A genetic circuit for epigenetic memory

Georg Fritz

Freiburg 2006

A genetic circuit for epigenetic memory

Georg Fritz

Diploma Thesis
submitted by
Georg Fritz

Supervisor:
Prof. Dr. Jens Timmer

Physikalisches Institut
Fakultät für Mathematik und Physik

October 2006

Albert-Ludwigs-Universität
Freiburg im Breisgau

Publications

[1] W. Schmidt and G. Fritz, On the geometry of sunspot penumbral filaments, *Astronomy & Astrophysics* **421**, 735 (2004)

[2] D.A.N. Mueller, R. Schlichenmaier, G. Fritz, and C. Beck, Net circular polarization of sunspot penumbrae, *Proceedings of SOHO-17* (2006)

[3] D.A.N. Mueller, R. Schlichenmaier, G. Fritz, and C. Beck, The multi-component field topology of sunspot penumbrae- A diagnostic tool for spectropolarimetric measurements, *Astronomy & Astrophysics* (accepted September 6th, 2006)

[4] G. Fritz, N.E. Buchler, T. Hwa, and U. Gerland, A genetic circuit for epigenetic memory (under review in Proc. Natl. Acad. Sci. USA)

Invited Talks and Selected Posters

[1] Poster at the International Conference on Systems Biology, Heidelberg (2004): M. Schilling, U. Klingmueller, T. Maiwald, K. Bartholome, G. Fritz, M. Kollmann, and J. Timmer, *Identification of damped oscillation as a crucial property of the MAP-kinase pathway revealed by data-based modeling*

[2] Talk in the Munich Systems Biology Forum (2006): *A genetic circuit that memorizes a signal on command*

[3] Talk in the Soft Condensed Matter Group of Prof. Rädler, Munich (2006): *A genetic circuit that memorizes a signal on command*

[4] Poster at the DPG conference in Dresden (2006): G. Fritz, N.E. Buchler, T. Hwa, U. Gerland, *A genetic circuit that memorizes a signal on command*

[5] Poster at the DPG conference in Dresden (2006): J. Leierseder, G. Fritz, K. Jung, and J. Rädler, *Fluorescence analysis of environmental stress response of single cells within a bacterial population*

[6] Poster at the Summer School for Quantitative Approaches to Gene Regulatory Systems, San Diego (2006): G. Fritz, N.E. Buchler, T. Hwa, U. Gerland, *A genetic circuit for epigenetic memory*

[7] Poster at the International Conference on Molecular Systems Biology, Munich (2006): C. Koller, K. Burdack, G. Fritz, C. Küper, U. Gerland and K. Jung, *Characterization of the pH-responsive module Cad of Escherichia coli*

Zusammenfassung

Lebende Zellen sind einer Vielzahl von fluktuierenden Umwelteinflüssen ausgesetzt, wie zum Beispiel variierende Nahrungsvorkommen, die Konfrontation mit toxischen Stoffen, oder sich ändernde Zelldichten. Um auf diese Signale in einer sinnvollen Weise zu reagieren, wurden durch die Evolution immer ausgefeiltere biochemische Netzwerke selektiert, die die Zelle befähigen verschiedenste Signale auf einer schnellen Zeitskala zu verarbeiten. Das Erforschen dieser regulatorischen Netzwerke ist von zentraler Bedeutung für unser Verständnis der Verbindung zwischen unbelebter und belebter Natur. Die verwickelten regulatorischen Strukturen dieser Netzwerke bringen jedoch solch eine Komplexität mit sich, dass angesichts der derzeitigen Datenlage eine Analyse der Netzwerke als ganzes nicht sinnvoll erscheint. Kleine, isolierte *Netzwerkmodule* können auf der anderen Seite sowohl theoretisch als auch experimentell sehr wohl untersucht werden und dieser Ansatz wurde auch in den jüngsten Jahren erfolgreich verfolgt.

Die vorliegende Arbeit beschäftigt sich mit der Frage, wie solche Netzwerke (basierend auf den Prinzipien der *Genregulation*) gewisse Signale aus ihrer Umwelt *speichern* können. Es wird ein minimaler genetischer Schaltkreis vorgeschlagen, welcher in der Lage ist, ein Signal in Form einer Proteinkonzentration zu speichern. Das Verhalten dieses Regelkreises ist dem Speicher in einem elektronischen Schaltkreis sehr ähnlich: Er liest und speichert das Eingangssignal nur, wenn ein Speichersignal vorliegt. Das Speichern geschieht also nur *bedingt* und wird deshalb als *konditional* bezeichnet. Die Architektur des vorgeschlagenen Schaltkreises basiert auf einem vormals schon experimentell realisierten "genetischen Schalter", welcher aus zwei, sich gegenseitig unterdrückenden Genen besteht. Es wurde gezeigt, dass dieser zentrale Baustein *bistabiles Verhalten* zeigt und somit geeignet ist, um die beiden logischen Zustände 0 und 1 einer Speichereinheit zu repräsentieren. Dieser Schalter wird durch ein regulierendes "Front-end" ergänzt, welches das Eingangs- und Speichersignal im Form von zwei weiteren Proteinen enthält. Diese Proteine können Homo- und Heterodimere bilden und in diesen beiden Konformationen selektiv die Produktion der jeweiligen "Schalter-Proteine" unterdrücken. Dadurch kann zwischen den Zuständen 0 und 1 umgeschaltet werden.

Die hochgradig stochastische Natur chemischer Reaktionen bei typischerweise sehr niedrigen Molekülzahlen (oft sind weniger als 100 Moleküle einer Spezies pro Zelle vorhanden) erfordert eine theoretische Analyse sowohl mit deterministischen als auch mit stochastischen Techniken. Zur Untersuchung der Gleichgewichtseigenschaften des Schaltkreises in Abhängigkeit von den realistisch gewählten Parametern wurden die deterministischen

Reaktions Raten-Gleichungen verwendet. Es wird gezeigt, dass der Schaltkreis für die Proteine TetR und LacI tatsächlich bistabiles Verhalten aufweist und das Zustandsdiagramm (analog zum Phasendiagramm in der Thermodynamik) wird präsentiert. Des Weiteren wird der Einfluss des regulierenden Front-ends auf den zentralen Schalter untersucht und es wird demonstriert, dass das resultierende Verhalten analog zu dem des "mean-field" Ising Modells in einem magnetischen Feld ist, wobei das Eingangs- und Speichersignal jeweils dem äusseren Feld und der Temperatur entsprechen. Es wird weiter dargelegt, wie man die Sensitivität des Schaltkeises an gegebene Signalamplituden durch Mutationen in den DNA-Bindestellen der Homo- und Heterodimere anpassen kann.

Das dynamische Verhalten des Speicherelements in Abhängigkeit von zeitabhängigen Eingangs- und Speichersignalen macht klar, dass die gewünschte Fähigkeit Signale auf schnellen Zeitskalen (ca. 30 min) zu lesen und zu speichern tatsächlich gegeben ist. Stochastische Simulationen mit dem Gillespie Algorithmus zeigen weiter, dass trotz der stochastisch induzierten Übergänge zwischen den beiden "stabilen" Zuständen, die Halbwertszeit des "Gedächtnisverlusts" bei mindestens 40 Stunden liegt. Zusätzlich werden noch neue stochastisch induzierte Fehler, die während des Speichervorgangs stattfinden, identifiziert. Diese Fehler können jedoch minimiert werden, wenn das Speichersignal nur ausreichend lange präsent ist.

Die Struktur dieser Arbeit ist wie folgt: Im ersten Kapitel wird eine kurze historische Einleitung in die Molekularbiologie gegeben und die Motivation für die vorliegende Arbeit wird aufgezeigt. Das zweite Kaptitel gibt einen Überblick über die derzeitigen quantitiven Zugänge zur Genregulation, führt den Leser in die Genregulations-Netzwerke ein und diskutiert die wichtigsten mathematischen Konzepte zu deren Beschreibung. Zusätzlich werden die Quellen biochemischer Fluktuationen in diesen Netzwerken diskutiert. Im dritten Kapitel werden die eingeführten Konzepte auf den oben beschriebenen "konditionalen Speicher" angewendet und dessen Verhalten sowohl mit deterministischen als auch mit stochastischen Methoden untersucht. Das letzte Kapitel diskutiert die Resultate und beleuchtet sie vom Standpunkt einer möglichen experimentellen Realisierung.

Summary

Living cells are exposed to a variety of fluctuating environmental signals, such as the nutrient supply, the concentration of toxics or the cell density. In order to respond to these signals in a reasonable way, evolution selected for sophisticated biochemical reaction networks that allow the cell to integrate and process a multitude of signals on a rapid timescale. The comprehension of these regulatory networks belongs to the most challenging tasks in understanding life and is important to build the bridge between the inanimate and the animate matter. However, the intricate structure of these networks confers an enormous complexity, that makes it in the lack of detailed biochemical data prohibitively difficult to analyze these networks as a whole. The theoretical and experimental characterization of small and compact network *modules* on the other hand, seems feasible and has been pursued in the recent years.

The present work deals with the question, how the capability to *memorize* certain environmental signals can be realized in a genetic network. A minimal genetic circuit is proposed, that conditionally memorizes a signal in the form of a protein concentration. The circuit behaves qualitatively similar to memory in an electronic circuit: it reads and stores an input signal only when conditioned to do so by a read signal. The circuit is based on a previous experimental realization of the "genetic toggle switch", which consists of two mutually repressing genes. This "core unit" exhibits bistable behavior, thus suitable for the implementation of the logic states 0 and 1. It is complemented with a "regulatory front end" comprising two additional proteins, which carry the input and the read signal, respectively. Moreover, these proteins may form homo- or heterodimers, which selectively repress the synthesis of the two toggle switch proteins. Thereby it is possible to switch between the states 0 and 1.

Due to the highly stochastic nature of the biochemical reaction events involved in these processes, the theoretical analysis is carried out by using both deterministic and stochastic simulation techniques. The deterministic rate equations are used to explore the steady state behavior of the circuit in dependence on the realistically chosen parameters. It is shown that the circuit exhibits bistable behavior for the proteins TetR and LacI and the state diagram (analogous to the phase diagram known from thermodynamics) is presented. Moreover the impact of the regulatory front end on the core toggle switch is investigated and it is found, that the resulting behavior is analogous to a mean-field Ising system in a magnetic field, where the read and input signals correspond to temperature and magnetic field, respectively. It is also shown, that the circuit can be adapted to varying amplitudes of

the input and the read signal by mutations in the binding sites of the regulatory proteins.

The dynamical response of the circuit with respect to time-dependent read and input signals shows that the desired ability to read and store signals on rapid timescales (around 30 min) is given. Stochastic simulations with Gillespie's algorithm reveal also, that even though there are stochastically induced transitions between the two 'stable' states, the resulting *memory loss* has a half-life of at least 40 hours. Additional to this, other noise induced errors are identified during the processes of "reading" the input signals. It is found, that these errors are an intrinsic property of the proposed design and can be minimized by increasing the duration of the read signal.

The structure of the thesis is as follows. In the first chapter a brief historical introduction to molecular biology is presented and the motivation for the current work is given. The second chapter provides a review of the present quantitative approaches to gene regulatory systems, gives an introduction to gene regulatory networks and to the most important mathematical modeling techniques. In addition, also the sources and descriptions of biochemical fluctuations are discussed. In the third chapter these concepts are applied to the 'conditional memory circuit' and its behavior is investigated from a deterministic and stochastic point of view. The last chapter discusses the results and gives an outlook on possible experimental applications of the presented work.

Contents

1 Motivation **1**

2 Gene regulation and genetic networks **3**
2.1 The central dogma of molecular biology 3
2.2 Quantitative description of gene regulation 4
2.2.1 Statistics of the bacterium *Escherichia coli* 5
2.2.2 Transcription and its control by regulated recruitment 6
2.2.3 Combinatorial transcription logic and cooperativity 13
2.2.4 Translation and turnover 15
2.3 Genetic networks 19
2.3.1 Synthetic biology and designer networks 20
2.3.2 Bistable systems for epigenetic memory 21
2.4 Modeling techniques 26
2.4.1 Chemical master equation 26
2.4.2 Gillespie's stochastic simulation algorithm 28
2.4.3 Reaction rate equations 29
2.4.4 Stochastic differential equations 30
2.5 Noise in gene expression 34
2.5.1 Intrinsic and extrinsic noise 34
2.5.2 Fluctuation-Dissipation Theorem 36
2.5.3 Suppression and exploitation of noise 39

3 A conditional memory circuit **41**
3.1 Illustration of the circuit's function 42
3.1.1 Conditional vs. non-conditional memory 42
3.1.2 Working principle 44
3.2 Quantitative model 46
3.2.1 Chemical reaction scheme 46
3.2.2 Deterministic and stochastic model of the full system 49
3.2.3 Parameter choice 49
3.3 Steady state characteristics 53
3.3.1 State diagram 53
3.3.2 Regulatory front end 55

3.3.3 Switchable hysteresis . . . 57
3.4 Dynamical behavior . . . 60
3.4.1 Proof of principle . . . 60
3.4.2 Switching timescale . . . 62
3.4.3 Stochastically driven memory loss . . . 66
3.4.4 Noise induced toggle errors . . . 73
3.5 Adaptation of the circuit's sensitivity to the environment . . . 76
3.6 Stochastic average vs. deterministic result . . . 78
3.7 Model reduction . . . 79
3.7.1 Adiabatic approximation and promoter activity function . . . 79
3.7.2 Rapid equilibrium assumptions . . . 81
3.7.3 Comparison of the full with the reduced model . . . 82

4 Discussion and outlook **83**

A Model details **87**
A.1 List of elementary chemical reactions . . . 87
A.2 Rate equations of the full model . . . 90

B A versatile implementation of the stochastic simulation algorithm **93**

Danksagung **106**

Chapter 1

Motivation

From the 17th century on the natural sciences separated into many diverging disciplines. The major fields of physics, chemistry, earth sciences and biology specialized themselves in countless sub-disciplines which typically embody their own terminology and nomenclature. The 20th century however, was marked by the reunion of many of these disciplines. In the 1930's the history of molecular biology begins with the convergence of previously distinct branches of biology, as for instance biochemistry, genetics, microbiology and virology. Interestingly many physicist made substantial contributions to this newly emerging field: Max Delbrücks (1906-1981) thoughts about the physical basis of life stimulated the writing of the highly influential little book "What is life?" [101] by Erwin Schrödinger (1887-1961). This in turn was an important guidance for the Nobel prize winners James D. Watson (*1928) and Francis Crick (1916-2004), who discovered the double helical structure of DNA by X-ray diffraction in 1953 [37].

In the past 60 years molecular biology has made tremendous progress, as not only the genetic code of humans was sequenced, but also the mechanisms by which cells regulate the activity of their genes and respond to external stimuli were discovered. With the fast-paced development of new experimental techniques as e.g. micro-arrays, flow-cytometry or single-cell fluorescence microscopy, modern biology is evolving from a *qualitative* and *descriptive* science to a *quantitative* and *predictive* one. At the same time as biology becomes quantitative the physicist gets the opportunity to extend his horizon to a completely new sphere: the animated matter.

The hope of the biophysicist is, that many concepts used in physics also apply for living systems. Cells are in principle genuine non-equilibrium chemical reaction systems with a confined volume defined by their cell membrane. Due to the small molecule numbers involved in these reactions, they are expected to be of highly stochastic nature. A gene for instance serves as the template for protein synthesis and is often only available in a single copy. Therefore on the microscopic scale the appropriate description seems to be the one by a probabilistic *master-equation*. However, on a larger scale also other concepts, like reaction rate equations for continuously changing concentrations of chemicals do apply. Thus also *nonlinear dynamics* can be applied to the life sciences: the mathematics of a nonlinear LR-circuit (van der Pol oscillator) is for example very similar to the one of

a biological oscillator controlling the day-night cycle of cyanobacteria or humans [53], since both result in limit-cycle behavior. As we will see in Chapter 2 also concepts from equilibrium thermodynamics find their application.

Whether the ultimate goal of a mathematically based comprehension of life will be attained, remains to be seen, but there are already numerous examples where the fusion of biology, physics and informatics has fostered exceptional results. The interdisciplinary cooperation among these fields has culminated in the foundation of a new discipline called *Systems Biology* [71], with the emphasis on a system-wide view on cells and organisms.

Of particular interest to Systems Biology are the *genetic networks* with their intricately connected regulatory structures (Section 2.3). Roughly speaking, they control which genes in a cell are ON and which are not, i.e. which genes are translated into proteins and which not. Thus by altering the *gene expression pattern* the intracellular composition of proteins can be modified. If environmental signals are coupled via *signal transduction pathways* to the genetic network, the cell can respond to varying external conditions by producing appropriate sets of proteins. The understanding of the *emergent* phenomena encoded in these networks might hence be seen as one of the major tasks in understanding life. Although a lot of progress was made in this direction in the past two decades, this field is still in its infancies and the comprehension of the entire regulatory structure is far beyond reach. In contrast, small *network motifs* consisting only of a few interacting genes are amenable to both theoretical and experimental characterization, as will be illustrated in Section 2.3.1.

The aim of this work is to theoretically predict a minimal network motif with the ability to store transient signals in form of protein concentrations. Such an implementation of *memory* in its simplest form provides a basis for the comprehension of *epigenetic memory* in large-scale genetic networks. By definition, *"epigenetics is the study of epigenetic inheritance, a set of reversible heritable changes in gene function (...) that occur without a change in DNA sequence"* (Wikipedia). In Section 2.3.2 we will go into some mechanisms of epigenetic memory.

With this study the first attempt to introduce *sequential* [68] rather than *combinatorial* logic [28] into gene regulatory circuitry is made. By sequential logic one means, that the response of the network not only depends on its present input signals, but also on the *history* of these signals. It hence provides the natural next step in our understanding of cells as "molecular computing machines" [28].

Chapter 2

Gene regulation and genetic networks

This chapter shall give a brief overview of our current understanding of biological cells as non-equilibrium chemical reaction systems. The first two sections introduce the basic biochemical processes necessary for protein synthesis and a quantitative framework for their description is presented. It is shown, that the mechanisms by which certain proteins regulate the synthesis of other proteins are crucial for the emergence of complex genetic networks. These are discussed in Section 2.3 and two prototypic networks are scrutinized in detail. Section 2.4 reviews the most important deterministic and stochastic modeling techniques, ranging from the chemical master equation and its simulation to reaction rate and Langevin equations. The last section finally discusses the origins and descriptions of biochemical noise associated with gene regulatory circuits.

2.1 The central dogma of molecular biology

With the discovery of the structure of DNA by Watson and Crick in the early 50's [37] the foundations of modern molecular biology were set. They could show that DNA is a double-helical polymer, that encodes the information necessary for protein synthesis in the sequence of its constituting bases. Triplets made up of the four existing bases A (adenine), G (guanine), C (cytosine) and T (thymine) code for one of the 20 amino acids. The *codon* GCA for instance, codes for the amino acid *alanine*. A chain of several hundred to thousand amino acids eventually folds up to form of a protein. The sequence of bases on DNA required for the synthesis of a protein is called a *gene*.

In Fig. 2.1 the schematic processes of the *central dogma of molecular biology* as formulated by Francis Crick [36] is shown. It is a statement about the flow of the genetic information: the information encoded on the DNA is replicated in a robust manner and thereby it is passed on to the daughter cells. From DNA a template molecule, called messenger RNA (mRNA) is *transcribed* and this single-stranded molecule contains a sequence of bases complementary to the one on DNA. In the majority of the cases the information flow is strictly from DNA to RNA and not back. The mRNA itself acts as a blueprint for protein synthesis, and many proteins are *translated* from one copy of mRNA.

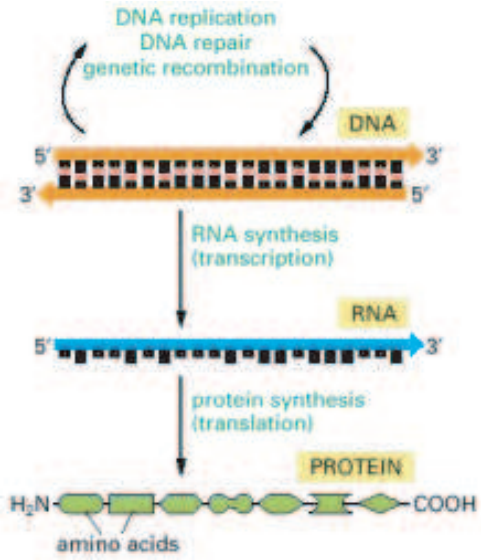


Figure 2.1: The central dogma of molecular biology is a statement about the flow of the genetic information. The figure was taken from [8].

Although the flow of genetic information is a "one-way-trip", this does not mean that it is an *unregulated* process. Indeed quite the contrary is the case, since every single step on the way from the gene to the protein is highly controlled: proteins regulate transcription and the lifetime of other proteins, small RNAs and RNA-binding proteins control translational initiation and the lifetime of mRNA, and so on. This results in an intricate feedback and feed-forward structure of the regulatory players, and gives rise to the emergence of complex systems behavior. Some of the regulation mechanisms and their quantitative description are introduced in the remainder of this chapter.

2.2 Quantitative description of gene regulation

In this section a quantitative view on gene regulation in bacterial cells (*prokaryotes*) will be developed. In contrast to *eukaryotic* cells (the cells of higher organisms as plants, mammals, etc.), prokaryotes lack a cell nucleus and the principles of gene regulation are much better understood. First, it is important to get aware of the size and the dimensions of the object of interest. This is in the present case the bacterium *E. coli* (see Fig. 2.2), which is one of the best-studied organisms of microbiology and serves in many cases as a paradigmatic example. Then we take a closer look at the previously mentioned processes of mRNA and protein production, i.e. transcription and translation, describe the biological details and show the most important mathematical descriptions of these processes. In addition to the mechanisms of mRNA and protein production, it is very important for living cells to degrade and recycle these substances once they are not needed any more. This property is crucial, since it turns them into highly dissipative *open* chemical reaction systems. We

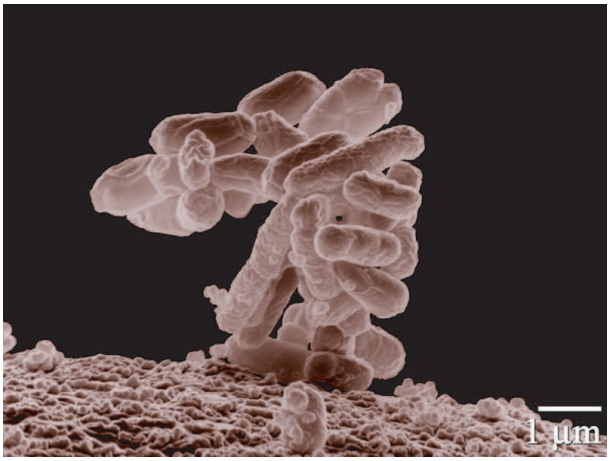


Figure 2.2: Low-temperature electron micrograph of a cluster of *Escherichia coli* bacteria, which lives in the lower intestines of mammals and is involved in the proper digestion of food. The photo was taken from [1].

will briefly introduce some of the mechanisms of active and passive degradation. Last, the mechanisms and implications of cooperative interactions will be discussed.

2.2.1 Statistics of the bacterium *Escherichia coli*

In order to develop a intuition for the quantitative description of gene regulation in bacterial cells it is instructive to have a closer look at the typical molecule numbers and the length scales found in these organisms. In Fig. 2.2 a colony of *E. coli* cells is shown. They are cylindrical cells with an average length of $1 - 2\,\mu m$, a diameter of $0.5\,\mu m$ and a resulting volume of about $10^{-15}\,\mathrm{L}$. Based on this estimation one molecule per cell corresponds to a concentration of $10^{-9}\,\mathrm{M} = 1\,\mathrm{nM}$ (one nano-mole). This is the canonical concentration unit for bacterial cells, since the typical molecule numbers vary from 1 to about 10^4. Table 2.1 gives an overview over the most important numbers and dimensions of *E. coli*'s constituents. One surprising example is the following: the intracellular pH level of *E. coli* is neutral (pH 7), which means that there is a concentration of $10^{-7}\,\mathrm{M} = 100\,\mathrm{nM}$ of H^+ ions in the cell. Hence there are only 100 soluted H^+ ions in each bacterium. These low numbers make it very evident, that the statistical fluctuations which are roughly proportional to the square root of the molecule number play a crucial role in the life of the cells. One can imagine that there have evolved sophisticated mechanisms of reducing these fluctuations if they threaten the survival of the bacteria. On the other hand one could also think of the possibility, that evolution made an advantage of these huge relative fluctuations and even amplified them in order to obtain a larger *population heterogeneity* (we will get back to that point in Section 2.3).

Item	Property
DNA	
size	5×10^6 base pairs
length	1.55 mm
number of open reading frames	4441
number of protein coding genes	4252
number of genes coding for transcriptional regulatory proteins	247
mRNA	
size	1100 bases
length	370 nm
total copy number	4000/cell
Proteins	
size	360 amino acids
diameter	5 nm
total copy number	3.6×10^6/cell
protein oligomerization state	4/complex
Housekeeping machinery	
number of ribosomes	18000/cell
diameter of a ribosome	20 nm
number of RNA polymerases	1000/cell
diameter of RNA polymerase	15 nm

Table 2.1: Average sizes and numbers of *E. coli*'s molecular constituents. Source: [2]

2.2.2 Transcription and its control by regulated recruitment

Transcription initiation, elongation and termination. As stated in Section 2.1 a copy of the genetic code in form of an mRNA molecule is synthesized in order to serve as a blueprint for the assembly of the respective protein. This process of transcription is carried out by a sophisticated molecular machine called RNA polymerase (RNAP), which forms together with a σ-factor a transcriptionally active *holoenzyme*, see Fig. 2.3 (a). The holoenzyme binds to a specific sequence of DNA, called the *promoter region*. It is located in front of the respective gene and is responsible for the initiation of transcription: Once the holoenzyme is bound to the promoter region, it separates the double stranded DNA locally into two single strands, releases the σ-factor and starts the assembly of the mRNA strand by complementary base pairing as it proceeds along the gene, see Fig. 2.3 (b). This *elongation* process continues until a terminator sequence is reached and the mRNA strand is released.

Physics of protein-DNA interaction

The mechanisms by which the holoenzyme recognizes the promoter region are general for a large class of DNA-binding proteins, i.e. also for the class of proteins named transcription factors, which will be discussed further below. In order to understand the physical interactions between proteins and DNA it is elucidating to have a glance at their physicochemical organization. In the lower part of Fig. 2.4 the three-dimensional structure of

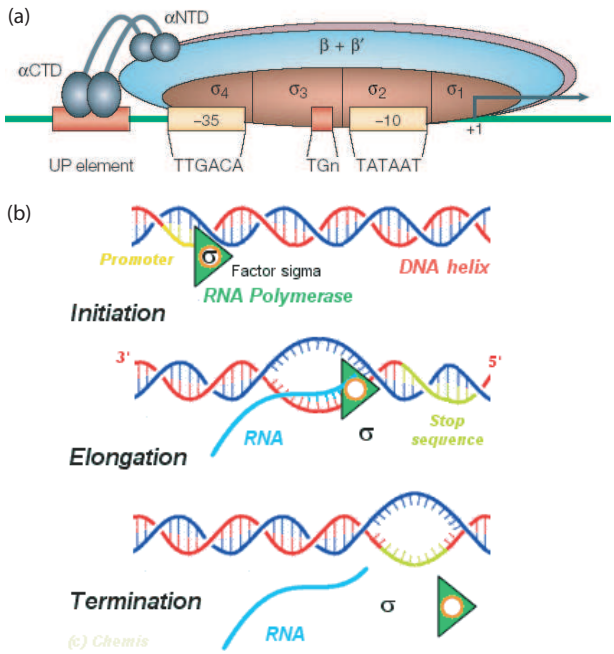


Figure 2.3: (a) The holoenzyme binds with its various sub-domains to distinct regions in the promoter region in order to initiate transcription. In (b) the processes of transcriptional initiation, elongation and translation are depicted schematically.(a) was taken from [26] and (b) from [3]

DNA is depicted. It consists of two strands of bases and the four existing bases A, G, C and T link the individual strands via hydrogen bond formation, but only the contacts A-T and G-C are energetically favorable. In addition, the individual base-pairs attract each other by hydrophobic forces. Together with the flexible sugar-phosphate backbone, which connects the bases along the axis of DNA, this leads to the double helical structure. Note that the hydrogen bonds between the bases are not arranged in a perfect angle of 180° and therefore double-helical DNA develops two grooves of different size: the major and the minor groove. This geometry has consequences for the binding of proteins to DNA, as for instance in the case of the 434 repressor protein (Fig. 2.4, upper part): The amino-acids of the two α-helices of 434 repressor (depicted in green) serve as "reading heads" and contact the bases in the major groove of the DNA. Thereby hydrogen bonds between proteins and DNA are formed, which is the physical basis of *specific* protein-DNA interaction. By specific it is meant, that the binding (free) energy of the protein to the DNA depends on the *sequence of base-pairs* on the DNA and a stable complex is only formed with the target sequence. This is due to the fact, that certain amino-acids "recognize" certain base-pairs on the DNA. For the Cro-repressor for instance, the amino acid glutamine recognizes an AT pair in the major groove of DNA while asparagine recognizes a GC pair in the minor groove [27]. This is called the *direct read-out* mechanism.

There exists a "best binder" sequence for every DNA-binding protein [47]. For the 434 repressor for instance the best binder sequence is depicted in Table. 2.2. The central bar signals the center of symmetry of the operator sequence and it indicates, that the 434

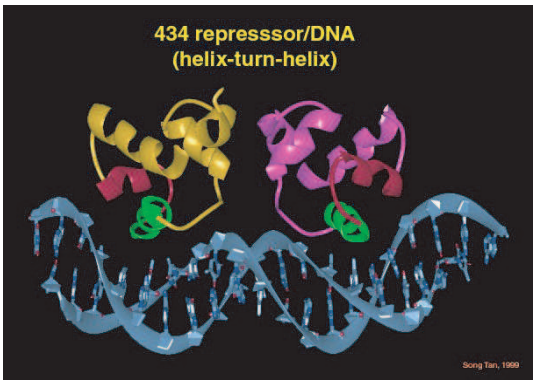


Figure 2.4: Schematic illustration of the dimeric 434 repressor (*top*) bound to the double stranded DNA (*bottom*). The two α-helices depicted in green mediate the contact with the bases in the major grooves of DNA. The illustration was taken from [4].

A	C	A	A	x	x	x	x	x	x	T	T	G	T
T	G	T	T	x	x	x	x	x	x	A	A	C	A

Table 2.2: Consensus operator sequence of the 434 repressor [61]. The x's denote non-conserved bases, which are not contacted by the repressor protein and yield therefore no contribution to the binding energy. The central bar represents the center of symmetry.

repressor binds as a *dimer* with anti-parallel orientation. Around the central region the x's stand for non-conserved bases. In terms of binding energy contributions, these base-pairs are somewhat arbitrary, since they are not contacted by the repressor and therefore don't confer any binding energy[1]. The exchange of a few base-pairs by non-ideal bases is penalized by a "mismatch energy" of $\varepsilon \approx 1 \ldots 3\,k_BT$ per base-pair and thereby the protein's binding affinity is reduced. Thus the affinity can be *tuned* or *programmed* over a large range through the choice of the nucleotide sequence and the resulting *equilibrium dissociation constant* K ranges from $K = 1\,\text{nM}$ to $1000\,\text{nM}$ for typical transcription factors, see [47].

Thermodynamic model of transcription

The various processes of transcriptional initiation, elongation and termination can be subsumed into an effective model, that only depends on a few parameters. It turns out, that under certain assumptions the *rate of transcription* is proportional to the equilibrium probability to find RNA polymerase (RNAP) on the promoter site [24]. One of these assumptions is for instance, that the open complex formation is the rate limiting step in the previously mentioned cascade of transcriptional initiation. Only then the assumption of

[1] Actually in some cases the non-contacted bases *do* matter and influence the binding energy. This is due to the fact, that the flexibility of DNA is determined by the sequence of the base-pairs [40] and the establishment of a TF-DNA contact often requires the *bending* of DNA.

an equilibrium between binding and unbinding to the promoter sequence is reasonable.

How can the probability of finding RNAP on the promoter be determined? In order to answer this question we will follow the reviews of Bintu *et al.* [24, 23]. There they employ the general framework of statistical mechanics to approach the problem: If one is able to compute the statistical weight Z_{bound} of all configurations, where RNAP is bound to the promoter site, the probability p_{bound} of finding RNAP on the promoter is simply the ratio of Z_{bound} and the total statistical weight Z_{tot}

$$p_{bound} = \frac{Z_{bound}}{Z_{tot}}. \tag{2.1}$$

These partition functions Z_i will be determined in the following. Under *in vivo* conditions, i.e. in living cells, most of the RNA polymerase molecules will be bound non-specifically to the DNA and only a negligible amount is free in solution [24]. There are $N_{NS} \approx 10^7$ possible non-specific binding sites on the DNA and the difference in binding energy $\Delta\varepsilon_{pd}$ between the non-specific and specific binding of RNAP to DNA is given by

$$\Delta\varepsilon_{pd} = k_B T \ln\left(\frac{K_{pd}^S}{K_{pd}^{NS}}\right). \tag{2.2}$$

Here the K's are the equilibrium dissociation constants for specific and non-specific binding taken from *in vitro* measurements of the lac promoter ($\Delta\varepsilon_{pd} \approx -3k_BT$) and the phage T7 promoter ($\Delta\varepsilon_{pd} \approx -8k_BT$).

With this notation the statistical weight $Z(P)$ of finding the promoter *unoccupied* is just the number of possibilities to distribute P RNAP molecules on N_{NS} non-specific sites times P Boltzmann factors of nonspecific binding:

$$Z(P) = \frac{N_{NS}!}{P!(N_{NS}-P)!} \times e^{-P\varepsilon_{pd}^{NS}/k_BT}. \tag{2.3}$$

Thus, the weight of the bound state is given by the weight of $P-1$ polymerases bound to nonspecific sites and one bound to the promoter

$$Z_{bound} = Z(P-1)e^{-\varepsilon_{pd}^S/k_BT}, \tag{2.4}$$

where ε_{pd}^S is the specific binding energy of RNAP binding to the promoter. The *total* statistical weight Z_{tot} is the sum of the two partition sums in Eqs. (2.3)and (2.4) and ultimately p_{bound} takes the simple form

$$p_{bound} = \frac{1}{1 + \frac{N_{NS}}{P} e^{-\Delta\varepsilon_{pd}/k_BT}}. \tag{2.5}$$

If we assume 1000 polymerases P and use the above mentioned values of N_{NS} and $\Delta\varepsilon_{pd}$ we obtain for the T7 promoter an equilibrium occupancy of 0.23 and for the lac promoter only 0.002. How can the lac promoter nevertheless yield an mRNA output comparable to the strong promoter in the T7 phage? This question will be addressed in the next section.

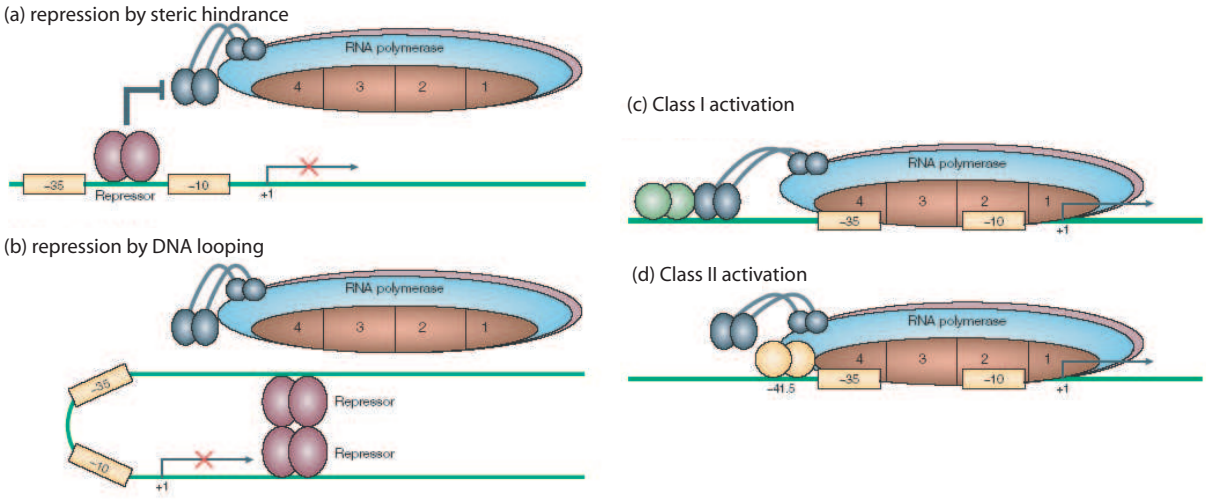


Figure 2.5: The most prominent mechanisms of bacterial gene regulation. For details see text. The figures were taken from [26].

