# Noise and Prediction in Molecular Systems




Luca Laurenti
Trinity College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary 2018

# Abstract

Living systems are inherently stochastic and operate in a noisy environment: in single cells, reactions that involve low numbers of molecules generate stochastic fluctuations that propagate to all dependent processes. Such a noise can be a disturbance, for example by disrupting cell cycle control, or could be advantageous by enabling probabilistic strategies when favourable. Unfortunately, the underlying principles behind these probabilistic phenomena are still not well understood. This also limits our capacity to engineer reliable synthetic biological circuits.

In this thesis we study Chemical Reaction Networks (CRNs) with a stochastic semantics as a probabilistic model for biochemical systems. In the first part of the thesis we consider probabilistic model checking of CRNs, which enables rigorous quantitative analysis of stochastic CRNs against a fragment of Continuous Stochastic Logic (CSL) with reward operators. Classical numerical algorithms for CSL model checking are hindered by the state space explosion problem. Here, we employ a continuous-space approximation of the stochastic model of a CRN in terms of a Gaussian process through the Central Limit Approximation (CLA). We then develop efficient and scalable approximate model checking algorithms on the resulting Gaussian process, where we restrict the probabilistic reachability problem to convex target regions. This allows us to derive an abstraction in terms of a time-inhomogeneous discrete-time Markov chain (DTMC), whose dimension is independent of the number of species on which model checking is performed, thus alleviating the state space explosion problem. We prove the correctness of our approach by demonstrating the convergence in distribution of our algorithms.

In the second part of the thesis we employ mathematical analysis and the previously developed model checking techniques to reverse engineer molecular systems. We investigate how CRNs propagate and reduce the

noise. Inspired by the concept of low-pass filters in electronics, we introduce the concept of molecular filter, which is a CRN that can reduce the noise of an input signal while still maintaining certain features of its time evolution. We investigate three different molecular noise filtering mechanisms, study their noise reduction capabilities and limitations, and show that non-linear dynamics, such as complex formation, are necessary for efficient noise reduction. We further suggest that the derived molecular filters are widespread in gene expression and regulation and, particularly, that microRNAs can serve as such noise filters.

# Acknowledgements

First and foremost, I would like to express my gratitude towards my supervisors, Luca Cardelli and Marta Kwiatkowska: without their guidance and support this thesis would have not be possible. I am particularly grateful to them for granting me the freedom to work on any problem I found interesting.

I thank Alessandro Abate and Luca Bortolussi for the many discussions we had, and for inspiring me with their work. I also thank Milan Ceska, Morteza Lahijanian, Nicola Paoletti, and Max Whitby for always finding time to discuss ideas and for their inspiring inputs. My thank go also to Luca Cardelli, Andrew Phillips, and the Biological Computation group at Microsoft Research, for hosting me many times in Cambridge, for their many inputs and for giving me many stimulating ideas. I should also thank all my co-authors, without whom my results would have not be possible. The list of my co-authors include: Attila Csikasz-Nagy (King's College & Pzmny Pter Catholic University), Nicola Paoletti (Stony Brook University), Milan Ceska (Brno University of Technology), Alessandro Abate (Oxford), Max Whitby (Oxford), Morteza Lahijanian (Oxford), Andrea Patane (Oxford), Luca Bortolussi (University of Trieste), Boyan Yordanov (Microsoft Research), and Martin Franzle (University of Oldenburg). I also acknowledge the other students in the Marta's group, whose presence made working during my DPhil a much more enjoyable experience. A final thanks goes to all my friend and to my family, who granted me unconditioned support.

Luca Laurenti

April 2018, Oxford

# Contents


**1 Introduction** **9**

1.1 Contributions . . . . . 11

1.2 Thesis Structure . . . . . 12

1.3 Publications . . . . . 13

1.4 Tools . . . . . 14

**2 Related Works** **15**

2.1 Chemical Reaction Networks . . . . . 16

2.2 Central Limit Approximation . . . . . 17

2.2.1 Limitations of the CLA . . . . . 18

2.3 Probabilistic Model Checking . . . . . 18

2.4 Noise in Biochemical Systems . . . . . 20

**3 Preliminaries** **22**

3.1 Probabilistic Models . . . . . 22

3.1.1 Discrete-Time Markov Chain (DTMC) . . . . . 23

3.1.1.1 Reachability for DTMCs . . . . . 24

3.1.2 Continuous-Time Markov Chain (CTMC) . . . . . 26

3.1.2.1 Transient-State and Steady-State probability . . . . 27

3.1.3 Discrete-Time Markov Process (DTMP) . . . . . 27

3.1.4 Gaussian Processes . . . . . 28

3.1.4.1 Properties of Gaussian processes . . . . . 29

3.2 Frequency Analysis of Stochastic Processes . . . . . 31

3.3 Probabilistic Model Checking . . . . . 31

3.3.1 Continuous Stochastic Logic (CSL) . . . . . 32

3.3.1.1 State Space Explosion Problem . . . . . 34

3.4 Summary . . . . . 34

**4 Chemical Reaction Networks** **36**
4.1 Syntax . . . 36
4.2 Deterministic Semantics . . . 37
4.3 Stochastic Model . . . 39
4.3.1 Moment Equations . . . 41
4.3.2 Relationship between Deterministic and Stochastic Models . . . 41
4.4 Central Limit Approximation . . . 42
4.5 CRNs with unspecified variables . . . 45
4.5.1 Synthesis of CRN sketches . . . 47

**5 Model Checking Chemical Reaction Networks** **49**
5.1 Continuous Stochastic Logic (CSL) . . . 51
5.2 The Reachability Operator . . . 52
5.2.1 Time Discretization Scheme . . . 53
5.2.2 Space Discretization . . . 56
5.2.3 Correctness . . . 57
5.2.4 Computation of Reachability Probabilities . . . 59
5.3 Until operator . . . 59
5.4 Correctness . . . 61
5.4.1 Complexity . . . 63
5.5 Rewards . . . 63
5.5.1 Extending CSL with Rewards . . . 65
5.5.2 Computing Expectation and Variance Using Reward Operators 66
5.5.3 Computing Rewards through Central Limit Approximation . . . 66
5.5.3.1 Instantaneous Rewards . . . 67
5.5.3.2 Cumulative Rewards . . . 69
5.5.3.3 Bounded Reachability Rewards . . . 70
5.6 Experimental Results . . . 73
5.6.1 Gene Expression . . . 73
5.6.2 Phosphorelay Network . . . 74
5.7 Further Related Works on CSL Model Checking . . . 76
5.8 Summary . . . 77

**6 Molecular Filters for Noise Reduction** **79**
6.1 Molecular Filters . . . . . 80
6.2 Linear Filters . . . . . 81
6.2.1 Classical Frequency Analysis of Linear Filters . . . . . 85
6.2.2 Stochastic analysis . . . . . 86
6.2.3 Higher Order Linear Filters . . . . . 89
6.3 Annihilation Module . . . . . 90
6.4 Annihilation Filter . . . . . 93
6.5 Numerical Analysis . . . . . 98
6.6 Molecular filters are widespread in gene expression . . . . . 101
6.6.1 Gene and protein expression can work as linear filters . . . . . 101
6.6.2 MicroRNAs can serve as annihilation filters . . . . . 103
6.7 Related Molecular Motifs . . . . . 105
6.8 Summary . . . . . 106

**7 Conclusions** **108**
7.1 Future Work . . . . . 108

**8 Proofs** **110**
8.1 Proofs of Chapter 5 . . . . . 110
8.2 Time discretization of continuous stochastic processes . . . . . 118

**Bibliography** **123**

# Chapter 1

# Introduction

Molecular interactions are inherently stochastic and biochemical systems operate in a noisy environment. Yet, despite all these uncertainties, one of the main features of natural systems is robustness [100]. Many biochemical systems have evolved to reduce noise in order to perform complex functions in a reliable way [144, 159], and others are able to take advantage of the noise in order to perform their computation [67]. However, the natural mechanisms used by biological systems to control noise are still unclear. As a result, many biological mechanisms, particularly when some of the interacting agents are in low numbers, are still not well understood. This gap of knowledge also limits our ability to engineer quantitative molecular systems that can perform computation reliably [157]. Such synthetic systems have the potential to offer tremendous benefits in the areas of health, energy, and the environment [125].

Recent advances in single cell measurements have enabled us to observe molecular interactions and measure variability at the single cell level [126, 122]. Consequently, to integrate such measurement techniques, effective automated methods that permit in-silico analysis of molecular systems are needed. These would allow one to reverse-engineer molecular systems from available data, and to support synthetic biologists to considerably reduce the cost of in vivo or in vitro experiments [27, 125]. As molecular interactions are probabilistic, such automated techniques must be able to capture the stochastic nature of biochemical systems. Unfortunately, existing formal approaches for biological systems are hindered by the state space explosion problem. That is, the dimension of the state space of the mathematical model of a biochemical system is exponential in the number of species or, more commonly, not finite [59]. Thus, obtaining reliable statistics or even performing simulations tends to be very time-consuming, and often not even feasible. As a result, stochasticity is often neglected, and deterministic or non-quantitative models are generally employed [37, 142].

**Aims of the thesis.** The general aim of this thesis is to obtain a deeper understanding of the underlying principles of how molecular systems perform their functions in a noisy environment and with probabilistic interactions. Since existing approaches for quantitative analysis of global properties of a stochastic biochemical system tends to be very time consuming and often unfeasible, we aim to fill this gap and develop a framework that allows one to perform scalable analysis of biochemical systems according to a stochastic model. The resulting methods are then used to reverse-engineer natural systems and study how these control noise. We aim to identify fundamental and general mechanisms that, at the molecular level, can reduce noise, and to understand their properties and limitations. This will provide new insights into how biochemical networks control noise and illuminate why some reoccurring molecular network modules may have been selected by evolution.

**Approach.** Biochemical systems, under the well-mixed assumption [9], are distributed systems that can be modeled as *Chemical Reaction Networks (CRNs)*. CRNs are a language for modelling biochemical interactions, which is equivalent to Petri Nets [132], Process Algebras [48], and Vector Addition Systems [97], and can be given semantics in terms of *Continuous-Time Markov Chains (CTMCs)* [69, 79]. Computing the probability distribution of the species of a CRN over time is achieved by solving the Kolmogorov equation, also known in the biochemical literature as the *Chemical Master Equation (CME)* [161]. Unfortunately, numerical solution methods of the CME based on uniformisation [17] are often infeasible. In fact, in the CME we have an *ordinary differential equation (ODE)* for each state in the state space of the resulting CTMC and the dimension of the state space is at least exponential in the number of species in the CRN, and infinite when species populations are unbounded. A more scalable transient analysis can be achieved by employing statistical model checking [78], but to obtain good accuracy large numbers of simulations are needed. Those are usually very time-consuming and often infeasible.

In this thesis, our approach is to approximate the CTMC induced by a CRN as a *continuous state space* stochastic process by means of the *Central Limit Approximation (CLA)* [69], also known in statistical physics as the *Linear Noise Approximation (LNA)*. That is, a Gaussian process is derived to approximate the original CTMC [161]. As a Gaussian process is fully determined by its expectation and covariance matrix, its complete characterization requires solving a number of differential equations that is quadratic in the number of species and independent of the population

size. As a consequence, the CLA is generally much more scalable than a discrete-state stochastic representation and has already been successfully used for the analysis of biochemical systems [40, 165]. To enable scalable model checking of CRNs with respect to properties expressed in Continuous Stochastic Logic (CSL) [13], we use the CLA and develop efficient model checking algorithms for the resulting Gaussian process. Since a Gaussian process is a stochastic process evolving in continuous space and time, analysis of its temporal properties cannot be computed in closed form and requires a numerical approach. Thus, we discretize time and space and derive abstractions of Gaussian processes in terms of *Discrete-Time Markov Chains (DTMCs)*. The resulting abstractions have the property that their dimension is independent of the number of molecular species in the system, hence alleviating the state space explosion problem. By making use of known inequalities for Gaussian processes, such as the Dudley's entropy integral [65] and the Borell-TIS inequality [3], and of the theory of weak convergence [31], we establish the convergence in distribution of the resulting abstractions.

In order to obtain general principles underlying molecular computation we use the model checking techniques developed in this this thesis to study how CRNs propagate and attenuate noise, accounting for intrinsic as well as extrinsic noise. In fact, it is well known that molecular processes must integrate mechanisms that, directly or indirectly, reduce noise or otherwise confine it. Examples include signaling cascades, which have been shown to reduce extrinsic fluctuations [159], and the role that microRNA plays in attenuating noise of protein expression [144]. To derive general principles for these phenomena, we formalize the concept of a *molecular filter*, that is, a CRN that reduces noise at the molecular level. Using mathematical analysis based on the CLA and the underlying CTMC, we introduce different classes of molecular filters, study their properties and limitations, and show that these are ubiquitous in natural systems.

## 1.1 Contributions

The main contributions of this thesis are briefly summarized below.

- We consider probabilistic model checking for Continuous-Time Markov Chains (CTMCs) induced from CRNs against a time-bounded fragment of Continuous Stochastic Logic (CSL) extended with reward operators. We employ a continuous-space approximation of the CTMC in terms of a Gaussian process based on the CLA. We then develop efficient and scalable approximate model

checking algorithms on the resulting Gaussian process, where we restrict the target regions for probabilistic reachability to be convex polytopes. This allows us to derive an abstraction in terms of a time-inhomogeneous DTMC, whose dimension is independent of the number of species of the CRN on which model checking is performed. Using results from probability theory, we prove the convergence in distribution of our algorithms to the corresponding measures on the original CTMC. We demonstrate that this approach allows us to overcome the state space explosion problem, while still correctly characterizing the stochastic behaviour of the system. In fact, the resulting algorithms, at the price of introducing an approximation, have a complexity that is polynomial in the number of species and is independent of the initial number of molecules. The CLA is introduced in Section 4.4, while model checking for probabilistic and reward properties based on the CLA is presented in Chapter 5.

- We study how biochemical systems control noise, accounting for intrinsic as well as extrinsic noise. We make use of analytical and numerical techniques, included those developed in this thesis (see, e.g. Figure 6.2), and investigate three molecular noise filtering mechanisms, study their noise reduction capabilities and limitations, and show that non-linear dynamics, such as complex formation, is necessary for efficient noise reduction. We further show that the derived molecular filters are widespread in gene expression and regulation and, particularly, that microRNAs can serve as such noise filters. This gives new insight into the fundamental principles used by natural systems to deal with noise. Molecular filters are the subject of Chapter 6.

## 1.2 Thesis Structure

Chapter 2 reviews related works, and in Chapter 3 we give the technical background and introduce the mathematical models used in this thesis. In Chapter 4 we formally introduce Chemical Reaction Networks (CRNs) and the different semantics of CRNs. The main theoretical contributions are presented in Chapters 5-6. Chapter 7 summarizes our work and highlighting possible future directions. We conclude with Section 8 which reports the proofs not included Chapters 5-6.

## 1.3 Publications

The work presented in this thesis has been previously published in jointly authored papers. In particular, the thesis covers [35, 36, 110] essentially completely, and it also discusses some results from [49, 109] partially. In [35] we study probabilistic reachability for CRNs using the CLA, while in [36] these results are extended to perform CSL model checking for CRNs using the CLA, also extended with reward operators. I collaborated in the conception of these papers and in the derivation of the results. I further contributed by performing the numerical evaluation. In [109] we study how continuous stochastic processes can be discretized and abstracted in terms of DTMCs for the purpose of formal analysis. I derived the error bounds for the time-discretization process. Synthesis for CRNs is the object of [49]. I contributed to the definition of sketches for CRNs and developed a semantics for the sketches using the CLA. Finally, in [110] we develop the concept of molecular filters, derive different molecular filters, and study their properties and limitations. I collaborated in the conception of the research ideas published in the paper and performed the research. Further, I derived all the technical details and performed numerical analysis.

The following co-authored papers have also been published but not included in this thesis. In [52], journal version of [50], a first study of the CLA and how this can be used to perform formal analysis of CRNs is presented. One of the limitations of the CLA is addressed in [53]. In fact, since the CLA is a Gaussian process, this does not allow one to represent multi-modal distributions. In [53] we tackled this issue by developing a stochastic hybrid process semantics for CRNs. Stochastic hybrid processes are then studied in [111], where a theoretical framework and a scalable computational method for formal analysis and control synthesis for stochastic hybrid systems against safety specifications is presented. In [1] we formalized the syntax of a core protocol language, which provides a unified description for the models of biochemical systems being experimented on, together with the discrete events representing the liquid-handling steps of biological protocols. We present a deterministic and a stochastic semantics to this language, both defined in terms of hybrid processes. In [51] and in the corresponding journal paper [54], the class of distributions that can be "programmed" with CRNs is explored.

## 1.4 Tools

CLA analysis of CRNs has been implemented in the tool Microsoft Visual GEC (https://www.microsoft.com/en-us/research/project/genetic-engineering-of-living-cells/). I have been involved in the implementation of the CLA and of the derivative matching algorithm for moment equations.

# Chapter 2

# Related Works

*Chemical Reaction Networks (CRNs)* is a language for modelling the behaviour of real-world chemical systems that has been used for analysis of bio-chemical interactions since the 19th century [164, 86]. CRNs have found versatile application in areas ranging from biological systems to sensor networks as a model of a distributed systems with asynchronous interactions [11, 48], and have been shown to be equivalent to stochastic *Petri Nets* [132], *Vector Addition Systems (VAS)* [97] and *population protocols* [12]. Fundamental for the development of the field has been the work of Feinberg, Horn and Jackson [70, 93]. They explored the relationship between the structure of a CRN and its underlying semantics, modelled as a set of autonomous ordinary differential equations (ODEs), the so-called *deterministic semantics* or *rate equations.* For many years, such a deterministic model has been considered as the universal model of CRNs. However, as the field developed, greater emphasis has been given to probabilistic models of CRNs. This was motivated by the inherently stochastic nature of molecular interactions, which requires a probabilistic model.

In Section 2.1 we give a general overview of current literature on Chemical Reaction Networks (CRNs); first, we consider existing works on the deterministic semantics associated to this language, then we focus on the stochastic models of CRNs and principal implications. Then, in Section 2.2 we review the main papers about the Central Limit Approximation (CLA), also called the Linear Noise Approximation (LNA), which models CRNs in terms of Gaussian processes. Section 2.3 presents an overview of the key results and applications of *Probabilistic Model Checking* methods for biochemical systems. Finally, Section 2.4 gives an overview of the current understanding about molecular computation. We show how, for low-number counts, natural systems attenuate noise (stochastic fluctuations) and have evolved to deal with it.

## 2.1 Chemical Reaction Networks

Mass action kinetics is a commonly used kinetic scheme for CRNs, which assumes that the rate of a chemical reaction is proportional to the product of the concentrations of the reacting chemical species [87]. Deterministic semantics and mass action kinetics have been traditionally used for the analysis of biochemical systems and CRNs, for example for studying oscillations [127] or multiple equilibria systems [61]. However, when the number of molecules is not large enough, one needs to take into account the stochastic behaviour of a biochemical system and this may lead to substantially different results. For instance, in [149], Feinberg and Shinar showed that, for a restricted class of CRNs, the equilibrium solution of these CRNs presents *absolute independence robustness*. That is, according to the deterministic semantics, the value of the equilibrium of the system is independent of the initial number of molecules of each species. However, in [8] it is shown that, for the same initial conditions and assuming conservation of the number of molecules, the stochastic model predicts a behaviour that is different, and for $t \to \infty$ there is an event of extinction with probability 1.

**Stochastic Semantics of CRN** If the number of molecules in a CRN is not large, then there is a need to take into account the stochasticity of molecular interactions. The stochastic semantics of a CRN, in its more common form, is given in terms of a CTMC, whose transient evolution is described by the *Kolmogorov Equations*, also called the *Chemical Master Equation* (CME) in the bio-physics literature [162]. An intuitive derivation of the CME is also given by Gillespie [79], where he shows that the CME is exact for any gas-phase chemical system that is kept well-mixed and thermally equilibrated. An equivalent stochastic characterization of the resulting Markov process in terms of the summation of independent unit Poisson processes, one for each reaction, has been considered by Kurtz [10]. The relationship between deterministic and stochastic models of CRNs has been discussed by Either and Kurtz [69]. In [69] they derive the deterministic semantics of a CRN as an approximation of the stochastic semantics, in the limit of an infinitely large count of molecules, thus showing how the deterministic semantics is a correct model only for molecular counts large enough.

The drawback of the stochastic models is their computational complexity. Indeed, the solution of the CME can require the solution of a large or even infinite number of differential equations, which is also dependent on the initial number of molecules. Therefore, solving the CME for real systems is generally infeasible (see [155] for an

overview). An important step towards a solution of this problem is the work of Gillespie; in [78] Gillespie introduces a stochastic method, the *Gillespie's stochastic simulation algorithm,* which is exact and much more scalable than the solution of the CME. The idea is to simulate a single realization of the CTMC modelling the system, instead of solving the CME, thus obtaining an exact *stochastic simulation* of the CRN for a given initial condition. This approach is much more efficient than solving the CME. However, if many simulations are required, for example to compute some properties on the temporal behaviour of a subset of the species, even this approach becomes infeasible. Improved versions of Gillespie's algorithm have been proposed, for example by Bruck et al. [77], and by Gillespie himself [81].

A stochastic equation, which is an approximation of the CME, has been developed by Gillespie [80]; this equation is called the *Chemical Langevin Equation* (CLE). The accuracy of the CLE can be checked by considering two conditions on the propensity rates of the reactions of the CRN, namely the *leap conditions*. Interesting is also the work of Hespanha and coauthors [150, 151], where they develop a set of moment closure techniques to study the time evolution of the moments of the CME up to a pre-determined order.

In some applications the focus is on the behaviour of the CTMC induced by a CRN at the equilibrium. In this setting, the work of Cracium et al. [7] is important. They show that if the conditions for the *Deficiency Zero Theorem* hold then the network admits a product-form stationary limit distribution. This is an important result: indeed, it permits the characterization of limit behaviour of the stochastic process by considering structural properties without any need of state space analysis. This result remains valid even in the case of infinite state spaces.

## 2.2 Central Limit Approximation

In Van Kampen's most famous work [162], the CME is Taylor expanded in the size of the system. The first order truncation of this expansion is given by the rate equations, while the second order truncation of the CME is the LNA. The LNA has also been derived by linearising the *two-term-truncated chemical Kramers-Moyal equation* [74]. In [165], a further derivation of the LNA has been presented. The authors showed that the LNA can be derived as an approximation of the *Chemical Langevin Equation (CLE)*, and therefore is valid if the two leap conditions are satisfied. The multi-dimensional solution of the LNA has been derived by Elf and Ehrenberg [68]; in their work, the authors also present a modified version of the LNA to improve the

behaviour near the critical points. An alternative derivation of the LNA can be obtained using the central limit theorem [69]. Thus, the LNA, in the mathematical community, is also called Central Limit Approximation (CLA). Such a derivation has been presented in [69], where the authors derived an equivalent representation of the LNA using the central limit theorem as a correction term of the deterministic semantics. This work is of particular interest to us, because the authors prove the convergence in distribution [31] of the CLA to the CTMC model of a CRN for a large number of molecules. This mathematically establishes the correctness of the CLA in the framework of convergence in distribution. To emphasize the relationship with the central limit theorem, in this thesis we will use the term CLA. Analysis of the CLA with applications can be found, for example, in the following review papers [84, 83].

### 2.2.1 Limitations of the CLA

The CLA gives a semantics to a CRN which is a Gaussian process. Thus, the CLA inherits not only the advantages, but also the limitations of Gaussian processes [138]: Gaussian processes are fully characterized by their first two moments (mean and covariance), and the resulting distribution is symmetric and uni-dimensional. Moreover, a Gaussian process takes values in a continuous space. As a result, the CLA is not accurate for multi-modal dynamics or for very small molecular populations. Unfortunately, these scenarios are not uncommon in biochemical systems (see, e.g. [75, 115]). To solve this issue stochastic hybrid approaches have been developed [90, 34, 53]. In these approaches only the species with relatively large molecular counts are approximated with a continuous space process. For instance, in [53] the author of this thesis and his co-authors presented a stochastic hybrid model for CRNs, where a subset of species and reactions is treated with the CLA, while the remaining species are treated as a discrete state-space stochastic process. The resulting approach, although accurate for multi-modal distributions, is still computationally very expensive compared to the CLA.

## 2.3 Probabilistic Model Checking

Model checking is a set of techniques to automatically verify properties of systems specified usually in *temporal logic*. The model checking approach to automated verification has been pioneered by Clarke, Emerson, and Sifakis [58]. For the analysis of systems whose behaviour can be represented by a CTMC, Aziz et al. introduced *Continuous Stochastic Logic (CSL)* [13]. CSL has then been extended with rewards

and it permits both transient and steady state analysis [104]. The *steady state* operator of CSL involves solving a set of linear equations, assuming a finite state space. Transient evolution of the CTMC requires instead the solution of a matrix exponential. However, in [16] the authors show that the *time bounded until operator*, and so transient analysis of the CTMC, can be solved by a technique called *uniformization*. Uniformization (or standard uniformization) is a numerical integration method introduced in 1953 by Jensen [95], which calculates the transient probabilities of a finite-state CTMC by approximating the process with a DTMC. For systems with very different time scales, such as biochemical reaction networks, uniformization is computationally expensive, and therefore a variant, the *adaptive uniformization*, has been introduced to reduce the number of iterations required in these cases [163]. Then, in [118], adaptive uniformization has been improved by introducing *fast adaptive uniformization* (FAU) method. CSL and uniformization variants have been implemented in the probabilistic model checker *PRISM* [106]. PRISM also permits model checking of DTMCs and *Markov Decision Processes (MDPs)* [134]. In the context of molecular programming, PRISM and CSL have already been used with success in [107], where the authors use PRISM for the verification of DNA strand displacement devices expressing properties using CSL. This work shows the usefulness of probabilistic model checking in the design phase of molecular systems. The drawback is that their approach is not scalable. PRISM and CSL have also been used for the verification of DNA walkers [63], employing also FAU.

Standard uniformization is not able to compute transient probabilities on infinite state space systems, which are likely to arise in biochemical systems. The *sliding window abstraction* method has been introduced to take into account this problem [89]. In the sliding window abstraction the authors suppose the probability mass is centered in a restricted region of the state space, so they consider only a subset of the (possibly infinite) state space to compute the transient probabilities. The sliding window approach relies on a partial exploration of the state space, and for biochemical systems can be useful only for relatively low number of molecules and quite compact CRNs (with small number of species and reactions) because of the state space explosion problem.

An alternative to uniformization techniques is that of statistical model checking. Statistical model checking solves the model checking problem by simulating the CTMC under analysis for finitely many executions, and use those simulations to infer whether the samples provide a statistical evidence for the satisfaction or violation of the specification [60, 168]. Statistical model checking for CTMCs is currently

supported by many tools such as PRISM [106], COSMOS [23], U-check [42], and PLASMA [44]. Statistical model checking for parametric CTMCs is discussed in [43], where a Gaussian process is built from observation data to approximate the satisfaction of a MITL formula. Note that in [43] a Gaussian processes is derived from observation data to describe the behavior of the Markov process under test against a given specification. This is a different approach than the one of the CLA, where the Gaussian Process is obtained directly as an approximation of the CTMC.

For the analysis of real time systems, *Signal Temporal Logic (STL)* has been introduced as a temporal logic variant tailored to specify desired properties of continuous signals [116]. STL extends *Metric Interval Temporal Logic* (MITL) [5] for the verification of real-valued signals. STL has been recently extended to STL* in [47], where the authors introduce the freeze operator, which enables freezing the value of a signal and then reusing it for the verification of a formula. The Boolean satisfaction value of an STL formula can be enriched by a quantitative measure of the robustness of the satisfaction of the formula. In [64] the authors present several variants of robustness measures that indicate how far a given trajectory lies, in space and time, from satisfying or violating a property. STL with its quantitative measure can then be used in order to analyze deterministic or hybrid systems. Robustness of STL properties for stochastic models has been studied in [26], where the authors address the system design problem. That is, the authors optimise some control parameters of a stochastic model in order to maximise the robustness of the desired specifications.

## 2.4 Noise in Biochemical Systems

Biochemical processes are inherently stochastic and must deal with noise, which presents itself as stochastic fluctuations. These fluctuations can be extrinsic, arising from interactions occurring with other processes in the environment, or intrinsic, resulting from the random timing of the reactions themselves [69, 119, 67, 22, 135, 96]. Molecular processes transform noisy input signals from the environment into output signals through a number of stages, with signals represented by chemical species and each stage implemented by a molecular reaction network. Since each stage can compound the noise, in order to obtain a reliable final output natural systems must integrate mechanisms that, directly or indirectly, reduce noise or otherwise confine it. Examples of such noise controlling mechanisms include signaling cascades, which have been shown to reduce extrinsic fluctuations [159], and the role microRNA plays in attenuating noise of protein expression [144].

Noise reduction has been studied extensively in electronics [146] and certain noise filtering principles have been successfully applied to molecular systems [141, 45, 169]. For instance, negative feedback and feed-forward loops have been shown to reduce noise [117, 28], and fundamental limits for noise suppression of feedback loops have been derived using techniques from control and information theory [112]. Analogues of mechanisms from signal processing such as *low-pass filters*, which transduce low-frequency signals while attenuating high-frequency signals, have also been observed in biochemical systems [136]. This low-pass filter mechanism, both theoretically and experimentally, has been shown to reduce stochastic fluctuations in gene expression [158, 32].

As illustrated by Eldar and Elowitz in their review paper [67], noise can also be functional for molecular systems, allowing the system to execute randomized algorithms and maximize chances of survival. One of the most commonly observed mechanisms in this sense is that of stochastic state switching, in which cells, randomly, switch their state. Theoretical and experimental analysis has shown that random switching between states can be advantageous in unpredictable environments [103, 166]. For instance, Escherichia coli switch spontaneously in and out of a slow-growing state in which they are insensitive to antibiotics [21]. At the same time, it has been established that positive feedback loops can generate such switching transitions [75].

# Chapter 3

# Preliminaries

In this chapter we introduce the technical background material, summarizing known results that we employ in this thesis. We first review the models of stochastic systems that we use. In Section 3.1.1 we introduce DTMCs. CTMCs are discussed in Section 3.1.2, while Gaussian processes are the subject of Section 3.1.4. Section 3.3 is dedicated to reviewing previously known results on model checking of CTMCs against Continuous Stochastic Logic (CSL) [14].

## 3.1 Probabilistic Models

Denote by $\mathcal{P}(Q)$ the powerset of a set $Q$. A measurable space $(Q, \mathcal{Q})$ is such that $\mathcal{Q} \subseteq P(Q)$ is a sigma-algebra of $Q$. That is, $Q \in \mathcal{Q}$ and $\mathcal{Q}$ is closed under complement and countable unions. Given measurable spaces $(Q, \mathcal{Q})$ and $(E, \mathcal{E})$, a function $f : Q \to E$ is $(\mathcal{Q}, \mathcal{E})$-*measurable* (or simply measurable if the context is clear) if, for all $e \in \mathcal{E}$, $f^{-1}(e) \in \mathcal{Q}$, that is, $f$ preserves measurability. A *probability space* is a tuple $(Q, \mathcal{Q}, Pr)$, where $Q$ is the the *sample space*, $\mathcal{Q}$ is a sigma-algebra of Q, and $Pr : \mathcal{Q} \to [0, 1]$ is a probability measure, which assigns a real between 0 and 1 to each set in $\mathcal{Q}$. Given a probability space $(Q, \mathcal{Q}, Pr)$, and a measurable space $(E, \mathcal{E})$, a *random variable* is a $(\mathcal{Q}, \mathcal{E})$-*measurable* function $V : Q \to E$, where $E$ is the space where the random variable takes values.

We now introduce *stochastic processes*. In this thesis, we are interested in dynamical systems. Hence, we restrict ourself to stochastic processes taking values in $T \subseteq \mathbb{R}^m, m > 0$. As a consequence, sometimes we refer to $T$ as a collection of time instants.

**Definition 1.** *(Stochastic Process) Let $(Q, \mathcal{Q}, Pr)$ be a probability space and $T$ a topological space. Then a $\mathbb{R}^m$-valued stochastic process for $m > 0$ is a measurable*

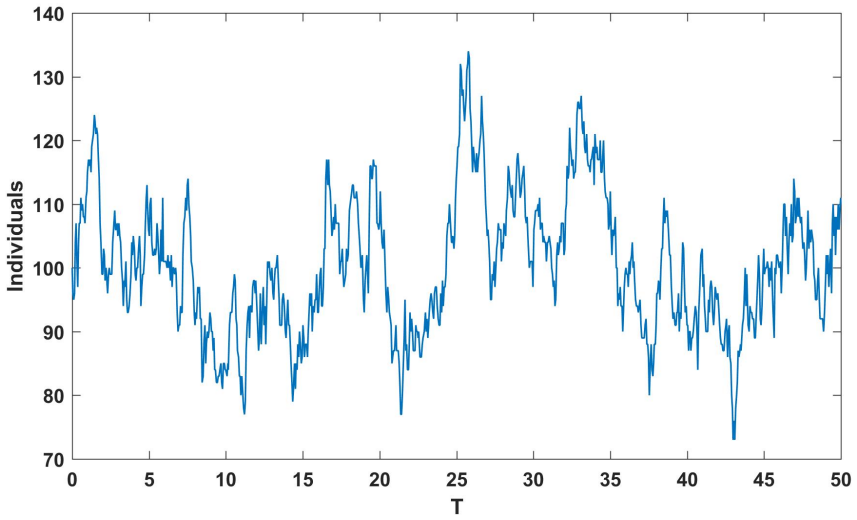



Figure 3.1: Single trajectory of Poisson process $X : Q \times T \to \mathbb{R}_{\geq 0}$, where $T = [0, 50]$ and for each $t \in T$, $X(t)$ is distributed according to a Poisson distribution with mean 100.

*function* $X : Q \times T \to \mathbb{R}^m$.

Thus, a stochastic process can either be seen as a collection of random variables, one for each point in $T$, or as a random variable on the space of the trajectories of the process [30]. That is, for each $t \in T$, $X(t)$ is a random variable, while for $q \in Q$, $X(q)$ is a function representing a trajectory of $X$. Hence, according to the latter perspective, a stochastic process can be seen as a random variable over the space of the trajectories of $X$. Often, viewing stochastic processes as random variables on a functional space can be advantageous. For instance, this is the case when studying the notion of convergence between stochastic processes, as we will do in Chapter 5. An example of stochastic process is shown in Figure 3.1 where a trajectory of a Poisson process [73] is shown.

### 3.1.1 Discrete-Time Markov Chain (DTMC)

The simplest class of stochastic processes we consider are *DTMCs*. A DTMC models a system as a countable set of discrete states, each representing a configuration of the system, together with transitions between states, which are determined via probability distributions depending only on the current state and on time, in case of time-inhomogenous processes. Transitions between states occur in discrete-time.

**Definition 2.** *(Discrete-time Markov chain) A (time-inhomogeneous) discrete-time Markov chain (DTMC)* $\mathcal{D}$ *is uniquely defined by a tuple* $(Q, \mathcal{T}, \gamma)$ *where* $Q$ *is a count-*

*able set of states,* $\gamma : Q \to [0,1]$ *is the initial distribution, and* $\mathcal{T} : Q \times Q \times \mathbb{N} \to [0,1]$ *is the* transition kernel *and it is such that* $\mathcal{T}(x_{k+1}, x_k, k)$ *is the probability that* $\mathcal{D}(k+1) = x_{k+1}$ *given* $\mathcal{D}(k) = x_k$.

Thus, in our definition, a DTMC is a countable set of random variables evolving according probabilistic laws. A path of a DTMC $\omega = x_0 x_1 x_2 ...$ is a (possibly infinite) sequence of states of $Q$ such that, for all $k \geq 0$, $\mathcal{T}(x_{k+1}, x_k, k) > 0$. We denote the sets of finite and infinite paths of a DTMC $\mathcal{D}$ respectively as $\Omega^{fin}$ and $\Omega$. To reason about the probabilistic behaviour of a DTMC $\mathcal{D}$, we need to define a probability measure over its infinite paths. We can do that by using the sigma-algebra generated by the cylinder sets of finite paths $\Omega^{fin}$, where the cylinder set of a finite path $\omega \in \Omega^{fin}$ is the set of infinite paths with prefix $\omega$. In what follows we give the intuitions behind this construction. For further details see [99]. Let $\mathcal{F}(\Omega^{fin})$ be the smallest sigma-algebra on $\Omega^{fin}$ which contains for all $\omega \in \Omega^{fin}$ its cylinder set $Cyl(\omega)$. On the measurable space $(\Omega, \mathcal{F}(\Omega^{fin}))$, for $\omega = x_0 ... x_n$, we can define the probability measure

$$Pr(Cyl(\omega)) = \gamma(x_0) \prod_{k=1}^{n} \mathcal{T}(x_k, x_{k-1}, k-1),$$

where $\gamma \in D(Q)$ and $D(Q)$ is the set of distributions on $Q$. We can now quantify the probability that, given an initial distribution, $\mathcal{D}$ behaves in a specified fashion by identifying the set of paths which satisfy a specification and, assuming that this set is measurable, using the measure $Pr$.

#### 3.1.1.1 Reachability for DTMCs

Given a path $\omega = x_0 x_1 ...$, we call $\omega(k) = x_k$ the state of the path at time $k$. We say that a subset of states $K \subseteq Q$ is *reachable from* $x_0 \in Q$ if there exists a path $\omega$ and $k > 0$ such that $\omega(k) \in K$ and $\omega(0) = x_0$. Computing reachable sets is one of the key properties for analysis of DTMCs. In fact, model checking of DTMCs against most commonly used logics such as the *Linear Temporal Logic (LTL)* or *Probabilistic Computation Tree Logic (PCTL)* reduces to computing the probability of reachable sets [104]. In what follows, we consider DTMCs with finite state spaces (i.e. $|Q|$ is finite) and show how the probability of the reachable sets can be computed on such DTMCs.

**Definition 3.** *( Bounded Reachability Property) Let* $\mathcal{D} = (Q, \mathcal{T}, \gamma)$ *be a DTMC with the set of infinite paths* $\Omega$. *Assume* $\gamma(x) = 1$ *if* $x = x_0$ *and* $0$ *otherwise. Then, given* $K \subseteq Q$, $[n_i, n_f] \subseteq \mathbb{N}$, *we define the* Probabilistic Reachability *problem as*

$$P_{reach}(\mathcal{D}, K, n_i, n_f, x_0) = Pr(\exists k \in [n_i, n_f] \, s.t. \, \omega(k) \in K | \omega \in \Omega)$$

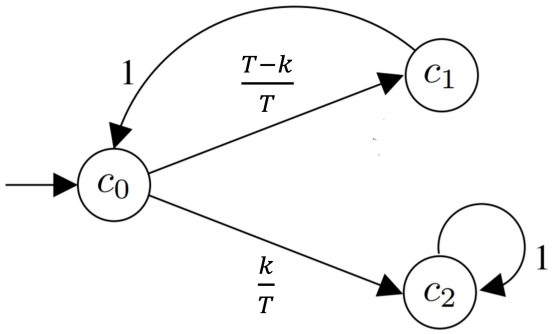



Figure 3.2: Graphical Representation DTMC $\mathcal{D}$ in Example 3.1.1.

$P_{reach}(\mathcal{D}, K, n_i, n_f, x_0)$ is the probability that $\mathcal{D}$ reaches a measurable set $K$ during $[n_i, n_f]$ given that the initial state is $x_0$. Then the probabilistic reachability problem is measurable for finite DTMCs, and can be computed as follows. For $k \in \mathbb{N}$, we have

$$Re(\mathcal{D}, K, n_i, n_f, x_0, k) = \begin{cases} 1 & \text{if } x_0 \in K \\ \sum_{x \in Q} \mathcal{T}(x, x_0, k) Re(\mathcal{D}, K, n_i, n_f, x, k+1) & \text{if } k < n_f \\ 0 & \text{if } k \geq n_f \end{cases} \tag{3.1}$$

Then, it follows that $P_{reach}(\mathcal{D}, K, n_i, n_f, x_0) = Re(\mathcal{D}, K, n_i, n_f, x_0, 0)$.

**Example 3.1.1.** *Consider the time-inhomogeneous DTMC $\mathcal{D}$ represented graphically in Figure 3.2 for $T = 10$ and $k \in [0, T]$. The initial state is $c_0$ and $\mathcal{T}$ is such that $\mathcal{T}(c_0, c_2, k) = \frac{k}{T}, \mathcal{T}(c_0, c_1, k) = \frac{T-k}{T}$, $\mathcal{T}(c_1, c_0, k) = \mathcal{T}(c_2, c_2, k) = 1$ for all $k \in [0, T]$ and 0 otherwise. We consider the property $P_{reach}(\mathcal{D}, \{c_2\}, 0, 5, c_0)$. We obtain*

$$\begin{aligned} P_{reach}(\mathcal{D}, \{c_1\}, 0, 5, c_0) = & \\ & 0 + 1(Re(\mathcal{D}, \{c_1\}, 0, 5, c_1, 1)) = \\ & 0 + 1(Re(\mathcal{D}, \{c_1\}, 0, 5, c_0, 2)) = \\ & 0 + (0.2 + 0.8Re(\mathcal{D}, \{c_1\}, 0, 5, c_1, 3)) = \\ & 0 + (0.2 + 0.8Re(\mathcal{D}, \{c_1\}, 0, 5, c_0, 4)) = \\ & 0 + (0.2 + 0.8(0.4 + 0.6Re(\mathcal{D}, \{c_1\}, 0, 5, c_1, 5)) = \\ & 0 + (0.2 + 0.8(0.4)) = 0.52. \end{aligned}$$

Note that Algorithm (3.1) is valid only for finite-time horizon. Extensions to infinite-time horizon exist and are well studied in the *time-homogeneous* case and they reduce in solving linear equation systems [104].

### 3.1.2 Continuous-Time Markov Chain (CTMC)

CTMCs generalize DTMCs to a continuous-time setting. A CTMC models a system as a countable set of states, each representing a configuration of the system, together with transitions between states, which are determined via probability distributions depending only on the current state. Transitions between states occur according to an exponential distribution.

**Definition 4.** *(CTMC) A (time-homogeneous) Continuous-Time Markov Chain $\mathcal{X}$ is uniquely defined by a tuple $(Q, \mathbf{R}, \gamma)$, where $Q$ is a countable set of states, $\gamma : Q \to [0,1]$ is the initial distribution, and $\mathbf{R} : Q \times Q \to \mathbb{R}_{\geq 0}$ is the* rate transition matrix *such that for $x \in Q$ it holds that $\sum_{x_k \in Q} \mathbf{R}(x, x_k) = 0$. Given $x_1, x_2 \in Q$, the probability that a transition between $x_1$ and $x_2$ occurs in the first $t$ seconds is $1 - e^{-t\mathbf{R}(x_2,x_1)}$. Given $x \in Q$, the time spent in $x$ is exponentially distributed with rate $r(x) = \sum_{x_k \neq x} \mathbf{R}(x_k, x)$, and, when a transition occurs, the probability of jumping to state $x'$ is $\frac{\mathbf{R}(x',x)}{r(x)}$.*

A probability measure for a CTMC $\mathcal{X}$ can be defined in a similar as that for DTMCs using cylinder sets [17]. A path of a CTMC $\mathcal{X}$ is defined as $\omega = x_0 t_0 x_1 t_1 ...$ where $x_k \in Q, t_k \in \mathbb{R}_{\geq 0}$ and such that for all $k \geq 0$ we have $\mathbf{R}(s_{k+1}, s_k) > 0$. That is, $\omega$ is a sequence of states with the residence time in that state. A path $\omega = x_0 t_0 x_1 t_1 ... x_n$ is finite if $x_n$ is absorbing. That is, there are no outgoing transitions from $x_n$. Let $\Omega$ be the set of all (finite and infinite) paths of $\mathcal{X}$, and $\Omega_x$ the set of paths starting from $x$. Call $\omega(t)$ the state of the path at time $t$ (i.e. $\omega(t) = x_n$ where $\sum_{k=0}^{n} t_k \leq t \leq \sum_{k=0}^{n+1} t_k$) and $\omega@k = t_k$ is the residence time in the $k$-th state. Let $x_0, x_1, ..., x_n \in Q$ such that $\mathbf{R}(x_i, x_{i-1}) > 0$ and $I_0, I_1, ..., I_{n-1} \subseteq \mathbb{R}_{\geq 0}$ non-empty time intervals. Then, $Cyl(x_0, I_0, ..., x_{n-1}, I_{n-1}, x_n)$ is the *cylinder set* consisting of all paths $\omega \in \Omega$ such that for $i \leq n$ we have $\omega@i \in I_i$ and $\omega(t) = x_k$ where $\sum_{j=0}^{k} t_j \leq t \leq \sum_{j=0}^{k+1} t_j$. Let $\mathcal{F}(\Omega)$ the smallest sigma-algebra on $\Omega$ which contains all cylinder sets $Cyl(x_0, I_0, ..., x_{n-1}, I_{n-1}, x_n)$ for $x_0, ...x_n$ ranging over all states, $I_0, .., I_{n-1}$ ranging over all non-empty intervals over $\mathbb{R}_{\geq 0}$. The probability measure $Pr$ is the unique probability measure $Pr(Cyl)$ such that $Pr(x_0) = \gamma(x_0)$ and for $n > 0$ defined inductively as follows

$$Pr(Cyl(x_0, I_0, ..., x_n, I', x')) = Pr(Cyl(x_0, I_0, ..., x_n)) \frac{\mathbf{R}(x', x_n)}{r(x_n)} (e^{-ar(x_n)} - e^{-br(x_n)})$$

for $I' = [a, b]$.

