

**Stochastic analysis of nonlinear dynamics
and feedback control
for gene regulatory networks
with applications to synthetic biology**

A thesis presented for
Doctor of Philosophy degree of Imperial College London
by

Natalja Strelkova

Department of Bioengineering
Imperial College London

2011

I certify that this thesis, and the research to which it refers, are the product of my own work, and that any ideas or quotations from the work of other people, published or otherwise, are fully acknowledged in accordance with the standard referencing practices of the discipline.

Signed:

Natalja Strelkova

Copyright

Copyright in text of this thesis rests with the Author. Copies (by any process) either in full, or of extracts, may be made **only** in accordance with instructions given by the Author and lodged in the doctorate thesis archive of the college central library. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the Author.

The ownership of any intellectual property rights which may be described in this thesis is vested in Imperial College, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the written permission of the University, which will prescribe the terms and conditions of any such agreement. Further information on the conditions under which disclosures and exploitation may take place is available from the Imperial College registry.

Table of contents

Abstract	10
1 Introduction: Modeling and engineering in Synthetic Biology	15
1.1 Gene expression in bacteria	18
1.2 Abstraction of the gene expression process and model derivation	23
1.2.1 Abstractions and simplifications	23
1.2.2 Derivation of the deterministic gene expression model	25
1.3 Stochastic dynamics for gene regulatory networks using Master equations	29
1.3.1 Linear one-step birth-death processes	31
1.3.2 Master equation for biochemical networks	33
1.3.3 Implemented numerical algorithms for stochastic analysis	34
1.3.4 Noise qualitatively changes gene network dynamics	39
1.3.5 From stochastic Master equations to ordinary differential equations: bridging the gap with Ω -expansion	40
1.3.6 SDEs and Master equations	42
1.4 Pioneering <i>in vivo</i> synthetic circuits	43
1.5 Transferring concepts between the disciplines	46
1.5.1 A condition for oscillation	47
1.5.2 A condition for switching behavior	48
1.6 Application of control theory in biology	49
1.6.1 Use of feedback control concepts to explain perfect adaptation in natural biological systems	50
1.6.2 Control theoretical concepts for synthetic biology engineering	52
2 Genetic oscillator design around unstable periodic orbits	54
2.1 Introduction	54
2.2 The generalized repressilator model	56
2.2.1 Ordinary differential equations	56
2.2.2 The stationary state	57
2.3 Numerical methods for oscillator design	61
2.3.1 Numerical simulations and analysis of the dynamics	61
2.3.2 Global robustness analysis and control aspects of quasi-stable oscillations	62
2.3.3 Robustness analysis and shape similarity	63

2.3.4	Basic Floquet theory and quasi-stable oscillating orbits under small perturbations	65
2.4	Results	67
2.4.1	Stable and quasi-stable oscillations in the generalized repressilator	67
2.4.2	Traveling wave structure of the periodic solutions	68
2.4.3	Robust induction of quasi-stable oscillations in the deterministic regime	71
2.4.4	Stochastic oscillations in even rings and readout-based control	73
2.5	Discussion	75
3	Analytical conditions for Hopf bifurcations in uni-directionally coupled rings	80
3.1	Introduction	81
3.2	Analytical bifurcation analysis of generalized repressilator and UPOs	83
3.2.1	Analytical conditions for the appearance of unstable periodic orbits	88
3.3	Symmetries and kinks in the generalized repressilator model	91
3.4	Spatio-temporal symmetries of unstable periodic orbits of the generalized repressilator model	91
3.4.1	Unstable periodic orbits as traveling kink solutions	93
3.5	Kink dynamics	95
3.6	Probabilities for long oscillating transients and distribution of their lengths	97
3.7	Comparison to unidirectionally coupled flux-gate magnetometers	98
3.8	Discussion	99
4	Influence of stochasticity on the dynamics of the generalized repressilator	104
4.1	Introduction	104
4.2	Time scale separation in the master equation	106
4.2.1	Hill-type variable reduction for activated transcription	107
4.2.2	Monod-type variable reduction for repressed transcription	111
4.2.3	Contributions from higher moments	113
4.2.4	Numerical validation of the approximation scheme	113
4.3	The generalized repressilator model in the Master equation formalism	116
4.4	Reaction rates dependence on the system size Ω for the generalized repressilator model	116
4.5	Transient oscillations in the stochastic setting	119
4.6	Analysis of Hopf bifurcations in the stochastic setting	120
4.6.1	Induction of Hopf bifurcations in stochastic setting	120
4.6.2	The expansion in the inverse system size Ω and stochastic resonance around the bifurcation point for stable limit cycles	121
4.7	The quality of stochastic oscillations	124
4.8	Discussion	126

5	Model-free algorithmic induction of optimized control strategies for genetic networks	129
5.1	Motivation for feedback control in biology	130
5.2	Methods Description	133
5.2.1	Fitted-Q - the reinforcement learning algorithm	135
5.2.2	Effective multi-dimensional regressor - extremely randomized trees	138
5.2.3	Classical control problems	140
5.3	Application on biological examples	144
5.3.1	Controlled reversal of a genetic switch	146
5.3.2	Controlling even gene rings around unstable periodic orbits	150
5.4	Discussion	153
5.5	Appendix: constants for algorithmic realization	156
6	Summary and perspectives	157
6.1	Synthetic engineering and oscillator design	157
6.2	Transient dynamics and transferable concepts	159
6.3	Stochastic analysis of bifurcations and (Ω, c) -plane	161
6.4	Readout-based control for gene transcription networks	162
	Bibliography	176

List of Figures

1.1	Gene expression in bacteria	19
1.2	One-step stochastic birth-death process	31
1.3	The algorithm Dominated Coupling From The Past (DCFTP)	38
1.4	Qualitative difference of the stochastic version for a genetic switch compared to the deterministic model.	41
1.5	Pioneering synthetic plasmid designs of the genetic toggle Gardner et al. (1) and a genetic oscillator Elowitz and Leibler (2)	45
1.6	Classical feedback control	50
2.1	Steady state of the generalized repressilator model	58
2.2	Periodic solutions and travelling waves in the generalized repressilator model	60
2.3	Global robustness analysis	63
2.4	Oscillation shape variations under global robustness	64
2.5	Poincaré map illustrating escape from a periodic orbit	65
2.6	Quasi-stable periodic orbits under small perturbations	66
2.7	Numerical bifurcation analysis and unstable periodic solutions of repressilator rings with even number of genes.	69
2.8	Numerical bifurcation analysis of repressilator rings with odd number of genes	70
2.9	Induction of oscillations in even rings in the deterministic regime	74
2.10	Stochastic oscillations in even rings and readout-based control	76
3.1	Global bifurcation analysis of the generalized repressilator model	83
3.2	Spatio-temporal symmetries in the generalized repressilator model	90
3.3	How symmetries and kinks are related to bifurcations in the generalized repressilator model	92
3.4	Kinks in the repressilator model.	94
3.5	Kink density in UPOs of the generalized repressilator model	96
3.6	Kink dynamics in the generalized repressilator model.	102
3.7	Numerical analysis of oscillating transients starting from random initial conditions	103
4.1	Numerical validation of time scale separation for activated transcription	115
4.2	Numerical validation of time scale separation for repressed transcription	117

4.3	From deterministic to stochastic regime in the generalized repressilator model	118
4.4	The length of transient oscillations in the stochastic setting	120
4.5	Onset of oscillating state in stochastic realization of the generalized repressilator model	122
4.6	Onset of oscillating state in stochastic realization of the generalized repressilator model	123
4.7	Stochastic resonance in the generalized repressilator model	125
4.8	Dependence of oscillation quality on the system size Ω	126
5.1	Readout-based control for genetic networks.	132
5.2	Algorithmic elucidation of optimized control techniques from a data matrix	136
5.3	The regression method "EXTremely RAndomized Trees (Extra-Trees)"	139
5.4	Algorithmic estimations of the optimal control for the benchmark "Left-Right"	141
5.5	Classical control problem "Car on the Hill"	142
5.6	Algorithmic estimations of the optimal control for the benchmark "Car on the Hill"	145
5.7	Optimized reversal of a genetic switch	149
5.8	Optimized tracking of an unstable periodic orbit in the generalized repressilator model	152
5.9	Partial observability: 2-genes keep a 10 gene ring in the oscillating state	154
6.1	Length of oscillating transients in rings of different sizes starting from comparable initial conditions.	160

Acknowledgements

In the first place, I owe my deepest gratitude to my adviser Prof Mauricio Barahona, with whom it has been a pleasure to write papers and have inspiring discussions. His brilliant ideas, encouragement, and support made the three years of my PhD a fruitful and challenging life-experience.

The members of Bioengineering department at Imperial provided wonderful environment for this interdisciplinary PhD offering a collaborative platform for theoreticians and experimentalists. In particular, I appreciated the discussions with Prof Kim Parker, Prof Michael Schneider, Dr Karen Polizzi, Dr Manos Drakakis, and Dr Guy-Bart Stan. I am also thankful to Martin Holloway and my adviser for the opportunity to participate in the departmental teaching.

I appreciate the criticism from Prof M. Stumpf, Prof P. Freemont, and Prof M. Sternberg, who have always questioned and challenged all my ideas and actions, it has helped me to make the concepts convincing and grow both personally and as a scientist.

It is a pleasure to acknowledge members of Mauricio's group for being good office mates and providing friendship. I have also enjoyed encouraging scientific and philosophical discussions with my colleague Andreas Ipsen, with whom we shared the ups and downs of the common MSc/PhD program. And for the relaxing coffee breaks and stimulating conversations about god-and-the-world throughout the time at Imperial I am grateful to Dr Heather Harrington. Thank you guys for your delightful companionship!

My special thank goes to Daniela Römer, who gave me starting help in London providing advice and letting me live in her hotel for the first weeks. For emotional support and holding my back free during the key moments, I am deeply obliged to my partner, Holger Pfeil. Liebster Puhbärchen, grössten Dank für Deine Unterstützung, ohne Dich hätte ich es nicht geschafft und ich habe Dich lieb!

For the generous funding and opportunities to go to conferences and workshops I am thankful to the funding agency The Wellcome Trust. The PhD grant scheme offered by the Trust creates the leading researchers of tomorrow and I am grateful for the opportunity that I could be part of it.

Thank you everybody, none of this would be possible without your care!

Abstract

The focus of the thesis is the investigation of the generalized repressilator model (repressing genes ordered in a ring structure). Using nonlinear bifurcation analysis stable and quasi-stable periodic orbits in this genetic network are characterized and a design for a switchable and controllable genetic oscillator is proposed. The oscillator operates around a quasi-stable periodic orbit using the classical engineering idea of read-out based control. Previous genetic oscillators have been designed around stable periodic orbits, however we explore the possibility of quasi-stable periodic orbit expecting better controllability.

The ring topology of the generalized repressilator model has spatio-temporal symmetries that can be understood as propagating perturbations in discrete lattices. Network topology is a universal cross-discipline transferable concept and based on it analytical conditions for the emergence of stable and quasi-stable periodic orbits are derived. Also the length and distribution of quasi-stable oscillations are obtained. The findings suggest that long-lived transient dynamics due to feedback loops can dominate gene network dynamics.

Taking the stochastic nature of gene expression into account a master equation for the generalized repressilator is derived. The stochasticity is shown to influence the onset of bifurcations and quality of oscillations. Internal noise is shown to have an overall stabilizing effect on the oscillating transients emerging from the quasi-stable periodic orbits.

The insights from the read-out based control scheme for the genetic oscillator lead us to the idea to implement an algorithmic controller, which would direct any genetic circuit to a desired state. The algorithm operates model-free, i.e. in principle it is applicable to any genetic network and the input information is a data matrix of measured time series from the network dynamics. The application areas for readout-based control in genetic networks range from classical tissue engineering to stem cells specification, whenever a quantitatively and temporarily targeted intervention is required.

Papers and Manuscripts

- N. Strelkowa, G-B Stan, D. Ernst, M. Barahona. Applications of optimal control to genetic networks. *In preparation*.
- N. Strelkowa, M. Barahona. Onset of oscillations and stochastic resonance in the generalized repressilator model. *In preparation*
- N. Strelkowa, M. Barahona. Transient dynamics around unstable periodic orbits in the generalized repressilator model. *Provisionally accepted by Chaos*.
- N. Strelkowa, M. Barahona. Switchable genetic oscillator operating in quasi-stable mode. *Journal of the Royal Society Interface*, 7:1071-1082, 2010.
- N. Strelkowa, M. Lassig. Clonal interference in the evolution of influenza. *In review*.
- T. Toni, D. Welch, N. Strelkowa, A. Ipsen and M. P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6, 187-202, 2009.

About the Thesis

Synthetic biology and forward design of gene expression networks has infused classical engineers and physicists with new ideas and possibilities. It offers them the opportunity to undust and adapt their over several years collected know-how and contribute to a new scientific discipline. The key research outcomes in synthetic biology coming from this physical/electrical engineering source include the implementation of oscillators in living bacteria by Elowitz and Leibler (2) and by Sticker, Hasty and others (3). The first design is based on a uni-directionally coupled rings, which have been exploited in electrics and the second design is based on the classical reasoning of time scale separations and delays, which are typically necessary for oscillations in physical systems. Further contributions came from computer science. The idea on hierarchical organization has inspired creation of logical gates in living cells, where fundamental work has been done by Andrianantoandro, Weiss and co-workers (4).

However classical physics or engineering ideas can not be directly applied to biological systems, there are some bridges to be crossed beforehand. Main gaps between the well-predictable and operating in steady state classical systems and a synthetic gene network are high levels of noise and nontrivial transient behaviour. Well, I am one these lucky physicists, who has the change to benefit from synthetic biology progress and to develop methods for these noisy and out-of-steady state operating gene networks.

After the Introduction (chap. 1) I introduce and analyze in research chapter 2 a standard genetic model and develop methods, which are of a potential use for synthetic biology. This model, also known under the name "The generalized repressilator", is a network of genes repressing each other and ordered in a ring structure. This configuration has outstanding symmetrical properties allowing analytical calculations of steady states and also analytical bifurcation analysis. I

will show the importance of Hopf bifurcations previously ignored in the model and propose a protocol, which allows the creation of a switchable and controllable oscillator based on the transients originating from these bifurcations (5). It seems to be feasible to implement this structure in a real biological cell and we have a collaborative project here at Imperial, with an experimental group who are concentrating on creation of parts for this structure.

The in dept mathematical analysis of this highly symmetric gene configuration will be continued in the chapter 3. It turns out not only being interesting on its own right, but also because of the underlying mathematical connection to uni-directionally coupled ring elements, which are common in maths and physics. There is a mechanical counterpart to the genetic repressilator, which is the recently developed magnetometer by Bulsara and co-workers (6) showing similar transient oscillating states. Our results are extend-able to these kind of systems and therefore also have impact outside of the synthetic biology field. The generalized repressilator model and its ring symmetry can be thought of as a universal classical concept for switches and oscillators, which is transferable across disciplines.

One of the challenges for the transfer of classical engineering concepts to synthetic networks is the omnipresence of noise. A lot of effort has been spent on understanding the biophysical mechanisms and mathematical description of noisy behaviour in the gene expression (7, 8). For example, work by Swain and co-workers (9), who proposed a theoretical description of noise sources in terms of two orthogonal components: extrinsic and the intrinsic. This theoretical concept was supported by the experimental proof in bacterial populations shortly after (10). Mathematical description of eukaryote gene expression and the influence of noise is more complex, since effects of stochastic memory and "burst"-like behavior need to be taken into account (11). The main objective of these studies was to reveal the naturally occurring biophysical mechanisms as sources of stochasticity. In the third chapter of the thesis I will revise classical analytical methods assessing the influence and possible control of noise in synthetic circuits. The well-known van Kampen Ω -expansion will be revised and the regulation of noise levels in the generalized repressilator model will be investigated.

In chapter 5 I will present a reinforcement learning algorithm, which can be used for a general control of biological systems, where only limited information

on the underlying dynamics in the biological system is available and the system is noisy. This algorithm is based on the fundamental linear programming technique firstly introduced by Bellman in 1957 (12) and recently revised and adapted to cope with larger uncertainties by Ernst and co-workers (13). The use of the algorithm is showcased on a genetic switch and a genetic ring, both examples are special cases of the generalized repressilator model. For the genetic switch an optimized strategy for the reversal is inferred and for the gene ring a control around an unstable periodic orbit is shown.

This thesis presents an interdisciplinary research, where existing mathematical, computational and analytical techniques are adapted and further developed. The main objective, which is to design and influence biological system requires crossing borders between the classical disciplines and adapting an arsenal of techniques. In the hope that the reader will enjoy the variety in methods and connections between the disciplines which became clear during this thesis, I wish you a pleasant journey through the rugged and noisy bio-landscape!

Chapter 1

Introduction: Modeling and engineering in Synthetic Biology

Can we engineer biological systems in a predictable manner to perform new functions or to optimize already existing ones? And what are the underpinning principles behind the forward design of gene regulatory networks? These questions gave rise to a new engineering discipline - the synthetic biology (4, 14, 15). Inspired in many cases by electronic elements, simple gene networks have been designed to perform reproducible, low-level functions. The pioneering synthetic circuit implementations in 2000 include the genetic ring oscillator known as the repressilator (2) and the toggle switch (1). Since then progress has been made in experimental implementation of engineered genetic transcription networks during the last decades. Simple synthetic modules have been integrated into the complex machinery of the cell, as in the oscillator recently implemented in a mammalian cell (16), or interfaced with cellular pathways to induce particular responses, as in the construct where the toggle switch was connected to the SOS pathway to induce DNA protection mechanisms in *E. coli* when exposed to UV light (17). Synthetic circuits have also been used for medically relevant research as for example the study where genetically engineered *E. coli* have selectively invaded cancer cells (18). Similar principles have been exploited in the rational design of internal negative feedback operated in conjunction with external arabinose-driven positive feedback to produce cell-synchronized oscillations (3). This recent experimental progress infuses the minds of mathematicians and engineers with new design ideas for more complex artificial gene networks, which

would perform more sophisticated functions.

A successful design of an artificial network is often preceded by thorough mathematical modelling (7, 19), which usually includes abstraction of the biochemical process and its analytical time scale estimations. The abstraction steps include several assumptions on the biochemical reactions such as that the laws of mass actions are approximately valid or that the number of DNA copies in a cell is constant. After the abstraction the biochemical process can be coded in an ODE (ordinary differential equation) model amenable to numerical integration. Very often inherent stochasticity of the gene expression process generates a qualitatively different dynamic behaviour as the according deterministic model. Then the stochasticity can be explicitly included into the model using either the Master Equation formalism (20) for analytical estimations or Gillespie algorithm (21) in order to obtain stochastic time traces numerically.

Even though standard modelling techniques exist and a lot of progress in the theoretical description has been achieved during the last years, there are still gaps remaining to be bridged for a predictable forward design of gene expression parts. One of the difficulties arises if simple networks consisting of few genes are hierarchically interconnected as it is envisioned by creators of the MIT Registry for re-usable biological parts (22). The convergence to the steady states of larger transcriptional networks will be slow compared to the cell life-time and other gene expression reactions and therefore not only steady states but also transient effects must be explicitly considered in the design. Further complications for the successful or predictable designs arise because of the inherent stochasticity due to low copy numbers involved into the transcription process. Therefore transcriptional network designs either require robustness to noise (23, 24) or new concepts, which would explicitly take advantage of the noise presence. Finally the lack of quantitative information, such as transcription and translation constants and their dependence on the environmental conditions as for instance fluctuations in cell-feeding media, temperature etc, makes predictability of synthetic designs very difficult. In particular for practical tissue engineering quantitative information about transcriptional networks is required in order to infer interference protocols, with which a genetic network can be directed to a desired state. Following such protocols colonies of cells can be conducted to desired phenotypes in an effective manner.

In this thesis the generalized repressilator model (uni-directionally coupled genetic repressors ordered in a ring) is used to illustrate these difficulties and to demonstrate solution strategies. The appearance of non-trivial and long-lived transient dynamics in larger synthetic networks is showcased with the characterization of unstable periodic orbits (UPOs) in the generalized repressilator (thesis chapter 2). These UPOs are shown to be observable, long-lived oscillating transients, which are likely to feature in time-limited environments such as bacterial cells. These transients are interesting not only for the illustration of the non-trivial behavior in large genetic networks, but also from the bioengineering point of view. The uni-directionally and cyclically coupled genetic repressors can be used for the design of switchable and controllable genetic oscillators.

More mathematically rigorous characterization of the UPOs in the generalized repressilator model and the analytical conditions for the emergence of a cascade of unstable periodic orbits (UPOs) via Hopf bifurcations is given in chapter 3. These unstable orbits can be understood with concepts of travelling waves and kinks from classical physics. In the electrical engineering there are other uni-directionally coupled ring systems, as for instance Duffing oscillators or magnetic flux gates, which are well-studied and used for practical applications. Via bifurcation analysis and group theory, parallels are shown between the generalized repressilator model and the well-established ring symmetric structures. The oscillator design as a cyclic coupling of repressors appears to be an example for a transferable engineering principle which works in electrical and biological systems.

In the following thesis chapter 4 biochemical parameters are identified, which tuning can induce Hopf Bifurcations in the stochastic version of the generalized repressilator model. The stochasticity has a strong effect on the onset of stable limit cycle oscillations, where stochastic fluctuations influence both the onset of bifurcations and their quality. For a regime far away from the oscillation onset we show that in the generalized repressilator model with even number of genes noise seems to be constructive and has a stabilizing effect on the unstable periodic orbits.

After the dynamical characterization of the generalized repressilator in deterministic and stochastic regimes performed in the previous chapters, we use it as an engineering example to propose optimized interference strategies via external

signals. In general, the mechanisms behind the gene expression in both bacterial and eukaryotic cells are reasonably-well understood in qualitative terms. The gene expression dogma that DNA information is transcribed into *mRNA* and subsequently followed by the translation into proteins is considered to be textbook knowledge (25, 26). However, the descriptive understanding of the process is too uncertain for engineering purposes. In conventional approaches, a mathematical model for the network is tried to be obtained (see (27) for a classical procedure) and then, analysing the inferred topology, the network states can be defined and interference strategies can be designed for effective tissue engineering. Instead of this conventional approach a less ambitious, but potentially more practically relevant technique is proposed in thesis chapter 5. The assumptions for the technique to work are a descriptive understanding of the genetic network, a possibility to interfere with the network components such for example UV light activation (28, 29) of certain proteins (input) and a possibility to estimate a state of the relevant proteins (readout) (30). Based on the input-output samples in form of a data matrix, a machine-learning algorithm is proposed.

Classical oscillator engineering and biochemical mechanisms during gene expression in bacteria have motivated the creation of the generalized repressilator model. This chapter contains background knowledge on regulation of bacterial gene expression derived from the first biochemical principles (26) and the review of pioneering implementation in living cells of special cases of the generalized repressilator model.

1.1 Gene expression in bacteria

Synthetic genetic circuits use the natural gene expression resources of the cells to perform functions. If designing synthetic regulatory elements for bacterial cells from scratch, mathematical models for gene expression are required. The derivation of the models is based on the background knowledge about gene regulation principles, the law of mass action, and the intrinsic time scales involved into the associated biochemical reactions. In the following we review background knowledge on the biochemical reactions necessary for the model construction.

Mathematical equations for gene expression are based on the dogma that protein synthesis occurs in two steps: transcription and translation. In the first step

information stored in the DNA sequence is used for the production of messenger RNA (*mRNA*) Fig. 1.1 (a). This transcription step is followed by the translation of the RNA information to construct the protein product Fig. 1.1 (b). During the

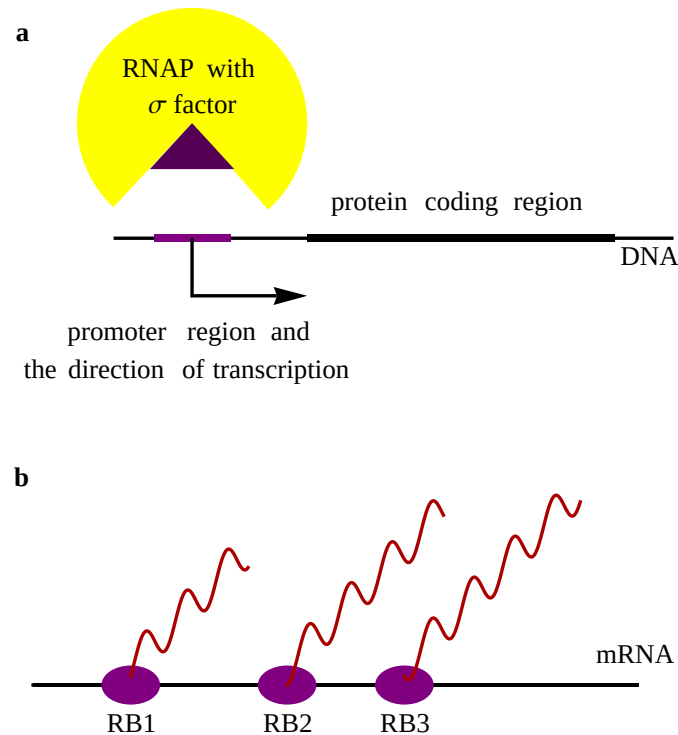


Figure 1.1: **Abstract view on gene expression in bacteria.** (a) Transcription initiation. The structure of a typical bacterial gene consisting of a promoter region (purple), which is -35 to -10 nb (nucleotide base pairs) upstream the coding sequence. The holoenzyme consisting of RNA polymerase, (RNAP yellow) and promoter recognizing σ factor (dark purple) is about to bind to the promoter region. The direction of transcription is indicated with the black arrow.

(b) Translation in bacteria. Three ribosomes RB1, BR2, RB3 are translating one *mRNA* template into proteins, which are indicated with dark red lines. Since ribosomes are in abundance, protein production rate can be approximated as a linear first order reaction, which is proportional to the *mRNA* concentration. The rate constant for the translation reaction can be larger than one, because several ribosomes can simultaneously translate one template.

first step a prokaryotic transcription initiation factor σ binds to RNA polymerase (RNAP) forming a holoenzyme. σ -factors direct the transcription machinery towards the gene promoter regions. There are 7 different σ -factors for *E. coli*, whereas one factor is highly conserved across bacteria. It is responsible for house

keeping during normal growth conditions. This house keeper finds the (-35 to -10 basepairs) upstream promoter region, whereas the other factors activate specific genes during extreme conditions: heat shock, salt stress etc. Engineered gene circuits can take advantage of the natural transcription machinery if they contain a promoter region recognizable by σ -factors.

For the mathematical modeling of biochemical reactions involved in gene expression we need to know in which amounts RNAP and σ -factors are present in the cell and if the complex formation is a limiting step in the transcription. RNAP is believed to be highly abundant protein (7000-10000 molecules per cell) and is not a rate limiting factor of any kind (26). The same is true for the σ -factor, so that the complex concentration can be assumed abundant.

The RNAP – σ -factor complex binds specifically to certain DNA sites - the promoter regions. The search for a promoter region is usually fast and takes less than 60s as well established by theoretical results and experimental single cell measurements (31). The promoter region is usually located a couple of nucleotide base pairs (-35 to -10) upstream the DNA coding region. Once the holoenzyme is attached to the promoter region the σ initiation factor is released and the elongation (transcription) starts. During the elongation the RNA polymerase moves down the DNA strand unwinding the double helix and allowing the ribonucleotide triphosphates to enter the complex and to elongate RNA transcript as a complement of DNA sequence. After the passing of the transcription complex the opened DNA re-winds to the original form and the RNA transcript of the passed region appears at the opposite site of ribonucleotide triphosphates entry. The elongation process continues until it is stopped by a ρ -protein or as in case of engineered plasmids by a terminating sequence. The time scale for the transcription of 100 long *mRNA* piece is about 100 seconds (19). We note that there is a small delay between the binding of the promoter site and the appearance of the final *mRNA* transcript, so that the *mRNA* production is a slow reaction. Finally, the transcription termination sequence causes the RNA polymerase to end the transcription process and to detach from the DNA. The detaching process occurs very fast $\approx 1s$. The number of *mRNA* transcripts for a particular protein is quite low compared to other chemicals involved in the transcription process, it is on the order of dozens.

After the schematic sketch of bio-molecular processes, gene regulation on the

microscopic level can be explained. The transcription of genes is regulated by activators such as for example σ initiation factor. The RNA polymerase (RNAP) must be complexed with the activator σ in order to recognize the binding site for the transcription initiation. And that means the gene is not expressed without the according activator. Moreover, the repression of transcription of certain genes can be achieved if an appropriate protein tightly binds to the promoter region and forbids in this way the binding of the holoenzyme for transcription initiation. This inhibition mechanism requires that the shape of the inhibitor protein is tailor-made for the according binding site. Such proteins develop through evolution or can be engineered. The expression inhibition mechanism can be and indeed was used in synthetic circuits to implement repression (1, 2).

Summarizing, the transcription step majorly depends on the concentration of the regulatory proteins and involves biochemical reactions on two different time scales: the attachment and the detachment of the regulatory proteins is fast compared to the *mRNA* production in the elongation step. The overall number of *mRNA* transcripts is low. We will use this information by the model derivation for the gene expression process.

The second step in the gene expression is the translation of messenger RNA to protein sequence. For the translation to take place several bio-molecular agents need to be assembled: 2 subunits (large and small), the messenger RNA produced during transcription, the tRNA (transfers a specific amino acid to a growing polypeptide chain) and the initialization factors (F1-3). All agents are present in the cytoplasm at approximately constant (on the time scale of interest) concentrations, but not the *mRNA*. *mRNA* is present in low copy numbers and the concentration changes on relevant time scales. So the number of *mRNAs* is the rate limiting factor of the translation process and we can neglect the dynamics of other molecules in the model and consider the translation being proportional to the amount of *mRNA*. Once the stepwise assembly is accomplished the elongation of the polypeptide chain can start 1.1 (b). As pointed out in 1.1 (b) more than one ribosome unit is attached to the *mRNA* template at the same time. The elongation of a 30 peptide long chain takes $\approx 30s$ (19), so it is a slow reaction occurring on similar time scales as the *mRNA* creation. Elongation stops whenever one of the 3 translation termination codons (*UAG*, *UAA*, *UGA*) enters the ribosome and the time scale involved is $\approx 1s$. The overall number of specific proteins is larger than

the number of *mRNAs* and is on the order of hundreds or thousands.

mRNAs and proteins produced in the transcription and translation reactions have limited lifetimes, which means they are destroyed regularly by cell internal processes. The messenger RNA transcripts are degraded by ribonucleases (RNase E,G). The exact process and which ribonucleases are more important is still disputed (32, 33), but there are strong indications that there is only one initiation process. Once this process happened *mRNA* can not be used as a protein transcript any longer and is degraded through a decay pathway. The decay molecules involved are present in approximately constant amounts in the cell cytoplasm and therefore the amount of transcripts degraded per time can be assumed proportional to the *mRNA* copies and the proportionality constant is reciprocal to the mean life time of *mRNA*, which is $\approx 2 - 4$ min in *E. coli* (34). Proteins are longer lived than *mRNAs* with the average lifetime of 40 – 60 min (2, 35, 36). Through similar degradation pathways as *mRNAs* the proteins are degraded by proteases and the degradation can be approximated as a biochemical single step reaction for the same reasons.

Gene expression involves several biochemical species. Therefore it makes sense from the synthetic biological point of view to use this naturally occurring machinery for synthetic genetic circuits. For transcription and translation to work with synthetic parts natural constraints such as start-, end codons, promoter and repressor sites, etc must be respected. Referring back to implemented examples 1.5 (a) the promoter regions must be designed in a way that they attract the transcription machinery if the inhibitor protein is not bounded to the promoter region. Also, each artificial gene must contain a transcription terminator (T_1T_2 black-lined boxes show the position of terminators explicitly for the toggle switch).

Summarizing, we can identify four major reactions dictating the time scale of the gene expression and regulation

1. production of *mRNA* involving fast binding and unbinding of activators/repressors to the DNA and majorly depending on the activator/repressor concentrations
2. production of proteins majorly depending on the *mRNA* copy numbers in the cell
3. *mRNA* degradation majorly depending on the amount of *mRNA* in the cell

4. protein degradation majorly depending on the amount of proteins in the cell

1.2 Abstraction of the gene expression process and model derivation

1.2.1 Abstractions and simplifications

The underlying gene expression process involves several bio-chemical steps, and some of them are more important for the qualitative dynamics than others. From the gene modelling perspective we would like to produce a manageable abstract model, which captures the main characteristics of the process. We assume the following:

1. **Laws of mass action.** Bacterial cytoplasm is highly packed with proteins (340mg/ml) (26), which can lead to molecular overcrowding and gel behavior of the cytoplasm. However, if molecules involved in the model process have a relative small size (smaller than 400kDa) for the assumption of free diffusion, random collision, and well-mixture of the molecules to hold in the first approximation. The effects of molecular overcrowding and gel behavior are second order effects.
2. **Synthetic genes are present in constant amounts per cell.** There are actually two ways how a synthetic genetic element can be incorporated into the bacteria. The first possibility is to encode the synthetic circuit onto a plasmid, which will be taken up by the cell. Plasmids do replicate inside the cell and this process must be controlled. For the classical examples of the toggle switch and the repressilator (1, 2) replication control *ECol1* was encoded on the synthesized gene. A different way to encode synthetic elements is using expression cassettes, which would incorporate the artificial DNA directly into the bacterial genome. This procedure is more controlled and ensures the constancy of the gene copy numbers.
3. **Engineered proteins bind inhibitory to the DNA regulatory regions without leakiness.** Basal transcription can easily be incorporated into the models

as an additive term, however this effect should be negligible, if the engineered proteins are designed in a reasonable way.

4. **Binding and unbinding of the regulatory proteins and DNA is much faster than transcription and translation.** The dynamics of the intermediate complex between the DNA and the regulatory protein occurs on the much faster time scale than other reactions. Taken the time scale separation into account we obtain Michaelis-Menten-like kinetics for the transcription rate. Standard mathematical derivations are provided in detail in Sec. 1.2.2, 1.2.2.
5. **Protein and *mRNA* degradation are first order reactions.** There are naturally occurring proteins, which degrade in several steps and in particular for eukaryote cells several steps are involved in folding/creation as well as degradation reactions (11). Engineered bacterial proteins are degraded by simpler mechanisms and we approximate their degradation as a first order reaction.
6. **Reaction constants are not time dependent.** Effectively we neglect cell growth and death and the change of the cell environment as the cells get older (33). These effects would influence the reaction constants and in order to account to some extent for these changes we perform robustness analysis as for instance in Sec. 2.4.3.
7. **All molecules are present in large amounts.** This constraint is only assumed for the deterministic description with ODEs. As it turns out stochastic fluctuations do have a qualitative effect on the overall dynamics and their role for the generalized repressilator model is explored in Chap. 4.
8. **Cell division is a renormalization step.** The "big elephant in the room" every modeller is aware of that probably has the strongest effect on the synthetic circuit is the cell division, which for *E. coli* happens approximately every 15-60 minutes depending on the feeding conditions. The dynamical features such as switch toggle (1) or oscillation period (2, 3) take hours, which means they are spread over several cell generations. Formally we can assume that the cell division is strictly symmetric and the reactions are renormalizable (see for instance (37) chap. 8 for definition of renormalization group), then

the deterministic equations continue to be valid for the daughter cells. Effectively it means that the circuit dynamics do not depend on the system size, but scale accordingly. These assumptions are often not fulfilled, because cell division and plasmid distribution between the daughter cells are not symmetric and it is not precisely clear how biochemical reactions scale with the cell size.

1.2.2 Derivation of the deterministic gene expression model

In the previous sections the biochemical reactions during the gene expression process as well as necessary abstractions lead to the conclusion that four major reactions, namely transcription of *mRNA*, degradation of *mRNA*, translation resulting in the protein production, and protein degradation, are the most important biochemical steps during the gene expression (see section 1.1). We will derive a deterministic ordinary differential equation model for gene expression based on these reactions.

Activated transcription

To start with, we will derive an expression for *mRNA* birth rates, where a protein acts as transcription enhancer or activator. The bio-molecular steps are as following: certain number of proteins *P* binds to the DNA regulatory region *D* with the rate k_1 (≈ 0.17 (38)) forming the intermediate complex $DP^m \equiv X$. The reversed reaction occurs approximately with the rate k_{-1} (≈ 10 (38)). Then the protein is complexed with the regulatory region, *mRNA* is produced at the much slower rate k_2 (≈ 0.015 (38)). This reaction is irreversible, since once *mRNA* is produced, it remains in the system until it is degraded. The steps for enzymatic activation with *P* can be summarized in the following chemical reaction scheme:



From the equation (1.1) it is clear that *mRNA* is produced proportionally to the complex concentration: $\frac{dmRNA}{dt} = k_2[X]$. The complex concentration is difficult to measure and in general we would like to express the time evolution of *mRNA* con-

centration in terms of its activating protein concentration. We make a reasonable assumption that the initial concentration of the proteins $[P]_0$ is much larger than the concentration of the DNA binding sites $[D]_0$, i.d. $[D]_0/[P]_0 \approx 0$, and assume the quasi stationarity of the complex $d[X]/dt \approx 0$, since the binding/unbinding (rates k_1, k_{-1}) occurs on the much faster time scales compared to the *mRNA* production (rate k_2):

$$\begin{aligned} \frac{d[X]}{dt} &= 0 = k_1[D][P]^m - [X](k_{-1} + k_2) \\ [X] &= \frac{k_1[D][P]^m}{k_{-1} + k_2} \\ &\equiv \frac{[D][P]^m}{K_m}. \end{aligned}$$

The constant K_m is usually referred to as **Michaelis-Menten constant**, after the authors (these are indeed two people :) who firstly introduced the concept for chemical enzymatic reactions. We can get rid of $[D]$ the unbounded DNA using the fact that the overall DNA concentration is equal to the sum of bounded and unbounded DNA $[D]_0 = [D] + [DP]$:

$$\begin{aligned} [X] &= \frac{([D]_0 - [X])[P]^m}{K_m} \\ \rightarrow \frac{d[mRNA]}{dt} &= k_2[X] = k_2[D]_0 \frac{[P]^m}{K_m + [P]^m} \equiv V_{max} \frac{[P]^m}{K_m + [P]^m}. \end{aligned}$$

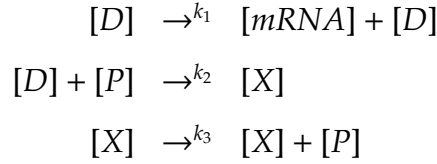
The time scale separation in this case allowed us to express the time evolution of *mRNA* concentration depending on the activator concentration.

