α-value of 0% compared to the FRM is worse. Since, in this case, COAST is in principle performing the same task as the FRM , but has an additional overhead to check the other two regimes, this is very reasonable.

This paragraph should have illustrated how the error-parameter α influences the performance of COAST. Therefore before setting α one has to consider that a higher α-value results in a better performance in terms of run time behavior, but also results in a lower accuracy. Furthermore a lower α-value slows down the processing, but increases the accuracy.

4.1.3 Circadian Clock

Background Information All eukaryotes (like plants, animals and fungi) and some prokaryotes (cyanobacteria) display changes in gene activity, biochemistry, physiol-

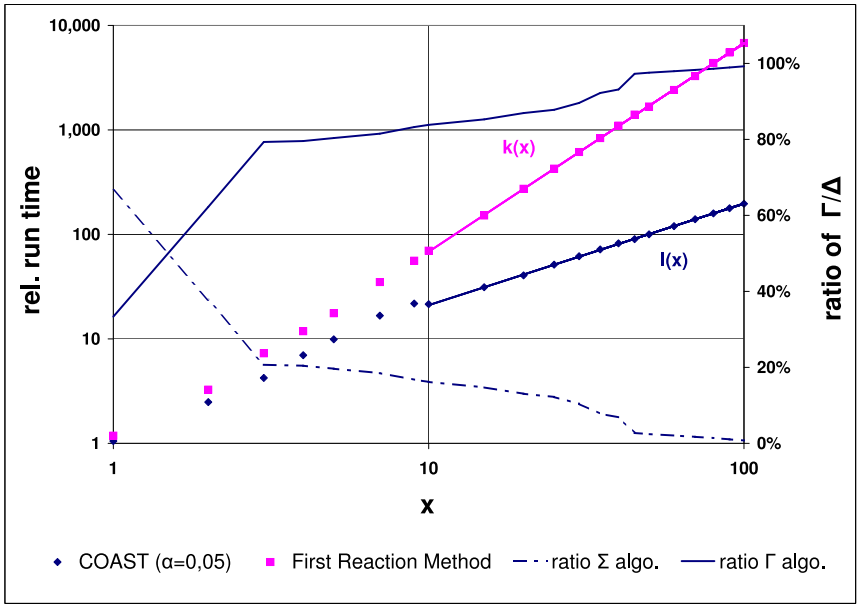

Figure 4.8: *Run time behavior of the* COAST- *($\alpha = 0.05$) and the* FRM-*simulations of the Oregonator in dependence of the initial values of the particle numbers parametrized by x (cf. Equation (4.8)). Additionally, the number of reactions calculated in* COAST *by the model classes Σ and Δ is shown. $k(x) := 0.7 \cdot x^{1,9}$ and $l(x) := 2.2 \cdot x^{1,0}$ are results of least mean square-fits to the run times of* FRM *($k(x)$) or* COAST *($l(x)$) in the interval $[10, 100]$, where in* COAST *more than 80 % evaluations are done by Γ.*

ogy and behavior through the cycle of days and nights. These endogen rhythms have a period length of approximately 24h and help the organism to adjust to daily repeating incidents, so called Circadian Clocks.
For most animals a pacemaker was able to be localized in the area of the visual system, but only for simple organisms the Circadian Clock behind these rhythms is already described.

The circadian model the following simulations are based on, was originally described by Barkai and Leibler [2000] and is founded on experimental results. Vilar et al. [2002], who did further research on this system describe the functionality of the Circadian Clock as following:

"The main characteristic is the presence of intracellular transcription regulation networks with a set of clock elements that give rise to stable oscillations in gene expression. A positive element activates genes coupled to the Circadian Clock. It simultaneously promotes the expression of a negative element, which in turn represses the positive element. The cycle completes itself upon degradation of the negative element and re-expression of the positive element."

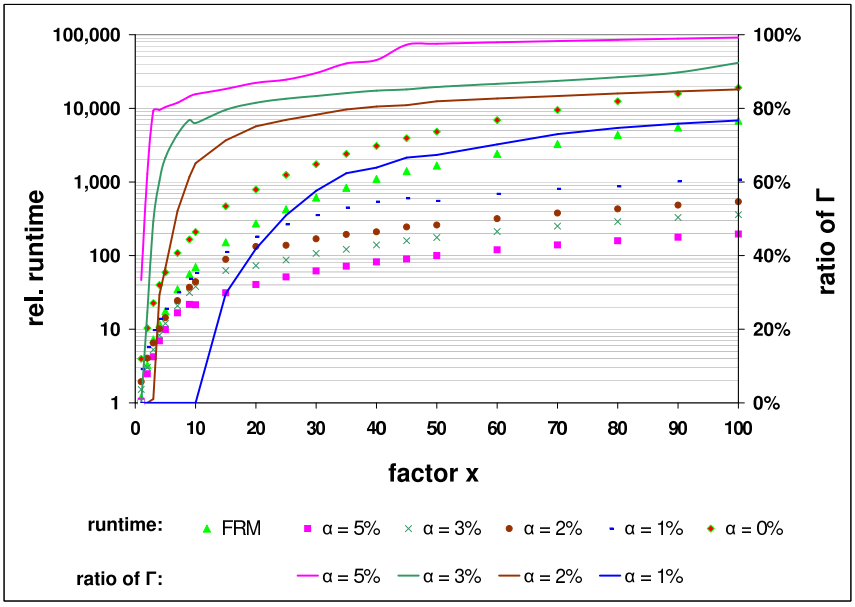

Figure 4.9: *Run time behavior of* COAST *($\alpha = 0, 0.01, 0.02, 0.03$ and 0.05) and the* FRM *of the Oregonator dependent on the initial values of the particle numbers parametrized by x (cf. Equation (4.8)). The ratio of reactions calculated in* COAST *by the model class Δ is presented by solid lines.*

The experimental System The model includes two genes, an activator-gene and a repressor-gene, which are transcribed into mRNA and translated into the products A and R. The two genes have promotor regions Pa and Pr. If the activator A binds to the promotors, the expression of the respective mRNAs (mRNAa, mRNAr) is enhanced. By forming a dimer with A, R is able to inhibit the activator. Figure 4.12 reflects the reaction channels of this system.

It may be worth noting that Pa, Pr, Pr-A and Pa-A are variables that can only take the values 0 or 1. Since I want to compare a deterministic simulation with a stochastic one, I allow continuous values between zero and one as also proposed by Vilar et al. [2002]. It has to be noted, that for ODEs the oscillations can disappear, but in a stochastic model the oscillations will persist. This phenomenon is a manifestation of "coherence resonance" and illustrates the crucial interplay between noise and dynamics.

To demonstrate that the Circadian Clock is also a multiple time scale-model, Figure 4.10 shows the expected length of the time steps Q_μ^{-1} (cf. Equation (4.10)) in the FRM-simulations; the illustration is restricted to the fastest and the slowest reaction channels. As one can easily see, there are five orders of magnitude between the fastest and the slowest reaction channels in the Circadian Clock.

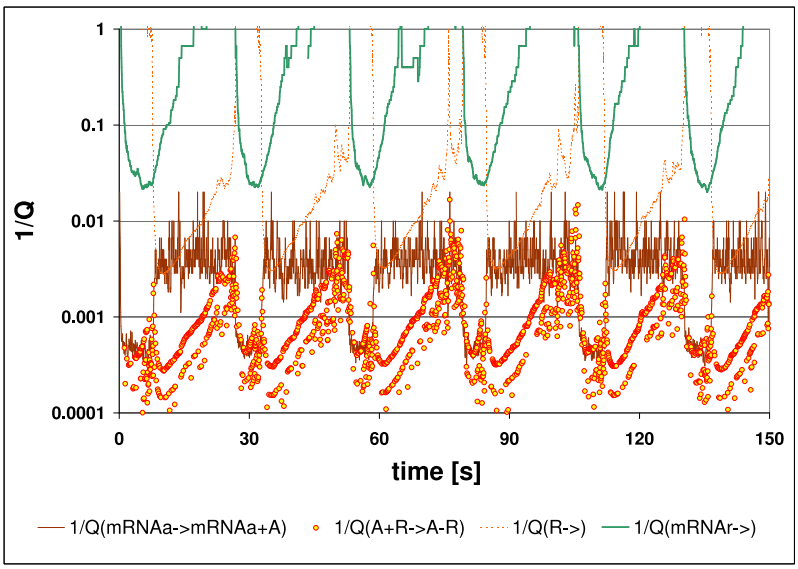

Figure 4.10: The Circadian Clock as a multiple time scale-model. The figure presents some channels with their corresponding Q_μ^{-1}-value, which corresponds to the expected length of a time step in the FRM.

Figure 4.11 includes three graphics representing three 150 second runs using the COAST-algorithm, Gillespie's FRM, and the deterministic solution, with the initial conditions

$$N_X(0) = \begin{cases} 1 & \text{if } X = Pa \text{ or } X = Pr\,, \\ 0 & \text{otherwise.} \end{cases} \tag{4.11}$$

In all three simulations, the Circadian Clock showed periodic oscillations. The periods and the amplitudes are given in Table 4.2. The COAST results coincide within 1.1% (amplitude) or 4.8 % (period) with the values of FRM. Since $\alpha = 0.05$ was chosen for the error control parameter, the obtained accuracy is in agreement with the estimated error.

The deterministic reaction kinetics deviates strongly from the results of FRM, which can be explained by the necessary modifications of the modeling mentioned above, i.e. deterministic models allow values between 0 and 1, while stochastic do not.

Table 4.2 shows the results for the initial values given by Figure 4.12.

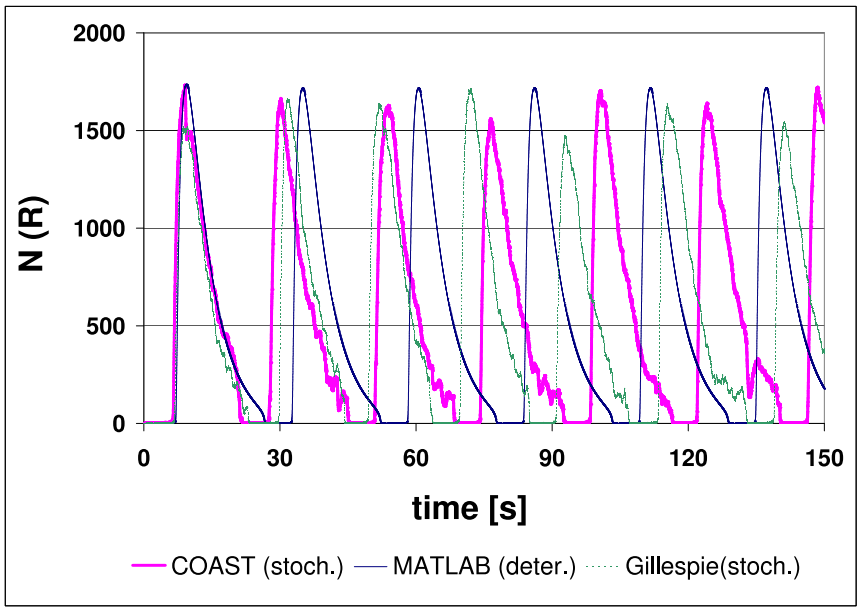

Figure 4.11: *Simulation of the Circadian Clock. 150 seconds using Gillespies* FRM, *COAST and a deterministic approach*

$$
\begin{array}{rcl}
Pa & \xrightarrow{50\,h^{-1}} & Pa + mRNAa \\
Pr & \xrightarrow{0.01\,h^{-1}} & Pr + mRNAr \\
mRNAa & \xrightarrow{50\,h^{-1}} & mRNAa + A \\
A + R & \xrightarrow{2\,g^{-1}} & A - R \\
Pa - A & \xrightarrow{50\,h^{-1}} & Pa + A \\
Pr - A & \xrightarrow{100\,h^{-1}} & Pr + A \\
R & \xrightarrow{0.2\,h^{-1}} & \emptyset \\
mRNAr & \xrightarrow{0.5\,h^{-1}} & \emptyset
\end{array}
\qquad
\begin{array}{rcl}
Pa - A & \xrightarrow{500\,h^{-1}} & Pa - A + mRNAa \\
Pr - A & \xrightarrow{50\,h^{-1}} & Pr - A + mRNAr \\
mRNAr & \xrightarrow{5\,h^{-1}} & mRNAr + R \\
Pa + A & \xrightarrow{1\,g^{-1}} & Pa - A \\
Pr + A & \xrightarrow{1\,g^{-1}} & Pr - A \\
A & \xrightarrow{1\,h^{-1}} & \emptyset \\
mRNAa & \xrightarrow{10\,h^{-1}} & \emptyset \\
A - R & \xrightarrow{1\,h^{-1}} & R
\end{array}
$$

Figure 4.12: *The reaction channels of the Circadian Clock (h=hour).*

	FRM	COAST	determ. solution
amplitude [N]	1599.8 ± 72.1	1617.4 ± 78.8	1717.2 ± 001
period [s]	23.0 ± 2.7	24.1 ± 1.8	25.± 0.002

Table 4.2: Amplitude and periodicity for the different simulations of the Circadian Clock.