#### 3.1.2.1 Transient-State and Steady-State probability

Transient probabilities consider the behaviour of the system for a given finite time $t$, while steady-state probabilities concern the long-run behaviour of the system. Formally, for CTMC $\mathcal{X}$, the transient probability

$$\pi_{\mathcal{X}}(t, x, \gamma) = Pr(\omega(t) = x | \omega \in \Omega) \tag{3.2}$$

is the probability that $\mathcal{X}$ is in state $x$ at time $t$ given initial distribution of states $\gamma$. We call $\pi_{\mathcal{X}}(t, \gamma) = (\pi_{\mathcal{X}}(t, x_1, \gamma), ..., \pi_{\mathcal{X}}(t, x_{|Q|}, \gamma))$ the vector of all the transition probabilities for all states. The transient probabilities can be computed by solving a system of differential equations known as *Chapman-Kolmogorov differential equations*:

$$\frac{d\pi_{\mathcal{X}}(t, \gamma)}{dt} = \pi_{\mathcal{X}}(t, \gamma)\mathbf{Q}, \quad \pi_{\mathcal{X}}(0, \gamma) = \gamma, \tag{3.3}$$

where $\mathbf{Q} = R - diag(E)$ is the *infinitesimal generator matrix* with $diag(E) : Q \times Q \to \mathbb{R}_{\geq 0}$ such that in the diagonal entries $diag(E)(x, x) = r(x)$ and 0 otherwise. Eqn (3.3) is a system of autonomous ODEs in number equal to $|Q|$, the size of $Q$.

*Steady-state probabilities* are defined as:

$$\pi^{\infty}_{\mathcal{X}}(x, \gamma) = \lim_{t \to \infty} \pi_{\mathcal{X}}(t, x, \gamma) \tag{3.4}$$

The above limit always exists for finite-space CTMCs [17], and can be computed by solving a system of linear equations [104, 17].

### 3.1.3 Discrete-Time Markov Process (DTMP)

Discrete-Time Markov Processes (DTMPs) are Markov processes evolving in discrete-time and continuous-space. They represent a generalization of DTMCs to general and possibly uncountable state spaces. In this thesis, DTMPs will be used as abstraction models for continuous-time, continuous-space stochastic processes.

**Definition 5.** *A* (time-inhomogeneous) discrete-time Markov process (DTMP) $(\mathcal{M}(k), k \in [0, I] \subseteq \mathbb{N})$ *is a stochastic process uniquely defined by a triple* $(S, \mathcal{B}(S), \mathcal{T})$*, where* $(S, \mathcal{B}(S))$ *is a measurable space and* $\mathcal{T} : \mathcal{B}(S) \times S \times \mathbb{N} \to [0, 1]$ *is a transition kernel such that, for any* $x \in S$*,* $A \in \mathcal{B}(S)$ *and* $k \in \mathbb{N}$*,* $\mathcal{T}(A, x, k)$ *is the probability that* $\mathcal{M}(k + 1) \in A$ *conditioned on* $\mathcal{M}(k) = z$*.*

From Definition 5, it follows that for $[0, I] \subseteq \mathbb{N}$ $\mathcal{M}$ is a discrete-time Markov stochastic process defined on the sample space given by the product space $\Omega = S^{I+1}$, and can be

endowed with the sigma-algebra $\mathcal{B}(\Omega)$ generated by the product topology, and with a probability measure $Pr^h$, which is uniquely defined by the transition kernel $\mathcal{T}$. We will not discuss further the measure theoretical properties of DTMPs in here. For additional details we refer to [29].

**Example 3.1.2.** *An example of time-inhomogenous DTMP is given by the solution of the following stochastic difference equation*

$$\mathcal{M}(k+1) = F\mathcal{M}(k) + G\omega(k),$$

*where* $F, G \in \mathbb{R}$ *and for* $k \in \mathbb{N}$, $\omega(k)$ *is a random variable with Gaussian distribution with mean* $\mu(k)$ *and variance* $\sigma^2(k)$*. Then, for any measurable set* $A \subseteq R$ *and* $x \in \mathbb{R}$ *we have*

$$\mathcal{T}(A, x, k) = \int_A \mathcal{N}(z \mid Fx + G\mu(k), G^2\sigma^2(k))dz,$$

*where* $\mathcal{N}(z \mid Fx + G\mu(k), G^2\sigma^2(k))$ *is the normal distribution with mean* $Fx + G\mu(k)$ *and variance* $G^2\sigma^2(k)$*.*

### 3.1.4 Gaussian Processes

A *Gaussian process* (GP) is a stochastic process such that, for every finite collection of time instants, the resulting random variables have a multivariate normal distribution, i.e. every finite linear combination of them is normally distributed. Gaussian processes will be widely used in this thesis. As we will see in Section 4.4, this is because, due to the CLA [69], CTMCs can be approximated in terms of Gaussian processes.

In order to formally define Gaussian Processes, we have to introduce the concept of finite-dimensional distributions. Let $(Q, \mathcal{Q}), (E, \mathcal{E})$ be measurable spaces, and $\mathcal{Y} : Q \times \mathbb{R}_{\geq 0} \to E$ be a stochastic process. Let $t_1 < t_2 < ... < t_n$ be a countable sequence of times, and $A_1, A_2, ..., A_n$ a sequence of elements of $\mathcal{E}$. Let $F$ be the event

$$F = \{\forall i \in \{1, ..., n\}, \mathcal{Y}(t_i) \in A_i\} \tag{3.5}$$

Then, assuming the existence of a *probability measure* $Pr$ for $\mathcal{Y}(t_1), ..., \mathcal{Y}(t_n)$, we have

$$Pr(F) = \int_A f(x; t_1, ..., t_n)dx,$$

where $x \in E \times .... \times E = E^N$ and $A = A_1 \times ... \times A_n$. $f$ is called a *finite-dimensional distribution* of $\mathcal{Y}$ of order $n$.

**Definition 6.** *Let $(Q, \mathcal{Q}), (E, \mathcal{E})$ be measurable spaces, and $\mathcal{Y} : Q \times \mathbb{R}_{\geq 0} \to E$ be a stochastic process. Assume $E \subseteq \mathbb{R}^m$. Then, we say that $\mathcal{Y}$ is a* Gaussian process *if all its finite-dimensional distributions (i.e. of any finite order $n \in \mathbb{N}$) have a multivariate normal distribution. That is, assuming $|\Sigma| \neq 0$, we have*

$$Pr(F) = \int_A \frac{\exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^k |\ \Sigma|}} dx, \tag{3.6}$$

*where $\mu = \mathbb{E}[\mathcal{Y}(t_1), ..., \mathcal{Y}(t_n)]$ is the expectation, $\Sigma : nk \times nk \to \mathbb{R}$ is the* covariance matrix *such that $\Sigma(t_k, t_j) = \mathbb{E}[(\mathcal{Y}(t_k) - \mathbb{E}[\mathcal{Y}(t_k)])((\mathcal{Y}(t_j) - \mathbb{E}[\mathcal{Y}(t_j)]))^T]$, $\Sigma^{-1}$ is the inverse of matrix $\Sigma$, and $|\Sigma|$ its determinant.*

Definition 6 completely defines a Gaussian process by considering only a countable number of time points. However, note that $\mathcal{Y}$ evolves in continuous-time. Thus, many path properties depend on uncountable many time instants, and thus may not be measurable for the probability measure $Pr$ as defined in Definition 6 [30]. A classic example is the supremum of the process. However, these issues can be solved under the assumption of *separability* of the process [30]. This assumption guarantees that the behavior of the system is completely characterized by a countable number of points. Thus, properties like the supremum of the process become measurable. Separability of the process is guaranteed, for example, under the assumption that all sample paths are continuous, so that they are determined by their values at all rational times (see [30] for a detailed discussion). Throughout this thesis, we implicitly assume the separability of the stochastic processes under consideration. As discussed in [3] this is not at all a strong assumption and automatically resolves many measurability issues of interest to us.

#### 3.1.4.1 Properties of Gaussian processes

Eqn (3.6) is completely characterized by its mean and covariance matrix. Thus, a key property of Gaussian processes is that they are fully characterized by their mean and covariance matrix. This is a very important property of Gaussian processes, and is the key for many other properties.

**Proposition 1.** *(Closure wrt Linear Combinations) Let $B \in \mathbb{Z}^{m \times n}$ be a matrix and $\mathcal{Y}$ a Gaussian process with values in $\mathbb{R}^n$. Then, $\mathcal{Z} = B \cdot \mathcal{Y}$ is a m-dimensional Gaussian process. For any $t \in \mathbb{R}_{\geq 0}$ we have that $\mathcal{Z}(t)$ is characterized by the following mean and covariance:*

$$\mathbb{E}[\mathcal{Z}(t)] = B\, \mathbb{E}[\mathcal{Y}(t)] \tag{3.7}$$

$$\Sigma_{\mathcal{Z}(t)} = B\, \Sigma_{\mathcal{Y}(t)}\, B^T \tag{3.8}$$

Another key property of Gaussian processes and Gaussian random variables is that for jointly normal-distributed random variable their conditional distribution is still Gaussian

**Proposition 2.** *(Closure with respect to conditioning) Let* $[\mathcal{Y}(t_1), \mathcal{Y}(t_2)]$ *a jointly normal distributed random variable with mean* $\mu = \begin{bmatrix} \mathbb{E}[\mathcal{Y}(t_1)] \\ \mathbb{E}[\mathcal{Y}(t_2)] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ *and covariance matrix* $\Sigma = \begin{bmatrix} \mathbb{E}[(\mathcal{Y}(t_1) - \mathcal{Y}(t_1))(\mathcal{Y}(t_1) - \mathcal{Y}(t_1))^T] & \mathbb{E}[(\mathcal{Y}(t_1) - \mathcal{Y}(t_1))(\mathcal{Y}(t_2) - \mathcal{Y}(t_2))^T] \\ \mathbb{E}[(\mathcal{Y}(t_1) - \mathcal{Y}(t_1))(\mathcal{Y}(t_2) - \mathcal{Y}(t_2))^T] & \mathbb{E}[(\mathcal{Y}(t_2) - \mathcal{Y}(t_2))(\mathcal{Y}(t_2) - \mathcal{Y}(t_2))^T] \end{bmatrix} = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{2,1} & \Sigma_{2,2} \end{bmatrix}$. *Then, the random variable* $[\mathcal{Y}(t_1)|\mathcal{Y}(t_2) = x]$ *is still a normal distributed random variable with mean and covariance matrix given by*

$$\mathbb{E}[\mathcal{Y}(t_1)|\mathcal{Y}(t_2) = x] = \mu_1 + \Sigma_{1,2}\Sigma_{2,2}^{-1}(x - \mu_2) \tag{3.9}$$

$$\Sigma_{t_1|t_2} = \Sigma_{1,1} - \Sigma_{1,2}\Sigma_{2,2}^{-1}\Sigma_{2,1}^T. \tag{3.10}$$

Gaussian processes are continuous-time and continuous-space stochastic processes. Thus, in this thesis, in order to develop general techniques for analysis of Gaussian processes, we need to bound the supremum of a Gaussian process. Given $\mathcal{Y}$ and $T \subset \mathbb{R}_{\geq 0}$ the *supremum of* $\mathcal{Y}$ is defined as

$$\mathcal{Y}_{sup}^T = sup_{t \in T}|\mathcal{Y}(t)|, \tag{3.11}$$

where $|\cdot|$ is a proper metric. A key property of Gaussian processes is that their supremum is still Gaussian distributed. This is due to the *Borell-TIS inequality* [33, 57], which is one of the most important results in all the theory of Gaussian processes.

**Theorem 1.** *(Borell-TIS inequality) Let* $\mathcal{Y}$ *be a zero-mean unidimensional Gaussian process with covariance matrix* $\Sigma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. *Assume* $\mathbb{E}[sup_{t \in T}\mathcal{Y}(t)] < \infty$. *Then, for any* $u > \mathbb{E}[sup_{t \in T}\mathcal{Y}(t)]$ *it holds that*

$$Pr(sup_{t \in T}\mathcal{Y}(t) > u) \leq e^{\frac{(u - \mathbb{E}[sup_{t \in T}\mathcal{Y}(t)])^2}{2\sigma_T^2}}, \tag{3.12}$$

*where* $\sigma_T^2 = sup_{t \in T}\Sigma(t)$.

$\mathbb{E}[sup_{t \in T}\mathcal{Y}(t)]$ can be bounded using the notion of entropy [4]. Thus, Theorem 1 offers a sharp bound for the supremum of the process.

## 3.2 Frequency Analysis of Stochastic Processes

Consider a stochastic process $X(t), t \in \mathbb{R}_{\geq 0}$, with values in $\mathbb{R}_{\geq 0}$. The autocorrelation function of $X$ at time $t$, for $\tau \in \mathbb{R}_{\geq 0}$, is defined as

$$R_X(t, \tau) = \mathbb{E}[X(t)X(t + \tau)].$$

Note that $R_X(t, 0)$ is the second moment of $X$ at time $t$. Under the assumption that $X$ is stationary, $R_X(t, \tau)$ does not depend on $t$, so we write directly $R_X(\tau)$. Provided that $R_X(\tau)$ is absolutely integrable, the *Power Spectral Density (PSD)* of $X$ is defined for each angular frequency $\omega$ in $\mathbb{R}$ as

$$S_X(\omega) = \int_{-\infty}^{+\infty} R_X(\tau) \exp^{-i\omega\tau} \, d\tau.$$

That is, the PSD of a signal is the Fourier transform of its autocorrelation function. From the definition of PSD the following dual relation can be derived

$$R_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) \exp^{i\omega\tau} \, d\omega.$$

Setting $\tau = 0$, we have

$$R_x(0) = \mathbb{E}[X^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) d\omega.$$

If $X$ is zero mean, we have that $R_x(0) = \mathbb{E}[X^2(t)] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_X(\omega) d\omega$ is the variance of $X$ at time $t$. This implies that the variance of $X(t)$ is given by the summation of all the frequency components of its spectrum. As a result, given two stochastic processes $X_1(t), X_2(t)$ with the same spectrum up to frequency $\omega^*$, but with all frequencies $\bar{\omega} > \omega^*$ such that $S_{X_1}(\bar{\omega}) < S_{X_2}(\bar{\omega})$, then, necessarily, the first signal has a smaller variance han the second one. Therefore, attenuating some frequency components of a signal without introducing any amplification at the other frequencies automatically reduces the variance.

## 3.3 Probabilistic Model Checking

Probabilistic model checking aims to, given a probabilistic model, automatically verify statements such as: "The system remains safe within 10000 seconds, with a probability of at least 99.999". Using a probabilistic model checking software, the user, after constructing the model, only has to specify the statement in a temporal logic,

and a model checking algorithm validates if the statement holds or not. In this way, the user does not have to make manual adjustments to their model to compute the desired properties. In this thesis, we employ probabilistic model checking to perform automated analysis of biochemical systems. In this section we review Continuous Stochastic Logic (CSL), which is the prevalent logic for describing continuous-time probabilistic systems, and discuss the state space explosion problem, which limits the applicability of classical model checking algorithms for biochemical systems. This motivates the results introduced in Chapter 5.

### 3.3.1 Continuous Stochastic Logic (CSL)

The prevalent logic to describe temporal properties for CTMC models is Continuous Stochastic Logic (CSL), an extension of Computation Tree Logic (CTL) [13]. CSL expresses three types of properties, which can be summarized as

- *Time-bounded properties* that reason about behaviour up until to a finite time $T$. For instance, $P_{>0.9}[A\, U^{[0,T]}\, B]$ is true if the probability of $B$ becoming true before or at time $T$, while $A$ is true until $B$ is true, is greater than 0.9.
- *Time-unbounded properties* that reason about *unbounded* behavior. An example is $P_{=?}[F\, A]$, which is the probability of $A$ eventually being true.
- *Steady-state properties.* For example, $S_{<0.5}[A]$ is true if, at steady-state, the probability of being in a state where $A$ holds is less than 0.5.

In this thesis we consider only time-bounded properties.

We will define CSL for stochastic process $\mathcal{X}$. Since $\mathcal{X}$ takes values in the real numbers, we will work with predicates over reals, similarly to how real-time signals are verified in *Signal Temporal Logic (STL)* [116], instead of the conventional atomic propositions defined over states of the Markov chain [105].

**Definition 7.** *(CSL Syntax) Given a CTMC $\mathcal{X}$ with values in $E \subseteq \mathbb{R}^m, m > 0$, we define the syntax of CSL as*

$$\Psi \ ::= \ \neg\Psi \mid \Psi_1 \ \wedge \Psi_2 \mid P_{\sim p}(F^{[t_1,t_2]}\, \eta) \mid P_{\sim p}(\eta_1 U^{[t_1,t_2]} \eta_2)$$

*where $\eta, \eta_1, \eta_2 : \mathbb{R}^m \to \{true, false\}$ are predicates and $\sim \in \{<, >\}$*

The above definition differs from the usual definition of CSL in that we do not consider nested CSL properties or unbounded time intervals [104]. Unbounded time

intervals are not considered due to the CLA, whose convergence is guaranteed only for finite time. Instead, nested formulae do not present any theoretical challenge and could be implemented as an extension of the methods presented in this thesis. In Chapter 5, using the CLA, we develop an approximate model checking algorithm for CSL, as presented in Definition 7 and extended with reward operators, which allow us to obtain an exponential speed up compared to existing algorithms.

**Definition 8.** *(Semantics of CSL.) Let $\Omega_x$ be the set of all paths of CTMC $\mathcal{X}$ with initial state $x$. Then, the semantics of CSL is defined as follows*

$$\begin{aligned}
\mathcal{X}, x \models \neg\Psi &\quad\leftrightarrow\quad \mathcal{X}, x \not\models \Psi \\
\mathcal{X}, x \models \Psi_1 \wedge \Psi_2 &\quad\leftrightarrow\quad \mathcal{X}, x \models \Psi_1 \wedge \mathcal{X} \models \Psi_2 \\
\mathcal{X}, x \models P_{\sim p}(F^{[t_1,t_2]}\eta) &\quad\leftrightarrow\quad Pr(\exists t \in [t_1, t_2]\, s.t.\, \eta(\omega(t)) | \omega \in \Omega_x) \sim p \\
\mathcal{X} \models P_{\sim p}(\eta_1 U^{[t_1,t_2]} \eta_2) &\quad\leftrightarrow\quad Pr(\exists t \in [t_1, t_2]\, s.t.\, \eta_2(\omega(t)) \wedge \forall t' \in [0, t)\eta_1(\omega(t') | \omega \in \Omega_x)) \sim p
\end{aligned}$$

Additionally, in the style of probabilistic model checking [92], $\sim$ can be replaced with $=?$, indicating the probability of satisfaction. Note that $F$, the reachability operator, can be expressed with the until. For example, $P_{>0.9}[F^{[0,1]}A] = P_{>0.9}[(1 > 0)\, U^{[0,1]} A]$.

**Example 3.3.1.** *Given a CTMC $\mathcal{X}$ consider the property $P_{=?}[(\mathcal{X}_1 < 100)\, U^{[0,100]}\, ({\mathcal{X}_2}^2 > 20)]$, which computes the probability that in the first 100 unit of times we reach a state where the second component of $\mathcal{X}$ square is greater that 20 and until then the first component of $\mathcal{X}$ remains bounded by 100.*

**Definition 9.** *(CSL Model Checking) Given a CSL specification $\Psi$, CTMC $\mathcal{X}$, initial state $x$, decide if $\mathcal{X}$ satisfies $\Psi$, that is, if*

$$\mathcal{X}, x \models \Psi$$

Model checking algorithms for CSL have been widely investigated [17, 19, 104], and reduce to computing the probability of the reachable sets. We illustrate model checking algorithms for the until operator, $U$, in the next example.

**Example 3.3.2.** *Consider the CSL formula introduced in Example 3.3.1, then model checking for the considered formula for a CTMC $\mathcal{X}$ taking values in $Q \subseteq \mathbb{N}^{|m|}$ can be performed as follows.*

- *Sat is the set of states defined as $Sat = \{x \in Q\, s.t.\, x_2^2 > 20\}$*

- $\bar{\mathcal{X}}$ *is the CTMC that is equal to* $\mathcal{X}$ *except that is absorbing for all* $x \in Sat$. *That is, for each* $x \in Sat$, $\mathbf{R}(x, x') = 0$ *for* $x' \neq x$. *Also* $\bar{\mathcal{X}}$ *has an additional state* $x_{dead}$ *such that, for all* $x \in Q\text{-}Sat$ *satisfying* $x_1 \geq 100$, $\mathbf{R}(x, x') = 1$ *if* $x' = x_{dead}$ *and* 0 *otherwise, where notation* $Q - Sat$ *represents all the states in* $Q$ *that are not in* $Sat$.

- $P_{=?}[(\mathcal{X}_1 < 100)\, U^{[0,100]}\, (\mathcal{X}_2{}^2 > 20)]$ *is equal to the transient probability of* $\bar{\mathcal{X}}$ *being in* $Sat$ *at time* 100. *That is, given* $\gamma$, *the initial distribution of* $\mathcal{X}$, *we have* $P_{=?}[\,(\mathcal{X}_1 < 100)\, U^{[0,100]}\, (\mathcal{X}_2{}^2 > 20)\,] = \pi_{\bar{\mathcal{X}}}(100, \gamma)$, *where*

$$\frac{d\pi_{\bar{\mathcal{X}}}(t,\gamma)}{dt} = \pi_{\bar{\mathcal{X}}}(t,\gamma)\mathbf{Q}, \quad \pi_{\bar{\mathcal{X}}}(0,\gamma) = \gamma,$$

*for* $\mathbf{Q}$ *being the Infinitesimal Generator Matrix of* $\bar{\mathcal{X}}$.

#### 3.3.1.1 State Space Explosion Problem

As shown in Example 3.3.2, model checking CSL requires the computation of Eqn (3.3), namely a system of ODEs in number equal to $|Q|$, the number of states of the CTMC. Although efficient algorithms exist to solve Eqn (3.3) [140], still the dimension of the problem is at least quadratic in $|Q|$, as we need to compute $\mathbf{Q}$. For many applications, $|Q|$ is already exponential in some of the components of the systems. For instance, in Markov population models, $|Q|$ is exponential in the number of agents and often infinite [37], which makes Eqn (3.3) not tractable in practice. This is the so-called *state space explosion problem*. This problem motivates us to develop novel model checking routines, which, at the price of introducing an approximation, allow us to evaluate properties without relying on solution of Eqn (3.3), thus guaranteeing much greater scalability.

## 3.4 Summary

In this chapter we presented the principal stochastic models that we use in this thesis. We introduced time-inhomogeneous DTMCs and discussed reachability analysis for such processes. CTMCs were the subject of Section 3.1.2, where we discussed transient and steady state analysis for CTMCs. We showed that computing the transient distribution for CTMCs requires solving a set of ODEs in number quadratic in the dimension of the state space. Since in many of the applications considered in this thesis the state space dimension is already exponential in some of the dimensions of the problem, solving the transient evolution may be time consuming, or even infeasible,

for many practical applications. Gaussian processes were introduced in Section 3.1.4. We showed that Gaussian processes have many convenient analytic properties, which make them very convenient to use in practice. Finally, in Section 3.3, we discussed current approaches and limitations for probabilistic model checking of CSL formulae for CTMCs.

# Chapter 4

# Chemical Reaction Networks

We define Chemical Reaction Networks (CRNs) and describe the most commonly used semantics for CRNs. We discuss deterministic models in Section 4.2, while stochastic models in terms of CTMCs are considered in Section 4.3. In Section 4.4, we discuss the Central Limit Approximation (CLA), which is a Gaussian model for CRNs. In this thesis CRNs are used to model molecular systems under the well-mixed assumption [9].

## 4.1 Syntax

A *chemical reaction network (CRN)* $\mathcal{C} = (\Lambda, R)$ is a pair of finite sets, where $\Lambda$ is a set of *chemical species*, $|\Lambda|$ denotes its size, and $R$ is a set of reactions. Species in $\Lambda$ interact according to the reactions in $R$. A *reaction* $\tau \in R$ is a triple $\tau = (r_\tau, p_\tau, k_\tau)$, where $r_\tau \in \mathbb{N}^{|\Lambda|}$ is the *source complex*, $p_\tau \in \mathbb{N}^{|\Lambda|}$ is the *product complex* and $k_\tau \in \mathbb{R}_{>0}$ is the coefficient associated to the rate of the reaction; $r_\tau$ and $p_\tau$ represent the stoichiometry of reactants and products. Often, we equivalently write reactions in a more intuitive arrow-based notation. For example, for $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$, given a reaction $\tau_1 = ([1,0,1]^T, [0,2,0]^T, k_1)$ where $\cdot^T$ is the transpose of a vector, we can refer to it as $\tau_1 : \lambda_1 + \lambda_3 \to^{k_1} 2\lambda_2$. The *net change* associated to a reaction $\tau$ is defined by $\upsilon_\tau = p_\tau - r_\tau$. For example, for $\tau_1$ as above, we have $\upsilon_{\tau_1} = [-1, 2, -1]^T$.

Given CRNs $\mathcal{C} = (\Lambda, R)$ and $\mathcal{C}' = (\Lambda', R')$, we define the union of $\mathcal{C}$ and $\mathcal{C}'$ as the CRN $\mathcal{C} \cup \mathcal{C}' = (\Lambda \cup \Lambda', R \cup R')$. That is, the union of two CRNs is the CRN given by the union of the species and reaction sets. A *chemical reaction system* (CRS) $\mathcal{C} = (\Lambda, R, x_0)$ is a tuple where $(\Lambda, R)$ is a CRN and $x_0 \in \mathbb{N}^{|\Lambda|}$ represents its initial condition. Since a CRS is a CRN with an initial condition, we sometimes use these terms interchangeably.

**Example 4.1.1.** *Consider the CRN* $\mathcal{C} = (\{\lambda_1, \lambda_2, \lambda_3\}, R)$*, where* $R = \{([1,1,0]^T, [0,2,0], 10), ([0,1,1]^T, [0,0,2], 10)\}$*. Then* $\mathcal{C}$ *can be expressed as*

$$\lambda_1 + \lambda_2 \rightarrow^{10} \lambda_2 + \lambda_2$$

$$\lambda_2 + \lambda_3 \rightarrow^{10} \lambda_3 + \lambda_3$$

**Well-Mixed Assumption** For all the CRNs considered in this thesis, we assume that the system is well-mixed, that is, the probability of the next reaction occurring between two molecules is independent of the location of those molecules, at fixed volume $V$ and temperature. Under these assumptions a *configuration* or *state* of the system $x \in \mathbb{N}^{|\Lambda|}$ is given by the number of molecules of each species. Given a configuration $x$ we define $\hat{x} = \frac{x}{N}$, where $N = V \cdot N_A$ is the *volumetric factor* where $V$ is the volume and $N_A$ Avogadro's number. Then, for $\lambda \in \Lambda$, $x_\lambda$ represents the number of molecules of $\lambda$ in the configuration $x$ and $\frac{x_\lambda}{N} = \hat{x}_\lambda$ is the concentration of $\lambda$ in the same configuration.

## 4.2 Deterministic Semantics

Often, CRNs are given a semantics in terms of a deterministic model given by a set of ordinary differential equations (ODEs). This is called deterministic or fluid semantics. Such deterministic semantics corresponds to the so called *rate equations*. Let $\mathcal{C} = (\Lambda, R)$ be a CRN, then the rate equations, modelling the concentration of the species in $\Lambda$ over time, are given by a set of autonomous polynomial first order ODEs:

$$\frac{\mathrm{d}\Phi(t)}{\mathrm{d}t} = F(\Phi(t)), \tag{4.1}$$

where $F(\Phi(t)) = \sum_{\tau=(r_\tau,p_\tau,k_\tau)\in R} \upsilon_r \bar{\alpha}_\tau(\Phi(t))$ is called the *drift term* and $\bar{\alpha}_\tau(\Phi(t))$ is the *propensity rate* of reaction $\tau$ expressed in terms of concentrations. Here, we assume $\bar{\alpha}_\tau : \mathbb{R}^{|\Lambda|} \rightarrow \mathbb{R}_{\geq 0}$ are real analytic functions [37]. That is, they possess derivatives of all orders and agree with their Taylor series in a neighborhood of every point. This guarantees that $\Phi$ exists and is unique [69]. This is general and includes all the most commonly used kinetics. A general and very common kinetics used in this thesis is *mass action*. According to mass action kinetics we have $\bar{\alpha}_\tau(\Phi(t)) = k_\tau \prod_{i=1}^{|\Lambda|} \Phi_i(t)^{r_{\tau,i}}$. That is, the propensity rate of a reaction is proportional to the product of the concentrations of the reactant species. $\Phi : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}^{|\Lambda|}$ describes the concentration of the species over time, therefore $\Phi(t) \in \mathbb{R}^{|\Lambda|}_{\geq 0}$ is the vector of the

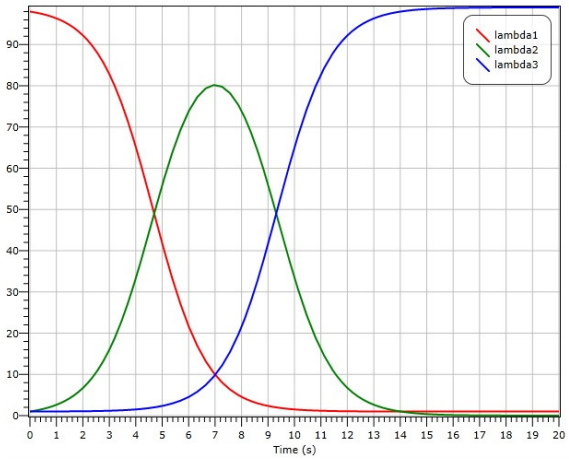



Figure 4.1: The figure plots the time evolution of the deterministic semantics of CRN $\mathcal{C}$ of Example 4.2.1. For each species the figure plots $N\Phi_i(t)$ for $t \in [0, 20]$ .

species concentrations at time $t$. Assuming $t_0 = 0$, the initial condition is $\Phi(0) = \frac{x_0}{N}$, where $x_0$ is the initial configuration of the system and $N$ is the volumetric factor.

A formal definition of the deterministic model (or semantics) of a CRN can be given as follows.

**Definition 10.** *(Deterministic Semantics) Let $\mathcal{C} = (\Lambda, R)$ be a CRN, $x_0 \in \mathbb{R}^{|\Lambda|}_{\geq 0}$, $N > 0$ be respectively initial molecular counts and the volumetric factor. Let $F : \mathbb{R}^{|\Lambda|} \to \mathbb{R}^{|\Lambda|}$ be the drift term for $\mathcal{C}$. Then, the deterministic semantics of the CRS $(\Lambda, R, x_0)$, for the volumetric factor $N$ and time horizon $T \in \mathbb{R}_{\geq 0} \cup \{\infty\}$,*

$$[\![\cdot]\!] : (CRS \times \mathbb{R}_{\geq 0}) \to \mathbb{R}_{\geq 0} \cup \{\infty\} \to \mathbb{R}_{\geq 0} \to \mathbb{R}^{|\Lambda|}_{\geq 0}$$

*is defined as*

$$[\![((\Lambda, R, x_0), N)]\!](T)(t) = \Phi(t)$$

*where $\Phi : [0...T) \to \mathbb{R}^{|\Lambda|}$ is the solution of $\Phi(t') = \frac{x_0}{N} + \int_0^{t'} F(\Phi(s))ds$.*

In Definition 10 we have explicitly introduced a dependence on a time horizon $T$, because it may happen that the solution of the rate equations is defined only for a finite time horizon [69].

**Example 4.2.1.** *Consider the CRN introduced in Example 4.1.1 and defined by the following reactions:*

$$\lambda_1 + \lambda_2 \to^{10} \lambda_2 + \lambda_2$$

$$\lambda_2 + \lambda_3 \to^{10} \lambda_3 + \lambda_3$$

*as introduced in Example 4.1.1. Consider the initial state $x$ such that $x_{\lambda_1} = 98$, $x_{\lambda_2} = 1$ and $x_{\lambda_3} = 1$, for a system with $N = 1000$. Assume mass action kinetics, then the deterministic semantics of $\mathcal{C}$ is given by the following set of ODEs:*

$$\begin{aligned}
\frac{d\Phi_1(t)}{dt} &= -10\Phi_1(t)\Phi_2(t) \\
\frac{d\Phi_2(t)}{dt} &= 10\Phi_1(t)\Phi_2(t) - 10\Phi_2(t)\Phi_3(t) \\
\frac{d\Phi_3(t)}{dt} &= 10\Phi_2(t)\Phi_3(t)
\end{aligned}$$

*whose time evolution is represented in Figure 4.1.*

## 4.3 Stochastic Model

In order to define a stochastic model for CRNs, given a reaction $\tau$, we define its *stochastic propensity rate* (or simply propensity rate if the context is clear) $\alpha_\tau$ as a function of the current configuration of the system $x$ such that $\alpha_\tau(x)dt$ is the probability that a reaction event occurs in the next infinitesimal interval $dt$. We make the simplifying assumption that for $x \in \mathbb{R}^{|\Lambda|}_{\geq 0}$, $\alpha_\tau(x) = N\bar{\alpha}_\tau(\frac{x}{N})$, where $\bar{\alpha}_\tau$ is the propensity of $\tau$ expressed in terms of concentration, as introduced in the preceding section. This property allows us to obtain simpler equations and holds for all the kinetics used in this thesis and for real-world biochemical systems [82]. However, as shown in [69], similar results can be obtained even for different scaling. Assuming fixed volume, $N$, the volumetric factor, is a constant factor. Therefore, sometimes in the rest of the paper, for the sake of a simpler notation, $N$ is considered embedded inside the coefficient $k$ for any reaction.

Under these modelling assumptions CRNs are well represented by *continuous-time Markov processes* (CTMCs) (see Section 3.1.2).

**Definition 11.** *Given a CRN $\mathcal{C} = (\Lambda, R)$ and the volumetric factor $N$, we define a time-homogeneous CTMC $(\mathcal{X}(t), t \in \mathbb{R}_{\geq 0})$ with discrete state space $Q \subseteq \mathbb{N}^{|\Lambda|}$. Given $x_0 \in Q$, the initial configuration of the system, then $Pr(\mathcal{X}(0) = x_0) = 1$. The transition rate from state $x_i$ to state $x_j$ is defined as $\mathbf{R}(x_i, x_j) = \sum_{\{\tau \in R | x_j = x_i + \upsilon_\tau\}} \alpha_\tau(x_i)$.*

$\mathcal{X}(t)$ describes the stochastic evolution of the molecular populations of each species in $\mathcal{C}$ at time $t$. For $x \in Q$, we define $P^{(t)}(x) = Pr(\mathcal{X}(t) = x | \mathcal{X}(0) = x_0)$, where $x_0$ is the initial configuration. The Kolomogorov Equations, also called Chemical Master Equation (CME) (Eqn 3.3), describes the time evolution of $\mathcal{X}$ as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(P^{(t)}(x)\right) = \sum_{\tau \in R}\{\alpha_\tau(x - \upsilon_\tau)P^{(t)}(x - \upsilon_\tau) - \alpha_\tau(x)P^{(t)}(x).\} \tag{4.2}$$

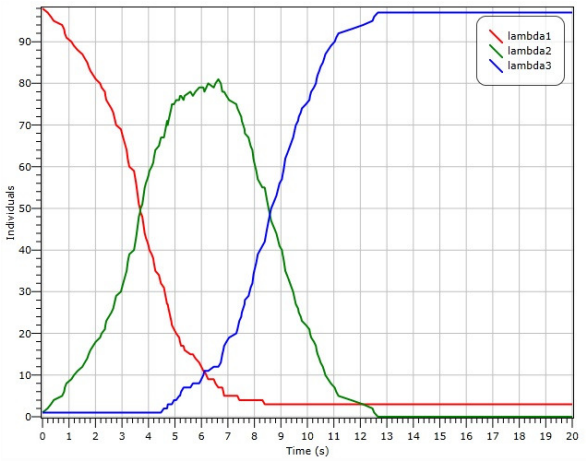



Figure 4.2: The figure plots a realization of the CTMC derived from the CRN $\mathcal{C}$ of Example 4.2.1 from $t \in [0, 20]$.

The CME can be equivalently defined in terms of the infinitesimal generator matrix $\mathbf{Q}$ as

$$\frac{\mathrm{d}}{\mathrm{d}t}P^{(t)}(x) = P^{(t)}(x)\mathbf{Q}$$

where $\mathbf{Q}$ is defined as in Section 3.1.2. Note that $|Q|$ can be countably infinite. This representation admits computing approximations of the CME using, for example, the sliding window method [89].

**Example 4.3.1.** *Consider the CRN of Example 4.2.1 with the same initial condition $x_0$ as in the referenced example. Figure 4.2 plots a trajectory of the resulting CTMC.*

**Definition 12.** *Given a CRN $\mathcal{C} = (\Lambda, R)$ and $\lambda \in \Lambda$, we define expectation, variance and Fano factor of $\lambda$ at time $t$ as*

$$\mathbb{E}_\lambda(t) = \mathbb{E}[\mathcal{X}_\lambda(t)]; \qquad \mathbb{V}_\lambda(t) = \mathbb{E}[(\mathcal{X}_\lambda(t) - \mathbb{E}[\mathcal{X}_\lambda^2(t)])^2]; \qquad F_\lambda(t) = \frac{\mathbb{V}_\lambda(t)}{\mathbb{E}_\lambda(t)}.$$

*Expectation, variance and Fano factor of $\lambda$ at steady state are defined as*

$$\mathbb{E}_\lambda^\infty = \lim_{t\to\infty} \mathbb{E}_\lambda(t); \quad \mathbb{V}_\lambda^\infty = \lim_{t\to\infty} \mathbb{V}_\lambda(t); \quad F_\lambda^\infty = \lim_{t\to\infty} F_\lambda(t)$$

$\mathbb{E}_\lambda^\infty, \mathbb{V}_\lambda^\infty, F_\lambda^\infty$ always exist if $\mathcal{X}$ has a finite state space, but may not be well defined for certain degenerate CRNs, such as the one given by the single reaction $\tau : \lambda+\lambda \to^1 \lambda + \lambda + \lambda$. In this thesis, we do not consider such degenerate cases and we assume $\mathbb{E}_\lambda^\infty, \mathbb{V}_\lambda^\infty, F_\lambda^\infty$ exist and are well defined.

### 4.3.1 Moment Equations

Computing the transient evolution of the CTMC induced by a CRN requires solving a system of ODEs whose size is equal to the number of possible states or configurations of the system (Eqn (4.2)). The number of possible states grows exponentially with the number of species if the species counts are bounded, and is infinite when the species counts are unbounded. As a consequence, obtaining a solution of the CME is generally infeasible, even numerically. This is the *state space explosion problem* described in Section 3.3.1.1. Given a CRN $\mathcal{C} = (\Lambda, R)$ and a polynomial function $\mathcal{G} : \mathbb{N}^{|\Lambda|} \to \mathbb{R}$ over the species of $\Lambda$, it is possible to show that the evolution of the expectation of $\mathcal{G}$ over time can be described by the following system of ODEs, which can be derived directly from the CME [150]

$$\frac{d(\mathbb{E}[\mathcal{G}(\mathcal{X}(t))])}{dt} = \sum_{\tau \in R} \mathbb{E}[\alpha_\tau(\mathcal{X}(t))(\mathcal{G}(\mathcal{X}(t) + \upsilon_\tau) - \mathcal{G}(\mathcal{X}(t)))]. \tag{4.3}$$

Eqn (4.3) can be used to describe the moments of $\mathcal{X}$. For instance, for $\lambda \in \Lambda$ and $\mathcal{G}(\mathcal{X}(t)) = \mathcal{X}_\lambda^2(t)$, where $\mathcal{X}_\lambda$ is the component of $\mathcal{X}$ relative to species $\lambda$, Eqn (4.3) describes the time evolution of the second moment of $\Lambda$ according to its stochastic model. If all the reactions in $R$ are at most uni-molecular, then the equations for the $k$th moment depend on moments of order smaller than or equal to $k$. This leads to a system of ODEs in number polynomial with respect to the number of species, and independent of the molecular counts, ensuring much greater scalability with respect to solving the CME. If, instead, there is at least one multi-molecular reaction (at least two reagents are present in a reaction), then the equations for the $k$th moment may depend on the higher-order moments. This leads to an infinite system of ODEs, which cannot be solved. However, approximations of the moments can still be obtained by using moment closure techniques [150], additionally guaranteeing much greater scalability with respect to numerical solutions of the CME even for non-linear systems.

### 4.3.2 Relationship between Deterministic and Stochastic Models

We define the CTMC $(\frac{\mathcal{X}(t)}{N}, t \in \mathbb{R}_{\geq 0})$ with discrete state space $Q \subseteq \mathbb{R}^{|\Lambda|}$. If $\hat{x}_0 \in Q$ is the initial state, then $Pr(\frac{\mathcal{X}(0)}{N} = \hat{x}_0) = 1$. The transition rate from state $\hat{x}_i$ to $\hat{x}_j$ is defined as $\mathbf{R}(\hat{x}_i, \hat{x}_j) = \sum_{\{\tau \in R | \hat{x}_j = \hat{x}_i + \frac{\upsilon_\tau}{N}\}} \alpha_\tau(N\hat{x}_i)$. $\frac{\mathcal{X}(t)}{N}$ is the random vector describing the system at time $t$ in terms of concentrations. In [69] it is proved that, assuming

the propensity rate of each reaction remains finite up to a finite time $T \in \mathbb{R}_{\geq 0}$, it holds that

$$\lim_{N\to\infty} \sup_{t\leq T} |\frac{\mathcal{X}(t)}{N} - \Phi(t)| = 0 \tag{4.4}$$

almost surely for every finite time $t$. This explains the relationship between the two different semantics: the stochastic model converges to the deterministic model when the number of molecules tends to infinity. Therefore, it can be used as an approximation of the system when the number of molecules is large enough to attenuate stochastic fluctuations.

## 4.4 Central Limit Approximation

Eqn (4.4) illuminates the relationship between deterministic and CTMC models for CRNs: when the molecular population is large, the stochasticity averages out and a deterministic models becomes an appropriate model. However, the CTMC model, due to the state space explosion problem, can be analyzed only in limited cases, where all species are in low population and the state space is finite. Consequently, there is a need for an alternative stochastic model, which remains scalable at the price of introducing some approximations. This is exactly the case of the CLA, which approximates the CTMC induced by a CRN in terms of a Gaussian process. The CLA is a continuous state space approximation of the CTMC in terms of a Gaussian process and is based on the central limit theorem [161, 69].

Given a CRN $\mathcal{C} = (\Lambda, R)$ the CLA at time $t$ approximates the distribution of the induced CTMC $\mathcal{X}(t)$ with the distribution of the random vector $\mathcal{Y}(t)$ such that

$$\mathcal{X}(t) \approx \mathcal{Y}(t) = N\Phi(t) + N^{\frac{1}{2}}G(t) \tag{4.5}$$

where $G(t) = (G_1(t), G_2(t), ..., G_{|\Lambda|})$ is a random vector, independent of $N$, representing the stochastic fluctuations at time $t$ around $\Phi(t)$, the solution of Eqn (4.1). The probability distribution of $G(t)$ is then given by the solution of a linear Fokker-Planck equation [165]. As a consequence, $G$ is a Gaussian process and, for any time instant $t$, $G(t)$ has a multivariate normal distribution whose expected value $\mathbb{E}[G(t)]$ and covariance matrix $\Sigma_{G(t)}$ are the solution of the following differential equations:

$$\frac{\mathrm{d}\mathbb{E}[G(t)]}{\mathrm{d}t} = J_F(\Phi(t))\mathbb{E}[G(t)] \tag{4.6}$$

$$\frac{\mathrm{d}\Sigma_{G(t)}}{\mathrm{d}t} = J_F(\Phi(t))\Sigma_{G(t)} + \Sigma_{G(t)}J_F^T(\Phi(t)) + W(\Phi(t)) \tag{4.7}$$

where $W(\Phi(t)) = \sum_{\tau \in R} \upsilon_\tau {\upsilon_\tau}^T \bar{\alpha}_\tau(\Phi(t))$, $J_F(\Phi(t))$ is the Jacobian of $F(\Phi(t))$, $J_F^T(\Phi(t))$ its transpose, and $F_j(\Phi(t))$ the $j$th component of $F(\Phi(t))$. We assume $\mathcal{X}(0) = x_0$ with probability 1. As a consequence, $\mathbb{E}[G(0)] = 0$ and $\Sigma_{G(0)} = 0^{|\Lambda|\times|\Lambda|}$, which implies $\mathbb{E}_{G(t)} = 0$ for every $t$.

The following theorem illustrates the nature of the approximation in using the CLA.

**Theorem 2.** *[69] Let $\mathcal{C} = (\Lambda, R)$ be a CRN and $\mathcal{X}$ the discrete state space Markov process induced by $\mathcal{C}$. Let $\Phi(t)$ be the solution of Eqns* (4.1) *and $G$ be the Gaussian process with expected value and variance given by Eqns* (4.6) *and* (4.7)*. Then, for any $\in \mathbb{R}_{\geq 0}$*

$$N^{\frac{1}{2}} \left| \frac{\mathcal{X}(t)}{N} - \Phi(t) \right| \Rightarrow_{N\to\infty} G(t). \tag{4.8}$$

In the above, $\Rightarrow_{N\to\infty}$ indicates convergence in distribution when increasing parameter $N$ [31]. The CLA is exact in the limit of high populations, but has also been successfully used in many different scenarios showing surprisingly results [85, 165]. To compute the CLA is necessary to solve $O(|\Lambda|^2)$ first order differential equations, and the complexity is independent of the initial number of molecules of each species. Therefore, one can avoid the exploration of the state space that methods based on uniformization rely upon, making an important step towards the scalable stochastic analysis of reaction systems.

