A Stochastic Framework to Model Extrinsic Noise in Gene Regulatory Networks

Ariane Leoni Hofmann

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Mathematics

Reinhard Laubenbacher, Chair
Stanca Ciupe
John A. Burns

July 31, 2012
Blacksburg, Virginia

A Stochastic Framework to Model Extrinsic Noise in Gene Regulatory Networks

Ariane Leoni Hofmann

Abstract

Stochastic modeling to represent intrinsic and extrinsic noise is an important challenge in molecular systems biology. There are numerous ways to model intrinsic noise. One framework for intrinsic noise in gene regulatory networks was recently proposed within the discrete setting. In contrast, extrinsic perturbations were rarely modeled due to the complex mechanisms that contribute to its emergence. Here a discrete framework to model extrinsic noise is proposed. The interacting species of the model are represented by discrete variables and are perturbed to represent extrinsic noise. In particular, they are subject to a discretized lognormal distribution. Additionally, a delay is imposed on the update with a certain probability. These two perturbations represent global extrinsic noise and pathway-specific extrinsic noise. It leads to large variations in the concentration of proteins, which is consistent with an existing continuous way of modeling extrinsic fluctuations. The framework is applied to three different published discrete models: the cell fate of lambda phage infection of bacteria, the lactose utilization system in E. coli, and a signaling network in melanoma cells. The framework captures factors that significantly contribute to the random decision between lysis and lysogeny as well as explains the bistable switch in the model of the lac operon. Finally, a feed-forward loop analysis is conducted by measuring and comparing the noise level in the target protein of feed-forward loops. This analysis reveals the ability of certain feed-forward loops to attenuate or amplify fluctuations, dependent upon various levels of noise. In conclusion, this thesis aims to resolve the question of how the extrinsic noise can be modeled and how biological systems are able to maintain functionality in the wake of such large variations.

Acknowledgements

I am grateful to many people for help, both direct and indirect, in conducting this research. I would like to express my greatest gratitude to my advisor and committee chairman, Dr. Reinhard Laubenbacher, who encouraged and challenged me throughout this program. The thesis could not have been written without his ongoing support and patience. I wish to thank Dr. Stanca Ciupe and Dr. John A. Burns for serving on my committee. Also, I would like to thank Dr. David Murrugarra for fruitful discussions and for sharing his Matlab code.

I would like to thank the German Academic Exchange Service (DAAD) for the financial support throughout the entire year of the program.

I wish to acknowledge many thanks to the Karlsruhe Institute of Technology for giving me the opportunity to be part of this great exchange program. In particular, I thank Dr. Wolfgang Reichel, whose dedication and support inspirits this exchange program. I would also like to express my thanks to Dr. Stefan Kühnlein for encouraging me to pursue mathematical biology.

I am deeply grateful to my family for their continuous support and encouragement. Thanks also to my research group, the Laubenbacher Group of the Virginia Bioinformatics Institute, for their friendship and cordiality. And, last but not least, I would like to thank all my friends for their motivation and faithfulness.

My apologies if I have inadvertently omitted anyone to whom acknowledgement is due.

Contents

List of Figures

List of Tables

Chapter 1

Biological Context

1.1 Introduction

"For the concert of life no-one gets a program." [27]

This statement illustrates the fascination of randomness in biological processes. Even in a population of genetically identical cells, every cell is unique [39]. For example, they can differ in their morphology or in the rate of expression of their genes [39]. The harmonizing of stochasticity and determinism are an inescapable part of life [32]. That is, living organisms must balance between random biochemical interactions and precise development [32].

The observed fluctuations in the concentrations of chemical species between genetically identical cells are indeed larger than expected, which implicates the question of how the processes function properly despite these variations [9]. Therefore, it is of great interest to understand how these large fluctuations arise and how the cells maintain their precise functioning [9]. In fact, the comprehension of stochastic gene expression is said to be one of the most important and exciting challenges in molecular biology [29]. A striking example of the significance of noise in gene expression is given by a bacterium: The stochastic expression of a specific protein actually predicts the life span of C. elegans [32].

Mathematical models help to gain understanding of biological systems. For example, by knowing how to represent the extrinsic variations in a single protein, the effects of extrinsic noise on whole gene regulatory networks can be analyzed. This in turn enables elucidation of how the cells can function properly with the noise. Generally, the notion "modeling" refers

to the mathematical representation of a system. The notion "simulation" denotes its computational solution [16]. One can distinguish between two classes of models, namely discrete and continuous models [26]. Continuous models can be systems of differential equations which are of advantage if the purpose of the model is to gain quantitative insights about the system [26, 16]. The downside of this approach is that it requires detailed knowledge of the reaction rates and the concentration of the reacting species, and such information is often difficult to obtain [16]. Additionally, an increase in complexity of the system comes along with an increase in the number of parameters [16]. In contrast, a discrete model enables one to gain insights about the qualitative behavior of a system [16]. The interactions between the chemical species can be visualized graphically and it is not necessary to have detailed information, as is the case for continuous models [16]. This discrete approach provides a description of the dynamics of a network without knowing the exact rate parameters. In turn, the result does not give us precise information about the concentrations of the involved chemical species, but rather represents the qualitative states of the system. That is, the outcome of the simulation tells us for example whether a certain protein is absent or present in low or in high concentrations. Hence, the decision between continuous or discrete models is dependent on whether one is interested in quantitative or qualitative results, and whether the required data is available [11]. Another classification of mathematical models is that of stochastic and deterministic models. A stochastic model of gene regulatory networks more accurately portrays a real biological system. While deterministic models can only predict the average behavior of the system, stochastic models are able to capture the observed variability and thus represent a more realistic model [24].

Here, we are interested in the qualitative behavior of systems and thus work with discrete models, where each variable of the model can have integer values in a finite set $\{0, 1, ..., p-1\}$, $p \in \mathbb{N}$. Also, the time is discrete and the variables change their values at each time step according to specific update rules as explained in Chapter 2. To introduce stochasticity and thereby enable more realistic modeling, we embed the respective model into the framework for extrinsic noise. This allows for comparison between the deterministic and stochastic dynamics of the system and analysis of the relations between network structures and robustness.

The thesis is structured as follows. Section 1.2 describes the difference between intrinsic and extrinsic noise and names the sources of the latter.

Chapter 2 provides the stochastic framework to model extrinsic noise and is divided into three sections. Section 2.1 specifies the biological background necessary for the framework, which is then introduced in Section 2.2. The chapter ends with remarks about the lognormal distribution and seeks to answer the question of how it emerged in biological systems.

Chapter 3 proceeds with specifying general characteristics of extrinsic noise. In particular, Section 3.1 is dedicated to the questions of how to measure fluctuations and develops a measure of dispersion, the AMSD, which will be used in Chapter 5. Section 3.2 aims to describe potential methods to distinguish extrinsic and intrinsic noise. Section 3.3 illustrates the effect of extrinsic noise on the expression level of a single protein. Finally, Section 3.4 closes with examining the advantages and disadvantages of extrinsic noise and noise in general.

Chapter 4 consists of three applications of the stochastic framework to model extrinsic noise. More precisely, Section 4.1 applies the framework to the model of infection of bacteria with phage lambda. Section 4.2 continues with the application to the model of the lactose utilization system in E. coli. Finally, the framework is applied to a signaling network in melanoma cells in Section 4.3.

Chapter 5 begins with Section 5.1, which discusses the effect of the extrinsic noise on the dynamics of networks as a whole. An analysis of certain network motifs is provided in Section 5.2. The thesis closes with a conclusion in Section 5.3.

The discrete stochastic framework to model extrinsic noise is a new concept developed in this thesis. Another new concept is the AMSD, which is as a measure of dispersion for discrete models. The AMSD is a modification of the multiplicative standard deviation. It was suggested to use the multiplicative standard deviation to measure the magnitude of variations when the observed variables obey a lognormal distribution [19]. We conduct a feed-forwad loop analysis and use the AMSD to compare the transmitted variations. Furthermore, we also measure the noise level with the AMSD in a network motif of the Wnt network and the lambda phage network.

1.2 Definition and Sources of Extrinsic Noise

The notions of intrinsic and extrinsic noise were introduced by Elowitz et al. [8]. The distinction between either source of noise is not always clear and depends on the definition of the network of interest. The term intrinsic or extrinsic refers to fluctuations inside the system or in the environment of the network, respectively [29]. Hence, it depends on the distinction between inside and outside of the system [10]. The environment of the system comprises not only the extracellular surroundings, but also the intracellular environment.

That is, extrinsic variations are caused by environmental factors that can be outside the cell, e.g. the temperature, or inside the cell such as the number of present reactant molecules [23]. For example in a network of proteins, fluctuations in the abundances of mRNAs count as extrinsic noise, since they indirectly cause variations in the protein concentrations [29]. Thus, extrinsic noise also consists of the fluctuations in the amount of other molecular components inside the cell [44]. Extrinsic noise is caused by variations in the abundances of molecular species that are not a node in the network, but their concentration differs between cells and causes cell-wide variations in the amount of proteins and mRNAs of interest [44]. The system of interest inherits extrinsic noise by interacting with other stochastic systems in the intra- or extracellular environment [38]. The expression level of a gene depends on the amount and location of regulatory proteins, which is why variations in their abundance causes extrinsic perturbations in the concentration of the gene [8]. Extrinsic variations mostly alter the parameters of a network. For example the rate of translation fluctuates due to different numbers of free ribosomes [28]. In fact, extrinsic variations contribute mainly to the total cellular noise in prokaryotes [8] and eukaryotes [34].

Recall that gene expression involves the enzyme RNA-polymerase that transcribes the gene on the DNA into mRNA [45]. The latter one is then translated into proteins, which requires enzymes called ribosomes [45]. The promoter is a part of the DNA, which controls transcription [45]. Transcription factors are proteins that can bind to operators, which are regions in the DNA, and activate or inhibit gene expression [45].

External factors that cause variations include fluctuations in general transcription factors [4], and enzymes indirectly involved in the network, cell-cycle-dependent periodic changes, random segregation of protein copies at mitosis [29, 17], variations in the amounts of RNA-polymerases, ribosomes, nucleotides [35], and amino acids [45], the cell's growth rate [38], as well as the cell's environment [44]. The cell can be exposed to different environments, which can vary for example in temperature and in the amount of nutrients available [45]. Not only differences in the cell size, but also variations in the cell's shape and organelle composition contribute to extrinsic fluctuations [21]. Furthermore, the chromosomal location of a gene counts as an origin of extrinsic noise [4, 47].

To further refine this definition, extrinsic noise can be subdivided into global noise and transmitted pathway-specific noise [21] [35]. The latter includes variations in the concentration of the protein or gene of interest that are caused by factors 'upstream' of this gene [21], such as disparity in the signaling pathway [21]. For example, if a transcription factor is regulated by an enzyme, the fluctuations in the concentration of this enzyme result in extrinsic noise in the concentration of the transcription factor [21]. Pathway-specific noise includes intrinsic and global variations in upstream components [30] and can be different to each node of

the network. The intrinsic noise in these components can potentially be averaged out over time and thus depends on the lifetime of the respective protein [30]. Additionally, genes can have a different susceptibility to the upstream noise [30]. The same upstream regulator can govern several genes such that fluctuations in the concentration of the regulator affect several proteins equally [15]. Global extrinsic noise encompasses all nonspecific fluctuations, potentially affecting all network nodes equally [38]. As an example, this includes differences in rates of the basic reactions [35] or the metabolic state of the cell [21] that have an impact on the expression of all genes [35]. The global noise is sometimes used as a synonym for extrinsic noise [6, 47].

In contrast, intrinsic noise is produced by differences in the timing of individual reactions, which arise due to intermolecular collisions caused by the random movement of molecules [38]. Intrinsic noise acts gene-specific [21]. It reflects the inherent stochasticity in biochemical reactions, due to the laws of physics and chemistry [37]. This can cause, for example, random transcription and translation initiation [44]. Intrinsic noise becomes especially pronounced when low numbers of protein molecules are involved [4, 26, 8]. The reacting species diffuse and collide frequently with other molecules, which in turn changes their propensity to react [33]. Researchers showed that the random transitions of the promoter between active and inactive states, resulting in mRNA bursts, is due to intrinsic noise [31] and this factor is specific for each gene [4]. The molecules involved in transcription, gene regulation, and signal transduction are mostly present in low molecule numbers [23]. Thus the timing of reactions can vary significantly. The fact that they are only available in smaller amounts reflects the cell's attempt to minimize its metabolic cost [32]. Especially the number of mRNA molecules in a cell is low, compared to the number of protein molecules [4]. As an example, certain types of mRNAs that encode transcriptional regulators in yeast cells are present in average with 0.005, 0.3 and 1 copy per cell [4]. In fact, it was demonstrated experimentally that intrinsic noise increases when the number of molecules involved decreases [8]. Intrinsic noise is sometimes referred to as thermal noise, since it emerges due to the thermal movement of the molecules [21, 17, 37, 43].

The chromosomal position of a gene has an impact on the gene's expression level since alterations in the chromatin structure determine whether the gene's promoter is accessible for potential transcription factors [4, 3]. Chromosomal positioning of a gene was classified as an extrinsic noise source [47]. Those genes, being controlled by the same promoter exhibit correlated fluctuations in their expression levels [4]. The concentration of transcription factors present in the intracellular environment influence the state of an operator of a certain gene [2]. Operator state fluctuations propagate to the expression level of the protein and thus variations in the concentration of transcription factors can be considered as extrinsic noise [2].

One may also consider the distinction of intrinsic and extrinsic as origin and propagation of noise [28]. Fluctuations in the concentration of regulatory components are themselves caused by probabilistic processes in the cell and thus propagate this noise to the proteins of the network. For example, a gene encoding a repressor could stochastically switch on or off, causing intrinsic fluctuations in the concentration of repressors and propagate to the target protein as extrinsic fluctuations [28].

Pertaining to the random segregation of molecules at cell division, the uneven partitioning, specifically of mitochondria, can cause large variations in the number of protein molecules between the two cells [14]. In addition to variations in their number, the mitochondria also vary in their functionality [14]. Since mitochondria are the producers of ATP, their variations affect biochemical reactions and gene expression levels [14]. Additional evidence for the impact on noise levels was discovered in an experiment: A population of cells was treated with antioxidants, which are known to enhance mitochondrial functionality. As a result, the noise level significantly decreased [14]. The variations in mitochondrial mass and functionality contribute to the extrinsic noise and dominate over intrinsic noise [14].

Moreover, the cell cycle constitutes a source of extrinsic noise, since the protein production rates oscillate with the cell cycle [36]. At the end of the cell cycle, the gene copy number is on average twice as high compared to the beginning of the cell cycle [36]. This type of variation in protein molecule numbers is thought to be predictable since it recurs periodically as opposed to the other purely stochastic sources of extrinsic fluctuations [35]. Even in this fixed process, there can be significant variations in the cell growth rate contributing to a great amount of extrinsic fluctuations in the concentration of proteins [47]. The concentration of proteins is determined by the rate of synthesis, the rate of degradation, and the dilution caused by the cell growth [47]. Among these factors, the dilution due to the noisy cell growth rate contributes a fundamental part to variations in the concentration of proteins [47].

In summary, extrinsic noise arises due to fluctuations in the amount of any molecular component that is indirectly involved in the network of interest, but is not a network node itself. These components may be intra- or extracellular. It furthermore comprises fluctuations caused by the system's environmental circumstances, such as the cell growth rate or the temperature. Extrinsic fluctuations act indirectly and globally, i.e. they potentially affect several nodes equally. Therefore, extrinsic variables would be the concentration of transcription factors, enzymes, ribosomes, polymerases, nucleotides, and amino acids, the chromosomal location of a gene, cell-cycle-dependent changes, the cell's size, shape, organelle composition and growth rate, random partitioning at mitosis, extracellular nutrients and the temperature.

Chapter 2

The Discrete Modeling Framework

2.1 Biological Reasoning

The concentration of a protein can vary due to extrinsic fluctuations. To appropriately model such fluctuations, it is important to find out how they affect the distribution of the protein concentrations. As described in the previous section, many factors contribute to extrinsic variations in protein abundances. Some extrinsic variables can be grouped together to a representative extrinsic variable, because their overall effect on the network can be modeled in the same way. To monitor the concentration of a protein in a single cell, it is common to use fluorescent reporters [45]. These are fluorescent proteins that are expressed together with the protein of interest and are detectable with a fluorescence microscopy [21].

The first extrinsic variable E_1 represents fluctuations in the amount of polymerases, ribosomes, nucleotides, and amino acids as well as the cell growth rate. It has a lognormal distribution and affects all nodes in the network, i.e. it acts globally. This type of distribution shows a longer right tail, indicating that greater chemical abundances occur more frequently compared to symmetric distributions [9]. Researchers investigated protein concentrations in cells and found a distribution, which was skewed to the right and thus can be fitted well with a lognormal distribution [36] [9] [47] [17] [37]. The variations were found to be mostly of extrinsic origin [36] [47].

Rosenfeld et al. [36] defined the gene regulation function as the function whose value represents the downstream protein production rate given a certain concentration of transcription factor. The authors built a so-called '$\lambda - cascade$' in E. coli, which consists of the λ repressor

and a downstream gene. The former is coupled on the genome with yellow fluorescent protein and the latter with cyan fluorescent protein. In this way, it was possible to monitor the expression levels of both, the repressor and the target protein. When exploring the deviations from the mean gene regulation function, the investigators found a standard deviation of 0.55. To avoid cell-cycle-dependent fluctuations, the protein production rate was normalized to the average cell-cycle phase, which resulted in a remaining standard deviation of 0.4. In fact, the protein concentration was distributed lognormally around the mean gene regulation function. To assess whether these fluctuations were mainly due to intrinsic or extrinsic fluctuations, the dual-reporter method was applied. Cyan and yellow fluorescent proteins were produced from two identical promoters. As a result, 35% of the variations in protein production rate were of extrinsic origin, and 20% intrinsic. Section 3.2 provides a precise description of the dual-reporter method and explains why it might even underestimate the percentage of extrinsic noise. The experiment was conducted with identified repressor concentrations and the cell-cycle-dependent changes were normalized. Thus, in this setup the extrinsic component of the noise consists of variations in global cellular components. This comprises RNA-polymerases, nucleotides, ribosomes and amino acids. These variables are grouped together into E_1. Due to this discovery, extrinsic fluctuations were modeled with a lognormal distribution by Shahrezaei et al. [38] [39]. They modeled the perturbations with a lognormal random variable with mean one that is multiplied to the extrinsic variables [38]. The next chapter contains an example from their work.

During the cell cycle, the protein concentrations vary periodically. In particular, the protein synthesis rate is approximately doubled in the course of a cell cycle [36]. However, taking into account the gain in cell volume these periodic variations become negligible since the proteins are in turn diluted out. Nevertheless, the increase in the amount of proteins during the cell cycle is not proportional to the growing cell volume [47]. In fact, Tsuru and colleagues in [47] argued that the protein production rate and the cell growth rate are decoupled and showed that the cell growth is itself a noisy process. Therefore, protein concentrations can vary due to variations in the cell growth rate. The dilution of proteins in a bacterial cell is mostly caused by the cell growth and protein degradation contributes a much smaller part. The authors investigated in which way the variations in the cell growth rate affect the protein concentrations. A gene circuit in E. coli was constructed comprising a repressor coupled with red fluorescent protein and the target gene co-expressed with green fluorescent protein. The concentration of the repressor was held constant. Thus the fluctuations in the green fluorescent protein were only due to global noise without transmitted noise. Since the green fluorescent protein concentrations increased with the growing cell volume, the ratio of protein concentration to cell volume was considered in order to avoid cell-cycle-related effects. The resulting distribution of the green fluorescent protein concentration exhibited a long right tail. It was furthermore plotted on a logarithmic scale displaying a distribution

which is approximately normal. This suggests a lognormal protein distribution. Tsuru et al. attributed the observed variations in protein concentrations to noise in the cell growth rate. Therefore, the effect of noisy cell growth rates on the amount of proteins can be absorbed into the extrinsic variable E_1, which obeys a lognormal distribution. The authors estimated that noise in the cell growth rate contributes to 27% of the total extrinsic noise.

Such lognormal protein distributions were reported by several other investigators [17, 9, 37], and mRNA distributions were also found to have a long right tail [31], which matches a lognormal distribution. Furthermore, protein concentration distributions in a growing bacterial population were examined, and a crossover from a long-tailed to a Gaussian distribution was found in [17]. From experiments, a lognormal distribution was documented for smaller protein concentrations and a Gaussian distribution for large amounts of proteins after approximately 15 hours of growth [17]. This growth hour corresponds to the late stationary phase of bacterial growth [2]. In the stationary phase, the protein distribution is more symmetric and might fit a Gaussian distribution [17].

Moreover, it was stated that bacterial cells in the exponential growth phase generally exhibit a lognormal protein distribution [9]. At this point, it is worth mentioning that the growth of a bacterial population is divided into several phases [1]: The first part is the lag phase, in which the cell mass only increases a little. The exponential growth phase then is characterized by unrestricted growth of the population. In this phase, the cells grow efficiently, having reached their critical growth speed, which is the maximal growth speed after which a cell would not grow any faster [9]. When the growth rate ceases and the metabolic activity declines, the stationary phase is reached [1].
Furusawa et al. restricted their experiments to cells that were in the exponential growth phase [9]. They showed experimentally and in silico that all chemical abundances in these cells obey a lognormal distribution. The authors conducted a variety of simulations and stated that theses results are true in general, independent of the details of the model such as the structure of the network. It was concluded that the lognormal protein distributions generally occur in cells that are in the exponential growth phase. To confirm these laws, experiments with E. coli were conducted. The protein distributions of the cells in the exponential phase of growth were measured. The protein abundances were normalized to the volume of each cell to eliminate the fluctuations due to differences in cell size. As a result, the protein distributions were lognormal. This distribution is not related to variations in plasmid copy number, since it also occurred when the gene was located on the genome.

Thus, whether or not a lognormal distribution truly reflects the protein distributions in a cell might depend on the stage in the bacterial growth phase. It was experimentally confirmed

that the lognormal distribution holds until 15 hours of bacterial growth [17]. For this modeling framework, we assume that the system of interest resides in these earlier growth phases.

To sum up, we define the representative extrinsic variable E_1 as lognormally distributed. It reflects the effects of fluctuations in global extrinsic factors such as the concentration of RNA-polymerases, nucleotides, ribosomes and amino acids as well as the noisy cell growth rate. Hence E_1 affects all network nodes equally.

It remains to consider the gene-specific extrinsic noise. This includes variations in the concentration of transcription factors or enzymes that are required for the expression of a certain protein [21], as well as the chromosomal positioning of a specific gene and the current chromatin structure, which affects its accessibility for RNA-Polymerases and transcription factors [4]. Additionally, the cells may have different shapes and constellations of the organelles [21] which may have an impact on the biochemical reactions. When considering how genes respond to variations in upstream transcription factors it is important to note that the transmitted global noise can antagonize the effect of the global noise on the gene of interest [30]. For example if the global noise calls for an overexpression of all genes, it also leads to an overexpression of a repressor. This in turn downregulates the gene of interest, counteracting the global call for overexpression on that gene. Furthermore, the susceptibility of a gene to upstream fluctuations can vary significantly, depending on the amount of active transcription factor [30]. Variations in the amount of transcription factors or enzymes might cause differences in the timing of biochemical reactions, independently of the differences in the timing caused by intrinsic noise. Various shapes and organelle constellations of cells as well as the location of a gene on the genome together with the current chromatin structure might also cause individual reactions to occur at different points of time. Thus in this modeling framework, the variations in protein concentrations due to the varying amounts of potential transcription factors and enzymes as well as due to the different chromosomal locations, cell shapes, and organelle compositions are represented by a probability that the update of a certain network variable does not happen. As will be specified in the next Section, each variable of the network x_i has a probability d_i that it remains at the current expression level at the next time step instead of being updated. The probabilities d_i are kept quite small in order to not distort the original dynamics too much.