4.1.4 Michaelis-Menten Kinetics

The Michaelis-Menten kinetics formulates an expression combining the velocity of catalysis with the concentrations of substrate and enzyme. It is the simplest model to describe the kinetic characteristics of many enzyme catalysed reactions. The model is named for Mr. Leonor Michaelis and Ms. Maud Leonora Menten who published their results in 1913 [Michaelis and Menten, 1913]. These kinetics are valid only when the concentration of the substrate is higher than the concentration of the enzyme, and in the particular case of a steady-state, where the concentration of the complex enzyme-substrate is constant. The described system is shown in Equation (4.12).

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P \tag{4.12}$$

E and S are the concentrations of the enzyme and the substrate, and ES and P the concentrations of the resulting complex and the product. By looking at the top of Figure 4.13 one can see the theoretical development of the concentrations in this system for initial enzyme and substrate concentrations, where the substrate is of higher concentration then the enzyme. It is observed in nature that k_{-2} is much smaller than k_2. Therefore the concentrations of all species are changing in a pre-steady state until they reach the equilibrium. There is no net change of product or substrate in the equilibrium. In this phase, the reaction from product to substrate can no longer be neglected.

The development of the pre-steady state can be described by the reaction equations in Table 4.3.

$$\begin{array}{rcl} \mathrm{E} + \mathrm{S} & \overset{c_1}{\rightarrow} & \mathrm{C} \\ \mathrm{ES} & \overset{c_2}{\rightarrow} & \mathrm{E} + \mathrm{S} \\ \mathrm{ES} & \overset{c_3}{\rightarrow} & \mathrm{E} + \mathrm{P} \end{array}$$

Table 4.3: Michelis-Menten: reaction equations

This system was simulated with COAST and has also been solved numerically using an implementation of the system in the form of differential equations using the mathematical software MATLAB (cf. Table 4.4). It can be demonstrated in the bottom left of Figure 4.13 that the implementation of a system of ordinary differential equations matches the stochastic approach with COAST. The stochastic representation of this biological process is much more realistic than the deterministic one, since the deterministic model allows continuous variables and the stochastic model does not. On the lower right figure one is able to observe an important difference between the deterministic model and the stochastic one. Although the mean values

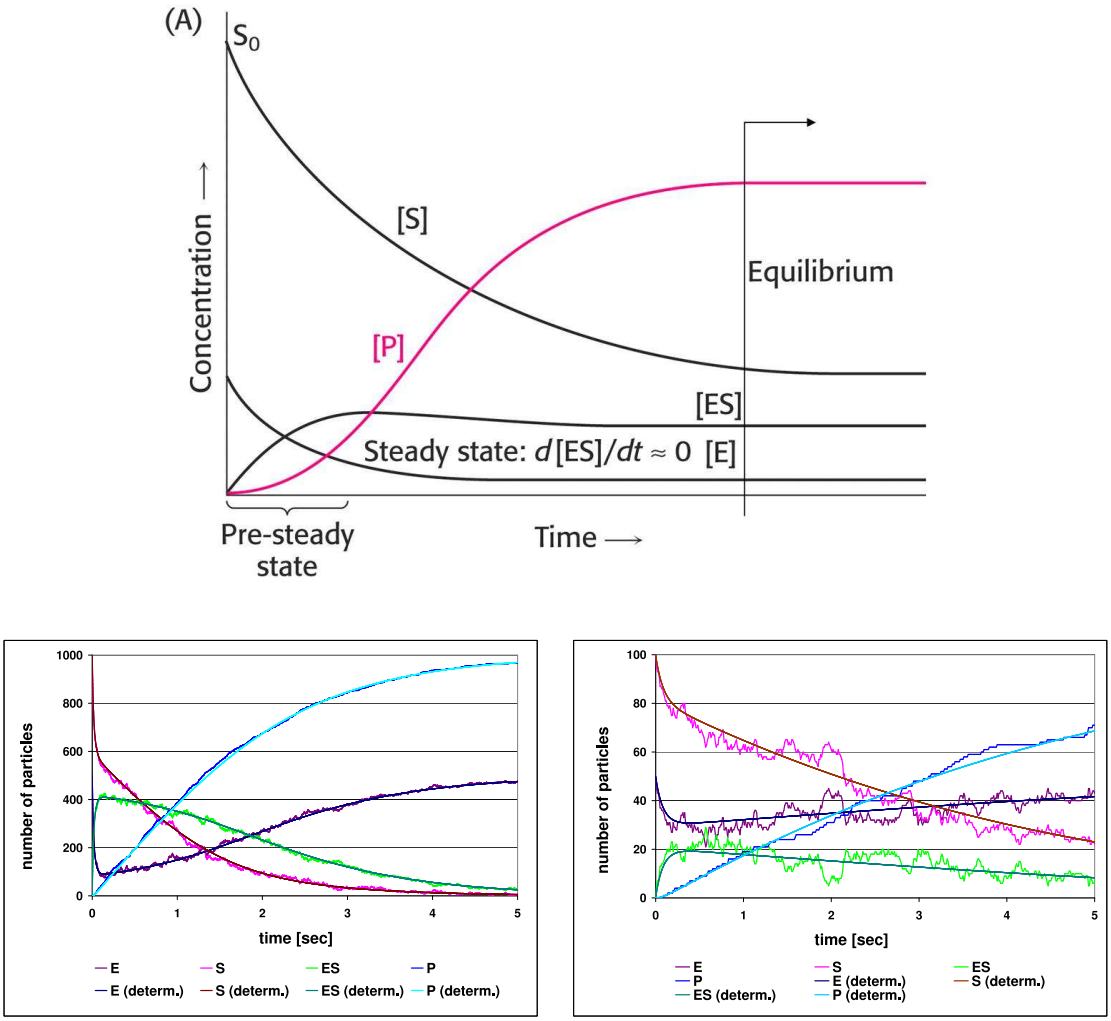

Figure 4.13: *Michaelis-Menten Kinetics: the top figure shows the development of concentrations in theory and the lower left one the results of the simulation with time =5sec; α=0.05; c_1=0.05, c_2=5.0, c_3=1.0; S_0=1000, E_0=500, $ES_0 = P_0 = 0$; the lower right figure illustrates the same simulation with only 1/10th of the initial particle numbers*

are the same, the curves representing the enzyme-concentration is always higher than the one for the enzyme-substrate complex. However, the stochastic model presents a different picture. Here, the enzyme substrate complex can exist in higher concentrations than the free enzyme. This is a good example to demonstrate the significant difference between stochastic and deterministic models. The rate parameters have not changed between the two experiments. Only the initial molecular concentrations were lowered.

$$\begin{array}{lcrcl} \frac{dE}{dt} & = & - & E \cdot S \cdot c_1 & + & ES \cdot (c_2 + c_3) \\ \frac{dS}{dt} & = & - & E \cdot S \cdot c_1 & + & ES \cdot c_2 \\ \frac{dES}{dt} & = & & E \cdot S \cdot c_1 & - & ES \cdot (c_2 + c_3) \\ \frac{dP}{dt} & = & & ES \cdot c_3 & & \end{array}$$

Table 4.4: Michelis-Menten: differential equations

4.2 Test Simulations Regarding COAST's Extension to Diffusion

4.2.1 Basic Systems

In this section, the reliability and accuracy of COAST as applied to diffusion is evaluated by test simulations. To this aim, COAST-simulations of the one-dimensional motion of a single substrate were compared with the predictions of the diffusion equations and with the results of random-walk simulations (cf. Equation (3.5)):

$$x(t+1) = x(t) + \frac{\Delta t\, f(x(t))}{\gamma} + \sqrt{2\, D\, \Delta t}\, W\,, \tag{4.13}$$

with a normally distributed random variable W. All simulations in this subsection were performed with a diffusion coefficient of $D = 10^{-13}\ \frac{m^2}{s}$, $T = 298$ K, and $R = 100$ nm.

Diffusion without external force Let us consider 0.75s-simulations of the Smoluchowski-equation (cf. Equation (3.31)) with $f = 0$ in the interval [-2000nm,+ 2000nm]. In doing so, two initial conditions were considered: First,

$$\varrho_1(x, 0) = N \cdot \delta(x)\,, \tag{4.14}$$

(N=total number of particles) i.e. at time $t = 0$ all particles have position $x = 0$. In this case, the Smoluchowski-equation has the solution

$$\varrho_2(x, t) := \frac{1}{\sqrt{4\,\pi\, D\, t}}\, e^{-\frac{x^2}{4\,D\,t}} \tag{4.15}$$

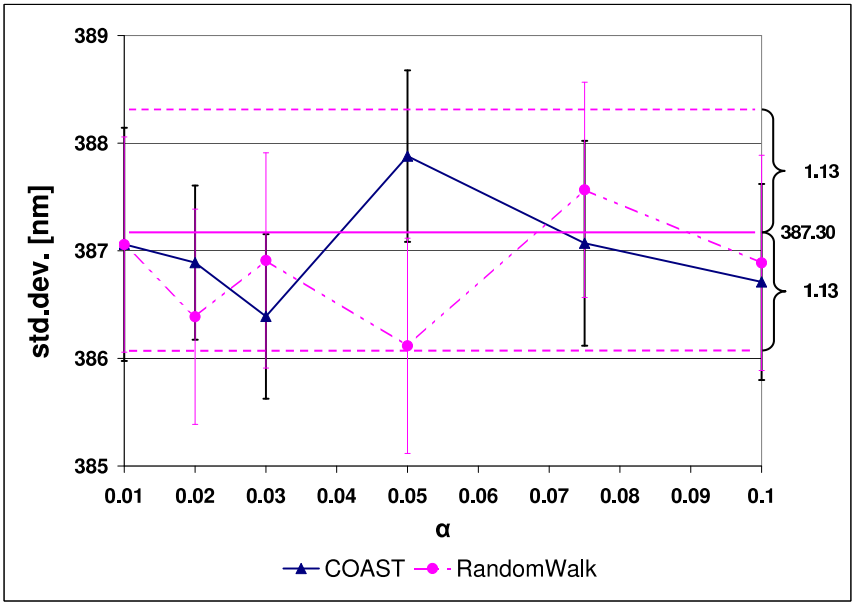

Figure 4.14: *The error bars of the standard-deviation for free diffusion with initial condition ϱ_1 -delta-distribution- (cf. Equation (4.14)). The diagram shows the outcome of* COAST*-simulations dependent on α. For comparison sake, the results of the random walk-simulations are also included, where, for each α, the time steps are identical with the time steps of the* COAST*-simulation.*
*t=0.75s; D=*1 $\cdot 10^{-13} \frac{m^2}{s}$; k=0 $\frac{kg}{s^2}$; R=100nm; N=100000; 25 repetitions

Due to its diminishing standard deviation, a δ-distribution can lead to additional numerical errors (cf. Section 3.6).

For all α-values, the value of the mean position of the particles was consistent with the exact value of 0. For example, for $\alpha = 0.05$ the averaged mean value from 25 runs was -1.53 nm. Accordingly, the focus will be on the second quantity necessary to characterize Gaussian-distributions, this is the standard deviation σ. To characterize the dependency of σ on α, simulations for both initial values with $N = 10^5$ were performed, where each simulation was repeated 25 times. Additionally, random walk-simulations of the same system with δt adjusted to the corresponding α-value by Equation (3.57) were also performed.

To illustrate the statistical effects, the standard deviation from the position of 100,000 particles randomly distributed according to ϱ_2 (cf. Equation (4.15)) were also computed. This experiment was repeated 10 times. The corresponding error bar of the standard deviation is also shown in Figure 4.14.