By Eqn (4.5), we have that the distribution of $\mathcal{Y}(t)$ is Gaussian with expected value and covariance matrix given by

$$\mathbb{E}[\mathcal{Y}(t)] = N\Phi(t)$$

$$\Sigma_{\mathcal{Y}(t)} = N^{\frac{1}{2}} \Sigma_{G(t)} N^{\frac{1}{2}} = N\Sigma_{G(t)}.$$

**Example 4.4.1.** *Consider the CRN of Example 4.2.1 with the same initial condition $x_0$ as in the referenced example. Figure 4.2 plots a path of the resulting CTMC.*

The CLA is an approximation of the CTMC induced by a CRN in terms of a Gaussian process. The resulting Gaussian process is completely characterized by expectation and covariance matrix. It is possible to show that expectation and covariance matrix as estimated by the CLA and those computed by using Eqn (4.3) are identical for any possible initial condition for linear systems (i.e. systems with at most uni-molecular reactions) [85]. Instead, in case of non-linear CRNs, they may differ, as the CLA does not depend on the correction terms given by higher order moments. In fact, since $\mathcal{Y}$

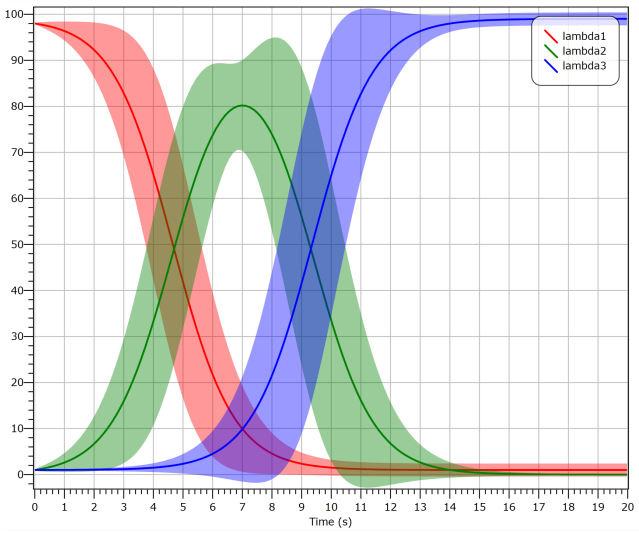



Figure 4.3: The figure plots expectation and standard deviation for the species of CRN $\mathcal{C}$ in Example 4.2.1 for $t \in [0, 20]$. The LNA is plotted using the tool *Microsoft Visual GEC* [130].

is a Gaussian process, the CLA can be not accurate when some of the species have a multi-modal distribution. To overcome this modelling issue, in different works not included in this thesis, we proposed stochastic hybrid models based on the CLA [53].

As the CLA depends only on mean and variance, it can be thought of as a *moment closure technique* (see Section 4.3.1). That is, higher order moments are approximated as a function of the lower order moments. Moment closure techniques have been widely studied in the context of CRNs [91, 150] to approximate the moments of a subset of the species of the CRNs, and efficient algorithms, like *derivative matching* [152], have been proposed to approximate Eqn (4.3) for non-linear systems for any order of moment. However, in this thesis, we will mostly use the CLA. In fact, although other moment closure techniques may be more precise on some specific examples and are not limited by 2nd order moments, the theoretical basis behind the CLA guarantees the convergence in distribution of the resulting Gaussian process to the original CTMC. Similar convergence results are generally missing in other moment closure techniques, whose correctness is typically guaranteed by heuristics. Moreover, many moment closures techniques, such as derivative matching [91], tend to be numerically more unstable than the CLA, especially when some of the species have molecular counts close to 0.

## 4.5 CRNs with unspecified variables

Often the knowledge of a modeller about a molecular process is partial, with little information about the reactions and the rates known. We extend CRNs to model such scenarios by introducing *CRN sketches*. CRN sketches are defined in a similar fashion to concrete CRNs, with the main difference being that species, stoichiometric constants and reaction rates are specified as unknown *variables*. The use of variables considerably increases the expressiveness of the language, allowing the modeller to specify *additional constraints* over them. Constraints facilitate the representation of key background knowledge of the underlying network, e.g. that a reaction is faster than another, or that it consumes more molecules than it produces.

Another important feature is that reactants and products of a reaction are lifted to *choices* of species (and corresponding stoichiometry). In this way, the modeller can explicitly incorporate in the reaction a set of admissible alternatives.

Further, a sketch distinguishes between *optional* and *mandatory* reactions and species. These are used to express that some elements of the network *might* be present and that, on the other hand, other elements *must* be present. CRN sketches will be used in Chapter 6 when studying molecular filters, when we will be interested in leaving part of the network unspecified to show properties that hold for all possible instantiations of the sketch.

**Definition 13** (Sketching language for CRNs)**.** *A* CRN sketch *is a tuple* $\mathcal{S} = (\Lambda, R, \mathsf{Var}, \mathsf{Dec}, \mathsf{Ini}, \mathsf{Con})$*, where:*

- $\Lambda = \Lambda_m \cup \Lambda_o$ *is a finite set of* species*, where* $\Lambda_m$ *and* $\Lambda_o$ *are sets of mandatory and optional species, respectively.*
- $\mathsf{Var} = \mathsf{Var}_\Lambda \cup \mathsf{Var}_c \cup \mathsf{Var}_r$ *is a finite set of* variable names*, where* $\mathsf{Var}_\Lambda$*,* $\mathsf{Var}_c$ *and* $\mathsf{Var}_r$ *are sets of* species, coefficient and rate variables*, respectively.*
- $\mathsf{Dec}$ *is a finite set of* variable declarations*. Declarations bind variable names to their respective domains of evaluation and are of the form* $x : D$*, where* $x \in \mathsf{Var}$ *and* $D$ *is the domain of* $x$*. Three types of declaration are supported:*
  - Species*, where* $x \in \mathsf{Var}_\Lambda$ *and* $D \subseteq \Lambda$ *is a finite non-empty set of species.*
  - Stoichiometric coefficients*, where* $x \in \mathsf{Var}_c$ *and* $D \subseteq \mathbb{N}$ *is a finite non-empty set of non-negative integers.*
  - Rate parameters*, where* $x \in \mathsf{Var}_r$ *and* $D \subseteq \mathbb{R}_{\geq 0}$ *is a bounded set of non-negative reals.*

- $\mathsf{Ini}$ *is the* set of initial states*, that is, a predicate on variables* $\{\Lambda_0\}_{\Lambda\in\Lambda}$ *describing the initial number of molecules for each species.*

- $\mathsf{Con}$ *is a finite set of* additional constraints*, specified as quantifier-free formulas over* $\mathsf{Var}$*.*

- $R = R_m \cup R_o$ *is a finite set of* reactions*, where* $R_m$ *and* $R_o$ *are sets of mandatory and optional reactions, respectively. As for a concrete CRNs, each* $\tau \in R$ *is a triple* $\tau = (r_\tau, p_\tau, k_\tau)$*, where in this case* $k_\tau \in \mathsf{Var}_r$ *is a rate variable; the reaction complex* $r_\tau$ *and the product complex* $p_\tau$ *are* sets of reactants and products*, respectively. A* reactant $R \in r_\tau$ (product $P \in p_\tau$) is a finite choice of species and coefficients*, specified as a (non-empty) set* $R = \{c_i\Lambda_i\}_{i=1,\dots,|R|}$*, where* $c_i \in \mathsf{Var}_c$ *and* $\Lambda_i \in \mathsf{Var}_\Lambda$*.*

  *We denote with* $f_{r_\tau}$ *the* uninterpreted choice function *for the reactants of* $\tau$*, that is, a function* $f_{r_\tau} : r_\tau \to \mathsf{Var}_c \times \mathsf{Var}_\Lambda$ *such that* $f_{r_\tau}(R) \in R$ *for each* $R \in r_\tau$*. The choice function for products,* $f_{p_\tau}$*, is defined equivalently.*

As an example, reaction $\tau = (\{\{c_1\lambda_1, c_2\lambda_2\}, \{c_3\lambda_3\}\}, \{\{c_4\lambda_4, c_5\lambda_5\}\}, k)$ is preferably written as $\{c_1\lambda_1, c_2\lambda_2\} + c_3\lambda_3 \ \to^k \ \{c_4\lambda_4, c_5\lambda_5\}$, using the shortcut $c_3\lambda_3$ to indicate the single-option choice $\{c_3\lambda_3\}$. A possible concrete choice function for the reactants of $\tau$ is the function $\overline{f_{r_\tau}} = \{\{c_1\lambda_1, c_2\lambda_2\} \mapsto c_1\lambda_1, \{c_3\lambda_3\} \mapsto c_3\lambda_3\}$ that chooses option $c_1\lambda_1$ as the first reactant.

**Holes and syntactic sugar.** Unknown information about the network can also be expressed using *holes*, i.e. portions of the model left "unfilled" and resolved by the synthesiser. Holes, denoted with ?, are implicitly encoded through sketch variables. To correctly interpret holes, we assume default domains, $D_r \subseteq \mathbb{R}$ bounded and $D_c \subseteq \mathbb{N}$ finite, for rate and coefficient variables, respectively. We also support the implicit declaration of variables, as shown in Example 4.5.1.

**Example 4.5.1.** *(Bell shape generator)*

*For a given species* $K$*, our goal is to derive a CRN such that the evolution of* $K$*, namely the expected number of molecules of* $K$*, has a bell-shaped profile during a time interval* $[0, 0.5]$*, i.e. during an initial interval the population* $K$ *increases, then reaches the maximum, and finally decreases, eventually dropping to* $0$*. Table 4.1 (left) defines a sketch for the bell-shape generator inspired by the CRN in Example 4.2.1.*

*This sketch reflects our prior knowledge about the control mechanism of the production/degradation of* $K$*. It captures that the solution has to have a reaction generating*

$\Lambda_m = \{K\}, \Lambda_o = \{A, B\}, \mathcal{R}_m = \{\tau_1, \tau_2\},$
$\mathcal{R}_o = \{\tau_3\}$, $\mathsf{Dec} = \{c_1, \ldots, c_4 : [0, 2],$
$k_1, k_2, k_3 : [0, 0.1],\ \lambda_1, \lambda_2 : \{A, B\}\},$
$\mathsf{Con} = \{\lambda_1 \neq \lambda_2,\ c_1 < c_2,\ c_3 > c_4\},$
$\mathsf{Ini} = \{K_0 = 1 \wedge A_0 \in [0, 100] \wedge B_0 \in [0, 100]\}$

$\tau_1 = \lambda_1 + c_1 K \xrightarrow{k_1} c_2 K$
$\tau_2 = \{0, 1\}\lambda_2 + c_3 K \xrightarrow{k_2} ?\lambda_2 + c_4 K$
$\tau_3 = \emptyset \xrightarrow{k_3} \{\lambda_2, [1, 2]K\}$

$A + K \rightarrow^{56} 2K; \quad K + B \rightarrow^{43} 2B$

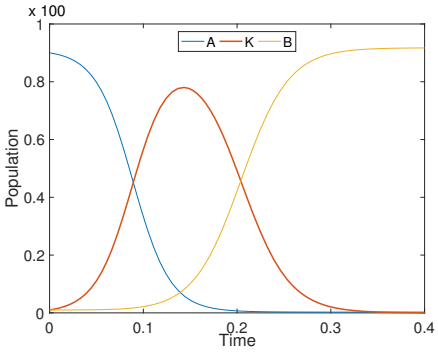



Table 4.1: **Left:** The sketch for bell-shape generator, with Volume $N = 100$. **Right:** CRN producing the bell-shape profile (species K) synthesized by the algorithm presented by the author of this thesis as a co-author in [49]

*$K$ ($\tau_1$) and a reaction where $K$ is consumed ($\tau_2$). We also know that $\tau_1$ requires a species, represented by variable $\lambda_1$, that is consumed by $\tau_1$, and thus $\tau_1$ will be blocked after the initial population of the species is consumed. An additional species, $\lambda_2$, different from $\lambda_1$, may be required. However, the sketch does not specify its role exactly: reaction $\tau_2$ consumes either none or one molecule of $\lambda_2$ and produces an unknown number of $\lambda_2$ molecules, as indicated by the hole ?. There is also an optional reaction, $\tau_3$, that does not have any reactants and produces either 1 molecule of $\lambda_2$ or between 1 and 2 molecules of $K$. The sketch further defines the mandatory and optional sets of species, the domains of the variables, and the initial populations of species. We assume the default domain $D_c = [0, 2]$, meaning that the hole ? can take values from 0 to 2. Note that many sketch variables are implicitly declared, e.g. term $[1, 2]K$ corresponds to $c'\lambda'$ with fresh variables $c' : [1, 2]$ and $\lambda' : \{K\}$.*

*Table 4.1 (right) shows a CRN that satisfies the constraints.*

### 4.5.1 Synthesis of CRN sketches

Our sketching language is also very well suited for synthesis of biological networks: it allows expressing key domain knowledge about the network, and, at the same time, it allows for network under-specification (holes, choices and variables). Given a CRN sketch, a specification in a given temporal logic, and eventually a cost constraint, the optimal synthesis problem seeks to find a concrete instantiation of the sketch (i.e. a CRN) that satisfies a given property and has a minimal cost. This problem has been consider by the author of this thesis in a jointly authored paper in [49]. The problem

has been encoded using satisfiability modulo ODEs [72] and algorithm schemes for its solution have been presented. As this argument will not be part of this thesis, we refer the interested reader to [49].

# Chapter 5

# Model Checking Chemical Reaction Networks

In this chapter we derive fast and scalable approximate model checking algorithms for CTMCs induced by CRNs against a time-bounded fragment of CSL extended with reward operators. The resulting algorithms are numerical and explore the continuous-space approximation of the CTMC in terms of the Gaussian process given by the CLA. One of the key results of this chapter is a novel scalable algorithm for computing probabilistic reachability for Gaussian processes over target regions of the state space that are assumed to be convex polytopes, i.e. intersections of a finite set of linear inequalities. More specifically, for a CTMC approximated as a Gaussian process, the resulting algorithm computes the probability that the system falls in a given target region within a specified time interval. Given that the target region is defined by a set of $m > 0$ linear inequalities, and relying on the fact that a linear combination of the components of a Gaussian distribution is still Gaussian, we discretize time and space for the resulting $m$-dimensional stochastic process. This allows us to derive an abstraction in terms of a time-inhomogeneous *discrete-time Markov chain (DTMC)*, whose dimension is independent of the number of species in the CRN, since a linear combination is a one-dimensional entity. The method ensures scalability, as in general we are interested in a small number, i.e., one or at most two, of linear inequalities. To compute the resulting abstraction, the most delicate aspect is to derive equations for the transition kernel of the resulting DTMC. This is formulated as the conditional probability at the next discrete-time step given the system in a particular state. Reachability probabilities are then computed by making the target set absorbing. Reachability properties are then used to perform model checking of time-bounded CSL properties [105, 18].

We then extend CSL with the reward operators as in [105]. We derive approximate reward measures using the CLA, and prove the convergence in distribution of our algorithms to the original measures when the size of the system (number of molecules) tends to infinity. We show the effectiveness of our approach on a set of case studies taken from the biological literature, also in cases where existing numerical model checking techniques are infeasible.

This chapter is structured as follows. In Section 5.1 we revise the syntax of Continuous Stochastic Logic (CSL) and restrict the operators to be defined on convex sets. Probabilistic model checking algorithms for CSL are the subject of Sections 5.2 and 5.3. In Chapter 5.2 we introduce numerical routines for approximate model checking reachability properties that are then extended in Section 5.3 to perform model checking of the Until operator. Correctness of the presented model checking algorithms is established in Chapter 5.4. CSL is then extended with rewards in Section 5.5, where approximate algorithms for rewards based on the CLA are presented. Two case studies are the subject of Section 5.6. Proofs of the statements of this chapter are reported in Chapter 8, while in Section 5.7 further related works are discussed. We conclude with a summary of the main results.

**Notation.** Since the CLA is correct in the limit of diverging system size $N$, given a CRN $\mathcal{C} = (\Lambda, R)$, in this chapter, we will make explicit the dependence on $N$ of the CTMC induced by $\mathcal{C}$. That is, the stochastic semantics of $\mathcal{C}$ will be called $\mathcal{X}^N$. Similarly, we will call the Gaussian process derived from the CLA of $\mathcal{X}^N$, $\mathcal{Y}^N$. Moreover, in order for the scaling processes to make sense in the limit of diverging $N$, in this chapter, we will work with the *normalized* processes $\hat{\mathcal{X}}^N = \frac{\mathcal{X}^N}{N}$ and $\hat{\mathcal{Y}}^N = \frac{\mathcal{Y}^N}{N}$, which describe the evolution of the species in $\Lambda$ in terms of concentrations. Note that $\hat{\mathcal{X}}^N$ is still a CTMC taking values in $\mathbb{R}^{|\Lambda|}_{\geq 0}$. A path of $\hat{\mathcal{X}}^N$ is defined as $\omega = \hat{x}_0 t_0 \hat{x}_1 t_1 ...$ where $\hat{x}_k \in \mathbb{R}^{|\Lambda|}_{\geq 0}, t_k \in \mathbb{R}_{\geq 0}$ and for all $k \geq 0$ there exists $\tau \in R$ such that $\bar{\alpha}(\hat{x}_k) > 0$ and $\hat{x}_k + \frac{\upsilon_\tau}{N} = \hat{x}_{k+1}$, where $\bar{\alpha}_\tau$ is the density dependent reaction rate and $\upsilon_\tau$ is the net change associated to reaction $\tau$. Thus, a probability measure $Pr$ over cylinder sets of the paths of $\hat{\mathcal{X}}^N$, as defined in Section 3.1.2, can be associated to $\hat{\mathcal{X}}^N$.

We stress that working in terms of concentrations instead of population counts is not a limitation: if we are interested in a fixed value of $N$, population densities can always be rescaled to population counts, and vice versa, by multiplying or dividing them by $N$. In the following, we will thus refer to states and concentrations interchangeably without loss of generality.

## 5.1 Continuous Stochastic Logic (CSL)

Since $\hat{\mathcal{X}}^N$ takes values in $\mathbb{R}^{|\Lambda|}_{\geq 0}$, we will work with predicates over concentrations.

**Definition 14.** *(Convex Predicate). Let $\eta : \mathbb{R}^{|\Lambda|} \to \{\text{true}, \text{false}\}$ be a predicate. We call $\eta$ a* convex predicate *if there exist $B_1, ..., B_m \in \mathbb{Z}^{|\Lambda|}, l_1, ..., l_m \in \mathbb{R}, m > 0$, such that for $\hat{x} \in \mathbb{R}^{|\Lambda|}$ it holds that:*

$$\eta(\hat{x}) = (B_1 \cdot \hat{x} \leq l_1) \wedge ... \wedge (B_m \cdot \hat{x} \leq l_m).$$

Hence, convex predicates are true for concentration $\hat{x}$ belonging to a, not necessarily bounded, convex polytope. We denote by $\Theta$ the set of all convex predicates with domain in $\mathbb{R}^{|\Lambda|}_{\geq 0}$.

We now define a time-bounded fragment of CSL for stochastic CRNs. We do not consider time-unbounded properties because of the nature of the convergence of CLA, which is guaranteed just for finite time. In Section 5.5 we extend this fragment with reward operators.

**Definition 15.** *(CSL Syntax) Given a CTMC $\hat{\mathcal{X}}^N$ induced by CRN $\mathcal{C} = (\Lambda, R)$, we define the syntax of CSL as:*

$$\Psi \ ::= \ \neg\Psi \,|\, \Psi_1 \wedge \Psi_2 \,|\, P_{\sim p}(F^{[t_1,t_2]}\, \eta) \,|\, P_{\sim p}(\eta_1\, U^{[t_1,t_2]}\, \eta_2)$$

*where $\eta, \eta_1, \eta_2 \in \Theta$, $t_1, t_2 \in \mathbb{R}_{\geq 0}$, $\in [0, 1]$ and $\sim \in \{<, >\}$.*

Note also that, in Definition 14, we restrict predicates to sets that are convex polytopes. This does not limit the expressivity of the logic. However, it is a fundamental requirement for our model checking algorithms, which allows us to obtain an exponential speed up compared to existing algorithms.

**Example 5.1.1.** *As a running example we consider the following simple model of gene expression [148], where the mRNA is produced by an always active promoter, and then catalyzes the production of the protein. We have $\Lambda = \{mRNA, Pro\}$ and the following set of reactions $R$:*

$$\to^{0.5} mRNA; \quad mRNA \to^{0.0058 \cdot mRNA} mRNA + Pro$$

$$mRNA \to^{0.0029 \cdot mRNA}; \quad Pro \to^{0.0001 \cdot Pro}.$$

*Fix $N = 100$, then the property "is the probability that the concentration of Pro remains below 0.1 until there is a concentration of mRNA of at least 0.3 in the first 50 time units greater than 0.6?" can be expressed as:*

$$P_{>0.6}[(\hat{Pro} < 0.1)\, U^{[0,50]}\, (\hat{mRNA} > 0.3)],$$

*where, with an abuse of notation, we call $\hat{Pro}$ and $\hat{mRNA}$ the components of vector $\hat{\mathcal{X}}^N$ relative to species $Pro$ and $mRNA$. Obviously, this property is equivalent to the following one, but expressed on the rescaled process $\mathcal{X}^N$:*

$$P_{>0.6}[(Pro < 10)\, U^{[0,50]}\, (mRNA > 30)].$$

Model checking procedures for CTMCs against CSL specifications are well known (see Section 3.3). They reduce to computing the probability of reaching a given set, and hence to solving Eqn (4.2), albeit resulting in the well known state space explosion problem. Here, we explore the usage of the CLA to derive approximate model checking procedures that converge to the original CTMC values but do not suffer from the state space explosion problem, therefore enabling fast stochastic characterization of the system.

## 5.2 The Reachability Operator

In this section we define our CLA-based algorithm to verify the probabilistic reachability operator $P_{\sim p}(F^{[t_1,t_2]}\eta)$, which is the key procedure for model checking of more complex CSL properties. As $\eta$ is a convex predicate, in order to check this property, for a convex polytope $A$ defined as $A = \{x \in \mathbb{R}^{|\Lambda|}_{\geq 0}\, s.t.\, Bx \leq b\}$ where $B \in \mathbb{Z}^{m\times|\Lambda|}, b \in \mathbb{R}^m$, we need to compute:

$$P^A_{reach}(\hat{x}_0, t_1, t_2) = Pr(\exists t \in [t_1, t_2]\, s.t.\, \omega(t) \in A \,|\, \omega \in \Omega_{\hat{x}_0}),$$

where $\Omega_{\hat{x}_0}$ is the set of paths of $\hat{\mathcal{X}}^N$ starting from $\hat{x}_0$ as defined in Section 3.3.1.

We will compute such a probability for $\hat{\mathcal{Y}}^N$, the CLA of $\hat{\mathcal{X}}^N$, and then show how the computed measure converges to the original process, but guaranteeing much greater scalability. Computing the reachability probability for $\hat{\mathcal{Y}}^N$ is not straightforward, because the system evolves in continuous-time and analytic solutions cannot be derived in general. As a consequence, we need to devise numerical algorithms and prove their correctness. Here, we derive a scalable numerical algorithm based on time and space discretization of linear projections of $\hat{\mathcal{Y}}^N$, and, using properties of Gaussian processes, we then prove the convergence of the algorithm to the original measure.

In order to exploit the CLA, we first discretize time for the Gaussian process given by the CLA, with a fixed (or adaptive) step size $h$, which we can do effectively owing to the Markov property and the knowledge of its mean and covariance. As a result, we obtain a *discrete-time, continuous-space*, Markov process with a Gaussian transition kernel. Then, by resorting to state space discretization with parameter

$\Delta z > 0$, we compute the reachability probability on this new process, obtaining an approximation in terms of time-inhomogeneous DTMC converging to the CLA approximation uniformly, when $h$ and $\Delta z$ go to 0. At first sight, there seems to be little gain, as we now have to deal with a $|\Lambda|$-dimensional continuous state space. Indeed, for general regions this can be the case. However, if we restrict to convex polytopes, we can exploit properties of Gaussian distributions (i.e. their closure with respect to linear combinations), reducing the dimension of the continuous space to the number of different linear combinations defining the polytope (in fact, the same hyperplane can be used to fix both an upper and a lower bound). As we are generally interested only in one or few projections, the complexity will then be dramatically reduced.

### 5.2.1 Time Discretization Scheme

Given $\hat{\mathcal{Y}}^N$, the CLA of $\hat{\mathcal{X}}^N$ expressed in terms of concentrations, and matrix $B \in \mathbb{Z}^{m\times|\Lambda|}$, we introduce an exact time discretization scheme for $\hat{\mathcal{Z}}^N = B\hat{\mathcal{Y}}^N$. For simplicity we assume $m = 1$, but all the results extend to $m > 1$. Fix a small time step $h > 0$. By sampling $\hat{\mathcal{Y}}^N$ at step $h$ and invoking the Markov property,[1] we obtain a *discrete-time Markov process* (DTMP) $\hat{\mathcal{Y}}^{h,N}(k) = \hat{\mathcal{Y}}^N(kh)$ on continuous space. Applying the linear projection mapping $\hat{\mathcal{Z}}^N$ to $\hat{\mathcal{Y}}^N(k)$, and leveraging its Gaussian nature, we obtain a process $\hat{\mathcal{Z}}^{h,N}(k) = \hat{\mathcal{Z}}^N(kh)$ which is also a DTMP, though with a kernel depending on time through the mean and variance of $\mathcal{Y}^N$.

From Definition 5 in the preliminaries, it follows that, for $[0, I] \subseteq \mathbb{N}$, $\hat{\mathcal{Z}}^{h,N}$ is a discrete-time stochastic process with a probability measure $Pr^h$, which is uniquely defined by the transition kernel $\mathcal{T}$ [29]. Thus, in order to characterize $\hat{\mathcal{Z}}^{h,N}$, we need to compute its transition kernel, $\mathcal{T}$. This is equivalent to computing $f_{\hat{\mathcal{Z}}^N(t+h)|\hat{\mathcal{Z}}^N(t)=\bar{z}}(z)$, i.e. the density function of $\hat{\mathcal{Z}}^N(t+h)$ given $\hat{\mathcal{Z}}^N(t) = \bar{z}$.

Consider the joint distribution $(\hat{\mathcal{Y}}^N(t), \hat{\mathcal{Y}}^N(t+h))$, which is Gaussian. Its projected counterpart $(\hat{\mathcal{Z}}^N(t), \hat{\mathcal{Z}}^N(t+h))$ is thus also Gaussian, with covariance function:

$$\Sigma_{\hat{\mathcal{Z}}^N(t),\hat{\mathcal{Z}}^N(t+h)} = B\,\Sigma_{\hat{\mathcal{Y}}^N(t),\hat{\mathcal{Y}}^N(t+h)}\,B^T = \frac{1}{N}B\,\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}\,B^T,$$

where $\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}$ is the covariance matrix of $\mathcal{Y}^N$ at times $t$ and $t+h$. It follows by the closure properties of Gaussian processes that $(\hat{\mathcal{Z}}^N(t+h)|\hat{\mathcal{Z}}^N(t) = \bar{z})$ is Gaussian

[1]The Gaussian process obtained by the Central Limit Approximation is Markovian, as it is the solution of a linear Fokker-Planck equation (stochastic differential equation) [161].

too, and thus fully characterized by its mean and variance. Hence, we need to derive $\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}$. Following [69], we introduce the following matrix differential equation:

$$\frac{d\Upsilon(t,s)}{dt} = J_F(\Phi(t))\Upsilon(t,s) \tag{5.1}$$

with $t \geq s$ and initial condition $\Upsilon(s,s) = Id$, where $Id$ is the identity matrix of dimension $|\Lambda|$ and $J_F$ is the Jacobian of matrix $F$ as introduced in Eqn (4.1). Then, as illustrated in [69], we have:

$$\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)} = \int_0^t \Upsilon(t,s)W(\Phi(s))[\Upsilon(t+h,s)]^T ds, \tag{5.2}$$

where $W$ is the matrix introduced in Eqn (4.7). This is an integral equation, which has to be computed numerically. To simplify this task, we derive an equivalent representation in terms of differential equations. This is given by the following lemma.

**Lemma 3.** Solution of Eqn (5.2) is given by the solution of the following differential equations:

$$\begin{aligned}\frac{d\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}}{dt} =& W(\Phi(t))\Upsilon^T(t+h,t) \; + \; J_F(\Phi(t))\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)} \\ &+ \; \Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}J_F^T(\Phi(t+h))\end{aligned} \tag{5.3}$$

with initial condition $\Sigma_{\mathcal{Y}^N(0),\mathcal{Y}^N(t)}$ computed as the solution of:

$$\frac{d\Sigma_{\mathcal{Y}^N(0),\mathcal{Y}^N(s)}}{ds} = \Sigma_{\mathcal{Y}^N(0),\mathcal{Y}^N(s)}J_F^T(\Phi(s)).$$

*Proof.* Applying the general form of the Fundamental Theorem of Calculus to Eqn (5.2) with respect to $t$ we get:

$$\begin{aligned}\frac{d\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}}{dt} &= \Upsilon(t,t)W(\Phi(t))\Upsilon(t+h,t)^T + \int_0^t \frac{d}{dt}(\Upsilon(s,t)W(\Phi(s))\Upsilon(t+h,s)^T)ds \\ &= Id \cdot W(\Phi(t))\Upsilon(t+h,t)^T + \int_0^t \frac{d\Upsilon(s,t)}{dt}W(\Phi(s))\Upsilon(t+h,s)^T ds \\ &\quad + \int_0^t \Upsilon(s,t)W(\Phi(s))\frac{d\Upsilon(t+h,s)}{dt}^T ds.\end{aligned}$$

As $\frac{d\Upsilon(t,s)}{dt} = J_F(\Phi(t))\Upsilon(t,s)$, we get

$$\begin{aligned}\frac{d\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}}{dt} =& W(\Phi(t))\Upsilon(t+h,t)^T + J_F(\Phi(t))\int_0^t \Upsilon(t,s)W(\Phi(s))\Upsilon(t+h,s)^T ds \\ &+ \int_0^t \Upsilon(s,t)W(\Phi(s))\Upsilon(t+h,s)^T ds J_F(\Phi(t+h))^T.\end{aligned}$$

By substituting Eqn (5.2) into the expression above we have the result. Similarly, to derive the initial condition $\Sigma_{\mathcal{Y}^N(0),\mathcal{Y}^N(t)}$ we can apply the Fundamental Theorem of Calculus to Eqn (5.2), but with respect to $h$. □

$\Upsilon(t+h,t)$ can be computed by solving Eqn (5.1). Knowledge of $\Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N(t+h)}$ allows us to directly compute:

$$\Sigma_{\hat{\mathcal{Z}}^N(t),\hat{\mathcal{Z}}^N t+h)} = \frac{1}{N} B\, \Sigma_{\mathcal{Y}^N(t),\mathcal{Y}^N t+h)}\, B^T.$$

Then, by using the law for conditional expectation of a Gaussian distribution, we finally have:

$$\mathbb{E}[\hat{\mathcal{Z}}^N(t+h)|\hat{\mathcal{Z}}^N(t)=\bar{z}] = \mathbb{E}[\hat{\mathcal{Z}}^N(t+h)] + \Sigma_{\hat{\mathcal{Z}}^N(t+h),\hat{\mathcal{Z}}^N(t)}(\Sigma_{\hat{\mathcal{Z}}^N(t)})^{-1}(\bar{z}-\mathbb{E}[\hat{\mathcal{Z}}^N(t)]) \tag{5.4}$$

$$\Sigma_{\hat{\mathcal{Z}}^N(t+h)|\hat{\mathcal{Z}}^N(t)=\bar{z}} = C[\hat{\mathcal{Z}}^N(t+h)] - \Sigma_{\hat{\mathcal{Z}}^N(t),\hat{\mathcal{Z}}^N(t+h)}(\Sigma_{\hat{\mathcal{Z}}^N(t)})^{-1}\Sigma_{\hat{\mathcal{Z}}^N(t),\hat{\mathcal{Z}}^N(t+h)}. \tag{5.5}$$

As the kernel is Gaussian, for $A' \subseteq \mathbb{R}$, $k \in \mathbb{N}$, $z \in \mathbb{R}$ we have that

$$\mathcal{T}(A',z,k) = \int_{A'} \mathcal{N}(s \,|\, \mathbb{E}[\hat{\mathcal{Z}}^N(kh+h)|\hat{\mathcal{Z}}^N(kh)=\bar{z}], \Sigma_{\hat{\mathcal{Z}}^N(kh+h)|\hat{\mathcal{Z}}^N(kh)=\bar{z}})ds,$$

where $\mathcal{N}(s \,|\, E, C)$ is the normal distribution with mean $E$ and covariance $C$. Note that the resulting kernel is time-inhomogeneous. The dependence on time is via the mean and covariance of $\mathcal{Y}^N$, which are functions of time and define completely the distribution of $\mathcal{Y}^N$. The following result guarantees the correctness of the approximation.

**Theorem 4.** *Given vector $B \in \mathbb{Z}^{|\Lambda|}$, $b \in \mathbb{R}$, measurable set $A = \{x \in \mathbb{R}^{|\Lambda|}_{\geq 0}\, s.t.\, Bx \leq b\}$ and process $\hat{\mathcal{Z}}^N = B\hat{\mathcal{Y}}^N$ with initial condition $z_0 = B\hat{x}_0$, call*

$$P^{\hat{\mathcal{Y}}^N,A}_{reach}(\hat{x}_0,t_1,t_2) = Pr^{\hat{\mathcal{Y}}^N}(\exists t \in [t_1,t_2]\, s.t.\, \hat{\mathcal{Y}}^N(t) \in A \,|\, \hat{\mathcal{Y}}^N(0) = \hat{x}_0),$$

*where $Prob^{\hat{\mathcal{Y}}^N}$ is the Gaussian probability measure of $\hat{\mathcal{Y}}^N$. Further, let $\hat{\mathcal{Z}}^{h,N}$ be the DTMP obtained by discretizing $\hat{\mathcal{Z}}^N$ at time step $h > 0$. Then, for $t_1, t_2 \in \mathbb{R}_{\geq 0}$, it holds that*

$$|P^{\hat{\mathcal{Y}}^N,A}_{reach}(\hat{x}_0,t_1,t_2) \;-\; Pr^h(\exists k \in [\lfloor \frac{t_1}{h} \rfloor, \lceil \frac{t_2}{h} \rceil]\, s.t.\, \hat{\mathcal{Z}}^{h,N}(k) \leq b)| \to_{h\to 0} 0,$$

*uniformly.*

Proof of Theorem 4 is reported in Section 8.2, where error bounds are also derived. It uses properties of Gaussian processes to bound the probability that $\mathcal{Z}^N$ reaches a particular region of the state spaces during a sampling interval $(t, t+h) \subseteq \mathbb{R}_{\geq 0}$. Since $[t, t+h]$ is a continuous-time interval, this is a very difficult property to compute in general. However, for Gaussian processes this can be bounded using the Borell-TIS inequality [3].

### 5.2.2 Space Discretization

In order to compute the reachability probability for the DTMP $\hat{\mathcal{Z}}^{h,N}$, we partition its continuous state space into a countable set of non-overlapping cells (regions) of constant size $\Delta z > 0$ (except for at most regions of measure 0, i.e. the boundaries of the cells), obtaining an abstraction in terms of a DTMC $\hat{\mathcal{Z}}^{\Delta z,h,N}$ with state space $S^{\Delta z}$. Specifically, given $S$, the state space of $\hat{\mathcal{Z}}^{h,N}$, and the target set $A = \{x \in \mathbb{R}^{|\Lambda|}\, s.t.\, Bx \leq b\}$, we partition $S \backslash A$ into a grid of cells of length $2\Delta z$, where $\Delta z$ defines how fine our space discretization is. For each of the resulting regions we consider a representative point, given by the median of the set. We call the set of representative points $\hat{S}^{\Delta z}$. Then, we have $S^{\Delta z} = \hat{S}^{\Delta z} \cup \{z_d^A\}$, where $z_d^A$ is the state representing the target set. Theorem 5 guarantees that for $\Delta z \to 0$ the error introduced by the space discretization tends to zero. However, for a fixed $N$, a possible choice of $\Delta z$ is $\Delta z = \frac{0.5}{N}$, which means that the rescaled process $N\hat{\mathcal{Z}}^{\Delta z,h,N}$ takes values in $\mathbb{Z}$. Nevertheless, when the population is of the order of hundreds or thousands, it can be beneficial to consider $\Delta z > \frac{0.5}{N}$, since a coarser state space aggregation is reasonable.

Similarly to the previous section (see Definition 5), as $\hat{\mathcal{Z}}^{\Delta z,h,N}$ is a discrete-time stochastic process, given $[0, I] \subseteq \mathbb{N}$ we can associate to $\hat{\mathcal{Z}}^{\Delta z,h,N}$ a probability space with sample space given by the product space $(S^{\Delta z})^{I+1}$, and with a probability measure $Pr^{\Delta z,h}$ uniquely defined by $\mathcal{T}^{\Delta z}$, the transition kernel of $\hat{\mathcal{Z}}^{\Delta z,h,N}$, which is defined as follows (see Section 3.1.1 for further details on the measure theoretic properties of DTMCs). For $z'_d, z_d \in \hat{S}^{\Delta z}$, $\mathcal{T}^{\Delta z}(z'_d, z_d, k)$ is defined as:

$$\mathcal{T}^{\Delta z}(z'_d, z_d, k) = \int_{z'_d - \Delta z}^{z'_d + \Delta z} f_{\hat{\mathcal{Z}}^N(hk+h)|\hat{\mathcal{Z}}^N(hk)=z_d}(x)dx, \tag{5.6}$$

where $h$ is the discrete-time step, assumed to be fixed to simplify the notation, and $f_{\hat{\mathcal{Z}}^N(hk+h)|\hat{\mathcal{Z}}^N(hk)=z_d}$ is the normal density function of $\hat{\mathcal{Z}}^N(hk+h)$ given $|\hat{\mathcal{Z}}^N(hk) = z_d$. For $z_d \in \hat{S}^{\Delta z}$, we have:

$$\mathcal{T}^{\Delta z}(z_d^A, z_d, k) = \int_A f_{\hat{\mathcal{Z}}^N(hk+h)|\hat{\mathcal{Z}}^N(hk)=z_d}(x)dz, \tag{5.7}$$

and for the last case, we have:

$$\mathcal{T}^{\Delta z}(z_d, z_d^A, k) = \begin{cases} 1 & \text{if } z_d = z_d^A \\ 0 & \text{otherwise} \end{cases}.$$

That is, $z_d^A$ is made absorbing. Finally, we define:

$$P^{\Delta z,h,A}_{reach}(z_d, t_1, t_2) = Pr^{\Delta z,h}(\exists k \in [\lfloor \frac{t_1}{h} \rfloor, \lfloor \frac{t_2}{h} \rfloor] \, s.t. \, \hat{\mathcal{Z}}^{\Delta z,h,N}(k) \in z_d^A \mid \hat{\mathcal{Z}}^{\Delta z,h,N}(0) = z_d).$$

The following theorem, which has been proved in [2] (Theorem 2), guarantees that the error introduced by the state space approximation tends to zero, decreasing $\Delta z$.

**Theorem 5.** *Let $\hat{\mathcal{Z}}^{h,N}$ be a DTMP, and $\hat{\mathcal{Z}}^{\Delta z,h,N}$ the DTMC obtained by space discretization of $\hat{\mathcal{Z}}^{h,N}$ with space discretization step $\Delta z > 0$. Call $z_0$ the initial state of $\hat{\mathcal{Z}}^{h,N}$ and $z_{d,0} \in S^{\Delta z}$ the discrete state representing the region containing $z_0$. Then, for $t_1, t_2 \in \mathbb{R}_{\geq 0}$, and convex polytope $A$,*

$$|Pr^h(\exists k \in [\lfloor \frac{t_1}{h} \rfloor, \lceil \frac{t_2}{h} \rceil] \, s.t. \, \hat{\mathcal{Z}}^{h,N}(k) \in A | \hat{\mathcal{Z}}^{h,N}(0) = z_0) - P^{\Delta z,h,A}_{reach}(z_{d,0}, t_1, t_2)| \rightarrow_{\Delta z} 0,$$

*uniformly.*

### 5.2.3 Correctness

To prove the correctness of our numerical algorithm we need to show that, for any measurable set, the reachability measure computed on $\hat{\mathcal{X}}^N$ converges to that computed on $\hat{\mathcal{Y}}^N$. This is guaranteed by the following theorem.

**Theorem 6.** *Let $\mathcal{C} = (\Lambda, R)$ be a CRN with induced CTMC $\hat{\mathcal{X}}^N$ and $\hat{\mathcal{Z}}^{\Delta z,h,N}$ be the DTMC obtained by space and time discretization of $B\hat{\mathcal{Y}}^N$. Assume $\hat{\mathcal{X}}^N(0) = \hat{x}_0$ and the corresponding initial state for $\hat{\mathcal{Z}}^{\Delta z,h,N}$ is $z_{d,0}$. Then, for $t_1, t_2 \in \mathbb{R}_{\geq 0}$, $B \in \mathbb{R}^{m \times |\Lambda|}$ and $b \in \mathbb{R}^m$ and $A = \{x \in \mathbb{R}^{|\Lambda|}_{\geq 0} \, s.t. \, Bx \leq b\}$, it holds that:*

$$\lim_{N \to \infty} \lim_{h \to 0} \lim_{\Delta z \to 0} |P^A_{reach}(\hat{x}_0, t_1, t_2) - P^{\Delta z,h,A}_{reach}(z_{d,0}, t_1, t_2)| = 0.$$

The proof of Theorem 6 is detailed in the Section 8.1. The main idea is to use Theorems 5 and 4 to show that the numerical model checking algorithms on the Gaussian process $\hat{\mathcal{Y}}^N$ are sound. Then, we employ Theorem 2 and the theory of weak convergence to show the convergence in distribution of the reachability measure on $\hat{\mathcal{X}}^N$ to that on $\hat{\mathcal{Y}}^N$. The proof is complicated by the fact that both $\hat{\mathcal{Y}}^N$ and $\hat{\mathcal{X}}^N$ depend on $N$. As we discussed in the previous sections, error bounds would be a viable companion to quantify the error introduced by our approach. However, although they can be computed for both space and time discretization of Gaussian processes (see Chapter 8), we are not aware of any explicit bound to quantify the convergence of a CTMC to the CLA.

**Algorithm 1** Compute Time-Bounded Probabilistic Reachability

**Input:** CRN $\mathcal{C} = (\Lambda, R)$ with initial concentration $\hat{x}_0$, $B \in \mathbb{Z}^{|\Lambda|}$, $H \subseteq \mathbb{R}$, a finite time interval $[t_1, t_2]$, a partition of the real numbers with the set of representative points $\mathbf{P}_{\Delta z}$, a target set $A = \{x \in \mathbb{R}\, s.t.\, Bx \in H\}$ and a threshold **Th**.

**Output:** $P^A_{reach}(\hat{x}_0, t_1, t_2)$.

Set $t = 0$, $k = 0$, $\mathcal{S} = \{B \cdot \hat{x}_0\}$ with $Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(0) = B \cdot \hat{x}_0) = 1$

**while** $t < t_1$ **do**

Compute time step $h$

$\mathcal{S} \leftarrow \{z_d \in \mathbf{P}_{\Delta z}\, s.t.\, Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d) \geq \mathbf{Th}$ where

$$Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d) = \sum_{z'_d \in \mathcal{S}} Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d | \hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z'_d) Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z'_d)\}$$

$t \leftarrow t + h$

$k \leftarrow k + 1$

**end**

**while** $t < t_2$ **do**

Compute time step $h$

$S' \leftarrow \mathcal{S} \setminus H$

$S_1 \leftarrow \{z_d \in \mathbf{P}_{\Delta z} \setminus H\, s.t.\, Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d) \geq \mathbf{Th}$, where

$$Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d) = \sum_{z'_d \in S'} Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d | \hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z'_d) Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z'_d)\}$$

$S_2 \leftarrow \{z_d \in \mathbf{P}_{\Delta z} \cap H\, s.t.\, Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d) \geq \mathbf{Th}$, where

$$Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d) = Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z_d) + \sum_{z'_d \in S'} Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z_d | \hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z'_d) Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z'_d)\}$$

$\mathcal{S} \leftarrow S_1 \cup S_2$

$t \leftarrow t + h$

$k \leftarrow k + 1$

**end**

**return** $P^A_{reach}(\hat{x}_0, t_1, t_2) = \sum_{z_d \in \mathcal{S} \cap H} Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z_d)$

### 5.2.4 Computation of Reachability Probabilities

Our approach for computing reachability probabilities is summarized in Algorithm 1. Intuitively, we use Definition 3 to compute probabilistic reachability on DTMC $\mathcal{Z}^{\Delta z,h,N}$.