Different cell sizes might also affect the timing of biochemical reactions. It is intuitive to assume that the reacting molecules have lower probabilities of collision in large cells, as compared to smaller cells. Thus, the probability of a delay in the updates could be increased when modeling larger cells. By contrast, in order to model a smaller cell, the probability for a delay could be decreased. However, it is assumed that the probabilities for a delay are

already quite small. In order to model a smaller cell, there might be the need for an additional adjustment. Investigators reported that proteins are diluted by the cell growth [47]. This stimulates the idea that proteins in larger cells tend to be distributed more sparsely in the cell. Whereas in smaller cells the proteins are more densely concentrated. Thus, the level of protein concentration could be increased in a model of smaller cells. The application of the modeling framework to the lambda phage model in Chapter 4 includes an attempt to model smaller cells. In this case, the discrete protein concentration levels were increased by one after each update for smaller cells. Additionally, the probabilities for a delay where decreased to 1% and 0%. It might be a simplified way of representing smaller cells, but in the case of the lambda phage model, the results of the simulations indeed exhibited a property of smaller cells, which was observed experimentally [40]. Note that the modification to increase the concentration levels by one is only appropriate in a system with multivalued variables. It would lead to an overrepresentation in a Boolean model.

The network of interest may also include extracellular nutrients, such as the network of the lac operon model. It was reported that nutrients provided externally to a cell or imported from the outside obey a normal distribution [9]. Hence, the nodes in the respective network, which represent extracellular nutrients, will be perturbed by an extrinsic variable E_2, which is normally distributed. This distribution was indeed used for external nutrients in a model of the lac operon [48].

Fluctuations in protein concentrations due to uneven segregation at mitosis have a normal distribution and affect all nodes in the network at cell division. It seems natural to assume that the proteins are distributed evenly to the two daughter cells at cell division. However, there are fluctuations in the amounts of proteins after segregation. These variations are termed cell partitioning noise [17]. In fact, the cell partitioning noise leads to a Gaussian distribution at the beginning of the cell cycle [17]. It was observed experimentally that the proteins are distributed binomially to each daughter cell at mitosis, which reflects equal probability for each molecule to go to one of the two daughter cells [36]. In fact, a binomial distribution was used previously to represent the segregation of proteins and mRNAs at mitosis [44, 17, 21].

The effect of uneven segregation of mitochondria was investigated and it was found that it significantly varies transcription rates and thus gene expression levels, as well as the cell cycle length [14]. Depending on the number and functionality of mitochondria, the ATP concentration in a cell in turn affects the rates of biochemical reactions and the cell growth rate [14]. The volume of the daughter cell and its mitochondrial mass right after cell division are normally distributed. This represents equal probability to be assigned to either one of

the two daughter cells. For the mRNAs and proteins, a binomial distribution was chosen, which, for high values of n, can be transformed into a normal distribution [14] according to the relation

$$B(n,p) \simeq N(np, \sqrt{np(1-p)}),$$

where $p = \frac{1}{2}$. Thus a normal distribution is suitable for modeling the effect of cell partitioning noise. However, in many models of gene regulatory networks, the experimental data is not related to the cell cycle. In such a case, the effects of the cell partitioning noise are excluded from the model. Here, we will consider the model of the infection of bacteria with phage lambda, the model of the lactose utilization system in E. coli, and a signaling network in melanoma. These models are based on data that is not related to the cell cycle, and thus we will neglect the cell partitioning noise in the modeling framework.

Furthermore, it should be taken into account that the growth rate of bacteria generally increases with a rising temperature, see [13]. The temperature-growth rate curves from five different types of microorganisms including E. coli showed an increase in the growth rate with temperature within certain temperature ranges. In particular, E. coli starts growing at approximately 8 °C, and accelerates the growth with rising temperatures until around 48 °C, which is the maximal growth temperature. This reflects a speeding up in reactions rates at higher temperatures, i.e. faster transcription, translation and accelerated interactions among the proteins. To incorporate this effect into the modeling framework, the probabilities d_i (i=1, ..., n) for a delay can be adjusted according to the temperature. That is, if these probabilities are enhanced, it is more likely that an update of a node does not happen. This delay represents a slower reaction time and thus a slower protein synthesis and degradation rate. As the temperature increases, the probabilities become slightly smaller, and inversely, as the temperature decreases, the probabilities slightly rise. Thus, higher probabilities represent a higher inertia in the network and accordingly a greater chance for a delay in reactions, resulting in slower reactions, which reflects low temperatures.

Not only the distribution of fluctuations, but also the time scale over which the fluctuations occur is an important characteristic. In fact, extrinsic variations last significantly longer than intrinsic ones [36, 21]. The autocorrelation time is defined as the time a fluctuation persists, namely until the perturbed gene expression rate returns to its mean [21]. Extrinsic fluctuations feature an autocorrelation time of around 40 minutes, which is about one cell cycle in fast dividing bacteria, such as E. coli, whereas intrinsic variations only last for approximately nine minutes [21]. Extrinsic noise causes the protein concentrations to vary considerably from their mean expression levels over relatively long periods [36]. If a cell overexpresses a certain protein, it will probably maintain this higher synthesis rate for the rest of the cell cycle [36]. In light of this property of extrinsic variations, it seems to be

suitable to alter the extrinsic impact on the network only for every few updates rather than for each update of the network. Here, this is modeled by changing the extrinsic noise every fourth update. This stems from the stochastic framework to model intrinsic fluctuations [26], where the impact of intrinsic noise is different at each time step. Assuming that extrinsic perturbations last approximately four times longer than intrinsic ones [21], we change the extrinsic variations every four time steps.

2.2 The Algorithm

In accordance to the modeling framework for intrinsic noise introduced by Murrugarra et al. [26], let the variables $x_1(t), x_2(t), ..., x_n(t)$ denote the concentration of n chemical species at time t. The modeling framework is discrete in time and in the concentration of the variables. That is, each variable $x_i(t)$ can take on values in some discrete finite set X_i (i=1,2,...,n), and its values change over time $t = 1, 2, ...$ The value of each variable is discrete and represents the current concentration of the respective variable. For example $X_i = \{0, 1, 2\}$ would indicate that the variable x_i can be absent ($x_i(t) = 0$), present in low concentration ($x_i(t) = 1$), or present in high concentration ($x_i(t) = 2$). Depending on the respective context, the various levels could also represent different conditions. For instance, instead of the concentration, it could also be distinguished among different locations of proteins (cytoplasmic/in the nucleus) or among different levels of activity (inactive/low active/high active), see [11].

A discrete dynamical system of the variables $x_1(t), x_2(t), ..., x_n(t)$ is defined as a mapping

$$f = (f_1, f_2, ..., f_n) : X \longrightarrow X$$

where $X = X_1 \times X_2 \times ... \times X_n$ and each coordinate function f_i is a function of $x = (x_1, x_2, ..., x_n)$ and dictates the deterministic value of the variable x_i at the next time step, that is $x_i(t + 1) = f_i(x_1(t), x_2(t), ..., x_n(t)) = f_i(x(t))$ and t denotes the time [26].

In the modeling setup described above, the coordinate functions describe the deterministic update of the respective variable. After identification of the deterministic update of the system, the states are perturbed with extrinsic fluctuations. The extrinsic variations are imposed on the network nodes first by introducing a probability for each variable to maintain its current expression level, and second by multiplying each variable with a representative extrinsic variable E_1 or E_2. Recall that the extrinsic random variable E_1 is lognormally distributed. As described in Section 2.1, it was discovered experimentally that global extrinsic noise results in protein distributions with a longer right tail, which match lognormal

distributions [36, 9, 47]. Such global factors were found to be, for example, varying concentrations of RNA-polymerases, nucleotides, ribosomes and amino acids, as well as the noisy cell growth rate. By multiplying each variable with the lognormal distributed extrinsic variable E_1, which has the mean one, a protein distribution with a longer right tail is obtained. The mean is still the original deterministic value. Hence this approach is appropriate to represent the global extrinsic noise. The external nutrients of a cell were said to be normally distributed [9], and thus they are multiplied by the normal random variable E_2, which has mean one. Therefore, the external nutrients obey a normal distribution with the mean being the deterministic update. After the perturbation of each variable with either E_1 or E_2, the states are rounded to integer values in order to maintain discrete states. The probability for a delay d_i for the update of variable x_i $(i = 1, 2, ..., n)$ represents the chance that the respective update may not happen. As explained in Section 2.1, such a delay is expected to represent a delay in the signaling pathway of the specific protein. This includes varying concentrations of gene-specific transcription factors and enzymes that are required for the gene's expression. Also, the location of a gene on the chromosome influences its accessibility for transcription factors and RNA-polymerases, and thus could be responsible for a delay. Furthermore, different cell sizes, shapes and organelle compositions can result in longer or shorter signaling paths, such that this factor is also expected to cause potential delays. The precise algorithm is described below.

Discrete Dynamical Delay System with Extrinsic Noise

Let $p_i \in \mathbb{N}$ ($p_i \geq 2$, $i = 1, ..., n$) be the number of possible expression levels that the variable x_i can have ($i = 1, ..., n$). Let $x(0) = x_0 = (x_1(0), x_2(0), ..., x_n(0))$ be the initial state of the system. Set $\tau = 0$ and repeat the following for increasing τ until $steps_{max} \in \{\tau, \tau + 1, \tau + 2, \tau + 3\}$, where $steps_{max} \in \mathbb{N}$ is the number of time steps of the simulation.
1. At time step $\tau \in \mathbb{N}_0$, a sample of the extrinsic variable $E_1(\tau)$ is taken, which is lognormally distributed with mean one. Also, a sample of $E_2(\tau)$ is taken, which obeys a normal distribution with mean one.

---

2. For the timesteps $t = \tau, \tau+1, \tau+2, \tau+3$, and as long as $t \leq steps_{max}$, the following is repeated:
2.1 The deterministic update $f_i(x(t))$ for each network variable x_i is determined ($i = 1, ..., n$).
2.2 It is decided for each node x_i ($i = 1, ..., n$), whether the update is on, or whether it is delayed. Here the probability for the update to take place is $1 - d_i \in [0, 1]$, whereas the update does not happen with probability $d_i \in [0, 1]$ ($i = 1, ..., n$). Hence

$$x_i(t+1) = \begin{cases} x_i(t), & \text{with probability } d_i \\ f_i(x(t)), & \text{with probability } 1 - d_i. \end{cases}$$

In the case of a delay for a certain network variable, the respective variable will not undergo steps 2.3, 2.4 and 2.5 of the algorithm.
2.3 If there is a node x_k ($k \in \{1, 2, ..., n\}$), which represents an external nutrient, its value is multiplied by $E_2(\tau)$, that is $x_k(t+1) := f_k(x(t))E_2(\tau)$.
2.4 The state of all other network nodes $x_i(t)$ is multiplied by $E_1(\tau)$, that is $x_i(t+1) := f_i(x(t))E_1(\tau)$.
2.5 The values of each node are rounded to keep them discrete. In particular, the threshold for a variable to be rounded to the next higher integer is a decimal place ≥ 0.5, and otherwise, it is rounded to the next smaller integer. Additionally, the values of each node x_i are kept within the range of possible expression levels $\{0, 1, ..., p_i - 1\}$ ($i = 1, ..., n$). Namely, if the resulting value of $x_i(t+1)$ ($i = 1, ..., n$) is greater than $p_i - 1$, the value is set to $p_i - 1$, and if $x_i(t+1)$ ($i = 1, ..., n$) is smaller than zero, the variable is set to zero. If $t < \tau + 3$, set $t := t + 1$ and go back to 2.

---

3. If $\tau + 3 \geq steps_{max}$, the algorithm ends. If $\tau + 3 < steps_{max}$, set $\tau := \tau + 4$ and go back to step 1.

The update $x_i(t+1)$ $(i = 1, ..., n)$ for a network variable $x_i(t)$ may also be formulated in mathematical terms. Here, we assume that the variable x_i is not an external nutrient, i.e. it is not multiplied by E_2, but by E_1. Let $\tau \in \mathbb{N}_0$, t $\in \{\tau, \tau+1, \tau+2, \tau+3\}$ and $i \in \{1, ..., n\}$. Let $\delta(t)$ be a random variable, which is uniformly distributed on the interval [0,1], that is $\delta \sim \mathcal{U}([0,1])$. Let $d_i \in [0,1]$ be the probability for a delay in the update of node x_i. The update $x_i(t+1)$ for node x_i reads as

$$x_i(t+1) = \begin{cases} x_i(t), & \text{if } \delta(t) < d_i, \\ \min\{p_i - 1, \lceil E_1(\tau) f_i(x(t)) \rceil\}, & \text{if}(E_1(\tau) f_i(x(t)) - \lfloor E_1(\tau) f_i(x(t)) \rfloor) \geq 0.5 \\ & \text{and } (\delta(t) \geq d_i), \\ \min\{p_i - 1, \lfloor E_1(\tau) f_i(x(t)) \rfloor\}, & \text{if } (E_1(\tau) f_i(x(t)) - \lfloor E_1(\tau) f_i(x(t)) \rfloor) < 0.5 \\ & \text{and } (\delta(t) \geq d_i), \end{cases}$$

where the number of possible states p_i can be different for each network variable. If the network variable x_i represents an external nutrient, $E_1(\tau)$ would be replaced by the normal random variable $E_2(\tau)$ in the equation above.
Note that this modeling framework allows adjustments to various temperatures, cell shapes, organelle compositions by altering the probabilities for a delay d_i $(i = 1, ..., n)$. This probability may be different for each network node to reflect gene-specific differences such as fluctuations in the amount of specific transcription factors and enzymes or the different individual locations of the genes on the chromosome. Thus, by choosing different probabilities for a delay it is possible to fine-tune the susceptibility of the genes to the upstream noise. It is furthermore possible to vary the level of the global extrinsic noise simply by choosing different variances for the lognormal distribution. Thus, even during a simulation, the magnitude of extrinsic perturbations can be changed.

The lognormal perturbation is represented by the product of the deterministic update with a lognormal random variable with mean one. The resulting distribution shows larger variations for higher deterministic update values (upper panel of Figure 2.1). Another possibility for the lognormal perturbation would be to take a sample of a lognormal random variable, whose mean is the deterministic update. This would lead to the same magnitude of variations for lower and higher deterministic values (lower panel of Figure 2.1). These two classifications of lognormal perturbations ultimately yield the same mean. However, for the upper panel, a much greater variability is produced for larger protein expression levels. In fact, it was observed experimentally that the variations for higher protein concentrations are indeed larger [47], which suggests the first way of modeling the lognormal perturbation. Additionally, it was shown in simulations that the standard deviation increases with the

average protein expression level [9]. Intrinsic noise is expected to decrease with the increasing protein concentrations [4]. By contrast, extrinsic noise typically leads to the same ratio of standard deviation to the mean [28], which implies greater variability for higher mean protein concentrations. The ratio of standard deviation to the mean is termed coefficient of variation and is explained in Section 3.1.

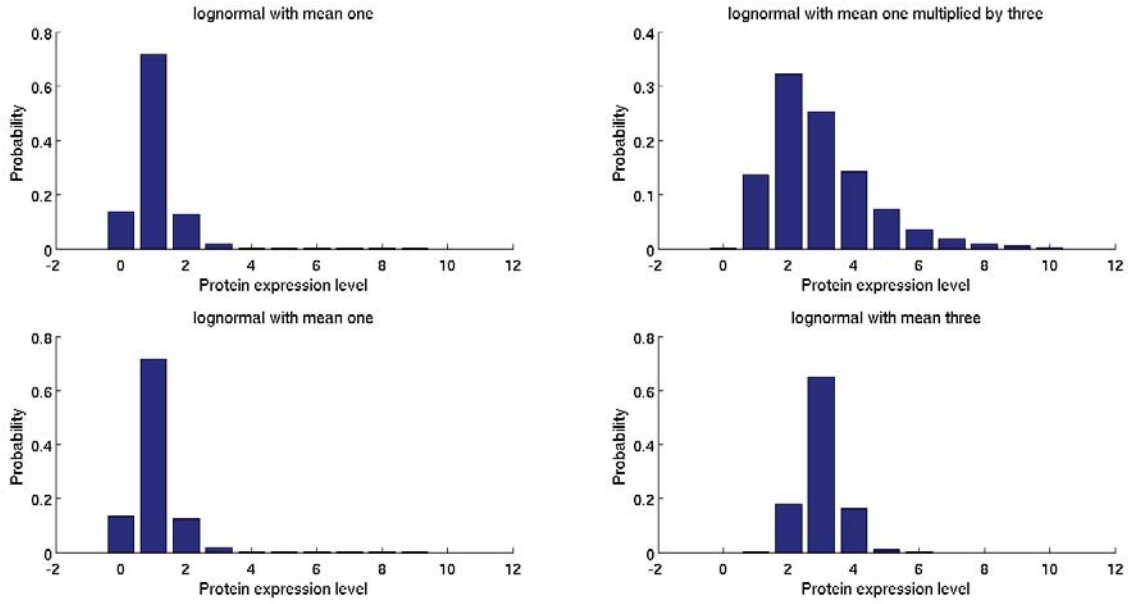

Figure 2.1: In the upper panel, the mean of the lognormal distribution is one and the variance 0.3. On the upper right, the distribution is multiplied by three. In the lower panel, the mean is one and three, respectively, and the variance 0.3.

In Chapter 1, it was mentioned that the interactions between the chemical species in gene regulatory networks could be visualized by a directed graph. It is important to explain this graphing method prior to its use in Chapter 4. This directed graph is called wiring diagram or simply network. The nodes are the interacting proteins or genes, i.e. the variables $x_1, x_2, ..., x_n$, and the edges represent the interactions among these molecular species [25]. An edge between two proteins can either be an inhibiting or an activating edge. For example the protein with the exiting edge can be a transcription factor that regulates the respective protein with the incoming edge. Inhibiting edges are usually specified by red arrows or black blocked lines, whereas the activating edges are green or black arrows. Examples for wiring diagrams are given in Chapter 4.

2.3 The Lognormal Distribution

This section introduces the basic concept of a lognormal distribution. Furthermore, the emergence of the observed lognormal protein distributions is examined. A variable X is lognormally distributed with parameters μ and σ^2 if $Y = \ln(X)$ is distributed normally [5]. Thereby, μ and σ^2 are the mean and variance of the normal distribution of Y. Briefly, if $Y = ln(X) \sim \mathcal{N}(\mu, \sigma^2)$, then $X \sim \Lambda(\mu, \sigma^2)$. Since $X = e^Y$, this random variable is always positive.
The probability density function for $X \sim \Lambda(\mu, \sigma^2)$ is given by

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\{-\frac{(\ln x-\mu)^2}{2\sigma^2}\}, & \text{for } x > 0, \\ 0, & \text{for} x \leq 0. \end{cases}$$

The mean m of X reads as

$$m = e^{\mu+\frac{1}{2}\sigma^2} \tag{2.1}$$

and the variance v is

$$v = e^{2\mu+\sigma^2}(e^{\sigma^2} - 1). \tag{2.2}$$

It is possible to compute μ and σ for given v and m. Equations (2.1) and (2.2) are converted such that μ and σ only depend on m and v. This results in

$$\mu = \ln\left(\frac{m^2}{\sqrt{v+m^2}}\right)$$

and

$$\sigma^2 = \ln\left(1 + \frac{v}{m^2}\right).$$

Thus, a lognormal random variable E_1 with mean $m = 1$ and variance v has the parameters $\mu = \ln(\frac{1}{\sqrt{v+1}})$ and $\sigma^2 = \ln(v + 1)$, that is $E_1 \sim \Lambda(\ln(\frac{1}{\sqrt{v+1}}), \ln(v + 1))$.

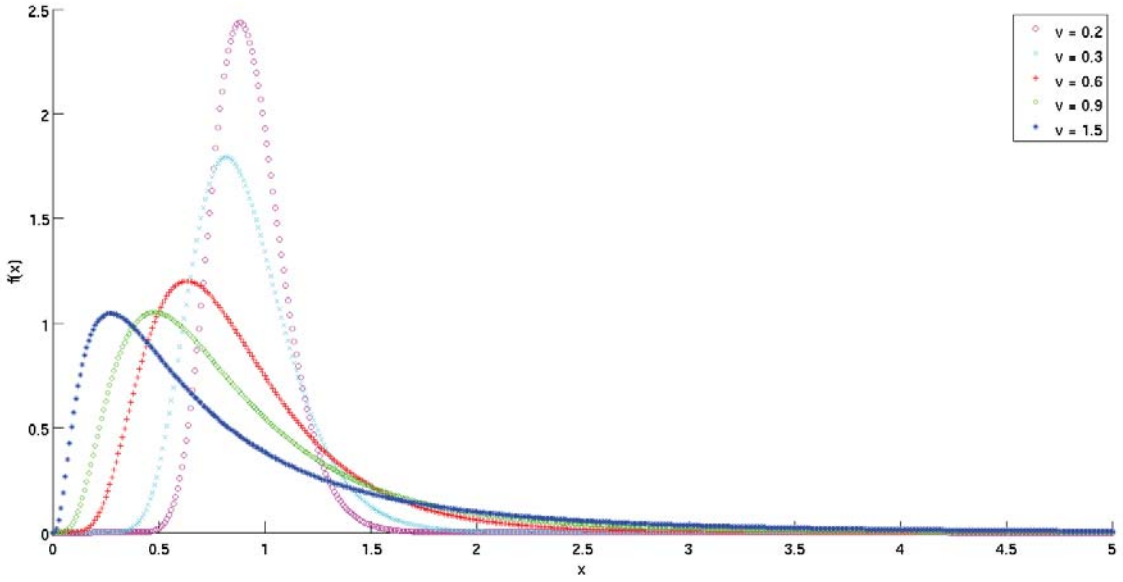

Figure 2.2: Lognormal probability density function of $X \sim \Lambda(\ln(\frac{1}{\sqrt{v+1}}), \ln(v+1))$, where $v = 0.2, 0.3, 0.6, 0.9, 1.5$.

Figure 2.2 shows the probability density function of the lognormal distribution with mean one and different variances.

The left panel of Figure 2.3 illustrates the trajectories of a lognormal random variable with mean one and variance $v = 0.4$ over 100 time steps. The corresponding rounded values are displayed in the central panel of Figure 2.3. They build a discretization of the lognormal distribution. The right panel of Figure 2.3 demonstrates the lognormal probability density function $f(x)$ as well as the associated normal probability density function $g(y)$. This figure is based on a generic simulation of lognormal random numbers to visualize the trajectories and the discretized probability density function.

The probability density function of the lognormal distribution shows a longer right tail, which implies that the abundance of interest occurs more frequently with high concentrations compared to a symmetric distribution [9]. Yet one might intuitively expect a symmetric distribution, where deviations to higher or smaller values are equally likely. Recall that the central limit theorem suggests an approximately normal distribution for the sum of several sources of fluctuations, which are independent and identically distributed [9]. Interestingly, biological phenomena often obey a lognormal distribution [19]. The underlying principle that leads to a lognormal distribution is that the factors do not sum up, but they combine multiplicatively [19]. Indeed, the major operation in the laws of physics and chemistry

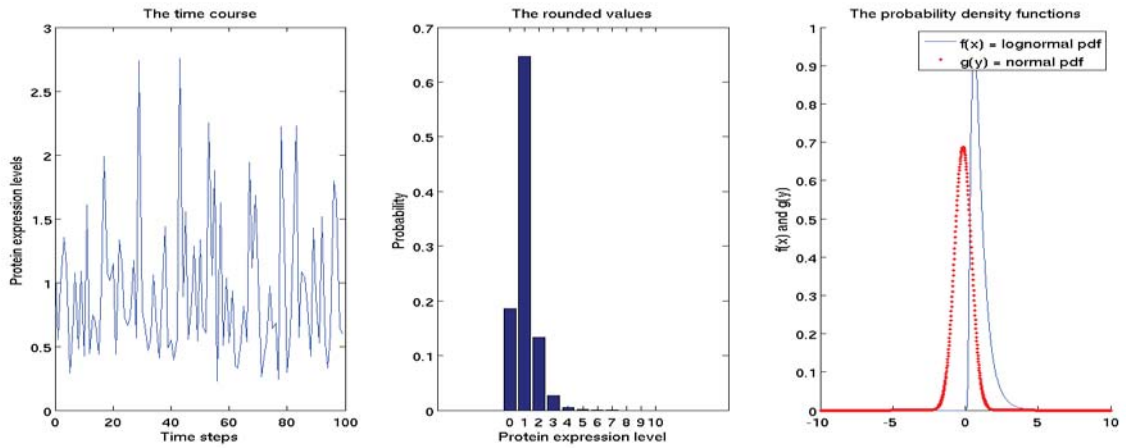

Figure 2.3: The trajectories of a lognormal random variable (left panel) with the corresponding probabilities for the rounded values (central panel) and the plot of the underlying lognormal probability density function $f(x)$ and the probability density function $g(y)$ of the associated normal distribution.

is multiplication [19]. In particular, the variations in chemical reactions are transmitted multiplicatively [9].

