Karlsruhe Institute of Technology

Analysis and Simulation of Hybrid Models for Reaction Networks

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

von der Fakultät für Mathematik des
Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION

von
Dipl.-Biol. Michael Kreim
aus
Darmstadt

Tag der mündlichen Prüfung:	02.10.2014
Referent:	Prof. Dr. Tobias Jahnke
Korreferent:	Prof. Dr. Wilhelm Huisinga

Acknowlegements

I like to thank my advisor Prof. Dr. Tobias Jahnke for his support and mentoring in the last years and I like to thank Prof. Dr. Wilhelm Huisinga for agreeing to be my second referee. Further, I like to thank Vikram Sunkara for tootsierolling a taylorswift, for his friendship, for two amazing years, many helpful discussions and all the math he taught me. I like to thank Tudor Udrescu, Andreas Arnold and Marcel Mikl for the helpful discussions, the support over the years and the fun we had. Also, I like to thank all my coworkers for the nice time together in the IANM3, the discussions, lunches, parties: Sonja Becker, Prof. Tobias Jahnke, Tim Kreutzmann, Daniel Maurer, Prof. Andreas Rieder, JProf. Katharina Schratz, Andreas Schulz, Ramin Shirazi-Nejad, Nathalie Sonnefeld, Ekkachai Thawinan, Daniel Weiß, Prof. Christian Wieners, Robert Winkler, Wolfgang Müller, Nicolas Neuß and Martin Sauter. I also like to thank Vikram, Theresa, Marcel, Tudor and Sonja for proofreading and commenting on this thesis. Further, I like to thank PD Stefan Kühnlein and Jochen Schröder for their patience in teaching me algebra and Prof. Peter Nick and Dr. Jan Maisch for the nice discussions in the small project we had together. Last, but most important, I like to thank Theresa and Konstantin Stang, for all.

Contents

1 Introduction **1**

2 Mathematical Models for Reaction Networks **5**
2.1 Physical Motivation 5
2.2 Reaction Networks 7
2.3 Mathematical Models for Reaction Networks 9
2.3.1 Kurtz Process and Chemical Master Equation 9
2.3.2 Chemical Langevin Equation and Fokker-Planck Equation 13
2.3.3 Reaction Rate Equation and Liouville Equation 16
2.4 Connections between the Different Models 17
2.4.1 Derivation of the CLE and the RRE from the Kurtz process 17
2.4.2 Derivation of the FPE and the LVE from the CME 19
2.5 Summary 20

3 Numerical Methods **23**
3.1 Stochastic Simulation Algorithm 23
3.2 The Finite State Projection Method 25
3.3 Further Methods 27
3.4 Examples 28
3.4.1 Example 1: An Inflow Reaction 28
3.4.2 Example 2: A Dimerisation Reaction 31
3.4.3 Example 3: A Protein Network 33

4 The Thermodynamic Limit and Convergence of the Fokker-Planck Equation **39**
4.1 Scaled Reaction Networks 39
4.2 The Scaled Kramers-Moyal Expansion 43
4.3 Convergence of the Kurtz process to the RRE 45
4.4 Convergence of the FPE 46
4.4.1 A Numerical Example 54
4.4.2 Future Research 56

5 A Hybrid Model Combining the Kurtz process and the RRE **59**
5.1 Motivation 59
5.2 An Example 62
5.3 Notation and Definitions 65
5.4 Derivation of the Hybrid Model 69
5.4.1 Derivation of the Hybrid Process 69
5.4.2 Derivation of the Liouville master equation 70
5.4.3 Connection of Hybrid Process and LME 73
5.5 Marginal Distributions and Conditional Expectations 74
5.6 An Error Bound 76

5.7 Summary and Discussion 86

6 A Hybrid Model Combining the Kurtz process and the CLE 89
6.1 Motivation 89
6.2 A general derivation of the scaled Hybrid Process 89
6.3 Derivation of the Fokker-Planck master equation 92
6.4 A Numerical Example 93
6.5 An Error Bound for the FPME 96
6.6 Summary and Discussion 102

7 The Lac Operon - A 10-Dimensional Numerical Example 105
7.1 The Lac Operon Network 105
7.2 Numerical Simulation of the Lac Operon Network 109
7.3 Results and Discussion of the Numerical Simulations 110

8 Conclusions and Outlook 119

Appendix 123
A Probability Distributions 123
B Multi-Index Notation 125
C Bounds for Propensity Functions 126

Bibliography 132

List of Theorems, Lemmata, Definitions and Assumptions 139

List of Algorithms 142

List of Figures 143

List of Tables 146

Index 147

1 | Introduction

Modern biology has developed an increasing interest in mathematical modeling over the recent decades. Emerging fields like systems biology, computational biology or quantitative biology opened up new possibilities in the study of biochemical interactions of different molecules, processes and pathways inside cells, tissues and even whole organisms. Research in these areas has mainly focused on the study of gene regulatory networks and metabolic / signaling pathways. Traditionally, such studies have been carried out *in-vitro*, but recently they are increasingly performed *in-silico*. Some of the models used in computer simulations originate from the mathematical description of population dynamics [Mur02]. Initially, the main focus lied on modeling the earth population in order to predict future demographic developments. Later on, other species were considered and the interaction of populations or the spread of diseases were analysed. Recently, these models have been successfully applied for simulating the biochemical processes mentioned above.

The research conducted in the last decades resulted in the development of a general framework for a large class of these models: *Reaction Networks*.

A reaction network consists of several species that interact via reaction channels. These networks allow us to represent the *interactions* between different species. The concept of species in these networks is very abstract and can be interpreted as actual biological species like plants or animals. Alternatively, they can represent individuals belonging to some population that exhibit certain traits, like people infected with a disease that interact with people that are not (yet) infected, or the species can be seen as molecules that interact via chemical reactions.

Reaction networks* are very common in life sciences and can be found for example in genetics [MA99], physiology [MPM10, SMF+13], biochemistry [BTS02], ecology [HDD+11, Mur02], medicine [GGJ13] or even astrophysics [DP11]. However, they only give a qualitative representation of interactions, i.e. they answer the question how the reagents are connected, but not how many particles exist. But as we pointed out, we are interested in the time evolution of the *quantities* of the species. We want to answer questions like: Given that we have 7 billions of people on earth now, how many will we have in 50 years? How long will it take for a disease to reach its peak? Is a strategy (vaccination, quarantine, drugs) capable of reducing the numbers of deaths? How fast does a virus spread in a

*Throughout this work, a reaction network is to be understood as the *connections* between several species. Quantitative values are calculated by *models* after a network has been compiled (see next pages).

tissue? How does the inhibition or activation of an enzyme affect a biochemical pathway? How does a repressor or an enhancer control the outcome of a gene regulatory network?

The first model capable of simulating reaction networks was the deterministic Reaction Rate equation (RRE), an N-dimensional system of coupled ordinary differential equations. However, in the last decades it became obvious that many networks are highly stochastic and that the RRE is not suitable for detecting such effects [MA99, McQ67]. The switching of ion channels between open and closed states is just one example of stochastic effects in biology [DKB10]. Also, networks with bi- or multi-modal probability density functions cannot be accurately simulated with the RRE, which only outputs the expectation of the network. An example of a network with a bi-modal behaviour is the infection of a bacterium by a phage [SYSY02]. This network can evolve according to two different scenarios. Either the virus cannot infect the cell and then becomes extinct or it can proliferate and the disease spreads. Examples for networks with multi-modal probability distributions are stem cells, which develop into a large variety of different types of tissues [MML09].

A thermodynamic interpretation of reaction networks as a gas phase system with freely moving particles led to the derivation of a Markov jump process and the Chemical Master equation (CME). Parallel to this development, models basing on stochastic differential equations (Chemical Langevin equation (CLE)) and the Fokker-Planck equation (FPE) were derived. The most important difference between these models is the observation that the RRE, the FPE and the CLE have a continuous state space, while the CME and the Markov jump process have a discrete state space.

In 1972, Thomas Kurtz proved convergence of the Markov jump process to the Reaction Rate equation [Kur72], by introducing scaled reaction networks and presenting the "thermodynamic limit" technique. However, the convergence of the solution of the FPE to the solution of the CME was intensively discussed in the last decades [Kam61, Gil80, Gil00, Gar04].

In this thesis, we take Kurtz's result as a motivation and analyse the behaviour of different models. This analysis requires the comparison of models with discrete state space and models with continuous state space. We physically interpret the discrete states as particle numbers and the continuous states as concentrations. Based on this interpretation we derive scaled reaction networks and scaled versions for the different models. Further, we define a mapping that allows the comparison of probability density functions on discrete state spaces with densities on continuous state spaces.

We show, as a first important results, that the FPE can be motivated as an approximation of the CME by combining scaled reaction networks with a truncation of the so-called Kramers-Moyal expansion [Kra40, Moy49]. Then, we prove the convergence of the solution of the FPE to the solution of the CME, by showing an error bound depending on the first derivative with respect to space of the solution of the Fokker-Planck equation (section 4.4).

In 2002, Haseltine and Rawlings proposed the combination of two different models to new hybrid models [HR02]. Since then, interest in the usage and further development of hybrid models has continued to grow. Hybrid models combine the advantages of the

underlying sub-models. For example, a network where some species have large and some species have very small particle numbers is hard to solve numerically. The species with low copy numbers require the usage of the Markov Jump process (or the CME) to handle stochastic effects. On the other hand, the species with large particle counts cause a high numerical complexity and could be simulated by the RRE. A hybrid model combines these two approaches and simulates the small particle numbers as a stochastic process and the species with large particle numbers with the RRE. This procedure reduces the numerical complexity and conserves stochastic effects.

In chapter 5 we derive a hybrid model combining the Liouville equation and the CME and prove a convergence rate for the marginal distributions and conditional expectations of this model to the chemical master equation. Although this hybrid model has been discussed in literature for a while, no convergence rate has been proven so far.

In addition, we derive a second hybrid model combining the FPE and the CME (chapter 6). The model has also been discussed in literature, but an analysis of the convergence properties is missing. Therefore, we derive the stochastic process and an equation of motion for the probability distribution and discuss convergence properties. This analysis leads to a theorem that shows convergence of the distribution of the hybrid model to the solution of the CME and gives an error bound depending on the first space derivative of the solution of the hybrid model.

Before we conclude this section, we have to point out the crucial difference between *models* and *methods*. In the context of this thesis, a model is a stochastic process or a differential equation that models the time evolution of the state of a reaction network or its distribution. A method is a numerical scheme that generates a numerical solution of a model. A model is a particular description / interpretation of a reaction network, a method solves a model.

This work is structured as follows: We define the notation and the different models in chapter 2 and summarise numerical methods to solve them in chapter 3. Further, we discuss the connections between the different models (the Markov jump process, the Chemical Langevin equation, the Reaction Rate equation, the Chemical Master equation, the Fokker-Planck equation and the Liouville equation) by deriving all these models starting from the Markov jump process. In chapter 4, we introduce the concepts of the thermodynamic limit and of scaled reaction networks and we conclude the chapter giving a proof for the convergence rate of the solution of the Fokker-Planck equation to the solution of the Chemical Master equation.

In chapter 5 we introduce hybrid models. These models combine two models to a new model that utilises the numerical advantages of the two sub models. Also we will show an error bound for a hybrid model combined from the Chemical Master equation and the Liouville equation. Further, we discuss an extension of this hybrid model obtained by combining the Chemical Master equation with the Fokker-Planck equation in chapter 6. And we will show an error bound for the probability density function of this hybrid model in comparison to the solution of the CME.

In chapter 7, these two hybrid models are applied to a gene-regulatory network composed of 10 species, the famous lac operon, and compared with an approximated solution of the Chemical Master equation. Finally, we will conclude this work with a summary of the results and an outlook on future research topics.

2 | Mathematical Models for Reaction Networks

In this chapter we will define and motivate the usage of reaction networks. Section 2.1 will shortly review the physical background of the models before we define the mathematical setup of reaction networks in section 2.2.

In section 2.3 we will introduce three different models for reaction networks that can be found in literature. We will start with a Markov Jump process and the Chemical Master equation that transports the density of this process (section 2.3.1). Then, we will describe two models that approximate the first model: the Chemical Langevin equation and its distribution transported by the Fokker-Planck equation (section 2.3.2) and the Reaction Rate equation and its distribution transported by the Liouville equation (section 2.3.3).

After the definition of these three models, we discuss their connections in section 2.4, as they can be found in recent publications. Furthermore, we present how the two other models can be derived from the Markov jump process and the Chemical Master equation. But we will also see that some very questionable approximations are needed to do these derivations. These approximations will be motivated in chapter 4.

2.1 Physical Motivation

A reaction network is a qualitative description of the interactions between several (chemical) substances. An example is the reaction of hydrogen with oxygen in the famous "knallgas" (oxyhydrogen) reaction

$$2\ H_2 + O_2 \longrightarrow 2\ H_2O. \tag{2.1}$$

We are now interested in the time evolution of this reaction: How many molecules of H_2O will be in the reaction volume at time t, if we start the reaction with n_1 molecules of H_2 and n_2 molecules of O_2? Or, from a stochastic point of view, how probable is it that we have n_1 molecules of H_2, n_2 molecules of O_2 and n_3 molecules of H_2O at time t?

Although simple, this example allows us to introduce the basic nomenclature of reaction networks. We denote the interacting molecules as species, e.g. H_2, O_2 and H_2O. Further, we denote a single reaction as reaction channel, e.g. we have a network with only one reaction channel. However, most networks in this thesis are constructed from several different reaction channels.

We assume that all reactions take place in a container with fixed volume Ω and constant temperature. The network consists of $N \in \mathbb{N}$ different species $\mathcal{S}_1, \ldots, \mathcal{S}_N$. At the beginning of the experiment or simulation, the system contains a fixed particle number $X(0) \in \mathbb{N}_0^N$ and we are interested in $X(t) \in \mathbb{N}_0^N$, the change of the particle numbers over time.

We are not interested in the position or speed of certain particles or which individual particles interact. We are only interested in the quantities of the different species over time. Therefore it is not necessary to use full atomistic models, like molecular dynamic simulations.

Different authors, like Gillespie or McQuarrie, made several assumptions and modeled reaction networks using a Markov jump process that simulates the quantities of the species without the need to keep track of the positions of all atoms in the reaction volume [Gil76, Gil92, McQ67]. In this model, each particle is seen as a sphere that moves freely in the "well stirred" reaction volume, i.e. the particles are uniformly distributed. The molecules meet at random and react with probability $\gamma(X(t))\Delta t$ in a time interval $[t, t+\Delta t]$ [Gil76, Gil92]. The function $\gamma(X(t))$ is called propensity function and is a measure of the reactivity of the different reaction channels. The function depends on the purpose of the network and can even be time dependent [SMF+13]. Further, the propensity function is proportional to a rate constant c_j, a constant that depends on the chemical properties of the reaction channel.

The propensity function used in this work (cf. eq. (2.6)) arises from the idea that the particles are uniformly distributed in the reaction volume and "meet" at random. This motivates the usage of the binomial coefficient, which describes the drawing of objects out of a well mixed box.

The network description is completed by the stoichiometric vector ϱ_j, which describes the change of particle numbers via a reaction j and the quantities in the system can be updated via $X(t_2) = X(t_1) + \varrho$. In example (2.1) the stoichiometric vector is given by $\varrho_1 = (-2, -1, 2)^T$.

Although we are using a terminology inspired by chemistry, the species can be understood in a very abstract way. For example, they can be proteins inside a cell but also individuals at different stages of a disease.

We also have to understand the difference between particle numbers and concentrations. Particle numbers are natural numbers that keep count of how many particles are present. Concentrations on the other hand are non-negative real numbers that give a relative particle count. Throughout this thesis we will denote particle numbers with $n \in \mathbb{N}_0^N$ or $m \in \mathbb{N}_0^N$ and concentrations with $x \in \mathbb{R}_+^N$ or $y \in \mathbb{R}_+^N$. In chemistry, these quantities are related by the definition of the average concentration

$$x = \frac{n}{\Omega}$$

where $\Omega \in \mathbb{R}_+$ is the reaction volume, Avogadro's constant or the product of both [Gil00, Gil92, BKPR06, Kur72, Wil06, KSM+12, chapter 6.1.2].

A presentation of the thermodynamic principles or a detailed physical derivation of the Markov jump process are out of the scope of this thesis, but we refer the reader to [Gil76, McQ67, BKPR06, Gil92] and [Udr12]. However, the main idea is to assume that

$$c_j \Delta t \text{ is the average probability (to first order in } \Delta t), \quad \text{that the } j\text{th reaction channel will fire in a small time interval } [t, t+\Delta t] \text{ [Gil76].} \tag{2.2}$$

This property is sometimes referred to as the "Fundamental Hypothesis of Chemical Kinetics" [Gil76].

It is further assumed that the particles behave like hard spheres which move in a "well stirred" reaction volume, i.e. they are uniformly distributed. As Gillespie showed, it is possible to estimate the probability that two particles collide, by using these assumptions and thermodynamic arguments [Gil76, Gil92].

Historically, the interaction of species through reaction channels was modeled using ordinary differential equations (ODEs), but increasing awareness in life sciences rendered these type of models as inadequate for networks containing very small particle numbers or for models that have stochastic effects [McQ67]. The reason is that small particle numbers lead to highly stochastic behaviour, which can not be captured by a deterministic ODE. Later on, several other types of reaction networks were discovered, where a deterministic approach is not suitable, like cell fate decision networks which must be modeled using bi- or even multimodal distributions [SYSY02, MML09]. Other examples for highly stochastic networks can be found in neurology, where the toggling of ion channels between an open and closed state turned out to be stochastic [DKB10]. Although the Markov jump process was historically not the first model for reaction networks, it turned out to be the most general one.

2.2 Reaction Networks

A reaction network contains several components that describe the interaction of $N \in \mathbb{N}$ species $\mathcal{S}_i$. The interaction of these species is accomplished through $R \in \mathbb{N}$ reaction channels. Each reaction channel $\mathfrak{R}_j$, $j \in \{1, \cdots, R\}$ consists of substrate species that react to form product species:

$$\mathfrak{R}_j : \sum_{i=1}^{N} \chi_{j,i}^{\text{in}} \mathcal{S}_i \xrightarrow{c_j} \sum_{i=1}^{N} \chi_{j,i}^{\text{out}} \mathcal{S}_i. \tag{2.3}$$

The stoichiometric factors $\chi_{j,i}^{\text{in}} \in \mathbb{N}_0$ and $\chi_{j,i}^{\text{out}} \in \mathbb{N}_0$, with $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, define the number of particles for a species $\mathcal{S}_i$ that go in and come out of a reaction channel $\mathfrak{R}_j$, respectively. The reaction rate $c_j \in \mathbb{R}_+$, with $\mathbb{R}_+ := \{x \in \mathbb{R} \colon x \geq 0\}$ is a constant that describes the reactivity of channel j.

The definition of the network is completed by the stoichiometric vectors,

$$\varrho_j = \begin{pmatrix} \chi_{j,1}^{\text{out}} - \chi_{j,1}^{\text{in}} \\ \vdots \\ \chi_{j,N}^{\text{out}} - \chi_{j,N}^{\text{in}} \end{pmatrix},\ \varrho_j \in \mathbb{Z}^N, \tag{2.4}$$

that characterise the quantitative change in the particle number via reaction channels, and by the propensity functions

$$\gamma_j(n) : \mathbb{N}_0^N \to \mathbb{R}_+. \tag{2.5}$$

We can find different definitions for the propensity functions in literature [Wil06, ch. 6.3]. However, in this work we use a very common propensity function that is motivated by thermodynamic observations [Gil76, Kur72]:

$$\gamma_j(n) = c_j \prod_{i=1}^{N} \binom{n_i}{\chi_{j,i}^{\text{in}}} \tag{2.6}$$

where

$$\binom{n}{k} := \begin{cases} \frac{1}{k!} \prod_{j=0}^{k-1}(n-j) & \text{if } n > k-1 \text{ and } n \in \mathbb{N}_0,\ k \in \mathbb{N}_0 \\ 0 & \text{if } n \leq k-1 \text{ and } n \in \mathbb{N}_0,\ k \in \mathbb{N}_0 \end{cases} \tag{2.7}$$

is the binomial coefficient.

The variable $n = (n_1, \cdots, n_N)^T,\ n_j \in \mathbb{N}_0$ is interpreted as the particle number of the different species.

Some models for reaction networks use a real state variable $x = (x_1, \cdots, x_N)^T,\ x_j \in \mathbb{R}_+$. These models are not using the propensity function (2.6), but the slightly different propensity function

$$\tilde{\gamma}_j(x) : \mathbb{R}_{0,+}^N \to \mathbb{R}_+, \tag{2.8}$$
$$\tilde{\gamma}_j(x) = c_j \prod_{i=1}^{N} \frac{x_i^{\chi_{j,i}^{\text{in}}}}{\chi_{j,i}^{\text{in}}!}.$$

We will refer to the state variable $n \in \mathbb{N}_0^N$ as the discrete (state) variable and to the propensity function $\gamma_j(n)$ as the discrete propensity. Analogously, $x \in \mathbb{R}_+^N$ will be called the continuous (state) variable and $\tilde{\gamma}_j(x)$ the continuous propensity.

The difference between these two quantities will be an important question in this work and will be discussed in section 4.1.

Before we conclude this section, we define an important subtype of reactions.

Definition 2.1 (Monomolecular Reaction Channels)**.**
Reaction channels of the types

$$\begin{array}{lrcll} \mathfrak{R}_{jk}: & \mathcal{S}_j & \xrightarrow{c_{jk}} & \mathcal{S}_k & \text{conversion} \\ \mathfrak{R}_{0k}: & \star & \xrightarrow{c_{0k}} & \mathcal{S}_k & \text{inflow} \\ \mathfrak{R}_{j0}: & \mathcal{S}_j & \xrightarrow{c_{j0}} & \emptyset & \text{outflow} \end{array}$$

are called monomolecular reactions [JH07].

After describing a complex biochemical process as a reaction network, the experimenter is usually interested in the evolution of the species over time. We will discuss several models for simulating the time evolution in section 2.3, after defining some mathematical principles in the next section.

2.3 Mathematical Models for Reaction Networks

2.3.1 Kurtz Process and Chemical Master Equation

We discuss now a first model (and the physically most accurate model) for the reaction network (2.3). We mentioned already that many reaction networks are highly stochastic [McQ67, SYSY02, MML09, DKB10]. Therefore this model is a continuous time Markov jump process on the state space $\mathbb{N}_0^N$. The state space results from the need to represent the particle numbers of N different species. We assume that a discrete number of particles moves freely in a container with fixed volume and constant temperature. The particles are "well stirred", i.e. their positions are uniformly distributed at any time. We further assume that all particles belong to exactly one of the N different species and particles of the same species are undistinguishable [JK12]. We already mentioned that we are not interested in the positions of the particles. Only the number of particles of each species $X(t) \in \mathbb{N}^N$ is modeled.

We discussed already that the state of a reaction network can be updated, if we know that a reaction $\mathfrak{R}_j$ has fired between two time points $t_1 < t_2$:

$$X(t_2) = X(t_1) + \varrho_j.$$

We generalise this and denote with the counting process $F_j(t)$ the number of times a reaction channel $\mathfrak{R}_j$ has fired in the time interval $[0, t]$. It follows that the state of a reaction network at time t is given by

$$X(t) = X(0) + \sum_{j=1}^{R} F_j(t)\varrho_j.$$

Gillespie discussed in detail that the probability that the reaction $\mathfrak{R}_j$ fires in a time interval $[t, t+\Delta t]$ is given by $\gamma_j(X(t))\Delta t + \mathcal{O}(\Delta t^2)$ (cf. [Gil76, Gil92]). Kurtz et al. and Anderson

et al. pointed out [BKPR06, AK12, AH12],[EK05, ch. 6.4] that independent Poisson processes are suitable to model

$$F_j(t) = \mathcal{P}_j\left(\int_0^t \gamma_j\left(X(s)\right)\mathrm{d}s\right)$$

and we refer to this process as:

Definition 2.2 (Kurtz process)**.**
The time continuous Markov jump process

$$X(t) = X(0) + \sum_{j=1}^{R} \mathcal{P}_j\left(\int_0^t \gamma_j\left(X(s)\right)\mathrm{d}s\right)\varrho_j$$

models the state of the reaction network (2.3) [AH12],[EK05, ch. 6.4]. As defined in section 2.1 and 2.2, $X(t) \in \mathbb{N}_0^N$ denotes the state of the network at time t, $X(0)$ denotes the initial condition, R is the number of reaction channels, $\mathcal{P}_j$ are independent Poisson processes, γ_j denotes the propensity function (2.6) and ϱ_j the stoichiometric vector (2.4).

This process has not been named in the literature so far. However, referring to it as "the time continuous Markov jump process that models the reaction network" is tedious. Therefore we name this process the "Kurtz process". This naming is consistent with Sunkara [Sun12] and was chosen to pay tribute to Thomas G. Kurtz, who was one of the first that analysed these kind of processes [Kur72, Kur78, EK05, BKPR06]*.

A general closed formula for the distribution of the Kurtz process is not known†, but we can derive an equation of motion that gives the time evolution of the corresponding probability distribution. This equation is known as the Chemical Master equation (CME).

Lemma 2.3 (Chemical Master Equation)**.**
Let $(\mathfrak{S}, \mathfrak{A}, \mathbb{P})$ be a probability space and let $X(t) : \mathfrak{S} \to \mathbb{N}_0^N$ be the Kurtz process as defined in def 2.2, then the probability density function (PDF) $p(t,n) = \mathbb{P}\left(X(t) = n\right)$ is given by the solution of the Chemical Master equation (CME)

$$\begin{aligned} p(t,n) :&\mathbb{R}_+ \times \mathbb{N}_0^N \to [0,1] \\ \partial_t p(t,n) &= \sum_{j=1}^{R} \gamma_j(n-\varrho_j)p(t,n-\varrho_j) - \gamma_j(n)p(t,n) \qquad (2.9) \\ p(0,n) = p_0 &= \delta_{X(0)}(n) = \begin{cases} 1, & \text{if } X(0) = n, \\ 0, & \text{otherwise,} \end{cases} \end{aligned}$$

with γ_j the propensity function defined by eq. (2.6) and $p(t, n-\varrho_j) = 0$ for $n - \varrho_j \notin \mathbb{N}_0^N$.

*Also we like to make sure that the process is not confused with the "Gillespie Algorithm" [Gil76].

†In lemma 2.4, we will state such a formula for monomolecular networks. However many important networks are more complicated and for these networks no closed formula exists.

Proof. By following the derivations made by Gillespie or Higham, we estimate the probability to reach a state $X(t+\Delta t)$ assuming state $X(t)$ is given [Gil76, Gil92, Hig08]. We denote with $\mathfrak{M}(t+\Delta t, t)$ the number of reaction events that occurred in the time interval $[t, t+\Delta t]$.

Gillespie and Higham distinguish three cases:

(i) The two states are already equal. We denote the probability of this event by $\mathbb{P}[\mathfrak{M}(t+\Delta t, t) = 0]$.

(ii) The probability that the state can be reached after one reaction has fired is denoted by $\mathbb{P}[\mathfrak{M}(t+\Delta t, t) = 1]$.

(iii) And finally, we denote the probability that the state $X(t+\Delta t)$ is reached after more than one reaction event with $\mathbb{P}[\mathfrak{M}(t+\Delta t, t) > 1]$.

We follow Gillespie and assume that the probability for more than one reaction in the time interval $[t, t+\Delta t]$ is of order $\mathcal{O}(\Delta t^2)$ [Gil76, Gil92].

We estimate the probability that one reaction fires in the time interval $[t, t+\Delta t]$ with $\sum_{j=1}^{R} \gamma_j(X(t))\,\Delta t + \mathcal{O}(\Delta t^2)$ (cf. section 2.1 and beginning of section 2.3.1) and the probability that no reaction fires is $1 - \sum_{j=1}^{R} \gamma_j(X(t))\,\Delta t + \mathcal{O}(\Delta t^2)$. A general and rigorous derivation of these transition probabilities can be found in [CM77, ch. 5.6] or [Sun12, appendix B].

In summary, the transition probabilities are estimated by

$$\begin{aligned}
&\textbf{(i)} \quad \mathbb{P}[\mathfrak{M}(t+\Delta t, t) = 0] \approx 1 - \sum_{j=1}^{R} \gamma_j(X(t))\,\Delta t + \mathcal{O}(\Delta t^2),\\
&\textbf{(ii)} \quad \mathbb{P}[\mathfrak{M}(t+\Delta t, t) = 1] \approx \sum_{j=1}^{R} \gamma_j(X(t))\,\Delta t + \mathcal{O}(\Delta t^2),\\
&\textbf{(iii)} \quad \mathbb{P}[\mathfrak{M}(t+\Delta t, t) > 1] \approx \mathcal{O}(\Delta t^2).
\end{aligned}$$

These are combined to form the transition probability $T_{\Delta t}\left(n^{(2)} \mid n^{(1)}\right)$ for going from state $X(t) = n^{(1)}$ to state $X(t+\Delta t) = n^{(2)}$ in a fixed time interval $[t, t+\Delta t]$ with fixed $n^{(1)}, n^{(2)} \in \mathbb{N}_0^N$:

$$\begin{aligned}
T_{\Delta t}\left(n^{(2)} \mid n^{(1)}\right) := &\underbrace{\left(1 - \sum_{j=1}^{R} \gamma_j\left(n^{(1)}\right)\Delta t\right)}_{\text{(i)}} \delta_{n^{(2)}, n^{(1)}} \\
&+ \underbrace{\sum_{j=1}^{R} \gamma_j\left(n^{(1)}\right)\Delta t\, \delta_{n^{(2)}, n^{(1)}+\varrho_j}}_{\text{(ii)}} + \underbrace{\mathcal{O}(\Delta t^2)}_{\text{(iii)}},
\end{aligned} \tag{2.10}$$

where δ is the Kronecker delta. We insert this into the Chapman-Kolmogorov equation (CKE) (cf. [Gar04])

$$T_{t+\Delta t}(n^{(2)}|n^{(0)}) = \sum_{n^{(1)}\in\mathbb{N}_0^N} T_{\Delta t}\left(n^{(2)}|n^{(1)}\right) T_t\left(n^{(1)}|n^{(0)}\right)$$

with $n^{(i)} \in \mathbb{N}_0^N,\ i \in \{0,1,2\}$ and $X(0) = n^{(0)}$.

By insertion of the transition probability ((2.10)) into the CKE we obtain

$$\begin{aligned} T_{t+\Delta t}(n^{(2)}|n^{(0)}) = \sum_{n^{(1)}\in\mathbb{N}_0^N} \Bigg[& \left(1 - \sum_{j=1}^{R} \gamma_j\left(n^{(1)}\right)\Delta t\right) \delta_{n^{(2)},n^{(1)}} \\ & + \sum_{j=1}^{R} \gamma_j\left(n^{(1)}\right)\Delta t \delta_{n^{(2)},n^{(1)}+\varrho_j} + \mathcal{O}(\Delta t^2)\Bigg] T_t\left(n^{(1)}|n^{(0)}\right) \\ = T_t\left(n^{(2)}|n^{(0)}\right) & - \sum_{j=1}^{R} \gamma_j\left(n^{(2)}\right)\Delta t T_t\left(n^{(2)}|n^{(0)}\right) \\ & + \sum_{j=1}^{R} \gamma_j\left(n^{(2)}-\varrho_j\right)\Delta t T_t\left(n^{(2)}-\varrho_j|n^{(0)}\right) + \mathcal{O}(\Delta t^2)\,. \end{aligned}$$

This can be rearranged by subtracting $T_t\left(n^{(2)}|n^{(0)}\right)$ and dividing by Δt:

$$\begin{aligned} \frac{T_{t+\Delta t}(n^{(2)}|n^{(0)}) - T_t\left(n^{(2)}|n^{(0)}\right)}{\Delta t} = & \sum_{j=1}^{R} \gamma_j\left(n^{(2)}-\varrho_j\right) T_t\left(n^{(2)}-\varrho_j|n^{(0)}\right) \\ & - \sum_{j=1}^{R} \gamma_j\left(n^{(2)}\right) T_t\left(n^{(2)}|n^{(0)}\right) + \mathcal{O}(\Delta t)\,. \end{aligned}$$

Now we take the limit $\Delta t \to \infty$ and obtain

$$\partial_t T_t\left(n^{(2)}|n^{(0)}\right) = \sum_{j=1}^{R} \gamma_j\left(n^{(2)}-\varrho_j\right) T_t\left(n^{(2)}-\varrho_j|n^{(0)}\right) - \gamma_j\left(n^{(2)}\right) T_t\left(n^{(2)}|n^{(0)}\right),$$

and by renaming $T_t\left(n^{(2)}|n^{(0)}\right)$ to $p(t,n)$ follows the assertion. □

In 2007, Jahnke and Huisinga showed that the CME can be solved for monomolecular reaction networks by a convolution of a product Poisson distribution with multinomial distributions [JH07].

Lemma 2.4 (Solution of the CME for monomolecuar reaction networks)**.**
The probability distribution for a monomolecular reaction network (cf. def. 2.1) (i.e. the solution of the CME for this network) at time $t > 0$ is given by

$$p(t,\cdot) = \mathcal{P}\left(\cdot,\lambda(t)\right) * \mathcal{M}\left(\cdot,\hat{n}_1,p^{(1)}(t)\right) * \cdots * \mathcal{M}\left(\cdot,\hat{n}_N,p^{(N)}(t)\right),$$

for a initial value $p(0,\cdot) = \delta_{\hat{n}}(\cdot),\ \hat{n} \in \mathbb{N}_0^N$. *The function* $\mathcal{P}(\cdot,\cdot)$ *denotes a product Poisson distribution (cf. def. A1) and* $\mathcal{M}(\cdot,\cdot,\cdot)$ *denotes a multinomial distribution (cf. def. A3). The vectors* $p^{(i)}(t) \in [0,1]^N$ *and* $\lambda(t) \in \mathbb{R}^N$ *are the solutions of*

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}p^{(i)}(t) &= Ap^{(i)}(t), & p^{(i)}(0) &= \varepsilon_i, \\ \frac{\mathrm{d}}{\mathrm{d}t}\lambda(t) &= A\lambda(t) + b, & \lambda(0) &= (0,\ldots,0)^T, \end{aligned}$$

with

$$\begin{aligned} A &= (a_{jk})_{j,k=0,\ldots,N}\,, \\ a_{jk} &= c_{kj}\ j \neq k \geq 1, \\ a_{kk} &= -\sum_{j=0}^{N} c_{kj}, \\ b &= (c_{01},\ldots,c_{0N})^T\,, \end{aligned}$$

and ε_i *the i-th column of the identity matrix. The constants* c_{kj} *are the reaction rates of the monomolecular network 2.1. The convolution of two PDFs* $P_1, P_2 : \mathbb{N}^N \to \mathbb{R}_+$ *is defined as*

$$(P_1 * P_2)(n) = \sum_z P_1(z)P_2(n-z) = \sum_z P_1(n-z)P_2(z)$$

with $z \in \mathbb{N}_0^N$ *such that* $(n-z) \in \mathbb{N}_0^N$.

Remark: This statement holds also for time dependent reaction rates.

Proof. We refer the reader to the complete and detailed proof in the original work by Jahnke and Huisinga [JH07] and denote only the basic idea of the proof. First of all, the authors observed that the time evolution of each molecule in the reaction volume is independent of the other molecules. This allowed them to sort the set of all molecules into $N+1$ different subsets, one for each species and the inflow reaction. The authors further proved two properties of this network, namely that the solution of a reaction system stays a Poisson distribution if the initial value was already a product Poisson distribution and that the solution of a reaction systems stays multinomial if the initial value was a multinomial distribution and no inflow reactions are present. Then, they used these properties to analyze the time evolution of the $N+1$ different random variables that denote the number of particles in the corresponding subsets and could show the assertion. □

We have seen that the CME operates on the state space $\mathbb{N}_0^N$. Numerically, this domain is often approximated by a finite state space $\mathbb{S} \subset \mathbb{N}_0^N$ on a bounded time interval. However, the size of $\mathbb{S}$ still depends exponentially on the dimension N. This connection of the problem size with the dimension is called the "*curse of dimensionality*" in literature. Many well known problems like the multi-dimensional Black-Scholes equation from financial math-

ematics, the Schrödinger equation from quantum physics or the Fokker-Planck equation suffer from this "curse" as well.

2.3.2 Chemical Langevin Equation and Fokker-Planck Equation

Definition 2.5 (Stochastic Differential Equation (SDE))**.**
The equation

$$\begin{aligned} \mathrm{d}X(t) &= a(X(t))\,\mathrm{d}t + b(X(t))\mathrm{d}W(t) \\ X(0) &= x_0, \end{aligned} \tag{2.11}$$

with $X(t) \in \mathbb{R}^N$, $a(X(t)) = (a_1(X(t), \ldots, a_d(X(t))) \in \mathbb{R}^N$, $b(X(t)) = (b_{ij}(X(t))_{i,j} \in \mathbb{R}^{N\times M}$ and $N, M \in \mathbb{N}$, is called a N-dimensional Stochastic Differential Equation (SDE) [KP99, ch. 3.3], [Øks07, ch. 5] and should be seen as an abbreviate way to write the Itō-integral equation

$$X(t) = X(0) + \int_0^t a(X(s))\,\mathrm{d}s + \int_0^t b(X(s))\mathrm{d}W(s). \tag{2.12}$$

In this section and all following ones, $W(t)$ denotes the (M-dimensional) Standard Wiener Process.

Lemma 2.6 (Fokker-Planck equation)**.**
Let $a(x) \in \mathcal{C}^1(\mathbb{R}^N)$, $b(x) \in \mathcal{C}^2(\mathbb{R}^N)$, $x \mapsto p(t,x) \in \mathcal{C}^2(\mathbb{R}^N)$ and $t \mapsto p(t,x) \in \mathcal{C}^1(\mathbb{R}_+)$ have continuous partial derivatives in space and time and let $a(x)$ and $b(x)$ and their first derivatives be bounded.

If $X(t)$ is the solution of the SDE (2.11)*, then the probability density function (PDF) $p(t,x)$, defined by*

$$\int_S p(t,x)\,\mathrm{d}x = \mathbb{P}\left(X(t) \in S\right),$$

for all measurable sets $S \subseteq \mathbb{R}^N$, solves the Fokker-Planck equation (FPE)

$$\begin{aligned} \partial_t p(t,x) &= -\sum_{i=1}^{N} \partial_{x_i}\left[a(x)p(t,x)\right] + \frac{1}{2}\sum_{i,j=1}^{N} \partial_{x_i}\partial_{x_j}\left[B_{ij}(x)p(t,x)\right] \\ p(0,x) &= p_0 \end{aligned}$$

with $B_{ij}(x) = \sum_{k=1}^{M} b_{ik}(x)b_{jk}(x)$ and initial value p_0.

Proof. Cf. [LM94, ch. 11.6] or [Gar04, ch. 4.3.4]. □

Because a further discussion of stochastic calculus is out of scope of this work, we refer the reader to the standard literature on this topic [LM94, Øks07, KP99, Gar04].

However, we note that the SDE (2.11) can be solved numerically using the Euler-Maruyama method

$$\hat{X}_{i+1} = \hat{X}_i + a(\hat{X}_i) \cdot \tau + b(\hat{X}_i)\Delta W_i, \tag{2.13}$$

which approximates the SDE at the time points $t_i = i \cdot \tau \in [0, T]$ with $i = 1, \cdots, n$, $\tau = \frac{T}{n}$ and $\Delta W_i = W_{t_{i+1}} - W_{t_i}$ [Hig01, Hig11].

Definition 2.7 (Chemical Langevin Equation)**.**
The SDE

$$\begin{aligned} \mathrm{d}x(t) &= \sum_{j=1}^{R} \varrho_j \tilde{\gamma}_j\left(x(t)\right)\ \mathrm{d}t + \sum_{j=1}^{R} \varrho_j \sqrt{\tilde{\gamma}_j\left(x(t)\right)}\ \mathrm{d}W_j(t), \\ x(0) &= x_0. \end{aligned}$$

is called the Chemical Langevin equation (CLE) in the literature [Gil00]. The CLE models the time evolution of reaction network (2.3) using a continuous state variable $x \in \mathbb{R}^N$ and the continuous propensity function (2.8).

Remark: We will see the connection between the CLE and the Kurtz process and how the CLE models a reaction network in the sections 2.4 and 4.

Lemma 2.8 (Fokker-Planck Equation)**.**
If the solution $x(t)$ of the CLE (cf. def 2.7) satisfies the assumptions of lemma 2.6, then the PDF $q(t,x)$, defined by $\int_S q(t,x)\,\mathrm{d}x = \mathbb{P}\left(x(t) \in S\right)$, for all measurable sets $S \subseteq \mathbb{R}^N$, is given by the solution of

$$\begin{aligned} q(t,x) :&\mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}_+, \\ \partial_t q(t,x) &= -\sum_{j=1}^{R} \nabla\left(\tilde{\gamma}_j(x) q(t,x)\right)^T \varrho_j + \frac{1}{2}\sum_{j=1}^{R} \varrho_j^T \nabla^2\left(\tilde{\gamma}_j(x) q(t,x)\right) \varrho_j, \\ q(0,x) &= q_0(x), \end{aligned} \tag{2.14}$$

with $\nabla\Big(f(x)\Big) := \left(\frac{\partial f(x)}{\partial x_i}\right)_{i=1,\ldots,N}$ the gradient and $\nabla^2\Big(f(x)\Big) := \left(\frac{\partial^2 f(x)}{\partial x_i \partial x_j}\right)_{i,j=1,\ldots,N}$ the Hessian matrix. This lemma is a special case of lemma 2.6. However, we will always refer to eq. (2.14) *if we speak of the Fokker-Planck equation (FPE) [Gil00].*

Proof. The assertion follows directly from lemma 2.6. □

In principle, we would expect that the initial condition q_0 of the FPE would be defined by

$$q(0,x) = q_0(x) = \delta_{x(0)}(x) = \begin{cases} 1, & \text{if } x(0) = x, \\ 0, & \text{otherwise.} \end{cases}$$

However, this raises the question if the solution $q(t,x)$ exists in such a case and how the first and second derivative of a delta peak can be interpreted. To avoid these questions we assume that $q_0 : \mathbb{R}_+^N \to \mathbb{R}_+$ is a function with the properties

- $q_0(x) = 0$ if $\|x - x(0)\|_\infty > \varepsilon$,
- $\int_{\mathbb{R}_+^N} q_0(x) = 1$,
- and $\mathbb{E}[q_0] = \int_{\mathbb{R}_+^N} x q_0(x) = x(0)$,

for a small constant $0 < \varepsilon \ll 1$.

Further, we note that the multi-dimensional state space $\mathbb{R}^N$ is no longer limited to positive states. However, negative particle numbers or concentrations have no physical interpretation. This means that trajectories that reach negative states should be avoided. In numerical simulations, these "negative" trajectories often occur if the average solution of the CLE is near zero. Often, the problem can be avoided on bounded time intervals $[0, t]$ if all state variables are $x_i(0) \gg 0$. Higham and Gillespie demand these "large particle numbers" as an initial condition for the CLE [Hig08, Gil00]. We will not formulate such a strict assumption, but we note that any CLE solution that reaches a negative state in any dimension or any FPE solution with non-zero probability for negative states on a bounded time interval, cannot be used to model a reaction network. In these cases, other models (like the Kurtz process and the CME) should be used.

2.3.3 Reaction Rate Equation and Liouville Equation

Definition 2.9 (Reaction Rate Equation)**.**
The ODE

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} x(t) &= \sum_{j=1}^{R} \tilde{\gamma}_j \left(x(t)\right) \varrho_j, \\ x(0) &= x_0. \end{aligned}$$

is called the Reaction Rate equation (RRE). The RRE models the time evolution of reaction network (2.3) using a continuous state variable $x \in \mathbb{R}_+^N$ and the continuous propensity function (2.8).

The RRE is heavily used in past and present literature [Wil06, ch. 6]. Because it is an ODE it is much easier to solve, simulate or analyse than the Kurtz process or the CLE. For simple networks it is possible to find closed analytical solution formulas. For complicated networks, we can make use of the rich spectrum of numerical ODE solvers. We can analyse the state space of the RRE or derive steady states. But we should always keep in mind that the RRE is only an approximation of the Kurtz process that has lost all stochasticity.

We have to clarify why we state a PDF for an ODE. Historically, the Liouville Equation (LVE) was used in statistical mechanics to simulate the evolution of an ensemble of different states over time. This can be translated to the reaction network framework. We assume that different initial states $x^{(i)}(0)$, $i = 1, 2, \ldots$ with given probabilities $\mathbb{P}\left(x^{(i)}(0)\right)$ are of interest. We define the PDF

$$u_0(x) = \mathbb{P}\left(x^{(i)}(0) = x\right)$$

and use this function as initial condition for the LVE. Now, the solution $u(t,x)$ of the LVE describes the evolution of $u_0(x)$.

In the reaction network literature, especially in the more applied context, the LVE is not used very often. However, we will see a very useful application of the LVE in chapter 5.

Lemma 2.10 (Liouville Equation)**.**
If we interpret the RRE as a SDE with zero stochasticity and if $x(t)$ is the solution of the RRE (cf. def 2.9) and its PDF $u(t,x)$ satisfies the assumptions of lemma 2.6, then it holds

$$\begin{aligned} u(t,x)&\colon \mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}_+, \\ \partial_t u(t,x) &= -\sum_{j=1}^{R} \nabla\Big(\tilde{\gamma}_j\left(x(t)\right) u(t,x)\Big)^T \varrho_j, \\ u(0,x) &= u_0(x). \end{aligned}$$

We call this equation the Liouville equation (LVE).

Proof. We interpret the RRE as a SDE with $b(x) \equiv 0$, then lemma 2.6 proves the assertion. □

2.4 Connections between the Different Models

As we have seen in the previous section, there exist several different models in the literature that model the time evolution of a reaction network. In total we have the models*:

$$\begin{aligned} \text{Kurtz process} &\longleftrightarrow \text{Chemical Master equation} \\ \text{Chemical Langevin equation} &\longleftrightarrow \text{Fokker-Planck equation} \\ \text{Reaction Rate equation} &\longleftrightarrow \text{Liouville equation} \end{aligned}$$

In this section we review *connections* between the models and see which approximations are needed to derive these connections. We will not discuss the motivation and legitimacy of these approximations in this chapter. But the goal of the following chapters is to introduce methods to analyse and rigorously prove the errors introduced by different approximations. Now we investigate the connection between the models found in the literature so far. To do this we have to slightly abuse notation from time to time and use disputable approximations. However, their motivation will become clearer in the subsequent chapters.

We have already seen how the CME, the FPE and the LVE can be derived from the Kurtz process, the CLE and the RRE, respectively. We discuss now the connection between the Kurtz process and the CLE and the RRE and how the CME can be approximated to derive the FPE and the LVE.

*We denote with $\longleftrightarrow$ that the stochastic procces and the ODE (on the left side) have PDFs (right side) and that these models should be seen as an unit, even if they have different equations.

2.4.1 Derivation of the CLE and the RRE from the Kurtz process

We motivate the CLE as an approximation of the Kurtz process by following a scheme that Gillespie introduced in 2000 [Gil00]. Similar derivations can be found in several other publications. The following derivations are not rigorous.

Gillespie started with the Kurtz process (def. 2.2) and approximated it on bounded time intervals $[t, t+\tau]$ [Gil00, Mac09]:

$$X(t+\tau) \approx \hat{n}_0 + \sum_{j=1}^{R} \mathcal{P}_j\left(\tau\gamma_j\left(X(t)\right)\right)\varrho_j,$$

where $\hat{n}_0$ is the state of the process at the beginning of the time interval.

Motivated by the law of large numbers, he approximated the Poisson process $\mathcal{P}_j(\xi)$ with a normal distributed stochastic variable $\mathcal{N}_j\left(\mu, \sigma^2\right)$ with expectation $\mu = \xi$ and variance $\sigma^2 = \xi$ [Mac09, Gil00, BTB04]. This yields a reasonable approximation if the expectation of the Poisson process $\mathbb{E}\left[\mathcal{P}_j(\xi)\right] = \xi \gg 1$ [Gil00]. Gillespie also used that

$$\mathcal{N}_j\left(\mu, \sigma^2\right) = \mu + \sigma\mathcal{N}_j(0,1), \tag{2.15}$$

which leads to

$$\mathcal{P}_j(\xi) \approx \xi + \sqrt{\xi}\mathcal{N}_j(0,1).$$

We should be aware that this is an abuse of notation because the Poisson process is operating on a discrete state space and the Itō process on a continuous one.

It follows that

$$X(t+\tau) \approx \hat{n}_0 + \sum_{j=1}^{R} \tau\gamma_j\left(X(t)\right)\varrho_j + \sum_{j=1}^{R} \sqrt{\tau\gamma_j\left(X(t)\right)}\varrho_j\mathcal{N}_j(0,1). \tag{2.16}$$

Simultaneously he changed the state vector from being discrete to being continuous. This is done in two steps, first Gillespie approximated the discrete state variable $X \in \mathbb{N}_0^N$ by a continuous state $\mathbb{R}_+^N \ni \hat{x}(t) \approx X(t)$ and second he replaced the discrete propensity function γ by its continuous counterpart $\tilde{\gamma}_j$ (cf. eq. (2.6) and (2.8)). Unfortunately this step is not that easy. For example, the quadratic propensity functions

$$\gamma_j(n) = c_j\binom{n_k}{2} = \frac{c_j}{2}(n_k^2 - n_k), \qquad \tilde{\gamma}_j(x) = c_j\frac{x_k^2}{2!} = \frac{c_j}{2}x_k^2$$

result in an error (if we evaluate them both at the same continuous state $x_k \in \mathbb{R}_+$)

$$\left|\gamma_j(x) - \tilde{\gamma}_j(x)\right| = \frac{c_j}{2}\left|x_k^2 - x_k - x_k^2\right| = \frac{c_j}{2}x_k.$$

Gillespie tried to motivate this replacement of propensity functions by introducing a scaling factor. Unfortunately, his approach is very vague and holds only up to linear propensity

functions. We will introduce a similar scaling in the next chapter and will show the resulting error rigorously.

For now, we follow the scheme of Gillespie and obtain

$$\hat{x}(t+\tau) = \hat{x}_0 + \sum_{j=1}^{R} \tau\tilde{\gamma}_j\left(\hat{x}(t)\right)\varrho_j + \sum_{j=1}^{R} \sqrt{\tau\tilde{\gamma}_j\left(\hat{x}(t)\right)}\varrho_j\mathcal{N}_j\left(0,1\right). \tag{2.17}$$

By taking the limit $\tau \to \infty$ (in principle, eq. (2.17) is a Euler-Maruyama discretization) and as discussed in [BMP96, Gil00, BTB04, Hig08, Gar04] and [Mac09] we obtain the Chemical Langevin equation (CLE):

$$\begin{aligned} \mathrm{d}x(t) &= \sum_{j=1}^{R} \varrho_j\tilde{\gamma}_j\left(x(t)\right)\ \mathrm{d}t + \sum_{j=1}^{R} \varrho_j\sqrt{\tilde{\gamma}_j\left(x(t)\right)}\ \mathrm{d}W_j(t), \\ x(0) &= x_0. \end{aligned}$$

We have seen that we need several approximations to derive the CLE. Therefore it is obvious that the CLE can only be an approximation of the Kurtz process.

We have seen how the CLE can be derived from the Kurtz process by approximating the discrete Poisson process by a continuous Itō process. Now, we approximate the Kurtz process to derive the Reaction Rate equation (RRE).

This is usually done by taking the expectation of the Kurtz process (def. 2.2), assuming that the expectation of the propensity function approximates the propensity function of the expectation ($\mathbb{E}\left[\gamma_j\left(X(s)\right)\right] \approx \gamma_j\left(\mathbb{E}\left[X(s)\right]\right)$), replacing the propensity functions on the discrete state space by the ones on a continuous state and defining $\hat{x}(t) := \mathbb{E}\left[X(t)\right]$. This leads to the approximation:

$$\hat{x}(t) \approx \hat{x}_0 + \sum_{j=1}^{R} \int_0^t \tilde{\gamma}_j\left(\hat{x}(s)\right)\mathrm{d}s\varrho_j$$

with initial condition $\hat{x}_0 = \hat{x}(0)$.

This is nothing else than the integral form of the RRE

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}x(t) &= \sum_{j=1}^{R} \tilde{\gamma}_j\left(x(t)\right)\varrho_j, \\ x(0) &= x_0. \end{aligned}$$

2.4.2 Derivation of the FPE and the LVE from the CME

We have seen in the last section that the Kurtz process can be approximated by estimating the Poisson process with an Itō process (CLE), or by approximating its expectation (RRE). We have also seen that the PDFs of the two processes are given by the FPE and the LVE, respectively.