In Line 1, we initialize the system at time 0. In the context of the algorithm, $\mathcal{S}$ is a set containing the states at a particular time with probability mass greater than the threshold **Th**. In our implementation we partition $\mathbb{R}$ with a grid of cells of constant size $\Delta z > 0$. Then, for each cell we select a representative point, and denote the set of representative points $\mathbf{P}_{\Delta z}$. $\mathcal{S}$, for any time $t > 0$, will be a subset of this set. **Th** equals $10^{-14}$ in all our experiments. The use of a threshold guarantees that the algorithm always terminates in finite time. This introduces a truncation error, which can be easily estimated at the run-time as shown in [167]. Initially, we have that $\mathcal{S}$ contains only one state $B \cdot \hat{x}_0$. Then, in Lines $3-7$, we propagate the probability for any discrete step while $t < t_1$, according to Definition 3. For generality, we assume that the time step $h$ is chosen adaptively, according to the system dynamics. Propagating probability is possible, as for any $z'_d \in \mathcal{S}$, $Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(k+1) = z'_d | \hat{\mathcal{Z}}^{\Delta z,h,N}(k) = z_d) = \mathcal{T}^{\Delta z}(z'_d, z_d, k)$. From Line 8 to 15, we compute the probabilistic reachability, $P^A_{reach}(\hat{x}_0, t_1, t_2)$, by propagating the probability only for states that are not in A. In fact, states in A are made absorbing. When we reach $t \geq t_2$, we have that $P^A_{reach}(\hat{x}_0, t_1, t_2) \approx \sum_{z \in S \cap H} Pr^{\Delta z,h}(\hat{\mathcal{Z}}^{\Delta z,h,N}(\lceil \frac{t_2}{h} \rceil) = z | \hat{\mathcal{Z}}^{\Delta z,h,N}(0) = z_{d,0})$.

**Example 5.2.1.** *We return to the CRN introduced in Example 5.1.1, and, for* $N = 100$, *we consider the following reachability property:*

$$P_{=?}(F^{[0,T]}\, \hat{m}RNA > \hat{P}ro + 0.2),\ T \in [0, 100]$$

*where* $=?$*, in the style of PRISM [106] or PEPA [56], represents the quantitative value of a property. The above formula asks for the probability that, during the first* 100 *seconds, the system reaches a state where the mRNA concentration exceeds the protein concentration by more than* 0.2*. In Figure 5.1 we compare the probability value computed by Algorithm 1 with the same property computed on the CTMC* $\hat{\mathcal{X}}^N$ *using PRISM for different values of* $h$*. We assume* $\Delta_z = \frac{0.5}{N}$*, which is justified by the fact that the number of molecules is an integer.*

## 5.3 Until operator

We show how to generalize the computation of the reachability probabilities of the previous section to the until operator. For $\hat{x} \in \mathbb{R}^{|\Lambda|}_{\geq 0}$, let $\eta_1(\hat{x}) = B_1\hat{x} \leq l_1$ and

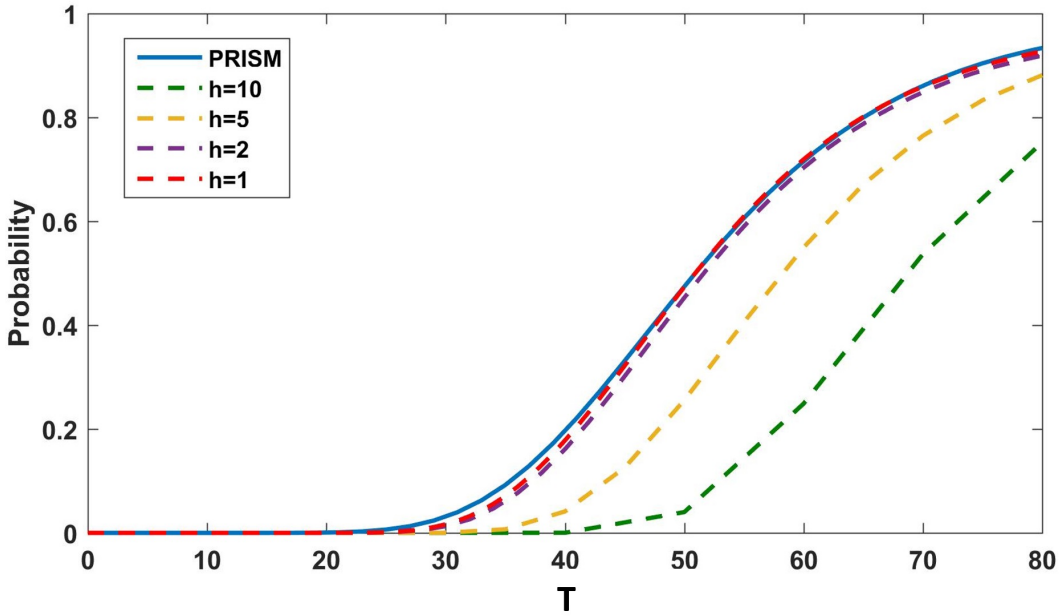



Figure 5.1: Comparison of the evaluation of $P_{=?}(F^{[0,K]} mRNA > Pro + 20), K \in [0, 100]$, on the CTMC as estimated by PRISM [106], and on the CLA approximation for a fixed $\Delta z$ and four different values of $h$.

$\eta_2(\hat{x}) = B_2\hat{x} \leq l_2$, then, by definition we have:

$$\hat{\mathcal{X}}^N, \hat{x} \models P_{\sim p}(\eta_1 U^{[t_1,t_2]} \eta_2) \quad \iff Pr(\exists t \in [t_1, t_2]\, s.t.\, \eta_2(\omega(t)) \wedge \forall t' \in [0, t), \eta_2(\omega(t)) | \omega \in \Omega_{\hat{x}}),$$

where $\Omega_{\hat{x}}$ is the set of paths of $\hat{\mathcal{X}}^N$ starting in $\hat{x}$. To solve this problem we can construct the following stochastic process:

$$\hat{\mathcal{Z}}^N = B\hat{\mathcal{Y}}^N$$

where $B = (B_1, B_2)^T$, and $\hat{\mathcal{Y}}^N$ is the CLA of $\hat{\mathcal{X}}^N$. By the properties of multivariate Gaussian distribution, $\hat{\mathcal{Z}}^N$ is still a Gaussian process with mean and covariance matrix given by

$$\mathbb{E}[\hat{\mathcal{Z}}^N(t)] = B\mathbb{E}[\mathcal{Y}^N(t)] \quad \Sigma_{\hat{\mathcal{Z}}^N(t)} = \frac{1}{N} B\, \Sigma_{\mathcal{Y}^N(t)}\, B^T,\ t \in \mathbb{R}_{\geq 0}.$$

Note that $\hat{\mathcal{Z}}^N$ is again a time-inhomogeneous Markov process, as its kernel depends on the statistics of $\mathcal{Y}^N$. Following the approach of the previous section, we can discretize time and space for $\hat{\mathcal{Z}}^N$, and thus obtain a DTMC $\hat{\mathcal{Z}}^{\Delta z,h,N}$. At this point, the problem reduces to computing the probability for the until operator on the DTMC. Algorithms for computing the resulting measure on a time-inomhogeneous DTMC exist and are well studied (See Section 3.1.1 and [55]). In fact, to compute $P_{\sim p}(\eta_1 U^{[t_1,t_2]} \eta_2)$, we can simply make the regions that do not satisfy $\eta_1$ and those for which $\eta_2$ holds absorbing,

and then compute the probability of reaching a region for which $\eta_2$ is satisfied. This can be computed by resorting to Algorithm 1, as presented in the previous section. Theorem 6 then guarantees the following proposition.

**Proposition 3.** *Let* $\eta_1(\hat{x}) = B_1\hat{x} \leq l_1$, $\eta_2(\hat{x}) = B_2\hat{x} \leq l_2$, *and* $B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$. *For* $\hat{x}_0 \in \mathbb{R}^{|\Lambda|}_{\geq 0}$, *let* $z_{d,0}$ *be the state in the state space of* $\mathcal{Z}^{\Delta z,h,N}$ *corresponding to the region containing* $B\hat{x}_0$. *Define*

$$
\begin{aligned}
&P_{until}((\hat{x}_0, \eta_1, \eta_2, \hat{\mathcal{X}}^N, t_1, t_2)) = \\
&\qquad Pr(\exists t \in [t_1, t_2]\, s.t.\, \eta_2(\omega(t)) \wedge \forall t' \in [0, t), \eta_1(\omega(t)) \,|\, \omega \in \omega_{\hat{x}_0}), \\
&P^{\Delta z,h}_{until}((z_{d,0}, \eta_1, \eta_2, \hat{\mathcal{Z}}^{\Delta z,h,N}, t_1, t_2)) = \\
&\qquad Pr^{\Delta z,h}(\exists k \in [\lfloor \frac{t_1}{h} \rfloor, \lfloor \frac{t_2}{h} \rfloor]\, s.t.\, \eta_2(\mathcal{Z}^{\Delta z,h,N}(k)) \\
&\qquad\qquad \wedge \forall k' \in [0, k-1], \eta_1(\mathcal{Z}^{\Delta z,h,N}(k')) \,|\, \mathcal{Z}^{\Delta z,h,N}(0) = z_{d,0}).
\end{aligned}
$$

*Then, it holds that*

$$
\lim_{N\to\infty} \lim_{h\to 0} \lim_{\Delta z \to 0} |P_{until}(((\hat{x}_0, \eta_1, \eta_2, \hat{\mathcal{X}}^N, [t_1, t_2])) - P^{\Delta z,h}_{until}((z_{d,0}, \eta_1, \eta_2, \hat{\mathcal{Z}}^{\Delta z,h,N}, [t_1, t_2]))| = 0.
$$

**Example 5.3.1.** *Let us return to the CRN introduced in Example 5.1.1. We consider the following quantitative property:*

$$
P_{=?}[(Pro < 10)\, U^{[0,T]}\, (mRNA > 30)], T \in [0, 100],
$$

*which is satisfied for those paths in which the mRNA population becomes greater than* 30 *before the protein population hits* 10 *molecules. In Figure 5.2 we evaluate the property for different values of* $h$ *and fixed* $N = 100$. *Already for* $h = 5$ *the property is surprisingly close to the same measure computed on the original CTMC using uniformization techniques as implemented in PRISM [106]. Note that here the property is expressed in terms of the number of molecules, rather than concentrations. In fact, as we explained in Section 5.1, for the CSL properties considered in here the two representations are equivalent.*

## 5.4 Correctness

The method we present is approximate. In particular, errors are introduced in two ways: by resorting to the CLA and by discretisation of time and space of the CLA. The quality of these two approximations is controlled by three parameters: (a) $N$, the system size, which influences the accuracy of CLA, (b) $h$, the time step size, and (c)

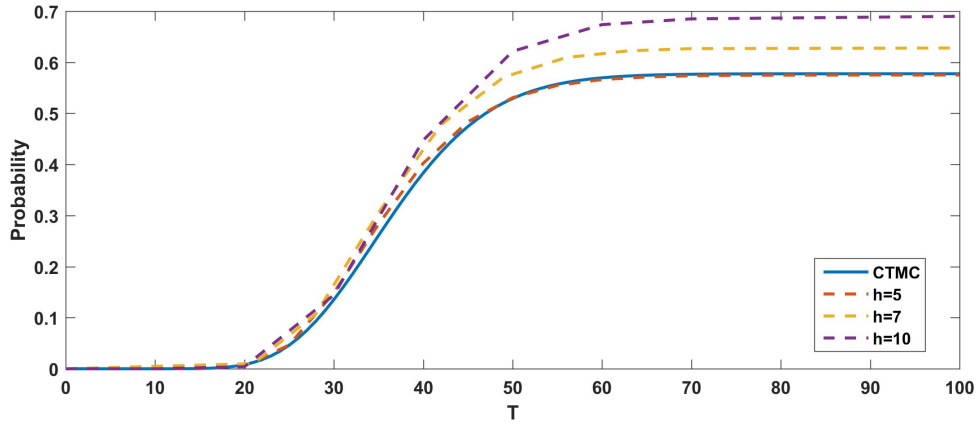



Figure 5.2: Comparison of the evaluation of $P_{[0,T]}[(Pro < 10)\, U^{[0,T]}\, (mRNA > 30)]$ on a CTMC as estimated by PRISM [106], and on the CLA approximation for $\Delta z = \frac{0.5}{N}$ and three different values of $h$.

$\Delta z$, the space discretization step, which influences the quality of the approximation of the reachability probability of the CLA. Then, we have the following result.

**Theorem 7.** *Let $\Psi$ be a CSL formula as defined in Definition 15, $\hat{x} \in \mathbb{R}^{|\Lambda|}_{\geq 0}$ and $z_{0,d}$ be the state in $\mathcal{Z}^{\Delta z,h,N}$ corresponding to the region containing $\hat{x}_0$. Then, for $N$ sufficiently large and $h$ and $\Delta z$ sufficiently small it holds that:*

$$\hat{\mathcal{X}}^N, \hat{x} \models \Psi \leftrightarrow \hat{\mathcal{Z}}^{\Delta z,h,N}, z_{d_0} \models \Psi,$$

*except for a set of thresholds of Lebesgue measure 0.*

*Proof.* The proof is by induction on the terms in Definition 15. The interesting cases are $\Psi = P_{\sim p}(F^{[t_1,t_2]}\eta)$ and $\Psi = P_{\sim p}(\eta_1\, U^{[t_1,t_2]}\, \eta_2)$. Theorem 6 guarantees that, for $N \to \infty, h \to 0, \Delta z \to 0$, the difference between the probability of the above properties computed on $\hat{\mathcal{Y}}^N$, the CLA of $\hat{\mathcal{X}}^N$, and on $\hat{\mathcal{X}}^N$ is equal to $\epsilon \to 0$. Assume $Pr(\exists t \in [t_1, t_2]\, s.t.\, \eta(\omega(t)) | \omega \in \Omega_{\hat{x}}) = q$, and consider the reachability property $P_{\sim q}(F^{[t_1,t_2]}\eta)$. In this case, no approximation algorithm can guarantee to give the right answer, as the threshold is exactly the value of the reachability property. However, the point $q$ is a set of Lebesgue measure zero with respect to the set of all possible thresholds, which is a subset of the reals. The same reasoning can be applied to the until case. □

The convergence stated in Theorem 7 means that, since $N$ is fixed for a given CRN, then, even if we have control over $h$ and $\Delta z$, the quality of the approximation depends on how well the CLA approximates the CRN. Error bounds would be a viable

companion to estimate the committed error, and although these could be extimated for time and space discretization following the approaches in [2, 109], we are not aware of any explicit formulation of them for the convergence of the CLA. However, experimental results in Section 5.6 show that the error committed is generally limited also for moderately small $N$ and quite large $h$.

### 5.4.1 Complexity

Complexity of the method depends on the following: (a) the equations we need to solve, (b) the time step size $h$, and (c) the space discretization step $\Delta z$. Algorithm 1 requires solving Eqns (5.3) and (4.7), that is, a set of differential equations quadratic in the number of species. In fact, solving these equations requires computing $J_F$, the Jacobian of $F$. However, the number of equations we need to solve is independent of the number of molecules in the system. This guarantees the scalability of our approach. An important point is that Eqn (5.3) requires solving Eqn (5.2) once for each sampling point of the numerical solution of Eqn (5.3). A possible way to avoid this is to consider approximate solutions to Eqn (5.2), which are accurate in the limit as $h \to 0$. However, to keep this approximation under control, $h$ has to be chosen really small, slowing down the computation. Moreover, for any sample point, Eqn (5.2) is solved only for a small time interval (between $t$ and $t+h$). As a consequence, in practice, the computational cost introduced in solving Eqn (5.2) is under control.

A smaller value of $h$ implies that, for a given time interval, we have a greater number of discrete-time steps, which can slow down the computation in some cases. The value of $\Delta z$ determines the number of states of the resulting DTMC. However, we stress that we discretize $\hat{\mathcal{Z}}^N(t)$, a uni-dimensional distribution (or $m$-dimensional in the case we have $m > 1$ linear inequalities). As a consequence, the number of reachable states with significant probability mass is generally limited and under control. Obviously, if the number of molecules is large and $\Delta z$ extremely small, then this is detrimental to performance.

## 5.5 Rewards

We extend CSL properties with reward operators as in [105]. As for probabilistic reachability, we will define the reward structure on the normalised process $\hat{\mathcal{X}}^N$. Formally, we define the *state reward* function $\rho : \mathbb{R}^{|\Lambda|} \to \mathbb{R}$, which associates a real-valued reward with any point of the normalised state space of $\hat{\mathcal{X}}^N(t), t \in \mathbb{R}_{\geq 0}$. In this paper, we make a few assumptions about the regularity of $\rho$:

- $\rho$ is bounded, i.e. there exists a constant $C > 0$ such that $\rho(\hat{x}) \leq C$ for each $\hat{x} \in \mathbb{R}^{|\Lambda|}$;

- $\rho$ is Lipschitz continuous on $\mathbb{R}^{|\Lambda|}$, i.e. there is a constant $L_\rho$ such that, for each $\hat{x}, \hat{x}' \in \mathbb{R}^{|\Lambda|}$, $|\rho(\hat{x}) - \rho(\hat{x}')| \leq L_\rho |\hat{x} - \hat{x}'|$.

These requirements are important to show the convergence of rewards computed on $\hat{\mathcal{X}}^N$ with the same measure but computed on the normalised CLA $\hat{\mathcal{Y}}^N$. Moreover, they do not limit the expressiveness of our framework: for a fixed $N$, we are interested only in the value of $\rho$ at a finite number of points. Such a function can always be extended to a Lipschitz continuous one. Boundedness is also not problematic, as we can always assume an upper bound on a physically meaningful population size, meaning that we can restrict ourselves to a compact set and define $\rho$ to be constant outside such a set.

Given a reward structure $\rho$, we consider the following three kinds of rewards.

- **Instantaneous Rewards** up to finite time $T$. $\rho_I(\hat{x}_0, \hat{\mathcal{X}}^N, T)$ is the expectation of $\rho(\hat{\mathcal{X}}^N(T))$, i.e., the expectation of the reward structure at time $T$ over all the trajectories of $\hat{\mathcal{X}}^N$ that start from state $\hat{x}_0$. More precisely, for $\Omega_{\hat{x}_0}$, the set of paths of $\hat{\mathcal{X}}^N$ starting from $\hat{x}_0$:

$$\rho_I(\hat{x}_0, \hat{\mathcal{X}}^N, T) = \sum_{\hat{x} \in \mathbb{R}^{|\Lambda|}} \rho(\hat{x}) Pr(\omega(T) = \hat{x} | \omega \in \Omega_{\hat{x}_0}). \tag{5.8}$$

- **Cumulative Rewards** up to a finite time $T$. Given $\omega : \mathbb{R}_{\geq 0} \to \mathbb{N}^{|\Lambda|}$, a path of $\hat{\mathcal{X}}^N$, the cumulative reward for $\omega$ is defined as:

$$\rho_C(\omega, T) = \int_0^T \rho(\omega(t))dt = \sum_{i=1}^{|jumps(\omega)|} \rho_(\omega(t_{i-1}))(t_i - t_{i-1}) + \rho_(\omega(T))(T - t_{|jumps(\omega,t)|})$$

  where $jumps(\omega, t) \subset \mathbb{R}_{\geq 0}$ is the set of time instants at which $\omega$ changes state between 0 and $T$. Then, we define:

$$\rho_C(\hat{x}_0, \hat{\mathcal{X}}^N, T) = \mathbb{E}[\rho_C(\omega, T) \,|\, \omega(0) = \Omega_{\hat{x}_0}],$$

  where the expectation is intended over the trajectories of $\hat{\mathcal{X}}^N$ starting from state $\hat{x}_0$.

- **Bounded Reachability Rewards.** Given a target set $A \in \mathbb{R}^{|\Lambda|}$, for the normalized process $\hat{\mathcal{X}}^N$, define $\rho_{reach}(\hat{x}_0, \mathcal{X}^N, A, T)$, the cumulative reward until we enter the target set $A$ within time $T$. Formally, we can define $\rho_{reach}(\hat{x}_0, \mathcal{X}^N, A, T)$ as the cumulative reward until time $T$ for the modified process $\bar{\mathcal{X}}^N$ where all states in $A$ are made absorbing, and where we consider the modified state rewards:

$$\bar{\rho}(\hat{x}) = \begin{cases} 0 & \text{if } \hat{x} \in A \\ \rho(\hat{x}) & \text{otherwise} \end{cases}$$

**Remark 1.** *Note that here $\rho$ is a state reward, that is, a function that associates a real value with any given state of the process. An alternative reward structure could be based on the* transition reward *function [38], which can be used for checking how many times a given reaction fires up to a certain time. However, in the context of CRNs, such a quantity can be easily expressed with an additional species counting how many times a subset of the transitions fire. Then, instantaneous rewards can be used to "read" its value. For instance, in Example 5.1.1, if we want to count the number of times a mRNA molecule is produced, we can consider an additional species $C$ and modify the CRN such that:*

$$\rightarrow^{0.5} mRNA + C.$$

*Then, for $\hat{x} \in \mathbb{R}^{|\Lambda|}_{\geq 0}$ we have $\rho(x) = \hat{x}_C$, where $\hat{x}_C$ is the component of vector $\hat{x}$ relative to species $C$, and $N\rho_I(\hat{x}_0, \hat{\mathcal{X}}^N, T)$ will give the desired measure at time $T$ for $\hat{x}_0$, the initial concentration of the species.*[2].

### 5.5.1 Extending CSL with Rewards

In order to incorporate rewards in our framework, given a CRN $\mathcal{C} = (\Lambda, R)$ and the induced CTMC $\mathcal{X}^N$, we extend CSL with the following formulae, whose semantics will depend on the particular reward structure $\rho$:

$$R_{\sim r}[C_\rho^{[\leq T]}] \quad | \quad R_{\sim r}[I_\rho^{=T}] \quad | \quad R_{\sim r}[F_\rho^{\leq T}\eta]$$

where $\eta : \mathbb{R}^{|\Lambda|} \rightarrow \{true, false\}$ is a convex predicate over $\hat{\mathcal{X}}^N$, $T \in \mathbb{R}_{\geq 0}$, $r \in \mathbb{R}_{\geq 0}$, and $\sim \in \{>, <\}$. For $\hat{x} \in \mathbb{R}^{|\Lambda|}_{\geq 0}$, the semantics of such formulae is as follows:

$$\hat{\mathcal{X}}^N, \hat{x} \models R_{\sim r}[C_\rho^{[\leq T]}] \quad \text{iff} \quad \rho_C(\hat{x}, \hat{\mathcal{X}}^N, T) \sim r$$

$$\hat{\mathcal{X}}^N, \hat{x} \models R_{\sim r}[I_\rho^{=T}] \quad \text{iff} \quad \rho_I(\hat{x}, \hat{\mathcal{X}}^N, T) \sim r$$

[2]The multiplication of $\rho_I$ by $N$ is needed to convert from the normalized process back to the integer population count.

$$\hat{\mathcal{X}}^N, \hat{x} \models R_{\sim r}[F_{\rho}^{\leq T}\eta] \quad \text{iff} \quad \rho_{reach}(\hat{x}, \mathcal{X}^N, A, T) \sim r,$$

where $A = \{\hat{x}' \in \mathbb{R}^{|\Lambda|} \, s.t. \, \eta(\hat{x}')\}$.

### 5.5.2 Computing Expectation and Variance Using Reward Operators

Two of the most common statistics needed when studying stochastic processes are expectation and variance (or covariance) of some set of variables. Suppose we have a CTMC $\mathcal{X}^N$ with values in $\mathbb{R}^{|\Lambda|}$, and we want to compute expectation and variance of one of its components $\mathcal{X}_i^N$ at time $t$. Then, for $\hat{x} \in \mathbb{R}^{|\Lambda|}$, we define the following reward structures on the normalised process:

$$\rho^{size}(\hat{x}) = \begin{cases} \hat{x}_i & \text{if } \hat{x}_i < K \\ K & \text{if } \hat{x}_i \geq K \end{cases} \quad \rho^{size^2}(\hat{x}) = \begin{cases} \hat{x}_i^2 & \text{if } \hat{x}_i < K \\ K^2 & \text{if } \hat{x}_i \geq K, \end{cases}$$

where $K \in \mathbb{R}$ can be any real number, typically an upper bound on the physically admissible population size. For instance, for biochemical processes, we can choose $K$ to be of the order of $10^{80}$, the estimated number of atoms of the universe. Then, we have

$$\mathbb{E}[\mathcal{X}_i^N(t)] = N R_{=?}[I^{=t}_{\rho_{size}}]$$

$$\Sigma_{\mathcal{X}_i^N(t)} = N(R_{=?}[I^{=t}_{\rho_{size^2}}] - (R_{=?}[I^{=t}_{\rho_{size}}])^2).$$

Since $T$ is finite and $K$ non-negative, the above equality holds for any CRN whose species count remains finite at least for a finite time interval. Note that, as rewards are defined for the normalised process, we need to rescale them back to population counts to compute variance and average of the non-normalised process.

### 5.5.3 Computing Rewards through Central Limit Approximation

Computing reward properties over $\hat{\mathcal{X}}^N$ is generally not possible because of the state space explosion problem. As a consequence, we compute such properties using $\hat{\mathcal{Y}}^N$, the CLA of $\hat{\mathcal{X}}^N$. We show that the values computed on $\hat{\mathcal{Y}}^N$ converge (weakly) to those computed on $\hat{\mathcal{X}}^N$.

We stress again how working in terms of the normalised processes is not a limitation. For instance, consider the reward for expectation. If we are interested in the expectation of population counts for a fixed $N$, we can either define the reward for $\hat{x}$ in the normalised space as $\rho(x) = N\hat{x}$, for $N$ fixed, or rescale the computed reward as done in the previous section.

#### 5.5.3.1 Instantaneous Rewards

Given a reward structure $\rho$, instantaneous rewards can be computed on $\hat{\mathcal{Y}}^N$ as:

$$\rho_I^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, t) = \mathbb{E}[\rho(\hat{\mathcal{Y}}^N(t))] \approx \int_K \rho(z)\mathcal{N}(z|\mathbb{E}[\hat{\mathcal{Y}}^N(t)], cov[\hat{\mathcal{Y}}^N(t)]])dz,$$

where $\mathcal{N}(z|\mathbb{E}[\hat{\mathcal{Y}}^N(t)], cov[\hat{\mathcal{Y}}^N(t)]])$ is the normal distribution with mean and covariance matrix respectively, $\mathbb{E}[\hat{\mathcal{Y}}^N(t)], cov[\hat{\mathcal{Y}}^N(t)]]$ for $\hat{\mathcal{Y}}^N(0) = \hat{x}$. Furthermore, $K \subseteq \mathbb{R}^{|\Lambda|}$ is a compact set in which we restrict integration for numerical purposes. The choice of $K$ is such that the error we commit is bounded by a chosen tolerance level. The following proposition guarantees that $\rho_I^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, t)$ converges to $\rho_I(\hat{x}, \hat{\mathcal{X}}^N, T)$.

**Proposition 4.** *Let $T \in \mathbb{R}_{\geq 0}$, then it holds that:*

$$\lim_{N\to\infty} \rho_I(\hat{x}, \hat{\mathcal{X}}^N, T) = \lim_{N\to\infty} \rho_I^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, t).$$

*Proof.* We want to prove that, for fixed $T > 0$, $\mathbb{E}[\rho(\hat{\mathcal{X}}^N)]$ converges to $\mathbb{E}[\rho(\hat{\mathcal{Y}}^N(T))]$ as $N$ tends to infinity. Using the triangular inequality, it holds that:

$$\begin{aligned} |\mathbb{E}[\rho(\hat{\mathcal{X}}^N(T))] &- \mathbb{E}[\rho(\hat{\mathcal{Y}}^N(T))]| \leq \\ |\mathbb{E}[\rho(\hat{\mathcal{X}}^N))] - \mathbb{E}[\rho(\Phi(T))]| &+ |\mathbb{E}[\rho(\Phi(T))] - \mathbb{E}[\rho(\hat{\mathcal{Y}}^N(T))]|, \end{aligned}$$

where $\rho(\Phi(T))$ is the reward $\rho$ evaluated on the fluid limit $\Phi(T)$, solution of the rate equations. Invoking the fluid approximation theorem [39], it holds that $\hat{\mathcal{X}}^N(T) \Rightarrow_{N\to\infty} \Phi(T)$ (as convergence in probability implies weak convergence). Furthermore, by Theorem 2

$$\hat{\mathcal{Y}}^N(T) = \frac{G(T)}{\sqrt{N}} + \Phi(T) \Rightarrow_{N\to\infty} \Phi(T),$$

as $G$ is independent of $N$ and it has a bounded covariance matrix for each $T$ (which implies convergence in probability). Therefore, recalling that $\rho$ is bounded and continuous by assumption, both terms on the right hand side of the triangular inequality converge to zero by virtue of the Portmanteau theorem [31] stating that, for any continuous and bounded functional $f$ on $\mathcal{D}$, the space of $\mathbb{R}^{|\Lambda|}-$valued Cadlag functions (i.e. right continuous functions with left limit) [31], it holds that $\mathbb{E}[f(\mathcal{X}^N)] \to_{N\to\infty} \mathbb{E}[f(X)]$ whenever $\mathcal{X}^N \Rightarrow X$. Thus, we can conclude:

$$\rho_I(\hat{x}, \hat{\mathcal{X}}^N, T) \to_{N\to\infty} \rho_I^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, T).$$

□

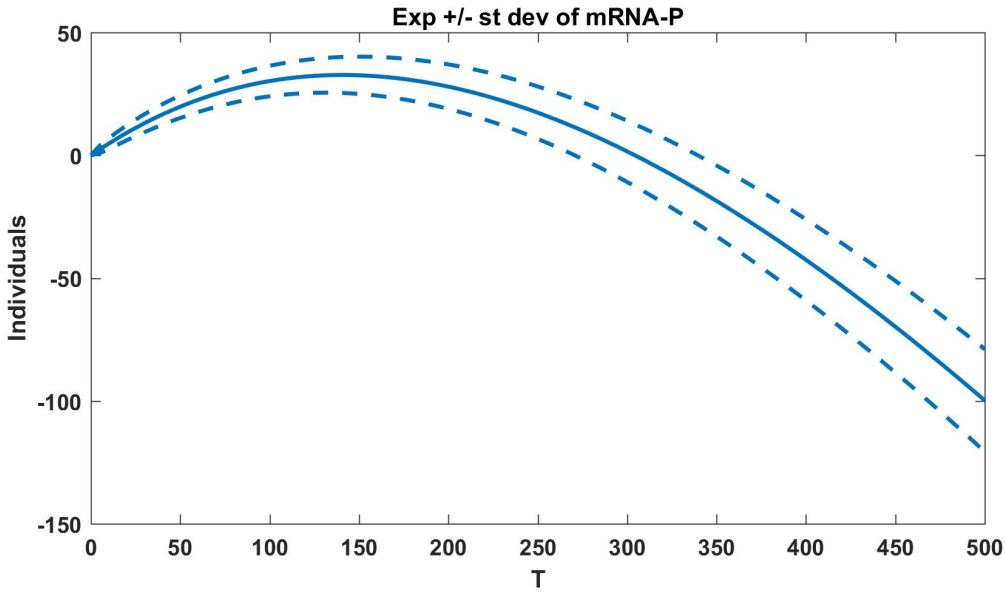



Figure 5.3: $N\rho_I(\hat{\mathcal{X}}^N, T), T \in [0, 100]$ for reward structure $\rho^{mRNA-P}$.

**Example 5.5.1.** *We consider the CRN introduced in Example 5.1.1. We are interested in knowing the expectation and variance of $mRNA - P$ over time. This can be computed using the following reward structures:*

$$\rho^{mRNA-P}(\hat{x}) = \begin{cases} \hat{x}(mRNA) - \hat{x}(P) & \text{if } \hat{x}(mRNA) - \hat{x}(P) < 10^{80} \\ 10^{80} & \text{otherwise} \end{cases}$$

$$\rho^{(mRNA-P)^2}(\hat{x}) = \begin{cases} (\hat{x}(mRNA) - \hat{x}(P))^2 & \text{if } (\hat{x}(mRNA) - \hat{x}(P))^2 < 10^{80} \\ 10^{80} & \text{otherwise} \end{cases}.$$

*Then, we have:*

$$\mathbb{E}[\mathcal{X}^N_{mRNA}(t) - \mathcal{X}^N_P(t)] = NR_{=?}[I^{=t}_{\rho^{mRNA-P}}], \quad t \in [0, 100],$$

$$\Sigma_{(\mathcal{X}^N_{mRNA}(t) - \mathcal{X}^N_P(t))} = N(R_{=?}[I^{=t}_{\rho^{(mRNA-P)^2}}] - (R_{=?}[I^{=t}_{\rho_size}])^2), \quad t \in [0, 100],$$

*where the rewards are computed on the normalised process $\hat{\mathcal{X}}^N$. The resulting expectation and variance is plotted in Figure 5.3. Note that, in this case, the resulting variance and expectation, as estimated by the CLA, are guaranteed to be exact for any $N$. This is because the CRN is linear [85]. It is easy to observe that our algorithms are exponentially faster than the computation of the same measures on the original CTMC, because of the continuous nature of the CLA.*

#### 5.5.3.2 Cumulative Rewards

Cumulative rewards can also be computed exploiting $\hat{\mathcal{Y}}^N$, the CLA approximation of $\hat{\mathcal{X}}^N$, as shown in the following proposition

**Proposition 5.** *Let $T \in \mathbb{R}_{\geq 0}$. Then, $\rho_C^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, T)$, the cumulative reward for $\hat{\mathcal{Y}}^N$ starting from $\hat{\mathcal{Y}}^N = \hat{x}$, can be computed as follows.*

$$\rho_C^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, T) = \int_0^T \rho_I^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, s)ds.$$

*Proof.* Let $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}^{|\Lambda|}$ be a path of $\hat{\mathcal{Y}}^N$. Then, we have that

$$\rho_C^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, T) = \mathbb{E}\left[\rho_C(\omega, T) | \omega(0) = \hat{x}\right] = \mathbb{E}\left[\int_0^T \rho(\omega(t))dt \,|\, \omega(0) = \hat{x}\right].$$

Now, in order to conclude, we can apply Fubini's theorem [143], which allows us to compute a double integral using iterated integrals, thus, switching the order of integration. A sufficient condition for application of Fubini's theorem is that $\mathbb{E}[\int_0^T |\rho(\omega(t))|dt]$ is finite. Owing to boundedness of $\rho$, there is an $M > 0$ such that, for all $x \in \mathbb{R}^{|\Lambda|}$, we have that $|\rho(x)| \leq M$. Thus, we have,

$$\mathbb{E}\left[\int_0^T |\rho(\omega(t))|dt\right] \leq \mathbb{E}\left[\int_0^T M dt\right] = MT,$$

which is finite as $T$ and $M$ are both finite. □

The following proposition, for $\hat{x} \in \mathbb{R}^{|\Lambda|}$, guarantees that $\rho_C^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, T)$ converges to $\rho_C(\hat{x}, \hat{\mathcal{X}}^N, T)$.

**Proposition 6.** *Let $T \in \mathbb{R}_{\geq 0}$, then it holds that*

$$\lim_{N\to\infty} \rho_C(\hat{x}, \hat{\mathcal{X}}^N, T) = \lim_{N\to\infty} \rho_C^{CLA}(\hat{x}, \hat{\mathcal{Y}}^N, T).$$

*Proof.* For a path $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}^{|\Lambda|}$, define the following functional $\mathcal{R}_C(T, \omega) = \int_0^T \rho(\omega(t))dt$, which is defined on $\mathcal{D}$, the space of $\mathbb{R}^{|\Lambda|}-$valued Cadlag functions (i.e. right continuous functions with left limit) [31]. $\rho_C(\hat{x}, \hat{\mathcal{X}}^N, T) = \mathbb{E}[\mathcal{R}_C(T, \omega)]$, where the expectation is taken over $\Omega_{\hat{x}}$, the paths of $\hat{\mathcal{X}}^N$ starting from $\hat{x}$. As $T$ and $\rho$ are bounded, for each $\omega$, $\mathcal{R}_C(T, \omega)$ is bounded. It is also continuous, due to the continuity of $\rho$. Thus, we can apply same reasoning as in the proof of Proposition 4, applying Portmanteau theorem to conclude. □

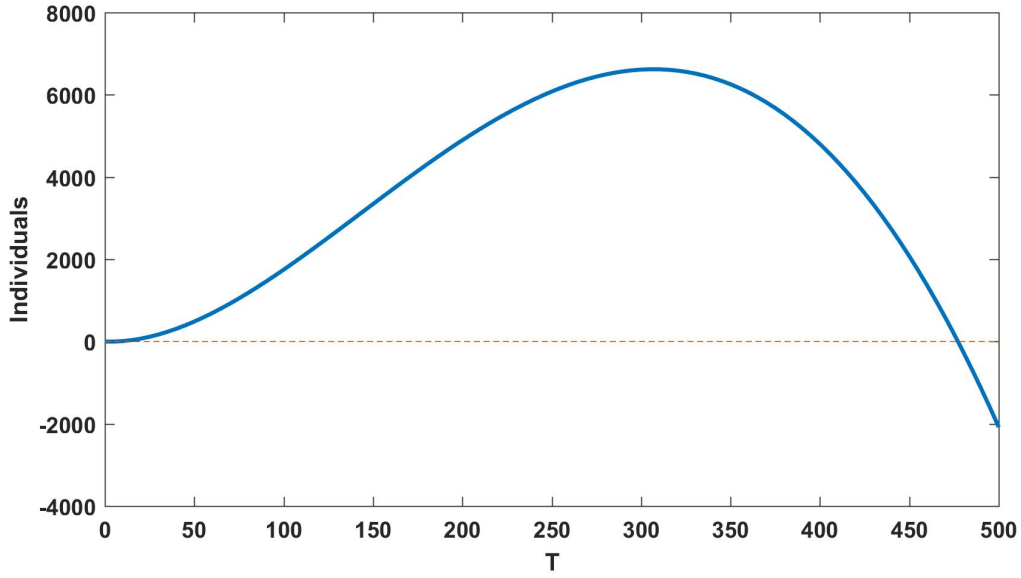



Figure 5.4: We plot $N\rho_C(\hat{\mathcal{X}}^N, T), T \in [0, 500]$, for reward structure $\rho^{mRNA-P}$.

**Example 5.5.2.** *We consider the CRN introduced in Example 5.1.1. We are interested in knowing the expected cumulative reward of $mRNA-P$ to understand if during the time interval there have been, on average, more $mRNA$ or more $P$ molecules in the system. This can be computed using the reward structure $\rho^{mRNA-P}$ introduced in Example 5.5.1, and the following cumulative reward:*

$$NR_{=?}[C^{\leq T}_{\rho^{mRNA-P}}],\ T \in [0, 500],$$

*where $R_{=?}[C^{\leq T}_{\rho^{mRNA-P}}]$ is intended to be computed on $\hat{\mathcal{X}}^N$. The resulting expectation and variance are plotted in Figure 5.4. We stress again how in this case, since the CRN is linear, the measure estimated by the CLA is exact for any $N$.*

#### 5.5.3.3 Bounded Reachability Rewards

Bounded reachability rewards can be computed efficiently on the CLA under a further assumption on the reward structure $\rho$. Specifically, for $\hat{x} \in \mathbb{R}^{|\Lambda|}_{\geq 0}$, consider the predicate $\eta(\hat{x}) = B\hat{x} \leq b$, $b \in (\mathbb{R} \cup \{\infty\})^m$, $m > 0$. We assume that the reward structure is defined on the projection of $\hat{\mathcal{X}}^N$ spanned by the colums of the matrix defining the $\eta$ predicate, namely $\rho : \mathbb{R}^m \to \mathbb{R}$ assigns a reward to each state of $B\hat{\mathcal{X}}^N$. Consider the CSL reward property $R_{\sim r}[F^{\leq T}_{\rho}\eta]$. That is, $R_{\sim r}[F^{\leq T}_{\rho}\eta]$ is the bounded reachability reward until reaching a state in $A = \{\hat{x} \in \mathbb{R}^{|\Lambda|}\ s.t.\ \forall i \in \{1, .., k\}, (B\hat{x})_i \geq b_i\}$ during the time interval $[0, T]$. Such a reward can be computed by exploring the approximation of the CLA in terms of the DTMC $\hat{\mathcal{Z}}^{\Delta z,h,N}$, which is obtained by time and

space discretization of the process $\hat{\mathcal{Z}}^N = B\hat{\mathcal{X}}^N$. We call $\rho_{reach}(\hat{x}_0, \hat{\mathcal{Z}}^{\Delta z,h,N}, \lfloor \frac{T}{h} \rfloor, A)$ the bounded reachability reward computed on $\hat{\mathcal{Z}}^{\Delta z,h,N}$ for a number of discrete steps $\lfloor \frac{T}{h} \rfloor$, where $h > 0$. Then $\rho_{reach}(\hat{x}_0, \hat{\mathcal{Z}}^{\Delta z,h,N}, \lfloor \frac{T}{h} \rfloor, A)$ can be computed by considering the modified process $\hat{\mathcal{Z}}^{\Delta z,h,N}$ where the target states are made absorbing, and modifying the reward structure $\rho$ to $\bar{\rho}$ such that $\bar{\rho}(\hat{x}) = 0$ for all absorbing states. Then, for $\hat{x}_0 \in \mathbb{R}^{|\Lambda|}$ and $z_{d,0}$, the state in the state space if $\hat{\mathcal{Z}}^{\Delta z,h,N}$ corresponding to the region containing $\hat{x}_0$, cumulative rewards can be computed with the following algorithm for $n > 0$:

$$\rho_{reach}(\hat{x}_0, \hat{\mathcal{Z}}^{\Delta z,h,N}, n, A) = \sum_{z \in S^{\Delta z}} \bar{\rho}(z) Prob(\hat{\mathcal{Z}}^{\Delta z,h,N}(n-1) = z | \hat{\mathcal{Z}}^{\Delta z,h,N}(0) = z_{d,0}) h + \rho_{reach}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \hat{x}_0, n-1, A) \tag{5.9}$$

and such that $\rho_{reach}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \hat{x}_0, 0, A) = 0$

**Proposition 7.** *For $T \in \mathbb{R}_{\geq 0}$ and $B \in \mathbb{R}^{m \times |\Lambda| \times k}$ let $A$ be the set defined as $A = \{x \in \mathbb{R}^{|\Lambda|} \, s.t. \, \forall i \in \{1, .., m\}, (Bx)_i \geq b_i\}$. Then, for $\hat{x}_0 \in \mathbb{R}^{|\Lambda|}$ and $z_{d,0}$, the state in the state space of $\hat{\mathcal{Z}}^{\Delta z,h,N}$ corresponding to the region containing $\hat{x}_0$, it holds that:*

$$\lim_{N \to \infty} \lim_{h \to 0} \lim_{\Delta z \to 0} |\rho_{reach}(\hat{x}_0, \hat{\mathcal{X}}^N, T, A) - \rho_{reach}(z_{d,0}, \hat{\mathcal{Z}}^{\Delta z,h,N}, \lfloor \frac{T}{h} \rfloor, A)| = 0.$$

Proof of Proposition 7 can be found in the Chapter 8. Intuitively, we use Theorem 4 and Theorem 5 to show the convergence of the numerical algorithms in the limit of $h \to 0$ and $\Delta z \to 0$. Then, we use the convergence of the $\mathcal{X}^N$ to its CLA for increasing $N$ to conclude.

**Example 5.5.3.** *We consider the CRN introduced in Example 5.1.1. We are interested in knowing the expected cumulative reward of $mRNA - P$ for all those paths for which the $mRNA$ does not reach 30 individuals within $[0, T]$ for $T \in [0, 50]$. We consider the reward structure $\rho^{mRNA-P}(x)$ introduced in Example 5.5.1, and the following cumulative reward $\rho_{reach}(\mathcal{X}^N, T, mRNA < 30) = N\rho_{reach}(\hat{\mathcal{X}}^N, T, mRNA < \frac{30}{N}), T \in [0, 50]$. To compute such a reward we explore the CLA approximation of $\mathcal{X}^N$. We consider $B_1 = [1, 0]$, $B_2 = [-1, 1]$ and $B = [B_1, B_2]$, where we assume the first component of $\mathcal{X}^N$ represents the number of protein molecules in that state. Then, we consider $\hat{\mathcal{Z}}^N$, the projection of the CLA of $\hat{\mathcal{X}}^N$ over B, namely, $\hat{\mathcal{Z}}^N = B\hat{\mathcal{Y}}^N$. At this point we discretize $\hat{\mathcal{Z}}^N$ with sampling time $h > 0$ and a grid of cells of size $dz > 0$, and compute the above rewards using Eqn (5.9). The solution to Eqn (5.9) is approximate, and errors are introduced by two factors: firstly, by the usage of the CLA*

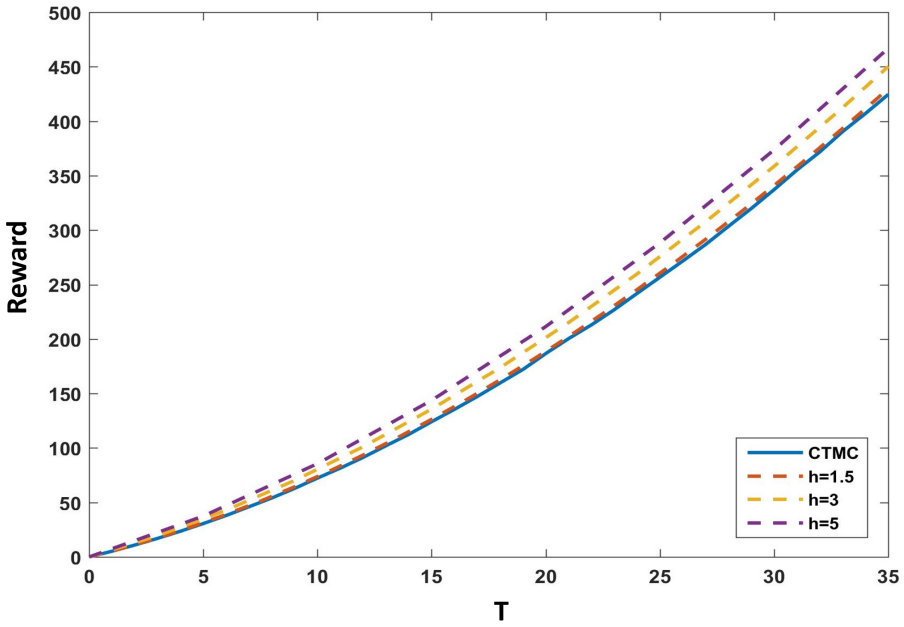



Figure 5.5: We plot the bounded reachability reward $\rho_{reach}(x_0, T), T \in [0, 35]$, for reward structure $\rho^{mRNA-P}$. We compare the CLA approximation for different values of $h$ and $dz = 0.5$ with the same measure estimated on the original CTMC using 10000 simulations.