The multiplicative version of the central limit theorem states that the product of many independent, identically distributed, positive random variables approximates a lognormal distribution [9]. That is, if $Z_1, Z_2, ..., Z_n$ are positive, independent and identically distributed random variables with

$$E(\ln Z_j) = \mu < \infty, \ Var(\ln Z_j) = \sigma^2 < \infty,$$

for $j = 1, ..., n$, then the product X converges in distribution as

$$X := \prod_{j=1}^{n} Z_j \ \xrightarrow{\text{d}} \ \Lambda(n\mu, n\sigma^2). \tag{2.3}$$

This can be realized by considering the random variable

$$Y := \ln(X) = \sum_{j=1}^{n} \ln(Z_j) \sim \mathcal{N}(n\mu, n\sigma^2),$$

which yields equation (2.3). Note that the product of independent lognormally distributed random variables obeys a lognormal distribution already if two variables are multiplied [19].

An explanation of how a lognormal distribution emerges is given by the so-called "law of proportionate effect" [5]: Let X_0 be a positive variable and consider the recursive equation

$$X_j = X_{j-1}(1 + \epsilon_j), \tag{2.4}$$

where the ϵ_j $(j \geq 1)$ are mutually independent and identical distributed random variables, which are also statistically independent of X_j [5]. This recursion obeys the law of proportionate effect [5]. The nth variable X_n can be written as

$$X_n = X_0 \prod_{j=1}^{n} (1 + \epsilon_j),$$

which is equivalent to

$$\ln(X_n) = \ln(X_0) + \sum_{j=1}^{n} \ln(1 + \epsilon_j). \tag{2.5}$$

The Taylor expansion of $\ln(1 + x)$ around a number a, is given by

$$\ln(1+x) = \sum_{k=0}^{\infty} \frac{\ln(1+a)^{(k)}}{k!}(x-a)^k = \ln(1+a) + \frac{\frac{1}{1+a}}{1!}(x-a) + \frac{-\frac{1}{(1+a)^2}}{2!}(x-a)^2 + \frac{\frac{2}{(1+a)^3}}{3!}(x-a)^3 + \ldots,$$

which for $a = 0$ results in

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} + ... \tag{2.6}$$

Suppose $|\epsilon_j| < 1$ $(j \geq 1)$. Replacing x by ϵ_j in (2.6) yields

$$\ln(1 + \epsilon_j) \approx \epsilon_j$$

and therefore, equation (2.5) becomes

$$\ln X_n \approx X_0 + \sum_{j=0}^{n} \epsilon_j.$$

The additive central limit theorem implies that $\ln X_n$ is asymptotically normally distributed. Thus, X_n is asymptotically lognormally distributed [5].
From a biological point of view, the recursion (2.4) could represent a cascade of catalytic reactions, where the chemical $j - 1$ catalyzes the chemical j $(j \geq 1)$ [9]. The multiplication with $(1 + \epsilon_{j-1})$ $(j \geq 1)$ perturbs the deterministic value of X_j. The variations combine multiplicatively in this recursion. The fact that protein concentrations exhibit a lognormal

distribution suggests that the process of gene expression is such a catalytic reaction [9]. Indeed, gene expression is influenced by a cascade of many factors such as transcription factors, RNA-polymerases, and ribosomes [9]. Recall that the perturbations ϵ_j ($j \geq 1$) from equation (2.5) are mutually independent and identically distributed random variables. Transferring this to the cascade of reactions involved in gene expression, we might conclude that variations in the concentration of transcription factors, RNA-polymerases and ribosomes are mutually independent and identically distributed.

It is worth mentioning that many phenomena in nature obey a lognormal distribution as presented in [19]. Examples are numerous, ranging from geological, environmental, and medical phenomena and even to occurrences in social sciences and economics. Researchers investigating the lognormal distribution were asked to find an original measurement that matches a normal but not a lognormal distribution. The term "original" was meant in the sense that the data is not a function of a measurement, such as the sum of original data. Interestingly, they could not find such an example.

Chapter 3

Properties of Extrinsic Noise

3.1 Quantification

The stochasticity in the expression level of a protein is typically measured with the coefficient of variation [44, 4, 38, 21, 49, 35, 47]. It is an empirical, dimensionless measure of the magnitude of variations [38, 21]. Let P be a protein. The coefficient of variation (CV) is defined as the standard deviation σ of the distribution of P divided by the mean concentration μ of P [21], that is

$$CV = \frac{\sigma}{\mu}.$$

To estimate the unknown CV of the distribution of some protein P in a cell, the concentration of the protein can be measured across a population of N genetically identical cells [44]. Let P_k be the concentration of the protein P found in the kth cell. Opposed to previous ways of approaching the CV [44], we shall use a different measure of dispersion based on the fact that lognormal protein distributions are expected.

Given the mean and the variance of the lognormal distribution in Section 2.3, one can compute its CV as

$$CV = \sqrt{e^{\sigma^2} - 1}, \tag{3.1}$$

which only depends on the variance of the associated normal distribution. Indeed, it was reported experimentally that the coefficient of variation of a protein concentration does not change with the mean [47]. However, the proposed measure of dispersion for the lognormal

distribution is not directly the CV itself [19]. To quantify the magnitude of fluctuations in a lognormal distribution, the multiplicative standard deviation s^* was suggested in [19], which is given by

$$s^* = e^{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[\ln(\frac{P_i}{m^*})]^2}}, \tag{3.2}$$

where m^* is the geometric mean, which reads as

$$m^* = \sqrt[N]{\prod_{i=1}^{N} P_i}.$$

In fact, the geometric mean m^* and the multiplicative standard deviation s^* are the maximum likelihood estimators for the lognormal distribution, because of its multiplicative emergence [19].

Note that the multiplicative standard deviation s^* to a certain extent is an approximation of the CV of the lognormal distribution. That is, by plugging in the geometric mean m^* into (3.1) and applying the basic rules of the natural logarithm, we get

$$s^* = e^{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}[\ln(P_i)-\frac{1}{N}\sum_{j=1}^{N}\ln(P_j)]^2}},$$

which can be further expanded into

$$s^* = e^{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\ln(P_i)^2-2\frac{1}{(N-1)N}(\sum_{i=1}^{N}\ln(P_i))(\sum_{j=1}^{N}\ln(P_j))+\frac{1}{(N-1)N^2}N(\sum_{j=1}^{N}\ln(P_j))^2}},$$

and this is just

$$s^* = e^{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\ln(P_i)^2+\frac{1}{(N-1)N}(\sum_{i=1}^{N}\ln(P_i))^2}}. \tag{3.3}$$

For high values of N we obtain from Equation (3.3) the approximation

$$s^* \approx e^{\hat{\sigma}},$$

where $\hat{\sigma}$ is the maximum likelihood estimator for the standard deviation of the normal distribution given by

$$\hat{\sigma} = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\ln(P_k)^2 - (\frac{1}{N}\sum_{k=1}^{N}\ln(P_k))^2}.$$

Thus, in a sense, s^* approximates the CV of the lognormal distribution given in equation (3.2). s^* includes division of the measured values P_k $(k = 1, ..., N)$ by the mean m^*. Hence, the fluctuations are normalized to the mean, which enables appropriate comparison of quantified fluctuations with different means.

Now, there is the need to find an appropriate measure of dispersion for the discrete modeling framework. For this reason, the multiplicative standard deviation proposed in [19] has to be modified.
First, it is important to notice that the deterministic updates $f_i(x(t))$ $(i = 1, ..., n)$ in the simulations are known. And they represent the mean trajectories of the network variable $x_i(t)$. Thus, we might use this as mean value instead of the geometric mean m^*.
Second, we have to take into account that the multiplicative standard deviation includes division by the mean value and also the logarithm of the protein levels. However, these values can be zero. Hence, we propose to shift all the observed values as well as the mean values by plus one. This modification does not distort the quantified fluctuations since the multiplicative standard deviation s^* normalizes the values to the mean and thus the measured magnitude of fluctuations is independent of the mean.
Third, the variables of the network X_i $(i = 1, 2, ..., n)$ may have different ranges of expression levels. That is, some proteins might have levels in $\{0, 1, 2, 3, 4\}$, while others can only take on values in $\{0, 1\}$ or $\{0, 1, 2\}$. If $\{0, 1\}$ are the states of the variable and there is a deviation from 0 to 1, we have a much larger fluctuation than if a variable that has 5 states $\{0, 1, 2, 3, 4\}$ fluctuates by one. Thus, the measurement for dispersion of network variable x_i $\in \{0, 1, ..., p_i - 1\}$ is multiplied by $\frac{1}{p_i}$ $(i = 1, ..., n)$, such that we have the variations relative to the whole range of possible values. This confers comparability even between network nodes with different ranges.
We want the measure of dispersion to be 0 for the deterministic run of the system. Thus we subtract $\frac{1}{p_i}$ at the end. This does not change the measured fluctuations, because we subtract it for all nodes.

After considering all the necessary modifications, we finally can define the measure of dispersion for the discrete modeling framework, which we will call adapted multiplicative standard deviation (AMSD).
Let the number of iterations of the system be m time steps. Let

$$(f_i(x(0)), f_i(x(1)), f_i(x(2))..., f_i(x(m-1)))$$

be the deterministic trajectories of protein x_i $(i \in \{1, ..., n\})$. Instead of measuring the protein concentration in N $(N \geq 2)$ cells, the simulations are run N times in order to have N samples of the protein concentration level at a given time point $t \in \{1, 2, ..., m\}$. Thus,

there will be N trajectories

$$(x_i^k(1), x_i^k(2), x_i^k(3), ..., x_i^k(m))$$

with $k = 1, 2, ..., N$, which were subject to extrinsic noise. Let p_i be the number of possible states for variable x_i. The resulting AMSD for time step $t \in \{1, ..., m\}$, denoted by $\bar{s}(t)$ is given by

$$\bar{s}(t) = \frac{1}{p_i} e^{\sqrt{\frac{1}{N-1}\sum_{k=1}^{N}[\ln(\frac{x_i^k(t)+1}{f_i(x(t-1))+1})]^2}}.$$

Thus, the AMSD for the whole run of the simulation, i.e. for $t = 1, 2, ..., m$, denoted by $\bar{s}$ is given by

$$\boxed{\bar{s} = \frac{1}{p_i}\sum_{t=1}^{m}\frac{1}{m} e^{\sqrt{\frac{1}{N-1}\sum_{k=1}^{N}[\ln(\frac{x_i^k(t)+1}{f_i(x(t-1))+1})]^2}} - \frac{1}{p_i}.} \tag{3.4}$$

Now, variations in the amount of a variable with a smaller range of possible values is larger, compared to the same variations in the amount of a variable with more possible values. Hence, the AMSD enables comparison of the magnitude of fluctuations between variables with different ranges.
Note that this measure of dispersion quantifies the general fluctuations in a variable x_i, which could be of both intrinsic and extrinsic nature. However, in the applications presented here, we only perturb the variables with extrinsic noise. Therefore, we can be sure that the amount of dispersion quantified with the AMSD is caused by extrinsic noise only. It is important to emphasize that N has to be greater than 1 for the formula to be defined.
The purpose of the AMSD is to measure the magnitude of fluctuations of a single protein. Section 5.2 will apply this to examine the effect of certain network motifs on a specific protein. A way of visualizing the stochasticity in the dynamics of a whole network is described in Section 5.1.

3.2 Distinction Between Intrinsic and Extrinsic Noise

Researchers came up with a method to experimentally distinguish between intrinsic and extrinsic noise sources [8]. That is, given the total noise level of a protein, the so-called dual-reporter method enables splitting up the total noise into intrinsic and extrinsic parts.

The clue is to introduce two fluorescent proteins of different colors into the genome of a cell, such that they are controlled by independent, identical promoters. Global extrinsic sources of noise lead to correlated fluctuations in the expression level of the two genes since they are in the same intracellular environment. By contrast, the noise of intrinsic origin results in uncorrelated fluctuations. To ensure that the measured extrinsic noise captures not only the global noise, but also the pathway-specific fluctuations, the two reporter genes can be placed under the control of the same signaling pathway [21]. It was demonstrated that intrinsic noise is particularly high when the abundance of transcripts is low, whereas extrinsic noise has its peak at medium rates of transcription [8]. The dual-reporter method was widely applied [31, 34, 3, 44, 31, 36, 38].

Yet there was recently an argument that the dual-reporter method might lead to an underestimation of the extrinsic noise [41]. Investigators reported that it is not possible to discriminate between intrinsic and extrinsic noise using the dual-reporter method if the two identical promoters share common regulatory proteins [41]. In this case, the two copies of the gene compete for the regulatory proteins what can contribute to converse fluctuations in the expression levels of the two genes, especially when the abundance of regulatory molecules is low [41]. Indeed, transcription factor molecules in bacteria generally occur in low numbers [33], for example in E. coli, the number of LacI repressors is around 10 copies per cell [41]. An extreme case was simulated, where the repressor occurs only fourfold and variations in the number of RNA polymerases or ribosomes are neglected [41]. The computed part of the extrinsic noise was negative [41]. The binding of a transcription factor to a target gene decreases the amount of transcription factors that can bind to the other target gene [41]. In this regard, the extrinsic noise is not only an input to the genes, but the latter ones in turn also affect the extrinsic variations [41]. A negative part of the extrinsic noise is problematic, because in the presence of additional positive correlated fluctuations due to, for example, varying numbers of RNA polymerases and ribosomes, the fluctuations of the two reporters might be in the end mainly uncorrelated. The extrinsic noise generated by the common regulatory proteins, which results in converse fluctuations in the expression level of the two identical gene copies counteracts the expected positive correlated fluctuations [41]. The effect is a potential underestimation of the extrinsic noise part [41].

Intrinsic noise is expected to decrease as the number of transcript molecules increases [4]. This enables an approach to determine the relative contribution of intrinsic noise to the total noise [4]. The variations in the expression of a gene can be measured, as well as the noise in the expression level of the same gene when inserted multiple times into the genome of the cell [4]. Then the intrinsic noise can be quantified by the ratio of the noise level in the latter measurement to the noise level of the former measurement [4]. However, the multiple copies of the same gene could also compete for the same regulatory proteins. Hence with a rise

in the number of copies of the gene, the competition for the transcription factors increases, too. This effect would enhance the overall uncorrelated extrinsic fluctuations and distort the expected drop in the noise level.

However, one could think of alternative ways of distinction between intrinsic and extrinsic sources of noise. It should be at least possible to detect the relative contribution of extrinsic global noise to the total noise level [21]. When using two different independent promoters that drive the expression of two fluorescent proteins of different colors, the correlated fluctuations in the expression level is due to global extrinsic noise [21]. Since they are two different promoters, they do not share common transcription factors. The remaining part of the noise, namely the extent to which the two reporters fluctuate uncorrelated consists of intrinsic noise and pathway-specific extrinsic noise.

Section 3.3 illustrates the dual-reporter method as well as the problem with positive and negative correlation.

3.3 Effects of Extrinsic Noise - Single Protein

Shahrezaei and colleagues explored the effects of extrinsic noise [38] [39]. First, they considered a model of gene expression comprising the promoter in the non-active and active state, the amount of mRNAs and the amount of proteins. The model incorporates promoter activation and deactivation, transcription, translation as well as degradation. Figure 3.1 shows the scheme of this basic reaction network. In this section, we apply the discrete modeling approach for extrinsic noise, which was introduced in Section 2.2, on this small reaction network. Let $D \in \{0, 1\}$ be the current state of the promoter, where 0 represents an inactive promoter and 1 indicates an active promoter. Let $M \in \{0, 1, ..., p_M - 1\}$ be the representative concentration of mRNA and $A \in \{0, 1, ..., p_A - 1\}$ be the concentration level of the protein. In this example we choose a high number of states to illustrate the effect of extrinsic noise. In this simplified model of gene expression, the promoter changes its state in each time step. The mRNA level is increased by one if the promoter was active in the last time step. The level is decreased by one if the promoter was inactive in the last time step. The protein molecule number is the mRNA level from the previous time step.

Shahrezaei et al. simulated the dual-reporter method, which was described in Section 3.2, and showed that the protein expression levels of the two reporters fluctuate independently in the presence of only intrinsic noise and exhibit a quite symmetric distribution. When adding extrinsic noise, the variations in the number of protein molecules become larger and most importantly, they are mainly correlated. The resulting protein distribution exhibits a longer right tail ([39, p.370]).

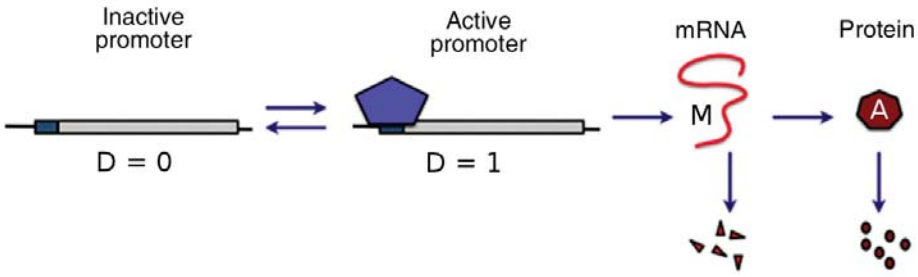

Figure 3.1: The small example includes promoter (inactive/active), mRNA, and protein. The Figure is adapted from [38]. Used under fair use guidelines, 2012.

Here, we show this effect using our discrete modeling approach introduced in Section 2.2. At first, only intrinsic noise is incorporated into the model, using the stochastic framework introduced by Murrugarra et al. [26]. Based on the idea that intrinsic noise affects the timing of biochemical reactions, this modeling framework uses propensities for activation and degradation for each variable [26]. Thereby, the propensity for activation/degradation represents the probability that the update of the respective variable actually happens, if the update increases/decreases the current expression level of the variable [26]. The propensities used in this example are displayed in Table 3.1. Two different proteins A and B are used, representing the two reporters. They were chosen to be relatively high to account for the fact that intrinsic noise is of much smaller magnitude than extrinsic noise [39].

	D = promoter	M = mRNA	A = protein
Propensity for activation	0.8	0.9	0.9
Propensity for degradation	0.8	0.9	0.9

Table 3.1: The propensities for activation and degradation to represent intrinsic stochasticity.

In the presence of only intrinsic noise (Figure 3.2 lower panel), as expected, the two proteins fluctuate independently and their distribution is quite symmetric.

Next, extrinsic noise was added to the system (Figure 3.2 upper panel). That is, the intrinsic noise framework is kept and additionally, the delay probability associated with the extrinsic noise caused by variations in the amounts of potential transcription factors, enzymes and variations in the cell's size, shape, organelle constellation and the chromosomal region of the gene, as described in Section 2.2, is introduced. The protein concentration is furthermore perturbed with the lognormal extrinsic variable E_1. The variance of the lognormal distribution was 0.8 and the probability for a delay was 5% for all three variables. The resulting

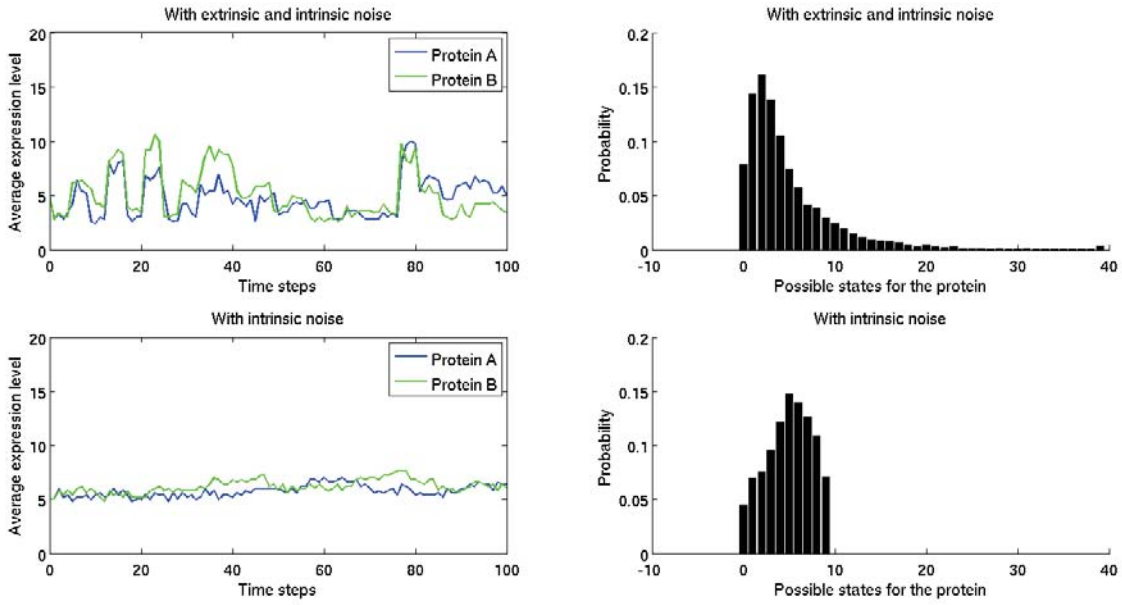

Figure 3.2: The upper panel shows the trajectory in the presence of both, intrinsic and extrinsic noise, whereas the lower panel the one without extrinsic noise. The average expression levels are the average over five simulations. The corresponding probabilities were generated using 100 simulations. As initial state we chose a random promoter state and $M = A = 5$. The number of expression levels of mRNA and proteins were $p_M = 10$ and $p_A = 40$ to compare the effect to the one generated by the continuous model.

average expression levels of protein A and B are now mainly correlated and the corresponding protein distribution exhibits a longer right tail (Figure 3.2 upper panel). The variations exhibit significantly larger magnitudes. This is in good agreement with Figure 1 in [39, p. 370].

A typical way to visualize the extent of correlated fluctuations between two reporters is to plot protein concentration A against concentration of protein B [32]. If fluctuations are highly correlated they tighten around the line A=B (as seen in [38, 21, 32]), whereas uncorrelated fluctuations result in a cloud of random points [32]. We reproduced this phenomenon with our discrete model (Figure 3.3). The intrinsic noise results in uncorrelated fluctuations, whereas the extrinsic variations mainly affect two identical independent proteins equally and therefore resulting in correlated fluctuations.