In this section, we see that these equations can also be derived as an approximation of the CME. We present the basic ideas of the approximations and see how the equations and models are connected, but we will not discuss if these approximations converge or if they are valid. In chapter 4, we will discuss the concept of the thermodynamical limit and show how these models can be derived in this limit. This will help us to understand the error behaviour of the different approximations.

The following steps were stated by Gillespie in 2000 [Gil00] and he assumed that the parameter $n \in \mathbb{N}_0^N$ of the CME can be approximated by the continuous state variable $x \in \mathbb{R}_+^N$ and replaced the discrete propensity function by the continuous one (cf. chapter 2.4.1). Further, he assumed that

$$\Big[x \mapsto \tilde{\gamma}_j(x)\hat{q}(\cdot, x)\Big] \in \mathcal{C}^\infty,$$

where $\hat{q}(t,x) : \mathbb{R}_+ \times \mathbb{R}_+^N \to \mathbb{R}_+$ is the approximation of the solution of the CME.

Then, Gillespie derived a Taylor expansion in space,

$$\tilde{\gamma}_j(x-\varrho_j)\hat{q}(t,x-\varrho_j) = \tilde{\gamma}_j(x)\hat{q}(t,x) - \nabla\Big(\tilde{\gamma}_j(x)\hat{q}(t,x)\Big)^T\varrho_j + \frac{1}{2}\varrho_j^T\nabla^2\Big(\tilde{\gamma}_j(x)\hat{q}(t,x)\Big)\varrho_j + \dots.$$

Insertion of the Taylor expansion and the above approximations into the CME yields:

Definition 2.11 (The Kramers-Moyal Expansion)**.**
The equation

$$\partial_t\hat{q}(t,x) = \sum_{j=1}^{R}\left(-\nabla\Big(\tilde{\gamma}_j(x)\hat{q}(t,x)\Big)^T\varrho_j + \frac{1}{2}\varrho_j^T\nabla^2\Big(\tilde{\gamma}_j(x)\hat{q}(t,x)\Big)\varrho_j + \sum_{|\boldsymbol{k}|\geq 0}\frac{1}{\boldsymbol{k}!}\varrho_j^{\boldsymbol{k}}\nabla^{\boldsymbol{k}}\left(\tilde{\gamma}_j(x)\hat{q}(t,x)\right)\right).$$

is called the Kramers-Moyal expansion [Kra40, Moy49, Gil00, Gar04], with $\boldsymbol{k}$ a multi-index (cf. def. B4). It is based on a Taylor expansion of the difference term in the CME (2.9).

Truncating the higher order terms of the Kramers-Moyal expansion up to the first derivative results in the Liouville equation:

$$\partial_t u(t,x) = -\sum_{j=1}^{R}\nabla\Big(\tilde{\gamma}_j(x)u(t,x)\Big)^T\varrho_j. \tag{2.18}$$

Truncating the higher order terms up to the second derivative results in the Fokker-Planck equation:

$$\partial_t q(t,x) = \sum_{j=1}^{R}\left(-\nabla\Big(\tilde{\gamma}_j(x)q(t,x)\Big)^T\varrho_j + \frac{1}{2}\varrho_j^T\nabla^2\Big(\tilde{\gamma}_j(x)q(t,x)\Big)\varrho_j\right).$$

2.5 Summary

We mentioned already that some of the steps we did in this chapter are highly disputable. Why can we replace the discrete state variable $n \in \mathbb{N}_0^N$ by a continuous variable $x \in \mathbb{R}_+^N$? Why can we replace the propensity function γ_j by $\tilde{\gamma}_j$? Why can we truncate the Kramers-Moyal expansion?

A way to answer these questions (or to give at least some meaning to the procedure) is to define a scaling based on the idea that the concentration is connected with the particle number by

$$x \approx \frac{n}{\Omega}.$$

Based on this observation we will define a set of scalings and transformations and then we will take the limit $n, \Omega \to \infty$. This ansatz is known as the "Thermodynamic Limit" in the literature (cf. [Kur72, Kur78, Gil00, BKPR06, JK12]).

This leads to a set of useful tools that allows the derivation of the Kramers-Moyal expansion in a scaled version. These scaled equations help to answer the question how the models on continuous state spaces approximate the models on discrete state spaces and motivate the truncation of the Kramers-Moyal expansion.

We conclude this chapter by summarizing the connection of the different models. Figure 2.1 displays this connection. The figure lists the processes and shows the basic idea that led to the approximations. To avoid confusion, we point out explicitly that only the downward facing arrows denote approximations, the horizontal arrows denote which PDE (on the right) transports the PDF of the process on the left.

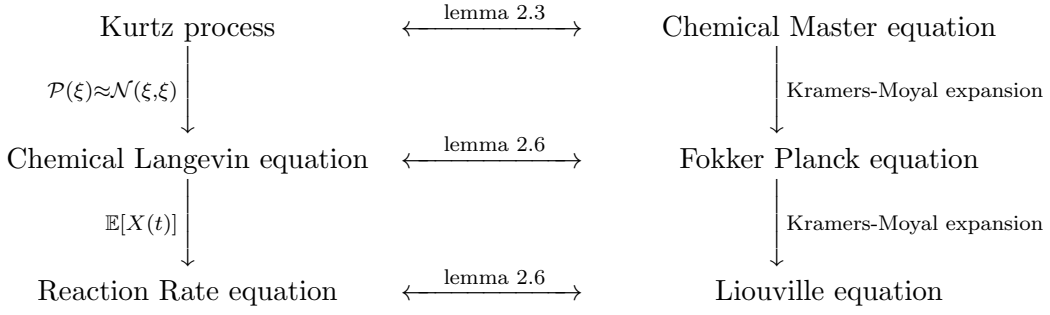


Figure 2.1: Relation between the different models. The column on the left lists the different stochastic processes. The Kurtz process is approximated by the CLE by approximating the Poisson processes. The CLE is approximated by the RRE by taking the expectation. The horizontal arrows list the lemmata that connect the processes in the left column and the PDFs in the right column. The downward facing arrows in the right column symbolise the approximation of the CME by the FPE and the LVE using the Kramers-Moyal expansion.

3 | Numerical Methods

In this chapter we will discuss two numerical *methods* for simulating the Kurtz process and for solving the CME, respectively. The first methods is the famous Stochastic Simulation Algorithm, which generates trajectories of the Kurtz process. We will motivate this algorithm in section 3.1 by reviewing Gillespies seminal work from 1976 [Gil76]. In section 3.2 we will state a method which approximates the solution of the CME, the Finite State Projection Method. This method was originally published by Munsky and Khammash in 2006 and recently improved by Sunkara and Hegland [MK06, Sun12].

Because this thesis is about the analysis of *models*, we will not discuss any further methods, but we will name a few important ones in section 3.3.

This chapter is concluded with three examples to gain a feeling how reaction networks look like, how the different models can be applied and how the corresponding solutions behave.

3.1 Stochastic Simulation Algorithm

In section 2.3.1, we introduced the Kurtz process (def. 2.2) as a model for the time evolution of reaction networks (cf. section 2.2). The value of a stochastic process over time is called trajectory, realisation, sample or path of a process. It is possible to calculate such paths for a given Kurtz process. Several paths can be used to approximate characteristics of the associated PDF. For example, the mean of several trajectories can be used to approximate the expectation. It is also possible to use a histogram of realisations as an approximation of the PDF. This means that many samples can be used to approximate the solution of the CME.

In 1976 Daniel T. Gillespie published a seminal paper, where he derived an algorithm to simulate reaction networks: the Stochastic Simulation Algorithm (SSA) [Gil76].

To derive the SSA, using assumption 2.2 and the CME (2.9), we define a so-called *reaction probability density function*.

Definition 3.1 (Reaction Probability Density Function)**.**

$P(\tau, \mu)\mathrm{d}\tau$ is the probability at time t that the next reaction
will take place in the time interval $(t + \tau, t + \tau + \mathrm{d}\tau)$
and that the firing reaction channel will be $\mathfrak{R}_\mu$.

Further we define

$\gamma_\mu(n)\mathrm{d}t$ is the probability up to $\mathcal{O}(\mathrm{d}t^2)$ that $\mathfrak{R}_\mu$ will fire in the next time interval $\mathrm{d}t$.

We calculate the probability, that $\mathfrak{R}_\mu$ will fire in $[t + \tau, t + \tau + \mathrm{d}t]$,

$$P(\tau, \mu)\mathrm{d}\tau = P_0(\tau) \cdot \gamma_\mu \mathrm{d}\tau \tag{3.1}$$

using the probability $P_0(\tau)$ that no reaction will occur in $[t, t+\tau]$. We state the probability that more than one reaction occurs in the time interval with $\mathcal{O}(\mathrm{d}t^2)$.

$P_0(\tau)$ is derived by dividing the interval $[t, t + \tau]$ into K small subintervals, multiplying the probabilities for reactions occurring in these subintervals, and taking the converse probability (cf. proof of lemma 2.3)

$$P_0(\tau) = \lim_{K\to\infty} \left(1 - \frac{1}{K}\sum_{j=1}^{R} \gamma_j \tau + \mathcal{O}\left(\frac{1}{K^2}\right)\right)^K = \exp\left(-\sum_{j=1}^{R} \gamma_j \tau\right). \tag{3.2}$$

So we get the reaction probability density function

$$P(\tau, \mu) = \gamma_\mu \exp\left(-\sum_{j=1}^{R} \gamma_j \tau\right). \tag{3.3}$$

A longer and more detailed derivation of equation (3.3) can be found in [Gil76]. It is also shown therein that $P(\tau, \mu)$ is a PDF with mass 1.

Now, we simulate a reaction network by using a random number generator to create values for τ and μ according to equation (3.3) [Gil76]. The actual time t has to be extended by $t + \tau$ to find the beginning of the next interval and the actual state of the network is changed by executing reaction channel $\mathfrak{R}_\mu$, i.e. by adding ϱ_μ to the actual state vector of the network. This approach is similar to the derivations of the Kurtz process and of the CME in chapter 2.3.1. First we find the next time interval with a reaction event, then we find the index of the reaction that fired. Algorithm 3.1 gives a possible implementation of SSA.

However, this approach is still too slow for the needs of many experimenters that are interested in analysing a large amount of trajectories for large reaction networks. Numerical experiments show that networks with large particle numbers or reaction channels with high fire frequencies often lead to slow SSA simulations. This results from an (expensive) execution of the while loop in algorithm 3.1 for each reaction event. Therefore, we briefly review some ideas published in recent years of ways to obtain faster simulations

Algorithm 3.1 Stochastic Simulation Algorithm (SSA): direct method

Require: initial condition n_0, time interval $[t_0, t_{\text{final}}]$, stoichiometric vectors $\varrho_1, \ldots, \varrho_N$, propensity functions $\gamma_1(n), \ldots, \gamma_N(n)$

$t \leftarrow t_0$
$n \leftarrow n_0$
while $t < t_{\text{final}}$ **do**
  $a \leftarrow \sum_{j=1}^{N} \gamma_j(n)$
  $r_1, r_2 \leftarrow \texttt{rand\_uniform}(0, 1)$
  $\tau \leftarrow \frac{1}{a} \log\left(\frac{1}{r_1}\right)$
  $t \leftarrow t + \tau$
  choose μ such that $\sum_{j=1}^{\mu-1} \gamma_j(n) < a r_2 \sum_{j=\mu}^{N} \gamma_j(n)$
  $n \leftarrow n + \varrho_\mu$
end while
return n

of the Kurtz process. Gibson and Bruck suggested a more efficient way to implement the algorithm called the "Next Reaction Method" which needs only one random number per reaction (due to a re-usage of random numbers), but this algorithm is often still too slow [GB00].

Gillespie introduced a new simulation scheme called τ-leaping [Gil01]. The basic idea is to predefine an equidistant time grid with step size τ and to estimate the number of reactions fired in a predefined time interval $[t, t + \tau]$, instead of calculating the time interval to the next reaction event. The method consists of two parts. First, Gillespie approximates the integral in the Kurtz process by

$$X(t + \Delta t) \approx \hat{X}(t + \Delta t) = \hat{X}(t) + \sum_{j=1}^{R} \mathcal{P}_j\left(\gamma_j\left(\hat{X}(t)\right) \Delta t\right) \varrho_j.$$

Second, the new process is sampled to create realisations of the stochastic reaction network.

The advantage of this method is that very frequent reactions are not simulated by many small time steps with only one reaction in between. They are combined to a single time step. A disadvantage of this method is the possibility to obtain negative particle numbers, if the chosen time step is to big. Some other authors proposed further "leaping" methods that are all based on similar ideas but introduce different approximations [TB04, BMT06, MLB07].

3.2 The Finite State Projection Method

As we have seen in the previous chapter, a reaction network can either be analysed using the Kurtz process or by inspecting the CME. A realisation of the process has a computational advantage, since it is quite easy to calculate paths of the process. However, we have also seen that we need several of these paths to approximate the moments or the underlying PDF. A different approach is to solve the CME and to use this solution to calculate the desired PDF directly (and then compute its moments, if needed). Unfortunately, no general

analytic solution* for the CME is known so far. A way to circumnavigate this problem is to consider the CME on a finite state space $\mathbb{S}$ only. We learned that the CME usually has an infinite state space, but by considering the problem numerically we assume that the CME starts with a finite inital condition that "grows" or "moves" around. However, this truncation of the state space makes it necessary to define proper boundary conditions in space. The approximation that the CME has a finite state space, i.e.

$$|\mathbb{S}| < \infty, \tag{3.4}$$

transforms the CME problem into:

Definition 3.2 (Finite CME)**.**
From assumption (3.4) follows the finite CME

$$\begin{aligned}
&\hat{p}(t,n) : \mathbb{R}_+ \times \mathbb{S} \to [0,1],\\
&\partial_t \hat{p}(t,n) = \left(\mathcal{A}_{\mathbb{S}}\hat{p}\right)(t,n) = \sum_{j=1}^{R} \gamma(n-\varrho_j)\hat{p}(t,n-\varrho_j) - \gamma(n)\hat{p}(t,n),\\
&\hat{p}(t,n) = 0, \qquad \forall n \notin \mathbb{S},\\
&\hat{p}(0,n) = \hat{p}_0(n),
\end{aligned}$$

with $\hat{p}_0(n)$ the original initial condition p_0 on the restricted domain $\mathbb{S}$.

This reduces the size of the problem and turns the infinite differential equation system into a finite systems of ODEs with unique solution [GY13]

$$\hat{p}(t,n) = e^{\mathcal{A}_{\mathbb{S}}\cdot t}\hat{p}_0(n),$$

with the operator $e^{\mathcal{A}_{\mathbb{S}}\cdot t}$ defined via the power series [LM94, ch. 7.8][EN00, chapter I, def. 2.2]

$$e^{\mathcal{A}_{\mathbb{S}}\cdot t} = \sum_{k=0}^{\infty} \frac{t^k}{k!}\mathcal{A}_{\mathbb{S}}^k.$$

We have seen in chapter 2.3.1 that the CME is suffering from the "*curse of dimensionality*" and so even the reduced state space $\mathbb{S}$ could be too large to solve. However, reducing the CME operator is the simplest known method to approximate solutions of the CME numerically. But finding $\mathbb{S}$ is a challenging task itself. A very trivial way could be to run several SSA simulations to estimate the size of the state space, but running these SSA simulations could be a time-consuming task by itself. Furthermore, such a method would not offer any possibility of error control [WGMH10].

Munsky and Khammash proposed the Finite State Projection (FSP) method [MK06] to estimate the truncated state space $\mathbb{S}$ of the CME solution. They have shown that, under

*Besides the one for monomolecular reaction networks given in lemma 2.4. However, most biological interesting networks are not monomolecular

certain conditions [MK06, Sun12], for any $\varepsilon > 0$ a reduced state space $\mathbb{S}$ can be found in a way that the approximated solution $\hat{p}_\varepsilon(t, n)$ on $\mathbb{S}$ has the properties

$$0 \leq \hat{p}(t, n) \leq \hat{p}_\varepsilon(t, n) + \varepsilon, \ \forall n \in \mathbb{S}, \ \forall t \in [0, t_{\text{end}}] .$$

The idea of the FSP can be summarised in the following steps, for each time step. The goal is to find a state space $\mathbb{S}_j$ such that the error of the solution of the CME on this state space $\hat{p}_j(t, n)$ is below a threshold $\varepsilon > 0$. In the following $\hat{\mathcal{A}}_j$ denotes the CME operator $\mathcal{A}_\mathbb{S}$ restricted to the state space $\mathbb{S}_j$. First we have to check if the error of the approximation is below a threshold value ε.

The calculation of this error is not obivous, because the solution of the CME is not known. However, Munsky and Khammash showed that this error can be estimated by introducing a so-called "sink state". This state "collects" the probability that flows out of the truncated state space. They showed that the absolute error of the actual iteration is bounded by the probability in this sink state (cf. [MK06] and [Sun12, ch. 2.3]). This sink state replaces the Dirichlet boundary conditions of the finite CME problem defined in def. 3.2.

If the error is too large, then we have to heuristically grow the actual state space $\mathbb{S}_j$ (for example by using the *N-step reachability* algorithm proposed by Munsky and Khammash [MK06]) to obtain a new space $\mathbb{S}_{j+1}$. Now the solution $\hat{p}_{j+1}$ has to be derived. This step is repeated until the error is below ε and it was shown that if $\mathbb{S}_j \subset \mathbb{S}_{j+1}$, then the error is monotonically decreasing [MK06]. But because the state space is only growing, it can occur that the new space $\mathbb{S}_{j+1}$ is (especially after several time steps) larger than needed to obtain a solution of error ε.

Sunkara and Hegland proposed an Optimal Finite State Projection (OFSP) [SH10, Sun12] that estimates the optimal size of $\mathbb{S}_j$. To achieve this, the FSP method is used to find a subspace $\mathbb{S}_j$ on which $\hat{p}_j$ has error $\frac{\varepsilon}{2}$. Next, this space is reduced to the new space $\mathbb{S}_{j+1}$ by using an algorithm called N-term approximation. The idea of this algorithm is to sort the values of the corresponding probabilities $\hat{p}_j$ in descending order on the natural numbers. Now the state space is truncated by removing corresponding probabilities until they accumulate to mass $\frac{\varepsilon}{2}$. Sunkara and Hegland showed in their work that this method finds the best (i.e. the smallest) state space with the solution having error ε.

The OFSP method is implemented in the software package `cmepy` [HFC10] and was used in this work to calculate several numerical examples.

3.3 Further Methods

We reviewed one numerical method for simulating the Kurtz process and one for solving the CME. Of course there is a wider spectrum of numerical methods for the Kurtz process or the CME. However, this thesis is about the connections, approximations and convergence errors of different *models*, so the analysis of numerical *methods* is out of scope of this work.

However, for general references, we list a few further methods here to given an overview of the wide spectrum of numerical methods for reaction networks:

- Burrage et al. and Wolf et al. discussed further methods for state space truncation [BHMS06, WGMH10].
- MacNamara et al. discussed the application of quasi-steady-state assumptions and splitting methods for the CME and the application of Krylov and splitting methods on the FSP algorithm [MBBS08].
- Engblom used a spectral approximation to solve the CME [Eng09].
- Jahnke and Udrescu used wavelets to solve the CME [Jah10, JU10, Udr12].
- Hegland et al. discussed the application of sparse grids to solve the CME [HHL08].
- Dolgov and Khoromskij, Hegland and Garcke, Jahnke and Huisinga and Kazeev et al. discussed tensor product approximations for reaction networks [DK12, HG11, JH08, KKNS12, KS13].
- Anderson and Higham extended the multilevel Monte Carlo approach developed by Giles to reaction networks [Gil08b, Gil08a, AH12].
- Deuflhard et al. discussed adaptive Galerkin methods for the CME [DHJW08].
- Several software tools exist that allow the modeling and simulation of reaction networks [HSG+06, BCMP11, LCPG08, WHK10].
- Several different hybrid methods [HR02, Jah11, MLSH12] will be discussed in chapter 5 and 6.

We will also not discuss numerical methods for solving the other models given in this work. We mentioned already the Euler-Maruyama scheme to solve the CLE, but there are further methods for solving SDEs numerically [KP99]. The FPE can be solved using a method of lines approach with difference quotients for the first and second derivative and a suitable ODE solver for the resulting problem. Moreover, there is a large spectrum of numerical methods for solving differential equations that include also methods that we can apply to the RRE and the LVE.

3.4 Examples

We conclude this chapter by stating three example networks. This allows us to see how the definitions given so far can be applied and offers an insight how the resulting solutions look like.

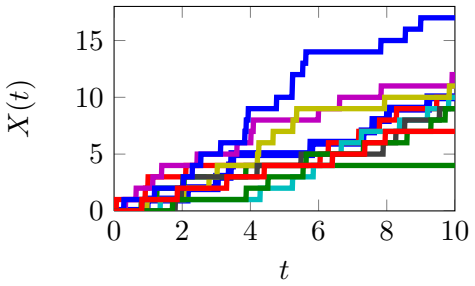
3.4.1 Example 1: An Inflow Reaction

First we examine a simple inflow reaction with a constant rate c_1

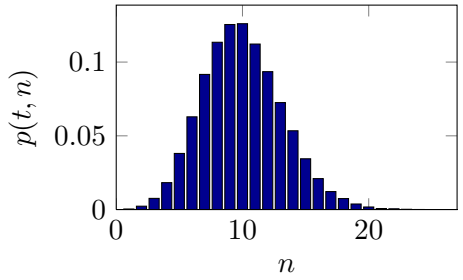
$$\mathfrak{R}_1 : \star \xrightarrow{c_1} \mathcal{S}_1.$$

The stoichiometry of this network is given by

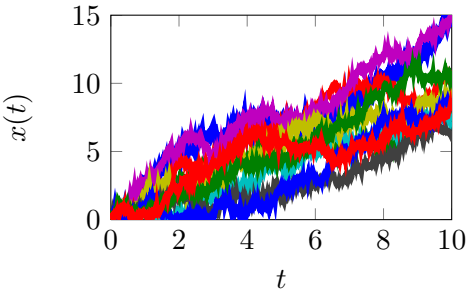
$$\varrho_1 = \chi_{1,1}^{\text{out}} - \chi_{1,1}^{\text{in}} = 1 - 0 = 1$$



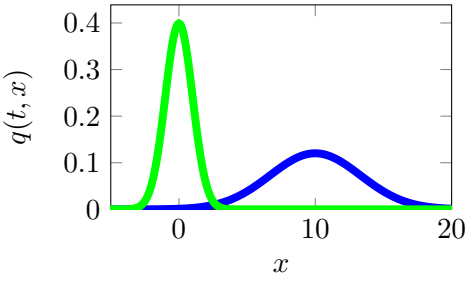
(a) Kurtz process



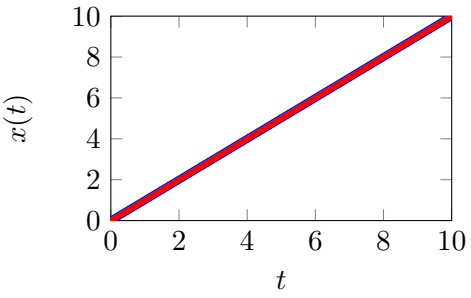
(b) CME solution at $t = 10$



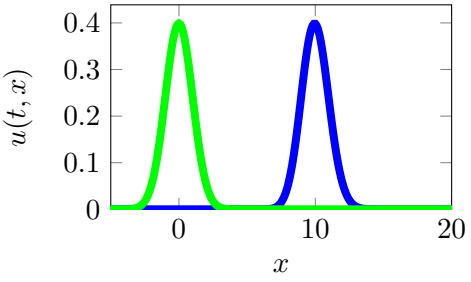
(c) CLE solution



(d) FPE solution at $t = 0$ (green) and $t = 10$ (blue)



(e) RRE solution



(f) LVE solution at $t = 0$ (green) and $t = 10$ (blue)

Figure 3.1: Solutions of the six different models for reaction networks for example 1, with $c_1 = 1$ and $\varepsilon = 1$.

and the propensity function by

$$\gamma_1(n) = c_1.$$

The Kurtz process for this reaction is

$$X(t) = n_0 + \mathcal{P}_1\left(\int_0^t c_1 \, \mathrm{d}s\right)(+1),$$
$$n_0 = 0$$

As one would expect, the random (Poisson distributed) jumps depend only on the rate constant and we observe that $X(t)$ only grows. We see this behaviour in figure 3.1a. The figure shows ten trajectories of the process and as one would expect, they evolve differently but their values are only growing over time.

The CME for this example is given by

$$\begin{aligned}\partial_t p(t,n) &= c_1 p(t,n-1) - c_1 p(t,n),\\ p(0,n) = \delta_0(n) &= \begin{cases} 1, & \text{if } n = 0,\\ 0, & \text{otherwise,}\end{cases}\end{aligned}$$

and figure 3.1b displays the solution at time $t = 10$. We see a discrete distribution with expectation 10.

The CLE is given by

$$\begin{aligned}\mathrm{d}x(t) &= c_1 \,\mathrm{d}t + \sqrt{c_1}\,\mathrm{d}W(t),\\ x(0) &= 0\end{aligned}$$

and the FPE by

$$\begin{aligned}\partial_t q(t,x) &= -c_1 \partial_x q(t,x) + \frac{c_1}{2}\partial_x^2 q(t,x),\\ q(0,x) &= \mathcal{N}(0,\varepsilon).\end{aligned}$$

Figure 3.1c and figure 3.1d show the solutions of these two models, respectively. In general the CLE solution looks similar to the Kurtz process, but we see the important difference between a discrete and a continuous model. We see the discrete jumps in the Kurtz process, while the pattern in the CLE solution is dominated by the noisy behaviour of the Wiener process. The FPE solution cannot be compared directly with the CME solution. We noted already in chapter 2.3.2 that we cannot use the same initial conditions for the CME and the FPE. The CME solution in figure 3.1b was started with a delta peak at $n = 0$. However, the FPE in figure 3.1d was started with a Gaussian initial PDF with expectation 0 and variance $\varepsilon = 1$. This reveals a further problem. Using a Gaussian centered around the initial condition of the CME leads here to an initial distribution that assigns a probability greater than zero for negative particle numbers. We discussed already that such a distribution has no biological interpretation. However, we discuss this example to call attention to these kind of problems. We will further discuss the comparability of CME and FPE in chapter 4. This initial condition is plotted in green, while the solution of the FPE at time $t = 10$ is plotted in blue. We note that the distribution moves to the right (inflow) and has a expectation of $\mathbb{E}\left[q(t,x)\right] = 10$. It is similar to the CME solution, but we have to keep in mind that the solution of the CME is a PDF on a discrete state space, while the solution of the FPE has a continuous state space. The FPE was solved using a method of lines approach, combining difference quotients in space with the Matlab® `ode45` solver in time.

The RRE is given by

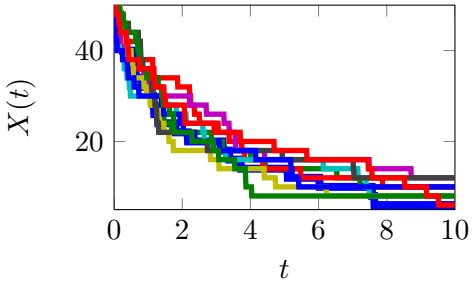
$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x(t) &= c_1\\ x(0) &= 0\end{aligned}$$

and the LVE by

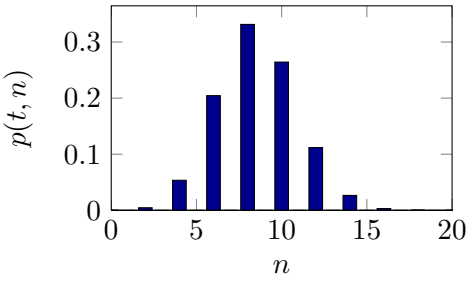
$$\partial_t u(t,x) = -c_1 \partial_x u(t,x),$$
$$u(0,x) = \mathcal{N}(0,\varepsilon).$$

Figure 3.1e and figure 3.1f show the solutions of these two models, respectively. Now we observe that all stochasticity is lost and all 10 solutions of the RRE are equal and follow the same deterministic trajectory. We expected this behaviour, because the RRE is an ODE. The LVE was started with the same initial condition as the FPE, a Gaussian with expectation 0 and variance $\varepsilon = 1$. This allows us to understand the effect of the second derivative term in the FPE. The first derivative term transports the solution (in this case to the right), the second derivative term diffuses the solution. In this case the variance increases.

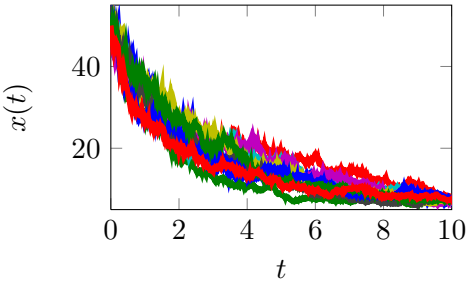
3.4.2 Example 2: A Dimerisation Reaction



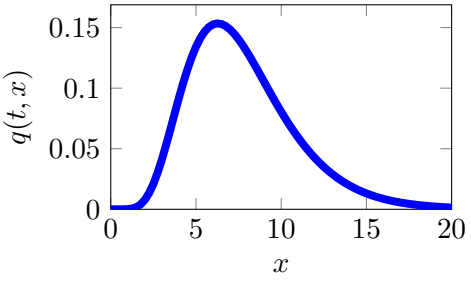
(a) Kurtz process



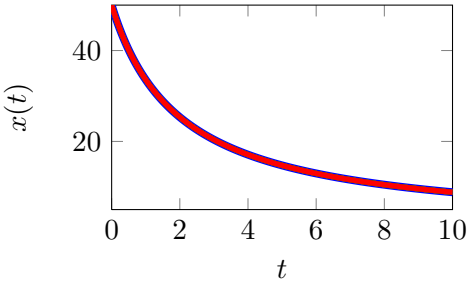
(b) CME solution



(c) CLE solution



(d) FPE solution



(e) RRE solution

Figure 3.2: Solutions of five different models for reaction networks for example 2, with $c_1 = 0.01$.

Our second example also consists of only one reaction but we examine a slightly more complicated propensity function

$$\mathfrak{R}_1 : \mathcal{S}_1 + \mathcal{S}_1 \xrightarrow{c_1} \emptyset.$$

The stoichiometry of this network is given by

$$\varrho_1 = \chi_{1,1}^{\text{out}} - \chi_{1,1}^{\text{in}} = 0 - 2 = -2$$

and the propensity function by

$$\gamma_1(n) = c_1 \binom{n_1}{2} = \frac{c_1}{2} n_1(n_1 - 1).$$

The Kurtz process for this reaction is

$$X(t) = n_0 + \mathcal{P}_1 \left(\int_0^t c_1 \binom{X(s)}{2} \mathrm{d}s \right) (-2) = n_0 + \mathcal{P}_1 \left(\int_0^t \frac{c_1}{2} X(s)(X(s) - 1) \mathrm{d}s \right) (-2),$$
$$n_0 = 50$$

We see that the Poisson process is now depending quadratically on $X(t)$ but the overall trend is decreasing, because of the non-negativity of the Poisson process and the negative stoichiometric vector. Again, the visualisation given in figure 3.2a supports this assertion.

The corresponding CME is given by

$$\partial_t p(t,n) = c_1 \binom{n+2}{2} p(t, n+2) - c_1 \binom{n}{2} p(t,n)$$
$$p(0,n) = \delta_{50}(n) = \begin{cases} 1, & \text{if } n = 50, \\ 0, & \text{otherwise,} \end{cases}$$

and figure 3.2b displays its solution. We can observe a very interesting discrete PDF in this case. Because we start with a delta peak at $n = 50$ and the stoichiometry is -2 we can only reach even numbers and therefore the probability for all odd states is zero.

The CLE is given by

$$\mathrm{d}x(t) = -c_1 (x(t))^2 \, \mathrm{d}t - 2\sqrt{\frac{c_1}{2}} x(t) \, \mathrm{d}W(t),$$
$$x(0) = 50$$

and the FPE by

$$\partial_t q(t,x) = c_1 \partial_x \left(x^2 q(t,x) \right) + c_1 \partial_x^2 \left(x^2 q(t,x) \right),$$
$$q(0,x) = \mathcal{N}(50, 0.1)$$

Figure 3.2c and figure 3.2d show the solutions of these two models, respectively. Again the global tendency of the CLE solution is similar to the Kurtz process, but the noise from the

Wiener process dominates the pattern. The diffusion resulting from the Wiener process or equivalently from the second derivative in the FPE is the reason why every real number in the shown interval has a non zero probability. The FPE is not able to reproduce the pattern observed in the CME: the different probability between odd and even states.

The RRE is given by

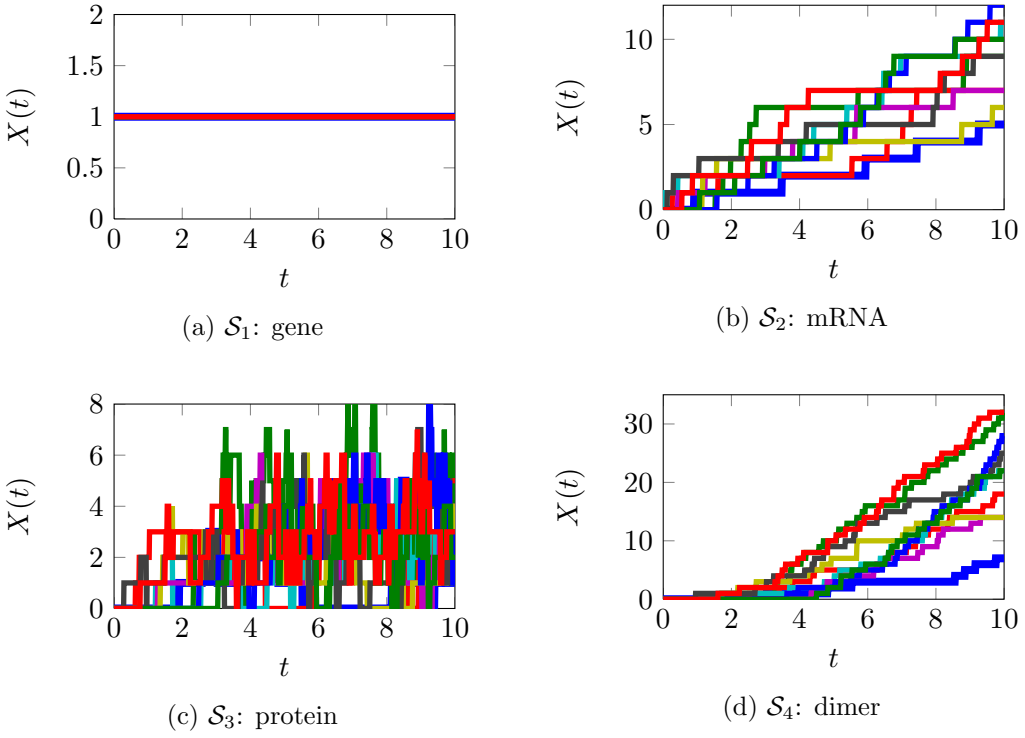
$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}x(t) &= -c_1 x(t)^2, \\ x(0) &= 50 \end{aligned}$$

and the LVE by

$$\begin{aligned} \partial_t u(t,x) &= c_1 \partial_x \left(x^2 u(t,x)\right), \\ u(0,x) &= \mathcal{N}(50, 0.1) \end{aligned}$$

Figure 3.2e shows the solution of the RRE. Again we see a deterministic solution that reflects the trajectory of the expectation of the process.

3.4.3 Example 3: A Protein Network



(a) $\mathcal{S}_1$: gene

(b) $\mathcal{S}_2$: mRNA

(c) $\mathcal{S}_3$: protein

(d) $\mathcal{S}_4$: dimer

Figure 3.3: Several trajectories of the Kurtz processes in examples 3, with $c_1 = c_2 = c_3 = 1$.

The last network models a biological problem in a very simplified way: The generation of a protein through a gene and the dimerisation of two proteins. $\mathcal{S}_1$ denotes a gene that

codes for a protein $\mathcal{S}_3$, $\mathcal{S}_2$ denotes the mRNA generated by the gene and $\mathcal{S}_4$ symbolises the dimer. A network describing this behaviour may look like this:

$$\begin{aligned} \mathfrak{R}_1: &\quad \mathcal{S}_1 &&\xrightarrow{c_1}& \mathcal{S}_1 + \mathcal{S}_2 \\ \mathfrak{R}_2: &\quad \mathcal{S}_2 &&\xrightarrow{c_2}& \mathcal{S}_2 + \mathcal{S}_3 \\ \mathfrak{R}_3: &\quad \mathcal{S}_3 + \mathcal{S}_3 &&\xrightarrow{c_3}& \mathcal{S}_4 \end{aligned}$$

From the above definitions follow the stoichiometric vectors

$$\varrho_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \qquad \varrho_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \qquad \varrho_3 = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}$$

and the propensity functions

$$\gamma_1(n) = c_1 n_1 \qquad \gamma_2(n) = c_2 n_2 \qquad \gamma_3(n) = c_3 \frac{n_3 \cdot (n_3 - 1)}{2}.$$

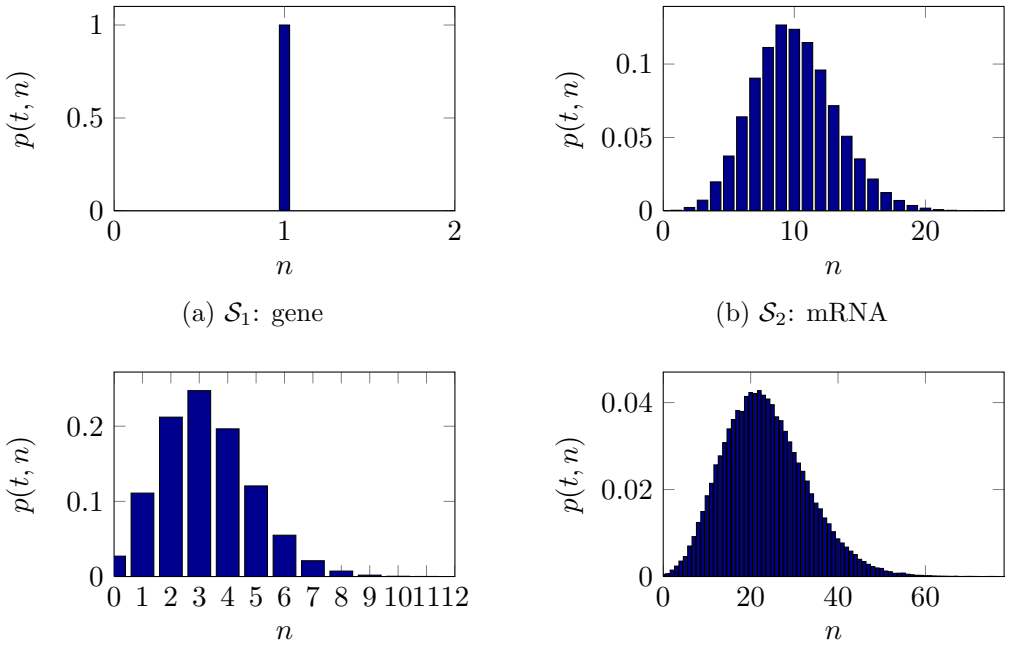
The Markov jump process (2.2) for this network is given by

$$\begin{aligned} X(t) = n_0 &+ \mathcal{P}_1 \left(\int_0^t c_1 X_1(s) \mathrm{d}s \right) \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \mathcal{P}_2 \left(\int_0^t c_2 X_2(s) \mathrm{d}s \right) \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \\ &+ \mathcal{P}_3 \left(\int_0^t c_3 \frac{X_3(s) \cdot (X_3(s) - 1)}{2} \mathrm{d}s \right) \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}, \\ n_0 &= (1, 0, 0, 0)^T, \end{aligned}$$

where $X_i(t)$ is the i-th entry of the vector $X(t)$. We note that not all reactions depend on all species. For example the first process only adds to the second species. Figure 3.3 shows the time evolution of the four different species in a single plot each. The gene is not changing at all. This follows directly from the observation that there are no stoichiometric vectors that change this species. The mRNA is growing over time. This could be changed, by adding a degradation reaction for this species. The protein is created by $\mathfrak{R}_2$ and reduced by $\mathfrak{R}_3$, therefore we see a raising and falling of the population values over time. Last, the dimer is also growing over time. It is the end product of this network and has no degradation reaction.

The CME for the network is given by

$$\begin{aligned} \partial_t p(t, n) = &\, c_1 n_1 p\left(t, n - \varrho_1\right) - c_1 n_1 p(t, n) \\ &+ c_2 n_2 p\left(t, n - \varrho_2\right) - c_2 n_2 p(t, n) \end{aligned}$$



(a) $\mathcal{S}_1$: gene

(b) $\mathcal{S}_2$: mRNA

(c) $\mathcal{S}_3$: protein

(d) $\mathcal{S}_4$: dimer

Figure 3.4: Marginal distributions of the solution of the CME in example 3, with $c_1 = c_2 = c_3 = 1$.

$$+ \frac{c_3}{2}\binom{n_3+2}{2} p\left(t, n-\varrho_3\right) - \frac{c_3}{2}\binom{n_3}{2} p(t,n),$$
$$p(0,n) = \delta_{n_0}(n)$$

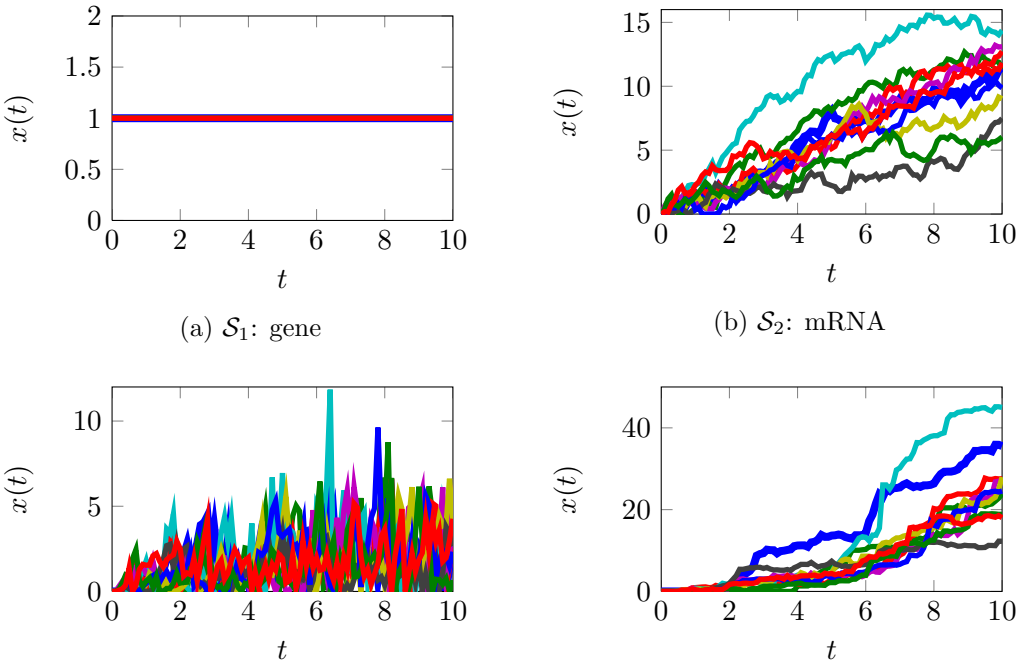
and figure 3.4 displays its solution at time $t = 10$. Because it is not possible to display a 4-dimensional PDF in a 2-dimensional figure we displayed the marginal distributions for each species. As we expected, the PDF for the first species is given by a probability of one for state $n_1 = 1$ and zero for all other states. The other PDFs express what we expect from the Kurtz processes. The CME solution was approximated using a histogram of 10^5 SSA realisations.

The CLE for this network is

$$\begin{aligned} \mathrm{d}x(t) = \varrho_1 &\left[c_1 x_1(t)\,\mathrm{d}t + \sqrt{c_1 x_1(t)}\,\mathrm{d}W_1(t)\right] \\ + \varrho_2 &\left[c_2 x_2(t)\,\mathrm{d}t + \sqrt{c_2 x_2(t)}\,\mathrm{d}W_2(t)\right] \\ + \varrho_3 &\left[\frac{c_3}{2} x_3^2(t)\,\mathrm{d}t + \sqrt{\frac{c_3}{2} x_3^2(t)}\,\mathrm{d}W_3(t)\right], \\ x(0) = &\left(1,0,0,0\right)^T, \end{aligned}$$

and the FPE is

$$\partial_t q(t,x) = -c_1 \nabla \left(x_1(t) q(t,x) \right)^T \varrho_1 + \frac{c_1}{2} \varrho_1^T \nabla^2 \left(x_1(t) q(t,x) \right) \varrho_1$$



(a) $\mathcal{S}_1$: gene

(b) $\mathcal{S}_2$: mRNA

(c) $\mathcal{S}_3$: protein

(d) $\mathcal{S}_4$: dimer

Figure 3.5: Solutions of the CLE in example 3, with $c_1 = c_2 = c_3 = 1$.

$$
\begin{aligned}
&- c_2 \nabla \left(x_2(t) q(t,x) \right)^T \varrho_2 + \frac{c_2}{2} \varrho_2^T \nabla^2 \left(x_2(t) q(t,x) \right) \varrho_2 \\
&- c_3 \nabla \left(\frac{1}{2} x_3^2(t) q(t,x) \right)^T \varrho_3 + \frac{c_3}{2} \varrho_3^T \nabla^2 \left(\frac{1}{2} x_3^2(t) q(t,x) \right) \varrho_3, \\
q(t,0) &= q_0.
\end{aligned}
$$

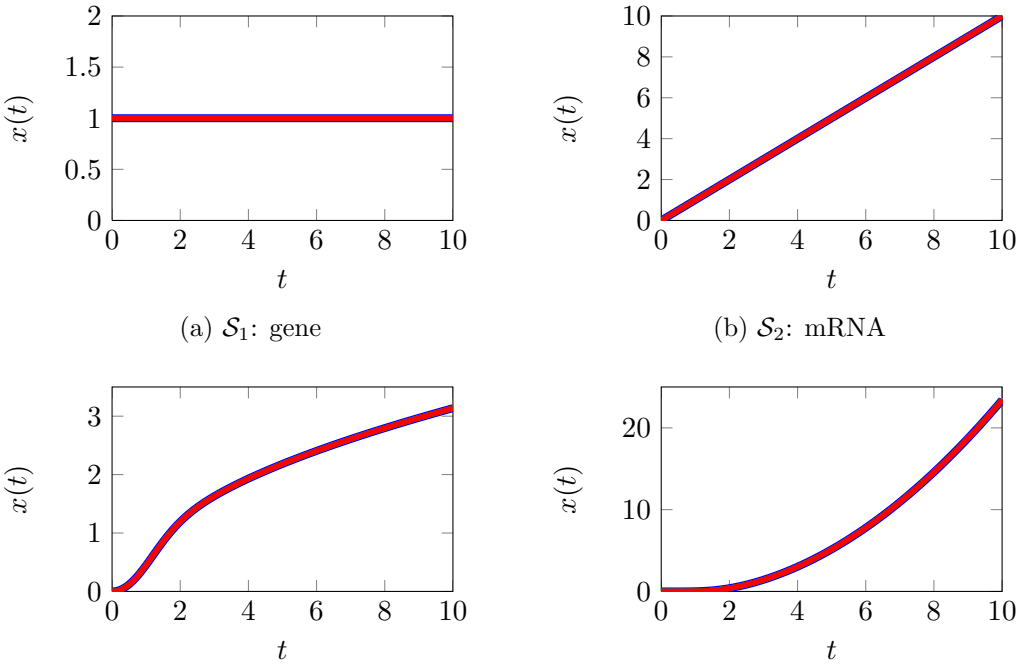
Figure 3.5 shows the solution of CLE. Both the FPE and CLE are no suitable models for this reaction network. The CLE solution was generated using the suggestion to set negative states to zero (cf. [Hig08]). However, these states are reached very often for species $\mathcal{S}_3$ and the generated data has no biological interpretation. Either the state zero would be over represented or the solutions would contain negative values.

The RRE and LVE are given by

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} x(t) &= c_1 \varrho_1 x_1(t) + c_2 \varrho_2 x_2(t) + \frac{c_3}{2} \varrho_3 x_3^2(t), \\
x(0) &= (1,0,0,0)^T,
\end{aligned}
$$

and

$$
\begin{aligned}
\partial_t u(t,x) = &- c_1 \nabla \left(x_1(t) u(t,x) \right)^T \varrho_1 - c_2 \nabla \left(x_2(t) u(t,x) \right)^T \varrho_2 \\
&- c_3 \nabla \left(\frac{1}{2} x_3^2(t) u(t,x) \right)^T \varrho_3, \\
u(t,0) &= u_0.
\end{aligned}
$$



(a) $\mathcal{S}_1$: gene

(b) $\mathcal{S}_2$: mRNA

(c) $\mathcal{S}_3$: protein

(d) $\mathcal{S}_4$: dimer

Figure 3.6: Solutions of the RRE for example 3, with $c_1 = c_2 = c_3 = 1$.

Figure 3.6 shows the solution of the RRE.

4 The Thermodynamic Limit and Convergence of the Fokker-Planck Equation

We concluded chapter 2 with a survey how the three models, based on a Markov jump process, a SDE and an ODE, respectively, are connected via different approximations. Further, we noted that some approximations are very questionable. Especially the connection between discrete and continuous state spaces, the change of propensity functions and the truncation of the Kramers-Moyal expansion was not motivated in the setting of section 2.4. We will discuss these questions in this chapter and motivate the connection of discrete and continuous states in section 4.1. In that section we will also define scaled reaction networks and prove the difference between the two propensity functions. These results will be used in section 4.2 to motivate the truncation of the Kramers-Moyal expansion.

In section 4.3 we will see that the Kurtz process converges to the RRE. This seminal work was originally published by Kurtz in 1972 and demonstrates how scaled reaction networks can be used for the analysis of convergence properties. We will use these techniques to prove the convergence of the FPE to the CME in section 4.4. This proof is concluded with a numerical example (section 4.4.1) and a discussion of possible future research projects (section 4.4.2).

4.1 Scaled Reaction Networks

The last chapter brought up the question, how the different models for reaction networks can be compared. We have seen that the CME and the corresponding Kurtz process are using natural vectors $n \in \mathbb{N}_0^N$ for the state variable modeling the species involved in the reaction network. However, the approximations arising from the Kramers-Moyal expansion – the LVE and the FPE – use real vectors$x \in \mathbb{R}^N$ for the state variables. This results in the problem that we have to compare discrete with continuous state spaces. We do this by physically interpreting x as a concentration

$$x = \frac{n}{\Omega}, \tag{4.1}$$

which gives us the physical connection between particle amounts $n \in \mathbb{N}_0^N$ and concentrations $x \in \mathbb{R}^N$. As we have seen (cf. chapter 2.1), the factor Ω is often interpreted as the volume of the reaction network, but here we interpret it as an abstract scaling factor.

Before we introduce the thermodynamic limit and the resulting scaled networks, we have to clarify some notation. The terms discrete model, state variable, PDF or propensity function denote that the model uses a discrete state variable $n \in \mathbb{N}_0^N$, therefore the term discrete PDF means that we are speaking of a PDF on a discrete state space. Analogously we denote with continuous model, state variable, PDF or propensity function the case where the model uses a continuous state variable $x \in \mathbb{R}^N$, i.e. a continuous PDF is a PDF on a continuous state space.

Thomas Kurtz introduced in 1972 a concept that became well known as the "Thermodynamic Limit" in literature [Kur72, Kur78, BKPR06, Gil00, JK12]. The idea behind this concept is to scale a reaction network and its models by Ω and study these scaled models for

$$X(t) \to \infty, \qquad \Omega \to \infty, \qquad \frac{X(0)}{\Omega} = \text{const.} \tag{4.2}$$

Before we return to the seminal paper of Kurtz later in this chapter, we have to define scaled reaction networks and derive the scaled CME, LVE and FPE and their processes.

Definition 4.1 (Scaled Propensity Functions)**.**
The scaled discrete and continuous propensity functions are defined by

$$\beta_j(m) = c_j \Omega \cdot \Omega^{(-|\chi_j^{\text{in}}|_1)} \prod_{j=1}^{N} \binom{m_i}{\chi_{j,i}^{\text{in}}},$$

$$\tilde{\beta}_j(x) = c_j \Omega \prod_{j=1}^{N} \frac{x_j^{\chi_{j,i}^{\text{in}}}}{\chi_{j,i}^{\text{in}}!},$$

respectively, with $|\chi_j^{\text{in}}|_1 = \sum_{i=1}^{N} \chi_{j,i}^{\text{in}}$. The discrete propensity function β_j is used for natural numbers $m \in \mathbb{N}_0^N$, but the definition of the binomial coefficient in eq. (2.7) would allow to use this function for real numbers, too.