In general it is possible that several proteins (lets denote their number by m) cooperatively bind to a promoter region of a gene. Their degree of cooperativity and the sequence in which the proteins bind is often not known. Therefore, the exponent does not necessarily mean that indeed m enhancer/repressor proteins are involved, but can be thought of as a fitting parameter and can be a positive real number. In this case, the exponent is called Hill coefficient and the activation/repression functions are called Hill-, Monodfunctions (39–41). Experimentally it was found that $m = 1$ often explains the results worse than larger exponents, see for example (2), who used $m \approx 2$.

This deterministic description has an analogue in the stochastic case and it is extensively treated in Sec. 4.2.1.

Repressed transcription

Similar derivation as for the activated transcription can be conducted for the repressed transcription. Let the concentration of the inhibitor protein be P , the number of DNA copies D_0 , the number of DNA regions not bound by the repressor D , the DNA-protein complex X . As for the case of activated transcription the equality $[D_0] = [X] + [D]$ holds here as well. A simplified model for enzymatic inhibition with P can therefore be described with the following chemical reaction scheme:



This leads to the following set of differential equations, according to the classical mass-action kinetics:

$$\begin{aligned} \frac{d[D]}{dt} &= -k_2[D][P] + k_3[X] \\ \frac{d[X]}{dt} &= k_2[D][P] - k_3[X] \\ \frac{d[mRNA]}{dt} &= k_1[D] \end{aligned}$$

Since the binding/unbinding of the inhibitor protein to the DNA is very fast compared to the $mRNA$ creation in the same way for the activated transcription, we assume that the complex will reach the equilibrium state fast $d[X]/dt \approx 0$. With the quasi-stationarity assumption, we can get rid of the variable $[X]$ and $[D]$ and describe the transcription inhibition in terms of inhibitor protein concentrations.

$$\begin{aligned} \frac{d[X]}{dt} &= 0 \\ k_2[D][P] &= k_3[X] \\ k_2([D_0] - [X])[P] &= k_3[X] \\ [X] &= \frac{k_2[D_0][P]}{[P]k_2 + k_3} \end{aligned}$$

where the last line is the average complex concentration, which will be reached quickly due to fast binding and unbinding.

Then the time evolution of the *mRNA* can be written in terms of inhibitor amount as:

$$\begin{aligned}\frac{d[mRNA]}{dt} &= k_1[D] = k_1([D_0] - [X]) \\ &= k_1([D_0] - \frac{k_2[D_0][P]}{[P]k_2 + k_3}) \\ &= \frac{[D_0]k_1}{1 + [P]k_2/k_3}\end{aligned}$$

which leads to a Hill-type non-linearity. This derivation is shown for the monomers, for dimers with cooperative binding the Hill coefficient can be approximately 2.

Degradation

We assume that *mRNAs* and proteins are degraded via first order reactions, which means linearly dependent on their current concentration:

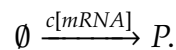


Here X is either *mRNA* or proteins concentration and d denotes the according degradation constants. The change of concentration due to this reaction can be described with the ODE

$$\frac{d[X]}{dt} = -d[X]$$

Translation

Translation of *mRNA* templates into proteins is an enzymatic reaction, where the translation is linearly dependent on *mRNA* concentration, the templates are unchanged after the reaction and protein concentration is increased. As pointed out earlier other necessary components besides *mRNA* are abundantly present in the cell and therefore, this reaction can be thought of as an enzymatic production from the source, where the limiting step is the amount of the *mRNA*:



The according concentration change per time due to this reaction is then

$$\frac{d[P]}{dt} = c[mRNA].$$

Full deterministic model

Putting together either activated or repressed transcription (the mutually exclusive terms are denoted in parenthesis $(\dots) : (\dots)$), translation, and degradation reactions, we obtain the full deterministic model for gene expression:

$$\begin{aligned}\frac{d[mRNA]}{dt} &= \left(\frac{V_{max}[P]^m}{K_m + [P]^m} \right) : \left(\frac{c_1}{K_m + [P']^m} \right) - c_2[mRNA] \\ \frac{d[P]}{dt} &= c_3[mRNA] - c_4[P].\end{aligned}$$

The deterministic description derived here will be used to construct models for networks of different topologies and study their steady state and transient dynamics. However, as already pointed out in the biochemical abstraction summary 1.2 gene expression is associated with non-negligible amount of noise due to low copy numbers of the molecules, in particular during the transcription reactions. There is a theoretical formalism describing the evolution of probability functions in time called Master equation and it will be used to model network topologies in this thesis taking the inherent gene expression stochasticity into account. We will use the notation introduced in the classical text book (20).

1.3 Stochastic dynamics for gene regulatory networks using Master equations

Molecular reactions occurring in classical chemical or physical systems can be seen as discrete events randomly occurring in time (see for instance (20), chap. V). If the number of molecules involved is very high, the system can be well described with ordinary differential equations, but if the number of molecules is low, the random nature becomes apparent and the deterministic description is no longer appropriate. To illustrate the difference imagine gas molecules, which are enclosed in a reservoir and react by collision. If the density of molecules is high, then several collision reactions will be happening per time period, then the fact

1.3 Stochastic dynamics for gene regulatory networks using Master equation 20

that the reactions are discrete events becomes irrelevant for the description of the process, and the reaction rate can be well approximated as a deterministic rate law. A different scenario occurs if the density of gas molecules in the reservoir is very low. Discrete collision events can be observed, which mathematically can be captured with the so called Master Equation, describing time evolving probability distributions instead of the deterministic ordinary differential equations.

Biochemical reactions for natural and engineered gene networks involve several variables such as the number of DNA copies ($\mathcal{O}(1)$) or the number of *mRNA* molecules ($\mathcal{O}(10)$), which are present at low densities in the cell. Therefore, the randomness due to the low copy numbers needs to be taken into account, if forward designing synthetic networks. In order to explore this aspect, we will use the Master equation formulation (see (20)). Instead of a real valued concentration vector in the deterministic formalism, we introduce the random variable $\mathbf{n}(t) \in \mathbb{N}^s$ describing the time dependent state of the system, which means the number of molecules for each of the s biochemicals (species) at the time t . A state $\mathbf{n}$ can appear at a time t with the probability $P(\mathbf{n}, t)$, so that the evolution of the system will be given as evolution of probability function $P(\mathbf{n}, t)$. The probability function is discrete in $\mathbf{n}$ the species space, but continuous in time. Hence, a general form of a Master equation:

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{\mathbf{n}'} W_{\mathbf{n}\mathbf{n}'} P(\mathbf{n}', t) - W_{\mathbf{n}'\mathbf{n}} P(\mathbf{n}, t)$$

can be seen as a gain-loss equation for the probabilities of the separate states. The transition probability from the state $\mathbf{n}'$ to the state $\mathbf{n}$ per unit time is given by the transition $W_{\mathbf{n}\mathbf{n}'}$. $W_{\mathbf{n}\mathbf{n}'}$ s are conditional probabilities given that the system state at time t is $\mathbf{n}'$ what is the probability that it will transform to the state $\mathbf{n}$. The first term of the equation is the gain to the state $\mathbf{n}$ by all other states and the second the loss of $\mathbf{n}$ caused by the transitions into the other states.

For modelling of the chemical reaction networks we require mainly four types of reactions Gadgil et al. (42):

1. **creation from the source** $\emptyset \rightarrow \mathcal{M}_j$ If we consider the cell resources being a large bath with environmental chemicals abundantly present in the cell, then reactions depending exclusively on abundant chemicals can be modeled as a creation from the source.

2. **deletion** $\mathcal{M}_j \rightarrow \emptyset$ The degradation of proteins and *mRNAs* can be modelled as a chemical deletion reaction.
3. **conversion** $\mathcal{M}_j \rightarrow \mathcal{M}_i$ The conversion of a chemical $\mathcal{M}_j$ into a chemical $\mathcal{M}_i$ either by itself or in dependence of abundantly present resources is associated with this reaction. For gene expression processes this reaction can be used if for example a post-translational protein modification occurs (26). However, we will not describe the gene expression process in such detail.
4. **catalysation** $\emptyset \xrightarrow{\mathcal{M}_i} \mathcal{M}_j$ A species $\mathcal{M}_i$ catalyzes the formation of $\mathcal{M}_j$ from a source and remains itself unchanged during the process. With this type of reaction the core gene regulation processes of transcription repression and transcription activation can be modeled. We will derive the stochastic version of the deterministic enzymatic-like reaction (see section 1.2.2) in the chapter 4.

1.3.1 Linear one-step birth-death processes

For the biochemical reactions involved in the gene regulation we need to consider a class of $W_{nn'}$, where the transitions occur between the neighbouring states. So in a time period δt a biochemical species is either created or deleted. Such processes are called one-step birth-death processes (see Fig. 1.2). In order to describe those

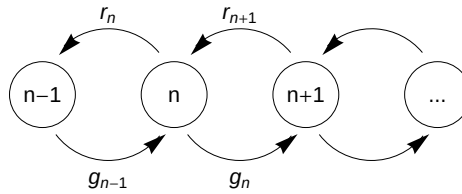


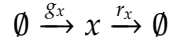
Figure 1.2: One-step stochastic birth-death process illustration as in van Kampen (20)

processes conveniently we introduce the creation/annihilation operators (20):

$$\begin{aligned}\mathbb{E}f(n) &= f(n+1) \\ \mathbb{E}^{-1}f(n) &= f(n-1)\end{aligned}$$

1.3 Stochastic dynamics for gene regulatory networks using Master equation 13

The schematically illustrated probability flows in Fig. 1.2 correspond to the reaction scheme:



where the rate functions r_x, g_x are linear in x . Using the creation/ annihilation operators the according master equation is then given by

$$\dot{p}_n(t) = (\mathbb{E} - 1)r_n p_n + (\mathbb{E}^{-1} - 1)g_n p_n \quad (1.2)$$

with $\mathbb{E}, \mathbb{E}^{-1}$ one dimensional step operators.

Stationary State Calculation

The stationary state in the stochastic setting is not given by a fixed point, but by a probability distribution. For a simple one-step birth-death process as illustrated by Eq. 1.2 and Fig. 1.2 with linear transitions: $r_n = k'n$ and $g_n = kn_A \equiv \text{const}$ the stationary distribution can be calculated analytically by recursion. For that we set the right hand side of the equation 1.2 to zero $\dot{p}_n(t) = 0$, similarly to the deterministic case, and obtain the following condition:

$$\begin{aligned} 0 &= k'(\mathbb{E} - 1)np_n + g_0(\mathbb{E}^{-1} - 1)p_n \\ &= (\mathbb{E} - 1)(k'np_n - \mathbb{E}^{-1}g_0p_n) \\ k'np_n &= \mathbb{E}^{-1}g_0p_n \\ k'np_n &= g_0p_{n-1} \\ p_n &= \frac{g_0p_{n-1}}{k'n} \\ p_n &= \left(\frac{g_0}{k'}\right)^n \frac{1}{n!}p_0^s. \end{aligned}$$

This recursion relationship is simple and the solution has been guessed and verified. From the normalization constrain we obtain an expression for p_0^s :

$$\begin{aligned} \frac{1}{p_0} &= 1 + \sum_{n=1}^{\infty} \left(\frac{g_0}{k'}\right)^n \frac{1}{n!} \\ \frac{1}{p_0} &= 1 + e^{g_0/k'} - 1, \text{ (boundary started by 1)} \\ p_0 &= e^{-g_0/k'} \end{aligned}$$

This leads to the well known result, that the stationary distribution of one-step processes has a Poissonian distribution:

$$p_n = \left(\frac{g_0}{k'}\right)^n \frac{1}{n!} e^{-g_0/k'} \equiv \text{Poisson}(g_0/k')$$

This recursion relationship was easy to solve, in general however, the solutions of recursion relationships are not obvious.¹.

If the stationary distribution can not be obtained analytically, there is a numerical technique called DCFTP (Dominated Coupling From the Past) (43), which allows the calculation of the stationary distribution for a stochastic system. The stationary state calculated here is also the necessary input function for the so called dominating process in that algorithm (see section 1.3.3).

1.3.2 Master equation for biochemical networks

Using the one-step birth-death processes and the four types of reactions introduced earlier, we can write the Master equation for a general biochemical network:

$$\begin{aligned} d_t P(\mathbf{n}, t) = & \sum_i \left(K_{ii}^I (\mathbb{E}_i^{-1} - 1) P(\mathbf{n}, t) \right. \\ & + \sum_j \left(K_{ij}^{IV} (\mathbb{E}_i^{-1} - 1) \right. \\ & + K_{ij}^{III} (\mathbb{E}_i^{-1} \mathbb{E}_j) \\ & \left. \left. - K_{ij}^{II} (\mathbb{E}_i - 1) \right) n_j P(\mathbf{n}, t) \right) \end{aligned}$$

where K^i are matrices according to the 4 reaction types and $\mathbb{E}, \mathbb{E}^{-1}$ are multi-dimensional step operators defined by

$$\begin{aligned} \mathbb{E}_i f(n_1, n_2, \dots, n_i, \dots) &= f(n_1, n_2, \dots, n_i + 1, \dots) \\ \mathbb{E}_i^{-1} f(n_1, n_2, \dots, n_i, \dots) &= f(n_1, n_2, \dots, n_i - 1, \dots) \end{aligned} \tag{1.3}$$

¹There are specific examples where hypo-geometric functions can be successfully used as an Ansatz. In addition to this sporadic special examples, there is a group of networks containing exclusively the following set of reactions, which can be treated analytically (42) as outlined before. Since the derived expression model contains non linear catalytic reaction it does not fall under this category and must be solved numerically.

1.3 Stochastic dynamics for gene regulatory networks using Master equation 24

on the support $\mathbf{n} \in \mathbb{N}^s$. The main advantage of the formalism is that the probability distribution is continuous in time and discrete in the number of molecules. Low copy number reactions can be modeled as discrete transitions from one state to the next. Therefore, this formalism will be used for the stochastic modelling of gene expression.

1.3.3 Implemented numerical algorithms for stochastic analysis

The introduced Master equation formalism capture finite number fluctuations in physical, chemical, and biological molecular reactions. Unfortunately, Master equations with nonlinear transition probabilities W_{mn} , which will arise in gene expression network models, can not be solved analytically and efficient numerical procedures are required in order to obtain samples from time evolution of Master equations.

Time traces and steady states of Master equations 1.3 can not be solved analytically in the most cases and an efficient numerical simulation techniques are required. The algorithm "Gillespie on the time grid" for stochastic time trace simulation and "DCFTP" for calculation of stochastic stationary probability distributions have been implemented in the programming language C and used throughout the thesis. In the following these techniques are reviewed.

Gillespie on the time grid

For numerical simulation of Master equations a classical Gillespie algorithm (21) was used with a minor change, which allowed time traces output in equidistant intervals. The overall program architecture is designed containing a basic model for a gene network, where only the connectivity structure for the network is a required input. The implementation was used to carry out all stochastic numerical simulations in this thesis and also as the subroutine program for DCFTP (Dominated Coupling Form The Past) algorithm, which is described below.

Algorithm description

Initialization

- specify the number of genes n of the network

- define initial conditions for the system, which means set the initial concentrations for *mRNAs* (m_i) and proteins (p_i) $\vec{x}_0 = \{m_1(0), p_1(0), m_2(0), p_2(0), \dots, m_n(0), p_n(0)\}$
- define the gene network matrix $genCon_{ij}$, which contains -1 if gene i represses gene j , 1 if gene i activates gene j and zero otherwise
- initialize the random number generator storing the seed

Classical implementation of the Gillespie algorithm produces time traces of the species counts (proteins, *mRNA*, or other molecules), which are evaluated at the process inherent stochastic times. That means the system state recordings are not taken at the equidistant intervals. There is a simple extension in Wilkinson (44) book, which works as follows: while the network is running, it calculates all reactions as the standard Gillespie, but dynamically decides to store the state of the system only at the specified time points. This simple trick makes the implementation much faster, less memory consuming and operation such as average taking over several runs, Fourier transformations can be efficiently performed.

There is a trade off to this simplification and that is the arbitrary choice of the time increment δt . It must be chosen small enough, so that the dynamics of the system are not missed, but at the same time large enough to minimize the size of the output. By experience appropriate values are obtained around $L/100$ with L being the characteristic time length of the system (for example time period of the inherent oscillations).

This leads to the following overall procedure, where gT is the counting "grid-time" variable and gO is the species output matrix containing the system state at equidistant points.

1. calculate the time dependent reaction propensities $a_v = c_v h_v$ according to the transcriptional model as well as their sum $A = \sum_{v=1}^{4n} c_v h_v$
2. generate two uniformly distributed random numbers r_1 and r_2
3. advance time: $t = t + 1/A \cdot \ln(1/r_1)$
4. This is the discussed extension of the classical algorithm (44): *in case the new time has exceeded the current grid time $t \geq gT$*
 - store the current system state to the output matrix $gO[gT/\delta t] = \vec{x}$

1.3 Stochastic dynamics for gene regulatory networks using Master equation 26

- *advance the grid time: $gT = gT + \delta t$*
 - *repeat until $t > gT$*
5. calculate the next reaction μ according to reaction propensities distribution:
 $\sum_{v=1}^{\mu-1} a_v > Ar_2 > \sum_{v=1}^{\mu} a_v$ (the reaction with the largest propensity will occupy the most space in the interval and is therefore most likely to be chosen)
 6. update the state according to the chosen reaction $x = x + st[\mu]$

A numerical characterization of stochastic steady states with DCFTP

We need a numerical procedure for a stochastic system, which would reliably distinguish between the transient behavior and the steady state. For the deterministic systems well-developed software exists (AUTO/MATCONT) and it will be used throughout the thesis for the numerical calculation of steady states and characterization of non-trivial transient behavior. If the stationary distribution of a stochastic model can not be calculated analytically, DCFTP (Dominating Coupling From The Past) (43) algorithm can be applied. This technique is only permitted for networks containing exclusively Hill-, Mono-type and/or first order reactions. The basic gene model used in this thesis falls under this category.

Description of the algorithm the Dominated Coupling From The Past (DCFTP).

Since we want samples from the stationary distribution of a stochastic process, we need to get rid of the transient or the "burn in" period. But how can we decide that the burn in is over? Obviously, a sample from the stationary distribution could be obtained if Gillespie chains would start at the time $t = -\infty$ in all possible initial conditions and run to time 0. Then at the time 0 we could collect a sample and that would be a stationary distribution sample. Numerically, it is not feasible to run something for infinitely long time and starting from infinitely long many initial conditions. So, the main idea of DCFTP is to run a Gillespie chain backwards in time starting from the upper boundary and the lower boundary to include all possible initial conditions between these boundaries and to do it for long enough ensuring that it has reached the steady state at the time 0. To decide that the chain has started long enough in the past we use the idea of coalescence

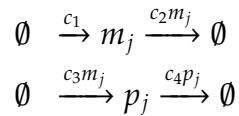
In case the state space is bounded we start two Gillespie chains in upper most state U and the lower most state L and run them coupled from time $-T$ to 0 choosing T so large that both chains reach the same state at time 0 (coalesce) see Fig. 1.3. The state at time 0 will then be an unbiased sample from the stationary distribution if the network fulfills the following properties

- sandwiching
- funnelling
- monotonicity or anti-monotonicity

An additional difficulty is that our state space does not have an upper bound ($n \in [0, \infty]$). But we can constrain the state space by a dominating process, which will always remain in the state space above the actual process and update the chains coupled. For the dominating process the stationary distribution has to be known, and if we start the 3 chains U, X, L in the stationary state of the dominating process, the dominating process will never leave it acting as a dynamic upper boundary and the above described properties will ensure that all chains will end up in the same state given we update them coupled.

As pointed out in Hemberg and Barahona (43) the described model fall under the type of networks, for which DCFTP can be applied, because it contains linear propensity functions for translation and degradation terms and the non-linear transcription propensity is a bounded anti-monotonic function for which the dominating process would be $\Phi_D(x) = \max \Phi(x) = \text{const.}$

The stationary distribution for dominating process is topology independent, since each gene was coupled to other genes through the non-linear monotonic function, which is replaced by a bounding constant in the dominating process:



It means that the stationary probability function will factorize

$$P(m_1, m_2, \dots, m_N, p_1, p_2, \dots, p_N) = \prod_i P_i(m_i, p_i)$$

and by a closer look at the equation we realize that the analytical solution for $P_i(x_i)$

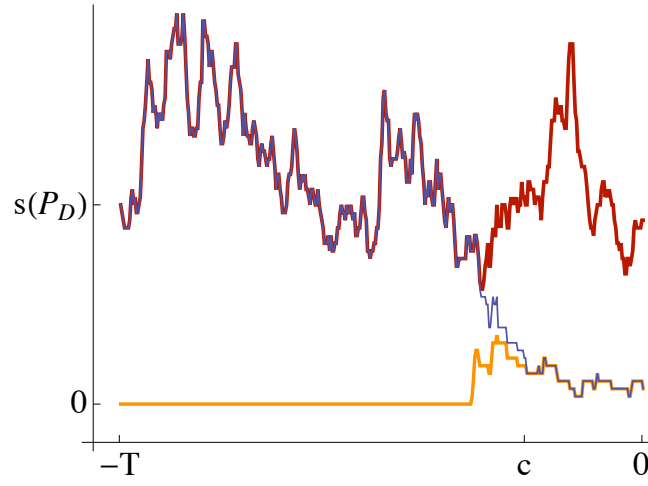


Figure 1.3: Illustration of DCFTP. The stochastically dominating process (red) starts with a sample $s(P_D)$ from its known stationary distribution P_D together with the process of interest (blue) and acts as an upper boundary. The lower boundary process (orange) starts at the ergodic atom 0 (lowest absorbing state). The states of all 3 processes are updated in a "coupled" way, which means whenever the dominating process goes into another state, a time mark m is generated and other chains are updated at these marks if their propensities Φ_L, Φ_X are high enough: $u_m < \Phi_L / \Phi_D$ & $u_m < \Phi_X / \Phi_D$, $u_m \in \text{Uniform}[0, 1]$, where u_m is a uniformly distributed random number drawn for each time mark m . Since L and X are updated according to the same rule, once they meet (**coalesce**) in the state space, they will stay together. This time point c is called coalescence point. If L, X have coalesced in $[-T, 0]$, the unbiased stationary sample can be obtained at time 0.

has already been obtained in section 1.3.1. That means for the *mRNA* steady state $P_i(m_i) = \text{Poisson}(c_1/c_2)$ protein steady state $P'_i(p_i) = \text{Poisson}(c_1c_3/c_2/c_4) \equiv \text{Poisson}(c)$. This distribution is the initial stationary distribution for the dominating process $\tilde{\pi}$ from which the initial state of the system in the DCFTP will be drawn.

The algorithms have been implemented in C by the author and the pseudo code description follows:

- set time to 1: $T \leftarrow 1$
- draw the initial state for all 3 chains from the known stationary distribution of the dominating process: $\tilde{D}_0^0 \sim \tilde{\pi}$, draw the initial random numbers from the uniform distribution: $\tilde{M}_0^0 \sim U(0, 1)$
- run the algorithm for the linearized network: $\tilde{D}_T^0 \leftarrow \text{Extend}(\tilde{D}_0^0, T)$
- $\tilde{M}_T^0 \leftarrow \text{GenerateMarks}(\tilde{M}_0^0, T)$
- loop
 - $(D_0^{-T}, M_0^{-T}) \leftarrow \text{Reverse}(\tilde{D}_T^0, \tilde{M}_T^0)$
 - $(L_0^{-T}, U_0^{-T}) \leftarrow \text{Evolve}(D_T^0, M_T^0)$
 - If $U_0 = L_0$ then
 - return L_0
 - end if
 - $T \leftarrow 2T$
 - $\tilde{D}_T^0 \leftarrow \text{Extend}(\tilde{D}_{T/2}^0, T/2)$
 - $\tilde{M}_T^0 \leftarrow \text{GenerateMarks}(\tilde{M}_{T/2}^0, T/2)$
- end loop

The stationary state computational tool became useful showing the qualitative differences between the deterministic and the stochastic stationary state as discussed in the following section 1.3.4.

1.3.4 Noise qualitatively changes gene network dynamics

Taking low copy number noise into account in gene expression does not only slightly randomize the deterministic dynamics, in a sense that the qualitative

1.3 Stochastic dynamics for gene regulatory networks using Master equation

behavior is still well approximated with the deterministic equation. With one of the simplest genetic network where two genes are repressing each other - a genetic switch - it can be shown that the qualitative dynamics in the deterministic regime are different from the equivalent dynamics in the stochastic setting.

To show that we compare the stationary state of a genetic switch in the deterministic and stochastic setting. In the deterministic version the stationary state of the switch is given by two fixed points, one where gene 1 is expressed and gene 2 is repressed and the other where gene 1 is repressed and gene 2 is expressed (see Fig. 1.4). Depending on the initial conditions deterministic system will reach one of the fixed points and remain there. In the equivalent stochastic version the stationary state is given by a stationary probability distribution. The stationary probability distribution is bimodal with two hubs around the deterministic fixed points. The main difference between the deterministic and stochastic model is that depending on the noise level there is a non-vanishing probability that the genetic switch can spontaneously reverse the state from gene 1 expressed and gene 2 repressed to gene 2 expressed and gene 1 repressed. This quality is excluded in the deterministic model and is only present in the stochastic model.

1.3.5 From stochastic Master equations to ordinary differential equations: bridging the gap with Ω -expansion

Even in a very simple genetic network there is a qualitative difference between the stochastic and the deterministic setting. Intuitively the stochastic description must go over into the deterministic one if the number of molecules involved becomes high as already intuitively illustrated with the classical gas molecules example in section 1.3. In the following a mathematically rigorous expansion will be introduced connecting the deterministic and the stochastic description and quantitatively estimating the errors for neglected terms.

Intuitively, it is clear that "the volume or the size of the system", which will be denoted with Ω , influences how well the deterministic dynamics approximate the single molecule reactions. And indeed, a classical Taylor expansion in the inverse of Ω can be used for the systematic transition from the stochastic description with the Master equation to the deterministic ODEs (20).

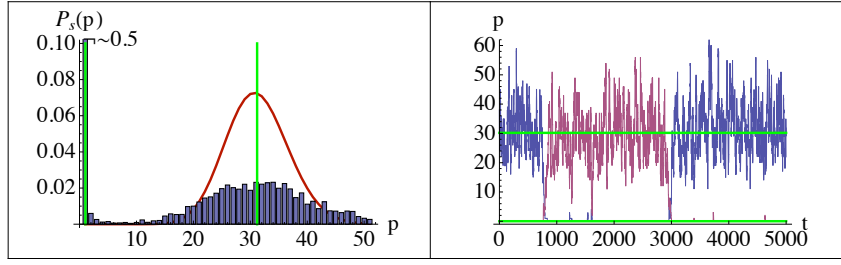


Figure 1.4: **Qualitative difference in the stochastic compared to the deterministic version of a genetic switch.** A numerical estimate of the stationary distribution for a genetic switch (blue bars) have been obtained from my implementation of the algorithm Dominated Coupling From The Past introduced in Hemberg and Barahona (43) and revised in the section 1.3.3. As expected we obtain a bimodal distribution $P_s(p)$ (blue bars) centered around deterministic fixed points marked by light green bars (the dominating process distribution is shown for the reference in magenta). The qualitative difference between the stochastic and the deterministic dynamics is that in the deterministic setting reaching the steady state is a final decision, the system dynamics will remain there for all times. In contrast for the stochastic setting there is a non-vanishing probability that switch spontaneously reverse. We show on the right a time trace of protein concentrations (p_1 in blue and p_2 in magenta) of the forward Gillespie simulation (see (21) and the section 1.3.3 for the algorithmic details) containing two such reversal events. The probability of the reversal can be estimated from the stationary bimodal distribution as the area between the hubs.

Starting from the general form of the Master equation

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{\mathbf{n}'} W_{\mathbf{n}\mathbf{n}'} P(\mathbf{n}', t) - W_{\mathbf{n}'\mathbf{n}} P(\mathbf{n}, t)$$

we rewrite the multi-dimensional step operators E_i, E_i^{-1} , the probability distribution $P(\mathbf{n}, t)$, the transitions $W_{\mathbf{n}\mathbf{n}'}$, and the state vectors $\mathbf{n} = (n_1, \dots, n_s)^t$ in terms of the system size Ω , a new variable ξ , and the continuous function $\phi(\mathbf{t})$ which is undetermined at the moment:

$$\begin{aligned} n_i &\rightarrow \Omega \phi_i + \Omega^{1/2} \xi_i \\ P(\mathbf{n}; t) &\rightarrow \Pi(\xi; t) \\ W_{\mathbf{n}\mathbf{n}'} &\rightarrow f(\Omega) \left[\Phi_0\left(\frac{n'}{\Omega}; n - n' \equiv r\right) + \Omega^{-1} \Phi_1\left(\frac{n'}{\Omega}; n - n' \equiv r\right) + \dots \right] \\ E_i &\rightarrow 1 + \Omega^{-1/2} \partial_{\xi_i} + 1/2 \Omega^{-1} \partial_{\xi_i}^2 + \dots \\ E_i^{-1} &\rightarrow 1 - \Omega^{-1/2} \partial_{\xi_i} + 1/2 \Omega^{-1} \partial_{\xi_i}^2 + \dots \end{aligned}$$

1.3 Stochastic dynamics for gene regulatory networks using Master equation 13

For the linear noise approximation and the derivation of the macroscopic equations we substitute the variables from the above into the Master equation and use the Taylor expansion

$$\begin{aligned}
\partial_t \Pi - \Omega^{1/2} d_t \phi \partial_\xi \Pi &= \\
&- \Omega^{-1/2} f(\Omega) \partial_\xi \int r \Phi_0(\phi(t) + \Omega^{-1/2} \xi; r) d\mathbf{r} \cdot \Pi(\xi, t) \\
&+ 1/2 \Omega^{-1} f(\Omega) \partial_{\xi^2}^2 \int r^2 \Phi_0(\phi(t) + \Omega^{-1/2} \xi; r) d\mathbf{r} \cdot \Pi(\xi, t) \\
&- 1/3 \Omega^{-3/2} f(\Omega) \partial_{\xi^3}^3 \int r^3 \Phi_0(\phi(t) + \Omega^{-1/2} \xi; r) d\mathbf{r} \cdot \Pi(\xi, t) \\
&- 1/3 \Omega^{-3/2} f(\Omega) \partial_\xi \int r \Phi_1(\phi(t) + \Omega^{-1/2} \xi; r) d\mathbf{r} \cdot \Pi(\xi, t)
\end{aligned}$$

where we denote the expansion of the step operators as defined in the equation above 1.4. To simplify the notation we define the rescaled moments

$$\alpha_{\nu, \lambda}(x) = \int r^\nu \Phi_\lambda(x, r) d\mathbf{r}$$

and also rescale time $\Omega^{-1} f(\Omega) t = \tau$ according to the size of the system. Depending on the level of description the inherent time scales of the process will be different. In this expansion we make them dependent on the system size Ω .

Now with the usual argument that each Ω power must vanish independently, for the $\Omega^{1/2}$ we obtain the macroscopic deterministic equations:

$$d_\tau \phi(\tau) = \alpha_{1,0}(\phi(\tau))$$

1.3.6 SDEs and Master equations

If collecting and summarizing the zeroth Ω power from the above expansion we will obtain equations for the second moments, the variances and covariances.

The according approximate PDE is called the linear Fokker-Planck equation whose coefficients depend on time through ϕ :

$$\partial_\tau \Pi(\xi, \tau) = -\alpha'_{1,0}(\phi) \partial_\xi \xi \Pi + \frac{1}{2} \alpha_{2,0}(\phi) \partial_{\xi^2}^2 \Pi \quad (1.4)$$

This equation is also often referred to as “linear noise approximation” because the solution is determined by a Gaussian (20), which is fully characterized with the first and the second moments, therefore it is equivalent to the SDE (45):

$$x(t) = x_0 + \int_0^t b(x_s) ds + \int_0^t \sigma(x_s) \cdot d\omega_s \quad (1.5)$$

with $d\omega_s$ Gaussian white noise, $b = \alpha'_{1,0}$ the drift term, and $\sigma = \alpha_{2,0}$ the diffusion term.

The Ω expansion therefore not only naturally connects deterministic and stochastic regimes, but through the intermediates (in form of the moments) offers the possibility to systematically control stochastic nature in a problem. We will use the formalism to assess the influence of stochastic fluctuations onto the oscillations in the thesis.

1.4 Pioneering *in vivo* synthetic circuits

Classical engineering has inspired the creation of key synthetic elements such as switches and oscillators. The progress and remaining difficulties are discussed in the following section for two special cases: 2 genes forming a switch implemented by Gardner et al. (1) and 3 genes forming an oscillator implemented by Elowitz and Leibler (2).

In the early 2000’s, synthetic circuits for bacterial cells were implemented on plasmids using cloning techniques as for instance described in Ausubel (46). The plasmids enter the cell by transformation, that means naked synthesized DNA enters the cell after a chemo-physical treatment. The treatment is to chill the *E. coli* cells in the presence of divalent cations such as Ca^{2+} (in CaCl_2), which makes the cell walls permeable to plasmid DNA. The cells are incubated on ice together with the synthetic plasmid DNA and then briefly heat shocked (eg 42 C for 30-120 seconds), which causes the DNA take in. This method works very well for circular plasmid DNAs. An excellent preparation of competent cells will give 10⁸ colonies per microgram of plasmid, whereas a rather poor preparation will be about 10⁴/μg or less. Once the plasmid is inside the cell, the available transcription machinery of bacteria ensures the expression of the genes encoded on the plasmid. The engineered plasmids contain in addition to the designed

network also resistance genes against antibiotics. If the cells then are grown in an antibiotic enriched medium, only those with the desired plasmids will survive.

The pioneering experimental implementation of this kind of network was performed by (2) creating a genetic oscillator in a network of three mutually repressing genes and by (1), who created a genetic toggle switch. Both experimentalists used *E. coli* as the model organism and ampicillin as antibiotic control, which was included in the plasmid. Both networks are based on the idea that each gene has a promoter region and a product of one gene binds to the promoter site of the other gene inhibiting the other gene's transcription. The plasmid design and abstract architecture are shown in fig. 1.5.

Both designs are based on classical engineering ideas of an oscillator and a switch. One of the differences between the biological implementations and the electrical or mechanical circuits are the uncertainties in the parameters. Their estimations are given in orders of magnitudes, and they are by far more difficult to control in biological systems compared to the classical engineering. Therefore the biological designs have to be robust to changes in parameters.

We give literature-based parameter ranges below and take these estimates as a reference point for later simulations of gene networks:

- c_4 protein degradation rate example studies $\approx 0.06 \text{ min}^{-1}$ (which is of the order of magnitude of protein half life of 40-60 *min*) (from Elowitz and Leibler (2), Mizusawa and Gottesman (35), Rao and Arkin (36))
- c_2 *mRNA* degradation rate typical rates 2-4 *min* in *E. coli*, the actual range is quite large 30 s to 50 min (from Belasco and Brawerman (34))
- n Hill coefficient and cooperativity 2-5 (from Müller et al. (47))
- $K_m \approx 1$ (from Müller et al. (47))
- $c_1 \approx 1-10 \text{ mRNAs per min}$ is calibrated to achieve the maximal concentration around decades
- c_3 the translation efficiency is taken to be low, around one protein per transcript. The value might be increased to 20 proteins per transcript (2)

Parameter uncertainty is not the only difference between the classical and biological engineering. Synthetic gene circuits operate in the biological environment

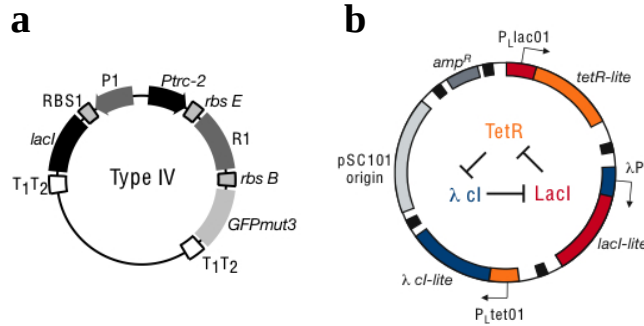


Figure 1.5: **(a) Plasmid design of the genetic toggle used by Gardner et al. (1).** The promoter binding sites ($P1$, $Ptrc - 2$) are marked by solid lines with arrowheads, where $P1$ controls the expression of $lacI$ gene and $Ptrc - 2$ of $R1$. The toggle behavior arises as the protein produced from $R1$ gene binds to $P1$ inhibiting the transcription of $lacI$ gene. At the same time $lacI$ protein inhibits the transcription of $R1$, binding to $Ptrc - 2$. The ribosomes binding sites ($RBS1$, $rbsE$, and $rbsB$) and the transcription terminators (T_1T_2) are marked with outlined boxes. Genetically fused with $R1$ is $GFPmut3$, a reporter gene producing green fluorescent protein, which concentration can be detected by a laser. **(b) Plasmid for a genetic oscillator designed by Elowitz and Leibler (2).** The activators (P_{Llac01} , λP_R , and P_{Ltet01}) induce the transcription of the genes following them on the ring. The products of $tetR - Lite$, $lacI - lite$, and $\lambda cl - lite$ suppress the transcription of P_{Ltet01} , P_{Llac01} , and λP_R respectively as indicated by matching colors.

that can not be considered static. Quoting again the designs discussed above, the switch and the oscillator are slow time scale dynamics compared to *E. coli* cell cycle and the main dynamics features such as switch reversal or a full oscillation embrace several cell generations occurring on the order of hours, while the cell division occurs on the order of minutes. So the dynamical features must be continued in the next cell generation, which means on top of the uncertainties also population evolution of the cells must be considered.