Transcriptional control by regulated recruitment of RNAP.

The genome of *E. coli* consists of approximately 4500 genes. Without any regulatory mechanisms all genes would be transcribed simultaneously with a probability proportional to the affinity of RNAP to their associated promoter site. Clearly, the living cell needs some means to orchestrate these transcriptional events in time and (in some cases) also in space. This is necessary in order to respond to environmental stimuli "intelligently" by the production of specific sets of proteins. Expressing different (disjunct) sets at the same time would have devastating consequences for the cell. If for instance the apoptosis pathway of a eukaryotic cell would not be regulated carefully, the cell could kill itself at any time.

The orchestration of the different genes is mediated by a subclass of proteins called transcription factors (TFs). These DNA-binding proteins have no other purpose than regulating the transcription of other genes by a simple principle: they either *increase* or *decrease* the affinity of RNAP to bind to the promoter site of the respective target gene. This principle is called *regulated recruitment* of RNA polymerase [93, 91] and the TFs are named *activators* or *repressors*, respectively. In Fig. 2.5 the most recurrent regulatory processes are summarized. Before they are discussed in detail, we highlight the thermodynamical implications of regulated recruitment for the probability to find RNAP bound to the promoter.

Regulated recruitment. If both TF and RNAP are bound adjacently on DNA, the interaction of a TF with RNAP involves an additional contribution ε_{tp} to the free energy. In the case of an attractive interaction (activator), mediated by hydrogen bonds or hydrophobic forces, we have $\varepsilon_{tp} < 0$. The binding or *operator* site of the activator may not overlap with the promoter, because otherwise the TF would sterically inhibit the binding

of RNAP. This is exactly the case for many repressors, i.e. their operator sites overlap at least partially with the promoter. If the repressor and RNAP mutually exclude each other, the interaction energy becomes for the repressor with RNAP becomes infinitely large ($\varepsilon_{tp} \to +\infty$). With the additional configurations of having RNAP and TFs bound and unbound one can again find expressions for the different partition functions entering Eq. (2.1). Ultimately one can show [24] that the probability to find RNAP on the promoter site becomes

$$p_{bound}(TF) = \frac{1}{1 + \frac{N_{NS}}{P\, F_{reg}(TF)} e^{-\Delta\varepsilon_{pd}/k_B T}} \,. \tag{2.6}$$

The only difference to Eq. (2.5) is the introduced *regulation factor* $F_{reg}(TF)$, which is a function of the TF abundance. It can be seen as the effective increase ($F_{reg} > 1$, activator) or decrease ($F_{reg} < 1$, repressor) of the number of RNAP molecules available for binding the promoter. The precise dependence of $F_{reg}(TF)$ on the TF abundance is specific for the respective regulation mechanism. We discuss the different mechanisms depicted in Fig. 2.5 and their regulation factors one by one in the following.

Transcriptional repression by competition. The simplest and most frequent way to *repress* transcription is by *competitive binding* (Fig. 2.5 (a)): if the binding site of the TF overlaps the promoter region, the bound TF *sterically inhibits* the RNAP to locate the promoter site ($\varepsilon_{tp} \to +\infty$). The regulation factor F_{reg} is in this case [24] given by

$$F_{reg}(R) = \frac{1}{1 + \frac{R}{N_{NS}} e^{-\Delta\varepsilon_{rd}/k_B T}}. \tag{2.7}$$

Here R is the number of repressor molecules and $\Delta\varepsilon_{rd}$ the difference between specific and nonspecific binding free energy of the repressor bound to DNA. Inserting Eq. (2.7) in Eq. (2.6) yields with the *in vivo* equilibrium dissociation constant of repressor-operator interaction $K_{O_R} := N_{NS} \exp(\Delta\varepsilon_{rd}/k_B T)$ and with the dimensionless factor
$\tilde{P} := \frac{P}{N_{NS}} \exp(-\Delta\varepsilon_{pd}/k_B T)$

$$p_{bound}(R) = \frac{1}{1 + \tilde{P}^{-1}\left(1 + \frac{R}{K_{O_R}}\right)}. \tag{2.8}$$

With $N_{NS} = 10^7$ nonspecific binding sites, $P = 10^3$ RNAP molecules and $\Delta\varepsilon_{pd} \approx -3 k_B T$ for the lac promoter, it is guaranteed that $\tilde{P} << 1$ (weak promoter limit) and thus

$$p_{bound}(R) \approx \tilde{P}\, \frac{1}{1 + \frac{R}{K_{O_R}}}. \tag{2.9}$$

This is the traditional Hill- form of the *promoter activity function* (PAF) with an Hill-exponent[2] $n = -1$. We note that the abundance of the RNAP molecules enters this formula only as a pre-factor and is usually assumed to be constant. Thus, in the later sections we will refer to the PAF only as the dimensionless second factor, call it p and subsume $\tilde{P}$ into the other multiplicative pre-factors.

[2]The negative sign is because $H(x) := \frac{(x/K)^n}{1+(x/K)^n}$.

Repression by DNA looping. Another interesting example of repressive transcriptional control is DNA looping. In some cases there is a second operator site located a few hundred base pairs upstream of the actual promoter region [113], see Fig. 2.5 (b). In the case of the Lac-repressor LacI one dimer binds to this upstream operator site and one to the second operator site located near the promoter. Since the distance between these operators is larger than the persistence length of the DNA, it is likely that the DNA bends and forms a loop. This loop is then stabilized by the heads of the LacI molecules, that glue together to form a stable tetramer. This mechanism confers an extremely robust way of repressing transcription for at least two reasons: first, the loop itself prevents RNAP to approach the promoter and second, once one of the two TFs unbinds, the possibility of reconnecting to the proper operator site is facilitated as can be seen in the following. The promoter activity function similar to Eq. (2.8) can be obtained via a statistical thermodynamics approach [98] and leads in the limit where the upstream operator is very strong to

$$p_{bound}(R) \sim \frac{1}{1 + \left(e^{-\Delta\epsilon_{loop}/k_BT} + R\right)/K_{O_R}}, \tag{2.10}$$

where $\Delta\epsilon_{loop}$ is the free energy difference for looping and K_{O_R} the *in vivo* dissociation constant for the repressor R binding to the operator site. Here the contribution of the loop effectively increases the concentration of R and thus repression is enhanced (for details see [98] or [113]).

Transcriptional activation. Although most of *E. coli*'s promoters are strong promoters which are restrained by the action of repressors, there are also several examples of *weak* promoters which are *activated* by transcription factors. The two main classes of activators are shown in Fig. 2.5 (c),(d) and they differ only in the location of their binding sites. The regulation factor is in both cases given by

$$F_{reg} = \frac{1 + \frac{A}{N_{NS}} e^{-\Delta\varepsilon_{ad}/k_BT} e^{-\varepsilon_{ap}/k_BT}}{1 + \frac{A}{N_{NS}} e^{-\Delta\varepsilon_{ad}/k_BT}}, \tag{2.11}$$

where $\varepsilon_{ap} < 0$ is the free energy of the interaction between activator A and RNAP. With the substitutions analogous to above and $f := e^{-\varepsilon_{ap}/k_BT}$ we get

$$p_{bound} = \frac{1 + \frac{A}{K_{O_A}} f}{1 + \frac{A}{K_{O_A}} f + \tilde{P}^{-1}\left(1 + \frac{A}{K_{O_A}}\right)} \tag{2.12}$$

$$\overset{\tilde{P}<<1}{=} \tilde{P}\, f\, \frac{A/K_{O_A}}{1 + A/K_{O_A}}. \tag{2.13}$$

Eq. (2.13) is again the 'weak promoter limit' and resembles the commonly used Hill-function for activation with an Hill-exponent of $n = 1$. In Section 2.2.3 we will see, that *cooperative binding* of multiple TFs to adjacent operator sites leads to $|n| > 1$.

2.2.3 Combinatorial transcription logic and cooperativity

In Section 2.2.2 the regulation of gene expression by a single transcription factor (TF) via regulated recruitment of RNA polymerase (RNAP) was introduced. The natural extension of this concept is the recruitment of RNAP by *multiple* TF's acting synergistically or competitively. As we will see in the subsequent section, these synergistic and cooperative effects are the underlying reason for the emergence of complex systems behavior on the network level.

Dimerization. Dimerization is one of the most common synergistic effects in gene regulation. It is known that many bacterial TFs bind DNA only in their dimeric form, as e.g. λ-repressor, Cro repressor [92] or the activator AraC [99]. There are plenty of possible reasons for this , but one of the most important functional implications is the following: If we consider for instance the dimerization reaction of a repressor $R+R \underset{k^-}{\overset{k^+}{\rightleftharpoons}} R_2$ in equilibrium, the 'law of mass action' applies and we have $K_R = R^2/R_2$. Here and in the remainder of the thesis the squared brackets, which usually denote the concentrations are omitted. This relation implies, that at low total protein concentrations mainly momoners R, and at high total concentrations mainly dimers R_2 are present. Thus, additional to the Hill-form of the promoter activity function a further *nonlinearity* is imposed into the system's response. In the limit of 'infinite cooperativity', i.e. where *only* the dimers can bind DNA, we have to replace the repressor in Eq. (2.9) by the dimer R_2. The promoter activity function is now given by

$$p_{bound} = \tilde{P}\,(1 + R_2/K_{O_R})^{-1}\,. \tag{2.14}$$

However, since actually the total protein abundance $R_{tot} = R + 2\,R_2$ of the repressor is the variable under the control of other cellular processes, one has to express R_2 in Eq. (2.14) in terms R_{tot}. This is again achieved via the law of mass action and we obtain

$$p_{bound}(R_tot) = \tilde{P}\left(1 + \frac{R_{tot} + K_R/4 - \sqrt{K_R^2 + 8K_R R_{tot}}/4}{2\,K_{O_R}}\right)^{-1}. \tag{2.15}$$

For many bacterial TFs the equilibrium dimerization constant is very low ($K_R = 1 \ldots 10\text{nM}$) and thus the equilibrium is strongly biased towards the dimers. The result is that $R_2 \approx R_{tot}/2$ and thus a Hill function with exponent $n = -1$ is the result:

$$p_{bound}(R_{tot}) \approx \tilde{P}\left(1 + \frac{R_{tot}}{2\,K_{O_R}}\right)^{-1}. \tag{2.16}$$

For large dimerization constants K_R on the other hand, one can expand the root in Eq. (2.15) in $x = 8R_{tot}/K_R$ and one obtains $R_2 \approx 2R_{tot}^2/K_R$. Thus, only in the limit of *weak dimerization* one obtains a Hill function with exponent $n = -2$ and $K = \sqrt{K_R K_{O_R}/2}$:

$$p_{bound}(R_{tot}) \approx \tilde{P}\frac{1}{1 + (R_{tot}/K)^2}. \tag{2.17}$$

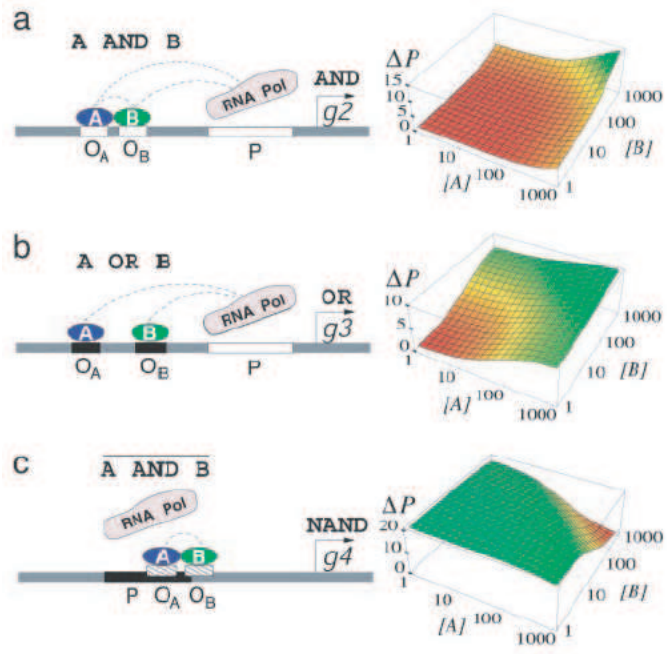


Figure 2.6: *Logic gates* are the result of combinatorial transcription control by multiple TFs. The dashed lines depict adhesive interactions between the molecules and the black and white boxes correspond to strong and weak binding-/promoter sites, respectively. The surface-plots on the right depict the promoter activity for a given combination of TFs in units of the minimal activity: $\Delta P = p([A],[B])/p_{min}$. The figure was taken from [28].

The *sigmoidal* shape of this function (for $|n| > 1$) is essential for the emergence of complex systems behavior, as will be revealed in Section 2.3.2. Anyhow, since the limit of large dimerization constants is not met in most of the cases, nature has invented more efficient cooperative mechanism for gene regulation and the most important ones will be introduced now.

Synergistic TF-DNA binding. Analogous to the adhesive or repulse interaction between TFs and RNAP molecules, also the (dimeric) TFs interact with each other. This was already implicitly assumed in the case of DNA-looping: The TF bound to the upstream operator site increased to effective concentration of repressor at the second operator site. Another example is the synergistic binding of λ-repressor CI to the adjacent operator sites O_{R1} and O_{R2} [91]. As soon as the operator O_{R1} is occupied by CI_2, it *facilitates* the occupation of the second operator by another CI_2 dimer. If ε_{tt} is the interaction free energy for the TF-TF contact, the regulation factor for two interacting repressors becomes for this specific example [24]

$$F_{reg}(R_1, R_2) = \left(1 + \frac{R_1}{K_{O_{R_1}}} + \frac{R_2}{K_{O_{R_2}}} + \frac{R_1}{K_{O_{R_1}}}\frac{R_2}{K_{O_{R_2}}}\omega\right)^{-1}, \tag{2.18}$$

with $\omega := \exp(-\varepsilon_{tt}/k_BT)$ and R_1 and R_2 being the two repressor concentrations (which are identical in the cited example). In the general framework of statistical mechanics it is possible to derive the regulation factor for *any* combination of activating and repressing TFs (see [24] for a comprehensive listing).

Combinatorial transcription logic. If we regard the promoter activity as the output signal and the TF abundances as the input signals, the process of *combinatorial* transcription regulation can be seen as the *computation* of the input signals with the result of an *analog* output signal [28, 23]. The above mentioned example for instance, corresponds - for weak operator affinities (large $K_{O_{R_1}}$, $K_{O_{R_2}}$) and strong cooperativity (large ω) - to the logic computation of an **NAND**-gate (see Fig. 2.6 (c)): while a single repressor-species is not sufficient to inhibit transcription substantially, only the presence of *both* factors R_1 and R_2 at the same time leads to a synergistic down-regulation of the promoter activity. In Fig. 2.6 (a) and (b) exemplarily the realizations of an **AND** and an **OR**-gate are depicted as well. Although a variety of possibilities exists to combine different input signals to distinct *logic gates*, there are nevertheless limitations due to the confined size of the promoter region. As we saw before, the promoter region is approximately 40 bp long, and only a limited number of operator sites can be placed in its vicinity.

However, an effective strategy to prevent promoter overcrowding is to take advantage of DNA looping. Thereby it is possible to integrate a variety of different signals and construct a "molecular computing machine" [28]. The parallel or serial connection of multiple gates, where the output of a gate is again a TF and serves as one input for an other gate, is the basis for complex transcriptional networks, as will be discussed in Section 2.3.

2.2.4 Translation and turnover

In addition to the complex mechanisms of transcriptional regulation, also the subsequent steps on the way to the functional protein are carefully administered. This *post-transcriptional control* ranges from translational control by small regulatory RNAs [55], involving translational initiation, peptide elongation and translational termination, to the quality control of protein folding by the proteolytic machinery [67]. However, this work is based only on the cis-regulatory mechanisms as discussed in the previous sections. The post-translational steps will be considered simplistically as unregulated processes, as described in the following.

Translation and mRNA degradation. The first bases downstream of the transcriptional start site include a sequence called the *Shine-Dalgarno* sequence or *ribosome binding site* (RBS) [81]. Soon after RNAP has synthesized this part of the mRNA, the first ribosome binds to the RBS and initiates the translation of the message into a chain of amino acids that eventually folds up to the final protein. As the ribosome proceeds downstream, it clears the RBS, the next ribosome can bind and thus a chain of parallely translating ribosomes is the consequence, see Fig. 2.7 (a). In addition to the ribosome the RNA degrading

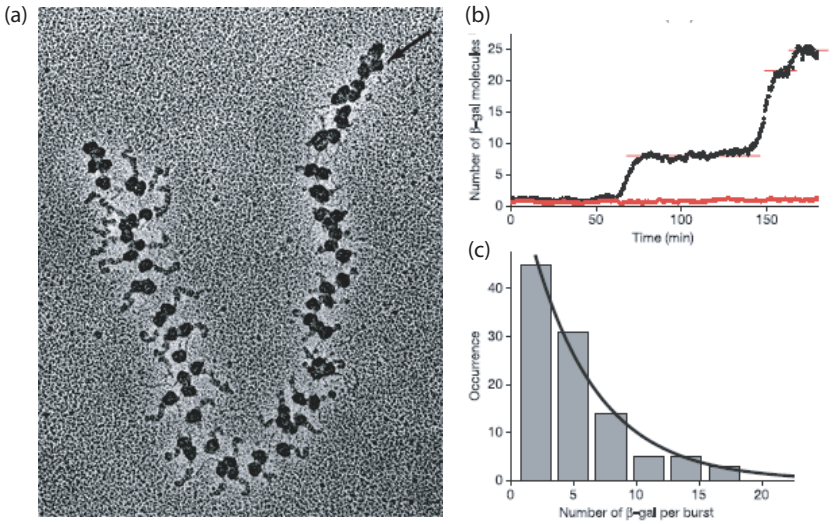


Figure 2.7: (a) Many ribosomes translate a strand of mRNA in parallel; the arrow indicates the transcriptional start site (upper right) and as the ribosomes progress along the mRNA one observes the elongated polypeptide chains (best visible at the left) (the image was taken from [5]). (b) shows the timecourse of a single protein measurement. It is clearly visible, that protein synthesis occurs at bursts. The distribution of burst sizes in (c) follows a geometric distribution (figures (b) and (c) were taken from [31])

RNase has its binding site in the vicinity of the RBS and therefore ribosomes and RNases compete for binding [81]. The number of proteins produced per mRNA can be calculated from a simple probabilistic model [81], if we assume independent trials of ribosome/RNase binding. Let p be the probability of a ribosome to bind the RBS and $1 - p$ the probability of the RNase to bind its recognition site. Then the probability of obtaining n proteins from one mRNA transcript follows a geometric distribution

$$P(n) = p^n (1 - p), \tag{2.19}$$

with mean $b = \frac{p}{1-p}$ and variance $\sigma = \frac{p}{(1-p)^2}$. The average number b of proteins produced per mRNA is usually > 1. Therefore the protein production is a bursty process, and b is called the *burst factor*.

Recent pioneering experiments on the *single molecule* level have supported this simple theory [31, 117], see Fig. 2.7 (b) and (c). They could show in *E. coli* that translation indeed follows a geometric distribution with a burst size of around 5 molecules per mRNA for the protein β-galactosidase. Other typical values for b range from 5 to 40 proteins per mRNA [106].

Dynamical model. The standard dynamical molecular model for translation [87] is a 'truncated' Poisson process, since degradation of mRNA terminates translation from the

respective mRNA:

$$m \xrightarrow{\nu_p} m + \text{protein} \tag{2.20}$$

$$m \xrightarrow{\lambda_m} \emptyset, \tag{2.21}$$

with translation rate ν_p and mRNA degradation rate λ_m. The burst factor is then simply given by $b = \nu_p/\lambda_m$. Due to the active degradation of mRNA via RNases, the resulting half-life of mRNA ($\tau_m = \ln 2/\lambda_m$) can be very short and is in these cases of the order of 1 to 5 minutes [21].

Protein dilution and degradation. Most of the bacterial proteins are stable and their concentrations are governed by the balance of synthesis and dilution. Their dilution rate is determined by the cell doubling time, which ranges in the exponential growth phase from 30 to 60 minutes. However, some selected proteins are targets of active proteolysis and their half-life is reduced to the order of a few minutes. What are the advantages for a cell to actively degrade its well folded and functional proteins, especially since *both* synthesis and active degradation are *energy consuming*?

As already mentioned, mRNA and protein synthesis are stochastic processes. Therefore it happens, that a protein burst accidentally takes place, although the gene is supposed to be OFF. If the respective protein plays a central regulatory role in the organism, it is advantageous to actively get rid of it in this case. Another very important advantage of active proteolysis is connected to the dynamical behavior of TF networks. In a later section we will see, that the time required to reach the steady state concentration of a TF is solely determined by its half-life (independent of whether the gene is turned from ON to OFF or *vice versa*). For this reason directed degradation of proteins via SsrA tags with half-lives of around 5 minutes [56, 67] speeds the response times of transcriptional networks [96].

Models of dilution and degradation. If we consider bacterial growth to be linear in time [13] the dilution with dilution rate λ_{dil} is a simple first order process and leads to a dilution flux f_{dil} of the proteins X

$$f_{dil}(X) = \lambda_{dil} X. \tag{2.22}$$

Although cell division involves the bisection of the *molecule numbers*, the *concentrations* in mother and daughter cells remain the same, since they are intensive variables (evidently also the volume is divided by 2 and thus the ratio of molecule number and volume remains constant). Thus the assumption of a continuous process for dilution is justified .

Active degradation on the other hand involves the enzymatic digestion of proteins by proteases. Since the proteases are in some cases even rate limiting, the degradation flux f_{deg} takes a Michaelis-Menten form [78]

$$f_{deg} = \tilde{\lambda}_{deg} \frac{\text{protease} \cdot X}{K_m + X}, \tag{2.23}$$

with a modified degradation rate $\tilde{\lambda}_{deg}$ and the Michaelis-Menten constant K_m. In many situations the proteases are not limiting ($K_m \gg X$) and active degradation can be approximated by a first order process [81, 13] as well

$$f_{deg}(X) \approx \lambda_{deg} X, \tag{2.24}$$

where $\lambda_{deg} := \frac{\tilde{\lambda}_{deg} \cdot \text{protease}}{K_m}$ is the degradation rate with dimension s^{-1}.

However, the saturating effect of the proteases seems to play a crucial role for the emergence of oscillations in biological clocks [76], as it introduces *cooperativity* into the system. We can consider this nonlinear degradation as an cooperative effect, since at high protein levels the proteases are saturated by some sacrificing proteins and they allow the rest of the population to 'survive'. Another nonlinear degradation effect is in many cases mediated by cooperative stability of multimeric protein complexes [29]. Often the protease recognition domains get hidden in the dimerization interfaces, thus conferring a much longer half-life of the complexes. The shift of the multimer equilibrium from the monomeric form at low total protein concentration to the multimeric form at high concentrations thus implies a concentration-dependence of the degradation rate. The consequences on the function of genetic circuits are similar to the ones discussed in the following section [29].

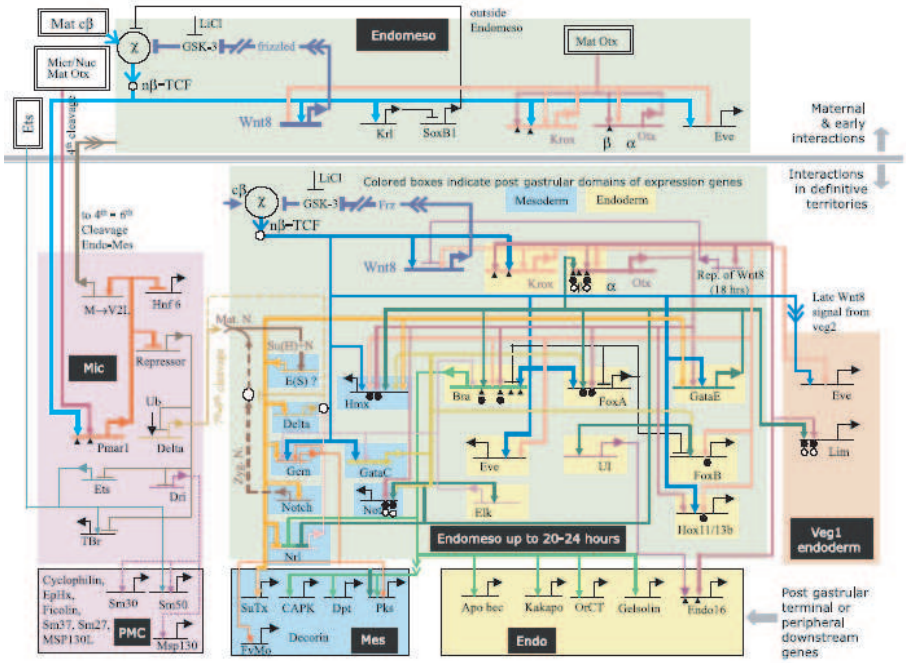


Figure 2.8: Gene regulatory network controlling the early development of the sea urchin embryo. The arrowheads represent positive regulation, whereas the arrow bars stand for repressive control of the respective gene. The figure was taken from [38].

2.3 Genetic networks

In the previous sections, we saw how the transcription of genes is regulated by transcription factors, which themselves are products of a certain gene. In this way, the activity of one gene is regulated indirectly by other genes, in combination with external stimuli. These interdependencies commonly form large regulatory networks of highly complex structure. The network of the developmental body plan of the sea urchin embryo [38] is shown in Fig. 2.8 and it reminds us strongly of an electronic circuit. Here the nodes correspond to genes and the positive or negative interactions between the nodes are mediated by TFs. The respective sign of the interaction is commonly abstracted by an arrow-head (activator) or an arrow-bar (repressor). In contrast to an electronic circuit that bears many nodes with a low degree of connectivity (one node - a transistor - is usually regulated by not more than one or two input signals), genetic networks have a low number of nodes that are highly connected. The gene Bra in the central part of Fig. 2.8 for instance, is regulated by five activators and one repressor. Evidently, the quantitative understanding of the signal integration at individual genes is crucial for the comprehension of the entire network. Without a detailed knowledge about that, it is not possible to predict what happens when e.g. an activator and a repressor are present at the same time.

Therefore a network diagram as depicted in Fig. 2.8 is not sufficient for a quantitative model. Especially the fact, that the network's qualitative response is highly dependent on

all involved biochemical parameters, which are in most of the cases unknown, poses severe problems. One approach to solve this problem is to separate the whole network into smaller sub-networks, called *network modules*. A module refers to a group of proteins (and genes) that work together to execute a distinct function [80]. The modules are often separated in time or space and have a low number of in- and output signals. The ultimate goal of this approach is to gain a systems-level understanding from the interplay of the individual modules. But before this may be done, it is crucial to characterize the individual modules in every detail.

In the recent years this *bottom-up* approach has made amazing advances, both theoretically and experimentally. Ultimately it culminated in the foundation of a new field, called *Synthetic Biology*.

2.3.1 Synthetic biology and designer networks

The goal of synthetic biology is to extend or modify the behavior of organisms and engineer them to perform new tasks [102, 11]. As we saw in the previous sections, there exist various ways of revising the naturally existing regulatory structures. The key feature that allows the experimentalist to comfortably modify biological function, is the great versatility and programmability of the genetic code [47]. The exchange of individual base pairs in the operator or promoter sequences permits the tuning of the protein binding affinities. Thereby the transcription rates (mutations in the promoter region) or sensitivities with respect to TF concentrations (mutations in the operator sites) can be adjusted. Similarly, the translation and degradation rates are affected by the sequence of the ribosome binding sites or the RNase recognition sites. In addition it is also practicable to replace entire operator sequences by the binding sites of other TFs. Thereby one is able to redirect the wires and create *synthetic transcription networks*. Other synthetic systems involve engineered RNA systems [63] or gene-metabolic circuits [43].

Such synthetic systems are of immense importance for the study of the *principles* of gene regulation. The limited number of network components gives rise to the hope, that the enormous complexity visible in Fig. 2.8, can be reduced. Some successful studies are summarized in the following. The fusion of multiple fluorescent *reporter proteins* to a synthetic construct including the tet-promoter P_{tet} allowed the determination of its promoter activity function [97]. By the step-wise engineering of a synthetic promoter it could be shown that the properties of regulatory subsystems can be used to predict the behavior of larger, more complex regulatory networks [57]. Small network modules were constructed and proved to exhibit bistable [46, 64, 15] or oscillatory behavior [41, 15, 43]. It was also shown, that the coupling of multiple oscillators leads to synchronization and confers an increased stability of the individual oscillators with respect to molecular noise [44]. Other work was dedicated to the engineering of bacterial colonies, that were able to create patterns of differently fluorescing bacteria [20]. This was possible by constructing 'band-detector' networks, reacting specifically to a narrow concentration range of a signaling molecule. By generating spatial gradients of these molecules bacterial bulls-eye patterns were created. In [77] light-receptors were coupled to a synthetic circuit, such that a colony of bacteria

could 'photograph' various light-patterns.

2.3.2 Bistable systems for epigenetic memory

In this section it is shown, that simple feedback mechanisms are able to confer interesting behavior of small genetic network motifs. If we consider the concentrations of the TFs as the state variables of a dynamical system, it will be shown that feedback may lead to *multiple steady states* in the phase space of TF concentrations. Of particular interest are the *bistable* systems, since their features are representative for the entire class of multistable circuits. The ability to exhibit two distinct states implies to some extend the 'digitalization' of circuit behavior: only LOW or HIGH concentration values for the respective TF-species are allowed and most importantly - these states are self-perpetuating. Therefore it is reasonable to regard these bistable circuits as digital *memory devices* in which the LOW and HIGH concentrations correspond to the logic states 0 and 1, respectively [33].

Epigenetics. The evolutionary advantage of the ability to switch between different types of behavior is demonstrated in an imposing manner in certain bacterial viruses as e.g. in temperate phages [110]. Upon infection of the host cell a *decision process* takes place: either the virus kills the cell and thereby replicates itself (lytic state) or it integrates its genome into the host chromosome and remains dormant (lysogentic state) [92]. Only if certain signals like e.g. cell death of the host arrive, the virus returns to the lytic state. On the microscopic scale this *switch* corresponds to a bistable system consisting basically of two mutually repressing TFs, the λ-repressor CI and the Cro-repressor. The lysogenic state corresponds to a state where CI is highly expressed and the transcription of Cro is inhibited. Only transient induction signals that promote the synthesis of Cro drive the switch to the lytic state in which Cro is ON and CI is OFF. A prototype of this mutually repressive network motif will be discussed later on.

The switch in phage λ has served as a paradigmatic system for gene regulation and can be seen as one of its 'hydrogen atoms'. The low number of involved TFs made it possible to study the basic regulatory mechanisms thoroughly and revealed many fundamental mechanistic details of gene regulation. But most importantly it shed light on how an organism with *identical* genomic material can produce *different phenotypes*, corresponding to distinct gene expression states. This phenomenon combined with with the ability to inherit the phenotype to the daughter cells is known as *epigenetic memory* [65]. In addition to stable gene expression states epigenetics involves also more durable forms of memory, as e.g. the chemical modification of chromatin [14], but this is beyond the scope of this work.

What are the minimal ingredients necessary for epigenetic memory? Many theoretical [34, 15, 115] and experimental studies [46, 64] have focused on this question and we will now give a brief overview over the most simple network modules that confer bistability.

Positive autoregulation

The most elementary circuit exhibiting bistability consists only of one gene that positively regulates its own production. The fairly simple cartoon is shown in Fig. 2.9 (a) and its dynamical behavior shall be scrutinized now. To demonstrate the main features of the system, it is in this case sufficient to consider only the time-evolution of the total protein concentration $\mathrm{A_{tot}}$. Under a number of simplifying assumptions[3] we can write the *reaction reaction rate equation* of the positive autoregulator as

$$\frac{dA_{tot}}{dt} = \alpha\, p(A_{tot}) - \lambda\, A_{tot}. \tag{2.25}$$

The first term represents the protein production flux and it is given by the maximal production rate α times the dimensionless promoter activity function. The second term is the degradation or dilution flux (compare Section. 2.2.4). In order to determine the steady state of the circuit, we have to specify the promoter activity function (PAF). If only the dimeric form of the TF is capable to bind the operator we obtain in the limit of large dimerization constants a Hill function with $n = 2$ (see Fig. 2.9, red graph): $p(A_{tot}) \approx A_{tot}^2/(\tilde{K}^2 + A_{tot}^2)$ with $\tilde{K} = \sqrt{K_{O_A} K_A/2}$. The steady state values A^*_{tot} of Eq. (2.25) are given by

$$A^*_{tot_1} = 0, \quad A^*_{tot_2} = \frac{1}{2}\Big(\frac{\alpha}{\lambda} + \sqrt{\Big(\frac{\alpha}{\lambda}\Big)^2 - 4\tilde{K}^2}\Big), \quad A^*_{tot_3} = \frac{1}{2}\Big(\frac{\alpha}{\lambda} - \sqrt{\Big(\frac{\alpha}{\lambda}\Big)^2 - 4\tilde{K}^2}\Big) \tag{2.26}$$

Linear stability analysis [109] yields, that for $\alpha/\lambda < 2\tilde{K}$ only $A^*_{tot_1}$ is stable (and the other two are imaginary). On the other hand, if the ratio α/λ is increased, at $\alpha/\lambda = 2\tilde{K}$ a *saddle-node bifurcation* additionally creates one stable and one unstable fixed point, see the bifurcation diagram in Fig. 2.9 (c). Thus, for $\alpha/\lambda > 2\tilde{K}$ the positive autoregulator exhibits *bistability* with two stable steady states $A^*_{tot_1}$ and $A^*_{tot_2}$.

From Fig. 2.9 (b) the fixed points can be determined graphically as well: the functions $p(A_{tot})$ and $f(A_{tot}) = \frac{\lambda}{\alpha} A_{tot}$ are plotted against A_{tot}. The intersection points correspond to the fixed points of Eq. (2.25) and their stability is determined by the ratio of the slopes of both curves. If the derivatives of f and g at the fixed point obey for instance $f'(A^*_{tot}) > p'(A^*_{tot})$, a small deviation from A^*_{tot} leads to a net flux back to the fixed point and thus it is a *stable fixed point*. For the red curve with $n = 2$ the sigmoidal shape of the function is responsible for three intersection points with the black curve $f(A_{tot})$, corresponding to the fixed points in Eq. (2.26). As the slope of $f(A_{tot})$ increases, e.g. due to a reduced promoter strength, the intersection points $A^*_{tot_2}$ and $A^*_{tot_3}$ approach each other until they finally annihilate at $\lambda/\alpha = 2\tilde{K}$.

For a Hill coefficient of $n = 1$ (see Fig. 2.9 (b), blue graph) on the other hand, we find maximally two intersection points. For $\alpha/\lambda > \tilde{K}$ only the A^*_{tot}-LOW state and for $\alpha/\lambda > \tilde{K}$ only the A^*_{tot}-HIGH state is stable. Thus the circuit is *monostable* for all parameters.