Further, we assume:

Assumption 4.2 (Bound for the Stoichiometric Factor)**.**
We assume that the stoichiometric factor of the substrate species is bounded by

$$\sum_{i=1}^{N} \chi_{j,i}^{\text{in}} \leq 2.$$

Remark: This means that we are mostly interested in four types of reactions:

1. Inflow reactions of type:

$$\chi_{j,i}^{\text{in}} = 0 \; \forall \; i \in [1, \ldots, N] : \quad \mathfrak{R} : \quad \star \quad \xrightarrow{c} \quad \ldots$$

This type occurs for example, when new individuals join a population through birth or immigration or when molecules enter the reaction volume through gating processes.

2. Conversion of a single molecule into something else.

$$\exists k : \chi^{\text{in}}_{j,k} = 1 \text{ and } \chi^{\text{in}}_{j,i} = 0 \; \forall \; i \neq k : \quad \mathfrak{R} : \quad \mathcal{S}_k \quad \xrightarrow{c} \quad \dots$$

This often occurs in death processes, when a molecule decomposes into one or more molecules or in birth processes.

3. Bimolecular reaction:

$$\exists k, l : \chi^{\text{in}}_{j,k} = \chi^{\text{in}}_{j,l} = 1 \text{ and } \chi^{\text{in}}_{j,i} = 0 \; \forall \; i \neq k, l : \quad \mathfrak{R} : \quad \mathcal{S}_k + \mathcal{S}_l \quad \xrightarrow{c} \quad \dots$$

Two different species react with each other. This may be the case when a protein binds to a DNA or an enzyme catalyses a substrate.

4. Dimerisation:

$$\exists k : \chi^{\text{in}}_{j,k} = 2 \text{ and } \chi^{\text{in}}_{j,i} = 0 \; \forall \; i \neq k : \quad \mathfrak{R} : \quad \mathcal{S}_k + \mathcal{S}_k \quad \xrightarrow{c} \quad \dots$$

This is a dimerisation reaction, where two molecules of the same species react with each other. An example is the aggregation of proteins (like ion channels) consisting of several alike subunits.

It is possible to think of reactions of higher order, like the process of translation, where the two subunits of the ribosome, the mRNA and tRNAs combine. Another example is the chain reaction inside an atomic bomb where a critical mass of uranium or plutonium is needed to start the reaction. But these reactions could also be modeled in a network, where the first two particles "meet", form a new intermediary species and then "meet" the next particle and so on [Gil00].

We are now able to derive the difference between the discrete and continuous propensity function.

Lemma 4.3 (Difference between discrete and continuous propensity functions)**.**
If $m \in \mathbb{N}_0^N$ and if there is a constant K such that $m_i \leq K\Omega \; \forall i \in \{1, \dots, N\}$ and if assumption 4.2 holds, then a constant $C \in \mathbb{R}_{>0}$ can be found, such that

$$\left|\beta_j(m) - \tilde{\beta}_j\left(\frac{m}{\Omega}\right)\right| \leq C \qquad \text{for } \max_{i \in [1,\dots,N]} \left(\chi^{in}_{j,i}\right) = 2,$$
$$\left|\beta_j(m) - \tilde{\beta}_j\left(\frac{m}{\Omega}\right)\right| = 0 \qquad \text{otherwise.}$$

Remark: This statement has been similarly proven in [Kur72, JK12].

Proof. We have to distinguish four cases, by assumption 4.2:

$|\chi_j^{\mathbf{in}}|_1 = \mathbf{0}$

Therefore $\chi_{j,i}^{\text{in}} = 0 \; \forall \; i \in \{1, \ldots, N\}$ with the propensity functions

$$\beta_j(m) = c_j \Omega \Omega^0 \prod_{i=1}^{D} \binom{m_i}{0} = c_j \Omega,$$
$$\tilde{\beta}_j\left(\frac{m}{\Omega}\right) = c_j \Omega \prod_{i=1}^{D} \frac{\left(\frac{m_i}{\Omega}\right)^0}{0!} = c_j \Omega,$$

and the error

$$\left|\beta_j(m) - \tilde{\beta}_j\left(\frac{m}{\Omega}\right)\right| = 0.$$

$|\chi_j^{\mathbf{in}}|_1 = \mathbf{1}$

Therefore $\exists k : \chi_{j,k}^{\text{in}} = 1$ and $\chi_{j,i}^{\text{in}} = 0 \; \forall \; i \neq k$. This results in the propensity functions

$$\beta_j(m) = c_j \Omega \Omega^{-1} \binom{m_k}{1} = c_j \Omega \frac{m_k}{\Omega} = c_j m_k,$$
$$\tilde{\beta}_j\left(\frac{m}{\Omega}\right) = c_j \Omega \frac{\left(\frac{m_k}{\Omega}\right)^1}{1!} = c_j \Omega \frac{m_k}{\Omega} = c_j m_k,$$

and the error

$$\left|\beta_j(m) - \tilde{\beta}_j\left(\frac{m}{\Omega}\right)\right| = 0.$$

$|\chi_j^{\mathbf{in}}|_1 = \mathbf{2}$

This can be obtained in two ways:

$\exists k, l : \chi_{j,k}^{\mathbf{in}} = \chi_{j,l}^{\mathbf{in}} = 1$ **and** $\chi_{j,i}^{\mathbf{in}} = 0 \; \forall \; i \neq k, l$

This results in the propensity functions

$$\beta_j(m) = c_j \Omega \Omega^{-2} \binom{m_l}{1} \binom{m_k}{1} = c_j \Omega \frac{m_l}{\Omega} \frac{m_k}{\Omega},$$
$$\tilde{\beta}_j\left(\frac{m}{\Omega}\right) = c_j \Omega \frac{\left(\frac{m_l}{\Omega}\right)^1}{1!} \frac{\left(\frac{m_k}{\Omega}\right)^1}{1!} = c_j \Omega \frac{m_l}{\Omega} \frac{m_k}{\Omega},$$

and the error

$$\left|\beta_j - \tilde{\beta}_j\left(\frac{m}{\Omega}\right)\right| = 0.$$

$\exists k : \chi_{j,k}^{\mathbf{in}} = 2$ **and** $\chi_{j,i}^{\mathbf{in}} = 0 \; \forall \; i \neq k$

This results in the propensity functions

$$\beta_j(m) = c_j \Omega \Omega^{-2} \binom{m_k}{2} = c_j \Omega \frac{m_k(m_k - 1)}{2\Omega^2},$$
$$\tilde{\beta}_j\left(\frac{m}{\Omega}\right) = c_j \Omega \frac{\left(\frac{m_k}{\Omega}\right)^2}{2!} = c_j \Omega \frac{m_k^2}{2\Omega^2},$$

and the error

$$\begin{aligned}\left|\beta_j(m) - \tilde{\beta}_j\left(\frac{m}{\Omega}\right)\right| &= \left|c_j\Omega\frac{m_k(m_k-1)}{2\Omega^2} - c_j\Omega\frac{m_k^2}{2\Omega^2}\right| \\ &= \left|\Omega\frac{c_j}{2}\left(\frac{m_k^2 - m_k - m_k^2}{\Omega^2}\right)\right| = \left|\Omega\frac{c_j}{2}\frac{m_k}{\Omega^2}\right| \\ &= \left|\frac{c_j}{2}\frac{m_k}{\Omega}\right| \leq C,\end{aligned}$$

because $\frac{m_k}{\Omega} \leq K$ by assumption..

□

Remark: It is possible to extend this proof for polynomial propensity functions of higher degree than 2. However, reactions with more than two substrate particles are unlikely. Moreover, we use assumption 4.2 several times later on and it is more interesting to see how to deal with it in proofs than to prove a statement that we do not "need".

Motivated by equation (4.2) and definition 4.1 we obtain the scaled Kurtz process

$$X(t) = m_0 + \sum_{j=1}^{R} \mathcal{P}_j\left(\int_0^t \beta_j\left(X(s)\right)\mathrm{d}s\right)\varrho_j \tag{4.3}$$

with $m_0 = n_0\Omega$. Using concentrations is only useful if the underlying particle amounts are very large..

From this point on, every model for stochastic networks will be scaled by Ω. Hence, we will not point this out every time a new equation is derived and will refer to the CME, the Kurtz process and not to the *scaled* CME or the *scaled* Kurtz process.

From lemma 2.3 follows the CME

$$\partial_t p(t,m) = \sum_{j=1}^{R}\left(\beta_j(m-\varrho_j)p(t,m-\varrho_j) - \beta_j(m)p(t,m)\right). \tag{4.4}$$

4.2 The Scaled Kramers-Moyal Expansion

Let $\hat{q}(t,x) : \mathbb{R}_+ \times \mathbb{R}_+^N \to \mathbb{R}$ be a PDF with the property

$$\begin{aligned}p(t,m) = \mathbb{P}\left(X(t) = m\right) &= \mathbb{P}\left(\frac{X(t)}{\Omega} \in I_m\right) \\ &=: \int_{I_m} \hat{q}\left(t, \frac{m}{\Omega}\right)\mathrm{d}x = \hat{q}\left(t,\frac{m}{\Omega}\right)\int_{I_m}\mathrm{d}x = \Omega^{-N}\hat{q}\left(t,\frac{m}{\Omega}\right),\end{aligned} \tag{4.5}$$

where I_m denotes the volume

$$I_m := \frac{m}{\Omega} + \left[0,\frac{1}{\Omega}\right)^N = \left[\frac{m_1}{\Omega},\frac{m_1+1}{\Omega}\right) \times \cdots \times \left[\frac{m_N}{\Omega},\frac{m_N+1}{\Omega}\right) \tag{4.6}$$

i.e. the hypercube that spans from point $\frac{m}{\Omega}$ with edge length $\frac{1}{\Omega}$. The integral over this volume is therefor defined by

$$\int_{I_m} f(x)\,\mathrm{d}x = \int_{\frac{m_1}{\Omega}}^{\frac{m_1+1}{\Omega}} \cdots \int_{\frac{m_N}{\Omega}}^{\frac{m_N+1}{\Omega}} f(x)\,\mathrm{d}x_N \ldots\, \mathrm{d}x_1 \tag{4.7}$$

for a fixed point $m \in \mathbb{N}_0^N$ with m_i the ith entry of m and a function $f : \mathbb{R}_{0,+}^N \to \mathbb{R}$.

We insert equation (4.4) in eq. (4.5), approximate the propensity function using lemma 4.3 and switch to concentrations using eq. (4.1). Furthermore, we assume that

$$\left[x \mapsto \tilde{\beta}_j\left(x\right)\hat{q}\left(t,x\right)\right] \in \mathcal{C}^\infty, \tag{4.8}$$

and state the Kramers-Moyal expansion

$$\begin{aligned}\tilde{\beta}_j\left(x-\frac{\varrho_j}{\Omega}\right)\hat{q}\left(t,x-\frac{\varrho_j}{\Omega}\right) - \tilde{\beta}_j\left(x\right)\hat{q}\left(t,x\right) = &-\frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)\hat{q}(t,x)\Big)^T\varrho_j \\ &+\frac{1}{2\Omega^2}\varrho_j^T\nabla^2\Big(\tilde{\beta}_j(x)\hat{q}(t,x)\Big)\varrho_j \\ &+\mathcal{G}^\star_{j,\geq 3},\end{aligned} \tag{4.9}$$

where $\mathcal{G}^\star_{j,\geq 3}$ denotes the remainder term of order ≥ 3 of the Taylor expansion. Using the multi-index notation defined in def. B4, we can express this remainder term as:

$$\mathcal{G}^\star_{j,\geq 3} = \sum_{|\boldsymbol{k}|\geq 3}\frac{1}{\boldsymbol{k}!}\left(\frac{\varrho_j}{\Omega}\right)^{\boldsymbol{k}}\nabla^{\boldsymbol{k}}\Big(\tilde{\beta}_j(x)\hat{q}(t,x)\Big), \tag{4.10}$$

with $\boldsymbol{k}$ using the multi-index notation stated in appendix B.

This leads to

$$\begin{aligned}\partial_t\hat{q}\left(t,x\right) &\approx \sum_{j=1}^{R}\left(\tilde{\beta}_j\left(x-\frac{\varrho_j}{\Omega}\right)\hat{q}\left(t,x-\frac{\varrho_j}{\Omega}\right) - \tilde{\beta}_j\left(x\right)\hat{q}\left(t,x\right)\right) \\ &= \sum_{j=1}^{R}\left(-\frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)\hat{q}(t,x)\Big)^T\varrho_j + \frac{1}{2\Omega^2}\varrho_j^T\nabla^2\Big(\tilde{\beta}_j(x)\hat{q}(t,x)\Big)\varrho_j + \mathcal{G}^\star_{j,\geq 3}\right).\end{aligned}$$

The legitimacy of these assumptions is not obvious, but we will see in the proof of theorem 4.5 the impact of these approximations and that it is possible to state an error bound for this derivation.

Truncating the Kramers-Moyal expansion after the first term yields the Liouville equation

$$\begin{aligned}\partial_t u(t,x) &= -\sum_{j=1}^{R}\frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)u(t,x)\Big)^T\varrho_j, \\ u(0,x) &= u_0\end{aligned} \tag{4.11}$$

and by following the steps in chapter 2.4.1 we obtain the RRE

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}y(t) &= \sum_{i=1}^{R} \frac{1}{\Omega}\tilde{\beta}_j\left(y(t)\right)\varrho_j, \\ y(0) &= y_0, \end{aligned} \tag{4.12}$$

Truncating after the second term yields the Fokker-Planck equation

$$\begin{aligned} \partial_t q(t,x) &= \sum_{i=1}^{R}\left(-\frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)q(t,x)\Big)^T\varrho_j + \frac{1}{2\Omega^2}\varrho_j^T\nabla^2\Big(\tilde{\beta}_j(x)q(t,x)\Big)\varrho_j\right), \\ q(0,x) &= q_0 \end{aligned} \tag{4.13}$$

and by lemma 2.6 the CLE

$$\begin{aligned} \mathrm{d}x(t) &= \sum_{i=1}^{R}\frac{1}{\Omega}\tilde{\beta}_j\left(x(t)\right)\varrho_j\mathrm{d}t + \sum_{i=1}^{R}\sqrt{\frac{1}{\Omega^2}\tilde{\beta}_j\left(x(t)\right)}\varrho_j\mathrm{d}W(t), \\ x(0) &= x_0. \end{aligned} \tag{4.14}$$

4.3 Convergence of the Kurtz process to the RRE

Kurtz showed in 1972 convergence between the Kurtz process and the RRE and cleared with his seminal work the way for analysing the errors which are induced by the different models for reaction networks. He assumed that the probability of molecules reacting in a time interval is proportional to the volume of the reaction system [Kur72]. This assumption allowed the definition of scaled propensity functions, a scaled Kurtz process and a scaled RRE. Then, he analysed the behaviour of these models in the thermodynamic limit (eq. (4.2)). Doing this, he generalised and proved the conjecture by Oppenheim et al. that the Kurtz process converges to the RRE in the thermodynamic limit [OSW69].

The scaled Kurtz process $\frac{X(t)}{\Omega}$ converges in probability to the solution of the RRE $y(t)$:

$$\lim_{\Omega\to\infty}\mathbb{P}\left(\sup_{s\leq t}\left|\frac{X(s)}{\Omega} - y(s)\right| > \varepsilon\right) = 0,$$

for every time t and a small, but positive constant $\varepsilon > 0$, if

$$y(s) = y_0 \qquad \text{and} \qquad \lim_{\Omega\to\infty}\frac{X(0)}{\Omega} = y_0. \tag{4.15}$$

4.4 Convergence of the FPE

> A careful inspection of our chemical Fokker-Planck equation reveals that it is precisely the equation that would be obtained by simply truncating the chemical Kramers-Moyal equation at $n = 2$. The temptation to make that truncation has long been recognized, but the legitimacy of doing so has been seriously doubted.[...] Gardiner has observed that "a confused history [has arisen] out of repeated attempts to find a limiting form of the master equation in which a Fokker-Planck equation arises."
>
> *quoted from [Gil00, Gar04, p. 246]*

This quote by Gillespie and Gardiner summarizes the ("confused") history of the FPE quite well. Although, Einstein was the first who mentioned the connection between the Kramers-Moyal expansion and the FPE [Ein05], it was Kramers in 1940 and later Moyal in 1949, who first analysed and discussed this expansion [Kra40, Moy49]. To motivate the truncation, Van Kampen introduced the approach we used in the beginning of this chapter and scaled the particle amount by the system volume to derive a model for the concentration, although he used a slightly different scaling. This lead the way to the motivation of the truncation of the Kramers-Moyal expansion [Kam61]. Kubo et al. used a scaling of the propensity functions to motivate the same result, but as Gillespie pointed out, backed away from their own result [KMK73]. Gillespie contributed to the analysis of the FPE with a study in 1980 and one in 2000. In his first paper, he analysed truncations of the Kramers-Moyal expansion for reaction networks where species can only be changed by one molecule per reaction, i.e. $\sum_{i=1}^{N} \varrho_{j,i}^{\text{in}} \leq 1,\ \forall j = 1, \cdots, R$. He could show that, by using higher order difference quotients, the FPE is only an approximation of the CME and suggested that a cutoff of the Kramers-Moyal expansion after the fourth order derivative would equal the CME. But he did not distinguish rigorously between particle amounts and concentrations [Gil80]. In his second contribution, Gillespie derived the CLE and the FPE from the CME using certain assumptions [Gil00]. But this work omits an accurate discussion of the differences between propensity functions for particle amounts and propensity functions for concentrations. However, we followed this scheme in chapter 2.4.1, 4.1 and 4.2 to derive the CLE and the FPE, but added a proper handling of the propensity functions and state variables.

Thomas Kurtz discussed that the solution of the CLE converges to a Markov jump process on a scaled state space $\left\{ \frac{k}{\Omega} : k \in \mathbb{N}_0^N \right\}$ with order $\mathcal{O}\left(\frac{\log \Omega}{\Omega}\right)$ [Kur78].

Crispin Gardiner stated that a diffusion process, like the CLE, can be approximated by a Markov jump process, like the Kurtz process, but not vice versa. He could show this result, by using a different scaling scheme than the one used in this thesis. Gardiner introduced a parameter δ that is proportional to the mean and variance of the jump probabilities of the Poisson process. The factor δ becomes small if the jump probabilities increase. He took the limit $\delta \to 0$ and used this to show the statement. Later he analysed the approaches

by Kramers, Moyal and van Kampen and claimed that the FPE approximates the CME with order $\mathcal{O}\left(\frac{1}{\sqrt{\Omega}}\right)$[Gar04, p. 251]. This result is consistent with the numerical example at the end of this chapter, but the generality of Gardiner's result is questionable. He did not distinguish between a discrete and a continuous state variable. Further he assumed equality for the change in propensity functions, while lemma 4.3 clearly shows an error for polynomial propensity functions of degree two or higher.

Sjöberg et al. chose a completely different approach. They abstain from motivating the cut off of the Kramers-Moyal expansion and take the FPE as given. To compare it with the CME, they discretized the FPE numerically using a Finite Volume approach in space and a backward differentiation formula of second order (BDF-2) in time. The discretized FPE could be compared with the already discrete CME and based on this experiment convergence was assumed [SLE09].

However, these studies have in common that none is combining a proper scaling, resulting from correctly distinguishing between particle counts and concentrations, with a proper handling of the propensity functions for discrete and continuous state variables. Also there are nearly no studies that introduce a mapping that allows the comparison between results obtained by the CME with results from the FPE.

In this thesis, we combine these three approaches to show convergence of the FPE to the CME and give an error bound for the convergence.

After this review of the history of convergence proofs for the FPE we will analyse the problem in the scaled setting defined at the beginning of this chapter.

First we have to define in what sense we like to see convergence. We want to compare the solution of the CME with the solution of the FPE

$$p(t,m) \overset{?}{\longleftrightarrow} q(t,x).$$

But the CME maps

$$p : \mathbb{R}_+ \times \mathbb{N}^N \to \mathbb{R}$$

while the FPE maps

$$q : \mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}.$$

Therefore, we must find a way to compare a discrete PDF with a continuous one.

Figure 4.1 visualises an example distribution for the CME solution (labeled $p(t,m)$) and on for the FPE solution (labeled $q(t,x)$). As mentioned, the state spaces differ by two main characteristics. The location of the essential supports and the discreteness of the CME in contrast to the continuous space of the FPE. In fig. 4.1 we choose $\Omega = 100$ and assume that the essential support (cf. fig. 4.2)

$$\operatorname{esssupp}_\varepsilon(p(t,\cdot)) = \mathbb{N}_0^N \setminus \left\{ m \in \mathbb{N}_0^N : p(t,m) < \varepsilon \right\}$$

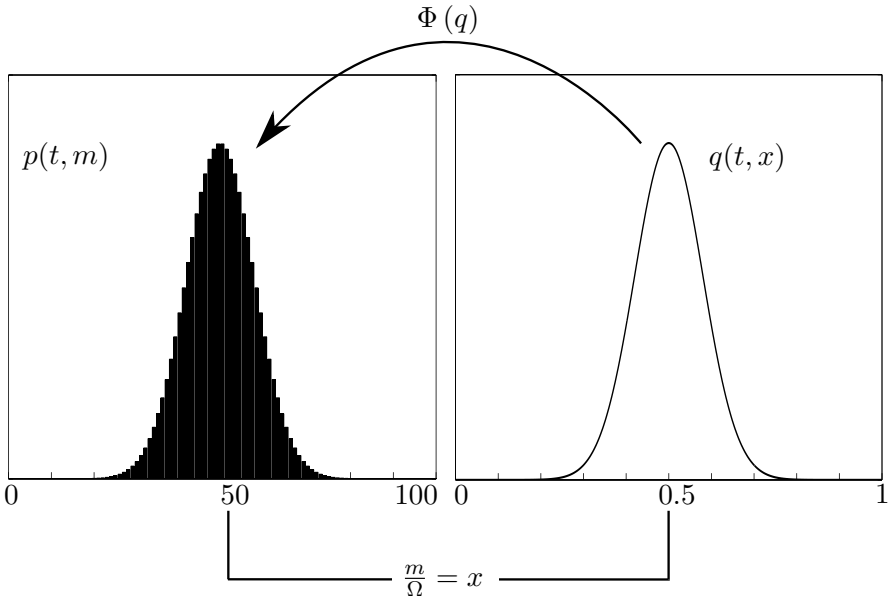


Figure 4.1: Connection between the scaled CME (left) and the scaled FPE (right). The transformation Φ maps the continuous values of q to a discrete state in p. We have to notice that not only the state space is changed between discrete and continuous, but also the position is changed by the relationship $x = \frac{m}{\Omega}$.

of the CME solution $p(t, m)$ is entirely located inside the interval $[0, \Omega]$. From the relation

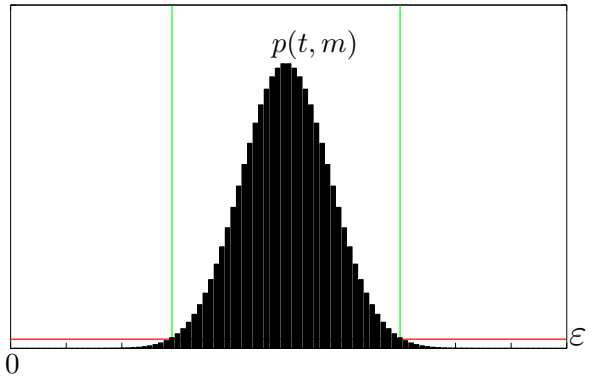


Figure 4.2: The essential support (green lines) of the solution $p(t, m)$ of the CME (black). The red line denotes the value of the thresholding parameter ε.

(cf. eq. (4.1))

$$x = \frac{m}{\Omega}$$

follows that the essential support of the FPE solution is located in the interval $[0, 1]$ and further

$$\mathbb{E}\left[p(t, \cdot)\right] \gg \mathbb{E}\left[q(t, \cdot)\right]. \tag{4.16}$$

However, we already motivated in the beginning of this chapter and in the last chapter why we cannot compare particle numbers with concentrations.

As noted, the CME and the FPE somehow solve the same reaction network, but the results cannot be compared directly. It is therefore impossible to derive a convergence rate based on the relation

$$\|p(t,m) - q(t,x)\|.$$

However, it is possible to define a transformation Φ that converts the solution of the FPE into a discrete PDF that can be compared with the CME solution. Figure 4.1 visualises the relation between the two PDFs and sketches the idea of the map Φ.

By using equation (4.5) and the interval I_m (cf. eq. (4.6)) we define

$$\Phi(q)(t,m) := \int_{I_m} q(t,x)\,\mathrm{d}x. \tag{4.17}$$

Further, we implicitly define the operators

$$\partial_t p(t,m) = (\mathcal{A}p)(t,m) = \sum_{j=1}^{R} \left(\beta_j(m-\varrho_j)p(t,m-\varrho_j) - \beta_j(m)p(t,m)\right), \tag{4.18}$$

$$\begin{aligned}\partial_t q(t,x) = (\mathcal{B}q)(t,x) = \sum_{j=1}^{R} \Big(&- \frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)q(t,x)\Big)^T \varrho_j \\ &+ \frac{1}{2\Omega^2}\varrho_j^T \nabla^2\Big(\tilde{\beta}_j(x)q(t,x)\Big)\varrho_j\Big),\end{aligned} \tag{4.19}$$

and state

$$\partial_t v(t,m) = (\mathcal{A}v)(t,m) + \mathcal{R}(t,m), \tag{4.20}$$

with $v(t,m) := \Phi(q)(t,m)$ and from eq. (4.18) and eq. (4.19) follows

$$\mathcal{R}(t,m) = \int_{I_m} (\mathcal{B}q)(t,x)\,\mathrm{d}x - (\mathcal{A}v)(t,m). \tag{4.21}$$

Assumption 4.4 (Bound for remainder terms)**.**
We assume that, on bounded time intervals $[0, t_{\mathrm{end}}]$,

$$\left\|\mathcal{G}^{\star}_{j,\geq 3}\right\|_{L_1} + \left\|\mathcal{G}^{\dagger}_{j,\geq 2}\right\|_{L_1} \leq \frac{C}{\Omega}$$

with the Taylor series remainder terms (cf. def. B4, eq. (4.9), eq. (4.10))

$$\mathcal{G}^{\star}_{j,\geq 3} := -\sum_{|\boldsymbol{k}|\geq 3} \frac{1}{\boldsymbol{k}!}\left(\frac{\varrho_j}{\Omega}\right)^{\boldsymbol{k}} \nabla^{\boldsymbol{k}}\Big(\tilde{\beta}_j(x)q(t,x)\Big),$$

$$\mathcal{G}^{\dagger}_{j,\geq 2} := -\sum_{|\boldsymbol{k}|\geq 2} \frac{1}{\boldsymbol{k}!}\left(\frac{\varrho_j}{\Omega}\right)^{\boldsymbol{k}} \nabla^{\boldsymbol{k}}\Big(q(t,x)\Big).$$

Theorem 4.5 (Convergence of the FPE)**.**
Let p be the solution of the CME (4.4) *and q be the solution of the FPE* (4.13)*, with non zero variance for all $t \in [0, t_{end}]$. Further, let $v(t,m) := \Phi(q)(t,m)$ be the transformed FPE solution. Then it holds, on bounded time intervals $[0, t_{end}]$, for $\Omega > \Omega_0$ and for Ω independent constants $C_1, C_2 > 0$, that*

$$\|p(t,\cdot) - v(t,\cdot)\|_{\ell_1} \leq \frac{C_1}{\Omega} \left\| \nabla\Big(q(t,\cdot)\Big) \right\|_{L_1} + \frac{C_2}{\Omega},$$

under the conditions of eq. (4.2)*, eq.* (4.8)*, eq.* (4.16)*, lemma 4.3, assumption 4.2, assumption 4.4 and assuming that*

$$\|p(0,\cdot) - v(0,\cdot)\|_{\ell_1} = \left\| p(0,\cdot) - \int_{I_m} q(0,\cdot)\,\mathrm{d}x \right\|_{\ell_1} = 0. \tag{4.22}$$

The constant Ω_0 depends on the conditions of eq. (4.2)*, lemma 4.3 and assumption 4.4.*

Remark: We should be aware that q depends on Ω and therefore $\|\nabla(q(t,\cdot))\|_{L_1}$ changes if Ω is changed.

Proof. We obtain, from eq. (4.20) and the variation of constants formula, that

$$\|p(t,\cdot) - v(t,\cdot)\|_{\ell_1} \leq \int_0^t \left\| e^{(t-s)\mathcal{A}} \right\|_{\ell_1} \cdot \|\mathcal{R}(s,\cdot)\|_{\ell_1}\,\mathrm{d}s.$$

The definition and discussion of $e^{(t-s)\mathcal{A}}$ for unbounded operators $\mathcal{A}$ can be found in [EN00, chapter II, section 3.3]. From the structure of the operator $\mathcal{A}$, it was shown that

$$\left\| e^{(t-s)\mathcal{A}} \right\|_{\ell_1} \leq 1,$$

cf. [Heg08, WGMH10, JA10, VK07, ch. V].

Hence the error only depends on $\mathcal{R}$:

$$\begin{aligned}
\mathcal{R}(t,m) &= \int_{I_m} (\mathcal{B}q)(t,x)\,\mathrm{d}x - (\mathcal{A}v)(t,m) \\
&= \sum_{j=1}^{R} \int_{I_m} \left(-\frac{1}{\Omega} \nabla\Big(\tilde{\beta}_j(x) q(t,x)\Big)^T \varrho_j + \frac{1}{2\Omega^2} \varrho_j^T \nabla^2\Big(\tilde{\beta}_j(x) q(t,x)\Big) \varrho_j \right) \mathrm{d}x \\
&\quad - \sum_{j=1}^{R} \left(\beta_j(m-\varrho_j) v(t, m-\varrho_j) - \beta_j(m) v(t,m) \right) \\
&= \sum_{j=1}^{R} \int_{I_m} \left(-\frac{1}{\Omega} \nabla\Big(\tilde{\beta}_j(x) q(t,x)\Big)^T \varrho_j + \frac{1}{2\Omega^2} \varrho_j^T \nabla^2\Big(\tilde{\beta}_j(x) q(t,x)\Big) \varrho_j \right) \mathrm{d}x \\
&\quad - \sum_{j=1}^{R} \left(\beta_j(m-\varrho_j) \int_{I_m - \varrho_j} q(t,x)\,\mathrm{d}x - \beta_j(m) \int_{I_m} q(t,x)\,\mathrm{d}x \right).
\end{aligned}$$

Using integration by substitution, it holds that

$$\int_{I_{m-\varrho_j}} q(t,x)\,\mathrm{d}x = \int_{I_m} q\left(t, x - \frac{\varrho_j}{\Omega}\right)\,\mathrm{d}x.$$

We use the Kramers-Moyal expansion to derive

$$\begin{aligned}-\frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)q(t,x)\Big)^T\varrho_j + \frac{1}{2\Omega^2}\varrho_j^T\nabla^2\Big(\tilde{\beta}_j(x)q(t,x)\Big)\varrho_j =& \tilde{\beta}_j\left(x-\frac{\varrho_j}{\Omega}\right)q\left(t,x-\frac{\varrho_j}{\Omega}\right)\\ &- \tilde{\beta}_j\left(x\right)q\left(t,x\right) + \mathcal{G}^\star_{j,\geq 3},\end{aligned}$$

with remainder term $\mathcal{G}^\star_{j,\geq 3}$ (defined in assumption 4.4). This can be combined to

$$\begin{aligned}\|\mathcal{R}(t,\cdot)\|_{\ell_1} &= \sum_{m\in\mathbb{N}_0^N}\left|\sum_{j=1}^R\left(\int_{I_m} -\frac{1}{\Omega}\nabla\Big(\tilde{\beta}_j(x)q(t,x)\Big)^T\varrho_j + \frac{1}{2\Omega^2}\varrho_j^T\nabla^2\Big(\tilde{\beta}_j(x)q(t,x)\Big)\varrho_j\,\mathrm{d}x\right.\right.\\ &\qquad\left.\left.-\beta_j(m-\varrho_j)\int_{I_{m-\varrho_j}} q(t,x)\,\mathrm{d}x + \beta_j(m)\int_{I_m} q(t,x)\,\mathrm{d}x\right)\right|\\ &= \sum_{m\in\mathbb{N}_0^N}\left|\sum_{j=1}^R\left(\int_{I_m}\tilde{\beta}_j\left(x-\frac{\varrho_j}{\Omega}\right)q\left(t,x-\frac{\varrho_j}{\Omega}\right) - \tilde{\beta}_j\left(x\right)q\left(t,x\right) + \mathcal{G}^\star_{j,\geq 3}\,\mathrm{d}x\right.\right.\\ &\qquad\left.\left.-\beta_j(m-\varrho_j)\int_{I_m} q\left(t,x-\frac{\varrho_j}{\Omega}\right)\,\mathrm{d}x + \beta_j(m)\int_{I_m} q(t,x)\,\mathrm{d}x\right)\right|\\ &\leq \sum_{j=1}^R\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\left[\tilde{\beta}_j\left(x-\frac{\varrho_j}{\Omega}\right)-\beta_j(m-\varrho_j)\right]q\left(t,x-\frac{\varrho_j}{\Omega}\right)\,\mathrm{d}x\right.\\ &\qquad\left.-\int_{I_m}\left[\tilde{\beta}_j\left(x\right)-\beta_j(m)\right]q(t,x)\,\mathrm{d}x + \int_{I_m}\mathcal{G}^\star_{j,\geq 3}\,\mathrm{d}x\right|.\end{aligned}\tag{4.23}$$

We represent the first term in $[\cdot]$-brakets by the second one plus a remainder term $\mathcal{H}_j(m,x)$

$$\left[\tilde{\beta}_j\left(x-\frac{\varrho_j}{\Omega}\right)-\beta_j(m-\varrho_j)\right] = \left[\tilde{\beta}_j\left(x\right)-\beta_j(m)\right] + \mathcal{H}_j(m,x).$$

We rearrange and obtain

$$\begin{aligned}\|\mathcal{R}(t,\cdot)\|_{\ell_1} \leq\quad & \sum_{j=1}^R\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\left[\tilde{\beta}_j\left(x\right)-\beta_j(m)\right]q\left(t,x-\frac{\varrho_j}{\Omega}\right)\,\mathrm{d}x\right.\\ &\qquad -\int_{I_m}\left[\tilde{\beta}_j\left(x\right)-\beta_j(m)\right]q(t,x)\,\mathrm{d}x\\ &\qquad\left.+\int_{I_m}\mathcal{H}_j(m,x)q\left(t,x-\frac{\varrho_j}{\Omega}\right)\,\mathrm{d}x + \int_{I_m}\mathcal{G}^\star_{j,\geq 3}\,\mathrm{d}x\right|\\ \leq\quad & \sum_{j=1}^R\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\left[\tilde{\beta}_j\left(x\right)-\beta_j(m)\right]\left[q\left(t,x-\frac{\varrho_j}{\Omega}\right)-q(t,x)\right]\,\mathrm{d}x\right|\\ &+\sum_{j=1}^R\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{H}_j(m,x)q\left(t,x-\frac{\varrho_j}{\Omega}\right)\,\mathrm{d}x\right|\end{aligned}$$

$$+\sum_{j=1}^{R}\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{G}^{\star}_{j,\geq 3}\,\mathrm{d}x\right|.$$

We choose for each of the three part the reaction channel with the biggest impact on the error. Therefor, we choose j_1 such that

$$\begin{aligned}&\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\left[\tilde{\beta}_{j_1}(x)-\beta_{j_1}(m)\right]\left[q\left(t,x-\frac{\varrho_{j_1}}{\Omega}\right)-q(t,x)\right]\,\mathrm{d}x\right|\\ \geq&\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\left[\tilde{\beta}_{j}(x)-\beta_{j}(m)\right]\left[q\left(t,x-\frac{\varrho_{j}}{\Omega}\right)-q(t,x)\right]\,\mathrm{d}x\right|\end{aligned}$$

and j_2 such that

$$\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{H}_{j_2}(m,x)q\left(t,x-\frac{\varrho_{j_2}}{\Omega}\right)\,\mathrm{d}x\right|\geq\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{H}_{j}(m,x)q\left(t,x-\frac{\varrho_{j}}{\Omega}\right)\,\mathrm{d}x\right|$$

and j_3 such that

$$\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{G}^{\star}_{j_3,\geq 3}\,\mathrm{d}x\right|\geq\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{G}^{\star}_{j,\geq 3}\,\mathrm{d}x\right|$$

for all $j=1,\ldots,R$. Then there exist constants C_1, C_2 and C_3 such that

$$\begin{aligned}\|\mathcal{R}(t,\cdot)\|_{\ell_1}\leq C_1&\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\left[\tilde{\beta}_{j_1}(x)-\beta_{j_1}(m)\right]\left[q\left(t,x-\frac{\varrho_{j_1}}{\Omega}\right)-q(t,x)\right]\,\mathrm{d}x\right|\\ +C_2&\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{H}_{j_2}(m,x)q\left(t,x-\frac{\varrho_{j_2}}{\Omega}\right)\,\mathrm{d}x\right|\\ +C_3&\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}\mathcal{G}^{\star}_{j_3,\geq 3}\,\mathrm{d}x\right|.\end{aligned}$$

Motivated by Hölders inequality we estimate

$$\left|\int_{I_m}f(m,x)g(x)\,\mathrm{d}x\right|\leq\int_{I_m}|f(m,x)|\cdot|g(x)|\,\mathrm{d}x\leq\sup_{x\in I_m}|f(m,x)|\int_{I_m}|g(x)|\,\mathrm{d}x,$$

for

$$\begin{aligned}&f(m,x)=\left[\tilde{\beta}_{j_1}(x)-\beta_{j_1}(m)\right]\qquad\text{and}\qquad g(x)=q\left(t,x-\frac{\varrho_{j_1}}{\Omega}\right)-q(t,x)\\ \text{or}\qquad&f(m,x)=\mathcal{H}_{j_2}(m,x)\qquad\text{and}\qquad g(x)=q\left(t,x-\frac{\varrho_{j_2}}{\Omega}\right).\end{aligned}$$

Further we note

$$\sum_{m\in\mathbb{N}_0^N}\left|\int_{I_m}a(x)\,\mathrm{d}x\right|\leq\sum_{m\in\mathbb{N}_0^N}\int_{I_m}|a(x)|\,\mathrm{d}x=\int_{\mathbb{R}_{0,+}}|a(x)|\,\mathrm{d}x=\|a(\cdot)\|_{L_1},$$

for a function $a(x)$.

We combine these results and estimate

$$\begin{aligned}
\|\mathcal{R}(t,\cdot)\|_{\ell_1} \leq & C_1 \sum_{m\in\mathbb{N}_0^N} \left| \int_{I_m} \left[\tilde{\beta}_{j_1}(x) - \beta_{j_1}(m)\right] \left(q\left(t, x - \frac{\varrho_{j_1}}{\Omega}\right) - q(t,x)\right) \mathrm{d}x \right| \\
& + C_2 \sum_{m\in\mathbb{N}_0^N} \left| \int_{I_m} \mathcal{H}_{j_2}(m,x) q\left(t, x - \frac{\varrho_{j_2}}{\Omega}\right) \mathrm{d}x \right| + C_3 \sum_{m\in\mathbb{N}_0^N} \left| \int_{I_m} \mathcal{G}^{\star}_{j_3,\geq 3} \mathrm{d}x \right| \\
\leq & C_1 \sum_{m\in\mathbb{N}_0^N} \sup_{x\in I_m} \left|\tilde{\beta}_{j_1}(x) - \beta_{j_1}(m)\right| \int_{I_m} \left| q\left(t, x - \frac{\varrho_{j_1}}{\Omega}\right) - q(t,x) \right| \mathrm{d}x \qquad (4.24) \\
& + C_2 \sum_{m\in\mathbb{N}_0^N} \sup_{x\in I_m} \left|\mathcal{H}_{j_2}(m,x)\right| \int_{I_m} \left| q\left(t, x - \frac{\varrho_{j_2}}{\Omega}\right) \right| \mathrm{d}x \\
& + C_3 \left\| \mathcal{G}^{\star}_{j_3,\geq 3} \right\|_{L_1}.
\end{aligned}$$

We estimate

$$\begin{aligned}
\sup_{x\in I_m} \left|\tilde{\beta}_{j_1}(x) - \beta_{j_1}(m)\right| &\leq C_4, \\
\sup_{x\in I_m} \left|\mathcal{H}_{j_2}(m,x)\right| &\leq \frac{C_5}{\Omega}.
\end{aligned}$$

These estimates are not obvious, but we postpone them to the lemmata C5 and C6, because we do not want to interrupt the actual line of thought.

We insert the estimates into equation (4.24) and get

$$\begin{aligned}
\|\mathcal{R}(t,\cdot)\|_{\ell_1} \leq & C_6 \sum_{m\in\mathbb{N}_0^N} \int_{I_m} \left| q\left(t, x - \frac{\varrho_{j_1}}{\Omega}\right) - q(t,x) \right| \mathrm{d}x \\
& + \frac{C_7}{\Omega} \underbrace{\sum_{m\in\mathbb{N}_0^N} \int_{I_m} \left| q\left(t, x - \frac{\varrho_{j_2}}{\Omega}\right) \right| \mathrm{d}x}_{=1} + C_3 \left\| \mathcal{G}^{\star}_{j_3,\geq 3} \right\|_{L_1} \\
\leq & C_6 \sum_{m\in\mathbb{N}_0^N} \int_{I_m} \left| q\left(t, x - \frac{\varrho_{j_1}}{\Omega}\right) - q(t,x) \right| \mathrm{d}x + \frac{C_7}{\Omega} + C_3 \left\| \mathcal{G}^{\star}_{j_3,\geq 3} \right\|_{L_1}. \qquad (4.25)
\end{aligned}$$

We derive by Taylor expansion

$$q\left(t, x - \frac{\varrho_{j_1}}{\Omega}\right) - q(t,x) = \frac{1}{\Omega} \nabla\Big(q(t,x)\Big)^T \varrho_{j_1} + \mathcal{G}^{\dagger}_{j_1,\geq 2},$$

with remainder term $\mathcal{G}^{\dagger}_{j_1,\geq 2}$ (defined in assumption 4.4).

This results in

$$\begin{aligned}
\|\mathcal{R}(t,\cdot)\|_{\ell_1} &\leq \frac{C_8}{\Omega} \left\| \nabla\Big(q(t,\cdot)\Big) \right\|_{L_1} + \frac{C_7}{\Omega} + C_3 \left\| \mathcal{G}^{\star}_{j_3,\geq 3} \right\|_{L_1} + \left\| \mathcal{G}^{\dagger}_{j_1,\geq 2} \right\|_{L_1} \\
&\leq \frac{C_8}{\Omega} \left\| \nabla\Big(q(t,\cdot)\Big) \right\|_{L_1} + \frac{C_9}{\Omega}
\end{aligned}$$

The two terms $\mathcal{G}^{\dagger}_{j_1,\geq 2}$ and $\mathcal{G}^{\star}_{j_3,\geq 3}$ are bounded by assumption 4.4, which is motivated using the same argument that justified the truncation of the Kramers-Moyal expansion: The Taylor series are scaled by a factor $\frac{1}{\Omega^k}$, which renders higher order terms negligible, if the derivatives are smooth.

These derivations result in the estimate

$$\begin{aligned}\|p(t,\cdot) - v(t,\cdot)\|_{\ell_1} &\leq \int_0^t \left[\frac{C_8}{\Omega}\left\|\nabla\Big(q(s,\cdot)\Big)\right\|_{L_1} + \frac{C_9}{\Omega}\right] \mathrm{d}s \\ &= \int_0^t \frac{C_8}{\Omega}\left\|\nabla\Big(q(s,\cdot)\Big)\right\|_{L_1} \mathrm{d}s + \int_0^t \frac{C_9}{\Omega}\mathrm{d}s \\ &\leq \frac{C_{10}}{\Omega}\left\|\nabla\Big(q(t,\cdot)\Big)\right\|_{L_1} + \frac{C_{11}}{\Omega},\end{aligned}$$

with C_{10}, C_{11} constants. These constants can be found because the time interval as well as the L_1 norm are bounded. Obviously they depend on the chosen time point t. □

4.4.1 A Numerical Example

Table 4.1: A one dimensional reaction network, consisting of four reaction channels and one species.

Reaction Channel				**Propensity**		**Stoichiometry**		**Rate constant**
$\mathfrak{R}_j$				$\beta_j(m)$	$\tilde{\beta}_j(x)$	χ_j^{in}	ϱ_j	c_j
$\mathfrak{R}_1$:	$\star$	$\xrightarrow{c_1}$	$\mathcal{S}$	Ω	Ω	0	$+1$	0.01
$\mathfrak{R}_2$:	$\mathcal{S}$	$\xrightarrow{c_2}$	$\emptyset$	m	Ωx	1	-1	0.01
$\mathfrak{R}_3$:	$\mathcal{S}+\mathcal{S}$	$\xrightarrow{c_3}$	$\mathcal{S}$	$\frac{m(m-1)}{2\Omega}$	$\Omega\frac{x^2}{2}$	2	-1	0.8
$\mathfrak{R}_4$:	$\mathcal{S}$	$\xrightarrow{c_4}$	$\mathcal{S}+\mathcal{S}$	m	Ωx	1	$+1$	0.005

Table 4.1 lists a reaction network consisting of four reaction channels and only one species. The network was designed to analyse the convergence of the FPE to the CME and has no biological interpretation. The first and second reaction channels are an inflow reaction and an outflow reaction, creating and decomposing particles of species $\mathcal{S}$, respectively. The third channel models a dimerisation reaction, where two particles of $\mathcal{S}$ react to only one particle. The last channel models the reverse reaction of the dimerisation. This network represents all possible reactions that are allowed by assumption 4.2 and exhibits polynomial propensity functions of zero, first and second order. Table 4.1 also displays the two propensity functions β_j and $\tilde{\beta}_j$, the stoichiometric values χ_j^{in} and ϱ_j and the rate constant c_j for each reaction.

The CME and the FPE for this network were derived and numerically solved for different values of Ω. Then, the mapping Φ was used to compute $v(t,m)$ from $q(t,x)$. Using these three functions it is possible to calculate the error in theorem 4.5 numerically.

To solve the CME, the network was constructed such that the essential support of the PDF always lies inside the interval $[0, \Omega]$. For this cases, the operator $\mathcal{A}$ can be approximated by a $(\Omega + 1) \times (\Omega + 1)$ matrix. This matrix was implemented in a Matlab® program and the resulting ODE was solved using the Matlab® ODE integrator `ode45`. The initial conditions where calculated transforming the continuous Gaussian $q_0 = \mathcal{N}(0.5, 0.05)$ on the unit interval to the support of the CME using $p_0 = \Phi(q_0)$.

Similarly, the FPE with initial condition q_0 and zero Dirichlet boundary conditions was solved by limiting the essential support to the unit interval $[0, 1]$ and discretising this interval with a small mesh size Δx. Similar to the previous case, the operator $\mathcal{B}$ can be approximated by a finite matrix. The derivatives were approximated using second order difference quotients. Again the resulting system was integrated over time using the Matlab® ODE routine `ode45`.

The integrals in the transformation Φ and in the L_1 norm were approximated by the Simpson rule for numerical quadrature.

The size of the computation intervals, the space step size Δx and the time step size of the `ode45` method where chosen such that the resulting numerical error $\varepsilon_{\text{num}} < \frac{1}{\Omega}$. The resulting loss of probability mass was less than $10^{-6} \ll 10^{-3} = \frac{1}{\Omega_{\max}}$. The maximal value of Ω in this example is $\Omega_{\max} = 10^3$.

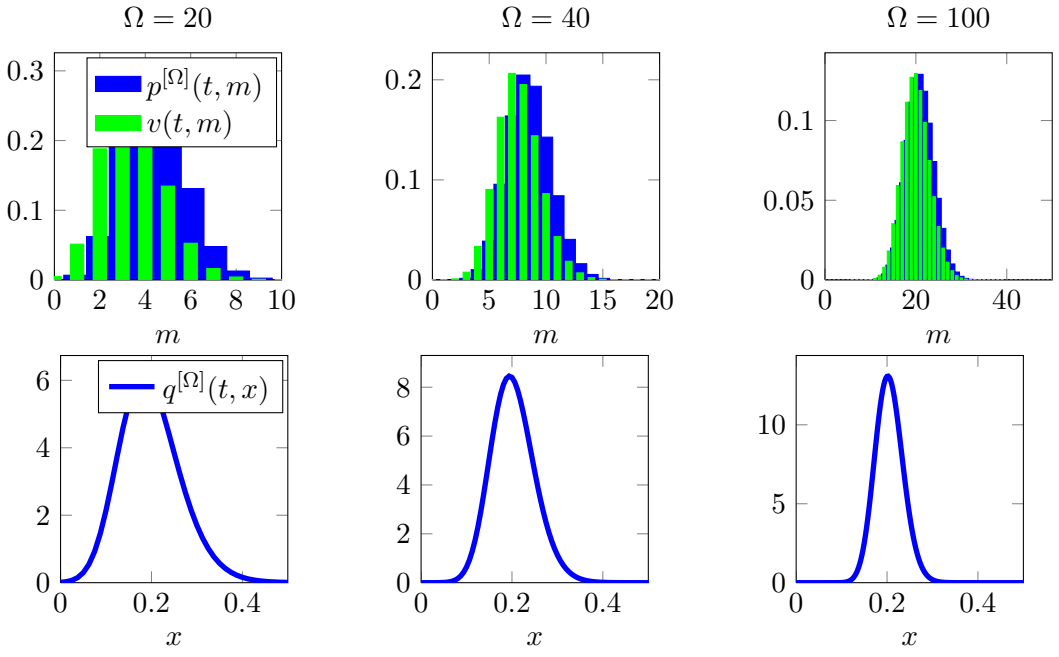


Figure 4.3: Numerical solutions of the CME (top, blue), the FPE (bottom, blue) and the transformed FPE solution v (top, green) for different values of Ω at time $t = 50$.

Figure 4.3 visualises the numerical solutions of these three quantities for different values of Ω. The FPE solution stays nearly constant, while the CME solution "moves" to the right for larger values of Ω. An interesting observation that may deserve future research, is that the transformed FPE solution always stays slightly left of the CME solution. The reaction network has the property that the PDF moves from its initial condition to the left. It seems that the FPE is a little faster in this movement than the CME solution. It is

possible that this is a property of the difference term $\beta_j(m-\varrho_j)p(t,m-\varrho_j)-\beta_j(m)p(t,m)$ in the CME in contrast to the derivatives in the FPE.

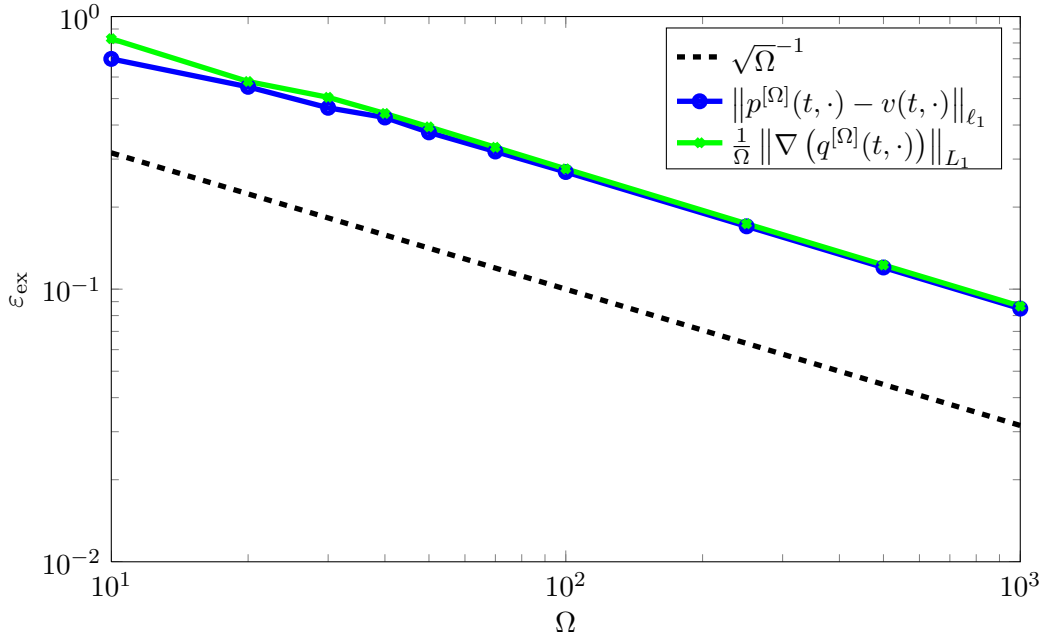


Figure 4.4: Convergence of the terms in theorem 4.5 using the solutions of the CME and the FPE for the reaction network given in table 4.1 at time point $t = 50$.