As one can easily see from Figure 4.14, the outcomes of COAST-simulations always

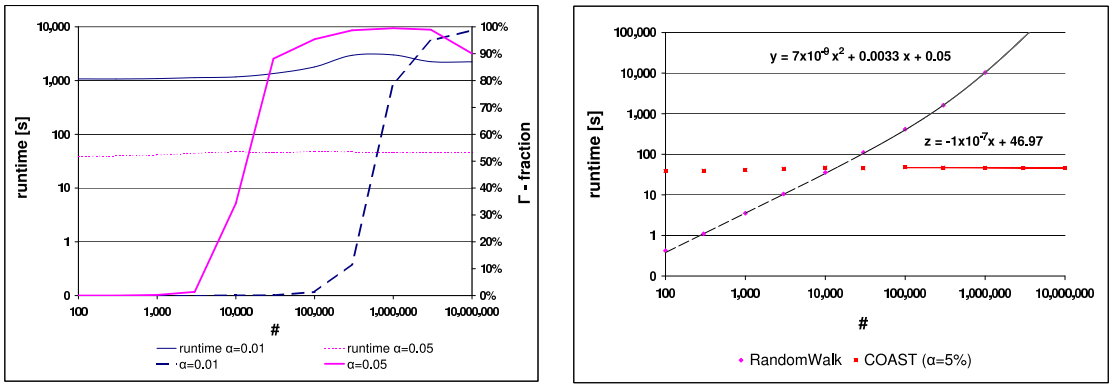

Figure 4.15: *Characterization of the run time behavior of* COAST *dependent on the number of particles N. The left figure shows for $\alpha = 0.01$ and $\alpha = 0.05$ the dependence of the run time from the modeling level used:* Γ*-fraction is the portion of evaluations done by Gaussian-distributions. In the right figure, the run time behavior of* COAST *($\alpha = 0.05$) is compared with the run time behavior of random walk-simulations with identical time steps. y and z are least mean square fits to the run times of random walk-simulations (y) or* COAST*-simulations (z) respectively.*

showed a similar accuracy as the results of the random walk-simulations. Furthermore, the mean values of both simulations were always within the error bar of the value computed from the placement of the particles according to the exact distribution (cf. Figure 4.14). Hence, in this case, COAST led to quite accurate simulation results.

As a next step, let us characterize the run time behavior of COAST dependent on the number of particles N. To this aim, COAST-simulations for $\alpha = 0.01$ and $\alpha = 0.05$ with initial condition ϱ_1 were considered and compared with random walk-simulations of the same system. The results are shown in Figure 4.15, which includes two diagrams: The left diagram shows the run times of the COAST-simulations together with the fractions of transition numbers computed by the modeling level Γ, which means the description by Gaussian-distributions (cf. Equation (3.59)). whereas the right diagram contains a comparison between the run times of the COAST-simulations for $\alpha = 0.05$ and the run times of the random walk-simulations.

It can be seen from the left diagram in Figure 4.15 that, for both $\alpha = 0.01$, the run time of the COAST-simulation is maximal if about $2/3$ of the transition numbers are computed by modeling level Γ. For larger portions of $\Gamma\,\&\,\Delta$, the run time becomes smaller and converges to a constant value. Similar observations can be found for other α-values.

To describe this asymptotic run time behavior quantitatively, a least mean square fit to the run times of the COAST-simulations with $\alpha = 0.05$ for $N > 10^5$ were performed, which resulted in the function $z(N) = (-1 \cdot 10^{-7}\, N + 46.97)\, s$. For comparison: A least mean square fit to the run time of the random walk-simulations led to the fit curve $y(N) = (7 \cdot 10^{-9}\, N^2 + 0.0033\, N + 0.05)\, s$, which is also shown in the right diagram in Figure 4.15. Accordingly, for $N = 10^6$, the run time of the random walk-simulations is about three hours, which is an enormous difference to the 18 seconds of COAST.

Diffusion with a linear external force Additionally, three-seconds-simulations of the Smoluchowski-equation with external force

$$f(x) := -k\, x \qquad \text{with} \qquad k = 10^{-7} \tfrac{kg}{s^2} \tag{4.16}$$

in the simulation interval to $[-1000\,\text{nm}, +1000\,\text{nm}]$ were investigated. As initial values, the uniform distribution

$$\varrho_2(x, 0) := \frac{N}{2000\,\text{nm}} \tag{4.17}$$

was chosen.

This results in the implementation for $N = 100000$ and $\alpha = 0.01$ in 197 volume elements with a width of $\Delta x = 1 \cdot 10^{-8} m$ and 507 particles in each volume element and 628 in the center. The solution of the Smoluchowski-equation with linear external force $f = -k\, x$ is given by [Schulten and Kosztin, 1999]:

$$\varrho_2(x, t) = \int_{-1000nm}^{+1000nm} p(x|y; t)\, \varrho_2(y, 0)\, dy\,, \tag{4.18}$$

where

$$p(x|y; t) \;:=\; \sqrt{\tfrac{k}{2\,\pi\, k_B\, T\, (1-s^2(t))}}\, \exp\left(-\tfrac{k\,(x - y\, s(t))^2}{2\, k_B\, T\, (1-s^2(t))}\right) \quad \text{with} \quad s(t) := e^{-\frac{k\,t}{\gamma}} \tag{4.19}$$

It follows that after three seconds the system has reached its state of equilibrium, which is given by a Gaussian-distribution with a mean value of 0 and a standard-deviation

$$\sqrt{\frac{k_B\, T}{k}} = 202.8\text{nm}\,. \tag{4.20}$$

To assess the accuracy of COAST-simulations, the standard deviation of the distributions at the end of the simulations were compared with the standard deviation of the positions of one hundred thousand particles randomly located over the thermodynamic equilibrium distribution given by

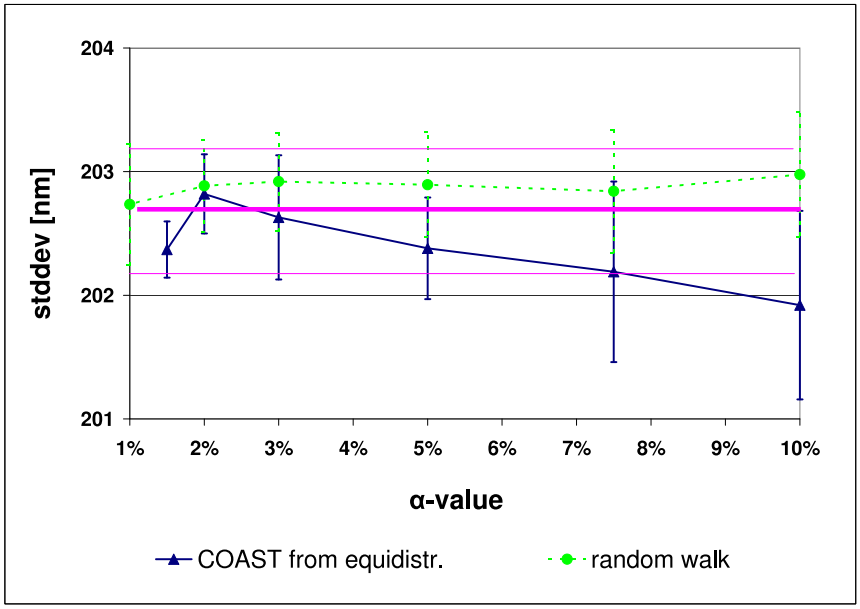

Figure 4.16: *The error bar of the standard deviations for the 25* COAST*-simulations and 25 random walk-simulations of the diffusion system with linear force (cf. Equation (4.16)) dependent on α, where, for all α, the random walk is based on the same time steps as* COAST*. The bold and the dotted lines correspond to the error bar of the standard deviation calculated from the position of 100000 particles randomly located over the thermodynamic equilibrium distribution (cf. Equation (4.21)).*
t=5s; D=$1 \cdot 10^{-13} \frac{m^2}{s}$; k=$1 \cdot 10^{-7} \frac{kg}{s^2}$; R=100nm; N=100000; 25 repetitions

$$\omega(x) := \sqrt{\frac{k}{2\,k_B\,T\,\pi}}\, e^{\frac{-k\,x^2}{2\,k_B\,T}}\,, \tag{4.21}$$

which coincides with Equation (4.19) in the limit $t \to \infty$.

As can be seen from Figure 4.16, the mean values from COAST-simulations lie for $\alpha \leq 0.075$ always within the error bar of the value derived from the thermodynamic equilibrium distribution. Furthermore, for all $\alpha \leq 0.1$, the COAST-results deviate less than 1% from the averaged values of the thermodynamic equilibrium, so that the COAST-results are found to coincide with the exact results.

4.2.2 Kramer's Transition State Theory

One of the most prominent applications of diffusion models is the prediction of decay rates for chemical bindings by Kramers' transition state theory [Kramers, 1940,

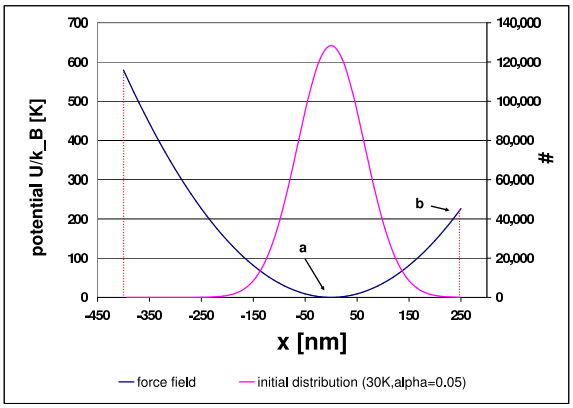

Figure 4.17: *The potential used for the simulation of Kramers theory and the corresponding thermodynamic equilibrium distribution. a and b are the minimum and the local maximum of the potential, where b reflects the transition between binded and dissociated state.*

Hänggi et al., 1990]. The basic idea of this theory is to describe the state of a molecule by a single (reaction) coordinate x, where the time-evolution of this coordinate is described by a Langevin-equation in the strong friction limit (i.e. the frictional force is much larger than the force of inertia). $x \in]-\infty, b[$ corresponds to an existing bond, where $x > b$ reflects a dissociated molecule. Thus, the molecule is protected against dissociation by a potential U, which has its local maximum at b (cf. Figure 4.17).

Furthermore, $a = 0$ is the minimum U for $x \in]-\infty, b[$. Thus, Kramer's theory predicts that, in thermodynamics equilibrium, the decay rate of this model is given by

$$r = c \cdot e^{-\frac{U^{\ddagger}}{k_B T}}, \tag{4.22}$$

where $U^{\ddagger} := U(b) - U(a)$ is the height of the potential barrier and where c depends only on the friction coefficient and on the curvatures of the potential, but not on temperature T.

In this section, simulations of the transitions of particles over such a potential barrier for different temperatures are presented, where COAST is used with $\alpha = 0.05$. The aim of these simulations is to test if COAST is able to reproduce the results of Kramers theory.

Each simulation was started with $N := 10^6$ particles, which were distributed according to the the thermal equilibrium of this system. More specifically, the probability distribution of particles is given by

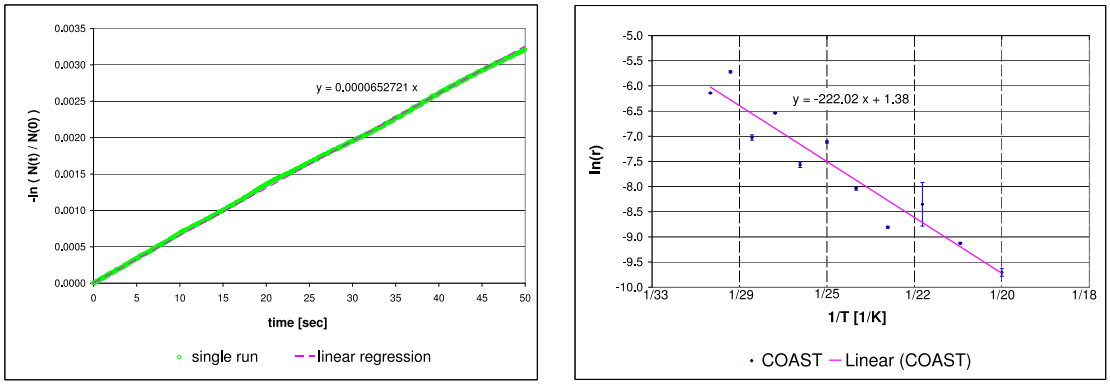

Figure 4.18: *Kramers-Theory: The left picture shows the number of transitions for a* COAST*-simulation ($\alpha := 0.05$) at 20 K. The figure to the right displays the result of 10 independent runs in the temperature span of 20K to 30K.*

$$\omega(x,k) := \sqrt{\frac{k}{2\,k_B\,T\,\pi}}\, e^{\frac{-k\cdot x^2}{2\,k_B\,T}}\,\Delta x\,, \tag{4.23}$$

a friction coefficient [1] $\gamma = 3.45165 \cdot 10^{-9}\frac{Ns}{m}$ and a potential

$$U := \frac{k\cdot x^2}{2} \qquad (k := 10^{-7}\tfrac{kg}{s^2})\,. \tag{4.24}$$

Furthermore, it is assumed that the dissociation occurs at $b = 250$ nm, so that the potential height $U^{\ddagger}$ was given by

$$\frac{U^{\ddagger}}{k_B} = \frac{U(250\,nm) - U(0\,nm)}{k_B} = 226.34\,K \tag{4.25}$$

Let $N(t)$ be the expected number of particles after time t, where $N_0 = N(0)$ is the initial value. If there exists any time-constant transition rate r, then one obtains:

$$N(t) = N_0\,e^{-r\,t} \quad \Leftrightarrow \quad r\,t = -\ln(N(t)/N_0)\,. \tag{4.26}$$

Accordingly, the quantity $-\ln(N(t)/N_0)$ was measured in each simulation, so that the transition rate r could be determined as the gradient of this straight line. In Figure 4.18, this is shown for a COAST-simulation ($\alpha = 0.05$) at 20 K. This procedure was performed for T=20, 21, 22, ..., 30 K, where for every T the simulation was repeated ten times. The simulations have been best according to the theory for low temperatures. Therefore simulations just above 0K would have been ideal, but the

[1] The friction coefficient γ and the diffusion coefficient D are connected by $\gamma = \frac{k_B\,T}{D}$ this results in $D = 1\ \ 10^{-13}$ for T=25K

computational effort increased dramatically so the area around 20K was chosen as a compromise.