[3] Under the premises that mRNA production and decay are in equilibrium all the time and monomers as well as dimers are degraded with the same rate λ (no cooperative stability) Eq. (2.25) is valid. The general details of model reduction are discussed is Section 3.7

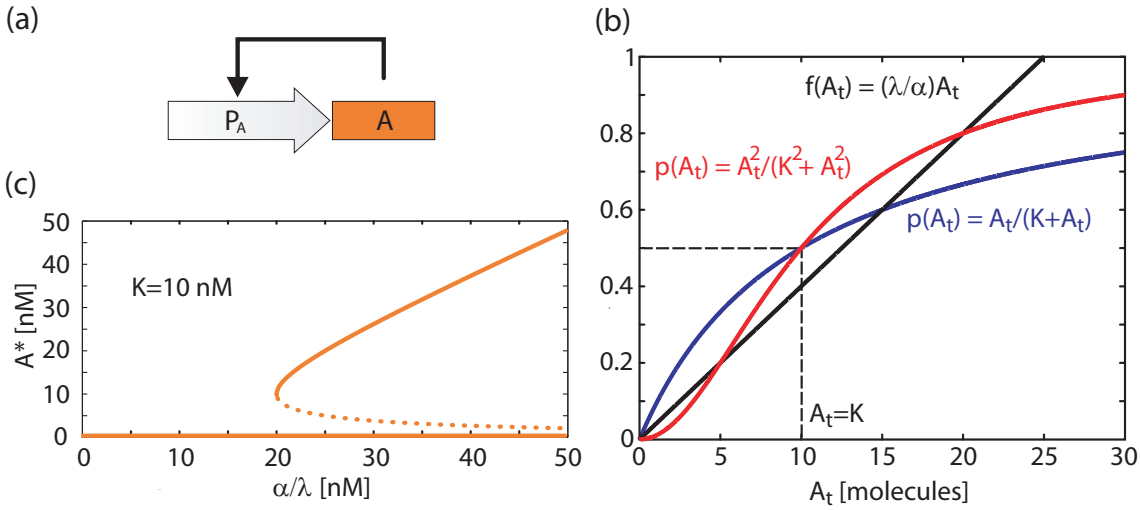


Figure 2.9: The positive autoregulatory network motif (a) exhibits bistability for parameter ratios $\alpha/\lambda > 2K$, as can be seen in the bifurcation diagram in (c) ($K = 10$ nM). The solid lines correspond to the stable fixed points, whereas the dotted one stands for the unstable one. With (b) the graphical solution of Eq. (2.25) can be determined, for details see text.

This important result can be generalized [34] and it turns out that for the emergence of bistability *always* some degree of cooperativity is required. The specific mechanisms are not of importance and can range from dimeric binding and non-additive effects from multiple operator sites [34] to nonlinear degradation [28].

Experimental validation of bistability in a positive autoregulatory module was given in [64]. In this work the authors studied a small synthetic autoregulator with a temperature-sensitive protein. By variations in temperature the degradation rate λ was altered and thereby one could drive the circuit from the monostable to the bistable regime. Another study focused on the positive feedback in the lactose utilization network [85] where the lactose uptake is autocatalytically enhanced. They could show, that this system exhibits *hysteresis*, an essential aspect of bistable systems - we will get back to this point later on. As a last example the investigation of the galactose-signalling network in [7] shall be mentioned. In this system the positive feedback leads to an extremely high stability of the HIGH and LOW states with almost irreversible behavior. Interestingly nature added a second, negative feedback loop that destabilizes this memory by allowing stochastically induced transitions between the two 'stable' states. It was speculated that this destabilized memory might amount to a selective advantage in environments with time-dependent nutrient supply. As theoretically shown in [75] the random switching between different *phenotypes* (or stable gene expression states) indeed increases the population fitness in fluctuating environments. Thus, bistable systems can not only act as persistent memory devices but might also be employed to generate *population heterogeneity*.

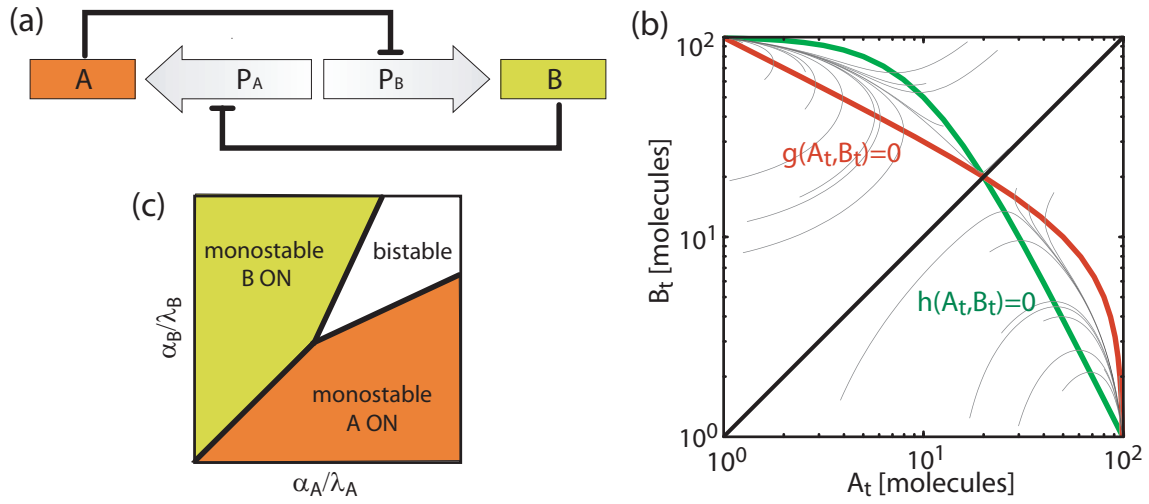


Figure 2.10: In (a) the mutually repressing network motif is depicted. (b) shows the nullclines $g(A_{tot}, B_{tot}) = 0$ and $h(A_{tot}, B_{tot}) = 0$ together with some trajectories under randomly distributed initial conditions (grey). The black line is the *separatrix*, the border which divides the two *basins of attraction*. The stability behavior in dependence of the parameter ratios α_A/λ_A and α_B/λ_B is summarized in the state diagram (c) (schematic).

Mutual repression

The next step in complexity is the mutual-repressor motif depicted in Fig. 2.10 (a). It consists of two genes A and B, that reciprocally repress the transcription of the other gene. As shown in the following, for properly chosen parameters the circuit exhibits bistability: one state with A HIGH and B LOW and another state with A LOW and B HIGH. The rate equations for the total protein numbers A_{tot} and B_{tot} are set up analogously to Eq. (2.25) and we have

$$\frac{dA_{tot}}{dt} = \frac{\alpha_A}{1 + (B_{tot}/\tilde{K_B})^2} - \lambda_A A_{tot} = g(A_{tot}, B_{tot}) \tag{2.27}$$

$$\frac{dB_{tot}}{dt} = \frac{\alpha_B}{1 + (A_{tot}/\tilde{K_A})^2} - \lambda_B B_{tot} = h(A_{tot}, B_{tot}). \tag{2.28}$$

The calculation of the fixed points however, is analytically not possible, since it involves the solution of a fifth order polynomial. They are either computed numerically, or for better intuition one might determine them graphically. This is done in Fig. 2.10 (b). If we plot the functions $g(A_{tot}, B_{tot}) = 0$ and $h(A_{tot}, B_{tot}) = 0$ of Eqs. (2.27) and (2.27) in the phase plane (A_{tot}, B_{tot}), the intersection points correspond to the fixed points of the circuit. Fig. 2.10 (b) shows additionally a few trajectories with random initial conditions $(A_{tot}(t = 0), B_{tot}(t = 0))$ in gray. We observe, that the initial condition determines fate of the individual trajectory: if $A_{tot}(t = 0) > B_{tot}(t = 0)$ the trajectory is attracted by the fixed point (A HIGH, B LOW) and to the other one else. One defines the *separatrix*

(depicted as the black line in Fig. 2.10 (b)) as the border between these two *basins of attraction*. However, one finds that the fixed point behavior again depends strongly on the parameter ratios α_A/λ_A and α_B/λ_B. This is summarized in the so called *state diagram* (see Fig. 2.10 (c)) - similar to the phase diagram known from thermodynamical systems. Experimental evidence for the occurrence of bistability in the mutual-repressor motif has been given by Gardner *et al.* [46] with the construction of a synthetic circuit implementing the described network topology. It is the basis of this work and will be discussed in greater detail in Section 3.1.1.

One might expect that other, more complicated networks also exhibit two or more stable states. *Multistability* can result from the combined effects of positive and negative regulators, or from the combined effects of regulators that each demonstrate bistability. Up to the current date no study came to the knowledge of the author, where multistability was investigated in gene regulatory circuits. Whether bacterial cells can only respond with 'yes' or 'no' or if they can also say 'perhaps' thus belongs to the realm of speculations.

2.4 Modeling techniques

In Section 2.3.2 we investigated the stability behavior of the positive autoregulator and the mutual repressor by means of reaction rate equations. We implicitly assumed, that the protein concentrations are *continuous* variables that follow *deterministic* kinetic laws. However, the concentration increments are actually 1 nM ('birth or death' of one molecule per cell) and the typical concentrations are of the order 1 to 1000 nM. Thus it is evident, that the assumption of continuous concentrations is very crude. But also the second assumption falls with low molecule numbers: the reaction of the type $A + B \rightarrow$ product is based on the molecular collision of the substrates. In the case of low molecule numbers this collision does no more occur with a constant rate but rather with a constant *reaction probability*. The times between two reaction events are then highly fluctuating random variables.

In this section an overview over the standard modeling techniques for chemical reaction systems is given. One of the major assumptions in all the presented formalisms is that we are dealing with a *well-stirred* solution of chemical reactants which is in thermal equilibrium. The question, to what extend this is really the case in the crowded cytoplasm of a cell is still a field of active research [112, 111].

2.4.1 Chemical master equation

We consider a reaction system of N chemical species $\{S_1, \ldots, S_N\}$ interacting through M elementary reaction channels $\{R_1, \ldots, R_M\}$. Elementary means, that the reaction is either *unimolecular* or *bimolecular* (more complicated reactions are actually coupled sequences of them) and that it happens instantaneously. Let $X_i(t)$ be the number of molecules of species S_i at time t. Our ultimate goal is to study the time-evolution of the state vector $\vec{X}(t) = (X_1(t), \ldots, X_N(t))$ under the initial condition that at $t = t_0$ the system was in the state $\vec{X}(t_0) = \vec{X}_0$.

Stoichiometric matrix and propensity function. Two quantities characterize the reaction channel R_j mathematically: The *state-change vector* $\vec{\Sigma}_j = (\Sigma_{1,j}, \ldots, \Sigma_{N,j})$ describes the change of the molecule numbers of $\{S_1, \ldots, S_N\}$ upon an reaction event R_j. If the system was in state $\vec{x}$ before, it jumps instantaneously to the state $\vec{x} + \vec{\Sigma}_j$ when the reaction R_j takes place. The $\{\Sigma_{i,j}\}$ are usually known as the elements of the *stoichiometric matrix* $\mathbf{\Sigma}$. The second characterizing quantity of the reaction channel R_j is its *propensity function* a_j. One can easily keep it in mind as 'probability density', since $a_j(\vec{x})\, dt$ yields the probability, that a reaction of type R_j occurs in the time interval $[t, t+dt]$, given $\vec{X}(t) = \vec{x}$. For unimolecular reactions $S_j \rightarrow$ products the underlying quantum mechanical processes imply basically identical reaction probabilities for all particles and thus the propensity function is linear in the particle number of the substrate: $a_j(\vec{x}) = c_j\, x_i$. The constant c_j turns out to be equal to the reaction rate constant k_j of the deterministic rate equations [49]. For bimolecular reactions there is an additional source of stochasticity imposed by the

uncertainty of the precise positions and velocities of the involved particles. One can only predict the probability of a collision between particles S_i and $S_{i'}$ in the next time interval dt, followed by a reaction of type R_j. The propensity function for bimolecular reactions is then given by $a_j(\vec{x}) = \tilde{c}_j\, x_i\, x_{i'}$, if $i \neq i'$ and $a_j(\vec{x}) = \frac{1}{2}\,\tilde{c}_j\, x_i\,(x_i - 1)$ of $i = i'$ [49]. The products of the molecule numbers account for the number of possible collisions between the reactants, and in the case of one species reacting with itself, every particle S_i has only $x_i - 1$ possible reactants. With Ω being the reaction volume, the constants $\tilde{c}_j$ correspond to either k_j/Ω if the reactants are different species or $2\,k_j/\Omega$ if they are the same [49].

Master equation. With this notation we are able to formulate an expression for the time-evolution of the *probability* $P(\vec{x}, t \mid \vec{x}_0, t_0)$ that the system will be in state $\vec{x}$ at time t, given that it was in the state $\vec{x}_0$ at time t_0. We can write $P(\vec{x}, t + dt \mid \vec{x}_0, t_0)$ as

$$\begin{aligned} P(\vec{x}, t + dt \mid \vec{x}_0, t_0) &= P(\vec{x}, t \mid \vec{x}_0, t_0) \times \left[1 - \sum_{j=1}^{M} a_j(\vec{x})\, dt\right] \\ &\quad + \sum_{j=1}^{M} P(\vec{x} - \vec{\Sigma}_j, t \mid \vec{x}_0, t_0) \times a_j(\vec{x} - \vec{\Sigma}_j)\, dt. \end{aligned} \tag{2.29}$$

Here the first term on the right corresponds to the probability, that the system was already in the state $\vec{x}$ at time t and *no reaction occurs* in $[t, t + dt)$. The individual terms in the second sum reflect the probability, that the system was exactly in the state $\vec{x} - \vec{\Sigma}_j$ at t_0 *and* an reaction of type R_j occurrs in $[t, t + dt)$. In the limit of $dt \to 0$ we obtain the *chemical master equation* (CME) [49]:

$$\frac{\partial P(\vec{x}, t \mid \vec{x}_0, t_0)}{\partial t} = \sum_{j=1}^{M} \Big[a_j(\vec{x} - \vec{\Sigma}_j) P(\vec{x} - \vec{\Sigma}_j, t \mid \vec{x}_0, t_0) - a_j(\vec{x}) P(\vec{x}, t \mid \vec{x}_0, t_0)\Big] \tag{2.30}$$

Although the CME completely determines the time evolution of $P(\vec{x}, t \mid \vec{x}_0, t_0)$, it is in practice almost impossible to solve it analytically. Actually Eq. (2.30) is not only one equation, but rather a set of *infinitely* many equations for the different state vectors $\vec{x}$. This is due to the fact that there are as many state vectors $\vec{x}$ as there are combinations of molecule numbers and in principle these numbers may range from zero to infinity. Therefore the CME can analytically only be solved for a very limited number of simple model systems [45, 62]. Also the numerical solution of the CME is prohibitively difficult for the same reason. However, there is an other way to obtain the desired function $P(\vec{x}, t \mid \vec{x}_0, t_0)$: if one succeeds to *simulate* a stochastic process, that obeys the same probabilistic laws as the master equation, it is possible to *estimate* $P(\vec{x}, t \mid \vec{x}_0, t_0)$ from numerous repeated simulation runs. The resulting probability density may then be used to derive the properties of interest, i.e. the various moments. However, there are further ways of deducing the moments, as will briefly be discussed in the context of the 'derivation' of the deterministic rate equations in Section 2.4.3. But before that, an algorithm for simulating the function $P(\vec{x}, t \mid \vec{x}_0, t_0)$ is presented.

1.	Initialize the system's state $\vec{X}(t_0) = x_0$ and its time $t = t_0$
2.	Compute all the $a_j(\vec{x})$ according to $\vec{x}$ and sum them up to get $a_0(\vec{x})$
3.	Generate τ and j according to Eqs. (2.35) and (2.36)
4.	Update the state by $\vec{x} \to \vec{x} + \vec{\Sigma}_j$ and the system-time by $t \to t + \tau$
5.	Record $(\vec{x}, t)$ as desired; if $t <$ simulation time go to Step 2, else quit

Table 2.3: Gillespie's stochastic simulation algorithm.

2.4.2 Gillespie's stochastic simulation algorithm

In order to account for the stochastic nature of chemical reaction events D. T. Gillespie formulated in 1977 an efficient algorithm for the simulation of these processes [48] - the stochastic simulation algorithm (SSA). The underlying stochastic process is identical with the one that was considered to formulate the master equation in Eq. (2.30) and therefore the method claims to be *exact* [48], i.e. it obeys the same laws as the chemical master equation. Although Gillepie has been the first to formulate this algorithm for chemical reactions, there was an earlier paper of Bortz, Kalos and Lebowitz in 1975, where they worked out an analogous algorithm for the simulation of the Ising-model [25]. This algorithm is in the physical literature known as the *BKL-* or simply the *Kinetic Monte Carlo*-algorithm.

The basic idea of all these algorithms is to simulate the time evolution of the system under investigation by answering two questions: 1. *When* will the next event occur? and 2. If an event takes place, *what type of event* is it? How these decisions can be made is now shown for the SSA.

Stochastic simulation algorithm. Let $p(\tau, j \mid \vec{x}, t)\, d\tau$ be the probability that given $\vec{X}(t) = \vec{x}$, the *next* reaction will occur in $[t + \tau, t + \tau + d\tau)$ *and* will be an reaction of type R_j. In order to derive an analytical expression for $p(\tau, j \mid \vec{x}, t)\, d\tau$ we first note that one can write it as the product of the probability of *no* reaction taking place in $[t, t + \tau)$, $P_0(\tau \mid \vec{x}, t)$, and the probability of a reaction R_j to occur in $[t + \tau, t + \tau + d\tau)$:

$$p(\tau, j \mid \vec{x}, t) = P_0(\tau \mid \vec{x}, t) \times a_j(\vec{x})\, d\tau. \tag{2.31}$$

Further, the probability of having *no* reaction in $[t, t + \tau + d\tau)$ is given by

$$P_0(\tau + d\tau \mid \vec{x}, t) = P_0(\tau \mid \vec{x}, t) \times [1 - \sum_{k=1}^{M} a_k(\vec{x})\, d\tau]. \tag{2.32}$$

In the limit of $d\tau \to 0$ this leads to a simple differential equation with the solution

$$P_0(\tau \mid \vec{x}, t) = \exp(-a_0(\vec{x})\,\tau), \quad \text{where} \quad a_0(\vec{x}) := \sum_{k=1}^{M} a_k(\vec{x}). \tag{2.33}$$

The resulting probability density is now given by

$$p(\tau, j \mid \vec{x}, t) = a_j(\vec{x})\, \exp(-a_0(\vec{x})\,\tau) = p_j(\vec{x}) \times p_\tau(\vec{x}), \tag{2.34}$$

which is a product of a "τ-density" function $p_\tau(\vec{x}) = a_0(\vec{x})\, \exp(-a_0(\vec{x})\,\tau)$ and a "j-density" function $p_j = a_j(\vec{x})/a_0(\vec{x})$. In order to generate random numbers according to these densities we have to draw two uniformly distributed random numbers r_1 and r_2 in $(0, 1]$ and compute τ according to [62]

$$\tau = \frac{1}{a_0(\vec{x})} \ln\left(1/r_1\right) \tag{2.35}$$

and take j to be that integer for which

$$\sum_{k=1}^{j-1} a_k(\vec{x}) \leq r_2\, a_0(\vec{x}) < \sum_{k=1}^{j} a_k(\vec{x}). \tag{2.36}$$

Gillespie's formulation of the stochastic simulation algorithm [48] is eventually shown in Table 2.3.

The price of the exactness of the SSA is in many cases payed by a huge 'computational cost'. If the molecule numbers of the reaction system become large, also the total propensity function $a_0(\vec{x})$ in Eq. (2.35) becomes large, such that the time steps of the simulation are rendered extremely small. In these cases other, approximating algorithms are needed, some of which are briefly discussed at the end of Section 2.4.4.

2.4.3 Reaction rate equations

With the solution of the CME or by simulations with the SSA one gets hands on the full probability density $P(\vec{x}, t \mid \vec{x}_0, t_0)$. The burden for this is in most of the cases a high computational cost. However, for many problems it is not necessary to have the total information contained in the probability density. Very often the time evolution of the first two moments, i.e. the mean and the variance, are sufficient. In the most simplistic description one is only interested in the dynamical behavior of the average molecule numbers of each chemical species. The average of each species S_i is defined as $\langle X_i(t)\rangle = \sum_{\vec{x}} x_i P(\vec{x}, t \mid \vec{x}_0, t_0)$ and its time-evolution is obtained by multiplying both sides of Eq. (2.30) by x_i and summing over all possible states $\vec{x}$. After rearranging the summation indices this leads to

$$\frac{d\langle \vec{X}_i(t)\rangle}{dt} = \sum_{j=1}^{M} \Sigma_{ij}\, \langle a_j(\vec{X}(t))\rangle \tag{2.37}$$

for the first moment [52]. If all reaction channels are linear, we have $\langle a_j(\vec{X}(t))\rangle = a_j(\langle\vec{X}(t)\rangle)$ and Eq. (2.37) turns into a closed set of ordinary differential equations (ODEs) for the first moments. Obviously, if any of the reaction channels is bimolecular, such that $a_j(\vec{X})$ contains quadratic terms in the X_i's, the right side of Eq. (2.37) depends on higher moments. The same holds for the ODEs for each higher moment and thus Eq. (2.37) becomes just the first in an infinite system of coupled ODEs for all the moments. Only in the hypothetical case where we can neglect the fluctuations and all higher moments, we get

$$\frac{d\langle\vec{X}_i(t)\rangle}{dt} = \sum_{j=1}^{M} \Sigma_{ij}\, a_j(\langle\vec{X}(t)\rangle). \tag{2.38}$$

Dividing these equations by the volume Ω gives us the *reaction rate equations* (RRE) of traditional deterministic chemical reaction kinetics. The average molecule abundances $\langle X_i\rangle$ get replaced by the concentrations $[X_i]$ and the rates for the bimolecular reactions transform as mentioned in Section 2.4.1. Although the assumption of negligible fluctuations is certainly not fulfilled in gene regulatory networks, in practice this approach proves to be very useful. Moreover, the RRSs are easy to set up, as long as the stoichiometric matrix and the propensity functions are given. Their solution on the other hand is due to the involved nonlinearities in most of the cases only possible by means of numerical integration.

2.4.4 Stochastic differential equations

Another approach that accounts for the stochastic nature of the reaction events dates back to Langevin's treatment of *Brownian motion* [45]. The irregular movement of small pollen grains in a solution of water inspired him to formulate an equation of motion based on Newton's law, but with an additional *random force* $\xi(t)$

$$M\frac{d^2y(t)}{dt^2} = -\gamma M\frac{dy(t)}{dt} + \xi(t), \tag{2.39}$$

with γ the friction coefficient and y and M the position and mass of the particle, respectively. Since the random force is the result of countless collisions with water molecules, the *central limit theorem* suggests that ξ behaves like gaussian white noise with zero mean and $\langle\xi(t)\,\xi(t')\rangle = \sigma^2\,\delta(t-t')$. The variance σ^2 of the random force is eventually determined by the *Einstein relation*[4] [108]

$$\sigma^2 = 2\gamma M k_B T, \tag{2.40}$$

which relates it to the friction coefficient.

The general Langevin approach can be summarized as follows:

- Formulate the macroscopic equations of motion (or rate equations in the case of chemical reactions) for a given system.

[4]With Eq. (2.39) one determines the expectation value of the squared velocity $\langle\left(\frac{dy}{dt}\right)^2\rangle$ under application of the 'uncorrelatedness' of the Gaussian noise. Then one assumes the system in equilibrium and applies the equipartition theorem from equilibrium thermodynamics $\frac{1}{2}M\langle\left(\frac{dy}{dt}\right)^2\rangle = \frac{1}{2}k_BT$

- Add gaussian white noise with variance σ^2.
- Adjust σ^2 such that the stationary solution reproduces the correct mean-square fluctuations known from statistical mechanics (or by some other considerations).

However, despite its widely spread use this approach should be used with care, since it might lead to wrong results. Especially in the case of the general Langevin equation

$$\frac{dy(t)}{dt} = A(y(t),t) + B(y(t),t)\xi(t), \tag{2.41}$$

with a generally *nonlinear* function $A(y(t),t)$ and *multiplicative noise*, which is reflected in the y(t)-dependence of B, difficulties arise: the differential form of Eq. (2.41) implies the existence of an integral form[5], which involves the solution of a *stochastic integral* (resulting from the second term). Depending on the different interpretations of this stochastic integral[6] different *Fokker-Planck equations* and thus non-equivalent results are obtained, for details see [45]. But even if the proper interpretation was chosen for the respective problem there remain unresolved problems. If the noise source is not an externally acting one with know properties but rather caused by the internal properties of the dynamical system, it is not clear how to choose the the functions A and B.

Chemical Langevin Equation. In some cases the Langevin equation can nevertheless be successfully applied to chemical reaction systems. Following Gillespie [52] one employs the same premises as for the 'derivation' of the CME and demands two additional dynamical conditions, ultimately leading to a time evolution of the Langevin type. It is shown in [51], that the resulting chemical Langevin equation is equivalent to a Fokker-Planck equation, which can in turn be obtained from a Kramers-Moyal expansion of the chemical master equation (*continuum limit*).

Using the same nomenclature as in the previous sections and further introducing $K_j(\vec{x}^t,\tau)$ as the number of reactions R_j occurring in a time interval $[t,t+\tau]$ given that $\vec{X}(t) = \vec{x}^t$, one can write for the state vector at time $t+\tau$

$$X_i(t+\tau) = x_i^t + \sum_{j=1}^{M} \Sigma_{ij}\, K_j(\vec{x}^t,\tau) \qquad (i = 1,\ldots,N). \tag{2.42}$$

The random variables $K_j(\vec{X},\tau)$ are generally *not independent* of each other and their computation is equally difficult as solving the master equation. We will get back to this point later on. If the two following conditions hold the $K_j(\vec{X},\tau)$ can be further approximated:

[5]The integral form of Eq. (2.41) is given by $dy = \int_t^{t+dt} dt' A(y(t'),t') + \int_t^{t+dt} dt' B(y(t'),t')\xi(t')$.

[6]Since the stochastic process $\xi(t)$ can be considered as a random succession of delta peaked kicks, it is not clear which value of y(t) has to be used in the evaluation of the stochastic integral. In *Itô's* interpretation y(t) is taken *before* the random kick arrives ($dy = A(y(t'),t')dt + B(y(t),t)\int_t^{t+dt} dt'\xi(t')$), whereas in *Stratonovich's* interpretation the mean value of y(t) before and after the kick is chosen ($dy = A(y(t'),t')dt + B(\frac{y(t)+y(t+dt)}{2}, \frac{t+[t+dt]}{2})\int_t^{t+dt} dt'\xi(t')$)

Condition I: Demand τ to be *small* enough, such that the change of the state vector during $[t, t+\tau]$ does not alter the propensity functions noticeably:

$$a_j(\vec{X}(t')) \cong a_j(\vec{X}(t)), \qquad \forall\, t' \in [t, t+\tau], \qquad (j = 1, \ldots, M). \tag{2.43}$$

This means that all reaction events occurring in $[t, t+\tau]$ will be basically *independent* of each other, implying that each $K_j(\vec{x}^t, \tau)$ can be approximated by a *Poisson random variable* $\mathcal{P}_j(a_j(\vec{x}^t)\tau)$ with mean and variance equal to $a_j(\vec{x}^t)\tau$. Most importantly, condition I also implies that the individual $\mathcal{P}_j(a_j(\vec{x}^t)\tau)$ are independent of each other as well. Thus, it allows to approximate Eq. (2.42) by

$$X_i(t+\tau) = x_i^t + \sum_{j=0}^{M} \Sigma_{ij}\, \mathcal{P}_j(a_j(\vec{x}^t)\tau) \qquad (i = 1, \ldots, N). \tag{2.44}$$

Condition II: Demand τ to be *large* enough that each reaction channel R_j fires much more than once, i.e.

$$\langle \mathcal{P}_j(a_j(\vec{x}^t)\tau) \rangle = a_j(\vec{x}^t)\tau \gg 1, \quad (j = 1, \ldots, M). \tag{2.45}$$

This condition seems to contradict condition I and in practice it is very likely, that not both can be satisfied simultaneously. However, in the limit of large molecule numbers both can be met: if for instance τ would be chosen such that the expectation values in Eq. (2.45) were around 50 molecules, this change in the state vector would be negligible compared to total molecule numbers of a few thousands.

Hence, if condition II holds each discrete *Poisson* random variable $\mathcal{P}_j(a_j(\vec{x}^t)\tau)$ can be approximated by a continuous *normal* random variable $\mathcal{N}_j(a_j(\vec{x}^t)\tau, a_j(\vec{x}^t)\tau)$ with equal mean and variance. By applying the linear combination theorem [50] for normal random variables, $\mathcal{N}(m, \sigma^2) = m + \sigma\mathcal{N}(0,1)$ and after replacing $\vec{x}^t$ by $\vec{X}(t)$, τ by dt and the unit normal random variables $\mathcal{N}_j(0,1)$ by $\Gamma_j(t)$ (which are δ-correlated in time) one obtains

$$X_i(t+dt) = X_i(t) + \sum_{j=0}^{M} \Sigma_{ij}\, a_j(\vec{X}(t))dt + \sum_{j=1}^{M} \Sigma_{ij} \sqrt{a_j(\vec{X}(t))}\Gamma_j(t)\sqrt{dt} \qquad (i = 1, \ldots, N). \tag{2.46}$$

This has the form of a "standard-form" Langevin equation for a *multivariate* continuous Markov process. It can be transformed to the equivalent "white-noise form" Langevin equation [45, 52]

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} \Sigma_{ij}\, a_j(\vec{X}(t)) + \sum_{j=1}^{M} \Sigma_{ij} \sqrt{a_j(\vec{X}(t))}\Gamma_j(t) \qquad (i = 1, \ldots, N), \tag{2.47}$$

which is called the *chemical Langevin equation* (CLE).

Evidently the validity of the CLE stands and falls with the conditions I and II. It was exemplified, that in the limit of large molecule numbers a *macroscopically infinitesimal* time

τ exists, such that both conditions can be met simultaneously. However, for *typical* molecule numbers in the range of about a few to a few hundred this is not the case. Especially for bistable systems where the expectation values of the respective 'off- state' is often even smaller than 1, the described scheme does not apply. We also saw that condition I allowed us to assume the Poissonian and later the Gaussian random variables $\Gamma_j(t)$ as basically independent of each other (for all j). However, at low molecule abundances one expects complicated correlations between the individual $\Gamma_j(t)$, since the change of the state vector upon a reaction R_j has in this case a non-negligible impact on the propensity functions $a_{j'}(\vec{X})$ and all other reactions $R_{j'}$ might be affected by fluctuations in the number of R_j reactions. This makes it in practice almost impossible to simulate random processes with these intricate correlation structures. For these reasons in this study simulations of the chemical master equation with the SSA are preferred.

As investigated recently by Tănase-Nicola *et al.* [105], the reactions that allow a network to detect biochemical signals, indeed induce correlations between the extrinsic noise[7] of the input signals and the intrinsic noise of the reactions that form the network. This in turn has consequences for the modular description of noise transmission in signaling cascades (for details see [105]).

Approximate algorithms. Besides the *exact* simulation of the stochastic process with the SSA, there exist also other, approximate algorithms, that rely on the 'update-rules' defined by Eqs. (2.44) and (2.46). If condition I is fulfilled, it is reasonable to advance the system by some *pre-selected* time-step τ through the generation of M Poisson random variables according to Eq. (2.44). This is called the *tau-leaping* method and for the optimal choice of τ in this procedure it shall be referred to Cao *et al.* [32]. If additionally condition II is met, the update-rule according to Eq. (2.46) is called the *Langevin leaping formula*, which corresponds to the straightforward simulation of the Langevin equation Eq. (2.47).

[7]For definitions of extrinsic and intrinsic noise, see Section 2.5.1.

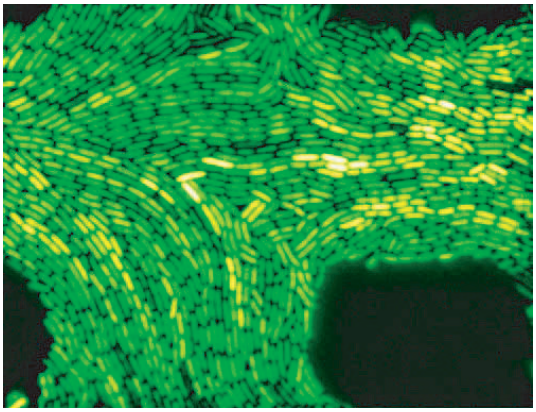


Figure 2.11: In a colony of genetically identical cells the synthesis of an artificially integrated green-fluorescence protein displays a high degree of variability. The image was taken from [6].

2.5 Noise in gene expression

In this section the different sources of variability in gene regulatory systems are presented. From the general description of stochastic gene expression by the master equation, one can derive a fluctuation-dissipation theorem, that allows for an easy and concise description of noise in gene regulatory networks. Last, a glance at the mechanisms for noise suppression or noise amplification is given and some examples are discussed.

2.5.1 Intrinsic and extrinsic noise

The classification of noise into intrinsic and extrinsic contributions has brought about some confusion under biologists and physicists. The origin of this confusion lies in the *variety of possible definitions* of this separation. Since intrinsic and extrinsic basically mean 'inside' and 'outside', the question is mainly how to discriminate the system from the environment. As summarized by Paulsson [87] there are many distinct (and reasonable) ways to distinguish the different contributions - depending on the respective perspective. One definition for instance terms all sources of variability arising from transcription and translation of a *single gene* as intrinsic and the impact of other regulatory proteins on its synthesis as extrinsic. Another definition on the other hand terms all noise arising from gene expression as intrinsic and only the influence of the *non-modeled variables*, which are subsumed into the reaction rates (as e.g. RNAP, RNase or protease abundances), as extrinsic [35]. This definition is very convenient from the modeling point of view. In many cases it is namely assumed, that the fluctuations in the reaction rates are either very slow or of negligible magnitude, such that the rates can be fixed to some constant values. Then all remaining fluctuations are intrinsic and stem from *low molecule number noise* of the modeled variables.

There are many other possible definitions [106, 97, 35, 103] and one should always keep in mind, that this classification is model-specific and therefore to some extend arbitrary.

No matter how one baptizes the fluctuations - one should be aware of their origins. In the following a summary of the most prominent sources of intrinsic and extrinsic noise according to the definition based on modeled and non-modeled variables are mentioned.

Intrinsic noise

In Section 2.2.2 it was stated, that the transcription rate of a gene is proportional to the *equilibrium* probability to find RNA polymerase (RNAP) bound to the respective promoter site. We saw that this probability can be regulated by the presence of activators or repressors and the resulting promoter activity function connected the transcription factor (TF) concentration with the gene activity. But if one scrutinizes this regulated recruitment from a stochastic point of view the situation changes: Although the operator occupation is on *average* given by the *equilibrium* dissociation constant of the TF to the operator, a permanent binding and unbinding of the individual TFs takes place. Only if the rates of binding and unbinding for the TFs are much faster than the association rate of RNAP to the promoter site, RNAP 'senses' the average operator occupancy. For slow attachment and detachment rates of the TFs the gene activity rather oscillates between HIGH and LOW values - following 'adiabatically' the operator occupation states [70, 94, 66]. As a result, the slow operator state fluctuations lead to *transcriptional bursting*. The burst size of mRNAs is analogously to the case of translational bursting (compare Section 2.2.4 on p. 15) given by a geometric distribution. Here we have again a Poisson process (transcriptional initiation by RNAP) that is truncated e.g. by the binding of a repressor into the promoter region . This was also confirmed experimentally in bacteria [54], where an average burst size of 4 transcripts per 'activity' period could be determined for the TetR repressible promoter P_{LtetO}.

Besides transcriptional and translational bursting additional intrinsic fluctuations are imposed by the association and dissociation events of dimerization and the randomly driven proteolysis. However, although these processes contribute to the overall stochasticity at very low protein abundances, in many cases the protein number is so high, that the major sources of noise are the previously mentioned bursting processes. This does not mean, that the protein abundance is not subject to drastic fluctuations. The origin of the protein fluctuations is rather the *amplified* low molecule number noise from the level of operator state fluctuations and random birth and death of mRNAs. Further below a general framework for the description of noise in *linear* chemical reaction systems is presented and the amplification of mRNA and operator-state noise will appear with more clarity.

Extrinsic noise

In a population of bacterial cells not all cells are identical. They vary in in many aspects as e.g. in size, in their doubling time or simply in their location inside the bacterial population. The last point results from non-ideally stirred bacterial populations and leads to gradients in signaling molecules, acidity or in the nutrient supply. The implications for gene regulation are, that every individual cell has a different intracellular composition of

its *housekeeping* protein machinery and of the available chemical building blocks. On the transcriptional level e.g. the number of RNAPs, the abundance of ribonucleotides (the building blocks of mRNA) and the available ATP (the energy source of the process) might be rate limiting [35]. But also on the translational level the abundance of ribosomes, the availability of amino acids (determined by the nutrient supply) or the number of RNases determine the respective transcription and degradation rates. Last, also the degradation rate of the proteins depends on the number of proteases and is thereby also subject to intercellular variations. These are only a few examples that demonstrate possible sources of population heterogeneity of the effective parameters used in the standard model of gene regulation as presented in the previous sections. It shall be mentioned as well, that the parameters not only vary from cell to cell, but also in time within individual cells [97].