Figure 4.4 visualises the term $\varepsilon_{\text{ex}} = \|p(t,\cdot)-v(t,\cdot)\|_{\ell_1}$ (blue) for different values of Ω plotted over Ω. The derivative term $\frac{1}{\Omega}\|\nabla(q(t,\cdot))\|_{L_1}$ (green) is also plotted over Ω. As we expected from theorem 4.1 these terms are of same order, i.e. they lie parallel in a double logarithmic plot. Interestingly, these terms are of order $\mathcal{O}\left(\frac{1}{\sqrt{\Omega}}\right)$, which results from the observation that $\|\nabla q\|_{L_1} = \mathcal{O}\left(\sqrt{\Omega}\right)$.

Another interesting observation is that the order of the error changes over time. So far, we discussed the error at the end of the time interval $[0, 50]$. But the error at the beginning of this time interval is of order $\mathcal{O}\left(\frac{1}{\Omega}\right)$. We observe a gradual change of this order from $\mathcal{O}\left(\frac{1}{\Omega}\right)$ to $\mathcal{O}\left(\frac{1}{\sqrt{\Omega}}\right)$ over time. This change is plotted in figure 4.5, accompanied by the scaled derivative of the FPE solution $\frac{1}{\Omega}\|\nabla(q(t,\cdot))\|_{L_1}$. However, this behaviour could also result from a change of the constant Ω_0 (cf. theorem 4.5) or from the numerical scheme used to compute this example.

4.4.2 Future Research

So far we have seen that there exist several models to simulate the time evolution of reaction networks. For a systems-biologist or any other user of these models is it important to know under which circumstances he can use each model. For example, it may be known that the measured data involved in creating the reaction network (e.g. the rate constants or initial values) have an error of order $\mathcal{O}(\varepsilon_{\text{bio}})$. Now the experimenter can decide to use the FPE, if

$$\varepsilon_{\text{bio}} \geq \frac{C_1}{\Omega}\left\|\nabla\left(q(t,\cdot)\right)\right\|_{L_1} + \frac{C_2}{\Omega}.$$

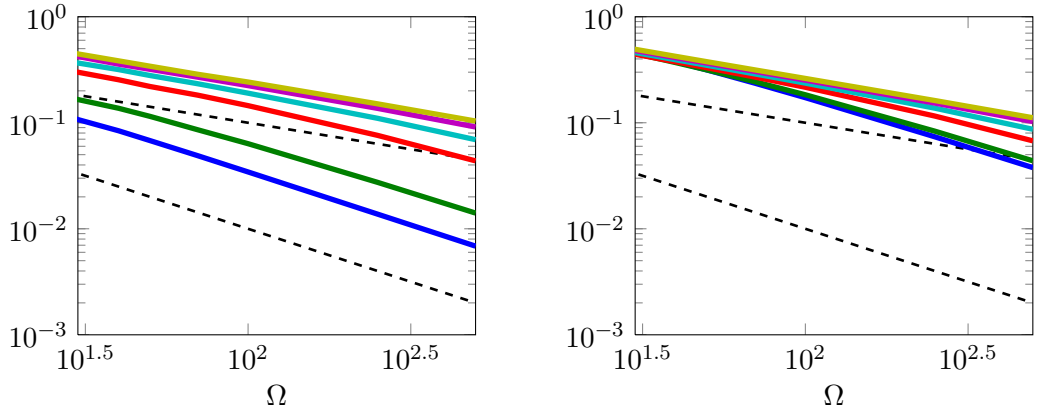


Figure 4.5: Evolution of the error of the FPE against the CME over time (left), in comparison to the evolution of the scaled derivative $\frac{1}{\Omega}\left\|\nabla\left(q(t,\cdot)\right)\right\|_{L_1}$ (right). The different time points are colour coded: $t = 0.5$ (blue), $t = 1$ (green), $t = 3$ (red), $t = 5$ (cyan), $t = 7.5$ (purple) and $t = 10$ (ochre). The black dashed lines plot the values of $\frac{1}{\sqrt{\Omega}}$ and $\frac{1}{\Omega}$ in logarithmic scaling.

In order to take the correct decision, one must know the order of the first space derivative of the FPE solution. A future research topic is therefore the analysis of the space derivatives of the FPE. In the numerical examples in the previous chapter, the derivative term changed between order $\mathcal{O}\left(\sqrt{\Omega}\right)$ and $\mathcal{O}(1)$. It should be reviewed, why and under which constraints and assumptions this result occurs and if other convergence orders can be observed for different reaction networks. Further, the values of the two constants C_1 and C_2 should be estimated. It would also be valuable to estimate the order of the convergence error *a priori*.

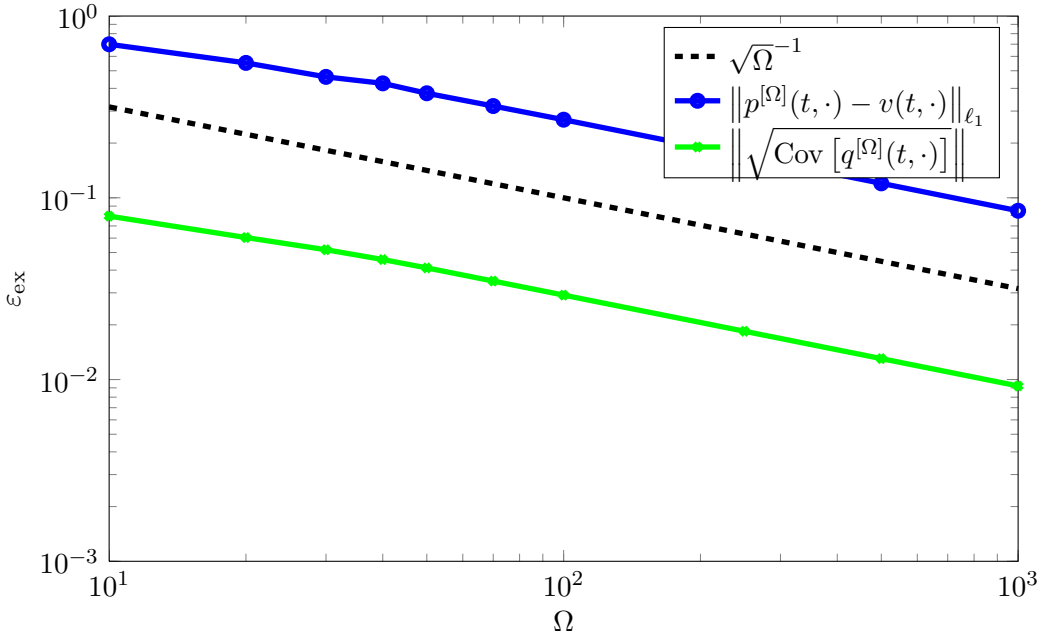


Figure 4.6: Convergence of the left hand side term ε in theorem 4.5 using the solutions of the CME and the FPE for the reaction network given in table 4.1. This term is compared with the order of the standard deviation of the FPE solution.

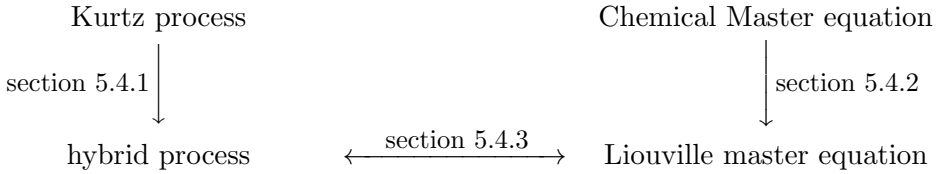
An interesting observation made for the numerical experiment in section 4.4.1 is that the standard deviation behaves with the same order as the convergence order of the FPE. Figure 4.6 plots the convergence error ε of the FPE with respect to Ω. Then the standard deviation $\sqrt{\mathrm{Cov}\left[q(t,x)\right]}$ was calculated for different values of Ω. These two terms are both, for this network, of order $\mathcal{O}\left(\sqrt{\Omega}\right)$. This poses immediately the question, whether the convergence error and the standard deviation are related. Future research should therefore include the question if it is possible to prove

$$\left\|p(t,\cdot)-v(t,\cdot)\right\|_{\ell_1} \overset{?}{\leq} C\left\|\sqrt{\mathrm{Cov}\left[q(t,\cdot)\right]}\right\|_{L_1}?$$

5 | A Hybrid Model Combining the Kurtz process and the RRE

The goal of this chapter is the motivation, derivation and analysis of a hybrid model. We will start with a short introduction and a historical overview in section 5.1. We will then derive and analyse the hybrid model with the help of an example in section 5.2. The notation used in the remainder of this chapter will then be introduced in section 5.3.

In section 5.4 we will present a general derivation of the hybrid model. This section is divided into three sub-sections. First, we will derive the hybrid process from the Kurtz process. Secondly, we will obtain the hybrid model (called Liouville master equation) from the CME. Concluding, we discuss the connection between the hybrid process and PDF. The following sketch visualises this procedure:



We will use some approximations in the derivation of the hybrid model. To verify the validity of this model, we will show convergence and prove an error bound in section 5.6. The hybrid model does not converge for the full distributions, but only for the marginal distributions and conditional expectations. These will be introduced in section 5.5.

5.1 Motivation

We have seen so far that simulating a reaction network is a complex problem, and we have discussed several models used for this task. We also derived error bounds for different models and showed the connection between them. These preliminary studies allow us to address an important class of models that were designed to reduce the simulation complexity of reaction networks that have a scale difference in the particle amounts of the different species. These models are combinations of two different models and we refer to them as hybrid models.

The Kurtz process converges to the RRE in the thermodynamic limit. This observation lends legitimacy to the usage of the RRE for reaction networks where all species have large particle numbers. On the other hand, reaction networks with small particle numbers are highly stochastic and the usage of the Kurtz process is crucial. Otherwise, important features of the underlying distribution could be missed. But if a reaction network contains some species with small particle numbers and some species with large particle numbers, we still have to choose the Kurtz process to simulate the network. However, simulating a network with large particle numbers is a time-consuming problem. High particle numbers lead to high values in the (linear or quadratic) propensity functions. This leads to short waiting times between the jumps of the process. We have seen this also in our discussion of the SSA, large particle numbers often result in an increased fire frequency of several reaction channels. Haseltine and Rawlings came up with the following idea to reduce the computational complexity of networks containing a scale difference in the species [HR02]. They recommended splitting the network into species with small particle numbers and species with large particle numbers. Now the sub-network with small amounts is simulated using the Kurtz process, hence stochastic effects are conserved. The sub-network with large particle numbers is simulated using the RRE. In this way, one can take advantage of the fast computational time of the RRE simulations to reduce the overall computation time.

Based on this initial idea, several authors discussed different hybrid schemes and their implementation. For example, Alfonsi et al. stated a hybrid method which adaptively partitions a network based on the firing frequency of reaction channels [ACT$^+$05]. Their approach differs from the setting in this chapter. The authors split the network by *reactions*. Reaction channels with small fire frequencies are simulated stochastically, while reactions with high fire frequencies are simulated using an ODE. In this thesis, we partition reaction networks based on the particle numbers of *species*. The species that are only present in small copy numbers are simulated stochastically, while species with large particle amounts are handled deterministically. However, the authors state that their approach leads to a CME coupled to a LVE (unfortunately without deriving this equation). We will see in section 5.4.2 that our hybrid model leads also to an equation, which combines the CME with the LVE. Therefore, we may assume that the error bounds proven in section 5.6 may also apply to the "reaction splitting" model. Furthermore, the approach by Alfonsi et al. is using an adaptive partitioning, i.e. changes in the fire frequencies of reaction channels result in a re-partitioning of the network. We partition the network *a priori* and keep the partitioning constant over the complete simulation time. They conclude their work by discussing three possible algorithms that solve this model and testing the speedup by means of several examples. They observe that hybrid simulations can be up to 100 times faster then fully stochastic simulations.

Kiehl and Mattheyses used the same approach and discussed possible implementations of their algorithm [KMS04]. Griffith et al. discussed a dynamic splitting strategy for reaction networks and implemented a Kurtz process / RRE algorithm to demonstrate their ideas [GCPS06]. Higham et al. combined the Kurtz process with the CLE and analysed the resulting hybrid algorithm [HIMS11]. Salis and Kaznessis analysed a sim-

ilar approach [SK05]. Takahashi et al. discussed the implementation details of a Kurtz process / SDE hybrid algorithm and Salis et al. presented a software tool that contains such a method [TKHT04, SSK06]. Recently, Lachor et al. compared several simulation techniques, including a hybrid method, for reaction networks and analysed their performance [LPP11]. A detailed review of hybrid algorithms can be found in [Pah09].

Henzinger et al. derived the same hybrid model analysed in this work and stated an equation of motion of its distribution [HMMW10]. Furthermore, they discussed a simulation algorithm to solve this equation. However, the convergence properties of the model were not analysed.

Several other hybrid models can be found in literature, but differ from the approach analysed here. For example, Hellander et al. used a splitting scheme to reduce the complexity of the CME [HHL08, HL07] and Jahnke showed an error bound for this model [Jah11]. Du and Parise combined the RRE with a moment approximation of the CME to create a hybrid algorithm [DP11]. Kaznessis combined the Kurtz process with a space continuous Markov process [Kaz06]. Chen et al. used a Taylor expansion of the CME to develop a hybrid piecewise deterministic Markov process [CWA09]. Hasenauer et al. and Menz et al. derived and analysed a hybrid model known as "model reduction based on conditional expectations" (MRCE), which uses the product of marginal distributions with conditional expectations (cf. section 5.5) [HWKT14, MLSH12]. Sunkara stated a further numerical method to solve the MRCE and showed an error bound for the model [Sun13]. This proof was extended for networks with a scale difference between the species by Jahnke and Sunkara [JS13]. The MRCE model splits the CME $p(t, n, m)$ in a marginal distribution $p_1(t, n)$ and a conditional distribution $p_2(t, m|n)$, where n denotes the state of the "stochastic / discrete species" and m the state of the "deterministic / continuous species":

$$p(t, n, m) = p_1(t, n)p_2(t, m|n).$$

The conditional distribution is turned into a continuous density on the discrete sub-space of the low copy number species. The two distributions p_1 and p_2 are then solved in a hybrid setting. For example Menz et al. use a CME to solve the marginal distributions p_1 and a differential algebraic equation for each state n to approximate the first moments of p_2 [MLSH12].

Hybrid methods suffer from different kinds of errors. Each of the numerical solvers used to simulate the two sub-networks, produces errors. For example, a histogram of several SSA realisations is an approximation of the CME solution or a numerical ODE integrator only approximates the RRE. Both approximations contain errors. As we will see, the hybrid model is often simulated using splitting methods (e.g. Trotter or Strang splitting scheme), i.e. two (or more [JA10]) components are decoupled, solved individually and then recoupled. In this work we use the Strang splitting scheme. Of course, such a scheme produces a numerical error. Finally there is the model error. This error results from the initial decision to separate the network into two different sub-networks and to use different models for each one. In this chapter, only the modelling error is analysed and we give an

error bound for a hybrid model combining the Kurtz process for species with small particle amounts with the RRE for species with large particle amounts. The analysis is done for the PDFs, therefore we call the model the Liouville master equation (LME), because the model combines the CME with the LVE. A derivation of this equation can be found in the work of Henzinger et al. [HMMW10] and in section 5.4.

Despite the large number of publications on hybrid models that indicate a high interest, only a few studies on the error behaviour can be found. Vasudeva and Bhalloa analysed the error of a Kurtz process / RRE hybrid algorithm numerically with respect to the size of the time step [VB04]. Crudu et al. and Riedler analysed hybrid methods from a stochastic process point of view [CDMR12, Rie10] and showed convergence of the distribution of the hybrid process against the solution of the CME, but the authors did not give a convergence rate.

To show an error bound and a convergence rate, we derive the LME, an equation of motion for the hybrid model, and extend the thermodynamic limit and the scaling concept from the last chapter to hybrid models. We introduce the definition of the partial thermodynamic limit. Next, we discuss in which sense convergence can be expected and give a numerical example. Finally, we prove an error bound and discuss the corresponding convergence rate.

The results of this chapter have already been published in [JK12].

5.2 An Example

Consider the reaction network

$$\begin{aligned} \mathfrak{R}_1: &\quad \mathcal{S}_1 &&\xrightarrow{c_1}& \mathcal{S}_1 + \mathcal{S}_3 \\ \mathfrak{R}_2: &\quad \mathcal{S}_1 + \mathcal{S}_3 &&\xrightarrow{c_2}& \mathcal{S}_2 \\ \mathfrak{R}_3: &\quad \mathcal{S}_2 &&\xrightarrow{c_3}& \mathcal{S}_1 + \mathcal{S}_3 \\ \mathfrak{R}_4: &\quad \mathcal{S}_3 + \mathcal{S}_3 &&\xrightarrow{c_4}& \emptyset. \end{aligned} \tag{5.1}$$

This network represents a simple self-regulating genetic network. We interpret $\mathcal{S}_1$ as a gene that generates a protein $\mathcal{S}_3$ via reaction $\mathfrak{R}_1$. Reaction $\mathfrak{R}_2$ describes the inhibition of the gene via the protein. The inhibited gene is named $\mathcal{S}_2$. The reverse reaction of the inhibition is represented by reaction $\mathfrak{R}_3$ and finally we have a dimerisation reaction. To keep the network simple the dimer is not explicitly stated as a species $\mathcal{S}_4$. The outcome of $\mathfrak{R}_4$ is simply decomposed.

For the network (5.1), the Kurtz process with initial condition $X(0) = X_0 = (10, 0, \Omega)^T$ is given as

$$\begin{aligned} X(t) = X_0 &+ \mathcal{P}_1\left(\int_0^t c_1 X_1(s)\mathrm{d}s\right)\begin{pmatrix}0\\0\\1\end{pmatrix} \\ &+ \mathcal{P}_2\left(\int_0^t c_2 \frac{1}{\Omega} X_1(s)X_3(s)\mathrm{d}s\right)\begin{pmatrix}-1\\1\\-1\end{pmatrix} \end{aligned} \tag{5.2}$$

$$+ \mathcal{P}_3\left(\int_0^t c_3 X_2(s)\mathrm{d}s\right)\begin{pmatrix}1\\-1\\1\end{pmatrix}$$
$$+ \mathcal{P}_4\left(\int_0^t \frac{c_4}{2\Omega} X_3(s)\left(X_3(s)-1\right)\mathrm{d}s\right)\begin{pmatrix}0\\0\\-2\end{pmatrix}.$$

We notice that on bounded time intervals

$$X_1(t) + X_2(t) = 10 \ll \mathbb{E}\left[X_3(t)\right] = \mathcal{O}(\Omega) \text{ for } \Omega \to \infty$$

holds. This scale difference makes the simulation of such reaction networks difficult*, but the application of Haseltine's and Rawlings's idea reduces the numerical complexity. We split the process into two sub-processes (note that the first and 4th reaction $\mathfrak{R}_1$ and $\mathfrak{R}_4$ only depend on the third species $\mathcal{S}_3$):

$$\begin{pmatrix}X_1(t)\\X_2(t)\end{pmatrix} = \begin{pmatrix}10\\0\end{pmatrix} + \mathcal{P}_2\left(\int_0^t c_2 X_1(s)\frac{X_3(s)}{\Omega}\mathrm{d}s\right)\begin{pmatrix}-1\\1\end{pmatrix} + \mathcal{P}_3\left(\int_0^t c_3 X_2(s)\mathrm{d}s\right)\begin{pmatrix}1\\-1\end{pmatrix},$$

$$\begin{aligned}X_3(t) = \Omega &+ \mathcal{P}_1\left(\int_0^t c_1 X_1(s)\mathrm{d}s\right) - \mathcal{P}_2\left(\int_0^t c_2 X_1(s)\frac{X_3(s)}{\Omega}\mathrm{d}s\right)\\ &+ \mathcal{P}_3\left(\int_0^t c_3 X_2(s)\mathrm{d}s\right) - 2\mathcal{P}_4\left(\int_0^t \frac{c_4}{2\Omega} X_3(s)\left(X_3(s)-1\right)\mathrm{d}s\right).\end{aligned}$$

The first two species, $\mathcal{S}_1$ and $\mathcal{S}_2$, should be handled using the Kurtz process, while $\mathcal{S}_3$ with its large copy numbers will be simulated using the RRE.

We find a small time interval $[t_1, t_2]$ where the parameters X_1 and X_2 of the first sub-process are given by the constants $x_1, x_2 \in \mathbb{N}_0$. The next step is motivated by the derivation of the RRE from the expectation of the Kurtz process (cf. section 2.4.1), the law of large numbers and from Kurtz famous "thermodynamic limit theorem" (cf. section 4.3).

We scale X_3 by Ω, assume

$$y(s) \approx \frac{X_3(s)}{\Omega} \approx \frac{X_3(s)-1}{\Omega}$$

and approximate the second sub-process on this time interval by the corresponding RRE (cf. section 2.4.1)

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}y(t) &= \frac{c_1 x_1}{\Omega} - \frac{c_2 x_1}{\Omega}y(t) + \frac{c_3 x_2}{\Omega} - c_4 y^2(t),\\ y(0) &= 1.\end{aligned}$$

*Obviously, this reaction network is very small and the simulation is not that difficult, but we use this network only as an example. We would also like to point out that the cost of simulating only slightly more complex networks quickly grows, e.g. [SYSY02].

In summary, our hybrid process is given by

$$Z(t) = \begin{pmatrix} 10 \\ 0 \end{pmatrix} + \mathcal{P}_2\left(\int_0^t c_2 Z_1(s)y(s)\mathrm{d}s\right)\begin{pmatrix} -1 \\ 1 \end{pmatrix} + \mathcal{P}_3\left(\int_0^t c_3 Z_2(s)\mathrm{d}s\right)\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \tag{5.3}$$

$$y(t) = 1 + \int_0^t \frac{c_1}{\Omega} Z_1(s)\mathrm{d}s - \int_0^t \frac{c_2}{\Omega} Z_1(s)y(s)\mathrm{d}s + \int_0^t \frac{c_3}{\Omega} Z_2(s)\mathrm{d}s - \int_0^t c_4 y^2(s)\mathrm{d}s, \tag{5.4}$$

with $Z_i \approx X_i$ for $i = 1, 2$.

We are now able to compare the two processes (5.2) and (5.3) / (5.4) numerically. We will give a detailed algorithm for the hybrid simulation scheme after the general derivation of the model (cf. algorithm 5.1). Because $X_1(t) + X_2(t) = 10$ by construction of the network, we can reduce the three-dimensional network to a two-dimensional one, where one of the species is is handled using the Kurtz process and the other one by the RRE. Figure 5.1 visualizes the solutions of the CME and of the PDF of the hybrid process. The plot on the left displays the solution of the CME while the right plot shows a histogram of several realisations of the hybrid process for our example. We see that the solution is "compressed" in x direction and does not span the same "width" as the CME solution. This example clearly shows that we cannot expect convergence for the hybrid PDF to the full CME solution.

However, we can expect convergence of the marginal distributions of $\mathcal{S}_1$ and $\mathcal{S}_2$ and of the conditional expectation times the marginal distribution for S_3. These quantities will be defined in chapter 5.5.

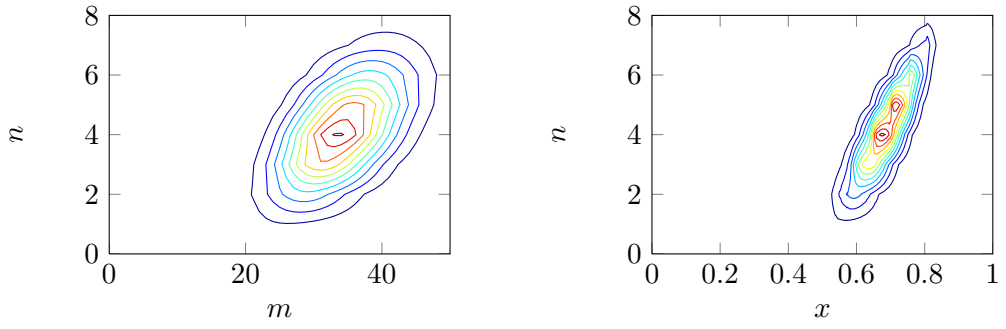


Figure 5.1: Solution of the CME (left) and PDF of the hybrid model (right) for example (5.1) for $\Omega = 50$ and $t = 0.5$.

Figure 5.2 visualises the marginal distributions of species $\mathcal{S}_1$ for the CME and of the PDF of the hybrid process for two different values of Ω. We observe that the difference of the two histograms becomes smaller as Ω becomes larger.

Motivated by these considerations we compute the difference of the marginal disitributions in the ℓ_1-norm for different values of Ω. Figure 5.3 visualises this experiment together with the error in conditional expectations. We see that both errors converge with order $\mathcal{O}\left(\frac{1}{\Omega}\right)$. We will prove this result in chapter 5.6 after a general derivation of the hybrid model and the definition of the different quantities used in this section.

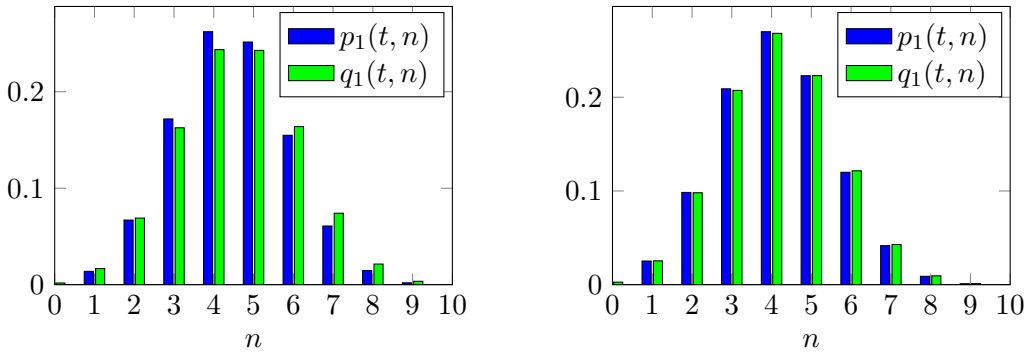


Figure 5.2: Marginal distribution of $\mathcal{S}_1$ of the CME (blue) and of the hybrid PDF (green) for example (5.1) for $\Omega = 10$ (left) and $\Omega = 100$ (right) at time $t = 0.5$.

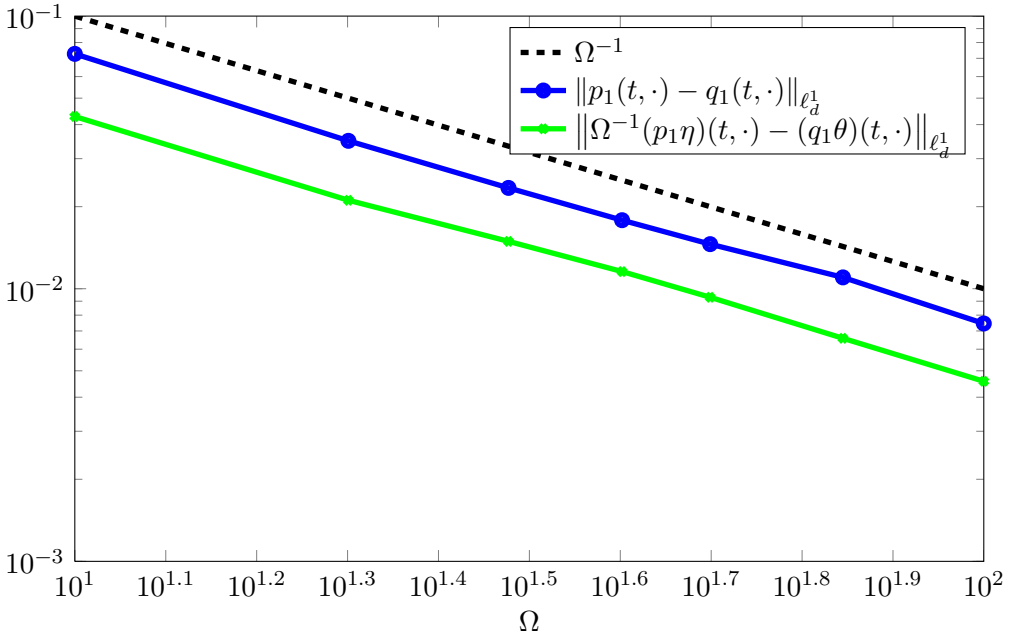


Figure 5.3: Convergence error of the marginal distribution (blue) and conditional expectation (green) for the reaction network in example (5.1) at time $t = 0.5$.

5.3 Notation and Definitions

We examine a reaction network consisting of $N = d + D$, $d, D \in \mathbb{N}$ species and $R \in \mathbb{N}$ reactions:

$$\mathfrak{R}_j : \quad \sum_{i=1}^{d} \kappa_{j,i}^{\text{in}} \mathcal{S}_i + \sum_{i=1}^{D} \lambda_{j,i}^{\text{in}} \mathcal{S}_{i+d} \xrightarrow{c_j} \sum_{i=1}^{d} \kappa_{j,i}^{\text{out}} \mathcal{S}_i + \sum_{i=1}^{D} \lambda_{j,i}^{\text{out}} \mathcal{S}_{i+d} \quad j = 1, \dots, R, \tag{5.5}$$

with the stoichiometric factors $\kappa_{j,i}^{\text{in}}, \kappa_{j,i}^{\text{out}}, \lambda_{j,i}^{\text{in}}, \lambda_{j,i}^{\text{out}} \in \mathbb{N}_0$.

We name the species $\mathcal{S}_1$ to $\mathcal{S}_d$ stochastic or discrete species and represent them by $n \in \mathbb{N}_0^d$. The species $\mathcal{S}_{d+1}$ to $\mathcal{S}_D$ are deterministic species and their particle counts are stored in $m \in \mathbb{N}_0^D$. Later, the deterministic species are supposed to be represented by continuous variables $x \in \mathbb{R}_{0,+}^D$ and will be referred to as continuous species.

We define the stoichiometric vectors

$$\nu_j = \begin{pmatrix} \kappa_{j,1}^{\text{out}} - \kappa_{j,1}^{\text{in}} \\ \vdots \\ \kappa_{j,d}^{\text{out}} - \kappa_{j,d}^{\text{in}} \end{pmatrix} \in \mathbb{Z}^d, \qquad \mu_j = \begin{pmatrix} \lambda_{j,1}^{\text{out}} - \lambda_{j,1}^{\text{in}} \\ \vdots \\ \lambda_{j,D}^{\text{out}} - \lambda_{j,D}^{\text{in}} \end{pmatrix} \in \mathbb{Z}^D. \tag{5.6}$$

We define two index sets

$$J_0 = \left\{ j \in \{1, \dots, R\} : \nu_j = (0, \dots, 0)^T \right\}, \tag{5.7}$$

$$J_1 = \{1, \dots, R\} \setminus J_0, \tag{5.8}$$

and the indicator function

$$\gamma(j) = \begin{cases} 0, & \text{if } j \in J_0 \\ 1, & \text{if } j \in J_1 \end{cases} \tag{5.9}$$

to distinguish if a reaction $\mathfrak{R}_j$ changes the stochastic species. We note that the case $j \in J_0$ does not mean that the reaction is independent of n. These species could still act as a catalyst, but their quantities are not changed.

The propensity function for each reaction channel $\mathfrak{R}_j$ is of the form $\alpha(n)\beta(m)$, with α, β defined as follows:

Definition 5.1 (Hybrid Propensity Function)**.**
The propensity functions of a hybrid reaction network are

$$\alpha_j : \mathbb{N}_0^d \to \mathbb{R}_{0,+}, \qquad \alpha_j(n) := c_j \prod_{i=1}^{d} \binom{n_i}{\kappa_{j,i}^{\text{in}}},$$

$$\beta_j : \mathbb{N}_0^D \to \mathbb{R}_{0,+}, \qquad \beta_j(m) := \Omega^{(1-\gamma(j))} \Omega^{-\left|\lambda_j^{\text{in}}\right|_1} \prod_{i=1}^{D} \binom{m_i}{\lambda_{j,i}^{\text{in}}},$$

$$\tilde{\beta}_j : \mathbb{R}_{0,+}^D \to \mathbb{R}_{0,+}, \qquad \tilde{\beta}_j(x) := \Omega^{(1-\gamma(j))} \prod_{i=1}^{D} \frac{x_i^{\lambda_{j,i}^{\text{in}}}}{\lambda_{j,i}^{\text{in}}!},$$

with $\left|\lambda_j^{\text{in}}\right|_1 = \sum_{i=1}^{D} \lambda_{j,i}^{\text{in}}$.

Definition 5.2 (Spaces and Norms for Hybrid Systems)**.**
Let

$$\ell_K^1 = \left\{ u : \mathbb{N}_0^K \longrightarrow \mathbb{R}^N \;\middle|\; \sum_{n \in \mathbb{N}_0^K} \|u(n)\| < \infty \right\}, \qquad \|u\|_{\ell_K^1} = \sum_{n \in \mathbb{N}_0^K} \|u(n)\|$$

be the multivariate and vector-valued version of the ℓ_1 space, with $K \in \mathbb{N}$ and $\|\cdot\|$ an arbitrary norm on $\mathbb{R}^N$.

Further, we define the spaces

$$\mathcal{X}_{d,D}^{i+1} = \left\{ u \in \mathcal{X}_{d,D}^i \middle| (n,m) \mapsto m_k u(n,m) \in \mathcal{X}_{d,D}^i \text{ for all } k \in \{1, \dots, D\} \right\}$$

via recursion, with $i \in \mathbb{N}_0$ and $\mathcal{X}^0_{d,D} = \ell^1_{d+D}$, Analogously, for functions on a discrete-continuous state space we define the spaces

$$\mathcal{Y}^0_{d,D} = \left\{ u : \mathbb{N}_0^d \times \mathbb{R}_+^D \longrightarrow \mathbb{R}^N \;\middle|\; \sum_{n \in \mathbb{N}_0^d} \int_{\mathbb{R}_+^D} \|u(n,x)\| \, dx < \infty \right\},$$
$$\mathcal{Y}^{i+1}_{d,D} = \left\{ u \in \mathcal{Y}^i_{d,D} \;\middle|\; (n,x) \mapsto x_k u(n,x) \in \mathcal{Y}^i_{d,D} \text{ for all } k \in \{1,\dots,D\} \right\}.$$

After these definitions, we derive now the process and PDF of the hybrid model. We start by defining a scaled and partitioned Kurtz process:

Definition 5.3 (Scaled and Partitioned Kurtz process and CME)**.**
The Kurtz process for the scaled and partitioned reaction network 5.5 is given by

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} n_0 \\ \Omega m_0 \end{pmatrix} + \sum_{j=1}^{R} \mathcal{P}_j \left(\int_0^t \alpha_j(X(s)) \beta_j(Y(s)) \mathrm{d}s \right) \begin{pmatrix} \nu_j \\ \mu_j \end{pmatrix},$$

with initial conditions $n_0 \in \mathbb{N}_0^d$ and $m_0 \in \mathbb{N}_0^D$. This Kurtz process has the PDF

$$p(t,n,m) = \mathbb{P}\left(X(t) = n, Y(t) = m \mid X(0) = n_0, Y(0) = \Omega m_0 \right)$$

which is the solution of the scaled and partitioned CME

$$\partial_t p(t,n,m) = \sum_{j=1}^{R} \left(\alpha_j(n-\nu_j)\beta_j(m-\mu_j) p(t, n-\nu_j, m-\mu_j) - \alpha_j(n)\beta_j(m) p(t,n,m) \right), \tag{5.10}$$
$$p(0,n,m) = \delta_{n_0}(n)\delta_{\Omega m_0}(m) = \begin{cases} 1, & \text{if } n = n_0 \text{ and } m = \Omega m_0, \\ 0, & \text{otherwise.} \end{cases},$$

with the convention that

$$\alpha_j(n-\nu_j)\beta_j(m-\mu_j) p(t, n-\nu_j, m-\mu_j) = 0, \qquad \text{if } n - \nu_j \notin \mathbb{N}_0^d \text{ or } m - \mu_j \notin \mathbb{N}_0^D.$$

To abbreviate the notation and make the forthcoming proofs more readable and compact, we define the shift operators:

Definition 5.4 (Shift Operators)**.**
We define the shift operators

$$\Delta_\nu u(n,m) = \begin{cases} u(n-\nu, m) & \text{if } n - \nu \in \mathbb{N}_0^d, \\ 0 & \text{otherwise,} \end{cases}$$
$$\Delta_\mu u(n,m) = \begin{cases} u(n, m-\mu) & \text{if } m - \mu \in \mathbb{N}_0^D, \\ 0 & \text{otherwise,} \end{cases}$$

with $\nu \in \mathbb{Z}^d$ and $\mu \in \mathbb{Z}^D$. They have the following properties

- commutativity, i.e.

$$\Delta_\nu \Delta_\mu u(n,m) = \Delta_\mu \Delta_\nu u(n,m) = \begin{cases} u(n-\nu, m-\mu) & \text{if } n-\nu \in \mathbb{N}_0^d, m-\mu \in \mathbb{N}_0^D \\ 0 & \text{else.} \end{cases}$$

- $(\Delta_\nu uv)(n,m) = (\Delta_\nu(uv))(n,m) = u(n-\nu,m)v(n-\nu,m) = (\Delta_\nu u)(\Delta_\nu v)(n,m)$.

Lemma 5.5 (Properties of the Shift Operators)**.**
Let $u \in \ell^1_{d+D}$ be a function with

$$u : \mathbb{N}_0^d \times \mathbb{N}_0^D \to \mathbb{R},$$

and $u(n,m) = 0$ for the cases $n+\nu \notin \mathbb{N}_0^d$ or $m+\mu \notin \mathbb{N}_0^D$, with $\nu \in \mathbb{Z}^d$ and $\mu \in \mathbb{Z}^D$.
Then it holds that:

$$\begin{aligned} (i) \quad & \sum_{n\in\mathbb{N}_0^d} ((\Delta_\nu - 1)u)(n,m) &&= 0 \\ (ii) \quad & \sum_{m\in\mathbb{N}_0^D} ((\Delta_\mu - 1)u)(n,m) &&= 0 \\ (iii) \quad & \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} ((\Delta_\nu\Delta_\mu - 1)u)(n,m) &&= 0. \end{aligned}$$

Further, if u is scalar-valued and $u \in \mathcal{X}^1_{d,D}$, it holds that

$$(iv) \quad \sum_{m\in\mathbb{N}_0^D} m\,((\Delta_\mu - 1)u)(n,m) = \mu \sum_{m\in\mathbb{N}_0^D} u(n,m).$$

Remark: The two operators allow a reformulation of the CME:

$$\partial_t p = \sum_{j=1}^{R} (\Delta_{\nu_j}\Delta_{\mu_j} - 1)\alpha_j\beta_j p.$$

Proof. A rigorous proof of this lemma can be found in [Eng09, Jah10], but we summarise the main idea here:

(i) Let $\tilde{n} = n - \nu$, then

$$\begin{aligned} \sum_{n\in\mathbb{N}^d} ((\Delta_\nu - 1)u)(n,m) &= \sum_{n\in\mathbb{N}^d} u(n-\nu,m) - \sum_{n\in\mathbb{N}^d} u(n,m) \\ &= \sum_{\tilde{n}+\nu\in\mathbb{N}^d} u(\tilde{n},m) - \sum_{n\in\mathbb{N}^d} u(n,m) = 0. \end{aligned}$$

(ii) The second statement can be shown by using the same argument as in (i).

(iii) Rearrangement and applying of (i) and (ii) gives

$$\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} (\Delta_\nu\Delta_\mu - 1)u = \sum_{n\in\mathbb{N}_0^d} \Delta_\nu \underbrace{\sum_{m\in\mathbb{N}_0^D} (\Delta_\mu - 1)\,u}_{=0 \text{ by (ii)}} + \sum_{m\in\mathbb{N}_0^D} \underbrace{\sum_{n\in\mathbb{N}_0^d} (\Delta_\nu - 1)\,u}_{=0 \text{ by (i)}}$$

(iv) Let $\tilde{m} = m - \mu$:

$$\begin{aligned}\sum_{m\in\mathbb{N}_0^D} m\left((\Delta_\mu - 1)u\right)(n,m) &= \sum_{m\in\mathbb{N}_0^D} mu(n,m-\mu) - \sum_{m\in\mathbb{N}_0^D} mu(n,m)\\ &= \sum_{\tilde{m}+\mu\in\mathbb{N}_0^D} (\tilde{m}+\mu)u(n,\tilde{m}) - \sum_{m\in\mathbb{N}_0^D} mu(n,m)\\ &= \mu \sum_{m\in\mathbb{N}_0^D} u(n,m).\end{aligned}$$

□

Lemma 5.6 (The solution of the partitioned CME is a PDF)**.**
If $\alpha_j\beta_j p \in \ell^1_{d+D}$ it follows that the solution $p(t,n,m)$ of the partitioned CME is a PDF at any time $t > 0$.

Proof.

$$\begin{aligned}\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} p(t,n,m) &= \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} p(0,n,m) + \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\int_0^t \partial_t p(s,n,m)\mathrm{d}s\\ &= 1 + \int_0^t \sum_{j=1}^R \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} \left((\Delta_{\nu_j}\Delta_{\mu_j} - 1)\alpha_j\beta_j p\right)(s,n,m)\mathrm{d}s\\ &= 1.\end{aligned}$$

It can then be shown that $p(t,n,m) \geq 0$ [Jah11, ch. 2.4]. □

5.4 Derivation of the Hybrid Model

5.4.1 Derivation of the Hybrid Process

To derive the hybrid process we start by motivating two approximations

$$Z(t) \approx X(t) \qquad\qquad \hat{Y}(t) \approx \mathbb{E}\left[Y(t)\right].$$

These steps are inspired by our derivation of the RRE (cf. section 2.4.1 and cf. section 4.3) and are similar to the example in section 5.2. The process (cf. def 5.3) will be scaled by $\frac{1}{\Omega}$ and the second sub-process is, addionaly, approximated by the RRE.

We assume

$$y(t) \approx \frac{\hat{Y}(t)}{\Omega},$$

and approximate the propensity function β by its continuous counterpart $\tilde{\beta}$:

$$\beta_j\left(\hat{Y}(t)\right) \approx \tilde{\beta}_j\left(\frac{\hat{Y}(t)}{\Omega}\right) \approx \tilde{\beta}_j\left(y(t)\right).$$

We apply these two assumptions to the original process and end up with the Kurtz / RRE hybrid process:

$$Z(t) = n_0 + \sum_{j=1}^{R} \mathcal{P}_j \left(\int_0^t \alpha_j \left(Z(s) \right) \tilde{\beta}_j \left(y(s) \right) \mathrm{d}s \right) \nu_j, \tag{5.11}$$

$$y(t) = m_0 + \frac{1}{\Omega} \sum_{j=1}^{R} \left(\int_0^t \alpha_j(Z(s)) \tilde{\beta}_j(y(s)) \mathrm{d}s \right) \mu_j. \tag{5.12}$$

5.4.2 Derivation of the Liouville master equation

It is also possible to derive the LME as an approximation of the CME by performing a partial Kramers-Moyal expansion on the partitioned CME with respect to the correct scaling.

The density $q\left(t, n, x\right)$ is the solution of the LME, if

$$\int_S q\left(t, n, x\right) \mathrm{d}x = \mathbb{P}\left(Z(t) = n, y(t) \in S \mid Z(0) = n_0, y(0) = m_0 \right) \tag{5.13}$$

for all measurable sets $S \subseteq \mathbb{R}_+^D$.

To derive the LME we start with the partitioned and scaled CME, scale by Ω and integrate over the D-dimensional unit cube $\mathcal{W}(0) = [0,1)^D$, to transform the argument $m \in \mathbb{N}^D$ into a real valued variable $x \in \mathbb{R}_{0,+}^D$ and the PDF $p(t,n,m)$ into $q(t,n,x)$. At this point of the derivation we will not really get a real vector $x \in \mathbb{R}^D$ but a rational number $\frac{m}{\Omega} \in \mathbb{Q}^D$. Later, this will be approximated as a real vector.

We apply the previously discussed strategy to equation 5.13 and get

$$\begin{aligned}
p(t,n,m) &= \mathbb{P}\left(X(t) = n, \frac{Y(t)}{\Omega} = \frac{m}{\Omega} \,\middle|\, X(0) = n_0, Y(0) = m_0 \right) \\
&= \mathbb{P}\left(X^{[\Omega]}(t) = n, \frac{Y^{[\Omega]}(t)}{\Omega} \in \frac{m + \mathcal{W}(0)}{\Omega} \,\middle|\, X^{[\Omega]}(0) = n_0, Y^{[\Omega]}(0) = m_0 \right) \\
&=: \int_{\frac{m+\mathcal{W}(0)}{\Omega}} \tilde{q}\left(t, n, x\right) \mathrm{d}x = \int_{\frac{m+\mathcal{W}(0)}{\Omega}} \tilde{q}\left(t, n, \frac{m}{\Omega}\right) \mathrm{d}x = \Omega^{-D} \cdot \tilde{q}\left(t, n, \frac{m}{\Omega}\right)
\end{aligned}$$

where the new probability density function should be understood as

$$\tilde{q}\left(t, n, \frac{m}{\Omega}\right) : \mathbb{R}_{0,+} \times \mathbb{N}_0^d \times \frac{\mathbb{N}_0^D}{\Omega} \to \mathbb{R}_{0,+}.$$

We apply the newly derived connection $\Omega^D \cdot p(t,n,m) = \tilde{q}\left(t, n, \frac{m}{\Omega}\right)$ to the partitioned and scaled CME

$$\begin{aligned}
\partial_t \tilde{q}\left(t, n, \frac{m}{\Omega}\right) &= \Omega^D \partial_t p(t,n,m) \\
&= \sum_{j=1}^{R} \Big(\alpha_j(n - \nu_j) \beta_j(m - \mu_j) \Omega^D p(t, n - \nu_j, m - \mu_j)
\end{aligned}$$

$$
\begin{aligned}
&\quad - \alpha_j(n)\beta_j(m)\Omega^D p(t,n,m)\Big) \\
&= \sum_{j=1}^{R} \Big(\alpha_j(n-\nu_j)\beta_j(m-\mu_j)\tilde{q}\left(t, n-\nu_j, \frac{m-\mu_j}{\Omega}\right) \\
&\quad - \alpha_j(n)\beta_j(m)\tilde{q}\left(t,n,\frac{m}{\Omega}\right)\Big),
\end{aligned}
$$

and replace the propensity function β by its continuous counterpart

$$
\begin{aligned}
\partial_t \tilde{q}\left(t,n,\frac{m}{\Omega}\right) \approx \sum_{j=1}^{R} \Big(&\alpha_j(n-\nu_j)\tilde{\beta}_j\left(\frac{m-\mu_j}{\Omega}\right)\tilde{q}\left(t,n-\nu_j,\frac{m-\mu_j}{\Omega}\right) \\
&- \alpha_j(n)\tilde{\beta}_j\left(\frac{m}{\Omega}\right)\tilde{q}\left(t,n,\frac{m}{\Omega}\right)\Big).
\end{aligned}
$$

Now we move to the real numbers and assume

$$
x \approx \frac{m}{\Omega},
$$

to obtain

$$
\partial_t \tilde{q}\left(t,n,x\right) \approx \sum_{j=1}^{R} \Big(\alpha_j(n-\nu_j)\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\tilde{q}\left(t,n-\nu_j,x-\frac{\mu_j}{\Omega}\right) - \alpha_j(n)\tilde{\beta}_j\left(x\right)\tilde{q}\left(t,n,x\right)\Big).
$$

Next, we expand by adding terms that sum up to zero

$$
\begin{aligned}
\partial_t \tilde{q}\left(t,n,x\right) \approx \sum_{j=1}^{R} \Big(&+ \alpha_j(n-\nu_j)\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\tilde{q}\left(t,n-\nu_j,x-\frac{\mu_j}{\Omega}\right) \\
&- \alpha_j(n)\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\tilde{q}\left(t,n,x-\frac{\mu_j}{\Omega}\right) \\
&+ \alpha_j(n)\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\tilde{q}\left(t,n,x-\frac{\mu_j}{\Omega}\right) \\
&- \alpha_j(n)\tilde{\beta}_j(x)\tilde{q}\left(t,n,x\right)\Big).
\end{aligned}
$$

We note that the first two lines in the sum are a CME scaled by $\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)$ and we transform the second part of the sum into a LVE by performing a Kramers-Moyal expansion.

First, we assume

$$
\left[x \mapsto \tilde{\beta}_j(x)\tilde{q}(t,n,x)\right] \in \mathcal{C}^\infty
$$

and perform the Kramers-Moyal expansion

$$
\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\tilde{q}\left(t,\cdot,x-\frac{\mu_j}{\Omega}\right) - \tilde{\beta}_j(x)\tilde{q}(t,\cdot,x) = -\frac{1}{\Omega}\nabla\left(\tilde{\beta}_j(x)\tilde{q}(t,\cdot,x)\right)^T \mu_j + \dots.
$$

We insert the result (note that some terms vanish due to different signs),

$$
\partial_t \tilde{q}\left(t,n,x\right) \approx \sum_{j=1}^{R} \Big(\alpha_j(n-\nu_j)\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\tilde{q}\left(t,n-\nu_j,x-\frac{\mu_j}{\Omega}\right)
$$

$$- \alpha_j(n)\tilde{\beta}_j\left(x - \frac{\mu_j}{\Omega}\right)\tilde{q}\left(t, n, x - \frac{\mu_j}{\Omega}\right)$$
$$- \alpha_j(n)\frac{1}{\Omega}\nabla\left(\tilde{\beta}_j(x)\tilde{q}(t, n, x)\right)^T \mu_j\Big).$$

We now assume that

$$x - \frac{\mu_j}{\Omega} \approx x,$$

in order to get rid of the shift term in the CME part of the LME. This assumption is motivated by the observation that $x_i \approx \frac{m_i}{\Omega} \gg \frac{\mu_{ji}}{\Omega}$ because $m_i \gg \mu_{ji}$ for all $i \in \{1, \ldots, D\}$ and $j \in \{1, \ldots, R\}$. This shift term arises from the two additional terms that sum up to zero and that were used to perform the Kramers-Moyal expansion. We could do this in another way but then the LVE part would be scaled by $\alpha_j(n - \nu_j)$ instead of $\alpha_j(n)$. It is more convenient to assume that the shift in the continuous variable can be neglected than to do the same in the discrete variable.

In a last step we rearrange the equation and note that the CME part becomes zero if $j \in J_0$ and motivated by these derivations we define:

Definition 5.7 (Liouville master equation)**.**
The PDF $q\,(t, n, x)$ is transported by the Liouville master equation (LME)

$$\begin{aligned}\partial_t q\,(t, n, x) = &\sum_{j \in J_1} \tilde{\beta}_j(x)\Big(\alpha_j(n - \nu_j)q\,(t, n - \nu_j, x) - \alpha_j(n)q\,(t, n, x)\Big) \\ &- \frac{1}{\Omega}\sum_{j=1}^{R} \alpha_j(n)\nabla_x\left(\tilde{\beta}_j(x)q\,(t, n, x)\right)^T \mu_j \\ q(0, n, x) = &\,\delta_{n_0}(n)q_0(x),\end{aligned}$$

with propensity functions $\alpha_j, \tilde{\beta}_j$ defined by def. 5.1, $n \in \mathbb{N}_0^d$, $x \in \mathbb{R}_{0,+}^D$, $t \geq 0$ and initial condition $q_0(x)$ defined by def. 5.8.

The initial condition of the LME deserves a discussion. Equation 5.13 implies that the initial condition is a Dirac delta peak located at (n_0, m_0). To avoid the question if a solution exists in this case and to circumnavigate numerical and analytical difficulties, we define the initial condition of the LME as a smooth function with the properties:

Definition 5.8 (Initial Condition of the LME)**.**
Let

$$q_0 : \mathbb{R}_{0,+}^D \to \mathbb{R}_{0,+}$$

be the initial condition of the LME, with the properties

- $q_0(x) = 0$ if $\|x - m_0\|_\infty > \varepsilon$,
- $\int_{\mathbb{R}_+^D} q_0(x) = 1$,
- and $\int_{\mathbb{R}_+^D} x q_0(x) = m_0$,

and a small constant $\varepsilon \in \mathbb{R}_+$ with the property $\varepsilon \ll 1$.

We used and discussed a similar initial condition in section 2.3.2, after the introduction of the Fokker-Planck equation.

5.4.3 Connection of Hybrid Process and LME

So far, we have seen that the hybrid process (5.11) / (5.12) can be derived from the Kurtz process and that the LME can be derived from the CME.