The aspect, that cell population is the operating environment, where the synthetic constructs are operating, raises a question of synchronization of circuit dynamics on the cell population level. In particular, oscillation phases may change both due to cell division and also due to stochasticity of the gene expression. Recent developments from the Hasty lab attempt to address that problem by creating more synchronized oscillations (3) or patterns on the cell population level (48). The genetic circuits inside of each cell are not autonomous oscillators, but can produce population pattern in form of propagating waves. The study illustrates the important aspect of population dynamics in the synthetic circuit creation.

What we are interested in are constructs in classical engineering, which have the potential to overcome the uncertainties and stochasticity and are controllable on the cell population level. There are mathematical analysis tools such as bifurcation or symmetry arguments helping identify topology-based designs which can function in challenging biological conditions.

1.5 Transferring concepts between the disciplines

The design of synthetic circuits is usually based on a simplified parametric mathematical model, which describes the behavior of a genetic network. In general, qualitative behavior of a dynamic system or a genetic network model can depend on parameter values (49), as for example an increase of a parameter can qualitatively change the system behavior from steady state convergence to oscillation (49). Before the physical implementation of a synthetic circuit, bifurcation analysis must be used to investigate, which qualitative dynamics can be exhibited by the proposed network topology depending on the parameter sets. The parameters can then be adjusted for the desired dynamics via synthetic biology engineering tools, as for example the degradation rates of the proteins can be changed with appropriate degradation tags (2), or the number of plasmids, i.e. the number of DNA copies in the cells can be controlled via replication control sequence (26).

In the following, we review shortly the main nonlinear dynamics tools from classical engineering, which have also preceded the implementation of the oscillator (2) and the switch (1). Interestingly, similar elements have been observed previously in other physical (50) and electronic (51) systems and bifurcation analysis along with symmetry arguments was used to describe the qualitative dynamics and their changes depending on the parameter sets. Therefore, the nonlinear analysis tools and symmetry arguments are not only helpful for description of qualitative dynamics depending on the parameter sets, but can also help finding other cross-disciplinary or universal concepts, which can be implemented in both physical and biological systems.

1.5.1 A condition for oscillation

On the abstract level the implemented genetic oscillator (2) is a ring of three identical cells. The ring topology is a classical system, which has been studied for other physical systems such as for instance the quantum harmonic oscillators (50) or even implemented as an sensor for the magnetic cores (51). The remarkable common characteristic of all these systems is the ring topology and the fact that for a particular set of parameters oscillatory modes appear and their number is directly connected to the number of ring elements N :

$$\begin{aligned}\#_{modes} &= (N - 1)/2 && \text{odd} - \text{rings} \\ \#_{modes} &= (N/2 - 1) && \text{even} - \text{rings}\end{aligned}\tag{1.6}$$

The connection between these systems can be made on the abstract level using the classical nonlinear bifurcation analysis and symmetry arguments (see for instance (49) or (52)). In general bifurcation analysis predicts the qualitative change of the system behavior as one or several parameter change. A parameter value at which such qualitative change occurs, is called bifurcation. In biochemical networks a frequently found bifurcation, where the behavior changes from approaching a constant value for time $\rightarrow \infty$ to stable oscillation is a Hopf bifurcation formally defined in the box below:

BOX 1

Definition of Hopf Bifurcation (Hopf 1942) Suppose that the parameter dependent differential system

$$\dot{x} = f(x, \mu)$$

$f : R^{m+1} \rightarrow R^m$, $f \in C^\infty$ has an equilibrium (x_0, μ_0) (i.e. $f(x_0, \mu_0) = 0$) at which the following properties are satisfied:

- **(H1)** The Jacobian $D_x f(x_0, \mu_0)$ has a simple pair of pure imaginary eigenvalues $\pm i\omega$ and there are no other eigenvalues with zero real part. Then (H1) implies that there is a smooth curve of equilibria $(x(\mu), \mu)$ with $x(\mu_0) = x_0$. There is pair of eigenvalues $\lambda(\mu), \bar{\lambda}(\mu)$ of the Jacobian $D_x f(x(\mu), \mu)$ along this curve with $\lambda(\mu_0) = i\omega, \bar{\lambda}(\mu_0) = -i\omega$ at $\mu = \mu_0$ which varies smoothly with the parameter. If, moreover,
- **(H2)** $\frac{d}{d\mu}(Re\lambda(\mu))|_{\mu=\mu_0} \neq 0$, then there exists a branch of periodic solutions bifurcating at (x_0, μ_0) .

In fact it is this type of bifurcation, which gave rise to the oscillating behavior in the ring of quantum harmonic oscillators (50) and other ring elements (53). We will assess later in the thesis in detail how Hopf bifurcations also give rise to oscillatory behavior in rings of repressed genes.

1.5.2 A condition for switching behavior

The dynamical feature, on which the genetic switch as realized by Gardner et al. (1) is based, is the Pitchfork bifurcation as formally defined below:

BOX 2

Definition of Pitchfork Bifurcation Suppose that the parameter dependent differential system is given by

$$\dot{x} = f(x, \mu)$$

$f : R^{m+1} \rightarrow R^m$, $f \in C^\infty$ where the following properties are satisfied:

- f is an odd function: $-f(x, \mu) = f(-x, \mu)$
- the first and second derivatives of f vanish, but not the third one: $\frac{\partial f}{\partial x}(0, \mu_0) = 0$, $\frac{\partial^2 f}{\partial x^2}(0, \mu_0) = 0$, and $\frac{\partial^3 f}{\partial x^3}(0, \mu_0) \neq 0$.
- $\frac{\partial f}{\partial \mu} = 0$, $\frac{\partial^2 f}{\partial x \partial \mu}(0, \mu_0) \neq 0$

then there is a Pitchfork bifurcation at $(0, \mu_0)$. If the third derivative is positive the Pitchfork bifurcation is sub-critical, if the third derivative is negative, it is super-critical.

The qualitative behavior of the system before the bifurcation is that it converges to a single fixed point, in contrast after the bifurcation the system can converge to one of the two stable fixed points or one unstable fixed point depending on the initial conditions (IC). The pitchfork bifurcation induces a symmetry breaking in the system, which then leads to the switching behavior.

1.6 Application of control theory in biology

Even though natural biological systems such as gene regulatory networks might appear to be very different from classical engineering examples as for instance flight control or nuclear reactor shutdown, several similarities on the system organization level can be revealed (54). In both systems robust functionality under changing and difficult to predict environments is achieved via hierarchical modular organization structure and feedback loops, between the hierarchical layers (see Fig. 1.6).

Insights and concepts from the control theory helped to explain naturally

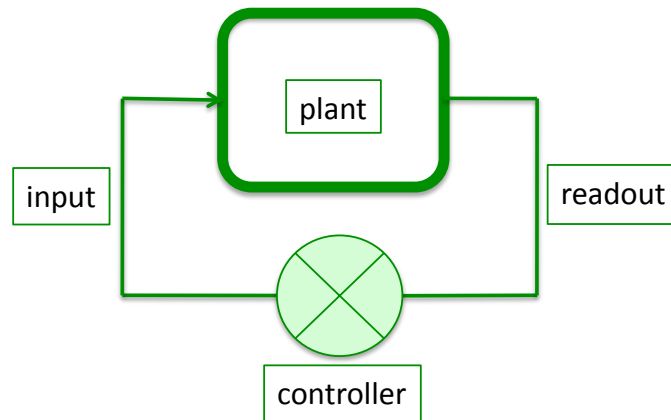


Figure 1.6: **Classical idea of feedback control.** The system, which we are trying to control is usually referred to as “plant”. The current state of the plant is estimated and this estimation is called readout. Based on the readout, the controller will calculate the appropriate control action and an input signal will be feed into the system.

occurring biological mechanisms such as for example the perfect adaptation, which we review in the following.

1.6.1 Use of feedback control concepts to explain perfect adaptation in natural biological systems

Control theoretical concept of integral feedback loop regulation has been proved useful to understand the ability of perfect adaptation in several natural biological systems. By perfect adaptation in this context we mean the ability of the system to assume the original state in the presence of an external signal. With this mechanism the steady-state output becomes independent from the steady-state input (55–58).

One of the classical examples is the integrative feedback control for maintaining the calcium concentration independent of the constant calcium demand - calcium homeostasis (55). A remarkable adaptation occurs in diary cows during the parturition (calving), as the calcium clearance adds up to 4-5 fold of the normal, due to both the static demand caused by the daily milking and step-like disturbance caused by calf feeding. The constant calcium concentrations are maintained in the animals via hormonal integrative feedback loops, which regulate the calcium transport from bones, kidney, and intestine to the milk plasma. Understanding of this mechanism and its failure in some animals is particularly

important for the agriculture.

Another classical and well-studied example is the perfect adaptation via integral feedback control during chemotaxis (57, 58). Chemotaxis is the process by which bacteria moves towards or away from chemicals, such as for instance food or poison respectively. During the directed movement, the bacteria exercises a random walk consisting of an alternation between the running and tumbling. In this process the bacteria senses the concentration gradient of the according chemical and decides upon the moving direction and tumbling frequency. Perfect adaptation returning to the prestimulus state is necessary in this system in order to ensure the directed movement responding to the concentration gradients. Moreover, based on the chemotaxis example, it can be argued that robust perfect adaptation requires some sort of an integral feedback loop (57). This conclusion has stimulated research into the identification of biochemical parts required for the implementation of this integral feedback mechanisms.

In this sense control theoretical concepts can also be used for biological model discrimination, whenever perfect adaptation is experimentally observed. A recent survey by Muzzey et al. (56), which concentrated on the response to step inputs of osmolyte in budding yeast, can be quoted as an example. Here, a perfect adaptation is observed on the single cell level in nuclear enrichment of Hog1, a gene involved into the MAPK signalling cascade, which is activated in response to the external osmotic disturbance. The precise model of the process is not known, but with the experimental proof of the perfect adaptation, all mathematical model, which do not contain integral feedback loops, could be discarded and a mechanism as the protein-protein interaction could be pinpointed.

Several other examples are given in El-Samad and Khammash (58), where robust functionality of genetic network topologies could be explained with control theoretical concepts. Understanding the natural systems in these terms also opens a perspective for the forward design of the synthetic genetic networks. Natural biochemical methods as well as the robust organization structure can be used in the design of new biological parts.

1.6.2 Control theoretical concepts for synthetic biology engineering

From a perspective of a control engineer, the design of a synthetic circuit should occur on two levels: the first is the design and characterization of biochemical modules and the second is the interconnection of the modules. The biochemical parts are genes, their promoters, proteins etc and there is a database for these synthetic biological parts (22). Ideally the parts should be well characterized: a promoter for instance should be characterized in terms of its activators and the strength in different growth conditions, a genetic sequence in terms of the transcription time, a protein or *mRNA* product the degradation times should be known, etc. In the summary, synthetic biology engineering requires in the first place robust, well characterized modules, which can be connected together to larger synthetic networks in order to perform sophisticated functions.

An engineer on the second - the part interconnection level - should ideally not care about the details of the biochemical part realization. The interconnection level design should connect the parts for the robust performance and effective adaption to external signals. An example of such an effective interconnection design is a feedback loop as discussed in the previous section. We will discuss in more detail in thesis chapter 5, how the control theoretical concept of a feedback loop can be adapted for a control of synthetic elements.

The main difficulty for an efficient design in synthetic biology is that parts are still lacking an appropriate description in order to be considered as independent modules with well-defined interfaces. The difficulties are the fluctuating environment leading to the change of system constants, the stochasticity, and cross-reactions between the modules. Furthermore, unlike the mechanical systems synthetic biology modules are expected to function within a cellular environment, which means they are not separated or isolated like mechanical parts. This leads to the cross-talks between the elements and the extent of the cross-talk sometimes can not be estimated prior the implementation.

For the implementation of a feedback control in thesis chapter 5, we take a model-free approach based on the measured time traces avoiding the difficulties such as the cross-talk and the parameter uncertainty. Our approach exemplifies how a classical control theoretical concept can be adapted for the use in synthetic

biology.

* * *

There is a genetic network model "The generalized repressilator", which has the reviewed synthetic oscillator (2) and the switch (1) as special cases. This model possesses ring symmetry and shows both Hopf and pitchfork bifurcations. We will analyze this model in the following using nonlinear dynamics tools in deterministic and stochastic settings. Finally, we will exemplify the adaptation of a classical feedback loop strategy on the generalized repressilator.

In the following chapter we show the presence of Hopf and pitchfork bifurcations in this model via numerical continuation and propose a design of a controllable genetic oscillator using classical engineering concept of readout-based feedback control.

Chapter 2

Genetic oscillator design around unstable periodic orbits

The following chapter is based on the publication

Switchable genetic oscillator operating in quasi-stable mode.

Natalja Strelkova and Mauricio Barahona

Journal of Royal Society Interface

July 6, 2010 7:1071-1082

2.1 Introduction

The central role played by oscillations in cellular function has made oscillatory circuits a primary target for the analysis and design of synthetic networks (59). A particular area of interest is the elucidation of strategies leading to robust timing and sequential activation in the cell. For instance, key stages in developmental biology and in cell differentiation may be controlled by so-called master regulators—a small set of transcription factors sequentially activating and driving several other genes with accurate timing (60–62). In addition, studies of both natural (63, 64) and engineered circuits (65) indicate that the correct timing and order of gene activation is a key characteristic of balanced, optimal cell function, as it reduces the metabolic burdening that ensues from the continuous presence of heterologous substances (66).

In this chapter, we consider the dynamics and control of noisy genetic oscillatory circuits in quasi-stable mode operation. We exemplify our results with a simple and widely studied synthetic network: the N -gene ring repressilator (Fig. 2.1a). Some natural networks of master regulators (60) contain such ring structures as subnetworks, making the exploration of their dynamic behavior relevant for both naturally occurring and synthetic systems. The underlying idea is well-known: when observing the dynamics of systems operating in highly variable environments, such as the cell, it might not be enough to characterize only the long-term attractors of the system since unstable solutions can play a significant role. For instance, quasi-stable transients might be so long-lived as to be the most significant feature of the observed noisy dynamics (67). Moreover, the presence of noise in nonlinear systems may induce non-stationary dynamics in systems with only fixed point attractors in the deterministic setting (68) or, conversely, noise may act as a stabilizer of unstable deterministic states (69).

In the generalized repressilator, results due to Smith (47, 70) show that rings with an even number of genes (e.g., the toggle switch (1) with $N = 2$) exhibit multistability and hence behave like switches in the stochastic regime. On the other hand, rings with an odd number of genes (like the standard repressilator (2) with $N = 3$) have a globally attracting limit cycle and are therefore oscillators both in the stochastic and deterministic regimes. However, we show here that generalized repressilators possess an intricate structure of unstable periodic orbits that play an important part in their observable noisy and transient dynamics. In particular, even rings have a quasi-stable limit cycle which, although unstable in terms of linear Floquet stability analysis, has only one unstable direction with a very slow escape rate. This means that trajectories are attracted to the limit cycle from all directions but one, hence leading to long-lived, inducible periodic transients in the deterministic setting and to sustained oscillations in the stochastic system. These effects become more pronounced as the number of genes grows. Therefore the finite-time, observable noisy dynamics of an even repressilator ring is not necessarily static (switch-like) but rather exhibits oscillatory characteristics.

In addition to their effect on the observable dynamics, quasi-stable oscillatory modes can be used as operating points to control the system around them. The advantage of such a scheme is that the oscillations can be switched on and off, unlike the limit cycle attracting behavior of odd repressilators. Operation

around unstable modes, usually illustrated with the example of the inverted pendulum (71), is a classic scenario in control theory for enhanced controllability and speed of response. It has a long and successful history of applications in fluid flow control (72) and in the steering of jet aircraft (73). Here, we illustrate the application of this concept to gene networks with a simple protocol of controlled interference based on an optical mechanism for readout and induction of gene expression. The current concept is based on an alternative mechanism to the chemical intervention proposed in (74) to produce switchable oscillations. Our simulations show that even repressilator rings in quasi-stable operation can behave as a robust and on-demand switchable oscillator in which genes become upregulated periodically in an ordered sequence according to a travelling-wave solution. This switchability, which is robust at both high and low copy numbers, could be used for synthetic biological applications such as accurately timed interference with naturally occurring networks.

2.2 The generalized repressilator model

2.2.1 Ordinary differential equations

The generalized repressilator consists of a ring of N genes in which transcription of each gene is repressed by the product of the preceding gene (Fig. 2.1a). A deterministic model of this circuit is given by the following set of ordinary differential equations:

$$\begin{aligned}\dot{m}_j &= \frac{c_1}{1 + p_{j-1}^2} - c_2 m_j \\ \dot{p}_j &= c_3 m_j - c_4 p_j,\end{aligned}\tag{2.1}$$

where p_j and m_j describe protein and mRNA concentrations for each gene, respectively (2). Here, $j = 1, \dots, N$ with the periodic boundary condition $p_0 = p_N$, and c_1 (c_3) is the creation rate and c_2 (c_4) is the degradation rate for the mRNAs (proteins). The production of mRNA is modeled as a source term that depends nonlinearly on the concentration of the inhibitor protein. Proteins are assumed to be produced at a rate linearly dependent on the amount of the corresponding mRNA. The degradation of mRNA and proteins is assumed to be linearly propor-

tional to their current amount. For the full derivation of these equations from the mass-action kinetics see Sec. 1.2.1 and for the outline of the underlying biophysical process see Sec. 1.1. For notational convenience we introduce a constant

$$c = \frac{c_1 c_3}{c_2 c_4}. \quad (2.2)$$

The toggle switch (1) and the repressilator oscillator (2), which have both been implemented in *E. coli*, are special cases with $N = 2$ and $N = 3$, respectively. Based on analytical results on monotone systems (i.e., systems in which partial derivatives do not change sign) due to Smith (47, 70), the stability analysis of this circuit reveals a fundamental difference between rings with odd and even numbers of genes. The main results are summarized in Fig. 2.1b–c.

2.2.2 The stationary state

The stability analysis characterizes the long-term dynamic behaviors of the deterministic system. An example of such behavior is given by the fixed points of the system, i.e., the states in which the dynamics is stationary. The variation of a parameter can produce a change in the stability or the existence of fixed points or other attractors. This is called a bifurcation and it leads to qualitative changes in the long-term behavior of the system. One can find the parameter values at which bifurcations are produced by performing a bifurcation analysis, which can be carried out analytically or numerically with the aid of continuation software packages such as AUTO (75), which can also track the stability of periodic solutions.

In the generalized repressilator model two types of bifurcations occur a **pitchfork bifurcation** (see section 1.5.2 and (49)) emerges in even numbered rings and **Hopf bifurcation** (see section 1.5.1 and (49)) leading to a stable limit cycle occurs in odd numbered rings (Fig. 2.1).

In the case of even rings, a pitchfork bifurcation takes place at $c = 2$ for all N (Fig. 2.1c). The two additional fixed points arising at that value of the parameter and we will derive analytical expressions for the points as well as their stability in section 3.2.1. The emerged fixed points of the system (3.1) correspond to two distinct dimerized states: one in which genes with odd indices are upregulated (p_u) while genes with even indices are downregulated (p_d); and another symmetric

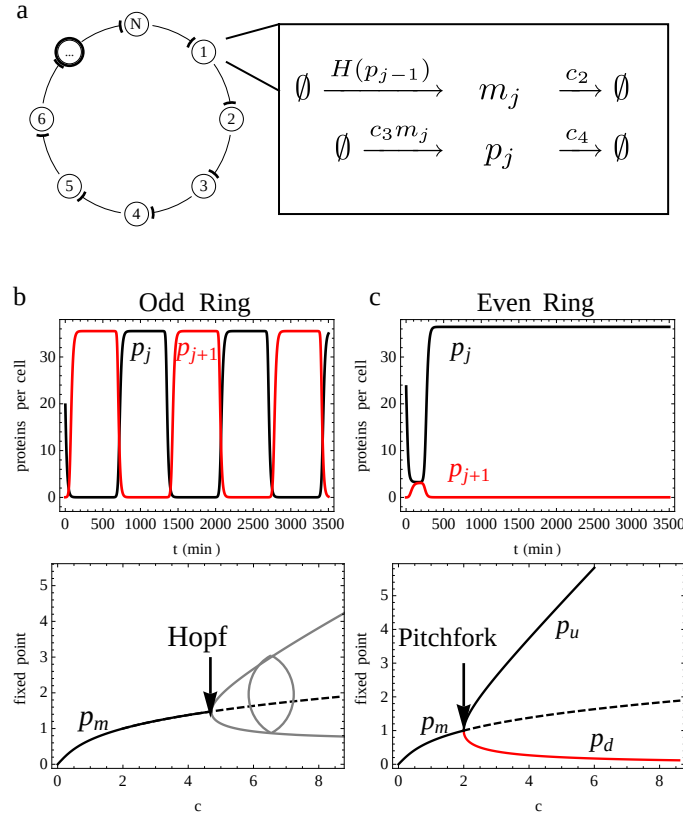


Figure 2.1: Attractors of the generalized repressilator model (a) Topology of the generalized repressilator: N genes in a cycle where each gene is repressed by the protein product of the preceding gene. Also shown is the reaction scheme underlying the dynamical system (3.1) with production and degradation terms for the mRNA (m_j) and protein (p_j) of each gene. The repression of the production of mRNA is modelled by a Hill-type term $H(p_{j-1})$. (b) Typical time traces of the long-term deterministic dynamics of an odd ring and an even ring above the bifurcation point $c = 2$: odd rings converge to a globally attracting periodic solution while even rings converge to fixed points. The time traces shown correspond to $N = 23$ and $N = 22$. (c) Stability of the fixed points of the system as a function of the bifurcation parameter c . Even rings undergo a pitchfork bifurcation at $c = 2$, leading to the emergence of two stable fixed points. Odd rings undergo a Hopf bifurcation leading to the emergence of a limit cycle. The critical parameter for the Hopf bifurcation depends on N but tends to $c = 2$ as N grows (see result figure 2.4.1).

state where the genes with odd and even indices exchange their patterns of regulation. These solutions are equivalent to tiling the ring with $N/2$ repeated copies of the up-down solution of the two-gene ring. Their structure is similar to that of other dimerized degenerate solutions in classic models of conjugated polymers and spin chains (76). Therefore, after the bifurcation, the system is bistable, i.e., it behaves like a switch.

In the case of odd rings (Fig. 2.1*b*), the symmetric solution p_m becomes unstable following in this case a Hopf bifurcation: no additional fixed points appear but rather the bifurcation signals the emergence of a periodic solution. We derive analytical conditions for the emergence of this Hopf Bifurcation in section 3.2.1. Smith (70) proved that in monotone systems such as the repressilator, the periodic solution that emerges is a globally attracting stable limit cycle. Therefore odd rings behave as stable oscillators following the Hopf bifurcation.

Floquet theory and unstable periodic orbits

The stability analysis summarized above does not provide information about unstable periodic solutions. Although, in principle, unstable periodic orbits are not relevant for the long-term deterministic dynamics, they can be key to the observed dynamics, especially if the orbits involve slow time scales. Such long-lived oscillations can appear as transients in deterministic simulations and are likely to be observed in the corresponding stochastic simulations. In fact, it was in numerical simulations that we first noticed the relevance of these modes in even repressilator rings.

Floquet theory can be used to find periodic solutions and quantify their linear stability in terms of their Poincaré map, i.e., the crossings of the orbit with a (hyper)plane in phase space (see Fig. 2.5). Under this analysis, a periodic solution (a closed orbit indicated in red in Fig. 2.5) becomes a fixed point of the Poincaré map and its stability is reformulated as the linear stability of this fixed point. The eigenvalues of the Poincaré map linearized around the fixed point constitute the Floquet multipliers. They indicate how an infinitesimal perturbation around the orbit decays or grows (exponentially). The periodic solution is linearly stable if all the Floquet multipliers have magnitudes smaller than unity (There is always a trivial multiplier which equals to 1 and corresponds to the perturbation along

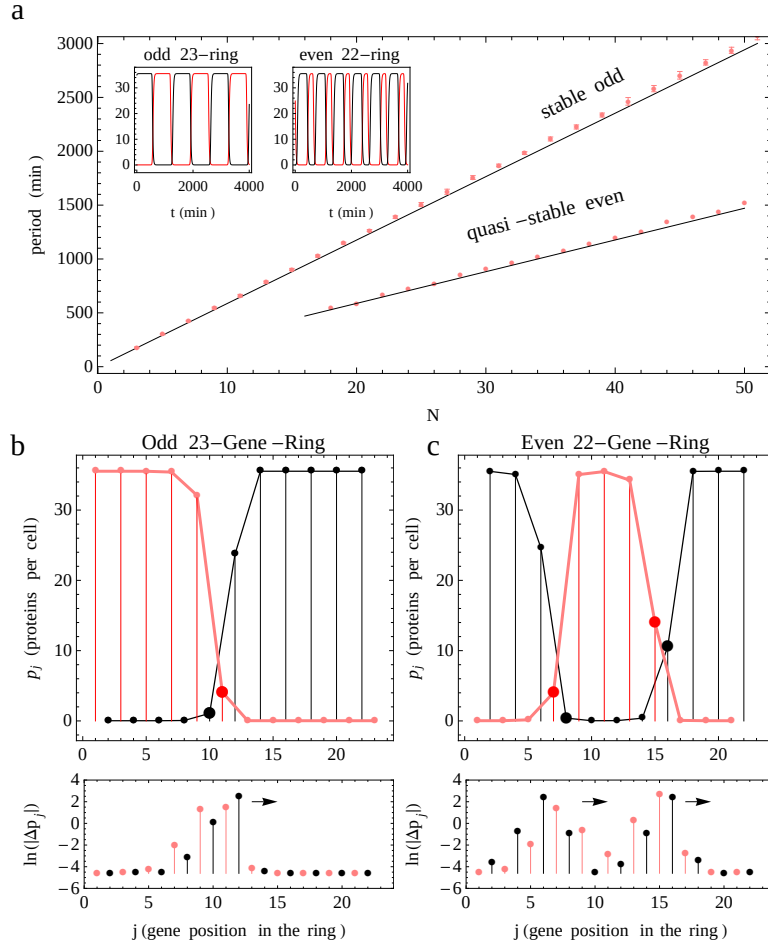


Figure 2.2: Periodic solutions and travelling waves in the generalized repressilator model (a) The period of the limit cycle of the deterministic model of odd rings (solid line) increases linearly with the length of the ring. Simulations of the stochastic version of this system using the Gillespie algorithm (see section 1.3.3) show that the period (shown as circles) follows the same trend, although they are slightly larger. The period of the quasi-stable solutions found in even rings (deterministic and stochastic) increases linearly with the number of genes but with a slope that is half that of the odd rings. The inset shows representative time traces of the periodic solutions in odd rings (stable) and in even rings (quasi-stable). (b) Time snapshot of the spatial distribution of the concentrations of two successive proteins for the periodic solution in the odd ring with $N = 23$. The solution has a travelling-wave structure with a kink-like perturbation propagating around the ring, indicated by the arrow in the bottom figure. The bottom figure represents the minimum distance $|\Delta p_j| = \min(|p_u - p_j|, |p_d - p_j|)$ between the traveling wave solution and the dimerized solution with an alternating pattern of protein expression given by p_u and p_d . The distance becomes large around the kink in the traveling wave solution. (c) Same as (b) for the quasi-stable periodic solution of the even ring with $N = 22$. In this case, the travelling wave solution has two kinks that propagate around the ring, as indicated by the arrows.

the direction of the orbit.)

In some cases, a few (possibly only one) Floquet multipliers will be slightly larger than one. We will then have ‘quasi-stable’ periodic solutions in that it takes a long time to diverge away from them. Quasi-stability in this sense is a local property. To assess if these solutions will be reachable (and therefore observable in the dynamics), one needs to employ global techniques, e.g. sampling the space of initial conditions. However, Floquet analysis provides an indication of the possibility of observable, yet unstable, periodic solutions. If a periodic orbit has a small number of very weakly unstable directions, it is likely that it could be observed as long-lived periodic transients in the deterministic system and that it could also play a role in the stochastic dynamics. Moreover, such quasi-stable orbits are good targets for a control mechanism that can make oscillations switchable, as shown below.

2.3 Numerical methods for oscillator design

In the following we summarize the methods, which have been used for the analysis and the design of our oscillator.

2.3.1 Numerical simulations and analysis of the dynamics

The deterministic system of ODEs (3.1) was solved numerically with an adaptive fourth-order Runge-Kutta integrator (77), in which the step-size automatically adapts to meet the required accuracy ϵ . We have checked that the inducibility, reachability and transient times of the quasi-stable oscillations are not affected by the accuracy of the integrator by using the Runge-Kutta integrator with accuracies ϵ between 10^{-2} and 10^{-8} (see Fig 2.6). In addition, the observability of the unstable orbits was confirmed by using a nonlinear integrator (78).

The bifurcation analysis and the calculation of the Floquet multipliers of the unstable periodic orbits shown in this chapter were carried out with the numerical continuation software AUTO (75).

Stochastic simulations of the generalized repressilator were performed using the Gillespie algorithm (see section 1.3.3). Random numbers and quasi-random numbers for numerical simulations were generated with the GSL Scientific Li-

brary (79).

2.3.2 Global robustness analysis and control aspects of quasi-stable oscillations

As part of our numerical evaluation of the generalized repressilator, we have developed a method to carry out a robustness and reachability analysis of its quasi-stable oscillations. This was necessary because available global robustness tools (78) quantify changes in fixed points induced by parameter variations and are therefore not directly applicable to oscillations. In order to evaluate the global robustness and inducibility of the quasi-stable oscillations, we attempt to induce sustained oscillations with a predetermined intervention and quantify changes in the observed response when the model parameters are varied. The method defines an operating point in parameter space (the reference set c_j^*), based on biologically appropriate estimates, and a hypercube around it to account for biological variability, temperature gradients and other noise. We then sample parameter sets from the hypercube using Reverse Halton Sequences (80), quasi-random sequences that have been shown to homogenically cover hypercubes faster than standard Monte Carlo sampling for high dimensional spaces (81). For each sampled parameter set, we attempt to induce oscillations with the STOP-KICK scenario described below and record if the system evolves towards sustained oscillations. If oscillations are observed, we calculate numerically the period of the oscillation and the change in shape of the oscillation, the latter defined as the minimal normalized mean square deviation (L_2 norm) of the observed oscillation with respect to the reference operation point (see next sections for details).

Characterizing the change in shape is essential to establish that the oscillation remains detectable and functionally recognizable in the biological system. Note that here we are only concerned with global robustness of the reachability of the solution. A modification of the same algorithm could be used to study the parameter combinations that contribute most strongly to the sensitivity of the network, a question relevant for the experimental tuning of the system that is not addressed here.

2.3.3 Robustness analysis and shape similarity

We have performed global robustness analysis for the generalized repressilator model, which means all parameters have been varied at the same time within a specified multidimensional hypercube. For instance, for the global robustness of 10 gene ring, the dimension of the according parameter space and the hypercube is 40, as there are 4 constants for each gene (c_1, c_2, c_3, c_4) and these are varied independently of each other. This procedure is fundamentally different to standard robustness assessments, where only one parameter is gradually varied at the same time.

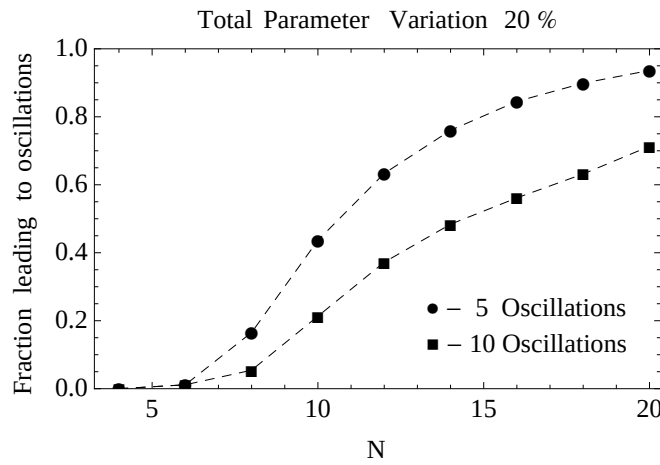


Figure 2.3: **Sample output for robustness analysis.** On the x-axis we plot the number of genes in the ring and on the y-axis the fraction of samples is shown, for which the model was in the oscillating state after 5 and 10 oscillations. The total parameter variation of 20% around the reference set was used for the figure.

The parameter hypercube has been defined as 5%, 10%, 20% variation around a reference state, where all parameters c_i are the same for all genes. We have sampled this parameter hypercube using pseudo-random numbers created with reverse Halton sequences. Halton sequences have been used instead of conventional random sampling, because they have been shown to be more efficient than the classical Monte-Carlo for large dimensional space sampling (81). (The C code for the global robustness analysis is available from the author on request.)

For all parameter samples we have assessed if the ODE system Eq. 3.1 was in the oscillating state after certain number of periods. The "yes"/"no"-decision for the generalized repressilator was based on the condition that all genes in the ring

show a non-decayed full amplitude. Using 10^4 samples for each N -gene-ring, we have assessed the robustness towards parameter variations of the STOP-KICK scenario (see section 2.4.3). The fraction of the parameter samples, which has lead to at least 5, 10, 20 oscillations has been plotted on the y-axis (see figure 2.3 and result2.9).

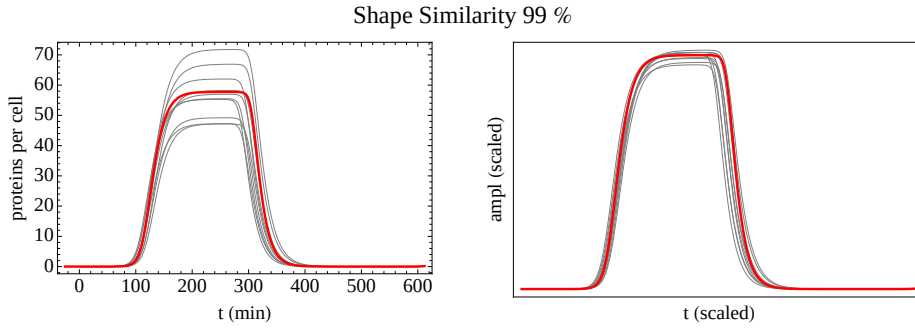


Figure 2.4: **Shape variations under the global parameter changes.** On the left we see the numerically observed oscillations, which differ moderately in the period length and amplitude. The reference shape for the unperturbed system is shown in red. On the right the time scale (x-Axis) has been scaled with the period and the maximal amplitude for each curve respectively. This demonstrates how the scaling in time and normalized L_2 score assesses the shape similarity.

In addition to the "yes"/"no"-decision about the system being in the oscillating state, we have also quantified the change in the shape of the oscillation. This measure is independent of the variation in period and amplitude. The measure **ssh** would give small similarity score if the oscillations became spiky, i.e. the time spent in the up state is much smaller than the time in the down state or vice versa. The similarity score would be intermediate if the oscillation shape became more like a Gaussian instead of rather quadratic reference shape (see Fig. 2.4). The similarity score is calculated by first scaling the time for the reference shape and the perturbed shape to the same length and then applying the standard normalized L_2 -norm:

$$\mathbf{ssh} = \langle \pi_{ref}(t) \pi_{pert}(t) \rangle / \langle \pi_{ref}(t) \rangle / \langle \pi_{pert}(t) \rangle \quad (2.3)$$

The score is independent of both variations in the period and amplitude, because through the first scaling the length variations disappear and through the normalization in the L_2 -norm the amplitude fluctuations disappear. Thus, our

measure assesses shape changes for moderate parameter perturbations and gives an additional information to the variation in the oscillation amplitude and period length.