However, if certain extrinsic fluctuations are suspected to play a major role for the behavior of the regulatory circuit, it is possible to simply include the respective processes into the model. Therefore standard modeling attempts [81, 13, 106, 82] do not consider fluctuations in the parameters.

2.5.2 Fluctuation-Dissipation Theorem

Measures of noise

In principle the fluctuations of a chemical reaction species S_i are determined by the second moments $\sigma_i^2(X_i) := \langle X_i^2 - \langle X_i \rangle^2 \rangle$ of the stochastic process determined by the master equation Eq. (2.30). Reasonable quantitative measures for noise are either given by the *relative magnitude of noise* $\eta_i(X_i)$

$$\eta_i(X_i) := \frac{\sigma_i(X_i)}{\langle X_i \rangle} \tag{2.48}$$

or by the *Fano Factor* $F_i(X_i)$

$$F_i(X_i) := \frac{\sigma_i^2(X_i)}{\langle X_i \rangle}. \tag{2.49}$$

Which measure is more appropriate, depends on the specific situation. The Fano Factor for instance equals one for the Poisson process, since variance equals the mean in this case. Therefore the Fano Factor is often employed to measure the deviation from a Poisson process. But this is only suitable for univariate random processes, for which the variance is at least proportional to the average. For the case of multivariate random processes the Fano Factor can be misleading, since the poisson distribution holds no special position [87]. In these cases a more useful measure is the dimensionless relative magnitude of noise η_i.

Fluctuation-Dissipation Theorem

As already mentioned, the computation of the η_i or F_i requires the derivation of the second moments from the master equation analogously to Eq. (2.37). Besides this straightforward method there exists a more generic approach that relies on a first order van Kampen's

system size- (or Ω-) expansion of the master equation [86, 87]. By this it is possible to derive a general *Fluctuation-Dissipation Theorem* (FDT), that connects the matrix of covariances $\boldsymbol{\sigma}$ ($\sigma_{ii} = \sigma_i^2$) with the Jacobian $\boldsymbol{A}$ of the dynamics of the averages and the diffusion matrix $\boldsymbol{B}$:

$$\frac{d\boldsymbol{\sigma}}{dt} = \boldsymbol{A}\boldsymbol{\sigma} + \boldsymbol{\sigma}\boldsymbol{A}^T + \boldsymbol{B}. \tag{2.50}$$

More precisely, $\boldsymbol{A}$ is given by

$$A_{ij} = \frac{\partial}{\partial\langle X_j\rangle}\frac{d\langle X_i\rangle}{dt} \tag{2.51}$$

$$= \frac{\partial}{\partial\langle X_j\rangle}\sum_{j=1}^{M}\Sigma_{ij}\, a_j(\langle\vec{X}(t)\rangle), \tag{2.52}$$

where in the second step the fluctuations were neglected and by this the average propensity function was replaced by the propensity function of the averages. As stated in Section 2.4.3 this is exact for linear systems and represents an approximation for nonlinear ones. The diffusion matrix $\boldsymbol{B}$ depends on the size of the random events and is given by

$$B_{ij} = \sum_k \Sigma_{ik}\Sigma_{jk}c_k, \tag{2.53}$$

where the Σ_{ij} are the elements of the stoichiometric matrix $\boldsymbol{\Sigma}$ and the c_k are the reaction probabilities of reaction channel R_k.

For many applications it is sufficient to calculate the steady state noise characteristics, given by setting the left side of Eq. (2.50) to zero. Then the calculation of $\boldsymbol{A}$ and $\boldsymbol{B}$ and the subsequent solution of the resulting set of equations for $\boldsymbol{\sigma}$ is in the linear case straightforward. For nonlinear systems this is generally not the case and leads to complicated expressions or is only numerically solvable. However, even in the linear case the rearrangement of the solution and its *interpretation* in terms of intuitive physical principles is the real conceptual challenge [87]. One example is briefly presented in the following.

Applications and limitations of the FDT

The standard model. For the simple linear gene-mRNA-protein motif as depicted in Fig. 2.12 the normalized stationary variance of the protein ($\eta_3^2 = \left(\frac{\sigma_3}{\langle n_3\rangle}\right)^2$) can be be computed [87] as shown in the lower part of Fig. 2.12. This expression is already arranged in intuitive terms: The first contribution to the noise of the proteins is the Poissonian term resulting from births and deaths of individual proteins. The two other terms have their origin in the fact that the rate of protein synthesis is itself a fluctuating random variable. The second one describes the transduced noise from the Poissonian nature of mRNA birth and death. It is multiplied by a dimensionless factor, ranging from 0 to 1, that measures the time-averaging imposed by the finite lifetime of the proteins. This time-averaging stems

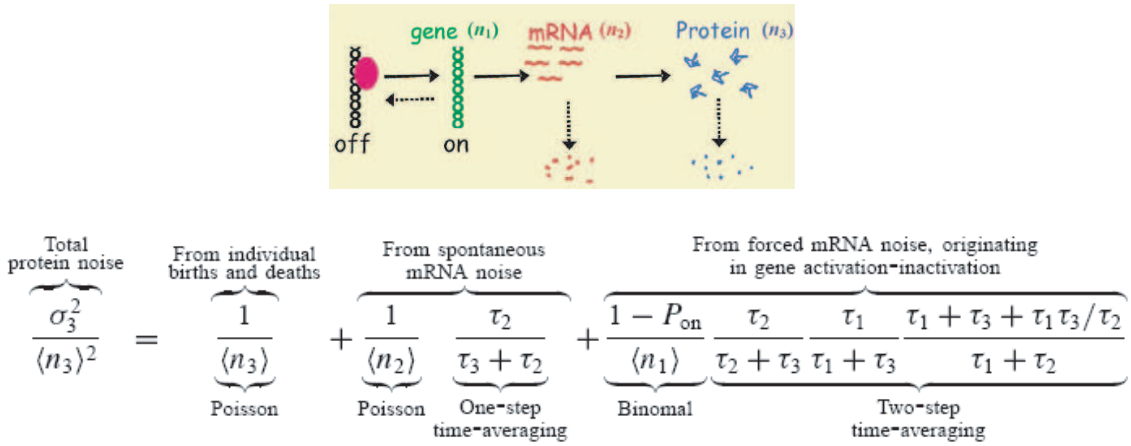


Figure 2.12: Application of the FDT to a linear system of gene activation and inactivation by a repressor, followed by subsequent transcription and translation. The involved timescales τ_i i=1,2,3 are given by the inverse inactivation or degradation rates of the individual species. The figure is taken from [87].

from the fact that the time to reach the steady state value is for any arbitrary variable X, following the rate equation

$$\frac{dX}{dt} = \alpha - \lambda X, \tag{2.54}$$

solely determined by the degradation rate λ. This is because the distance from the steady state $X_\infty = \alpha/\lambda$ follows

$$(X(t_1) - X_\infty) = (X(t_0) - X_\infty)\, e^{-\lambda(t_1 - t_0)}. \tag{2.55}$$

In the above mentioned example the mRNA abundance controls the synthesis rate α of the proteins. Consequently, changes in the number of mRNAs alter the protein steady state level X_∞. If the protein lifetime is much longer than the lifetime of mRNA ($\tau_3 >> \tau_2$ or $\lambda_3 << \lambda_2$), the timescale of the mRNA fluctuations would be so fast, that the proteins could not adjust to the newly defined steady state values. Thus the mRNA fluctuations would be averaged out and the second term in the lower part of Fig. 2.12 disappears. This means that for slow protein degradation the spontaneous mRNA fluctuations cannot influence of the protein level. In the other limit, in which the protein degradation is infinitely fast and therefore the new steady state is approached instantaneously ($\tau_3 = 0$), spontaneous mRNA noise is fully transduced to the protein level.

The last term in the lower part of Fig. 2.12 has its origin activation and inactivation of the gene by the previously mentioned operator state fluctuations. The binomial fluctuations of the operator state induce an enforced mRNA noise. But before it enters the protein noise a time averaging by mRNA and proteins take place (two-step time-averaging).

Other applications. With the FDT a simple theoretical framework for the computation of different contributions to the noise of an individual variable is readily available. It has

been applied for instance to study the dependence of protein noise on the burst factor b [106, 87], as introduced in Section 2.2.4 on p.15. In the case where the gene is always ON and the protein lifetime is much longer than the one of its mRNAs ($\tau_3 >> \tau_2$) one obtains for the Fano Factor of the proteins

$$F_3 = \frac{\sigma_3^2}{\langle n_3 \rangle} \cong 1 + b. \tag{2.56}$$

Thus, only in the limit of continuous or deterministic protein production ($b \to 0$) the Fano factor really approaches unity, as expected for a Poisson process. The linear dependence of the Fano factor on b in Eq. (2.56) was confirmend by altering the translational efficiency through mutations in the ribosome binding site [84].

The FDT was also used to study the noise propagation in a synthetic transcriptional cascade [90] or to discriminate between global or local sources of fluctuations [90, 18].

Limitations of the FDT. The derivation of the FDT in Eq. (2.50) assumes small Gaussian fluctuations [87], while general models of gene regulatory circuits usually produce very broad and skewed distributions. The situation gets even worse when the circuit involves intricate feedback and feedforward loops since they can confer multistability, leading to multipeaked distributions. In these cases the results generated by the FDT are not valid.

2.5.3 Suppression and exploitation of noise

It is astounding, that a *single copy* of a gene with a burst-like production of proteins can generate reliable functionality of genetic circuits. Already Schrödinger was fascinated by this empirical fact [101]:

> *" (...) Eine einzelne nur in einem einzigen Exemplar vorhandene Atomgruppe ist Ausgangspunkt geordneter Vorgänge, die in wunderbarer Weise und nach höchst subtilen Gesetzen aufeinander und auf die Umwelt abgestimmt sind. (...) Es bedarf keiner dichterischen Vorstellungskraft, sondern nur klarer und nüchterner Überlegung, um zu erkennen, daß die gesetzmässige und ordnungsgemäße Abwicklung dieser Vorgänge von einem ganz anderen 'Triebwerk' bestimmt wird als vom 'Wahrscheinlichkeitsmechanismus' der Physik. (...) Es ist gleichgültig, ob wir es erstaunlich oder selbstverständlich finden, daß eine kleine, aber hochorganisierte Atomgruppe fähig ist, in dieser Weise zu wirken; das ändert nichts an der Einmaligkeit dieses Tatbestandes, der ausschließlich bei der lebenden Substanz vorkommt."*

Today it seems as if the additional mechanisms ('Triebwerk') required for the explanation of the highly ordered and proper function of these processes are given by the regulatory structure of the genetic networks themselves. It has been shown, that a negative autoregulatory network module is able to reduce the normalized noise η, when compared to the circuit without negative feedback [106]. Also DNA looping has proved to suppress noise

[113]. But it was also speculated, that the intrinsic determinants of noise themselves as e.g. the operator state fluctuations and the burst size have evolved as a trade-off between energy efficiency and noise reduction [106, 84]. If for instance the burst factor was reduced while the transcriptional efficiency simultaneously was increased, the same average protein abundance would be obtained with a much lower variance. The burden would be however, that the cell had to synthesize many mRNAs, which in most of the cases would be degraded before any protein was translated from them.

Another possible exploitation of noise is 'stochastic focusing' [88] - a phenomenon similar to stochastic resonance [58]. Stochastic focusing refers to the paradox situation, where the noise of biochemical reactions can reduce the noise of the overall network. This is because under certain conditions noise leads to an increased sensitivity of the promoter activity function and thereby fluctuations can be suppressed [89, 107].

One crucial aspect for gene regulatory and signaling networks was highlighted by Kollmann *et al.* [73]. It was shown, that the signaling network responsible for chemotaxis in *E. coli* not only attenuates the noise of individual network components, but is also *robust* against globally acting *co-variations* of the parameters. The up-regulation of the transcription rates (by fluctuations in the RNAP abundance) for instance could have devastating effects for the stability of gene expression patterns, if it would not be compensated for by the network topology itself. It was hypothesized that this concept of robustness against co-variations would apply to a large class of signaling and gene regulatory networks.

Besides the high reliability in the production of genes that are crucial for the proper function of cell fate decisions as for instance in development [12], other genes exhibit remarkable fluctuations. Some of these genes are responsible for the induction of phenotypic diversity, which was speculated to generate evolutionary advantages on the population level [75].

Chapter 3

A conditional memory circuit

The ability to learn and respond to recurrent events can provide a selective advantage to an organism. Learning depends on the capacity to remember transient biological events that occurred in the past. In this chapter a simple genetic circuit that conditionally memorizes a signal in the form of a transcription factor concentration is proposed. The circuit behaves qualitatively similar to memory in an electronic circuit; it reads and stores an input signal only when conditioned to do so by a read signal. It is based on the genuine physical and chemical principles of gene regulation introduced in the last chapter and employs a bistable switch as the 'core memory module'.

In the following we will investigate the "device physics" of this conditional memory circuit. First, the proposed network design is introduced and its difference to a traditional non-conditional memory circuit is elucidated with a simple coarse-grained model. Then a full description including all the microscopic details of gene regulation is developed for a set of experimentally characterized genes and proteins and the employed models are presented. Next, the results of the steady state characteristics are presented. It is shown that the behavior of the conditional memory circuit is analogous to what we know from magnetic memory. After that the deterministic and stochastic dynamics of the circuit are presented. The determining timescales are scrutinized in detail and moreover the impeding effects of molecular noise on the reliability of the circuit are characterized. In a next step we investigate how the circuit's sensitivity can be adjusted to given environmental signals. The last two sections deal with two modeling-specific issues that arose during this work.

3.1 Illustration of the circuit's function

In this section the basic design of the proposed conditional memory circuit is presented. First, the difference with respect to the traditional genetic toggle switch is elucidated. Then the physical working principle of the memory circuit is explained by a simplified mathematical model.

3.1.1 Conditional vs. non-conditional memory

Our starting point for a circuit that exhibits conditional memory is the mutually repressing network motif, which exhibits bistability for properly chosen parameters (see Section 2.3.2). This means, that the circuit will be either in the state where A is high and B is low or in the state where A is low and B is high, depending on the initial concentrations. But for being useful as a memory circuit, it is important that the state of this bistable device can be dictated by some external condition. This means that one has to couple it to some external signal in order to toggle intendedly between the two states, i.e. one has to make the circuit *addressable*.

Unconditional memory. This addressability was achieved in the groundbreaking work of Gardner *et al.* [46] by the coupling of the inducers I_1 and I_2 to the mutually repressing network as depicted in Fig. 3.1 (a). I_1 prevented gene A to repress gene B and so did I_2 with the repression of B on A. Thereby they could toggle between the ON-state (A high) and the OFF-state (A low) by adding one of the inducer signals. The addition of I_2 for instance blocked the repression of gene B on gene A, proteins A could be synthesized and thus the switch was set to the A ON-state. The logic rules of this 'toggle switch' are summarized in Fig. 3.1 (b), and its dynamic behavior is sketched in Fig. 3.1 (c). It is illustrated, how the switch is driven to the A ON state by a transient I_2 exposure (Fig. 3.1 (c), left column) and reset to the A OFF state by a pulse of I_1 (Fig. 3.1 (c), right column). Clearly, the toggle switch always stores the state determined by its *latest* inducer exposure similar to a light switch responding to the latest finger pressing it. In that sense the behavior of the toggle switch is *unconditional*, as no higher hierarchical signal "allows" or "prevents" this storage.

Conditional memory. To obtain a conditional memory we also start with the mutually repressing network motif, but introduce combinatorial control of two other proteins R and S, see Fig. 3.1 (d). The protein R can form homodimers R_2 which repress the expression of gene A via an additional binding site in the promoter P_A. It can also bind to S to form heterodimers RS which bind specifically to another operator site in the promoter region P_B and thereby repress the expression of gene B.

Qualitatively we expect that if R is absent, neither R_2 nor RS can form. In this state the existing memory of the circuit is maintained regardless of the level of S. When a significant amount of R is present, at a low concentration of S mostly homodimers R_2 are formed. Gene A is repressed and the switch is forced into the OFF-state. Conversely, when S is

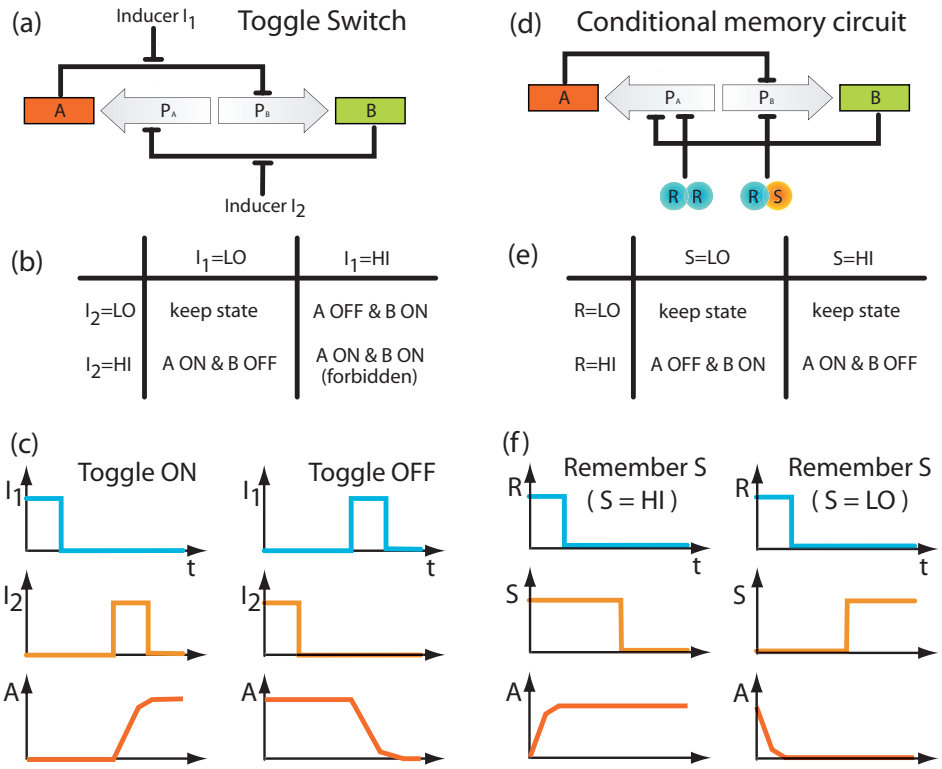


Figure 3.1: The schematic design and function of (a–c) the toggle switch of Gardner *et al.* [46] compared to (d–f) the "conditional memory circuit". In the toggle switch (a) the bistable circuit of two mutually repressing genes A, B is controlled by the two inducers I_1 and I_2 according to the rules (b). The diagrams (c) illustrate how the switch is "set" to the ON state, i.e. high (HI) expression of A, and "reset" to the OFF state, i.e. low (LO) expression of A, by pulses of I_2 and I_1, respectively. In contrast, the conditional memory circuit (d) is regulated by the transcription factors R and S. They form hetero- and homodimers RS and R_2 repressing the transcription of gene A and B, respectively. Effectively, the circuit remembers the expression state of S during the last pulse of R. Hence, R functions as a read signal for the information contained in S, as illustrated in the table (e) and the diagrams (f).

highly expressed, mostly the heterodimers RS form and force the switch into the ON-state, see Fig. 3.1 (e) and (f). Hence at a high level of R, the state of gene A reflects the state of S and at low levels, S does not affect the state of the circuit. This means that R corresponds to the command (or the condition) to "read" the input signal S, which is then "memorized" when R is set to a low level in the following.

3.1.2 Working principle

The working principle of the conditional memory circuit of Fig. 3.1 (d) is best understood within a reduced deterministic description. It considers only the time evolution of the total concentrations of the proteins A and B. In such a description one assumes that all biochemical processes which do not change the total concentrations $A_{\rm tot}$ and $B_{\rm tot}$ are so rapid that they remain equilibrated at almost all times (see Section 3.7). The net change of $A_{\rm tot}$ and $B_{\rm tot}$ due to protein synthesis and degradation then follows rate equations of the form

$$\begin{aligned} \frac{d}{dt}A_{\rm tot} &= \alpha_A P_A(B_2, R_2) - \lambda_p A_{\rm tot} \\ \frac{d}{dt}B_{\rm tot} &= \alpha_B P_B(A_2, RS) - \lambda_p B_{\rm tot} \; . \end{aligned} \tag{3.1}$$

Here, it is assumed that protein degradation occurs at the constant rate λ_p. In contrast, the synthesis of proteins A and B is regulated. Their maximal synthesis rates are denoted by α_A and α_B, while the form of the regulatory control is described by the promoter activity functions $P_A(B_2, R_2)$ and $P_B(A_2, RS)$. The promoter activity is the fraction of time the promoter is not blocked by a repressor and thereby free to bind RNA polymerase (P_A and P_B take on dimensionless values between 0 and 1), see Section 2.2.2. The arguments of P_A and P_B are the concentrations of the dimeric repressors which down-regulate the transcription. Within the thermodynamic model for transcription regulation (Section 2.2.3), the promoter activity function P_A takes the form

$$P_A(B_2, R_2) = \Big(1 + \frac{B_2}{K_{O_B}}\Big)^{-2}\Big(1 + \frac{R_2}{K_{O_{R_2}}}\Big)^{-1} \tag{3.2}$$

and similarly,

$$P_B(A_2, RS) = \Big(1 + \frac{A_2}{K_{O_A}}\Big)^{-2}\Big(1 + \frac{RS}{K_{O_{RS}}}\Big)^{-1} \; . \tag{3.3}$$

Here, the K's denote the equilibrium dissociation constants for the dimer-operator interaction *in vivo* [47]. To achieve the cooperativity required for the bistability of the toggle switch [34], we introduce two binding sites each for the repressor dimers A_2 and B_2, which is reflected in the square of the first factor [24, 23]. Note that the dimer concentrations A_2 and B_2 in Eqs. (3.2) and (3.3) must be expressed in terms of the total protein concentrations $A_{\rm tot}$ and $B_{\rm tot}$ to close the rate equations (3.1). The explicit expressions will be derived in Section 3.7. Similarly, the concentrations of the control proteins, R_2 and RS, are functions of the total protein concentrations $R_{\rm t}$ and $S_{\rm t}$, which we use to quantify the strengths of the read and input signals.

Without read signal ($R_{tot} = 0$), the second factor on the right hand side of Eqs. (3.2) and (3.3) disappears and the conditional memory circuit behaves like the regular toggle switch [46]. The toggle switch shows three different types of behavior depending on the ratio of the maximal promoter activities α_A and α_B: The switch is bistable only when α_A and α_B are similar and sufficiently strong; otherwise it is monostable, either always ON

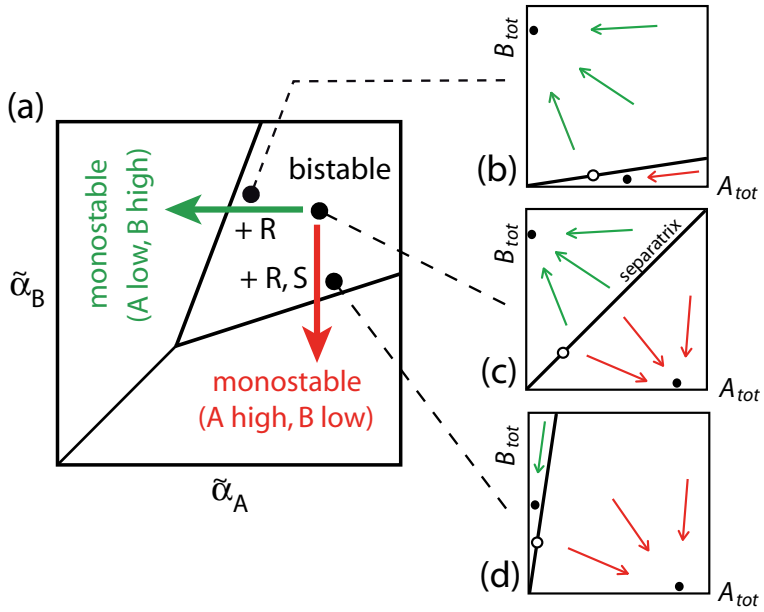


Figure 3.2: Working principle of the conditional memory circuit. (a) When the concentrations of the control proteins R and S change, the circuit moves in the state diagram of the toggle switch, which displays either bistable or monostable behavior depending on the (effective) promoter strengths of gene A and B. (b-d) The control proteins R and S tilt the separatrix, which separates the basins of attraction of the two stable fixedpoints (filled circles). When the circuit reaches the borderline to a monostable regime in (a), one of the stable fixedpoints "annihilates" with the unstable fixedpoint (empty circle).

(A high, B low) or always OFF (A low, B high), see Fig. 3.2(a) and Section 2.3.2. In the conditional memory circuit, the maximal promoter activities get replaced by the effective activities $\tilde{\alpha}_A$ and $\tilde{\alpha}_A$:

$$\tilde{\alpha}_A = \alpha_A/(1 + R_2/K_{O_{R_2}}) \tag{3.4}$$

and

$$\tilde{\alpha}_B = \alpha_B/(1 + RS/K_{O_{RS}}). \tag{3.5}$$

Hence, variation in the concentrations of the control proteins R and S effectively change the maximal promoter activities, and thus can be interpreted as regulated shifts within the state space of the toggle switch. Without read signal, the conditional memory circuit is in the bistable regime and its state is saved. When it receives a read signal, but no S signal, it moves into the monostable low-A regime as indicated by the green arrow in Fig. 3.2 (a). In contrast, when receiving an R and S signal together, it moves into the monostable high-A regime as indicated by the red arrow. The mechanism underlying these operations is illustrated in Figs. 3.2 (b-d), which represent the dynamic properties of the circuit at the three indicated points in the state diagram (Fig. 3.2 (a)): At any given time, the state of the two-gene circuit is specified by the two concentrations A_{tot} and B_{tot} and its dynamics is therefore represented by trajectories in the A_t-B_t phase-plane. These

trajectories are attracted by the stable fixedpoints shown as filled circles in Figs. 3.2 (b-d). Within the bistable regime, there are two stable fixedpoints and the state space is divided by a separatrix into two basins of attraction for these fixedpoints (the empty circle indicates the unstable fixedpoint). When a read signal is given, this separatrix tilts either towards the A-axis or the B-axis, depending on the signal S. Thereby, the basin of attraction of one of the fixedpoints is eliminated, so that the circuit is ultimately forced towards the remaining fixedpoint. The upward/downward tilting of the separatrix is the physical working principle underlying the two functional operations of Fig. 3.1(f). This simple picture holds only within the reduced model in Eq. (3.1), but we will see that the qualitative behavior survives in a more realistic quantitative description developed in the following.

3.2 Quantitative model

In this section a quantitative model model for the conditional memory circuit is developed. All involved reaction mechanisms are described and the different levels of modeling are presented. Last, the used parameters are stated.

3.2.1 Chemical reaction scheme

A quantitative model of Fig. 3.1 must involve all details of gene regulation introduced in Section 2.2. Fig. 3.3 shows the reaction mechanisms and introduces the parameter and variable notations used for the remainder of this work.

Toggle switch module. The upper part of Fig. 3.3 depicts the toggle switch module: genes A and B are transcribed from the promoters P_X into mRNA m_X at a rate ν_{m_X} (X = A, B for the rest of the paragraph). These mRNAs are either translated at rate ν_p or actively degraded with a degradation rate λ_m. Monomeric and dimeric proteins associate or dissociate with on- and off-rates k_X^+ and k_X^- respectively[1] and are both degraded at rates λ_p. Finally the dimers bind and unbind independently to the two operator sites O_{X_1} and O_{X_2} inside the promoter region of the respective antagonist at rates $k^+_{O_{X_j}}$ and $k^-_{O_{X_j}}$ (j=1,2). We introduce *two* identical binding sites for A and B on each promoter in order to guarantee a sufficient degree of cooperativity needed for the emergence of bistability (see Section 2.3.2). However, cooperative interactions between the two TFs are *not* assumed, since this would confer harder experimental constraints.

Regulatory front-end. In the lower part of Fig. 3.3 the interactions of the "regulatory front-end" are shown. As above we equally include the processes of transcription, translation and turnover of genes R and S. Protein R can either form homodimers R_2 with itself

[1] In Fig. 3.3 only the equilibrium constant is depicted, but the model includes all on- and off-processes explicitly.

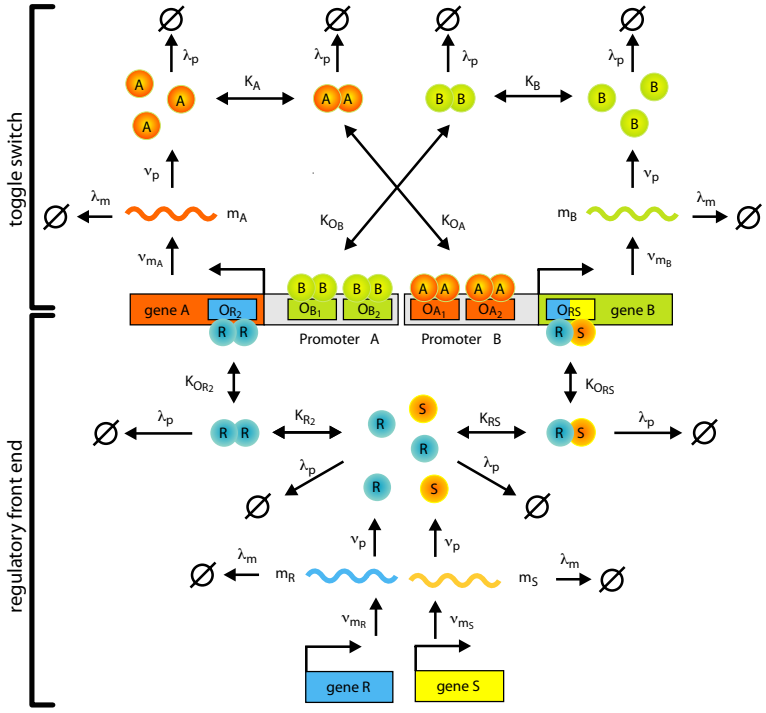


Figure 3.3: Detailed reaction mechanisms included in the model of the conditional memory circuit: the translated proteins form dimers, which bind dual, but independently to their operator sites. If at least one dimer is bound, transcription of the downstream gene is inhibited.

or heterodimers RS with S. In this model we forbid the possible formation of S_2 homodimers. They don't play a functional role in our design because there is no binding site for S_2. Thus, the only effect of S_2 homodimerization is the reduction of the monomeric form of S. The dimers R_2 and RS can bind to their respective operator site *downstream* of the transcriptional start site, see Fig. 3.3. This downstream binding is necessary, because there are spatial constraints for the binding to the promoter region and no more than two TFs can bind to one promoter site.

Promoter states and elementary reactions. From Fig. 3.3 one can read off the elementary reactions between the reactants. One obtains exemplarily for translation, dimerization and turnover of gene A the following reactions

$$\begin{aligned} m_A &\xrightarrow{\nu_p} m_A + A \\ A + A &\xrightarrow{k_A^+} A_2 \\ A_2 &\xrightarrow{k_A^-} A + A \\ m_A &\xrightarrow{\lambda_p} \emptyset \\ A &\xrightarrow{\lambda_p} \emptyset \\ A_2 &\xrightarrow{\lambda_p} \emptyset. \end{aligned}$$

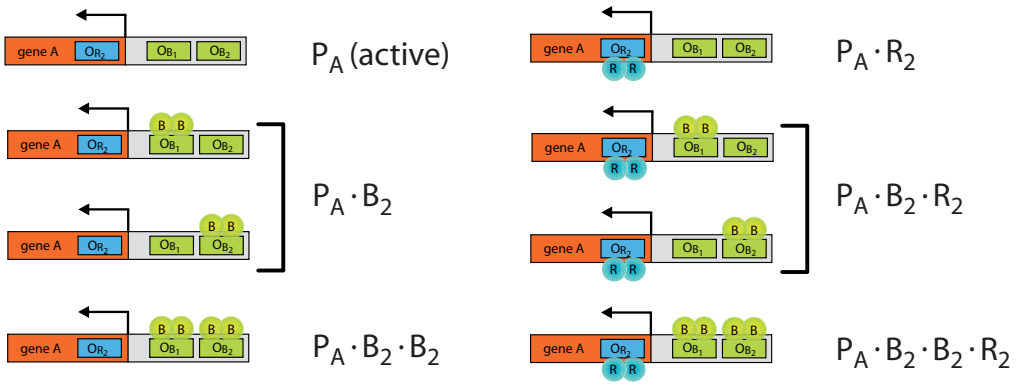


Figure 3.4: In this model the eight possible promoter states are captured by 6 distinct variables (exemplarily only the promoter P_A is shown). We account for the two fold degeneracy of the states $P_A \bullet B_2$ and $P_A \bullet B_2 \bullet R_2$ by doubling the respective on- and off-rates, see text for details.

The full list of 56 reactions for the 24 distinct species (including the reactions between the different promoter states, see below) can be found in Appendix A.1, Eqs. (A.1) - (A.56).

Transcriptional regulation is modeled by the introduction of 6 different promoter occupation states for each of the promoters P_A and P_B as depicted for P_A in Fig. 3.4. We treat each of these promoter states as individual variables, which are transformed into each other by binding and unbinding of the allowed dimeric transcription factors (compare Fig. 3.3 and Refs. [13, 70]). Fig. 3.4 shows that there are two distinct microscopic configurations for the variables $P_A \bullet B_2$ and $P_A \bullet B_2 \bullet R_2$: one with B_2 bound to O_{B_1} and the other one with B_2 bound to O_{B_2}. Since the operators O_{B_1} and O_{B_2} are identical, we account for this two fold degeneracy by doubling the rates of the reactions, that lead to one of these states, e.g.

$$P_A + B_2 \xrightarrow{2\,k^+_{O_B}} P_A \bullet B_2 \tag{3.6}$$

$$P_A \bullet B_2 \bullet B_2 \xrightarrow{2\,k^-_{O_B}} P_A \bullet B_2 + B_2 \tag{3.7}$$

but

$$P_A \bullet B_2 + B_2 \xrightarrow{k^+_{O_B}} P_A \bullet B_2 \bullet B_2 \tag{3.8}$$

$$P_A \bullet B_2 \xrightarrow{k^-_{O_B}} P_A + B_2. \tag{3.9}$$

Input signals. The aim is to control the time dependent transcription rates of genes R ($\nu_{m_R}(t)$) and S ($\nu_{m_S}(t)$), and to use them as input signals to the conditional memory circuit. For instance, the circadian rhythm may impose a periodic transcription rate $\nu_{m_R}(t)$ due to other regulatory processes in the cell. The transcription of S might be coupled via a signal transduction pathway or a quorum sensing mechanism [19] to some time-dependent

external signal, such as the cell density or the nutrient supply. Therefore our aim is to characterize the response of the circuit for different forms of the input signals $\nu_{m_R}(t)$ and $\nu_{m_S}(t)$.

3.2.2 Deterministic and stochastic model of the full system

Deterministic rate equations Following the general framework of Section 2.4.3 we set up now the rate equations from the chemical reaction system in Eqs. (A.1) - (A.56). This leads, again exemplarily for mRNA and monomers A, to the following equations

$$\frac{dm_A}{dt} = \nu_{m_A} P_A - \lambda_m m_A \tag{3.10}$$

$$\frac{dA}{dt} = \nu_p m_A - 2k_A^+ A^2 + 2k_A^- A_2 - \lambda_p A. \tag{3.11}$$

The full list of equations for the 24 variables are stated in Appendix A.2. With a number of steady state assumptions and in the limit of large molecule abundances this vast number of degrees of freedom (24) can be reduced to two: we obtain the simplified (or reduced) model for the total number of proteins A_{tot} and B_{tot} (Eqs. (3.1); for derivation see Section 3.7 on p.79). In Section 3.7 it is also shown, that the assumptions leading to the reduced model are not really met for the used set of parameters. Therefore only the full model is appropriate to serve as a reference for the stochastic simulations and all quantitative statements related to deterministic issues will be obtained from numerical integration[2] of the full model.

Stochastic simulation algorithm. The vast number of reactants (24) and reaction channels (56) demanded a versatile implementation of the stochastic simulation algorithm (Section 2.4.2). This was essential to this work, especially because the goal was to investigate different modifications of the model. Therefore a comprehensive program with the stoichiometric matrix and the reaction probabilities as input parameters was coded in C. The source code can be found in Appendix B.