We discuss now if the solution of the LME is the PDF of the hybrid process. From lemma 2.3 follows that the PDF of the first sub-process $Z(t)$ is characterised by

$$\begin{aligned}\partial_t q^{(1)}(t,n,x) &= \left(\mathcal{A}^{(1)} q^{(1)}\right)(t,n,x) \\ &= \sum_{j\in J_1} \tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j) q^{(1)}\left(t,n-\nu_j,x\right) - \alpha_j(n) q^{(1)}\left(t,n,x\right)\Big)\end{aligned}$$

and from lemma 2.6 follows that the PDF of the second sub-process $y(t)$ is specified by

$$\begin{aligned}\partial_t q^{(2)}(t,n,x) &= \left(\mathcal{A}^{(2)} q^{(2)}\right)(t,n,x) \\ &= -\frac{1}{\Omega}\sum_{j=1}^{R} \alpha_j(n)\nabla_x\left(\tilde{\beta}_j(x) q^{(2)}\left(t,n,x\right)\right)^T \mu_j.\end{aligned}$$

We examine now a Lie-Trotter splitting scheme for the hybrid process:

$$\begin{pmatrix} Z(t_0) \\ y(t_0)\end{pmatrix} \xrightarrow{\Delta t} \begin{pmatrix} Z(t_1) \\ y(t_0)\end{pmatrix} \xrightarrow{\Delta t} \begin{pmatrix} Z(t_1) \\ y(t_1)\end{pmatrix}.$$

The sub-process $Z(t)$ is simulated first on the time interval $[t_0, t_1 = t_0 + \Delta t]$, while $y(t_0)$ is held constant. Then, the value of $Z(t_1)$ is understood as a constant for the process $y(t)$ which is now simulated on the same time interval. We examine now the PDF $\hat{q}$ of our hybrid model at three time points:

- t_{00} is the time at the beginning of the interval,
- t_{10} is the time when the first process Z was simulated but y was not,
- t_{11} is the end of the interval:

$$\hat{q}\left(t_{00},\cdot,\cdot\right) \xrightarrow{\Delta t} \hat{q}\left(t_{10},\cdot,\cdot\right) \xrightarrow{\Delta t} \hat{q}\left(t_{11},\cdot,\cdot\right).$$

We use the two operators $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$ to calculate the solutions at these three time points

$$\begin{aligned}\hat{q}\left(t_{10},\cdot,\cdot\right) &= e^{\Delta t \mathcal{A}^{(1)}} \hat{q}\left(t_{00},\cdot,\cdot\right), \\ \hat{q}\left(t_{11},\cdot,\cdot\right) &= e^{\Delta t \mathcal{A}^{(2)}} \hat{q}\left(t_{10},\cdot,\cdot\right),\end{aligned} \quad \Rightarrow \hat{q}\left(t_{11},\cdot,\cdot\right) = e^{\Delta t \mathcal{A}^{(2)}} e^{\Delta t \mathcal{A}^{(1)}} \hat{q}\left(t_{00},\cdot,\cdot\right).$$

Finally, we choose a fixed t, define $\Delta t = \frac{t}{M}$ and take the limits $M \to \infty$ and $\Delta t \to 0$. Now, we use the Trotter product formula to derive,

$$\hat{q}\,(t,\cdot,\cdot) = \left(e^{\Delta t \mathcal{A}^{(2)}} e^{\Delta t \mathcal{A}^{(1)}}\right)^M \hat{q}\,(t_{00},\cdot,\cdot) \xrightarrow[M\to\infty]{} e^{\Delta t(\mathcal{A}^{(2)}+\mathcal{A}^{(1)})}\hat{q}\,(t_{00},\cdot,\cdot)$$

These steps should be understood as a *motivation* why the two operators of the two sub-processes can be added to the LME

$$\begin{aligned}\partial_t q\,(t,n,x) &= \sum_{j\in J_1} \tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\,(t,n-\nu_j,x) - \alpha_j(n)q\,(t,n,x)\Big) \\ &\quad - \frac{1}{\Omega}\sum_{j=1}^{R} \alpha_j(n)\nabla_x\left(\tilde{\beta}_j(x)q\,(t,n,x)\right)^T \mu_j \\ q(0,n,x) &= \delta_{n_0}(n)q_0(x).\end{aligned}$$

The Trotter product formula is discussed in detail in [EN00, chapter III, Corollary 5.8]. This reference lists also all prerequisites for the usage of the formula (especially regularity assumptions).

We have now seen how to construct a stochastic / deterministic hybrid model from the Kurtz process. Also we have seen how the LME can be derived from this process and how it is associated with the CME.

We conclude this section by specifying algorithm 5.1, which simulates the hybrid process (5.11) / (5.12) and returns one trajectory of the process. The algorithm is based on the Strang splitting scheme that splits the problem into two sub-problems, solves them on small time intervals and combines the results. The two sub-problems are our two sub-processes. We propose the usage of SSA to solve the Kurtz sub-process but the usage of other solvers (e.g. τ-leaping) is possible. The function `ODE_INTEGRATOR` denotes any suitable numerical ODE integrator for this problem. We like to point out that this scheme returns only an approximation of the hybrid process (5.11) / (5.12) at certain time points. And we remember that the total error of any simulation performed using the methods proposed in this chapter depends not only on the *model* error, for which we will give an error bound in chapter 5.6, but also on the *numerical* error induced by algorithm 5.1. The analysis of the numerical error is out of the scope of this work, but a discussion can be found in [JA10]

5.5 Marginal Distributions and Conditional Expectations

We expected in our analysis of the example in chapter 5.2 convergence in the marginal distributions and conditional expectations of the LME to the CME. However, we did not define these terms so far.

The marginal distribution of the discrete species is defined as

$$p_1(t,n) = \sum_{m\in\mathbb{N}_0^D} p(t,n,m),$$

Algorithm 5.1 A Sampling Algorithm for the Hybrid process

Require: initial condition n_0 and m_0, time interval $[t_0,\ t_{\text{final}}]$, stoichiometric vectors $\nu_j, \mu_j,\ j = 1, \dots, R$, propensity functions $\alpha_j(n), \tilde{\beta}_j(x)\ j = 1, \dots, R$, number of time steps $K \in \mathbb{N}$, scaling parameter Ω

$\Delta t = \frac{t_{\text{final}} - t_0}{K}$
$Z(0) \leftarrow n_0$
$y(0) \leftarrow m_0$
for $i \leftarrow 0, \dots, K$ **do**
  $t_i \leftarrow t_0 + i \cdot \Delta t$
  $\hat{y} \leftarrow y(t_i)$
  $\hat{z} \leftarrow \texttt{SSA}\left[Z(t_i) + \sum_{j=1}^{R} \mathcal{P}_j\left(\int_{t_i}^{t_i + \frac{\Delta t}{2}} \alpha_j\left(Z(s)\right) \tilde{\beta}_j\left(\hat{y}\right) \mathrm{d}s\right) \nu_j\right]$
  $\hat{y} \leftarrow \texttt{ODE\_INTEGRATOR}\left[y(t_i) + \frac{1}{\Omega} \sum_{j=1}^{R} \left(\int_{t_i}^{t_i + \Delta t} \alpha_j(\hat{z}) \tilde{\beta}_j(y(s)) \mathrm{d}s\right) \mu_j\right]$
  $\hat{z} \leftarrow \texttt{SSA}\left[\hat{z} + \sum_{j=1}^{R} \mathcal{P}_j\left(\int_{t_i + \frac{\Delta t}{2}}^{t_i + \Delta t} \alpha_j\left(Z(s)\right) \tilde{\beta}_j\left(\hat{y}\right) \mathrm{d}s\right) \nu_j\right]$
  $Z(t_{i+1}) = \hat{z}$
  $y(t_{i+1}) = \hat{y}$
end for
return $Z(t_i), y(t_i)$ for $t_i = t_0 + i \cdot \Delta t,\ i = 0, \dots, K$

with p the solution of the CME. The corresponding marginal distribution of the LME solution q is given by

$$q_1(t, n) = \int_{\mathbb{R}^D_{0,+}} q(t, n, x)\, \mathrm{d}x.$$

Further, we define the conditional probability of state m given state n, for the CME solution

$$p_2(t, m|n) = \begin{cases} \frac{p(t,n,m)}{p_1(t,n)}, & \text{if } p_1(t, n) > 0, \\ 0, & \text{otherwise} \end{cases}$$

and for the LME solution

$$q_2(t, x|n) = \begin{cases} \frac{q(t,n,x)}{q_1(t,n)}, & \text{if } q_1(t, n) > 0 \\ 0, & \text{else,} \end{cases}$$

respectively. Based on these definitions we define the conditional expectations

$$\eta(t, n) = \sum_{m \in \mathbb{N}_0^D} m p_2(t, m|n),$$
$$\theta(t, n) = \int_{\mathbb{R}^D_{0,+}} x q_2(t, x|n)\, \mathrm{d}x,$$

and the conditional covariance matrices

$$\sigma(t, n) = \sum_{m \in \mathbb{N}_0^D} \left(m - \eta(t, n)\right) \left(m - \eta(t, n)\right)^T p_2(t, m|n),$$

$$\varsigma(t,n) = \int_{\mathbb{R}^D_{0,+}} (x-\theta(t,n))\,(x-\theta(t,n))^T\, q_2(t,x|n)\,\mathrm{d}x.$$

5.6 An Error Bound

Before we give the main theorem of this chapter, we have to make a few assumptions and show two lemmata that are needed in the proof.

Assumption 5.9 (Bound of the Stoichiometric Factor)**.**
We assume that

$$\left|\lambda_j^{\mathrm{in}}\right|_1 \leq 2,\ \forall j = 1,\dots,R.$$

Assumption 5.10 (Scale Difference of a Reaction Network)**.**
We assume a scale difference between the particle amounts of the first d and the last D species, such that for $\Omega \to \infty$

$$\begin{aligned}
&\mathbb{E}\left[X_i(t)\right] \approx \mathbb{E}\left[Z_i(t)\right] = \mathcal{O}(1)\,, && \forall i \in [1,\dots,d]\,,\\
&\mathbb{E}\left[Y_j(t)\right] \approx \Omega\mathbb{E}\left[y_j(t)\right] = \mathcal{O}(\Omega)\,, && \forall j \in [d+1,\dots,D]\,,\\
&\mathbb{E}\left[X_i(t)\right] \ll \mathbb{E}\left[Y_j(t)\right].
\end{aligned}$$

Assumption 5.11 (Solution of CME and LME)**.**
We assume that the CME 5.3 has a unique classical solution $p(t,\cdot,\cdot) \in \mathcal{X}^3_{d,D}$ for $t \in [0,t_{\mathrm{end}}]$ and that

$$(n,m) \mapsto \alpha_j(n)p(t,n,m) \in \mathcal{X}^3_{d,D}\ \forall j \in \{1,\dots,r\}.$$

Further, we assume that the LME (cf. def. 5.7) with initial condition given by def. 5.8 has a unique classical solution $q(t,\cdot,\cdot) \in \mathcal{Y}^3_{d,D}$ for $t \in [0,t_{\mathrm{end}}]$ and that

$$(n,x) \mapsto \alpha_j(n)q(t,n,x) \in \mathcal{Y}^3_{d,D}\ \forall j \in \{1,\dots,r\}.$$

Remark: This assumption guarantees the existence of $\eta(t,n)$, $\sigma(t,n)$, $\theta(t,n)$, and $\varsigma(t,n)$ for all $n \in \mathbb{N}_0^d$ and $t \in [0,t_{\mathrm{end}}]$.

Assumption 5.12 (Bound for the Conditional Moments)**.**
There is a constant $C > 0$ such that

$$\|\eta(t,n)\| \leq C\cdot\Omega, \qquad \|\theta(t,n)\| \leq C, \qquad \forall t \in [0,t_{\mathrm{end}}], n \in \mathbb{N}_0^d$$

and

$$\|\sigma(t,n)\| \leq C\cdot\Omega, \qquad \|\varsigma(t,n)\| \leq \frac{C}{\Omega}, \qquad \forall t \in [0,t_{\mathrm{end}}], n \in \mathbb{N}_0^d.$$

Further, we assume that all third moments are bounded by

$$\left\|\mathbb{E}\left[p_2^3(t,\cdot|n)\right]\right\| \leq C\cdot\Omega^2, \qquad \left\|\mathbb{E}\left[q_2^3(t,\cdot|n)\right]\right\| \leq \frac{C}{\Omega}, \qquad \forall t \in [0,t_{\mathrm{end}}], n \in \mathbb{N}_0^d.$$

Remark: From def. 5.3, def 5.7, and def. 5.8 follows that

$$\eta(0,n) = \mathcal{O}(\Omega)\,, \quad \theta(0,n) = \mathcal{O}(1)\,, \quad \sigma(0,n) = 0, \quad \varsigma(0,n) = \mathcal{O}\big(\epsilon^2\big)\,, \quad \text{for } \Omega \to \infty.$$

We find for each Ω a time point $t_{\text{end}}^{[\Omega]} > 0$ such that the bounds in assumption 5.12 hold for every $t \in [0, t_{\text{end}}^{[\Omega]}]$. But assumption 5.12 is stronger, because we assume that t_{end} is not depending on Ω.

Assumption 5.13 (Bound for the Propensity Function)**.**
We assume that a constant $C > 0$ exists, so that

$$\|\alpha_j(\cdot)u(t,\cdot)\|_{\ell_d^1} \leq C\,\|u(t,\cdot)\|_{\ell_d^1}$$

holds for $t \in [0, t_{\text{end}}]$, $j \in \{1,\dots,R\}$ and

$$\begin{aligned} u(t,n) &= \Big(\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\Big)\,(t,n) \\ \text{or } u(t,n) &= \Big(\Omega^{-1}\beta_j(\eta)\eta p_1 - \tilde{\beta}_j(\theta)\theta q_1\Big)\,(t,n). \end{aligned}$$

Remark:

- By assumption 5.11 it follows that $\alpha_j(\cdot)u(t,\cdot) \in \ell_n^1$ for the two possible cases of u.
- This assumption holds, if u goes fast enough to zero for $n \to \infty$. This is, for example, the case for the network in chapter 5.2, because the state space is bounded by 10 in n-direction.

Lemma 5.14 (Difference of the Propensity Functions)**.**
If $x \in \mathbb{R}_{0,+}^D$ and $x_k \geq \frac{1}{\Omega}\ \forall k = 1,\dots,D$ and the assumptions 5.9 and 5.10 hold, then there exists a constant $C > 0$ such that

$$\Big|\beta_j(\Omega x) - \tilde{\beta}_j(x)\Big| \leq C\Omega^{-\gamma(j)}.$$

If $\lambda_{jk}^{in} \leq 1\ \forall k = 1,\dots,D$,

$$\beta_j(\Omega x) = \tilde{\beta}_j(x)$$

holds.

Remarks:

- This lemma is similar to lemma 4.3, C5 and C6. The proof is stated in appendix C7.
- The propensity functions β_j and $\tilde{\beta}_j$ are polynomials of degree two or less, due to assumption 5.9. Therefore, the Hessians $\nabla^2\beta_j$ and $\nabla^2\tilde{\beta}_j$ are both constant.

Lemma 5.15 (A Bound for Propensity Functions)**.**
Let

$$y : \mathbb{N}^d \to \mathbb{R}^d, \qquad\qquad z : \mathbb{N}^d \to \mathbb{R}^d$$

with upper bounds

$$\max_{n\in\mathbb{N}^d} \|y(n)\| \leq C\cdot\Omega, \qquad \max_{n\in\mathbb{N}^d} \|z(n)\| \leq C.$$

Further, let $u \in \ell^1_d$ and $v \in \ell^1_d$. Then there is a a constant $C > 0$ for all $j = 1, \cdots, R$ such that

$$\left\|\beta_j(y)u - \tilde{\beta}_j(z)v\right\|_{\ell^1_d} \leq C\Omega^{1-\gamma(j)}\left(\left\|\frac{1}{\Omega}yu - zv\right\|_{\ell^1_d} + \|u-v\|_{\ell^1_d}\right) + C\Omega^{-\gamma(j)}\|u\|_{\ell^1_d}.$$

It follows that

$$\frac{1}{\Omega}yu - zv \in \ell^1_d.$$

Proof. We investigate, as in the proof of lemma 5.14, the three different cases resulting from assumption 5.9:

$|\lambda^{\mathsf{in}}_j|_1 = 0$

We have the propensity functions

$$\beta_j(y) = \Omega^{1-\gamma(j)},$$
$$\tilde{\beta}_j(y) = \Omega^{1-\gamma(j)}$$

and

$$\left\|\beta_j(y)u - \tilde{\beta}_j(z)v\right\|_{\ell^1_d} = \Omega^{1-\gamma(j)}\|u-v\|_{\ell^1_d}.$$

$|\lambda^{\mathsf{in}}_j|_1 = 1$

In this case exists a $k \in \{1,\ldots,d\}$ with

$$\beta_j(y) = \Omega^{-\gamma(j)}y_k,$$
$$\tilde{\beta}_j(y) = \Omega^{1-\gamma(j)}y_k$$

and

$$\left\|\beta_j(y)u - \tilde{\beta}_j(z)v\right\|_{\ell^1_d} = \Omega^{1-\gamma(j)}\left\|\frac{1}{\Omega}y_k u - z_k v\right\|_{\ell^1_d} \leq \Omega^{1-\gamma(j)}\left\|\frac{1}{\Omega}yu - zv\right\|_{\ell^1_d}.$$

$|\lambda^{\mathsf{in}}_j|_1 = 2$

In this case exists $k, l \in \{1,\ldots,d\}$ with

$$\tilde{\beta}_j(y) = \hat{c}_j\Omega^{1-\gamma(j)}y_k y_l \qquad \text{with } \hat{c}_j = \begin{cases} c_j & \text{if } k \neq l \\ \frac{1}{2}c_j & \text{if } k = l \end{cases}$$

Using the assumption

$$\max_{n\in\mathbb{N}^d} \|y(n)\| \leq C \cdot \Omega, \qquad \max_{n\in\mathbb{N}^d} \|z(n)\| \leq C,$$

and applying lemma 5.14, gives us

$$\left\| \beta_j(y)u - \tilde{\beta}_j\left(\frac{y}{\Omega}\right) u \right\|_{\ell_d^1} \leq C\Omega^{-\gamma(j)} \|u\|_{\ell_d^1}. \tag{5.14}$$

Then, we derive

$$\begin{aligned}
\tilde{\beta}_j\left(\frac{y}{\Omega}\right) u - \tilde{\beta}_j(z)v &= \hat{c}_j\Omega^{-\gamma(j)}\left(\frac{1}{\Omega^2} y_k y_l u - z_k z_l v\right) \\
&= \hat{c}_j\Omega^{-\gamma(j)}\Bigg(\frac{1}{\Omega} y_k\left(\frac{1}{\Omega} y_l u - z_l v\right) + z_l\left(\frac{1}{\Omega} y_k u - z_k v\right) \\
&\qquad - \frac{1}{\Omega} y_k z_l (u - v)\Bigg).
\end{aligned}$$

The assumption of the lemma gives us

$$\left|\frac{y_k(n)}{\Omega}\right| \leq C \qquad \text{and} \qquad |z_l(n)| \leq C.$$

Then, it follows that

$$\left\| \tilde{\beta}_j\left(\frac{y}{\Omega}\right) u - \tilde{\beta}_j(z)v \right\|_{\ell_d^1} \leq C\Omega^{1-\gamma(j)} \left(\left\| \frac{1}{\Omega} yu - zv \right\|_{\ell_d^1} + \|u - v\|_{\ell_d^1} \right).$$

Finally, applying eq. (5.14) proves the assertion.

□

Theorem 5.16 (An Error Bound for the LME)**.**
Let $p(t, n, m)$ be the solution of the partitioned CME (cf. def. 5.3) and let $q(t, n, x)$ be the solution of the LME (cf. def 5.7) with initial condition given in def. 5.8. If the assumptions 5.9, 5.10, 5.11, 5.12 and 5.13 hold, there exists a constant $C > 0$ such that the errors in marginal distributions and conditional expectations are bounded by

$$\|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell_d^1} \leq \frac{C}{\Omega}, \tag{5.15}$$

$$\left\| \frac{1}{\Omega}\eta(t,\cdot)p_1(t,\cdot) - \theta(t,\cdot)q_1(t,\cdot) \right\|_{\ell_d^1} \leq \frac{C}{\Omega}, \tag{5.16}$$

for all $t \in [0, t_{end}]$.

Proof. The main idea of this proof is the application of Gronwall's lemma. We postpone the estimate of

$$E(t) = \|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell_d^1} + \left\| \frac{1}{\Omega}\eta(t,\cdot)p_1(t,\cdot) - \theta(t,\cdot)q_1(t,\cdot) \right\|_{\ell_d^1}$$

to lemma 5.17 and lemma 5.18. These lemmata show the Gronwall inequality

$$E(t) \leq \frac{C_1}{\Omega} + C_2 \int_0^t E(\tau)\mathrm{d}\tau$$

for constants $C_1, C_2 > 0$. The Gronwall lemma states that

$$E(t) \leq \frac{C_1}{\Omega} e^{C_2 t}$$

and therefore proves the assertion. □

Remark:

- We note that the parameter ε (cf. def 5.8) influences the constants and the time interval, but not the convergence rate.
- The constant C depends on the assumptions 5.12 and 5.13 and on the size of t_{end}.

Lemma 5.17 (A Bound for the Marginal Distributions)**.**
Under the same conditions as in theorem 5.16, there exists a constant $C > 0$, for all $t \in [0, t_{end}]$, such that,

$$\begin{aligned}\|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell_d^1} \leq \frac{C}{\Omega} &+ C\int_0^t \|(p_1 - q_1)(\tau,\cdot)\|_{\ell_d^1}\,\mathrm{d}\tau \\ &+ C\int_0^t \left\| \left(\frac{1}{\Omega}\eta p_1 - \theta q_1\right)(\tau,\cdot)\right\|_{\ell_d^1} \mathrm{d}\tau.\end{aligned}$$

Proof.

Step 1: By the definition of the marginal distribution of the partitioned CME it follows that

$$\partial_t p_1 = \sum_{j\in J_0}\sum_{m\in\mathbb{N}_0^D} (\Delta_{\nu_j}\Delta_{\mu_j} - 1)\alpha_j\beta_j p + \sum_{j\in J_1}\sum_{m\in\mathbb{N}_0^D} (\Delta_{\nu_j}\Delta_{\mu_j} - 1)\alpha_j\beta_j p.$$

We find that the first sum equals zero, because $\Delta_{\nu_j} = I\ \forall j \in J_0$ and because lemma 5.5 shows that

$$\sum_{m\in\mathbb{N}_0^D} (\Delta_{\mu_j} - 1)\alpha_j\beta_j p = 0.$$

Further, lemma 5.5 also shows us that

$$\sum_{j\in J_1}\sum_{m\in\mathbb{N}_0^D} \Delta_{\nu_j}(\Delta_{\mu_j} - 1)\alpha_j\beta_j p = 0$$

and together with

$$\Delta_{\nu_j}\Delta_{\mu_j} - 1 = \Delta_{\nu_j}(\Delta_{\mu_j} - 1) + (\Delta_{\nu_j} - 1)$$

we have that

$$\partial_t p_1 = \sum_{j\in J_1}\sum_{m\in\mathbb{N}_0^D}(\Delta_{\nu_j}-1)\alpha_j\beta_j p = \sum_{j\in J_1}(\Delta_{\nu_j}-1)\alpha_j\sum_{m\in\mathbb{N}_0^D}\beta_j p_2 p_1 \tag{5.17}$$

holds. We derive a Taylor series of β_j at $\eta\in\mathbb{R}_{0,+}^D$ [Eng06]:

$$\beta_j(m) = \beta_j(\eta) + \nabla\beta_j(\eta)^T(m-\eta) + \frac{1}{2}(m-\eta)^T\left(\nabla^2\beta_j\right)(m-\eta). \tag{5.18}$$

Higher order terms are equal to zero, because of assumption 5.9. Furthermore, if we consider η to be the conditional expectation of p_2, we have

$$\sum_{m\in\mathbb{N}_0^D}\beta_j(m)p_2(t,m|n) = \beta_j(\eta) + \operatorname{trace}\left(\sigma(t,n)\nabla^2\beta_j\right) \tag{5.19}$$

because of $\sum_{m\in\mathbb{N}_0^D}p_2(t,m|n)=1$ and $\sum_{m\in\mathbb{N}_0^D}(m-\eta(t,n))p_2(t,m|n)=0$. By definition, we have that

$$\nabla^2\beta_j = \begin{cases}0 & \text{if } |\lambda_j^{in}|_1 < 2\\ \Omega^{-1-\gamma(j)} & \text{if } |\lambda_j^{in}|_1 = 2.\end{cases} \tag{5.20}$$

Further, because $\gamma(j)=1\ \forall j\in J_1$ and $\sigma=\mathcal{O}(\Omega)$ by assumption (5.12), it follows that $\sigma\nabla^2\beta_j=\mathcal{O}\left(\Omega^{-1}\right)$. Inserting in eq. (5.19) yields

$$\sum_{m\in\mathbb{N}_0^D}\beta_j(m)p_2(t,m|n) = \beta_j(\eta) + \mathcal{O}\left(\Omega^{-1}\right). \tag{5.21}$$

Combining eq. (5.21) and eq. (5.17) results in an equation of motion for the marginal distribution of the CME

$$\partial_t p_1 = \sum_{j\in J_1}(\Delta_{\nu_j}-1)\alpha_j\beta_j(\eta)p_1 + \mathcal{O}\left(\Omega^{-1}\right). \tag{5.22}$$

Step 2: We perform the same steps to derive an equation of motion for the marginal distribution of the LME. By definition and due to the boundary conditions of the LME it holds that

$$\begin{aligned}\partial_t q_1 &= \sum_{j\in J_1}\int_{\mathbb{R}_{0,+}^D}(\Delta_{\nu_j}-1)\alpha_j\tilde{\beta}_j q\,\mathrm{d}x - \frac{1}{\Omega}\sum_{j=1}^{r}\alpha_j\int_{\mathbb{R}_{0,+}^D}\nabla\left(\tilde{\beta}_j q\right)^T\mu_j\,\mathrm{d}x\\ &= \sum_{j\in J_1}(\Delta_{\nu_j}-1)\alpha_j\int_{\mathbb{R}_{0,+}^D}\tilde{\beta}_j q\,\mathrm{d}x\end{aligned}$$

We derive a Taylor series for $\tilde{\beta}_j$ (again higher terms vanish due to assumption 5.9):

$$\tilde{\beta}_j(x) = \tilde{\beta}_j(\theta) + \nabla\tilde{\beta}_j(\theta)^T(x-\theta) + \frac{1}{2}(x-\theta)^T\left(\nabla^2\tilde{\beta}_j\right)(x-\theta), \tag{5.23}$$

which gives us

$$\int_{\mathbb{R}^D_{0,+}} \tilde{\beta}_j(x) q_2(t,x|n)\,\mathrm{d}x = \tilde{\beta}_j(\theta) + \operatorname{trace}\left(\varsigma \nabla^2 \tilde{\beta}_j\right). \tag{5.24}$$

Because assumption 5.12 gives us $\varsigma\nabla^2\tilde{\beta}_j = \mathcal{O}\big(\Omega^{-1}\big)$, we obtain

$$\partial_t q_1 = \sum_{j\in J_1} (\Delta_{\nu_j} - 1)\alpha_j \tilde{\beta}_j(\theta) q_1 + \mathcal{O}\big(\Omega^{-1}\big). \tag{5.25}$$

Step 3: After deriving the two equations of motions eq. (5.22) and eq. (5.25), we combine them to estimate

$$\begin{aligned}\|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell^1_d} &\leq 2\int_0^t \sum_{j\in J_1} \left\|\alpha_j\left(\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\right)(\tau,\cdot)\right\|_{\ell^1_d} \mathrm{d}\tau + \mathcal{O}\big(\Omega^{-1}\big) \\ &\leq C\int_0^t \sum_{j\in J_1} \left\|\left(\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\right)(\tau,\cdot)\right\|_{\ell^1_d} \mathrm{d}\tau + \mathcal{O}\big(\Omega^{-1}\big).\end{aligned}$$

This holds because

$$p_1(0) = q_1(0), \qquad \left\|(\Delta_{\nu_j} - 1)\right\|_{\ell^1_d} \leq 2,$$

and by applying assumption 5.13. Finally, we apply lemma 5.15 and the assertion follows, because $\Omega^{1-\gamma(j)} = 1\ \forall j \in J_1$.

□

Lemma 5.18 (A Bound for the Conditional Expectations)**.**
Underthe same conditions as in theorem 5.16, there exists a constant $C > 0$, for all $t \in [0, t_{end}]$, such that,

$$\begin{aligned}\left\|\Omega^{-1}\eta p_1 - \theta q_1\right\|_{\ell^1_d} \leq \frac{C}{\Omega} &+ C\int_0^t \|(p_1 - q_1)(\tau,\cdot)\|_{\ell^1_d}\,\mathrm{d}\tau \\ &+ C\int_0^t \left\|\left(\frac{1}{\Omega}\eta p_1 - \theta q_1\right)(\tau,\cdot)\right\|_{\ell^1_d} \mathrm{d}\tau.\end{aligned}$$

Proof.

Step 1: By definition we have that

$$\begin{aligned}\partial_t\left(\eta p_1\right)(t,n) &= \sum_{m\in\mathbb{N}_0^D} m\partial_t p(t,n,m) \\ &= \sum_{j=1}^R \sum_{m\in\mathbb{N}_0^D} m\left((\Delta_{\nu_j}\Delta_{\mu_j} - 1)\alpha_j\beta_j p\right)(t,n,m) \\ &= \sum_{j=1}^R \sum_{m\in\mathbb{N}_0^D} m\left(\Delta_{\nu_j}(\Delta_{\mu_j} - 1)\alpha_j\beta_j p\right)(t,n,m)\end{aligned}$$

$$+\sum_{j=1}^{R}\sum_{m\in\mathbb{N}_0^D} m\left((\Delta_{\nu_j}-1)\alpha_j\beta_j p\right)(t,n,m)$$

$$=\sum_{j=1}^{R}\mu_j\sum_{m\in\mathbb{N}_0^D}\left(\Delta_{\nu_j}\alpha_j\beta_j p\right)(t,n,m) \tag{5.26}$$

$$+\sum_{j\in J_1}\sum_{m\in\mathbb{N}_0^D} m\left((\Delta_{\nu_j}-1)\alpha_j\beta_j p\right)(t,n,m). \tag{5.27}$$

As in the last proof we apply lemma 5.5 and $\Delta_{\nu_j}-1=0\ \forall j\in J_0$ to derive the statement. By using eq. (5.19), which was derived in the last proof and resulted from the Taylor expansion of the propensity function, we obtain for the first sum (5.26)

$$\sum_{j=1}^{R}\mu_j\sum_{m\in\mathbb{N}_0^D}\left(\Delta_{\nu_j}\alpha_j\beta_j p\right)(t,n,m)=\sum_{j=1}^{R}\mu_j\left(\Delta_{\nu_j}\alpha_j\left[\beta_j(\eta)+\operatorname{trace}\left(\sigma\nabla^2\beta_j\right)\right]p_1\right)(t,n,m)$$

$$=\sum_{j=1}^{R}\mu_j\left(\Delta_{\nu_j}\alpha_j\beta_j(\eta)p_1\right)(t,n,m)+\mathcal{O}(1) \tag{5.28}$$

since eq. (5.20) and assumption 5.12 show that for all $t\in[0,t_{\text{end}}]$ and $n\in\mathbb{N}_0^d$

$$\left\|\operatorname{trace}\left(\sigma(t,n)\nabla^2\beta_j\right)\right\|\le C\Omega^{-\gamma(j)}\le C.$$

The similar trace term (5.21) in the proof of lemma 5.17 caused there an error of $\mathcal{O}\left(\Omega^{-1}\right)$, because $j\in J_1$ which results in $\gamma(j)=1$. However, here we have the case $j\in J_0\cup J_1=\{1,\dots,R\}$ and $\gamma(j)$ can also be zero.

Insertion of the Taylor series (5.18) in the second term (5.27) yields

$$\sum_{m\in\mathbb{N}_0^D} m\beta_j(m)p(t,n,m)=\left(\beta_j(\eta)\eta+R_1(t,n)\right)p_1(t,n) \tag{5.29}$$

with $\eta=\eta(t,n)$ and

$$R_1(t,n)=\sum_{m\in\mathbb{N}_0^D} m\left(\nabla\beta_j(\eta)^T(m-\eta)+\frac{1}{2}(m-\eta)^T\left(\nabla^2\beta_j\right)(m-\eta)\right)p_2(t,m|n)$$

$$=\sigma\nabla\beta_j(\eta)+\frac{1}{2}\eta\operatorname{trace}\left(\sigma\right)\nabla^2\beta_j)$$

$$+\frac{1}{2}\sum_{m\in\mathbb{N}_0^D}(m-\eta)\left((m-\eta)^T\left(\nabla^2\beta_j\right)(m-\eta)\right)p_2(t,m|n).$$

The last sum is the third moment of $p_2(t,\cdot|n)$ and by assumption 5.12 the term vanishes, because

$$\nabla\beta_j(\eta)=\mathcal{O}\left(\Omega^{-1}\right),\qquad\qquad \nabla^2\beta_j=\mathcal{O}\left(\Omega^{-2}\right),$$

and thus obtain

$$\|R_1(t,n)\| \leq C \qquad \forall n \in \mathbb{N}^d \qquad \text{and} \qquad t \in [0, t_{\text{end}}]. \tag{5.30}$$

By combining eq. (5.28) and eq. (5.30) and substitution into eq. (5.26) we obtain the equation of motion,

$$\begin{aligned} \partial_t \left(\eta p_1\right)(t,n) = &\sum_{j=1}^{r} \mu_j \left(\Delta_{\nu_j} \alpha_j \beta_j(\eta) p_1\right)(t,n) \\ &+ \sum_{j \in J_1} \left((\Delta_{\nu_j} - 1)\alpha_j \beta_j(\eta)\eta p_1\right)(t,n) + \mathcal{O}(1)\,. \end{aligned} \tag{5.31}$$

Later, this equation will be scaled by Ω^{-1} which reduces the error term $\mathcal{O}(1)$ to $\mathcal{O}\big(\Omega^{-1}\big)$.

Step 2: We derive now an equation of motion for

$$\begin{aligned} \partial_t \left(\theta q_1\right) &= \int_{\mathbb{R}^D_{0,+}} x \partial_t q \,\mathrm{d}x \\ &= \sum_{j \in J_1} \int_{\mathbb{R}^D_{0,+}} x \tilde{\beta}_j (\Delta_{\nu_j} - 1)\alpha_j q \,\mathrm{d}x - \frac{1}{\Omega} \sum_{j=1}^{R} \int_{\mathbb{R}^D_{0,+}} x \alpha_j \nabla \left(\tilde{\beta}_j q\right)^T \mu_j \,\mathrm{d}x, \end{aligned}$$

with $\tilde{\beta}_j = \tilde{\beta}_j(x)$ and $q = q(t,n,x)$. Integration by parts gives us

$$\int_{\mathbb{R}^D_{0,+}} x \nabla \left(\tilde{\beta}_j q\right)^T \mu_j \,\mathrm{d}x = -\mu_j \int_{\mathbb{R}^D_{0,+}} \tilde{\beta}_j q \,\mathrm{d}x$$

due to the boundary conditions of the LME. Further, it holds that

$$\lim_{x \to \infty} \tilde{\beta}_j(x) q(t,n,x) = 0,$$

because by assumption 5.11 $q(t,n,\cdot) \in \mathcal{Y}^3_{d,D}$.

These considerations result in

$$\partial_t \left(\theta q_1\right) = \sum_{j \in J_1} (\Delta_{\nu_j} - 1)\alpha_j \int_{\mathbb{R}^D_{0,+}} x \tilde{\beta}_j q \,\mathrm{d}x \tag{5.32}$$

$$+ \frac{1}{\Omega} \sum_{j=1}^{r} \mu_j \int_{\mathbb{R}^D_{0,+}} \alpha_j \tilde{\beta}_j q \,\mathrm{d}x. \tag{5.33}$$

Insertion of the Taylor series (5.23) into the first sum (5.32) yields

$$\int_{\mathbb{R}^D_{0,+}} x \tilde{\beta}_j q \,\mathrm{d}x = \left(\tilde{\beta}_j(\theta)\theta + R_2(t,n)\right) q_1(t,n)$$

with $\theta = \theta(t,n)$ and

$$R_2(t,n) = \int_{\mathbb{R}^D_{0,+}} x \left(\nabla \tilde{\beta}_j(\theta)^T (x - \theta) + \frac{1}{2}(x-\theta)^T \left(\nabla^2 \tilde{\beta}_j\right)(x-\theta)\right) q_2 \,\mathrm{d}x$$

$$= \varsigma(t,n)\nabla\tilde{\beta}_j(\theta) + \frac{1}{2}\theta \operatorname{trace}\left(\varsigma(t,n)\nabla^2\tilde{\beta}_j\right) + \frac{1}{2}\int_{\mathbb{R}^D_{0,+}} (x-\theta)(x-\theta)^T\left(\nabla^2\tilde{\beta}_j\right)(x-\theta)q_2\,\mathrm{d}x.$$

We can estimate the remainder term in a similar way to the procedure used in step 1 of the proof:

$$\|R_2(t,n)\| \leq \frac{C}{\Omega} \qquad \forall n \in \mathbb{N}_0^d \qquad \text{and} \qquad t \in [0, t_{\text{end}}],$$

because $j \in J_1$ and because assumption 5.12 bounds the derivatives and the third moment.

Now, we have a look at term (5.33). The Taylor series (5.23) gives

$$\begin{aligned}\frac{1}{\Omega}\sum_{j=1}^{R}\mu_j\int_{\mathbb{R}^D_{0,+}}\alpha_j\tilde{\beta}_j q\,\mathrm{d}x &= \frac{1}{\Omega}\sum_{j=1}^{R}\mu_j\alpha_j\left[\tilde{\beta}_j(\theta) + \operatorname{trace}\left(\varsigma\nabla^2\tilde{\beta}_j\right)\right]q_1 \\ &= \frac{1}{\Omega}\sum_{j=1}^{R}\mu_j\alpha_j\tilde{\beta}_j(\theta)q_1 + \mathcal{O}\big(\Omega^{-1}\big) \qquad (5.34)\end{aligned}$$

by assumption 5.12

$$\left\|\operatorname{trace}\left(\varsigma(t,n)\nabla^2\tilde{\beta}_j\right)\right\| \leq C\Omega^{-\gamma(j)} \leq C \qquad (5.35)$$

for all $t \in [0, t_{\text{end}}]$ and $n \in \mathbb{N}_0^d$.

Substitution of eq. (5.34) and eq. (5.35) into eq. (5.32) gives,

$$\partial_t\left(\theta q_1\right) = \sum_{j\in J_1}(\Delta_{\nu_j} - 1)\alpha_j\tilde{\beta}_j(\theta)\theta q_1 + \frac{1}{\Omega}\sum_{j=1}^{R}\mu_j\left(\alpha_j\tilde{\beta}_j(\theta)q_1\right) + \mathcal{O}\big(\Omega^{-1}\big). \qquad (5.36)$$

Step 3: We combine the equations of motion, by scaling (5.31) with Ω^{-1} and subtracting (5.36):

$$\begin{aligned}\left\|\Omega^{-1}\eta p_1 - \theta q_1\right\|_{\ell^1_d} &\leq \frac{C}{\Omega} + \frac{C}{\Omega}\int_0^t\sum_{j=1}^{R}\left\|\alpha_j\left[\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\right](\tau,\cdot)\right\|_{\ell^1_d}\mathrm{d}\tau \\ &\quad + C\int_0^t\sum_{j\in J_1}\left\|\alpha_j\left[\Omega^{-1}\beta_j(\eta)\eta p_1 - \tilde{\beta}_j(\theta)\theta q_1\right](\tau,\cdot)\right\|_{\ell^1_d}\mathrm{d}\tau \\ &\leq \frac{C}{\Omega} + \frac{C}{\Omega}\int_0^t\sum_{j=1}^{R}\left\|\left[\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\right](\tau,\cdot)\right\|_{\ell^1_d}\mathrm{d}\tau \\ &\quad + C\int_0^t\sum_{j\in J_1}\left\|\left[\Omega^{-1}\beta_j(\eta)\eta p_1 - \tilde{\beta}_j(\theta)\theta q_1\right](\tau,\cdot)\right\|_{\ell^1_d}\mathrm{d}\tau, \qquad (5.37)\end{aligned}$$

because $\left\|(\Delta_{\nu_j} - 1)\right\|_{\ell^1_d} \leq 2$ and as a result of applying assumption 5.13. From Lemma 5.15 follows

$$\frac{1}{\Omega}\left\|\left[\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\right](\tau,\cdot)\right\|_{\ell^1_d} \leq C\left\|\left[\Omega^{-1}\eta p_1 - \theta q_1\right](\tau,\cdot)\right\|_{\ell^1_d} + C\left\|\left[p_1 - q_1\right](\tau,\cdot)\right\|_{\ell^1_d} + \frac{C}{\Omega}.$$

To construct a bound for eq. (5.37), we consider the following algebraic expansion

$$\begin{aligned}\Omega^{-1}\beta_j(\eta)\eta p_1 - \tilde{\beta}_j(\theta)\theta q_1 = \beta_j(\eta)\cdot\left[\Omega^{-1}\eta p_1 - \theta q_1\right] + \beta_j(\eta)\theta\left[q_1 - p_1\right] \\ + \theta\left[\beta_j(\eta)p_1 - \tilde{\beta}_j(\theta)q_1\right].\end{aligned}$$

We take the ℓ^1-norm on both sides, use the triangle inequality and apply

$$\begin{aligned}\max_{\tau\in[0,t_{\text{end}}]}\max_{n\in\mathbb{N}_0^d}\left|\beta_j\big(\eta(\tau,n)\big)\right| \leq C \qquad\qquad \forall j \in J_1, \\ \max_{\tau\in[0,t_{\text{end}}]}\max_{n\in\mathbb{N}_0^d}\left|\theta(\tau,n)\right| \leq C\end{aligned}$$

(cf. assumption 5.12).

Finally we use lemma 5.15 and note that $j \in J_1$, to show

$$\begin{aligned}\left\|\left[\Omega^{-1}\beta_j(\eta)\eta p_1 - \tilde{\beta}_j(\theta)\theta q_1\right](\tau,\cdot)\right\|_{\ell^1_d} \leq \frac{C}{\Omega} + C\left\|\left[\Omega^{-1}\eta p_1 - \theta q_1\right](\tau,\cdot)\right\|_{\ell^1_d} \\ + C\left\|p_1(\tau,\cdot) - q_1(\tau,\cdot)\right\|_{\ell^1_d}\end{aligned} \tag{5.38}$$

for all $\tau \in [0, t_{\text{end}}]$.

In summary, we have shown that

$$\begin{aligned}\left\|\Omega^{-1}\eta p_1 - \theta q_1\right\|_{\ell^1_d} \leq \frac{C}{\Omega} + C\left\|\left[\Omega^{-1}\eta p_1 - \theta q_1\right](\tau,\cdot)\right\|_{\ell^1_d} \\ + C\left\|p_1(\tau,\cdot) - q_1(\tau,\cdot)\right\|_{\ell^1_d}.\end{aligned}$$

□

5.7 Summary and Discussion

We have seen that it is possible to combine the CME and the LME to model networks with a scale difference between the species. The advantage of this model lies in the reduced complexity of the numerical simulation. A model with small and large particle amounts is hard to solve, because high particle numbers reduce the simulation speed dramatically and result in large state spaces. On the other hand, the usage of the RRE would result in wrong results, because the RRE is not capable of dealing with the stochasticity of the species with small particle amounts. However, by combining the two models, we obtain the advantages of both models. The large species are handled using RRE and the small species are handled using the Kurtz process. Therefore the simulation time is reduced, but the stochasticity of the small species is still captured.

These considerations can be found in literature for several years, now. However, nearly no analysis of the hybrid model has been performed so far and no convergence rate for this model has been shown.

We have seen how the hybrid model can be derived in terms of a stochastic process and its PDF. Finally, we have shown the error bound

$$\begin{aligned} \|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell_d^1} &\leq \frac{C}{\Omega}, \\ \left\| \frac{1}{\Omega}\eta(t,\cdot)p_1(t,\cdot) - \theta(t,\cdot)q_1(t,\cdot) \right\|_{\ell_d^1} &\leq \frac{C}{\Omega}, \end{aligned}$$

given in theorem 5.16. .

6 | A Hybrid Model Combining the Kurtz process and the Chemical Langevin Equation

In this chapter we will expand the hybrid model from the last chapter. In section 6.1, we will shortly motivate the usage of the FPE and the CLE to replace the continuous part of the hybrid model presented in the previous chapter. We will then derive the new hybrid process (section 6.2) and its PDF (section 6.3).

We will examine an example network in section 6.4 and analyse what convergence properties we can expect from the new hybrid model. These properties are then proven in section 6.5.

6.1 Motivation

In the previous chapter, we discussed the properties of a hybrid model composed of the Kurtz process and the RRE. We have seen that the two models converge only in the marginal distributions and conditional expectations in the sense of theorem 5.16. Figure 5.1 indicated that we cannot expect convergence in the full distributions.

However, we showed that the FPE converges against the CME in distribution. Motivated by this observation we combine the CME with the FPE to a new hybrid model named Fokker-Planck master equation (FPME).

We like to point out that several other authors combined these two models [Pah09], but these publications mostly state hybrid algorithms or implementations. We are not aware of any publication that states an equation of motion for the PDF of the hybrid model.

6.2 A general derivation of the scaled Hybrid Process

This derivation can also be done in a general way. We examine the scaled and partitioned Markov Jump process

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} n_0 \\ \Omega m_0 \end{pmatrix} + \sum_{j=1}^{R} \mathcal{P}_j \left(\int_0^t \alpha_j \left(X(s) \right) \beta_j \left(Y(s) \right) \mathrm{d}s \right) \begin{pmatrix} \nu_j \\ \mu_j \end{pmatrix}$$

and first discuss the continuous subprocess.

By following the derivations introduced by Gillespie and MacNamara [Gil00, Mac09], we examine the process on bounded time intervals $[t, t+\tau]$ (cf. ch. 2.4):

$$Y(t+\tau) \approx \hat{Y}(t+\tau) = \Omega \hat{m}_0 + \sum_{j=1}^{R} \mathcal{P}_j \left(\tau \alpha_j \left(Z(t) \right) \beta_j \left(\hat{Y}(t) \right) \right) \mu_j,$$

where $\hat{m}_0$ is the state of the subprocess at the beginning of the time interval.

As motivated in [Mac09, Gil00, BTB04], we approximate the Poisson process via

$$\mathcal{P}_j(\xi) \approx \mathcal{N}_j(\xi, \xi) = \xi + \sqrt{\xi} \mathcal{N}_j(0,1).$$

It follows that

$$\begin{aligned} \hat{Y}(t+\tau) \approx \Omega \hat{m}_0 &+ \sum_{j=1}^{R} \tau \alpha_j \left(Z(t) \right) \beta_j \left(\hat{Y}(t) \right) \mu_j \\ &+ \sum_{j=1}^{R} \sqrt{\tau \alpha_j \left(Z(t) \right) \beta_j \left(\hat{Y}(t) \right)} \mu_j \mathcal{N}_j(0,1). \end{aligned}$$

Now we scale the process by $\frac{1}{\Omega}$, assume

$$\tilde{y}(t) \approx \frac{\hat{Y}(t)}{\Omega},$$

and use lemma 5.14

$$\beta_j \left(\hat{Y}(t) \right) \approx \tilde{\beta}_j \left(\frac{\hat{Y}(t)}{\Omega} \right) \approx \tilde{\beta}_j \left(\tilde{y}(t) \right)$$

to obtain

$$\begin{aligned} \tilde{y}(t+\tau) = \hat{m}_0 &+ \sum_{j=1}^{R} \tau \frac{1}{\Omega} \alpha_j \left(Z(t) \right) \tilde{\beta}_j \left(\tilde{y}(t) \right) \mu_j \\ &+ \sum_{j=1}^{R} \sqrt{\tau \frac{1}{\Omega^2} \alpha_j \left(Z(t) \right) \tilde{\beta}_j \left(\tilde{y}(t) \right)} \mu_j \mathcal{N}_j(0,1). \end{aligned} \tag{6.1}$$

By taking the limit $\tau \to 0$ (in principle, eq. 6.1 is a Euler-Maruyama discretization) we obtain the CLE [BMP96, Gil00, BTB04, Hig08, Gar04, Mac09]:

$$\begin{aligned} \mathrm{d}y(t) &= \sum_{j=1}^{R} \mu_j \frac{1}{\Omega} \alpha_j(n) \tilde{\beta}_j \left(y(t) \right) \, \mathrm{d}t + \sum_{j=1}^{R} \mu_j \sqrt{\frac{1}{\Omega^2} \alpha_j(n) \tilde{\beta}_j \left(y(t) \right)} \, \mathrm{d}W_j(t), \\ y(0) &= m_0. \end{aligned}$$

Finally, we have to apply the assumptions

$$y(t) \approx \frac{\hat{Y}(t)}{\Omega}$$
$$\beta_j\left(\hat{Y}(t)\right) \approx \tilde{\beta}_j\left(y(t)\right)$$

to the first subprocess and we end up with the stochastic process for the scaled hybrid process:

Definition 6.1 (Kurtz / CLE Hybrid Process)**.**

$$\begin{aligned} Z(t) &= n_0 + \sum_{j=1}^{R} \mathcal{P}_j\left(\int_0^t \alpha_j\left(Z(s)\right)\tilde{\beta}_j\left(y(s)\right)\mathrm{d}s\right)\nu_j \\ y(t) &= m_0 + \sum_{j=1}^{R}\left(\int_0^t \frac{1}{\Omega}\alpha_j\left(Z(s)\right)\tilde{\beta}_j\left(y(s)\right)\mathrm{d}s\right)\mu_j \\ &\quad + \sum_{j=1}^{R}\left(\int_0^t \sqrt{\frac{1}{\Omega^2}\alpha_j\left(Z(s)\right)\tilde{\beta}_j\left(y(s)\right)}\mathrm{d}W_j(s)\right)\mu_j. \end{aligned}$$

Algorithm 6.1 A Sampling Algorithm for the Hybrid process in def. 6.1

Require: initial condition n_0 and m_0, time interval $[t_0,\ t_{\text{final}}]$, stoichiometric vectors $\nu_j, \mu_j,\ j = 1,\ldots,R$, propensity functions $\alpha_j(n), \tilde{\beta}_j(x)\ j = 1,\ldots,R$, number of time steps $K \in \mathbb{N}$, scaling parameter Ω

$\Delta t = \frac{t_{\text{final}} - t_0}{K}$
$Z(0) \leftarrow n_0$
$y(0) \leftarrow m_0$
for $i \leftarrow 0,\ldots,K$ **do**
  $t_i \leftarrow t_0 + i \cdot \Delta t$
  $\hat{y} \leftarrow y(t_i)$
  $\hat{z} \leftarrow \texttt{SSA}\left[Z(t_i) + \sum_{j=1}^{R}\mathcal{P}_j\left(\int_{t_i}^{t_i+\frac{\Delta t}{2}}\alpha_j\left(Z(s)\right)\tilde{\beta}_j\left(\hat{y}\right)\mathrm{d}s\right)\nu_j\right]$
  $\hat{y} \leftarrow \texttt{SDE\_INTEGRATOR}\Big[y(t_i) + \frac{1}{\Omega}\sum_{j=1}^{R}\left(\int_{t_i}^{t_i+\Delta t}\alpha_j\left(\hat{z}\right)\tilde{\beta}_j\left(y(s)\right)\mathrm{d}s\right)\mu_j$
    $+\frac{1}{\Omega}\sum_{j=1}^{R}\left(\int_{t_i}^{t_i+\Delta t}\sqrt{\alpha_j\left(\hat{z}\right)\tilde{\beta}_j\left(y(s)\right)}\mathrm{d}W_j(s)\right)\mu_j\Big]$
  $\hat{z} \leftarrow \texttt{SSA}\left[\hat{z} + \sum_{j=1}^{R}\mathcal{P}_j\left(\int_{t_i+\frac{\Delta t}{2}}^{t_i+\Delta t}\alpha_j\left(Z(s)\right)\tilde{\beta}_j\left(\hat{y}\right)\mathrm{d}s\right)\nu_j\right]$
  $Z(t_{i+1}) = \hat{z}$
  $y(t_{i+1}) = \hat{y}$
end for
return $Z(t_i), y(t_i)$ for $t_i = t_0 + i \cdot \Delta t,\ i = 0,\ldots,K$

This process can be simulated using the hybrid algorithm 6.1 which follows the same ideas as algorithm 5.1, but instead of the RRE now the CLE has to be solved. The resulting realisations sample the probability distribution that is implicitly defined by the Fokker-Planck master equation (FPME). We will discuss this in the next section.

6.3 Derivation of the Fokker-Planck master equation

It is also possible to derive the FPME as an approximation of the CME by performing a partial Kramers-Moyal expansion on a partitioned CME with respect to the correct scaling.