Figure 4.18 illustrates how $\ln(r(T))$ responds to changes in T^{-1}. Particularly, a least mean square fit was performed to the curve, which resulted in

$$\ln(r) = -222.02\,K\,\tfrac{1}{T} + 1.38\,. \tag{4.27}$$

For comparison, the prediction of Kramers theory (cf. Equation (4.22)) is

$$\ln(r) = -\tfrac{U^{\ddagger}}{k_B}\,\tfrac{1}{T} + \ln(c) = -226.34\,K\,\tfrac{1}{T} + \ln(c)\,. \tag{4.28}$$

Hence, the COAST-simulation ($\alpha = 0.05$) was able to reproduce the prediction of Kramers theory within 1.9 %.

4.2.3 Linear Diffusion

This section applies COAST to a common biophysical problem and describes how a concentration gradient reaches the equilibrium by linear diffusion. This process can be described by Fick's Second Law of diffusion:

$$\left(\frac{\partial c}{\partial t}\right)_x = D\left(\frac{\partial^2 c}{\partial x^2}\right) \tag{4.29}$$

D is assumed to be independent from the concentration of the substrate c and therefore from the location of the particles. We further assume the initial condition to be $c = c_0$ in the interval $]-\infty,0]$ and $c = 0$ in $]0,\infty]$. Thus, the solution for Equation (4.29) is

$$c(x_0,t) = \frac{c_0}{2}\,[1-\phi(u)] \quad with \;\; u = \frac{x}{2\sqrt{Dt}} \tag{4.30}$$

$\phi(u)$ is the so called error function

$$\phi(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{\frac{-x^2}{4Dt}} \tag{4.31}$$

with this the quotient c/c_0 can be formulated as following

$$\frac{c(x,t)}{c_0} = \frac{1}{2}\,[1-\phi(u)] \tag{4.32}$$

In Figure 4.19 one can observe how well the theoretical curve of the gradient after the timespan of t=16d matches with the simulations done with COAST. The

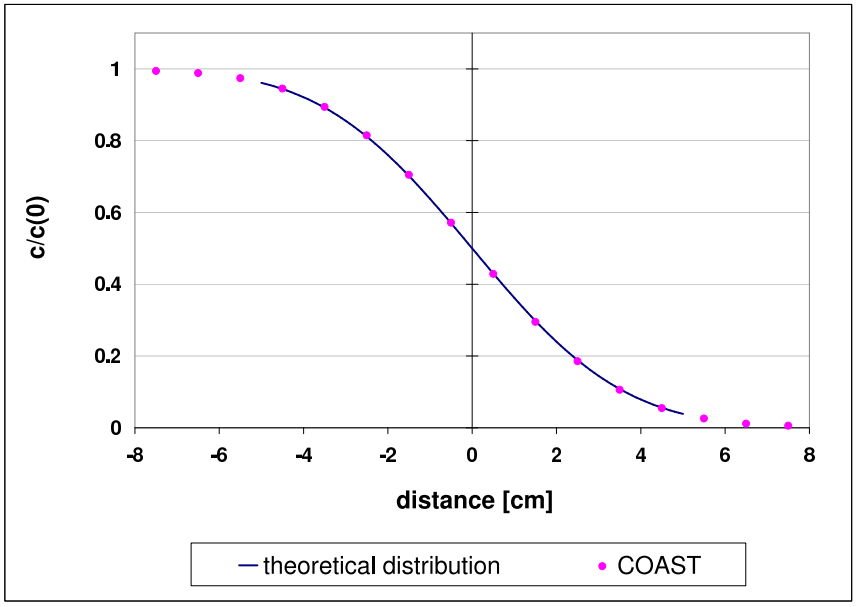

Figure 4.19: *Linear Diffusion:* D=$2.9 \cdot 10^{-6} \frac{cm^2}{s}$, t=$16d$=$1382400s$, δx=$1cm$, α=0.05

theoretical curve is limited to the left and right side by the limited amount of values tabled for the error function. This problem did not occur as a positive side effect for the application of COAST.

4.3 General Technical Considerations

4.3.1 Run time analyses

In this chapter several time depending simulations have been discussed. To obtain the most accurate results all simulations should have been run on the same system under exactly the same conditions without the influence of any counter processes.

This is not the case for the simulations performed here. Usually the tasks were transfered to a cluster of computers and processors and it has been up to this grid to choose an appropriate machine. However, due to the fact that most simulation runs took several hours and by that taking much more time than the usual counter processes occurring. I am positively convinced that on an average the measured run times are comparable. Very short simulations have been performed on an IBM notebook with an Intel Pentium III central processing unit running on 700Mhz and 768Mb of memory using the Windows2000 operating system.

4.3.2 Used Software

COAST was implemented in JAVA 1.4 using the integrated development environment *eclipse* in the version 3.0, which is freely available via internet. To make the process of implementation easier I chose the programming language offered by *Matlab* in the Version 6.5 to apply *COAST* to diffusion processes.

5 Discussion and Conclusion

After outlining the task of efficiently modeling cellular processes in the introduction, I gave an overview on existing strategies for simulating reaction and diffusion processes in the second chapter. In chapter three I introduced with the *Controllable Approximative Stochastic Reaction Algorithm* a hybrid algorithm for simulating reaction and diffusion. After setting one error parameter α, COAST adjusts itself according to the development of the system. Its three modeling levels are used to be as accurate as necessary and as fast as possible. In the last chapter I have presented the application of COAST to a variety of problems related to reaction and diffusion.

COAST was able to show its reliability and accuracy for reaction and diffusion processes for different settings of α. In this last chapter I will sum up my findings and discuss COAST for the background of existing tools available and the possibilities offered by science.

5.1 Reflecting on COAST

Good algorithms have four common features: they are fast, accurate, simple to implement, and they can be applied without too much knowledge of the details of the basic methods and concepts. In this section, I will discuss how the COntrollable Approximative STochastic reaction algorithm (COAST) fulfills these criteria.

The runtime behavior of COAST in the simulations was composed of two different parameter ranges; one range, in which the *First Reaction Method* (FRM)-like modeling level Σ dominates, and one in which mainly the regimes by Gaussian-distributions (Γ) or by deterministic reaction kinetics (Δ) were used.

If Σ was predominantly used, the runtime of COAST was nearly identical to the runtime of the FRM, on the contrary I found for higher particle numbers qualitative differences between the runtime behaviors of COAST and the FRM. If the total number of particles N in the system was large enough so that Γ and Δ were predominantly used, then the runtime of COAST increased with N^a with $0 \leq a \leq 1$. The runtime of FRM increased with N^b, where $1 \leq b \leq 2$ (cf. Section 4.1.1). Since the *τ-leap method* showed a similar behavior to the FRM, we can conclude that COAST is fast in comparison to the FRM and the *τ-leap method*. This can be easily explained by the length of a single time step and the amount of random numbers generated; for the FRM, the mean length of a time step is proportional to N_A^{-1} for first order reactions and proportional to N_A^{-2} (cf. Table 2.3) for second order reactions [Gillespie, 1977], so that the quantity of random numbers is proportional to N_A and N_A^2,

respectively. For the *τ-leap method* and COAST, the length of the time steps, or the τ-leaps, are proportional to N_A^{-1} for second order reactions, but independent from N_A for first order reactions (cf. Equation (3.22)).

However, for each time step or τ-leap, the evaluation of a Poisson-distribution in the *τ-leap method* requires[1] a quantity of random numbers proportional to the expected number of reactions, or equivalently to the number of particles. Whereas the evaluation of a Gaussian-distribution can always be performed by generating a single random number [Box and Muller, 1958].

In contrast, an optimization of the *τ-leap method* would require a method for evaluating the Poisson-distribution with computational costs independent from the particle number. The only method to my knowledge is the approach of Ahrens and Dieter [1982], who approximated Poisson-distributions for large particle numbers by using a Gaussian-distribution. However, this is equivalent to replacing the Poisson-distribution in the *τ-leap method* by the the modeling level Γ of COAST, with the exception that COAST is based on probabilities that are more realistic for the long time steps used in both algorithms.

As mentioned above, the runtime for both modeling levels Γ and Δ always increases with the same exponent of n. In Section 4.1.1 I was able to show for a single reaction channel, that Δ reduces the runtime of simulations by about 10% as compared with Γ. Neglecting fluctuations by using deterministic reaction kinetics Δ leads to additional inaccuracies. Thus, it is quite difficult to globally answer the question if one should use Δ, or if one should reduce COAST to the two other modeling levels Σ and Γ. Instead, it is recommended to introduce an option in the implementation of COAST , so that the user can adjust this according to the needs of the given system.

To check the accuracy of COAST, I considered systems with relatively small numbers of particles: the initial values were 20,000 for A+B $\rightleftharpoons$ C and A+B $\rightleftharpoons$ 2 C, 35,000 for the Oregonator (cf. Section 4.1.2), and 2 for the Circadian clock (cf. Section 4.1.3). Note that the inaccuracies of COAST decrease for larger particle numbers, because the approximation of binomial-distributions by Gaussian-distributions improves. That is why the usage of relatively small systems (cf. Section 4.1.1) is the best test for the reliability of COAST. In all these systems, the values of the COAST-simulations with error parameter $\alpha \leq 0.05$ coincided with the corresponding values of the FRM-simulations within 1%, except for the period of the Circadian Clock.

As a result, I conclude that COAST is a fast and accurate algorithm, not only for elementary systems with smooth dynamics, but also for complex systems such as the Oregonator and the Circadian Clock. Here, a "complex system" refers to multiple time-scale systems with rapidly and strongly fluctuating particle numbers. Another definition of a "complex system" refers to the fact that some of the substrates contribute to a large number of different reaction channels. It is worth noting that in

[1] The implementation I used in my simulations has been Cellware [Dhar et al., 2005], it uses the rejection method.

this case, the critical number of reactions l_μ (cf. Equation (3.24)), and thus the length of the time steps is reduced, so that the accuracy remains constant. Accordingly, one can observe that the results for the Oregonator and the Circadian Clock, where a part of the substrates contributes to several different reaction channels, are no less accurate than for the simple models. Furthermore, for every fixed set of reaction channels, the exponential dependency of the runtime from the particle number is not changed by the reduction of the time steps. Thus, COAST also works accurately for a second kind of complex systems, namely systems in which substrates contribute to large number of reaction channels.

Furthermore, COAST is quite simple to implement and its usage does not require a deep insight into its foundations. It is recommended users perform the first COAST-simulation of their system with error control parameter $\alpha = 0.05$ and usage of modeling level Δ. This has led to quite accurate results without too long run times for all simulations performed so far.

It should be mentioned, that it is suboptimal to run COAST with α=0, since then the algorithm requires the same amount of random numbers like the FRM, but has a larger computational overhead than the FRM, which should be used then instead. For systems composed of many reaction channels, it would be of course helpful to reduce the number of computations necessary for the determination of the length of the time-steps and the succession of the evaluations. As a summary of these considerations, COAST can be considered as a good reaction algorithm in the sense described at the beginning of this chapter.

5.2 The Adoption of COAST to Diffusion

The aim of Chapter 3.4 was the modification of COAST towards an efficient algorithm for the simulation of thermal motions of particles. The starting point of COAST is the Smoluchowski-equation [Smoluchowski, 1917], which is a diffusion-model based on two essential approximations. The first approximation is the strong friction limit, which is a good approximation if the moment of inertia of the particles is small compared to the forces acting on the particles. The other approximation is that the interactions between the described substrates are considered as much smaller than the interactions between the described substrates and their environment (cell compartments, water,...), which results in a linear diffusion model. Obviously, this second approximation can always be applied if the concentration of the described substrates is low enough.