In Section 3.2, it was reasoned why the dual-reporter method may lead to an underestimation of the extrinsic noise. Recall that this can happen if the two reporters share regulatory proteins. To illustrate this principle, we simulated two extreme cases. At first, the hypothet-

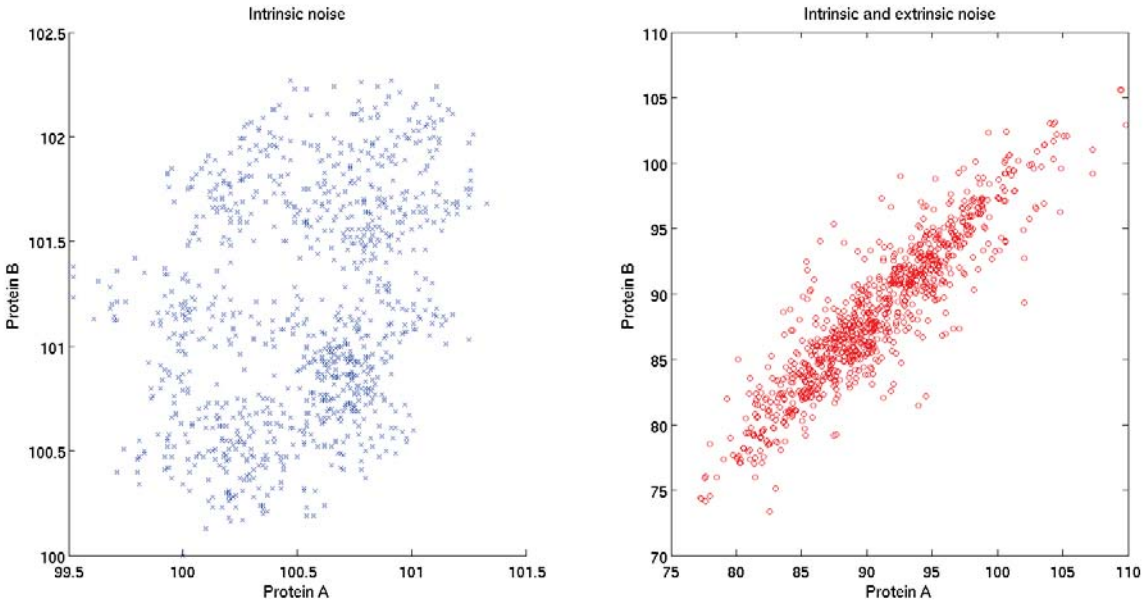

Figure 3.3: The idea of visualizing the dual reporter method. The extrinsic noise causes greater variabilities in the protein expression levels, as can be seen by comparing the range of values of the left and right panel. For this simulation, the number of different expression levels of mRNA and protein was $p_M = p_A = 200$. It was necessary to choose that many states to display the effect of the correlations.

ical case, where the two promoters of the reporter proteins are subject to identical delays and intrinsic noise is the only part of the noise that affects the promoters different. Such a case can occur, when the two promoters have the same activator and it is assumed that the activator is abundant. For this simulation, the intrinsic noise was kept at a low level and the probabilities for a delay for all variables were only 1%. The variance of the extrinsic lognormal random variable was $v = 0.2$. The resulting fluctuations are, as expected, mainly positive correlated (top left in Figure 3.4) what can be verified by plotting protein concentration A against protein concentration B (top right in Figure 3.4). However, as described in Section 3.2, it is often the case that regulatory proteins occur in low abundances. The second extreme case represents the scenario, where the two promoters of the reporter proteins are again under the control of the same activator, but only one copy of the activator is present. This setting implies that the regulator can only activate one of the two promoters at a time step. Both reporters are still subject to the same extrinsic variable E_1, but their promoters have converse delay occurrences. That is, if there is no delay in the update of the first promoter, then there is a delay in the update of the second promoter and vice versa. For this simulation, all parameters remained the same as before except for the probability

for a delay in the promoters, which was set to 50%. The average expression levels fluctuate mainly uncorrelated, due to the negative (delay) and positive (E_1) correlations that concur (bottom left in Figure 3.4). The resulting plot of protein concentration A versus protein concentration B reflects this effect (bottom right in Figure 3.4).

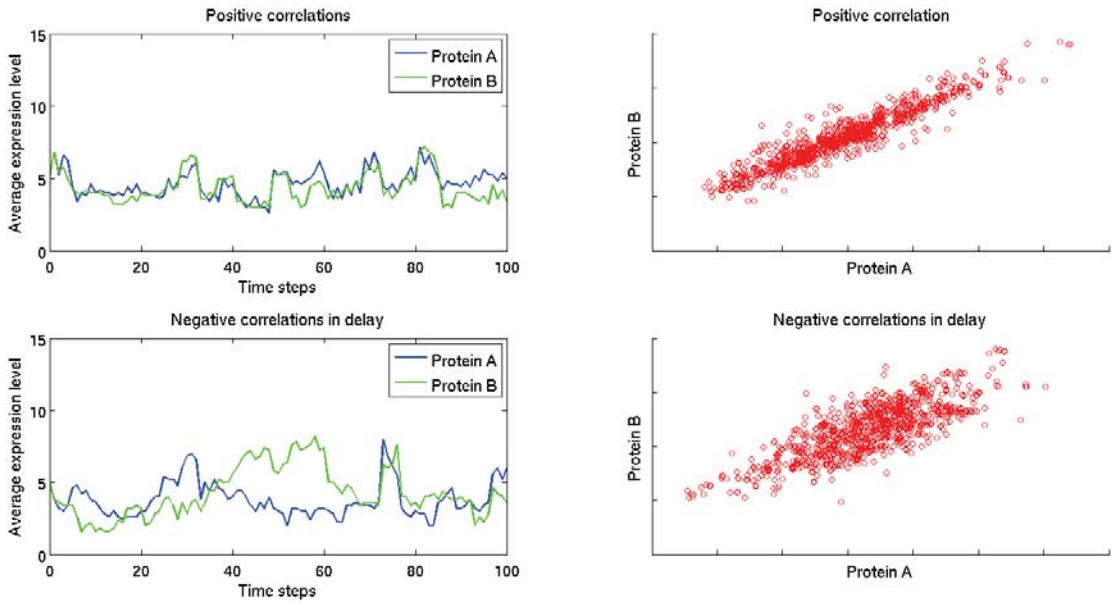

Figure 3.4: The upper panel depicts the results from the first extreme case, where the extrinsic perturbations are mainly positive correlated. The lower panel depicts the simulation of the second extreme case, where the two proteins A and B are activated by the same transcription factor, which occurs only once.

This simulation demonstrates the case, where the delay in the update of the two reporters is negatively and positively correlated at the same time. Thus, the resulting correlations appear to be uncorrelated to a great extent. The dual-reporter method would identify uncorrelated fluctuations as intrinsic noise. Indeed the resulting cloud of points in Figure 3.4 bottom right is less tightened around the line A=B, when compared to the plot in Figure 3.4 top right.

Variations of extrinsic nature are significantly larger in magnitude than intrinsic variations, which was exhibited in Figure 3.2. This is consistent with the different effects that intrinsic and extrinsic noise exert on the state space and thus the dynamics of a whole system, which is the issue of Section 5.1.

3.4 Advantages and Disadvantages

Noise might appear to be a nuisance that hampers proper functioning of a cell, however, it was reported that noise can actually be of advantage [7]. For example a population of identical cells can randomly differentiate due to noise, which permits division of labor and bet-hedging [7]. A metaphor for the latter one would be the common idiom "Don't put all your eggs in one basket," which implies for evolutionary biology that variability is of advantage in the light of an unpredictable future [42]. Noise can enhance the number of different phenotypes that emerge from a certain genotype [7].
Examples for bet-hedging strategies are numerous. First, a bacterial population of genetically identical cells can switch stochastically between two metabolic states, what was shown to be a fitness advantage in fluctuating environments [7]. A popular case of stochastic state switching is the model of the lactose utilization of the bacterium E. coli addressed in Section 4.2. The stochastic switching between the two states of metabolizing either glucose or lactose ensures that the population is composed of two subpopulations, one that has certain proteins highly expressed and is able to metabolize lactose, and the other subpopulation, which does not express these proteins at high levels and thus relies on glucose being available in the environment. The population is then bimodal [17]. Second, the soil bacterium Bacillus subtilis can stochastically switch to a state, where it is able to uptake DNA from the environment, due to noise [20]. This potentially confers a fitness advantage since the cells are provided with new genetic material [20]. Third, it was discovered that E. coli cells can switch to a state, in which they are resistant to antibiotics [32]. This leads to a bimodal population with the property that one part of the population is resistant to antibiotics. Even if the growth rate of this subpopulation is much slower, it enables surviving of antibiotic treatment [32].
An example of labor division would be the case of stem cells, since they are able to differentiate into more specialized cells and the choice of the cell fate can be stochastic due to noise [7]. For example a mouse has over a thousand different odorant receptors and instead of developing a complex regulatory network to ensure development of all different receptors, the system relies on stochastic decisions due to noise in gene expression [32].

Section 3.3 revealed that extrinsic noise leads to protein distributions with a longer right tail and thus increases the variability, when comparing it to a hypothetical system with only intrinsic noise. In this sense, extrinsic noise fosters the bet-hedging strategy. A lognormal protein distribution provides a broader range of possible thresholds for selection that lead to a bimodal population compared to a normal distribution [17]. It was concluded that the cell might have established long-tailed protein distributions, when variability is advantageous and symmetric protein distributions with a smaller variance, when robustness is more beneficial [17].

Noise also comes along with disadvantages [32]. For example it was revealed that noise and aging are related [32]. In particular, researchers discovered that aging in mice is associated with enhanced noise levels [32]. When precise functioning is required, noise can be detrimental [32]. In this case, living system have evolved mechanisms to suppress the variations [32]. It was shown that genes with a higher rate of expression usually have less fluctuations [32]. But a high expression rate also comes along with a high metabolic cost [32]. Indeed, researchers revealed that many molecular components occur in such small abundances that it cannot just be assumed the noise would be averaged out [40]. Therefore, the key for reducing noise might be a different one. It was for example hypothesized that genetic networks, i.e. the specific interactions among the genes or proteins, have evolved to reduce noise [32]. Section 5.2 addresses this issue.

Chapter 4

Applications

4.1 Infection of Bacteria with Phage Lambda

The lambda phage is the best known temperate bacteriophage, which is a virus that can live in symbiosis with its bacterial host [46]. It infects the bacterium E. coli, which then becomes lytic or lysogenic [40]. The lytic pathway is characterized by the proliferation of the virus, which kills and lyses the cell [46]. In contrast, in the lysogenic pathway, the DNA of the phage is inserted into the bacterial genome, where it stays and thus remains for future generations [46]. In that case, the cell is called a lysogen [46] and the virus a prophage [40]. The decision for either cell fate is due to noise, because even isogenic cells in identical environments and all infected by a single virus choose different pathways [40].

The random cell fate of E. coli after being infected by phage lambda can be explained by two alternative theories as proposed in [40]. Figure 4.1 illustrates the two possible explanations. The first theory suggests that variations in the timing of biochemical reactions during the infection drive the random decision for either cell fate. The alternative theory asserts that physical variations between different cells prior to infection have a significant impact on the outcome of the lambda phage infection.

In particular, the probability for a cell to become a lysogen decreases as the cell size increases [40]. Additionally, the amount of phages infecting a cell influences the decision between lysis and lysogeny. The likelihood for lysogeny increases with the amount of phages attacking a cell. It was suggested that lysogeny takes place if certain pro-lysogeny factors exceed a threshold in their expression level, which can be due to multiple phages infecting the cell or due to a small cell size. cII is an example of such a pro-lysogeny gene. Investigators explored the effects of different cell sizes using a mean cell volume ranging from $0.82\mu m^3 - 1.33\mu m^3$

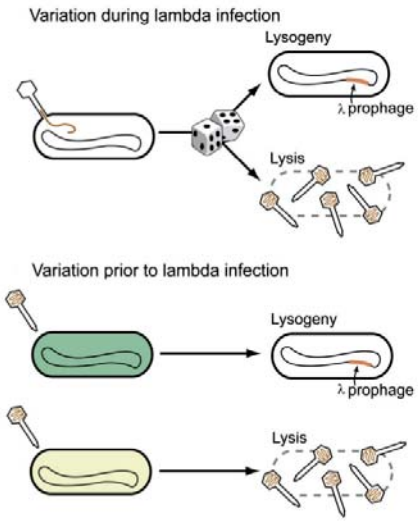

Figure 4.1: The decision between lysogeny and lysis can be attributed to either variations during the infection, which may be of intrinsic nature. But it might also be possible to ascribe the cell fate to physical differences between the cells that existed prior to infection. The figure is taken from [40]. Used under fair use guidelines, 2012.

and exhibited a 3-fold decrease in the probability of lysogeny with an increase in mean cell volume. That is, approximately 66.8% of the smaller cells became lysogens, whereas the same is true of only 22% of the larger cells.

However, researches underlined that there is no critical volume where all cells that are smaller become lysogens and all cells that are larger lyse [40]. Thus, the decision between either pathway depends not only on the cell size, but also on other factors which might be of both intrinsic and extrinsic nature. In the course of this section, we will show that the different contributors to the stochastic cell fate can be explained by either intrinsic or extrinsic noise, suggesting that both sources of variations together might drive the random decision.

Murrugarra et al. modeled the phage lambda infection with intrinsic noise and showed that it successfully captures the stochastic decision between lysis and lysogeny [26]. Hence, the work of Murrugarra et al. approached the first alternative depicted in Figure 4.1. This section can be seen as the supplementary part to their approach. The phage lambda model is incorporated into the framework to model extrinsic noise and it is shown that it can also explain the random decision. In particular the cell size is shown to contribute to the decision between lysis and lysogeny. Thus the second theory of Figure 4.1 is included.

Like the approach to explain the stochastic cell fate with intrinsic noise [26], the framework

for extrinsic noise is applied to the four variable model introduced by Thieffry and Thomas [46]. The four viral genes cI, cro, cII and N are crucial in determining which of the two pathways a cell will choose. The core regulatory network with the basic interactions between these genes is depicted in Figure 4.2.

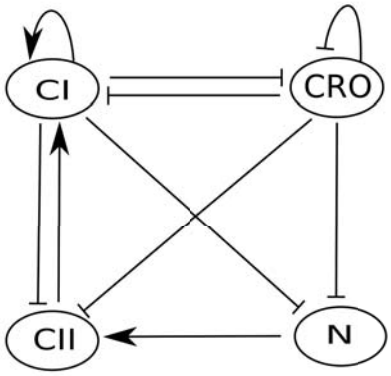

Figure 4.2: The basic interactions between the viral genes cI, cro, cII and N, which are the most important regulatory genes in the decision between lysis and lysogeny. The arrows represent activation whereas the blocked lines indicate repression. The figure is from [26]. Used under fair use guidelines, 2012.

The network includes a positive feedback loop between cI and cro, indicating that the system is bistable, with cI and cro as the two antagonists [46]. In fact, there is a bistable switch between lysis and lysogeny, where either cro or cI is highly expressed, respectively [26]. The most important lysogenic regulator, the viral gene cI, represses all other genes of the phage and thereby enables the virus to live in the bacterial cell [46]. Moreover, the lysogen is resistant against infection with further lambda phages [46]. In the lytic pathway, the gene cro represses cI and cII [26].

Adapted from Murrugarra et al [26], the discrete model for the interactions between cI, cro, cII and N is comprised of the following: The variables $x_1 = cI,\ x_2 = cro,\ x_3 = cII$ and $x_4 = N$ represent the four viral genes and can take on the five values $\{0, 1, 2, 3, 4\}$. Where the state 0 means no expression of the respective gene, 4 is full expression and the values 1,2,3 represent medium expression levels. The coordinate functions of this discrete dynamical system can be found in the Appendix.

As already mentioned, the dynamics of the system comprise a bistable switch between lysis and lysogeny [26]. The stable state $(x_1, x_2, x_3, x_4) = (2, 0, 0, 0)$ represents lysogeny, where cI is highly expressed and all other genes are off. In contrast, the limit cycle $(x_1, x_2, x_3, x_4) = (0, 2, 0, 0) \longleftrightarrow (0, 3, 0, 0)$ indicates lysis. All genes but cro, which is the most important lytic

gene [27], are off. Due to the negative autoregulation, the expression level of cro fluctuates between two states. The simulations are initiated with the state (0000), representing the moment when the cell is infected with the virus [26].

As described in Section 3.3, the framework to model intrinsic noise included propensities for activation and degradation for each protein [26]. The propensities leading to mostly lysogeny are shown in Table 4.1.

	cI	cro	cII	N
Activation	0.8	0.2	0.9	0.9
Degradation	0.2	0.8	0.9	0.9

Table 4.1: The propensities for activation and degradation for the four proteins cI, cro, cII and N, which result in lysogeny being the most likely cell fate.

The propensity for activation is increased for protein cI and decreased for cro. Alternatively, the propensity for degradation is small for cI and high for cro. This actually reflects the infection of multiple phages per cell: as described earlier, when multiple viruses infect a cell, certain pro-lysogeny factors are higher, such as cII. Increased levels of cII certainly come along with a higher propensity of cI being activated by cII. Higher levels of cI in turn result in a higher chance of cro being repressed and thus are represented by the high propensity for degradation. The opposite holds for the propensities resulting in mostly lysis, as shown in Table 4.2. The small propensity for activation of gene cI might reflect the infection of only a single phage. As mentioned earlier, infection by a single phage results in smaller levels of the pro-lysogeny factor cII, which in turn leads to a smaller propensity for cI being activated and thus the propensity for activation of cro is higher. With only rare exceptions, this results in lysis exclusively [26].

	cI	cro	cII	N
Activation	0.3	0.7	0.9	0.9
Degradation	0.7	0.3	0.9	0.9

Table 4.2: The propensities for acivation and degradation for the four proteins cI, cro, cII and N, which result in lysis being the most likely cell fate.

One may infer that the stochastic framework introduced by Murrugarra et al. also captures the effects related to the amount of infecting phages, in addition to the effects of stochasticity due to differences in the timing of biochemical reactions.

Now, we will show that the framework to model extrinsic noise completes the whole ensemble, as it captures the missing parts of the theories from the various contributors concerning stochastic cell fate.
First, it is important to note that the lytic pathway is generally more likely than the lysogenic one [46]. Moreover, it was reported that cells in the early exponential growth phase that are infected by one phage almost always lyse [40]. Contrarily, bacteria in the stationary growth phase will choose either pathway with equal probability [40]. This might be related to the fact that cells in the stationary phase are generally smaller [1]. Our modeling framework is mostly based on experiments, in which the cells were in the exponential growth phase [9] [17]. In fact, when simulating the lambda phage infection, almost all cells underwent lysis. The trajectories for this simulation are depicted in Figure 4.3. In the simulations, a cell is counted as lytic if the final state of the iterations is (cI, cro, cII, N)=(0,cro,0,0) and cro $\in \{2, 3, 4\}$. A lysogen was a cell that had (cI, cro, cII, N)=(cI,0,0,0) with cI $\in \{2, 3, 4\}$ as final state. Note that the percentages for lysis and lysogeny do not necessarily add up to 100%. This might be related to the large fluctuations caused by extrinsic noise. In the experiments there were indeed some cells that could neither be counted as lytic nor as lysogens [40]. It was observed that some cells would divide and one of their daughters becomes a lysogen, whereas another lyses. Such an occurrence was neither counted as lytic nor as lysogenic. Considering the fluctuations with increasing variances in Figure 4.3, it appears that the magnitude of fluctuations does not significantly increase. Also, the qualitative behavior of the system does not change, i.e. we do not get more lysogeny instead of lysis. This quite robust feature might be related to the network topology of the model. It is striking that the wiring diagram, which only consists of four nodes, contains five feed-forward loops. Section 5.2 is dedicated to a further investigation of feed-forward loops.

We will attempt to capture the trend that cells exhibit higher percentages for lysogeny when they are smaller. Since there were hardly any lysogens among the cells in the simulation so far, we assumed that the logical rules that govern the update of the system are designed to represent rather large cells. To consider smaller cell sizes, we introduced a parameter that determines the fraction of small cells in the population. To model a smaller cell, the protein expression levels were increased by one after each iteration. As described in Section 2.1, this is based on the idea that a smaller cell has the same amount of proteins, but they are more densely concentrated. Similarly, the proteins are more diluted in a larger cell. Besides this adjustment, also the probabilities for a delay were decreased to account for the lower chance of a delay in a smaller cell. In fact, when running the simulations with various fractions of smaller cells, the percentage of lysogeny increased, as shown in Figure 4.4. This is a simple way to model the effect of smaller cells, but it successfully enhances the percentage for lysogens, as predicted in [40]. From an evolutionary point of view, the dependence on the cell size could have evolved because larger cells can generate larger bursts of phages, which

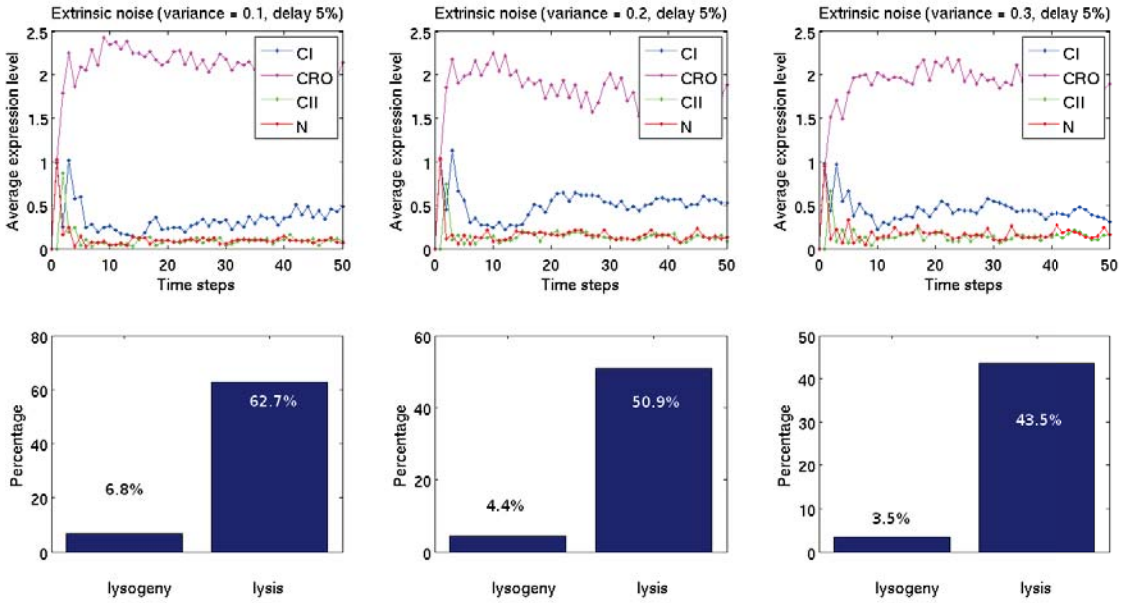

Figure 4.3: The trajectories for the simulations using different variances for the lognormal perturbation (upper panel) and the corresponding probabilities for lysogeny and lysis (lower panel). The probability for a delay was 5%. As seen, the extrinsic noise distorts the states of the variables, such that an increase in the noise level contributes to indeterminacy of lysis or lysogeny.

would be advantageous for the virus [40].

To sum up, we may conclude that the stochastic decision between either pathway is likely to be driven by the interplay of intrinsic and extrinsic sources of noise. Each contributor to the random cell fate mentioned in [40] could be captured either by the framework to model intrinsic noise proposed by Murrugarra et al. or by the framework to represent extrinsic fluctuations. In particular, the modeling framework for intrinsic noise could comprise the stochastic cell fate due to intrinsic variations during the infection and the number of phages infecting the cell [26]. Supplementary to that, the framework for extrinsic noise could capture the random decision between lysis and lysogeny due to extrinsic variations during the infection and the cell size.

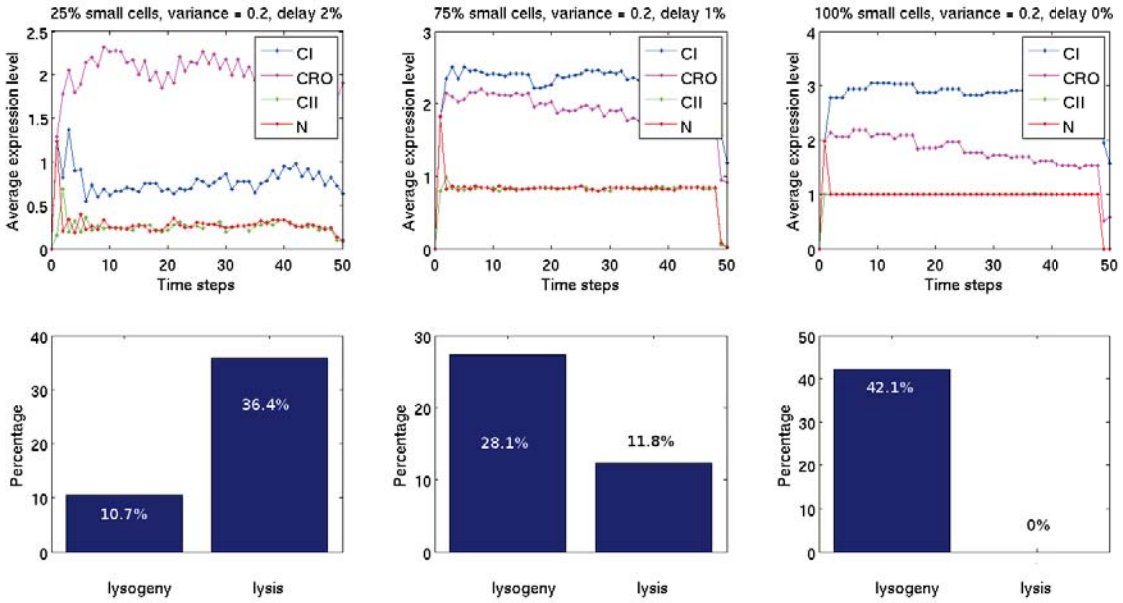

Figure 4.4: The trajectories for the simulations with variance 0.2 and different fractions of smaller cells (upper panel) together with the associated percentages for lysogeny and lysis (lower panel).