The steps required for this derivation are mostly the same steps used in the derivation of the LME (cf. section 5.4.2). The difference is that we use one more term in the Kramers-Moyal expansion

$$\begin{aligned}\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)\hat{q}\left(t,\cdot,x-\frac{\mu_j}{\Omega}\right)-\tilde{\beta}_j(x)\hat{q}(t,\cdot,x) &= -\frac{1}{\Omega}\nabla\left(\tilde{\beta}_j(x)\hat{q}(t,\cdot,x)\right)^T\mu_j \\ &\quad+\frac{1}{2\Omega^2}\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)\hat{q}(t,\cdot,x)\right)\mu_j \\ &\quad+\dots.\end{aligned}$$

In the end these derivations result in the FPME:

$$\begin{aligned}\partial_t q\left(t,n,x\right) &= \sum_{j\in J_1}\tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\left(t,n-\nu_j,x\right)-\alpha_j(n)q\left(t,n,x\right)\Big) \\ &\quad-\frac{1}{\Omega}\sum_{j=1}^{R}\alpha_j(n)\nabla\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)^T\mu_j \\ &\quad+\frac{1}{2\Omega^2}\sum_{j=1}^{R}\alpha_j(n)\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)\mu_j,\end{aligned} \tag{6.2}$$

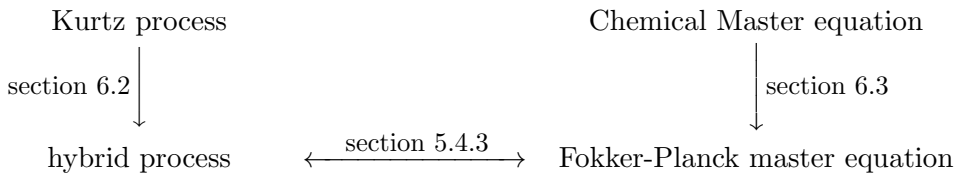
where the density $q\left(t,n,x\right)$ is the solution of the FPME and

$$\int_S q\left(t,n,x\right)\mathrm{d}x=\mathbb{P}\left(Z(t)=n,y(t)\in S\middle|Z(0)=n_0,y(0)=m_0\right).$$

We have seen so far how the new hybrid model can be constructed by splitting the Kurtz process and transforming one sub-process into a CLE. We also discussed how the FPME can be derived from the CME. The connection between the hybrid process and the FPME can now be motivated by following the same arguments as in section 5.4.3. In principle only the definition of the operator $\mathcal{A}^{(2)}$ must be expanded by the diffusion term of the FPE:

$$\begin{aligned}\partial_t q^{(2)}(t,n,x) &= \left(\mathcal{A}^{(2)}q^{(2)}\right)(t,n,x) \\ &= -\frac{1}{\Omega}\sum_{j=1}^{R}\alpha_j(n)\nabla_x\left(\tilde{\beta}_j(x)q^{(2)}\left(t,n,x\right)\right)^T\mu_j \\ &\quad+\frac{1}{2\Omega^2}\sum_{j=1}^{R}\alpha_j(n)\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)q^{(2)}\left(t,n,x\right)\right)\mu_j.\end{aligned}$$

Summarizing, we have shown in this section how the new hybrid model is connected with the Kurtz process and the CME:



6.4 A Numerical Example

Table 6.1: A reaction network that can be interpreted as the translation of a protein that inhibits its own gene. In this example, we have $d = 2$, $D = 1$, $R = 4$, $J_0 = \{1, 4\}$, $J_1 = \{2, 3\}$.

Reaction Channel $\mathfrak{R}_j$				**Propensity** $\alpha_j(n)$	$\beta_j(m)$	$\tilde{\beta}_j(x)$	**Stoichiometry** ν_j	μ_j	$\gamma(j)$
$\mathfrak{R}_1$:	$\mathcal{S}_1$	$\xrightarrow{c_1=0.5}$	$\mathcal{S}_1 + \mathcal{S}_3$	$c_1 \cdot n_1$	Ω	Ω	$\begin{pmatrix}0\\0\end{pmatrix}$	1	0
$\mathfrak{R}_2$:	$\mathcal{S}_1 + \mathcal{S}_3$	$\xrightarrow{c_2=3}$	$\mathcal{S}_2$	$c_2 \cdot n_1$	$\frac{m_1}{\Omega}$	x_1	$\begin{pmatrix}-1\\1\end{pmatrix}$	-1	1
$\mathfrak{R}_3$:	$\mathcal{S}_2$	$\xrightarrow{c_3=1}$	$\mathcal{S}_1 + \mathcal{S}_3$	$c_3 \cdot n_2$	1	1	$\begin{pmatrix}1\\-1\end{pmatrix}$	1	1
$\mathfrak{R}_4$:	$\mathcal{S}_3 + \mathcal{S}_3$	$\xrightarrow{c_4=5}$	$\emptyset$	c_4	$\frac{m_1(m_1-1)}{2\Omega}$	$\frac{\Omega x_1^2}{2}$	$\begin{pmatrix}0\\0\end{pmatrix}$	-2	0

At the beginning of this chapter, we claimed that we will construct a hybrid model with better convergence properties than the model in the previous chapter. To corroborate this claim we perform a numerical experiment and solve the CME and the FPME numerically for the example in table 6.1 (cf. section 5.2). Again we reduce the dimension of the problem by using that $n_1 + n_2 = 10$. The resulting network is 2-dimensional: one discrete and one continuous dimension. The CME is solved using the OFSP method (cf. 3.2) implemented in the software package `cmepy` [HFC10] on the time interval $[0, 0.5]$. The FPME is solved by generating 50 million realisations of the hybrid process with algorithm 6.1. We generate a histogram from the resulting trajectories for the time point $t = 0.5$. Based on these approximations for $p(t, n, m)$ and $q(t, n, m)$, we compute the quantities $p_1(t, n)$, $q_1(t, n)$, $\eta(t, n)$ and $\theta(t, n)$.

First of all, we reproduce the result from theorem 5.16 and compute the terms

$$\varepsilon_1 = \|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell^1_d}, \qquad \varepsilon_2 = \left\| \frac{1}{\Omega}\eta(t,\cdot)p_1(t,\cdot) - \theta(t,\cdot)q_1(t,\cdot) \right\|_{\ell^1_d},$$

for different values of Ω. We expect both terms to be of order $\mathcal{O}\left(\frac{1}{\Omega}\right)$. This experiment is visualised in figure 6.1. The figure supports our expectation that the error rate of the first hybrid model is conserved by our new model.

Corollary 6.2 (An Error Bound for the FPME)**.**
Let $p(t, n, m)$ be the solution of the partitioned CME (5.10) *and let $q(t, n, x)$ be the solution of the FPME* (6.2)*. Under the same assumptions as in theorem 5.16, there exists a constant*

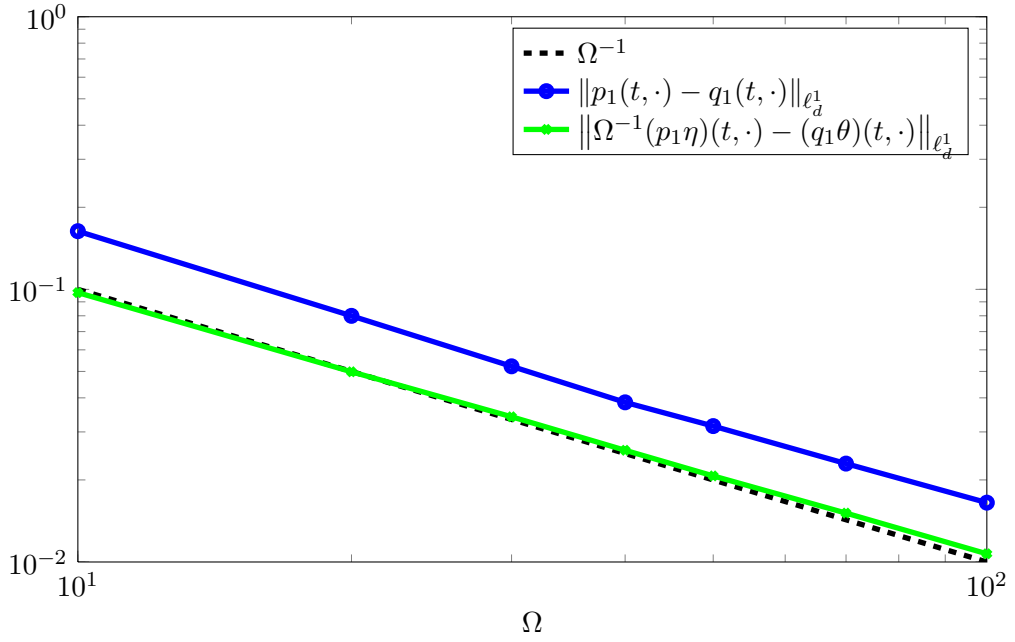


Figure 6.1: Convergence of the marginal distributions (blue) and conditional expectations (green) of the FPME against the CME. The black line plots the function $\Omega \mapsto \Omega^{-1}$. Both error terms are of order $\mathcal{O}\left(\frac{1}{\Omega}\right)$, i.e. they are parallel to the black line.

$C > 0$ such that the errors in marginal distributions and conditional expectations are bounded by

$$\begin{aligned} \|p_1(t,\cdot) - q_1(t,\cdot)\|_{\ell^1_d} &\leq \frac{C}{\Omega}, \\ \left\|\frac{1}{\Omega}\eta(t,\cdot)p_1(t,\cdot) - \theta(t,\cdot)q_1(t,\cdot)\right\|_{\ell^1_d} &\leq \frac{C}{\Omega}, \end{aligned}$$

for all $t \in [0, t_{end}]$.

Proof. The new term $\frac{1}{2\Omega^2}\sum_{j=1}^{R}\alpha_j(n)\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)\mu_j$ is of order $\mathcal{O}\left(\Omega^{-2}\right)$. Then the statement follows directly from the proof of theorem 5.16. □

However, we further expect that the full distributions of the two models converge. We compare therefore the solutions of the CME and the FPME. Figure 6.2 visualises these two solutions for $\Omega = 100$ and $t = 0.5$. We observe that the shape of the two distributions is very similar, but we also see that a direct comparison of the PDFs is not possible. The domain of the CME is $\mathbb{R}_{0,+}\times\mathbb{N}_0^d\times\mathbb{N}_0^D$ while the FPME maps from $\mathbb{R}_{0,+}\times\mathbb{N}_0^d\times\mathbb{R}_{0,+}^D$. Further, the essential support of the CME is located in the interval $[0,\Omega]$ for the m parameter, while for the FPME the essential support of x is located in the interval $[0,1]$.

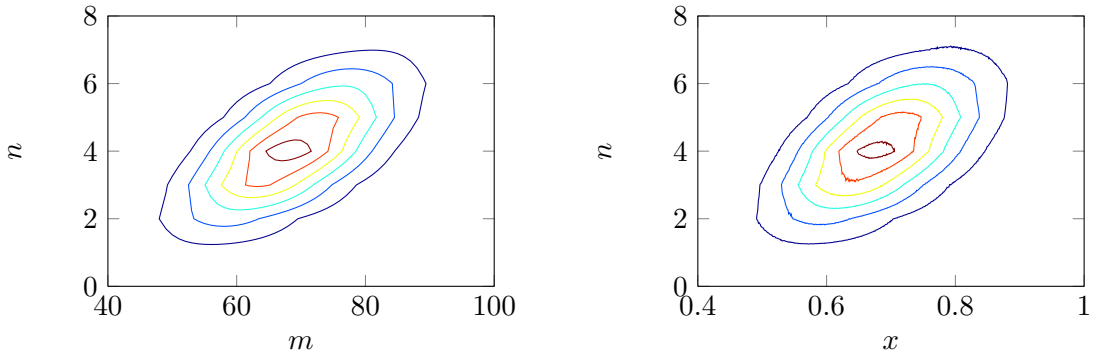


Figure 6.2: Solution of the CME (left) and the FPME (right) for the network in table 6.1 at time $t = 0.5$ and $\Omega = 100$.

We have seen this problem in chapter 4.4 before. There, we compared the FPE with the CME and defined a mapping (cf. eq. (4.17)) that integrates piecewise over the hypercubes between the points $\frac{m}{\Omega}$ and $\frac{m+1}{\Omega}$. Motivated by this, we define

$$\Psi\left(q\right)\left(t,n,m\right) := \int_{I_m} q(t,n,x)\,\mathrm{d}x \tag{6.3}$$

We derive Ψ for our example network and compare it with the CME solution p for different values of Ω. These results are visualized in figure 6.3. We observe that for increasing values

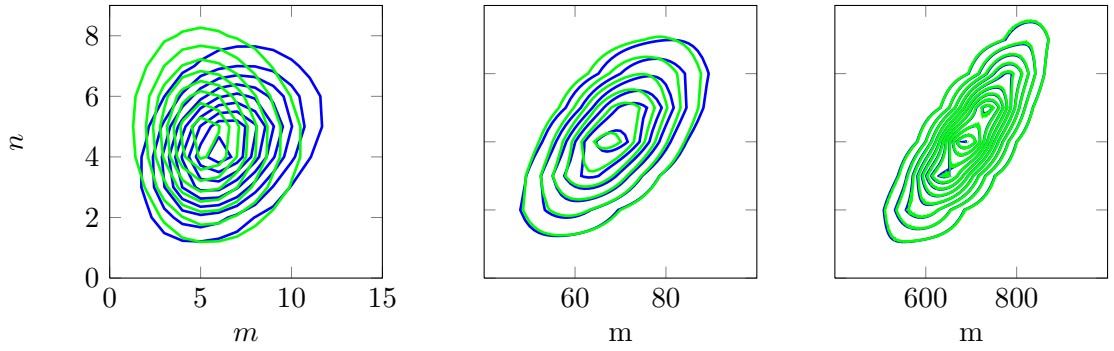


Figure 6.3: Solution of the CME $p(t,n,m)$ (blue) and of the transformed FPME $\Psi\left(q\right)\left(t,n,m\right)$ (green) for the network in table 6.1 at time $t = 0.5$ for $\Omega = 10$ (left), $\Omega = 100$ (center) and $\Omega = 1000$ (right).

of Ω the two solutions become more and more similar and for $\Omega = 1000$ they are nearly indistinguishable. These results motivate us to define the error

$$\varepsilon_{\mathrm{FPME}} := \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} \left|p(t,n,m) - \Psi\left(q\right)\left(t,n,m\right)\right|. \tag{6.4}$$

We solve the network from table 6.1 with algorithm 6.1 on the time interval $[0, 0.5]$ for different values of Ω. For each Ω we construct a histogram out of $50 \cdot 10^6$ realisations to approximate the FPME at time $t = 0.5$. As a reference solution, the CME (5.10) is solved using the OFSP implementation `cmepy` [HFC10] for the same values of Ω and

$t = 0.5$. The result is visualised in figure 6.4 and we observe that the transformed FPME solution Ψ converges to the CME solution p with order $\mathcal{O}\left(\frac{1}{\Omega}\right)$, at least for this network. Motivated from the convergence rate of the FPE we also compute the derivative term $\frac{1}{\Omega}\left\|\nabla\left(q(t,\cdot,\cdot)\right)\right\|_1$, with the norm defined later in equation (6.9), which converges with order $\mathcal{O}\left(\frac{1}{\sqrt{\Omega}}\right)$.

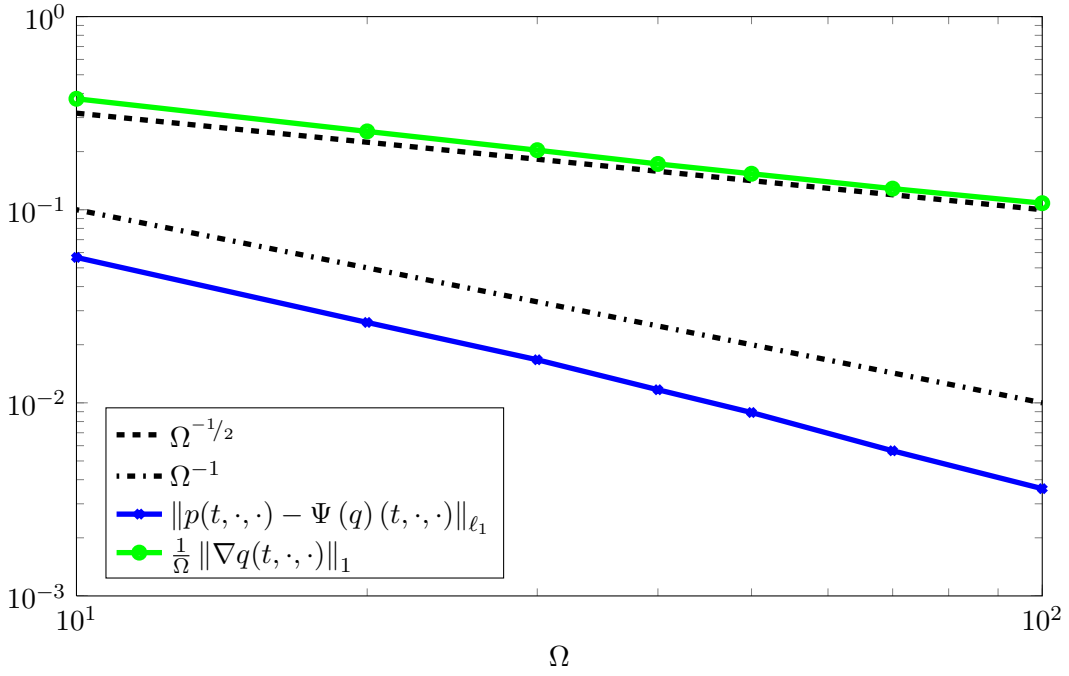


Figure 6.4: Convergence of $\varepsilon_{\text{FPME}}$ for different values of Ω (blue). The black, dashed lines denotes the function $\Omega \mapsto \Omega^{-1/2}$, the black, dash-pointed line the function $\Omega \mapsto \Omega^{-1}$. The green line shows the convergence of the first derivative $\frac{1}{\Omega}\left\|\nabla\left(q(t,\cdot,\cdot)\right)\right\|_1$.

6.5 An Error Bound for the FPME

.

As in the numerical example (cf. eq. (6.3)), we compare the CME with the FPME

$$\|p(t,\cdot,\cdot) - v(t,\cdot,\cdot)\|_{\ell_1},$$

with $v(t,n,m) := \Psi(q)(t,n,m) = \int_{I_m} q(t,n,x)\,\mathrm{d}x$, in the norm

$$\|f(t,\cdot,\cdot)\|_{\ell_1} := \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D} |f(t,n,m)|.$$

We define the CME operator

$$\partial_t p(t,n,m) = (\mathcal{A}p)(t,n,m) = \sum_{j=1}^{R}\Big(\alpha_j(n-\nu_j)\beta_j(m-\mu_j)p(t,n-\nu_j,m-\mu_j) - \alpha_j(n)\beta_j(m)p(t,n,m)\Big)$$

$$
\begin{aligned}
&= \sum_{j=1}^{R} \Big(\alpha_j(n-\nu_j)\beta_j(m-\mu_j)p(t,n-\nu_j,m-\mu_j) \\
&\qquad - \alpha_j(n)\beta_j(m-\mu_j)p(t,n,m-\mu_j) \\
&\qquad + \alpha_j(n)\beta_j(m-\mu_j)p(t,n,m-\mu_j) \\
&\qquad - \alpha_j(n)\beta_j(m)p(t,n,m) \Big) \\
&= \sum_{j\in J_1} \left(\Delta_{\mu_j}\left(\Delta_{\nu_j}-1\right)\alpha_j\beta_j p\right) + \sum_{j=1}^{R} \left(\left(\Delta_{\mu_j}-1\right)\alpha_j\beta_j p\right). \quad (6.5)
\end{aligned}
$$

The last equality holds because $\left(\Delta_{\nu_j}-1\right)\alpha_j\beta_j p = 0 \ \forall j \in J_0$. We use the same shift operators and index sets as defined in chapter 5.3 (cf. def. 5.4 and eq. (5.7)).

Further, we define the FPME operator

$$
\begin{aligned}
\partial_t q(t,n,x) = \left(\mathcal{D}q\right)(t,n,x) &= \sum_{j\in J_1} \tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\left(t,n-\nu_j,x\right) - \alpha_j(n)q\left(t,n,x\right)\Big) \\
&\qquad - \frac{1}{\Omega}\sum_{j=1}^{R} \alpha_j(n)\nabla\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)^T \mu_j \\
&\qquad + \frac{1}{2\Omega^2}\sum_{j=1}^{R} \alpha_j(n)\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)\mu_j \\
&= \sum_{j\in J_1} \tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\left(t,n-\nu_j,x\right) - \alpha_j(n)q\left(t,n,x\right)\Big) \\
&\quad + \sum_{j=1}^{R}\alpha_j(n)\Bigg(\tilde{\beta}_j\left(x-\frac{\mu_j}{\Omega}\right)q\left(t,n,x-\frac{\mu_j}{\Omega}\right) \\
&\qquad\qquad - \tilde{\beta}_j\left(x\right)q\left(t,n,x\right)\Bigg) + \mathcal{G}^{\star}_{\geq 3} \\
&= \sum_{j\in J_1}\Big(\left(\Delta_{\nu_j}-1\right)\alpha_j\tilde{\beta}_j q\Big) + \sum_{j=1}^{R}\Big(\Big(\tilde{\Delta}_{\mu_j}-1\Big)\alpha_j\tilde{\beta}_j q\Big) + \mathcal{G}^{\star}_{\geq 3},
\end{aligned} \quad (6.6)
$$

with remainder term

$$
\mathcal{G}^{\star}_{\geq 3} := \sum_{j=1}^{R}\alpha_j(n)\left(-\sum_{|\boldsymbol{k}|\geq 3}\frac{1}{\boldsymbol{k}!}\left(-\frac{\mu_j}{\Omega}\right)^{\boldsymbol{k}}\nabla^{\boldsymbol{k}}\Big(\tilde{\beta}_j(x)q(t,n,x)\Big)\right), \quad (6.7)
$$

and shift operator

$$
\tilde{\Delta}_{\mu}u(n,x) = \begin{cases} u\left(n,x-\frac{\mu}{\Omega}\right), & \text{if } x-\frac{\mu}{\Omega}\in\mathbb{R}^D_{0,+}, \\ 0, & \text{otherwise.}\end{cases}
$$

We represent $\partial_t v$ by

$$
\partial_t v(t,n,m) = \left(\mathcal{A}v\right)(t,n,m) + \mathcal{R}(t,n,m) \quad (6.8)
$$

and it follows that

$$\mathcal{R}(t,n,m) = \int_{I_m} (\mathcal{D}q)(t,n,x)\,\mathrm{d}x - (\mathcal{A}v)(t,n,m).$$

Further, we state the norm

$$\|f(t,\cdot,\cdot)\|_1 = \sum_{n\in\mathbb{N}_0^d} \int_{x\in\mathbb{R}_{0,+}^D} |f(t,n,x)|\,\mathrm{d}x, \tag{6.9}$$

and assume:

Assumption 6.3 (Bound for the Propensity Function in J_1)**.**
We assume that a constant $C > 0$ exists, so that

$$\left\| \int_{I_m} \tilde{\beta}_j(x) \left(q\left(t,n,x-\frac{\mu_j}{\Omega}\right) - q(t,n,x)\right)\mathrm{d}x \right\|_{\ell_1} \leq C \left\| \int_{I_m} q\left(t,n,x-\frac{\mu_j}{\Omega}\right) - q(t,n,x)\mathrm{d}x \right\|_{\ell_1},$$

holds for $t \in [0, t_{\mathrm{end}}]$ and $j \in J_1$.

Remark: This assumption is motivated by the observation that the propensity function $\tilde{\beta}_j$ reduces, for $j \in J_1$ and therefore $\gamma(j) = 1$, to

$$\tilde{\beta}_j(x) = \prod_{i=1}^{D} \frac{x_i^{\lambda_{j,i}^{\mathrm{in}}}}{\lambda_{j,i}^{\mathrm{in}}!}.$$

Assumption 6.4 (Bound for remainder terms)**.**
We assume that, on bounded time intervals $[0, t_{\mathrm{end}}]$,

$$\sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \mathcal{G}_{\geq 3}^{\star}\,\mathrm{d}x \right| + \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \mathcal{G}_{j,\geq 2}^{\dagger}\,\mathrm{d}x \right| \leq \frac{C}{\Omega}$$

with the Taylor series remainder terms (cf. def. B4 (page 125), eq. (6.7), eq. (6.10))

$$\mathcal{G}_{\geq 3}^{\star} := \sum_{j=1}^{R} \alpha_j(n) \left(- \sum_{|\boldsymbol{k}|\geq 3} \frac{1}{\boldsymbol{k}!} \left(-\frac{\mu_j}{\Omega}\right)^{\boldsymbol{k}} \nabla^{\boldsymbol{k}} \Big(\tilde{\beta}_j(x) q(t,n,x) \Big) \right),$$

$$\mathcal{G}_{j,\geq 2}^{\dagger} := - \sum_{|\boldsymbol{k}|\geq 2} \frac{1}{\boldsymbol{k}!} \left(-\frac{\mu_j}{\Omega}\right)^{\boldsymbol{k}} \nabla^{\boldsymbol{k}} \Big(q(t,n,x) \Big).$$

Remark: The remainder terms $\mathcal{G}_{\geq 3}^{\star}$ and $\mathcal{G}_{j,\geq 2}^{\dagger}$ arise from Taylor expansions in different parts of the proof in this section and the assumption is similar to assumption 4.4 used in the proof of the FPE convergence theorem (cf. section 4.4). This assumption is motivated by the observation that the remainder terms are dominated by the term $\frac{1}{\Omega^3}$, for those cases where the derivatives of $\tilde{\beta}q$ or q, respectively, are smooth. So the bound could even be "stronger" but we only "need" order $\mathcal{O}\big(\Omega^{-1}\big)$ here.

Theorem 6.5 (Convergence of the FPME)**.**
Let p be the solution of the CME (5.10) *and q be the solution of the FPME* (6.2)*. Then it holds, on bounded time intervals $[0, t_{end}]$ and for $\Omega > \Omega_0$ independent constants $C_1, C_2 > 0$, that*

$$\left\| p(t,\cdot,\cdot) - \int_{I_m} q(t,\cdot,x)\,\mathrm{d}x \right\|_{\ell_1} \leq \frac{C_1}{\Omega} \left\| \nabla\Big(q(t,\cdot,\cdot)\Big) \right\|_1 + \frac{C_2}{\Omega}$$

under the conditions of eq. (4.2)*, eq.* (4.8)*, eq.* (6.8)*, lemma 4.3, assumption 4.2, assumption 5.13, assumption 6.3, assumption 6.4, assumption 5.10, and assuming that*

$$\|p(0,\cdot,\cdot) - v(0,\cdot,\cdot)\|_{\ell_1} = \left\| p(0,\cdot) - \int_{I_m} q(0,\cdot,\cdot)\,\mathrm{d}x \right\|_{\ell_1} = 0,$$

with $v(t,n,m) = \Psi\left(q\right)(t,n,m)$ and Ω_0 depending on eq. (4.2)*, lemma 4.3, assumption 6.4 and assumption 5.10.*

Remark:

- We should be aware that q depends on Ω and therefore $\left\| \nabla\Big(q(t,\cdot,\cdot)\Big) \right\|_1$ changes if Ω is changed.
- For the example in table 6.1 the term $\frac{1}{\Omega}\left\| \nabla\Big(q(t,\cdot,\cdot)\Big) \right\|_1$ is of order $\mathcal{O}\big(\Omega^{-1/2}\big)$.
- This proof is similarly structured and uses the same techniques as the proof of theorem 4.5, but extends it by an estimate of the "CME part" of the FPME.

Proof. From the variation of constants formula follows that

$$\|p(t,\cdot,\cdot) - v(t,\cdot,\cdot)\|_{\ell_1} \leq \int_0^t \left\| e^{(t-s)\mathcal{A}} \right\|_{\ell_1} \cdot \|\mathcal{R}(s,\cdot,\cdot)\|_{\ell_1}\,\mathrm{d}s.$$

From the structure of the operator $\mathcal{A}$, it was shown that [Heg08, WGMH10, JA10, VK07, ch. V],

$$\left\| e^{(t-s)\mathcal{A}} \right\|_{\ell_1} \leq 1.$$

Hence, the error only depends on $\mathcal{R}$:

$$\begin{aligned}
\|\mathcal{R}(s,\cdot,\cdot)\|_{\ell_1} &= \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \left(\mathcal{D}q\right)(t,n,x)\,\mathrm{d}x - \left(\mathcal{A}v\right)(t,n,m) \right| \\
&= \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \left(\sum_{j\in J_1} \Big(\left(\Delta_{\nu_j} - 1\right) \alpha_j \tilde{\beta}_j q \Big) + \sum_{j=1}^{R} \Big(\Big(\tilde{\Delta}_{\mu_j} - 1 \Big) \alpha_j \tilde{\beta}_j q \Big) + \mathcal{G}^{\star}_{\geq 3} \right) \mathrm{d}x \right. \\
&\qquad \left. - \left(\sum_{j\in J_1} \left(\Delta_{\mu_j} \left(\Delta_{\nu_j} - 1\right) \alpha_j \beta_j v\right) + \sum_{j=1}^{R} \left(\left(\Delta_{\mu_j} - 1\right) \alpha_j \beta_j v \right) \right) \right| \\
&\leq \underbrace{\sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \sum_{j\in J_1} \left| \left(\Delta_{\nu_j} - 1\right) \alpha_j \left(\int_{I_m} \tilde{\beta}_j q\,\mathrm{d}x - \Delta_{\mu_j} \beta_j v \right) \right|}_{=:\Psi_1}
\end{aligned}$$

$$+ \underbrace{\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j=1}^{R}\left|\alpha_j\left(\int_{I_m}\left(\tilde{\Delta}_{\mu_j}-1\right)\tilde{\beta}_j q\,\mathrm{d}x-\left(\Delta_{\mu_j}-1\right)\beta_j v\right)\right|}_{=:\Psi_2}$$
$$+\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\left|\int_{I_m}\mathcal{G}^{\star}_{\geq 3}\,\mathrm{d}x\right|.$$

In the first term Ψ_1, we estimate the shift operator by $\|(\Delta_\nu - 1)\|_{\ell_1} \leq 2$. Further, we extend assumption 5.13 to this case, insert the definition of v and using integration by substitution we obtain that

$$\begin{aligned}\Psi_1 &= \sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j\in J_1}\left|\left(\Delta_{\nu_j}-1\right)\alpha_j\left(\int_{I_m}\tilde{\beta}_j q\,\mathrm{d}x-\Delta_{\mu_j}\beta_j v\right)\right|\\ &\leq C_1\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j\in J_1}\left|\int_{I_m}\tilde{\beta}_j(x)q(t,n,x)-\beta_j(m-\mu_j)q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\mathrm{d}x\right|\end{aligned}$$

holds. Rearranging, similarly to the technique used in the proof of theorem 4.5, and obtain that

$$\begin{aligned}\Psi_1 &\leq C_1\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j\in J_1}\left|\int_{I_m}\tilde{\beta}_j(x)q(t,n,x)-\beta_j(m-\mu_j)q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\mathrm{d}x\right|\\ &= C_1\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j\in J_1}\Bigg|-\int_{I_m}\tilde{\beta}_j(x)\left(q\left(t,n,x-\frac{\mu_j}{\Omega}\right)-q(t,n,x)\right)\mathrm{d}x\\ &\qquad\qquad+\int_{I_m}\left(\tilde{\beta}_j(x)-\beta_j(m-\mu_j)\right)q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\mathrm{d}x\Bigg|\\ &\leq C_1\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j\in J_1}\left|\int_{I_m}\tilde{\beta}_j(x)\left(q\left(t,n,x-\frac{\mu_j}{\Omega}\right)-q(t,n,x)\right)\mathrm{d}x\right|\\ &+ C_1\sum_{n\in\mathbb{N}_0^d}\sum_{m\in\mathbb{N}_0^D}\sum_{j\in J_1}\left|\int_{I_m}\left(\tilde{\beta}_j(x)-\beta_j(m-\mu_j)\right)q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\mathrm{d}x\right|.\end{aligned}$$

The first term is estimated with the help of assumption 6.3, because all reaction channels $\mathfrak{R}_j$ are in $j \in J_1$:

$$\left\|\int_{I_m}\tilde{\beta}_j(x)\left(q\left(t,n,x-\frac{\mu_j}{\Omega}\right)-q(t,n,x)\right)\mathrm{d}x\right\|_{\ell_1}\leq C\left\|\int_{I_m}q\left(t,n,x-\frac{\mu_j}{\Omega}\right)-q(t,n,x)\mathrm{d}x\right\|_{\ell_1}.$$

The second term can be estimated by

$$\begin{aligned}&\left|\int_{I_m}\left(\tilde{\beta}_j(x)-\beta_j(m-\mu_j)\right)q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\mathrm{d}x\right|\leq\\ &\int_{I_m}\left|\tilde{\beta}_j(x)-\beta_j(m-\mu_j)\right|\left|q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\right|\mathrm{d}x\leq\\ &\sup_{x\in I_m}\left|\tilde{\beta}_j(x)-\beta_j(m-\mu_j)\right|\int_{I_m}\left|q\left(t,n,x-\frac{\mu_j}{\Omega}\right)\right|\mathrm{d}x\end{aligned}$$

The estimate of the last term in the above expression is postponed to lemma C8, because it uses similar arguments as in the lemmata C5 and C6 and we do not want to interrupt the exposition ot the actual line of thought. Further, as in the proof of theorem 4.5 we assume that reaction channel $\mathfrak{R}_{j_1}$ is the dominant one in the sum over all j.

Because $\left\| \int_{I_m} |q(t,\cdot,\cdot)| \, \mathrm{d}x \right\|_{\ell_1} = 1$, we finally obtain

$$\Psi_1 \leq C_2 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \int_{I_m} \left| q\left(t,n,x-\frac{\mu_{j_1}}{\Omega}\right) - q(t,n,x) \right| \mathrm{d}x + \frac{C_3}{\Omega}$$

The estimation of the term Ψ_2 uses the same argument as in the proof of theorem 4.5, by extending assumption 5.13 to this case and by integration by substitution:

$$\begin{aligned}
\Psi_2 &= \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \sum_{j=1}^{R} \left| \alpha \left(\int_{I_m} \left(\tilde{\Delta}_\mu - 1 \right) \tilde{\beta} q \, \mathrm{d}x - \left(\Delta_\mu - 1 \right) \beta v \right) \right| \\
&\leq C_4 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \left(\int_{I_m} \left(\tilde{\Delta}_\mu - 1 \right) \tilde{\beta} q \, \mathrm{d}x - \left(\Delta_\mu - 1 \right) \beta v \right) \right| \\
&= C_4 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \Bigg| \int_{I_m} \tilde{\beta}_{j_1}\left(x-\frac{\mu_{j_1}}{\Omega}\right) q\left(t,n,x-\frac{\mu_{j_1}}{\Omega}\right) - \tilde{\beta}_{j_1}(x)\, q(t,n,x) \\
&\qquad\qquad - \beta_{j_1}(m-\mu_{j_1})\, q\left(t,n,x-\frac{\mu_{j_1}}{\Omega}\right) + \beta_{j_1}(m)\, q(t,n,x) \, \mathrm{d}x \Bigg| \\
&= C_4 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \Bigg| \int_{I_m} \left[\tilde{\beta}_{j_1}\left(x-\frac{\mu_{j_1}}{\Omega}\right) - \beta_{j_1}(m-\mu_{j_1}) \right] q\left(t,n,x-\frac{\mu_{j_1}}{\Omega}\right) \mathrm{d}x \\
&\qquad\qquad - \int_{I_m} \left[\tilde{\beta}_{j_1}(x) - \beta_{j_1}(m) \right] q(t,n,x) \, \mathrm{d}x \Bigg|
\end{aligned}$$

We represent the first term in $[\cdot]$-brakets by

$$\left[\tilde{\beta}_{j_1}\left(x-\frac{\mu_{j_1}}{\Omega}\right) - \beta_{j_1}(m-\mu_{j_1}) \right] = \left[\tilde{\beta}_{j_1}(x) - \beta_{j_1}(m) \right] + \mathcal{H}_{j_1}(m,x).$$

We rearrange and obtain

$$\begin{aligned}
\Psi_2 \leq\ & C_4 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \left[\tilde{\beta}_{j_1}(x) - \beta_{j_1}(m) \right] \left(q\left(t,n,x-\frac{\mu_{j_1}}{\Omega}\right) - q(t,n,x) \right) \mathrm{d}x \right| \\
& + C_4 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \mathcal{H}_{j_1}(m,x) q\left(t,n,x-\frac{\mu_{j_1}}{\Omega}\right) \mathrm{d}x \right|.
\end{aligned}$$

Then, analogously to the proof of theorem 4.5, we apply Hölders inequality. Further, we have already estimated

$$\sup_{x\in I_m} \left| \tilde{\beta}_{j_1}(x) - \beta_{j_1}(m) \right| \leq C_5, \qquad \sup_{x\in I_m} |\mathcal{H}_{j_1}(m,x)| \leq \frac{C_6}{\Omega},$$

in the lemmata C5 and C6, respectively. We apply these considerations and get

$$\Psi_2 \leq C_7 \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} q\left(t, n, x - \frac{\mu_{j_1}}{\Omega}\right) - q\left(t, n, x\right) \, \mathrm{d}x \right| + \frac{C_8}{\Omega}.$$

We derive by Taylor expansion that

$$q\left(t, n, x - \frac{\mu_{j_1}}{\Omega}\right) - q(t, x) = \frac{1}{\Omega} \nabla\Big(q(t, n, x)\Big)^T \mu_{j_1} + \mathcal{G}^{\dagger}_{j_1,\geq 2},$$

with remainder term given by

$$\mathcal{G}^{\dagger}_{j_1,\geq 2} := - \sum_{|\boldsymbol{k}|\geq 2} \frac{1}{\boldsymbol{k}!} \left(-\frac{\mu_{j_1}}{\Omega}\right)^{\boldsymbol{k}} \nabla^{\boldsymbol{k}}\Big(q(t, n, x)\Big). \tag{6.10}$$

Putting now all the pieces together, we obtain

$$\begin{aligned}
\|\mathcal{R}(s,\cdot,\cdot)\|_{\ell_1} \leq & \underbrace{\sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \sum_{j\in J_1} \left| (\Delta_\nu - 1)\, \alpha \left(\int_{I_m} \tilde{\beta} q \, \mathrm{d}x - \Delta_\mu \beta v \right) \right|}_{=:\Psi_1} \\
& + \underbrace{\sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \sum_{j=1}^{R} \left| \alpha \left(\int_{I_m} \left(\tilde{\Delta}_\mu - 1\right) \tilde{\beta} q \, \mathrm{d}x - (\Delta_\mu - 1)\, \beta v \right) \right|}_{=:\Psi_2} \\
& + \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \mathcal{G}^{\star}_{\geq 3} \, \mathrm{d}x \right| \\
\leq & \frac{C_9}{\Omega} \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \int_{I_m} \left| \nabla\Big(q(t, n, x)\Big) \right| \, \mathrm{d}x \\
& + \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \mathcal{G}^{\star}_{\geq 3} \, \mathrm{d}x \right| + \sum_{n\in\mathbb{N}_0^d} \sum_{m\in\mathbb{N}_0^D} \left| \int_{I_m} \mathcal{G}^{\dagger}_{j_1,\geq 2} \, \mathrm{d}x \right| + \frac{C_{10}}{\Omega}.
\end{aligned}$$

We note that, the remainder terms $\mathcal{G}^{\star}_{\geq 3}$ and $\mathcal{G}^{\dagger}_{j_1,\geq 2}$ are estimated using assumption 6.4. Finally, by applying the definition of the $\|\cdot\|_1$-norm in eq. (6.9) and by taking the boundedness of the time interval into account we prove the assertion:

$$\begin{aligned}
\left\| p(t,\cdot,\cdot) - \int_{I_m} q(t,\cdot,x) \, \mathrm{d}x \right\|_{\ell_1} &\leq \int_0^t \left[\frac{C_9}{\Omega} \left\| \nabla\Big(q(s,\cdot,\cdot)\Big) \right\|_1 + \frac{C_{11}}{\Omega} \right] \mathrm{d}s \\
&\leq \frac{C_{12}}{\Omega} \left\| \nabla\Big(q(t,\cdot,\cdot)\Big) \right\|_1 + \frac{C_{13}}{\Omega},
\end{aligned}$$

with C_{12}, C_{13} constants. These constants can be found because the time interval as well as the norm are bounded. Obviously they depend on the chosen time point t. □

6.6 Summary and Discussion

We started this chapter with the goal to derive a hybrid method with "better" convergence properties than the Liouville master equation (LME) analysed in chapter 5. To achieve

this goal we combined the CME with the FPE and derived the Fokker-Planck master equation (FPME)

$$\begin{aligned}\partial_t q\left(t,n,x\right) = &\sum_{j\in J_1} \tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\left(t,n-\nu_j,x\right) - \alpha_j(n)q\left(t,n,x\right)\Big) \\ &- \frac{1}{\Omega}\sum_{j=1}^{R} \alpha_j(n)\nabla\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)^T \mu_j \\ &+ \frac{1}{2\Omega^2}\sum_{j=1}^{R} \alpha_j(n)\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)\mu_j.\end{aligned}$$

This equation extends the LME by adding a diffusion term. The construction of the hybrid model was completed by the derivation of the hybrid process in def. 6.1.

We have already seen in section 4.4 that the solution of the FPE converges to the solution of the CME and therefore we analysed an example and found that the solution of the FPME converged to the solution of the CME. This result motivated us to use the techniques developed in the preliminarily discussion of the FPE (cf. theorem 4.5) to prove convergence of the FPME to the CME:

$$\left\|p(t,\cdot,\cdot) - \int_{I_m} q(t,\cdot,x)\,\mathrm{d}x\right\|_{\ell_1} \leq \frac{C_1}{\Omega}\left\|\nabla\Big(q(t,\cdot,\cdot)\Big)\right\|_1 + \frac{C_2}{\Omega}.$$

However, the observation in the example in table 6.1 that the derivation term is of order

$$\frac{C_1}{\Omega}\left\|\nabla\Big(q(t,\cdot,\cdot)\Big)\right\|_1 \sim \mathcal{O}\Big(\Omega^{-1/2}\Big)$$

should be analysed in future studies.

7 | The Lac Operon - A 10-Dimensional Numerical Example

7.1 The Lac Operon Network

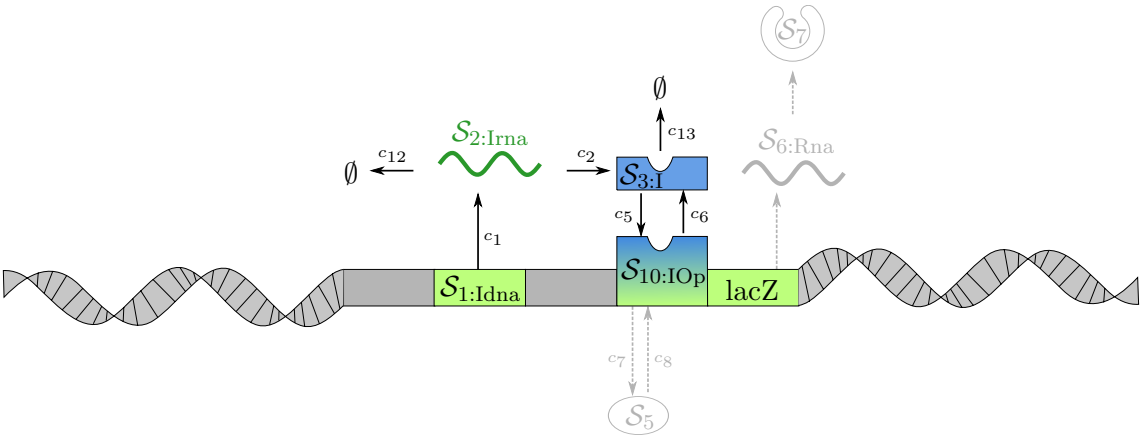
François Jacob and Jacques Monod published in 1961 a seminal paper with their discoveries on genetic regulatory principles in bacteria [JM61]. In 1965 they were awarded the Nobel prize for their work [Nob65]. Jacob and Monod developed a gene regulatory network for the so-called lac operon. The network describes the up and down regulation of a Lactose exploting pathway in *Escherichia coli* (*E. coli*).

The bacterium *E. coli* lives in the intestine of mammals. Normally, *E. coli* feeds on the monosaccharide glucose, but it can also progress the disaccharide lactose, if present. The bacterium transports lactose inside the cell using the membrane-protein β-galactoside permease, then it reduces the disaccharide lactose to the monosaccharides galactose and glucose using the enzyme β-galactosidase. These sugars are then processed in further metabolic pathways. The enzyme is also capable of turning lactose into allolactose. However, the proteins β-galactoside permease and β-galactosidase are only in the presence of lactose produced in high copy numbers, otherwise the genes are down regulated by a repressor protein.

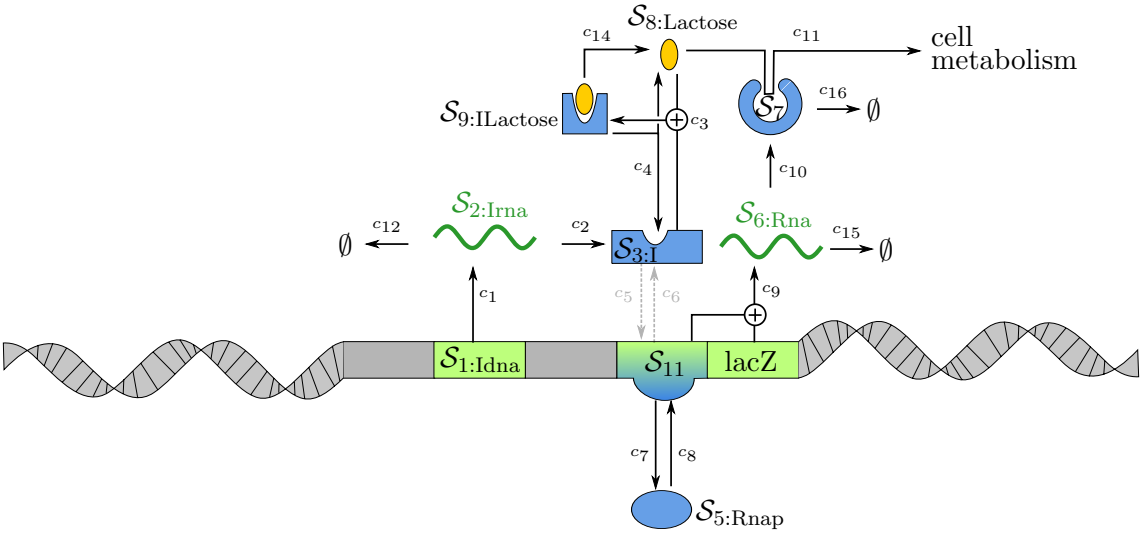
The lac operon is structured as follows. On the 5' end of the operon lies the promotor, the binding site of the RNA polymerase protein. The promotor is followed downstream by three operators. These are directly followed by the genes lacZ, which codes for the β-galactosidase, lacY, which codes for the β-galactoside permease, and lacA, which encodes the enzyme β-galactoside transacetylase. This enzyme seems not to be important for the lactose metabolism of *E. coli*. However, there are some hints that the protein is decomposing some potential poisoning side products of the metabolism. The operon is closed with a terminator on the 3' end.

Upstream of the lac operon the gene lacI, together with its promotor and terminator region, is located. This gene codes for the repressor protein of the lac operon. The repressor binds on the operator of the lac operon and blocks the RNA-polymerase.

(a) The simplified structure of the lac operon with operator ($\mathcal{S}_{4:\mathrm{Op}}$) and structure gene lacZ. The gene lacI ($\mathcal{S}_{1:\mathrm{Idna}}$) is located upstream of the operon.



(b) The lac operon network with no lactose present.



(c) The lac operon network with lactose (yellow) present.

Figure 7.1: Visualisation of the lac operon network

If *E. coli* enters a medium containing lactose, some of the molecules are transported inside the cell through the β-galactoside permease symporter, and some are transformed into allolactose by β-galactosidase. These two enzymes are always present in low copy numbers, because the repressor protein is detaching the operator from time to time. The allolactose is inhibiting the repressor protein. Therefor the probability for the lac operon being repressed goes down by a factor of order 10^3. The structural genes lacZ, lacY and lacA are now no longer down regulated and their proteins are produced in high copy numbers.

When all lactose is processed, allolactose cannot be produced any longer and the repressor proteins are no longer inhibited. Therefor the repressor binds again to the operator and the operon is down regulated again.

In a second inhibiting process, glucose inhibits the transport of lactose through β-galactoside permease.

In summary, these processes make sure that the cell only produces the proteins encoded by the lac operon if lactose is present and if no glucose is present. The enzyme β-galactosidase is also capable of processing other galactosides. These occur, for example, when animals regenerate the cell layers in their intestine. It seems that these galactosides are also inhibiting the repressor protein [BTS02].

Table 7.1: Species names, symbols and initial conditions for the lac operon network.

species	**description**	**network symbol**	**model symbol**	**initial condition**
lacI	DNA of the lacI gene	$\mathcal{S}_{1:\mathrm{Idna}}$	n_1	1
Irna	RNA transcribed from lacI	$\mathcal{S}_{2:\mathrm{Irna}}$	n_2	0
I	repressor protein	$\mathcal{S}_{3:\mathrm{I}}$	n_3	50
Op	operator of the lac operon	$\mathcal{S}_{4:\mathrm{Op}}$	n_4	1
Rnap	RNA polymerase	$\mathcal{S}_{5:\mathrm{Rnap}}$	n_5	100
Rna	RNA transcribed from lacZ	$\mathcal{S}_{6:\mathrm{Rna}}$	n_6	0
Z	β-galactosidase	$\mathcal{S}_{7:\mathrm{Z}}$	n_7	0
Lactose	lactose	$\mathcal{S}_{8:\mathrm{Lactose}}$	m_8, x_8	Ω
ILactose	lactose inhibiting I	$\mathcal{S}_{9:\mathrm{ILactose}}$	n_9	0
IOp	I repressing the operator	$\mathcal{S}_{10:\mathrm{IOp}}$	n_{10}	0
RnapOp	RNA polymerase bound to the operator	$\mathcal{S}_{11:\mathrm{RnapOp}}$	n_{11}	0

In this work, we simulate a simplified version of the lac operon network published by Wilkinson [Wil06]. The gene regulatory network consists of 11 species and 16 reactions. The names and symbols of the different species are listed in table 7.1. The reactions, together with the propensity functions and rate constants, are listed in table 7.3. In this network, the lac operon is reduced to the gene lacZ and one operator site ($\mathcal{S}_{4:\mathrm{Op}}$). The gene lacI ($\mathcal{S}_{1:\mathrm{Idna}}$) is still located upstream of the lac operon. The gene lacY is removed from the network, because we are not interested in spatial phenomenons and therefor we will not simulate transporter proteins. The gene lacA is removed because of the unknown role of the protein β-galactoside transacetylase. The DNA is visualised in figure 7.1a.

In the absence of lactose (fig. 7.1b) the gene lacI ($\mathcal{S}_{1:\mathrm{Idna}}$) is translated via the reactions $\mathfrak{R}_1$ and $\mathfrak{R}_2$ and the repressor protein ($\mathcal{S}_{3:\mathrm{I}}$) binds on the operator via reaction $\mathfrak{R}_5$. The resulting species is named $\mathcal{S}_{10:\mathrm{IOp}}$. This bound is reversible ($\mathfrak{R}_6$). As a result the RNA polymerase cannot longer transcribe lacZ and no β-galactosidase is translated.

If lactose is present (fig. 7.1b), the repressor protein can also react with lactose ($\mathfrak{R}_3$) and becomes inhibited ($\mathcal{S}_{9:\mathrm{ILactose}}$). Now, the RNA-Polymerase binds on the operator and β-galactosidase ($\mathcal{S}_{7:\mathrm{Z}}$) is translated ($\mathfrak{R}_9$ and $\mathfrak{R}_{10}$). The binding of the RNA-polymerase

Table 7.3: The lactose operon network.