A related problem to this is, how good is our knowledge about the cell structure. This is not a limitation of the algorithm but a problem of modeling itself. The common picture of the cell as a wet space with some organelles and some floating enzymes is far from reality. Luby-Phelps et al. [1986] was able to show in experiments using "fluorescence recovery after photobleaching" (FRAP), that the structure of the

cytoplasm has a deep impact on diffusion. It is a well known fact that diffusion coefficients are usually measured for enzymes in vitro and so there is a big difference between these results and the real values in vivo and therefore the usability of the listed in vitro values for simulations is very limited. Luby-Phelps also clarified that, on average, the viscosity of the cytoplasm is four times as high as the one of water. Even more she found that the diffusibility of macromolecules is limited by their size. This is due to structural barriers within the cytoplasm. There are three types of filaments, which are made responsible for this: F-actin, microtubules and intermediate filaments, and an assortment of accessory proteins that cross-link these filaments. They leave a pore size of about 300 to 400Å. Knowing the cellular structure it is possible to formulate a forcefield for the diffusion of the particles so COAST would be able to handle this problem, since this is a question of input. However, so far our knowledge in this area is very limited.

The situation becomes even more complicated if one considers that not only the size of the diffusing protein is responsible for the cytoplasmic diffusion coefficient, but the structure of that protein also has an effect [Luby-Phelps, 2000]. Hydrophobic domains and ionizable surface groups influence the mobility of proteins significantly. This implies that the diffusion coefficient is not a constant, but rather depends on the surroundings. Consequently one would have to model the whole cellular structure as well. Again, with detailed information, I am convinced this can be done, but the problem is the generation of reliable data. Luby-Phelbs points out that some investigators of cellular diffusion have come to the conclusion that most enzymes are immobilized by cellular structure, i.e., they are attached to membranes or cytoskeletal surfaces. In this case one does not have to model diffusion for those enzymes anymore, but now the whereabouts of these enzymes become important. One needs the exact localization within the cell. Even if the assumption of spatially fixed enzymes cannot be verified, this demonstrates one of the basic dilemma modeling has to get along with and this is "few information".

Accordingly, the Smoluchowski-equation cannot possibly reflect all aspects of thermal motions in cells [Agutter et al., 1995], but is suitable in quite general scenarios. On the other hand, it is quite likely that for many systems, the Smoluchowski-equation is the best diffusion model for which simulations can be performed. It should be noted that if one does not use the "strong friction limit", then one has to consider double the number of dimensions, (these are positions and velocities), so that, in a first approximation, the computational costs are no longer proportional to n^d (strong friction limit) but proportional to n^{2d}, where d is the number of dimensions and n is the number of lattice points (or voxels) in each direction.

Note that nonlinear diffusion corresponds to second and higher order reactions. Since COAST allows one to treat all types of reactions, a derivation of an algorithm for nonlinear diffusion model from COAST can easily be done, where one can use analogous estimations as presented here for the application of COAST to diffusion.

However, in this case, one cannot use constant time steps; one has to compute the appropriate length of such time steps for every simulation step (analogously to the reaction version of COAST). As a result, the runtime would be nearly doubled and an application of parallel computing to such an algorithm for nonlinear diffusion is much more difficult and less efficient than for COAST. Hence, the application of nonlinear diffusion models will not likely be possible in the generic case.

For the current simulation, a discrete version of the Smoluchowski-equation was used, so that the thermal motions of particles were identified with jumps between adjacent lattice points (or analogously volume elements). This discrete diffusion model obviously looks quite different than the Langevin-equation, from which the diffusion model was originally derived. Thus, to achieve that the discrete diffusion model can be used as an adequate description of the thermal motion of particles, the conditions were imposed that the first (expectation) and the second moment (variance) as well as the flux of particles of the discrete diffusion model is, up to $\alpha \cdot 100\,\%$ ($\alpha \in [0,1]$), in agreement with the values of of the Langevin-equation. These requirements led to a large part of the conditions for the discretization parameters. The other criteria were derived from the constraint that the transition probabilities must have positive values and that in every time step the changes of particle numbers at each lattice point must be small- at least on average.

A discrete diffusion model allows a very natural interpretation of thermal motion in terms of chemical reactions, by identifying transitions between adjacent lattice points as reaction channels. Due to the linearity of the diffusion model, all transitions correspond to first-order reactions.

The correspondence between discrete diffusion model and chemical reaction systems makes it very natural to adapt an algorithm for the simulation of chemical reaction systems to the needs of a discrete diffusion model. In this thesis, this adaption process was performed for the COntrollable Approximative STochastic reaction-algorithm (COAST).

One important feature of COAST is its usage of three different levels of modeling: for small particle numbers an exact stochastic model, for intermediate particle numbers an approximative stochastic model by Gaussian distributions, and for large particle numbers deterministic reaction kinetics. Thereby, the criteria for the application of the different modeling levels are, as all errors due to used approximation, formulated dependently from one single error control parameter $\alpha \in [0,1]$, which helps one to easily find an optimal balance between accuracy and computational efficiency for each individual system.

This formulation dependently from a single control parameter is, in general, also used in the adaption of COAST to linear diffusion – with the exception that one criterion includes a second parameter R, which reflects the spatial resolution of the model. Howerver, the value of this second parameter is also very easy to choose, so that the simple controllability of COAST is also given here.

To test COAST, simulations of one-dimensional diffusion without external and with

linear external force were performed. In both cases, the deviations of the COAST-results from the exact values were always in the range of purely statistical errors, suggesting that COAST works pretty accurate for these models.

On the other hand, COAST showed a much more advantageous runtime behavior for intermediate and large particle numbers as compared to random walk-simulations: The runtime of COAST was asymptotically independent from the particle number, whereas the runtime of random walk-simulations is asymptotically proportional to the number of particles. Only for very small particle numbers, random walk-simulations are faster than COAST. However, in this range, COAST is also quite fast, so that this disadvantage of COAST is not really problematic.

To further support the credibility of COAST, the escape rates of particles from a local potential minimum (metastable state) over a potential barrier were computed from COAST-simulations and compared with the predictions from the well-known Kramers-theory [Kramers, 1940]. The essential statement of the Kramers-theory is that the logarithm of the escape rate is, up to a constant term, given by $\frac{U^\ddagger}{k_B T}$, where $U^\ddagger$ is the height of the potential barrier. From the COAST-simulations ($\alpha = 0.05$), a value for $U^\ddagger$ was obtained, which coincided with the correct value up to 1.9 %. Thus, the result was much better than expected by the choice of the error control parameter α.

Throughout this thesis, all considerations were restricted to a one-dimensional model. This restriction was used to simplify the notation. An application of the presented methods to higher dimensions is analogously possible. However, one has to take into consideration that the computational costs (of diffusion and of reaction-diffusion models) are, as mentioned above, are proportional to n^d. Although the power of modern computers is rapidly increasing, a full 3D-description of such complex systems as biological cells is not possible in the generic case.

5.3 Combining Reaction and Diffusion Algorithms

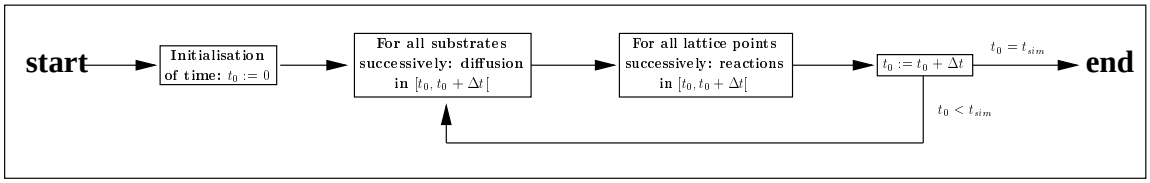

Figure 5.1: *Scheme of the module for reaction-diffusion systems. The keywords 'reaction' and 'diffusion' mean the application of an arbitrary algorithm for reaction- or diffusion models.* t_{sim} *is the duration of the whole simulation.*

As mentioned in the introduction, it can be very helpful to consider reaction and diffusion in the same time interval successively, because this allows the subdivision

of the system in independent subunits: For diffusion, the different substrates can be treated independently, whereas for reactions the different lattice points are independent. Hence, one can use the algorithm illustrated in Figure 5.1.

To avoid errors caused by the successive treatment of reactions and diffusion, one has to use suitable conditioned probabilities. Analogously, to the treatment of the different directions of thermal motions in COAST (cf. Figure 3.8), one has to use the reaction probabilities under the condition that the particles do not jump from one lattice point to an adjacent point in the same time interval. This correction can, in a first approximation, be performed by replacing the reaction constants c [Gillespie, 1977] by

$$\tilde{c} := \begin{cases} \frac{c}{1-Q_A(i)}\,, & \text{for } A \to P, \\ \frac{c}{(1-Q_A(i))\,(1-Q_B(i))} & \text{for } A + B \to P, \\ \frac{c}{(1-Q_A(i)}\big)^2\,, & \text{for } 2\,A \to P\,, \end{cases}$$

where $Q_S(i) := q_S(i+1|i) + q_S(i-1|i)$ is the total transition probability of the substrate S at a lattice-point i and P an arbitrary product.

A crucial point to note here, is the choice of an appropriate value Δt for the length of a time step in which diffusion and reactions can be treated successively. A possible choice would be to compute the times (cf. Equation (3.57)):

$$\Delta t_{diff} := \min\left\{ \tfrac{\tilde{\tau}_A}{2} \mid A \in \mathbb{S} \right\} = \min\left\{ \tfrac{\alpha\, k_B\, T\, \gamma_A}{f^2_{max}(A)}, \tfrac{\alpha\,(\Delta x)^2}{4\,D_A\,(1+\alpha)} \mid A \in \mathbb{S} \right\} \tag{5.1}$$

and Δt_{reac} in which all transition or reaction probabilities respectively are smaller than α and to define

$$\Delta t := \min\{\, \Delta t_{diff}\,,\, \Delta t_{reac}\,\}\,. \tag{5.2}$$

In this case, one can easily show that all errors due to the successive treatment of thermal motions and reactions are of the order $\mathcal{O}(\alpha^2)$, so that the algorithm will work quite accurately. But, on the other hand, if at any point the reaction rate of a single reaction channel or the diffusion rate of a single substrate is very large, one would use this small Δt for all transitions and reactions so that the algorithm will work very inefficiently. As a consequence, the choice of Δt presented here is only suitable for reaction-diffusion systems with similar reaction and diffusion rates.

5.4 Final Conclusions

In the last section, I presented a method to combine COAST and its application to linear diffusion to an algorithm for the stochastic simulation of reaction-diffusion models. Thereby, the crucial step is an appropriate choice of the time step Δt in which thermal motions and reactions can be computed successively.

In this thesis, I defined Δt as the time span, in which all changes of probabilities for reactions and transitions are smaller than α. In this case, one can easily show that

the errors are of the order of $\mathcal{O}(\alpha^2)$. On the other hand, one fast reaction channel is, thus, sufficient to use very small time steps for all reaction channels and all transitions, so that this definition of Δt can only be good if the occurring probabilities are not too inhomogeneous. For general systems, one has to look for better choices. To justify such choices, however many test simulations will be necessary, so that a discussion of these choices is beyond the view of the present study.

COAST is the first hybrid algorithm, to my knowledge, that spans three different regimes of modeling and, therefore, the whole spectrum of occurring particle amounts in the most efficient way. All other algorithms (cf. Table 3.3) only consider small and large particle numbers. COAST is the only algorithm designed for strongly fluctuating systems covering a large variety of molecular abundances, which are likely to occur in signal transduction pathways. Since intermediate particle numbers are the most common scenario, the idea of using a Gaussian distribution and reducing the amount of random numbers is new in this scientific field. So far the most efficient stochastic algorithms in this area are derivatives of the *τ-leap method*. Here several reactions are allowed to take place in one time step, but these are calculated by binomial or Poisson-distributions; therefore for every reaction occurring one random number is used. COAST uses only one random number in a time step for one reaction, this results in an enormous cut down of computational effort.
COAST, as an hybrid algorithm, has characteristics of deterministic and stochastic approaches (cf. Table 2.4). Small volumes have the resulting effect that noise becomes significantly important (cf. Section 2.3.1). COAST pays respect to that, because a decreasing volume makes it more likely the stochastic Σ-regime is activated.
This is the most important advantage of COAST referring to implementation and usability. COAST is the only algorithm so far, that uses an intuitive error control parameter α. Other algorithms (cf. Section 2.4.2 and Table 3.3) like the hybrid methods of Kiehl et al. [2004] or Haseltine [2002] demand a direct intervention by the user to divide reactions in "slow" and "fast" reactions. Kiehl is treating reactions with low reaction probabilities by the *Next Reaction Method* and "fast" reactions (reactions with high reaction probabilities) by a deterministic approach, while Haseltine uses the *First Reaction Method* for the "slow" reactions. Nevertheless, the user is the one to do a fixed separation. If one of the reaction channels changes during the process from "slow" to "fast" there is now way to change the setting.

Therefore an automatic division of the reaction channels in every time step is superior. The *Maximum Time Step Method* [Puchalka and Kierzek, 2004] is an example for these group of algorithms. It uses the *Next Reaction Method* for single reactions and a *tau-leap method* for faster reactions. Puchalka uses three parameters r, n and κ. Only κ which is the maximal time step, has an intuitive meaning. The other two values are supposed to be selected "empirically". On the other hand COAST is only using one parameter α (for diffusion a spatial resolution is needed) that is defining the accuracy of the results. This value defined by the user is in direct manner defin-

ing the maximal allowed divergence between the experimental results and the exact values.