4.2 The Lac Operon Model

As mentioned in Section 3.4, the model of the lactose utilization of E. coli is an example of a bet-hedging strategy in a bacterial population. In this section we will at first explain the biological background and then apply the discrete modeling framework to this model. We will finally demonstrate the fitness advantage for a population that comes along with extrinsic noise.

The lac operon appears in E. coli bacteria and its dynamics were modeled extensively since it was first introduced in 1961 by Jacob and Monod [48]. After their discovery of the dynamics of the lac operon, Jacob and Monod won the Nobel Prize in Medicine in 1965 [18]. An operon consists of several genes side by side on an organism's DNA and transcribed together into a single mRNA [18]. The lac operon encodes for the proteins β-galactoside permease, β-galactosidase, and β-galactoside transacetylase [48]. The first one is a protein, which is bound to the membrane and transports lactose into the cell [48]. The second one decomposes lactose into glucose and galactose and also transforms lactose into allolactose [48]. Briefly, the lac operon encodes for all proteins that E. coli needs to metabolize lactose [18]. However, E. coli prefers to uptake the sugar glucose directly and as long as glucose is available in the environment, it is not necessary to have the ability to convert lactose into glucose; thus, the lac operon is repressed [18]. This mechanism minimizes the metabolic cost of the cell

and reflects the fundamental principle that energy, which is required for gene expression, is valuable for all living systems and therefore only those proteins currently necessary are transcribed [50] [32]. As long as glucose is available, a molecule called cAMP, which is required for transcription of the lac operon, is not synthesized and the repressor protein LacI inhibits transcription of the lac operon, which is called catabolite repression [18]. Glucose also prevents lactose from being transported into the cell, which is called inducer exclusion [48]. When glucose is not present any longer, but lactose is available, the cell needs to be able to metabolize lactose because it would starve otherwise [18]. Some of the lactose is transformed into its isomer allolactose, which in turn prevents LacI from repressing the lac genes [18]. Additionally, the catabolite activator protein (CAP) and cAMP build a complex and enable transcription of the lac operon [48].

The mathematical model was derived from a Boolean model of the lac operon [48]. The variables of the model are the mRNA of the lac operon ($M \in \{0,1\}$), the protein β-galactoside permease ($P \in \{0,1\}$), the repressor LacI ($R \in \{0,1,2\}$), the sugar lactose inside the cell ($L \in \{0,1,2\}$), the protein CAP ($C \in \{0,1\}$), the protein β-galactosidase ($B \in \{0,1\}$), and the isomer allolactose ($A \in \{0,1,2\}$). Additionally, the extracellular nutrients are glucose ($G_e \in \{0,1\}$), and lactose ($L_e \in \{0,1,2\}$). Originally, the variables were all Boolean and for those having three states, additional variables were introduced indicating whether the respective variable is in the low, medium or high state. For example allolactose A was represented by the tuple (A_m^*, A^*). Both representative variables were Boolean and $(A_m^*, A^*) = (0,0)$ denotes low concentration, i.e. here $A = 0$, $(A_m^*, A^*) = (1,0)$ is medium concentration, i.e. here $A = 1$, and $(A_m^*, A^*) = (1,1)$ denotes high concentration, i.e. here $A = 2$. In this way, the authors were able to keep the model Boolean and still distinguish between multiple concentration levels when necessary [48]. For our purposes, we instead chose a multivalued model. In Section 2.3 it was demonstrated that extrinsic noise results in fluctuations of large magnitudes and thus it is more appropriate to have multivalued variables, which enables representing larger variations. For this reason, the variables R, L, A and L_e were converted back to multivalued; the truth tables were converted as well. An example of how the truth table was transformed is the following. The original truth tables for the repressor are shown in Tables 4.3 and 4.4.

A_m^*	A^*	R^*
0	0	1
1	0	0
1	1	0

Table 4.3: The truth table for high concentrations of the repressor, R^*. A_m^* and A^* denote medium and high concentrations of allolactose, respectively.

A_m^*	A^*	R^*	R_m^*
0	0	0	1
1	0	0	0
1	1	0	0
0	0	1	1
1	0	1	1
1	1	1	1

Table 4.4: The truth table for medium concentrations of the repressor, R_m^*. A_m^* and A^* denote medium and high concentrations of allolactose, respectively.

At first, the two Boolean variables for medium and high concentrations of allolactose (A_m^*, A^*) were fused to one variable A. This process transformed the truth tables for the repressor to the Tables 4.5 and 4.6.

A	R^*
0	1
1	0
2	0

Table 4.5: The new truth table for high concentrations of the repressor, R^*. A denotes the new variable for allolactose.

A	R^*	R_m^*
0	0	1
0	1	1
1	0	0
1	1	1
2	0	0
2	1	1

Table 4.6: The new truth table for medium concentrations of the repressor, R_m^*. A denotes the new variable for allolactose.

Finally, the truth tables for R_m^* and R^* were merged to build the truth table for the new variable for the repressor R. This resulted in the truth table depicted in Table 4.7.

The truth tables for all other variables can be found in the Appendix. They were derived based on the biological interactions among the involved variables [48]. Note that in this application, not all network variables are proteins. That is, M represents mRNA and glucose

R	A	R
0	0	2
0	1	0
0	2	0
1	0	2
1	1	0
1	2	0
2	0	2
2	1	1
2	2	1

Table 4.7: The truth table for the new variable for the repressor, R. A denotes the new variable for allolactose.

as well as lactose are both carbohydrates, which represent the nutrients for the cell. These extracellular variables G_e and L_e do not have a truth table. They are parameters in the model, according to which the cell switches the lac operon on or off. The wiring diagram of the model is depicted in Figure 4.5.

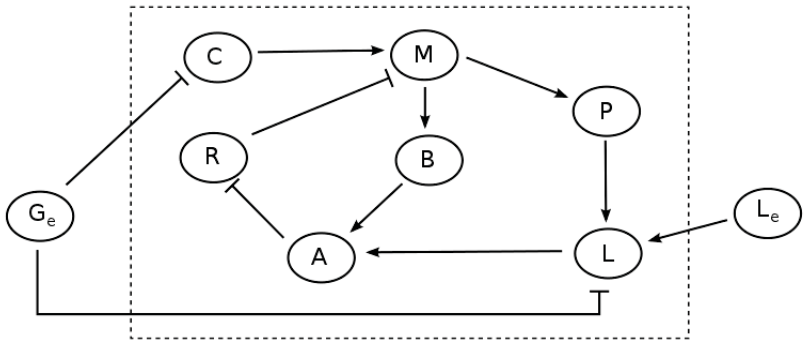

Figure 4.5: The wiring diagram of the lac operon reflects the basic interactions between the variables of the network. Two positive feedback loops involving the lac operon mRNA can be identified: The first one involving M, B, A, R, M and the second one with M, P, L, A, R, M. It is known that positive feedback loops can include multistability [21]. This scheme is drawn according to the wiring diagram in [48].

At first, deterministic simulations for all combinations of concentrations of the extracellular nutrients were run, to compare the dynamics with the Boolean model in [48]. The simulations were run 10^4 times with random initial settings of the variables. This represents a population of 10^4 E. coli cells. After each simulation, the state of the tuple (M, B, P, R) was used to determine whether the lac operon is switched on or off. $(M, B, P, R) = (0, 0, 0, 2)$

indicates that the repressor is at a high concentration, the mRNA of the lac operon is not transcribed and the proteins which it encodes for are absent. Hence the lac operon is off. The state $(M, B, P, R) = (1, 1, 1, 0)$ represents the lac operon being switched on, because the repressor is absent, but the mRNA as well as the proteins it encodes for are present. Thus we simulated the exposure of the bacterial population to various extracellular environments and checked after a sufficient amount of time (here 50 time steps) how many of the cells expressed the lac operon and how many did not.
When glucose is present, i.e. $G_e = 1$, the dynamics of the system lead to the lac operon being repressed for the whole bacterial population for all concentration levels of extracellular lactose $L_e \in \{0, 1, 2\}$. When glucose is absent, i.e. $G_e = 0$, the final state depends on the concentration of extracellular lactose. In the absence of both nutrients, the whole population does not express the lac operon genes. In this case, the cells would all starve after a certain amount of time and it requires additional energy to switch the lac operon on, which would be even more detrimental. When extracellular lactose is present in high concentrations, i.e. $L_e = 2$, the whole population expresses the lac genes. When external lactose is available in medium concentrations, that is $L_e = 1$, the bacterial population becomes bimodal, meaning that one subpopulation has the lac operon switched on and the other one does not. Thus the system is bistable, because there are two steady states in the state space.
The behavior of the system described here conforms exactly to that of the original Boolean model [48].

Next, extrinsic noise is introduced into the model as described in Section 2.2. In this case, the external nutrients are perturbed by the extrinsic random variable E_2, which obeys a normal distribution. The remaining variables are disrupted with the lognormal random variable E_1. Additionally, each variable is imposed with a probability for a delay in the update.
The external variables are set to a fixed mean and fluctuate around these values. The values $G_e = 1$ and $L_e = 2$ and the initial state $(M, B, R, A, L, P, C, Ge, Le) = (1, 1, 0, 2, 2, 1, 1, 1, 2)$ lead to the trajectories depicted in Figure 4.6. The values $G_e = 0$ and $L_e = 2$ together with the initial state $(M, B, R, A, L, P, C, Ge, Le) = (1, 1, 0, 1, 1, 1, 1, 0, 2)$ result in the trajectories shown in Figure 4.7.

It is remarkable that the system is much more susceptible to the noise when the lac operon is switched on. In both, Figure 4.6 and Figure 4.7, the variance of the distributions was 0.6, but the noise results in much larger variations in the latter case. The reason for this behavior can be found in the network topology. The repressor R and the lac operon mRNA M are both involved in two positive feedback loops as depicted in Figure 4.5. As will be discussed in Section 5.2, positive feedback loops are known to amplify fluctuations [21]. The crucial difference is that both positive feedback loops are only activated when glucose is absent. In the presence of glucose, intracellular lactose cannot enter the cell (inducer exclusion) and the lac operon mRNA cannot be transcribed because cAMP is inhibited [18]. This not only

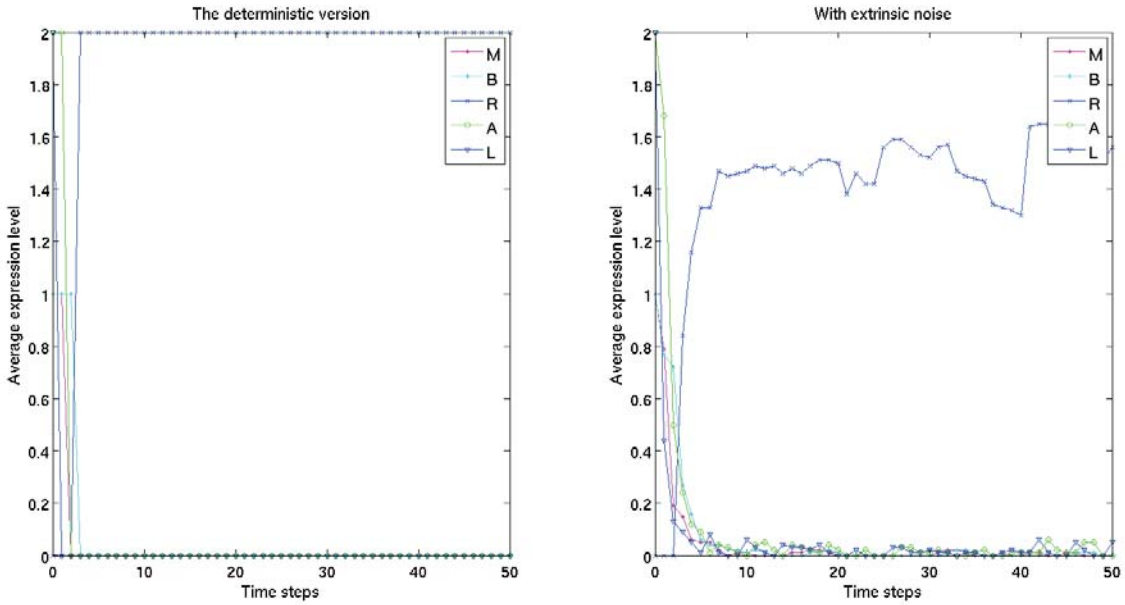

Figure 4.6: The trajectories for the deterministic system (left panel) and the system with extrinsic noise (right panel). The variance of the distributions was 0.6 and the probability for a delay was 20% for each variable. The average was taken over 100 simulations. Initially, the cell was in the state, where the lac operon was expressed. Due to the available glucose, the system changes to having the lac operon switched off. Even the noise cannot significantly change the qualitative behavior. We only illustrate the trajectories of M, B, R, A and L because knowing their expression levels suffices to determine whether the lac operon is on or off.

leads to repression of the lac operon mRNA M, but also to a suppression of the positive feedback loops. When glucose is absent in the environment, the positive feedback loop is active as soon as there is lactose available. This process amplifies the fluctuations and leads to the observed large variations.

In the bistable region, i.e. when glucose is absent and lactose is at a medium concentration, the population is bimodal. This can be interpreted as bet-hedging: the signals are not clear in favor of lactose utilization, since it is only present at medium concentrations. So there is a trade-off between risking the additional metabolic cost to express the lac genes in the hope of the lactose remaining available at least at a medium concentration and between saving this extra energy consumption and remaining in a state where the lac operon is switched off, to hope for glucose to become available again soon. This dilemma is solved by a bet-hedging strategy. In order to ensure overall survival of the population, one subpopulation is committed to switch the lac operon on, and the other subpopulation does not express the lac

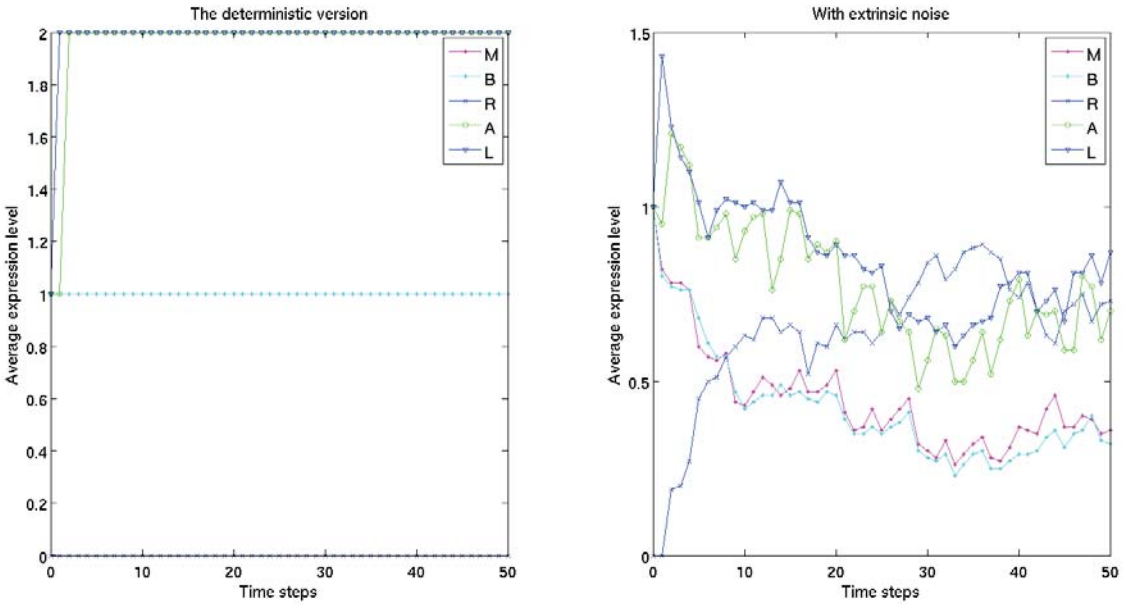

Figure 4.7: The trajectories for the deterministic system (left panel) and the system with extrinsic noise (right panel). The variance of the distributions was again 0.6 and the probability for a delay was 20% for each variable. The average was taken over 100 simulations. Initially, the cell was in the state where the lac operon was expressed. The deterministic run clearly leads to the lac operon being switched on. However, the extrinsic noise introduces large variations such that the population of the 10^4 E. coli cells contains a subpopulation with the lac genes on and a subpopulation with the lac genes off that are about the same size.

genes. Hence, the dilemma is solved by sacrificing one subpopulation to suboptimal growth or even starvation, just to ensure that the other subpopulation will survive. When running the deterministic simulation for the bistable region, it appears striking that only 92 of the 10^4 bacteria switched the lac operon on. Taking into account that only lactose is available, this implies that in a hypothetical deterministic system, not even 1% of the population would survive. Section 3.4 addressed the issue that noise promotes the bet-hedging strategy and actually forms a fitness advantage for the population in fluctuating environments. We will mimic this principle with the example of the lactose utilization system in the bistable region. Recall that for the external parameters $G_e = 0$ and $L_e = 1$, it depended on the initial state in the deterministic simulation whether the lac operon is switched on or off. Upon introducing the extrinsic noise, this decision is much more stochastic. Even for an initial state that leaded to either state of the lac operon in the deterministic simulation, the extrinsic noise can disrupt this path and change the qualitative behavior.

The population of 10^4 E. coli cells was simulated to be in an environment where glucose and lactose have the fixed means $G_e = 0$ and $L_e = 1$, respectively. A cell was counted as survived if it had the lac operon switched on in the presence of lactose or if glucose was available. The latter case is actually impossible because glucose cannot become available due to the noise since the extrinsic variable E_2 multiplied by zero does not change $G_e = 0$. However, the amount of external lactose fluctuated, as did that of the remaining variables. The noise was increased gradually and the survival rate was indeed enhanced as listed in Table 4.8. The cells that survived are those that are in the right state at the end of the 50 iterations, i.e. in this case, when the lac operon is switched on and lactose is available. The delay probability was neglected here (i.e. set to zero). It is remarkable that the survival rates actually can increase up to four-fold with the noise.

Variance	# of cells lac operon OFF	# of cells lac operon ON	# of survived cells
0	9908	92	92
0.1	7802	76	76
0.3	5444	195	183
0.5	4168	482	403
0.7	3625	525	398
0.9	3371	479	344
1.1	3220	503	352

Table 4.8: With different variances, i.e. different amounts of extrinsic noise, the chances of survival change.

The simulations were also performed with different delay probabilities. The resulting numbers of survivors in the population are illustrated in Figure 4.8. The simulations revealed that an increased delay probability contributes to the survival rate. For example, researchers stated that the number of LacI repressors in E. coli is only around 10 per cell, which potentially contributes to a higher probability for delays [41]. Except in one case (variance = 0.1, delay probability = 0%), the noise always increased the number of survived cells.

Thus, in the bistable region, the extrinsic noise generally enhances the chances for the lac operon to be switched on, which is a fitness advantage since lactose is the only available nutrient. Here, the external nutrients fluctuate in the same pace as the extrinsic noise. This was found to be the best condition for the survival chances [7]. We may conclude that noise promotes the bet-hedging strategy. Note that this model is quite coarse-grained as it only comprises 2 or 3 different levels of concentration of the proteins. It would be interesting to examine the effect of extrinsic noise on a model of the lactose utilization system, which distinguishes between even more states.

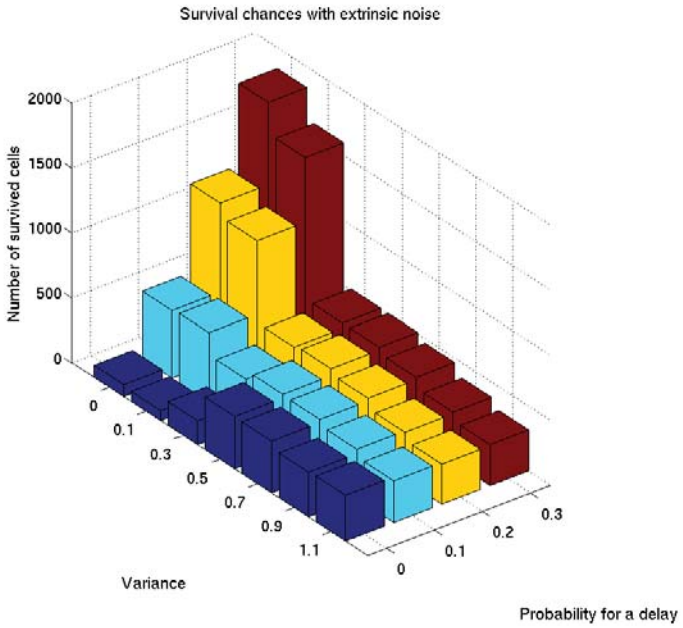

Figure 4.8: The number of survived cells in the population was monitored with various amounts of extrinsic noise. It is distinguished between various probabilities for a delay and different variances for the extrinsic variables E_1 and E_2.

4.3 Signaling Pathways in Melanoma Cells

Finally, in contrast to the previous applications in E. coli, the discrete modeling framework for extrinsic noise will now be applied to a gene regulatory network, which occurs in vertebrates and invertebrates [16]. Namely, this section works with the Wnt signaling network designed by Hinkelmann et al. [11]. The model and the biological background comes from the article [11]. Other references used here are denoted in the text.

Wnt proteins are known to activate intracellular signaling pathways and play a crucial role in regulating processes such as differentiation or proliferation [16]. If the signals are distorted, a tumour or metastasis can emerge [16]. Wnt proteins can activate different signaling pathways by interacting with various transmembrane proteins [16]. Examples for such pathways are the Wnt/β-catenin or canonical Wnt pathway and the planar cell polarity pathway [16]. Since these pathways share regulatory components, it was suggested that they can be activated at once in the same cell type, building together a Wnt signaling network [16].

85% of skin cancer deaths can be attributed to melanoma skin tumors, whose metastases

can often not be cured with chemotherapy. Researchers investigating potential therapeutic strategies found that the Wnt signaling pathway is abnormally activated in melanoma, which is a likely contributor to the proliferation. For early stage melanoma cells, it was reported that Retinoic Acid (RA), a vitamin A derivative, can lead to reduced cell growth by down-regulating the Wnt signaling. In particular, the canonical Wnt pathway was found to be active in melanoma cells, resulting in activation of the proteins Cyclin D and c-myc what in turn promotes cell proliferation. Treatment with RA inhibits the canonical Wnt pathway. However, more aggressive melanoma cell lines do not respond to treatment with RA. Hence, a potential therapy for these cells is to be found. Hinkelmann et al. built a discrete model of an early stage melanoma cell involving the canonical Wnt pathway and the non-canonical PCP Wnt pathway. The wiring diagram of their model is depicted in Figure 4.9. The model includes 25 variables representing the proteins involved in the pathways. Twelve of the variables are Boolean, eight of them have three different states and five of them have four states. The truth tables and their derivation from the biological processes can be found in the article "Downregulation of LRP6 inhibits growth of melanoma cells" [11].

The PCP Wnt pathway results in expression of Daam 1 and JNK, whereas the canonical Wnt pathway leads to activation of Cyclin D and c-myc. These proteins form the focus of the simulations, since they enable identification of the currently activated pathway. Cyclin D and c-myc are regulated identically, thus it suffices to monitor the states of Cyclin D, JNK, and Daam1. Cyclin D has the three states $\{0, 1, 2\}$ representing absent, present at medium concentrations and present at high concentrations. JNK has the states $\{0, 1\}$ indicating absence or presence. Daam1 can take on the values $\{0, 1, 2, 3\}$ meaning low inactive, low active, high inactive and high active, respectively. The simulated therapy results in the states shown in Table 4.9.

Protein	no RA added	RA added	RA added with SFRP1 siRNA
Cyclin D	present (2)	absent (0)	medium (1)
JNK	absent (0)	present (1)	present (1)
Daam1	low inactive (0)	high active (3)	high active (3)

Table 4.9: Treatment of the early stage melanoma cell lines with RA leads to the desired redirection of the canonical Wnt pathway to the PCP Wnt pathway. This can be monitored by focusing on the states of the three proteins Cyclin D, JNK, and Daam1.