2.3.4 Basic Floquet theory and quasi-stable oscillating orbits under small perturbations

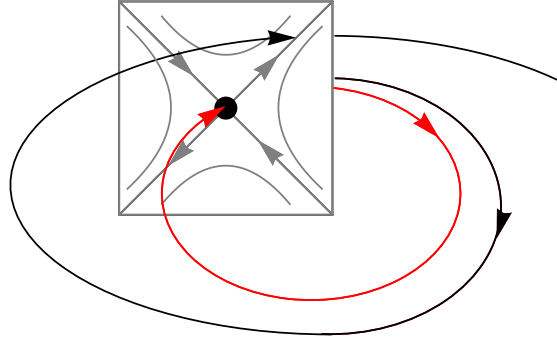


Figure 2.5: **Poincare map with quasi-stable fixed point (black) and possible trajectories.** The periodic orbit (red trajectory) is linearly stable along the left diagonal, where the arrows are pointing towards the fixed point and unstable along the right diagonal, where the arrows are pointing away from the fixed point. The escape scenario from unstable periodic orbit due to perturbations is illustrated with black trajectory.

Oscillating modes of an ODE system can be assessed regarding their stability with the Floquet multipliers theory (49). Here, the question about stability of an orbit can be reformulated to the question for the stability of the corresponding fixed point x^* in the Poincaré map. If v_0 is an infinitesimal perturbation such that $x^* + v_0$ is in S , then after the first return to S

$$x^* + v_1 = P(x^* + v_0) \approx P(x^*) + [DP(x^*)]v_0 \quad (2.4)$$

The Eigenvalues λ_j of the linearized Poincare map $[DP(x^*)]$ obtained this way are nontrivial Floquet multipliers. The closed orbit is linearly stable if and only if $|\lambda_j| < 1$ for all j . We illustrate the case where the closed orbit has unstable directions and how the system may escape from it in Figure 2.5 above. The Poincare section (shown in gray) has stable directions (left diagonal) and unstable directions (right diagonal). The system may escape the closed orbit (black trajectory) in a directed manner once perturbed along the unstable direction. If for a periodic orbit in a

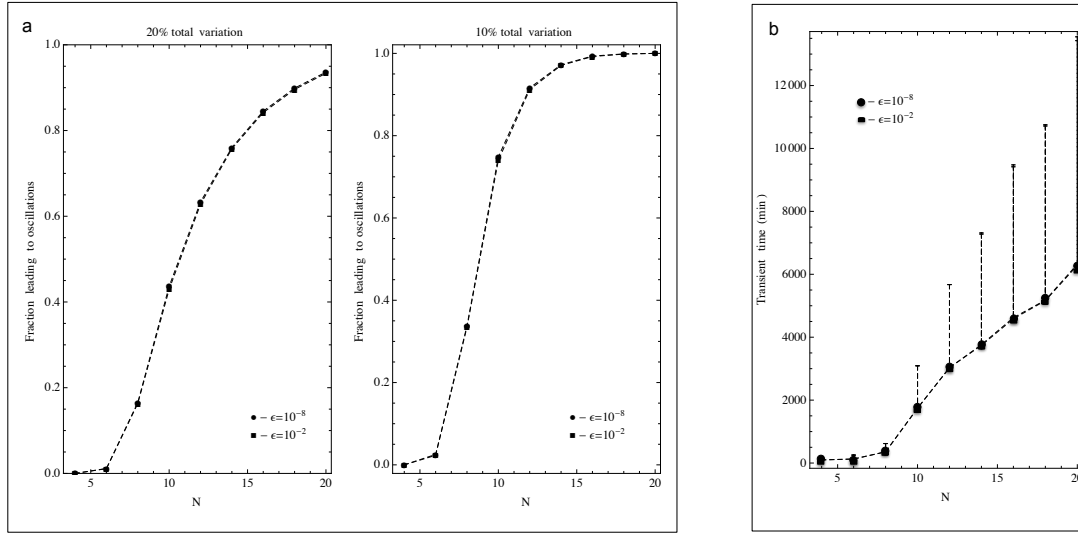


Figure 2.6: Quasi-stable periodic orbits under small perturbations. (a) Less accurate integration with $\epsilon = 10^{-2}$ (circles) does not change the numerical outcomes of the global robustness analysis, which has also been performed for the integrational accuracy $\epsilon = 10^{-8}$ (squares). All other parameters are identical for the inspected simulations. Less accurate integration is a simple way to simulate the small noise and with the figure we show that the robustness of quasi-stable oscillations is not influenced by small fluctuations. Larger variations in form of internal noise due to the low copy number can be simulated with Gillespie algorithm (see section 1.3.3) and we show later (see chapter 4) that the quasi-stable oscillations are also observed under larger fluctuations. (b) Different integrator accuracy also does not affect the length of transient oscillations. The figure illustrates the mean lengths of oscillations, which occur if starting from randomly chosen initial conditions and inspect the region for the first 50 oscillations. The error bars are first and third quartiles.

multi-dimensional system only few directions are unstable and the escape rate is slow, we would observe long-lasting oscillating transients, which we call quasi-stable oscillations.

A relevant question to assess is how the escape rate along an unstable direction would be influenced by noise. We can simulate small noise effects by simply using less accurate integration with larger maximal error ϵ per step (see Fig. 2.6). It turns out that the duration of transient quasi-stable oscillations is as long as with the accurate integration. So small noise effects do not influence the quasi-stable oscillations. The investigation of larger perturbations, which will be carried out in chapter 4 shows that noise even extends the duration of quasi-stable oscillations.

We try to give an intuitive explanation for the observed effect. Errors resulting from less accurate integration affect all variables in the ODE system. It has the effect that the system is perturbed in all directions due to this noise. However,

a quasi-stable orbit can be escaped through directed movement along one out of many possible directions (see Figure 2.5 above). If the system is perturbed along other directions, it is pulled back to the limit cycle, because all other directions are attractive. Therefore random perturbations do not necessarily kill the transient oscillations. On the contrary, random perturbations may interfere with the directed escape out of the limit cycle and this might even stabilize the oscillations. In fact in the following chapter 4 we assess the influence of internal noise on the onset and the duration of the quasi-stable oscillations and we show numerically that the oscillations are stabilized in the noisy system.

2.4 Results

2.4.1 Stable and quasi-stable oscillations in the generalized repressilator

As pointed out in the section 2.2 above, odd repressilator rings are globally attracted to stable limit cycle oscillations for $c > 2$. Numerical simulations show that the period of these solutions increases linearly with the number of genes in the ring (Fig. 2.2a). The stability analysis also shows that, in contrast, even rings only support fixed points as stable solutions. However, direct dynamical simulations of even repressilator rings reveal the existence of long-lived periodic solutions, which are easily reachable, as checked by extensive sampling (not shown) of the space of initial conditions. The period of these oscillations also increases linearly with the number of genes, albeit with a slope that is approximately half of that in odd rings (Fig. 2.2a).

These numerical observations do not pose a contradiction with the stability analysis above: the observed oscillations in even rings are periodic solutions yet unstable. Unstable solutions can be studied using the numerical bifurcation detection software AUTO (75), a continuation package that does not rely on dynamical simulations. We have used AUTO to find bifurcations in the biologically relevant range of the parameter c and to assess the linear stability of fixed points and periodic solutions—the latter through Floquet analysis.

The result for even rings is presented in Fig. 2.7. In agreement with the analytical results, which we will show in the next chapter, a pitchfork bifurcation

is found numerically as a branching point at $c = 2$, above which Hopf bifurcations leading to the appearance of unstable periodic solutions, are detected in all rings with more than 4 genes. The Floquet stability analysis indicates that the first unstable periodic orbit to emerge has only one unstable direction, regardless of the number of genes. The only positive Floquet multiplier, which indicates how fast the trajectory diverges away from the orbit, is small and decreases as the length of the ring increases. This is the signature of quasi-stability: if this periodic orbit is reached, it will be long-lived. We have also checked that this solution is reachable through numerical sampling of the space of initial conditions (see Fig. 2.9a–b). Such reachable quasi-stable modes affect significantly the observed transient dynamics and also play a central role in stochastic dynamics where unstable solutions are explored under the effect of noise. Both these conditions are relevant for dynamics of genetic circuits inside the cell.

The existence of quasi-stable modes provides us with the opportunity to design a distinct control strategy. If we operate the system to revolve around a quasi-stable mode, the result is an oscillator that can be switched on, kept oscillating and switched off on demand. Below, we introduce a simple implementation of such a scenario and evaluate its robustness of operation. Note that an intricate family of unstable periodic orbits with high symmetry exists both in odd and even rings (see Fig. 2.7 and Fig. 2.4.1). However, these additional periodic solutions have several unstable directions that make them essentially unobservable and uncontrollable.

2.4.2 Traveling wave structure of the periodic solutions

The spatio-temporal structure of the periodic solutions, both in the odd and even cases, corresponds to a travelling-wave solution propagating around the ring. The snapshots in Figs. 2.2b–c show that this propagation occurs against the backdrop of the dimerized fixed point solution of the even ring, where all odd (even) numbered genes are ‘up’ while the even (odd) numbered genes are ‘down’ (3.3). Clearly, a dimerized configuration cannot be accommodated in an odd ring. This leads to a kink-like (frustrated) solution where two consecutive genes have similar expression levels. This local imbalance of repression induces a dynamical instability that makes the kink propagate around the odd ring in a periodic fash-

N	Bifurcation	c	Stable/all directions	Max. Floquet	Period (min)
2	P	2.00	-	-	-
4	P	2.00	-	-	-
6	P	2.00	-	-	-
	HB	4.68	11/12	6.2	132
8	P	2.00	-	-	-
	HB	3.00	15/16	3.9	186
10	P	2.00	-	-	-
	HB	2.56	19/20	3.0	239
	HB	9.62	17/20	9.0	104
12	P	2.00	-	-	-
	HB	2.36	23/24	2.5	291
	HB	4.68	21/24	6.2	132
14	P	2.00	-	-	-
	HB	2.24	27/28	2.2	342
	HB	3.51	25/28	4.8	160
	HB	16.9	23/28	10.5	94
16	P	2.00	-	-	-
	HB	2.18	31/32	2.0	393
	HB	3.00	29/32	3.9	186
	HB	6.91	27/32	7.8	113
18	P	2.00	-	-	-
	HB	2.15	35/36	1.8	444
	HB	2.71	33/36	3.4	213
	HB	4.71	31/36	6.2	132
	HB	27.33	29/36	11.5	89

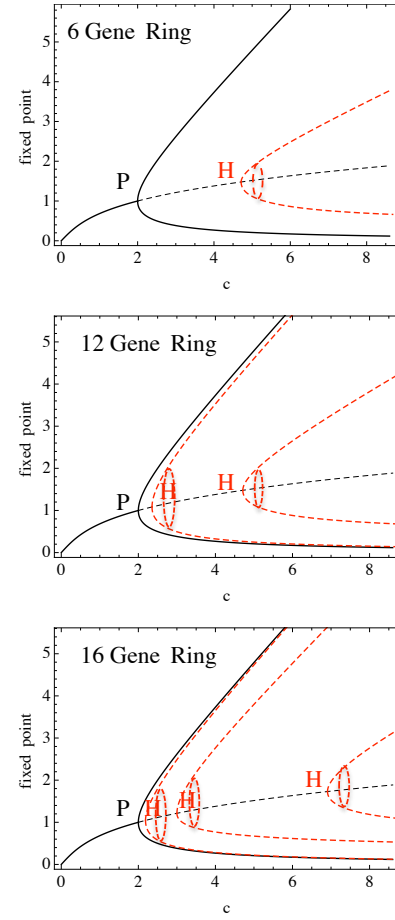


Figure 2.7: Numerical bifurcation analysis and unstable periodic solutions of repressor rings with even number of genes. We use the continuation package AUTO (75) to obtain the bifurcations of rings of size N (3.1). The parameter c , defined in Eq. (2.2), is swept in the biologically relevant range $c \in [0.001, 30]$ by changing c_1 with $c_2 = 0.12$, $c_3 = 0.16$ and $c_4 = 0.06$ constant. A branching point corresponding to a pitchfork bifurcation (P) is found at $c = 2$. A series of Hopf bifurcations (HB) linked to the emergence of unstable periodic solutions are found subsequently. Floquet analysis indicates that the first unstable orbit to emerge has only one unstable direction, regardless of the dimension of the system, and that the maximal Floquet multiplier decreases with increasing N . Hence this periodic solution is quasi-stable: if it is reached, the divergence away from it is slow, and gets slower for longer rings. Other unstable orbits are present but their high instability makes them irrelevant to the observed dynamics. A similar structure of unstable orbits exists in odd rings (see Fig. 2.4.1). The figure on the right shows the bifurcation diagrams for even rings of length $N = 6, 12, 16$. The unstable periodic orbits, shown in red dashed lines, emerge through Hopf bifurcations.

N	Bifurcation	c	Stable/all directions	Max. Floquet	Period (min)
3	HB	4.689	6/6	1	132
5	HB	2.556	10/10	1	239
7	HB	2.244	14/14	1	342
	HB	16.91	12/14	8.0	94
9	HB	2.156	18/18	1	444
	HB	4.711	16/18	5.0	132
11	HB	2.089	22/22	1	545
	HB	3.311	20/22	3.75	169
	HB	42.84	18/22	11.0	86
13	HB	2.067	26/26	1	645
	HB	2.800	24/26	3.05	204
	HB	7.778	22/26	7.55	110
15	HB	2.044	30/30	1	746
	HB	2.533	28/30	2.63	239
	HB	4.711	26/30	5.76	132
17	HB	2.044	34/34	1	847
	HB	2.400	32/34	2.35	274
	HB	3.667	30/34	4.68	154
	HB	11.75	28/34	9.13	100

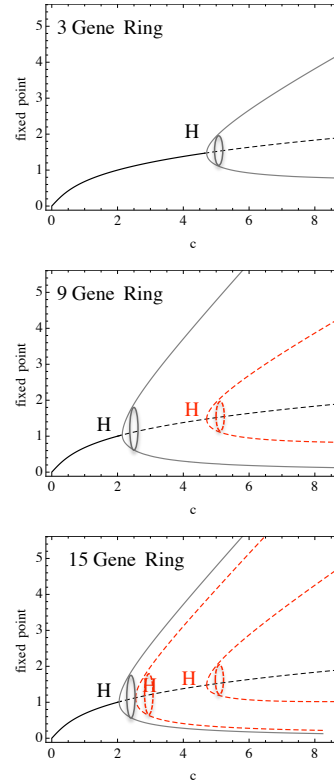


Figure 2.8: **Numerical bifurcation analysis of repressilator rings with odd number of genes and corresponding bifurcation diagrams.** We use the continuation package AUTO (75) to obtain the bifurcations of rings of size N (3.1). The parameter c , defined in Eq. (2.2), is swept in the biologically relevant range $c \in [0.001, 30]$ by changing c_1 with $c_2 = 0.12$, $c_3 = 0.16$ and $c_4 = 0.06$ constant, as in the previous figure 2.7.

ion (Fig. 2.2b). This spatio-temporal structure underlies the limit cycle solution in odd rings. The fact that the period of the limit cycle increases (roughly) linearly with the number of genes indicates that the speed of propagation of the kink is (roughly) constant.

The quasi-stable periodic solution in even rings can be interpreted under the same prism. Fig. 2.2c shows that it corresponds to *two interacting kinks* propagating around the ring at a roughly constant speed with a period that is approximately one half of that of the closest odd ring (Fig. 2.2a). The instability of this periodic solution has a clear meaning in this picture: if the two kinks ‘collide’, they annihilate each other and the system returns to the stable fixed point, i.e., the dimerized solution. Figure 2.2c also shows that each kink has a minimum spatial width that depends on the parameters of the model. Hence it is more difficult to find these oscillatory solutions in rings that are not large enough to fit two such perturbations although they can still be observed in smaller, biologically realizable rings. For clarity, we have chosen to illustrate the spatio-temporal structure of the solutions with long rings. However, we have checked that the quasi-stable periodic orbits in rings with $N = 6, 8, 10$ (not shown) maintain the features of the two-kink structure and operate under the same principles as the long rings shown in Fig. 2.2c.

The spatio-temporal structure of the periodic solutions in repressilator rings shows a strong parallelism with similar dynamical solutions observed in classical models of discrete lattices (82). The travelling-wave nature of the oscillations could have potential biological applicability if one were to use this circuit as a control element for genes that must be activated in a particular order and for a pre-defined time interval.

The spatio-temporal structure will be assessed in more detail in the following chapter by connecting it to the symmetries in the generalized repressilator model in its solutions.

2.4.3 Robust induction of quasi-stable oscillations in the deterministic regime

To test that the oscillations described above are reachable, and therefore observable, we have carried out a sampling of the convergence to such solutions from

random initial conditions. The results are presented in Fig. 2.9*a–b*, where we show that the proportion of initial conditions leading to sustained oscillations in even rings increases as the ring becomes longer but are already substantial for $N = 10$. This indicates that such oscillations could be observable in transient dynamics, especially in the presence of noise. We will assess systematically the influence of noise on the transient length of quasi-stable oscillations in the chapter 4 and show an increased observability of these oscillations in the corresponding stochastic system.

The causality imposed by the travelling-wave structure means that the manipulation of one gene will have a predictable effect on the others. This means that we can use such causality to devise a protocol to induce and stop oscillations in even rings reliably by activation of one gene for a short time span. We illustrate this simple scenario in Fig. 2.9*c*. First, the even ring is forced to converge to one of the fixed point solutions with a STOP signal that consists of the external activation of gene j for a time interval longer than the period of the oscillation. This signal is used to ‘initialize’ the system, suppressing any transient oscillations present in the system. Once the system is at rest, the oscillation can be started with a KICK signal, consisting of the external activation of gene $j + 1$ with a step function of width and amplitude similar to those of the oscillatory pattern. Such signals can be elicited non-invasively through an optical mechanism that uses UV or red light to activate the production of mRNAs of particular genes (28, 29).

The STOP-KICK scenario induces long-lived oscillations with absolute reliability in rings with $N > 6$ identical genes. In order to check that the proposed protocol is robust to parameter variations, we have carried out a global robustness analysis as outlined in the Methods section. We construct a hypercube by taking variations of 5%, 10% and 20% around the reference values of the parameters in Eq. (3.1) and take 10^4 samples in this hypercube varying all parameters simultaneously. Sampling is performed with quasi-random Reverse Halton Sequences for improved convergence (80, 81). Figure 2.9*d* shows that the fraction of parameter samples that lead to oscillations with this protocol converges to 1 for large rings. Our numerics also show that oscillations can be elicited with significant robustness in rings with $N > 6$. When oscillations are present, the period shows very small variation with respect to the reference set, as shown by the coefficient of variation (Fig. 2.9*d*, inset). We have also quantified the change in the shape of oscillations through a normalized mean square measure and found that the shapes in the perturbed system exhibit very high similarity ($\approx 99\%$) to the reference set (see section 2.3.3). In summary, our global robustness analysis indicates that in the deterministic regime, long-lived quasi-stable periodic solutions are reliably inducible in larger rings with a single intervention. For moderate size rings, oscillations can still be induced for a large fraction of the parameter hypercube but are shorter lived. This suggests that repeated interventions could be used in order to keep the ring in the quasi-stable oscillating state. A simple control protocol that implements these ideas is proposed in the following section and shown to be applicable for rings as small as $N = 6$ operating in the stochastic regime.

2.4.4 Stochastic oscillations in even rings and readout-based control

We have used the slightly modified Gillespie algorithm (see section 1.3.3) to study the generalized repressilator in the stochastic regime, i.e., when intrinsic noise is high due to low copy numbers. It is well known that stochastic models of odd rings behave as oscillators and that the travelling wave structure is preserved (2, 83). In the case of even rings, we have performed long stochastic simulations (not shown) and found bistability and switching events, as expected from the switch-like stationary probability function (see sections 1.3.3 and 1.3.4).

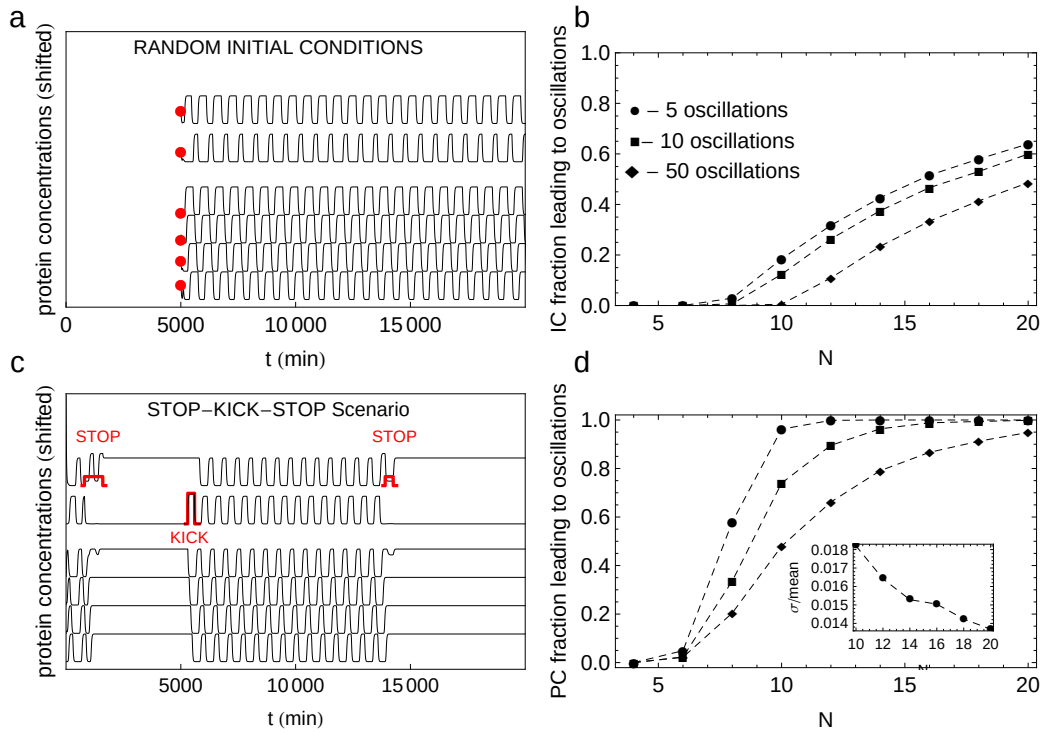


Figure 2.9: Induction of oscillations in even rings in the deterministic regime **a)** Quasi-stable periodic solutions in even repressilator rings can be observed starting from random initial conditions. The dynamical trajectories shown here correspond to a ring with $N = 18$ and parameters $c_1 = 2.6$, $c_2 = 0.12$, $c_3 = 0.2$, $c_4 = 0.06$. **b)** Percentage of initial conditions (IC) leading to 5(●), 10(■) and 50(◆) oscillations. The percentages are obtained from a sampling of 10^4 initial conditions. **c)** The quasi-stable oscillations can be induced reliably with a simple sequence of signals. First, apply a STOP signal to gene j to force the system to approach a fixed point solution. Second, apply a KICK signal to gene $j + 1$ to drive the ring into oscillation. The oscillation can be terminated at will by applying another STOP signal. The signals can be implemented via on-demand UV or red light gene transcription activation (28). This STOP-KICK-STOP protocol is shown here for the same ring as in (a). **d)** Global robustness of the inducibility of the quasi-stable oscillations. The STOP-KICK scenario is applied to 10^4 random parameter combinations (PC) c_j for each even ring of length N and we record the proportion of parameter sets that lead to five oscillations. The parameters are sampled with reverse Halton sequences from a hypercube with 5%(●), 10%(■) and 20%(◆) variation around the reference set. Quasi-stationary oscillations are robustly induced for $N \geq 10$ (see Fig. 2.3 for 10 oscillations). Small rings can be kept in the oscillating state applying repeated interventions in a simple control protocol, as seen in Fig. 2.10. (Inset) The oscillations are robust in shape (see Fig. 2.4) and in period with respect to changes in the parameters. The relative variability (coefficient of variation) of the period of the induced oscillations is small and decreases with the length of the ring.

Additionally, the simulations show sustained oscillatory behavior, especially in longer rings although they are also observable in rings as small as $N = 6$. The oscillations appear in a variety of settings: as transients from a variety of initial conditions; spontaneously emerging from one of the stable points; or associated with switching events. It is also easy to induce such oscillations with localized interventions in particular genes; to sustain them with periodic driving; and to terminate them with a prolonged induction of a gene (as in the STOP signal above). We have examined the structure of these oscillations and they correspond well with the quasi-stable periodic solutions in the deterministic system: their period is approximately half the period of the closest odd ring (Fig. 2.2a) and the spatio-temporal travelling-wave structure is maintained.

Our numerical simulations confirm the relevance of the underlying quasi-stable oscillations for the observed stochastic dynamics of even rings (Fig. 2.10). Similarly to the deterministic case, the quasi-stable mode can also be used as a control operating point such that the system becomes switchable. The robust reachability of this mode allows us to use an extremely simplified feedback mechanism that could be implemented through an optical read-out (GFP, YFP or luciferase protein labeling) and response (on-demand UV or red light gene transcription activation) (28).

The simple control scheme illustrated in Fig. 2.10a uses the optical read-out from two successive proteins in the ring to introduce optical KICK signals that sustain the oscillation based on a threshold rule (Fig. 2.10b). The oscillation can be started and terminated using the same optical signals. Although we have chosen to illustrate the possible implementation of the scheme with light sensitive inducers, it is worth remarking that any suitable mechanism capable of precisely-timed gene induction with good spatial resolution at the cell population could be used. A potential advantage of a switchable mode of operation is the economical and targeted use of the transcriptional resources without overburdening the cell with unnecessary mRNA production (66).

2.5 Discussion

In this chapter, we have studied how the presence of quasi-stable periodic solutions affects the observable dynamics of even repressilator rings. Previously, even

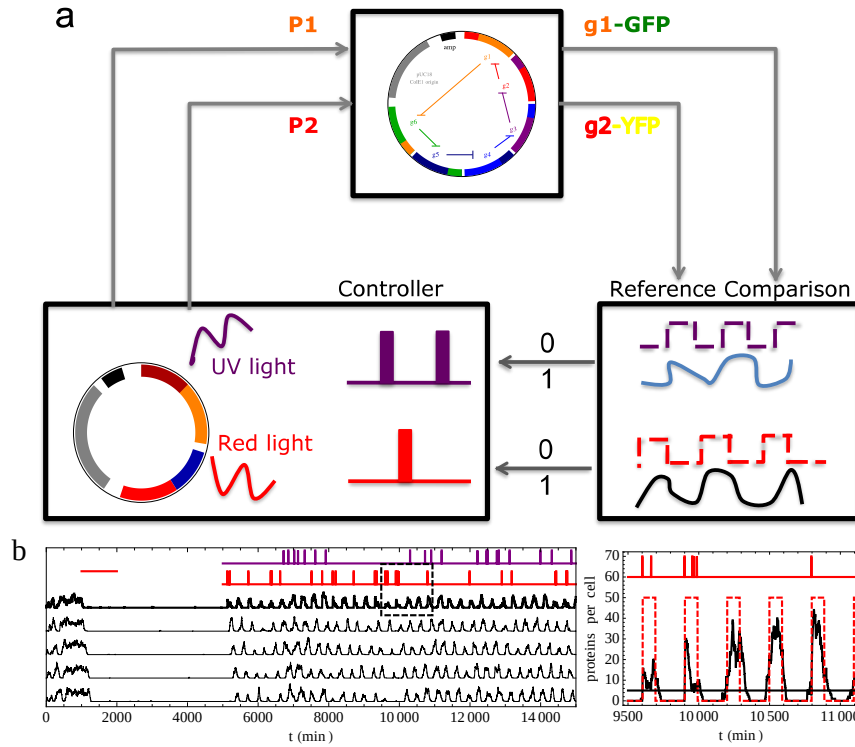


Figure 2.10: Stochastic oscillations in even rings and readout-based control. (a) Illustration of the readout-based control scheme for a ring of 6 genes. Two proteins of the ring are read out with fluorescent tags. This readout is then compared to a reference defined according to the oscillating behavior of the ring, with similar period and a shift between consecutive genes. The reference comparison is threshold-based and leads to an ON-OFF (1-0) control for the KICK signals. These can be implemented with light responsive genes promoters. In the numerical simulations shown in (b), the KICK signals are indicated with the red markings in the upper panels. (b) A simple readout-based control reliably switches on the oscillations, sustains them and switches them off. The control mechanism functions by monitoring two successive proteins in the ring. Whenever each of them falls below a threshold, a KICK signal for the corresponding protein is given. These threshold-based KICK signals are indicated with red and magenta markings in the upper panels. The oscillation can be terminated with a STOP signal as in the deterministic state. The optical read-out can be based on GFP or YFP protein labeling while the response can be implemented with on-demand UV or red light that enhances the production of the corresponding mRNAs (28). The figure shows the application of this mechanism to a ring with $N = 10$. The stochastic time traces correspond to the protein expression of proteins p_j with $j = 1, 3, 5, 7, 9$ and the corresponding control (top) in response to proteins p_1 and p_2 (trace not shown). The right figure is a magnification of the dashed square inside the main figure. We have also checked that this control protocol is applicable for rings with as low as $N = 6$ genes (not shown).

rings have been thought of as switches due to the fact that they only support fixed point solutions. However, our bifurcation analysis reveals the existence of a set of unstable orbits, some of which have slow timescales associated with them. These quasi-stable periodic solutions are both reachable and long-lived, thus playing a role in the observed dynamics, both transient and stochastic. This suggests that oscillatory behavior might be more widespread than expected in genetic models since it could feature in systems that possess only static attractors.

The presence of quasi-stable solutions provides us with the possibility of designing control protocols that operate the system around such modes so that the oscillations can be turned on and off reliably. Our numerics indicate that a robust mechanism could be implemented based on appropriate optical feedback to switch the system between stable fixed points and quasi-stable oscillations. Although the proposed pared-down control scheme is only intended to provide an illustration of the potential implementation and its performance could be improved with the use of optimized strategies for stochastic and robust control that take into account specific details of the experimental setup, it is worth discussing some of its limitations. The challenge for the dynamical control algorithm is to deliver the optical interference signal necessary for the induction of gene expression for a short time period, to a particular spatial area of the cell population, and with a well-controlled delay following the fluorescent signal of the proteins. The proposed scheme shows both enough spatial and time resolution to address individual cells in a population with well-defined pulses (28). The scheme would need to rely on proper calibration of life-times of fluorescent proteins affected by phototoxic and photo-bleaching effects, as reviewed in detail by Bennett and Hasty (30). Finally, we note that the delay between the actual protein concentration and the corresponding fluorescent signal introduced by the maturation time (~ 2 -8 min) is short compared to the period of the oscillation (~ 130 min), hence acceptable for the control scheme.

A synthetic circuit operating under such principles could be interfaced with a naturally occurring network to induce an intrinsic interference that is interruptible on-demand. The switchability of this regulatory element can help avoid the appearance of adverse cumulative effects. The NF κ B pathway is an example where such a regulator could provide controlled activation over short time intervals as an alternative to conventional knock-downs and other functional interventions

that modify the balance of important proteins for the cell cycle (84, 85). The underlying travelling wave structure of the observed periodic solutions could also be potentially useful for design purposes. It allows for coordinated intervention when the timing and order of activation of different pathways is crucial. Examples of cellular networks, e.g. in developmental biology, indicate that timed patterns of sequential activation are at the heart of the function of families of master regulators (61, 64, 86, 87) and, in the case of the vertebrate segmentation clock (88, 89), the associated oscillations are well-defined but do not survive in the long-term. The importance of heterogeneously timed gene induction has also been highlighted in a model of Arabinose uptake in *E. coli* (63). Experiments with genetically engineered yeast have also shown that pulsed activation of chaperons followed by pulsed activation of the associated heterologous proteins is more efficient at maximizing the production of particular metabolites (66). These applications hint at potential uses for circuits that can produce sequential patterns of activation on demand, such as the even repressilator studied here, which interact with other cellular pathways via intrinsic proteins, thus avoiding the timed delivery of external agents through the cell membrane.

The design of control strategies for the operation of systems around an inherently unstable state has a long history in other disciplines (e.g., flight and fluid control) since it affords enhanced responsiveness and controllability with relative ease and simplicity of design (71). This strategy differs fundamentally from the biochemical alteration of the network topology proposed by Atkinson et al. (74) based on a smaller gene circuit but with a complex regulatory scheme involving promoter and repressor sites regulating one gene. The molecular kinetics of such regulators are less well understood than those with single regulatory sites due to unavoidable cross-talk and compound logic. The ring topology relies on simple regulation to provide a sequence of causal signals but at the expense of involving a larger number of genes.

The present scheme is also in contrast with previously engineered gene circuits, such as odd repressilators, which possess globally attracting limit cycles leading to behavior that is robust yet not controllable. Quasi-stable operation, on the other hand, is robustly switchable. The switchability of the oscillator coupled with dynamic control that affords good spatial resolution could be used to elicit localized oscillations in cell populations as an aid to examine mechanisms of cell

synchronization.

It is an open area of current research to elucidate the role of a design concept based on control around unstable behavior, similar to the inverted pendulum in classic control theory, to further our understanding of cell strategies and its potential use in the design of synthetic topologies that can interfere with naturally occurring pathways. We explore in more detail the control strategies and numerical algorithms for their optimization in thesis chapter 5.

Chapter 3

Analytical conditions for Hopf bifurcations in uni-directionally coupled rings

Previously we have explored the generalized repressilator model for design purposes of a controllable genetic oscillator. The intuitive threshold-based control strategy proved successful for the example, because the generalized repressilator model has ring symmetries. Using these inherent symmetries we establish connections to engineering designs in other scientific disciplines and derive mathematically more rigorous analytical conditions for the emergence of Hopf bifurcations in this chapter.

The following chapter is based on the manuscript

Transient dynamics around unstable periodic orbits in the generalized
repressilator model

Natalja Strelkowa and Mauricio Barahona

Provisionally accepted by

Chaos

3.1 Introduction

Ring structures of coupled elements are an appealing mathematical concept, because of the inherent symmetry and therefore amenability to analytical treatment (53, 90, 91). Ring-coupled elements are also of general practical interest for engineering and biology (2, 51, 92). There are several examples of well-studied ring-coupled systems, as for instance the Duffing oscillators (53) and magnetic flux-gates (51). The ring structures were reported to have rich dynamical behavior, including multi-stability, unstable periodic solutions (UPOs), quasi-periodicity and even chaos and hyperchaos.

Recently an extended numerical and analytical treatment of the autonomous unidirectionally coupled Duffing oscillators has been performed by Perlikowski et al. (53). The study focuses on investigation of chaos and hyperchaos in the unidirectionally and linearly coupled oscillator system. It also explains the mechanism of chaos emergence, which in this case occurs via transient oscillations. The authors explore the role of unstable periodic orbits (UPOs) in chaos emergence and therefore spend considerable effort on the analysis of transient dynamics, which in many studies is rather neglected due to their instability.

The unidirectionally coupled elements have also been built and used as sensors. A series of papers appeared on a successful design of magnetic field sensors, which is based on the unidirectionally ring coupled magnetic flux-gates (6, 51, 93, 94). The system has been well analyzed, including deterministic steady state dynamics (94), and numerical bifurcation analysis (51). A more recent study in this area (6), reports that longer even-numbered rings of unidirectionally coupled flux-gates, which are bi-stable according to the steady state analysis, show observable, extremely long-lived oscillating transients and this behaviour is more prevalent in the presence of noise. The study is an example of an electronic design, where transient behavior is observable (in this context we mean likely to occur starting from random initial conditions) and extremely long-lived. Investigation and utilization of long-lived transients is a research direction complementing the study of steady states and the nontrivial dynamics appear in biological systems as well as in the classical electronics.

Previous studies on the generalized repressilator model were mainly devoted to the steady state analysis and stable limit cycles, which are characteristics deter-

mining the system property for infinite time $t \rightarrow \infty$. Analysis of how the system approaches the fixed point, i. e. transient unstable behavior in particular unstable periodic orbits were rather neglected. However, steady state analysis alone is not sufficient in cases where transients are non-trivial, qualitatively different from steady states, and occur on slow time scales compared to the inherent time scales of the system. As shown in (5) and the first chapter, even-rings of the generalized repressilator model are examples of a genetic network with nontrivial transient behavior. The oscillating transients are long lived and the globally attracting stable fixed points of even-rings are approached through the very long lasting transient oscillations. Life-times of these unstable periodic orbits are large compared to inherent time scales of the system, as for instance generation time of bacterial population. Therefore the exploration of unstable periodic orbits in the generalized repressilator model is necessary to complement the classical steady state analysis.

In this chapter unstable periodic orbits of the generalized repressilator model are characterized in terms of their stability, duration and their inherent symmetries. We provide analytical conditions for Hopf bifurcations, from which unstable periodic orbits are originating as a function of the length of the ring N . We also establish connections to classical physics concepts of kink-like propagating perturbations. It turns out that dynamical symmetries of these unstable periodic orbits can be understood in terms of moving kinks and that the symmetries are related to the stability of these unstable periodic orbits. We compare our system with the magnetic sensors (93) using classical group theoretical approach (52) and establish the relationship between this known real-world implemented system and the generalized repressilator model. Finally, we show that the unstable periodic orbits in even-rings are reachable if starting from random initial conditions and explain the duration distribution of oscillating transients using kink concept.