3.2.3 Parameter choice

To render this theoretical study as realistic as possible, the parameters of all involved processes were collected carefully from existing literature. In Table 3.1 and Table 3.2 the complete list of parameters used here is presented. However, some of them were not experimentally characterized and therefore additional assumptions had to be made, as will be shown in the following.

[2]The solver routine 'ode23s' for stiff sets of ordinary differential equations from MATLAB was used.

Reaction/event	Parameter	Reference and comments
Transcription		
$P_A \xrightarrow{\nu_{m_A}} P_A + m_A$	$\nu_{m_A} = 5\,\mathrm{min}^{-1}$	strong promoters; programmable through choice of nucleotide
$P_B \xrightarrow{\nu_{m_B}} P_B + m_B$	$\nu_{m_B} = 5\,\mathrm{min}^{-1}$	sequence [79]
$P_R \xrightarrow{\nu_{m_R}} P_R + m_R$	$P_R\nu_{m_R} = (0.01 - 1)\,\mathrm{nM\,min}^{-1}$	inducible promoter; [79]
$P_S \xrightarrow{\nu_{m_S}} P_S + m_S$	$P_S\nu_{m_S} = (0.01 - 2)\,\mathrm{nM\,min}^{-1}$	inducible promoter; [79]
Translation		
$m_i \xrightarrow{\nu_p} p_i$	$\nu_p = 2.3\,\mathrm{min}^{-1}$	for all translation events i; selected to match average burst size of 10 proteins/mRNA for the given mRNA half-life [106]
Turnover		
$m_i \xrightarrow{\lambda_m} \emptyset$	$\lambda_m = 0.23\,\mathrm{min}^{-1}$	for all transcripts; corresponds to a half-life T_m of 3 min [21]; ($\lambda_m = ln(2)/T_m$)
$p_i \xrightarrow{\lambda_p} \emptyset$	$\lambda_p = 0.138\,\mathrm{min}^{-1}$	for all proteins; corresponds to a half-life T_p of 5 min achieved by SsrA-tags [56, 83];

Table 3.1: Parameters for transcription, translation and turnover.

Reaction/event	Parameter	Reference and comments
Dimerization		
$A + A \underset{k_A^-}{\overset{k_A^+}{\rightleftharpoons}} A_2$	$k_A^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_A^- = 0.02\ \mathrm{min}^{-1}$	parameters of TetR; for details see paragraph **Equilibrium constant of TetR dimerization**
$B + B \underset{k_B^-}{\overset{k_B^+}{\rightleftharpoons}} B_2$	$k_B^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_B^- = 0.003\ \mathrm{min}^{-1}$	parameters of LacI [104]
$R + R \underset{k_{R_2}^-}{\overset{k_{R_2}^+}{\rightleftharpoons}} R_2$	$k_{R_2}^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_{R_2}^- = 2\ \mathrm{min}^{-1}$	parameters of 434 repressor; In [40] *in vitro* measurements yielded $K = 1000\,nM$ but the results in [61] suggested much lower values for K *in vivo*. Therefore we used $K = 10\,nM$ for the *in vivo* value, in agreement with [74].
$R + S \underset{k_{RS}^-}{\overset{k_{RS}^+}{\rightleftharpoons}} RS$	$k_{RS}^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_{RS}^- = 2\ \mathrm{min}^{-1}$	parameters of 434 and 434[α3($P22R$)] repressor; since the dimerization domain is unchanged, we used same K as for 434 repressor (for reference see above)
TF-DNA binding		
$A_2 + O_{A_i} \underset{k_{O_A}^-}{\overset{k_{O_A}^+}{\rightleftharpoons}} A_2 \cdot O_{A_i}$	$k_{O_A}^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_{O_A}^- = 2\ \mathrm{min}^{-1}$	for both binding sites i=1,2: $(TetR)_2$ interacting with $tet-O_1$; the given *in vitro* equilibrium dissociation constant of $K_d = 0.45\,nM$ [22] was altered in order to obtain a more symmetric toggle switch
$B_2 + O_{B_i} \underset{k_{O_B}^-}{\overset{k_{O_B}^+}{\rightleftharpoons}} B_2 \cdot O_{B_i}$	$k_{O_B}^+ = 0.308\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_{O_B}^- = 2.22\ \mathrm{min}^{-1}$	for both binding sites i=1,2: $(LacI)_2$ interacting with $lac-O_1$; [116]
$R_2 + O_{R_2} \underset{k_{O_{R_2}}^-}{\overset{k_{O_{R_2}}^+}{\rightleftharpoons}} R_2 \cdot O_{R_2}$	$k_{O_{R_2}}^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_{O_{R_2}}^- = 0.32\ \mathrm{min}^{-1}$	$(434)_2$ interacting with $434 - O_1$; [40]
$RS + O_{RS} \underset{k_{O_{RS}}^-}{\overset{k_{O_{RS}}^+}{\rightleftharpoons}} RS \cdot O_{RS}$	$k_{O_{RS}}^+ = 0.2^*\ \mathrm{nM}^{-1}\mathrm{min}^{-1}$ $k_{O_{RS}}^- = 1\ \mathrm{min}^{-1}$	$(434[\alpha3(P22R)])_2$ interacting with hybrid operator, see [61]

Table 3.2: Chosen parameters for dimerization and TF-DNA binding. *: In these cases only the equilibrium dissociation constant K was given and the off-rate was calculated from $k^- = k^+K$.

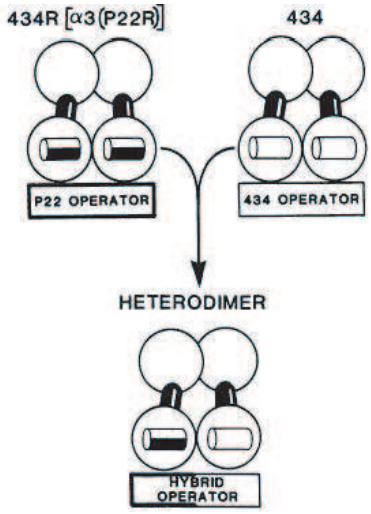


Figure 3.5: *Top:* The homodimers $(434\text{R} \bullet 434\text{R}[\alpha3(\text{P22R})])_2$ and 434R_2 bind to the native P22 and 434 operators, respectively. Their amino-terminal recognition helices are depicted as cylinders. *Bottom:* After mixing, a heterodimer of the two proteins is shown bound to a hybrid P22/434 operator, where each repressor monomer is bound to its cognate half-site. The figure was taken from [61].

Concrete genes and proteins

Toggle switch. For the proteins A and B of the toggle switch an obvious choice is to use the same proteins as in the toggle switch of Gardner et al. [46]. These are Tet repressor (TetR) for A and Lac repressor (LacI) for B, both belonging to the best characterized TFs in *E. coli*. Besides the fact that almost all thermodynamic and kinetic parameters of these proteins are known and thus ideal modeling premises are given, this pairing has already proven to exhibit bistabililty [46]. Another possibility would be the usage of the well studied bacteriophage λ repressor CI [46, 92] instead of one of the other proteins. However, the strong cooperative binding of CI_2 to the adjacent operators O_{R1} and O_{R2} imposes additional experimental constraints. The spacing between the operator sites for instance has to be kept at its native value.

Regulatory front end. Hollis *et al.* characterized a pair of proteins, the 434 repressor (434R) and its synthetic mutant 434R[α3(P22R)], which behaves like R and S described above, see Fig. 3.5. They replaced the amino-terminal domain of 434R repressor by the recognition helix of the *Salmonella* phage P22 repressor, resulting in the mutant protein 434R[α3(P22R)]. Typically the corboxyl-terminal domain (which is unaffected by the mutation) mediates the dimerization between the proteins, whereas the amino-terminal domain is responsible for the recognition of the DNA sequence. They could show that a mixed solution of 434R and 434R[α3(P22R)] yields homodimers $(434R)_2$ and heterodimers $434\text{R} \bullet 434\text{R}[\alpha3(\text{P22R})]$.

Additional to the mutant protein they engineered a hybrid operator site, which consisted

one half of the 434R operator and the other half of the P22R operator. The resulting *chimeric operator* bound the heterodimer with high affinity and specificity. In the following we will refer to the 434 repressor as protein R and to the mutant protein as S. Alternatively one could have used the pair of TFs described by Dmitrova *et al.* in [39], but since the parameter details are not known in that case, they were not considered here.

Protein degradation

The protein half-life determines the timescale the conditional memory circuit, as we will see in Section 3.4.2. Therefore it is favorable to use proteins, that are not only diluted by cell division but are rather actively degraded (compare Section 2.2.4). This can be achieved by the addition of SsrA tags [67] to the end of the sequence of the proteins. Natively, the SsrA system is used to clear stalled ribosomes. This can happen if the lack of an amino acid or a tRNA prevents the ribosome to continue the assembly of the protein. The SsrA system involves a tRNA-like structure of mRNA which enters in this case the ribosome and thereby forces the translation of its own mRNA. This sequence, once translated, serves as a recognition site for the proteases ClpXP or Tsp and thus degradation is enforced. The resulting protein half-life is of the order of a few (3-5) minutes [56, 83], so we will use a half-life of 5 minutes for all proteins, see Table 3.1.

Diffusion limited on-rate

For almost all bimolecular reactions, i.e. for TF dimerization and TF-DNA binding, only the equilibrium dissociation constants were experimentally determined, see Table 3.2. Since our aim is to model the binding and unbinding processes explicitly, we had to find an estimate for either the on- or the off-rate and calculate the other one ($K = \frac{k^-}{k^+}$).

A natural assumption is, that at least the order of magnitude of the on-rate is determined by diffusion, because the elementary reaction requires the molecular collision of the two reactants. We assume in the lack of data, that only one out of ten molecular collisions leads to a successful reaction and thus $k^+ = \frac{k^+_{DS}}{10}$ for *all* on-rates. Here k^+_{DS} is the diffusion limited on-rate given by the *Debye-Smoluchowski theory* [27]

$$k^+_{DS} = 4\pi D_3 b, \tag{3.12}$$

where D_3 is the *in vivo* diffusion constant for three-dimensional diffusion in the cytoplasm of a cell and b is the scattering cross-section of the two molecules. Typical values are $D_3 \approx 7\,\frac{\mu\mathrm{m}^2}{\mathrm{s}}$ (for the GFP protein) [42] and $b = 5\,\mathrm{nm}$, leading to an on-rate of $k^+_{DS} = 2\,\mathrm{nM}^{-1}\mathrm{min}^{-1}$. Thus we use $k^+ = 0.2\,\mathrm{nM}^{-1}\mathrm{min}^{-1}$, which seems to be reasonable since the only known on-rate reported for the Lac repressor is $k^+_{LacI} \approx 0.3\,\mathrm{nM}^{-1}\mathrm{min}^{-1}$ [116].

Equilibrium constant of TetR dimerization

The folding and dimerization of Tet repressor seems to be concerted, which means that there is no evidence for folded monomers [60]. However, by indirect measurements it was

shown that the free energy difference for the dimerization reaction (at 22° C and standard conditions) is larger than

$$\Delta G_0 > 54 \frac{\mathrm{kJ}}{\mathrm{mol}} \tag{3.13}$$

and thus the equilibrium is strongly biased towards the dimer. This results in an equilibrium dissociation constant of

$$K_{TetR} = V^{-1}\, e^{-\Delta G_0/RT} < 0.27\,\mathrm{nM}, \tag{3.14}$$

where V is the reaction volume at standard conditions (1 liter). Thus we choose $K_{TetR} = 0.1\,\mathrm{nM}$ for convenience.

Transcription, translation and mRNA turnover

The translation rates and half-lifes of the used mRNA's were not given in the literature. Therefore we assumed in all cases typical values of 3 minutes half-life and adjusted the transcription rates such that an average burst factor of 10 proteins per mRNA was obtained, see Table 3.1.

For the transcriptional processes we assumed in all cases strong promoters with maximal transcription rates of $\nu_{m_A} = \nu_{m_B} = 5\,\mathrm{min}^{-1}$ for $\mathrm{P_A}$ and $\mathrm{P_B}$ and tunable transcription rates for the control proteins $\nu_{m_R} = (0.01 - 1)\,\mathrm{min}^{-1}$ and $\nu_{m_S} = (0.01 - 2)\,\mathrm{min}^{-1}$, see also Table 3.1. Since the promoter sequence is *programmable* and therefore the transcription rate tunable over a wide range (compare Section 2.2.2) it is reasonable to study the behavior of the circuit in dependence of ν_{m_A} and ν_{m_B} in the following. We could alternatively tune the translation rates via mutations in the ribosome binding site, but for simplicity we will only vary the experimentally most convenient parameters.

3.3 Steady state characteristics

In this section the steady state behavior in dependence of the programmable transcription rates of A and B, ν_{m_A} and ν_{m_B} respectively, is characterized. First, it is demonstrated that the circuit may exhibit bistability for the chosen genes and proteins - a necessary condition for the emergence of *memory*. Then the influence of the regulatory front end on the toggle switch module is elucidated in quantitative terms and it is shown, that this eventually results in a switchable hysteresis as known from magnetic memory.

3.3.1 State diagram

In the simple descriptions of the circuits in Section 2.3.2 we saw, that their steady state behavior crucially depends on the parameters. As the ratios of production and degradation of the proteins were altered, bifurcations from one to two stable steady states occurred. This stability behavior in dependence on the parameters could be summarized in a *state diagram* analogously to the phase diagram in thermodynamics. Fig. 3.6 shows the result

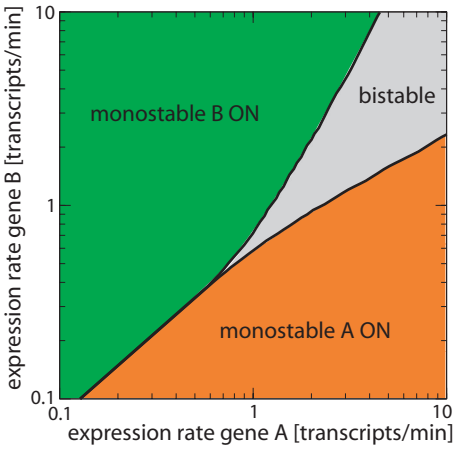


Figure 3.6: State diagram of the conditional memory circuit in the absence of the control proteins R and S. Data obtained from the full deterministic model (Section 3.2.2).

for the full model of the conditional memory circuit in the *absence* of the control proteins R and S, i.e. $\nu_{m_R} = \nu_{m_S} = 0$. As mentioned before, in this limit the behavior of conditional memory circuit reduces to that of the simple genetic toggle switch [46]. If the promoter strength (or maximal transcription rate ν_{m_A} and ν_{m_B}) of one gene dramatically dominates over the other, all initial conditions of concentrations will end in a state in which the dominating gene is highly expressed and the other one is only basally expressed. Since this behavior corresponds to an attractive fixedpoint in the 24 dimensional phase space, one speaks of monostable behavior. On the other hand, if the transcription of both genes is comparably large *and* sufficiently strong the circuit exhibits bistable behavior and two attractive fixedpoints are present - one with A high and B low and the other one with A low and B high.

Note that the 'phase' borders are somewhat asymmetric due to asymmetries in the parameter values of TetR and LacI. The width of the bistability region is determined as in [46, 34] by the degree of cooperativity. Increasing cooperativity widens the bistability region in Fig. 3.6 (data not shown).

Stability analysis

In the case of the autoregulatory positive feedback circuit the steady state solution could be obtained analytically and the stability of these fixedpoints could be determined by linear stability analysis. But already for the simple description of the mutually repressive circuit it was evident that both the fixedpoints and their stability had to be computed numerically.

However, in the current model with 24 state variables it is even worse: it turned out, that also the numerical stability analysis[3] failed. Since the scope of this work was not

[3]The stability analysis with MATHEMATICA yielded results for the reduced description in Eqs. (3.1). However, for the full model the computations for *one* parameter set were aborted after the runtime of a day.

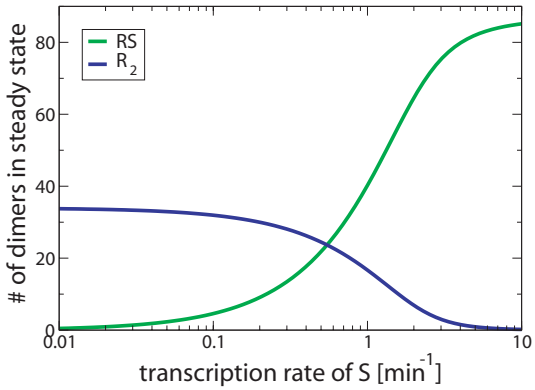


Figure 3.7: Steady state characteristics of the regulatory front end. At a constant transcription level of R ($\nu_{m_R} = 1.2/min$) the transcription rate of S was varied. At low levels of S formation of R_2 is dominant, whereas at high transcription rates of S the formation of RS is dominant. The data was obtained by numerical integration of the full model.

the inversion of high dimensional matrices, a pragmatic way of obtaining the fixed-point behavior had to be chosen: Two initial conditions $\vec{X}_1$ and $\vec{X}_2$ were prepared - one with A high and B low and one with A low and B high. The respective high and low values were chosen such that the two initial conditions belonged to the different basins of attraction in the bistable regime[4]. In the monostable case there is only one basin and all trajectories are attracted by the same fixedpoint. The two simulations were run for long times t_∞ and it was checked, whether their resulting state vectors differed by a significant amount. If

$$\left|\vec{X}_1(t_\infty) - \vec{X}_2(t_\infty)\right| > \varepsilon(t_\infty) \tag{3.15}$$

the system was considered to be bistable and monostable else. As depicted in Eq. (3.15), the discrimination 'distance' $\varepsilon(t_\infty)$ between the state vectors has to be adjusted according to t_∞, such that two trajectories that converge to the same fixed point are not erroneously identified as trajectories that converge to different fixed points.

3.3.2 Regulatory front end

Dimer balance in steady state. In Section 3.2.1 it was mentioned, that the input signals to the conditional memory circuit are the time-dependent transcription rates of the proteins R and S, ν_{m_R} and ν_{m_S} respectively. It is expected that for a given read signal (R high) the balance between homo- and heterodimers R_2 and RS can be adjusted by the amount of proteins S. But is this really the case for the physiologically relevant parameters given in Table 3.1 and Table 3.2?

[4]We chose for the high-values of A and B total protein numbers of 5000 and for their low-values 0. The high-values were estimated by increasing them until the resulting state diagram didn't change any more. The resulting value was multiplied by a factor 5.

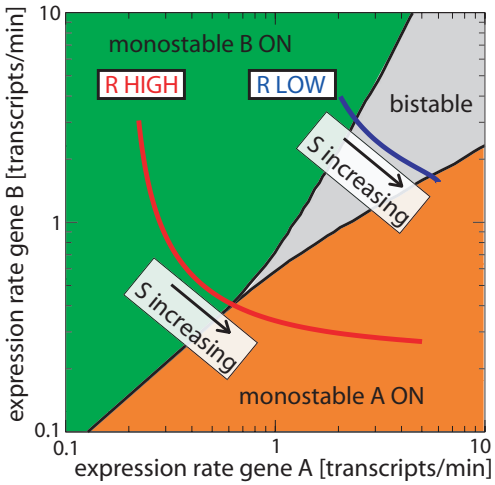


Figure 3.8: Impact of the regulatory front end on the stability of the toggle switch. For R HI ($\nu_{m_R} = 1\,\mathrm{min}^{-1}$, red curve) the variation of S from basal level ($\nu_{m_S} = 0.01\,\mathrm{min}^{-1}$) to a HI level ($\nu_{m_S} = 2\,\mathrm{min}^{-1}$) drives the memory circuit from the monostable (A OFF, B ON) to the monostable (A ON, B OFF) regime. When R is LOW ($\nu_{m_R} = 0.1\,\mathrm{min}^{-1}$, blue curve) the circuit is driven through the bistable regime under variation of S over the *same* range as in the R HI case, resulting in hysteretic behavior (the stable fixed points of A_{tot} along these traces are plotted in Fig. 3.9 (a)). The simulations were carried out with the full deterministic model. The promoter strengths of P_A and P_B are $\nu_{m_A} = \nu_{m_B} = 5\,\mathrm{min}^{-1}$.

Fig. 3.7 shows the steady state behavior of homo- and heterodimers R_2 and RS as a function of the store signal strength ν_{m_S} for a constant read signal ($\nu_{m_R} = 1.2\,\mathrm{min}^{-1}$). It can be obtained by solving the determining equations

$$R_{tot} = RS + 2R_2 + \sqrt{\tilde{K}_{R_2} R_2} \tag{3.16}$$

$$S_{tot} = RS + \frac{\tilde{K}_{RS} RS}{\sqrt{\tilde{K}_{R_2} R_2}} \tag{3.17}$$

for R_2 and RS. Here $\tilde{K}_{RS} = (k^-_{RS} + \lambda_p)/k^+_{RS}$ and $\tilde{K}_{R_2} = (k^-_{R_2} + \lambda_p)/k^+_{R_2}$ are the *in vivo* dimerization constants [29] and $R_{tot} = (\nu_{m_R}\nu_p)/(\lambda_m\lambda_p)$ and $S_{tot} = (\nu_{m_S}\nu_p)/(\lambda_m\lambda_p)$ the total R and S abundances.

At basal level of S ($\nu_{m_S} = 0.01\,\mathrm{min}^{-1}$) almost only homodimers R_2 form, whereas at high levels of S ($\nu_{m_S} = 10\,\mathrm{min}^{-1}$) the heterodimers dominate. Without changing the dimerization constants the crossover point at which the two species have the same abundance, can be adjusted by changing the level of R: a reduction of R shifts this point to lower values of ν_{m_S} and an increase of R shifts it to higher values of ν_{m_S}.

Impact on the toggle switch. It was stated in the context of the reduced model, that the control dimers R_2 and RS reduce the effective expression rates α_A and α_B of the promoters P_A and P_B in Eqs. (3.4) and (3.5). It turns out that the same holds for the full

model, if we compute the equilibrium probability to find the promoters P_A and P_B *not* occupied by the control dimers. Thus they reduce the maximal transcription rates ν_{m_A} and ν_{m_B} to effective transcription rates $\tilde{\nu}_{m_A}$ and $\tilde{\nu}_{m_A}$:

$$\tilde{\nu}_{m_A} = \nu_{m_A}/(1 + R_2/K_{O_{R_2}}) \tag{3.18}$$

and

$$\tilde{\nu}_{m_B} = \nu_{m_B}/(1 + RS/K_{O_{RS}}). \tag{3.19}$$

Therefore the addition of different ratios of R_2 and RS results in shifts in the state diagram, and hence the fixed point behavior can be tuned. How do the effective transcription rates in Eqs. (3.18) and (3.19) change if we keep the read signal constant and vary the store signal, i.e. if we vary the ratio of R_2 and RS as in Fig. 3.7? In Fig. 3.8 this behavior in the two dimensional parameter space $(\tilde{\nu}_{m_a}, \tilde{\nu}_{m_B})$ is embedded into the state diagram of the toggle switch, since this depends on the same parameters (compare Fig. 3.6). We used $\nu_{m_A} = \nu_{m_B} = 5\,\mathrm{min}^{-1}$ for the maximal transcription rates in Eqs. (3.18) and (3.19) and varied the store signal for either R LOW ($\nu_{m_R} = 0.1\,\mathrm{min}^{-1}$, blue trace) or R HIGH ($\nu_{m_R} = 1\,\mathrm{min}^{-1}$, red trace) over the physiological range of $\nu_{m_S} = 0.01\,\mathrm{min}^{-1}$ to $\nu_{m_S} = 2\,\mathrm{min}^{-1}$.

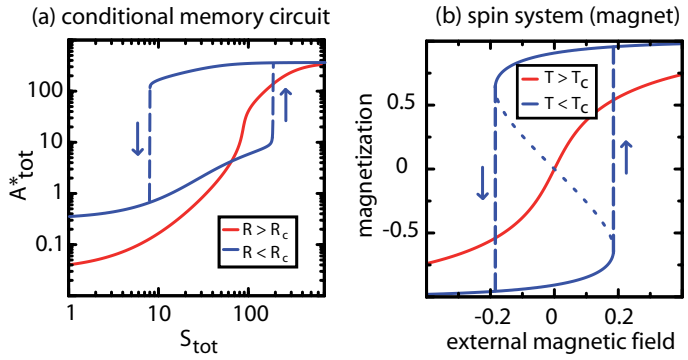
For a high level of the read signal (red trace) this variation of S has dramatic effects: at LOW S (upper left part of the curve) predominantly homodimers R_2 are formed and thus basically the transcription of A is repressed from its maximal rate of $\nu_{m_A} = 5\,\mathrm{min}^{-1}$ to $\tilde{\nu}_{m_A} \approx 0.2\,\mathrm{min}^{-1}$. This results in a shift of the circuit's stability to the monostable B region. If the amount of S is increased, the stability of the conditional memory circuit follows the red trace. Ultimately, if S is highly expressed (lower right part of the curve), RS is dominant over R_2 and the production rate of B is decreased from $\nu_{m_B} = 5\,\mathrm{min}^{-1}$ to $\tilde{\nu}_{m_B} \approx 0.3\,\mathrm{min}^{-1}$, thus leading to the monostable A state.

At a low level of the read signal (blue trace) the circuit is affected less drastically, since the maximal numbers of R_2 and RS are limited by the amount of total R. Therefore a variation of S over the *same* range as above can hardly force the circuit out of the bistable region. The implications of this will be discussed in the next section.

3.3.3 Switchable hysteresis

A key property of the conditional memory circuit is to be sensitive to the input S, if the read signal R is high, and to be insensitive to S when R is low. The behavior described above results in a hysteresis effect, which exists at low R, but is switched off when R is raised. Fig. 3.9 (a) demonstrates the switchable hysteresis in a plot of the steady-state concentration A^*_{tot} of gene A against S_{tot} for the same traces as in Fig. 3.8. The red curve shows the behavior at high R: Over the entire interval, gene A has a unique steady state, and the steady state concentration increases monotonically with S. Hence the state of gene A follows that of signal S.

At a low R level (blue curve), gene A has two steady state values, one low and one high, over most of the plotted range of S (bistable regime). When S is raised from a low value,



(c)

	control param.	external condition	state variable
magnet	T	external magn. field H	magnetization M
epigenetic switch	R	S_{tot}	A^*_{tot}

Figure 3.9: Comparison of the behavior of the conditional memory circuit with the thermodynamic behavior of a (single-domain) magnet.

gene A remains in the lower steady state and switches to the higher steady state only at the upper end of the bistable regime. Conversely, if S is lowered from maximal expression, A remains in the higher steady state until the lower end of the bistable regime is reached.

In the remainder of the study, we will use a LOW level of R that is still a factor 10 smaller than in the Figs. 3.8 and 3.9. Then the hysteresis loop extends to a much broader range, such that for physiological concentrations of S the bistability borders cannot be crossed.

Analogy to thermodynamical systems

The switchable hysteresis is similar to that of magnetic memory. Although the underlying physics of a magnet is fundamentally different from that of a genetic circuit, it is instructive to compare their behavior: Fig. 3.9 (b) shows the corresponding plot for a single-domain magnet, which was obtained from the rescaled state equation of a Weissian magnet

$$x = \tanh\left(\frac{y+x}{t}\right), \tag{3.20}$$

where x, y and t are proportional to the magnetization M, the external magnetic field B_0 and temperature T, respectively [100].

Figs. 3.9 (a) and (b) both show the external condition for the respective system on the vertical axis, the store signal strength S_{tot} for the conditional memory circuit and the external magnetic field for the magnet. The horizontal axis shows a state variable of the system: the magnetization for the magnet, which is the conjugate variable to the external field, and the level of protein A for the conditional memory circuit, which is the

state variable that is linked to the store signal S (the table in Fig. 3.9 summarizes the corresponding quantities of the two systems). Both plots show two curves with different values of a control parameter: for the magnet the control parameter is the temperature, while the level of protein R assumes the role of temperature in the conditional memory circuit. In Fig. 3.9 (b), the red curve shows the magnetization for a temperature above the critical temperature T_c (paramagnetic phase), where the magnetization directly follows the external field. The blue curve shows the well-known hysteresis below T_c (ferromagnetic phase), where the magnetization remembers its state: upon application of an external field in the opposite direction, it inverts its direction only when the field has at least a certain minimal strength. Clearly, the conditional memory circuit displays essentially the same behavior as magnetic memory, even though it is an dissipative dynamical system and does not have a well-defined Gibbs free energy. Here, we compare only the bulk behavior of the magnet with the deterministic behavior of the conditional memory circuit.

However, on a qualitative level, the similarity extends beyond this "mean-field" level: fluctuations in the biochemical reactions, which are prominent at low concentrations, lead to spontaneous switching between the two stable states in the bistable regime, and similarly fluctuations in the orientation of single spins can induce a spontaneous flip in the magnetization for small systems of few spins. More precisely, at external fields different from zero (but still small enough to be inside the hysteresis loop), the spin system has two minima in the Gibbs free energy, that are not symmetric: for external fields larger than zero the state with positive magnetization has a lower free energy than the negative magnetization. Thus on timescales much larger than an intrinsic relaxation timescale the spin system will flip from the *locally* stable to the *globally* stable state. This timescale depends on the potential barrier height separating the two minima and the relative magnitude of the fluctuations. Thus, for systems with only a low number of interacting spins, these fluctuations will become larger compared to the barrier height and the spontaneous switching rate increases.

The van der-Waals equation of state also exhibits a hysteresis loop for the volume of a real gas as a function of the pressure. However, in this system the relaxation time is so short, that the metastable branch of the hysteresis loop can hardly be observed - not even for macroscopic systems. Therefore the isotherms are better described by a *Maxwell-construction* in this phase coexistence region [100]. One assumes that the system is at *all values of p* in the globally stable potential minimum and switches to the other state only if the free energies of both states equal each other. This corresponds formally to a phase transition of first order between the gas and the liquid phase with a corresponding discontinuity of the first derivative of the free energy.

As mentioned before, the conditional memory circuit resembles the behavior of the magnet rather than the behavior of the van der-Waals gas. This is due to a spontaneous switching rate corresponding to a half-life time of the order of 40 hours and longer, as will be shown in Section 3.4.3. Nevertheless one could also make a Maxwell-type of construction for this dissipative system [69].

3.4 Dynamical behavior

3.4.1 Proof of principle

To perform its intended function, the conditional memory circuit must respond to time-dependent input signals if and only if the protein R is present, as illustrated in Fig. 3.1 (f). To verify this response within the quantitative model, we prepare the circuit in the ON state (A high) and then impose time-dependent transcription rates $\nu_{m_R}(t)$ for R and $\nu_{m_S}(t)$ for S with shapes as shown in Fig. 3.10 (a). The protocol of Fig. 3.10 (a) tests the complete set of fundamental circuit operations in the order: (i) remember S high, (ii) read S low, (iii) remember S low, and (iv) read S high. Fig. 3.10 (b) shows the response for both genes, A and B, within the deterministic description (dark green and red curve) as well as exemplary trajectories from a stochastic simulation (light green and orange curve, respectively). Qualitatively, the curves exhibit the desired behavior: From $t = 0-150$ min, when both R and S are in the low state (basal transcription rate of 0.01/min), gene A remains in the ON state. Shortly after the transcription of R is turned on at $t = 150$ min, the memory switches to the state of S, i.e. the OFF state. When the transcription of R is stopped at $t = 210$ min, gene A remains OFF, even after S is switched to the ON state at $t = 250$ min. The change in the state of S affects gene A only after the transcription of R is turned on again at $t = 300$ min.

However, after the validation of the proper functionality of the circuit, now its device physics shall be characterized. From Fig. 3.10 a lot of interesting questions arise and the following ones are addressed in the remainder of the chapter:

- How long must the read signal be presented until the store signal is stored? What determines this timescale?
- Some of the trajectories flip even in the absence of the read signal to the other state. What is the spontaneous switching rate for these events? Is the timescale of the induced *memory loss* sufficiently long compared to the time it takes to read a signal?
- Are there noise induced "toggle errors"? In Fig. 3.10 (c) and (d) it is visible, that after the end of the second read pulse at t = 260 min a considerable percentage does not reach the intended OFF state, even though the read signal was applied for 60 min. What is the chance to reach the desired state in dependence on the read signal length?
- What happens if the transcription rates of the control proteins shown in Fig. 3.10 (a) deviate from the chosen values? Can the sensitivity of the circuit be adjusted to different levels of external signal amplitudes?

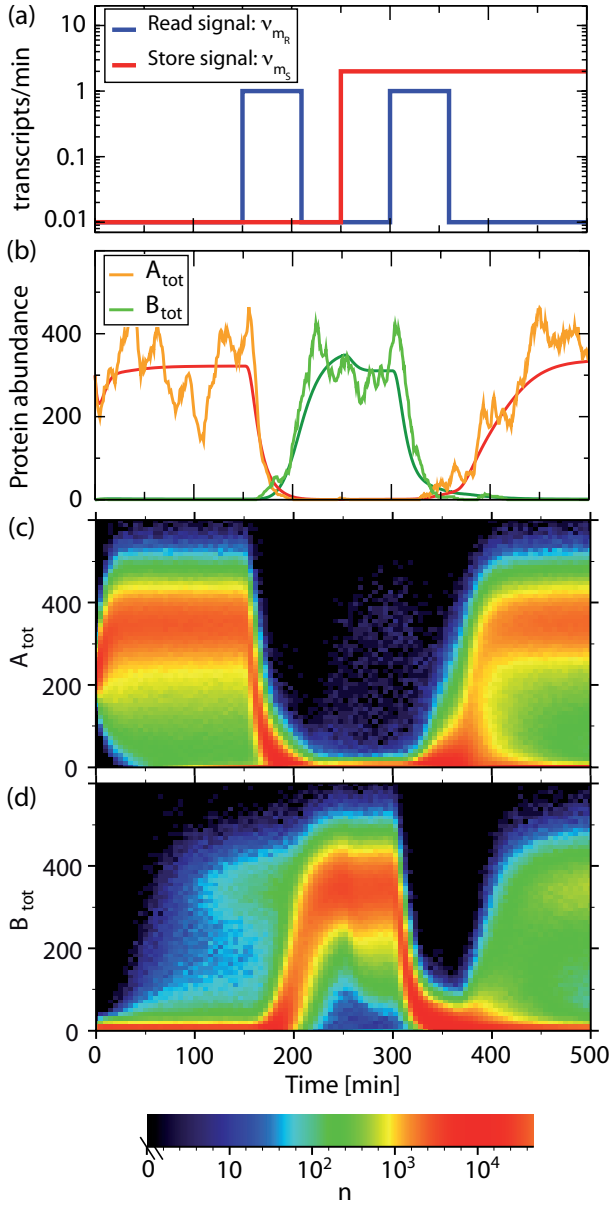


Figure 3.10: The time-dependent response of the conditional memory circuit to the input signals (a) is shown in (b-d). The dark red and dark green line in (b) show the total concentrations A_t and B_t calculated from the deterministic rate equations, whereas the light red and green lines are obtained from a single stochastic simulation run. The time evolution of the probability densities for A_t and B_t are shown in (c) and (d), respectively (the densities are estimated from 50000 stochastic simulation runs; the color codes for the observed number of trajectories inside each bin).

3.4.2 Switching timescale

Here we want to investigate the factors that determine the timescale of the conditional memory circuit. From Fig. 3.10 (b) we can get a rough idea of the characteristic time it takes to flip the switch: The duration from the onset of the read signal to the point where gene A and B have reached the same expression level is about 35 min when the switch is flipped from the ON to the OFF state and approximately one hour for the reverse direction.

The toggle time

In order to find a quantitative measure for this 'toggle time' one would naively look at the time it takes to reach and cross the separatrix. However, in our system it is hard to make such a generic definition because the separatrix as well as the fixed points of A and B are themselves a function of time (compare Section 3.1.2 on p. 44). If the read pulse is applied, the separatrix tilts until the control proteins reach the critical value at which the stable and unstable fixed point annihilate and only one attractive fixed point is left. Evidently, at this stage we cannot determine whether the switching event was successful or not, since the system is not in the bistable regime. Thus, we have to release the read pulse first and check after the decay of the control proteins, which state is reached. Here, the crucial point is that the control proteins themselves have a certain dynamical timescale that is not infinitely fast, but actually on the same order as the toggle switch.