The model was implemented in Matlab to apply the extrinsic noise framework to this example. The deterministic simulations were consistent with the expected behavior of the system, as shown in Figure 4.10. The extrinsic noise caused fluctuations in the expression levels. It stands out that the fluctuations of Cyclin D in the first panel are particularly high. This becomes pronounced when taking into account that Cyclin D has less states than Daam1. So we would expect Daam1 to exhibit the largest fluctuations. The answer for this phe-

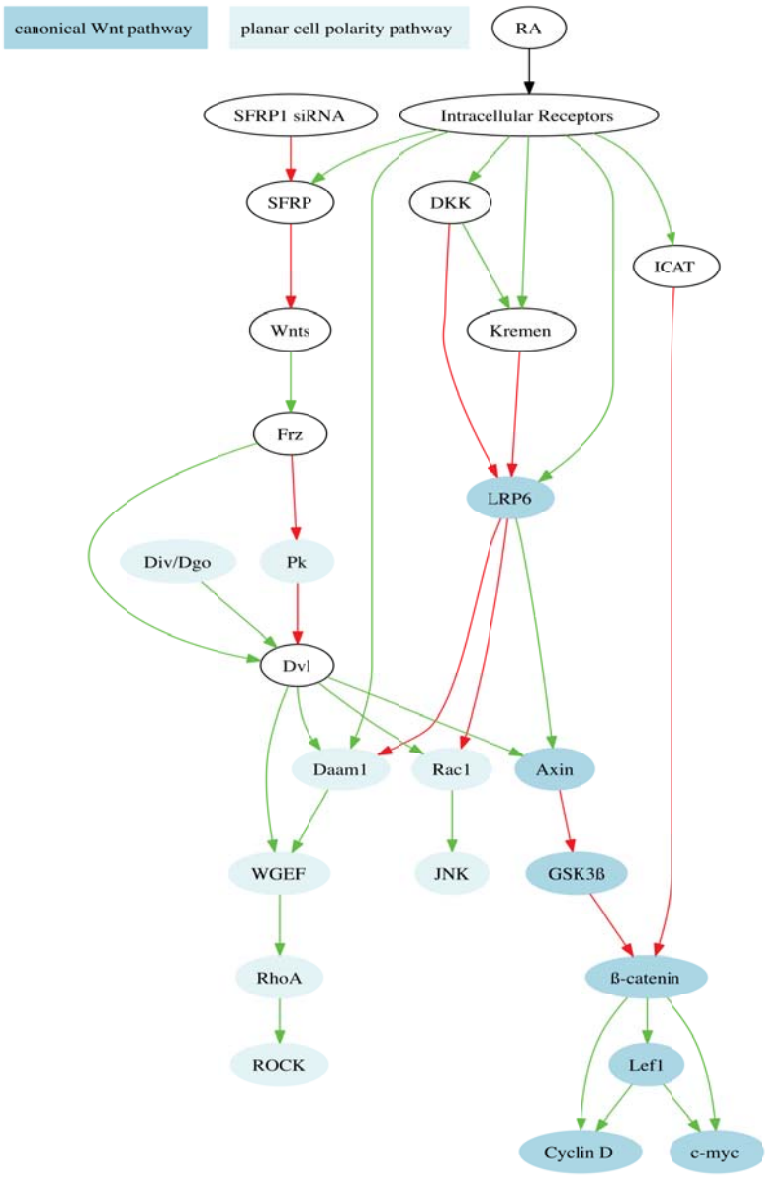

Figure 4.9: The wiring diagram of the Wnt signaling network. Red arrows indicate inhibition and green arrows represent activation. This scheme was designed by Hinkelmann et al. [11]. Used under fair use guidelines, 2012.

nomenon can be found in the network topology: Cyclin D is the target of a so-called positive

feed-forward loop. Section 5.2 will discuss feed-forward loops. However, the extrinsic noise does not change the qualitative behavior of the system.

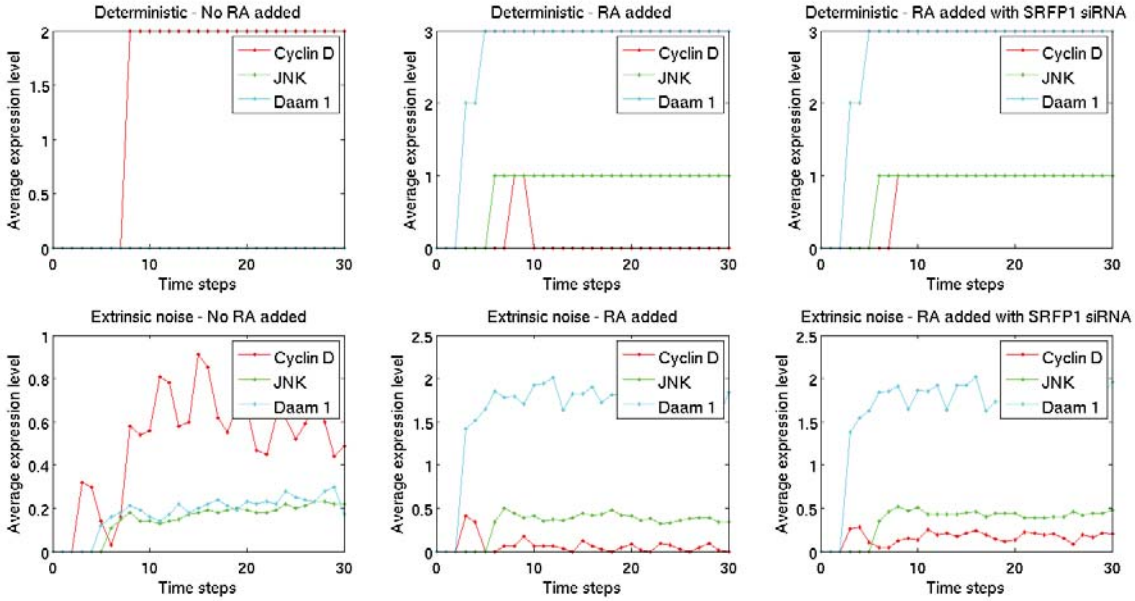

Figure 4.10: From left to right the trajectories of the simulation, where no RA is added, RA is added and RA is added together with SFRP1 siRNA. The upper panel depicts the average expression level of Cyclin D, JNK and Daam1 in the deterministic run. The lower panel shows the trajectories with extrinsic noise. Here, the variance of the lognormal random variable E_1 was 0.5 and the probability for a delay was 5%.

By knocking out the intracellular receptors and SFRP, the model can be modified to represent more aggressive cell lines, the so-called vertical growth phase melanoma cells. Hinkelmann et al. revealed that LRP6 might be the important regulator that can redirect the canonical Wnt pathway into the non-canonical PCP Wnt pathway. In particular, downregulation of LRP6 can reduce the tumor growth in the aggressive melanoma cells. This discovery was verified experimentally. The simulations of the aggressive cell lines showed that Cyclin D is present, whether RA or SFRP1 siRNA are added or not. Also, JNK is always absent and Daam1 low inactive. This reflects the fact that the vertical growth phase melanoma cells are resistant to RA. As soon as LRP6 is knocked out, the dynamics change, i.e. Cyclin D is absent, JNK present, and Daam1 low active. The dynamics are illustrated in Figure 4.11. It appears again that Cyclin D is exhibits large variations due to the positive feed-forward loop.

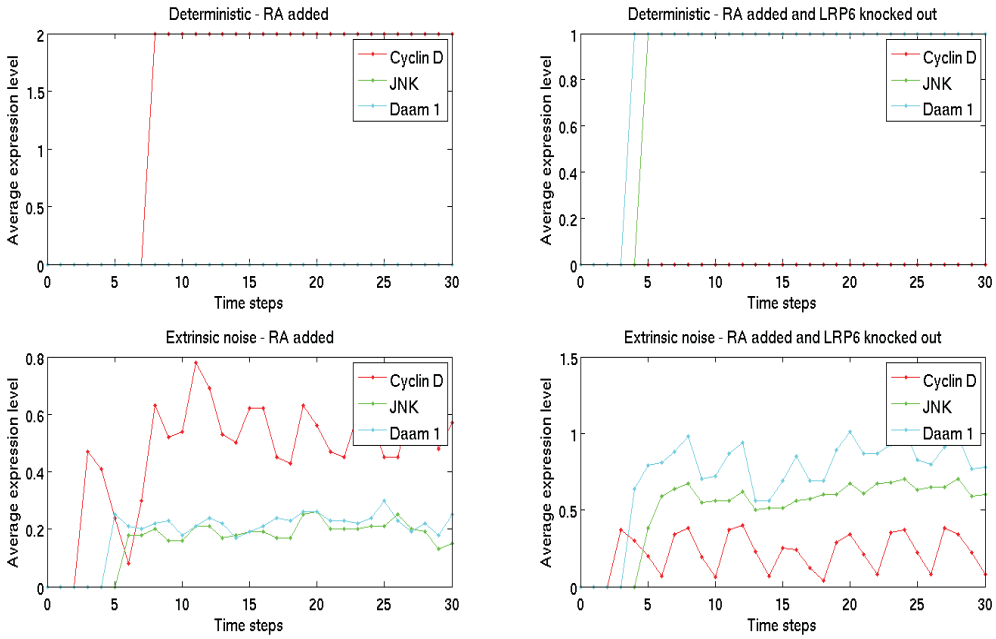

Figure 4.11: The trajectories of the simulation where RA is added are shown on the left. They are exactly similar to those when RA is not added or with SFRP1 siRNA. On the right, the average expression levels are depicted in the case, where LRP6 is downregulated. As can be seen, the canonical pathway is deactivated, since Cyclin D is absent. The upper panel depicts the average expression levels of the deterministic simulation. The lower panel shows the trajectories disrupted by extrinsic noise. Here, the variance of the lognormal random variable E_1 was again 0.5 and the probability for a delay was 5%. With the noise, Daam1 fluctuates quite close around its deterministic value, as can be seen in the right panel. In contrast, Cyclin D exhibits large deviations from the deterministic value, especially in the left panel. This phenomenon is related to the network topology.

Having conducted the simulations with extrinsic noise, one notices that the trajectories fluctuate, but the noise does not change the qualitative dynamics. This was additionally verified by assessing whether the noise could redirect one pathway to the other. But this was, with some exceptions, not possible. One may conclude that the Wnt signaling network is quite robust against noise. This might be attributed to the network topology. Therefore, Section 5.2 also examines a network motif of the Wnt signaling network.

Chapter 5

Effects of Extrinsic Noise

5.1 State Space

Section 3.3 demonstrated the effect of extrinsic perturbations on the expression level of one protein. After three applications of the discrete modeling framework for extrinsic noise in Chapter 4, this section aims to analyze the general effect of the extrinsic variations on gene regulatory networks.

The dynamics of a gene regulatory network correspond to the state transitions of the whole system, as described in [48]. This can be visualized with the "state space", which is a directed graph. The nodes are n-tuples, representing the states of the system. A directed edge $\langle x_1(t), x_2(t), ..., x_n(t)\rangle \longrightarrow \langle x_1(t+1), x_2(t+1), ..., x_n(t+1)\rangle$ indicates that the first state turns into the second state at the next time step. Directed cycles in the state space are denoted as limit cycles. Fixed points are those nodes in the state space, having a directed edge exiting and the same edge entering the node again, i.e. they are mapped to themselves. They are also referred to as "stable states" or "steady states".

The discrete framework to model intrinsic noise preserved the fixed points of the deterministic state space [26]. By contrast, extrinsic noise can eliminate the stable states of the deterministic state space. This can be seen when considering the state space of a small example.
Let x_1, $x_2 \in \{0, 1, 2, 3\}$, i.e. $p = 4$. Let the wiring diagram be the one depicted in Figure 5.1 and the truth table governing the updates of the system the one given in the Appendix. Moreover, the deterministic state space is depicted in Figure 5.2.

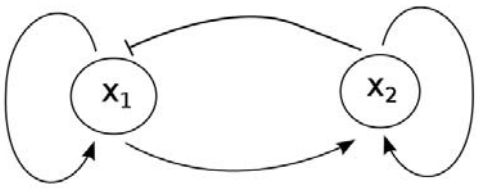

Figure 5.1: The wiring diagram of the small example.

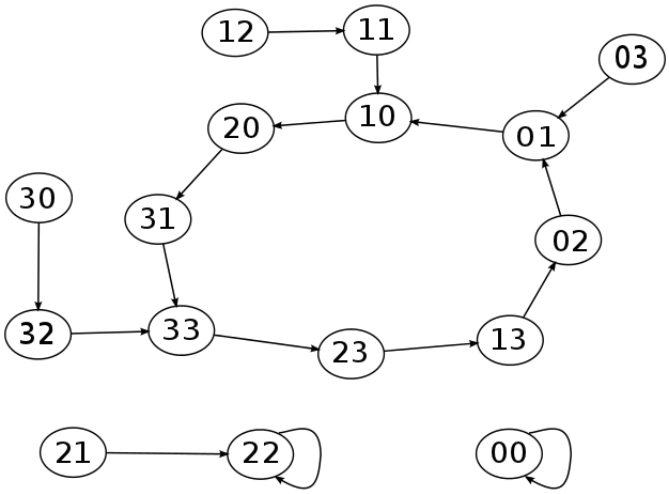

Figure 5.2: The deterministic state space of the small example. There are two fixed points in the state space, namely (00) and (22). Also one can identify a limit cycle. The state space consists of three components.

Extrinsic noise is now introduced. Let the variance of the extrinsic variable be $v = 0.3$ and let the probability for a delay be 5%. Let $\tau \geq 0$ and let $t \in \{\tau, \tau+1, \tau+2, \tau+3\}$ indicate the time step. As described in Section 2.2, $E_1(\tau)$ is the lognormal extrinsic variable for the time scale $\{\tau, \tau+1, \tau+2, \tau+3\}$ and $d_i \in [0, 1]$ is the probability for a delay in the update of node x_i ($i \in \{1, 2\}$). Here, $d_1 = d_2 = 0.05$. According to the truth table, the deterministic update $f_1(x_1(t), x_2(t))$ for $(x_1(t), x_2(t)) = (1, 0)$ is $x_1(t+1) = f_1(x_1(t), x_2(t)) = 2$, such that $x_1(t) = 1$ is mapped to $x_1(t+1) = 2$. Now, with extrinsic noise, this update is perturbed. That is, the update becomes

$$x_1(t+1) = \begin{cases} x_1(t), & \text{with probability } d_1 = 0.05, \\ x_1(t+1)^*, & \text{with probability } (1-d_1) = 0.95, \end{cases}$$

where

$$x_1(t+1)^* = \begin{cases} \min\{3, \lceil E_1(\tau) \cdot 2 \rceil\}, & \text{if } (E_1(\tau) \cdot 2 - \lfloor E_1(\tau) \cdot 2 \rfloor) \geq 0.5, \\ \min\{3, \lfloor E_1(\tau) \cdot 2 \rfloor\}, & \text{if } (E_1(\tau) \cdot 2 - \lfloor E_1(\tau) \cdot 2 \rfloor) < 0.5, \end{cases}$$

is the case, where the update is not delayed and the variable is perturbed with E_1 and finally rounded to keep the states discrete. Multiplying the deterministic update with the variable E_1 can lead to a variety of other updates, representing the extrinsic noise. In the example where the deterministic update for $x_1(t) = 1$ is $x_1(t+1) = 2$, the two is multiplied by the extrinsic variable E_1. Upon rounding, we get at on average the values depicted in the left panel of Figure 5.3. They build a "discretized" lognormal distribution. The percentages represented by the bars reflect the increased probability for other values to be chosen as the actual update. Especially the update to one is equally likely as the update to two. The potential updates have to be kept within the range $0, 1, \ldots, p-1$. The central panel in Figure 5.3 displays the probabilities after this adjustment. It has the side effect of increasing the probability for the highest state. Finally, the probability for a delay d_1 is imposed on the update. All probabilities were multiplied by $(1 - d_1)$ and d_1 was added to the probability for the current state. The pie chart in the right panel of Figure 5.3 illustrates the final probabilities for the different states to be the actual update. Note that this update scheme is not tied to a specific network. It shows in general, how the update $1 \longmapsto 2$ is perturbed by extrinsic noise, when the variance is $v = 0.3$, the probability for delay is 5% and the number of states is $p = 4$.

The striking characteristic is that extrinsic noise can render the deterministic update to be less or equally likely than other updates, especially those adjacent to the deterministic update. This is the case in Figure 5.3, even if a variance of 0.3 is quite small. Figure 5.4 illustrates the effect of the extrinsic noise on the updates of the level two, when the variance increases. With an increasing variance, the probability for an update other than the deterministic update increases. The update to zero forms the only exception, since multiplication of zero by the extrinsic variable E_1 always results in zero. The only case where an update to zero is perturbed by extrinsic noise occurs when there is a delay in the update and the current value is not zero. The first, second, third and fourth panel in Figure 5.4 demonstrate the noisy update, when the deterministic update is zero, one, two or three, respectively. The first row shows the resulting probabilities when the variance is

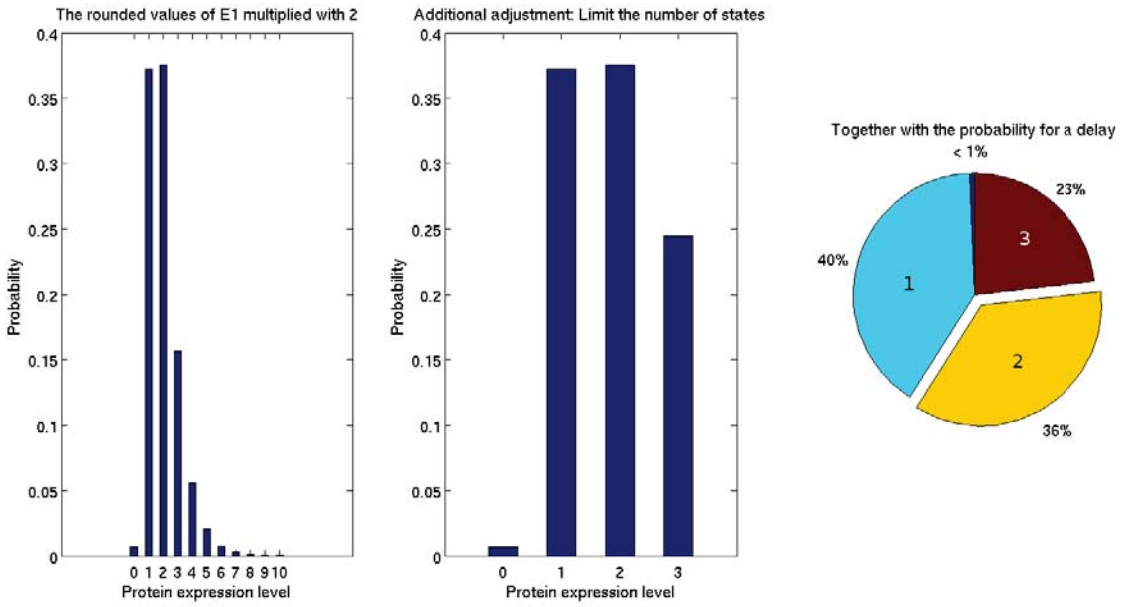

Figure 5.3: The process of how a deterministic update is distorted by the extrinsic noise.

0.3, the second row when the variance is 0.5, and the third row, when the variance is 0.7. The probability for a delay was 5% in all simulations. These fractions are again not tied to a specific network. It shows, in general, how the update can be perturbed by extrinsic noise, when the current value is 2, the probability for delay is 5% and the number of states is $p = 4$.

Having considered the update of one variable, we can now examine the effect on the whole state space. It is already clear, that the state (00), if a fixed point in the deterministic state space, will remain a fixed point in the state space with extrinsic noise. This can be realized by considering the update to zero in Figure 5.4 and replacing the two in the pie of the left panel by a zero. Furthermore, one can expect that the "extrinsic state space" will have many more edges, since each update can be disrupted with a significant probability. This has the effect that fixed points other than (00) will disappear. Potential limit cycles might also vanish. The extrinsic state space will consist of one component.

The extrinsic state space for the example was computed and can be found in Figure 5.5. For simplicity, only the arrows with percentages greater than 5% are shown. Especially the former fixed point (22) suffers an extreme perturbation by the extrinsic noise. Instead of being update to (22) with 100% probability, like in the deterministic case, it can be updated to many other states as listed in Table 5.1. The percentage for a certain transition in the

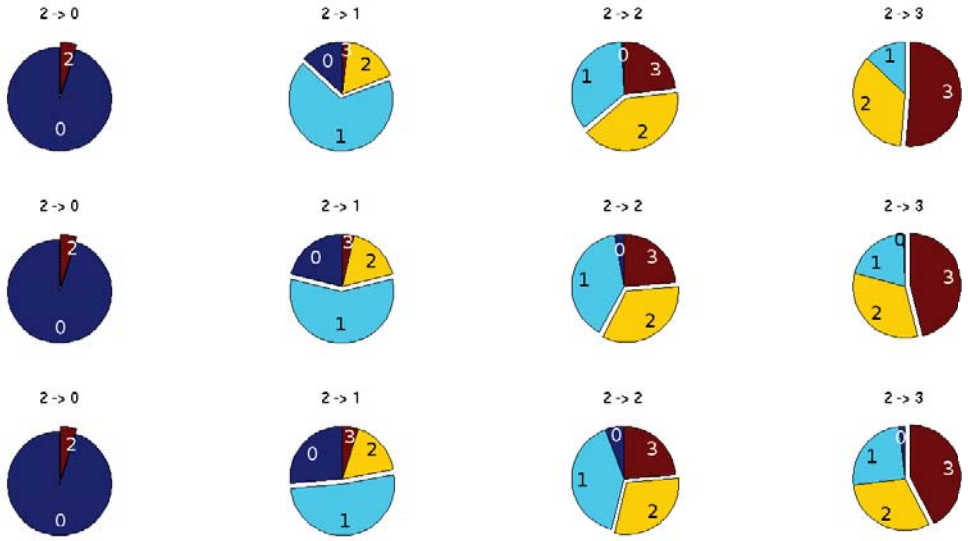

Figure 5.4: The pie charts indicate the final probabilities for the different states to be the actual update. In this example, the current value is two.

state space is the product of the percentages of the single variables in the respective update.

Update	(11)	(12)	(13)	(21)	(22)	(23)	(31)	(32)	(33)
Probability	12.6%	14.4%	8.2%	14.4%	16.5%	9.4%	8.2%	9.4%	5.4%

Table 5.1: The update of the state (22) was (22) with 100% probability, i.e. (22) was a fixed point. With extrinsic noise, this update is significantly disrupted. This table exhibits the different possible updates for (22) together with the likelihood in percent. Only the transitions with probability higher than 5% are listed.