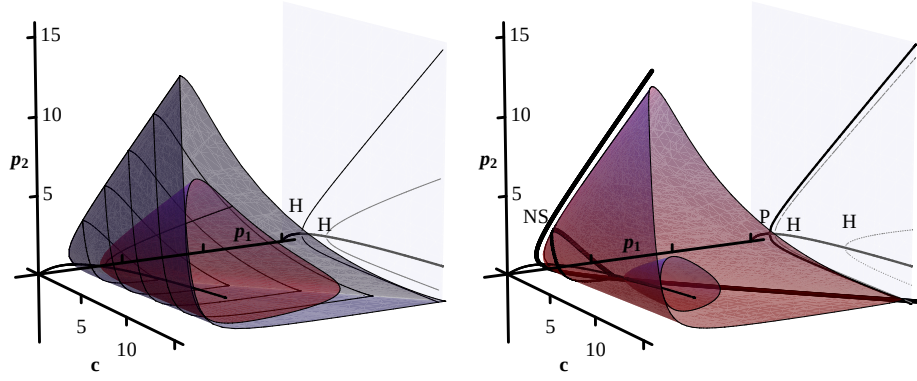


Figure 3.1: **Global bifurcation analysis of the generalized repressilator model** Bifurcation diagrams showing the cascade of Hopf Bifurcations in the generalized repressilator model. Increase of the parameter c (see Eq. 3.2) a cascade of limit cycles emerges, which we show in p_1 vs. p_2 phase plots. In odd rings (exemplified by $N = 9$) an attracting limit cycle (meshed style) emerges from the first Hopf bifurcation (H). In even-rings (exemplified by $N = 10$) a Pitchfork bifurcation (P) leads to two symmetrical stable fixed points (full black lines), which we will refer to as up/down solutions. In even- and odd-rings further Hopf bifurcation lead to emergence of unstable periodic orbits. Some limit cycles in even rings undergo a Neimark-Sacker bifurcations (NS), which is a global bifurcation, leading to an unstable torus.

3.2 Analytical bifurcation analysis of generalized repressilator and unstable periodic orbits

The generalized repressilator model is given by the following ODEs:

$$\begin{aligned}\dot{m}_j &= f(p_{j-1}) - c_2 m_j \\ \dot{p}_j &= c_3 m_j - c_4 p_j,\end{aligned}\tag{3.1}$$

where p_j and m_j describe protein and mRNA concentrations for each gene, and $f(p)$ is a monotonic function acting as a repressor (70). For simulation results and figures throughout the thesis a special case of a Hill function has been used with the Hill coefficient $h = 2$ for the repression function $f(x) = \frac{c_1}{1+x^2}$. However the outlined calculations are similar for any $f : U(\in \mathbb{R}^+) \rightarrow U(\in \mathbb{R}^+)$, which are bounded and monotonically decreasing in this region. The symmetric solution, where all proteins have the same concentration $p_j = p_{j-1} \forall j$ is guaranteed for such $f(x)$. For a Hill type repression function $f(x) = \frac{c_1}{1+x^h}$ a complete stationary solution ($\mathbf{f}(\mathbf{x}) = \mathbf{0}$) can be formulated in terms of a recursion relation for the proteins, which

then determine stationary state of $mRNA$:

$$p_j^* = \frac{c_1 c_3}{c_2 c_4} \cdot \frac{1}{1 + p_{j-1}^{*h}}$$

$$m_j^* = \frac{c_4}{c_3} p_j^*$$

In general, the full solution has to be obtained from nested fractions. For illustration purposes we show the 4 genes ring:

$$p = \frac{c}{1 + \frac{c^2}{\left(1 + \frac{c^2}{\left(1 + \frac{c^2}{\left(1 + p^2\right)^2}\right)^2}\right)^2}}$$

Although in principle it is thinkable that the above equation has several solutions for larger N s, there are only two types of solutions for this symmetric system, which have been found numerically: the symmetric fixed point solution p_m , where all p_j 's have the same concentration $p_j = p_{j+1} \equiv p_m$ and for even rings only the dimerized solutions $p_{u/d}$ where every other gene has the same concentration $p_j^* = p_{j+2}^* \neq p_{j+1}^*, \forall j$. Numerically we have tried to find other positive and real solutions for the system using the symbolic solver in *Mathematica* for rings up to 8 genes. But in agreement with findings in (47, 70), no further solutions have been found numerically or analytically.

The symmetric fixed point solution p_m from the equation

$$p_m + p_m^{h+1} = c \quad (3.2)$$

exists for both even and odd rings ($c \equiv \frac{c_1 c_3}{c_2 c_4}$ as before). It is stable for smaller c 's, but eventually becomes unstable in both even and odd rings. In odd rings as already discussed earlier via a Hopf Bifurcation (see section 2.2.2).

In the case of even rings, a pitchfork bifurcation occur at $c = 2$ for all N and this condition can be derived by requiring that $p_j^* = p_{j+2}^* \neq p_{j+1}^*, \forall j$, which essentially means tiling a larger even ring into gene pairs, where each pair is equivalent to a two gene ring.

$$p_1(1 + p_2^h) = c \quad \& \quad p_2(1 + p_1^h) = c$$

Positive and real solutions satisfying the two equations are given either by ($p_1 = p_u, p_2 = p_d$) or the symmetric alternative ($p_1 = p_d, p_2 = p_u$), where $p_u \approx c$ and $p_d \approx 0$. These condition lead to dimerized solutions, where if one gene is expressed the following gene in the ring will be suppressed or vice versa.

For the specific case of the Hill coefficient $h = 2$, which is practically mostly relevant (2), the explicit solutions can be given by the following:

$$\begin{aligned} p_j^* &= \frac{c}{2} + \sqrt{\frac{c^2}{4} - 1} \equiv p_u \\ p_{j+1}^* &= \frac{c}{2} - \sqrt{\frac{c^2}{4} - 1} \equiv p_d. \end{aligned} \quad (3.3)$$

Note that $p_u \rightarrow c$ and $p_d \rightarrow 0$ for large c . The emerged fixed points for larger c 's correspond to two distinct dimerized states: one in which genes with odd indices are upregulated (p_u) while genes with even indices are downregulated (p_d); and another symmetric state where the genes with odd and even indices exchange their patterns of regulation. These solutions are equivalent to tiling the ring with $N/2$ repeated copies of the up-down solution of the two-gene ring. Their structure is similar to that of other dimerized degenerate solutions in classic models of conjugated polymers and spin chains (76). Therefore, after the bifurcation, the system is bistable, i.e, it behaves like a switch.

Linearization around switch-like stationary states

We assess the stability of the fixed points $p_{u/d}$ and show that they are stable after the pitchfork bifurcation. Because of the symmetry we demonstrate only the first branch, where $p_1^* = p_{2j+1}^* = p_u$ and $p_2^* = p_{2j}^* = p_d$ (first gene and odd numbered genes are up-regulated, second gene and all the even numbered genes are down-regulated). For convenience we define: $\partial_{p_i \frac{c_1}{1+p_i^h}} \Big|_{p_{u/d}} = -\frac{hp_{u/d}^{h-1}c_1}{(1+p_{u/d}^h)^2} \equiv u/d$ with $p_{u/d}$ as previously defined. First we eliminate the mRNA equations taking the derivative of the protein equations with respect to time and substituting from the mRNA equations:

$$\ddot{p}_j = c_3 f(p_{j-1}) - (c_2 + c_4)\dot{p}_j - c_2 c_4 p_j. \quad (3.4)$$

To obtain a second order differential equation for protein concentrations, which can be rewritten into a system of first order differential equations:

$$\begin{aligned}\dot{p}_j &= q_j \\ \dot{q}_j &= c_3 f(p_{j-1}) - (c_2 + c_4)q_j - c_2 c_4 p_j,\end{aligned}\tag{3.5}$$

The system can then be written separating the linear and the non-linear parts as:

$$\dot{\mathbf{z}} = (I \otimes A)\mathbf{z} + H(\mathbf{z})\tag{3.6}$$

where

$$A = \begin{pmatrix} 0 & 1 \\ -c_2 c_4 & -(c_2 + c_4) \end{pmatrix}, \mathbf{z} = \begin{pmatrix} \vdots \\ p_j \\ q_j \\ \vdots \end{pmatrix}$$

and $H(\mathbf{z})$ containing the non-linear coupling terms between the elements. Linearizing around the symmetrical fixed point p_m , we obtain the Jacobian, which contains the typical topological ring feature namely the circulant matrix

$$G = \begin{pmatrix} 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \\ . & . & . & . & . \\ 0 & \dots & 1 & 0 & 0 \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

For a convenient notation we introduce a symmetry breaking function f_{sym} , which evaluates the genes with odd index at p_u and the the ones with the even index at p_d . Then the Jacobian evaluated at the up/down solution is the following:

$$J|_{p_{u/d}} = (I \otimes A) + (G \otimes H_{p_{u/d}})\tag{3.7}$$

where $H_{p_{u/d}} = \begin{pmatrix} 0 & 0 \\ f'(f_{sym}(p)) & 0 \end{pmatrix}$.

The characteristic polynomial χ factorizes and, taking into account that the eigenvalues of the circulant matrix are $\gamma_j = e^{i\frac{2\pi j}{N}}$ with $j \in \{0, \dots, N-1\}$, we obtain

the following complex quadratic equation for the eigenvalues:

$$\begin{aligned}
 (c_2 + \lambda)(c_4 + \lambda) &= c_3 \sqrt{ud} \gamma_j \\
 \lambda_j &= -\frac{c_2 + c_4}{2} \pm \sqrt{\frac{1}{4}(c_2 - c_4)^2 + \sqrt{ud}c_3 \gamma_j} \\
 &= -\frac{c_2 + c_4}{2} \pm \sqrt{\underbrace{\frac{1}{4}(c_2 - c_4)^2 + \sqrt{ud}c_3 \cos(\frac{2\pi j}{N})}_x + i \underbrace{\sqrt{ud}c_3 \sin(\frac{2\pi j}{N})}_y}
 \end{aligned}$$

with $j \in \{0, 1, \dots, N-1\}$.

In the following we show that the fixed points are stable after the pitchfork bifurcation for the Hill coefficient $h = 2$. The line of arguments is similar for higher coefficients. The fixed points are stable because the real negative contributions to the Eigenvalues are always larger than the positive ones, so the inequality to be shown is $\frac{c_2+c_4}{2} > \text{Re}\{\sqrt{\frac{1}{4}(c_2 - c_4)^2 + \sqrt{ud}c_3 \gamma_j}\}$:

- $\sqrt{ud}c_3 = \frac{2c_1c_3}{c^2} = \frac{2c_2c_4}{c}$ following directly from definitions for c, u, d, p_u, p_d with simple algebra
- using the standard formula for real and imaginary parts separation under the complex square root $\sqrt{x + iy} = \sqrt{\frac{1}{2}(\sqrt{x^2 + y^2} + x)} + i \frac{y}{\sqrt{2(\sqrt{x^2 + y^2} + x)}}$
after some algebra it can be shown that $\text{Re}\{\sqrt{\frac{1}{4}(c_2 - c_4)^2 + \sqrt{ud}c_3 \text{Exp}(\frac{2\pi ik}{N})}\} > \sqrt{\frac{1}{4}(c_2 - c_4)^2 + \sqrt{ud}c_3}$
- leading to inequality to be shown $\frac{c_2+c_4}{2} > \sqrt{\frac{1}{4}(c_2 - c_4)^2 + \frac{2c_2c_4}{c}}$
- for $c = 2$ both sides of the inequality would be equal
- $c > 2$ would diminish the positive contribution to the monotonically increasing root function and therefore $\frac{c_2+c_4}{2} > \text{Re}\{\sqrt{\frac{1}{4}(c_2 - c_4)^2 + \frac{2c_2c_4}{c} \text{Exp}(\frac{2\pi ik}{N})}\}$

Confirming the numerical observation of a classical supercritical pitchfork bifurcation with c being the critical parameter is shown analytically in agreement with Müller et al. (47).

3.2.1 Analytical conditions for the appearance of unstable periodic orbits

In the previous section we have assessed the stability of the symmetry breaking $p_{u/d}$ branches, which appear in even rings. Now we will linearize around the symmetric solution p_m , which occurs in both even and odd rings and derive analytical conditions for the emergence of unstable periodic orbits in the bifurcation cascades as observed numerically in section 2.4.1.

$$J|_{p_m} = (I \otimes A) + (G \otimes H^*) \quad (3.8)$$

where $H^* = \begin{pmatrix} 0 & 0 \\ f'(p_m) & 0 \end{pmatrix}$. The characteristic polynomial χ factorizes and, taking into account that the eigenvalues of the circulant matrix are $\gamma_j = e^{i\frac{2\pi j}{N}}$ with $j \in \{0, \dots, N-1\}$, we obtain

$$\chi(\lambda, j) = \det(A + \gamma_j H^* - \lambda I) \quad (3.9)$$

We therefore obtain the following complex quadratic equation for the eigenvalues

$$\lambda_l^2 + (c_2 + c_4)\lambda_l + c_2c_4(1 + \alpha \exp(i\beta_l(N))) = 0 \quad (3.10)$$

$\beta_l(N) = \frac{2\pi l}{N}$, $l = \{0, \dots, N-1\}$, $\alpha = \frac{hp_m^{h+1}}{c}$. The necessary conditions for a Hopf bifurcation are purely imaginary roots. After some algebra (see next section), it turns out that for the most interesting case in synthetic biology (2), where repressor proteins form dimers and Hill coefficient h is 2, odd-rings possess $N_{HB} = \left\lfloor \frac{1}{2} \frac{(N-1)}{2} \right\rfloor$ Hopf bifurcations and even-rings possess $N_{HB} = \left\lfloor \frac{1}{2} \left(\frac{N}{2} - 1 \right) \right\rfloor$.

There are in principle $N_{HB} = (N-1)/2$ (odd-rings) and $N_{HB} = (N/2 - 1)$ (even-rings) possible solutions, predicted by standard group theory (52) for strong enough couplings in rings. For the set of constants and suppression $h = 2$ we do not observe all possible solutions with the discrete rotation symmetry $\tilde{\mathbf{Z}}_k$, however more solutions are numerically observed for stronger repression, for instance for larger Hill coefficient $h = 4$ also the 4-gene ring shows a Hopf bifurcation. The number of unstable periodic orbits is therefore dependent on the parameter set and the repression function.

Analytical Conditions for the Number of Hopf bifurcations

Non-linearity conditions and the length of the ring N will determine how many Hopf bifurcations and which unstable periodic orbits are expected for a parameter set $\{c_1, c_2, c_3, c_4\}$ of the generalized repressilator model. We will go through the case with the Hill repressor function $f(p_m) = \frac{c_1}{1+p_m^h}$, where h is the Hill coefficient.

The characteristic polynomial given in the main text is

$$\lambda_l^2 + (c_2 + c_4)\lambda_l + c_2c_4(1 + \alpha \exp(i\beta_l(N))) = 0 \quad (3.11)$$

$\beta_l(N) = \frac{2\pi l}{N}, l = \{0, \dots, N-1\}$, $\alpha = \frac{hp_m^{h+1}}{c}$. The necessary condition for the Hopf bifurcation to occur is purely imaginary eigenvalues of the Jacobian. For a complex quadratic polynomial $p(z)$

$$(a + iA)z^2 + 2(b + iB)z + (c + iC) = 0 \quad (3.12)$$

the necessary condition for the purely imaginary root is (95)

$$C^2 - 4bBC - 4b^2c = 0. \quad (3.13)$$

In our case $A = B = 0$, $C = c_2c_4\alpha \sin(\beta_l(N))$, $c = c_2c_4(1 + \alpha \cos(\beta_l(N)))$ and $b = -(c_2 + c_4)/2$, which leads to the condition:

$$\alpha^2 c_2 c_4 \sin^2(\beta_l(N)) - (c_2 + c_4)^2 (1 + \alpha \cos(\beta_l(N))) = 0 \quad (3.14)$$

This equation determines, which $\beta_l(N)$'s can satisfy the condition for Hopf bifurcation for a particular parameter set. From the fixed point equation $p_m + p_m^{h+1} = c$, the inequality $0 \leq \alpha < h$ follows trivially. Biologically relevant parameters are given close to the following set: $c_1 \in [0, 3]$ is the bifurcation parameter, and $c_2 = 0.12 \text{ min}^{-1}$, $c_3 = 0.16 \text{ min}^{-1}$, $c_4 = 0.06 \text{ min}^{-1}$ for linear terms. Then the set $h = 2$ and $\beta_l(N) \in]\frac{3}{2}\pi, 2\pi[$ leads to solutions starting from $N = 6$. For the same cs and $h = 4$, the condition is already fulfilled for 4 gene ring and for $l = 1$ a Hopf bifurcation and an unstable periodic orbit are observed, as expected from the above inequality and Eq. 3.14.

Although unstable periodic orbits do not lead to long term attractors, they

can still be important for practical applications if they are likely to be reached from random initial conditions (or as a result of random fluctuations) and they dictate long-lived dynamics that will therefore feature in finite-time observations. As shown below, the unstable periodic orbits that emerge from the first Hopf bifurcation in even rings have these characteristics for parameter ranges consistent with biological experiments. In particular, with mRNA and protein life-times on the order of minutes and Hill coefficients $h \in \{2, \dots, 6\}$, the transient oscillatory behavior would extend over several hours in rings of moderate size $N \sim 10$ when the cellular doubling time is on the order of one hour. Therefore the transient behavior dominates the observable dynamics.

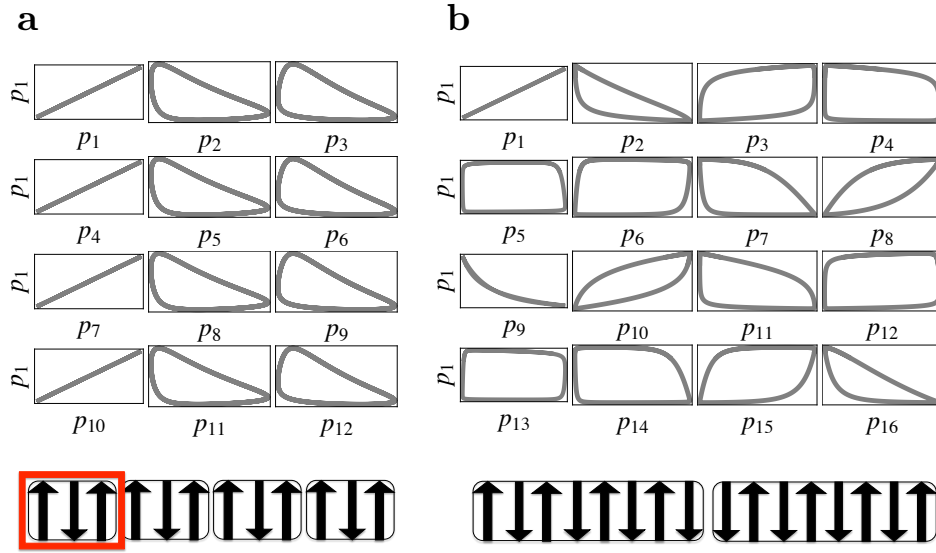


Figure 3.2: Spatio-temporal symmetries in the generalized repressilator model. (a) Phase plots for the symmetric unstable periodic orbit originating from the second Hopf bifurcation in the 12-gene ring. We plot the time evolution of the first gene against other genes in the ring. Genes that are dynamically equivalently are plotted column-wise. Each of the four rows shows a group of genes, which behave dynamically in the same way as the 3-gene ring. The solution has four kinks corresponding to the domain walls between the 3-gene tiles, as shown in the bottom diagram (the first tiling corresponding to the classical repressilator is framed in red). (b) Phase plots for the non-symmetric unstable periodic orbit originating from the first Hopf bifurcation in the 16-gene ring. In this case there is no obvious symmetry between the different genes. As shown in the bottom diagram, this a basic solution that contains two kinks.

3.3 Characteristics of unstable periodic orbits

We now explore the different dynamical facets of the emerging unstable periodic orbits. In particular, we describe their spatio-temporal symmetry, and their relationship to tilings of the ring, and present their linear stability. We also show that these solutions can be understood in terms of traveling waves that correspond to propagating interacting kinks, analogous to solutions from classical physics such as domain walls. We also show that, although unstable, these periodic solutions are reachable from random initial conditions and long-lived and, as such, they are relevant in finite-time dynamics in real-world applications.

3.4 Spatio-temporal symmetries of unstable periodic orbits of the generalized repressilator model

The discrete rotational symmetry of the equations (3.1) induces non-trivial symmetries in the unstable periodic orbits which can be directly related to the linear stability of the solutions. Fig. 3.3 presents a summary of the bifurcation analysis and emerging solutions for $N \in [3, 18]$ carried out with the continuation software MATCONT (96), which was done in addition to the continuation with AUTO shown in the previous chapter. The analysis reveals that some of the higher order bifurcations for larger rings have the same periods as lower order bifurcations in smaller rings. This is the result of a rotational symmetry of such solutions that correspond to tilings of the rings, i.e., gene p_j has exactly the same time behavior as gene p_{j+s} . This can be seen, for instance, in the unstable periodic solution that emerges from the second Hopf bifurcation in the $N = 12$ gene ring, where p_j, p_{j+3}, p_{j+6} and p_{j+9} have the same time traces (see Fig. 3.2 a). Each triplet of consecutive elements follows a dynamics equivalent to that of a 3-gene ring, i.e., the solution corresponds to the tiling of the 12-gene ring by four 3-gene repressilators. Therefore, its period is the same as that of the 3-gene ring for the same set of constants.

Figure 3.3 summarizes the Hopf bifurcations and some of the properties of the ensuing periodic solutions (limit cycles and unstable periodic orbits) obtained through numerical continuation and from the analysis of numerical simulations. We have considered rings of length between $N = 3$ and 18 of the model with

N	Unstable/all Floquet directions	Period	kinks	basic solution
11	4/22	86	5	11(5)
18	7/36	89	8	18(8)
7	2/14	94	3	7(3)
14	5/28	94	6	7(3)
17	6/34	100	7	17(7)
10	3/20	104	4	10(4)
13	4/26	110	5	13(5)
16	5/32	113	6	16(6)
3	0/6	132	1	3(1)
6	1/12	132	2	3(1)
9	2/18	132	3	3(1)
12 ^b	3/24	132	4	3(1)
15	4/30	132	5	3(1)
18	5/36	132	6	3(1)
17	4/34	154	5	17(5)
14	3/28	160	4	14(4)
11	2/22	169	3	11(3)
8	1/16	186	2	8(2)
16	3/32	186	4	8(2)
13	2/26	204	3	13(3)
18	3/36	213	4	18(4)
5	0/10	239	1	5(1)
10	1/20	239	2	5(1)
15	2/30	239	3	5(1)
17	2/34	274	3	17(3)
12	1/24	291	2	12(2)
7	0/14	342	1	7(1)
14	1/28	342	2	7(1)
16 ^c	1/32	393	2	16(2)
9	0/18	444	1	9(1)
18	1/36	444	2	9(1)
11	0/22	545	1	11(1)
13	0/26	645	1	13(1)
15	0/30	746	1	15(1)
17	0/34	847	1	17(1)

Figure 3.3: **How symmetries and kinks are related to bifurcations in the generalized repressilator model.** The analytical conditions following Eq. 3.10 for the emergence of unstable periodic orbits are confirmed by numerical analysis (MATCONT) for rings of the lengths $N=[3-18]$ and are summarized above. We have followed the unstable periodic orbits numerically calculating their Floquet exponents (col. 2), period (col. 3), and shapes. The number of kinks in each solution has been numerically extracted from the shapes using the Eq. 3.15 (col. 4). The unstable periodic orbits are sorted by the period and symmetric tiling solutions are highlighted by the same colors. There are solutions in the generalized repressilator model, which can be thought of as being glued together from the solutions of shorter rings. For example, the 12-genes unstable periodic orbit (marked with b in the table) can be thought of as being glued together from 4 basic solutions with 3-genes (see Fig. 3.2a). The basic solutions are given in col. 5 and the number of kinks in the basic solution is shown in parenthesis below.

Hill-type repression with $h = 2$ for a range of the bifurcation parameter $c_1 \in (0, 3]$. The observed bifurcations follows the number and pattern of Hopf bifurcations

previously found analytically (see section 3.2.1). The solutions are sorted by increasing period and therefore symmetric tiling solutions appear grouped together (since they have the same period). Not all the solutions correspond to tilings of the ring. Some of the unstable periodic orbits are ‘basic solutions’ with no strict rotational symmetry, i.e., no two genes behave in the same way (Figure 3.2b). The fact that some solutions of the system can be understood in terms of tilings of basic solutions indicates the importance of domain walls and kink-like excitations in the understanding of the observed solutions. We explore this alternative dynamical picture below.

3.4.1 Unstable periodic orbits as traveling kink solutions

An alternative viewpoint to understand the dynamics of gene rings is to view the periodic solutions in terms of propagating kinks against the backdrop of a dimerized (‘up-down’) solution. Analogues to this picture are abundant in both the nonlinear waves literature and in classical examples in condensed matter physics, such as propagation of domain walls in antiferromagnetic chains or soliton-like excitations in dimerized polymer chains (50, 97, 98).

The link between the dynamical bifurcations presented above and the kink-domain picture follows straightforwardly from the examination of the fixed points of the system. As explained above, the uniform solution, where all proteins have the same value $p_j = p_m, \forall j$, is always a fixed point of the system and is the global attractor of the equations for small enough values of the coupling parameter c .

When the coupling between the genes is large enough, the uniform solution becomes unstable either through a Hopf bifurcation (for odd rings) or through a pitchfork bifurcation (for even rings). The two stable fixed points that emerge from the pitchfork bifurcation in even rings correspond to dimerized solutions (‘up-down’) where genes are over-expressed ($p_u > p_m$) or under-expressed ($p_d < p_m$) in an alternating sequence around the ring. This arrangement leads to two degenerate solutions, equivalent by a one-gene rotation, which correspond to the sequences ‘up-down’ and ‘down-up’, respectively. Clearly, such a dimerized solution is compatible with an even ring but not with an odd ring.

In the case of an odd ring, trying to fit an ‘up-down’ solution leads to frustration due to the boundary conditions, originating a *kink* mismatch when wrapping up

3.4 Spatio-temporal symmetries of unstable periodic orbits of the generalized repressilator model 94

the sequence. Indeed, the first Hopf bifurcation in odd rings can be viewed as the emergence of a propagating solution of this type. Hence the attracting limit cycles in odd rings correspond to the fundamental solution with one traveling kink propagating against a background of the dimerized solution. The localized structure of the kink (domain wall) is shown in Figure 3.4a for a ring of size $N = 51$. A quantitative description of the dynamics of the kinks can be obtained from numerical simulations by defining the threshold function:

$$v(p_j) = \begin{cases} 1 & \text{if } p_j \geq p_m \quad \equiv \text{'up'} \\ -1 & \text{if } p_j < p_m \quad \equiv \text{'down'}, \end{cases} \quad (3.15)$$

which allows us to define the domain regions as sequences of alternating genes. The location of the kinks is then defined by the condition $v(p_j) = v(p_{j+1})$.

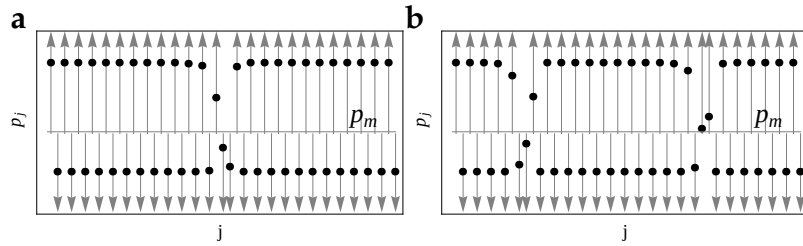


Figure 3.4: Kinks in the repressilator model. The kink perspective is useful to characterize periodic orbits. Numerically observed time snapshots in gene space (black dots) are mapped into the qualitative kink using function $v(p_j)$ Eq. 3.15 (gray arrows). **(a)** Frustrated solution in odd-numbered ($N=51$) gene ring, which emerges from the first Hopf bifurcation and is a stable attractive limit cycle. With kink concept the stability is intuitively clear: the kink can propagate through the ring forever. **(b)** A pair of kinks in the even-numbered rings ($N=50$) is on the other hand unstable. This solution originates from the first Hopf bifurcation in an even-ring. The dynamics is the movement of the two kinks along the rings with a slightly different speed. As soon as one kink gains on the other, they annihilate and the ring becomes quiescent. This happens only after a very large number of periods ($O(100)$).

The cascade of Hopf bifurcations corresponds to the emergence of solutions with an increasing number of kinks in the ring. The first Hopf bifurcation in *even* rings corresponds to the emergence of a solution with two traveling kinks which can be seen as localized excitations in the ring, as seen in Figure 3.4b for a ring of size $N = 50$. The kinks (or walls) originate from the mis-matching of two 'up-down' domains to give an arrangement that can be fitted into an even ring, as shown at the bottom of Figure 3.2b. As indicated by Floquet analysis, this

solution is an unstable periodic orbit in which the two kinks get closer as time evolves until they annihilate each other. As the two kinks conflate, the domain separation unravels and we recover the 'up-down' dimerized solution. Similarly, the second Hopf in *odd* rings gives rise to unstable periodic orbits that correspond to solutions containing three traveling kinks, while the second Hopf in *even* rings leads to the emergence of unstable periodic orbits with four propagating kinks in the ring.

In even-numbered rings only configurations with even-number of kinks are possible therefore these solutions are always unstable. In odd-rings only odd number of kinks is possible, therefore after a possible unstable periodic orbit with several kinks, all kinks but one will annihilate. This one remaining frustration will propagate through the ring indefinitely, corresponding to a stable limit cycle in odd-rings and making them stable oscillators. Note also that when the solutions correspond to tilings of basic solutions, the total number of kinks is the result of multiplying the number of kinks of the basic solution by the number of tilings. Moreover, the number of Hopf bifurcations, which follows from symmetry constraints, is therefore linked to the maximum number of kinks that can be accommodated in the rings. A summary of the periodic solutions in Figure 3.3 is presented in terms of their kink density in Figure 3.5.

3.5 Kink dynamics

The unstable periodic orbits originating from the first Hopf in even-rings have a strong influence on their dynamics, even though even-rings act as switches in the steady state. Often if starting from an initial condition even-rings get trapped into a long-lived oscillating transient before they eventually approach the stable up/down solution. The observed transients can be assigned to first Hopf bifurcations.

A typical transient will quickly settle down into a configuration with two kinks separated by a domain, which length will depend on initial conditions (see Fig. 3.6a). Then the two kinks propagate through the ring for a large number of periods, slowly approaching each other.

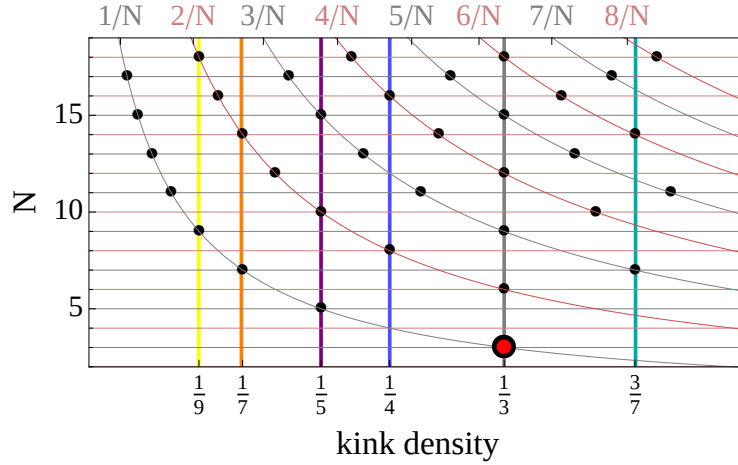


Figure 3.5: Tilings and kink density of UPOs of the generalized repressilator model. Summary of the solutions presented in Fig. 3.3 using the same color scheme. According to a numerical observation for this set of parameters the kink density is always smaller than $1/2$. Solutions with the same number of kinks are connected by the grey (odd) and pink (even) curves. The curves correspond to the decaying functions $\frac{i}{N}$ as indicated on the top of the figure. The kink density for the first Hopf bifurcation in an odd ring behaves as $\frac{1}{N}$ if plotted against N , since stable limit cycle solutions always have one kink. The kink density of the first Hopf bifurcation in even rings behaves as $\frac{2}{N}$, since quasi-stable solutions always have two kinks, and so on. The solutions of odd-numbered rings appear at the crossings of the grey horizontal lines with the grey curves. Similarly, the intersections of the pink horizontal lines and the pink curves mark the UPOs for even rings. The colored vertical lines connect the families of symmetry tilings (colors matched to those in Fig. 3.3). The kink density of the symmetry solutions is obviously the same.

Slow convergence to quiescent steady state as kink annihilation event

We use the transformation Eq. 3.15 to locate the initial position of the two kinks in the ring 3.6b). At the beginning of the simulation the kinks are separated by a domain, which is five genes long. As the time passes the domain separating the kinks decreases due to slow drift. This drift is also observable in the phase plot perspective 3.6b) and in the torus phase perspective 3.6d).

These numerics suggest that length of the oscillating transient will strongly depend on the initial separation of the moving perturbations. Which also means that longer rings, where the kinks can be further separated can support longer transients.

3.6 Probabilities for long oscillating transients and distribution of their lengths

For applications and element design the emergence probability for the oscillating transients is of interest as well as the distribution shape for the length of oscillating states. For even-rings starting from random initial conditions the probability to end up in an oscillating transient corresponding to the unstable periodic orbit originating from the first Hopf bifurcation converges to 50% as has been shown in the previous chapter (see Fig. 2.9b).

In even-rings all numerically observed transients leading to oscillating states, which duration is larger than five periods, correspond to unstable periodic orbit emerged from first Hopf Bifurcation. Unstable periodic orbits with more than 2 kinks are possible in even-rings as shown by the bifurcation analysis, however starting from random initial conditions they are less likely combinatorially than the ones with two kinks. Even more unlikely is a configuration where the kinks will be widely enough separated to produce a longer oscillating transient. Such solutions leading to longer transient than 5 oscillations have not been observed numerically for our biologically inspired constants set. Even though all initial conditions leading to oscillating states can be assigned to the first Hopf Bifurcation, their oscillation duration is different. As has been indicated in the previous section the distance between the kinks is the essential characteristic determining the length of the transient.

In even-rings the duration of the oscillating transient depends on the kink distance k_d

In the previous section we have shown that the transients are reachable from random initial conditions. In this section we explain the transient duration distribution starting from random initial conditions. As an example we use $N = 16$ to show possible transients configurations, which we map to the qualitative dynamics (Eq. 3.15), and explain the oscillation length distribution based on the kink concept.

The pie-chart and the kink configuration table (see Fig. 3.7a) show the numerically observed distribution of kink configurations with kink distances k_d , to

which the system settles down after the initial "burn-in" of five periods of oscillation. Depending on the nonlinearity there will be a minimal size of the domain, which if separating two kinks would lead to an observable unstable periodic orbit. For our example, $N = 16$ -gene-ring with biologically inspired constants set, a minimal kink distance k_d is three. Therefore, there will be on total four different kink configurations (see Fig. 3.7a), to which the system settles down if starting from arbitrary initial conditions. Around each kink configuration we observe a distribution of oscillating transient lengths T_{tr} . The mean $\langle T_{tr} \rangle$ of the distribution correlates with the kink distance (see Fig. 3.7a).

The duration distribution of the oscillating transients has a heavy tail

As expected configurations with different k_d s, $k_d = 4, 5, 6$ are reached with comparable probabilities from random initial states. The states $k_d = 2, 3$ show smaller percentage fraction, because these configurations do not always oscillate for longer than 5 periods and are therefore are partially assigned into the <2 fraction. Hence, according to the numerics the transients with different kink distances are all reachable with comparable probabilities from random initial conditions.

That means that very long oscillating transients are reachable from a non-vanishing set of initial conditions and therefore we obtain a broad oscillating transient distribution with a heavy tail (see Fig. 3.7b). The y-Axis is chosen to be logarithmic showing that the probability to end up in a very long transient >500 oscillations is of the same order of magnitude as the probability to end up in the stable up down solution.

3.7 Comparison to unidirectionally coupled flux-gate magnetometers

There are several systems with a directed ring topology where unstable periodic orbits play a significant role in the dynamics. A recent example is the study of unstable periodic orbits in relation to the emergence of chaotic behavior in rings of Duffing oscillators (53). Closest to the dynamics of the gene system studied here are the rings of unidirectionally coupled magnetic cores (51) that have been not

only studied numerically but also implemented experimentally by Bulsara *et al.* . In the antiferromagnetic case, the system of coupled magnetic cores is described by the following ODEs (51):

$$\dot{x}_j = -x_j + \tanh\left(c \left(x_j - x_{j+1}\right)\right), \quad (3.16)$$

which falls under the same framework as the generalized repressilator if linearizing around the unstable symmetric fixed point.

The bifurcations branching from the unstable symmetric solution follow from a Jacobian of the same form as that in Eq. (3.8). Hence the unstable periodic orbit characteristics follow the same pattern as those of the gene system. For even rings, the system shows symmetry breaking when c is increased giving rise first to the two degenerate dimerized solutions followed by the emergence of unstable periodic orbit from the now unstable symmetric branch. For odd rings, there is also a similar cascade of Hopf bifurcations. The maximal number of Hopf bifurcations due to ring symmetry is $(N/2 - 1)$ for even-numbered rings and $(N - 1)/2$ for odd-numbered rings, as is the case for the generalized repressilator model.

Such rings of magnetic cores have been implemented as real-world devices and the odd-numbered versions are used as sensitive magnetic sensors. Long even-numbered rings have also been built and their dynamics has been explored under noise perturbations and under random initial conditions (6). It turns out that long oscillatory transients feature heavily in long even-numbered rings, even if the steady state is quiescent (6).