We therefore define the toggle time T_t as the minimal read pulse duration that is needed to flip successfully between the states. This observable is of relevance for the experimental implementation since the read pulse duration is a controllable external parameter and also in the *in silico* model it can easily be determined by an Interval Bisection routine.

Now that we have a reasonable measure for the response time of the toggle switch, we want to investigate its determinants. Consider the switch to be in either one of the states, such that one of the concentrations is high and the other one is low. If the appropriate read pulse is applied (with either S high or low, opposite to the state of the switch), the production of the gene that was in the ON-state will be blocked. However, for the expression of the other gene to start, the proteins of the former ON state have to be degraded in order to clear the promoter of its antagonist. Therefore we expect the protein half-lives of A and B to be the major determinants of the toggle time. But also the production rates of A and B as well as the binding thresholds of O_A and O_B are expected to influence the toggle time substantially. The former defines the steady state protein level of the ON-state and therefore the initial value of the exponential decay (X_{HI}) and the latter determines the protein level below which the production of the new gene takes off (X_{LO}). With the protein half-life $T_{1/2}$ we expect a behavior of the following form

$$T_t \sim T_{1/2} \log_2 \left(\frac{X_{HI}}{X_{LO}} \right). \tag{3.21}$$

The logarithmic factor corresponds to the number of half-lives required for a exponential decay from X_{HI} to X_{LO}.

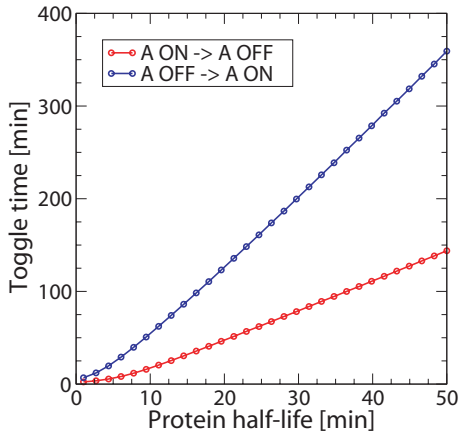


Figure 3.11: Toggle time as a function of the protein half-life for both switching directions. The connecting lines are meant as a guide to the eye.

Fig. 3.11 shows the toggle time as a function of the protein half-life[5] of A and B for both switching directions, while the half-lives of R and S were not altered. It increases for $T_t > 10\,\text{min}$ in both cases linearly with the protein half-life as expected. However, for short protein half-lives it saturates and does not approach zero. This behavior can be understood if we recall, that also the control proteins have a certain dynamics. It turns out, that the offset in the toggle time for small protein half-lives is due to the time it takes the control proteins to fully block the transcription of the active gene.

Another interesting aspect is the evident asymmetry between the both switching directions. For the half-life of 5 min used throughout this study, the toggle time for switching from the ON- to the OFF-state is only 10.5 min, whereas the switching into the other direction takes a read pulse duration of at least 36 min. We will investigate the origin of this asymmetry in the next section, where the full state space dynamics is explored.

Full state space dynamics

Up to now only the dynamics of the total proteins A_t and B_t (Fig. 3.10 (b)) was shown. But in order to fully understand the behavior of the circuit one has to scrutinize the dynamics of all the components of the state vector, as done in Fig. 3.12[6]. There, the 'information propagation' through the network for both switching directions is compared: the left column corresponds to the circuit's answer upon a pulse of R for a low store signal (switching from A ON to A OFF) and the right column shows the same for a high store signal (switching from A OFF to A ON). In both cases the desired switching is successful, but however, the answer upon a (R high, S low) pulse is considerably faster than the answer upon the other pulse. This asymmetry will be investigated in the following.

[5]The increased half-life was compensated by increasing the transcription rates such that the steady state levels remained constant.

[6]Not all different promoter states are shown, since they are not essential for the understanding of the message.

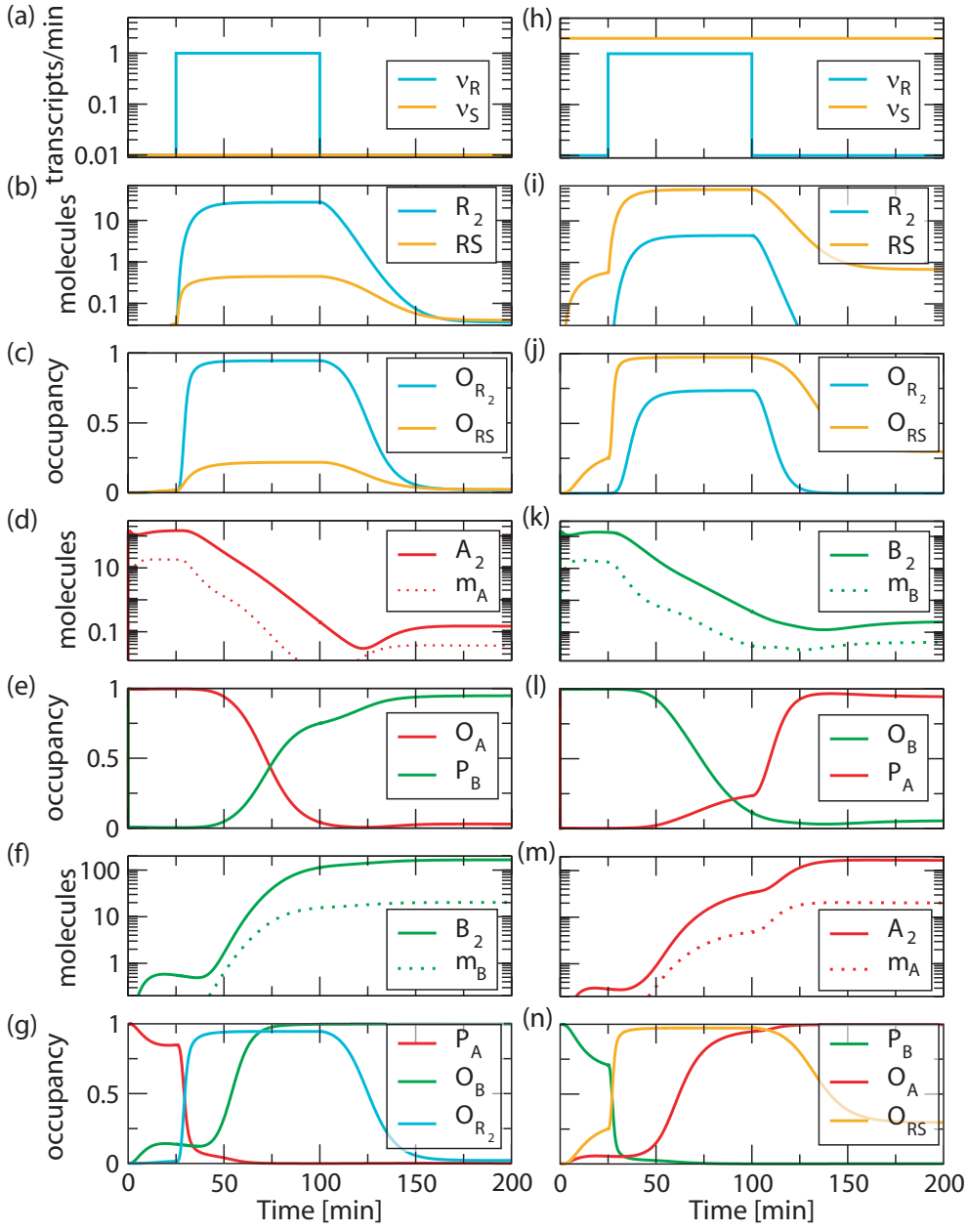


Figure 3.12: Information propagation through the network. The left column (a)-(g) shows the response upon a read pulse at basal store signal level $\nu_{m_S} = 0.01\,\mathrm{min}^{-1}$ (with initial condition A ON) whereas the right column (h)-(m) shows the same at high store signal level $\nu_{m_S} = 2\,\mathrm{min}^{-1}$ (with initial condition A OFF). For details see text.

Switching from ON to OFF. In this and in the next paragraph all figure labels (a),(b), etc. refer to Fig. 3.12. In (a) the time-dependent read and store signals are depicted and the resulting abundances of functional control protein dimers is shown in (b). Within about 10 minutes after the start of the read pulse at $t = 25$ min they approach their steady states and, subsequently also their respective operator sites become occupied. The occupancy of O_{R_2} of about 0.95 leads to a down-regulation of gene A's effective transcription rate to 5% of the unrepressed value and consequently, the transcript and protein abundance of gene A decrease exponentially to their new steady state levels (d). The kink observed in the m_A- curve will become clear soon.

As soon as the dimer abundance A_2 reaches the *apparent binding threshold*[7] for the TF-DNA binding $\hat{K}_A = 4.1nM$, the occupancy[8] of O_A drops below 0.5. At the same time the promoter activitiy P_B increases (e), allowing the synthesis of m_B and B_2 in (f). In (g) one observes, that additional to the occupancy of O_{R_2}, also the occupancy of O_B increases at $t \approx 50$ min. This is the negative feedback mechanism of gene B on gene A and it further decreases the promoter activity function P_A (g). Without this feedback, the m_A abundance in (d) would saturate at a level of about 5% of its maximal value (due to the impact of R_2), which would be at $m_A \approx 1$ transcript. Only the additional feedback of gene B further represses A's transcription. This is the origin of the kink in the m_A- curve in (d).

Switching from OFF to ON. For the reverse switching direction the time evolution of the corresponding state variables is shown in (h)-(n). In contrast to the other direction, one now observes in (i), that there is a substantial amount of R_2 dimers formed, although ideally only RS should be available (compare Section 3.1.2 on p. 44). Since the binding threshold for R_2 is lower ($K_{O_{R_2}} = 1.6$ nM) than the one for RS binding ($K_{O_{RS}} = 5$ nM) this amount of R_2 is sufficient to occupy O_{R_2} significantly, see (j). Nevertheless, the high occupancy of the proper operator site O_{RS} results in a shut-off of transcription of gene B as can be seen in (k). Although the subsequent drop in O_B occupancy is visible in (l), the promoter activity of gene A P_A does only increase slowly while the read pulse is present (until $t = 100$ min). This is due to the high and spurious occupation of O_{R_2} in (j). Consequently, the synthesis of gene A is hampered (m) and therefore the establishment of the negative feedback of gene A back to gene B gets impeded.

Thus, the reason for the asymmetry of the two switching directions can be directly related to the spurious overproduction of R_2 dimers, while the store signal is ON. This leads to a faulty repression of gene A's synthesis and as a consequence, the negative feedback of gene A on gene B is hampered.

[7]The apparent binding threshold is defined as the transcription factor concentration, at which the promoter activity function equals 50% of its maximal value. Note that in the case of independent operator binding (Eqs. (3.2) and (3.3) on p. 44) used here, the apparent binding threshold $\hat{K}$ is not equal to the equilibrium dissociation constant K, but is rather given by $\hat{K} = (\sqrt{2} - 1)K \approx 0.41K$.

[8]Here the occupancy of O_A is taken to be the sum of all promoter states, where an A_2 dimer is bound to P_B. The same definition is made for the occupancy of O_B.

3.4.3 Stochastically driven memory loss

In Section 3.3.3 we discussed in the context of magnetic memory the spontaneous switching of the magnetization in systems of only a few interacting spins. Similar events occur in the conditional memory circuit, as can be seen in Fig. 3.10 during the first 150 minutes: even though the read signal is at a basal level ($\nu_{m_R} = 0.01\,\mathrm{min}^{-1}$) a considerable fraction of trajectories flips to the OFF state. Interestingly, when the system is in the OFF state ($t = 210$min to $t = 300$min) only a very small percentage flips spontaneously to the ON state. Why do we observe noise induced toggle events for molecule numbers as large as about 350? What determines the spontaneous switching rates and what is their magnitude? Where does the observed asymmetry stem from?

Lifetime estimation via two state model

Before these questions will be addressed, it is first shown how the spontaneous switching rates are determined operationally. Theoretically, we are dealing with a stochastically driven 'first exit problem' and the task is to compute the probability of getting from each of the fixed points across the separatrix. This has been thoroughly investigated in the work of Aurell and Sneppen [16], where they studied the spontaneous switching of the λ-switch with a simplified Langevin-approach. Another recent study of Roma *et al.* [95] estimated the transition-rates in a reduced model of the toggle switch model based on an Eikonal approximation of the chemical master equation. However, both approaches rely on effective models using only two state variables and heuristical sources of noise. The full description of the conditional memory circuit by 24 state variables cannot be approached on this way.

Another possibility is to run many simulations for a long time and from the outcome one can *estimate* the probability of switching between the states. In the case of very rare switching events this brute-force procedure fails due to computational limitations: Reliable estimates of the rates require many events. Therefore extremely long simulation times are required for rare switching events. Although there have been successful attempts to simulate rare switching events very efficiently and accurately [10, 9], these algorithms were not employed here. It turns out, that the brute-force method yields a sufficient number of events for the switching rates under investigation.

Kramers reaction-rate theory. In order to develop an intuition for the stochastic switching between the two states of the conditional memory circuit, it is instructive to have a look at an analogous problem: In 1940 Kramers [59] developed a theory for chemical reactions based on the model depicted in Fig. 3.13. He considered a classical particle with mass M moving in an asymmetric double well potential U(y). y corresponds to the reaction coordinate and the respective minima of the potential denote the substrate and product states. In order to cross the potential barrier between both states a certain activation energy is required. This energy is provided by the coupling of the system to a heat bath of temperature T. The stochastic dynamics of the system is described by a Langevin equation

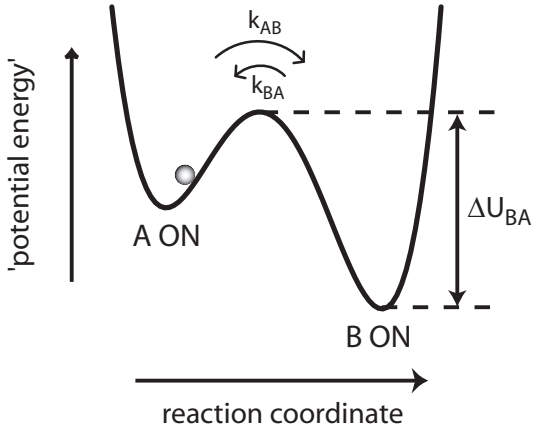


Figure 3.13: Illustration of the stochastically driven switching between the states A ON and B ON by a particle moving in an asymmetric double-well potential. This sketch serves only as a qualitative visualization. We can *not* define a potential of the flux for the conditional memory circuit.

with damping rate γ and a random force $\xi(t)$

$$M\frac{d^2y}{dt^2} = -\frac{dU}{dy} - \gamma M\frac{dy}{dt} + \xi(t). \tag{3.22}$$

The force $\xi(t)$ denotes Gaussian white noise with zero mean, which fulfills the fluctuation-dissipation theorem $\langle \xi(t)\xi(t')\rangle = 2M\gamma k_B T\delta(t-t')$. The Fokker-Planck equation corresponding to Eq. (3.22) can then be employed to calculate the stationary probability flux from one potential well to the other and this is in turn proportional to the Kramers rate for barrier crossing [59]. In the overdamped limit (large γ) the Kramers rate for getting from B to A writes

$$k_{BA}^{\text{Kramers}} = \frac{\omega_0\omega_b}{2\pi\gamma}e^{-\Delta U_{BA}/(k_BT)}, \tag{3.23}$$

where the ω are the negative curvatures of the potential, $\omega_0 = -M\frac{d^2U(y_0)}{dy^2}$ at the respective minimum and $\omega_b = -M\frac{d^2U(y_b)}{dy^2}$ at the barrier maximum. The pre-factor of the exponential is often called the *attempt frequency* and the exponential itself is referred to as the *Arrhenius factor*.

Dynamical two state model for lifetime estimation. Here a simple two state model is presented, which is used to estimate the rates of switching back and forth between the states from the simulated data. Although we cannot define a potential for the system[9] [70], we want to visualize the stability of each of the states by a heuristic 'free energy' as shown in Fig. 3.13. Let k_{AB} be the rate for hopping from A to B and k_{BA} the rate for

[9]Even for the simple description in Eqs. (3.1) on p. 44 the flux is not curl-free and thus no potential can be defined.

hopping from B to A. Then we define the *lifetime* τ_X of a state $X = A, B$ as the inverse rate of exiting it:

$$\tau_A := \frac{1}{k_{AB}}, \quad \tau_B := \frac{1}{k_{BA}}. \tag{3.24}$$

If we think of an ensemble of N realizations of our stochastic process, i.e. N different simulation runs, we can set up a rate equation for the average number of runs being e.g. in the A ON state $N_A(t)$:

$$\begin{aligned} \frac{d}{dt}N_A(t) &= k_{BA}N_B(t) - k_{AB}N_A(t) && (3.25)\\ N &= N_A(t) + N_B(t) = const && (3.26)\\ \Rightarrow \frac{d}{dt}N_A + (k_{BA} &+ k_{AB})N_A - k_{BA}N = 0\,. && (3.27) \end{aligned}$$

The solution is given by elementary integration

$$N_A(t) = N\frac{k_{BA}}{k_{BA}+k_{AB}} + ce^{-(k_{BA}+k_{AB})t} \tag{3.28}$$

with an integration constant c, which is determined by the initial conditions. If all runs are initialized e.g. in the A ON state, we have

$$\begin{aligned} N_A(t) &= N\left(\frac{k_{BA} + k_{AB}e^{-(k_{BA}+k_{AB})t}}{k_{BA}+k_{AB}}\right) && (3.29)\\ N_B(t) &= N - N_A(t)\,. && (3.30) \end{aligned}$$

Lifetime estimation. In order to estimate the lifetime of the states A ON and A OFF, 50000 simulation runs were initialized in the respective state. Analogous to Fig. 3.10 the time evolution of these trajectories was recorded and thereby the probability densities $n(A_{tot}, t)$ and $n(B_{tot}, t)$ of the total protein concentrations could be estimated as a function of time. In Fig. 3.14 the A ON state was initialized and the resulting probability density of gene A $n(A_{tot}, t)$ is shown over 280 hours. From this the time dependent number of simulation runs in the A OFF state could be defined as

$$N_B(t) = \int_0^{A_{sep}} n(A,t)dA\,, \tag{3.31}$$

where $A_{sep} = 75$ molecules is the approximate location of the high-dimensional separatrix projected onto the one dimensional A-axis. However, it turns out that the exact value of A_{sep} does not influence our results dramatically, because the vicinity of the separatrix is populated sparsely anyway. Finally, by fitting the theoretical prediction of Eq. (3.30) to the results from Eq. (3.31) one can determine the rates k_{AB} and k_{BA} and thereby the lifetimes τ_A and τ_B.

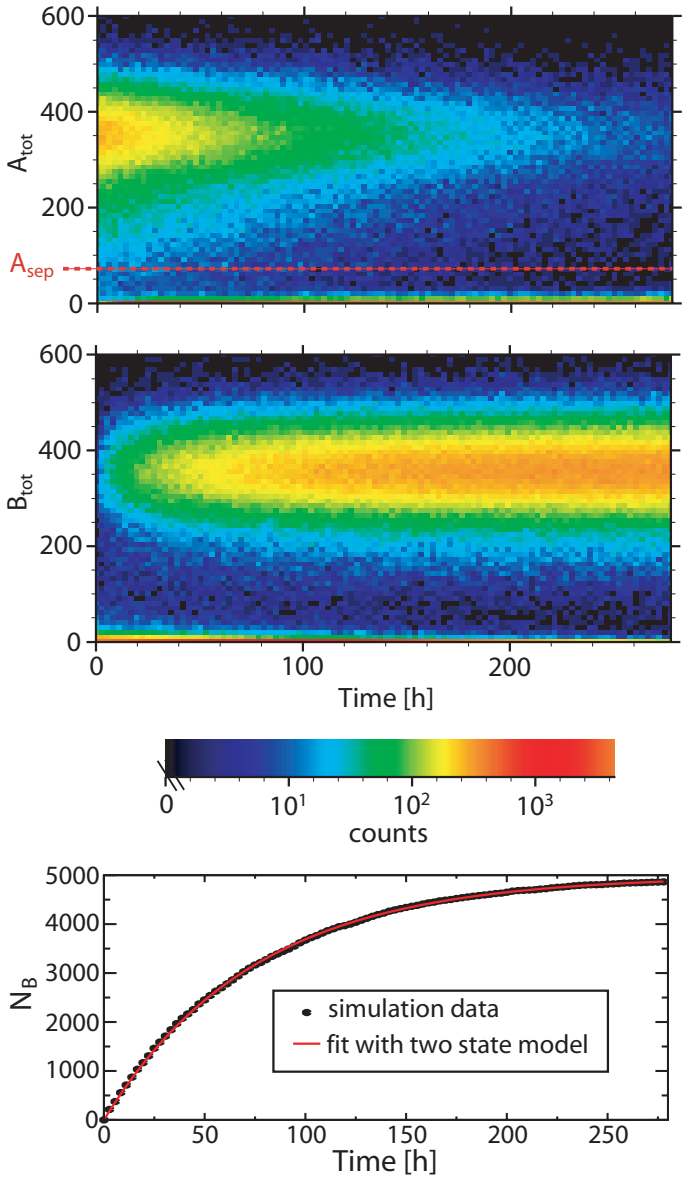


Figure 3.14: *Top:* The time dependent probability density of proteins A (estimated from $N = 5000$ simulation runs) is the basis of the lifetime estimation. The color codes for the number n of simulation runs at time t observed in an interval of $[A, A + \Delta A]$, $\Delta A = 5$ molecules. The number of trajectories in the (A OFF, B ON) state (which assumed to correspond to an interval of 0 to 75 proteins of A) was measured as a function of time (*bottom*) and fitted to the theoretical prediction of Eq. (3.30), see text.

τ_A	R absent	R basal
S absent	(73.01 ± 0.03) h	
S basal	(73.01 ± 0.03) h	(37.71 ± 0.015)h
S high	(73.01 ± 0.03) h	(100.67 ± 0.13)h

τ_B	R absent	R basal
S absent	$(2.27 \pm 0.05)\,10^4$h	
S basal	$(2.27 \pm 0.05)\,10^4$h	$(1.85 \pm 0.07)\,10^4$h
S high	$(2.27 \pm 0.05)\,10^4$h	$(614.3 \pm 1.6)\,h$

Table 3.3: *Top:* lifetime of state A ON under different amounts of the control proteins; *Bottom:* the same for state A OFF; errors were obtained from the asymptotic standard errors of the parameter fits: for the state A ON the relative errors were smaller than 10^{-3}. The basal values of R and S correspond to transcription rates of $\nu_{m_R} = \nu_{m_S} = 0.01/min$ and the S HIGH value to $\nu_{m_S} = 2/min$.

The lifetimes depend on the basal level of the control proteins

Table 3.3 summarizes the results of the described procedure for different values of the control proteins. We make three findings: first, in the absence of the read signal the state A OFF is about 2 orders of magnitude more stable than the state A ON. This is based on the asymmetry in the parameters: LacI (protein B) has a slightly higher dimerization affinity (equilibrium biased to the dimer form) as well as a higher operator affinity (more dimers bound to the operator) than TetR. At equal transcription rates of both proteins their average levels are equally high, but due to the mentioned differences in their dimerization- and binding affinities the repression of TetR by LacI is stronger than the repression of LacI by TetR. Therefore the state with LacI ON (B ON), is more stable than the other state. In other words the heuristic 'potential energy' has two minima of different depth, corresponding to one globally stable fixed point (A OFF) and one locally stable fixed point (A ON), as depicted in Fig. 3.13.

The second observation is, that the lifetime of both states depends on the control protein abundance, as summarized in Table 3.3. The addition of basal levels of R and S ($\nu_{m_R} = \nu_{m_S} = 0.01/min$) destabilizes state A and stabilizes state B. If we recall that the control proteins tilt the separatrix towards one of the axes, it is clear that the probability of one of the states to cross the separatrix increases, while the probability of the other one decreases. This qualitative description also explains the increase of state A's lifetime if we add R at basal and S at HIGH amounts ($\nu_{m_R} = 0.01/min,\ \nu_{m_S} = 2/min$): the separatrix is tilted towards the B axis and the probability of state A to cross the barrier decreases. In the 'free energy' analogy the addition of different amounts of control proteins distorts the energy landscape, such that the depth of the minima and the height of the barrier can be regulated. Note that the control proteins are at basal synthesis rates themselves random

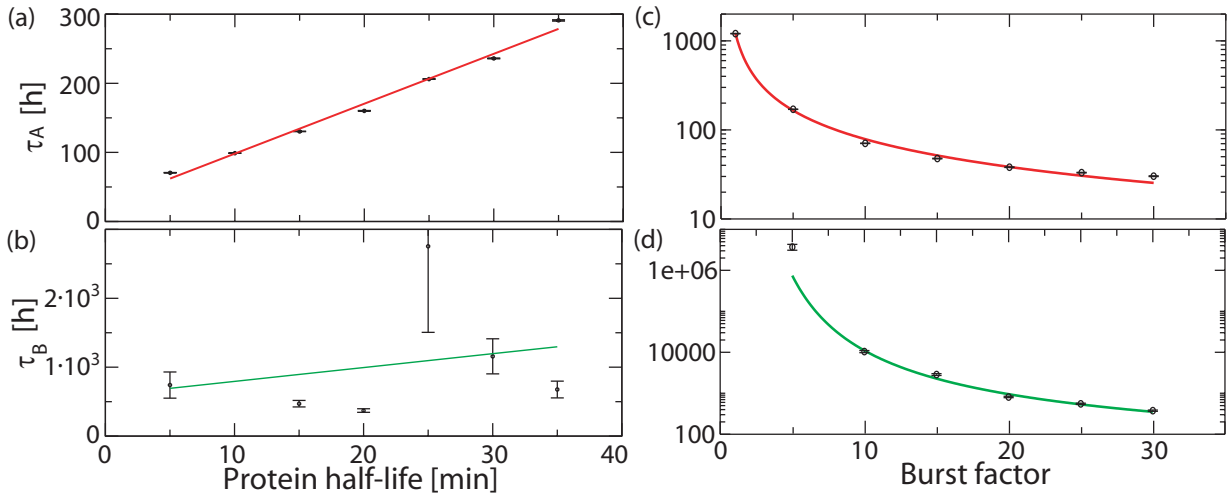


Figure 3.15: The lifetime of the states A ON (a) and B ON (b) are shown as a function of the protein half-life (the degradation rates of proteins A and B were altered simultaneously while their translation rates were adjusted to match a constant burst factor of 10 [recall: $b = \frac{\nu_p}{\lambda_m}$]). The lifetime of the state A ON (c) and B ON (d) as a function of the burst factor. The mRNA half-life was kept constant (3 min) while the translation rate ν_p was altered in order to change the burst factor. At the same time we adjusted the transcription rate ν_{m_i} in order to keep the average protein level, which is proportional to $\frac{\nu_{m_i}\nu_p}{\lambda_m} = \nu_{m_i} \cdot b$, at a constant value.

variables and thus they generate a heavily fluctuating potential-landscape.

The third and concluding remark is that, although we find a certain asymmetry in the lifetimes, even the shortest lifetime of around 40 hours is sufficiently long to fulfill the requirements of persistent memory.

Influence of the shape of noise

In addition to the 'potential-landscape', also the *size* and the *frequency* of the random events are crucial for the stochastically driven barrier crossing [118].

Burst factor. In the analogy to Kramers' model, the size of the random events corresponds to the temperature of the heat bath. Since the major determinant of the size of the random events is expected to be the burst factor b, a high temperature would correspond in this picture to a genetic circuit with a bursty protein synthesis. Recall that the burst factor b is defined as the average number of proteins synthesized from one mRNA. The simulated dependence of the lifetime on b is shown in Fig. 3.15 (c) and (d) for both states. We observe a monotonic decrease of the lifetimes with increasing burst factor. This is analogous to increasing barrier crossing rates at high temperatures. In order to verify the correspondence between burst factor and temperature, it was checked whether the dependence of the Kramers rate on temperature could predict the dependence of the lifetimes on the burst factor. Since the fluctuation-dissipation theorem imposes a dependence of the

friction coefficient on temperature, $\gamma \sim T^{-1}$, the Kramers rate suggests a dependence of the lifetime on the burst factor b of the following form

$$\tau(b) = \frac{\alpha e^{\beta/b}}{b}, \tag{3.32}$$

with the fit-parameters α and β. Fits of Eq. (3.32) to the data are shown in Fig. 3.15 (c) and (d) and the agreement seems plausible . However, this result should not be over-interpreted and serves mainly as a rough explanation of the behavior shown in Fig. 3.15 (c) and (d).

Protein degradation rate. As we saw in Section 3.4.2, the timescale of the circuit to react to environmental changes in the parameters, which are here the transcription rates of R and S, is determined by the protein half-life. This suggests that the timescale of the spontaneous fluctuations is determined by the same quantity as well. It is therefore reasonable to assume, that the attempt frequency entering the Kramers rate in Eq. (3.23) is proportional to the degradation rate. Thus the lifetime, which is the inverse Kramers rate, is expected to depend linearly on the protein half-life given by the inverse degradation rate.

Indeed, as shown in Fig. 3.15 (a), this is precisely the case[10]: The longer the half-life is, the slower is the timescale of the fluctuations and thus less attempts to cross the potential barrier are made. Fig. 3.15 (b) shows the same for state A OFF and although the relative errors are immense[11], the general linear trend of increasing lifetimes can be verified as well. This effect of the protein half-life on the frequency of the fluctuations was confirmed recently by single-cell fluorescence microscopy [17, 35]: an increased degradation rate leads to a shift in the power spectrum to higher frequencies.

In summary, the lifetimes of states A ON and A OFF are significantly longer than the the time it takes to address the memory (given by the toggle time). The asymmetry in the lifetimes stems from the intrinsic parameter asymmetries of TetR and LacI. Additionally the basal levels of R and S bias the 'potential-landscape' such that the stability of the states can be altered. Furthermore higher magnitudes and frequencies of the fluctuations destabilize both states, thus leading to an increased stochastically driven barrier crossing. These quantities are accessible by tuning the burst factor and the protein degradation rates, respectively.

While the spontaneous switching has been characterized in different types of toggle switches before, we now want to investigate a new feature of the conditional memory circuit.

[10]The transcription rates were adjusted such that the average protein abundance remained constant.

[11]These huge errors stem from the fact, that the extremely low switching rates generate in these cases only a low number of switching events. However, with a lifetime of more that 10^4 hours, the precise values do not matter for the current study, since from an experimental point of view, they are close to infinity.

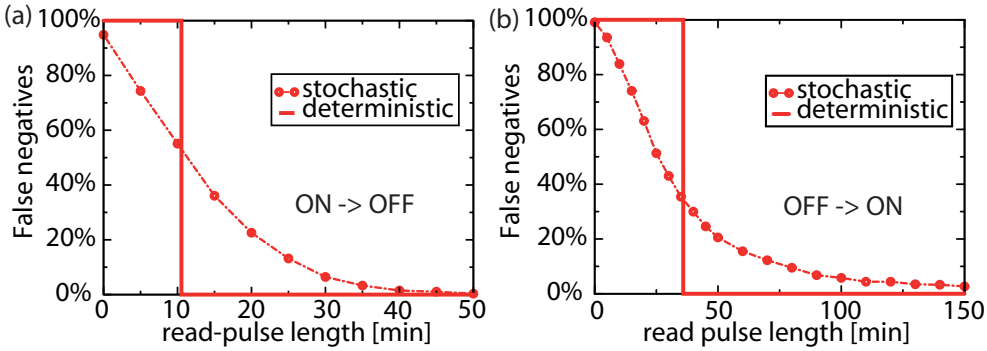


Figure 3.16: The false negative error fraction as a function of the read pulse duration for switching (a) into the OFF state and (b) into the ON state. The circles indicate data from stochastic simulations and the solid lines indicate the toggle time, which is the minimum read pulse duration for switching in the absence of noise. The interpolating dashed lines are only guides to the eye.

3.4.4 Noise induced toggle errors

The new aspect of the conditional memory circuit is that biochemical noise leads to two additional types of noise-induced errors: During a read pulse (high R), the switch may not flip even though it is triggered to do so (false negative), or the switch may flip, even though it was already in the correct state (false positive).

False negatives

False negative errors are visible in Figs. 3.10 (c) and (d) on p. 61 at $t \approx 400$ min right after the second read pulse, where a certain fraction of the A_{tot} density remains in the low state, while the same fraction of the B_{tot} density erroneously ends up in the high state. In contrast, for the inverse switching direction (after the first read pulse), we observe hardly any false negatives. Fig. 3.16 shows the fraction of false negatives as a function of the read pulse duration for both switching directions[12]. We observe that the error fraction decreases rapidly with increasing read pulse duration. For long read pulses, it drops below 0.5 % when switching OFF and below 4 % when switching ON. Note however that the timescales of Fig. 3.16 (a) and (b) differ by a factor of 3. This asymmetry is equally reflected in the deterministic 'toggle times' (solid lines in Fig. 3.16), as discussed in Section 3.4.2.

The origin of the asymmetry in the false negatives is again the asymmetry of the regulatory front end: The spurious overproduction of R_2 during the presence of the store signal leads to a repression of gene A, which in turn only slowly approaches the value required for a stable establishment of the negative feedback loop. Thus the switching from A OFF to A ON (driven by this 'imperfect' read pulse) demands a much longer read pulse

[12] We allow for a relaxation time of 60 min after the end of the read pulse and then determine the error fraction. Since the rate of spontaneous flipping is very low, the result depends only very weakly on the precise value of the relaxation time, provided it is not too short.

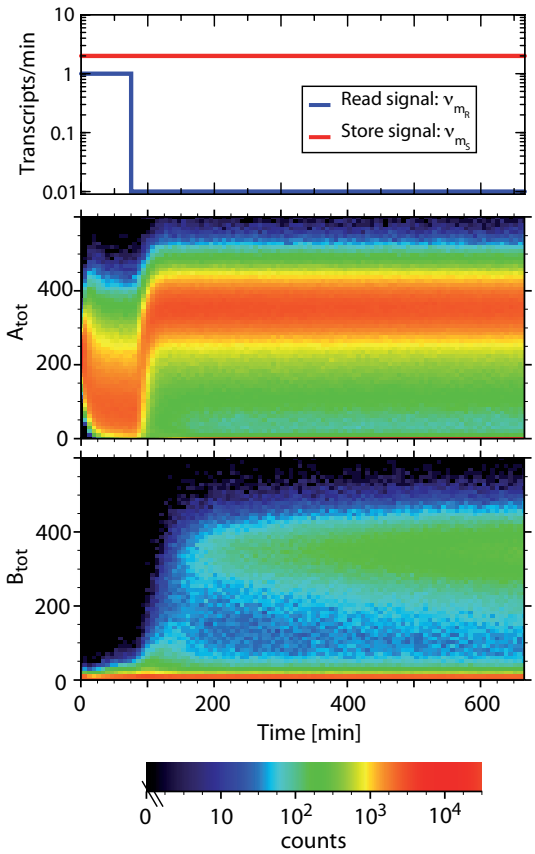


Figure 3.17: Estimation of the false positives for a read pulse duration of 75 minutes. Even though the read pulse is applied while S is HI (a) and all trajectories are initially in the correct state, some of the them flip after the end of the read pulse erroneously to the state A OFF. The densities were estimated from 50000 repeated simulation runs and the color codes for the number of trajectories in each bin.

duration than the reverse direction for the same number of false negatives.

False positives

The protocol depicted in Fig. 3.10 (a) did not test for false positives: The read pulse was only applied, when the present store signal was contrary to the state of the system, i.e. R was turned ON when A was high AND S was low (at $t = 150$ min) or when A was low AND S was high (at $t = 300$ min). Therefore another protocol had to be used to quantify the percentage of false positives. As an example the estimation for the A ON state is shown in Fig. 3.17: The circuit was prepared in the ON state and while the store signal S was ON as well, a read pulse was applied for a defined duration. The false positives were measured 60 minutes after the end of the read signal, analogously to the false negatives.