After the demonstration of the effects of extrinsic perturbations with this small example, general conclusions can be drawn about any network. Consider a network with n variables, which can each have p different expression level. In the presence of extrinsic noise, a simple deterministic transition from one state to the other in the state space

$$\langle x_1(t), x_2(t), ..., x_n(t)\rangle \longrightarrow \langle x_1(t+1), x_2(t+1), ..., x_n(t+1)\rangle,$$

where $t \geq 0$ indicates the time step, becomes much more uncertain in the presence of extrinsic noise. In particular, instead of one deterministic arrow that leaves a certain state, there can be up to p^n exiting edges depending on the underlying deterministic update. This

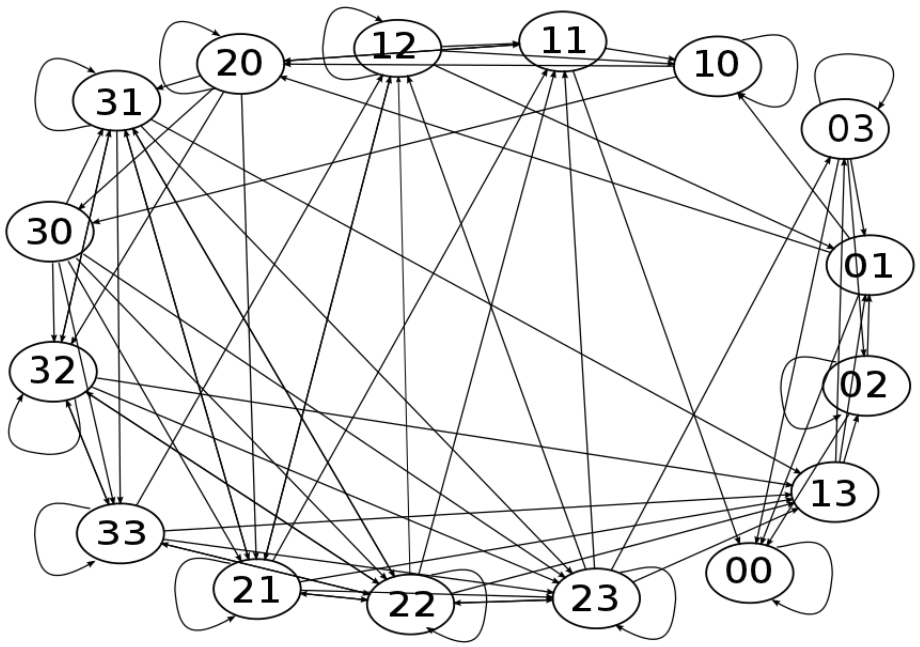

Figure 5.5: The extrinsic state space of the small example. The percentages for each arrow are not shown in the graphic. Only arrows with likelihood greater than 5% are shown. The limit cycle disappeared and the only conserved fixed point is (00). The state space consists of one component. Note that there are some transitions, having an arrow in both directions. Such an arrow indicates that the transition is possible in both directions.

has the effect that the number of edges in the state space will increase dramatically and the probabilities on the edges will decline, especially with large values for p and n. This was also observed in the extrinsic state space of the example shown in Figure 5.5. Only the arrows with probability higher than 5% were illustrated.

One may infer that the extrinsic noise significantly increases the connectivity of the state space. Consequently, the probabilities on the edges decrease and also the number of components declines. The state space will only consist of one component of size p^n in the presence of sufficient extrinsic noise, since there enough additional edges. Thus, the underlying undirected graph is connected and therefore the state space is weakly connected. However, the state space is not necessarily strongly connected: If $(0, 0, ..., 0)$ is a fixed point in the deter-

ministic state space, it will remain a fixed point in the extrinsic state space. Other fixed points will disappear.

This section also showed that the state space enables visualization of how stochastic a network is. The number of exiting arrows of a state and the associated probabilities indicate the extent of stochasticity in the respective update. With extrinsic noise, the edges become numerous and the probabilities decrease, with the feature that the highest probability is not necessarily the deterministic update.

One may draw the conclusion that extrinsic noise leads to larger fluctuations as well as more chaotic and unpredictable behavior. The fact that biological processes still function properly points towards some mechanism that must have evolved to prevent the system from behaving too chaotic. In Section 4.2, the example of the lac operon model already suggested that fluctuations might be linked to the network topology. This motivates Section 5.2, where specific network motifs are analyzed.

5.2 Network motif: Feed-forward Loop

As pointed out in Section 3.4, the clue of how biological systems deal with the noise could be the way the proteins interact with each other. Gene regulatory networks might have evolved in a way that fluctuations are attenuated when necessary.

The Wnt signaling network, examined in Section 4.3, as well as the lambda phage network, analyzed in Section 4.1, exhibited a quite robust behavior in response to the introduced stochasticity. This raises the question about which feature in the network confers this robustness. When considering the topology of the wiring diagram, several network motifs can be found.

A network motif is a specific pattern in the wiring diagram of gene regulatory networks, which occurs much more often compared to random networks with the same number of nodes and edges [25]. Such motifs might have evolved because they confer some advantageous features to the gene network. Interestingly, gene regulatory networks in E. coli and the yeast S. cerevisiae share the same motifs [25]. Therefore, it is of particular interest to examine the effects of noise on such motifs.

An example for that would be positive or negative feedback loops, which are directed cycles

in the wiring diagram [48]. Whether such a feedback loop is positive or negative is determined by multiplying the signs of the edges in the cycle [48]. Thereby, activating edges have a positive sign, whereas repressing edges have a negative sign [48]. The sign of the product then indicates the sign of the feedback loop [48]. It was shown theoretically and experimentally that negative feedback dampens fluctuations [21, 35]. This seems to be natural, since large variations are suppressed by the negative feedback [21]. In particular, the ability to reduce variations increases with the strength of the negative feedback [38]. However, negative feedback also reduces the number of transcripts and thus leads to increased levels of intrinsic noise, whereas extrinsic noise does not depend on the number of transcripts [38]. Indeed, researchers showed in simulations that the total noise decreases with autorepression if extrinsic noise dominates. By contrast, the overall noise level increases with autorepression if intrinsic noise is the major contributor. It was deduced that negative feedback evolved in gene regulatory networks, where extrinsic variations dominates [38]. In contrast, positive feedback amplifies fluctuations and thus can lead to multistability of a system [21]. It was reported that the random switching between two states is caused by the combination of noise and a positive feedback loop [35]. For example the wiring diagram of the lactose system contains two positive feedback loops, which involve the lac operon, and thus lead to the observed bistability [48, 7]. Indeed, the results in Section 4.2 exhibited that larger variations can be ascribed to the positive feedback loops. Also, the wiring diagram of the phage-lambda model contains a positive feedback loop, which, together with noise, drives the stochastic cell fate decision [35].

Here, we will focus on the feed-forward loop (FFL), which is a common network motif involving three molecular components, as described for example in [22]. We will now refer to these components by proteins, but part of the molecular components involved in such a feed-forward loop might as well be microRNAs [12]. It consists of two proteins, say proteins A and B that both regulate a third protein C. In addition, protein A also regulates protein B. Each edge between two proteins can be either an inhibiting or activating edge. Since there are three edges involved, it is distinguished between eight types of FFLs. They are divided into two groups, the coherent and incoherent FFLs. The former ones have the feature to direct the same input signs into the third protein C. That is, the direct edge A $\longrightarrow$ C has the same sign as the indirect path A $\longrightarrow$ B $\longrightarrow$ C. The four different types of coherent FFLs are depicted in Figure 5.6. By contrast, incoherent FFLs are characterized by different signs for the two paths from A to C. That is, if the edge A $\longrightarrow$ C is a repressing edge, i.e. with a negative sign, then the overall sign of the indirect path A $\longrightarrow$ B $\longrightarrow$ C has a positive sign, and vice versa. The four different types of incoherent FFLs are shown in Figure 5.7.

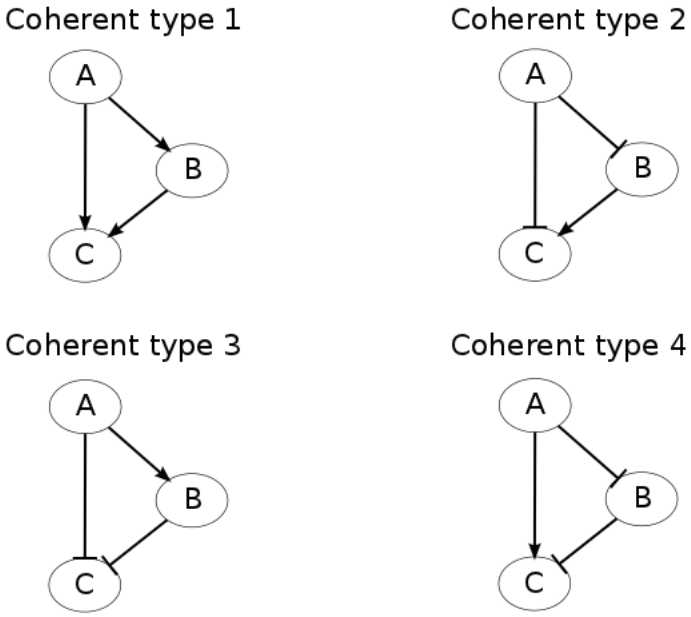

Figure 5.6: The four different types of coherent feed-forward loops.

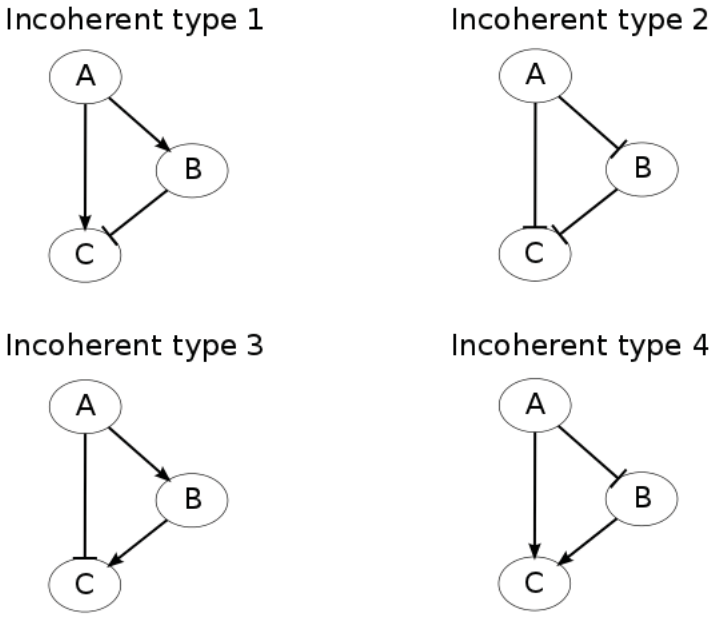

Figure 5.7: The four different types of incoherent feed-forward loops.

This network motif is of particular interest because it recurs often in many gene regulatory networks. For example the coherent FFL type 1 can be found 28 times in the whole gene regulatory network of E. coli and 26 times in the yeast S. cerevisiae [22]. Moreover, it can also be found in neural networks [25]. Interestingly, FFLs can be found in the wiring diagram of the lambda phage model. It is striking that this small network of four nodes contains five FFLs, where cII is the target protein of four of these FFLs. In particular, cII is the target of two coherent FFLs of type 2, and of two incoherent FFLs of type 2. N is the target of an incoherent FFL type 2. One might conjecture that this could be related to the observations from Section 4.1. Namely, an increasing level of extrinsic noise could not change the qualitative behavior of the system. Also, the wiring diagram of the Wnt signaling network features seven FFLs. In particular, the coherent FFL type 1 recurs four times. The other three FFLs are the coherent FFL of type 3, 4 and the incoherent FFL of type 1.

Prior to an examination of the network motifs in the Wnt signaling or the lambda phage network, a general analysis of the noise in FFLs is conducted. The fluctuations in the target protein C of each FFL are compared with the variations in a protein C, which is just regulated by protein A and protein B does not exist. We will refer to this with "two-node network," where A either activates or inhibits C. The average expression level for the coherent FFL type 1, and the two-node network, where A activates C are illustrated in Figure 5.8. This FFL analysis is based on a three node network with $p = 5$, i.e. all three variables can have the same amount of expression levels. The truth tables were designed according to the interactions defined by the FFLs and taking into account that all five states are equally likely in the deterministic update, in order to not have any bias in the update towards a specific value. The truth tables can be found in the Appendix.

As can be seen in Figure 5.8, it is not clear, whether the amount of fluctuations in protein C are larger in the FFL motif or in the two-node network. A similar picture was obtained when the fluctuations of the other coherent FFLs were computed. A way to gain more insight into which fluctuations are actually larger is enabled by the measure of dispersion, the AMSD, defined in equation (3.4) in Section 3.1.

We may plot the measured values for the AMSD versus the increasing variance. This results in the curves of Figure 5.9. The left panel shows the AMSD for the coherent FFL 1 and 4 together with simple activation, inhibition and the single node. The right panel illustrates the AMSD for the coherent FFL type 2 and 3 together with simple activation, inhibition and the single node. The AMSD was evaluated for increasing variances (x-axis). The AMSD is computed based on 100,000 iterations of the system with the probability for a delay of 20%.

The comparison between the different types of coherent FFLs, as well as simple inhibition and activation reveals how the noise is transmitted in the construct of the FFL motif. Based on Figure 5.9, one may draw the conclusion that the coherent FFL of type 1 amplifies the fluctuations. Coherent FFLs type 2 and 3 attenuate the variations. Interestingly, the

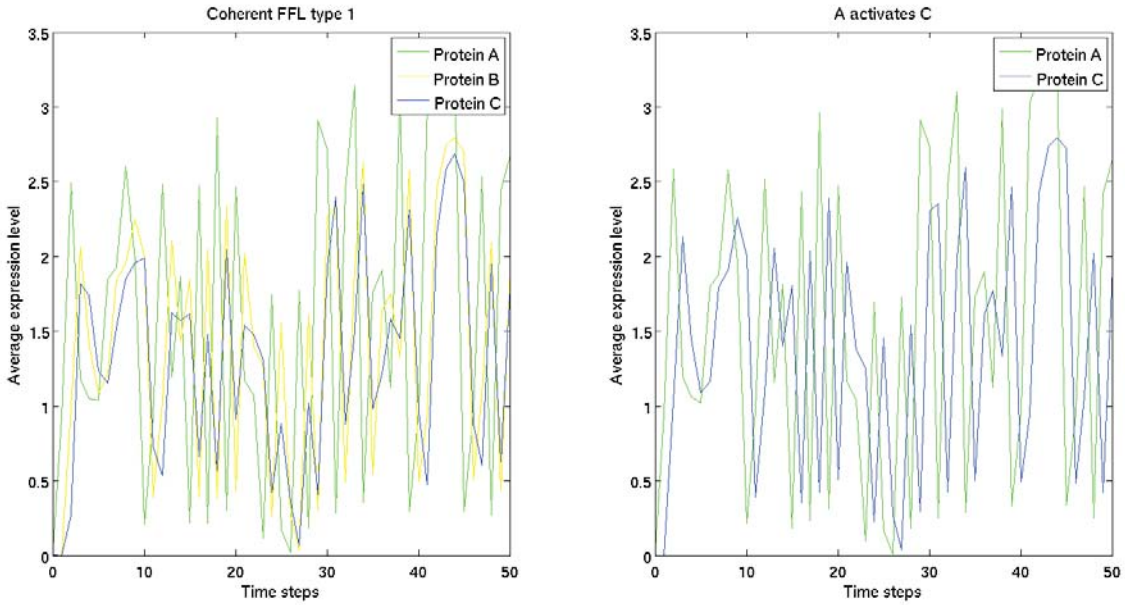

Figure 5.8: The left panel demonstrates the average expression level of the proteins involved in the coherent FFL of type 1. The right panel depicts the average expression level of the proteins A and C, when A activates C. This simulation was done using a variance of $v = 0.3$ and a probability for a delay of 10%.

coherent FFL 4 exhibits smaller variations in the target protein than a simple inhibition until a variance of $v = 0.3$, after which it amplifies fluctuations compared to a simple inhibition. When contrasting the coherent FFL 4 with the simple activation, it is clear that the former dampens noise. Simple activation increases the noise in the target protein compared to inhibition for variances higher than $v = 0.2$. This matches the effect of a positive or negative feedback loop on the noise level, where the former enhances and the latter dampens fluctuations.
The same was done for the incoherent FFLs. Figure 5.10 illustrates the average expression levels of the proteins involved in an incoherent FFL type 1 compared to the fluctuations in the number of proteins in a simple activation two-node-network.

The fluctuations in the number of the target protein C in the incoherent FFL 1 appear to be smaller than the fluctuations generated by the simple activation scheme. To precisely compare the effects of the FFLs on the magnitude of variations in the target protein, the AMSD for all incoherent FFLs with increasing variances was computed.

Figure 5.11 displays the different capabilities of the incoherent FFLs in propagating the noise. The left panel shows the AMSD for the incoherent FFL 1 and 4 together with simple

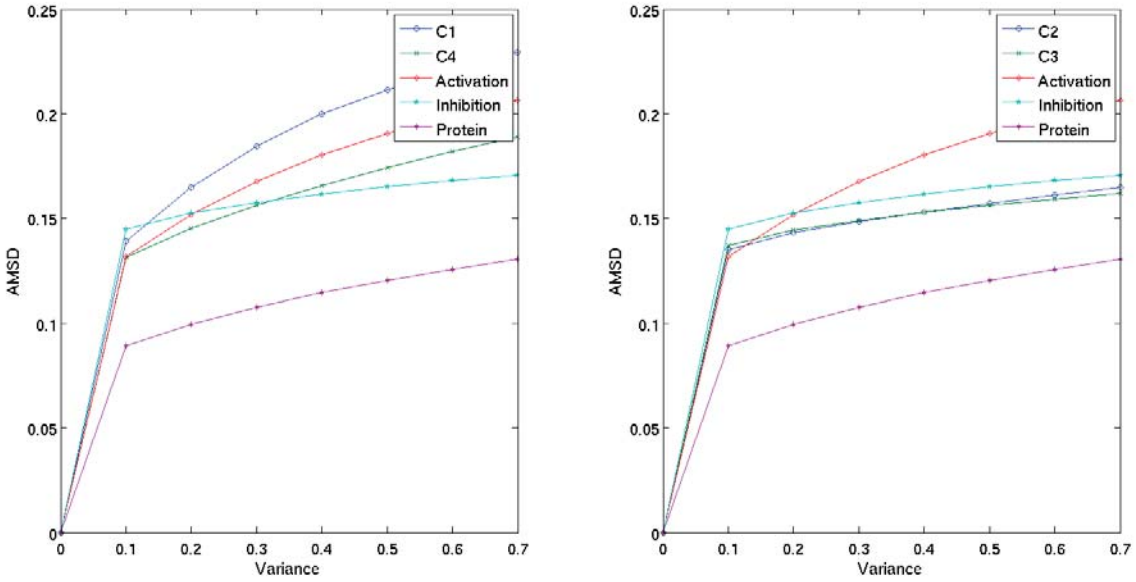

Figure 5.9: The AMSD was measured in the target protein C. This was done for all types of coherent FFLs, the two-node network of activation and inhibition, as well as in the protein C when it is not regulated at all, denoted by "Protein".

activation, inhibition and the single node. The right panel illustrates the AMSD for the incoherent FFL type 2 and 3 together with simple activation, inhibition and the single node. The AMSD was evaluated for increasing variances (x-axis). The AMSD is computed based on 100,000 iterations of the system and the probability for a delay of 20%. This illustration reveals that the incoherent FFL 1 decreases fluctuations for variances higher than $v = 0.2$, compared to simple activation. The incoherent FFLs 3 and 4 amplify the variations. Interestingly, the incoherent FFL 2 exhibits the same noise level in the target protein as a simple inhibition. Again, it becomes pronounced that a simple activation increases the variations for variances higher than $v = 0.2$, when compared to a simple repression, similar to what would be expected from positive or negative feedback loops. Furthermore, the incoherent FFL 2 exhibits smaller noise levels for variances higher than $v = 0.2$ when compared to simple activation.
This analysis also demonstrated that the noise level in a single unregulated protein without inputs is still less than the noise level in the target protein of an FFL or a two-node network of simple activation or inhibition.

Now the insight of this FFL analysis will help to investigate the Wnt signaling network. An especially interesting motif can be found between the four variables Intracellular Receptors,

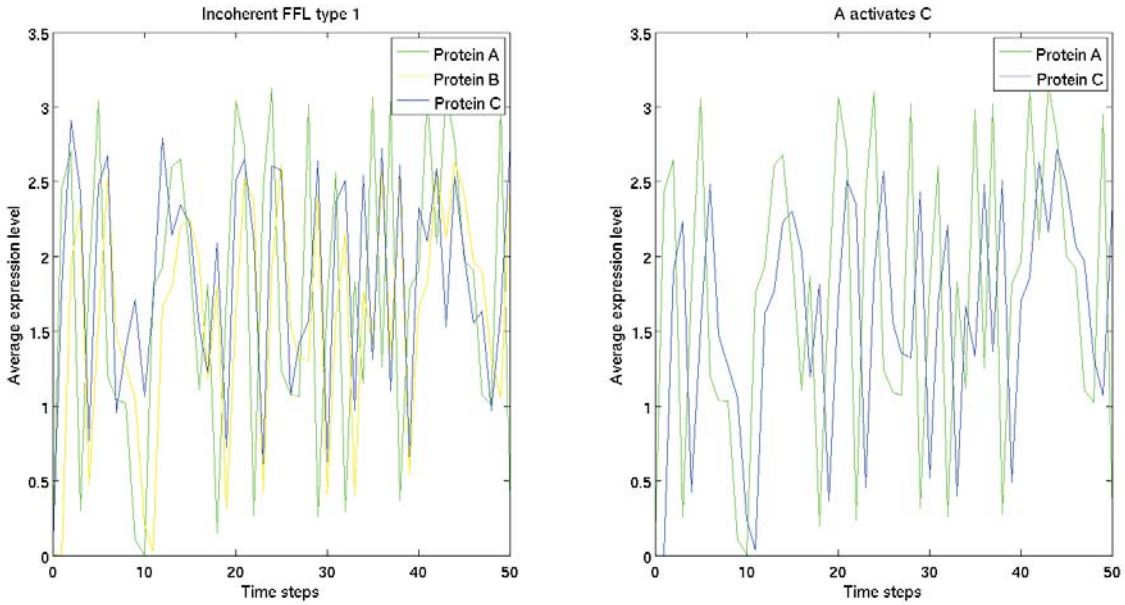

Figure 5.10: The left panel illustrates the average expression level of the proteins involved in the incoherent FFL of type 1. The right panel shows the average expression level of the proteins A and C, when A activates C. This simulation was done using a variance of $v = 0.3$ and a probability for a delay of 10%.

DKK, Kremen and LRP6. These four nodes are intertwined with three FFLs. The triangle consisting of Intracellular Receptors, DKK and Kremen builds a coherent FFL of type 1. Additionally, the three proteins DKK, Kremen and LRP6 form a coherent FFL type 3. Finally, an incoherent FFL of type 1 is detectable between the Intracellular receptors, Kremen and LRP6. It is striking that LRP6 is the target of two FFLs. To compare the amount of fluctuations in these proteins and to assess whether it is possible to find similar effects in these FFLs as observed in the general FFL analysis, the AMSD was computed.
In Figure 5.12, the average expression levels are shown for this network motif. The corresponding AMSD values are depicted in table 5.2. Note that these proteins have different ranges of expression levels, for example Kremen can have four states, whereas DKK is only either zero or one. Thus, the values of the AMSD might tend to be higher for proteins with smaller ranges of expression levels. However, in this case, the results correspond to the ones from the general FFL analysis.

Table 5.2 reveals that the fluctuations in Kremen, which is the target of a coherent FFL 1, are indeed of highest magnitude in this motif. This is consistent with the general FFL analysis, where the coherent FFL 1 was found to amplify fluctuations. Interestingly, LRP6, which is the target protein of two intertwined FFLs exhibits the lowest noise level. That

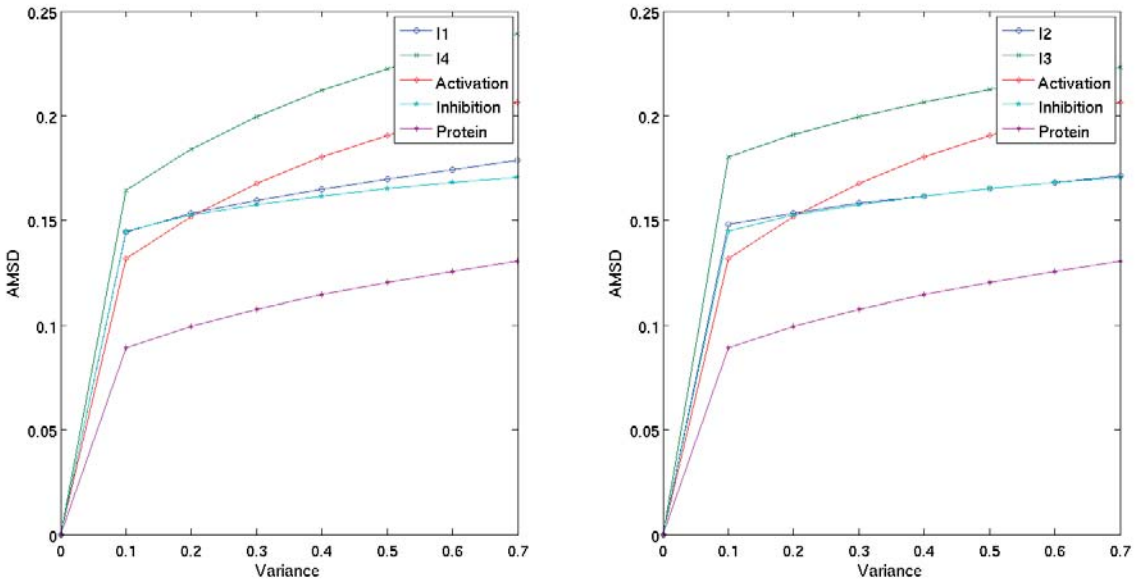

Figure 5.11: The AMSD was measured in the target protein C. This was done for all types of incoherent FFLs, the two-node network of activation and inhibition, as well as in the protein C when it is not regulated at all, denoted by "Protein".