3.8 Discussion

In this chapter we have analyzed the dynamics of the generalized repressilator model and have characterized the existence of long-lived oscillatory transients in even rings—systems that would otherwise be expected to behave as bistable switches based on strict stability analysis. These transients correspond to slowly decaying dynamics around particular unstable periodic orbits (UPOs) that emerge as part of a cascade of Hopf bifurcations. The UPOs have spatio-temporal symmetries that can be understood in terms of tilings and propagating kinks, corre-

sponding to the moving walls separating different domains of the stable dimerized solution. In particular, the most probable long-lived observable transients correspond to propagation of kink pairs along the ring. We showed numerically that in the presence of such unstable periodic orbits the dynamics of the system is funneled to its vicinity onto a constrained manifold from which it is difficult to escape, thus leading to observable very long-lived transient oscillations of hundreds of periods when starting from random initial conditions. The likelihood of such behavior can be understood from the fact that when a pair of kinks is well separated, they will propagate along the ring for a long time until they conflate to restore the fixed point solution. This process results in a heavy-tailed distribution of the duration of oscillatory transients.

The indept analytical treatment of the genetic repressilator model revealed that a gene ring with just 4 genes has a parameter set which supports the unstable oscillations. This opens a new perspective on how the control design stabilizing this orbit can be implemented experimentally: for the structure to work we would need a very efficient repressors potentially with a more complicated repression mechanism resulting in an overall repressor with the Hill coefficient 4. For design perspective it would mean that we are trading the more complex repression mechanism for less number of genes, which are involved into the structure.

The generalized repressilator system has been compared to flux-gates magnetometers, which have been implemented as electronic devices and are in commercial use. The main common feature between these devices and the repressilator model is, of course, the ring topology. In both cases nonlinearity should be strong enough, monotonic, and bounded, however its precise functional form is not important for the emergence of unstable periodic orbits. This is because the observed effect is a topological phenomenon and originates from the circular topology. Therefore the linearized versions of both systems around the symmetric unstable fixed point are very similar. The main dynamical feature is the existence of the circular matrix, leading to complex eigenvalues symmetrically ring ordered in the complex plane and therefore leading to necessary conditions for Hopf bifurcations. These ideas are mathematically formally expressed in classical group theoretical terms in Golubitsky et al. (52).

The classified Hopf bifurcations lead to unstable periodic orbits and these solutions are observable starting from random initial conditions and also in the

presence of noise. Unstable periodic orbits are likely to feature in the symmetry breaking systems, if there is a mechanism, by which the separatrix between the multiple stable states can be crossed locally, resulting in a global "twist" or the creation of two kinks. This can either happen by starting from random initial conditions or by addition of moderate noise to dynamics, as has been observed in real-world devices by Lindner and Bulsara (6). Therefore the exploration and classification of transients for real-world applications and especially for time limited and noisy biological systems is worth considering in addition to the steady state analysis.

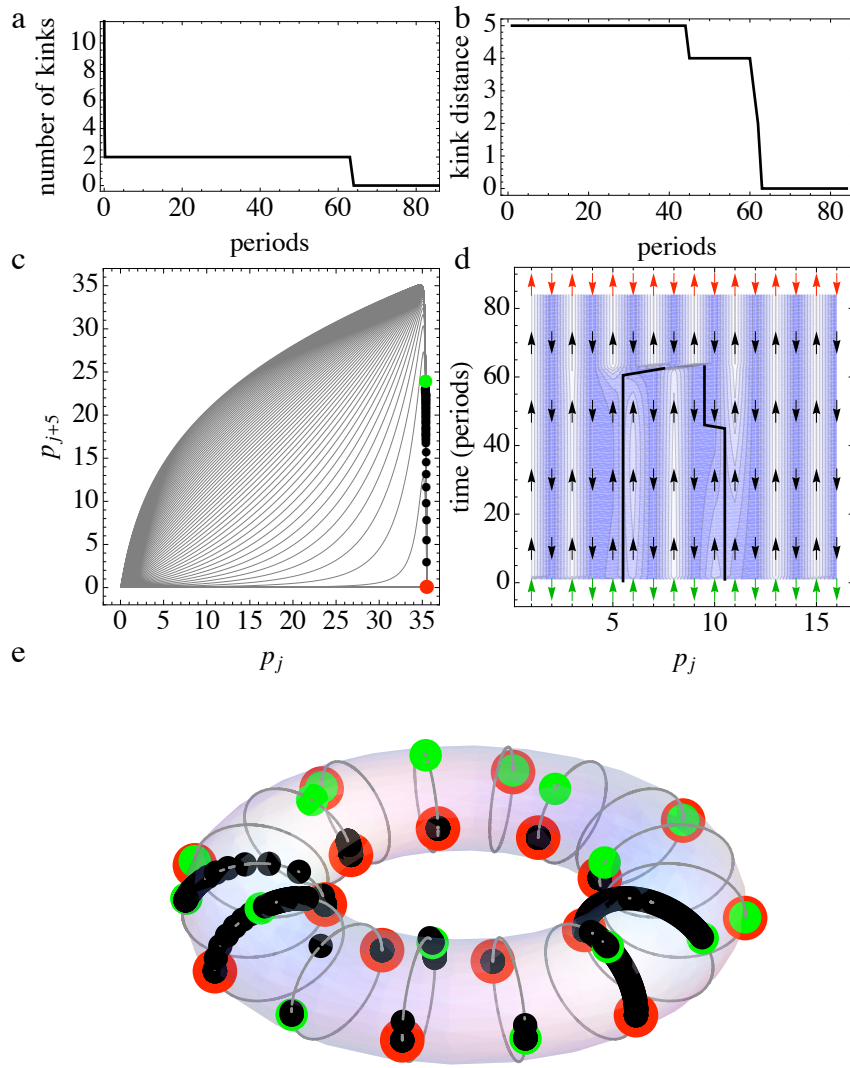


Figure 3.6: Slow kink-kink annihilation leads to long oscillatory transients. Dynamics around the unstable periodic orbit arising from the first Hopf bifurcation in the 16-gene ring to illustrate the long oscillating transient that eventually reaches the stable ‘up-down’ solution. **(a)** Starting from a random initial condition, the ring quickly settles into a configuration with two kinks. **(b)** The distance between the kinks, k_d , goes from $k_d = 5$ to 4 and quickly through 3, 2 until the kink-kink annihilation producing a long oscillatory transient of ~ 80 periods. **(c)** Phase plot (p_1 versus p_5) from the initial condition (light green) to the final point at the stable up-down solution (red). The black dots correspond to Poincaré crossings at equidistant time intervals (every period) and illustrate the very slow initial drift which only accelerates rapidly at the end of the transient, when the unravelling of the solution takes place. Note also the uniform spreading of the trajectories on the surface of the phase plot, an indication of the reduced dimensionality of the transient dynamics. **(d)** Space-time diagram of the transient shows the slow approach of the kinks and their annihilation. The kink traces are indicated by gray lines. **(e)** Poincaré sections in gene space. The value of each protein at the Poincaré section is represented as a phasor $\exp i2\pi\phi_j$ where $\phi_j = p_j/p_{max}$. The kink annihilation event in this representation corresponds to the unravelling of a “twist”, which is a global event involving several ring elements and therefore occurring on very slow time scales.

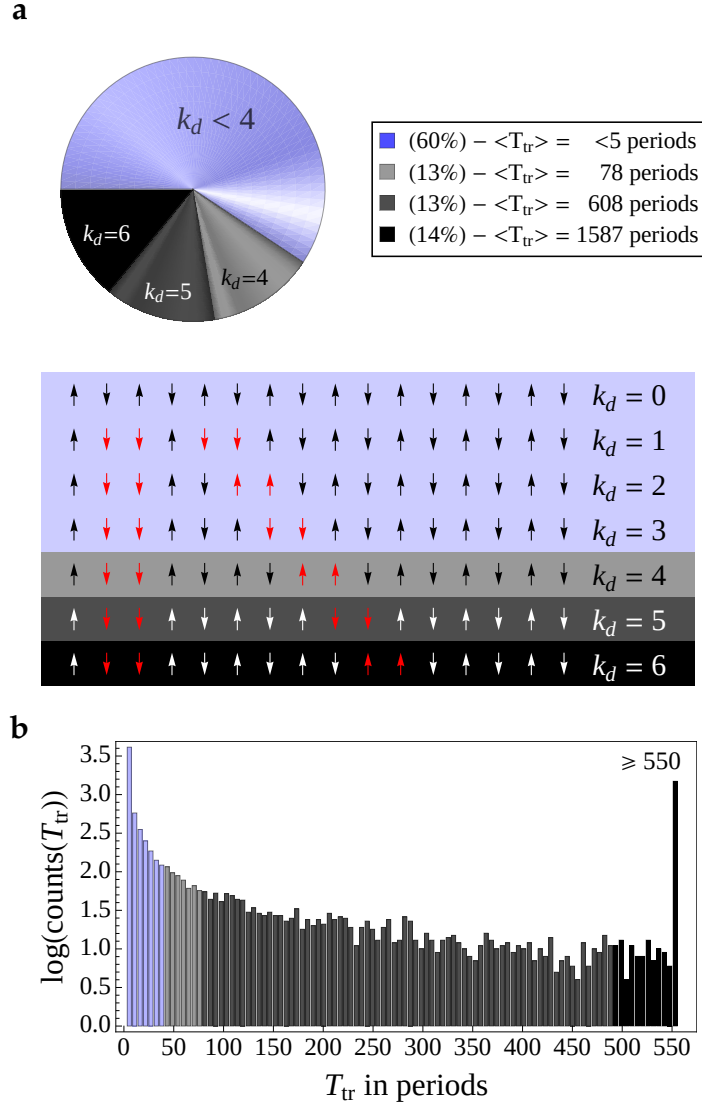


Figure 3.7: Numerical analysis of oscillating transients starting from random initial conditions for 16-gene ring. (a) The pie-chart represents the proportions of observed conformations collected from 10^4 initial conditions covering the hyper cube $([0, m_u] \times [0, p_u])^N$. For each random initial condition, we measure the distance k_d between the kinks after a burn-in time of five periods. Combinatorially we expected that starting from random initial conditions each configuration should be reachable with comparable probabilities and this is confirmed numerically, as seen in the similar proportions of $k_d = 4, 5, 6$. The proportions for the distances $k_d < 4$ are pooled together due to the burn-in time of the simulation. The table also includes the mean of the distribution of numerically observed lengths of the oscillating transients for all the two-kink conformations for the 16-gene ring, as represented below. **(b)** Distribution of the length of the transients showing a heavy tail due to the likelihood of observing the very long transients that ensue from the configurations with inter-kink distances $k_d = 4, 5, 6$.

Chapter 4

Influence of stochasticity on the dynamics of the generalized repressilator

In the previous two chapters Hopf bifurcations resulting in UPOs and stable limit cycles have been shown to introduce non-trivial oscillating behavior in the deterministic description of the generalized repressilator model. But deterministic description is not taking the stochastic fluctuations into account, which occur due to low copy numbers involved in the gene expression process as already mentioned in the introduction section 1.1. In this chapter we will use the Master equation formalism to examine the effects of noise on the oscillations in the generalized repressilator model. The following chapter is based on the manuscript

Onset of oscillations and stochastic resonance in the generalized repressilator model

Natalja Strelkowa and Mauricio Barahona

Intended for the submission to

The Journal of Theoretical Biology

4.1 Introduction

That noise is an important mechanism in naturally occurring and engineered gene expression networks is undoubted (7, 68, 99–101). Naturally occurring noise in

biology is far from being a nuisance on the contrary necessary noise seems to be an evolutionary strategy keeping genetically identical, but phenotypically diverse populations. This strategy has been reported in several microbial populations (68, 102, 103). A well-characterized example is a soil bacterial population *Bacillus subtilis*, which switches between the "dormant" and "competent" (able to take up DNA from the environment) states governed by a stochastic fluctuations of a single protein (68, 99). The origins of noise in genetically identical populations are attributed to the inherent stochastic nature of biochemical reactions (8, 104).

Therefore, noise will also strongly influence synthetically engineered systems, because these networks use the natural transcription machinery and operate in the regime, where copy numbers of the involved molecules are very low. Therefore, we will address the influence of the biochemical noise on the oscillations in an engineered synthetic circuit. Intuitively, we expect noise to excite oscillating behaviour in genetic circuits (5, 69, 105, 106), either effectively stabilizing unstable periodic orbits (5, 69) or even lead to the effect of coherence resonance in the otherwise monostable system (105).

The generalized repressilator model undergoes Hopf bifurcations and it would be interesting to determine a biochemical parameter, which would control the bifurcations, and explore the onset of Hopf bifurcations in the stochastic setting. The onset of Hopf bifurcations has been addressed earlier in the naturally occurring systems attempting to pinpoint and control parameters for the induction of oscillating behavior in cell cycle models. The studied organisms include yeast and *Xenopus* eggs. By varying an external parameter and bifurcation analysis, Marlovits et al. (107) have shown how to induce oscillations in a naturally occurring system of *Xenopus* eggs. For this system a deterministic formalism has been used and the bifurcation parameters and their values have been roughly estimated from a deterministic nonlinear model.

In this chapter, we first derive a master equation for the generalized repressilator model, then establish the dependence of the reaction constants on the system size Ω , so that we can systematically explore the influence of internal noise on the dynamics. We show numerically that the duration of oscillating transients due to unstable periodic orbits (UPOs) is not destroyed by low copy number fluctuations. On the contrary, internal noise effectively prolongs the duration of the oscillating transients. The expectation for this behavior was already indicated

in the section 2.3.4, where small perturbations, simulated with less accurate ODE integration, have not destroyed the quasi-stable oscillations in even rings.

In the following we will explore the effects of noise around the Hopf bifurcations and the onset of stable oscillations. A biochemical parameter which can be used for Hopf bifurcation induction in the generalized repressilator model is the gene copy number. The gene copy number is directly proportional to the bifurcation parameter c_1 used for the deterministic bifurcation analysis in previous chapters. However, in the stochastic system, the onset of stable oscillations is influenced by the noise and in order to explore the effects of noise systematically, we apply the classical Ω expansion as introduced in section 1.3.5.

The linear noise approximation, which is obtained via the Ω -expansion, uncovers an interesting effect known as stochastic resonance which appears at the parameter values around the onset of stable oscillations. The precise definition of the stochastic resonance is disputed, but what we mean in this context is the amplification of microscopic fluctuations leading to coherent oscillations on the macroscopic level (108). Stochastic resonance has been characterized in several systems, for example in a simple chemical equation model – the Brusselator (108), in the SIR (Susceptible-Infected-Recovered hosts) models (109), and also in the predator-prey models (110). We explore here a similar scenario appearing during the emergence of stable limit cycles in the generalized repressilator model.

Finally, we explore how internal noise influences the quality of the oscillations. We show that the quality of oscillation is not only dependent on a bifurcation parameter, but also on the system size Ω . Pinpointing such dependencies is in particular interesting for engineering purposes of synthetic oscillators.

4.2 Michaelis-Menten-like time scale separation in master equations for genetic transcription regulation

For the master equation formalism we would like to derive a stochastic analogue to the classical deterministic Hill-reduction based on the time scale separation. In this way we will obtain a reduced master equation containing only the regulatory protein and *mRNA* dynamics. The quality of the approximation can be calculated analytically for the first two moments and assessed with the stationary coefficient of variation. In addition, the distributions of the reduced and the full master

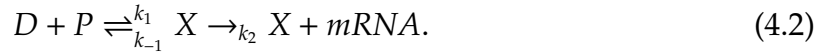
equations will be assessed numerically using the standard Kullback-Leibler divergence:

$$KL = \sum_k P(x_k) \log(P(x_k)/Q(x_k)) \quad (4.1)$$

where P is the distribution of the full model and Q the distribution of the reduced model.

4.2.1 Hill-type variable reduction for activated transcription

We consider the simplified reaction scheme for the activator regulated genetic transcription reaction:



where D is the amount of DNA not occupied by the activator protein P , X is the DNA-activator complex and the conservation equation for the overall DNA present in the system $D_0 = D + X$ holds. Then, following the formalism for enzymatic reactions in (111), the corresponding master equation can be written as:

$$\begin{aligned} \frac{dP(d, p, x, m; t)}{dt} = & k_1(d+1)(p+1)P(d+1, p+1, x-1, m; t) \\ & + k_{-1}(x+1)P(d-1, p-1, x+1, m; t) \\ & + k_2(x+1)P(d-1, p, x+1, m-1; t) \\ & - (k_1dp + k_{-1}x + k_2x)P(d, p, x, m; t) \end{aligned}$$

The dynamics of the complex X happen on the considerably faster time scales compared to the $mRNA$ production, because the binding and unbinding reaction rates of the activator to the DNA are much larger compared to the transcription rate: $k_1, k_{-1} \gg k_2$. We are mainly interested in slow dynamics of the $mRNA$ production and its dependence on the activator dynamics, which also occurs on much slower rate than k_1 and k_{-1} . Therefore, we will decouple fast and slow time scales obtaining a reduced Master equation for the $mRNA$ production, which will only depend on the dynamics of the activator.

We assume that the number of activator proteins is much larger than the num-

ber of DNA copies: $P \gg D_0$, so that the activator dynamics are almost unaffected by the binding/unbinding reactions and can be considered slow. For the slow stochastic variables P and $mRNA$, we assume intermediate values $P = \mathbf{p}^*$, $mRNA = \mathbf{m}^*$ and write the fast dynamics probability distribution for the complex X , $p_n \equiv P(X = n)$:¹

$$\begin{aligned} d_t p_n &= k_1(d_0 - n - 1)\mathbf{p}^* p_{n-1} \\ &+ (k_{-1} + k_2)(n + 1)p_{n+1} \\ &- (k_1(d_0 - n)\mathbf{p}^* + (k_{-1} + k_2)n)p_n \\ &= (\mathbb{E} - 1)r_n p_n + (\mathbb{E}^{-1} - 1)g_n p_n \end{aligned}$$

with $g_n = k_1\mathbf{p}^*(d_0 - n)$ and $r_n = (k_{-1} + k_2)n$. This master equation is a typical stochastic one-step birth-death reaction confined between two absorbing boundaries. The number of available DNA copies d_0 for the enzymatic-like activation reaction is the upper boundary for the number of complexes X , where all DNA copies are occupied by the activator. The state $n = 0$, in which none DNA promoter site is occupied by the activator is the lower boundary l . The state $n = d_0$, in which all DNA promoter sites are occupied by the activator is upper boundary u . Respecting the absorbing boundary conditions $r_l = 0$ and $g_u = 0$, the fast dynamics of the complex X are given by the following set of equations:

$$\begin{aligned} d_t p_l &= r_{l+1}p_{l+1} - g_l p_l \\ d_t p_n &= (\mathbb{E} - 1)r_n p_n + (\mathbb{E}^{-1} - 1)g_n p_n, \quad l < n < u \\ d_t p_u &= g_{u-1}p_{u-1} - r_u p_u \end{aligned} \tag{4.3}$$

We can use the above master equation for stochastic dynamics of the complex in order to derive ODEs for the time evolution of moments (20):

$$\langle n^j \rangle \equiv \sum_n n^j p_n.$$

Analytically we will solve the ODEs for the first and the second moment, find their steady state values and the time scale on which the steady states are approached.

¹Note: $\mathbf{p}^*$ and $\mathbf{m}^*$ are stochastic variables, not mean values, which change slowly compared to the dynamics of the complex X . We assume that the time scale to reach the steady state of X is so fast, that the slow stochastic variables do not change during this period.

The dynamics of the random variable X will then be approximated by its steady state mean value x_{st} . The justification for the approximation is that the time scales on which the variable is fluctuating are fast compared to *mRNA* dynamics. The steady state time average of X is then approximately equivalent to its steady state mean value.

The quality of the approximation can be assessed by showing that the fluctuations around the steady state mean are small. We will calculate the steady state coefficient of variation and determine the conditions, by which the approximation is valid. By this analytical calculation we consider only the first two moments: the mean and the variance. The contributions from higher moments will then be explored numerically in section 4.2.4.

The moment equation for the first and second moment of the master equation 4.3 can be solved analytically following the formalism by van Kampen (20), Goel and Richter-Dyn (111):

$$\begin{aligned} d_t \langle n \rangle &= \sum n(\mathbb{E} - 1)r_n p_n + n(\mathbb{E}^{-1} - 1)g_n p_n \\ &= \sum r_n p_n (\mathbb{E}^{-1} - 1)n + g_n p_n (\mathbb{E} - 1)n \\ &= -\langle r_n \rangle + \langle g_n \rangle \end{aligned}$$

Using the definitions from the above for r_n and g_n we obtain a first order ODE for the mean of the complex X

$$d_t \langle n \rangle + (k_1 \mathbf{p}^* + (k_{-1} + k_2)) \langle n \rangle = d_0 k_1 \mathbf{p}^*$$

The equation is solved analytically by time integration leading to the solution

$$\langle n \rangle = \frac{d_0 \mathbf{p}^* k_1}{\mathbf{p}^* k_1 + (k_{-1} + k_2)} (1 - e^{-(k_1 \mathbf{p}^* + (k_{-1} + k_2))t}) - IC * e^{-(k_1 \mathbf{p}^* + (k_{-1} + k_2))t}$$

with IC denoting initial condition. The intermediate complex X will approach the steady state exponentially fast on the time scale, which is proportional to the fast

reaction variables $\propto (k_1 \mathbf{p}^* + (k_{-1} + k_2))$:

$$\lim_{t \rightarrow \infty} \langle n \rangle = \frac{d_0 \mathbf{p}^* k_1}{\mathbf{p}^* k_1 + (k_{-1} + k_2)} = \frac{d_0 \mathbf{p}^*}{\mathbf{p}^* + K_m}$$

That means that the relaxation time $1/[k_1 \mathbf{p}^* + k_{-1} + k_2]$ is short for the mean of the complex, since it is inversely proportional to the fast reaction rates k_1 and k_{-1} .

We now substitute the steady state mean value x_{st} into the Master equation for the *mRNA* production:

$$d_t p_m \approx (\mathbb{E} - 1) k_2 x_{st} p_m = \frac{d_0 k_2 \mathbf{p}^*}{K_m + \mathbf{p}^*} (\mathbb{E} - 1) p_m \quad (4.4)$$

in correspondence to the deterministic case.

In addition we ensure that the second moment also quickly converges to its equilibrium and that the stationary fluctuations are small compared to the mean. Therefore we will calculate the stationary coefficient of variation.

The time evolution of the variance $\langle (n - \langle n \rangle)^2 \rangle = \langle n^2 \rangle - \langle n \rangle^2$ can also be addressed analytically and we obtain with a similar calculation:

$$\begin{aligned} \text{var}(n) = & (IC - k_1 \mathbf{p}^* d_0) \frac{(2\mathbf{p}^* k_1 d_0 + k_{-1} + k_2 - k_1 \mathbf{p}^*)}{(k_{-1} + k_2 + k_1 \mathbf{p}^*)} e^{-(k_{-1} + k_2 + k_1 \mathbf{p}^*)t} (1 - e^{-(k_{-1} + k_2 + k_1 \mathbf{p}^*)t}) \\ & + \frac{k_1 \mathbf{p}^* d_0 (k_{-1} + k_2)}{(k_{-1} + k_2 + k_1 \mathbf{p}^*)^2} (1 - e^{-2(k_{-1} + k_2 + k_1 \mathbf{p}^*)t}) + IC_{\text{var}} e^{-2(k_{-1} + k_2 + k_1 \mathbf{p}^*)t} \end{aligned}$$

The time scales for the steady state approach are of the same order of magnitude as for the mean, which means that the fluctuations reach the stationary solution fast. The stationary variance is then

$$\text{var}(x)_{st} = \frac{(k_{-1} + k_2) k_1 \mathbf{p}^* d_0}{(k_{-1} + k_2 + k_1 \mathbf{p}^*)^2}.$$

This finally leads to the expression of the stationary coefficient of variation

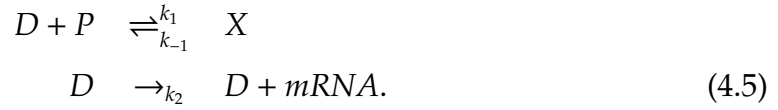
$$\lim_{t \rightarrow \infty} \frac{\sqrt{\text{var}(x)}}{\langle x \rangle} = \frac{\sqrt{d_0 (k_{-1} + k_2) k_1 \mathbf{p}^*}}{d_0 k_1 \mathbf{p}^*} = \sqrt{\frac{K_m}{d_0 \mathbf{p}^*}}$$

In addition to the assumption that the binding/unbinding rates must be faster than the *mRNA* translation rate $k_2 \ll k_{-1}, k_1$ as in the deterministic time scale separation,

ration, we need to make sure in the stochastic regime the stationary coefficient of variation is small. That means that the approximation is better if the number of the activator proteins is large.

4.2.2 Monod-type variable reduction for repressed transcription

As above we consider the simplified reaction scheme for the repressed genetic transcription reaction:



where D is the amount of DNA not occupied by the repressor protein P , X is the DNA-repressor complex and the conservation equation for the overall DNA present in the system $D_0 = D + X$ holds. Then the resulting master equation can be written as:

$$\begin{aligned} \frac{dP(d, p, x, m; t)}{dt} = & k_1(d+1)(p+1)P(d+1, p+1, x-1, m; t) \\ & + k_{-1}(x+1)P(d-1, p-1, x+1, m; t) \\ & + k_2dP(d, p, x, m-1; t) \\ & - (k_1dp + k_{-1}x + k_2d)P(d, p, x, m; t) \end{aligned}$$

As for the case of the promoted transcription binding and unbinding reactions are much faster compared to the transcription: $k_1, k_{-1} \gg k_2$ and therefore the dynamics of the complex X happen on the considerably faster time scales than the $mRNA$ production. As before, we assume intermediate values for the slow variables $P = \mathbf{p}^*, mRNA = \mathbf{m}^*$ and write the fast dynamics probability distribution for the complex X :

$$\begin{aligned} d_t p_n &= k_1(d_0 - n - 1)\mathbf{p}^* p_{n-1} \\ &+ k_{-1}(n+1)p_{n+1} \\ &- (k_1(d_0 - n)\mathbf{p}^* + k_{-1}n)p_n \\ &= (\mathbb{E} - 1)r_n p_n + (\mathbb{E}^{-1} - 1)g_n p_n \end{aligned}$$

with $g_n = k_1 \mathbf{p}^*(d_0 - n)$ and $r_n = k_{-1}n$.

Following the same considerations for a stochastic process confined between the two absorbing boundaries, we obtain an analytically solvable first order ODE for the mean of the complex X

$$d_t \langle n \rangle + (k_1 \mathbf{p}^* + k_{-1}) \langle n \rangle = d_0 k_1 \mathbf{p}^*$$

The solution is

$$\langle n \rangle = \frac{d_0 \mathbf{p}^* k_1}{\mathbf{p}^* k_1 + k_{-1}} (1 - e^{-(k_1 \mathbf{p}^* + k_{-1})t}) - IC * e^{-(k_1 \mathbf{p}^* + k_{-1})t}$$

with IC denoting initial condition. The intermediate complex X will approach the steady state exponentially fast on the time scale, which is proportional to the fast reaction variables $\propto (k_1 \mathbf{p}^* + k_{-1})$:

$$x_{st} = \lim_{t \rightarrow \infty} \langle n \rangle = \frac{d_0 \mathbf{p}^* k_1}{\mathbf{p}^* k_1 + k_{-1}} = \frac{d_0 \mathbf{p}^*}{\mathbf{p}^* + k_{-1}/k_1}$$

leading to the reduced Master equation for the repressed transcription:

$$\begin{aligned} d_t p_m &\approx (d_0 - x_{st}) k_2 (\mathbb{E} - 1) p_m \\ &= \frac{d_0 k_2}{\mathbf{p}^* + k_{-1}/k_1} (\mathbb{E} - 1) p_m \end{aligned}$$

which again is in agreement with reduction result for the purely deterministic system.

Analogous calculations for the variance lead to the similar result for the stationary coefficient of variation:

$$\lim_{t \rightarrow \infty} \frac{\sqrt{\text{var}(x)}}{\langle x \rangle} = \sqrt{\frac{k_{-1}/k_1}{\mathbf{p}^* d_0}} \approx \sqrt{\frac{1}{\mathbf{p}^* d_0}}$$

In the deterministic description the coefficient of variation is implicitly assumed to be equal to zero. It means that this dynamical aspect is not accounted for if using deterministic description.

4.2.3 Contributions from higher moments

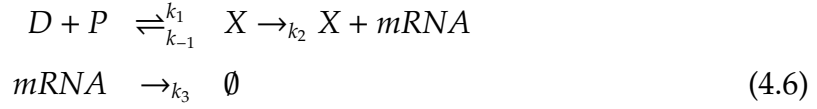
Time scale separation has been applied to other gene expression models as for example in Kepler and Elston (112), Shahrezaei and Swain (113), where *mRNA* and protein dynamics have been lumped in order to obtain steady state distributions for the proteins. The protein steady state distributions have been shown to possess a heavy tail or being bimodal. For such distributions the consideration of the first two moments (mean and variance) is not sufficient for validation of the approximation scheme and might be misleading. In our derivation the time scale separation is applied at the binding/unbinding of the activator or repressor proteins. The model reduction step in this scheme occurs on a different point, where the *mRNA* distribution is not expected to become bimodal or heavy tailed as the consequence of the approximation scheme. We have also chosen not to use a continuous variable for *mRNAs* as it is often done during the approximation schemes for the proteins, since *mRNA* copy numbers are low, i.e. on the order of dozens. In addition to the analytical conditions on the first and the second moments, we validate the approximation by numerical simulations of the full and the reduced model comparing the resulting distributions with the Kullback-Leibler measure and accounting in this way for higher moments.

4.2.4 Numerical validation of the approximation scheme

Analytically we have addressed the mean and variance during the quasi-steady state approximation. The contributions of other moments can be assessed numerically by comparing the probability distributions for the number of *mRNA* copies for the full and the reduced system.

We consider a scenario where the *mRNA* production depends on an activating protein $\mathbf{p}^*$, which is in the steady state. Note that the number of the regulator proteins $\mathbf{p}^*$ is not assumed constant, but is a stochastic variable fluctuating on slow time scales as mentioned earlier (see section 4.2.1). We simulate the stochasticity of the transcription regulators $\mathbf{p}^*$ by drawing their values from steady state probability distributions, which can be considered common for genetic transcription regulation. We use the following protein probability distributions Π_{common} for numerical simulations: (1) gamma distribution with low copy numbers, (2) gamma distribution with high copy numbers, (3) a bimodal distribution (1, 114).

In Fig. 4.1, the full and the reduced models are compared for the activated transcription, where the activator protein is sampled from one of the common distributions $P = \mathbf{p}^* \sim \Pi_{\text{common}}$. The following reaction scheme

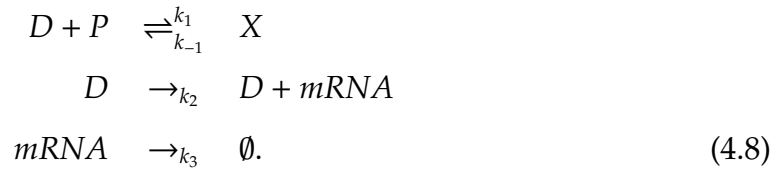


is used for the full model of the activated transcription and



for the reduced model. The constants are $k_1 = 1 \text{ min}^{-1}; k_{-1} = 0.75 \text{ min}^{-1}; k_2 = 0.05 \text{ min}^{-1}; k_3 = 0.06 \text{ min}^{-1}, d_0=2$ and $g_a(\mathbf{p}^*) = \frac{d_0 k_2 \mathbf{p}^*}{(k_{-1}+k_2)/k_1 + \mathbf{p}^*}$. (Note that $k_1, k_{-1} \gg k_2$ is the necessary condition for the time separation in the approximation scheme). The *mRNA* concentration reaches steady state distributions in the full and the reduced model and we compare the distributions with Kullback-Leibler measure. In this regime the reduced system approximates well the full system, and as expected the approximation is slightly better for the regime with large activator copy numbers.

Similar set up is used in Fig. 4.2 for the repressed transcription described by the following reaction scheme for the full model:



and the reduced model:



with $g_r(\mathbf{p}^*) = \frac{d_0 k_2}{\mathbf{p}^* + k_{-1}/k_1}$. In order to demonstrate the dependence of the approximation accuracy on the condition $k_1, k_{-1} \gg k_2$, the constants $k_1 = 1 \text{ min}^{-1}; k_{-1} = 0.75 \text{ min}^{-1}; k_2 = 0.1 \text{ min}^{-1}$ have been chosen closer together compared to Fig. 4.1 resulting in the larger Kullback-Leibler divergence as expected (see Fig. 4.2).

Interestingly, in both cases of the approximation slightly overestimates the true *mRNA* numbers, shifting the approximate distribution to the right. Similar

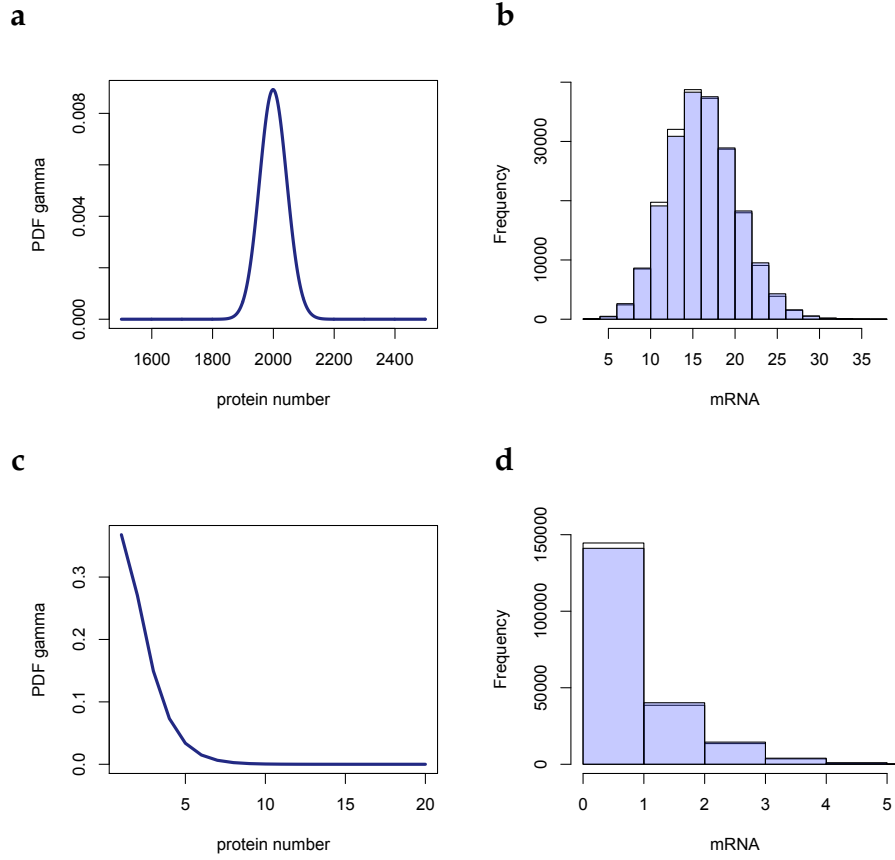


Figure 4.1: Numerical assessment of the stochastic quasi-steady state approximation for activated transcription. In a plausible gene expression scenario the regulator numbers will be constantly fluctuating. In order to explore how well the approximation works under the slow changing regulator numbers, we sample the values for the stochastic variable $\mathbf{p}^*$ with an adjusted sampling rate from the distributions shown in (a,c). The sampling rate is adjusted according to the typical time scale on which the proteins are expected to be changing ($\approx$ ten times slower than the *mRNA* time scales). The steady state distribution of the activators $\mathbf{p}^*$ are shown in (a,c) and the accordingly resulting *mRNA* distributions in (b,d). The *mRNA* distributions in (b,d) are reconstructed from $2 \cdot 10^4$ steady state samples and the full model is shown in empty bars and the approximation in blue filled bars. The reaction schemes Eq. 4.6 and 4.7 has been used to obtain the steady state distributions for *mRNAs*. The Kullback-Leibler measure for d, $KL = 1.2 \cdot 10^{-3}$, is slightly larger than for b, $KL = 5.0 \cdot 10^{-4}$, as expected since the approximation is better if the activators are present in large amounts. (Kullback-Leibler measure has been calculated using the R package "flexmix"). The overall agreement of the full and the reduced model for both scenarios is very good.

4.3 The generalized repressilator model in the Master equation formalism 116

systematic bias has been noticed in Rao and Arkin (36), where the comparison of the distribution means has been explored.

The approximation also works for binomial steady state distributions for the repressor protein (see Fig. 4.1e-f). The Kullback-Leibler measure is of the same order of magnitude as for the unimodal distributions.