*approximation of* $\mathcal{X}^N$*, and, secondly, by the discretization of the resulting Gaussian process. Thus, we compare our reward value with the value computed on* $\mathcal{X}^N$ *using* 10000 *simulations for each time point. The resulting values are plotted in Figure 5.5. To perform a further comparison we employ the following metrics,* $\epsilon^{rel}_{max}$ *and* $\epsilon^{rel}_{avg}$*, defined as follows:*

$$\epsilon^{rel}_{max} = max_{T\in\Sigma_h} \frac{|Rew^{CLA}_T - Rew_T)|}{|Rew|}, \quad \epsilon^{rel}_{avg} = \sum_{T\in\Sigma_h} \frac{|Rew^{CLA}_T - Rew_T)|}{|Rew|} \frac{1}{|\Sigma_h|}$$

*where* $\Sigma_h$ *is the set of sampling points for sampling time* $h$*,* $Rew^{CLA}_T = \rho^{mRNA-P}_{reach}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \hat{x}_0, \lfloor \frac{T}{h} \rfloor, A)$ *and* $Rew_T = \rho^{mRNA-P}_{reach}(\hat{X}, \hat{x}_0, T, A)$*. The calculated metrics are summarised in the table below for three different values of* $h$*.*

| $h$ | $\epsilon^{rel}_{avg}$ | $\epsilon^{rel}_{max}$ |
|---|---|---|
| 5 | 1.5468 | 7.96 |
| 3 | 0.196 | 0.88 |
| 1.5 | 0.041 | 0.24 |

*It is possible to observe how the two measures converge very fast. In fact, already for* $h = 1.5$*, which corresponds to* 25 *sampling times, the two measures have an average relative error of less than* 0.041*.*

## 5.6 Experimental Results

We implemented our algorithms in Matlab and evaluated them on two case studies. All the experiments were run on an Intel Dual Core i7 machine with 8 GB of RAM. The first case study is a Gene Expression Network as introduced in Example 5.1.1. We use this example to demonstrate that our approach is more powerful than existing approximate techniques. Specifically, we show how our CLA approach, based on a Gaussian process approximation, is able to correctly evaluate properties that methods based on Fluid Limit Approximation [37] cannot, while still guaranteeing comparable scalability. The second example is a Phospohorelay Network with 7 species. We use this example to show the trade-off between the different parameters and the molecular population. More precisely, we show that the accuracy of our approach increases as the number of molecules grows, but can still give fast and accurate results when the molecular population is relatively small. We validate our results by comparing our method with statistical model checking (SMC) as implemented in PRISM [106]. In fact, for both examples, exact numerical computation of the reachability probabilities using uniformisation techniques on the induced CTMC is not feasible because of state space explosion.

### 5.6.1 Gene Expression

We consider the following gene expression model, as introduced in Example 5.1.1 with initial counts of all the species equal to 0. We consider the property $P_{=?}(F^{[0,Time]}(mRNA \geq 175)$, which quantifies the probability that at least 175 $mRNA$ molecules are produced during the first $Time$ seconds, for $Time \in [0, 1000]$. This is a particularly difficult property because the trajectory of the mean-field of the model, and so the expected value of the CLA, does not enter the target region. As a consequence, approximate approaches introduced in [69] and [41], which are based on the hitting times of the mean-field model, fail and evaluate the probability as always equal to 0.

Conversely, our approach is able to correctly evaluate such a property. Figure 5.6 compares the value computed by our method with statistical model checking of the same property as implemented in PRISM over 30000 simulations for each time point and confidence interval 0.01. In Figure 5.6 we consider $h = 1.8$ and $\Delta z = 0.5$ and demonstrate that our approach is able to correctly estimate such a difficult property. Note that, as the mean-field does not enter the target region, for each time point the probability to enter the target region depends on a portion of the tail of the Gaussian given by the CLA. As a consequence, the accuracy of our results strictly depends on

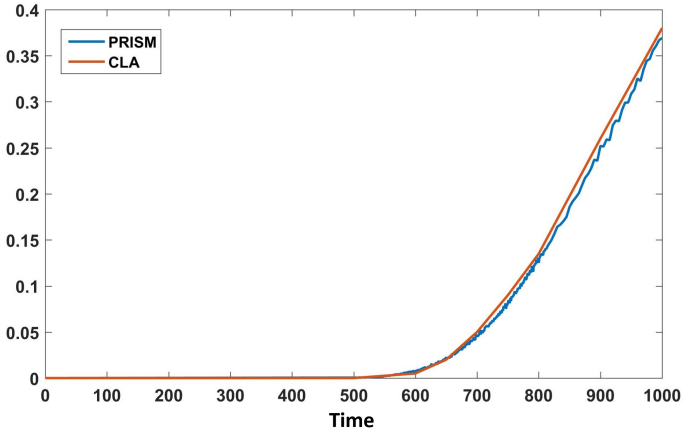



Figure 5.6: The figure plots $P_{=?}(F^{[0,Time]}\ (mRNA \geq 174)$ for $h = 1.85$ and $\Delta z = 0.5$. The x-axis represents the value of $Time$ and the y-axis the quantitative value of the formula for that value of $Time$.

how well the CLA approximates the original CTMC, much more than for properties where the mean-field enters the target region. In the following table, we evaluate our results for two different values of $h$ and $\Delta z = 0.5$. In order to compare the accuracy we consider the following metrics as defined in Example 5.5.3, $\epsilon^{rel}_{avg}$ and $\epsilon^{rel}_{max}$. The comparison is shown in the table below.

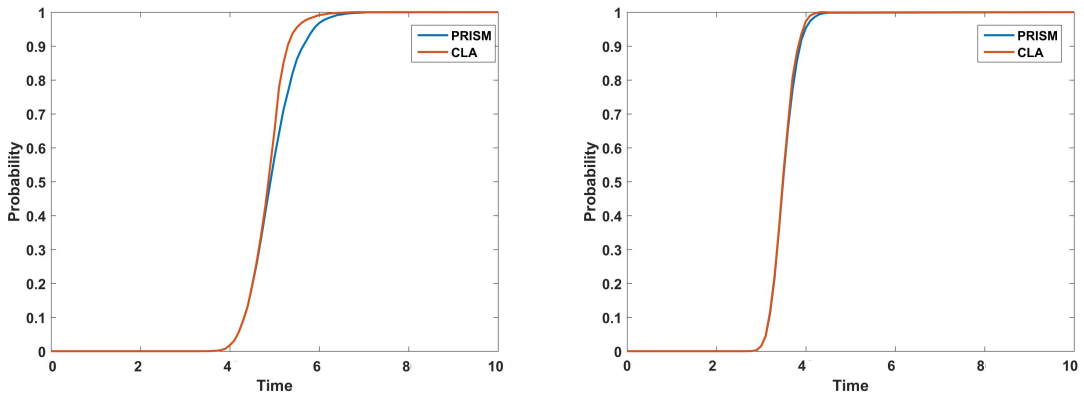
| Property | Ex. Time | h | $\epsilon^{rel}_{avg}$ | $\epsilon^{max}_{avg}$ |
|---|---|---|---|---|
| $P_{=?}(F^{[0,Time]}\ (mRNA \geq 174))$, $Time \in [0, 100]$ | 298 sec | 1.85 | 0.0075 | 0.022 |
| $P_{=?}(F^{[0,Time]}\ (mRNA \geq 174))$, $Time \in [0, 100]$ | 152 sec | 5 | 0.0147 | 0.13 |

### 5.6.2 Phosphorelay Network

We now consider a three-layer phosphorelay network consisting of 7 species given by the following reactions:

$$L1 + B \to^{\frac{0.01}{N} \cdot L1 \cdot B} B + L1p; \quad L2 + L1p \to^{\frac{0.01}{N} \cdot L2 \cdot L1p} L1 + L2p;$$
$$L3 + L2p \to^{\frac{0.01}{N} \cdot L3 \cdot L2p} L3p + L2; \quad L3p \to^{0.1 \cdot L3p} L3;$$
$$L2p \to^{0.01 \cdot L2p} L2; \quad L1p \to^{0.01 \cdot L1p} L1.$$

There are 3 layers, $(L1, L2, L3)$, which can be found in phosphorylate form $(L1p, L2p, L3p)$, and the ligand $B$. We consider the initial condition $x_0 \in \mathbb{N}^7$ such that $x_0(L1) = x_0(L2) = x_0(L3) = 0.25N$, $x_0(L1p) = x_0(L2p) = x_0(L3p) = 0$ and $x_0(B) = 0.15N$. In Figure 5.7, we compare the estimates obtained by our approach for two different




(a) (b)

Figure 5.7: Comparison of the evaluation of $P_{=?}(F^{[0,Time]}\,(L3p > 80))$ (a) with $N = 400$ and $P_{=?}(F^{[0,Time]}\,(L3p > 180))$ (b) with $N = 800$ using statistical model checking as implemented in PRISM and our approach. In both figures we considered $h = 0.1$, $\Delta z = 0.5$.

initial conditions ($N = 400$ and $N = 800$) with statistical model checking as implemented in PRISM [106], with 30000 simulations and confidence interval equal to 0.01. In both experiments we set $\Delta z = 0.5$.

In Figure 5.7a we can see that, if we increase the time interval of interest, the error tends to increase. This is because, for $N = 400$, the CLA and CME do not have perfect convergence. As a consequence, at every step of the discretized DTMC, a small error is introduced. This source of error is present until we enter the target region with probability 1. If we increase $N$ the error disappears, and the inaccuracies are due to the finiteness of $h$ and $\Delta z$. However, already for $h = 0.1$ and $N = 800$, the CLA produces a fast and reasonably accurate approximation. In the following table we compare our approach and PRISM evaluations for different values of $N$ and $h$ and $\Delta z = \frac{0.5}{N}$ in the normalised space, which implies the resulting discrete state process takes values in $\mathbb{Z}$.

| Property | Ex. Time | h | N | $\epsilon_{avg}^{rel}$ | $\epsilon_{avg}^{max}$ |
|---|---|---|---|---|---|
| $P_{=?}(F^{[0,Time]}\,(L3p > 80))$, $Time \in [0, 10]$ | 97 sec | 0.1 | 400 | 0.0088 | 0.11 |
| $P_{=?}(F^{[0,Time]}\,(L3p > 180))$, $Time \in [0, 10]$ | 130 sec | 0.1 | 800 | 0.0015 | 0.0217 |
| $P_{=?}(F^{[0,Time]}\,(L3p > 80))$, $Time \in [0, 10]$ | 28 sec | 0.5 | 400 | 0.0381 | 0.24 |
| $P_{=?}(F^{[0,Time]}\,(L3p > 180))$, $Time \in [0, 10]$ | 39 sec | 0.5 | 800 | 0.0289 | 0.14 |

The results show that the best accuracy is obtained for $h = 0.1$ and $N = 800$, where $h = 0.1$ induces a finer time discretization, whereas the worst are for $h = 0.5$ and $N = 400$. We comment that the numerical solution of the CME using PRISM is

not feasible for this model, and our method is several orders of magnitude faster than statistical model checking with PRISM (30000 simulations for each time point), which takes several hours for each property.

## 5.7 Further Related Works on CSL Model Checking

Algorithms for model checking CSL properties for continuous-time Markov chains have been introduced and then improved with techniques based on uniformization [16] (essentially a discretisation of the original CTMC), and reward computation [105]. The analysis typically involves computing the transient probability of the system residing in a state at a given time, or, for a model annotated with rewards, the expected reward that can be obtained. Despite improvements such as symmetry reduction [88], sliding window [167] and fast adaptive uniformisation [62], their practical use for chemical reaction networks is severely hindered by state space explosion [88], which in a CRN grows exponentially with the number of molecules when finite, and may be infinite, in which case finite projection methods have to be used [123]. As a consequence, approximate but faster algorithms are appealing. The mainstream solution is to rely on simulations combined with statistical inference to obtain estimates [106, 43]. These methods, however, are still computationally expensive. A recent trend of works explored as an alternative whether estimates could be obtained by relying on approximations of the stochastic process based on mean-field [37] or linear noise [41, 40, 52]. However, CSL and reward properties, like those considered here, are very challenging. In fact, most approaches consider either local properties of individual molecules [37], or properties obtained by observing the behaviour of individual molecules and restricting the target region to an absorbing subspace of the (modified) model [40]. The only approach dealing with more general subsets, [41], imposes restrictions on the behaviour of the mean-field approximation, whose trajectory has to enter the reachability region in a finite time. Our approach, as shown in the case study in Section 5.6.1, does not suffer from this problem. Another interesting approach has been developed in [145, 121], where model checking of time-bounded properties for CTMCs is expressed as a Bayesian inference problem, and approximated model checking algorithms are derived. However, no guarantees on the convergence of the resulting algorithms are given.

## 5.8 Summary

We presented a framework for approximate model checking of a time-bounded fragment of CSL extended with rewards for CTMCs that are induced from chemical reaction networks. Our approach employs an abstraction based on the Central Limit Approximation to approximate the CTMC as a Gaussian process. Then, numerical procedures for model checking CSL formulae on the resulting Gaussian process are derived. We demonstrate that our approach does not suffer from the state space explosion problem, thus enabling formal analysis of CTMCs that cannot be analysed using classical methods based on uniformization and with infinite state space [62, 167]. Deriving model checking algorithms was challenging because the CLA yields a continuous-time stochastic process with an uncountable state space. As a consequence, existing methods that rely on finite state spaces cannot be used directly and discretizing the uncountable state space induced by the CLA naturally leads to state space explosion. In order to overcome these issues, we considered reachability regions defined as polytopes. Using the fact that the CLA is a Gaussian Markov process, for a given linear combination of the species of a CRN we are able to project the original, multi-dimensional Gaussian process onto a uni-dimensional stochastic process. We then derived an abstraction in terms of a time-inhomogeneous DTMC, whose state space is independent of the number of the species of a CRN, as it is derived by discretizing linear combinations of the species. This ensures scalability. On different case studies, we showed that our approach outperforms existing methods and permits fast and scalable probabilistic analysis of CRNs. The accuracy depends on parameters controlling space and time discretization, as well as on the accuracy of the CLA. Using the theory of convergence in distribution we proved the convergence of our algorithms in the limit of high populations. A viable companion to estimate the committed error would be error bounds, but we are not aware of any explicit formulation of them for the convergence of the CLA.

As discussed in Section 2.2.1, one of the limitations of CLA, and thus of the model checking routines presented in this chapter, is that multi-modal dynamics cannot be represented. This may be limiting for applications where low number of molecules are involved. Nevertheless, when small counts are involved, an exact analysis based on the CTMC may be possible due to the fact that the dimension of the state space would be under control. Moreover, even in these limiting scenarios, our approach based on the CLA can still be useful to obtain a fast stochastic characterization of

the system, which will then require further validation, for instance via stochastic simulations.

# Chapter 6

# Molecular Filters for Noise Reduction

In this chapter we study how CRNs propagate and attenuate noise. We are interested in identifying fundamental and general mechanisms that, at the molecular level, could reduce noise and in understanding their properties and limitations. This would provide new insights into how biochemical networks control noise and illuminate why some reoccurring molecular network modules may have been selected by evolution.

In Section 6.1 we introduce and formalize the concept of a *molecular filter*, as a CRN that reduces the noise of a given biochemical signal while still maintaining certain features of its time evolution, and overview the results of this chapter. In Section 6.2 we discuss linear filters (molecular filters composed by at most uni-molecular reactions), and show their properties and limitations. Then, in Section 6.3 and 6.4 we discuss two non-linear filters, the *annihilation module* and the *annihilation filter*, and show that non-linear dynamics, such as complex formation, are necessary for efficient noise reduction. Finally, in Section 6.6, we discuss how the proposed filters are widespread in natural systems, and could give new insights into the mechanisms of gene expression.

The techniques presented in Chapter 5 are central for the derivation of the results presented in this chapter. In fact, as illustrated in Figure 6.2, the design of molecular filters relies on effective tools and automated model checking techniques for model selection of CRNs with an appropriate set of species and reactions. As a result, without a usable tool such as VisualGEC [108], which allowed us to compare different CRNs against complex reward properties in few seconds and while taking into account their stochastic behaviour, selecting the right morphology for our filters would have been a much more difficult and time consuming task.

**Notation.** In this chapter we always consider mass action kinetics, assume system size $N = 1$, and work in terms of molecular numbers. Thus, we consider the following shortcut when writing reactions: for bi-molecular reactions when we write $\tau : \lambda_1 + \lambda_2 \to^k ...$, for $\lambda_1 \neq \lambda_2$, we assume a propensity rate $\alpha_\tau(x) = kx_{\lambda_1}x_{\lambda_2}$, while for $\lambda_1 = \lambda_2$ we have $\alpha_\tau(x) = k\frac{x_{\lambda_1}(x_{\lambda_1}-1)}{2}$. Similarly, for uni-molecular reactions, for $\tau : \lambda \to^k ...$, we have $\alpha_\tau(x) = kx_{\lambda_1}$. Instead, for $\tau :\to^k ...$, we have $\alpha_\tau(x) = k$.

## 6.1 Molecular Filters

To model biological networks we focus on the underlying molecular interactions represented as Chemical Reaction Networks (CRNs). Given a CRN, we assume there are two species, whose time evolution represents input and output signals. In this thesis, species **A** will always represent the input signal/species, while species **O** the output. We assume **A** is a noisy input, with the noise identified by its Fano factor, $F_{\mathbf{A}}$. Thus, a *molecular filter* is a CRN with input **A** and whose output has a reduced Fano factor compared to **A**, but still maintains certain features of its time evolution. In this thesis, we focus on filters that maintain the same long term behaviour of **A**. Formally, we define molecular filters as follows

**Definition 16.** *Let $\mathcal{C} = (\Lambda, R)$ and $\mathcal{F} = (\Lambda', R')$ be two CRNs such that $\Lambda \cap \Lambda' = \{\mathbf{A}\} \cup \Lambda''$, where $\Lambda''$ is a possibly empty set of species. Assume $\{\mathbf{A}, \mathbf{O}\} \subseteq \Lambda'$ and that, for any $\tau \in R'$, $\upsilon_{\tau,\mathbf{A}} = 0$. That is, **A** is not changed by reactions in $\mathcal{F}$. Then, $\mathcal{F}$ is called a molecular filter for $\mathcal{C} \cup \mathcal{F}$, if, at steady state, **O** has a smaller Fano factor compared to **A** and they have the same mean. That is,*

$$F^\infty_{\mathbf{O}} < F^\infty_{\mathbf{A}}, \qquad \mathbb{E}^\infty_{\mathbf{O}} = \mathbb{E}^\infty_{\mathbf{A}}. \tag{6.1}$$

In this chapter, we propose three different molecular filters, shown in Figure 6.1. First, we consider *linear* filters (Figure 6.1*A*), implemented by means of at most uni-molecular reactions, and we show that, in case of positive correlation between the elements of the network, they are limited by Poisson levels, that is, the variance of the output signal is lower bounded by its mean ($F^\infty_{\mathbf{O}} \geq 1$). We then show how the presence of feedback loops may improve the performance of linear filters below Poisson levels. The limitations in the noise suppression capabilities of linear filters motivate us to consider *non-linear* filters. Therefore, we introduce a *non-linear* filter mechanism given by the co-expression of two species that then bind together, called the *annihilation module* (Figure 6.1*B*), which we demonstrate is able to reduce the

noise to below Poisson levels. We then propose the *annihilation filter* (Figure 6.1$C$), which combines the properties of the linear filter and annihilation module. We show how the annihilation filter in particular can greatly reduce molecular noise. We observe that co-expression and non-linear degradation are key requirements for such noise reduction.

## 6.2 Linear Filters

We start this section with an illustrative example to show the noise reduction capabilities of linear filters.

**Example 6.2.1.** *Consider an input signal given by the following CRN sketch*

$$\rightarrow^{k_p} \mathbf{A} + \mathbf{A}; \qquad \mathbf{A} \rightarrow^{k_d}, \tag{6.2}$$

*where $k_p, k_d$ are rate variables with values in $\mathbb{R}_{\geq 0}$. This may model a noisy signal $\mathbf{A}$ from the environment, with production rate $k_p$ and degradation rate $k_d$. The expectation and variance of $\mathbf{A}$ at steady state can be derived by solving Eqn* (4.3) *at steady state, and we get*

$$\mathbb{E}^{\infty}_{\mathbf{A}} = 2\frac{k_p}{k_d}; \quad \mathbb{V}^{\infty}_{\mathbf{A}} = 3\frac{k_p}{k_d}.$$

*Note that the fact that expectation and variance stops changing does not imply that a stochastic realization of $\mathbf{A}$ will eventually stop changing. We ask if it is possible to somehow generate a noiseless version of $\mathbf{A}$, that is, a copy of the $\mathbf{A}$ signal preserving the mean but with reduced variance. In order to answer this question, consider the following CRN, which we call $\mathcal{F}_L$*

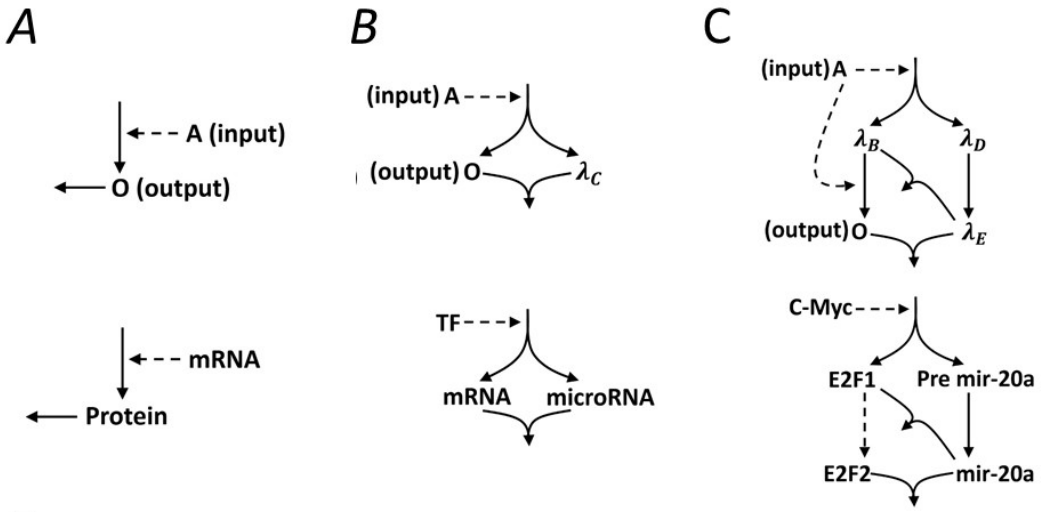
$$\mathcal{F}_L : \quad \mathbf{A} \rightarrow^{k_1} \mathbf{A} + \mathbf{O} \quad \mathbf{O} \rightarrow^{k_1},$$

*where $k_1$ is the rate constant. That is, $\mathbf{A}$ catalyzes the production of $\mathbf{O}$, which is then degraded with the same rate. Then, composing $\mathcal{F}$ and $\mathcal{C}$, we have*

$$\mathbb{E}^{\infty}_{\mathbf{O}} = 2\frac{k_p}{k_d}; \quad \mathbb{V}^{\infty}_{\mathbf{O}} = \frac{k_p(5k_1 + 2k_d)}{k_d(k_1 + k_d)}.$$

*The steady state expectation of $\mathbf{O}$ does not depend on the choice of $k_1$: it is always equal to $\mathbb{E}^{\infty}_{\mathbf{A}}$. However, $\mathbb{V}^{\infty}_{\mathbf{O}}$ does depends on $k_1$. As a consequence, we ask if there are values of $k_1$ for which $\mathbf{O}$ has less variance than $\mathbf{A}$. This would yield a molecular filter according to Definition 16. It is easy to show that*

$$\textit{If } 2k_1 < k_d \textit{ then } \mathbb{V}^{\infty}_{\mathbf{O}} < \mathbb{V}^{\infty}_{\mathbf{A}}.$$




D

| linear filter | annihilation module | annihilation filter |
|---|---|---|
| Transcr. lev. *S. cerevisiae* [133] | *c-Myc miR-17-5p E2F1* [124] | *c-Myc-E2F1-E2F2* [135] |
| Transl. lev. *B. subtilis* [125] | *c-Myc miR-20a E2F1* [124] | |
| Transl. lev. *S. cerevisiae* [30] | *Ubx expression* [91, 137] | |
| Transl. lev. *E.Coli* [69] | Drosophila eye [109] | |
| Transl. lev. *Yeast* [133] | | |

Figure 6.1: Molecular filters (Top row) along with biological examples where such components can be found (Middle row). We stress that for many of these examples it has yet to be verified experimentally that those systems behave as our filters. We simply present them here to facilitate understanding (see Section 6.6). Dashed arrows represent catalytic production/activation effects and continuous arrows molecular transitions. ($A$) First-order low-pass filter implemented by means of slow production and degradation reactions (linear filter, Eqn (6.3)). These networks are present both at the translation and transcription level to increase robustness of gene expression. ($B$) annihilation module (Eqn (6.10)) given by co-expression of two species that are then degraded together. These networks arise in microRNA regulated gene expression. ($C$) annihilation filter (Eqn (6.16)) based on the annihilation module. Such a system may be a suitable model for the c-Myc—E2F1—hsa-miR-20a circuit with its extension to E2F2 [139] under the assumption that E2F1 activates E2F2, except for the feed-forward loop between $\mathbf{A}$ and $C$ which has no influence on noise reduction (see Section 6.6). We stress that this assumption needs further experiments to be confirmed. ($D$) Biological systems where the filters may play a role in reducing noise.

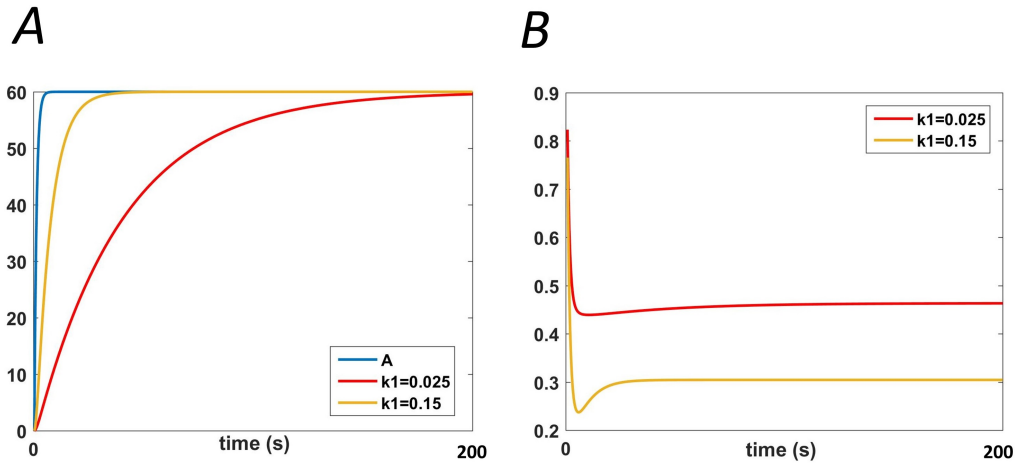



Figure 6.2: Comparison of $\mathcal{F}_L$ for different values of $k_1$ ($k_1 = 0.025$ and $k_1 = 0.15$) for an input process $\mathbf{A}$ generated by reactions in (6.2) for $k_p = 30$ and $k_d = 1$. $(A)$ we plot $\mathbb{E}_{\mathbf{A}}(t)$ and $\mathbb{E}_{\mathbf{O}}(t), t \in [0, 200]$ for different values of $k_1$. $(B)$ We plot $F_{\mathbf{A}}(t) - F_{\mathbf{O}}(t), t \in [0, 200]$ for two different values of $k_1$. Both rewards are obtained using the CLA approximation of instantaneous rewards presented in Section 5.5.3. $\mathcal{F}_L$ works as a low pass filter: the more delay it introduces in tracking the mean, the more noise reduction it introduces in the system.

*An interesting question is what is the minimum value of* $\mathbb{V}^{\infty}_{\mathbf{O}}$ *that can be obtained. By minimizing* $\mathbb{V}^{\infty}_{\mathbf{O}}$ *with respect to* $k_1$ *we have*

$$\inf_{k_1 \in \mathbb{R}_{\geq 0}} (\mathbb{V}^{\infty}_{\mathbf{O}}) = 2\frac{k_p}{k_d},$$

*which is obtained for* $k_1 \to 0$*. This means that the best that we can achieve with* $\mathcal{F}_L$ *is to produce an output species with variance equal to expectation, hence behaving as a Poisson process [160]. In Section 6.2.2 we will show that this is general for* $\mathcal{F}_L$*, and holds for all CRNs such that* $\mathbf{A}$ *and* $\mathbf{O}$ *are positively correlated.*

*We can also derive an analytic solution for the transient evolution of the expectation of* $\mathbf{A}$ *and* $\mathbf{O}$*. For* $\mathbf{A}$ *we have that for* $t \in \mathbb{R}_{\geq 0}$

$$\mathbb{E}_{\mathbf{A}}(t) = 2\frac{k_p}{k_d}(1 - \exp^{k_d t}),$$

*while for* $\mathbf{O}$

$$\mathbb{E}_{\mathbf{O}}(t) = 2\frac{k_p}{k_d(k_1 - k_d)}(-k_d + k_d exp^{-k_1 t} + k_1 - k_1 \exp^{-k_d t}).$$

*As expected, for* $t \to \infty$ *we have* $\mathbb{E}_{\mathbf{O}}(t) = \mathbb{E}_{\mathbf{A}}(t)$*. Moreover,* $k_1$ *and* $k_d$ *control how fast* $\mathbb{E}_{\mathbf{O}}(t)$ *tracks* $\mathbb{E}_{\mathbf{A}}(t)$*. This shows a trade-off between how fast* $\mathbf{O}$ *keeps tracking* $\mathbf{A}$ *and the quantity of noise that is filtered out.* $\mathcal{F}_L$ *acts as a buffer, which introduces a*

*delay to filter the fast dynamics of* $\mathbf{A}$*.* $\mathcal{F}_L$ *is indeed a molecular implementation of a* first-order low-pass filter*, that is, a molecular circuit that enables attenuating the high frequency components of a signal by introducing a delay [146].*

*To illustrate this point more clearly, we consider the following reward structures*

$$\rho_1^{\mathbf{A}}(\mathcal{X}) = \mathcal{X}_{\mathbf{A}} \quad \rho_1^{\mathbf{O}}(\mathcal{X}) = \mathcal{X}_{\mathbf{O}} \quad \rho_2^{\mathbf{A}} = \mathcal{X}_{\mathbf{A}}^2 \quad \rho_2^{\mathbf{O}} = \mathcal{X}_{\mathbf{O}}^2.$$

*Then, we have*

$$\mathbb{E}_{\mathbf{A}}(t) = R_{=?}[I^{=t}_{\rho_1^{\mathbf{A}}}] \quad F_{\mathbf{A}}(t) = \frac{R_{=?}[I^{=t}_{\rho_2^{\mathbf{A}}}] - (R_{=?}[I^{=t}_{\rho_1^{\mathbf{A}}}])^2}{R_{=?}[I^{=t}_{\rho_1^{\mathbf{A}}}]},$$

*and similarly for* $\mathbf{O}$*. In Figure 6.2, by making use of the techniques developed in Chapter 5, we compute those rewards and compare the action of* $\mathcal{F}_L$ *for different values of* $k_1$ *over time. We observe that, to obtain a smaller Fano factor in* $\mathbf{O}$*, a greater delay is needed.*

*Despite its simplicity, the performance of* $\mathcal{F}_L$ *is limited by Poisson noise. By using a more complex filter with non-linear dynamics (in the form of bi-molecular reactions), we can obtain much better performance in term of noise reduction at the cost of a more complex circuit. For instance, consider CRN* $\mathcal{F}_A$*, given by the following reactions*

$$\mathcal{F}_A: \quad \mathbf{A} \to^{r_1} \mathbf{A} + \lambda_B + \lambda_D \qquad \lambda_D \to^{r_1} \lambda_E \qquad \lambda_B + \lambda_E \to^{r_2}$$

$$\lambda_B + \mathbf{A} \to^{r_3} \mathbf{A} + \mathbf{O} \qquad \lambda_E + \mathbf{O} \to^{r_3},$$

*where* $\lambda_B, \lambda_D, \lambda_E$ *are auxiliary species initialized at 0 molecules.* $\mathcal{F}_A$ *will be discussed in detail in Section 6.4. Here, in Figure 6.3, we compare the performance of* $\mathcal{F}_L$ *and* $\mathcal{F}_A$ *for* $k_1 = 0.025$*,* $r_1 = 1, r_2 = 10, r_3 = 0.0008$*. This shows that* $\mathcal{F}_A$ *can achieve a much better noise reduction at the price of a more complex structure with non-linear dynamics.*

We define the following CRN, which we call a *linear filter* and is composed of a production and a degradation reaction with output species $\mathbf{O}$.

**Definition 17.** *A CRN* $\mathcal{F}_L$ *is called linear filter if, for some* $k_1, k_2 \in \mathbb{R}_{>0}$*,* $\mathcal{F}_L$ *is composed from the following reactions*

$$\mathcal{F}: \qquad \mathbf{A} \to^{k_1} \mathbf{A} + \mathbf{O} \qquad \mathbf{O} \to^{k_2}, \tag{6.3}$$

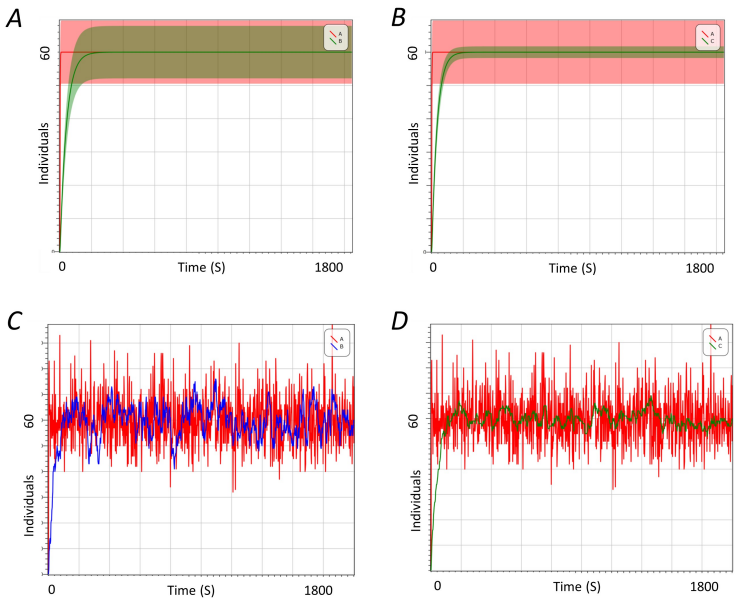



Figure 6.3: Comparison of the actions of $\mathcal{F}_L$ and $\mathcal{A}$ on CRN $\mathcal{C}$ for $k_1 = 0.025$, $r_1 = 1, r_2 = 10, r_3 = 0.0008$. Plots compare expected value and variance of $\mathbf{A}$ with species $\mathbf{O}$ in $\mathcal{F}$ $(A)$ and with species $C$ in $\mathcal{A}$ $(B)$. ($C$ and $D$) Stochastic simulation plots of $\mathbf{A}$ (red) with $\mathbf{O}$ (blue) and $C$ (green).

We consider the general scenario, shown in Figure 6.4, where $\mathcal{F}_L$ is embedded within an arbitrarily complex, possibly non-linear, reaction network with the only constraint that the output species $\mathbf{O}$ is changed only by the reactions in $\mathcal{F}$. We allow $\mathbf{O}$ to act as a catalyst in arbitrarily many reactions, and $\mathbf{A}$ to interact with the larger network with no constraint. This scenario is very general and includes the case where $\mathbf{A}$ is a function of $\mathbf{O}$ with feedback loops.

### 6.2.1 Classical Frequency Analysis of Linear Filters

The transfer function of $\mathcal{F}_L$ is obtained by applying the Fourier transform to the mass action rate equation corresponding to $\mathbf{O}$

$$\frac{d\Phi_{\mathbf{O}}(t)}{dt} = k_1\Phi_{\mathbf{A}}(t) - k_2\Phi_{\mathbf{O}}(t),$$

where $\Phi_{\mathbf{A}}$ and $\Phi_{\mathbf{O}}$ are the deterministic signals modeling the time evolution of $\mathbf{A}$ and $\mathbf{O}$. In the frequency domain, we get

$$i\omega\hat{\Phi}_{\mathbf{O}}(\omega) = k_1\hat{\Phi}_{\mathbf{A}}(\omega) - k_2\hat{\Phi}_{\mathbf{O}}(\omega)$$

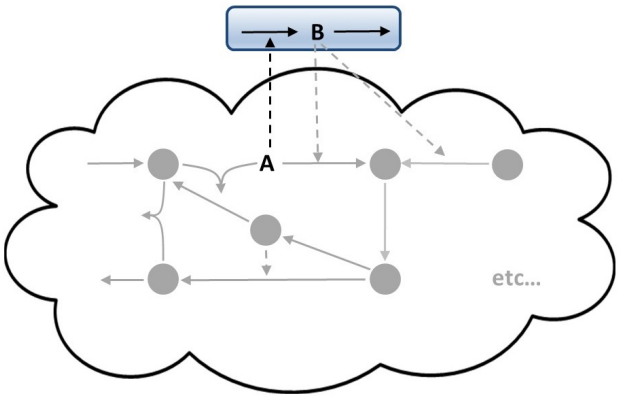



Figure 6.4: The linear filter (Eqn (6.3)) is embedded within an arbitrary reaction network. **A** can interact with any other species of the network with no restrictions, while **O** production is catalyzed by **A**, and **O** may act as a catalyst in arbitrarily many reactions.

where $\omega$ is the angular frequency, and $\hat{\Phi}_{\mathbf{O}}(\omega), \hat{\Phi}_{\mathbf{A}}(\omega)$ are the Fourier transforms of signals $\Phi_{\mathbf{O}}$ and $\Phi_{\mathbf{A}}$. For $k_1 = k_2$, we obtain

$$\frac{\hat{\Phi}_{\mathbf{O}}(\omega)}{\hat{\Phi}_{\mathbf{A}}(\omega)} = \frac{1}{1 + \frac{i\omega}{k_1}}, \tag{6.4}$$

Eqn (6.4) is the transfer function of a *first-order low-pass filter* [154]. This network attenuates frequencies higher than the cutoff frequency by introducing a delay and integrating the fast dynamics. From Eqn (6.4) the cutoff frequency is exactly $\bar{\omega} = k_1$. This means that, the higher the value of $k_1$, the less noise is filtered out, but the faster **O** tracks **A**.

### 6.2.2 Stochastic analysis

The classical frequency analysis of $\mathcal{F}_L$, based on the Fourier transform of the rate equations, does not take into account the intrinsic noise introduced by the reactions firing in $\mathcal{F}$: it only considers the extrinsic noise modeled as fast fluctuations of the input. However, the intrinsic noise cannot be simply neglected, as it may drive the behavior of biological systems [136]. This is the case in gene expression, where low molecular counts are often involved, and deterministic modeling is generally unsatisfactory [120, 69]. Thus, a deterministic analysis, clearly, cannot guarantee a certain reaction network is a molecular filter. To resolve this, we need to consider the CTMC induced by $\mathcal{F}_L$. In the following theorem we compute expectation and variance of

**O** in the general scenario shown in Figure 6.4. That is, Theorem 8 holds independently of what happens in the rest of the network and includes the cases where **A** is a non-linear function of **O** with feedback loops.

**Theorem 8.** *Let $\mathcal{C} = (\Lambda, R)$ and $\mathcal{F}_L = (\Lambda', R')$ be CRNs. Assume $\mathbf{A} \in \Lambda \cap \Lambda'$ and that $\mathbf{O}$ is changed only through $\mathcal{F}$. That is, $\upsilon_{\tau,\mathbf{O}} = 0$ for $\tau \neq R'$. Then, for $\mathcal{C} \cup \mathcal{F}_L$ it holds that*

$$\mathbb{E}^{\infty}_{\mathbf{O}} = \frac{k_1}{k_2}\mathbb{E}^{\infty}_{\mathbf{A}} \tag{6.5}$$

$$\mathbb{V}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\mathbf{O}} + \frac{k_1}{k_2} Cov[\mathbf{A}, \mathbf{O}]_{\infty}, \tag{6.6}$$

*where $Cov[\mathbf{A}, \mathbf{O}]_{\infty} = \mathbb{E}^{\infty}_{\mathbf{A}\cdot\mathbf{O}} - \mathbb{E}^{\infty}_{\mathbf{A}}\mathbb{E}^{\infty}_{\mathbf{O}}$ is the covariance of $\mathbf{A}$ and $\mathbf{O}$ at steady state with $\mathbb{E}^{\infty}_{\mathbf{A}\cdot\mathbf{O}}$ expectation of $\mathbf{A}\cdot\mathbf{O}$ at steady state.*

*Proof.* The steady state solution of the $k$th moment of **O** can be derived solving Eqn (4.3) at steady state. Thus,

$$0 = \mathbb{E}[\sum_{\tau\in R} \alpha_\tau(\mathcal{X})((\mathcal{X}_{\mathbf{O}} + \upsilon_{\tau,\mathbf{O}})^k - \mathcal{X}^k_{\mathbf{O}})]_{\infty} =$$
$$\mathbb{E}[\sum_{\tau\in R\, s.t.\, \upsilon_{\tau,\mathbf{O}}\neq 0} \alpha_\tau(\mathcal{X})((\mathcal{X}_{\mathbf{O}} + \upsilon_{\tau,\mathbf{O}})^k - \mathcal{X}^k_{\mathbf{O}})]_{\infty} + \mathbb{E}[\sum_{\tau\in R\, s.t.\, \upsilon_{\tau,\mathbf{O}}=0} \alpha_\tau(\mathcal{X})(\mathcal{X}^k_{\mathbf{O}} - \mathcal{X}^k_{\mathbf{O}})]_{\infty} =$$
$$\mathbb{E}[k_1\mathcal{X}_{\mathbf{A}}((\mathcal{X}_{\mathbf{O}} + 1)^k - \mathcal{X}^k_{\mathbf{O}})]_{\infty} + \mathbb{E}[k_2\mathcal{X}_{\mathbf{O}}((\mathcal{X}_{\mathbf{O}} - 1)^k) - \mathcal{X}^k_{\mathbf{O}}]_{\infty}.$$

where the notation $\mathbb{E}[f((\mathcal{X}))]_{\infty}$ indicates that the expectation is taken for $\mathcal{X}$ at steady state. For $k = 1$, we obtain exactly Eqn (6.5). The second moment of **O** at steady state can be derived for $k = 2$. Moreover, as $\mathbb{E}^{\infty}_{\mathbf{O}} = \frac{k_1}{k_2}\mathbb{E}^{\infty}_{\mathbf{A}}$, we have

$$\mathbb{E}^{\infty}_{\mathbf{O}^2} = \frac{k_1}{k_2}\mathbb{E}^{\infty}_{\mathbf{A}\cdot\mathbf{O}} + \mathbb{E}^{\infty}_{\mathbf{O}},$$

which leads to the following form for the variance of **O** at steady state

$$\mathbb{V}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\mathbf{O}^2} - (\mathbb{E}^{\infty}_{\mathbf{O}})^2 = \mathbb{E}^{\infty}_{\mathbf{O}} + \frac{k_1}{k_2} Cov[\mathbf{A}, \mathbf{O}]_{\infty} \tag{6.7}$$

□

Note that in Theorem 8 we do allow the use of **O** as a catalyst in arbitrarily many reactions.

Eqn (6.5) guarantees that the expected value of **O** always tracks the expectation of **A**, no matter which biochemical system is producing **A** and what happens in the rest

of the system. Perhaps more interesting is Eqn (6.6), which shows that, for any input signal, the filtered signal $\mathbf{O}$ has variance that is equal to its mean plus the covariance between $\mathbf{A}$ and $\mathbf{O}$. Assuming $\mathbf{A}$ and $\mathbf{O}$ are non-negatively correlated, then we have $\mathbb{E}^{\infty}_{\mathbf{A}\cdot\mathbf{O}} \geq \mathbb{E}^{\infty}_{\mathbf{A}}\mathbb{E}^{\infty}_{\mathbf{O}}$. As a result, in case of non-negative correlation between $\mathbf{A}$ and $\mathbf{O}$, for any $k_1, k_2 \in \mathbb{R}_{\geq 0}$, the following *lower bound* holds

$$F_{\mathbf{O}} \geq 1. \tag{6.8}$$

$\mathbf{A}$ and $\mathbf{O}$ being non-negatively correlated is natural since $\mathbf{A}$ catalyzes the production of $\mathbf{O}$. For instance, this is the case in Example 6.2.1. However, if a negative feedback loop between $\mathbf{O}$ and $\mathbf{A}$ is present, this may change the sign of their correlation, potentially leading to noise reduction to below Poisson levels, as shown in the next subsection.