	Intracellular Receptors	DKK	Kremen	LRP6
AMSD	0.2205	0.2486	0.2932	0.1110

Table 5.2: Interestingly, LRP6, which is the target protein of the intertwined FFLs exhibits the lowest noise level. In contrast, Kremen, which is the target of a coherent FFL features the largest fluctuations.

is, it is the target of the coherent FFL 3 and the incoherent FFL 1. When recalling the abilities of these two FFLs in attenuating noise by considering the values for the AMSDs in Figures 5.9 and 5.11, the following pattern becomes clear: Both types of FFL propagate less noise than simple activation for variances greater or equal $v = 0.3$. The presence of a sufficient level of extrinsic noise might actually be the case since the noise level in higher eukaryotes was found to be remarkably high [32]. Additionally, the coherent FFL 3 reveals to dampen the fluctuations compared to both, simple activation and inhibition. In light of these properties, both FFLs might actually be able to attenuate stochasticity. Thus, instead of just a simple regulation from the Intracellular Receptors to LRP6, this complex network motif might have evolved to enable precise coordination of the regulation of LRP6. And this explains the low value of AMSD found for LRP6. When considering the wiring diagram of

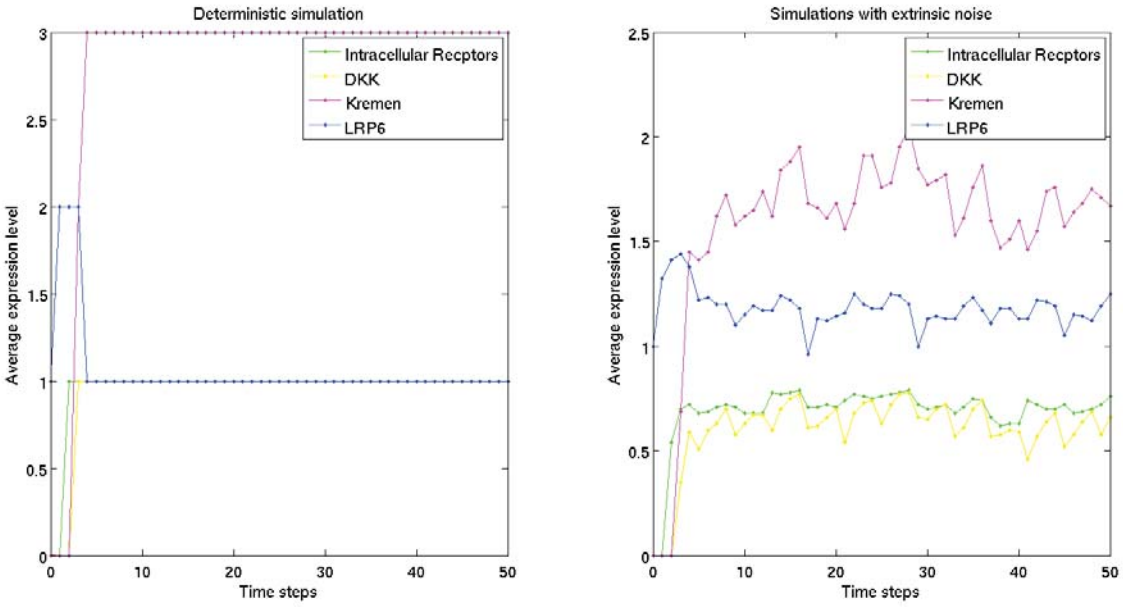

Figure 5.12: The network motif of the four variables Intracellular Receptors, DKK, Kremen and LRP6, which are connected with three FFLs. The average expression level without and with extrinsic noise (variance = 0.6) is shown. The delay probability is 30%. This simulation represents the early stage melanoma cells, when RA and SFRP siRNA are added.

the Wnt signaling network, it becomes clear that LRP6 is an important regulator for the decision for either the canonical or the PCP pathway. This might justify, why the network has evolved in a way to reduce fluctuations in the state of LRP6.

Concerning the wiring diagram of the lambda phage model, it was observed that cII is the target of two coherent FFLs of type 2 and of two incoherent FFLs of type 2. Figure 5.9 revealed that the coherent FFL type 2 attenuates fluctuations compared to simple activation or inhibition. Likewise, the incoherent FFL type 2 was found to dampen variations compared to simple activation for variances greater than $v = 0.2$, as depicted in Figure 5.11. Hence, the interactions among the lambda phage genes might have evolved to ensure precise levels of cII. The AMSD values computed for the four viral genes cI, cro, cII, and N are listed in Table 5.3. It is remarkable that the AMSD for N, and especially for cII is very low. This becomes pronounced when comparing it to the noise level of a single protein without any inputs. Recall that this exhibited the smallest AMSD values in the Figures 5.9 and 5.11. However, the AMSD for cII is even smaller. Indeed, cII was identified as a pro-lysogeny gene, and a critical regulator of the decision between lysis and lysogeny [40].

	cI	cro	cII	N	single protein
AMSD	0.1303	0.1396	0.0585	0.0664	0.1076

Table 5.3: These AMSD values were computed in a simulation with variance $v = 0.3$ and a probability for a delay of 20% for all variables. The rightmost column displays the AMSD value for a single protein, which is not regulated at all.

Conclusively, these insights support the hypothesis that gene regulatory networks might have evolved to decrease noise when precise coordination is necessary. It would be interesting to conduct further investigation.

5.3 Conclusion

We introduced a way to model extrinsic noise in gene regulatory networks, which arises due to both, global extrinsic sources of noise and pathway-specific extrinsic variations. Furthermore, the emergence of the observed lognormal protein distributions was examined and attributed to the multiplicative effect of biochemical reactions. Also, ways of distinguishing between intrinsic and extrinsic sources of noise were discussed. The effect of extrinsic noise on the concentration of a single protein was illustrated and found to be consistent with a continuous way of modeling extrinsic stochasticity. Next, the modeling framework was applied to three different gene regulatory networks. First, the model to represent the outcome of infection of bacteria by phage lambda was embedded in the extrinsic noise framework. It successfully captured the effect of certain factors that significantly contribute to the random decision between lysis and lysogeny. Second, the extrinsic noise was introduced into the gene regulatory network that describes the lactose utilization system in E. coli. We demonstrated that it is able to explain the bistable switch between the two states of the lac operon being expressed or inactivated. It could also demonstrate the fitness advantage in fluctuating environments that comes along with noise. Third, the extrinsic framework was applied to a recently designed signaling network in melanoma cells, which revealed the robustness of this network. Furthermore, the effect of extrinsic noise on a gene regulatory network in general was analyzed. It was demonstrated that extrinsic perturbations significantly increase the connectivity of the state space and thus can lead to large variations in the dynamics of a system. Based on the quite robust behavior of the Wnt signaling network, which includes seven feed-forward loops, a general analysis of these network motifs was conducted to elucidate the effects of noise. In particular, we proposed a measure of dispersion for the discrete modeling framework and applied it to the eight types of feed-forward loops to gain insight into the ability of each of these motifs to propagate noise. It was revealed that some feed-forward loops attenuate the fluctuations, while others have an amplifying effect. For some of the

motifs, this characteristic was demonstrated to change with increasing variances. It would be interesting to further analyze the effect of feed-forward loops on the whole state space. The features of the feed-forward loops revealed here could serve as a basis for this examination. Moreover, it might also be promising to analyze other network motifs and measure the transmitted fluctuations. The analysis of network motifs enables to draw conclusions of why certain networks evolved in a particular way. The knowledge obtained thereby about the ability of certain network structures to reduce noise could potentially be transferred to any network, for example networks of computers that transmit information.
Likewise, it would be fascinating to combine both discrete frameworks, the intrinsic and the extrinsic one together, since both types of noise arise in a realistic biological system. Also, the effects of feed-forward loops to attenuate or amplify fluctuations might become more pronounced with both types of noise.
Finally, the fact that extrinsic perturbations actually cause large variations suggests the contribution of extrinsic noise to bet-hedging strategies. This points towards the paradigm that evolution might be not only the interplay between mutation and selection but rather the combination of mutation and noise counter-regulated by selection.

Chapter 6

Bibliography

[1] T. Akerlund, K. Nordstrom, and R. Bernander. Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of Escherichia coli. *J. Bacteriol.*, 177(23):6791–6797, Dec 1995.

[2] B. Banerjee, S. Balasubramanian, G. Ananthakrishna, T. V. Ramakrishnan, and G. V. Shivashankar. Tracking operator state fluctuations in gene expression in single cells. *Biophys. J.*, 86(5):3052–3059, May 2004.

[3] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai. Noise in protein expression scales with natural protein abundance. *Nat. Genet.*, 38(6):636–643, Jun 2006.

[4] A. Becskei, B. B. Kaufmann, and A. van Oudenaarden. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.*, 37(9):937–944, Sep 2005.

[5] Edwin Crow, L. and Kunio Shimizu. History, genesis, and properties. In Edwin Crow, L. and Kunio Shimizu, editors, *Lognormal distributions. Theory and applications*, pages 1–25. Marcel Dekker, Inc., New York, 1988.

[6] M. J. Dunlop, R. S. Cox, J. H. Levine, R. M. Murray, and M. B. Elowitz. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat. Genet.*, 40(12):1493–1498, Dec 2008.

[7] A. Eldar and M. B. Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, Sep 2010.

[8] Michael Elowitz, B., Arnold Levine, J., Eric Siggia, D., and Peter Swain, S. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584):1183–1186, 2002.

[9] Chikara Furusawa, Takao Suzuki, Akiko Kashiwagi, Tetsuya Yomo, and Kunihiko Kaneko. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1:25–31, 2005.

[10] A. Hilfinger and J. Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc. Natl. Acad. Sci. U.S.A.*, 108(29):12167–12172, Jul 2011.

[11] Franziska Hinkelmann, Beverly Delidow, and Reinhard Laubenbacher. Downregulation of LRP6 inhibits growth of melanoma cells. *Under review*, 2012.

[12] E. Hornstein and N. Shomron. Canalization of development by microRNAs. *Nat. Genet.*, 38 Suppl:S20–24, Jun 2006.

[13] L. Huang, A. Hwang, and J. Phillips. Effect of temperature on microbial growth rate-mathematical analysis: the arrhenius and eyring-polanyi connections. *J. Food Sci.*, 76(8):E553–560, Oct 2011.

[14] I. G. Johnston, B. Gaal, R. P. Neves, T. Enver, F. J. Iborra, and N. S. Jones. Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Comput. Biol.*, 8(3):e1002416, 2012.

[15] J. P. Junker and A. van Oudenaarden. When noisy neighbors are a blessing: analysis of gene expression noise identifies coregulated genes. *Mol. Cell*, 45(4):437–438, Feb 2012.

[16] H. A. Kestler and M. Kuhl. From individual Wnt pathways towards a Wnt signalling network. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 363(1495):1333–1347, Apr 2008.

[17] S. Krishna, B. Banerjee, T. V. Ramakrishnan, and G. V. Shivashankar. Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 102(13):4771–4776, Mar 2005.

[18] Reinhard Laubenbacher and Bernd Sturmfels. Computer Algebra in Systems Biology. *The American Mathematical Monthly*, 116(10):882–891, Dec 2009.

[19] Eckhard Limpert, Werner Stahel, A., and Markus Abbt. Log-normal Distributions across the Sciences: Keys and Clues. *BioScience*, 51(5):341–352, 2001.

[20] H. Maamar, A. Raj, and D. Dubnau. Noise in gene expression determines cell fate in Bacillus subtilis. *Science*, 317(5837):526–529, Jul 2007.

[21] N. Maheshri and E. K. O'Shea. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu Rev Biophys Biomol Struct*, 36:413–434, 2007.

[22] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *Proc. Natl. Acad. Sci. U.S.A.*, 100(21):11980–11985, Oct 2003.

[23] T. C. Meng, S. Somani, and P. Dhar. Modeling and simulation of biological systems with stochasticity. *In Silico Biol. (Gedrukt)*, 4(3):293–309, 2004.

[24] J. T. Mettetal, D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden. Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. U.S.A.*, 103(19):7304–7309, May 2006.

[25] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, Oct 2002.

[26] D. Murrugarra, A. Veliz-Cuba, B. Aguilar, S. Arat, and R. Laubenbacher. Modeling stochasticity and variability in gene regulatory networks. *EURASIP J Bioinform Syst Biol*, 2012(1):5, Jun 2012.

[27] A. B. Oppenheim, O. Kobiler, J. Stavans, D. L. Court, and S. Adhya. Switches in bacteriophage lambda development. *Annu. Rev. Genet.*, 39:409–429, 2005.

[28] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415–418, Jan 2004.

[29] Johan Paulsson. Models of stochastic gene expression. *Physics of Life Reviews*, 2(2):157–175, 2005.

[30] J. M. Pedraza and A. van Oudenaarden. Noise propagation in gene networks. *Science*, 307(5717):1965–1969, Mar 2005.

[31] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.*, 4(10):e309, Oct 2006.

[32] A. Raj and A. van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, Oct 2008.

[33] S. Ramanathan and P. S. Swain. Tracing the sources of cellular variation. *Dev. Cell*, 9(5):576–578, Nov 2005.

[34] J. M. Raser and E. K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, Jun 2004.

[35] J. M. Raser and E. K. O'Shea. Noise in gene expression: origins, consequences, and control. *Science*, 309(5743):2010–2013, Sep 2005.

[36] Nitzan Rosenfeld, Jonathan W. Young, Uri Alon, Peter Swain, S., and Michael Elowitz, B. Gene Regulation at the Single-Cell Level. *Science*, 307(5715):1962–1965, Mar 2005.

[37] K. Sato, Y. Ito, T. Yomo, and K. Kaneko. On the relation between fluctuation and response in biological systems. *Proc. Natl. Acad. Sci. U.S.A.*, 100(24):14086–14090, Nov 2003.

[38] V. Shahrezaei, J. F. Ollivier, and P. S. Swain. Colored extrinsic fluctuations and stochastic gene expression. *Mol. Syst. Biol.*, 4:196, 2008.

[39] V. Shahrezaei and P. S. Swain. The stochastic nature of biochemical networks. *Curr. Opin. Biotechnol.*, 19(4):369–374, Aug 2008.

[40] F. St-Pierre and D. Endy. Determination of cell fate selection during phage lambda infection. *Proc. Natl. Acad. Sci. U.S.A.*, 105(52):20705–20710, Dec 2008.

[41] M. Stamatakis, R. M. Adams, and G. Balazsi. A common repressor pool results in indeterminacy of extrinsic noise. *Chaos*, 21(4):047523, Dec 2011.

[42] J. Starrfelt and H. Kokko. Bet-hedging-a triple trade-off between means, variances and correlations. *Biol Rev Camb Philos Soc*, 87(3):742–755, Aug 2012.

[43] J. Stewart-Ornstein and H. El-Samad. Stochastic modeling of cellular networks. *Methods Cell Biol.*, 110:111–137, 2012.

[44] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 99(20):12795–12800, Oct 2002.

[45] P. S. Swain and A. Longtin. Noise in genetic and neural networks. *Chaos*, 16(2):026101, Jun 2006.

[46] D. Thieffry and R. Thomas. Dynamical behaviour of biological regulatory networks–II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.*, 57(2):277–297, Mar 1995.

[47] S. Tsuru, J. Ichinose, A. Kashiwagi, B. W. Ying, K. Kaneko, and T. Yomo. Noisy cell growth rate leads to fluctuating protein concentration in bacteria. *Phys Biol*, 6(3):036015, 2009.

[48] A. Veliz-Cuba and B. Stigler. Boolean models can explain bistability in the lac operon. *J. Comput. Biol.*, 18(6):783–794, Jun 2011.

[49] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439(7078):861–864, Feb 2006.

[50] A. Wagner. Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.*, 22(6):1365–1374, Jun 2005.

Appendix A

Coordinate functions of the lambda phage model

The coordinate functions of the discrete dynamical system that describes the infection of the bacterium with phage lambda read as [26]

$$f_1 = (1 + x_1^4 + 4x_2^4 + x_1x_2^4 + x_1^2x_2^4 + x_1^3x_2^4 + x_1^4x_2^4 + x_2^4x_3^4 + 4x_1x_2^4x_3^4 + 4x_1^2x_2^4x_3^4 + 4x_1^3x_2^4x_3^4 + \\ + 4x_1^4x_2^4x_3^4) \mod 5,$$

$$f_2 = (1 + 4x_1 + 4x_1^2 + 4x_1^3 + 3x_1^4 + 2x_2 + 3x_1x_2 + 3x_1^2x_2 + 3x_1^3x_2 + x_1^4x_2 + x_2^2 + 3x_2^3 + x_1x_2^3 + \\ + x_1^2x_2^3 + x_1^3x_2^3 + 2x_1^4x_2^3) \mod 5,$$

$$f_3 = (x_4^4 + 4x_1x_4^4 + 4x_1^2x_4^4 + 4x_1^3x_4^4 + 3x_1^4x_4^4 + x_2x_4^4 + 4x_1x_2x_4^4 + 4x_1^2x_2x_4^4 + 4x_1^3x_2x_4^4 + 3x_1^4x_2x_4^4 + \\ + 2x_2^3x_4^4 + 3x_1x_2^3x_4^4 + 3x_1^2x_2^3x_4^4 + 3x_1^3x_2^3x_4^4 + x_1^4x_2^3x_4^4 + 2x_2^4x_4^4 + 3x_1x_2^4x_4^4 + 3x_1^2x_2^4x_4^4 + \\ + 3x_1^3x_2^4x_4^4 + x_1^4x_2^4x_4^4) \mod 5,$$

$$f_4 = (1 + 4x_1^4 + 4x_2 + x_1^4x_2 + 4x_2^2 + x_1^4x_2^2 + 4x_2^3 + x_1^4x_2^3 + 3x_2^4 + 2x_1^4x_2^4) \mod 5,$$

and based on these functions, the truth tables can be generated.

Appendix B

Truth tables of the lac operon model

The truth table for the lac operon mRNA can be found in table B.1.

R	C	M
0	0	0
0	1	1
1	0	0
1	1	0
2	0	0
2	1	0

Table B.1: Truth table for the lac operon mRNA, denoted with M.

The truth table for the repressor LacI can be found in table B.2.

The truth table for intracellular lactose can be found in table B.3.

The truth table for the protein β-galactoside permease can be found in table B.4.

The truth table for β-galactosidase can be found in table B.5.

R	A	R
0	0	2
0	1	0
0	2	0
1	0	2
1	1	0
1	2	0
2	0	2
2	1	1
2	2	1

Table B.2: Truth table for the repressor LacI, abbreviated with R.

P	Ge	Le	L
0	0	0	0
0	0	1	0
0	0	2	1
0	1	0	0
0	1	1	0
0	1	2	0
1	0	0	0
1	0	1	1
1	0	2	2
1	1	0	0
1	1	1	0
1	1	2	0

Table B.3: The truth table for intracellular lactose, denoted by L.

M	P
0	0
1	1

Table B.4: The truth table for β-galactoside permease, abbreviated with P.

The truth table for the protein CAP can be found in table B.6.

The truth table for allolactose can be found in table B.7.

M	B
0	0
1	1

Table B.5: The truth table for β-galactosidase, denoted by B.

Ge	C
0	1
1	0

Table B.6: The truth table for protein CAP, abbreviated with C.

B	L	A
0	0	0
0	1	1
0	2	1
1	0	0
1	1	1
1	2	2

Table B.7: The truth table for allolactose, denoted by A.

Appendix C

Truth table of the example in Section 5.1

x_1	x_2	x_1	x_2
0	0	0	0
0	1	1	0
0	2	0	1
0	3	0	1
1	0	2	0
1	1	1	0
1	2	1	1
1	3	0	2
2	0	3	1
2	1	2	2
2	2	2	2
2	3	1	3
2	0	3	2
2	1	3	3
2	2	3	3
2	3	2	3

Table C.1: The truth table for x_1 and x_2.

Appendix D

Truth tables for the feed-forward loop analysis in Section 5.2

The FFL analysis is based on a three node network with the variables A, B and C and $p = 5$, i.e. all the three variables can have the same amount of expression levels, which are $\{0, 1, 2, 3, 4\}$. The truth tables were designed according to the interactions defined by the FFLs and taking into account that all five states are equally likely in the deterministic update, in order to not have any bias in the update towards a specific value. The node A acts as a pure input and is updated randomly according to a fixed sequence, which is determined prior to each simulation. The random selection of the states ensures that all states are equally likely for A.
The truth table for the case, where A activates C can be found in Table D.1.

A	C
0	0
1	1
2	2
3	3
4	4

Table D.1: The truth table for the simple activation, where A activates C.

The truth table for the simple inhibition, where A represses C is shown in Table D.2.

The truth tables for simple activation and inhibition are the same as the ones for the update of protein B, since B is only regulated by A. Now, the truth tables for the target protein C for the different coherent FFLs are shown. Table D.3 shows the truth table for the target

A	C
0	4
1	3
2	2
3	1
4	0

Table D.2: The truth table for the simple inhibition, where A represses C.

protein C of the coherent FFL 1 and the incoherent FFL 4. Table D.4 displays the truth table for the target protein C in the coherent FFL type 2 and the incoherent FFL type 3. The truth table for the target protein in the coherent FFL 3 and the incoherent FFL 2 is depicted in Table D.5. Finally, the Table D.6 depicts the updates of C for the coherent FFL of type 4 and the incoherent FFL of type 1.

A	B	C
0	0	0
0	1	0
0	2	0
0	3	0
0	4	1
1	0	0
1	1	1
1	2	1
1	3	2
1	4	2
2	0	1
2	1	1
2	2	2
2	3	2
2	4	3
3	0	2
3	1	3
3	2	3
3	3	4
3	4	4
4	0	3
4	1	3
4	2	4
4	3	4
4	4	4

Table D.3: The truth table for the target protein C in the coherent FFL type 1 and the incoherent FFL type 4. The target proteins are identically regulated in these two FFLs.

A	B	C
0	0	2
0	1	3
0	2	4
0	3	4
0	4	4
1	0	1
1	1	2
1	2	3
1	3	3
1	4	4
2	0	0
2	1	1
2	2	2
2	3	3
2	4	4
3	0	0
3	1	1
3	2	1
3	3	2
3	4	3
4	0	0
4	1	0
4	2	0
4	3	1
4	4	2

Table D.4: The truth table for the target protein C in the coherent FFL type 2 and the incoherent FFL type 3. They are identically regulated in these two FFL motifs.

A	B	C
0	0	4
0	1	4
0	2	4
0	3	3
0	4	3
1	0	4
1	1	4
1	2	3
1	3	3
1	4	2
2	0	3
2	1	2
2	2	2
2	3	1
2	4	1
3	0	2
3	1	2
3	2	1
3	3	1
3	4	0
4	0	1
4	1	0
4	2	0
4	3	0
4	4	0

Table D.5: The truth table for the target protein in the coherent FFL type 3 and the incoherent FFL type 2. They regulate the target protein C in the same way.

A	B	C
0	0	2
0	1	1
0	2	0
0	3	0
0	4	0
1	0	3
1	1	2
1	2	1
1	3	1
1	4	0
2	0	4
2	1	3
2	2	2
2	3	1
2	4	0
3	0	4
3	1	3
3	2	3
3	3	2
3	4	1
4	0	4
4	1	4
4	2	4
4	3	3
4	4	2

Table D.6: The truth table for the protein C in the coherent FFL type 4 and the incoherent FFL type 1. Both motifs regulate C identically.