4.3 The generalized repressilator model in the Master equation formalism

Similarly to the deterministic description we consider model transcription, translation and degradation of proteins and *mRNAs* in the full Master equation resulting in 4 reactions per gene. We will use the creation (annihilation) $\mathbb{E}_{p_j}(\mathbb{E}_{p_j}^{-1})/\mathbb{E}_{m_j}(\mathbb{E}_{m_j}^{-1})$ for creation (annihilation) events of proteins and *mRNAs* as introduced in the introduction section 1.3. Then the Master equation for the generalized repressilator amounts to:

$$\begin{aligned} \partial_t P(\mathbf{m}, \mathbf{p}; t) = & \left[\frac{c_1}{p_{j-1}^2 + 1} (\mathbb{E}_{m_j}^{-1} - 1) \right. \\ & + c_2 (\mathbb{E}_{m_j} - 1) m_j \\ & + c_3 \cdot m_j (\mathbb{E}_{p_j}^{-1} - 1) \\ & \left. + c_4 (\mathbb{E}_{p_j} - 1) p_j \right] P(\mathbf{m}, \mathbf{p}; t) \end{aligned}$$

using as usually the index $j \in \{1, \dots, N\}$ and respecting circular boundary conditions.

4.4 Reaction rates dependence on the system size Ω for the generalized repressilator model

The Master equation for the generalized repressilator model contains 3 linear reactions, which are the death rates of *mRNAs* and protein birth rate, and a nonlinear term for the repressor dependent transcription rate. The constants c_j in linear terms do not change with the size of the system Ω , because the biochemical rate is just proportional to the number of species in the system and if the size

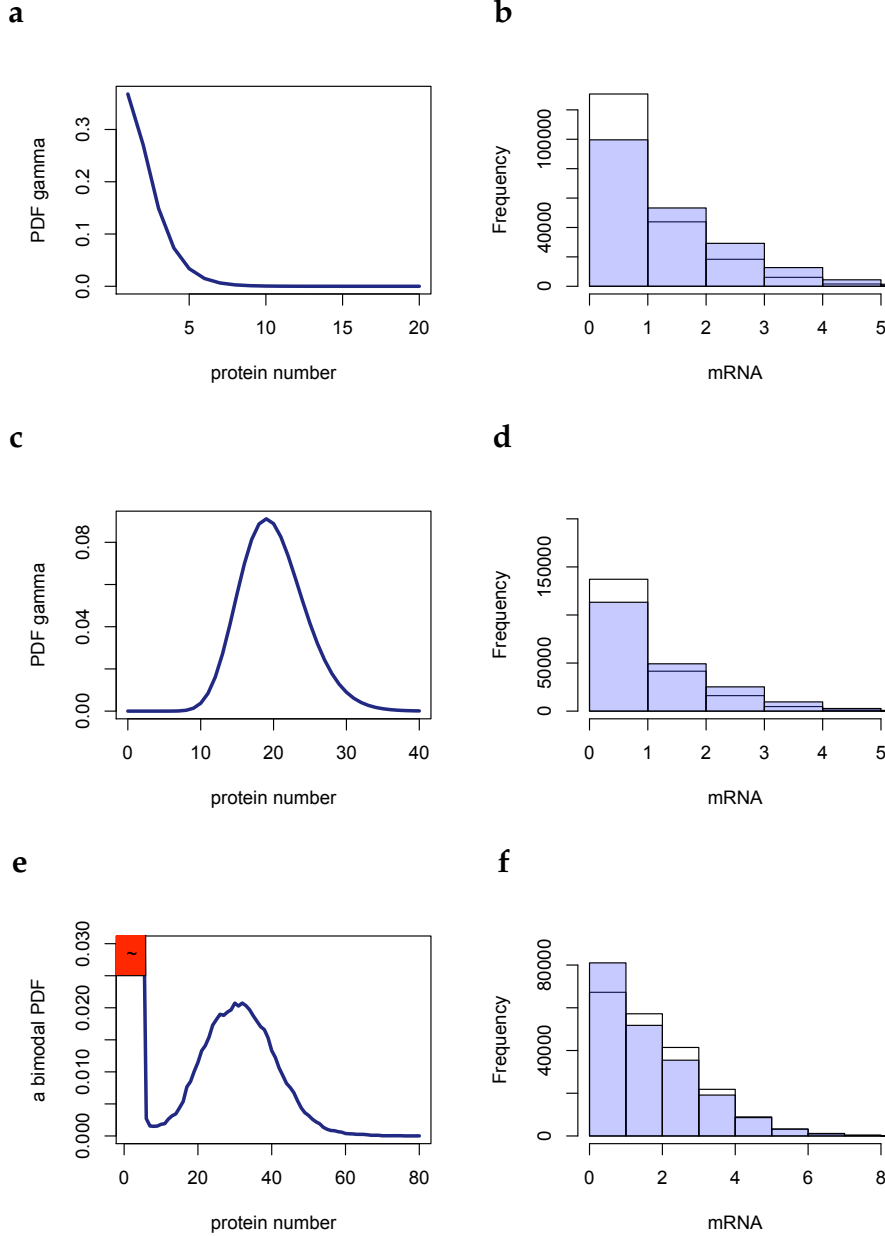


Figure 4.2: **Numerical assessment of the stochastic quasi-steady state approximation for repressed transcription.** Similarly to Fig. 4.1, we explore how well the approximation works under the slow changing repressor numbers sampling the values for the stochastic variable p^* with an adjusted sampling rate from the distributions shown in (a,c,e). The common steady state distribution for the repressors p^* are shown in (a,c,e) and the accordingly resulting *mRNA* distributions in (b,d,f). The distributions in (b,d,f) are reconstructed from $2 \cdot 10^4$ steady state samples and the full model is shown in white and the approximation in blue filled bars. The reaction schemes Eq. 4.8 and 4.9 has been used to obtain the steady state distributions for *mRNAs*. The Kullback-Leibler measure between the full and the Monod-type approximated model is $KL = 6 \cdot 10^{-2}$ for d and b. The agreement is worse than Fig. 4.1, since the constants k_1, k_{-1}, k_2 are closer together than in Fig. 4.1. This is to demonstrate the dependence of the approximation scheme accuracy on the condition $k_1, k_{-1} \gg k_2$. (e-f) A bimodal stationary state probability distribution obtained numerically using the DCFTP (see section 1.3.3) was used as a steady state distribution for the repressor (the red part at the beginning of the distribution has been truncated for illustration clarity). The *mRNA* distributions for the full and the reduced model agree well ($KL = 2 \cdot 10^{-2}$) also in this scenario.

is increased, so will be the absolute number of species. The biochemical rate of linear reactions scales therefore automatically with the size of the system (20).

However, for the nonlinear term we need to find an appropriate transformation. An easy intuitive way would be to use the "scaling argument" typical in physics (see for instance (115)): the overall contribution to the change of the species number or concentration should be multiplied by Ω to adjust for the size increase. If other dependencies on the number of proteins appear as for example in the Hill- or Monod-functions, then they have to be scaled by Ω . Alternatively derivations from Guantes and Poyatos (116) or van Kampen (20) for extensive and intrinsic variables can be consulted. Both argument lines result in the following transformation rule for our non-linear terms:

$$\frac{c_1}{1 + p^2} \rightarrow \frac{c_1 \Omega}{1 + (\frac{p}{\Omega})^2}$$

The figure 4.3 below shows simulations of the stochastic system carried out with the Gillespie algorithm at different values of Ω . As expected if we increase the size of the system Ω the fluctuations disappear and the system becomes more dependent on initial conditions.

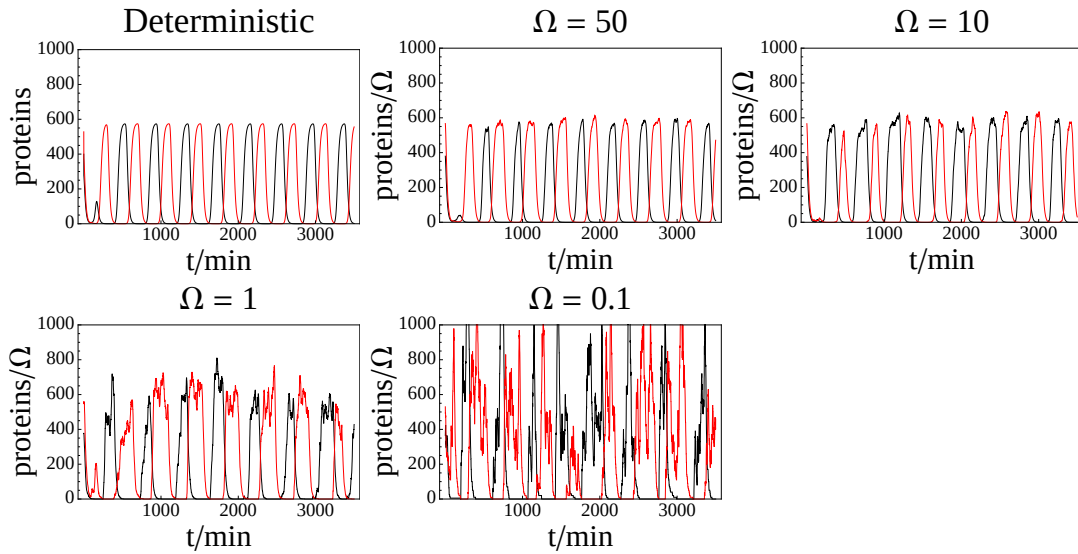


Figure 4.3: From deterministic to stochastic regime in the generalized repressilator model using the dependence of reaction rates on the system size Ω .

We will use the established parameter dependence on the system size in order to address how the system size will influence the onset, the duration of transient

oscillations, and the quality of oscillations in the generalized repressilator model.

4.5 Transient oscillations in the stochastic setting

The first question to address with the Master equation formalism is, if internal noise would affect the oscillations in the generalized repressilator model. In particular, it is interesting how the quasi-stable oscillations will be affected by the internal noise, in different regimes and for different system sizes.

First we assess the effects of internal noise on the transient oscillations in a biologically relevant regime far away from the bifurcation point. We carried out numerical simulations using Gillespie (section 1.3.3) and compared the length of oscillating transients starting from the same random initial conditions in stochastic and deterministic regime (see Fig. 4.4).

In the stochastic setting the noise due to small copy numbers does not destroy the quasi-stable oscillations in even rings. On the contrary, transient oscillations last longer under noise. It means that the quasi-stable attractor identified with the deterministic bifurcation theory becomes even more prevalent under internal noise. The figure 4.4 shows the direct comparison using the following parameters $c_1 = 1.6, c_2 = 0.12, c_3 = 2.6, c_4 = 0.06$. The initial conditions for the stochastic simulation are chosen to be exactly the same as in the deterministic case and the statistics from 10 runs for each starting value have been collected for the stochastic simulation.

Interestingly, the $N = 6$ gene ring is quiescent in the deterministic realization, but in the stochastic setting transient oscillations have been observed. The effective volume of the attractor to the quasi-stable oscillations seems to be increased in the stochastic regime by fluctuations for parameter values far away from the bifurcation point. This result is a numerical hint that internal noise influences the quasi-stable oscillations in the generalized repressilator model and might even to some extent excite them.

The influence of noise can be explored in a systematic manner by using the reaction rate scaling in the system size – Ω , which has been introduced in the section 4.4. We will use the system size Ω as an independent variable together with the bifurcation variable c_1 and systematically explore the influence of noise in the (Ω, c_1) -plane.

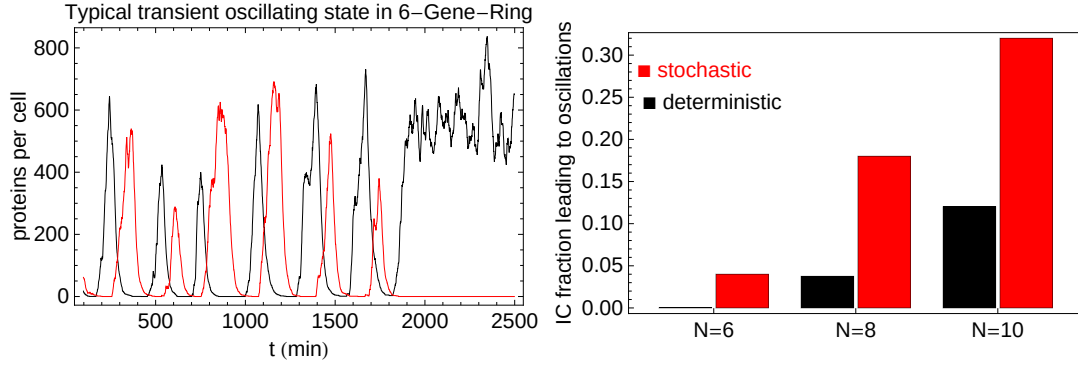


Figure 4.4: **Transient oscillations are more prevalent in the stochastic setting.** For the same initial conditions (IC) and the same constants ($c_1 = 1.6, c_2 = 0.12, c_3 = 2.6, c_4 = 0.06$) the stochastic system goes more frequently into the oscillating state than the deterministic system.

4.6 Analysis of Hopf bifurcations in the stochastic setting

From the model derivations (see section 4.2) it becomes clear that the parameter c_1 for the *mRNA* production rate, which has been used as bifurcation parameter in the nonlinear deterministic analysis is directly proportional to d_0 the number of DNA copies in the system also in the stochastic setting. d_0 can be manipulated experimentally and its increase would induce Hopf bifurcations and the according limit cycle oscillations. The gene copy number d_0 is a “biological knob”, which can be tweaked to *in vivo* induce oscillations in a bacterial population.

The knob d_0 also influences strongly the stochastic characteristics of the system, so that the oscillatory behavior as well as the quality of the oscillations are expected to be affected. Therefore, we analyze the onset of oscillations in the stochastic model, not only in dependence of a parameter c_1 , but in the (Ω, c_1) -plane, systematically taking the influence of noise into account.

4.6.1 Induction of Hopf bifurcations in stochastic setting

The numerical Gillespie simulations show that system size plays a significant role by the onset of the oscillating behavior (see Fig. 4.5 and 4.6). In the low- Ω regime, where the number of molecules involved in the biochemical reactions are low and the stochastic fluctuations are strong, the oscillating pattern are less distinct even after the onset of bifurcations according to the deterministic bifurcation analysis (see Fig. 4.5a). (We will quantify the quality argument using a metric in the

following.)

An interesting effect occurs in large- Ω regime right before the onset of oscillations in the deterministic model (see Fig. 4.5(f,g)). Oscillating pattern are appearing already at smaller parameter values in the stochastic setting. In general, it is an indication that the stochastic resonance effect might be present in the model (108). We will explore this hint in more detail in the following section 4.6.2.

The onset of quasi-stable oscillations in the even $N = 6$ gene ring shows the appearance of transient oscillations, however these disappear after a few periods. The stable up/down solution is then approached (see Fig. 4.6). The oscillation enhancing effect, which we have seen in section 4.5, does not seem to set on right at the bifurcation point. This can be explained with the following: in the deterministic bifurcation figure 2.7 we see that the UPOs and the stable up/down solution approach each other for larger c_1 values. Therefore, for larger c_1 values and far from the bifurcation point the UPO can be reached more easily in the stochastic setting using the attractive directions from the stable up/down solution. However, right at the onset of oscillations at the Hopf bifurcation point, the distance between the stable up/down solution and the quasi-stable attractor is quite large, the up/down attractive directions are distinct from the UPO. The net effects of noise are destructive for the oscillations in this regime.

4.6.2 The expansion in the inverse system size Ω and stochastic resonance around the bifurcation point for stable limit cycles

Following the numerical hint from the previous section (see Fig. 4.5(f,g)) we will investigate the occurrence of stochastic resonance around the bifurcation point for stable limit cycles. Before the occurrence of the first Hopf Bifurcation, odd-numbered generalized repressilators possess a single stable fixed point (where all genes have the same concentrations (see Fig. 2.4.1) and in this regime the system dynamics are dominated by the macroscopic term so that the fluctuations do not grow. In this regime, the linear noise approximation discussed in section 1.3.5 can be applied (see (20) Chap. X.4). We linearized the Master equation around the analytically found deterministic Hopf bifurcation, i. e. the right edge of the parameter regime, where the linear noise approximation is valid. Then taking the deterministic and first fluctuation terms from the Ω expansion into account, we

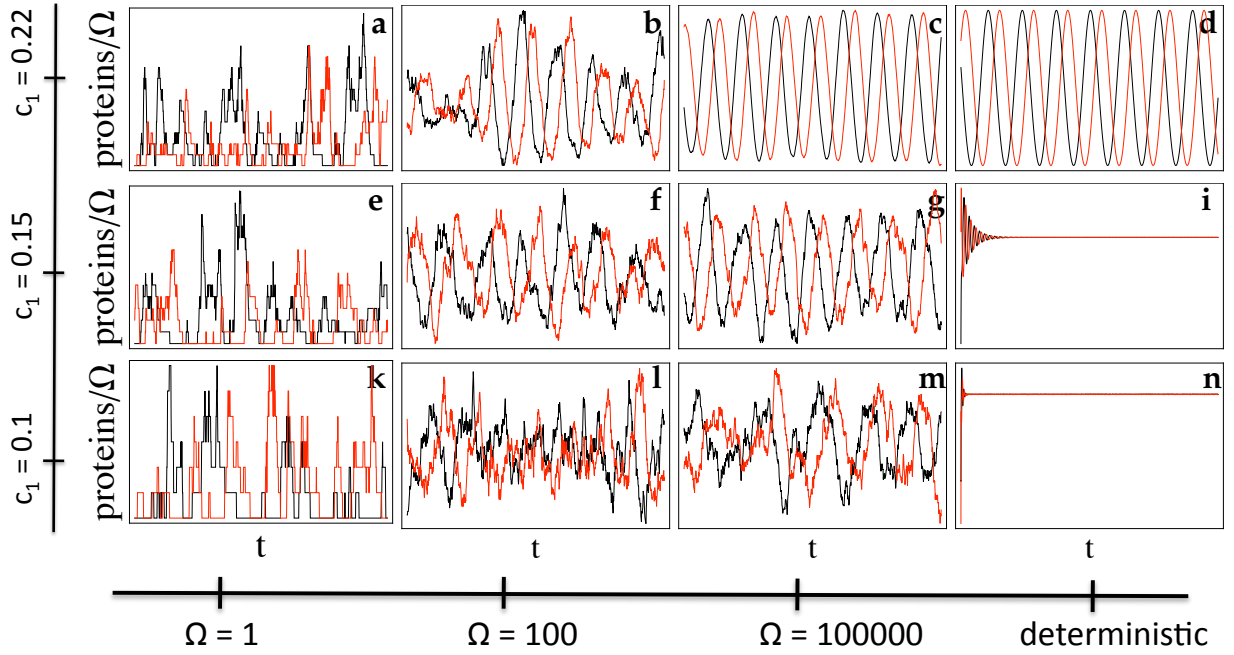


Figure 4.5: **Onset of oscillations in the stochastic repressilator model for $N=3$ genes nearby the bifurcation point under the variation of the system size Ω .** In the deterministic regime the limit cycle behaviour starts, then the bifurcation parameter c_1 reaches the value 0.21 so that the sub-figures **a-d** show the behavior of the system after the Hopf bifurcation and the sub-figures **e-n** before the Hopf bifurcation. Depending on the system size Ω oscillatory behavior seems to appear already earlier in the stochastic version of the generalized repressilator model as shown in sub-figures **(f,g)**. This is a numerical indication that stochastic resonance might be present around these parameter values and further investigation (see section 4.6.2) has confirmed the hint.

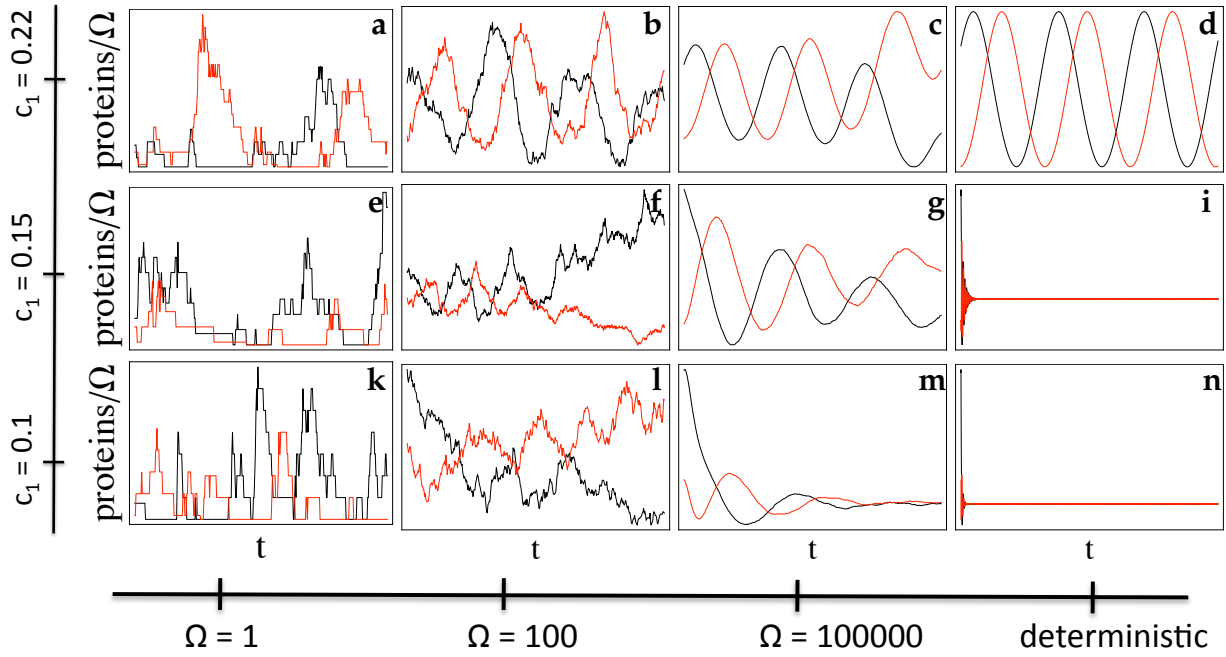


Figure 4.6: **Onset of oscillations in the stochastic generalized repressilator model for $N=6$ genes nearby the first Hopf bifurcation under the variation of the system size Ω .** In the deterministic regime the limit cycle behaviour starts, then the bifurcation parameter c_1 reaches the value 0.21 (same as for $N=3$ due to the symmetry arguments in chapter 3). Similarly to the 3 gene ring example we have assessed the oscillating behavior starting on the deterministic limit cycle. For even rings the system shows a couple of oscillations and then approaches the stable up down solution. It seems the oscillations do not last, i.e. are too unstable, to collect statistics in form of Fourier spectra and assess their quality.

show the effect of stochastic resonance appearing at the first Hopf Bifurcation.

Denoting the system state $(\mathbf{m}, \mathbf{p}) \equiv \mathbf{x}$, the according SDE is

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^t A \cdot \mathbf{x} ds + \int_0^t \sigma \cdot \mathbf{x} d\omega_s \quad (4.10)$$

where $A = J|_{x_t}$ is often referred to as drift matrix and is equal to the deterministic Jacobian evaluated at the Hopf bifurcation point (see section 3.2.1). The matrix σ is often referred to as the diffusion matrix, its calculation follows from the linear noise approximation (see section 1.3.6), and we evaluate this function also at the Hopf bifurcation point. The stationary stochastic power spectrum of this linear SDE is calculated with the well-known formula for multi-variable linear SDEs (117), chap. 4.4.6:

$$S(\omega) = (A + i\omega)^{-1} \sigma \cdot \sigma^T (A - i\omega)^{-1} \quad (4.11)$$

The power spectra are obtained from the diagonal contributions on $S(\omega)$ and shown for values around the bifurcation point in Fig. 4.7. The power spectrum peak can be made divergent to ∞ around the resonance values. The stochastic amplification effect due to low copy numbers is found in all odd-numbered rings, in figure 4.7 we exemplify the effect with rings $N=3$ and $N=7$.

4.7 The quality of stochastic oscillations

Somehow we would like to assess the quality of oscillations in a more quantitative manner. Which means finding a preferably dimensionless measure, which would say that Fig. 4.5(g,c,d) shows a qualitatively good oscillation regime and Fig. 4.5(a,i,k) does not oscillate. Intuitively, the power spectrum for the time traces should be peaked around a frequency and the variance of the power spectrum should be small, compared to the mean. A useful measure with theoretical boundaries has been proposed by Risau-Gusman and Abramson (118):

$$Q(\omega_{peak}) = \frac{\omega_{peak} S(\omega_{peak})}{\int S(\omega) d\omega} \quad (4.12)$$

We have used this score in order to assess the quality of oscillations in the

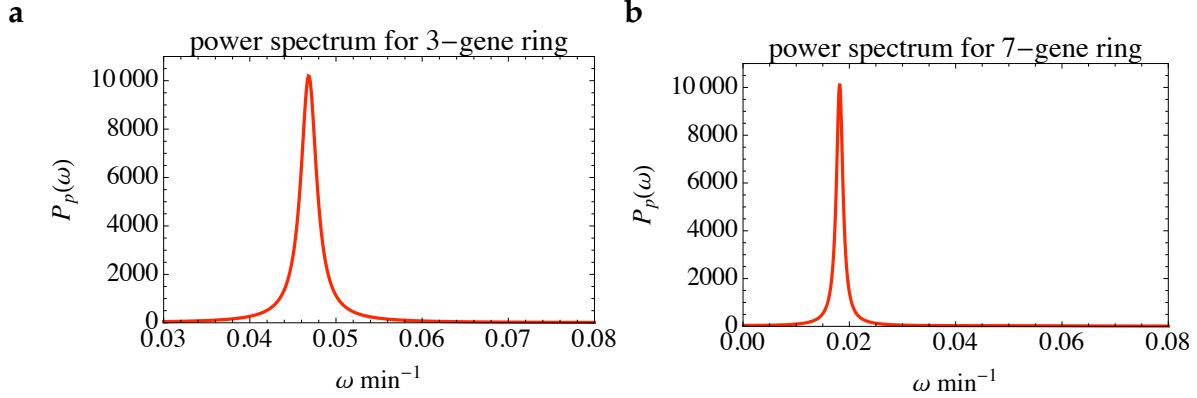


Figure 4.7: Power spectra obtained from the linearization around Hopf bifurcations giving rise to stable periodic orbits in the generalized repressilator model. In the stochastic regime the oscillations emerge already for smaller parameter values, where there are no oscillations in the deterministic setting. We are therefore expecting the appearance of stochastic resonance at the bifurcation point (108–110). The parameter value for resonance can be estimated from the linearized system for parameter values around the deterministic oscillation onset. According to the linearized model, we expect stochastic resonance to appear around $c_1 = 0.162$ for the stable limit cycle in $N = 3$ **(a)** and around $c_1 = 0.126$ for the stable limit cycles in $N = 7$ **(b)**. The frequency peak can be made divergent around these parameter values, which physically means that stochastic fluctuations occurring on the microscopic level lead to coherent oscillations on the macroscopic level (108). Direct Gillespie simulations around these values for $N = 3$ and $N = 7$ have confirmed the numerical hints of the linearized spectra. **Note:** As a direct consequence of the tiling discussed in chapter 3, the linearized spectra of all symmetric solutions, i.e. 6-, 12-, 18-gene unstable periodic orbits with the same symmetry class (indicated by the same color in Fig. 3.3), will have a similar linearized spectra as $N = 3$ stable periodic orbit. However, in the unstable solutions the effect of stochastic resonance has not been observed and these solutions approach after few oscillations the stable up/down solution also at the resonance parameter value.

generalized repressilator model for the simple case of three genes $N = 3$. The results are summarized in the Table 4.1. We have used three Ω regimes as before $\Omega \in \{1, 100, 100000\}$ and three values for the parameter $c_1 \in \{0.1, 0.162, 1.6\}$, i.e. before the bifurcation, at the stochastic resonance and far away from the bifurcation.

Before the bifurcation and the onset of stochastic resonance the power spectrum is flat, and the quality score assumes its lowest values. The stochastic resonance as expected is more prominent in larger systems and a steady increase of the quality score is observed in the second row of the result table 4.1. In the regime far away from the bifurcation ($c_1 = 1.6$) the quality of oscillations accord-

ing to this score strongly depends on the system size (Ω). In particular in small systems the quality of oscillations according to this score does not become better even if we move far away from the bifurcation point.

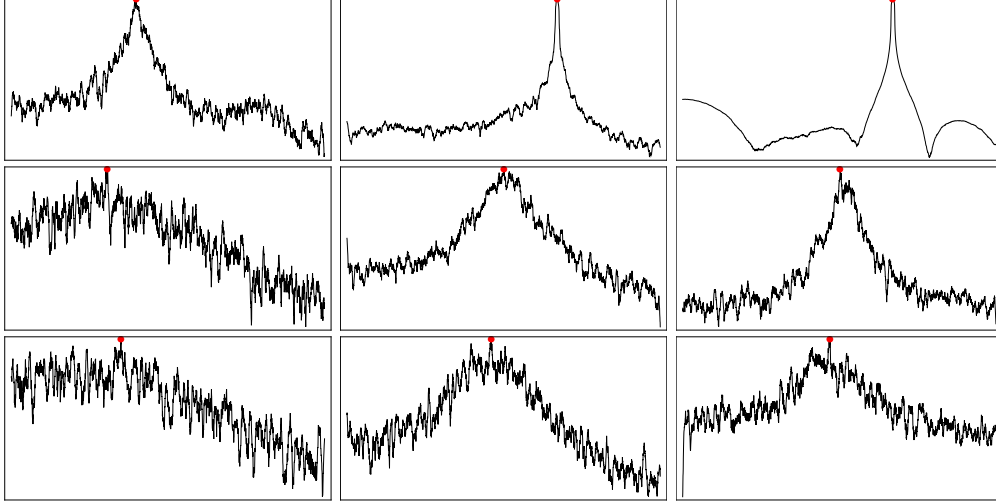


Figure 4.8: **Power spectra for the corresponding quality scores shown in table 4.1.** The spectra have been obtained from Gillespie simulations (see section 1.3.3), are normalized and shown on the logarithmic scale to provide an intuition for oscillation quality score Eq. 4.12 and the result table 4.1. The peak ω_{peak} as defined in Eq. 4.12 and is marked by a dot for each spectrum.

	$\Omega = 1$	$\Omega = 100$	$\Omega = 100000$
$c_1 = 1.6$	$2.4 \cdot 10^{-3}$	$8.0 \cdot 10^{-2}$	$1.3 \cdot 10^{-1}$
$c_1 = 0.162$	$1.0 \cdot 10^{-3}$	$2.0 \cdot 10^{-3}$	$3.0 \cdot 10^{-3}$
$c_1 = 0.1$	$1.4 \cdot 10^{-4}$	$1.4 \cdot 10^{-3}$	$1.3 \cdot 10^{-3}$

Table 4.1: **Ω -dependent quality of oscillations in the generalized repressilator model** The quality score defined in Eq. 4.12 has been applied for the repressilator model with $N = 3$ genes at the parameter value before the bifurcation ($c_1 = 0.1$), at the stochastic resonance value ($c_1 = 0.162$) and further away from the bifurcation point ($c_1 = 1.6$) for the different system sizes Ω . As expected, better oscillation quality is observed in larger Ω regimes further away from the bifurcation point and at the stochastic resonance value.

4.8 Discussion

We have explored the effects of internal noise on the dynamics of the generalized repressilator model at the bifurcation point and far away from the bifurcation

point. To begin with, we observed via numerical simulations that transient oscillations in even rings last longer in the stochastic model than in the deterministic setting in the regime far away from the bifurcation point. Internal noise does not seem to destroy the quasi-stable oscillations in this regime, on the contrary it effectively prolongs transient behavior. This has motivated us to explore the influence of noise in a more systematic manner.

To this end, we have established the dependence of the rate constants on the system size Ω , which controls the amount of internal noise in the system. The constants of linear reactions do not change with Ω , but the nonlinear terms as usually needed an according transformation. The derivation allowed us to systematically investigate the influence of the parameter change c_1 and the system size Ω on the quality and the onset of oscillating behavior in the generalized repressilator model.

We have identified a biochemical "knob" – the number of *DNA* copies in the cell, which triggers the emergence of Hopf bifurcations in the generalized repressilator model. Changing the copy numbers in the cell in this way, can induce the oscillations *in vivo* in living cells. In the deterministic description this operation would in principle reproduce the previously shown bifurcation diagrams. However, in the stochastic setting the increase of gene copies and the increase of the parameter c_1 has an effect on two levels in the systems dynamics: on the deterministic level it strongly affects the mean number of molecules present in the system and on the stochastic level it affects the noise level. The increase of gene copies strongly impacts the number of *mRNAs* and consequently also the number of proteins. The overall effect on the system dynamics is strong and in particular the noise characteristics change. In the (c_1, Ω) -plane indicated by Fig. 4.5, the increase of gene copy number induces the diagonal (c_1, Ω) -plane navigation and not vertical along the c_1 direction as could wrongly be assumed from the pure deterministic analysis.

An interesting effect of stochastic resonance is observed by the navigation on the (c_1, Ω) -plane. We see the oscillations in the stochastic system for large system size Ω before the onset of oscillations in the deterministic system in odd-numbered rings, i.e. already for smaller c_1 values. This effect has been documented in other stochastic systems around the bifurcation point (108–110), and is explained as a collective phenomenon of stochastic fluctuations on the microscopic level

leading to coherent oscillations on the macroscopic level. Essentially, it is like a driven oscillator, where the external driving force are the microscopic fluctuations. The effect of stochastic resonance is confirmed for stable limit cycles via Master equation linearization around the Hopf bifurcation point.

In the following, we have investigated the quality of oscillations depending on the system size Ω and the bifurcation parameter c_1 . We have carried out numerical simulations for $N=3$ gene model varying c_1 and Ω values. c_1 values prior the Hopf bifurcation, at the stochastic resonance value, and further away from the bifurcation have been used. We then applied numerical Fourier analysis assigning quality scores. As expected, in large Ω regimes the oscillations are of better quality and more regular. For small system sizes the quality of the oscillations remains poor even for the values further away from the bifurcation point. This finding again emphasizes the importance of the system size as an independent variable, which needs to be considered for synthetic biology design purposes.

The conclusion for the engineering point of view is that for regular oscillations on the macroscopic level, synthetic oscillators should be designed in sufficiently large Ω regimes. However, there is always be a trade off between the quality of the oscillations and the amount of metabolic burden due to high levels of expressed heterologous proteins. The (c_1, Ω) -plane approach can be used to suggest a good compromise, which for our model is by Ω values around 100.

Chapter 5

Model-free algorithmic induction of optimized control strategies for genetic networks

We have explored the dynamics of the generalized repressilator model in the deterministic and stochastic formulation assessing both the steady state and the transient behavior. Based on the acquired knowledge about the system we want to influence and control gene ring structures via external signals directing them to the desired dynamics. In chapter 2 of the thesis we have already shown an oscillator design around an UPO, where the desired oscillating state has been influenced by application of external interventions, i.e. light flashes. In this chapter we generalize the idea of control for genetic networks and present numerically induced optimized strategies for external interference with the generalized repressilator model.

The following chapter is based on the manuscript

Natalja Strelkowa, Guy-Bart Stan, Damien Ernst, and Mauricio Barahona
intended for the submission to
Molecular Systems Biology

5.1 Motivation for feedback control in biology

As discussed in the “Introduction” chapter (see section 1.6), control theoretical concepts have proved useful for the understanding of naturally occurring biological mechanisms, such as for example perfect adaptation. These concepts can also be used to design optimized intervention strategies with natural systems. In particular, feedback control based on the input-output relationships seems promising for epidemiological applications, for instance non-pharmaceutical interventions as first response to a pandemic threat (119) or optimized drug treatment of HIV-infected patients (120). Control theoretical concepts have also been widely applied in neuroscience, as for example in (121) the well-known Kalman filter has been used for estimation of the system state and parameters in neural cortex model.

We consider the application of feedback control to gene regulatory networks. Perturbations or intervention signals for genetic networks can be implemented biochemically via conditional gene knock outs (61, 122), heat shocks (123), light pulses (28, 124), etc. Application of control theoretical concepts would provide intervention protocols for these biochemical perturbations, which would conduct genetic networks to desired states in an optimal way by accumulation and integration of information from previous experiments. The optimality criterion, which is usually the minimal energy in classical engineering problems, can be defined in a system tailored manner for the genetic networks. It can for example be time, i.e. use the control signals to reach the desired state as quickly as possible, or cell burden, i.e. use the control signals to reach the desired state minimizing the expression of heterologous proteins, or similar.

Previously discussed feedback control strategies based on the input-output relationships can now also be applied to genetic networks due to the recent progress in technology for observing and influencing gene regulatory networks. Quantitative *in vivo* estimates of biological system states can for example be obtained via fluorescent markers (30, 114) and spatially targeted transcription induction can be performed in living cells using monochromatic light (28, 124). With these tools classical engineering concepts of optimal feedback control, where a quantitative system state estimate (read-out) and targeted intervention (control action) are required, is likely to become *in vivo* realizable also for genetic circuits (Fig. 5.1). This would open a new perspective for the experiment design in synthetic biol-

ogy as well as systematic information accumulation based on the input-output relationships providing rules for targeted and system-tailored interventions.

Classical control theory is a mature field and there are several computational algorithms for optimal feedback control, but what kind of characteristics should an ideal algorithmic controller for a gene regulatory network have? There are three main challenges, which a designer of a genetic network controller is facing: inherent non-linearity, stochasticity, and model uncertainty. Gene expression for both synthetic and natural networks is based on the law of mass-action (19, 40) and therefore polynomial and Hill nonlinearities are common features in gene expression models and that is why an ideal algorithmic gene network controller should effectively handle nonlinearities in the underlying problems. The second challenge is the stochasticity in form of internal and external noise during the gene expression (9), which requires that the proposed control strategy is robust to both internal and external noises in the read-out. And finally, the controller should be able to tackle the generally omnipresent model uncertainty in genetic networks, which despite efforts such as for instance world-wide reverse engineering competitions (125) is still an unsolved question.