Reaction Channel $\mathfrak{R}_j$				Propensity $\alpha_j(n)$	Propensity $\beta_j(m)$	Propensity $\tilde{\beta}_j(x)$	Rate c_j
$\mathfrak{R}_1$:	$\mathcal{S}_{1:\mathrm{Idna}}$	$\xrightarrow{c_1}$	$\mathcal{S}_{1:\mathrm{Idna}} + \mathcal{S}_{2:\mathrm{Irna}}$	$c_1 n_1$	1	1	0.02
$\mathfrak{R}_2$:	$\mathcal{S}_{2:\mathrm{Irna}}$	$\xrightarrow{c_2}$	$\mathcal{S}_{2:\mathrm{Irna}} + \mathcal{S}_{3:\mathrm{I}}$	$c_2 n_2$	1	1	0.1
$\mathfrak{R}_3$:	$\mathcal{S}_{3:\mathrm{I}} + \mathcal{S}_{8:\mathrm{Lactose}}$	$\xrightarrow{c_3}$	$\mathcal{S}_{9:\mathrm{ILactose}}$	$c_3 n_3$	m_8/Ω	x_8	0.005
$\mathfrak{R}_4$:	$\mathcal{S}_{9:\mathrm{ILactose}}$	$\xrightarrow{c_4}$	$\mathcal{S}_{8:\mathrm{Lactose}} + \mathcal{S}_{3:\mathrm{I}}$	$c_4 n_9$	1	1	0.1
$\mathfrak{R}_5$:	$\mathcal{S}_{3:\mathrm{I}} + \mathcal{S}_{4:\mathrm{Op}}$	$\xrightarrow{c_5}$	$\mathcal{S}_{10:\mathrm{IOp}}$	$c_5 n_3 n_4$	1	1	1.0
$\mathfrak{R}_6$:	$\mathcal{S}_{10:\mathrm{IOp}}$	$\xrightarrow{c_6}$	$\mathcal{S}_{3:\mathrm{I}} + \mathcal{S}_{4:\mathrm{Op}}$	$c_6 n_{10}$	1	1	0.01
$\mathfrak{R}_7$:	$\mathcal{S}_{4:\mathrm{Op}} + \mathcal{S}_{5:\mathrm{Rnap}}$	$\xrightarrow{c_7}$	$\mathcal{S}_{11:\mathrm{RnapOp}}$	$c_7 n_4 n_5$	1	1	0.1
$\mathfrak{R}_8$:	$\mathcal{S}_{11:\mathrm{RnapOp}}$	$\xrightarrow{c_8}$	$\mathcal{S}_{5:\mathrm{Rnap}} + \mathcal{S}_{4:\mathrm{Op}}$	$c_8 n_{11}$	1	1	0.01
$\mathfrak{R}_9$:	$\mathcal{S}_{11:\mathrm{RnapOp}}$	$\xrightarrow{c_9}$	$\mathcal{S}_{4:\mathrm{Op}} + \mathcal{S}_{5:\mathrm{Rnap}} + \mathcal{S}_{6:\mathrm{Rna}}$	$c_9 n_{11}$	1	1	0.03
$\mathfrak{R}_{10}$:	$\mathcal{S}_{6:\mathrm{Rna}}$	$\xrightarrow{c_{10}}$	$\mathcal{S}_{6:\mathrm{Rna}} + \mathcal{S}_{7:\mathrm{Z}}$	$c_{10} n_6$	1	1	0.1
$\mathfrak{R}_{11}$:	$\mathcal{S}_{8:\mathrm{Lactose}} + \mathcal{S}_{7:\mathrm{Z}}$	$\xrightarrow{c_{11}}$	$\mathcal{S}_{7:\mathrm{Z}}$	$c_{11} n_7$	m_8	Ωx_8	10^{-5}
$\mathfrak{R}_{12}$:	$\mathcal{S}_{2:\mathrm{Irna}}$	$\xrightarrow{c_{12}}$	$\emptyset$	$c_{12} n_2$	1	1	0.01
$\mathfrak{R}_{13}$:	$\mathcal{S}_{3:\mathrm{I}}$	$\xrightarrow{c_{13}}$	$\emptyset$	$c_{13} n_3$	1	1	0.002
$\mathfrak{R}_{14}$:	$\mathcal{S}_{9:\mathrm{ILactose}}$	$\xrightarrow{c_{14}}$	$\mathcal{S}_{8:\mathrm{Lactose}}$	$c_{14} n_9$	1	1	0.002
$\mathfrak{R}_{15}$:	$\mathcal{S}_{6:\mathrm{Rna}}$	$\xrightarrow{c_{15}}$	$\emptyset$	$c_{15} n_6$	1	1	0.01
$\mathfrak{R}_{16}$:	$\mathcal{S}_{7:\mathrm{Z}}$	$\xrightarrow{c_{16}}$	$\emptyset$	$c_{16} n_7$	1	1	0.001

is modeled in a simplified way, because the operator and promoter are combined to a single species $\mathcal{S}_{4:\mathrm{Op}}$. The transcription process is also simplified to a single reaction $\mathfrak{R}_9$. In this reaction the combined RNA-polymerase / operator species $\mathcal{S}_{11:\mathrm{RnapOp}}$ is producing the RNA for the translation of β-galactosidase and the combined species separates into its components. The enzyme $\mathcal{S}_{7:\mathrm{Z}}$ splits lactose ($\mathcal{S}_{8:\mathrm{Lactose}}$) up ($\mathfrak{R}_{11}$) and the resulting monosaccharides are delivered to further pathways in the cell metabolism. However, we are not interested in these pathways, therefor reaction $\mathfrak{R}_{11}$ is modeled as a death process that decomposes lactose using the enzyme $\mathcal{S}_{7:\mathrm{Z}}$ as a catalyst. The network is completed by several reverse reactions and decomposing reactions for the RNA molecules and proteins.

Although the network consists of 11 species, we consider the network as 10-dimensional. Species $\mathcal{S}_{1:\mathrm{Idna}}$ should be interpreted as a parameter, because only one copy of the gene is present and there are no reaction channels that change the value of $X_1(t)$. Therefor the number does not change and only one particle is present over the whole time interval:

$$X_1(t) = X_1(0) = 1.$$

Thus, we will not visualise this species in any of the following plots.

7.2 Numerical Simulation of the Lac Operon Network with SSA and the two Hybrid Models

We simulate the reaction network in table 7.3 on the time interval

$$t \in \left[0, 10^3\right]$$

with initial conditions given in table 7.1. We choose a scaling parameter

$$\Omega = 10^4.$$

This is also the initial number of lactose molecules present.

The network is partitioned into the two sets

$$\begin{aligned} S_1 &= \{\mathcal{S}_{1:\mathrm{Idna}}, \ldots, \mathcal{S}_{7:\mathrm{Z}}, \quad \mathcal{S}_{9:\mathrm{ILactose}}, \ldots, \mathcal{S}_{11:\mathrm{RnapOp}}\}, \\ S_2 &= \{\mathcal{S}_{8:\mathrm{Lactose}}\}. \end{aligned}$$

The species in set S_1 are modeled as discrete state variables, the species in set S_2 as continuous state variables.

The network is sampled using three different methods:

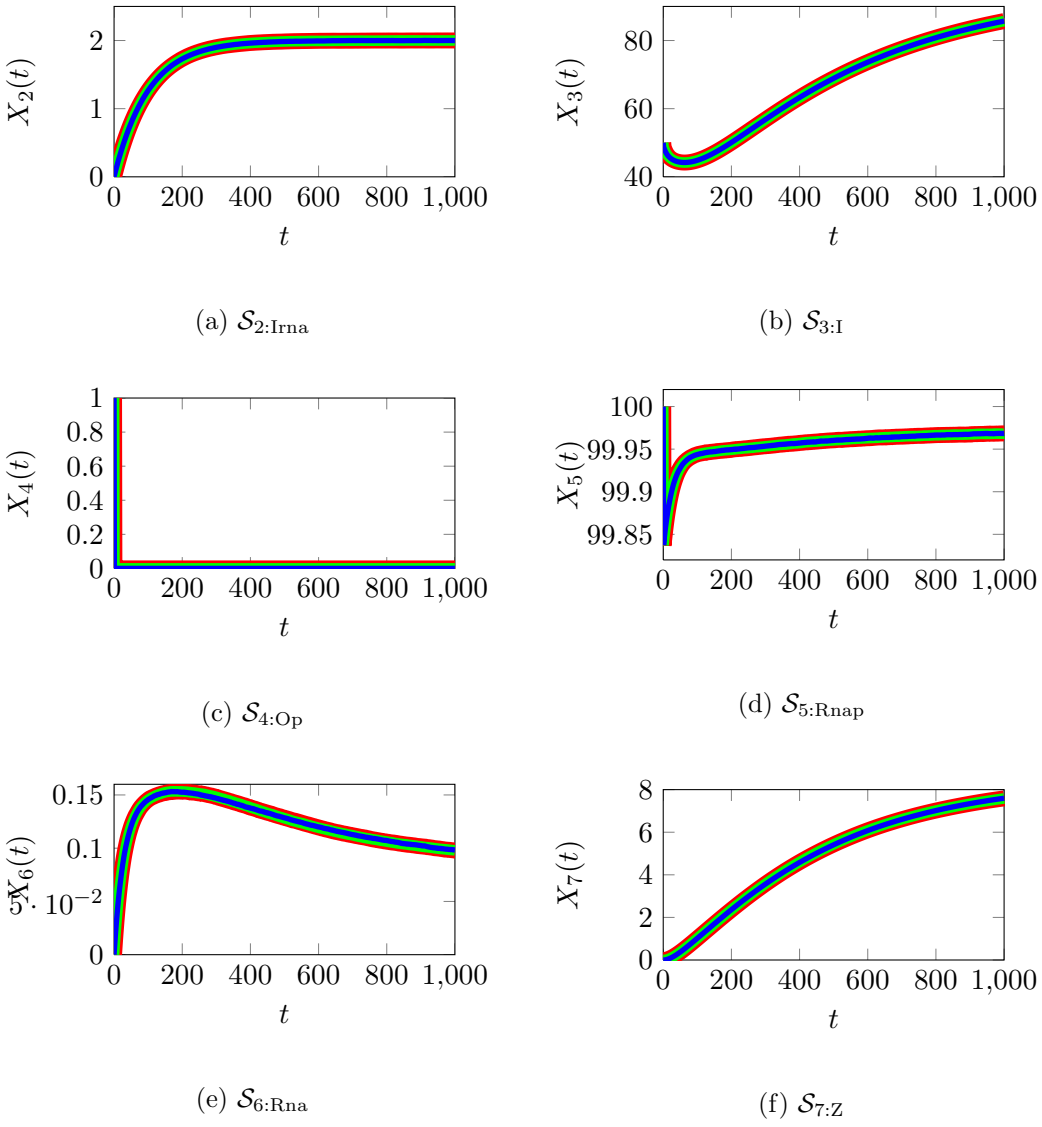
1. SSA (cf. algorithm 3.1) to approximate a solution of the partitioned CME (cf. eq. (5.10)),

2. Strang splitting for the first hybrid model (cf. algorithm 5.1) to approximate a solution of the LME (cf. def 5.7).

3. Strang splitting for the second hybrid model (cf. algorithm 6.1) to approximate a solution of the FPME (6.2).

For each method several trajectories are generated and averaged to approximate the expectation over time for each species. Furthermore the values of the trajectories at the end of the time interval are used to generate histograms to approximate the marginal distributions of each species. A full PDF of the solution is not approximated, because it is not possible to visualise a 10-dimensional PDF in a 2-dimensional figure.

We simulate the network in the state immediately after a large amount of lactose was added. It would also be interesting to "see" a simulation of the network with an initial number of zero lactose molecules and to see the response of the network if this value is increased after a certain time point. However, we like to simulate the network using the two hybrid models derived in the chapters 5 and 6. Therefor we need a scale difference between the species and so cannot simulate the network with a low copy number of lactose.



(a) $\mathcal{S}_{2:\mathrm{Irna}}$ (b) $\mathcal{S}_{3:\mathrm{I}}$

(c) $\mathcal{S}_{4:\mathrm{Op}}$ (d) $\mathcal{S}_{5:\mathrm{Rnap}}$

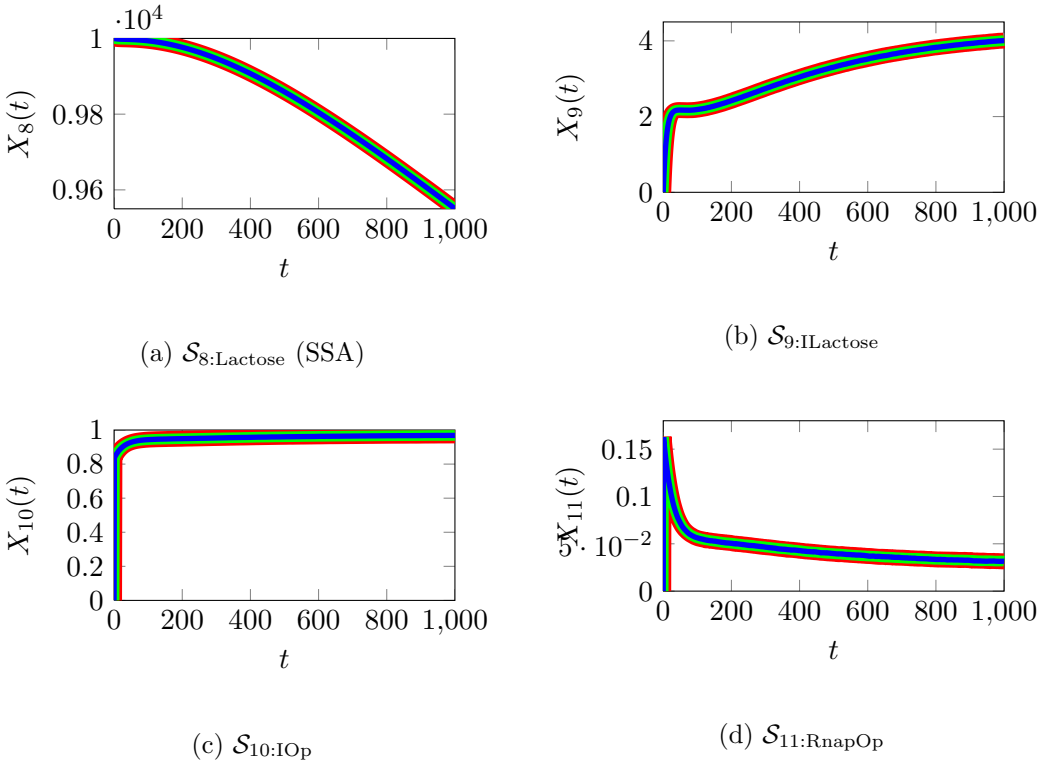
(e) $\mathcal{S}_{6:\mathrm{Rna}}$ (f) $\mathcal{S}_{7:\mathrm{Z}}$

Figure 7.2: Average trajectory of the 10 species of the lac operon reaction network defined in table 7.3. For each species the results of the SSA (red), the Kurtz process / RRE hybrid model (green) and the Kurtz process / CLE hybrid model (blue) are plotted. The averaged trajectory for species $\mathcal{S}_{8:\mathrm{Lactose}}$ was scaled with Ω for the hybrid models. This allows a direct visual comparison of the results. The average was taken over $3.2 \cdot 10^6$ realisations for the SSA, $3.2 \cdot 10^6$ realisations for the Kurtz process / RRE hybrid model and $1.5 \cdot 10^6$ realisations for the Kurtz process / CLE hybrid model. This figure is continued in figure 7.3.

7.3 Results and Discussion of the Numerical Simulations

We generate, for the lac operon network in table 7.3, $3.2 \cdot 10^6$ realisations with the SSA, $3.2 \cdot 10^6$ realisations with the Kurtz / RRE hybrid method and $1.5 \cdot 10^6$ realisations* with the Kurtz / CLE hybrid method for the time interval $[0, 10^3]$. The hybrid methods are simulated using a time step of $\Delta t = 1$ for the Strang Splitting scheme and a time step of $\tau = 0.1$ for the ODE and SDE integrator, respectively.

*Due to technical difficulties (the computer broke down irreversibly), we generated less trajectories for the second hybrid model.



(a) $\mathcal{S}_{8:\text{Lactose}}$ (SSA)

(b) $\mathcal{S}_{9:\text{ILactose}}$

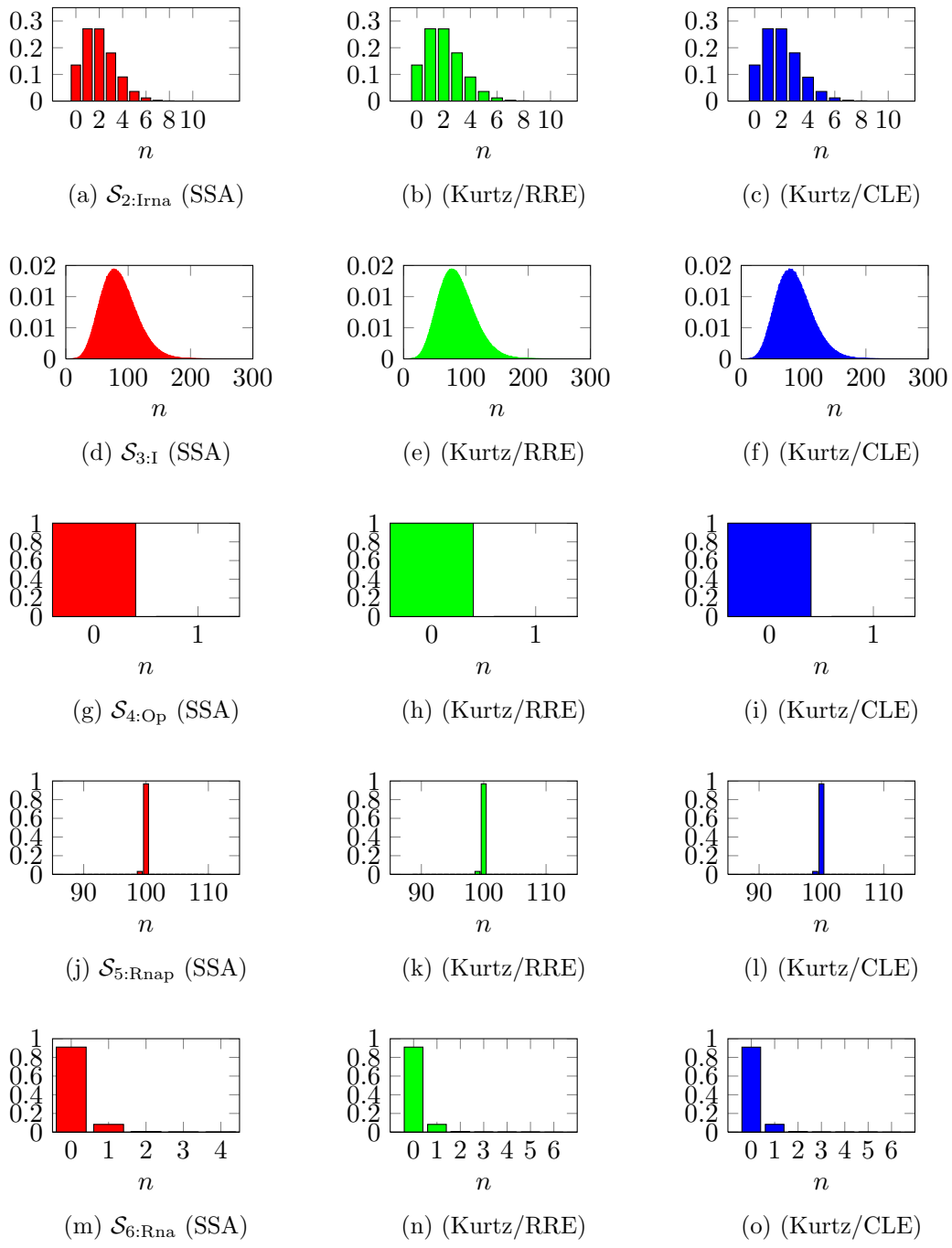
(c) $\mathcal{S}_{10:\text{IOp}}$

(d) $\mathcal{S}_{11:\text{RnapOp}}$

Figure 7.3: Continuation of figure 7.2.

Figure 7.2 shows the averaged trajectories of these simulations. As we can see the trajectories are visually not distinguishable for all three models. The histograms for each species can be found in the figure 7.4 for the SSA and the two hybrid models.

The average trajectory of species $\mathcal{S}_{2:\text{Irna}}$ grows from its initial value 0 to a value of 2 particles and plateaus there. However, the histograms show that also higher values are reached. This behaviour result completely from the reaction channels $\mathfrak{R}_1$ and $\mathfrak{R}_{12}$. The first reaction models the transcription of the RNA while the second reaction models its decomposition. Species $\mathcal{S}_{3:\text{I}}$ is the repressor protein and is produced via the translation reaction $\mathfrak{R}_2$. The protein can then block the operator of the lacZ gene via $\mathfrak{R}_5$. This reaction is reversed in channel $\mathfrak{R}_6$. Alternatively, the repressor can be inhibited by lactose via $\mathfrak{R}_3$ (with reverse reaction $\mathfrak{R}_4$). Finally, the repressor protein is decomposed in reaction channel $\mathfrak{R}_{13}$. In the averaged trajectory (cf. fig. 7.2b), we observe that the initial amount of 50 molecules is reduced at first due to the high lactose number in the reaction volume. When the amount of lactose is decreasing the particle number of $\mathcal{S}_{3:\text{I}}$ molecules increases. This is due to the unregulated transcription and translation of its DNA (species $\mathcal{S}_{1:\text{Idna}}$) and because the number of molecules it can interact with (the number of lactose plus one operator side) is decreasing. The histogram of $\mathcal{S}_{3:\text{I}}$ shows that this species reaches a wide spectrum of states. The operator species $\mathcal{S}_{4:\text{Op}}$ interacts via the reactions $\mathfrak{R}_5$ and $\mathfrak{R}_6$ with the repressor protein. Further, it reacts via the channels $\mathfrak{R}_7$, $\mathfrak{R}_8$ and $\mathfrak{R}_9$ with the RNA polymerase. The averaged trajectory of the operator (cf. fig. 7.2c) drops immediately from the initial value 1 to zero. The operator is, in average, bound to the repressor protein or the RNA polymerase over the whole time interval. The histogram also reflects this



(a) $\mathcal{S}_{2:\mathrm{Irna}}$ (SSA) (b) (Kurtz/RRE) (c) (Kurtz/CLE)

(d) $\mathcal{S}_{3:\mathrm{I}}$ (SSA) (e) (Kurtz/RRE) (f) (Kurtz/CLE)

(g) $\mathcal{S}_{4:\mathrm{Op}}$ (SSA) (h) (Kurtz/RRE) (i) (Kurtz/CLE)

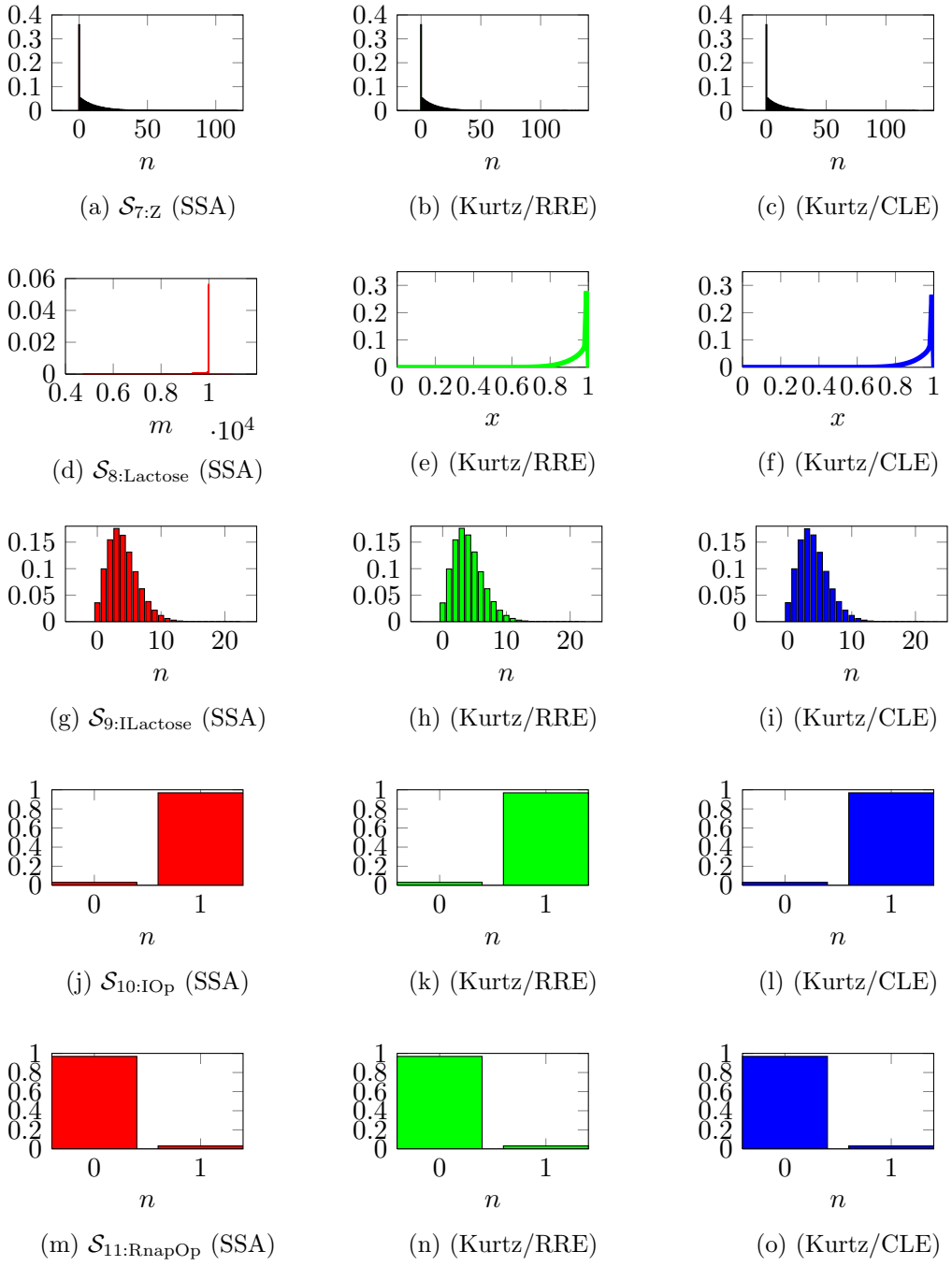
(j) $\mathcal{S}_{5:\mathrm{Rnap}}$ (SSA) (k) (Kurtz/RRE) (l) (Kurtz/CLE)

(m) $\mathcal{S}_{6:\mathrm{Rna}}$ (SSA) (n) (Kurtz/RRE) (o) (Kurtz/CLE)

Figure 7.4: Histograms of the marginal distributions of the 10 species of the lac operon reaction network defined in table 7.3. The underlying trajectories were generated using $3.2 \cdot 10^6$ realisations of the SSA (left column, red), $3.2 \cdot 10^6$ realisations of the Kurtz / RRE hybrid algorithm (middle column, green) and $1.5 \cdot 10^6$ realisations of the Kurtz / CLE hybrid algorithm (right column, blue). This figure is continued in figure 7.5.

behaviour and shows a probability of nearly one for the operator bound at the end of the time interval.

The number of RNA polymerase molecules (species $\mathcal{S}_{5:\mathrm{Rnap}}$) drops off at the beginning of the time interval. This species only interacts with the operator side via the reaction channels $\mathfrak{R}_7$, $\mathfrak{R}_8$ and $\mathfrak{R}_9$ and has no birth or death reactions. The repressor protein is inhibited



(a) $\mathcal{S}_{7:\mathrm{Z}}$ (SSA) (b) (Kurtz/RRE) (c) (Kurtz/CLE)

(d) $\mathcal{S}_{8:\mathrm{Lactose}}$ (SSA) (e) (Kurtz/RRE) (f) (Kurtz/CLE)

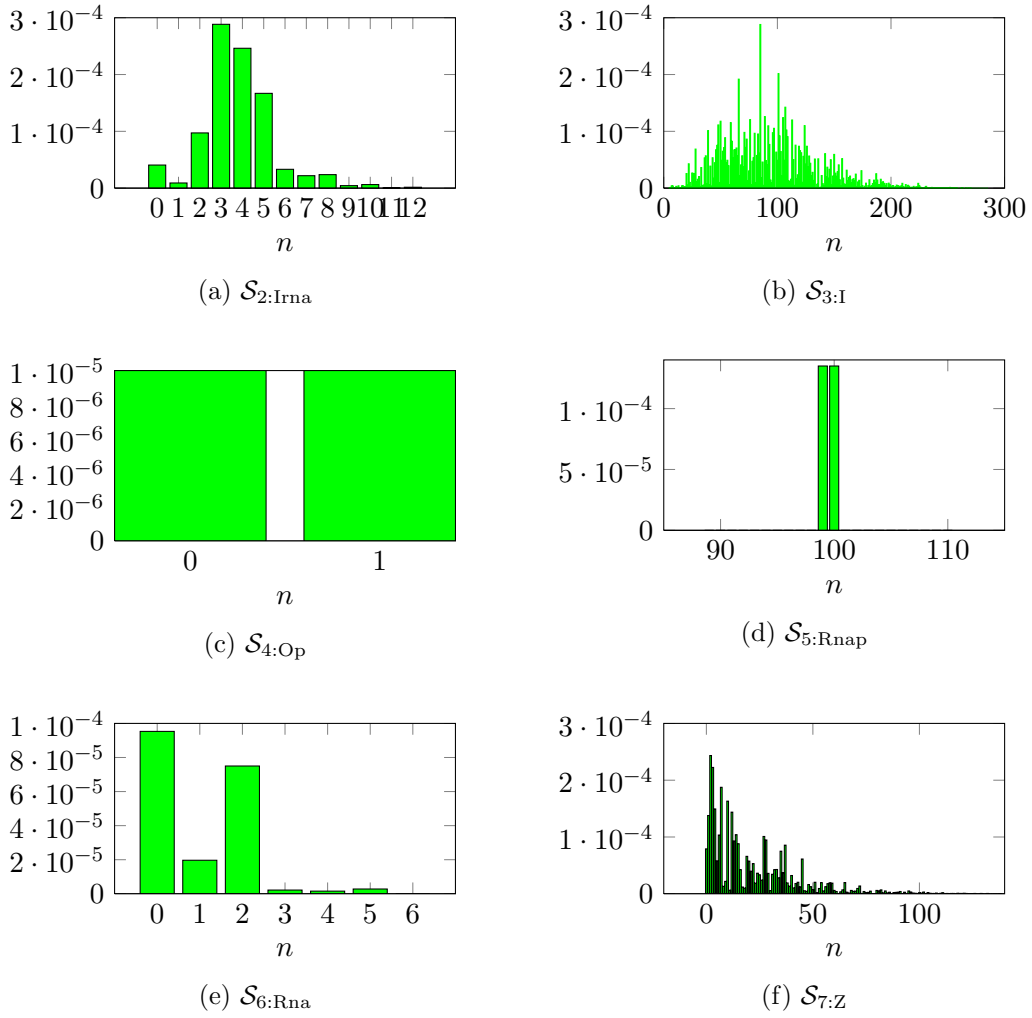
(g) $\mathcal{S}_{9:\mathrm{ILactose}}$ (SSA) (h) (Kurtz/RRE) (i) (Kurtz/CLE)

(j) $\mathcal{S}_{10:\mathrm{IOp}}$ (SSA) (k) (Kurtz/RRE) (l) (Kurtz/CLE)

(m) $\mathcal{S}_{11:\mathrm{RnapOp}}$ (SSA) (n) (Kurtz/RRE) (o) (Kurtz/CLE)

Figure 7.5: Continuation of figure 7.4.

by Lactose, therefor RNA polymerase binds to the operator and lacZ is transcribed. The probability of the repressor protein inhibition through lactose sinks with declining lactose numbers and it becomes more likely that the operator is blocked by the repressor protein. As a result the level of "free" RNA polymerase molecules rises. The histogram shows that most of the RNA polymerase molecules are unbound at the end of the time interval. The inhibition of the repressor protein results also in an increase of the lacZ RNA (species $\mathcal{S}_{6:\mathrm{Rna}}$, cf. fig 7.2e). The amount of $\mathcal{S}_{6:\mathrm{Rna}}$ declines later in the time interval, due to the decompositon of lactose. The lacZ gene is no longer transcribed (cf. reaction $\mathfrak{R}_9$), because the docking of RNA polymerase to the operator side is inhibited by the repressor protein.



(a) $\mathcal{S}_{2:\mathrm{Irna}}$ (b) $\mathcal{S}_{3:\mathrm{I}}$

(c) $\mathcal{S}_{4:\mathrm{Op}}$ (d) $\mathcal{S}_{5:\mathrm{Rnap}}$

(e) $\mathcal{S}_{6:\mathrm{Rna}}$ (f) $\mathcal{S}_{7:\mathrm{Z}}$

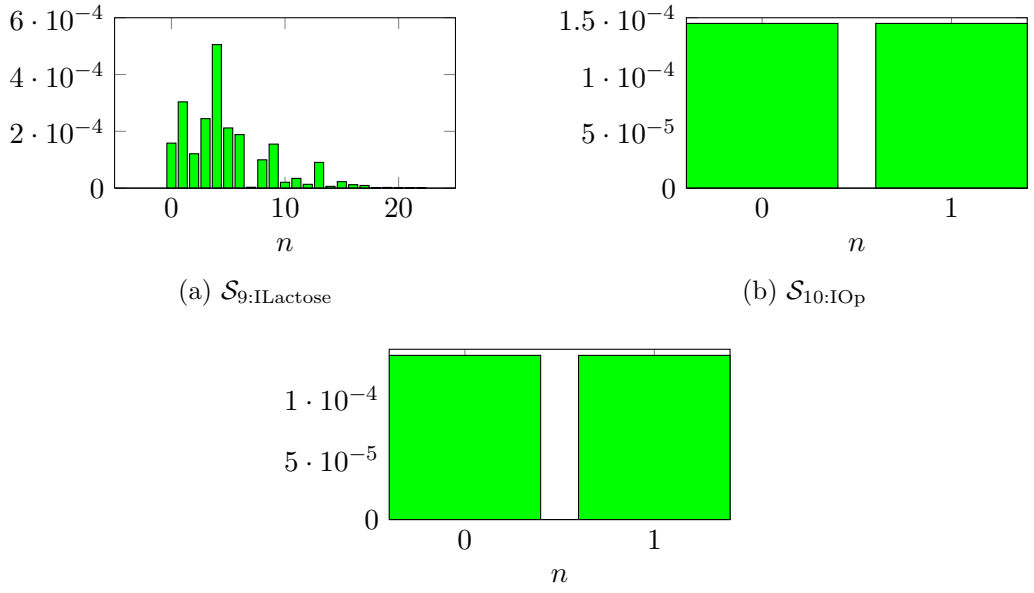
Figure 7.6: Error $\varepsilon_1^{(i)}(n)$ of the Kurtz / RRE hybrid model at time $t = 10^3$. The model was solved using $3.2 \cdot 10^6$ realisations of algorithm 5.1 to approximate the marginal distributions. Histograms from $3.2 \cdot 10^6$ SSA realisations were used as a reference solutions. This figure is continued in figure 7.7.

The influence of the decomposition reaction $\mathfrak{R}_{15}$ takes over and the averaged amount of species $\mathcal{S}_{6:\mathrm{Rna}}$ is falling. The amount of β-galactosidase (species $\mathcal{S}_{7:\mathrm{Z}}$) increases over the whole time interval. The protein is translated via reaction $\mathfrak{R}_{10}$ and decomposed via $\mathfrak{R}_{16}$. The molecule numbers are not declining, because the decomposition reaction has a very small rate ($c_{16} = 0.001$). The histogram shows some variability in the distribution of species $\mathcal{S}_{7:\mathrm{Z}}$.

The derivations in the chapters 5 and 6 teach us that

$$\mathbb{E}\left[Y^{(\mathrm{Kurtz})}(t)\right] \approx \Omega\mathbb{E}\left[y^{(\mathrm{Kurtz\ /\ RRE})}(t)\right] \approx \Omega\mathbb{E}\left[y^{(\mathrm{Kurtz\ /\ CLE})}(t)\right].$$

Motivated by this observation, we shift the averaged trajectories resulting from the two hybrid models by Ω for species $\mathcal{S}_{8:\mathrm{Lactose}}$ in figure 7.3a. This allows a direct visual comparison of the three models. We see that the initial amount of lactose is falling over the complete time interval. The slope is not very large at the beginning but increases over



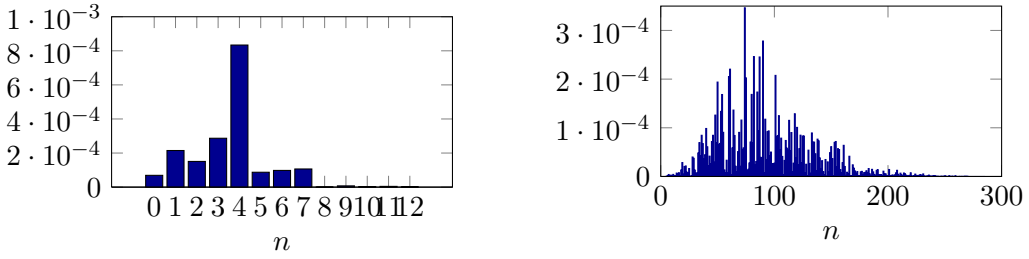
(a) $\mathcal{S}_{9:\mathrm{ILactose}}$ (b) $\mathcal{S}_{10:\mathrm{IOp}}$

(c) $\mathcal{S}_{11:\mathrm{RnapOp}}$

Figure 7.7: Continuation of figure 7.6.

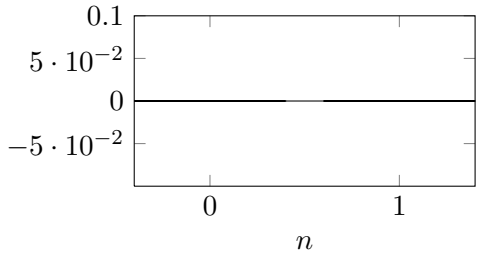
time. This can be interpreted as a small time interval that is needed by the reaction network to "adapt" to the new ressource. The lactose molecules are mostly reduced by β-galactosidase via $\mathfrak{R}_{11}$. The enzyme splits the disaccharide into glucose and galactose. For simplicity, these two monosaccharides are not modeled in this network. We display also the histograms for the marginal distribution of species $\mathcal{S}_{8:\mathrm{Lactose}}$ (cf. fig. 7.5d, fig. 7.5e and fig. 7.5f). In chapters 5.6, we have only shown convergence of the marginal distributions for the discrete species of the Kurtz / RRE hybrid model. However, we observe that for this network the shape of the distributions is conserved, but we have to keep in mind that figure 7.5d shows a histogram on a discrete state space with essential support located around Ω, while the corresponding plots for the hybrid models show histograms that approximate a PDF on a continuous state space with essential support located between zero and one (cf. fig. 7.5e and 7.5f).

The average trajectory for species $\mathcal{S}_{9:\mathrm{ILactose}}$ jumps, from its initial value zero, to a value of 2 very quickly and then grows to a value of 4 more slowly. The species models the inhibited repressor protein and is created via reaction $\mathfrak{R}_3$ and decomposed via the channels $\mathfrak{R}_4$ and $\mathfrak{R}_{14}$. The histogram shows, that the support at the end of the time interval covers a wider spectrum of states. The species $\mathcal{S}_{10:\mathrm{IOp}}$ increases very fast, in the first few time steps, and then slowly grows to a value shortly less than 1. This species models the operator blocked by the repressor protein. Because only one operator side is present, the species can not grow larger than 1. The operator is blocked via reaction $\mathfrak{R}_5$ and released via $\mathfrak{R}_6$. The last species $\mathcal{S}_{11:\mathrm{RnapOp}}$ models the species resulting from the binding of the RNA polymerase to the operator side. The polymerase bind via reaction $\mathfrak{R}_7$ and releases via $\mathfrak{R}_8$. In reaction channel $\mathfrak{R}_9$, the polymerase transcribes the lacZ gene and creates one copy of the β-galactosidase RNA. The averaged trajectory increases very quickly at the beginning of the time interval which correlates with the increased production of the RNA

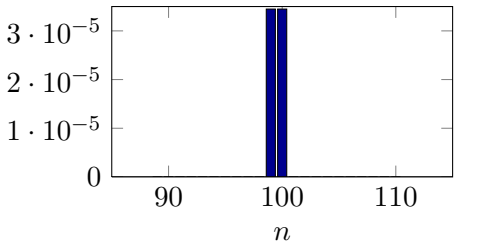


(a) $\mathcal{S}_{2:\mathrm{Irna}}$

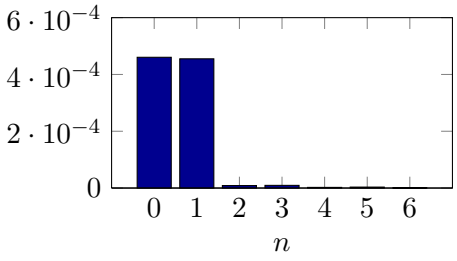
(b) $\mathcal{S}_{3:\mathrm{I}}$



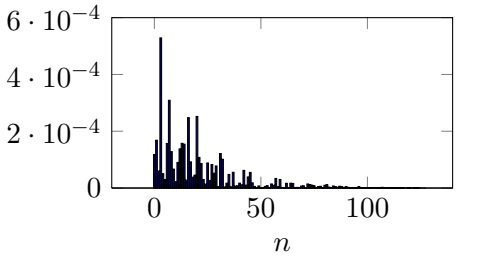
(c) $\mathcal{S}_{4:\mathrm{Op}}$



(d) $\mathcal{S}_{5:\mathrm{Rnap}}$



(e) $\mathcal{S}_{6:\mathrm{Rna}}$



(f) $\mathcal{S}_{7:\mathrm{Z}}$

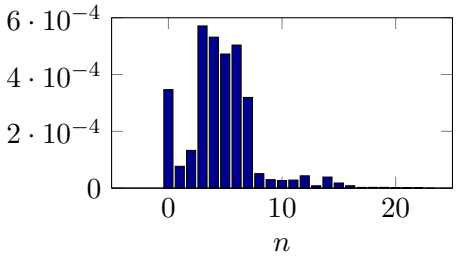
Figure 7.8: Error $\varepsilon_2^{(i)}(n)$ of the Kurtz / CLE hybrid model at time $t = 10^3$. The model was solved using $1.5 \cdot 10^6$ realisations of algorithm 6.1 to approximate the marginal distributions. Histograms from $3.2 \cdot 10^6$ SSA realisations were used as reference solutions. This figure is continued in figure 7.9.

at the beginng of the interval. After a few time steps the values decrease quickly and then decline slowly. It is very interesting to observe that this short burst of increased β-galactosidase RNA transcriptions is sufficient to start the lactose metabolism.

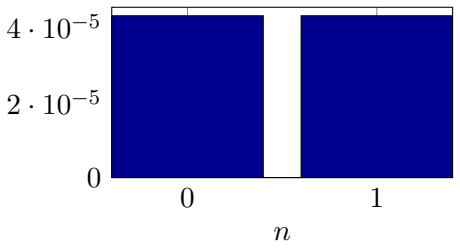
Figure 7.6 visualises the errors of the different marginal distributions by plotting the function

$$\varepsilon_1^{(i)}(n) = \left| p_1^{(i)}(t,n) - q_1^{(i)}(t,n) \right|,$$

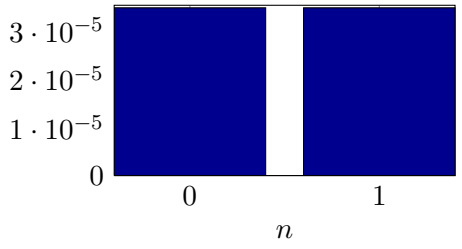
for each species $\mathcal{S}_i$. The functions $p_1^{(i)}(t,n)$ and $q_1^{(i)}(t,n)$ denote the marginal distribution corresponding to species $\mathcal{S}_i$, $i \in \{2,\ldots,7, \quad 9,\ldots,11\}$. The function $p_1^{(i)}(t,n)$ is derived from the solution of the CME and the function $q_1^{(i)}(t,n)$ from the solution of the LME. The CME solution is approximated by generating $3.2 \cdot 10^6$ realisation of the Kurtz process using SSA on the time interval $\left[0, 10^3\right]$. Then histograms for the marginal distributions are computed for the last time point $t = 10^3$ in the time interval. Analogously, the



(a) $\mathcal{S}_{9:\mathrm{ILactose}}$



(b) $\mathcal{S}_{10:\mathrm{IOp}}$



(c) $\mathcal{S}_{11:\mathrm{RnapOp}}$

Figure 7.9: Continuation of figure 7.8.

LME solution is computed from $3.2 \cdot 10^6$ realisations of the hybrid process generated by algorithm 5.1.

We also compute the analogue error terms for the second hybrid model. Unfortunately, it is not possible to visualise the error in the full 10-dimensional distribution here. Therefor we concentrate on the errors of the marginal distributions of the discrete species again and define the function

$$\varepsilon_2^{(i)}(n) = \left| p_1^{(i)}(t,n) - q_1^{(i)}(t,n) \right|,$$

for each species $\mathcal{S}_i$. The functions $p_1^{(i)}(t,n)$ and $q_1^{(i)}(t,n)$ denote the marginal distributions of the CME solution and the FPME solution, respectively. The functions $p_1^{(i)}$ and $q_1^{(i)}$ are approximated by histograms from $3.2 \cdot 10^6$ SSA realisations and $1.5 \cdot 10^6$ realisations of the hybrid algorithm 6.1. The error functions $\varepsilon_2^{(i)}(n)$ are plotted in figure 7.8. We observe an error of zero for the species $\mathcal{S}_{4:\mathrm{Op}}$. The species $\mathcal{S}_{2:\mathrm{Irna}}$, $\mathcal{S}_{3:\mathrm{I}}$, $\mathcal{S}_{6:\mathrm{Rna}}$, $\mathcal{S}_{7:\mathrm{Z}}$ and $\mathcal{S}_{9:\mathrm{ILactose}}$ have errors of scale 10^{-4} and the remaining species $\mathcal{S}_{5:\mathrm{Rnap}}$, $\mathcal{S}_{10:\mathrm{IOp}}$ and $\mathcal{S}_{11:\mathrm{RnapOp}}$ show errors of scale 10^{-5}.

In the numerical examples in chapter 5 and 6 we computed the errors

$$\begin{aligned}
\varepsilon_1(\Omega) &= \| p_1(t,\cdot) - q_1(t,\cdot) \|_{\ell_d^1}, \\
\varepsilon_2(\Omega) &= \left\| \frac{1}{\Omega} \eta(t,\cdot) p_1(t,\cdot) - \theta(t,\cdot) q_1(t,\cdot) \right\|_{\ell_d^1}, \\
\varepsilon_3(\Omega) &= \left\| p(t,\cdot,\cdot) - \int_{I_m} q(t,\cdot,x)\,\mathrm{d}x \right\|_{\ell_1}.
\end{aligned}$$

for different values of Ω (cf. fig 5.3 and 6.4) and plotted the convergence of these errors. This helped us to predict the result of the theorems 5.16 and 6.5. Unfortunatly, we cannot

do this here because it is not possible to solve the CME for 10-dimensions and ca. $1.6 \cdot 10^{12}$ (for a value of $\Omega = 10^4$) states numerically. Therefor, we cannot compute a proper reference solution to compare the solutions of the hybrid models with. We determined in previous numerical experiments (data not shown) that the approximation of CME solutions with SSA is not suitable for this task. The sampling error is usually too high.

In summary, we learned in this chapter how a high dimensional and biological relevant reaction network is structured. Further, we saw that the connection of gene regulatory networks and metabolic networks lead to networks which allow the application of hybrid models. In this case we modeled the species $\mathcal{S}_{8:\text{Lactose}}$ as a continuous species and the other species as discrete. This partition is motivated by the observation that lactose is present in very high particle numbers in the system, while the other species exist in significant smaller copy numbers.

We have also seen, that it is possible to simulate this network with both hybrid models and that all features of the network are conserved. It is possible to interpret the complete simulation results and to understand how the network works.

8 | Conclusions and Outlook

Reaction networks are an often used framework that describe the interactions of different species connected via reaction channels. We have seen that three different models have been brought up in literature over the recent decades. The Kurtz process (a Markov jump process) and the Chemical Master equation (CME) are very accurate models that are derived by thermodynamical arguments. However, the computation of paths of the Kurtz process can be quite time consuming, especially for networks with many dimensions or species with high particle counts. The solution of the high dimensional CME is also a challenging task. To reduce this complexity, the Chemical Langevin equation (a stochastic differential equation) can be derived from the Kurtz process by replacing the Poisson process and the discrete propensity function. Also the Fokker-Planck equation (FPE) can be derived by truncating the Kramers-Moyal expansion. The models can even be more simplified by taking the expectation of the Kurtz process to derive the Reaction Rate equation (RRE) or by truncating the Kramers-Moyal expansion further.

We have seen that the validity of these simplifications is widely discussed in the literature. We followed the approaches from Kurtz and van Kampen and interpreted the discrete states $n \in \mathbb{N}_0^N$ of the Kurtz process and the CME as particle numbers and the continuous states $x \in \mathbb{R}_+^N$ as concentrations. This resulted in the derivation of scaled versions of the three different models.

The first important result was to see that the truncation of the Kramers-Moyal expansion can be motivated by scaling the CME with Ω. Furthermore, the convergence of the FPE to the CME is an important question, which was often discussed in the literature in the last decades (cf. section 4.4 or [Gil00]). We derived an answer, proposing that

$$\left\| p(t,\cdot) - \int_{I_m} q(t,x)\,\mathrm{d}x \right\|_{\ell_1} \leq \frac{C_1}{\Omega} \left\| \nabla \left(q(t,\cdot) \right) \right\|_{L_1} + \frac{C_2}{\Omega},$$

with $p(t,m)$ the solution of the scaled CME and $q(t,x)$ the solution of the scaled FPE.

In section 4.4.2 we discussed that for our example network it is possible to compute the error also if the covariance of the FPE is known. We stated that this observation should be discussed in future studies.

Motivated by these preliminary results we turned our focus on the derivation and analysis of hybrid models. These models combine two other models to a new one, taking the advantages of both. Since the initial idea was published by Haseltine and Rawlings, several

hybrid models have been stated in literature, but nearly no analysis was done for these models.

The frequent usage and the good achievements made with hybrid models raised the important question if these models converge to the CME. Furthermore, if hybrid models converge, can an error bound be stated?

To answer these questions, we analysed two hybrid models. First, we constructed a hybrid model combining the Kurtz process with the RRE, or on the level of the PDFs the CME with the LVE. Section 5.4 resulted in the derivation of the Liouville master equation (LME)

$$\begin{aligned}\partial_t q\left(t,n,x\right) &= \sum_{j\in J_1}\tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\left(t,n-\nu_j,x\right)-\alpha_j(n)q\left(t,n,x\right)\Big)\\ &\quad-\frac{1}{\Omega}\sum_{j=1}^{R}\alpha_j(n)\nabla_x\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)^T\mu_j,\\ q(0,n,x) &= \delta_{n_0}(n)q_0(x).\end{aligned}$$

Often, hybrid models are only stated in terms of algorithms that solve the models. The knowledge what these models actually look like is needed to allow a rigorous analysis of them. Therefore, our derivation of the LME allowed a rigorous analysis of the hybrid model and we proved convergence for

$$\begin{aligned}\|p_1(t,\cdot)-q_1(t,\cdot)\|_{\ell_d^1} &\leq \frac{C}{\Omega},\\ \left\|\frac{1}{\Omega}\eta(t,\cdot)p_1(t,\cdot)-\theta(t,\cdot)q_1(t,\cdot)\right\|_{\ell_d^1} &\leq \frac{C}{\Omega},\end{aligned}$$

with p_1 and η the marginal distribution and conditional expectation of the CME solution and q_1 and θ the marginal distribution and conditional expectation of the LME, respectively.

After these very important results, the question came up if it is possible to construct a hybrid model whose full distribution converges to the solution of the CME. We answered this question in chapter 6. We combined the CME with the FPE (or the Kurtz process with the CLE) and derived a differential equation for this model:

$$\begin{aligned}\partial_t q\left(t,n,x\right) &= \sum_{j\in J_1}\tilde{\beta}_j(x)\Big(\alpha_j(n-\nu_j)q\left(t,n-\nu_j,x\right)-\alpha_j(n)q\left(t,n,x\right)\Big)\\ &\quad-\frac{1}{\Omega}\sum_{j=1}^{R}\alpha_j(n)\nabla\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)^T\mu_j\\ &\quad+\frac{1}{2\Omega^2}\sum_{j=1}^{R}\alpha_j(n)\mu_j^T\nabla^2\left(\tilde{\beta}_j(x)q\left(t,n,x\right)\right)\mu_j,\\ q(0,n,x) &= \delta_{n_0}(n)q_0(x).\end{aligned}$$

We called this model Fokker-Planck master equation (FPME). We used the same techniques as in the convergence proof of the FPE to show

$$\left\| p(t,\cdot,\cdot) - \int_{I_m} q(t,\cdot,x)\,\mathrm{d}x \right\|_{\ell_1} \leq \frac{C_1}{\Omega} \left\| \nabla\Big(q(t,\cdot,\cdot)\Big) \right\|_1 + \frac{C_2}{\Omega}.$$

We saw that for this hybrid model the full solution of the FPME converges to the solution of the CME and therefore shows significant better convergence properties than the first hybrid model which converges only for the marginal distributions and the conditional expectations. An interesting future research topic is the observation that the FPME for the network in section 6.4 converges with order $\mathcal{O}\big(\frac{1}{\Omega}\big)$ while the derivative term is of order $\frac{1}{\Omega}\left\|\nabla\Big(q(t,\cdot,\cdot)\Big)\right\|_1 = \mathcal{O}\Big(\frac{1}{\sqrt{\Omega}}\Big)$. Future studies should therefore analyse under which conditions this improved convergence rate can be found or if the result is even general.