**Example 6.2.2.** *We consider the following CRN where there is a feedback between* $\mathbf{A}$ *and* $\mathbf{O}$ *and* $\lambda_L$ *is an auxiliary species.*

$$\begin{aligned} &\to^{1} \lambda_L \quad \lambda_L \to^{100} \lambda_L + \mathbf{A} + \mathbf{A}; \quad \mathbf{A} \to^{0.1} \\ &\mathbf{A} \to^{0.01} \mathbf{A} + \mathbf{O}; \quad \mathbf{O} \to^{0.01}; \quad \mathbf{O} + \lambda_L \to^{k_f} \mathbf{O}, \end{aligned} \tag{6.9}$$

*where* $k_f > 0$ *is a rate constant. That is, a feedback between* $\mathbf{A}$ *and* $\mathbf{O}$ *is present. The strength of the feedback can be controlled by changing the rate* $k_f$*.* $\mathbf{O}$ *can be thought as a protein that inhibits its expression. The above CRN meets the condition of validity of Eqns* (6.5) *and* (6.6)*. Thus, Eqn* (6.5) *guarantees that for any possible initial condition and value of* $k_f$

$$\mathbb{E}^{\infty}_{\mathbf{A}} = \mathbb{E}^{\infty}_{\mathbf{O}},$$

*while Eqn* (6.6) *guarantees that* $F^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\mathbf{A}} + Cov[\mathbf{A}, \mathbf{O}]_{\infty}$*. Thus, to compute the Fano factor of* $\mathbf{O}$ *at steady state we need to estimate* $Cov[\mathbf{A}, \mathbf{O}]_{\infty}$*. The system is non-linear. As a consequence,* $Cov[\mathbf{A}, \mathbf{O}]_{\infty}$ *cannot be computed exactly, but can be estimated using the CLA. We obtain*

$$\begin{aligned} Cov[\mathbf{A}, \mathbf{O}]_{\infty} = \\ &\frac{10\big(-2.2231415 \cdot 10^9 + \frac{541202.2407982067}{\sqrt{k_f}} + 9.282074699360052 \cdot 10^{10}\sqrt{k_f}\big)}{121 - 1.9802 \cdot 10^9 k_f + 4.84 \cdot 10^{14} k_f^2} \\ &+ \frac{10\big(4.4027 \cdot 10^{14} k_f - 8.362894235849215 \cdot 10^{14} k_f^{\frac{3}{2}}\big)}{121 - 1.9802 \cdot 10^9 k_f + 4.84 \cdot 10^{14} k_f^2}. \end{aligned}$$

*We obtain that, for* $k_f > 0.277368$, $Cov[\mathbf{A}, \mathbf{O}]_{\infty} < 0$*. Thus,* $F^{\infty}_{\mathbf{O}} < 1$*. This shows how strong feedback can reduce the Fano factor of* $\mathbf{O}$ *to below Poisson levels. However,*

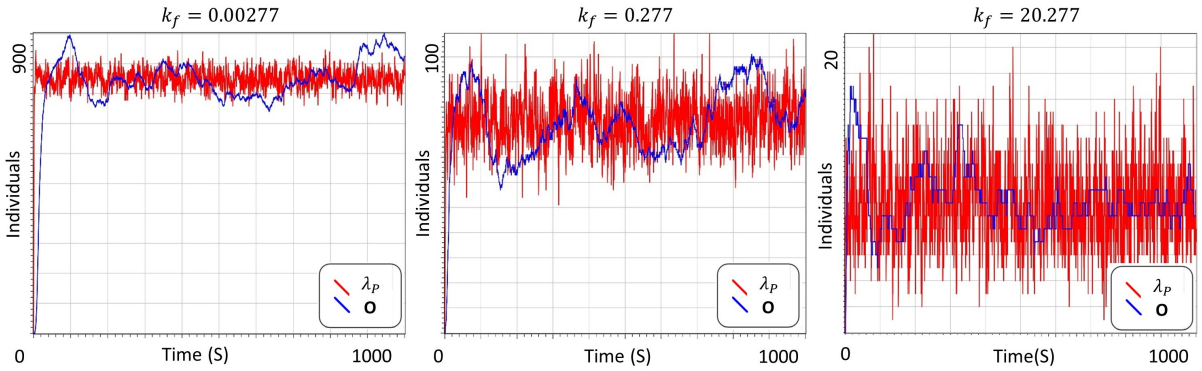



Figure 6.5: For the CRN in Eqn (6.9) we plot a stochastic simulation of **O** for three different values of $k_f$ and assuming initial condition of all the species are 0. In all the plots $\lambda_P$ is a signal with Poisson noise and such that $\mathbb{E}^\infty_{\lambda_P} = \mathbb{E}^\infty_{\mathbf{A}}$. When the feedback is weak ($k_f = 0.00277$) the Fano factor of **O** is greater than one. For $k_f = 0.277$ our analysis predicts a Fano foctor for **O** of approximately 1. When the the feedback is strong ($k_f = 20.277$) the Fano factor of **O** is smaller than 1, but there is also a strong repression of $\mathbb{E}^\infty_B$.

*strong feedback means strong repression of the mean of $B$ and* **A** *by Eqn* (6.5)*. To confirm the mathematical analysis, in Figure 6.5, for different values of $k_f$, we plot a single stochastic simulation of $B$ compared with a signal with the same mean of $B$ but affected by Poisson noise.*

### 6.2.3 Higher Order Linear Filters

A possible way to improve first-order filters could be to consider second or higher order filters. That is, filters with transfer function of the type

$$\frac{c}{(i\omega + a)(i\omega + b)}.$$

In fact, these filters are known to guarantee better performance because of their sharper attenuation of the higher frequencies [154]. Second or higher order filters (in case non real and non-positive roots) can be implemented as a cascade of linear filters. For instance, for the CRN given by the following reactions

$$\mathbf{A} \to^{k_1} \mathbf{A} + \lambda_C \quad \lambda_C \to^{k_1} \quad \lambda_C \to^{k_2} \lambda_C + \mathbf{O} \quad \mathbf{O} \to^{k_2},$$

in the Fourier domain we have

$$\frac{\hat{\Phi}_{\mathbf{O}}(i\omega)}{\hat{\Phi}_{\mathbf{A}}(i\omega)} = \frac{k_1 \cdot k_2}{(i\omega + k_1)(i\omega + k_2)}$$

However, a frequency characterization of such systems does not take into account the intrinsic noise introduced by the filter itself. As a result, each linear filter in the cascade is still limited by Poisson noise (Theorem 8). Therefore, higher-order filters, if linear, do not improve the performance of first-order filters.

An example of such a mechanism can be found in multi-step models of gene expression [131], where protein expression is a sequence of linear reactions, thus limited by Eqn (6.6). Another example can be observed in signaling cascades, such as the mitogen-activated protein kinase (MAPK) cascade [147], where non-linear filtering mechanisms are necessary in order to reduce stochastic fluctuations of a downstream process below Poisson levels [159].

## 6.3 Annihilation Module

The noise filtering capability of linear filters can be improved by using *higher-order reactions*. The simplest second-order reaction is complex formation. We show how complex formation of two molecules that are positively correlated in their expression can indeed work as an efficient molecular filter. The following CRN, which we call *annihilation module* (Figure 6.1$B$), is based on binding and degradation of two parallel synthesized molecules. Thus, the annhilation module, $\mathcal{F}_M$, is composed of the following reactions

$$\mathcal{F}_M : \quad \mathbf{A} \rightarrow^{r_1} \mathbf{A} + \mathbf{O} + \lambda_C; \quad \mathbf{O} + \lambda_C \rightarrow^{r_2}, \tag{6.10}$$

where $\lambda_C$ is an auxiliary species. In this module, $\mathbf{O}$ and $\lambda_C$ are co-expressed and then they inhibit each other. To study how the annihilation module behaves with respect to intrinsic and extrinsic noise, we consider a general birth-death process $\mathbf{A}$ affected by Poisson noise. That is, $\mathbf{A}$ is generated and removed by the following CRN, which we call $\mathcal{A}$ and composed by the following two reactions for $k_p, k_d \in \mathbb{R}_{\geq 0}$

$$\rightarrow^{k_p} \mathbf{A}; \quad \mathbf{A} \rightarrow^{k_d} .$$

Thus, we have $\mathbb{E}_{\mathbf{A}}(t) = \mathbb{V}_{\mathbf{A}}(t), t \in \mathbb{R}_{\geq 0}$. Since $\mathcal{F}_M$ is non-linear, a general and exact analysis, as in the linear case, cannot be performed, as the moment equations cannot be solved. Consequently, in the following theorem we make use of the CLA and derive analytic solution for the expectation and Fano factor of $\mathbf{O}$ at steady state for such an input process.

**Theorem 9.** *Let $\mathcal{C} = \mathcal{F}_M \cup \mathcal{A}$ be the CRN union of CRNs $\mathcal{F}_M$ and $\mathcal{A}$. Then, it holds that*

$$\mathbb{E}^{\infty}_{\mathbf{O}} = \sqrt{\frac{r_1 \mathbb{E}^{\infty}_{\mathbf{A}}}{r_2}} \tag{6.11}$$

$$F^{\infty}_{\mathbf{O}} = \frac{2r_1^{3/2}\sqrt{r_2 k_p k_d} + 4r_1 r_2 k_p - r_1 k_d^2 - k_d^3}{8 r_1 r_2 k_p - 2k_d^3}. \tag{6.12}$$

*Proof.* Solving Eqn (4.6) we obtain the following relation for the expectation

$$r_1 \mathbb{E}^{\infty}_{\mathbf{A}} = r_2 \mathbb{E}^{\infty}_{\mathbf{O}} \mathbb{E}^{\infty}_{\lambda_C}. \tag{6.13}$$

In order to obtain a finite number of solutions we introduce the following additional constraint, which holds assuming $\mathbf{O}$ and $\lambda_C$ start from the same initial condition

$$\mathbb{E}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\lambda_C}.$$

Solving these togheter we obtain (6.11). For the variance, solving Eqn (4.7) at steady state, we obtain the following equations

$$\begin{aligned}
&- \mathbb{E}^{\infty}_{\mathbf{O}} Cov[\mathbf{A}, \lambda_C]_{\infty} r_2 - \mathbb{E}^{\infty}_{\lambda_C} Cov[\mathbf{A}, \mathbf{O}]_{\infty} r_2 + \mathbb{V}^{\infty}_{\mathbf{A}} r_1 - Cov[\mathbf{A}, \mathbf{O}]_{\infty} k_P = 0 \\
&- \mathbb{E}^{\infty}_{\mathbf{O}} Cov[\mathbf{A}, \lambda_C]_{\infty} r_2 - \mathbb{E}^{\infty}_{\lambda_C} Cov[\mathbf{A}, \mathbf{O}]_{\infty} r_2 + \mathbb{V}^{\infty}_{\mathbf{A}} r_1 - Cov[\mathbf{A}, \lambda_C]_{\infty} k_P = 0 \\
&\mathbb{E}^{\infty}_{\mathbf{A}} r_1 + \mathbb{E}^{\infty}_{\mathbf{O}} \mathbb{E}^{\infty}_{\lambda_C} r_2 - \mathbb{E}^{\infty}_{\mathbf{O}} \mathbb{V}^{\infty}_{\mathbf{O}} r_2 - \mathbb{E}^{\infty}_{\mathbf{O}} V[C]_{\infty} r_2 - \\
&\qquad \mathbb{E}^{\infty}_{\lambda_C} \mathbb{V}^{\infty}_{\mathbf{O}} r_2 - \mathbb{E}^{\infty}_{\lambda_C} Cov[\mathbf{O}, \lambda_C]_{\infty} r_2 + Cov[\mathbf{A}, \mathbf{O}]_{\infty} r_1 + Cov[\mathbf{A}, \lambda_C]_{\infty} r_1 = 0 \\
&\mathbb{E}^{\infty}_{\mathbf{A}} r_1 + \mathbb{E}^{\infty}_{\mathbf{O}} \mathbb{E}^{\infty}_{\lambda_C} r_2 - 2\mathbb{E}^{\infty}_{\mathbf{O}} Cov[\mathbf{O}, \lambda_C]_{\infty} r_2 - 2\mathbb{E}^{\infty}_{\lambda_C} \mathbb{V}^{\infty}_{\mathbf{O}} r_2 + 2Cov[\mathbf{A}, \mathbf{O}]_{\infty} r_1 = 0 \\
&\mathbb{E}^{\infty}_{\mathbf{A}} r_1 + \mathbb{E}^{\infty}_{\mathbf{O}} \mathbb{E}^{\infty}_{\lambda_C} r_2 - 2\mathbb{E}^{\infty}_{\mathbf{O}} \mathbb{V}^{\infty}_{\lambda_C} r_2 - 2\mathbb{E}^{\infty}_{\lambda_C} Cov[\mathbf{O}, \lambda_C]_{\infty} r_2 + 2Cov[\mathbf{A}, \lambda_C]_{\infty} r_1 = 0
\end{aligned}$$

In order to derive a system of ODEs with a finite number of solutions we need to introduce an additional constraint, which will depend on the initial conditions. Since initially all the variances and covariances are 0 (we start from a known initial condition), from Eqn (4.7) we can add the following constraint

$$2\mathbb{V}^{\infty}_{\mathbf{O}} = Cov[\mathbf{O}, \lambda_C]_{\infty} + \mathbb{E}^{\infty}_{\lambda_C}.$$

Solving the resulting system of equations, we obtain (6.12). □

**Corollary 9.1.** Assume $r_1 = r\gamma$ and $r_2 = r$, with $r, \gamma \in \mathbb{R}_{>0}$, then it holds that

$$\lim_{r \to 0} F^{\infty}_{\mathbf{O}} = \frac{1}{2}, \tag{6.14}$$

$$\lim_{r \to \infty} F^{\infty}_{\mathbf{O}} = \frac{2+n}{4}, \tag{6.15}$$

where $n = \frac{\mathbb{E}^{\infty}_{\mathbf{O}}}{\mathbb{E}^{\infty}_{\mathbf{A}}}$.

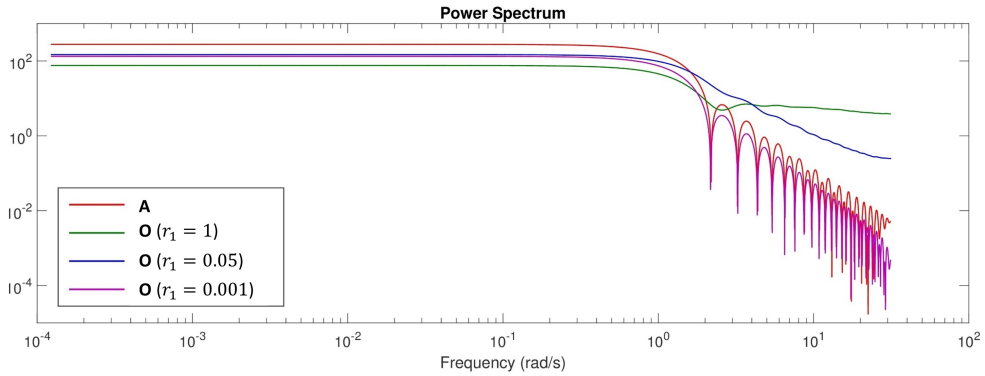



Figure 6.6: Comparison of the PSD of the input signal **A** affected by Poisson noise for different implementations of the annihilation filter, varying the parameter rates, and for $\mathbb{E}^{\infty}_{\mathbf{A}} = \mathbb{E}^{\infty}_{\mathbf{O}}$. The power spectrum is estimated using the Blackman-Tukey Spectral Estimate algorithm with frequency deep resolution, as implemented in Matlab, over 300000 data points with sampling time of 0.1 seconds. In all the different cases, the annihilation module attenuates the low frequency components of the input signal, while it can amplify the high frequency if the rates are faster than those of the input signal.

Corollary 9.1 guarantees that in the limit of an arbitrarily long delay introduced $F^{\infty}_{\mathbf{O}}$ converges to $\frac{1}{2}$, halving Poisson values. Perhaps, Eqn (6.15) may appear more surprising. It guarantees that for $n = 1$, that is, $\mathbb{E}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\mathbf{A}}$ (perfect tracking of the mean), the Fano factor is always smaller than 1 for arbitrarily large values of $r_1$ and $r_2$. Thus, $\mathcal{F}_M$ can act as a molecular filter and reduce the noise even without introducing a delay in its buffering action. This can be justified because this architecture, where **O** and $\lambda_C$ are co-expressed, enables attenuation of the low-frequency components of the input signal. Therefore, we obtain noise reduction even if the high-frequency components are not necessarily attenuated (Figure 6.6), where we compare the Power Spectral Density (PSD) of **A** with that of **O** of the annihilation filter, for different rate parameters. The relationship between the PSD and the Fano factor of a signal is explained in Section 3.2. We select $r_1, r_2$ such that $\mathbb{E}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\mathbf{A}}$. For any choice of $r_1$ we have that **O** attenuates the low frequency components of the input signal. This causes a reduction of the Fano factor independently of the value of $r_1$. However, if $r_1$ is large, the high frequency components tend to be more amplified, while for $r_1$ small enough these are attenuated. This explains why noise attenuation is greater for small rates. Therefore, the annihilation module reduces the noise by attenuating the low frequency components of a signals, and, if its dynamics are slow enough, it reduces the noise further by integrating the fast dynamics.

We note that, in the annihilation module, the steady-state value of the output

signal is proportional not to the steady-state value of the input signal, but to its square root (Eqn (6.11)). This may be beneficial in molecular networks, where it may help maintain regulatory stability under changes in initial conditions. However, this mechanism would not be appropriate in cases where the long-term evolution of the upstream component should be followed, because changes to the input would not be followed proportionally (Figure 6.7*B*).

## 6.4 Annihilation Filter

We propose a general architecture, called the *annihilation filter*, which is based on the annihilation module but also guarantees $\mathbb{E}^{\infty}_{\mathbf{O}} = s\mathbb{E}^{\infty}_{\mathbf{A}}$ for a given constant $s \geq 0$, independently of $\mathbb{E}^{\infty}_{\mathbf{A}}$ (in what follows, without any loss of generality, we assume $s = 1$). We show how the annihilation filter can asymptotically reduce molecular noise to zero, that is, for appropriate limiting values for the rates, the Fano factor of the output converges to zero (Eqn (6.22)).

Our annihilation filter, $\mathcal{F}_A$, is illustrated in Figure 6.7*C*. $\mathcal{F}_A$ is composed of the following reactions

$$\begin{aligned} &\mathbf{A} \rightarrow^{r_1} \mathbf{A} + \lambda_B + \lambda_D \quad \lambda_D \rightarrow^{r_1} \lambda_E \quad \lambda_B + \lambda_E \rightarrow^{r_2} \\ &\lambda_B + \mathbf{A} \rightarrow^{r_3} \mathbf{A} + \mathbf{O} \quad \lambda_E + \mathbf{O} \rightarrow^{r_3} \end{aligned} \tag{6.16}$$

**A** is the input species and **O** is the output filtered species. The first three reactions are similar to the annihilation module, but with an additional delay introduced by the reaction $\lambda_D \rightarrow^{r_1} E$. Eqn (6.5) guarantees that $\lambda_D$ is a copy of **A**, and the number of times that $\lambda_D$ molecules have been produced or destroyed is stored respectively in $\lambda_B$ and $\lambda_E$. As the role of these reactions is to act as a sensor, high values of $r_1$ are more informative than small ones. If $r_2$ is large enough, the count of **O** is modified not any time a $\lambda_B$ or $\lambda_E$ molecule is produced, but just by their difference. The fourth and fifth reactions increase or decrease **O**. The rate $r_3$ controls the delay introduced by the filter, and thus also the noise reduction.

Since the system is non-linear, to study the noise reduction capabilities of the annihilation filter we make use of the CLA. We assume **A** is a general input process, with extrinsic noise modeled by a Poisson process. That is, **A** is generated by the CRN $\mathcal{S}$, which is composed by the following reactions

$$\rightarrow^{k_p} \mathbf{A}; \quad \mathbf{A} \rightarrow^{k_d} .$$

Thus, we have $\mathbb{E}_{\mathbf{A}}(t) = \mathbb{V}_{\mathbf{A}}(t), t \in \mathbb{R}_{\geq 0}$. Using the CLA equations we can derive the following theorem.

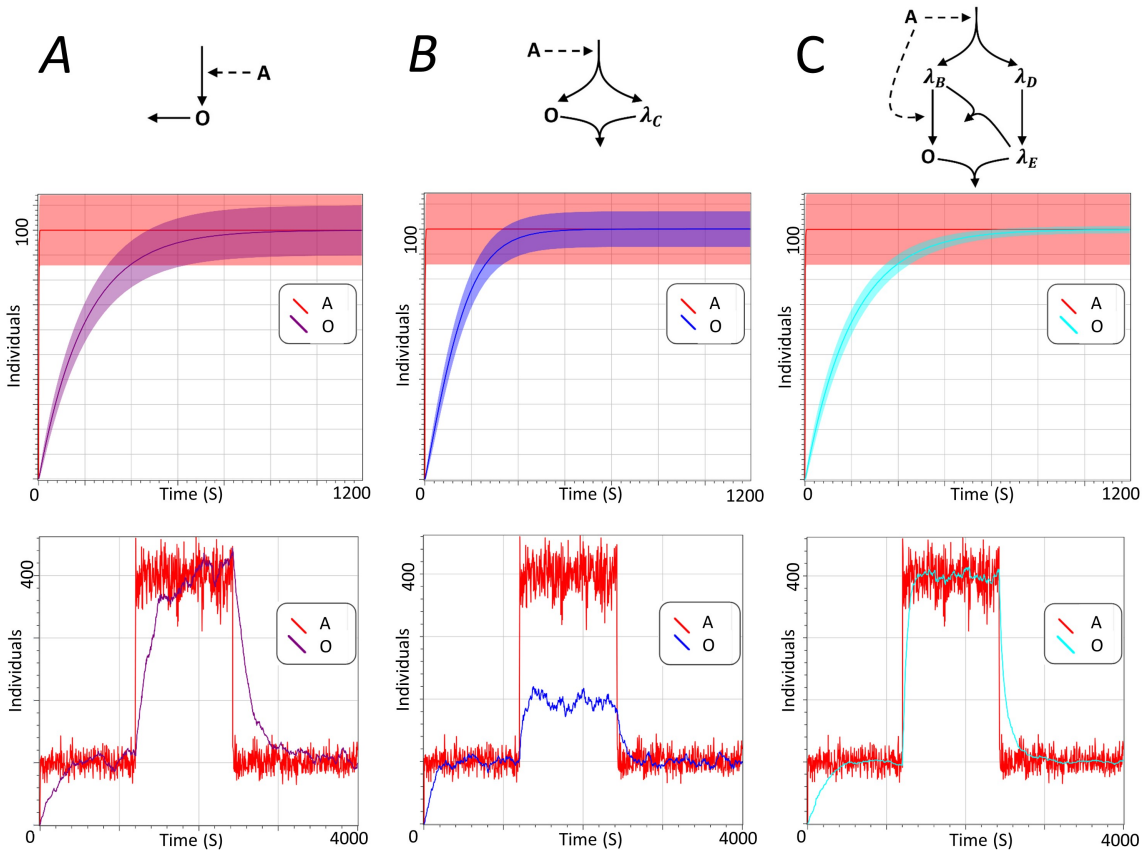



Figure 6.7: Comparison of the actions of $\mathcal{F}_L$ (linear filter, CRN (6.3)), $\mathcal{F}_M$ (annihilation module, Eqn (6.10)), and $\mathcal{F}_A$ (annihilation filter, Eqn (6.16)) on noisy input $\mathbf{A}$ generated by reactions $\to^{k_p} \mathbf{A} + \mathbf{A}$; $\mathbf{A} \to^{k_d}$ . (Top row). (Middle row) Plots of expectation and standard deviation of the respective species until time 1200, as estimated by the CLA. (Bottom row) Plots of the effect of each filter obtained from a single stochastic simulation until time 4000. *(A)*The linear filter (Eqn (6.3)), for $k_1 = 0.0064, k_2 = 0.0064$, introduces a delay to buffer stochastic fluctuations and reduce the noise, but only to Poisson levels. *(B)* For $r_1 = 0.005, r_2 = 0.00005$, the annihilation module (Eqn (6.10)) improves noise reduction performance but cannot proportionally follow changes in the input. In fact, Eqn (6.11) predicts that the expectation of the output of the annihilation module changes with the square root of input. Thus, changes on the average value of the input are attenuated in the output. *(C)* For $r_1 = 100, r_2 = 1000, r_3 = 0.000055$, the annihilation filter (Eqn (6.16)) not only improves the noise reduction capabilities compared to the other modules, but also proportionally follows changes in the input.

**Theorem 10.** *Let $\mathcal{C} = \mathcal{F}_A \cup \mathcal{A}$ be the CRN union of CRNs $\mathcal{F}_A$ and $\mathcal{A}$. Assume $\mathbb{E}_{\lambda_B}(0) = \mathbb{E}_{\lambda_E}(0)$. Then, it holds that*

$$\mathbb{E}^{\infty}_{\mathbf{A}} = \mathbb{E}^{\infty}_{\mathbf{O}}.$$

*Moreover, assume $r_2 = \frac{\gamma}{r}, r_3 = r$, where $\gamma, r$ are constants. Then,*

$$\lim_{r\to 0} F^{\infty}_{\mathbf{O}} = \frac{k_p}{k_p + r_1}.$$

*Proof.* Using the CLA equations for the mean (Eqn (4.6)) we obtain the following ODEs

$$\begin{aligned}
\frac{d\mathbb{E}_{\lambda_D}(t)}{dt} &= r_1\mathbb{E}_{\mathbf{A}}(t) - r_1\mathbb{E}_{\lambda_D}(t)\\
\frac{d\mathbb{E}_{\lambda_B}(t)}{dt} &= r_1\mathbb{E}_{\mathbf{A}}(t) - r_2\mathbb{E}_{\lambda_B}(t)\mathbb{E}_{\lambda_E}(t) - r_3\mathbb{E}_{\lambda_B}(t)\mathbb{E}_{\mathbf{A}}(t)\\
\frac{d\mathbb{E}_{\lambda_E}(t)}{dt} &= r_1\mathbb{E}_{\lambda_D}(t) - r_2\mathbb{E}_{\lambda_B}(t)\mathbb{E}_{\lambda_E}(t) - r_3\mathbb{E}_{\lambda_E}(t)\Phi_C(t)\\
\frac{d\mathbb{E}_{\mathbf{O}}(t)}{dt} &= r_3\mathbb{E}_{\lambda_B}(t)\mathbb{E}_{\mathbf{A}}(t) - r_3\mathbb{E}_{\lambda_E}(t)\mathbb{E}_{\mathbf{O}}(t).
\end{aligned}$$

From these equations we can derive the following relations at steady state

$$\mathbb{E}^{\infty}_{\mathbf{A}} = \mathbb{E}^{\infty}_{\lambda_D} \tag{6.17}$$

$$\mathbb{E}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\lambda_D}\frac{\mathbb{E}^{\infty}_{\lambda_B}}{\mathbb{E}^{\infty}_{\lambda_E}} \tag{6.18}$$

$$0 = r_1\mathbb{E}^{\infty}_{\lambda_D} - r_2\mathbb{E}^{\infty}_{\lambda_B}\mathbb{E}^{\infty}_{\lambda_E} - r_3\mathbb{E}^{\infty}_{\lambda_E}\mathbb{E}^{\infty}_{\mathbf{O}}. \tag{6.19}$$

It is easy to see that we have an infinite number of possible solutions depending on the initial conditions. By integrating the differential equations for $\mathbb{E}_{\lambda_B}(t)$ and $\mathbb{E}_{\lambda_E}(t)$, we get

$$\begin{aligned}
\mathbb{E}_{\lambda_B}(t) - \mathbb{E}_{\lambda_E}(t) =&(\mathbb{E}_{\lambda_B}(0) - \mathbb{E}_{\lambda_E}(0)) + \int_0^t r_1\mathbb{E}_{\mathbf{A}}(s) - r_1\mathbb{E}_{\lambda_D}(s)ds -\\
&\int_0^t r_3\mathbb{E}_{\lambda_B}(s)\mathbb{E}_{\mathbf{A}}(s) - r_3\mathbb{E}_{\lambda_E}(t)\mathbb{E}_{\mathbf{O}}(s)ds,
\end{aligned}$$

which, assuming $\mathbb{E}_{\lambda_B}(0) = \mathbb{E}_{\lambda_E}(0)$, leads to the following relation

$$\mathbb{E}_{\lambda_B}(t) - \mathbb{E}_{\lambda_E}(t) = \mathbb{E}_{\lambda_B}(0) - \mathbb{E}_{\lambda_E}(0) + \mathbb{E}_{\lambda_D}(t) - \mathbb{E}_{\mathbf{O}.}(t) \tag{6.20}$$

By solving Eqn (6.20), Eqn (6.17), Eqn (6.18), and Eqn (6.19) at steady state for $E[B(0)] - E[E(0)] = 0$, we obtain that $\mathbb{E}^{\infty}_{\lambda_D} = \mathbb{E}^{\infty}_{\mathbf{O}} = \mathbb{E}^{\infty}_{\mathbf{A}}$.

Now, from the moment equations (Eqn (4.3)), we get the following exact expression for $\mathbb{E}_{\mathbf{O}}(t)$,

$$\frac{d\mathbb{E}_{\mathbf{O}}(t)}{dt} = r_3\mathbb{E}[\mathcal{X}_{\lambda_B}(t)\mathcal{X}_{\mathbf{A}}(t)] - r_3\mathbb{E}[\mathcal{X}_{\lambda_E}(t)\mathcal{X}_{\mathbf{O}}(t)]],$$

where $\mathbb{E}[\mathcal{X}_{\lambda_B}(t)\mathcal{X}_{\mathbf{A}}(t)]$ and $\mathbb{E}[\mathcal{X}_{\lambda_E}(t)\mathcal{X}_{\mathbf{O}}(t)]]$ are the expectation of the components of $\mathcal{X}(t)$ relative to respectively species $\lambda_B$ and $\mathbf{A}$ and $\lambda_E$ and $\mathbf{O}$. The above relation, at steady state, for $\mathbb{E}_{\lambda_D}(0) = \mathbb{E}_{\mathbf{O}}(0) = \mathbb{E}_{\lambda_B}(0) = \mathbb{E}_{\lambda_E}(0) = 0$, togheter with Eqn (6.18), leads to

$$Cov[B, A]_\infty = Cov[E, C]_\infty. \tag{6.21}$$

Solving (6.21) with Equations (4.7) for CRN $\mathcal{F}_A$, we can derive a closed form for the Fano factor of $\mathbf{O}$ at steady state. However, the resulting expression is too complex to be directly useful. Nevertheless, we can assume $r_2 = \frac{T}{r}, r_3 = r$, where $T, r$ are constants. Then, for $r \to 0$, we obtain

$$\lim_{r\to 0} F_{\mathbf{O}} = \frac{k_p}{k_p + r_1}.$$

□

**Corollary 10.1.** Under the assumptions of Theorem 10 it holds that

$$\lim_{r_1\to\infty}(\lim_{r\to 0} F^\infty_{\mathbf{O}}) = 0. \tag{6.22}$$

Thus, Theorem 10 guarantees that, under appropriate scaling of the rate constants, $\mathcal{F}_A$ can always work as a molecular filter. Moreover, increasing $r_1$, and for $r_2$ large compared to $r_3$, the noise can be made arbitrarily small, showing how this architecture has ideal noise reduction capabilities: independently of the intrinsic noise introduced the total noise can be made arbitrarily close to 0.

Note that in the CRN (6.16) it is assumed that some reactions have same rates. This assumption allows us to obtain simpler analytic results. In order to show that the above analysis remains valid also in the more general scenario where the reactions have different rates, we consider the following CRN, $\mathcal{F}'_A$, modifying $\mathcal{F}_A$

$$\begin{aligned} &\mathbf{A} \to^{r_{1,1}} \mathbf{A} + \lambda_B + \lambda_D \quad \lambda_D \to^{r_{1,2}} \lambda_E \quad \lambda_B + \lambda_E \to^{r_2} \\ &\lambda_B + \mathbf{A} \to^{r_{3,1}} \mathbf{A} + \mathbf{O} \quad \lambda_E + \mathbf{O} \to^{r_{3,2}} \end{aligned} \tag{6.23}$$

In Figure 6.8 we compute $F^\infty_{\mathbf{O}}$ and $\mathbb{E}^\infty_{\mathbf{O}}$ for different values of $r_{1,1}, r_{1,2}$ and $r_{3,1}, r_{3,2}$. It is easy to observe that Eqn (6.22) is confirmed: if $r_{1,1}, r_{1,2}$ are big enough compared

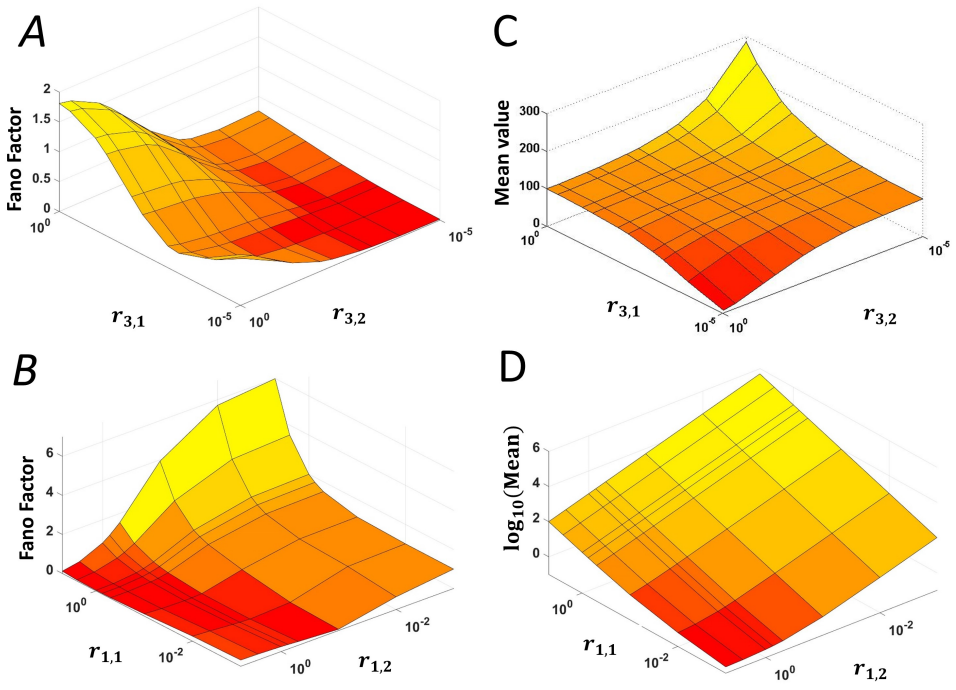



Figure 6.8: For the CRN (6.23), the figure shows how the Fano factor of $\mathbf{O}$,$F^{\infty}_{\mathbf{O}}$, and the mean value of $C$ at steady state, $\mathbb{E}^{\infty}_{\mathbf{O}}$ depends on $r_{1,1}, r_{1,2}$ and $r_{3,1}, r_{3,2}$. For all figures we consider an input species $\mathbf{A}$ such that $\mathbb{E}^{\infty}_{\mathbf{A}} = 100$ and $\mathbf{A}$ affected by Poisson noise. It is possible to observe that, as predicted by the theoretical analysis, $F^{\infty}_{\mathbf{O}}$ tends to decrease either when both $r_{1,1}, r_{1,2}$ increase or when both $r_{3,1}, r_{3,2}$ decrease. Interestingly, note that $F^{\infty}_{\mathbf{O}}$ tends to be smaller when there is a strong suppression of the mean. Thus, when a low number of molecules is involved, the noise has more influence on the behaviour of the system. *(A,C)* We plot $F^{\infty}_{\mathbf{O}}$ and $\mathbb{E}^{\infty}_{\mathbf{O}}$ for $r_{1,1} = r_{1,2} = 10$ and $r_2 = 100$. The non-linearity of the reactions involved is such that $F^{\infty}_{\mathbf{O}}$ and $\mathbb{E}^{\infty}_{\mathbf{O}}$ are robust with respect to parameters variation. *(B,D)* We plot $F^{\infty}_{\mathbf{O}}$ and $\mathbb{E}^{\infty}_{\mathbf{O}}$ for $r_{3,1} = r_{3,2} = 0.001$ and $r_2 = 100$. In this case, $\mathbb{E}^{\infty}_{\mathbf{O}}$ changes linearly with the rates. $F_C$ tends to increase when there is a strong amplification of $\mathbb{E}^{\infty}_{\mathbf{O}}$.

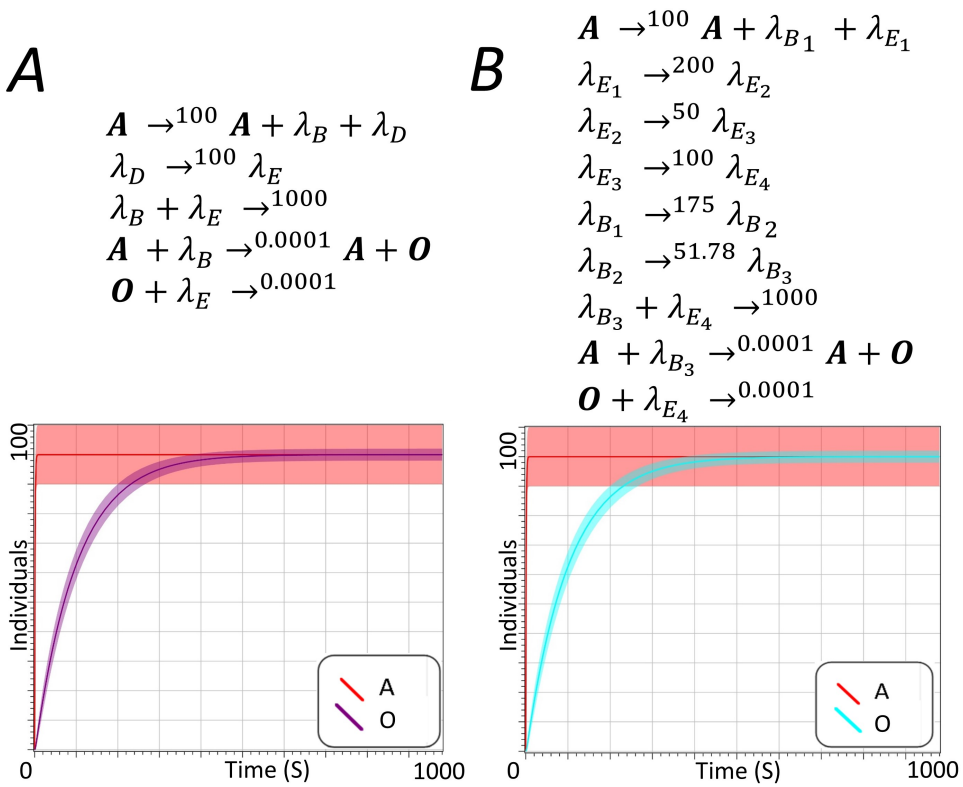



Figure 6.9: $(A)$ Annihilation filter for an input process $A$ affected by Poisson noise. $(B)$ Generalization of the annihilation filter where co-expression of molecules requires more intermediate steps. Reaction rates are chosen such $\mathbb{E}_{\mathbf{A}}^{\infty} = \mathbb{E}_{\mathbf{O}}^{\infty}$. We can observe that the Fano factor of the output remains constant for both networks.

to $r_{3,1}, r_{3,2}$ the Fano factor will decrease converging to a value of 0. Nevertheless, if $r_{1,1} \neq r_{1,2}$ or $r_{3,1} \neq r_{3,2}$, the noise will still be reduced, but this will affect $\mathbb{E}_{\mathbf{O}}^{\infty}$ , which may be different from $\mathbb{E}_{\mathbf{A}}^{\infty}$.

The first two reactions of the annihilation filter (Eqn (6.16)) can be thought as a model for co-expression of molecules $\lambda_B$ and $\lambda_E$ at different rates. Such a model of co-expression can be generalized, having $\lambda_B$ and $\lambda_E$ co-expressed, and then interacting after a pathway of linear reactions. For instance, this is the case of mRNA and microRNA that, when co-expressed, undergo a series of maturation steps before of interacting [131, 24]. In Figure 6.9 we show that these auxiliary reactions do not influence the noise reduction capabilities of the annihilation filter.

## 6.5 Numerical Analysis

While the mathematical analysis performed on linear filters is exact and general, for the annihilation module and annihilation filter, since they are non-linear, our claims

are based on the CLA and for birth-death input processes. Thus, we need to support our results about the noise suppression capabilities of such networks with stochastic simulations for different classes of input. In order to do that, in Figure 6.7, we consider a step-like perturbation, and in Figure 6.10 we consider oscillatory inputs. In both cases, the annihilation filter outperforms the other filters: for the same delay introduced it suppresses more noise at the high frequencies while still maintaining similar long-term behaviour. Instead, the annihilation module does not follow the long term behaviour of the input proportionally (it works as a high-pass filter). In Figure 6.10, the Fano factor of **A** is due to two components: phase and amplitude of the oscillations, which may change at any oscillation period, and the fast fluctuations that corrupt the oscillatory behaviour. In this case, the goal of the filtering process should be to produce an output process with the same oscillatory behaviour of **A**, but with reduced fast fluctuations. Therefore, the annihilation module cannot perform well for such a goal. In fact, we showed that the annihilation module reduces the low frequency components of the spectrum of **A**. This implies that the output of the annihilation module will have attenuated oscillations compared to **A**. The annihilation filter, instead, correctly reduces the fast fluctuations while still maintaining similar profile to the oscillations and not introducing any other noise, as it happens for the linear filter.

To further confirm the mathematical analysis, using stochastic simulations we compare the power spectral density (PSD) (see Section 3.2) of the input species affected by Poisson noise with the PSD of the output species of the annihilation filter ($Eqn(6.16)$), and of the linear filter ($Eqn(6.3)$), both for the same input (see Figure 6.11$A$). We can see that both filters are indeed low-pass filters in the sense that they attenuate the high frequencies. However, although they behave similarly at the high-frequencies, the linear filter is less robust to intrinsic noise, and such intrinsic noise amplifies the low frequencies, resulting in noise reduction being lower bounded by Poisson dynamics. That is, the reactions of the linear filter introduce slow and medium time variations of the output leading to an amplification of the low frequency components of its spectrum. The annihilation filter enables a much better reduction of the intrinsic noise, leading to a smaller amplification of the low frequencies. The connection between the noise of a process and its PSD is explained in detail in the Section 3.2.

In Figure 6.11$B$, we consider again input species **A** affected by Poisson noise, and, on this input, we compare the action of the annihilation filter (Eqn (6.16)), annihilation module (Eqn (6.10)), and linear filter (Eqn (6.3)). In order to reduce the

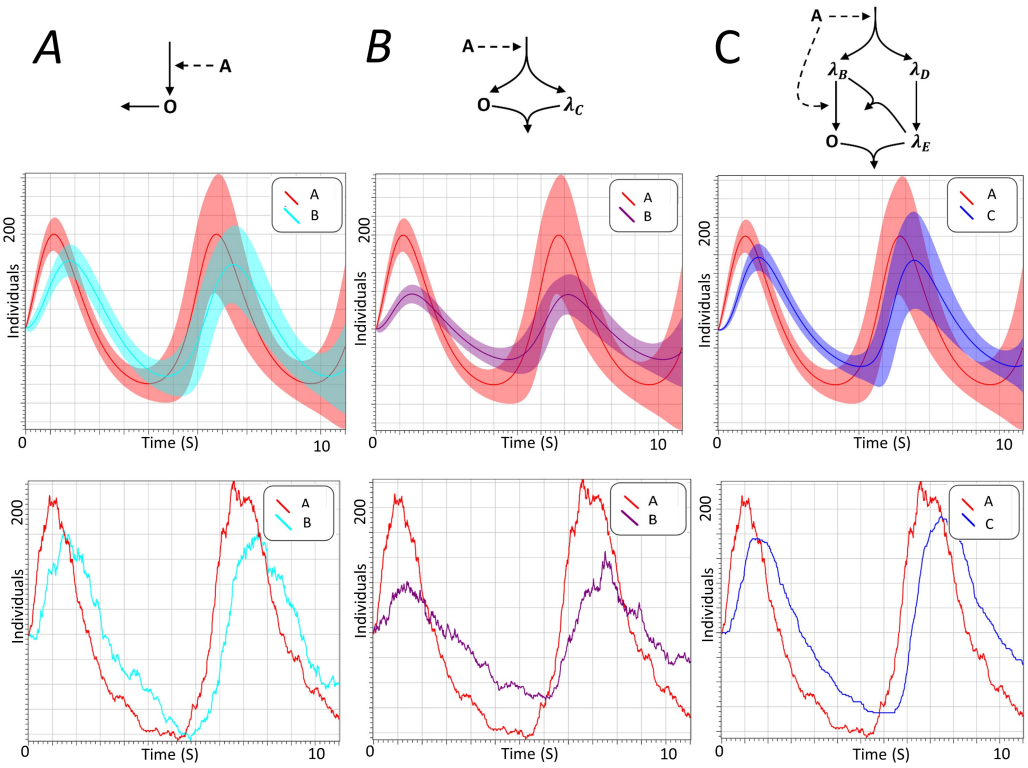



Figure 6.10: Comparison of linear filter for $k_1 = 0.0064, k_2 = 0.0064$ $(A)$, annihilation module for $r_1 = 0.005, r_2 = 0.00005$ *(B)*, and annihilation filter for $r_1 = 100, r_2 = 1000, r_3 = 0.000055$ *(C)* for an oscillatory input **A**. In the second row we plot the expectation and standard deviation of input and output of each filter according to the CLA. In the third row we show a stochastic simulation of input and output of each filter.

number of free variables in the system, we constrain the output of the filters to have the same mean as **A**. Then, we plot the Fano factor as a function of the remaining free rate parameters. As expected, our key observations are that the annihilation module is the only mechanism that guarantees noise reduction for any value of the parameter rate, confirming the theoretical result of Eqn (6.12) that it is able to reduce the noise even without introducing a delay in its buffering action. Also, for an arbitrarily long delay, the annihilation filter converges to a Fano factor of 0, showing the ability of this network to remove all the noise (variance tends to 0 when delay tends to infinity). On the other hand, the linear filter converges to a Fano factor of 1, corresponding to Poisson levels, thus confirming Eqn (6.8).