In the experiments performed, COAST was able to show that its performance and accuracy is even better than expected by the set error parameter α. COAST was designed to be fast, therefore one has to make a compromise presented in the form of the allowed error parameter. The runtime experiments were able to show that even a small α-value has a deeply positive impact on the speed of the calculations.

COAST worked well for reaction and diffusion problems. A next developing step would be a comprehensive model combining the application of COAST to reaction and to diffusion to model reaction-diffusion systems in one dimension.

Finally, I would like to address a topic that is a crucial aspect for modeling chemical processes within cellular structures. In the literature it is known as macromolecular crowding, molecular crowding and also as macromolecular confinement [Chebotareva et al., 2004]. However, it is more accurately termed as the "excluded volume effect". Biochemical processes in living systems occur in media containing high concentrations of macromolecules (50-400 mg/ml) [Ellis and Minton, 2003]. The present molecules are packed in such a way that they do not leave enough space for other molecules of their kind. This excluded volume is lost for these molecules. The theoretical aspects of excluded volume on chemical reactions has been discussed by Hall et al., Winzor et al. and others [Hall and Minton, 2003, Winzor and Wills, 1995]. Crowding has a complex effect on the rate of biochemical reactions. Simply put, as the activity of a specific particle in a crowded environment is increased, the diffusibility is reduced, and the probability of two particles meeting and reacting decreases. Of course, the overall result of these opposing factors depends on the nature of each reaction [Chebotareva et al., 2004].

So far we know the excluded volume effect cannot be neglected, even though the consequences are still the topic of present studies. Therefore, the only way to currently model the excluded volume effect is to model single molecules with a distinct volume and shape, which is computationally very inefficient.

As mentioned before, COAST was originally designed for diluted environments. Nevertheless, if a better understanding of the cellular processes is known, a better algorithm based on COAST can be developed. To date, however, the current degree of understanding is still not enough.

I would like to finish with a citation by Luby-Phelps [2000]:

"The potential impact of actual intracellular conditions on the kinetics, mechanisms, and regulation of metabolism make it imperative to reexamine continuum descriptions of cellular biochemistry that have been extrapolated from reductionist experiments carried out in dilute solution."

Modeling relies on accurate information. Mathematical models can only be developed on the base of reliable data. So far there is still a big leak of details on exact cellular structure and the interactions between all the parts of a cell. Only with

better knowledge, more background information, and maybe a better approach on cellular kinetics, can modeling approach reality.

However, COAST provides a new and significant step toward that goal.

Bibliography

P.S. Agutter, P.C. Malone, and D.N. Wheatley. Intracellular transport mechanisms: A critique of diffusion theory. *J. theor. Biol.*, 176:261–272, 1995. 63, 109

P.S. Agutter, P.C. Malone, and D.N. Wheatley. Diffusion theory in biology: a relic of mechanistic materialism. *J Hist Biol*, 33(1):71–111, 2000. 21

J.H. Ahrens and U. Dieter. Computer generation of poisson deviates from modified normal distributions. *ACM Transactions on Mathematical Software*, 8(2):163–179, June 1982. 107

M. Ander, P. Beltrao, and al. Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Systems Biology*, 1(1):129–138, 2004. 24, 37

A. Arkin, J. Ross, and H.H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells. *Genetics*, 149 (4):1633–1648, Aug 1998. 13, 19

F. Baras and M.M. Mansour. Reaction-diffusion master equation: A comparison with microscopic simulations. *Physical Review E Statistical Physics, Plasmas, Fluids, and related interdisciplinary Topics*, 54(6):6139–6148, Dec 1996. 25

N. Barkai and S. Leibler. Circadian clocks limited by noise. *Nature*, 403(6767): 267–268, Jan 2000. 90

Brian M. Baynes and Bernhardt L. Trout. Rational design of solution additives for the prevention of protein aggregation. *Biophys. J.*, 87(3):1631–1639, 2004. 23

B. P. Belousov. A periodic reaction and its mechanism. *Sbornik Referatov po Radiatsionni Meditsine*, page 145, 1958. 81, 83

U.S. Bhalla. The chemical organization of signaling interaction. *Bioinformatics*, 18 (6):855–863, 2002. 13

U.S. Bhalla. Signaling in small subcellular volumes. ii. stochastic and diffusion effects on synaptic network properties. *Biophys J*, 87(2):745–753, Aug 2004. 8, 36

U.S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283(5400):381–387, Jan 1999. 2

G.E.P. Box and M.E. Muller. A note on the generation of random normal deviates. *Ann. Math. Stat*, 29:610–611, 1958. 61, 107

A Boyarsky, PB Noble, and SC Peterson. Chemotaxis in vitro. quantitation of human granulocyte movement using a stochastic differential equation. *Biophys. J.*, 16: 249–259, 1976. 22

P. Buchner. Bemerkungen zur sterlingschen formel. *Elem. d. Math*, 6:8–11, 1951. 43

R. Bundschuh, F. Hayot, and C. Jayaprakash. Fluctuations and slow variables in genetic networks. *Biophysical Journal*, 84:1606–1615, 2003. 4

J. C. Butcher. *The Numerical Analysis of Ordinary Differential Equations.* John Wiley, New York, 1987. 16

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, Sep 2004a. 29

Y. Cao, L.R. Petzold, M. Rathinam, and D.T. Gillespie. The numerical stability of leaping methods for stochastic simulation of chemically reacting systems. *J Chem Phys*, 121(24):12169–12178, Dec 2004b. 5

A. Chatterjee, K. Mayawala, J.S. Edwards, and D.G. Vlachos. Time accelerated monte carlo simulations of biological networks using the binomial τ-leap method. *Bioinformatics*, 21(9):2136–2137, May 2005. 30

N.A. Chebotareva, B.I. Kurganov, and N.B. Livanova. Biochemical effects of molecular crowding. *Biochemistry (Mosc)*, 69(11):1239–1251, Nov 2004. 114

L. Chong and L.B. Ray. Whole-istic biology. *Science*, 295:1661, March 2002. 11

B.G. Cox. *Modern Liquid Phase kinetics.* Oxford university press, 1994. 17

B. deFinetti. *Theory of Probability.* John Wiley & Sons, 1974. 43, 48

M. Delbrück. Statistical fluctuations in autocatalytic reactions. *Chemical Physics*, 8 (1):120–124, 1940. 5

D.H. Deutsch. Did robert brown observe brownian motion: probably not. *Bulletin of the American Physical Society*, 36(4):1374, April 1991. reported in scientific american, 265: 20, 1991. 21

Pawan Dhar, Tan Chee Meng, Sandeep Somani, Li Ye, Anand Sairam, Mandar Chitre, Zhu Hao, and Kishore Sakharkar. Cellware–a multi-algorithmic software for computational systems biology. *Bioinformatics*, 20(8):1319–1321, 2004. 35

P.K. Dhar, T.C. Meng, S. Somani, L. Ye, K. Sakharkar, A. Krishnan, A.B. Ridwan, S.H. Wah, M. Chitre, and Z. Hao. Grid cellware: the first grid-enabled tool for modelling and simulating cellular processes. *Bioinformatics*, 21(7):1284–1287, Apr 2005. 78, 107

J. Doyle. Computational biology: Beyond the spherical cow. *Nature*, 411(6834): 151–152, May 2001. 1, 14

J. Elf and M. Ehrenberg. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biology*, 2:230–236, 2004. 37

J. Elf, A. Doncic, and M. Ehrenberg. Mesoscopic reaction-diffusion in intracellular signaling. *Proc. of SPIE*, 5110:114–125, 2003. 62

R.J. Ellis and A.P. Minton. Cell biology: join the crowd. *Nature*, 425(6953):27–28, Sep 2003. 114

Leibler S.A. Elowitz, M.B. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000. 2

M.B. Elowitz, A.J. Levine, and et al. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, Aug 2002. 6, 19

D. Endy and R. Brent. Modelling cellular behaviour. *Nature*, 18(409):391–395, January 2001. 3, 8, 13, 62

I. Epstein and J.A. Pojman. *An Introduction to Nonlinear Chemical Dynamics*. Oxford University Press, 1998. 16

SN Ethier and TG Kurtz. *Markov Processes - Characterization and Convergence*. John Wiley & Sons, 2005. 48

M.G.Th. Fechner. Ueber umkehrungen der polarität in der einfachen kette (sur la polarité inversée des circuits électriques). *Schweigger Journal für Chemie und Physik*, 53:129–151, 1828. 83

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, 3rd edition, 1970. 43, 66, 79

J.E. Ferrell and E.M. Machleder. The biochemical basis of an all-or-none cell fate switch in xenopus oocytes. *Science*, 280:895–898, 1998. 2

J.E. Jr. Ferrell. Building a cellular switch: more lessons from a good egg. *BioEssays*, 21:866–870, 1999. 2, 14

A. Fick. Ueber diffusion. *Poggendorf's Annalen der Physik*, 94:59–86, 1855. 21

R.J. Field and R.K. Noyes. Oscillations in chemical systems. iv. limit cycle behaviour in a model of a real chemical reaction. *The Journal of Chemical Physics*, 60(5): 1877–1884, March 1974. 85

C.C. Fink, B. Slepchenko, II. Moraru, J. Watras, J.C. Schaff, and L.M. Loew. An image-based model of calcium waves in differentiated neuroblastoma cells. *Biophys J*, 79(1):163–183, Jul 2000. 7, 9

M. Friedel and J.E. Shea. Self-assembly of peptides into a β-barrel motif. *Journal of Chemical Physics*, 120(12):5809–5823, 2004. 23

M.A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A.*, 104:1876–1889, 2000. 5, 13, 29, 62, 63

D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comp. Phys.*, 22:403–434, 1976. 8, 13, 25, 59, 77, 78

D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Chem. Phys.*, 81(25):2340–2361, 1977. 5, 8, 27, 59, 62, 63, 77, 78, 85, 106, 112

D.T. Gillespie. The chemical langevin equation. *Journal of Chemical Physics*, 113 (1):297–306, July 2000. 20

D.T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *Journal of Chemical Physics*, 115(4):1716–1733, July 2001. 8, 30

D.T. Gillespie and L.R. Petzold. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.*, 119:8229–8234, 2003. 30

N.C. Giri. *Introduction to Probability and Statistics (Part I).* Marcel Dekker, New York, p:65, 1974. 30

P Glansdorff and I Prigogine. *Thermodynamics of structure, stanility and fluctuations.* New York Wiley, 1971. 83

A. Goldbeter. *Biochemical Oscillations and Cellular Rhythms: The molecular bases of periodic and chaotic behaviour.* Cambridge University Press, 1996. 81

A. Goldbeter and D.E. Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl. Acad. Sci. USA*, 78:6840–6844, 1981. 2

A Goldbeter and DE Koshland. Ultrasensitivity in biochemical systems controlled by covalent modification. *The Journal of Biological Chemistry*, 259(23):14441–14447, December 1984. 18

D. Goodsell. *The Machinery of Life*. Springer-Verlag, Berlin, Germany, 1993. 3

D.S. Goodsell. Inside a living cell. *Trends Biochem Sci*, 16(6):203–206, Jun 1991. 62

J. Górecki, A.L. Kawczyñski, and Nowakowski B. Master equation and molecular dynamics simulations of spatiotemporal effects in a bistable chemical system. *J. Phys. Chem. A*, 103:3200–3209, 1999. 25

I.S. Gradshteyn and I.M. Ryshik. *Table of Integrals, Series, and Products - Corrected and Enlarged Edition*. Academic Press, New York, 1980. 53

Olli Haavisto. *Modeling and Simulation in Cellware*, November, 3 2004. 34

J.B.S. Haldane. *Enzymes*. Longmans, Green and Co., London, 1930. 5

D. Hall and A.P. Minton. Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochim Biophys Acta*, 1649(2):127–139, Jul 2003. 114

P.J. Halling. Do laws of chemistry apply to living cells? *Trends in Biochemical Sciences*, 14:317–318, 1989. 5

P. Hänggi, P.Talkner, and M.Borkovec. Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys.*, 62:251–341, 1990. 101

J.B. Haseltine, E.L. ; Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical reactions. *J. Chem. Phys.*, 117(15):6959–6969, 2002. 6, 31, 37, 113

J. et al. Hasty. Noise-based switches and amplifiers for gene expression. *Proc. Natl. acad. Sci. USA*, 97:2075–2080, 2000. 19

J. Hattne, D. Fange, and J. Elf. Stochastic reaction-diffusion simulation with mesord. *Bioinformatics*, 21(12):2923–2924, Jun 2005. 24, 25, 37