Finally, we have seen in chapter 7 how a 10-dimensional reaction network can be simulated using the two hybrid algorithms and that the simulation results are comparable with an approximation generated with the SSA. Furthermore, we found the error to be of the expected order. We have also seen that the results can be fully biological interpreted and that the simulation of such a high dimensional system provides insight in the biochemical functionality of the network.

We already mentioned two topics that should be analysed in future studies. The connection of the convergence of FPE in dependence of the covariance and the question if the convergence rate of the FPME can be further improved. Another interesting research topic is the question if it is possible to construct hybrid models with rates higher than one, i.e. models for that the solutions converge with order $\mathcal{O}\big(\frac{1}{\Omega^2}\big)$ or better to the CME solution. Maybe it is possible to construct models that combine adaptively the three different models CME, FPE and LVE or even further models or exact solutions. This model could be a combination of the adaptive partitioning approach discussed by Alfonsi et al. and the splitting method for several different sub-problems discussed by Jahnke and Altıntan [ACT+05, JA10].

Appendix

A Probability Distributions

Definition A1 (Poisson Distribution)**.**
The PDF of a Poisson distributed random variable $X \in \mathbb{N}$ with parameter $\lambda \in \mathbb{R}_+$ is given by

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}.$$

The expectation and variance are given by

$$\mathbb{E}[X] = \lambda \qquad \text{and} \qquad \mathbb{V}[X] = \lambda.$$

Such a variable is often denoted as

$$X \sim \mathcal{P}(\lambda).$$

The high dimensional generalisation of the Poisson PDF is given by the product Poisson distribution

$$\mathcal{P}(x, \lambda) = \begin{cases} e^{-\|\lambda\|_1} \prod_{i=1}^{N} \frac{\lambda_i^{x_i}}{x_i!} & x \in \mathbb{N}_0^N, \\ 0 & x \notin \mathbb{N}_0^N, \end{cases}$$

with $x \in \mathbb{N}^N$ and $\lambda \in \mathbb{R}_+^N$ [JH07, Wil06, ch. 3.6.4].

Definition A2 (Normal Distribution)**.**
The PDF of a Gauß distributed random variable $Xx \in \mathbb{R}$ with parameters $\mu, \sigma \in \mathbb{R}$ is given by

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

The expectation and variance are given by

$$\mathbb{E}[X] = \mu \qquad \text{and} \qquad \mathbb{V}[X] = \sigma^2.$$

Such a variable is often denoted as

$$X \sim \mathcal{N}\left(\mu, \sigma^2\right).$$

The high dimensional generalisation of the Gauß PDF is given by

$$p(x) = \frac{1}{\sqrt{(2\pi)^N \det(\sigma)}} \exp\left\{-\frac{1}{2}(x-\mu)^T \sigma^{-1}(x-\mu)\right\},$$

with $x \in \mathbb{R}^N$, expectation $\mu = \mathbb{E}[X]$ and covariance matrix $\sigma = \mathrm{Cov}[X]$.

Definition A3 (Multinomial Distribution)**.**
A multinomial PDF is given by

$$\mathcal{M}(x, n, r) = \begin{cases} n! \frac{(1-\|r\|_1)^{(n-\|x\|_1)}}{(n-\|x\|_1)!} \prod_{k=1}^{N} \frac{r_k^{x_k}}{x_k!} & \text{if } \|x\|_1 \leq n \text{ and } x \in \mathbb{N}^N \\ 0 & \text{otherwise,} \end{cases}$$

with $x \in \mathbb{N}^N$ and the parameters n and $r \in [0,1]^N$, $\|r\|_1 \leq 1$ [JH07]. This PDF can be interpreted as a high dimensional generalisation of the binomial distribution. This models the probability that in an event the combination $X_1 = x_1, \ldots, X_N = x_N$ occurs.

B Multi-Index Notation

Definition B4 (Multi-index notation)**.**
A tuple

$$\boldsymbol{k} = (k_1, k_2, \ldots, k_N)\,,\ k_i \in \mathbb{N}_0\ \forall i = 1, \ldots, N$$

is called a N-dimensional multi-index. We define

$$x^{\boldsymbol{k}} := \prod_{i=1}^{N} x_i^{k_i}, \qquad \boldsymbol{k}! := \prod_{i=1}^{N} k_i!,$$

$$|\boldsymbol{k}| := \sum_{i=1}^{N} k_i, \qquad \nabla^{\boldsymbol{k}} f(x) := \frac{\partial^{|\boldsymbol{k}|}}{\partial x_1^{k_1} \ldots \partial x_N^{k_N}} f(x),$$

with $x \in \mathbb{R}^N$ and $f : \mathbb{R}^N \to \mathbb{R}$, $f \in \mathcal{C}^\infty$. This allows the definition of the high-dimensional Taylor series

$$f(x+h) = \sum_{|\boldsymbol{k}| \geq 0} \frac{1}{\boldsymbol{k}!} \nabla^{\boldsymbol{k}} \left(f(x)\right) (h)^{\boldsymbol{k}}\,,$$

where $|\boldsymbol{k}| \geq 0$ denotes all multi-indices with $|\cdot|$-norm greater or equal to 0 and $h \in \mathbb{R}_+^N$.

C Bounds for Propensity Functions

Lemma C5 (Difference between propensity functions for shifted parameters)**.** *Under the conditions of theorem 4.5, there exists a constant $C \in \mathbb{R}_+$ such that*

$$\sup_{x \in I_m} \left| \tilde{\beta}_j(x) - \beta_j(m) \right| \leq C,$$

for all $\Omega > 0$.

Proof. We find for every interval I_m a constant $\delta \in [0,1)^N$ such that

$$\sup_{x \in I_m} \left| \tilde{\beta}_j(x) - \beta_j(m) \right| = \sup_{\delta \in [0,1)^N} \left| \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m) \right|$$

and by assumption 4.2, we distinguish the four cases:

$|\chi_j^{\text{in}}|_1 = 0$

Therefore $\chi_{j,i}^{\text{in}} = 0 \; \forall \, i \in [1, \ldots, N]$. This results in the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = c_j \Omega, \qquad \beta_j(m) = c_j \Omega$$

and we can estimate

$$\sup_{\delta \in [0,1)^N} \left| \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m) \right| = 0.$$

$|\chi_j^{\text{in}}|_1 = 1$

Therefore $\exists k : \chi_{j,k}^{\text{in}} = 1$ and $\chi_{j,i}^{\text{in}} = 0 \; \forall \, i \neq k$. This effects the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = c_j \Omega \frac{m_k + \delta_k}{\Omega} = c_j (m_k + \delta_k), \qquad \beta_j(m) = c_j \Omega \frac{m_k}{\Omega} = c_j m_k$$

which leads to

$$\sup_{\delta \in [0,1)^N} \left| \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m) \right| = \sup_{\delta \in [0,1)^N} |c_j (m_k + \delta_k - m_k)| = \sup_{\delta \in [0,1)^N} (c_j \delta_k) \leq C.$$

$|\chi_j^{\text{in}}|_1 = 2$

This can be obtained in two ways:

$\exists k, l : \chi_{j,k}^{\text{in}} = \chi_{j,l}^{\text{in}} = 1$ and $\chi_{j,i}^{\text{in}} = 0 \; \forall \, i \neq k, l$

This results in the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = c_j \Omega \frac{m_k + \delta_k}{\Omega} \frac{m_l + \delta_l}{\Omega} \qquad \beta_j(m) = c_j \Omega \frac{m_k}{\Omega} \frac{m_l}{\Omega}$$

which leads to

$$\begin{aligned} \sup_{\delta \in [0,1)^N} \left| \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m) \right| &= \sup_{\delta \in [0,1)^N} \frac{c_j}{\Omega} \left| (m_k + \delta_k)(m_l + \delta_l) - m_k m_l \right| \\ &= \sup_{\delta \in [0,1)^N} \frac{c_j}{\Omega} \left| m_k \delta_l + m_l \delta_k + \delta_k \delta_l \right| \\ &\leq C. \end{aligned}$$

The last estimate holds, because $m_k, m_l = \mathcal{O}(\Omega)$.

$\exists k : \chi^{\mathbf{in}}_{j,k} = 2$ **and** $\chi^{\mathbf{in}}_{j,i} = 0 \; \forall \; i \neq k$

This effects the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = \frac{c_j}{2\Omega}\left(m_k + \delta_k\right)^2, \qquad \beta_j(m) = \frac{c_j}{2\Omega}\left(m_k^2 - m_k\right).$$

Thus we obtain

$$\begin{aligned}\sup_{\delta\in[0,1)^N}\left|\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m)\right| &= \sup_{\delta\in[0,1)^N}\frac{c_j}{2\Omega}\left|(2\delta_k+1)m_k + \delta_k^2\right| \\ &\leq C.\end{aligned}$$

The last estimate holds, because $m_k = \mathcal{O}(\Omega)$.

□

Lemma C6 (A Bound for the Remainder Term)**.**
Under the conditions of theorem 4.5, there exists a constant $C \in \mathbb{R}_+$ such that

$$\sup_{x\in I_m}\left|\mathcal{H}_j\left(m,x\right)\right| \leq \frac{C}{\Omega},$$

with

$$\begin{aligned}\mathcal{H}_j(m,x) &= \left[\tilde{\beta}_j\left(x - \frac{\varrho_j}{\Omega}\right) - \beta_j(m-\varrho_j)\right] - \left[\tilde{\beta}_j\left(x\right) - \beta_j(m)\right] \\ &= \tilde{\beta}_j\left(x - \frac{\varrho_j}{\Omega}\right) - \beta_j(m-\varrho_j) - \tilde{\beta}_j\left(x\right) + \beta_j(m).\end{aligned}$$

Proof. We find for every interval I_m a vector $\delta \in [0,1)^N$ such that

$$\sup_{x\in I_m}\left|\mathcal{H}_j\left(m,x\right)\right| = \sup_{\delta\in[0,1)^N}\left|\mathcal{H}_j\left(m,\frac{m+\delta}{\Omega}\right)\right|$$

and by assumption 4.2, we distinguish the four cases:

$|\chi^{\mathbf{in}}_j|_1 = \mathbf{0}$

Therefore $\chi^{\text{in}}_{j,i} = 0 \; \forall \; i \in [1,\ldots,N]$. This results in the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta-\varrho_j}{\Omega}\right) = c_j\Omega, \quad \beta_j(m-\varrho_j) = c_j\Omega, \quad \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = c_j\Omega, \quad \beta_j(m) = c_j\Omega,$$

and we find that

$$\sup_{\delta\in[0,1)^N}\left|\mathcal{H}_j\left(m,\frac{m+\delta}{\Omega}\right)\right| = 0.$$

$|\chi^{\mathbf{in}}_j|_1 = \mathbf{1}$

Therefore $\exists k : \chi^{\text{in}}_{j,k} = 1$ and $\chi^{\text{in}}_{j,i} = 0 \; \forall \; i \neq k$. This effects the propensity functions

$$\begin{aligned}\tilde{\beta}_j\left(\frac{m+\delta-\varrho_j}{\Omega}\right) &= c_j\Omega\frac{m_k+\delta_k-\varrho_{jk}}{\Omega} = c_j(m_k+\delta_k-\varrho_{jk}), \\ \beta_j(m-\varrho_j) &= c_j\Omega\Omega^{-1}(m_k-\varrho_{jk}) = c_j(m_k-\varrho_{jk}), \\ \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) &= c_j\Omega\frac{m_k+\delta_k}{\Omega} = c_j(m_k+\delta_k),\end{aligned}$$

$$\beta_j(m) = c_j \Omega \Omega^{-1} m_k = c_j m_k.$$

Thus we obtain

$$\sup_{\delta \in [0,1)^N} \left| \mathcal{H}_j \left(m, \frac{m+\delta}{\Omega} \right) \right| = 0.$$

$|\chi_j^{\mathsf{in}}|_1 = 2$

This can be obtained in two ways:

$\exists k, l : \chi_{j,k}^{\mathsf{in}} = \chi_{j,l}^{\mathsf{in}} = 1$ and $\chi_{j,i}^{\mathsf{in}} = 0 \; \forall \, i \neq k, l$

This results in the propensity functions

$$\begin{aligned}
\tilde{\beta}_j \left(\frac{m+\delta-\varrho_j}{\Omega} \right) &= c_j \Omega \frac{m_k + \delta_k - \varrho_{jk}}{\Omega} \frac{m_l + \delta_l - \varrho_{jl}}{\Omega}, \\
\beta_j(m - \varrho_j) &= c_j \Omega \Omega^{-2} (m_k - \varrho_{jk})(m_l - \varrho_{jl}), \\
\tilde{\beta}_j \left(\frac{m+\delta}{\Omega} \right) &= c_j \Omega \frac{m_k + \delta_k}{\Omega} \frac{m_l + \delta_l}{\Omega}, \\
\beta_j(m) &= c_j \Omega \Omega^{-2} m_k m_l
\end{aligned}$$

and we find that

$$\sup_{\delta \in [0,1)^N} \left| \mathcal{H}_j \left(m, \frac{m+\delta}{\Omega} \right) \right| = \sup_{\delta \in [0,1)^N} \frac{c_j}{\Omega} \left| \delta_k \delta_l - \delta_k \varrho_{jl} - \delta_l \varrho_{jk} \right| \leq \frac{C}{\Omega}.$$

$\exists k : \chi_{j,k}^{\mathsf{in}} = 2$ and $\chi_{j,i}^{\mathsf{in}} = 0 \; \forall \, i \neq k$

This effects the propensity functions

$$\begin{aligned}
\tilde{\beta}_j \left(\frac{m+\delta-\varrho_j}{\Omega} \right) &= c_j \Omega \frac{1}{2} \left(\frac{m_k + \delta_k - \varrho_{jk}}{\Omega} \right)^2, \\
\beta_j(m - \varrho_j) &= c_j \Omega \Omega^{-2} \frac{1}{2} (m_k - \varrho_{jk})(m_k - \varrho_{jk} - 1), \\
\tilde{\beta}_j \left(\frac{m+\delta}{\Omega} \right) &= c_j \Omega \frac{1}{2} \left(\frac{m_k + \delta_k}{\Omega} \right)^2, \\
\beta_j(m) &= c_j \Omega \Omega^{-2} \frac{1}{2} m_k (m_k - 1)
\end{aligned}$$

and we find that

$$\sup_{\delta \in [0,1)^N} \left| \mathcal{H}_j \left(m, \frac{m+\delta}{\Omega} \right) \right| = \sup_{\delta \in [0,1)^N} \frac{c_j}{2\Omega} \left| -2\delta_k \varrho_{jk} - \varrho_{jk} - \delta_k^2 \right| \leq \frac{C}{\Omega}.$$

□

Lemma C7 (Difference of the Propensity Functions)**.**
If $x \in \mathbb{R}_{0,+}^D$ and $x_k \geq \frac{1}{\Omega} \; \forall k = 1, \ldots, D$ and the assumptions 5.9 and 5.10 hold, then there exists a constant $C > 0$ such that

$$\left| \beta_j(\Omega x) - \tilde{\beta}_j(x) \right| \leq C \Omega^{-\gamma(j)}.$$

If $\lambda_{jk}^{in} \leq 1 \; \forall k = 1, \ldots, D$,

$$\beta_j(\Omega x) = \tilde{\beta}_j(x)$$

holds.

Proof. We have to distinguish four cases, by assumption 5.9:

$|\lambda_j^{\mathbf{in}}|_1 = \mathbf{0}$
Therefore $\lambda_{j,i}^{\text{in}} = 0 \; \forall \; i \in [1, \ldots, D]$ with the propensity functions

$$\beta_j(\Omega x) = \Omega^{(1-\gamma(j))}\Omega^0 \prod_{i=1}^{D} \binom{\Omega x_i}{0} = \Omega^{(1-\gamma(j))},$$
$$\tilde{\beta}_j\left(x\right) = \Omega^{(1-\gamma(j))} \prod_{i=1}^{D} \frac{x_i^0}{0!} = \Omega^{(1-\gamma(j))}$$

and the error

$$\left|\beta_j(\Omega x) - \tilde{\beta}_j\left(x\right)\right| = 0.$$

$|\lambda_j^{\mathbf{in}}|_1 = \mathbf{1}$
Therefore $\exists k : \lambda_{j,k}^{\text{in}} = 1$ and $\lambda_{j,i}^{\text{in}} = 0 \; \forall \; i \neq k$. This results in the propensity functions

$$\beta_j(\Omega x) = \Omega^{(1-\gamma(j))}\Omega^{-1} \prod_{i=1}^{D} \binom{\Omega x_k}{1} = \Omega^{(1-\gamma(j))} x_k,$$
$$\tilde{\beta}_j\left(x\right) = \Omega^{(1-\gamma(j))} \prod_{i=1}^{D} \frac{x_k^1}{1!} = \Omega^{(1-\gamma(j))} x_k$$

and the error

$$\left|\beta_j(\Omega x) - \tilde{\beta}_j\left(x\right)\right| = 0.$$

$|\lambda_j^{\mathbf{in}}|_1 = \mathbf{2}$
This can be obtained in two ways:

$\exists k, l : \lambda_{j,k}^{\mathbf{in}} = \lambda_{j,l}^{\mathbf{in}} = 1$ **and** $\lambda_{j,i}^{\mathbf{in}} = 0 \; \forall \; i \neq k, l$
This effects the propensity functions

$$\beta_j(\Omega x) = \Omega^{(1-\gamma(j))}\Omega^{-2} \binom{\Omega x_k}{1} \binom{\Omega x_l}{1} = \Omega^{(1-\gamma(j))} x_k x_l,$$
$$\tilde{\beta}_j\left(x\right) = \Omega^{(1-\gamma(j))} \frac{x_k^1}{1!} \frac{x_l^1}{1!} = \Omega^{(1-\gamma(j))} x_k x_l$$

and the error

$$\left|\beta_j(\Omega x) - \tilde{\beta}_j\left(x\right)\right| = 0.$$

$\exists k : \lambda_{j,k}^{\mathbf{in}} = 2$ **and** $\lambda_{j,i}^{\mathbf{in}} = 0 \; \forall \; i \neq k$
This results in the propensity functions

$$\begin{aligned}\beta_j(\Omega x) &= \Omega^{(1-\gamma(j))}\Omega^{-2} \binom{\Omega x_k}{2} = \Omega^{(1-\gamma(j))}\Omega^{-2} \frac{1}{2} \Omega x_k (\Omega x_k - 1) \\ &= \Omega^{(1-\gamma(j))} \frac{1}{2} (x_k^2 - \Omega^{-1} x_k),\end{aligned}$$

$$\tilde{\beta}_j(x) = \Omega^{(1-\gamma(j))}\frac{x_k^2}{2!} = \Omega^{(1-\gamma(j))}\frac{1}{2}x_k^2$$

and the error

$$\begin{aligned}\left|\beta_j(\Omega x) - \tilde{\beta}_j(x)\right| &= \Omega^{(1-\gamma(j))}\frac{1}{2}\left|x_k^2 - \Omega^{-1}x_k - x_k^2\right| \\ &= \Omega^{(1-\gamma(j))}\frac{1}{2}\left|-\Omega^{-1}x_k\right| \\ &= \Omega^{(-\gamma(j))}\frac{1}{2}x_k\Omega^{(-\gamma(j))} \\ &\leq C\Omega^{(-\gamma(j))}.\end{aligned}$$

□

Lemma C8 (Difference between propensity functions with shifted parameters)**.** *Under the conditions of theorem 6.5 and for $j \in J_1$ exists a constant $C \in \mathbb{R}_+$ such that*

$$\sup_{x\in I_m}\left|\tilde{\beta}_j(x) - \beta_j(m-\mu_j)\right| \leq \frac{C}{\Omega}.$$

Proof. We find for every interval I_m a constant $\delta \in [0,1)^N$ such that

$$\sup_{x\in I_m}\left|\tilde{\beta}_j(x) - \beta_j(m-\mu_j)\right| = \sup_{\delta\in[0,1)^N}\left|\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m-\mu_j)\right|$$

and by assumption 4.2, we distinguish four cases:

$|\lambda_j^{\text{in}}|_1 = 0$

Therefore $\lambda_{j,i}^{\text{in}} = 0 \;\forall\, i \in [1,\ldots,N]$. This effects the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = 1 \qquad\qquad \beta_j(m-\mu_j) = 1$$

and we find

$$\sup_{\delta\in[0,1)^N}\left|\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m-\mu_j)\right| = 0.$$

$|\lambda_j^{\text{in}}|_1 = 1$

Therefore $\exists k : \lambda_{j,k}^{\text{in}} = 1$ and $\lambda_{j,i}^{\text{in}} = 0 \;\forall\, i \neq k$. This results in the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = \frac{m_k+\delta_k}{\Omega}, \qquad\qquad \beta_j(m-\mu_j) = \frac{m_k-\mu_{jk}}{\Omega},$$

which results in

$$\begin{aligned}\sup_{\delta\in[0,1)^N}\left|\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m-\mu_j)\right| &= \sup_{\delta\in[0,1)^N}\left|\frac{m_k+\delta_k}{\Omega} - \frac{m_k-\mu_{jk}}{\Omega}\right| \\ &= \sup_{\delta\in[0,1)^N}\left|\frac{\delta_k+\mu_{jk}}{\Omega}\right| \leq \frac{C}{\Omega}.\end{aligned}$$

$|\lambda_j^{\text{in}}|_1 = 2$

This can be obtained in two ways:

$\exists k, l : \lambda^{\mathbf{in}}_{j,k} = \lambda^{\mathbf{in}}_{j,l} = 1$ **and** $\lambda^{\mathbf{in}}_{j,i} = 0 \; \forall \; i \neq k, l$
This effects the propensity functions

$$\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) = \frac{m_k + \delta_k}{\Omega}\frac{m_l + \delta_l}{\Omega} \qquad \beta_j(m - \mu_j) = \frac{m_k - \mu_{jk}}{\Omega}\frac{m_l - \mu_{jl}}{\Omega}$$

which results in

$$\sup_{\delta \in [0,1)^N} \left| \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m - \mu_j) \right| \leq \frac{C}{\Omega}.$$

$\exists k : \lambda^{\mathbf{in}}_{j,k} = 2$ **and** $\lambda^{\mathbf{in}}_{j,i} = 0 \; \forall \; i \neq k$
This results in the propensity functions

$$\begin{aligned}
\tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) &= \frac{1}{2\Omega^2}\left(m_k + \delta_k\right)^2 = \frac{1}{2\Omega^2}\left(m_k^2 + 2m_k\delta_k + \delta_k\right) \\
\beta_j(m - \mu_j) &= \frac{1}{2\Omega^2}\left(\left(m_k - \mu_{jk}\right)^2 - \left(m_k - \mu_{jk}\right)\right) \\
&= \frac{1}{2\Omega^2}\left(m_k^2 - 2m_k\mu_{jk} + \mu_{jk}^2 - m_k + \mu_{jk}\right)
\end{aligned}$$

which results in

$$\sup_{\delta \in [0,1)^N} \left| \tilde{\beta}_j\left(\frac{m+\delta}{\Omega}\right) - \beta_j(m - \mu_j) \right| \leq \frac{C}{\Omega}.$$

□

Bibliography

[ACT+05] A. Alfonsi, E. Cancès, G. Turinici, B. Di Ventura, and W. Huisinga. Adaptive simulation of hybrid stochastic and deterministic models for biochemical systems. *ESAIM: Proceedings*, 14:1–13, 2005.

[AH12] D. F. Anderson and D. J. Higham. Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. *Multiscale Model. Simul.*, 10(1):146–179, 2012.

[AK12] D. F. Anderson and M. Koyama. Weak error analysis of numerical methods for stochastic models of population processes. *SIAM Multiscale Modeling and Simulation*, 10(4):1493–1524, 2012.

[BCMP11] D. Besozzi, P. Cazzaniga, G. Mauri, and D. Pescini. BioSimWare: A software for the modeling, simulation and analysis of biological systems. *Lecture Notes in Computer Science*, 6501:119–143, 2011.

[BHMS06] K. Burrage, M. Hegland, S. MacNamara, and R.B. Sidje. A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete modelling of biological systems. In *A.N.Langville, W.J.Stewart (eds.) Markov Anniversary Meeting: Aninternational conference to celebrate the 150th anniversary of the birth of A.A. Markov, Boson Books*, pages 21–38, 2006.

[BKPR06] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Annals of Applied Probability*, 16(4):1925–1961, 2006.

[BMP96] F. Baras, M. Malek Mansour, and J. E. Pearson. Microscopic simulation of chemical bistability in homogeneous systems. *The Journal of Chemical Physics*, 105(18):8257–8261, 1996.

[BMT06] K. Burrage, S. Mac, and T. H. Tian. Accelerated leap methods for simulating discrete stochastic chemical kinetics. *Positive Systems, Proceedings*, 341:359–366, 2006.

[BTB04] K. Burrage, T. H. Tian, and P. Burrage. A multi-scaled approach for simulating chemical reaction systems. *Progress In Biophysics & Molecular Biology*, 85(2-3):217–234, 2004.

[BTS02] J.M. Berg, J.L. Tymoczko, and L. Stryer. *Biochemistry.* W.H. Freeman, New York, 2002.

[CDMR12] A. Crudu, A. Debussche, A. Muller, and O. Radulescu. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *Annals of Applied Probability*, 22(5):1745–2164, 2012.

[CM77] D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes.* Chapman & Hall Ltd/CRC, Boca Raton, 1977.

[CWA09] L. N. Chen, R. Q. Wang, and K. Aihara. Stochastic hybrid system for chemical master equation. Optimization and Systems Biology Book Series: Lecture Notes in Operations Research Volume: 11 Pages: 475-481 Published: 2009 Conference: 3rd International Symposium on Optimization and Systems Biology, 2009.

[DHJW08] P. Deuflhard, W. Huisinga, T. Jahnke, and M. Wulkow. Adaptive discrete Galerkin methods applied to the chemical master equation. *SIAM Journal on Scientific Computing*, 30(6):2990–3011, 2008.

[DK12] S.V. Dolgov and B.N. Khoromskij. Tensor-product approach to global time-space-parametric discretization of chemical master equation. Technical report, Max-Planck-Institut für Mathematik in den Naturwissenschaften, 2012.

[DKB10] C.E. Dangerfield, D. Kay, and K. Burrage. Stochastic models and simulation of ion channel dynamics. *Procedia Computer Science*, 1(1):1587–1596, 2010.

[DP11] F. Du and B. Parise. A hybrid moment equation approach to gas-grain chemical modeling. *Astronomy & Astrophysics*, 530:A131, 2011.

[Ein05] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, 1905.

[EK05] S. N. Ethier and T. G. Kurtz. *Markov processes. Characterization and convergence*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[EN00] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer, Berlin / Heidelberg, 2000.

[Eng06] S. Engblom. Computing the moments of high dimensional solutions of the master equation. *Applied Mathematics and Computation*, 180(2):498–515, 2006.

[Eng09] S. Engblom. Spectral approximation of solutions to the chemical master equation. *Journal of Computational and Applied Mathematics*, 229(1):208–221, 2009.

[Gar04] C.W. Gardiner. *Handbook of Stochastic Methods*. SSSyn - An Interdisciplinary Series on Complex Systems. Springer, Berlin / Heidelberg, 3 edition, 2004.

[GB00] M.A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *Journal of Physical Chemistry A*, 104(9):1876–1889, 2000.

[GCPS06] M. Griffith, T. Courtney, J. Peccoud, and W. H. Sanders. Dynamic partitioning for hybrid simulation of the bistable HIV-1 transactivation network. *Bioinformatics*, 22(22):2782–2789, 2006.

[GGJ13] W. Garcia-Gabin and E. W. Jacobsen. Multilevel model of type 1 diabetes mellitus patients for model-based glucose controllers. *Journal of Diabetes Science and Technology*, 7(1):193–205, 2013.

[Gil76] D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[Gil80] D. T. Gillespie. Approximating the master equation by Fokker-Planck-type equations for single-variable chemical systems. *The Journal of Chemical Physics*, 72(10):5363–5370, 1980.

[Gil92] D. T. Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1-3):404–425, 1992.

[Gil00] D.T. Gillespie. The chemical Langevin equation. *Journal of Chemical Physics*, 113(1):297–306, 2000.

[Gil01] D. T. Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[Gil08a] M. B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In A. Keller, S. Heinrich, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer, Berlin / Heidelberg, 2008.

[Gil08b] M. B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3):607–617, 2008.

[GY13] L. Gauckler and H. Yserentant. Regularity and approximability of the solutions to the chemical master equation. *Matheon Preprint, Berlin*, 1010, 2013.

[HDD+11] D. He, J. Dushoff, T. Day, J. Ma, and D. Earn. Mechanistic modelling of the three waves of the 1918 influenza pandemic. *Theoretical Ecology*, 4:283–288, 2011.

[Heg08] M. Hegland. Approximating the solution of the chemical master equation by aggregation. *ANZIAM Journal*, 50:C371–C384, 2008.

[HFC10] M. Hegland and R. Fletcher-Costin. Cmepy documentation. `http://fcostin.github.io/cmepy/`, 2010.

[HG11] M. Hegland and J. Garcke. On the numerical solution of the chemical master equation with sums of rank one tensors. In *W. McLean, A.J. Roberts (eds.) Proceedings of the 15th Biennial Computational Techniques and Applications Conference, CTAC-2010*, pages C628–C643, 2011.

[HHL08] M. Hegland, A. Hellander, and P. Lötstedt. Sparse grids and hybrid methods for the chemical master equation. *BIT Numerical Mathematics*, 48(2):265–283, 2008.

[Hig01] D.J. Higham. An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546, 2001.

[Hig08] D. J. Higham. Modeling and simulating chemical reactions. *SIAM Review*, 50(2):347–368, 2008.

[Hig11] D. J. Higham. Stochastic ordinary differential equations in applied and computational mathematics. *IMA Journal of Applied Mathematics*, 76(3):449–474, 2011.

[HIMS11] D. Higham, S. Intep, X. Mao, and L. Szpruch. Hybrid simulation of autoregulation within transcription and translation. *BIT Numerical Mathematics*, 51:177–196, 2011.

[HL07] A. Hellander and P. Lötstedt. Hybrid method for the chemical master equation. *Journal of Computational Physics*, 227(1):100 – 122, 2007.

[HMMW10] T. A. Henzinger, L. Mikeev, M. Mateescu, and V. Wolf. Hybrid numerical solution of the chemical master equation. In *Proceedings of the 8th International Conference on Computational Methods in Systems Biology*, pages 55–65, Trento, Italy, 2010. ACM.

[HR02] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *Journal of Chemical Physics*, 117(15):6959–6969, 2002.

[HSG+06] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI - a complex pathway simulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[HWKT14] J. Hasenauer, V. Wolf, A. Kazeroonian, and F.J. Theis. Method of conditional moments (MCM) for the chemical master equation. *Journal of Mathematical Biology*, 69:687–735, 2014.

[JA10] T. Jahnke and D. Altintan. Efficient simulation of discrete stochastic reaction systems with a splitting method. *BIT*, 50(4):797–822, 2010.

[Jah10] T. Jahnke. An adaptive wavelet method for the chemical master equation. *SIAM Journal on Scientific Computing*, 31(6):4373–4394, 2010.

[Jah11] T. Jahnke. On reduced models for the chemical master equation. *SIAM Multiscale Modeling and Simulation*, 9(4):1646–1676, 2011.

[JH07] T. Jahnke and W. Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. *Journal of Mathematical Biology*, 54(1):1–26, 2007.

[JH08] T. Jahnke and W. Huisinga. A dynamical low-rank approach to the chemical master equation. *Bulletin of Mathematical Biology*, 70(8):2283–2302, 2008.

[JK12] T. Jahnke and M. Kreim. Error bound for piecewise deterministic processes modeling stochastic reaction systems. *SIAM Multiscale Modeling and Simulation*, 10(4):1119–1147, 2012.

[JM61] F. Jacob and J. Monod. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3(3):318–356, 1961.

[JS13] T. Jahnke and V. Sunkara. Error bound for hybrid models of two-scaled stochastic reaction systems. *Submitted*, 2013.

[JU10] T. Jahnke and T. Udrescu. Solving chemical master equations by adaptive wavelet compression. *Journal of Computational Physics*, 229(16):5724–5741, 2010.

[Kam61] N. G. van Kampen. A power series expansion of the master equation. *Canadian Journal of Physics*, 39(4):551–567, 1961.

[Kaz06] Y. N. Kaznessis. Multi-scale models for gene network engineering. *Chemical Engineering Science*, 61(3):940–953, 2006.

[KKNS12] V. Kazeev, M. Khammash, M. Nip, and C. Schwab. Direct solution of the chemical master equation using quantized tensor trains. Technical report, ETH Zurich, 2012.

[KMK73] R. Kubo, K. Matsuo, and K. Kitahara. Fluctuation and relaxation of macrovariables. *Journal of Statistical Physics*, 9(1):51–96, 1973.

[KMS04] T. R. Kiehl, R. M. Mattheyses, and M. K. Simmons. Hybrid simulation of cellular behavior. *Bioinformatics*, 20(3):316–322, 2004.

[KP99] P. E. Kloeden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, Berlin / Heidelberg, 1999.

[Kra40] H.A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284 – 304, 1940.

[KS13] V. Kazeev and C. Schwab. Tensor approximation of stationary distributions of chemical reaction networks. Technical report, ETH Zurich, 2013.

[KSM+12] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival Jr., N. Assad-Garcia, J. I. Glass, and M. W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[Kur72] T.G. Kurtz. Relationship between stochastic and deterministic models for chemical reactions. *Journal of Chemical Physics*, 57(7):2976–2978, 1972.

[Kur78] T. G. Kurtz. Strong approximation theorems for density dependent Markov chains. *Stochastic Processes and their Applications*, 6(3):223–240, 1978.

[LCPG08] H. Li, Y. Cao, L. R. Petzold, and D. T. Gillespie. Algorithms and software for stochastic simulation of biochemical reacting systems. *Biotechnol Progress*, 24(1):56–61, 2008.

[LM94] A. Lasota and M.C. Mackey. *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*. Springer, Berlin / Heidelberg, 1994.

[LPP11] P. Lachor, K. Puszynski, and A. Polanski. Deterministic models and stochastic simulations in multiple reaction models in systems biology. *BioTechnologia – Journal of Biotechnology, Computational Biology and Bionanotechnology*, 92(3):265–280, 2011.

[MA99] 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.

[Mac09] S. MacNamara. *Krylov and Finite State Projection methods for simulating stochastic biochemical kinetics via the Chemical Master Equation*. PhD thesis, The University of Queensland, 2009.

[MBBS08] S. MacNamara, A. M. Bersani, K. Burrage, and R. B. Sidje. Stochastic chemical kinetics and the total quasi-steady-state assumption: Application to the stochastic simulation algorithm and chemical master equation. *The Journal of Chemical Physics*, 129(9):095105–1–095105–13, 2008.

[McQ67] D. A. McQuarrie. Stochastic approach to chemical kinetics. *Journal of Applied Probability*, 4(3):413–478, 1967.

[MK06] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, 124(4):044104–1–044104–13, 2006.

[MLB07] T.T. Marquez-Lago and K. Burrage. Binomial tau-leap spatial stochastic simulation algorithm for applications in chemical kinetics. *Journal of Chemical Physics*, 127(10):104101–1–104101–9, 2007.

[MLSH12] S. Menz, J. Latorre, C. Schütte, and W. Huisinga. Hybrid stochastic–deterministic solution of the chemical master equation. *Multiscale Modeling & Simulation*, 10(4):1232–1262, 2012.

[MML09] B. D. MacArthur, A. Ma'ayan, and I. R. Lemischka. Systems biology of stem cell fate and cellular reprogramming. *Nat Rev Mol Cell Biol*, 10(10):672–681, 2009.

[Moy49] J. E. Moyal. Stochastic processes and statistical physics. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):150–210, 1949.

[MPM10] A. J. Marks, D. Pillay, and A. R. McLean. The effect of intrinsic stochasticity on transmitted HIV drug resistance patterns. *Journal of Theoretical Biology*, 262(1):1 – 13, 2010.

[Mur02] J.D. Murray. *Mathematical Biology. I: An Introduction.* Springer, Berlin / Heidelberg, 2002.

[Nob65] Nobelprize.org. The nobel prize in physiology or medicine 1965. `http://www.nobelprize.org/nobel_prizes/medicine/laureates/1965/`, 1965.

[Øks07] B. Øksendal. *Stochastic Differential Equations.* Springer, Berlin / Heidelberg, 2007.

[OSW69] I. Oppenheim, K. E. Shuler, and G. H. Weiss. Decay of correlations. II. relaxation of momentum correlations and momentum distributions in harmonic oscillator chains. *The Journal of Chemical Physics*, 50(9):3662–3669, 1969.

[Pah09] J. Pahle. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches. *Briefings in Bioinformatics*, 10(1):53–64, 2009.

[Rie10] M. G. Riedler. Almost sure convergence of numerical approximations for piecewise deterministic Markov processes. *Heriot-Watt Mathematics Report*, HWM10-34, 2010.

[SH10] V. Sunkara and M. Hegland. An optimal finite state projection method. *Procedia Computer Science*, 1(1):1579–1586, 2010.

[SK05] H. Salis and Y. Kaznessis. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *The Journal of Chemical Physics*, 122(5):054103–1–054103–13, 2005.

[SLE09] P. Sjöberg, P. Lötstedt, and J. Elf. Fokker-Planck approximation of the master equation in molecular biology. *Computing and Visualization in Science*, 12(1):37–50, 2009.

[SMF+13] A. Scialdone, S. T. Mugford, D. Feike, A. Skeffington, P. Borrill, A. Graf, A. M. Smith, and M. Howard. Arabidopsis plants perform arithmetic division to prevent starvation at night. *eLife*, 2:e00669, 2013.

[SSK06] H. Salis, V. Sotiropoulos, and Y. N. Kaznessis. Multiscale hy3s: Hybrid stochastic simulation for supercomputers rid a-2495-2008. *BMC Bioinformatics*, 7:93, 2006.

[Sun12] V. Sunkara. *Analysis and Numerics of the Chemical Master Equation.* PhD thesis, Australian National University, 2012.

[Sun13] V. Sunkara. Finite state projection method for hybrid models. *Submitted*, 2013.

[SYSY02] R. Srivastava, L. You, J. Summers, and J. Yin. Stochastic vs. deterministic modeling of intracellular viral kinetics. *Journal of Theoretical Biology*, 218(3):309 – 321, 2002.

[TB04] T. H. Tian and K. Burrage. Binomial leap methods for simulating stochastic chemical kinetics. *Journal of Chemical Physics*, 121(21):10356–10364, 2004.

[TKHT04] K. Takahashi, K. Kaizu, B. Hu, and M. Tomita. A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics*, 20(4):538–546, 2004.

[Udr12] T. Udrescu. *Numerical Methods for the Chemical Master Equation.* PhD thesis, Karlsruhe Institute of Technology (KIT), 2012.

[VB04] K. Vasudeva and U. S. Bhalla. Adaptive stochastic-deterministic chemical kinetic simulations. *Bioinformatics*, 20(1):78–84, 2004.

[VK07] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier, Amsterdam, 3 edition, 2007.

[WGMH10] V. Wolf, R. Goel, M. Mateescu, and T. A. Henzinger. Solving the chemical master equation using sliding windows. *BMC Systems Biology*, 4:42, 2010.

[WHK10] E. Weeding, J. Houle, and Y. N. Kaznessis. Synbioss designer: a web-based tool for the automated generation of kinetic models for synthetic biological constructs. *Briefings in Bioinformatics*, 11(4):394–402, 2010.

[Wil06] D. J. Wilkinson. *Stochastic modelling for systems biology*. Mathematical and computational biology series. Chapman & Hall / CRC, Boca Raton, Fla. [u.a.], 2006.

List of Theorems, Lemmata, Definitions and Assumptions

2.1 Definition (Monomolecular Reaction Channels) . . . 8
2.2 Definition (Kurtz process) . . . 9
2.3 Lemma (Chemical Master Equation) . . . 10
2.4 Lemma (Solution of the CME for monomolecuar reaction networks) . . . 12
2.5 Definition (Stochastic Differential Equation (SDE)) . . . 13
2.6 Lemma (Fokker-Planck equation) . . . 14
2.7 Definition (Chemical Langevin Equation) . . . 14
2.8 Lemma (Fokker-Planck Equation) . . . 15
2.9 Definition (Reaction Rate Equation) . . . 16
2.10 Lemma (Liouville Equation) . . . 16
2.11 Definition (The Kramers-Moyal Expansion) . . . 20

3.1 Definition (Reaction Probability Density Function) . . . 23
3.2 Definition (Finite CME) . . . 26

4.1 Definition (Scaled Propensity Functions) . . . 40
4.2 Assumption (Bound for the Stoichiometric Factor) . . . 40
4.3 Lemma (Difference between discrete and continuous propensity functions) . . . 41
4.4 Assumption (Bound for remainder terms) . . . 49
4.5 Theorem (Convergence of the FPE) . . . 49

5.1 Definition (Hybrid Propensity Function) . . . 66
5.2 Definition (Spaces and Norms for Hybrid Systems) . . . 66
5.3 Definition (Scaled and Partitioned Kurtz process and CME) . . . 67
5.4 Definition (Shift Operators) . . . 67
5.5 Lemma (Properties of the Shift Operators) . . . 68
5.6 Lemma (The solution of the partitioned CME is a PDF) . . . 69
5.7 Definition (Liouville master equation) . . . 72
5.8 Definition (Initial Condition of the LME) . . . 72
5.9 Assumption (Bound of the Stoichiometric Factor) . . . 76
5.10 Assumption (Scale Difference of a Reaction Network) . . . 76
5.11 Assumption (Solution of CME and LME) . . . 76
5.12 Assumption (Bound for the Conditional Moments) . . . 76
5.13 Assumption (Bound for the Propensity Function) . . . 77
5.14 Lemma (Difference of the Propensity Functions) . . . 77
5.15 Lemma (A Bound for Propensity Functions) . . . 77
5.16 Theorem (An Error Bound for the LME) . . . 79
5.17 Lemma (A Bound for the Marginal Distributions) . . . 80
5.18 Lemma (A Bound for the Conditional Expectations) . . . 82

6.1 Definition (Kurtz / CLE Hybrid Process) . . . 91
6.2 Corollary (An Error Bound for the FPME) . . . 93
6.3 Assumption (Bound for the Propensity Function in J_1) . . . 98
6.4 Assumption (Bound for remainder terms) . . . 98
6.5 Theorem (Convergence of the FPME) . . . 99

A1 Definition (Poisson Distribution) . . . 123
A2 Definition (Normal Distribution) . . . 124
A3 Definition (Multinomial Distribution) . . . 124
B4 Definition (Multi-index notation) . . . 125
C5 Lemma (Difference between propensity functions for shifted parameters) . . 126
C6 Lemma (A Bound for the Remainder Term) . . . 127
C7 Lemma (Difference of the Propensity Functions) . . . 128
C8 Lemma (Difference between propensity functions with shifted parameters) . 130

List of Algorithms

3.1 Stochastic Simulation Algorithm (SSA): direct method 25

5.1 A Sampling Algorithm for the Hybrid process 75

6.1 A Sampling Algorithm for the Hybrid process in def. 6.1 91

List of Figures

2.1 Relation between the different models. The column on the left lists the different stochastic processes. The Kurtz process is approximated by the CLE by approximating the Poisson processes. The CLE is approximated by the RRE by taking the expectation. The horizontal arrows list the lemmata that connect the processes in the left column and the PDFs in the right column. The downward facing arrows in the right column symbolise the approximation of the CME by the FPE and the LVE using the Kramers-Moyal expansion. 21

3.1 Solutions of the six different models for reaction networks for example 1, with $c_1 = 1$ and $\varepsilon = 1$. 29
3.2 Solutions of five different models for reaction networks for example 2, with $c_1 = 0.01$. 31
3.3 Several trajectories of the Kurtz processes in examples 3, with $c_1 = c_2 = c_3 = 1$. 33
3.4 Marginal distributions of the solution of the CME in example 3, with $c_1 = c_2 = c_3 = 1$. 35
3.5 Solutions of the CLE in example 3, with $c_1 = c_2 = c_3 = 1$. 35
3.6 Solutions of the RRE for example 3, with $c_1 = c_2 = c_3 = 1$. 36

4.1 Connection between the scaled CME (left) and the scaled FPE (right). The transformation Φ maps the continuous values of q to a discrete state in p. We have to notice that not only the state space is changed between discrete and continuous, but also the position is changed by the relationship $x = \frac{m}{\Omega}$. 48
4.2 The essential support (green lines) of the solution $p(t, m)$ of the CME (black). The red line denotes the value of the thresholding parameter ε. . . 48
4.3 Numerical solutions of the CME (top, blue), the FPE (bottom, blue) and the transformed FPE solution v (top, green) for different values of Ω at time $t = 50$. 55
4.4 Convergence of the terms in theorem 4.5 using the solutions of the CME and the FPE for the reaction network given in table 4.1 at time point $t = 50$. 56
4.5 Evolution of the error of the FPE against the CME over time (left), in comparison to the evolution of the scaled derivative $\frac{1}{\Omega} \|\nabla (q(t, \cdot))\|_{L_1}$ (right). The different time points are colour coded: $t = 0.5$ (blue), $t = 1$ (green), $t = 3$ (red), $t = 5$ (cyan), $t = 7.5$ (purple) and $t = 10$ (ochre). The black dashed lines plot the values of $\frac{1}{\sqrt{\Omega}}$ and $\frac{1}{\Omega}$ in logarithmic scaling. 57
4.6 Convergence of the left hand side term ε in theorem 4.5 using the solutions of the CME and the FPE for the reaction network given in table 4.1. This term is compared with the order of the standard deviation of the FPE solution. 57

5.1 Solution of the CME (left) and PDF of the hybrid model (right) for example (5.1) for $\Omega = 50$ and $t = 0.5$. 64
5.2 Marginal distribution of $\mathcal{S}_1$ of the CME (blue) and of the hybrid PDF (green) for example (5.1) for $\Omega = 10$ (left) and $\Omega = 100$ (right) at time $t = 0.5$. 65
5.3 Convergence error of the marginal distribution (blue) and conditional expectation (green) for the reaction network in example (5.1) at time $t = 0.5$. 65

6.1 Convergence of the marginal distributions (blue) and conditional expectations (green) of the FPME against the CME. The black line plots the function $\Omega \mapsto \Omega^{-1}$. Both error terms are of order $\mathcal{O}\left(\frac{1}{\Omega}\right)$, i.e. they are parallel to the black line. 94
6.2 Solution of the CME (left) and the FPME (right) for the network in table 6.1 at time $t = 0.5$ and $\Omega = 100$. 95
6.3 Solution of the CME $p(t, n, m)$ (blue) and of the transformed FPME $\Psi(q)(t, n, m)$ (green) for the network in table 6.1 at time $t = 0.5$ for $\Omega = 10$ (left), $\Omega = 100$ (center) and $\Omega = 1000$ (right). 95
6.4 Convergence of $\varepsilon_{\text{FPME}}$ for different values of Ω (blue). The black, dashed lines denotes the function $\Omega \mapsto \Omega^{-1/2}$, the black, dash-pointed line the function $\Omega \mapsto \Omega^{-1}$. The green line shows the convergence of the first derivative $\frac{1}{\Omega} \|\nabla (q(t, \cdot, \cdot))\|_1$. 96

7.1 Visualisation of the lac operon network 106
7.2 Average trajectory of the 10 species of the lac operon reaction network defined in table 7.3. For each species the results of the SSA (red), the Kurtz process / RRE hybrid model (green) and the Kurtz process / CLE hybrid model (blue) are plotted. The averaged trajectory for species $\mathcal{S}_{8:\text{Lactose}}$ was scaled with Ω for the hybrid models. This allows a direct visual comparison of the results. The average was taken over $3.2 \cdot 10^6$ realisations for the SSA, $3.2 \cdot 10^6$ realisations for the Kurtz process / RRE hybrid model and $1.5 \cdot 10^6$ realisations for the Kurtz process / CLE hybrid model. This figure is continued in figure 7.3. 110
7.3 Continuation of figure 7.2. 111
7.4 Histograms of the marginal distributions of the 10 species of the lac operon reaction network defined in table 7.3. The underlying trajectories were generated using $3.2 \cdot 10^6$ realisations of the SSA (left column, red), $3.2 \cdot 10^6$ realisations of the Kurtz / RRE hybrid algorithm (middle column, green) and $1.5 \cdot 10^6$ realisations of the Kurtz / CLE hybrid algorithm (right column, blue). This figure is continued in figure 7.5. 112
7.5 Continuation of figure 7.4. 113
7.6 Error $\varepsilon_1^{(i)}(n)$ of the Kurtz / RRE hybrid model at time $t = 10^3$. The model was solved using $3.2 \cdot 10^6$ realisations of algorithm 5.1 to approximate the marginal distributions. Histograms from $3.2 \cdot 10^6$ SSA realisations were used as a reference solutions. This figure is continued in figure 7.7. 114
7.7 Continuation of figure 7.6. 115
7.8 Error $\varepsilon_2^{(i)}(n)$ of the Kurtz / CLE hybrid model at time $t = 10^3$. The model was solved using $1.5 \cdot 10^6$ realisations of algorithm 6.1 to approximate the marginal distributions. Histograms from $3.2 \cdot 10^6$ SSA realisations were used as reference solutions. This figure is continued in figure 7.9. 116
7.9 Continuation of figure 7.8. 117

List of Tables

4.1 A one dimensional reaction network, consisting of four reaction channels and one species. 54

6.1 A reaction network that can be interpreted as the translation of a protein that inhibits its own gene. In this example, we have $d = 2$, $D = 1$, $R = 4$, $J_0 = \{1, 4\}$, $J_1 = \{2, 3\}$. 93

7.1 Species names, symbols and initial conditions for the lac operon network. . 107
7.3 The lactose operon network. 108

Index

β-galactoside permease, 105
$\star$, *see* inflow reaction
τ-leaping, 25

allolactose, 105

binomial coefficient, 8

Chapman-Kolmogorov equation, 11
Chemical Langevin equation, 14, 45
Chemical Master equation, 10, 12, 43
CKE, *see* Chapman-Kolmogorov equation
CLE, *see* Chemical Langevin equation
CME, *see* Chemical Master equation
concentration, 6, 39
conditional expectation, 74
continuous model, 40
continuous PDF, 40
continuous propensity, 8
continuous propensity function, 40
continuous state variable, 8, 40
conversion, 8
curse of dimensionality, 13

dimerisation, 41
direct method, 25
disaccharide, 105
discret state variable, 8
discrete model, 40
discrete PDF, 40
discrete propensity, 8
discrete propensity function, 40
discrete state variable, 40

essential support, 76
Euler-Maruyama method, 14

finite CME, 26
Finite State Projection, 26
Fokker-Planck equation, 14, 15, 45
Fokker-Planck master equation, 89, 92
FPE, *see* Fokker-Planck equation
FPME, *see* Fokker-Planck master equation

FSP, *see* Finite State Projection

galactose, 105
glucose, 105

hybrid process, 70, 91

indicator function γ, 66
inflow, 8
inflow reaction, 8, 40

Kramers-Moyal expansion, 19
Kurtz process, 9, 10

lac operon, 105, 107
lactose, 105
lacZ, 105
Liouville equation, 16, 44
Liouville master equation, 62, 72
LME, *see* Liouville master equation
LVE, *see* Liouville equation, 16

marginal distribution, 74
monomolecular reaction, 8, 12
Monomolecular Reaction Channel, 8
monosaccharide, 105
multi-index, 125

Next Reaction Method, 25

OFSP, *see* Optimal Finite State Projection
operator
  CME, 49
  FPE, 49
  FPME, 97
Optimal Finite State Projection, 27
outflow, 8
outflow reaction, 8

particle number, 6, 8
PDF, *see* probability density function
probability density function, 10
probability distribution

Multinomial, 124
Poisson, 123
product Poisson, 123
promotor, 105
propensity function, 8
hybrid, 66
scaled, 40

reaction, 7
reaction channel, 5, 7
reaction network, 7
hybrid, 65
Reaction Probability Density Function, 23
Reaction Rate equation, 16, 45
repressor, 105
RRE, *see* Reaction Rate equation, 16

scale difference, 76
SDE, *see* Stochastic Differential Equation
shift operator, 67
species, 5, 7
continuous, 65
deterministic, 65
discrete, 65
stochastic, 65
splitting scheme
Trotter, 74
SSA, *see* Stochastic Simulation Algorithm
Stochastic Differential Equation, 13
Stochastic Simulation Algorithm, 23, 25
stoichiometric factor, 7
stoichiometric vector, 7

thermodynamic limit, 40
transformed FPE, 49
Trotter product formula, 74

Wiener process, 14