## 6.6 Molecular filters are widespread in gene expression

To motivate the study of molecular filters, we show how these are widespread in gene expression and have a role in increasing the robustness of the system.

### 6.6.1 Gene and protein expression can work as linear filters

Gene expression is often modeled as a two or three-stage process, where mRNA is transcribed from a transcription factor TF, and the protein P is translated from the mRNA [148]

$$TF \rightarrow^{k_m} TF + mRNA; \qquad mRNA \rightarrow^{d_m};$$
$$mRNA \rightarrow^{k_P} mRNA + P; \qquad P \rightarrow^{d_P};$$

Under this modeling assumption, the linear filtering mechanism (CRN (6.3)) may be present both at the transcription (top two reactions) and translation level (bottom two reactions) to buffer noise and increase robustness by slowing down transcription or translation. At the transcription level, inefficient transcription that follows fast promoter activation is a mechanism that has been widely observed to buffer fluctuations in the mRNA time evolution [137]. At the translation level, the linear filtering mechanism reduces the noise in protein expression, as supported by experimental evidence on B. subtilis and S. cerevisiae [129, 32]. Eqn (6.6) guarantees that, independently of the presence of arbitrarily many feed-forward or feedback loops between $P$ and transcription factor or mRNA, we have

$$\mathbb{V}_P^\infty = \mathbb{E}_P^\infty + \frac{k_P}{d_P} Cov[\text{mRNA,P}]_\infty. \tag{6.24}$$

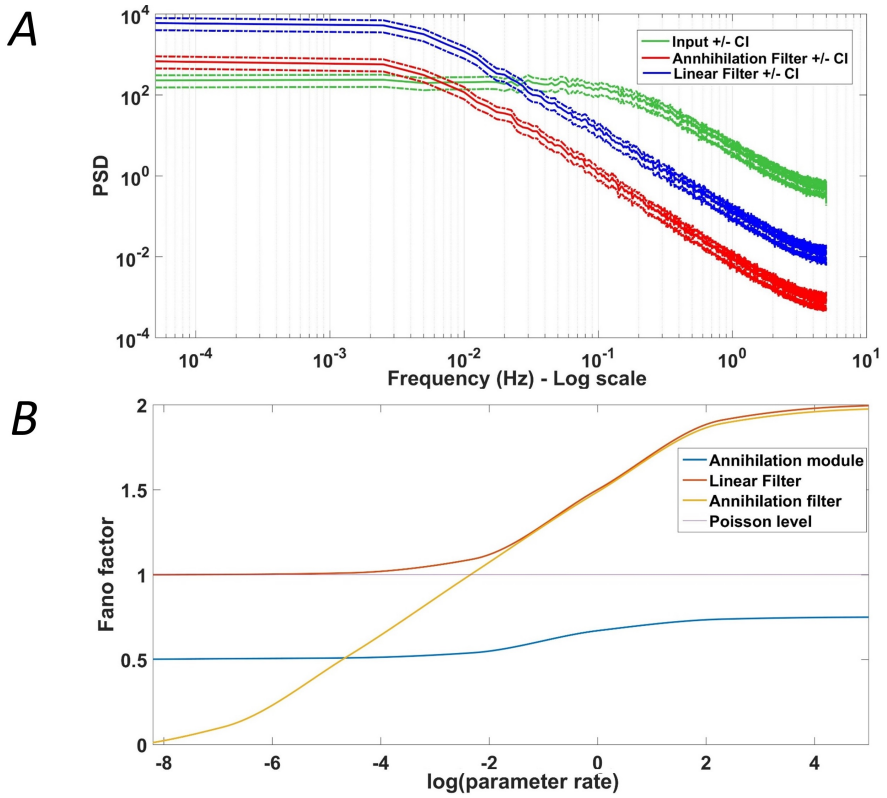



Figure 6.11: ($A$) Power Spectral Density (PSD) of input affected by Poisson noise (green), compared with the output of the linear filter (Eqn (6.3)) applied on such input (blue), and of the annihilation filter (Eqn (6.16)) (red). For the linear filter we fix the following parameters $k_1 = k_2 = 0.008$, while for the annihilation filter we have $r_1 = 1, r_2 = 10, r_3 = 0.00008$. Poisson input is generated by the following reactions: $\rightarrow^{100} Input$; $Input \rightarrow^{1}$. The power spectrum is estimated using the Blackman-Tukey Spectral Estimate algorithm with frequency deep resolution [156], as implemented in Matlab, over 300000 data points with sampling time of 0.1 seconds. For each system, we plot the 99% confidence interval (CI). ($B$) Fano factor of annihilation module (blue), annihilation filter (yellow), and linear filter (red) for an input affected by Poisson noise (dotted line). Outputs of the filters are constrained to have the same expectation as the input species at steady state. Plots are of the function of the remaining free rate parameters. Fano factor is estimated by means of the CLA.

Thus, if mRNA and protein are positively correlated, then the stochastic fluctuations in protein expression cannot be reduced below Poisson levels, and this limit is approached for slow translation. Hence, Eqn (6.24) may explain why in yeast and E. coli the translation rates tend to be slower than the transcription rates [71, 137], and also suggests that mechanisms to induce a negative correlation between protein and mRNA may have been selected to enhance robustness. This is the case for negative feedback, which may thus enable noise reduction below Poisson levels, as again confirmed by experimental evidence [28]. We note that a more realistic model of gene expression requires representation of transcription and translation as multi-step processes. However, since such processes can still be modeled as a sequence of first-order reactions [131], our analysis and the linear filter mechanism still applies.

### 6.6.2 MicroRNAs can serve as annihilation filters

Slow translation/transcription is a very simple mechanism of noise reduction. Since gene expression involves low molecular counts and highly stochastic signals, different (and more complex) network architectures may have been selected to deal with scenarios where greater noise reduction is needed [94]. A simple example can be found in microRNA-regulated post-transcriptional regulation. MicroRNAs (miRNAs) are short RNAs that are widely conserved in biological networks [25]. In animals, it is common that miRNAs and their target mRNAs are co-expressed or positively correlated in their expression [139, 66, 133]. For example, c-myc induces the expression of the microRNAs miR-17-5p and miR-20a together with their target E2F1 [128]. Furthermore, the system where miRNAs repress gene expression by binding with the target mRNAs, and either inhibiting translation of mRNA or promoting mRNA decay [6, 102], leads to a pattern that can be modelled with the annihilation module (Figure 6.1*B*). Although it is well accepted that microRNAs confer robustness on gene expression [66, 144, 98], it is still not clear what aspects of their inhibitory mechanisms are used to gain efficient noise reduction [20], and previous analysis has focused on microRNAs that are not co-expressed with the target proteins [144].. One hypothesis is that co-expression of microRNAs with their targets has a role in increasing robustness of gene expression [94]. This is also supported by experimental evidence [157]. Our mathematical analysis confirms such a hypothesis, and shows that correlated expression of microRNAs and mRNA, followed by translational repression of the mRNA when bound to the microRNA, may lead to noise reduction below Poisson levels. This result suggests that microRNA regulation may have been selected to post-regulate highly noisy genes. One specific example of such a pattern can be

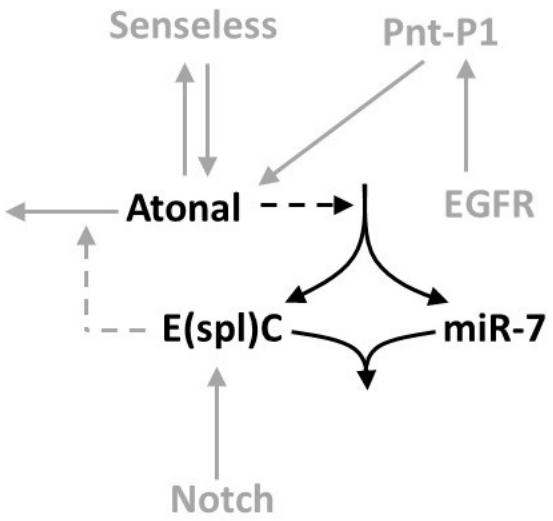



Figure 6.12: The network controlling sensory organ precursor (SOP) fate [113]. miR-7 participates in an annhilation module highlighted in black. The annhilation module is also interconnected with a double-negative feedback loop between Ato and E(spl), with miR-7 as an effector of Ato, and E(spl) directly inhibiting Ato.

found in the *Drosophila eye*, where miR-7 and its target protein are co-expressed, and experimental studies have suggested the role of miR-7 in buffering fluctuations [113, 133]. A particular network involving miR-7 and verified experimentally in [113] is shown in Figure 6.12. This network is responsible for sensory organ precursor (SOP) fate. An annihilation module between Ato, miR-7, and E(slp) genes may have a role in increasing robustness. We stress that it has yet to be verified experimentally that miR-7 works as an annihilation filter, especially due to the lack of experiments concerning the joint degradation of micro-RNA and mRNA in different organisms [15].

We have also found that the annihilation filter (Eqn (6.16)) may be a suitable model for the c-Myc/E2F1/hsa-miR-20a circuit, with its extension to E2F2 in the case when E2F1 activates E2F2 [139, 128] (Figure 6.1$C$). This may suggest that miR-20a, by repressing both E2F1 and E2F2, confers greater robustness on E2F2 expression. We note, however, that there is no experimental evidence yet for whether E2F1 activates or represses E2F2. Note that the annihilation filter link between A (c-Myc) and C (E2F2) is not present here: its role is to regulate the mean of the output, but it has no influence on noise reduction.

## 6.7 Related Molecular Motifs

Noise reduction has been studied extensively in electronics [146] and certain noise filtering principles have been successfully applied to molecular systems [141, 45, 169]. For instance, negative feedback and feed-forward loops have been shown to reduce noise [117, 28], and fundamental limits for noise suppression of feedback loops have been derived using techniques from control and information theory [112].

Specific models of linear filters have already been studied in the context of closed models of gene expression [158, 124]. $\mathcal{F}_L$, the linear filter, generalizes such a networks taking also into account the more general and common case where the dynamics of the other components of the networks are left unspecified, and possibly involving non-linearity and feedback loops. Thus, our analysis, while being exact and based on the CME, derives bounds that hold for the rest of the network left unspecified.

The fact that linear filters are limited by Poisson noise (Eqn (6.8)) has already been observed and studied in the context of specific closed models of gene expression for which mRNA and protein are positively correlated [131, 158, 124]. However, in the more general scenario we consider, we can observe that Eqn (6.8) holds only in the case of non-negative correlation of the species, meaning that a simple birth-death process of a downstream component cannot reduce the noise of an input signal below Poisson levels. Thus, our analysis gives a further explanation of why negative feedback regulation in gene expression is a widely selected mechanism to reduce noise and increase robustness. Such evidence has been confirmed by experiments [28].

The annihilation module is closely related to the incoherent Feed Forward Loop motif (iFFL) [117], where two species are co-expressed and one inhibits the other. However, it is possible to show that an iFFL with mass action kinetics cannot reduce the noise below Poisson levels [52], hence, showing how the non-linear degradation of the species is essential for efficient noise reduction.

The annihilation module and the annihilation filter (see next section) are also related to the *antithetical integral feedback motif* [45]. The main similarity lies in the fact that all these modules have an annihilation reaction, thus suggesting a key role for such a reaction in dealing with noisy dynamics. However, we stress here how the mechanisms differ: a key requirement of both our filters for efficient noise reduction is the co-expression of the molecules that will participate in the annihilation reaction. This requirement cannot be implemented in the antithetical integral feedback schema. Moreover, the antithetical integral feedback motif is known to increase the noise of the controlled network [46]. In fact, we argue that one of the reasons why in the

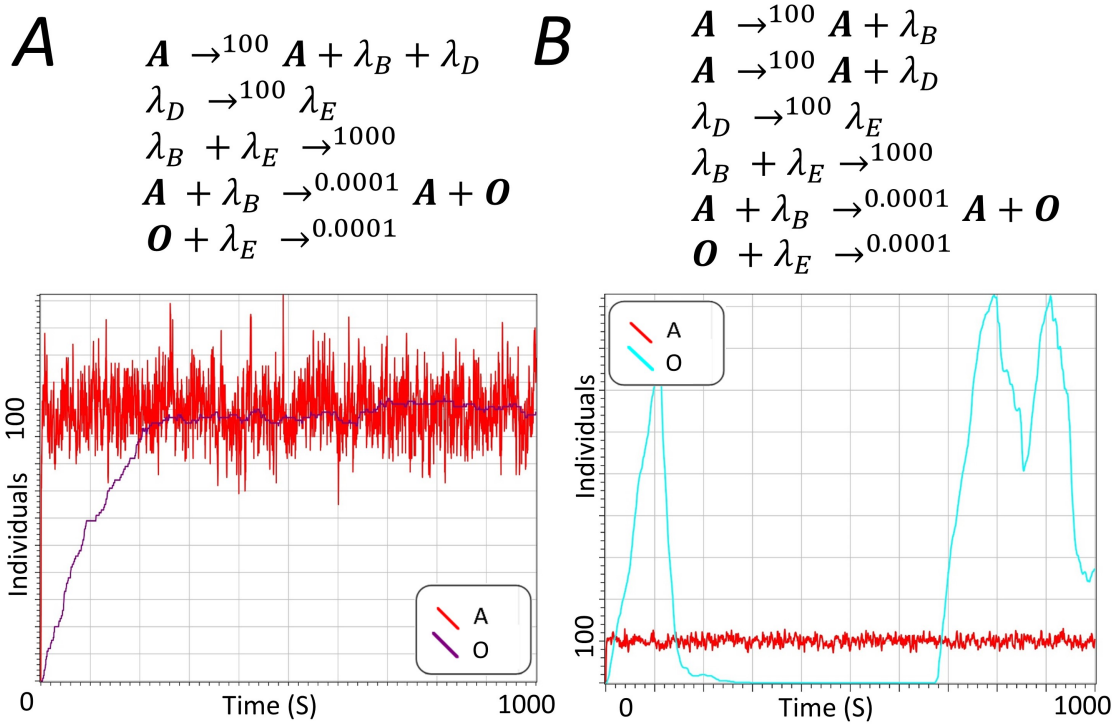



Figure 6.13: We consider two networks with the same rate equations that differ only in how $\lambda_B$ and $\lambda_D$ are expressed. For both figures we plot a stochastic simulation of an input signal **A** affected by Poisson noise with the output of the filter. $(A)$ considers the annihilation filter (Eqn (6.16)). $(B)$ considers a network that is identical to the one in Figure A, except for the fact that $\lambda_B$ and $\lambda_D$ are not co-expressed.

antithetical integral feedback the noise increases is that the species that undergo an annihilation reaction are not co-expressed. In order to illustrate this point, in Figure 6.13, we compare two networks: one is the annihilation filter, the other is identical to the annihilation filter except for the fact that B and D are not co-expressed, but only positively correlated in their expression. The two networks, deterministically, behave identically (they have same rate equations). However, interestingly, the stochastic behavior is completely different, thus demonstrating the importance of co-expression.

## 6.8 Summary

In this chapter we formalized the notion of molecular filters, that is, CRNs that are able to reduce the noise of a molecular signal while still maintaining certain features

of its time evolution. We introduced three different molecular filters and derived their properties and limitations. To motivate the study of molecular filters, we related the introduced filters to biological mechanisms, and show how these can give new insights about the fundamental principles about how natural systems perform their functions. We show, for example, that gene expression with slow translation/transcription can reduce noise, but only down to Poisson levels (Eqn (6.24)), even when considering linear multiple-step models [131]. In contrast, the annihilation module (Eqn (6.10)), which relies on co-expression and joint degradation, can drive the noise below Poisson levels. Such mechanisms can be seen in microRNA regulated systems, where a key requirement for effective noise reduction is the co-expression of mRNA with its microRNAs.

# Chapter 7

# Conclusions

In this thesis we focused on noise in biochemical systems. In order to enable automated probabilistic analysis of stochastic CRNs we presented a framework for approximate model checking of a time-bounded fragment of CSL extended with rewards. Our approach employed an abstraction based on the CLA to approximate the CTMC induced by a CRN as a Gaussian process. Then, numerical procedures for model checking CSL formulae on the resulting Gaussian process were derived. We demonstrated that our approach does not suffer from the state space explosion problem. Thus, it enables formal analysis of CTMCs that cannot be analyzed using classical methods based on uniformization and with infinite state space [62, 167]. The developed techniques can then be used to reverse engineer biochemical systems and study how natural systems deal with noise. While biological systems deal with noise in a variety of ways, in this thesis we focused on scenarios where noise should be controlled. We introduced the concept of molecular filter, and studied three different molecular filters, which we showed are used to reduce the noise at the molecular level. To derive our results we employed mathematical analysis and model checking algorithms based on the CLA. It remains an interesting endeavor to similarly discover and analyze the basic principles that allow biological systems to exploit noise functionally and use it to their benefit [67]. We believe that a systematic analysis of noise reduction in molecular systems, together with evidence of widespread noise reduction capabilities in biological systems, such as in gene expression, are fundamental to obtain new insights into the structure and evolutionary origin of noise reduction mechanisms.

## 7.1 Future Work

Several directions for future research arise from the results of this thesis. In this section, we highlight some directions of particular relevance.

**Gaussian Processes Model Checking.** The techniques developed in Chapter 5 are general and can be employed to perform formal analysis of Gaussian processes and Gaussian Random Fields (GRFs) [3]. As future work we target computation of probabilistic safety (the probability that the output of the system changes less than a given threshold for small variations of the input) for these processes. Such a property can be computed locally, without the need to rely on numerical approaches, and could be used to guarantee the safety of those applications where Gaussian processes are widely employed, such as cyber-physical systems and machine learning [138].

**Spatial Molecular Filters.** In both unicellular and multicellular organisms the well-mixed assumption often does not hold. Thus, there is a need to explicitly model the space location of the molecules in the system, taking also into account how molecules may diffuse in different locations. A mathematical model that takes into account all of these components may have a very different behaviour compared to one where space is not considered. For instance, spatial patterns can form and it has been shown that a diffusion component between cells can greatly influence the noise at the single cell level [153]. The underlying principles of how noise is controlled are still unknown. To fill this gap, we plan to develop a theory for molecular filters including also a spatial component in the mathematical model underlying a CRN. The resulting filters would represent a natural extension of the results in Chapter 6.

**Experimental Evaluation.** We plan to implement the molecular circuits derived in Chapter 6 in synthetic circuits or living cells. In particular, the biological analysis performed in Chapter 6 suggests that microRNA can work as an annihilation module and filter, and can further infer robustness to gene expression in case of co-expression with its target gene and joint degradation. To validate this hypothesis we plan to collaborate with experimental biologists to perform experiments in synthetic organisms and in living cells.

# Chapter 8

# Proofs

## 8.1 Proofs of Chapter 5

**Theorem 6** Let $\mathcal{C} = (\Lambda, R)$ be a CRN with induced CTMC $\hat{\mathcal{X}}^N$ and $\hat{\mathcal{Z}}^{\Delta z,h,N}$ be the DTMC obtained by space and time discretization of $B\hat{\mathcal{Y}}^N$. Assume $\hat{\mathcal{X}}^N(0) = \hat{x}_0$ and the corresponding initial state for $\hat{\mathcal{Z}}^{\Delta z,h,N}$ is $z_{d,0}$. Then, for $t_1, t_2 \in \mathbb{R}_{\geq 0}$,$B \in \mathbb{R}^{m\times|\Lambda|}$ and $b \in \mathbb{R}^m$, and $A = \{x \in \mathbb{R}^{|\Lambda|}_{\geq 0}\, s.t.\, \forall i \in \{1, ..., m\}(Bx)_i \leq b_i\}$, it holds that:

$$\lim_{N\to\infty}\lim_{h\to 0}\lim_{\Delta z\to 0} |P^A_{reach}(\hat{x}_0, t_1, t_2) - P^{\Delta z,h,A}_{reach}(z_{d,0}, t_1, t_2)| = 0.$$

**Proof.** Without any loss of generality, we assume $t_1 = 0, t_2 = T$. Call

$$P^{h,A}_{reach}(\hat{x}_0, 0, T) = Pr^h(\exists k \in [0, \lceil\frac{T}{h}\rceil]\, s.t.\, \hat{\mathcal{Z}}^{h,N}(k) \leq b \,|\, \hat{\mathcal{Z}}^{h,N}(0) = B\hat{x}_0),$$

and

$$P^{\hat{\mathcal{Y}}^N,A}_{reach}(\hat{x}_0, 0, T) = Pr^{\hat{\mathcal{Y}}^N}(\exists t \in [0, T]\, s.t.\, \hat{\mathcal{Y}}^N(t) \in A \,|\, \hat{\mathcal{Y}}^N(0) = \hat{x}_0),$$

where $Pr^{\hat{\mathcal{Y}}^N}$ is the probability measure of the Gaussian process $\hat{\mathcal{Y}}^N$ as defined in Section 3.1.4.

By application of the triangular inequality we have that:

$$\begin{aligned}|P^A_{reach}(\hat{x}_0, 0, T) - P^{\Delta z,h,A}_{reach}(z_{d,0}, 0, T)| \leq \\ |P^A_{reach}(\hat{x}_0, 0, T) - P^{\hat{\mathcal{Y}}^N,A}_{reach}(\hat{x}_0, 0, T)| + \\ |P^{\hat{\mathcal{Y}}^N,A}_{reach}(\hat{x}_0, 0, T) - P^{h,A}_{reach}(\hat{x}_0, 0, T)| + \\ |P^{h,A}_{reach}(\hat{x}_0, 0, T) - P^{\Delta z,h,A}_{reach}(z_{d,0}, 0, T)|.\end{aligned}$$

The convergence of the third and second components is a consequence of Theorems 5 and 4. We need to show that:

$$\lim_{N\to\infty} |P^{A}_{reach}(\hat{x}_0, 0, T) - P^{\hat{\mathcal{Y}}^N, A}_{reach}(\hat{x}_0, 0, T)| = 0.$$

Note that we removed the limits for $\Delta z$ and $h$, as this term is independent of time and space discretization. In what follows we assume that $B\mathcal{X}^N$ is a uni-dimensional process. Generalization for $m > 1$ follows from this case. Intuitively, this holds due to the convergence of $\mathcal{X}^N$ to its CLA $\mathcal{Y}^N$. A formal proof requires a more involved machinery. In fact, Theorem 4.8 states that:

$$\sqrt{N}\left(\hat{\mathcal{X}}^N(t) - \Phi(t)\right) \Rightarrow G(t),$$

hence, to rely on it, we need to reason on the modified stochastic model:

$$G^N(t) = \sqrt{N}\left(\hat{X}_N(t) - \Phi(t)\right),$$

rather than on the original CTMC $\hat{\mathcal{X}}^N(t)$. Now, consider the reachability objective $B\hat{\mathcal{X}}^N \leq b$; rephrasing it in terms of $G^N$ we get:

$$B\hat{\mathcal{X}}^N(t) \leq b \text{ iff } BG^N(t) \leq \sqrt{N}(b - B\Phi(t)) = b^N(t).$$

As we can see, the reachability problem for $G^N$ has a different nature: the threshold $b$ becomes both $N$ dependent and time dependent! In addition, we see that for the CLA, $B\mathcal{Y}^N(t) \leq b$ iff $BG(t) \leq b^N(t)$. Let's look at this reachability problem from the point of view of the trajectory space, i.e. the space of cadlag function $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}$. Both $G^N$ and $G$ can be seen as probability measures over this space. The reachable set in the trajectory space depends on $N$, precisely being $R_N = \{\omega \mid \exists t \in [0,T] : \omega(t) \leq b^N(t)\}$. We also consider the complement of this set, $R^c_N = \{\omega \mid \forall t \in [0,T] : \omega(t) > b^N(t)\}$, and the boundary of the set $\partial R_N = \{\omega \mid \forall t \in [0,T] : \omega(t) \geq b^N(t) \wedge \exists t \in [0,T] : \omega(t) = b^N(t)\}$.

Before proceeding further, we need to understand how the set $R_N$ changes as $N$ goes to infinity. Consider the threshold $b^N(t) = \sqrt{N}(b - B\Phi(t))$. There are three cases:

1. if $b > B\Phi(t)$, then $b^N(t) \to +\infty$;
2. if $b < B\Phi(t)$, then $b^N(t) \to -\infty$;
3. if $b = B\Phi(t)$, then $b^N(t) = 0$.

In the first case, the reachable set at time $t$ converges to $\mathbb{R}$, in the second case to the empty set, in the third case to $(-\infty, 0]$. Therefore, the limit reachable set $R$ in the trajectory space will be the union for each $t$ of one of these three kind of sets.

By the assumption that rate functions are real analytic, it follows that $\Phi(t)$ is also a real analytic function, and therefore $B\Phi(t)$ will equal $b$ only in a finite number of points of $[0, T]$, or in the whole interval (a degenerate case which is easily tractable) [101]. It then follows that $b(t) = \lim_{N\to\infty} b^N(t)$ changes value a finite number of times, say at times $t_1, \ldots, t_n$, where it equals zero. Outside these points, it is either plus or minus infinity. The reachable set $R$, in the limit of infinite $N$, is thus a finite union of sets of the form $t_i \times (-\infty, 0]$ at times $t_i$ and either $\emptyset$ or $\mathbb{R}$ for each $t$ in between $t_{i-1}$ and $t_i$.

Now, if in such a union the set $(t_{i-1}, t_i) \times \mathbb{R}$ is present at least once, then the reachability probability in the limit equals exactly one. This is because any trajectory $\omega$ will enter the set $R$ in that subregion. In this case, convergence is easily shown. In fact, being the Skorokhod space a Polish space, any converging sequence $G^N \Rightarrow G$ of random variables in that space is uniformly tight, meaning that for each $\epsilon$ there is a compact space $K_\epsilon$ such that, outside it, all random variables and the limit have probability less than $\epsilon$. In particular, a compact set of trajectories is bounded in $[0, T]$ with respect to the sup norm [3], meaning that for each $\epsilon$ there is a $k_\epsilon > 0$ such that the probability that a trajectory $\omega$ has modulus $|\omega(t)| \leq k$ uniformly in $[0, T]$ is more than $1 - \epsilon$ for all $N$. Now, consider the time interval $(t_{i-1}, t_i)$ where the reachable set converges to $(t_{i-1}, t_i) \times \mathbb{R}$ in the limit. As the threshold $b^N(t)$ is an analytic function of $t$, removing a region of length $\Delta$ near $t_{i-1}$ and $t_i$ (i.e. restricting to $[t_{i-1} + \Delta, t_i - \Delta]$), we can find an $N_0$ such that, for $N > N_0$, $b^N(t)$ is greater than $k_\epsilon$ uniformly in $[t_{i-1} + \Delta, t_i - \Delta]$. Then the limit of the reachability for $G^N$ is greater than $1 - \epsilon$ for any epsilon, that is, it equals one. The case in which the limit region $R$ is the empty set for every $t$ is easily proved along the same lines.

The interesting case is the one in which there are some $t_i$'s where $b^N(t_i) = 0$ for all $N$, but it is always negative outside them, implying the reachable region $R$ converges to the empty set everywhere but in the $t_i$'s, where it equals $(-\infty, 0]$. This corresponds to the scenario in which the fluid limit $\Phi(t)$ is tangent to the reachable set, but never enters it, a scenario known to cause trouble in the use of mean field to estimate hitting times [34].

To deal with this last case, let us denote with $P^N$ the probability in the trajectory space for $BG^N$, and with $P$ the probability for $BG$.

As before, denote with $R_N$ the reachability set for $G^N$ and with $R$ the limit set, taking the threshold $b^N$ to infinity. We now introduce a set which over-approximates $R_N$ for $N$ large. This set is defined as follows: invoking uniform tightness, we fix a large value $k_\epsilon$ as before, so that trajectories of $BG^N$ and of $BG$ are contained in $[-k_\epsilon, k_\epsilon]$ with probability $1-\epsilon$, uniformly for $t \in [0,T]$. Furthermore, we consider points $t_i$ where $b^N(t_i)$ is zero, and take a small neighborhood $B_i^\Delta$ of width $\Delta$ around them. Define the set $R_\epsilon$ in the trajectory space as:

$$R_\epsilon = \{\omega(t) | \omega(t) \leq 0, \text{for } t \in B_i^\Delta, \omega(t) \leq -k_\epsilon \text{ elsewhere in } [0,T]\}.$$

By relying on the continuity of the set $R$ for $G$, we can choose $\Delta$ small enough so as to enforce that $|P(R_\epsilon) - P(R)| \leq \epsilon$. The continuity of $R$ for $G$ follows from the fact that $\omega \in R$ if and only if $\omega(t_i) \leq 0$ for $i = 1, \ldots, n$, i.e. $R$ is a finite dimensional projection on $t_i$'s. Therefore, its boundary is a set of topological dimension less than $n$ in $\mathbb{R}^n$, which has probability zero under the finite dimensional projection of $G$ on $t_i$'s (which is Gaussian). Now, using triangular inequality, we get:

$$\begin{aligned} |P^N(R_N) - P(R)| &\leq |P^N(R_N) - P^N(R)| + |P^N(R) - P(R)| \\ &\leq |P^N(R_\epsilon) - P^N(R)| + |P^N(R) - P(R)| \\ &\leq |P^N(R_\epsilon) - P(R_\epsilon)| \\ &+ |P(R) - P(R_\epsilon)| + 2|P^N(R) - P(R)|. \end{aligned}$$

The second inequality above follows from the monotonic behaviour of probability distributions, as for each $\Delta$ and $k_\epsilon$ there is an $N_0$ such that, for all $N \geq N_0$, $R \subset R_N \subset R_\epsilon$, hence $|P^N(R^N) - P^N(R)| \leq |P^N(R_\epsilon) - P^N(R)|$.

Furthermore, $|P^N(R) - P(R)| \to 0$, by the continuity of the set $Y$. In $R$, by virtue of Lemmas 1 and 2 below, if also follows that $|P^N(R_\epsilon) - P(R_\epsilon)| \to 0$, and hence:

$$\limsup_{N \to \infty} |P^N(R_N) - P(R)| \leq \epsilon,$$

which holds for any $\epsilon > 0$, allowing us to conclude that:

$$\lim_{N \to \infty} |P^N(R_N) - P(R)| = 0,$$

as desired.

**Lemma 1.** Let $b \in \mathbb{R}$, and consider the reachable set $R = \{\omega | \exists t : \omega(t) \leq b\}$. Then $P^N(R) \to P(R)$, with $P^N$, $P$ as above.

**Proof.** The boundary of the reachable set $R$ is the set of trajectories $\omega$ such that $\inf_{t\in[0,T]} \omega(t) = b$. In order to conclude, we need to show that this set has measure 0. As $G$ is a Gaussian process, assuming the covariance function is non-zero, we have that the distribution of the infimum (or equivalently the supremum) is absolutely continuous with respect to the Lebesgue measure [114], which implies that the set of trajectories for which $\inf_{t\in[0,T]} \omega(t) = b$ has measure 0. Hence, $R$ is a continuity set for $G$, which prove the thesis due to the Portmanteau theorem.

**Lemma 2.** Consider a reachable set $R$ defined by a piecewise constant threshold. Hence, fix $0 = t_1, \dots t_{n+1} = T \in [0, T]$, and $b_i \in \mathbb{R}$, for $i = 1, \dots, n$, and let $R = \{\omega | \exists i \in \{1, \dots, n\}, \exists t \in [t_i, t_{i+1}] : \omega(t) \leq b_i\}$. Then $P^N(R) \to P(R)$, with $P^N$, $P$ as above.

**Proof.** We proceed by induction on $j$, showing that $R$ is a continuity set for $G$. The case for $j = 1$ follows from Lemma 1 above. Suppose we proved it up to $j - 1$. Then, conditioned on an initial trajectory $\omega$ from time zero to $t_j$, with $\omega(t_j) = y$, $G$ restricted in $[t_j, t_{j+1}]$ is a Gaussian process, and we can apply Lemma 1 to show that the probability of $\partial R$, restricted in this time span, is zero. Now, the probability of $\partial R$ restricted to $[0, t_{j+1}]$ can be bounded by the sum of two terms. The first is the probability of $\partial R$ in $[0, t_j]$ , which is zero, the second is probability of $\partial R \cup R^c$ up to time $t_j$ times the probability of $\partial R$ in $[t_j, t_{j+1}]$, conditional on being in $\partial R \cup R^c$ up to time $t_j$. Also this second term is zero, as the conditional probability is zero for any initial trajectory $\omega$. The bound on the probability of $\partial R$ follows because any trajectory in $\partial R$ up to time $t_{j+1}$ is either touching $b_j$ between $[t_j, t_{j+1}]$ (second term), or before $t_j$ (first term). The second case overlaps with the first for all trajectories that touch the threshold both before $t_j$ and between $[t_j, t_{j+1}]$.

**Proposition 7.** For $T \in \mathbb{R}_{\geq 0}$ and $B \in \mathbb{R}^{m\times|\Lambda|}$ let $A$ be the set defined as $A = \{x \in \mathbb{R}^{|\Lambda|} s.t. \forall i \in \{1, .., m\}, (Bx)_i \leq b_i\}$. Then, for $\hat{x}_0 \in \mathbb{R}^{|\Lambda|}$ and $z_{d,0}$, the state in the state space of $\hat{\mathcal{Z}}^{\Delta z,h,N}$ corresponding to the region containing $\hat{x}_0$, it holds that:

$$\lim_{N\to\infty} \lim_{h\to 0} \lim_{\Delta z\to 0} |\rho_{reach}(\hat{x}_0, \hat{\mathcal{X}}^N, T, A) - \rho_{reach}(z_{d,0}, \hat{\mathcal{Z}}^{\Delta z,h,N}, \lfloor \frac{T}{h} \rfloor, A)| = 0.$$

**Proof.** In order to prove the convergence, we start by introducing some notation. First of all, $B\hat{\mathcal{X}}^N$ and $B\hat{\mathcal{Y}}^N$ are the CTMC and its CLA projected on the inequalities defining the region $A$. Additionally we denote by $\hat{\mathcal{Z}}^{h,N}$ the DTMP obtained by time discretization of $B\hat{\mathcal{Y}}^N$), and $\hat{\mathcal{Z}}^{\Delta z,h,N}$ is the space discretization of $\hat{\mathcal{Z}}^{h,N}$.

We now introduce the following stopping times, which are random variables on $\mathbb{R}_{\geq 0}$ denoting the random time in which a certain event happens. In particular, we

are interested in the stopping times corresponding to the event of entering into the region $A$, usually known as hitting times, for the different processes we consider:

- $\mathtt{T}^N$ is the hitting time for $\hat{\mathcal{X}}^N$;
- $\bar{\mathtt{T}}^N$ is the hitting time for $\hat{\mathcal{Y}}^N$;
- $\mathtt{T}^{N,h}$ is the hitting time for $\hat{\mathcal{Z}}^{h,N}$;
- $\mathtt{T}^{N,h,\Delta z}$ is the hitting time for $\hat{\mathcal{Z}}^{\Delta z,h,N}$.

Hitting times are strictly related to the reachability probability. For instance, $Pr\{\exists t \leq T : \hat{\mathcal{X}}^N(t) \in A\} = Pr\{\mathtt{T}^N \leq T\}$. Furthermore, we introduce also the stopping time $\mathtt{T}^G$, which is the hitting time for the Gaussian process $G(t)$ to enter the rescaled region $A^\infty$, which is the limiting region, similarly to what we do in the proof of the Theorem 6. We have the following weak convergence relationships for such hitting times:

- $\mathtt{T}^{N,h,\Delta z} \Rightarrow \mathtt{T}^{N,h}$ as $\Delta z \to 0$;
- $\mathtt{T}^{N,h} \Rightarrow \bar{\mathtt{T}}^N$ as $h \to 0$;
- $\bar{\mathtt{T}}^N \Rightarrow \mathtt{T}^G$ as $N \to \infty$;
- $\mathtt{T}^N \Rightarrow \mathtt{T}^G$ as $N \to \infty$;

To show these relationships, one just has to use the correspondence of hitting times with the reachability probability, and the convergence of the latter by virtue of the proof of Theorem 4. For instance $Pr\{\mathtt{T}^N \leq T\} = Pr\{\exists t \leq T : \hat{\mathcal{X}}^N(t) \in A\} \to_{N\to\infty} Pr^G\{\exists t \leq T : G(t) \in A^\infty\} = Pr^G\{\mathtt{T}^G \leq T\}$, where $Pr^G$ is the Gaussian probability measure of $G$. The pointwise convergence of the cumulative distributions function of $\mathtt{T}^N$ to that of $\mathtt{T}$ implies weak convergence by the Portmanteau theorem [31].

In order to prove the convergence of rewards, given a reward structure $\rho$ on $\mathbb{R}^m$ and a path $\omega : \mathbb{R}_{\geq 0} \to \mathbb{R}^m, m > 0$, we define the functional $\mathcal{R}(\omega, T) = \int_0^T \rho(\omega(s)))ds$. In order to evaluate the desired reward, we need to stop the integration as soon as the process enters the target region $A$, hence $\rho_{reach}(\hat{\mathcal{X}}^N, T, A) = \mathbb{E}[\mathcal{R}(\hat{\mathcal{X}}^N, \mathtt{T}^N)]$, where the expectation is taken with respect to both $\mathcal{X}^N$ and $\mathtt{T}^N$. Then, by triangular inequality, we have:

$$
\begin{aligned}
|\mathbb{E}[\mathcal{R}(\hat{\mathcal{X}}^N, \mathtt{T}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \mathtt{T}^{N,h,\Delta z})]| \leq \\
|\mathbb{E}[\mathcal{R}(B\hat{\mathcal{X}}^N, \mathtt{T}^N)] - \mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)]| + \\
|\mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \mathtt{T}^{N,h})]| + \\
|\mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \mathtt{T}^{N,h})] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \mathtt{T}^{N,h,\Delta z})]|.
\end{aligned}
$$

We will prove the proposition by showing that all three terms on the right hand side of the above inequality converge to zero. In particular, the third term can be sent to zero for only $\Delta z \to 0$, and the second term by sending only $h \to 0$, as both are related to the discretization of $B\hat{\mathcal{Y}}^N$. Instead, the first term depends only on $N$.

We will start with the second term. First, Corollary 11.1, in the Appendix, implies that $\hat{\mathcal{Z}}^{h,N} \to B\hat{\mathcal{Y}}^N$ in probability as $h \to 0$. Furthermore, Theorem 4 gives us weak convergence of the hitting times: $\mathtt{T}^{N,h} \Rightarrow \bar{\mathtt{T}}^N$. The challenge in the second term lies in the fact that it depends on two random variables, so we need to rely again on triangular inequality to separate them:

$$\begin{aligned}|\mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \mathtt{T}^{N,h})]| \leq \\ |\mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \bar{\mathtt{T}}^N)]| + \\ |\mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \bar{\mathtt{T}}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \mathtt{T}^{N,h})]|\end{aligned}$$

Consider now a term appearing in the right hand side, e.g. $\mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)]$. As the expectation is taken with respect to both $B\hat{\mathcal{Y}}^N$ and $\bar{\mathtt{T}}^N$, we can rely on the following conditional expectation decomposition:

$$\mathbb{E}_{B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)] = \mathbb{E}_{\bar{\mathtt{T}}^N}[\mathbb{E}_{B\hat{\mathcal{Y}}^N}[\mathcal{R}(B\hat{\mathcal{Y}}^N, t) \mid \bar{\mathtt{T}}^N = t]].$$

Furthermore, recall that:

$$\mathbb{E}_{B\hat{\mathcal{Y}}^N}[\mathcal{R}(B\hat{\mathcal{Y}}^N, t)] = \int_0^t \mathbb{E}[\rho(B\hat{\mathcal{Y}}^N(s)]ds.$$

Now, consider the term $|\mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \bar{\mathtt{T}}^N)]|$. Applying the previous decomposition, we can upper bound it by:

$$\mathbb{E}_{\bar{\mathtt{T}}^N}\left[\int_0^t |\mathbb{E}[\rho(B\hat{\mathcal{Y}}^N(s) \mid \bar{\mathtt{T}}^N = t] - \mathbb{E}[\rho(\hat{\mathcal{Z}}^{h,N}(s) \mid \bar{\mathtt{T}}^N = t]|ds\right],$$

where we assume that $\hat{\mathcal{Z}}^{h,N}(s)$ is a piecewise constant function in between each step at distance $h$, to write its cumulative reward as an integral.

We have that $\sup_{s \leq t} |B\hat{\mathcal{Y}}^N(s) - \hat{\mathcal{Z}}^{h,N}(s)|$ converges to zero in probability as $h \to 0$ [109]. From this, we can deduce that $\mathbb{E}[\sup_{s \leq t} |B\hat{\mathcal{Y}}^N(s) - \hat{\mathcal{Z}}^{h,N}(s)|]$ converges to zero. This proof presented in [109] is consequence of the *Borell-TIS inequality* [3], which guarantees that the supremum of a Gaussian process is still normally distributed. Hence: $|\mathbb{E}[\rho(B\hat{\mathcal{Y}}^N(s) \mid \bar{\mathtt{T}}^N = t] - \mathbb{E}[\rho(\hat{\mathcal{Z}}^{h,N}(s) \mid \bar{\mathtt{T}}^N = t]| \leq \mathbb{E}[|\rho(B\hat{\mathcal{Y}}^N(s)) - \rho(\hat{\mathcal{Z}}^{h,N}(s))| \ \bar{\mathtt{T}}^N = t] \leq L_\rho \mathbb{E}[|B\hat{\mathcal{Y}}^N(s) - \hat{\mathcal{Z}}^{h,N}(s)| \ \bar{\mathtt{T}}^N = t] \leq L_\rho \mathbb{E}[\sup_{s \leq T} |B\hat{\mathcal{Y}}^N(s) - \hat{\mathcal{Z}}^{h,N}(s)| \ \bar{\mathtt{T}}^N = t] = L_\rho \Delta_h$, which converges to zero by the discussion above. In the

above $L_\rho$ is the Lipschitz constant of rewards $\rho$ Hence we can bound the first term by $\mathbb{E}[\int_0^t \Delta_h ds] \leq \Delta_h T$, which goes to zero as $h \to 0$.

Consider now the term $|\mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \bar{\mathtt{T}}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \mathtt{T}^{N,h})]|$: it tends to zero by application of the Portmanteau theorem, owing to the weak convergence of $\mathtt{T}^{N,h}$ to $\bar{\mathtt{T}}^N$, and the fact that $\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, t)$ is a bounded and continuous function of $t$ (being the cumulative reward up to time $t$ of a bounded function $\rho$).

The third term in the main inequality, $|\mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{h,N}, \mathtt{T}^{N,h})] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \mathtt{T}^{N,h,\Delta z})]|$, can be shown to converge to zero using a similar approach, owing to the convergence of the space discretization to the DTMP $\hat{\mathcal{Z}}^{h,N}$, and the convergence of the hitting times.

What is left is the first term of the main inequality of the theorem, namely $|\mathbb{E}[\mathcal{R}(B\hat{\mathcal{X}}^N, \mathtt{T}^N)] - \mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, \bar{\mathtt{T}}^N)]|$, which has to converge to zero as $N$ diverges.

To simplify the notation below, let us define:

- $g^N(t) = \mathbb{E}[\mathcal{R}(B\hat{\mathcal{X}}^N, t)]$ is the cumulative reward for $B\hat{\mathcal{X}}^N$ up to time $t$
- $\gamma^N(t) = \mathbb{E}[\mathcal{R}(B\hat{\mathcal{Y}}^N, t)]$ is the cumulative reward for $B\hat{\mathcal{Y}}^N$ up to time $t$.

Then the first term can be bounded by:

$$\begin{aligned}
|\mathbb{E}[g^N(\mathtt{T}^N)] - \mathbb{E}[\gamma^N(\bar{\mathtt{T}}^N)]| \;\leq\; & |\mathbb{E}[g^N(\mathtt{T}^N)] - \mathbb{E}[g^\infty(\mathtt{T}^N)]| \\
+ \; & |\mathbb{E}[\gamma^\infty(\mathtt{T}^N)] - \mathbb{E}[\gamma^\infty(\mathtt{T})]| \\
+ \; & |\mathbb{E}[\gamma^\infty(\mathtt{T})] - \mathbb{E}[\gamma^\infty(\bar{\mathtt{T}}^N)]| \\
+ \; & |\mathbb{E}[\gamma^\infty(\bar{\mathtt{T}}^N)] - \mathbb{E}[\gamma^N(\bar{\mathtt{T}}^N)]|
\end{aligned}$$

where $g^\infty = \gamma^\infty$ is the cumulative reward for the fluid limit $B\hat{X}^\infty = B\Phi$.