D.J. Hebert. Simulation of stochastic reaction-diffusion systems. *Mathematics ans Computers in Simulation*, 34:411–342, 1992. 59

R. Heinrich and S. Schuster. *The regulation of cellular systems*. Chapman and Hall, New York, 1996. 16

B. Hille. *Ionic Channels of Excitable Membranes*. Sinauer Associates, Inc., Sunderland, MA, 1984. 14

A. L. Hodgkin and A. F. Huxley. A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J. Physiol. (Lond.)*, 117:500–544, 1952. 14

C.Y.F. Huang and J.E. jr. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. USA*, 93:10078–10083, September 1996. 2, 18

J.P. Joule. *Philos.Mag.*, 24:106, 1844. 83

Ingrid M. Keseler, Julio Collado-Vides, Socorro Gama-Castro, John Ingraham, Suzanne Paley, Ian T. Paulsen, Martin Peralta-Gil, and Peter D. Karp. Ecocyc: a comprehensive database resource for escherichia coli. *Nucl. Acids Res.*, 33 (suppl-1):D334–337, 2005. 3

B.N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascade. *Eur. J. Biochem.*, 267:1583–1588, 2000. 14, 18

T.R. Kiehl, R.M. Mattheyses, and M.K. Simmons. Hybrid simulation of cellular behavior. *Bioinformatics*, 20(3):316–322, 2004. 6, 31, 37, 113

Barakat R. Kiester A.R. Exact solutions to certain stochastic differential equation models of population growth. *Theor Popul Biol.*, 6(2):199–216, October 1974. 22

H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, Mar 2002. 11, 12

H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemial reactions. *Physica*, VII:284–304, 1940. 5, 100, 111

M.S. Ladinsky, D.N. Mastronarde, J.R. McIntosh, K.E. Howell, and L.A. Staehelin. Golgi structure in three dimensions: functional insights from the normal rat kidney cell. *J Cell Biol*, 144(6):1135–1149, Mar 1999. 3

A. Levchenko, J. Bruck, and P.W. Sternberg. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *PNAS*, 97:5818 – 5823, May 2000. 18

M.D. Levin, C.J. Morton-Firth, W.N. Abouhamad, R.B. Bourret, and D. Bray. Origins of individual swimming behavior in bacteria. *Biophys J*, 74(1):175–181, Jan 1998. 6

David G. Levitt. Modeling of ion channels. *J. Gen. Physiol.*, 113(6):789–794, June 1 1999. 3

K. Lipkow, S.S. Andrews, and D. Bray. Simulated diffusion of phosphorylated chey through the cytoplasm of escherichia coli. *J Bacteriol*, 187(1):45–53, Jan 2005. 24, 36

J. Lippincott-Schwartz, E. Snapp, and A. Kenworthy. Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol*, 2(6):444–456, Jun 2001. 9

L.M. Loew and J.C. Schaff. The virtual cell: a software environment for computational cell biology. *Trends Biotechnol*, 19(10):401–406, Oct 2001. 14

K. Luby-Phelps. Cytoarchitecture and physical properties of cytoplasm: Volume, viscosity, diffusion, intracellular surface area. *International Review of Cytology*, 192:189–221, 2000. 109, 114

K. Luby-Phelps, D. Lansing Taylor, and F. Lanni. Probing the structure of cytoplasm. *The Journal of Cell Biology*, 102:2015–2022, June 1986. 108

H.H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A*, 94(3):814–819, Feb 1997. 6, 19

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. 5

H. McIlwain. The magnitude of microbial reactions involving vitamin-like compounds. *Nature*, 158:898–902, 1946. 5

I. Mellman and T. Misteli. Computational cell biology. *The Journal of Cell Biology*, 161(3):463–464, 2003. 1

P. Mendes. Gepasi: a software package for modelling the dynamics, steady states and control of biochemical and other systems. *Comput. Appl. Biosci.*, 9(5):563–571, 1993. 5

T.C. Meng, S. Somani, and P. Dhar. Modeling and simulation of biological systems with stochasticity. *In Silico Biol*, 4(2):0024–0024, Apr 2004. 15, 19

L. Michaelis and M.L. Menten. Die kinetik der invertinwirkung. *Biochem. Z.*, 49: 333–369, 1913. 8, 94

A.P. Minton. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem*, 276(14): 10577–10580, Apr 2001. 13

M. Möller and H. Wagner. An efficient stochastic algorithm for linear diffusion models. *submitted*, 2005. 7, 59

D. Noble. Modeling the heart–from genes to cells to the whole organ. *Science*, 295 (5560):1678–1682, Mar 2002. 14

R.M. Noyes, R.K. Field, and E Körös. Oscillations in chemical systems detailed mechanism in a system showing temporal oscillations. *J. Am. Chem. Soc.*, 94: 1394, 1972. 83

D Nuhr. ich bin's nuhr. CD, 2004. v

B.P. Oelveczky and A.S. Verkman. Monte carlo analysis of obstructed diffusion in three dimensions: Application to molecular diffusion in organelles. *Biophysical Journal*, 74-:2722–2730, May 1998. 8, 9

B. Oksendahl. *Stochastic Differential Equations.* Springer-Verlag, New York, 1985. p. 20 ff. 64

E. Ozbudak and al. Regulation of noise in the expression of single gene. *Nat. Genet.*, 1:69–73, 2002. 6, 19

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, Mar 2004. 31, 113

C.V. Rao and A.P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *Journal of Chemical Physics*, 118(11):4999–5010, 2003. 6, 31

C.V. Rao, D.M. Wolf, and A.P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, Nov 2002. 5, 39

A. Renyi. Treatment of chemical reactions by means of theory of stochastic processes. *Magyar Tud. Akad. Alkalm. Mat. int. Kozl.*, 2:93–101, 1954. 5

H. Resat, H.S. Wiley, and Dixon D.A. Probability-weighted dynamic monte carlo method for reaction kinetics simulations. *J. Phys. Chem. B*, 105:11026–11034, 2001. 32

I.L. Ross and al. Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell. Biol.*, 72:177–185, 1994. 19

F.F Runge. *Zur Farben-Chemie. Musterbilder für die Freunde des Schönen und zum Gebrauch für Zeichner, Maler, Verzierer und Zeugdrucker.* Berlin Sebstverlag, 1850. 83

M.A. Savageau. Biochemical systems analysis. ii. the steady-state solutions for an n-pool system using a power-law approximation. *J. Theor. Biol.*, 25:370–379, 1969. 18, 19

J. Schaff, C.C. Fink, B. Slepchenko, J.H. Carson, and L.M. Loew. A general computational framework for modeling cellular structure and function. *Biophys J*, 73 (3):1135–1146, Sep 1997. 5, 22, 36

C. Schönbein. *Archives d'electricité*, 2:269, 1842. 83

K. Schulten and I. Kosztin. *Lectures in Theoretical Biophysics*. University of Illinois at Urbana-Champaign, 1999. 99

M. Schwehm. Parallel stochastic simulation of whole-cell models. In *Proc. 2nd Int. Conf Systems Biology (ICSB 2001)*, pages 333–341. ICSB, Omnipress, 2001. 3, 8, 28

B.M. Slepchenko, J.C. Schaff, J.H. Carson, and L.M. Loew. Computational cell biology: spatiotemporal simulation of cellular events. *Annu Rev Biophys Biomol Struct*, 31:423–441, 2002. 36

A.E. Smith, B.M. Slepchenko, J.C. Schaff, L.M. Loew, and I.G. Macara. Systems analysis of ran transport. *Science*, 295(5554):488–491, Jan 2002. 14

M. von Smoluchowski. Stability and instability in dispersed systems. *Z. Phys. Chem.*, 92:129–136, 1917. 108

R. Srivastava, L. You, J. Summers, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *J Theor Biol*, 218(3):309–321, Oct 2002. 5

J.R. Stiles and al. *Computational Neuroscience*, chapter Monte Carlo Simulation of Neurotransmitter release using MCell, A General Simulator of Cellular Physiological Processes, pages 279–284. J.M. Plenum, NewYork, 1998. 24

D.E. Strier, C. Ventura, and al. Saltatory and continous calcium waves and the rapid buffering approximation. *Biphys. J.*, 85:3575–3586, 2003. 9

A.B. Stundzia and C.J. Lumsden. Stochastic simulation of coupled reaction-diffusion processes. *Journal of Computational Physics*, 127:196–207, 1996. 8, 24, 62

K. Takahashi and al. Computational challenge in cell simulation: A software engineering approach. *IEEE Intelligent Systems*, 17:64–71, 2002. 4

K. Takahashi, K. Kaizu, B. Hu, and M. Tomita. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics*, 20(4):538–546, Mar 2004. 6, 31, 37, 39

K. Takahashi, S.N. Arjunan, and M. Tomita. Space in systems biology of signaling pathways–towards intracellular molecular crowding in silico. *FEBS Lett*, 579(8): 1783–1788, Mar 2005. 25

T. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *J Chem Phys*, 121(21):10356–10364, Dec 2004. 30

M. Tomita, K. Hashimoto, K. Takahashi, TS. Shimizu, Y. Matsuzaki, F. Miyoshi, and K. Saito. E-cell: software environment for whole-cell simulation. *Bioinformatics*, 15(1):72–84, Jan 1999. 5

R.Y. Tsien. The green fluorescent protein. *Annu. Rev. Biochem.*, 67:509–544, 1998. 9

T.E. Turner, S. Schnell, and K. Burrage. Stochastic approaches for modelling in vivo reactions. *Comput Biol Chem*, 28(3):165–178, Jul 2004. 20, 26

J.J. Tyson. *The Belousov-Zhabotinskii reaction*. Springer Verlag Berlin, 1976. 83

J.J. Tyson. Modeling the cell division cycle: cdc2 and cyclin interactions. *Proc Natl Acad Sci U S A*, 88(16):7328–7332, Aug 1991. 14

L. Vereecken, G. Huyberechts, and J. Peeters. Stochastic simulation of chemically unimolecular reactions. *Journal of Chemical Physics*, 106(16):6564–6573, April 1997. 5

JM. Vilar, HY. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci U S A*, 99(9):5988–5992, Apr 2002. 90, 91

J.M. Vilar, C.C. Guet, and S. Leibler. Modeling network dynamics: the lac operon, a case study. *J Cell Biol*, 161(3):471–476, May 2003. 1

E. O. Voit. *Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity*. New York: Van Nostrand Reinhold, 1991. 18

E.O. Voit. Canonical modeling: review of concepts with emphasis on environmental health. *Environ Health Perspect*, 108 Suppl 5:895–909, Oct 2000. 14, 15

E.O. Voit. Metabolic modeling: a tool of drug discovery in the post-genomic era. *Drug Discov Today*, 7(11):621–628, Jun 2002. 3, 12, 13

Larry Wasserman. *All of Nonparametric Statistics*. Springer, 2006. 50

JR Weimar. Cellular automata for reaction-diffusion systems. *Parallel Computing*, 23(11):1699–1715, 1997. 24

N. Wiener. *Cybernetics: or Control and Communication in the Animal and the Machine*. The MIT Press, Cambridge, MA, 1948. 12

D.J. Winzor and P.R. Wills. Thermodynamic nonideality of enzyme solutions supplemented with inert solutes: yeast hexokinase revisited. *Biophys Chem*, 57(1): 103–110, Dec 1995. 114

A. M. Zhabotinsky. Periodical oxidation of malonic acid in solution. *Biofizika*, 9: 306–311, 1964. 83

Robert Zwanziger. A chemical langevin equation with non-gaussian noise. *J. Phys. Chem. B*, 105:6472–6473, 2001. 20

A Numerical solution of chemical differential equations

Here the solutions of the differential equations for basic chemical reactions are derived. A, B, C and D are substrates. k and l are the deterministic velocity constants. To simplify the writing in the equations $S = [S]$, which means that all specie-symbols characterize concentrations, and S_0 is the initial concentration of a substrate.