For the A ON state the dependency of the false positives on the read pulse duration is depicted in Fig. 3.18. With increasing read pulse duration the fraction of false positives

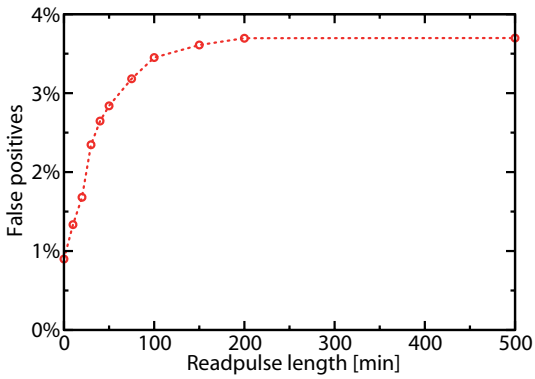


Figure 3.18: False positives as a function of the read pulse duration T_R for the state A ON. Even though the read pulse is applied while S is high (and A is supposed to be switched ON), for long pulses approximately 4 % of the trajectories switch to the A OFF state. The results were generated by 50000 simulation runs, all initialized in the state A ON, with a read pulse of $\nu_R = 1/\text{min}$ applied from $t = 0\text{min}$ to $t = T_R$ (compare Fig. 3.17).

increases until it eventually saturates at about 3.7 % and a read pulse length of 150 min. For the reverse direction the fraction of false positives is very small (< 0.1 %) and therefore not shown here. The explanation for the false positives is again given by the asymmetric front end: During the presence of the read pulse (0-75 min) there is a small amount of spuriously produced R_2 homodimers although ideally only heterodimers RS should be abundant and these R_2 dimers repress the transcription of A. Therefore one observes a down-regulation of the density of A in Fig. 3.17 and the lower the amount of A at the end of the read signal becomes, the more likely it is to relax to the state A OFF instead of A ON. This explains the time dependence of the false positives in Fig. 3.18: The longer the read pulse is applied, the lower drops the average abundance of A. Therefore the false positives increase with read pulse duration. They saturate at a level, that is determined by the steady state distribution of A_{tot} in the presence of the read signal. This distribution is in turn determined by the amount of overproduced R_2.

Note that the resulting curve in Fig. 3.18 does not approach zero for a read pulse length of 0 minutes, since there is a noticeable amount of spontaneous flips during the relaxation time of 60 min after the end of the read pulse. However, this systematic overestimation of the false positives is approximately the same for all measured read pulse durations and thus the general trend of a saturation at $T_R \approx 150\,\text{min}$ is not affected by this.

For the reverse direction (A OFF $\rightarrow$ A ON) under the control scenario (R HI, S LO) almost no false positives ($< 0.1\%$) could be detected. This makes perfect sense, since in this case no spurious formation of RS is possible and thus gene B cannot be suppressed.

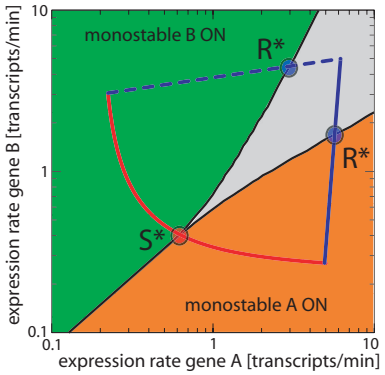


Figure 3.19: The set points S^* and R^* are defined as the threshold concentrations of R and S, at which the toggle switch is driven to an other 'phase'. For R there exist two qualitatively distinct set points: when R is increased while S is LOW (dashed blue curve) the toggle switch is driven into the A OFF state and similarly, when R is increased while S is HIGH, the toggle switch is set to the A ON state.

3.5 Adaptation of the circuit's sensitivity to the environment

Above, the duration of the read signal was varied, while the concentrations for the high and low levels of the R and S proteins were assumed to be given. However, when the conditional memory circuit is embedded into the cellular environment, it must be adjustable to work with a variety of input signals, the level of which depends on the specific context: in one situation an S concentration of 50 molecules per cell might correspond to the ON state of a signaling process, while in another situation this could be the basal level in the OFF state.

Set points. For a given set of circuit parameters, there exists a certain threshold concentration (or "set point") for S, below which the memory flips to the OFF state and above which it flips to the ON state when a read signal is given. Similarly, there is a set point for the read signal, above which the circuit reads the input and below which it ignores the input. These set points are depicted in Fig. 3.19.

For the circuit design to be versatile, the set points must be *programmable*, so that they can be adjusted to lie between the typical high and low levels of S and R, respectively. For the proposed circuit design this can be achieved by exploiting the programmability of operator binding affinities through simple changes in their nucleotide sequence [47]: the response of the toggle switch of Fig. 3.3 (top) on p. 47 to the regulatory front end (bottom) critically depends on the binding thresholds of the R_2 and RS binding sites. Altering these binding thresholds corresponds to an effective increase or decrease of the R_2 and RS abundances. Therefore these mutations lead to shifts of the red and blue "control curves" in the state diagram in Fig. 3.19. If for instance the affinity of the R_2 binding site O_{R_2} was increased, less R_2 would be sufficient to generate the same degree of repression of gene A as before. Thus, the curves depicted in Fig. 3.19 would be collectively shifted

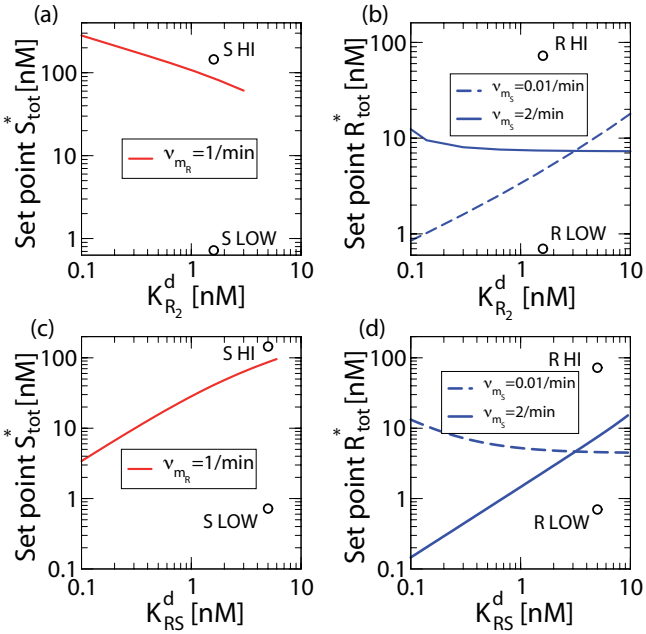


Figure 3.20: Tuning the set points as defined in Fig. 3.19 by mutations in the R_2 (a,b) and RS (c,d) binding sites. Similarly, there is a set point R^* for the read signal, above which the circuit will read the input and below which it will ignore the input. There exist two set points R^*: one for reading S HI (blue solid lines) and one for reading S LO (blue dashed lines).

to the left. Evidently, these shifts affect the set points, since the intersection points with the phase borders will vary. An alternative approach to adapt the circuits sensitivity to environmental signal amplitudes would be the adjustment of the underlying phase diagram in Fig. 3.19. Clearly this also leads to shifts in the set points, as now the phase borders are shifted themselves. In the following we will merely investigate the effect of the R_2 and RS binding sites, keeping in mind that the other option exists as well.

In Fig. 3.20 (a) and (b) the binding threshold for the R_2 binding site is varied from 0.1 nM to 10 nM and we observe, that all set points can be tuned over 1-2 orders of magnitude. Variation of the binding threshold for the RS binding site in (c) and (d) leads to similar results. Note, that in both cases one of the R^* set points is almost unaffected by the mutation. This is due to the fact, that e.g. in (b) the mutation of the R_2 binding site has no effect on the binding of RS and thus the case where S is high and thereby mainly RS is formed (solid line) is not affected. The same (with exchanged roles of R_2 and RS) holds for the opposite case in(d), where S LO is read (dashed line). The two large dots in all plots indicate the high and low level of the total concentrations of R or S (ordinate) and the binding thresholds (abscissa) used for all simulations reported here. For proper functionality of the circuit, the setpoints must always be adjusted to lie between these high and low levels of R and S, which are usually determined by the environment. Thus, one expects that the circuit can easily be adapted to work under a wide range of conditions.

3.6 Stochastic average vs. deterministic result

From Fig. 3.10 (b) on p. 61 one gets the impression that the deterministic reaction rate equations (RREs) mirror the average of the stochastic simulations. But already the mere view at the probability densities of the total protein abundances in Fig. 3.10 (c) and (d) suggests, that this is not the case.

In Fig. 3.21 the results of the RREs are plotted together with the averages of the time dependent probability densities of A_{tot} and B_{tot}, $\langle A_{\mathrm{tot}} \rangle$ and $\langle B_{\mathrm{tot}} \rangle$. It is clearly visible, that already during the first 150 min, where no read signal is present, the spontaneous switching leads to significant deviations between both results. The divergence becomes much more pronounced after the end of the second read pulse ($t = 360$ min), when many false negatives are found.

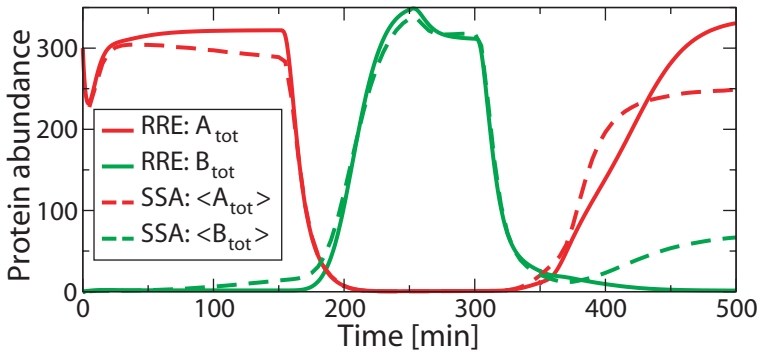


Figure 3.21: Comparison of the results obtained by the deterministic reaction rate equations (solid lines) with the averages of the stochastic simulations (dashed lines). The control protocol as well as the probability densities used for the averaging are the same as in Fig. 3.10 on p. 61.

This reveals what was already mentioned briefly in Section 2.4.3: as soon as we are dealing with nonlinear chemical reaction systems, the assumptions leading to the RREs loose their validity. Especially when the involved molecule numbers become low or the system gets close to a branching point (as is the case during the second read pulse) the differences are most significant.

Nevertheless we saw in the previous sections, that the RREs prove to be very useful in many circumstances and should be considered as the first approach to a new system. Deterministic stability analysis as well as phase space dynamics yield the indispensable guidance necessary for the examination of stochastic phenomena.

3.7 Model reduction

In this section the connection between the full model, consisting of 24 state variables, and the reduced description in Eqs. (3.1) shall be elucidated. First, the relationship between the dynamical model of transcriptional regulation by six distinct promoter states and the equilibrium promoter activity function in Eqs. (3.2) and (3.3) is derived. Then it is shown which additional equilibrium assumptions have to be made in order to further reduce the degrees of freedom. Last, the steady state characteristics and the dynamics of the full and the reduced model are juxtaposed.

3.7.1 Adiabatic approximation and promoter activity function

In the current approach the dynamics of transcriptional regulation is modeled by six distinct promoter states (compare Fig. 3.4 on p.48), which is described by the RREs in Eqs. (A.89) - (A.101). In order to compare this with the promoter activity function (PAF) in Eqs. (3.2) and (3.3), one assumes the binding and unbinding reactions to be in equilibrium for any given number of TFs (*adiabatic approximation*, [45, 114]) and calculates the fraction of free promoter sites. This fraction corresponds to the equilibrium promoter activity function. However, calculating the equilibrium of Eqs. (A.89) - (A.101) results in expressions, which extend to three MATHEMATICA notebook lines. Much more concise expressions can be derived with with an analogous approach, as shown in the following. Plotting both solutions yields identical results.

In order to derive an expression for the probability of the promoter being in an unoccupied state, we consider the following: The three operator sites on each of the promoters can be described by their occupancy number n_1, n_2, n_3. For the remainder of this section we will focus on the promoter P_A, since the problem is symmetric in A and B. n_1 is the occupancy for O_{B_1}, n_2 for O_{B_2} and n_3 for O_{R_2}. The probability for a single operator being *unoccupied* is $(1 - n_i)$ and therefore the probability for the entire promoter being unoccupied writes

$$P_A = \prod_{i=1}^{3} (1 - n_i) \tag{3.33}$$

Now we calculate the occupancy numbers of the operators. Assuming that the total abundance of the two species of TFs (e.g. B_2 and RS) in the cell are constant, the molecule number conservation writes

$$B_2 = n_{B_{2_{cell}}} + n_1 + n_2 \tag{3.34}$$
$$R_2 = n_{R_{2_{cell}}} + n_3, \tag{3.35}$$

where the species with the *cell* index correspond to the number of dimers in the cellular cytoplasm, i.e. not bound to the operators. The dynamical behavior is under the assumption of identical operator sites O_{B_1} and O_{B_2} with $k^+_{O_{B_1}} = k^+_{O_{B_2}} = k^+_{O_B}$ and $k^-_{O_{B_1}} = k^-_{O_{B_2}} = k^-_{O_B}$

described by 4 ODE's

$$\begin{aligned}
\partial_t n_1 &= k^+_{O_B} n_{B_{2_{cell}}}(1-n_1) - k^-_{O_B} n_1 && (3.36)\\
\partial_t n_2 &= k^+_{O_B} n_{B_{2_{cell}}}(1-n_2) - k^-_{O_B} n_2 && (3.37)\\
\partial_t n_3 &= k^+_{O_{R_2}} n_{R_{2_{cell}}}(1-n_3) - k^-_{O_{R_2}} n_3 && (3.38)\\
\partial_t n_{B_{2_{cell}}} &= k^-_B(n_1+n_2) - k^+_B n_{B_{2_{cell}}}(2-n_1-n_2). && (3.39)
\end{aligned}$$

With the molecule number conservation we find for the steady state solution, that $n_1 = n_2 = n$ and

$$\begin{aligned}
n &= \frac{1}{4}\left(2 + K_{O_B} + B_2 - \sqrt{(2+K_{O_B}+B_2)^2 - 8\,B_2}\right) && (3.40)\\
n_3 &= \frac{1}{2}\left(1 + K_{O_{R_2}} + R_2 - \sqrt{(1+K_{O_{R_2}}+R_2)^2 - 4\,R_2}\right). && (3.41)
\end{aligned}$$

One should not be confused by the dimensions of the single terms in the solution, since the total number of each operator site is set to 1 and its explicit dimension was suppressed. The probability for an unoccupied promoter thus reads

$$\begin{aligned}
P_A(B_2,R_2) &= \frac{1}{32}\left(2 - K_{O_B} - B_2 + \sqrt{(2+K_{O_B}+B_2)^2 - 8\,B_2}\right)^2 && (3.42)\\
&\quad \times \left(1 - K_{O_{R_2}} - R_2 + \sqrt{(1+K_{O_{R_2}}+R_2)^2 - 4\,R_2}\right)
\end{aligned}$$

and similarly for P_B

$$\begin{aligned}
P_B(A_2,RS) &= \frac{1}{32}\left(2 - K_{O_A} - A_2 + \sqrt{(2+K_{O_A}+A_2)^2 - 8\,A_2}\right)^2 && (3.43)\\
&\quad \times \left(1 - K_{O_{RS}} - RS + \sqrt{(1+K_{O_{RS}}+RS)^2 - 4\,RS}\right).
\end{aligned}$$

This form of the promoter activity function looks slightly more complicated than the previously used expressions (Eqs. (3.2) and (3.3)), which were based on purely thermodynamical considerations

$$\begin{aligned}
P_A(B_2,R_2) &= \left(1 + \frac{B_2}{K_{O_B}}\right)^{-2}\left(1 + \frac{R_2}{K_{O_{R_2}}}\right)^{-1}\\
P_B(A_2,RS) &= \left(1 + \frac{A_2}{K_{O_A}}\right)^{-2}\left(1 + \frac{RS}{K_{O_{RS}}}\right)^{-1}.
\end{aligned}$$

The difference between these functions stems from the fact, that the former explicitly takes into account, that the binding of one TF to an operator site reduces the number of TF's in solution. Therefore the binding of a second TF while one is bound already is less likely than in the case, where the TFs in solution is kept constant. This is explicitly assumed in the derivation of the latter expression. This effect is most pronounced if the TF number is small (data not shown). In Section 3.7.3 the circuit behavior using these two different forms of the PAF is compared.

3.7.2 Rapid equilibrium assumptions

Similar to the adiabatic approximation for the TF-DNA binding one can also assume most of the other cellular processes as equilibrated. The remaining *slow degrees of freedom* are the total protein abundances $A_{\rm tot}$ and $B_{\rm tot}$ [30]. Such an approximation is only valid, if the timescales really separate, and as we will see in the next section this is not completely the case for the parameters under investigation.

First, we focus on the toggle switch and do not explicitly consider the time evolution of the control proteins R and S. Let's consider R_2 and RS as external control parameters of the toggle switch, the levels of which are governed by Eqs. (3.17). The time evolution of the total protein number, here only shown exemplarily for $A_{\rm tot}$, is given by

$$\begin{aligned}\frac{d}{dt}A_{\rm tot} &= \frac{d}{dt}\big(A + 2\,A_2 + 2\,P_B \bullet A_2 + 2\,P_B \bullet A_2 \bullet RS \\ &\qquad +4\,P_B \bullet A_2 \bullet A_2 + 4\,P_B \bullet A_2 \bullet A_2 \bullet RS\big) \\ &= \nu_A m_A - \lambda_A A - 2\lambda_{A_2} A_2\ . \end{aligned} \tag{3.44}$$

By assuming to have no cooperative stability ($\lambda_A = \lambda_{A_2} = \lambda_p$) and by neglecting the TFs bound to their operator sites in the limit of large protein abundances ($A_{\rm tot} \approx A + 2A_2$) we can write

$$\frac{d}{dt}A_{\rm tot} = \nu_p m_A - \lambda_p A_{\rm tot}\ .$$

If further the relaxation time of mRNA is much faster than the one of the proteins ($\lambda_m >> \lambda_p$), mRNA synthesis and decay can be considered to be equilibrated for any TF concentration and we have, now for both genes:

$$\frac{d}{dt}A_{\rm tot} = \frac{\nu_A \nu_{m_A}}{\lambda_{m_A}} P_A(B_2, R_2) - \lambda_A A_{\rm tot} \tag{3.45}$$

$$\frac{d}{dt}B_{\rm tot} = \frac{\nu_B \nu_{m_B}}{\lambda_{m_B}} P_B(A_2, RS) - \lambda_B B_{\rm tot}\ . \tag{3.46}$$

The only remaining task is to express the dimers in the promoter activity function by the total protein concentrations. This is achieved by assuming that dimerization equilibrates prior to synthesis and degradation and the result is

$$A_2 = \frac{A_{tot}}{2} + \frac{\tilde{K}_A - \sqrt{\tilde{K}_A^2 + 8\,\tilde{K}_A\,A_{tot}}}{8} \tag{3.47}$$

$$B_2 = \frac{B_{tot}}{2} + \frac{\tilde{K}_B - \sqrt{\tilde{K}_B^2 + 8\,\tilde{K}_B\,B_{tot}}}{8}\ , \tag{3.48}$$

with the *in vivo* dimerization constants $\tilde{K}_A = \frac{k_A^- + \lambda_{A_2}}{k_A^+}$ and $\tilde{K}_B = \frac{k_B^- + \lambda_{B_2}}{k_B^+}$ [29]. Together with the steady state relations for the control dimers R_2 and RS of Eqs. (3.17) the rate equations in Eqs. (3.45) and (3.46) can be closed and related to the input signals ν_{m_R} and ν_{m_S}.

3.7.3 Comparison of the full with the reduced model

How do the assumptions made in the last two sections influence the steady state and dynamical behavior of the models? Here we juxtapose

1. the full model,
2. the reduced model in Eqs. (3.1) on p. 44 with the 'exact' promoter activity function in Eqs. (3.42) and (3.43) on p. 80 and
3. the reduced model with the simple promoter activity function in Eqs. (3.2) and (3.3).

Fig. 3.22 (a) displays the phase borders of the state diagram similar to Fig. 3.6 for all three models. Here only the model 3) shows minor deviations of the phase border between bistable and monostable B regime. The good agreement can easily be understood, if we recall that the assumptions leading to the reduced models 2) and 3) were primarily steady state assumptions. Therefore it is just reasonable that their steady state behavior is close to identical.

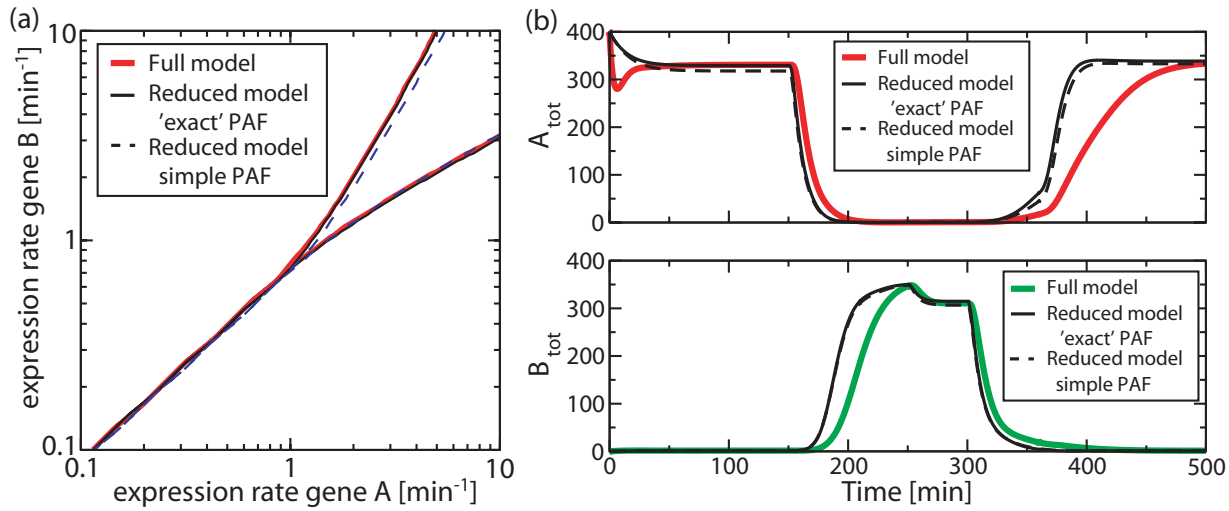


Figure 3.22: In (a) the phase borders of the state diagram similar to Fig. 3.6 on p. 54 are displayed for the three models, see labels. The dynamics of the models are juxtaposed in (b).

In Fig. 3.22 (b) the dynamical behavior of all models under the same control protocol as in Fig. 3.10 (a) is presented. The reduced models 2) and 3) react considerably faster with respect to the onset of the read pulse than the full model. Again, this is not surprising: with the elimination of all 'fast' variables from the full model, many processes that actually require at least some time to equilibrate are assumed to happen instantaneously. By comparing the half-lives of mRNAs (3 minutes) and proteins (5 minutes) it becomes clear, that the assumption of mRNA being the fast variable and proteins being the slow variables is somewhat crude. It turns out that the elimination of the mRNA dynamics is indeed the major reason for the delay between the reduced and the full model.

Chapter 4

Discussion and outlook

Discussion

In this work, the design of a simple genetic circuit to implement conditional memory in bacteria is described, and various properties of this circuit are characterized theoretically. The circuit is based on the genetic toggle switch which has been demonstrated to function *in vivo* in *E. coli* [46, 72], and contains an additional control module involving two transcription factors R and S. The additional layer of control dictates the condition (level of R) by which transient information (level of S) may be stored in the toggle switch. Conditional memory would then enable organisms to manipulate information "collected" under different conditions at different times. Such capabilities may provide selective advantages to organisms in time varying environments. For instance, under repeated cycles of famine and feast, bacteria which can remember certain environmental trait during feast may formulate better survival strategies at the time of famine [75].

In the proposed design, the input signals were taken to be the transcriptional rates of R and S. This provides a versatile interface of the circuit to other cellular processes. For instance, the transcription of R or S may be driven by the output of a natural or synthetic two-component signaling system that senses an environmental trait, e.g. the light intensity [77] or density of bacteria [19]. Alternatively it may be driven by metabolic or growth regulators that signal the internal state of the cell.

According to the presented design and analysis, one expects the circuit to be able to respond rapidly to variations in input signals, on a time scale as short as 30 min. This response time scale is dictated by the half-life of the regulatory proteins which was shortened in this design by the use of protein degradation tags [67]. Despite the short basic time scale, this analysis suggests that a broad parameter regime can be found for which the memory is stable to stochastic fluctuations in gene expression for many hours. It turned out that the stability of the states depends on the basal levels of R and S - an aspect where stochastic resonance might be important, see next paragraph. It was also shown that the circuit can easily be adjusted to function with a broad range of input signal amplitudes, e.g. the transcription rates for R and S, by mutations in the R_2 and RS binding sites.

This study suggests that conditional memory such as that envisaged by the presented

design should be implementable in bacteria using typical protein regulators and transcriptional control mechanisms. The basic toggle switch involves the mutual repression of two simple repressor proteins for which the only property demanded is specific binding to DNA. The control module involves only adding one operator site each to repress the two promoters of the toggle switch, for homodimer or heterodimer of the transcription factors R and S. The analysis was performed for parameters associated with a specific choice for the R/S pair, the 434 repressor and its mutant $434R[\alpha 3(P22R)]$, because their properties had already been characterized quantitatively [61]. However, it is believed that such pairs may be readily generated by modifying known bacterial transcription factors: As demonstrated by Dmitrova *et al.* [39], not only can DNA binding domains of the transcription factors be altered to enable different binding specificities, the dimerization domain can also be manipulated to enable the desired homo- and hetero-dimerization required by this design for the conditional memory circuit.

From a technical point of view this work shows, that the full description of the genetic circuit by a vast number of state variables can be reduced by rapid equilibrium assumptions. The resulting simple model accounts for the dynamics of the remaining slow degrees of freedom, which are here the total protein abundances. It could be shown, that the steady state characteristics of the simple model almost perfectly resembles the full model, whereas the dynamics differ significantly. This suggested, that the major assumption of mRNA equilibrium was not justified, since the timescales of mRNA and protein dynamics are similar. Thus, whenever a simple model is used, one has to be aware that it relies on adiabatic approximations and the separation of timescales has to be ensured.

Outlook

This work is intended to stimulate the construction and experimental characterization of the conditional memory circuit. It represents a concrete example for a nontrivial sequential logic circuit, the implementation of which would be a milestone in synthetic biology and a necessary first step for the construction of circuits with more advanced functions, e.g. “genetic counters”.

It might be used as a *diagnostic tool* as well, capturing the transient state of single cells under a given condition. If for instance the store signal would be coupled to some cell-cell signaling messenger, bacteria equipped with the conditional memory circuit could act as ”agent-cells”, locally detecting the ”state” of other cells only if triggered to do so by an external read signal which could be e.g. a pulse of light.

The conditional memory circuit might also serve as a perfect model system for the study of *stochastic resonance* in gene regulatory circuits. Quoting Hänggi, *”(...) Stochastic resonance refers to a situation where the mere addition of random noise to the dynamics improves a system's sensitivity to discriminate weak information-carrying signals.”* [58]. The mechanism underlying stochastic resonance, can be exemplified by using the ”potential-picture” of the conditional memory circuit (see Section 3.4.3). It was illustrated, that the effect of the control proteins R and S could be interpreted as a distortion of the ”double-well potential” of the bare toggle switch. At basal levels of R we saw, that the spontaneous

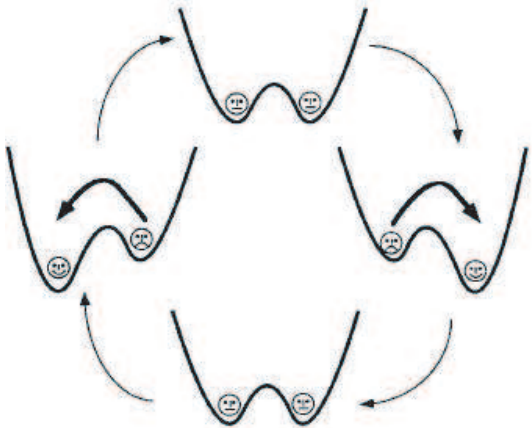


Figure 4.1: Schematic mechanism of stochastic resonance, for details see text. The illustration was taken from [58].

switching rate from one state to the other depends on the level of S, corresponding to varying depths of the potential minima. Thus a weakly oscillating store signal generates a periodically distorted energy landscape as depicted in Fig. 4.1. In this case *"(...) the presence of an optimal dose of noise (so that the average stochastic escape time approximately equals half the period of the signal) will statistically induce synchronized hopping events between the two locally stable states."* [58].

Through the regulation of the basal read signal strength the impact of S on the toggle switch could be readily tuned. This would in turn allow for a gradual activation of the stochastic resonance effect. Such a versatile "detection-tool" for small signal amplitudes would also permit the adjustment to different *periods* of the external signal. Due to the read signal dependence of the circuit's spontaneous switching time also the optimal frequency for stochastic resonance would be shifted.

Appendix A

Model details

A.1 List of elementary chemical reactions

Degradation of mRNAs, protein monomers and dimers:

$$R_1: \quad m_A \xrightarrow{\lambda_m} \emptyset \tag{A.1}$$

$$R_2: \quad m_B \xrightarrow{\lambda_m} \emptyset \tag{A.2}$$

$$R_3: \quad A \xrightarrow{\lambda_p} \emptyset \tag{A.3}$$

$$R_4: \quad B \xrightarrow{\lambda_p} \emptyset \tag{A.4}$$

$$R_5: \quad A_2 \xrightarrow{\lambda_p} \emptyset \tag{A.5}$$

$$R_6: \quad B_2 \xrightarrow{\lambda_p} \emptyset \tag{A.6}$$

Protein dimerization:

$$R_7: \quad A + A \xrightarrow{k_A^+} A_2 \tag{A.7}$$

$$R_8: \quad B + B \xrightarrow{k_B^+} B_2 \tag{A.8}$$

$$R_9: \quad R + R \xrightarrow{k_{R_2}^+} R_2 \tag{A.9}$$

$$R_{10}: \quad R + S \xrightarrow{k_{RS}^+} RS \tag{A.10}$$

$$R_{11}: \quad A_2 \xrightarrow{k_A^-} A + A \tag{A.11}$$

$$R_{12}: \quad B_2 \xrightarrow{k_B^-} B + B \tag{A.12}$$

$$R_{13}: \quad R_2 \xrightarrow{k_{R_2}^-} R + R \tag{A.13}$$

$$R_{14}: \quad RS \xrightarrow{k_{RS}^-} R + S \tag{A.14}$$

TF-Operator binding and unbinding:

$$R_{15}: \quad P_A + B_2 \xrightarrow{2\,k^+_{O_B}} P_A \cdot B_2 \tag{A.15}$$

$$R_{16}: \quad P_A + R_2 \xrightarrow{k^+_{O_{R_2}}} P_A \cdot R_2 \tag{A.16}$$

$$R_{17}: \quad P_B + A_2 \xrightarrow{2\,k^+_{O_A}} P_B \cdot A_2 \tag{A.17}$$

$$R_{18}: \quad P_B + R_2 \xrightarrow{k^+_{O_{R_2}}} P_B \cdot R_2 \tag{A.18}$$

$$R_{19}: \quad P_A \cdot B_2 + B_2 \xrightarrow{k^+_{O_B}} P_A \cdot B_2 \cdot B_2 \tag{A.19}$$

$$R_{20}: \quad P_A \cdot B_2 + R_2 \xrightarrow{k^+_{O_{R_2}}} P_A \cdot B_2 \cdot R_2 \tag{A.20}$$

$$R_{21}: \quad P_A \cdot R_2 + B_2 \xrightarrow{2\,k^+_{O_B}} P_A \cdot B_2 \cdot R_2 \tag{A.21}$$

$$R_{22}: \quad P_B \cdot A_2 + A_2 \xrightarrow{k^+_{O_A}} P_B \cdot A_2 \cdot A_2 \tag{A.22}$$

$$R_{23}: \quad P_B \cdot A_2 + RS \xrightarrow{k^+_{O_{RS}}} P_B \cdot A_2 \cdot RS \tag{A.23}$$

$$R_{24}: \quad P_B \cdot RS + A_2 \xrightarrow{2\,k^+_{O_A}} P_B \cdot A_2 \cdot RS \tag{A.24}$$

$$R_{25}: \quad P_A \cdot B_2 \cdot B_2 + R_2 \xrightarrow{k^+_{O_{R_2}}} P_A \cdot B_2 \cdot B_2 \cdot R_2 \tag{A.25}$$

$$R_{26}: \quad P_A \cdot B_2 \cdot R_2 + B_2 \xrightarrow{k^+_{O_B}} P_A \cdot B_2 \cdot B_2 \cdot R_2 \tag{A.26}$$

$$R_{27}: \quad P_B \cdot A_2 \cdot A_2 + RS \xrightarrow{k^+ O_{RS}} P_B \cdot A_2 \cdot A_2 \cdot RS \tag{A.27}$$

$$R_{28}: \quad P_B \cdot A_2 \cdot RS + A_2 \xrightarrow{k^+_{O_A}} P_B \cdot A_2 \cdot A_2 \cdot RS \tag{A.28}$$

$$R_{29}: \quad P_A \cdot B_2 \xrightarrow{k^-_{O_B}} P_A + B_2 \tag{A.29}$$

$$R_{30}: \quad P_A \cdot R_2 \xrightarrow{k^-_{O_{R_2}}} P_A + R_2 \tag{A.30}$$

$$R_{31}: \quad P_B \cdot A_2 \xrightarrow{k^-_{O_A}} P_B + A_2 \tag{A.31}$$

$$R_{32}: \quad P_B \cdot RS \xrightarrow{k^-_{O_{RS}}} P_B + RS \tag{A.32}$$

$$R_{33}: \quad P_A \cdot B_2 \cdot B_2 \xrightarrow{2\,k^-_{O_B}} P_A \cdot B_2 + B_2 \tag{A.33}$$

$$R_{34}: \quad P_A \cdot B_2 \cdot R_2 \xrightarrow{k^-_{O_{R_2}}} P_A \cdot B_2 + R_2 \tag{A.34}$$

$$R_{35}: \quad P_A \cdot B_2 \cdot R_2 \xrightarrow{k^-_{O_B}} P_A \cdot R_2 + B_2 \tag{A.35}$$

$$R_{36}: \quad P_B \cdot A_2 \cdot A_2 \xrightarrow{2\,k^-_{O_A}} P_B \cdot A_2 + A_2 \tag{A.36}$$

$$R_{37}: \quad P_B \cdot A_2 \cdot RS \xrightarrow{k^-_{O_{RS}}} P_B \cdot A_2 + RS \tag{A.37}$$

$$R_{38}: \quad P_B \cdot A_2 \cdot RS \xrightarrow{k^-_{O_A}} P_B \cdot RS + A_2 \tag{A.38}$$

$$R_{39}: \quad P_A \cdot B_2 \cdot B_2 \cdot R_2 \xrightarrow{k^-_{O_{R_2}}} P_A \cdot B_2 \cdot B_2 + R_2 \tag{A.39}$$

$$R_{40}: \quad P_A \cdot B_2 \cdot B_2 \cdot R_2 \xrightarrow{2\,k^-_{O_B}} P_A \cdot B_2 \cdot R_2 + B_2 \tag{A.40}$$

$$R_{41}: \quad P_B \cdot A_2 \cdot A_2 \cdot RS \xrightarrow{k^-_{O_{RS}}} P_B \cdot A_2 \cdot A_2 + RS \tag{A.41}$$

$$R_{42}: \quad P_B \cdot A_2 \cdot A_2 \cdot RS \xrightarrow{2\,k^-_{O_A}} P_B \cdot A_2 \cdot RS + A_2 \tag{A.42}$$

Transcription and translation:

$$R_{43}: \quad P_A \xrightarrow{\nu_{m_A}} P_A + m_A \tag{A.43}$$
$$R_{44}: \quad P_B \xrightarrow{\nu_{m_B}} P_B + m_B \tag{A.44}$$
$$R_{45}: \quad m_A \xrightarrow{\nu_p} m_A + A \tag{A.45}$$
$$R_{46}: \quad m_B \xrightarrow{\nu_p} m_B + B \tag{A.46}$$