Consider the first term in the above inequality:

$$\begin{aligned}
|\mathbb{E}[g^N(\mathtt{T}^N)] - \mathbb{E}[g^\infty(\mathtt{T}^N)]| \;&\leq\; \mathbb{E}_{t\sim\mathtt{T}^N}\left[\mathbb{E}\left[\int_0^t |\rho(\hat{\mathcal{X}}^N(s)) - \rho(\Phi(s))ds|\right]\right] \\
&\leq\; \mathbb{E}_{t\sim\mathtt{T}^N}\left[\int_0^t L_\rho \mathbb{E}[|\mathcal{X}^N(s) - \Phi(s)|]ds\right] \\
&\leq\; \mathbb{E}_{t\sim\mathtt{T}^N}\left[\int_0^t L_\rho \sup_{s\leq T} \mathbb{E}[|\mathcal{X}^N(s) - \Phi(s)|]ds\right].
\end{aligned}$$

Now $\sup_{s\leq T} E[|\mathcal{X}^N(s) - \Phi(s)|]$ converges to zero by virtue of a corollary of the fluid approximation theorem on the rate of convergence of expectations [76], meaning that there is $N_1$ such that, for $N \geq N_1$, it is less than $\epsilon/(4T)$. For all such $N$, it follows that $|E[g^N(\mathtt{T}^N)] - E[g^\infty(\mathtt{T}^N)]| \leq \epsilon/4$.

Le us deal with the fourth term:

$$|\mathbb{E}[\gamma^{\infty}(\bar{\mathtt{T}}^N)] - \mathbb{E}[\gamma^N(\bar{\mathtt{T}}^N)]| \leq \mathbb{E}_{t\sim\bar{\mathtt{T}}^N}[|\gamma^{\infty}(t) - \gamma^N(t)|].$$

For a fixed $t$, we have that $|\gamma^{\infty}(t) - \gamma^N(t)| \leq \int_0^t \mathbb{E}[|\rho(\Phi(s) + G(s)/\sqrt{N}) - \rho(\Phi(s))|] \leq \int_0^t \mathbb{E}[L_\rho |G(s)/\sqrt{N}|] = L_\rho \int_0^t \mathbb{E}[\sup_{s\leq T} |G(s)|]/\sqrt{N}$. Now, as $G(t)$ has bounded convariance matrix in $[0,T]$, $\mathbb{E}[\sup_{s\leq T} |G(s)|]$ is finite, say equal to $M_G$, hence $|\gamma^{\infty}(t) - \gamma^N(t)| \leq L_\rho M_G t/\sqrt{N}$, and so $|\mathbb{E}[\gamma^{\infty}(\bar{\mathtt{T}}^N)] - E[\gamma^N(\bar{\mathtt{T}}^N)]| \leq L_\rho M_G T/\sqrt{N}$ which is less than $\epsilon/4$ for $N \geq N_4$, for some $N_4 > 0$.

Terms two and three in the inequality above, instead, converge by virtue of the Portmanteau theorem and of the weak convergence of $\mathtt{T}^N$ or $\bar{\mathtt{T}}^N$ to $\mathtt{T}^G$, hence there is $N_2$ such that they are less that $\epsilon/4$ for $N \geq N_2$. It then follows that:

$$\limsup_{N\to\infty} |\mathbb{E}[g^N(\mathtt{T}^N)] - \mathbb{E}[\gamma^N(\bar{\mathtt{T}}^N)]| < \epsilon$$

for an arbitrary $\epsilon$, implying:

$$\lim_{N\to\infty} |E[g^N(\mathtt{T}^N)] - E[\gamma^N(\bar{\mathtt{T}}^N)]| = 0.$$

Thus, we showed that $|\mathbb{E}[\mathcal{R}(B\hat{\mathcal{X}}^N, \mathtt{T}^N)] - \mathbb{E}[\mathcal{R}(\hat{\mathcal{Z}}^{\Delta z,h,N}, \mathtt{T}^{N,h,\Delta z})]|$ converges to zero for $\Delta z, h$ tending to zero and $N$ diverging, as so do all the three terms bounding it.

## 8.2 Time discretization of continuous stochastic processes

In this section we consider the time discretization of Gaussian processes into DTMPs introduced in Section 5.2.1. We compute error bounds and prove the convergence for the discretization process.

Given a Gaussian process $\mathcal{Z}$ taking values in $\mathbb{R}^m, m > 0$ and a sampling time $h > 0$, we consider the discretization of $\mathcal{Z}$ into the DTMP $\mathcal{Z}^h$, as described in Section 5.2.1, and compute error bounds for such a discretization process with respect to probabilistic reachability properties. As reachability probabilities are dual to safety properties [2], we consider bounds with respect to the following property

$$P_{safe}(\mathcal{Z}, A, [0,T], z_0) = Pr^{\mathcal{Z}}(\forall t \in [0,T],\, \mathcal{Z}(t) \in A \,|\, \mathcal{Z}(0) = z_0),$$

where $A \subseteq \mathbb{R}^m$ and $Pr^{\mathcal{Z}}$ is the Gaussian probability measure associated to $\mathcal{Z}$.

For $i \in \{1, ..., m\}$, define the pseudometric $d_i$ as

$$d_i(t_1, t_2) = \sqrt{\mathbb{E}[(\bar{\mathcal{Z}}_i(t_2) - \bar{\mathcal{Z}}_i(t_1))^2]},$$

where $\bar{\mathcal{Z}}_i$ is the centered Gaussian process with zero mean and variance equals to $\mathcal{Z}_i$, the $i$-th component of $\mathcal{Z}$. Fix a set of sampling times $\Sigma = \{k_1, ..., k_{|\Sigma|}\} \subset [0, T]$, with fixed step distance $h > 0$. Then, call $K_d$ the constant such that for all $k \in \Sigma$ and $t_1, t_2 \in [k, k + h]$

$$\max_{i \in \{1,...,m\}} d_i(t_1, t_2) \leq K_d \cdot |t_2 - t_1|.$$

For $\epsilon > 0$ define the set:

$$A^\epsilon = \{x \in A \ : \ |x - \partial A|_1 \geq \epsilon\},$$

where $\partial A$ is the boundary of $A$ and

$$|x - \partial A^\epsilon|_1 = \min\{|x - \bar{x}|_1 \, s.t. \, \bar{x} \in \partial A^\epsilon\}$$

is the $L_1$ norm between a point and a set. That is $A^\epsilon$ is a subset of $A$ where all the points are distant at least $\epsilon$ from $\partial A$. Define the events:

$$\mathcal{A}^\epsilon = \{\forall k_i \in \Sigma, \, \mathcal{Z}(k_i) \in A^\epsilon\}$$

and

$$\mathcal{B} = \{\exists t \in [0, T] \ s.t. \ \mathcal{Z}(t) \notin A\}.$$

That is, $\mathcal{A}^\epsilon$ is the probability that the system is in $A^\epsilon$ at the sampling times, and $\mathcal{B}$ is the probability to reach an unsafe state within the time interval of interest. Clearly, it holds that

$$P_{safe}(\mathcal{Z}, A, [0, T], z_0) \geq Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon \wedge \neg\mathcal{B}); \quad P_{safe}(\mathcal{Z}, A, [0, T], z_0) = \lim_{\epsilon \to 0} Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon \wedge \neg\mathcal{B})$$

and similarly by the definition of $Pr^{\mathcal{Z}^h}$, the probability space of $\mathcal{Z}^h$ (see Definition 5), it holds that

$$P_{safe}(\mathcal{Z}^h, A, \Sigma, z_0) \geq Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon) \quad P_{safe}(\mathcal{Z}^h, A, \Sigma, z_0) = \lim_{\epsilon \to 0} Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon),$$

where

$$P_{safe}(\mathcal{Z}^h, A, \Sigma, z_0) = Pr^{\mathcal{Z}^h}(\forall k_i \in \Sigma, \, \mathcal{Z}^h(i) \in A \,|\, \mathcal{Z}^h(0) = z_0).$$

For a finite $\epsilon > 0$, $Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon \wedge \neg\mathcal{B})$ is the lower bound for the safety probability computed on $A$, since it requires the system to be inside $A^\epsilon \subseteq A$ at sampling times in $\Sigma$. Call $E$ the non-negative constant defined as

$$E = \max_{k \in \Sigma} \sup_{z_0 \in A} \sup_{t \in [k_i, k_i + h]} |\mathbb{E}[\mathcal{Z}(t)] - \mathbb{E}[\mathcal{Z}(0)]|_1$$

where $E[\mathcal{Z}(0)] = z_0$. For instance, if $\mathbb{E}[\mathcal{Z}]$ is a Lipschitz continuous with Lipschitz constant $L_{\mathbb{E}}$, that is, for $t \in \mathbb{R}_{\geq 0}$ we have that

$$|\mathbb{E}[\mathcal{Z}(t+h)] - \mathbb{E}[\mathcal{Z}(t)]|_1 \leq L_{\mathbb{E}} h,$$

for $\mathbb{E}[\mathcal{Z}(0)] \in A$, then we have $E \leq L_{\mathbb{E}} h$.

We are now ready to state the following result

**Theorem 11.** *Introduce constant*

$$\Gamma = 12 \int_0^{K_d h} \sqrt{ln\left(\frac{2K_d h}{x} + 1\right)}\, dx.$$

*Assume $\mathcal{Z}$ is almost surely bounded and that $\frac{\epsilon - E}{m} - \Gamma > 0$. Then, under the further assumption that*

$$|\frac{2mT}{h} \exp^{-\frac{(\frac{\epsilon-E}{m} - \Gamma)^2}{2\sigma_I^2}}| \leq 1,$$

*it holds that*

$$Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon) \geq Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon \wedge \neg\mathcal{B}) \geq Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon) \cdot \left(1 - \frac{2mT}{h} \exp^{-\frac{(\frac{\epsilon-E}{m} - \Gamma)^2}{2\sigma_I^2}}\right).$$

*where $\sigma_T^2 = \max_{k_i \in \Sigma} \max_{j \in \{1,\dots,m\}} \sup_{t \in [0,h]} \mathbb{V}[\mathcal{Z}_j(t)]$.*

*Proof.* For each $i \in \{1, ..., m\}$, $\mathcal{Z}_i$ is an almost surely bounded and uniformly continuous Gaussian process (GP) in $[0, T]$, time interval of interest. Using the rules of probability we get

$$Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon \wedge \neg\mathcal{B}) = Pr^{\mathcal{Z}}(\mathcal{A}^\epsilon) \cdot (1 - Pr^{\mathcal{Z}}(\mathcal{B}|\mathcal{A}^\epsilon))$$

Then, by definition of probability we have that

$$\begin{aligned}
&0 \leq Pr^{\mathcal{Z}}(\mathcal{B}|\mathcal{A}^\epsilon) \\
&\leq Pr^{\mathcal{Z}}(\exists k_i \in \Sigma_n \sup_{t \in [k_i, k_i+h]} |\mathcal{Z}(t) - \mathcal{Z}(k_i)|_1 > \epsilon) \\
&\leq Pr^{\mathcal{Z}}(\exists k_i \in \Sigma_n \, s.t. \sup_{t \in [k_i, k_i+h]} |\bar{\mathcal{Z}}(t) + \mathbb{E}[\mathcal{Z}(t)] - \bar{\mathcal{Z}}(k_i) - \mathbb{E}[\mathcal{Z}(k_1)]|_1 > \epsilon) \\
&\leq Pr^{\mathcal{Z}}(\exists k_i \in \Sigma_n \, s.t. \sup_{t \in [k_i, k_i+h]} |\bar{\mathcal{Z}}(t) - \bar{\mathcal{Z}}(k_i)|_1 > \epsilon - E),
\end{aligned}$$

where the first inequality holds because all paths $\omega$ of $\mathcal{Z}$ that are in $\mathcal{B}|\mathcal{A}^\epsilon$ necessarily are such that $\sup_{t \in [k_i, k_i+h]} |\omega(t) - \omega(k_i)|_1 > \epsilon$ for some $k_i \in \Sigma$ (note that $\omega(k_i) \in A^\epsilon$). By the properties of closure of Gaussian processes with respect to linear combinations

we have that $\mathcal{Z}(t) = \bar{\mathcal{Z}}(t) + \mathbb{E}[\mathcal{Z}(t)]$, where $\bar{\mathcal{Z}}(t)$ is a Gaussian process with zero mean and same covariance matrix of $\mathcal{Z}(t)$. The last inequality above holds because the covariance matrix of $\mathcal{Z}(t)$ is independent of the initial state. Then, it follows that

$$
\begin{aligned}
& Pr^{\mathcal{Z}}(\exists k_i \in \Sigma_n \, s.t. \sup_{t\in[k_i,k_i+h]} |\bar{\mathcal{Z}}(t) - \bar{\mathcal{Z}}(k_i)|_1 > \epsilon - E) && (8.1)\\
& \leq \sum_{i=1}^{|\Sigma_n|} Pr^{\mathcal{Z}} \left( \sup_{t\in[k_i,k_i+h]} |\bar{\mathcal{Z}}(t) - \bar{\mathcal{Z}}(k_i)|_1 > \epsilon - E \right) && (8.2)\\
& = \sum_{i=1}^{|\Sigma_n|} Pr^{\mathcal{Z}} \left( \sup_{t\in[k_i,k_i+h]} \sum_{j=1}^{m} |\bar{\mathcal{Z}}_j(t) - \bar{\mathcal{Z}}_j(k_i)|_1 > \epsilon - E \right) && (8.3)\\
& \leq \sum_{i=1}^{|\Sigma_n|} Pr^{\mathcal{Z}} \left( \exists j \in \{1,...,m\} \, s.t. \sup_{t\in[k_i,k_i+h]} |\bar{\mathcal{Z}}_j(t) - \bar{\mathcal{Z}}_j(k_i)|_1 > \frac{\epsilon - E}{m} \right) && (8.4)\\
& \leq m \sum_{i=1}^{|\Sigma_n|} \max_{j\in\{1,...,m\}} Pr^{\mathcal{Z}} \left( (\sup_{t\in[k_i,k_i+h]} |\bar{\mathcal{Z}}_j(t) - \bar{\mathcal{Z}}_j(k_i)|_1) > \frac{\epsilon - E}{m} \right) && (8.5)\\
& \leq 2m|\Sigma_n| \max_{k_i\in\Sigma} \max_{j\in\{1,...,m\}} Pr^{\mathcal{Z}} \left( \sup_{t\in[k_i,k_i+h]} (\bar{\mathcal{Z}}_j(t) - \bar{\mathcal{Z}}_j(k_i)) > \frac{\epsilon - E}{m} \right), &&
\end{aligned}
$$

where (8.2) and (8.5) come from the union bound and (8.3) is the definition of $L_1$ norm. Being $\bar{\mathcal{Z}}_j$ a centered uni-dimensional GP, we can make use of the Borell-TIS inequality [3, 4]. Given an interval $I = [0, h]$ with $\sigma_I = \sup_{t\in I}(\sigma(t))$, supremum of the standard deviation of the process, the Borell-TIS inequality guarantees that

$$
Prob\left( \sup_{t\in I} \bar{\mathcal{Z}}_i(t) > u \right) \leq \exp^{-\frac{(u - E[sup_{t\in I} \bar{\mathcal{Z}}_i(t)])^2}{(\sigma_I^2)}}
$$

Applying this result to our case for intervals of the type $[k_i, k_i + h]$ and assuming $\mathcal{Z}(k_i)$ has zero variance, we have that

$$
\begin{aligned}
& Pr^{\mathcal{Z}} \left( \sup_{t\in[k_i,k_i+h]} (\bar{\mathcal{Z}}_j(t) - \bar{\mathcal{Z}}_j(k_i)) > \frac{\epsilon - E}{m} \right)\\
& \leq \exp^{-\frac{\left(\frac{\epsilon - E}{m} - \mathbb{E}\left[sup_{t\in I}(\bar{\mathcal{Z}}_j(t))\right]\right)^2}{\sigma_I^2}}.
\end{aligned}
$$

At this point, we need to bound $\mathbb{E}\left[\sup_{t\in[k_i,k_i+h]}(\bar{\mathcal{Z}}_i(t))\right]$ from above. We can do that by making use of the *Dudley's entropy integral* [3]. The Dudley entropy integral guarantees that

$$
E[\sup_{t\in[k_i,k_i+h]} \mathcal{Z}_j(t)] \leq 12 \int_0^{\frac{diam([k_i,k_i+h])}{2}} \sqrt{\ln(N([k_i, k_i + h], d, \eta))} d\eta,
$$

where $N([k_i, k_i + h], d, \eta)$ represents the smallest number of balls of radius $\eta$ which covers $[k_i, k_i + h]$, under pseudo-metric $d$. That is, the maximum number of partitions of $[k_i, k_i + h]$ such that in each set $d(t_1, t_2) \leq \eta$. $diam([k_i, k_i + h])$ is defined as

$$diam([k_i, k_i + h]) = sup_{t',t \in [k_i, k_i+h]} d(t', t)$$

and with our assumptions, it is possible to show that for each $k_i \in \Sigma$

$$diam([k_i, k_i + h]) \leq K_d \cdot h$$

Moreover, we have

$$N([k_i, k_i + h], d, \epsilon) \leq \frac{2K_d h}{\epsilon} + 1,$$

This can be easily understood thinking at the geometry of the problem. In fact, $[k_i, k_i + h]$ can be divided in $\frac{2h}{r} + 1$ balls of radius $r$, and $r \geq \frac{\epsilon}{K_d}$. As a consequence, we have that for each $k_i \in \Sigma$

$$\max_{j \in \{1,\dots,m} E\left[\sup_{t \in [k_i, k_i+h]} (\hat{\mathcal{Z}}(t))\right] \leq 12 \int_0^{K_d h} \sqrt{ln\left(\frac{K_d h}{2\epsilon} + 1\right)}\, d\epsilon = \Gamma$$

□

**Corollary 11.1.** Under the assumptions of Theorem 11 it holds that

$$P_{safe}(\mathcal{Z}, A, [0, T], z_0) = \lim_{h \to 0} P_{safe}(\mathcal{Z}^h, A^\epsilon, \Sigma, z_0).$$

*Proof.* For $n \in \mathbb{N}$, such that $n > 3$, select $\epsilon = 2^{-log(n)}$ and $h = 2^{-n}$. As expectation and variance of $\mathcal{Z}$ are Lipschitz continuous (they are a solution of ODEs with bounded terms by assumption), we have that $E \leq L_{\mathbb{E}} h$ and $\sigma_T^2 \leq L_\sigma h$ for certain constants $L_{\mathbb{E}}$ and $L_\sigma$. For such an $h$ we have that $\Gamma \leq \sqrt{2} K_d h$. Thus, we obtain

$$\lim_{n \to \infty} \frac{2mT}{2^{-n}} e^{-\frac{(\frac{2^{-log(n)} - L_{\mathbb{E}} 2^{-n}}{m} - \sqrt{2} K_d 2^{-n})^2}{L_\sigma 2^{-n}}} = 0.$$

□

# Bibliography


[1] Alessandro Abate, Luca Cardelli, Marta Kwiatkowska, Luca Laurenti, and Boyan Yordanov. Experimental biological protocols with formal semantics. *CMSB*, 2018.

[2] Alessandro Abate, Joost-Pieter Katoen, John Lygeros, and Maria Prandini. Approximate model checking of stochastic hybrid systems. *European Journal of Control*, 16(6):624–641, 2010.

[3] Robert J Adler. *The geometry of random fields*, volume 62. Siam, 2010.

[4] Robert J Adler and Jonathan E Taylor. *Random fields and geometry*. Springer Science & Business Media, 2009.

[5] Rajeev Alur, Tomás Feder, and Thomas A Henzinger. The benefits of relaxing punctuality. *Journal of the ACM (JACM)*, 43(1):116–146, 1996.

[6] Stefan L Ameres and Phillip D Zamore. Diversifying microrna sequence and function. *Nature reviews Molecular cell biology*, 14(8):475–488, 2013.

[7] David F Anderson, Gheorghe Craciun, and Thomas G Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bulletin of mathematical biology*, 72(8):1947–1970, 2010.

[8] David F Anderson, Germán A Enciso, and Matthew D Johnston. Stochastic analysis of biochemical reaction networks with absolute concentration robustness. *Journal of The Royal Society Interface*, 11(93):20130943, 2014.

[9] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*, volume 1. Springer.

[10] David F Anderson and Thomas G Kurtz. Continuous time markov chain models for chemical reaction networks. In *Design and analysis of biomolecular circuits*, pages 3–42. Springer, 2011.

[11] Dana Angluin, James Aspnes, Zoë Diamadi, Michael J Fischer, and René Peralta. Computation in networks of passively mobile finite-state sensors. *Distributed computing*, 18(4):235–253, 2006.

[12] Dana Angluin, James Aspnes, David Eisenstat, and Eric Ruppert. The computational power of population protocols. *Distributed Computing*, 20(4):279–304, 2007.

[13] Adnan Aziz, Kumud Sanwal, Vigyan Singhal, and Robert Brayton. Verifying continuous time markov chains. In *Computer Aided Verification*, pages 269–276. Springer, 1996.

[14] Adnan Aziz, Kumud Sanwal, Vigyan Singhal, and Robert Brayton. Model-checking continuous-time Markov chains. *ACM Transactions on Computational Logic*, 1(1):162–170, 2000.

[15] Alessia Baccarini, Hemangini Chauhan, Thomas J Gardner, Anitha D Jayaprakash, Ravi Sachidanandam, and Brian D Brown. Kinetic analysis reveals the fate of a microrna following target regulation in mammalian cells. *Current biology*, 21(5):369–376, 2011.

[16] Christel Baier, Boudewijn Haverkort, Holger Hermanns, and Joost-Pieter Katoen. Model checking continuous-time markov chains by transient analysis. In *Computer Aided Verification*, pages 358–372. Springer, 2000.

[17] Christel Baier, Boudewijn Haverkort, Holger Hermanns, and Joost-Pieter Katoen. Model-checking algorithms for continuous-time Markov chains. *Software Engineering, IEEE Transactions on*, 29(6):524–541, 2003.

[18] Christel Baier, Boudewijn Haverkort, Holger Hermanns, and Joost-Pieter Katoen. Model-checking algorithms for continuous-time markov chains. *Software Engineering, IEEE Transactions on*, 29(6):524–541, 2003.

[19] Christel Baier, Joost-Pieter Katoen, et al. *Principles of model checking*, volume 26202649. MIT press Cambridge, 2008.

[20] Rachael Bakker and Richard W Carthew. MicroRNAs make a difference in cardiovascular robustness. *Developmental Cell*, 40(6):515–516, 2017.

[21] Nathalie Q Balaban, Jack Merrin, Remy Chait, Lukasz Kowalik, and Stanislas Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305(5690):1622–1625, 2004.

[22] Gábor Balázsi, Alexander van Oudenaarden, and James J Collins. Cellular decision making and biological noise: from microbes to mammals. *Cell*, 144(6):910–925, 2011.

[23] Paolo Ballarini, Hilal Djafri, Marie Duflot, Serge Haddad, and Nihal Pekergin. Cosmos: a statistical model checker for the hybrid automata stochastic logic. In *2011 Eighth International Conference on Quantitative Evaluation of SysTems*, pages 143–144. IEEE, 2011.

[24] David P Bartel. Micrornas: genomics, biogenesis, mechanism, and function. *cell*, 116(2):281–297, 2004.

[25] David P Bartel. Micrornas: target recognition and regulatory functions. *cell*, 136(2):215–233, 2009.

[26] Ezio Bartocci, Luca Bortolussi, Laura Nenzi, and Guido Sanguinetti. System design of stochastic models using robustness of temporal properties. *Theoretical Computer Science*, 587:3–25, 2015.

[27] Ezio Bartocci and Pietro Lió. Computational modeling, formal analysis, and tools for systems biology. *PLoS computational biology*, 12(1):e1004591, 2016.

[28] Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, 2000.

[29] Dimitir P Bertsekas and Steven Shreve. *Stochastic optimal control: the discrete-time case*. 2004.

[30] Patrick Billingsley. *Probability and measure*. John Wiley & Sons, 2008.

[31] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

[32] William J Blake, Mads Kærn, Charles R Cantor, and James J Collins. Noise in eukaryotic gene expression. *Nature*, 422(6932):633–637, 2003.

[33] Christer Borell. The brunn-minkowski inequality in gauss space. *Inventiones mathematicae*, 30(2):207–216, 1975.

[34] Luca Bortolussi. Hybrid behaviour of markov population models. *Information and Computation*, 247:37–86, 2016.

[35] Luca Bortolussi, Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Approximation of probabilistic reachability for chemical reaction networks using the linear noise approximation. In *International Conference on Quantitative Evaluation of Systems*, pages 72–88. Springer, 2016.

[36] Luca Bortolussi, Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Central limit model checking. *ACM Transactions on Computational Logic (TOCL)*, In Press.

[37] Luca Bortolussi and Jane Hillston. Fluid model checking. In *CONCUR 2012–Concurrency Theory*, pages 333–347. Springer, 2012.

[38] Luca Bortolussi and Jane Hillston. Efficient checking of individual rewards properties in markov population models. *arXiv preprint arXiv:1509.08561*, 2015.

[39] Luca Bortolussi, Jane Hillston, Diego Latella, and Mieke Massink. Continuous approximation of collective system behaviour: A tutorial. *Performance Evaluation*, 70(5):317–349, 2013.

[40] Luca Bortolussi and Roberta Lanciani. Model checking markov population models by central limit approximation. In *Quantitative Evaluation of Systems*, pages 123–138. Springer, 2013.

[41] Luca Bortolussi and Roberta Lanciani. Stochastic approximation of global reachability probabilities of Markov population models. In *Computer Performance Engineering*, pages 224–239. Springer, 2014.

[42] Luca Bortolussi, Dimitrios Milios, and Guido Sanguinetti. U-check: Model checking and parameter synthesis under uncertainty. In *International Conference on Quantitative Evaluation of Systems*, pages 89–104. Springer, 2015.

[43] Luca Bortolussi, Dimitrios Milios, and Guido Sanguinetti. Smoothed model checking for uncertain continuous-time Markov chains. *Information and Computation*, 2016.

[44] Benoît Boyer, Kevin Corre, Axel Legay, and Sean Sedwards. Plasma-lab: A flexible, distributable statistical model checking library. In *International Conference on Quantitative Evaluation of Systems*, pages 160–164. Springer, 2013.

[45] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Antithetic integral feedback ensures robust perfect adaptation in noisy bimolecular networks. *Cell systems*, 2(1):15–26, 2016.

[46] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Variance reduction for antithetic integral control of stochastic reaction networks. *arXiv preprint arXiv:1711.08291*, 2017.

[47] Luboš Brim, Petr Dluhoš, David Šafránek, and Tomáš Vejpustek. Stl*: Extending signal temporal logic with signal-value freezing operator. *Information and Computation*, 236:52–67, 2014.

[48] Luca Cardelli. On process rate semantics. *Theoretical Computer Science*, 391(3):190–215, 2008.

[49] Luca Cardelli, Milan Češka, Martin Fränzle, Marta Kwiatkowska, Luca Laurenti, Nicola Paoletti, and Max Whitby. Syntax-guided optimal synthesis for chemical reaction networks. In *International Conference on Computer Aided Verification*, pages 375–395. Springer, 2017.

[50] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Stochastic analysis of chemical reaction networks using linear noise approximation. In *Computational Methods in Systems Biology*, pages 64–76. Springer, 2015.

[51] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Programming discrete distributions with chemical reaction networks. In *International Conference on DNA-Based Computers*, pages 35–51. Springer, 2016.

[52] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Stochastic analysis of chemical reaction networks using linear noise approximation. *Biosystems*, 149:26–33, 2016.

[53] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. A stochastic hybrid approximation for chemical kinetics based on the linear noise approximation. In *International Conference on Computational Methods in Systems Biology*, pages 147–167. Springer, 2016.

[54] Luca Cardelli, Marta Kwiatkowska, and Luca Laurenti. Programming discrete distributions with chemical reaction networks. *Natural Computing*, 17(1):131–145, 2018.

[55] Taolue Chen, Tingting Han, Joost-Pieter Katoen, and Alexandru Mereacre. Ltl model checking of time-inhomogeneous markov chains. *Automated Technology for Verification and Analysis*, pages 104–119, 2009.

[56] Federica Ciocchetta and Jane Hillston. Bio-pepa: A framework for the modelling and analysis of biological systems. *Theoretical Computer Science*, 410(33-34):3065–3084, 2009.

[57] BS Cirel'soN, IA Ibragimov, and VN Sudakov. Norms of gaussian sample functions. In *Proceedings of the Third JapanUSSR Symposium on Probability Theory*, pages 20–41. Springer, 1976.

[58] Edmund M Clarke and E Allen Emerson. *Design and synthesis of synchronization skeletons using branching time temporal logic*. Springer, 1982.

[59] Edmund M Clarke, William Klieber, Miloš Nováček, and Paolo Zuliani. Model checking and the state explosion problem. In *Tools for Practical Software Verification*, pages 1–30. Springer, 2012.

[60] Edmund M Clarke and Paolo Zuliani. Statistical model checking for cyber-physical systems. In *International Symposium on Automated Technology for Verification and Analysis*, pages 1–12. Springer, 2011.

[61] Gheorghe Craciun and Martin Feinberg. Multiple equilibria in complex chemical reaction networks: semiopen mass action systems. *SIAM Journal on Applied Mathematics*, 70(6):1859–1877, 2010.

[62] Frits Dannenberg, Ernst Moritz Hahn, and Marta Kwiatkowska. Computing cumulative rewards using fast adaptive uniformization. volume 25, page 9. ACM, 2015.

[63] Frits Dannenberg, Marta Kwiatkowska, Chris Thachuk, and Andrew J Turberfield. Dna walker circuits: Computational potential, design, and verification. In *DNA Computing and Molecular Programming*, pages 31–45. Springer, 2013.

[64] Alexandre Donzé and Oded Maler. *Robust satisfaction of temporal logic over real-valued signals*. Springer, 2010.

[65] Richard M Dudley. The sizes of compact subsets of hilbert space and continuity of gaussian processes. *Journal of Functional Analysis*, 1(3):290–330, 1967.

[66] Margaret S Ebert and Phillip A Sharp. Roles for micrornas in conferring robustness to biological processes. *Cell*, 149(3):515–524, 2012.

[67] Avigdor Eldar and Michael B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 2010.

[68] Johan Elf and Måns Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome research*, 13(11):2475–2484, 2003.

[69] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[70] Martin Feinberg. Lectures on chemical reaction networks. *Notes of lectures given at the Mathematics Research Center, University of Wisconsin*, 1979.

[71] Hunter B Fraser, Aaron E Hirsh, Guri Giaever, Jochen Kumm, and Michael B Eisen. Noise minimization in eukaryotic gene expression. *PLoS Biol*, 2(6):e137, 2004.

[72] Sicun Gao, Soonho Kong, and Edmund M Clarke. Satisfiability modulo odes. In *Formal Methods in Computer-Aided Design (FMCAD), 2013*, pages 105–112. IEEE, 2013.

[73] Crispin Gardiner. *Stochastic methods*, volume 4. springer Berlin, 2009.

[74] Crispin W Gardiner et al. *Handbook of stochastic methods*, volume 4. Springer Berlin, 1985.

[75] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403(6767):339, 2000.

[76] Nicolas Gast. Expected values estimated via mean-field approximation are 1/n-accurate. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(1):17, 2017.

[77] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000.

[78] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[79] Daniel T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404–425, 1992.

[80] Daniel T Gillespie. The chemical langevin equation. *The Journal of Chemical Physics*, 113(1):297–306, 2000.

[81] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[82] Daniel T Gillespie. Deterministic limit of stochastic chemical kinetics. *The Journal of Physical Chemistry B*, 113(6):1640–1644, 2009.

[83] John Goutsias and Garrett Jenkinson. Markovian dynamics on complex reaction networks. *Physics Reports*, 529(2):199–264, 2013.

[84] Ramon Grima. Construction and accuracy of partial differential equation approximations to the chemical master equation. *Physical Review E*, 84(5):056109, 2011.

[85] Ramon Grima. Linear-noise approximation and the chemical master equation agree up to second-order moments for a class of chemical systems. *Physical Review E*, 92(4):042124, 2015.

[86] Cato Maximilian Guldberg and Peter Waage. *Etudes sur les affinités chimiques*. Brøgger & Christie, 1867.

[87] CM Guldberg, P Waage, and EW Lund. Studies concerning affinity. cm forhandlinger: Videnskabs-selskabet i christiana, 35. *1965) Journal of Chemical Education*, 42:548–550, 1864.

[88] John Heath, Marta Kwiatkowska, Gethin Norman, David Parker, and Oksana Tymchyshyn. Probabilistic model checking of complex biological pathways. *Theoretical Computer Science*, 391(3):239–257, 2008.

[89] Thomas A Henzinger, Maria Mateescu, and Verena Wolf. Sliding window abstraction for infinite markov chains. In *Computer Aided Verification*, pages 337–352. Springer, 2009.

[90] Thomas A Henzinger, Linar Mikeev, Maria Mateescu, and Verena 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. ACM, 2010.

[91] Joao Hespanha. Moment closure for biochemical networks. In *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, pages 142–147. IEEE, 2008.

[92] Jane Hillston. *A compositional approach to performance modelling*, volume 12. Cambridge University Press, 2005.

[93] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for rational mechanics and analysis*, 47(2):81–116, 1972.

[94] Eran Hornstein and Noam Shomron. Canalization of development by micrornas. *Nature genetics*, 38:S20–S24, 2006.

[95] A Jenson. Markov chains as an aid in the study of markov processes. *Skand. Aktuarietidskr*, 1953.

[96] Mads Kærn, Timothy C Elston, William J Blake, and James J Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6(6):451–464, 2005.

[97] Richard M Karp and Raymond E Miller. Parallel program schemata. *Journal of Computer and system Sciences*, 3(2):147–195, 1969.

[98] Dionna M Kasper, Albertomaria Moro, Emma Ristori, Anand Narayanan, Guillermina Hill-Teran, Elizabeth Fleming, Miguel Moreno-Mateos, Charles E Vejnar, Jing Zhang, Donghoon Lee, et al. Microrna s establish uniform traits during the architecture of vertebrate embryos. *Developmental Cell*, 40(6):552–565, 2017.

[99] John G Kemeny, J Laurie Snell, and Anthony W Knapp. *Denumerable Markov chains: with a chapter of Markov random fields by David Griffeath*, volume 40. Springer Science & Business Media, 2012.

[100] Hiroaki Kitano. Biological robustness. *Nature Reviews Genetics*, 5(11):826–837, 2004.

[101] Steven G Krantz and Harold R Parks. *A primer of real analytic functions*. Springer Science & Business Media, 2002.

[102] Jacek Krol, Inga Loedige, and Witold Filipowicz. The widespread regulation of microrna biogenesis, function and decay. *Nature Reviews Genetics*, 11(9):597–610, 2010.

[103] Edo Kussell and Stanislas Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743):2075–2078, 2005.

[104] Marta Kwiatkowska, Gethin Norman, and David Parker. Stochastic model checking. In *Formal methods for performance evaluation*, pages 220–270. Springer, 2007.

[105] Marta Kwiatkowska, Gethin Norman, and David Parker. Stochastic model checking. pages 220–270, 2007.

[106] Marta Kwiatkowska, Gethin Norman, and David Parker. Prism 4.0: Verification of probabilistic real-time systems. In *Computer aided verification*, pages 585–591. Springer, 2011.

[107] Matthew R Lakin, David Parker, Luca Cardelli, Marta Kwiatkowska, and Andrew Phillips. Design and analysis of dna strand displacement devices using probabilistic model checking. *Journal of the Royal Society Interface*, page rsif20110800, 2012.

[108] Matthew R Lakin, Simon Youssef, Filippo Polo, Stephen Emmott, and Andrew Phillips. Visual dsd: a design and analysis tool for dna strand displacement systems. *Bioinformatics*, 27(22):3211–3213, 2011.

[109] Luca Laurenti, Alessandro Abate, Luca Bortolussi, Luca Cardelli, Milan Ceska, and Marta Kwiatkowska. Reachability computation for switching diffusions: Finite abstractions with certifiable and tuneable precision. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, pages 55–64. ACM, 2017.

[110] Luca Laurenti, Attila Csikasz-Nagy, Marta Kwiatkowska, and Luca Cardelli. Molecular filters for noise reduction. *Biophysical Journal*, 114(12):3000–3011, 2018.

[111] Luca Laurenti, Morteza Lahijanian, Alessandro Abate, Luca Cardelli, and Marta Kwiatkowska. Formal and efficient control synthesis for continuous-time stochastic processes. *Under Submission*, 2018.

[112] Ioannis Lestas, Glenn Vinnicombe, and Johan Paulsson. Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467(7312):174–178, 2010.

[113] Xin Li, Justin J Cassidy, Catherine A Reinke, Stephen Fischboeck, and Richard W Carthew. A microrna imparts robustness against environmental fluctuation during development. *Cell*, 137(2):273–282, 2009.

[114] MA Lifshits. Absolute continuity of functionals of supremum type for gaussian processes. *Journal of Mathematical Sciences*, 27(5):3103–3112, 1984.

[115] Hédia Maamar, Arjun Raj, and David Dubnau. Noise in gene expression determines cell fate in bacillus subtilis. *Science*, 317(5837):526–529, 2007.

[116] Oded Maler and Dejan Nickovic. Monitoring temporal properties of continuous signals. In *Formal Techniques, Modelling and Analysis of Timed and Fault-Tolerant Systems*, pages 152–166. Springer, 2004.

[117] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985, 2003.

[118] Maria Mateescu, V Wolf, F Didier, Thomas Henzinger, et al. Fast adaptive uniformisation of the chemical master equation. volume 4, pages 441–452. IET, 2010.

[119] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, 94(3):814–819, 1997.

[120] Harley H McAdams and Adam Arkin. Itsa noisy business! genetic regulation at the nanomolar scale. *Trends in genetics*, 15(2):65–69, 1999.

[121] Dimitrios Milios, Guido Sanguinetti, and David Schnoerr. Probabilistic model checking for continuous time markov chains via sequential bayesian inference. *arXiv preprint arXiv:1711.01863*, 2017.

[122] Victoria Moignard, Steven Woodhouse, Laleh Haghverdi, Andrew J Lilly, Yosuke Tanaka, Adam C Wilkinson, Florian Buettner, Iain C Macaulay, Wajid Jawaid, Evangelia Diamanti, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nature biotechnology*, 33(3):269–276, 2015.

[123] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of chemical physics*, 124(4):044104, 2006.

[124] Brian Munsky, Gregor Neuert, and Alexander van Oudenaarden. Using gene expression noise to understand gene regulation. *Science*, 336(6078):183–187, 2012.

[125] Chris J Myers. Computational synthetic biology: progress and the road ahead. *IEEE Transactions on Multi-Scale Computing Systems*, 1(1):19–32, 2015.

[126] John RS Newman, Sina Ghaemmaghami, Jan Ihmels, David K Breslow, Matthew Noble, Joseph L DeRisi, and Jonathan S Weissman. Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise. *Nature*, 441(7095):840–846, 2006.

[127] Béla Novák and John J Tyson. Design principles of biochemical oscillators. *Nature reviews Molecular cell biology*, 9(12):981–991, 2008.

[128] Kathryn A O'donnell, Erik A Wentzel, Karen I Zeller, Chi V Dang, and Joshua T Mendell. c-myc-regulated microrna modulate e2f1 expression. *Nature*, 435(7043):839–843, 2005.

[129] Ertugrul M Ozbudak, Mukund Thattai, Iren Kurtser, Alan D Grossman, and Alexander Van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature genetics*, 31(1):69–73, 2002.

[130] Michael Pedersen and Andrew Phillips. Towards programming languages for genetic engineering of living cells. *Journal of the Royal Society Interface*, pages rsif–2008, 2009.

[131] Juan M Pedraza and Johan Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861):339–343, 2008.

[132] Carl Adam Petri. Communication with automata. 1966.

[133] Diana M Posadas and Richard W Carthew. MicroRNAs and their roles in developmental canalization. *Current opinion in genetics & development*, 27:1–6, 2014.

[134] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

[135] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.

[136] Christopher V Rao, Denise M Wolf, and Adam P Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, 2002.

[137] Jonathan M Raser and Erin K O'shea. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, 2005.

[138] Carl Edward Rasmussen. Gaussian processes in machine learning. In *Advanced lectures on machine learning*, pages 63–71. Springer, 2004.

[139] Angela Re, Davide Corá, Daniela Taverna, and Michele Caselle. Genome-wide survey of microrna–transcription factor feed-forward regulatory circuits in human. *Molecular BioSystems*, 5(8):854–867, 2009.

[140] Andrew Reibman and Kishor Trivedi. Numerical transient analysis of markov models. *Computers & Operations Research*, 15(1):19–36, 1988.

[141] Michael Samoilov, Adam Arkin, and John Ross. Signal processing by simple chemical systems. *The Journal of Physical Chemistry A*, 106(43):10205–10221, 2002.

[142] Marc A Schaub, Thomas A Henzinger, and Jasmin Fisher. Qualitative networks: a symbolic approach to analyze biological signaling networks. *BMC systems biology*, 1(1):4, 2007.

[143] René L Schilling. *Measures, integrals and martingales*. Cambridge University Press, 2017.

[144] Jörn M Schmiedel, Sandy L Klemm, Yannan Zheng, Apratim Sahay, Nils Blüthgen, Debora S Marks, and Alexander van Oudenaarden. Microrna control of protein expression noise. *Science*, 348(6230):128–132, 2015.

[145] David Schnoerr, Botond Cseke, Ramon Grima, and Guido Sanguinetti. Efficient low-order approximation of first-passage time distributions. *Physical review letters*, 119(21):210601, 2017.

[146] Adel S Sedra and Kenneth Carless Smith. *Microelectronic circuits*, volume 1. New York: Oxford University Press, 1998.

[147] Rony Seger and Edwin G Krebs. The mapk signaling cascade. *The FASEB journal*, 9(9):726–735, 1995.

[148] Vahid Shahrezaei and Peter S Swain. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*, 105(45):17256–17261, 2008.

[149] Guy Shinar and Martin Feinberg. Structural sources of robustness in biochemical reaction networks. *Science*, 327(5971):1389–1391, 2010.

[150] Abhyudai Singh and João P Hespanha. Approximate moment dynamics for chemically reacting systems. *IEEE Transactions on Automatic Control*, 56(2):414–418, 2011.

[151] Abhyudai Singh and Joao Pedro Hespanha. Lognormal moment closures for biochemical reactions. In *Decision and Control, 2006 45th IEEE Conference on*, pages 2063–2068. IEEE, 2006.

[152] Abhyudai Singh and João Pedro Hespanha. A derivative matching approach to moment closure for the stochastic logistic model. *Bulletin of mathematical biology*, 69(6):1909, 2007.

[153] Stephen Smith and Ramon Grima. Single-cell variability in multicellular organisms. *Nature Communications*, 9(1):345, 2018.

[154] Elias M Stein and Guido Weiss. *Introduction to Fourier analysis on Euclidean spaces (PMS-32)*, volume 32. Princeton university press, 2016.

[155] Williams J Stewart. *Introduction to the numerical solutions of Markov chains*. Princeton Univ. Press, 1994.

[156] Petre Stoica and Randolph L Moses. *Spectral analysis of signals*, volume 452. Pearson Prentice Hall Upper Saddle River, NJ, 2005.

[157] Timothy J Strovas, Alexander B Rosenberg, Brianna E Kuypers, Richard A Muscat, and Georg Seelig. Microrna-based single-gene circuits buffer protein synthesis rates against perturbations. *ACS synthetic biology*, 3(5):324–331, 2014.

[158] Mukund Thattai and Alexander Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

[159] Mukund Thattai and Alexander van Oudenaarden. Attenuation of noise in ultrasensitive signaling cascades. *Biophysical journal*, 82(6):2943–2950, 2002.

[160] Henk C Tijms. *A first course in stochastic models*. John Wiley and sons, 2003.

[161] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[162] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[163] Aad PA van Moorsel and William H Sanders. Adaptive uniformization. *Stochastic Models*, 10(3):619–647, 1994.

[164] Peter Waage and Cato Maximilian Gulberg. Studies concerning affinity. *Journal of chemical education*, 63(12):1044, 1986.

[165] EWJ Wallace, DT Gillespie, KR Sanft, and LR Petzold. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET systems biology*, 6(4):102–115, 2012.

[166] Denise M Wolf, Vijay V Vazirani, and Adam P Arkin. Diversity in times of adversity: probabilistic strategies in microbial survival games. *Journal of theoretical biology*, 234(2):227–253, 2005.

[167] Verena Wolf, Rushil Goel, Maria Mateescu, and Thomas A Henzinger. Solving the chemical master equation using sliding windows. *BMC systems biology*, 4(1):1, 2010.

[168] Håkan LS Younes and Reid G Simmons. Probabilistic verification of discrete event systems using acceptance sampling. In *International Conference on Computer Aided Verification*, pages 223–235. Springer, 2002.

[169] Christoph Zechner, Georg Seelig, Marc Rullan, and Mustafa Khammash. Molecular circuits for dynamic noise filtering. *Proceedings of the National Academy of Sciences*, 113(17):4729–4734, 2016.