A.1 Bimolecular Reaction; One Specie

$2\,C \xrightarrow{k} A + B \cong A + B \xrightarrow{k} C \quad (\textbf{if } A_0 = B_0)$

$$
\begin{aligned}
\frac{dC}{dt} &= -k \cdot C^2 \\
dC &= -k\,C^2\,dt \\
\frac{dC}{C^2} &= -k\,dt \\
\int_{C_0}^{C} C^{-2}\,dC &= -k \int_{t_0}^{t} dt \\
-C^{-1}\Big|_{C_0}^{C} &= -k\,t \\
-\frac{1}{C} + \frac{1}{C_0} &= -kt \\
\frac{1}{C_0} + kt &= \frac{1}{C} \\
\text{which results in:} \qquad C &= \frac{1}{\frac{1}{C_0} + kt}
\end{aligned}
\tag{A.1}
$$

A.2 Unimolecular Reaction

$A \xrightarrow{k} 2C$

$$
\begin{aligned}
\frac{dA}{dt} &= -k \cdot A \\
\frac{dA}{A} &= -k\,dt \\
\int_{A_0}^{A} dlnA &= -k \int_{t_0}^{t} dt \\
\ln A - \ln A_0 &= -kt \\
\ln A &= -kt + \ln A_0 \\
\text{which results in:} \qquad A &= A_0\,e^{-kt}
\end{aligned}
\tag{A.2}
$$

A.3 Bimolecular Reaction; Two Species (Part 1)

$A + B \overset{k}{\rightleftharpoons} C$

$$
\begin{aligned}
\frac{dC}{dt} &= k\,A\,B - k\,C \,/A = B = A_0 - C \\
\frac{dC}{dt} &= k\,(A_0 - C)^2 - k\,C \\
\frac{dC}{dt} &= k\,(A_0^2 - 2CA_0 + C^2 - C) \\
\int_{t_0=0}^{t} k dt &= \int_{C_0=0}^{C} \frac{dC}{A_0^2 + C(-1-2A_0) + C^2} \\
with \int \frac{dx}{ax^2+bx+c} &= \frac{1}{\sqrt{-\Delta}} \ln \frac{2ax+b-\sqrt{-\Delta}}{2ax+b+\sqrt{-\Delta}} \\
& \quad a = 1 \quad b = -1-2A_0 \\
& \quad c = A_0^2 \quad \Delta = 4ac - b^2 = -4A_0 - 1
\end{aligned}
$$

$$
kt = \frac{1}{\sqrt{4A_0+1}} \ln \frac{2C-2A_0-1-\sqrt{4A_0+1}}{2C-2A_0-1+\sqrt{4A_0+1}} - \frac{1}{\sqrt{4A_0+1}} \ln \frac{-2A_0-1-\sqrt{4A_0+1}}{-2A_0-1+\sqrt{4A_0+1}}
$$

$$
\begin{aligned}
kt\sqrt{4A_0+1} + \ln \frac{-2A_0-1-\sqrt{4A_0+1}}{-2A_0-1+\sqrt{4A_0+1}} &= \ln \frac{2C-2A_0-1-\sqrt{4A_0+1}}{2C-2A_0-1+\sqrt{4A_0+1}} \\
\underbrace{kt\,e^{\sqrt{4A_0+1}} + \frac{-2A_0-1-\sqrt{4A_0+1}}{-2A_0-1+\sqrt{4A_0+1}}}_{\mu} &= \frac{2C-2A_0-1-\sqrt{4A_0+1}}{2C-2A_0-1+\sqrt{4A_0+1}} \\
2\mu C + \mu\,(-2A_0-1+\sqrt{4A_0+1}) &= 2C - 2A_0 - 1 - \sqrt{4A_0+1} \\
C\,(2\mu-2) &= -2A_0-1-\sqrt{4A_0+1} - \mu(-2A_0-1+\sqrt{4A_0+1}) \\
C &= \frac{-2A_0-1-\sqrt{4A_0+1}-\mu(-2A_0-1+\sqrt{4A_0+1})}{2\mu-2} \qquad \text{(A.3)}
\end{aligned}
$$

For t tending toward infinity the system reaches an equilibrium with the following concentration for C:

$$
\lim_{\substack{t\to\infty \\ =\mu\to\infty}} C = \frac{2A_0+1-\sqrt{4A_0+1}}{2}.
$$

A.4 Bimolecular Reaction; Two Species (Part 2)

$A + B \overset{k,l}{\rightleftharpoons} 2C$

assumption: $A_0 = B_0 \Rightarrow A = B\,; C_0 = 0 \Rightarrow C = 2(A_0 - A)$

$$\begin{aligned}
\frac{dA}{dt} &= -kA^2 + 4l(A_0 - A) \\
\frac{dA}{dt} &= 4lA_0^2 - 8lA_0A + (4l - k)A^2 \\
\int_{t_0=0}^{t} dt &= \int_{A_0=0}^{A} \frac{dA}{(-k+4l)A^2 + (-8lA_0)A + 4lA_0^2} \\
with \int \frac{dx}{ax^2+bx+c} &= \frac{1}{\sqrt{-\Delta}} \ln \frac{2ax + b - \sqrt{-\Delta}}{2ax + b + \sqrt{-\Delta}} \\
& \quad a = -k + 4l \quad b = -8lA_0 \\
& \quad c = 4lA_0^2 \qquad \Delta = 4ac - b^2 = -16lkA_0
\end{aligned}$$

$$\begin{aligned}
t &= \frac{1}{4A_0\sqrt{lk}} \ln \frac{2(-k+4l)\frac{A}{A_0} - 8lA_0 - 4A_0\sqrt{lk}}{2(-k+4l)\frac{A}{A_0} - 8lA_0 + 4A_0\sqrt{lk}} \\
&\quad - \frac{1}{4A_0\sqrt{lk}} \ln \frac{(-k+4l)\frac{A_0}{2} - 8lA_0 - 4A_0\sqrt{lk}}{(-k+4l)\frac{A_0}{2} - 8lA_0 + 4A_0\sqrt{lk}} \\
4tA_0\sqrt{lk} &= \ln \frac{(-k+4l)\frac{A}{A_0} - 4l - 2\sqrt{lk}}{(-k+4l)\frac{A}{A_0} - 4l + 2\sqrt{lk}} - \ln \frac{-k - 2\sqrt{lk}}{-k + 2\sqrt{lk}} \\
e^{4tA_0\sqrt{lk}} &= \frac{(4l-k)\frac{A}{A_0} - 4l - 2\sqrt{lk}}{(4l-k)\frac{A}{A_0} - 4l + 2\sqrt{lk}} \, \frac{k - 2\sqrt{lk}}{k + 2\sqrt{lk}} \\
\underbrace{e^{4tA_0\sqrt{lk}} \frac{k + 2\sqrt{lk}}{k - 2\sqrt{lk}}}_{\mu} &= \frac{(4l-k)\frac{A}{A_0} - 4l - 2\sqrt{lk}}{(4l-k)\frac{A}{A_0} - 4l + 2\sqrt{lk}}
\end{aligned}$$

$$\begin{aligned}
A\frac{\mu}{A_0}(4l-k) + \mu(-4l + 2\sqrt{lk}) &= A\frac{4l-k}{A_0} - 4l - 2\sqrt{lk} \\
A\left(\frac{\mu}{A_0}(4l-k) - \frac{4l-k}{A_0}\right) &= -\mu(-4l + 2\sqrt{lk}) - 4l - 2\sqrt{lk} \\
A &= A_0 \, \frac{\mu(4l - 2\sqrt{lk}) - 4l - 2\sqrt{lk}}{(4l-k)\,(\mu - 1)}
\end{aligned} \tag{A.4}$$

$$\text{if t=0} \Rightarrow \underline{A = A_0}$$

$$\lim_{\substack{t\to\infty \\ =\mu\to\infty}} A = A_0 \frac{4l - 2\sqrt{lk}}{4l - k}$$

$$\text{if } l = k \Rightarrow \underline{A = \tfrac{2}{3} A_0}$$

A.5 Bimolecular Reaction; Two Species (Part 3)

$A + B \xrightarrow{k} P$, where P is an arbitrary product.

$$
\begin{aligned}
I. \qquad \frac{dA}{dt} &= -A \cdot B \cdot k \\
II. \quad B_0 - B &= A_0 - A \\
\Rightarrow \; B &= (B_0 - A_0) + A
\end{aligned}
$$

$$
\begin{aligned}
\frac{dA}{dt} &= -A \;\; [(B_0 - A_0) + A] \;\; k \\
\frac{dA}{dt} &= [A \;\; (B_0 - A_0) + A^2] \;\; k \\
\int_{A_0}^{A} \frac{dA}{A(B_0 - A_0) + A^2} &= -k \int_0^t dt \\
\left[\frac{1}{B_0 - A_0} \;\; \ln \frac{2A}{2A + 2(B_0 - A_0)} \right]_{A_0}^{A} &= -kt \\
\ln \frac{A}{A + (B_0 - A_0)} - \ln \frac{A_0}{B_0} &= -kt(B_0 - A_0) \\
e^{\ln \frac{A}{A+(B_0-A_0)} - \ln \frac{A_0}{B_0}} &= e^{-kt(B_0-A_0)} \\
e^{\ln \frac{A}{A+(B_0-A_0)}} \;\; e^{-\ln \frac{A_0}{B_0}} &= e^{-kt(B_0-A_0)} \\
\frac{A}{A + (B_0 - A_0)} &= e^{-kt(B_0-A_0)} \;\; \frac{A_0}{B_0} \\
A &= \frac{A_0}{B_0} \;\; e^{-kt(B_0-A_0)} \;\; (A + (B_0 - A_0)) \\
A &= \frac{A_0}{B_0} \;\; e^{-kt(B_0-A_0)} \;\; A + \frac{A_0}{B_0}(B_0 - A_0) \;\; e^{-kt(B_0-A_0)} \\
A \left(1 - \frac{A_0}{B_0} \;\; e^{-kt(B_0-A_0)} \right) &= \left(A_0 - \frac{A_0^2}{B_0} \right) \;\; e^{-kt(B_0-A_0)} \\
A &= \left(A_0 - \frac{A_0^2}{B_0} \right) \;\; e^{-kt(B_0-A_0)} \;\; \frac{1}{1 - \frac{A_0}{B_0} \;\; e^{-kt(B_0-A_0)}} \\
A &= \frac{(A_0 B_0 - A_0^2) \;\; e^{-kt(B_0-A_0)}}{B_0 - A_0 \;\; e^{-kt(B_0-A_0)}}
\end{aligned}
$$

B Belousov-Zhabotinsky Reaction

This appendix includes details on the Belousov-Zhabotinsky reaction presented in Figure 4.5.

B.1 The Composition

- 0.50M sodiumbromate solution ($NaBrO_3$)
- 1.50M malonic acid ($HOOCCH_2COOH$)
- 5.00M sulphuric(VI) acid (H_2SO_4)
- 0.30M sodiumbromide solution ($NaBr$)
- 0.01M ferroin solution

B.2 Reaction System

I.	$2Br^- + BrO_3{}^- + 3H^+ + 3H_2Mal$	$\rightarrow$	$3HBrMal + 3H_2O$
II.	$BrO_3^- + 4\,ferroin^{2+} + H_2Mal + 5H^+$	$\rightarrow$	$4\,ferriin^{3+} + HBrMal + 3H_2O$
III.	$4\,ferriin^{3+} + HBrMal + 2H_2O$	$\rightarrow$	$4\,ferroin^{2+} + HCOOH + 2CO_2 + 5H^+ + Br^-$
IV.	$3BrO_3^- + 5H_2Mal + 3H^+$	$\rightarrow$	$3HBrMal + 2HCOOH + 4CO_2 + 5H_2O$

Table B.1: The chemical reactions describing the oscillating system leading toward Figure 4.5
$BrO_3{}^-$ (bromate); H_2Mal (malonic acid); $HBrMal$ (bromomalonic acid); Br^- (bromide); CO_2 (carbondioxyde); H_2O (water)

B.3 Chemical Structures of Ferroin and Brome Malonic Acid

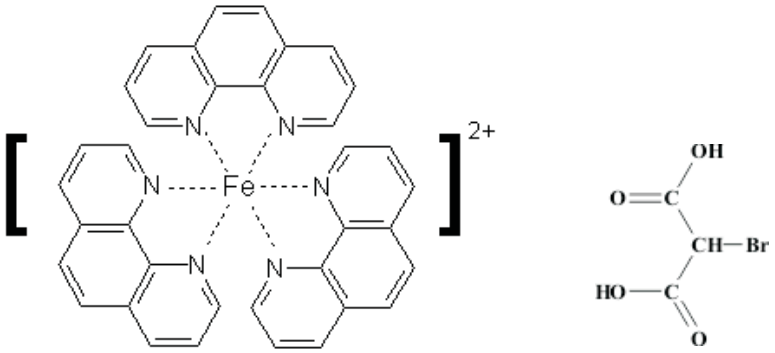

Figure B.1: *On the left one can see the red iron(II)-1,10-phenanthroline complex* $[Fe(C_{12}H_8N_2)_3]^{2+}$ *and on the right side bromomalonic acid.*

C Additional Information

C.1 URL-List of Cellular Simulators

Simulator	URL
SmartCell	http://smartcell.embl.de/
NEURON	www.neuron.yale.edu
GEPASI	http://www.gepasi.org/
StochSim	www.anat.cam.ac.uk/ compcell/StochSim.html
MesoRD	http://mesord.sourceforge.net/
MCell	www.mcell.cnl.salk.edu/
GENESIS	www.genesis-sim.org/GENESIS/
VirtualCell	www.nrcam.uchc.edu/

Table C.1: *Internet Representation of mentioned Simulators*