The regulatory front end:

$$R_{47}: \quad P_R \xrightarrow{\nu_{m_R}} P_R + m_R \tag{A.47}$$
$$R_{48}: \quad P_S \xrightarrow{\nu_{m_S}} P_S + m_S \tag{A.48}$$
$$R_{49}: \quad m_R \xrightarrow{\nu_p} m_R + R \tag{A.49}$$
$$R_{50}: \quad m_S \xrightarrow{\nu_p} m_S + S \tag{A.50}$$
$$R_{51}: \quad m_R \xrightarrow{\lambda_m} \emptyset \tag{A.51}$$
$$R_{52}: \quad m_S \xrightarrow{\lambda_m} \emptyset \tag{A.52}$$
$$R_{53}: \quad R \xrightarrow{\lambda_p} \emptyset \tag{A.53}$$
$$R_{54}: \quad S \xrightarrow{\lambda_p} \emptyset \tag{A.54}$$
$$R_{55}: \quad R_2 \xrightarrow{\lambda_p} \emptyset \tag{A.55}$$
$$R_{56}: \quad RS \xrightarrow{\lambda_p} \emptyset \tag{A.56}$$

The numbered chemical species S_i are:

$$S_1: \quad A \tag{A.57}$$
$$S_2: \quad A_2 \tag{A.58}$$
$$S_3: \quad B \tag{A.59}$$
$$S_4: \quad B_2 \tag{A.60}$$
$$S_5: \quad R \tag{A.61}$$
$$S_6: \quad S \tag{A.62}$$
$$S_7: \quad R_2 \tag{A.63}$$
$$S_8: \quad RS \tag{A.64}$$
$$S_9: \quad P_A \tag{A.65}$$
$$S_{10}: \quad P_A \cdot B_2 \tag{A.66}$$
$$S_{11}: \quad P_A \cdot R_2 \tag{A.67}$$
$$S_{12}: \quad P_A \cdot B_2 \cdot R_2 \tag{A.68}$$
$$S_{13}: \quad P_A \cdot B_2 \cdot B_2 \tag{A.69}$$
$$S_{14}: \quad P_A \cdot B_2 \cdot B_2 \cdot R_2 \tag{A.70}$$
$$S_{15}: \quad P_B \tag{A.71}$$
$$S_{16}: \quad P_B \cdot A_2 \tag{A.72}$$
$$S_{17}: \quad P_B \cdot RS \tag{A.73}$$
$$S_{18}: \quad P_B \cdot A_2 \cdot RS \tag{A.74}$$
$$S_{19}: \quad P_B \cdot A_2 \cdot A_2 \tag{A.75}$$
$$S_{20}: \quad P_B \cdot A_2 \cdot A_2 \cdot RS \tag{A.76}$$
$$S_{21}: \quad m_A \tag{A.77}$$
$$S_{22}: \quad m_B \tag{A.78}$$
$$S_{23}: \quad m_R \tag{A.79}$$
$$S_{24}: \quad m_S \tag{A.80}$$

A.2 Rate equations of the full model

$$\frac{dA}{dt} = \nu_p m_A - 2k_A^+ A^2 + 2k_A^- A_2 - \lambda_p A \tag{A.81}$$

$$\begin{aligned}\frac{dA_2}{dt} = {} & -\lambda_p A_2 - k_A^- A_2 - 2k_{O_A}^+ A_2 P_B - k_{O_A}^+ A_2 P_B \bullet A_2 - 2k_{O_A}^+ A_2 P_B \bullet RS \\ & -k_{O_A}^+ A_2 P_B \bullet A_2 \bullet RS + k_A^+ A^2 + k_{O_A}^- P_B \bullet A_2 + 2k_{O_A}^- P_B \bullet A_2 \bullet A_2 \\ & +k_{O_A}^- P_B \bullet A_2 \bullet RS + 2k_{O_A}^- P_B \bullet A_2 \bullet A_2 \bullet R_2\end{aligned} \tag{A.82}$$

$$\frac{dB}{dt} = \nu_p m_B - 2k_B^+ B^2 + 2k_B^- B_2 - \lambda_p B \tag{A.83}$$

$$\begin{aligned}\frac{dB_2}{dt} = {} & -\lambda_p B_2 - k_B^- B_2 - 2k_{O_B}^+ B_2 P_A - k_{O_B}^+ B_2 P_A \bullet B_2 - 2k_{O_B}^+ B_2 P_A \bullet R_2 \\ & -k_{O_B}^+ B_2 P_A \bullet B_2 \bullet R_2 + k_B^+ B^2 + k_{O_B}^- P_A \bullet B_2 + 2k_{O_B}^- P_A \bullet B_2 \bullet B_2 \\ & +k_{O_B}^- P_A \bullet B_2 \bullet R_2 + 2k_{O_B}^- P_A \bullet B_2 \bullet B_2 \bullet RS\end{aligned} \tag{A.84}$$

$$\frac{dR}{dt} = -2k_{R_2}^+ R^2 - k_{RS}^+ R\, S - \lambda_p R + 2k_{R_2}^- R_2 + k_{RS}^- RS + \nu_p m_R \tag{A.85}$$

$$\frac{dS}{dt} = -k_{RS}^+ R\, S - \lambda_p S + k_{RS}^- RS + \nu_p m_S \tag{A.86}$$

$$\begin{aligned}\frac{dR_2}{dt} = {} & -k_{R_2}^- R_2 - k_{O_{R_2}}^+ R_2 P_B - k_{O_{R_2}}^+ R_2 P_A \bullet B_2 - k_{O_{R_2}}^+ R_2 P_A \bullet B_2 \bullet B_2 - \lambda_p R_2 \\ & +k_{R_2}^+ R^2 + k_{O_{R_2}}^- P_A \bullet R_2 + k_{O_{R_2}}^- P_A \bullet B_2 \bullet B_2 + k_{O_{R_2}}^- P_A \bullet B_2 \bullet B_2 \bullet R_2\end{aligned} \tag{A.87}$$

$$\begin{aligned}\frac{dRS}{dt} = {} & -k_{RS}^- RS - k_{O_{RS}}^+ RSP_A - k_{O_{RS}}^+ RSP_B \bullet A_2 - k_{O_{RS}}^+ RSP_B \bullet A_2 \bullet A_2 - \lambda_p RS \\ & +k_{RS}^+ R\, S + k_{O_{RS}}^- P_B \bullet RS + k_{O_{RS}}^+ P_B \bullet A_2 \bullet RS + k_{O_{RS}}^+ P_B \bullet A_2 \bullet A_2 \bullet RS\end{aligned} \tag{A.88}$$

$$\frac{dP_A}{dt} = -2k_{O_B}^+ B_2 P_A - k_{O_{R_2}}^+ R_2 P_A + k_{O_B}^- P_A \bullet B_2 + k_{O_{R_2}}^- P_A \bullet R_2 \tag{A.89}$$

$$\begin{aligned}\frac{dP_A \bullet B_2}{dt} = {} & -k_{O_B}^+ B_2 P_A \bullet B_2 - k_{O_{R_2}}^+ R_2 P_A \bullet B_2 - k_{O_B}^- P_A \bullet B_2 + k_{O_B}^+ B_2 P_A \\ & +2k_{O_B}^- P_A \bullet B_2 \bullet B_2 + k_{O_{R_2}}^+ P_A \bullet B_2 \bullet R_2\end{aligned} \tag{A.90}$$

$$\frac{dP_A \bullet R_2}{dt} = -k_{O_B}^+ B_2 P_A \bullet R_2 - k_{O_{R_2}}^- P_A \bullet R_2 + k_{O_{R_2}}^+ R_2 P_A + k_{O_B}^- P_A \bullet B_2 \bullet R_2 \tag{A.91}$$

$$\begin{aligned}\frac{dP_A \bullet B_2 \bullet R_2}{dt} = {} & -k_{O_B}^+ B_2 P_A \bullet B_2 \bullet R_2 - k_{O_B}^- P_A \bullet B_2 \bullet R_2 - k_{O_{R_2}}^- P_A \bullet B_2 \bullet R_2 \\ & +k_{O_{R_2}}^+ R_2 P_A \bullet B_2 + k_{O_B}^+ B_2 P_A \bullet R_2 + 2k_{O_B}^- P_A \bullet B_2 \bullet B_2 \bullet R_2\end{aligned} \tag{A.92}$$

$$\begin{aligned}\frac{dP_A \bullet B_2 \bullet B_2}{dt} = {} & -k_{O_{R_2}}^+ R_2 P_A \bullet B_2 \bullet B_2 - 2k_{O_B}^- P_A \bullet B_2 \bullet B_2 + k_{O_B}^+ B_2 P_A \bullet B_2 \\ & +k_{O_{R_2}}^- P_A \bullet B_2 \bullet B_2 \bullet R_2\end{aligned} \tag{A.93}$$

$$\begin{aligned}\frac{dP_A \bullet B_2 \bullet B_2 \bullet R_2}{dt} = {} & -k_{O_B}^- P_A \bullet B_2 \bullet B_2 \bullet R_2 - k_{O_{R_2}}^- P_A \bullet B_2 \bullet B_2 \bullet R_2 \\ & +k_{O_{R_2}}^+ R_2 P_A \bullet B_2 \bullet B_2 + k_{O_B}^+ B_2 P_A \bullet B_2 \bullet R_2\end{aligned} \tag{A.94}$$

$$\frac{dP_B}{dt} = -2k_{O_A}^+ A_2 P_B - k_{O_{RS}}^+ RSP_B + k_{O_A}^- P_B \bullet A_2 + k_{O_{RS}}^- P_B \bullet RS \tag{A.95}$$

$$\begin{aligned}\frac{dP_B \bullet A_2}{dt} = {} & -k_{O_A}^+ A_2 P_B \bullet A_2 - k_{O_{RS}}^+ RSP_B \bullet A_2 - k_{O_A}^- P_B \bullet A_2 + 2k_{O_A}^+ A_2 P_B \\ & +k_{O_A}^- P_B \bullet A_2 \bullet A_2 + k_{O_{RS}}^- P_B \bullet A_2 \bullet RS\end{aligned} \tag{A.96}$$

$$\frac{dP_B \bullet RS}{dt} = -2k_{O_A}^+ A_2 P_B \bullet RS - k_{O_{RS}}^- P_B \bullet RS + k_{O_{RS}}^+ RSP_B + k_{O_A}^- P_B \bullet A_2 \bullet RS \tag{A.97}$$

$$\begin{aligned}\frac{dP_B \bullet A_2 \bullet RS}{dt} = {} & -k_{O_A}^+ A_2 P_B \bullet A_2 \bullet RS - k_{O_A}^- P_B \bullet A_2 \bullet RS - k_{O_{RS}}^- P_B \bullet A_2 \bullet RS \\ & +k_{O_{RS}}^+ RSP_B \bullet A_2 + 2k_{O_A}^+ A_2 P_B \bullet RS + 2k_{O_A}^- P_B \bullet A_2 \bullet A_2 \bullet RS\end{aligned} \tag{A.98}$$

$$\frac{dP_B \bullet A_2 \bullet A_2}{dt} = -k^+_{O_{RS}} RSP_B \bullet A_2 \bullet A_2 - 2k^-_{O_A} P_B \bullet A_2 \bullet A_2 + k^+_{O_A} A_2 P_B \bullet A_2 + k^-_{O_{RS}} P_B \bullet A_2 \bullet A_2 \bullet RS \quad \text{(A.99)}$$

$$\frac{dP_B \bullet A_2 \bullet A_2 \bullet RS}{dt} = -2k^-_{O_A} P_B \bullet A_2 \bullet A_2 \bullet RS - k^-_{O_{RS}} P_B \bullet A_2 \bullet A_2 \bullet RS + k^+_{O_{RS}} RSP_B \bullet A_2 \bullet A_2 + k^+_{O_A} A_2 P_B \bullet A_2 \bullet RS \quad \text{(A.100)}$$

$$\frac{dm_A}{dt} = \nu_{m_A} P_A - \lambda_m m_A \quad \text{(A.101)}$$

$$\frac{dm_B}{dt} = \nu_{m_B} P_B - \lambda_m m_B \quad \text{(A.102)}$$

$$\frac{dm_R}{dt} = \nu_{m_R} P_R - \lambda_m m_R \quad \text{(A.103)}$$

$$\frac{dm_S}{dt} = \nu_{m_S} P_S - \lambda_m m_S \quad \text{(A.104)}$$

Appendix B

A versatile implementation of the stochastic simulation algorithm

Listing B.1: C Implementation of the stochastic simulation algorithm

```
#include <iostream>
#include <fstream>
#include <cstdlib>
#include <string>
#include <cmath>
#include <sys/time.h>

using namespace std;

#define N    24
#define M    56
#define N_sim 100000        //number of sim. runs used for density estimation
#define R_max 600           //maximal molecule number of each species
#define sampling_time 600   //samping time of the density plot
#define total_time 42000    //must match the maximal time in the frontendfile!

int factorial(int);
int get_mu(double, double *);
double propens(int *,int substrate[M+1][N+1],int ,double);
void checkfile(int open, char filename[]);

int main(int argc, char** argv) {

  char model[70]="memory3";
  char spec[70]="dens";
  char frontfile[70]="front_new_1.txt";
  int N_time=int(total_time/sampling_time);
  double c[M+1],a[M+1];
  int int_time,x0[N+1],x[N+1],substrate[M+1][N+1],product[M+1][N+1],
    densityA[N_time+1][R_max+1],densityB[N_time+1][R_max+1];

  //outfiles
  ofstream myof1,myof2;
  int N_spec = strlen(spec);
  strcat(spec,"_densA.dat");
  myof1.open(spec,ios::out);
  spec[N_spec] ='\0';
  strcat(spec,"_densB.dat");
  myof2.open(spec,ios::out);
  spec[N_spec] ='\0';
```

```
41
42  // reading the model information
43  ifstream sta,sub,prod,prob,front;
44  int N_string = strlen(model);
45  strcat(model,"_state.txt");
46  sta.open(model,ios::in);checkfile(! sta,model);
47  model[N_string]='\0';
48  strcat(model,"_substrate.txt");
49  sub.open(model,ios::in);checkfile(! sub,model);
50  model[N_string]='\0';
51  strcat(model,"_product.txt");
52  prod.open(model,ios::in);checkfile(! prod,model);
53  model[N_string]='\0';
54  strcat(model,"_prob_autogenerated.txt");//_increased_R2_offrate
55  prob.open(model,ios::in);checkfile(! prob,model);
56  model[N_string]='\0';
57  front.open(frontfile,ios::in);checkfile(! front,frontfile);
58  front.close();
59
60  for(int j=1;j<=M;j++){
61    for(int i=1;i<=N;i++){
62      sub >> substrate[j][i];
63      prod>> product[j][i];
64    }
65  }
66  for(int i=1;i<=M;i++){prob >> c[i];}
67  for(int i=1;i<=N;i++){sta >> x0[i];}
68
69  //compose the propensity update matrix
70  int tick_update,update[M+1][N+1];
71  update[0][0]=0;update[0][1]=0;update[1][0]=0;
72  for(int i=1;i<=M;i++){for(int j=1;j<=N;j++){update[i][j]=0;}}
73  for(int i=1;i<=M;i++){
74    tick_update=1;
75    for(int j=1;j<=N;j++){
76      if(substrate[i][j]!=0 || product[i][j]!=0){
77        for(int k=1;k<=M;k++){
78          if(substrate[k][j]!=0){update[i][tick_update]=k; tick_update++;}
79        }
80      }
81    }
82  }
83
84  //generate random numbers
85  long idum;
86  long millisekunden = 0;
87  struct timeval time;
88  gettimeofday(&time, NULL);
89  milliseconds = time.tv_usec;
90  idum = -milliseconds;
91  drand48(idum);              //initialization of the random number generator
92  long double a_0,t;
93  long unsigned int mu,n_A,n_B,tt,tickt;
94  for(int i=1;i<=N_time;i++){
95    for(int j=1;j<=R_max;j++){
96      densityA[i][j]=0; densityB[i][j]=0;
97    }
98  }
99
100 for(int sim=1;sim<=N_sim; sim++){
101   for(int i=1;i<=N;i++){x[i]=x0[i];}
102   front.open(frontfile,ios::in);
103   front.clear();
104   front.seekg(0,ios::beg);
105   tt=sampling_time;t=0.;tickt=0;
```

```
106     while(front.good()){
107       front>>int_time; front>>c[47]; front>>c[48];
108       if(int_time>total_time){cerr<<"Time in frontendfile is not consistent
109 with time in code";exit(1);}
110       for(int i=1;i<=M;i++){
111         a[i]=propens(x,substrate,i,c[i]); //initial propensities
112       }
113       while(t<int_time){
114         while(t<double(tt)){
115           a_0=0.;
116           for(int i=1;i<=M;i++){a_0+=a[i];}
117           mu=get_mu(a_0,a);
118           for(int i=1;i<=N;i++){
119             x[i]+=(product[mu][i]-substrate[mu][i]);//state update
120           }
121           tick_update=1;
122           while(update[mu][tick_update]!=0 && tick_update<M){
123             a[update[mu][tick_update]]=propens(x,substrate,
124                 update[mu][tick_update],c[update[mu][tick_update]]);
125             tick_update++;
126           }
127           t+=log(1./(1.-drand48()))/a_0;//time update 1-drand48() is in (0,1]
128         }
129         tt+=sampling_time;tickt++;
130         n_A=int(x[1]+2.*x[2]+2.*x[16]+2.*x[18]+4.*x[19]+4.*x[20]);
131         n_B=int(x[3]+2.*x[4]+2.*x[10]+2.*x[12]+4.*x[13]+4.*x[14]);
132         if(n_A > R_max || n_B > R_max ){cout<<"Warning: Molecule number
133 exceeds matrix dimenions"<<endl;}
134         densityA[tickt][n_A]++;
135         densityB[tickt][n_B]++;
136       }
137     }
138     front.close();
139   }
140   double binsum_A,binsum_B;
141   int n_bin=10;
142
143   for(int j=0;j<R_max/double(n_bin);j++){
144     for(int i=1;i<=N_time;i++){
145       binsum_A=0.; binsum_B=0.;
146       for(int k=0;k<n_bin;k++){binsum_A+=densityA[i][j*n_bin+k];
147        binsum_B+=densityB[i][j*n_bin+k];}
148       myof1<<binsum_A<<" ";
149       myof2<<binsum_B<<" ";
150     }
151     myof1<<endl;
152     myof2<<endl;
153   }
154
155   return 0;
156 }
157
158 int factorial (int num)
159 {
160   int result=1;
161   for (int i=1; i<=num; ++i)
162     result*=i;
163   return result;
164 }
165
166 double propens(int *state, int substrate[M+1][N+1],int index,double c)
167 {
168   double result=c;
169   for(int i=1;i<=N;i++){
170     if(substrate[index][i]!=0){
```

```
171         if(state[i]>=substrate[index][i]){
172           for(int j=0;j<substrate[index][i];j++){
173             result*=double(state[i]-j);
174           }
175           result/=factorial(substrate[index][i]);
176         }
177         else result=0.;//else return 0.;//
178       }
179     }
180     return result;
181 }
182 
183 int get_mu(double a_0, double *a)
184 {
185     int mu=0;
186     double sum=0.;
187     double auxran = 1.-drand48();
188     while( sum/a_0 <= auxran ){
189       sum+=a[mu+1];
190       mu++;
191     }
192     return mu;
193 }
194 
195 void checkfile(int open, char filename[])
196 {
197     if (open){cout << "Error: Can't open the file "<<filename<<endl;exit(1);}
198     else cout<<"Opened file "<<filename<<endl;
199 }
```

Bibliography

[1] http://emu.arsusda.gov/default.html.

[2] http://redpoll.pharmacy.ualberta.ca/CCDB/cgi-bin/STAT_NEW.cgi.

[3] http://www.geneticengineering.org/chemis/Chemis-NucleicAcid/RNA.htm.

[4] http://www.bmb.psu.edu/faculty/tan/lab/gallery_protdna.html.

[5] http://nobelprize.org/educational_games/medicine/dna/a/translation/polysome_em.html.

[6] http://www.jacobsschool.ucsd.edu/news_events/releases/release.sfe?id=518.

[7] M. Acar, A. Becskei, and A. van Oudenaarden. Enhancement of cellular memory by reducing stochastic transitions. *Nature*, 435:228–232, 2005.

[8] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, 2002.

[9] R. J. Allen, D. Frenkel, and P. R. ten Wolde. Simulating rare events in equilibrium or nonequilibrium stochastic systems. *J. Chem. Phys.*, 124:024102–16, 2006.

[10] R. J. Allen, P. B. Warren, and P. R. ten Wolde. Sampling rare switching events in biochemical networks. *Phys. Rev. Lett.*, 94:018104–4, 2005.

[11] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, 2:msb4100073–E1–msb4100073–E14, 2006.

[12] A. M. Arias and P. Hayward. Filtering transcriptional noise during development: concepts and mechanisms. *Nat. Rev. Genet.*, 7:34–44, 2006.

[13] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, 149:1633–1648, 1998.

[14] K. L. Arney and A. G. Fisher. Epigenetic aspects of differentiation. *J. Cell. Sci.*, 117:4355–4363, 2004.

[15] M. R. Atkinson, M. A. Savageau, J. T. Myers, and A. J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell*, 113:597–607, 2003.

[16] E. Aurell and K. Sneppen. Epigenetics as a first exit problem. *Phys. Rev. Lett.*, 88:048101, 2002.

[17] D. W. Austin, M. S. Allen, J. M. McCollum, R. D. Dar, J. R. Wilgus, G. S. Sayler, N. F. Samatova, C. D. Cox, and M. L. Simpson. Gene network shaping of inherent noise spectra. *Nature*, 439:608–611, 2006.

[18] 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.*, 38:636–643, 2006.

[19] B. L. Bassler. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Curr. Opin. Microbiol.*, 2:582–587, 1999.

[20] S. Basu, Y. Gerchman, C. H. Collins, F. H. Arnold, and R. Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434:1130–1134, 2005.

[21] J. A. Bernstein, A. B. Khodursky, P.-H. Lin, S. Lin-Chao, and S. N. Cohen. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl. Acad. Sci. USA*, 99:9697–9702, 2002.

[22] M. Biburger. *Characterisierung intragenischer Suppressoren nichtinduzierbarer TetR(B) Mutanten und thermodynamische Analyse der Repressor-Operator Wechselwirkung*. PhD thesis, Universität Erlangen, 1997.

[23] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips. Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev.*, 15:125–135, 2005.

[24] L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15:116–124, 2005.

[25] A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comp. Phys.*, 17:10–18, 1975.

[26] D. F. Browning and S. J.W. Busby. The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, 2:1–9, 2004.

[27] R. F. Bruinsma. Physics of protein-DNA interaction. *Physica A*, 313:211–237, 2002.

[28] N. E. Buchler, U. Gerland, and T. Hwa. On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. USA*, 100:5136–5141, 2003.

[29] N. E. Buchler, U. Gerland, and T. Hwa. Nonlinear protein degradation and the function of genetic circuits. *Proc. Natl. Acad. Sci. USA*, 102:9559–9564, 2005.

[30] R. Bundschuh, F. Hayot, and C. Jayaprakash. Fluctuations and slow variables in genetic networks. *Biophys. J.*, 84:1606–1615, 2003.

[31] L. Cai, N. Friedman, and X. S. Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440:358–362, 2006.

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

[33] J. Casadesus and R. D'Ari. Memory in bacteria and phage. *Bioessays*, 24(6):512–518, 2002.

[34] J. L. Cherry and F. R. Adler. How to make a biological switch. *J. Theor. Biol.*, 203:117–133, 2000.

[35] C. D. Cox, J. M. McCollum, D. W. Austin, M. S. Allen, R. D. Dar, and M. L. Simpson. Frequency domain analysis of noise in simple gene circuits. *Chaos*, 16:026102, 2006.

[36] F. Crick. Central dogma of molecular biology. *Nature*, 227:561–563, 1970.

[37] F. Crick and J. Watson. Molecular structure of nucleic acids. *Nature*, 171:737738, 1953.

[38] E. H. Davidson, J. P. Rast, P. Oliveri, A. Ransick, C. Calestani, C.-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C. T. Brown, C. B. Livi, P. Y. Lee, R. Revilla, A. G. Rust, Z. Pan, M. J. Schilstra, P. J. C. Clarke, M. I. Arnone, L. Rowen, R. A. Cameron, D. R. McClay, L. Hood, and H. Bolouri. A genomic regulatory network for development. *Science*, 295:1669–1678, 2002.

[39] M. Dmitrova, G. Younes-Cauet, P. Oertel-Buchheit, D. Porte, M. Schnarr, and M. Granger-Schnarr. A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in Escherichia coli. *Mol. Gen. Genet.*, 257:205–212, 1998.

[40] A.L. Donner, K. Paa, and G.B. Koudelka. Carboxyl-terminal domain dimer interface mutant 434 repressors have altered dimerization and DNA binding specificities. *J. Mol. Biol.*, 283:931–946, 1998.

[41] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[42] M. B. Elowitz, M. G. Surette, P.-E. Wolf, J. B. Stock, and S. Leibler. Protein mobility in the cytoplasm of Escherichia coli. *J. Bacteriol.*, 181:197–203, 1999.

[43] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S. Lee, and J. C. Liao. A synthetic gene-metabolic oscillator. *Nature*, 435:118–122, 2005.

[44] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. *Proc. Natl. Acad. Sci. USA*, 101:10955–10960, 2004.

[45] C. W. Gardiner. *Handbook of stochastic methods*. Springer, 3 edition, 2004.

[46] T. S. Gardner, C. R. Cantor, and J. J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403:339–342, 2000.

[47] U. Gerland, J.D. Moroz, and T. Hwa. Physical constraints and functional characteristics of transcription factor-DNA interaction. *Proc. Natl. Acad. Sci. USA*, 99:12015–12020, 2002.

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

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

[50] D. T. Gillespie. The mathematics of Brownian motion and Johnson noise. *Am. J. Phys.*, 64:225, 1996.

[51] D. T. Gillespie. The multivariate Langevin and FokkerPlanck equations. *Am. J. Phys.*, 64:1246, 1996.

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

[53] A. Goldbeter. Computational approaches to cellular rhythms. *Nature*, 420:238–245, 2002.

[54] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123:1025–1036, 2005.

[55] S. Gottesman. The small RNA regulators of Escherichia coli: Roles and mechanisms. *Annu. Rev. Microbiol.*, 58:303–328, 2004.

[56] S. Gottesman, E. Roche, Y. Zhou, and R. T. Sauer. The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.*, 12:1338–1347, 1998.

[57] N. J. Guido, X. Wang, D. Adalsteinsson, D. McMillen, J. Hasty, C. R. Cantor, T. C. Elston, and J. J. Collins. A bottom-up approach to gene regulation. *Nature*, 439:856–860, 2006.

[58] P. Haenggi. Stochastic resonance in biology how noise can enhance detection of weak signals and help improve biological information processing. *ChemPhysChem*, 3:285–290, 2002.

[59] P. Haenggi, P. Talkner, and M. Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, 62:251–342, 1990.

[60] W. Hillen. private communication.

[61] M. Hollis, D. Valenzuela, D. Pioli, R. Wharton, and M. Ptashne. A repressor heterodimer binds to a chimeric operator. *Proc. Natl. Acad. Sci. USA*, 85:5834–5838, 1988.

[62] J. Honerkamp. *Statistical Physics*. Springer, 2 edition, 2002.

[63] F. J. Isaacs, D. J. Dwyer, and J. J. Collins. RNA synthetic biology. *Nat. Biotech.*, 24:545–554, 2006.

[64] F. J. Isaacs, J. Hasty, C. R. Cantor, and J.J. Collins. Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*, 100:7714–7719, 2003.

[65] R. Jaenisch and A. Bird. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.*, 33 Suppl:245–254, 2003.

[66] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.*, 6:451–464, 2005.

[67] A. W. Karzai, E. D. Roche, and R. T. Sauer. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol.*, 7:449–455, 2000.

[68] R. Katz and G. Boriello. *Contemporary Logic Design*. Prentice Hall, 2004.

[69] J. Keizer. Maxwell-type constructions for multiple nonequilibrium steady states. *Proc. Natl. Acad. Sci. USA*, 75:3023–3026, 1978.

[70] T. B. Kepler and T. C. Elston. Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys. J.*, 81:3116–3136, 2001.

[71] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002.

[72] H. Kobayashi, M. Kaern, M. Araki, K. Chung, T. S. Gardner, C. R. Cantor, and J. J. Collins. Programmable cells: Interfacing natural and engineered gene networks. *Proc. Natl. Acad. Sci. USA*, 101:8414–8419, 2004.

[73] M. Kollmann, L. Lovdok, K. Bartholomé, J. Timmer, and V. Sourjik. Design principles of a bacterial signalling network. *Nature*, 438:504–507, 2005.

[74] G. Koudelka. private communication.

[75] E. Kussell and S. Leibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309:2075–2078, 2005.

[76] J.-C. Leloup, D. Gonze, and A. Goldbeter. Limit cycle models for circadian rhythms based on transcriptional regulation in Drosophila and Neurospora. *J. Biol. Rhythms*, 14:433–448, 1999.

[77] A. Levskaya, A. A. Chevalier, J. J. Tabor, Z. B. Simpson, L. A. Lavery, M. Levy, E. A. Davidson, A. Scouras, A. D. Ellington, and E. M. Marcotte. Synthetic biology: engineering Escherichia coli to see light. *Nature*, 438:441–442, 2005.

[78] F. Luciani, C. Kesmir, M. Mishto, M. Or-Guil, and R. J. de Boer. A mathematical model of protein degradation by the proteasome. *Biophys. J.*, 88:2422–2432, 2005.

[79] R. Lutz and H. Bujard. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucl. Acids Res.*, 25:1203–1210, 1997.

[80] H. H. McAdams and L. Shapiro. A bacterial cell-cycle regulatory network operating in time and space. *Science*, 301:1874–1877, 2003.

[81] H.H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, 94:814–819, 1997.

[82] 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. USA*, 103:7304–7309, 2006.

[83] S. B. Neher, R. T. Sauer, and T. A. Baker. Distinct peptide signals in the UmuD and UmuD' subunits of UmuD/D' mediate tethering and substrate processing by the ClpXP protease. *Proc. Natl. Acad. Sci. USA*, 100:13219–13224, 2003.

[84] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. Regulation of noise in the expression of a single gene. *Nat. Genet.*, 31:69–73, 2002.

[85] E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden. Multistability in the lactose utilization network of Escherichia coli. *Nature*, 427:737–740, 2004.

[86] J. Paulsson. Summing up the noise in gene networks. *Nature*, 427:415–418, 2004.

[87] J. Paulsson. Models of stochastic gene expression. *Physics of Life Reviews*, 2:157–175, 2005.

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

[89] J. Paulsson and M. Ehrenberg. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys. Rev. Lett.*, 84:5447, 2000.

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

[91] M. Ptashne. Regulated recruitment and cooperativity in the design of biological regulatory systems. *Philos. Transact. A Math. Phys. Eng. Sci.*, 361:1223–34, 2003.

[92] M. Ptashne. *A Genetic Switch: Phage Lambda Revisited*. Cold Spring Harbor Laboratory Press, 2004.

[93] M. Ptashne and A. Gann. Transcriptional activation by recruitment. *Nature*, 386:569–577, 1997.

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

[95] D. M. Roma, R. A. O'Flanagan, A. E. Ruckenstein, A. M. Sengupta, and R. Mukhopadhyay. Optimal path to epigenetic switching. *Phys. Rev. E*, 71:011902–5, 2005.

[96] N. Rosenfeld, M. B. Elowitz, and U. Alon. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.*, 323:785–793, 2002.

[97] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, and M. B. Elowitz. Gene regulation at the single-cell level. *Science*, 307:1962–1965, 2005.

[98] L. Saiz, J. M. Rubi, and Jose M. G. Vilar. Inferring the in vivo looping properties of DNA. *Proc. Natl. Acad. Sci. USA*, 102:17642–17645, 2005.

[99] R. Schleif. Regulation of the L-arabinose operon of Escherichia coli. *Trends Genet.*, 16:559–565, 2000.

[100] J. Schnakenberg. *Thermodynamik und Statistische Physik*. Wiley-VCH, 2002.

[101] E. Schroedinger. *Was ist Leben?* Cambridge University Press, Cambridge, 1944.

[102] D. Sprinzak and M. B. Elowitz. Reconstruction of genetic circuits. *Nature*, 438:443–448, 2005.

[103] P. S. Swain, M. B. Elowitz, and E. D. Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*, 99:12795–12800, 2002.

[104] L. Swint-Kruse, H. Zhan, and K. S. Matthews. Integrated insights from simulation, experiment, and mutational analysis yield new details of LacI function. *Biochemistry*, 44:11201–11213, 2005.

[105] S. Tanase-Nicola, P. B. Warren, and P. R. ten Wolde. Signal detection, modularity, and the correlation between extrinsic and intrinsic noise in biochemical networks. *Phys. Rev. Lett.*, 97:068102, 2006.

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

[107] M. Thattai and A. van Oudenaarden. Attenuation of noise in ultrasensitive signaling cascades. *Biophys. J.*, 82:2943–2950, 2002.

[108] J. Timmer. Lecture notes: "Nichtlineare Dynamik". 2002.

[109] J. Timmer. Lecture notes: "Mathematische Modelle in der Biologie". 2003.

[110] A. Trusina, K. Sneppen, I. B. Dodd, K. E. Shearwin, and J. B. Egan. Functional alignment of regulatory networks: a study of temperate phages. *PLoS Comput. Biol.*, 1:e74, 2005.

[111] J. S. van Zon and P. R. ten Wolde. Green's-function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.*, 123:234910, 2005.

[112] J. S. van Zon and P. R. ten Wolde. Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.*, 94:128103, 2005.

[113] J. M. G. Vilar and L. Saiz. DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.*, 15:136–144, 2005.

[114] A. M. Walczak, J. N. Onuchic, and Peter G. Wolynes. Absolute rate theories of epigenetic stability. *Proc. Natl. Acad. Sci. USA*, 102:18926–18931, 2005.

[115] P. B. Warren and P. R. ten Wolde. Chemical models of genetic toggle switches. *J. Phys. Chem. B*, 109:6812–6823, 2005.

[116] H. Xu, M. Moraitis, R. J. Reedstrom, and K. S. Matthews. Kinetic and thermodynamic studies of purine repressor binding to corepressor and operator DNA. *J. Biol. Chem.*, 273:8958–8964, 1998.

[117] J. Yu, J. Xiao, X. Ren, K. Lao, and X. S. Xie. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311:1600–1603, 2006.

[118] X. Zhu, L. Yin, L. Hood, D. Galas, and P. Ao. Efficiency, robustness and stochasticity of gene regulatory networks in systems biology: lambda switch as a working example. `http://arxiv.org/abs/q-bio.SC/0512007`, 2006.

Danksagung

Zunächst einmal möchte Prof. Dr. Jens Timmer und Prof. Dr. Ulrich Gerland herzlich dafür danken, dass Sie mir die Möglichkeit gegeben haben, an einem solch spannenden und aktuellen Thema zu arbeiten. Jens Timmer möchte ich ganz besonders dafür danken, dass er mir den Weg in die Systembiologie gewiesen, die grundlegenden Konzepte vermittelt und zudem den Pfad durch den Dschungel der Bürokratie geebnet hat. Der Enthusiasmus, das umfassende Wissen und die grossartige Intuition von Ulrich Gerland waren der Motor dieser Arbeit und haben mich immer wieder inspiriert und motiviert. Zudem standen mir Prof. Dr. Terence Hwa und Dr. Nicolas E. Buchler mit Rat und Tat zur Seite.

Ich möchte mich auch bei meinen experimentellen Kollaborationspartnern und Kollegen aus der Physik, Judith Leierseder und Prof. Dr. Joachim Rädler, und aus der Mikrobiologie, Christiane Koller und Prof. Dr. Kirsten Jung, für eine inspirierende Zusammenarbeit bedanken. Inspirierend und witzig war auch die Atmosphäre in der Gruppe - ein besonderer Dank gilt hier Richard Neher und Nico Geisel, die mir mit dem Manuskript und bei vielen Problemen weiterhalfen.

Für den großen Rückhalt, das Vertrauen in mich und die bedingungslose Unterstützung möchte ich meiner Familie von Herzen danken: meinen Eltern Angelika und Norbert Fritz, meiner Schwester Theresa Fritz, meinen Großeltern Thekla und Herbert Wegmann und Uwe von Trotha. Ohne sie wäre diese Arbeit wohl niemals entstanden.

Zuletzt möchte ich Kirstin Seidel für ihre Hilfe mit dem Manuskript, aber vor allem für all ihre Liebe und Unterstützung danken!

Hiermit versichere ich, diese Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur angefertigt zu haben.

Freiburg, den 24. Oktober 2006 Georg Fritz