In this chapter, we introduce a framework of a reinforcement-learning method for deduction of optimized control strategies (126), which has the ability to handle the expected gene network challenges. This algorithm uses time traces to induce a multi-dimensional function $Q(\mathbf{n}, \mathbf{u})$, which is the score of a control action $\mathbf{u} \in U$ for a system state $\mathbf{n} \in \mathbb{N}^s$. The optimized intervention strategy for direction of the network to a desired state is obtained by taking the maximum of this score function at each state $\mathbf{n}$: $\max_{\mathbf{u} \in U} Q(\mathbf{n}, \mathbf{u})$. The formulation of the algorithm is model-free and can therefore overcome both challenges: nonlinearities as well as model uncertainties often encountered in genetic network models. The robustness towards noise in the dynamics is achieved via a regressor EXTremely RAndomized Trees (Extra-Trees) (127), which is used for estimation of the score function Q .

We demonstrate the application of the algorithm on two synthetic biologically inspired examples: toggling a switch and operating our recently proposed synthetic circuit around a quasi-stable periodic orbit. The optimization direction, which is usually the minimal energy in classical engineering problems, is posed here for biological problems as minimization of metabolic cell burden due to production of heterologous proteins for control actions. The first problem "toggling

a switch” is majorly used for demonstration of the difference between the previously used naive technique for the switch reversal (1), which over-expresses synthetic proteins and the algorithmically induced optimized control strategy. The second example is the control of an even gene ring around an UPO, as has been used for the design of a switchable genetic oscillator in thesis chapter 2, which is used to demonstrate how a control problem of tracking a curve can be formulated and the strategy induced via reinforcement learning. None of these examples has previously been considered as a control theoretical example.

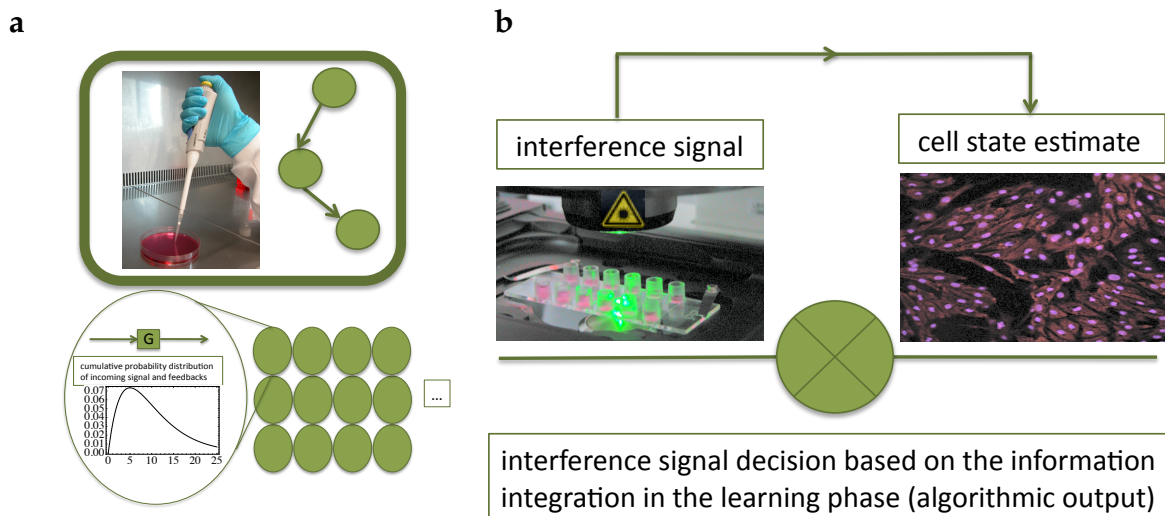


Figure 5.1: General idea on how classical control-theoretical approach can be adapted to *in silico* and in principle also to *in vivo* genetic networks. (a) In the learning phase time traces of key variables, for instance protein concentrations, are collected. This can be done either by numerical simulation of a model or experimentally preferably on the single cell resolution. (b) Based on the integrated information in the training set, the algorithm will provide a state dependent interference in order to conduct a genetic network to a desired state. (The photographs for this figure have been taken when I visited the experimental lab of Prof Rob Krams, and his PhD student Jennifer Frueh assisted me by taking the photographs and providing the cell image.)

5.2 Methods Description

Learning from a sample of trajectories: the Reinforcement Learning approach

Reinforcement learning approaches have been proved to be useful for a wide range of applications, where only few information on the underlying process is available (126, 128). The major requirement on the underlying process is that it is approximately Markovian, which means that the following state is fully determined by the information on the current state. In particular classical noisy mass-action kinetics underpinning gene expression processes can be formulated as control theoretical problems and solved with the reinforcement learning algorithms.

Formal System Description

Reinforcement learning can be used to find optimized control strategies for problems, which can be approximated as general first order Markov processes. Using otherwise notation as in van Kampen (20) and Introduction section 1.3.5 Markov property for successive times (i. e. $t_0 < (t_0 + \delta t) < \dots < t$) and control signals $\mathbf{u} \in U$ can be formulated as:

$$\begin{aligned} & Pr(\mathbf{n}(t), \mathbf{u}(t), t | \mathbf{n}_0, \mathbf{u}_0, t_0; \dots; \mathbf{n}(t - \delta t), \mathbf{u}(t - \delta t), (t - \delta t)) \\ &= Pr(\mathbf{n}(t), \mathbf{u}(t), t | \mathbf{n}(t - \delta t), \mathbf{u}(t - \delta t), (t - \delta t)) \end{aligned} \quad (5.1)$$

That means a process is first order Markovian if the state transition probability $Pr(\mathbf{n}(t), \mathbf{u}(t) | \cdot)$ at the time t only depends on the state of the system at the time $t - \delta t$ and is not dependent or affected by other previous states.

Master equation formulation for gene expression networks dynamics fulfills this requirement since the time evolution of the probability P on discrete set of states $\mathbf{n} \in \mathbb{N}^s$ is a PDE

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_{\mathbf{n}'} W(\mathbf{n} | \mathbf{n}', \mathbf{u}) P(\mathbf{n}', t) - W(\mathbf{n}' | \mathbf{n}, \mathbf{u}) P(\mathbf{n}, t)$$

where $W(\cdot | \cdot)$ are transition probabilities per unit time, which are now also depen-

dent on the external control signals $\mathbf{u} \in U$.

Approximate dynamics of the Master equation can be formulated as a stochastic differential equation (SDE) or as an ordinary differential equation (ODE) $d_t x = f(x)$ (see Introduction section 1.3.5 and (20) for connections between the formalisms via Ω expansion). The deterministic approximation $x = \phi(x_0, t - t_0)$ is a special case for a Markov process with a δ -peak transition probability $\delta(x - \phi(x_0, t - t_0))$. Similarly an SDE is a Markov process, where the transition probabilities are modulated by Gaussian noise. Therefore all three formulations are Markov processes and can be used to generate *in silico* time traces for the algorithmic trainings set. In principle the state transition data can also be obtained directly from the experiment as time traces taken at small enough intervals.

At each state transition a control signal $\mathbf{u}(t)$ can be applied to the system. What we would like to know is: if the system is in a state $\mathbf{n}$, what is the best action on the long term to be applied? This question is answered by a control strategy $\mu(\cdot) : \mathbb{N}^s \rightarrow U$, which is the mapping from the state space into the action space. Algorithmically we will evaluate possible strategies by score assignment and obtain a numerically optimized solution. This will be the final output of the algorithm, it assigns to each system state $\mathbf{n}$ the action $\mathbf{u}$, which is the best on the long-term scale.

Reinforcement learning needs to start with the definition of desired and forbidden/undesired states and actions. We use the function $r(\mathbf{n}(t), \mathbf{u}(t))$, which must be bounded, in order to associate an instantaneous reward to each state transition. The reward for the transition to desired target states will be positive and for the transitions to the undesired system states and actions will be negative. The reward function is part of the control problem formulation. In essence, it defines boundary conditions and can be used to summarize prior knowledge about the system. At the beginning of the algorithmic procedure system transitions, for which we do not know the score will be assigned a neutral reward. The algorithm will then induce scores for the transitions from each state $\mathbf{n} \in \mathbb{N}^s$ if the control action $\mathbf{u}$ is applied. The action with the highest score will identify the optimized control strategy μ .

Score functions for strategies μ will be denoted as J_∞^μ , i.e. the expected score obtained over an infinite time horizon when the system is controlled using the

control policy μ : $\mathbf{u}(t) = \mu(\mathbf{n}(t))$, $\forall t$. For a given initial condition $\mathbf{n}_0$, J_∞^μ is given by:

$$J_\infty^\mu(\mathbf{n}) = E \left[\int_{t=0}^{\infty} \gamma^t r(\mathbf{n}, \mu(\mathbf{n}(t))) dt \middle| \mathbf{n}_0 = \mathbf{n} \right] \quad (5.2)$$

where γ is a discount factor ($0 \leq \gamma < 1$) that weights short-term rewards more than long-term ones, and where the conditional expectation is taken over all trajectories starting with the initial condition $\mathbf{n}_0$. We target to find an approximation of the optimal policy μ^* , which maximizes the expectation J_∞^μ for all system states in the considered hypercube $\mathbf{n} \in \mathbb{N}^s$.

5.2.1 Fitted-Q - the reinforcement learning algorithm

The control policy as well as the overall reward function Eq. 5.2 can be induced numerically from a trainings set $\mathcal{F}$, which contains time series observations and control signals. The time series should be taken in equidistant time intervals and record the so called one-step transitions - system measurements containing the system state $\mathbf{n}(t)$ at the time t , the control action $\mathbf{u}(t)$, the system state at the next time step $\mathbf{n}(t + \Delta t)$ and the instantaneous reward according to the bounded reward function $r(\mathbf{n}(t), \mathbf{u}(t))$. These variables will form the input four-tuple $((\mathbf{n}(t), \mathbf{u}(t), r, \mathbf{n}(t + \Delta t)))$ and the trainings set $\mathcal{F}$ consists of such tuples.

From this measurement set $\mathcal{F}$, the fitted Q iteration algorithm computes a converging sequence of functions $\hat{Q}_1, \hat{Q}_2, \dots, \hat{Q}_N$ (see Fig. 5.2 for the pseudo code of the fitted Q algorithm). These functions are effectively system state dependent time integrations of the instantaneous rewards r discounted with a factor γ and they approximate the cumulative reward function J as defined in Eq. 5.2 above.

Thus, the objective of the fitted-Q iteration algorithm is to find an approximation of the optimal strategy μ^* , i.e. a strategy that maximises J^μ for all system states without knowledge of the dynamical equations. The only information available to the fitted Q iteration algorithm is contained in the trainings set $\mathcal{F}$ of one step transitions and the reward function $r(\mathbf{n}, \mathbf{u})$.

The deduction of the series of Q functions from the data set can be formulated as iterative solution of batch-mode supervised learning problems. A standard supervised learning problem is the induction of a rule or a function from input/output relationships. For the fitted Q the input consists of the four tuples

Figure 5.2: Fitted Q iteration algorithm

Inputs: a set of four-tuples $\mathcal{F}$ and a regression algorithm.

Initialization:

Set N to 0.

Let $\hat{Q}_N$ be a function equal to zero everywhere on $X \times U$.

Iterations:

Repeat until stopping conditions are reached

- $N \leftarrow N + 1$.

- Build the training set $\mathcal{TS} = (i^l, o^l), l = 1, \dots, \#\mathcal{F}$ based on the the function $\hat{Q}_{N-1}$ and on the full set of four-tuples $\mathcal{F}$:

$$i^l = (\mathbf{n}^l, \mathbf{u}^l), \quad (5.3)$$

$$o^l = E_{\mathbf{n}' \propto f(\mathbf{n}, \mathbf{u}, \cdot)} \left[\tilde{\rho}(\mathbf{n}^l, \mathbf{u}^l, \mathbf{n}') + \gamma \max_{\mathbf{u}' \in U} \hat{Q}_{N-1}(\mathbf{n}', \mathbf{u}') \right]. \quad (5.4)$$

- Use a noise robust regression algorithm to induce from $\mathcal{TS}$ the function $\hat{Q}_N(\mathbf{n}, \mathbf{u})$.

$\{\dots; (\mathbf{n}(t), \mathbf{u}(t), r(\mathbf{n}, \mathbf{u}), \mathbf{n}(t + \Delta t)); \dots\}$ and the output is the estimated reward o^l , which changes after each algorithmic step.

The fitted Q -iteration starts by approximating the first Q -function in the sequence $\hat{Q}_1(\mathbf{n}, \mathbf{u}) \equiv E[r(\mathbf{n}, \mathbf{u})] = \tilde{\rho}$ by averaging of the instantaneous rewards in $\mathcal{F}$ over the different states $\mathbf{n}(t + \Delta t)$, which have originated from the same $\mathbf{n}(t)$. In the case of the purely deterministic description no averaging is necessary, for the SDE's and Master equations, which stochastic characteristics are well approximated with a white noise on the small enough time increments Δt , we would effectively average the Gaussian noise away. The following Q -functions in the sequence are estimated from the training sample $\mathcal{TS}$, which is constructed from the input trainings set $\mathcal{F}$ and updated rewards with the following equation:

$$\mathcal{TS} = \left\{ \left((\mathbf{n}^l, \mathbf{u}^l), \tilde{\rho}(\mathbf{n}^l, \mathbf{u}^l, \mathbf{n}') + \gamma \max_{\mathbf{u}' \in U} \hat{Q}_{k-1}(\mathbf{n}(t + \Delta t)^l, \mathbf{u}') \right) \right\}_{l=1}^{\#\mathcal{F}}$$

In this way a sequence of function $\hat{Q}_k$ is generated and the outputs o^l from the step k are used for the regression of the next function in the sequence.

For the properly defined control problems the function sequence $\hat{Q}_1, \hat{Q}_2, \dots, \hat{Q}_N$ converges to the stationary scores J^μ and then the approximated control strategy can be inferred by taking the control action $\mathbf{u} \in U$ at the time point t , which has

the maximum approximated reward $\hat{Q}_N$:

$$\hat{\mu}_N^*(\mathbf{n}) = \arg \max_{\mathbf{u} \in U} \hat{Q}_N(\mathbf{n}, \mathbf{u}) \quad (5.5)$$

$\hat{\mu}_N^*(\mathbf{n})$ is by definition the approximation of the optimal control strategy $\mu^*(\mathbf{n})$.

Number of iterations and algorithmic convergence We iterate the algorithm until no significant changes in the Q score function appear. This can be assessed either with the Bellman residual as described in Ernst et al. (126) or other metrics. As it is the case in many numerical procedures an appropriate threshold value as the stopping condition needs to be chosen. Our results are not strongly dependent on the number of the iterations and the precise threshold value.

Theoretical conditions for the convergence Theoretical existence of the optimal control policy was addressed in classical dynamic programming theory. The optimal strategy μ^* can in principle be induced by analytical solution of the classical Bellman equation (see (129)), but this is only possible if the dynamics of the system are known in detail, for example in form of an ODE, SDE, or a Master equation and the minimization/maximization needed for the inference of the optimal control strategy can be done analytically.

However, the precise information in form of equations is not available for the most biological problems and therefore we have chosen here a model-free approach. We assume that there are time series measurements from which the required input in form of four-tuples can be constructed, but we do not assume the knowledge of a model.

To induce the function from these input/output relationships we need to use a robust and computationally cheap regressor, which is able to handle hundred thousands of input vectors during several hundreds of iterations. Decision trees are considered to be the most promising candidates for their computationally cheapness. We will use a particular type of decision trees, which are computationally effective and also have a parameter to tune the robustness to noise.

5.2.2 Effective multi-dimensional regressor - extremely randomized trees

In our implementation we have used EXTremely RAnimized Trees (Extra-Trees) (127) as the regression method, which has been shown to perform in several regression and classification problems better than the classical tree bagging (130), random forests (131), or single CART regressors (132).

In general, tree-based regression methods recursively part data sets with the goal to construct a predictive model or a multi-dimensional function approximator in form of binary decision trees. The partitioning of the data needs to be done in an informative manner. Intuitively the most informative partitioning is achieved if similar elements are grouped within a tree branch, which means the variance between the tree branches is maximized (see Fig. 5.3). The earliest method CART (132) does exactly that, it constructs a single decision tree from the data recursively partitioning the set and selecting the splits in the most informative way, which means maximizing the variance. A data point can be described by more than one characteristic or attribute, as for example each variable is an attribute for a multi-dimensional function. In the classical CART algorithm the splits are chosen to maximize the variance given all attributes. The algorithm has one variable: n_{min} the minimal number of items in the leaf and the number is chosen larger if there is a significant amount of noise in the data. The main disadvantage of the method is that it is too rigid and error-prone if there are errors in the data.

Random forests are in some cases less error prone, although it is a somewhat unintuitive way to construct a model. Here several decision trees are constructed randomly selecting both the attribute as well as the cut point (131). The prediction is made via the averaging of all the trees in the forest. In addition to the minimal number of the elements in the leaf n_{min} as in the CART method, random forests have a second algorithmic variable namely the number of trees M in the forest from which the average will be taken. Random forests are not always good approximators, and might even become unstable, probably because there is too much randomness in the method.

The Extremely Randomized Trees used here as a function approximator can be seen as a tunable intermediate between the purely deterministic CART method

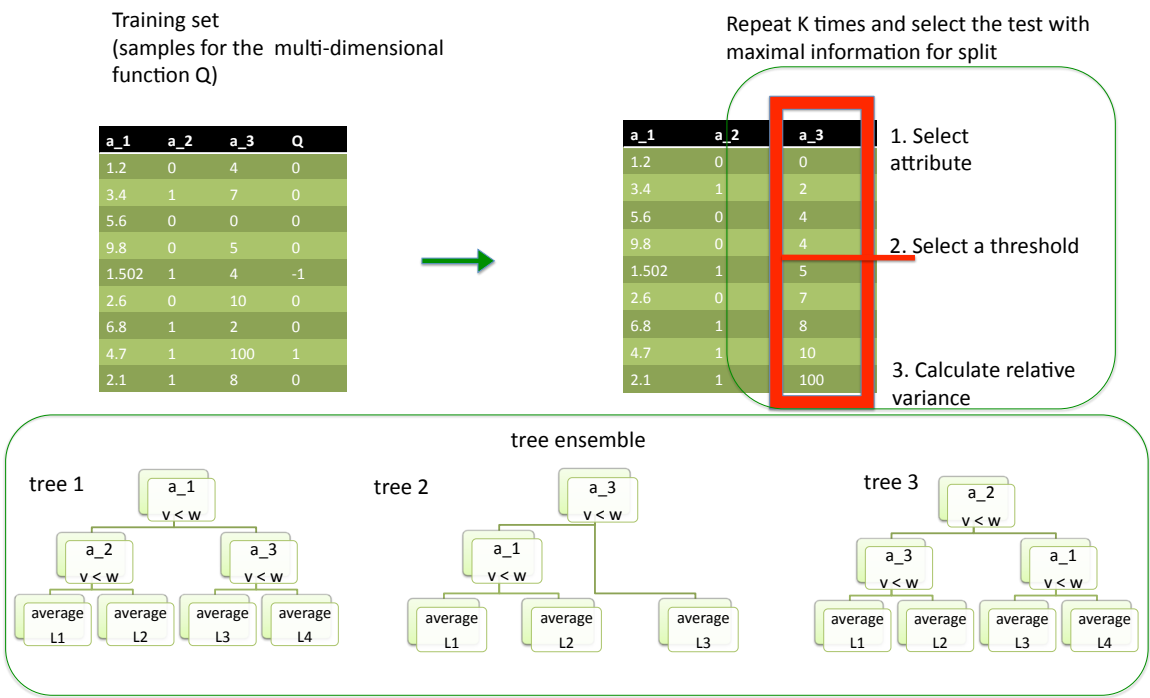


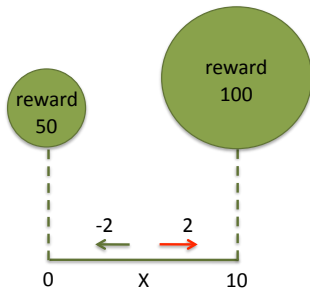
Figure 5.3: Illustration of the regression method “EXTremely Randomized Trees (Extra Trees)”.

and random forests (see Fig. 5.3). In order to adjust the randomness according to the problem a third parameter K is introduced. At each step K cut-points are assessed regarding the variance, which the split would introduce. The cut which maximizes the split variance, which means is the most informative one, is assigned to a node. The number of trials K is the tuning parameter between the classical CART and the random forests. If K is set to one, then we are in the regime of random forests, if it is set to ∞ we are in the CART regime. The main advantage of using this regressor for biological problems and in particular gene expression is that we can adjust K according to the credibility we have about the data and the experiment.

5.2.3 Classical control problems

There are classical control examples, which are used to illustrate the general purpose of the control technique and validate or benchmark an algorithmic implementation. Two typical examples are "Left-Right" and the "Car on the Hill" problem, which we also will be using here. The "Left-Right" problem is intuitively illustrative, relatively easy to learn, and highly stochastic. The "Car on the Hill" problem is less intuitive, deterministic, and the numerical derivation of the optimized policy is considered to be difficult, because the success rate in the trainings set is extremely low.

Classical control problem – Left-Right



A simple example to illustrate the control purpose in classical engineering is a Brownian motion of a particle confined between two absorbing boundaries $x \in [0, 10]$. The particle can be moved to the left or to the right by application of discrete control signals $u \in U = \{-2, 2\}$. If it crosses the left boundary a reward of a value 50 is given, if it crosses the right boundary a reward of value 100 is given. The dynamics of the particle can be described via a simple SDE Eq. 5.6. So the aim of the

control policy is to direct the stochastic particle to one of the boundaries within a given time interval or within a constraint finite time horizon maximizing the reward.

$$x_t = x_0 + \int_0^t d\omega_s + \int_0^t u(s)ds \quad (5.6)$$

with $d\omega_s$ Gaussian white noise.

Intuitively we expect that reward within a finite time horizon is maximized directing the particle to the left boundary if it has already started close to the left boundary. And if the particle starts somewhere in the middle then it would be better to control it to the right boundary associated with the larger reward. We expect a threshold rule as the optimized control strategy, but which value is the threshold value?

For this simple example the algorithm will produce a control strategy identi-

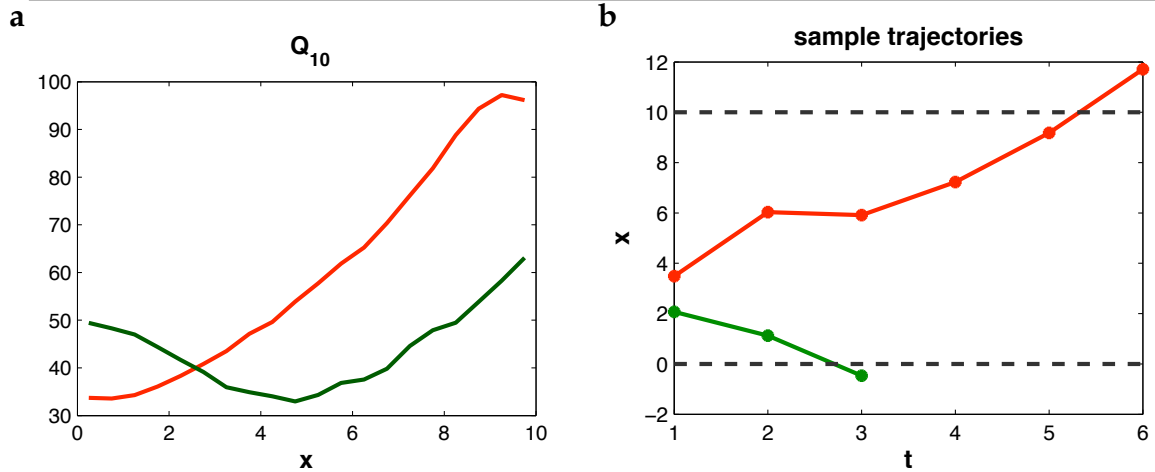


Figure 5.4: **Algorithmic estimations of the optimal control for the benchmark "Left-Right"**. **a)** Numerical estimation of the optimal control policy with fitted-Q algorithm obtained after 10 iterations. The score for left kick is shown in green for the right in red. The estimated control policy is state dependent and is obtained for each x by taking the action with the largest Q - score. The lines cross at the expected threshold value $x=2.7$. **b)** Two trajectories starting at the initial conditions below (green) and above (red) the threshold, which are controlled using the estimated policy shown in a). The absorbing boundaries at $x=0$ and $x=10$ are shown in gray dashed lines.

fying the most suitable threshold value given the data set. It turns out it is 2.7 for the discounting factor $\gamma=0.75$. Even though this is just a one-dimensional problem and we can almost guess the solution, the learning of the threshold rule is impeded by the noise. Therefore this example is often chosen as a benchmark to prove both that the proposed algorithm can learn under noise and that the implementation of the algorithm is correct. The results of our algorithmic implementation on this classical example are shown in Fig. 5.4 and demonstrate both that the fitted-Q is a suitable framework for learning under noise and that the implementation is correct.

Classical control problem – Car on the Hill

Another classical problem often used to illustrate the purpose of optimal control in engineering is the "Car on the Hill". Consider a landscape as illustrated in Fig. 5.5 containing a dip between a hill and a sheer and a car positioned at the minimum. The control aim is to direct the car to the top of the hill without letting it fall over the sheer by applying discrete control signals $u \in U = \{-4, 4\}$ "left acceleration" or "right acceleration". Intuitively, it is clear that if we just applied

the signals at random then the probability of falling down the sheer is by far higher than reaching the top of the hill, because the hill is higher. Therefore we expect that the estimation of the optimal strategy will be hard to induce due to low success rate in the trainings set. And indeed the trainings set containing 60 000 observations had only 18 successes, where the reward 1 was assigned, because the car achieved the top of the hill.

The dynamics are described by the ODE Eq. 5.7 on the support $(p, s) \in [-1, 1] \times [-3, 3]$:

$$\begin{aligned} d_t p &= s \\ d_t s &= \frac{u}{m(1 + h'(p)^2)} - \frac{h'(p)g}{1 + h'(p)^2} - \frac{s^2 h'(p)h''(p)}{1 + h'(p)^2}. \end{aligned} \quad (5.7)$$

with the hill potential $h(p)$

$$h(p) = \begin{cases} p^2 + p & \text{if } p < 0 \\ \frac{p}{\sqrt{1+5p^2}} & \text{if } p \geq 0. \end{cases}$$

The instantaneous reward function is defined on the boundaries: the hill top $p \geq 1$ & $s \leq 3$ = 1 and on the sheer $(p \leq -1, \cdot) = -1$ and is zero everywhere else on the support $(p, s) \in [-1, 1] \times [-3, 3]$.

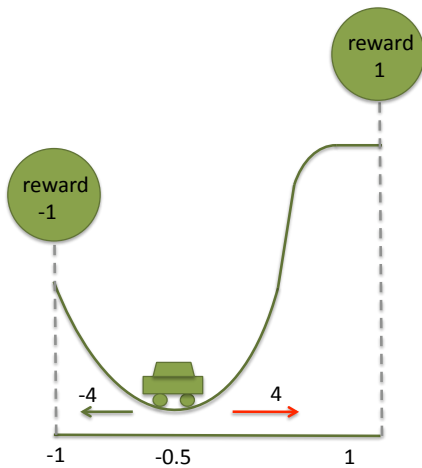


Figure 5.5: Classical control problem "Car on the Hill".

Figure 5.6 shows the algorithmic output of this control problem. The algorithmically optimized trajectory 5.6 c) has an interesting feature: the optimized use of the applied signals is achieved by a combination of the intuitively obvious acceleration towards the maximum point $u = 4$ and the less intuitive contra force $u = -4$. Energetically it is favorable to accelerate towards the other end in order to take advantage of the gravitational force (second term in Eq. 5.7). This is one of the useful characteristics of the approach that the estimated policy takes advantage of system's inherent characteristic

although the algorithmic input does not contain the equations of motion or any kind of explicit information on this force, but only sample trajectories.

This is also the reason why we want to apply the algorithm to a biologically motivated examples, since in many biological applications only limited information is directly accessible and inherent characteristics are often unknown. This approach would take advantage of inherent systems dynamics using time trajectories of the key variables and control signals.

5.3 Application on biological examples

We formulate and investigate biologically motivated control problems, which to the best of our knowledge have not been considered as control-theoretical examples before. We start with the optimization of the strategy for the switch reversal, where the control actions are the light pulses activating a light-sensitive promoter.

One of the major differences between the classical engineering examples such as the "Car on the Hill"-benchmark and typical biological examples are qualitative characteristics of possible control signals. While for the "Car on the Hill" the control signals ($u \in \{4, -4\}$) can explore the state space in both directions, the control signals for gene regulation given by the current state of biotechnology can only increase the concentration of a protein, but can not *in vivo* decrease a concentration of a protein in a targeted manner. This produces two additional complications. On the one hand we explore the state space only in one direction and on the other hand the algorithm has to differentiate between the signals "no change"/"increase" and in particular under noise this difference is less obvious compared to the classical "increase"/"decrease" alternative.

In general however, successful learning outcome depends solely on the Markov property of the underlying process (see previous section). Therefore the optimization of the control strategy for noisy and restricted biological examples is possible but likely to be more demanding compared to the classical problems i. e. requiring larger training sets. Also the formulation "optimal strategy" in the classical sense can not be used then speaking about biological models, because in most cases the system description is incomplete and it is difficult to judge if the numerically inferred strategy is optimal in the classical sense.

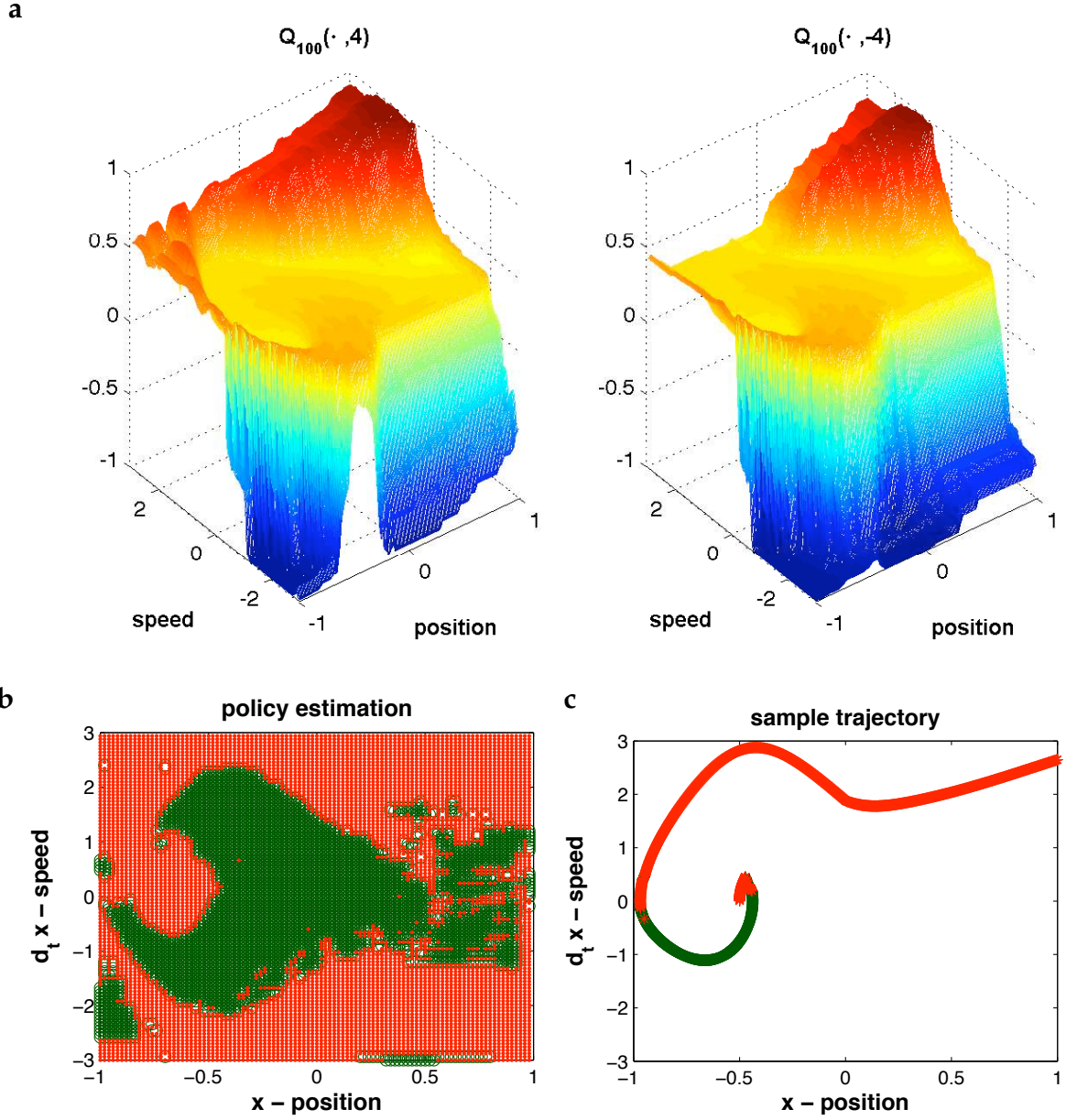
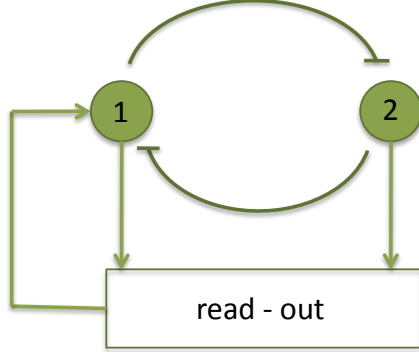


Figure 5.6: Algorithmic estimations of the optimal control for the benchmark “Car on the Hill”. **a)** Estimation of the Q - score function for the pair $(x, d_t x)$ position and speed obtained after 100 iterations. **b)** Induced state dependent optimized control policy (Eq. 5.5). The green markers indicate states for which Q - score for control signal $u = -4$ is larger than for $u = 4$, which means the set $\{Q(\cdot, -4) > Q(\cdot, 4)\}$. The complementary set within the boundary conditions is shown in red. **c)** A trajectory for initial condition $(-0.5, 0)$, where the car starts the motion at the potential minimum and at rest. It is controlled using the algorithmically estimated policy to the top of the hill respecting the boundary conditions $x \in [-1, 1]$ & $d_t x \in [-3, 3]$. The trajectory is plotted in red, whenever the control signal $u = 4$ and in dark green whenever $u = -4$ is applied.

5.3.1 Controlled reversal of a genetic switch



Switches are typical and important constructs for the forward designed (4) as well as for the naturally occurring circuits. One example, where cascades of switches occur and are believed to dominate the dynamics are antagonistic switches in developmental biology (133, 134). The cell fate is believed to be determined by successive switch reversals of the phenotype dominating genes. For the stem cells tissue engineering intervention protocols

are desired, which would direct the cells into a particular fate both fast and without protein over-expression, which reduces the fitness of the cells. An optimized strategy for switch reversals avoiding over-expression can be sought using the reinforcement learning algorithm.

We use a simple design for a switch to demonstrate the formulation of this biological problem in control theoretical terms. A classical topology for a biological switch are two genes mutually repressing each other (figure above). If the repression is strong enough then this system will show bi-stability (5, 70), which means in the steady state either the first gene is highly expressed and the second gene is suppressed ($p_1 \uparrow, p_2 \downarrow$) or vice versa ($p_2 \uparrow, p_1 \downarrow$). (see chapters 2, 3 for the bifurcation analysis of such systems.)

A Markov process for a simple switch can be formulated as a Langevin equation based on the basic mass-action kinetics and using time scale separation between the slow protein and fast mRNA dynamics (see for instance (111, 116)):

$$d_t p_1 = g_1(p_2) - d_1 p_1 + \sigma_1 d\omega_1 + q_s u \quad (5.8)$$

$$d_t p_2 = g_2(p_1) - d_2 p_2 + \sigma_2 d\omega_2 \quad (5.9)$$

where $g_1(x) = \frac{e_x}{1+x^2}$ $g_2(x) = \frac{e_y}{1+x^2}$ and ω terms correspond to the white noise increment. The order of magnitude for the constants is based on Gardner et al. (1), Guantes and Poyatos (116) and to illustrate the procedure we assume a symmetric switch $e_y = e_x = 400, d_1 = d_2 = 1, \sigma_1 = \sigma_2 = 0.005$. The number of proteins per unit time produced as the result of the light control signal is denoted by q_s

and u as usually denotes the control actions $u \in \{0, 1\}$.

Formal definition of the control problem for a switch reversal We assume that for both genes the protein concentrations are given as readouts for instance via fluorescent markers. The initial condition is the vicinity of the fixed point $(p_1 \downarrow, p_2 \uparrow)$, where gene 1 is silent and gene 2 is expressed. The goal of the control strategy is to direct the switch to the second fixed point $(p_1 \uparrow, p_2 \downarrow)$ by applying control signals to the first gene. The control signals increase the concentration of the first gene in small increments, which biochemically corresponds to light flashes activating a photo-sensitive promoter. Biochemically the response occurs on fast time scales (order of seconds (28)) compared to the gene expression time scales occurring on the order of minutes or even hours and therefore we can approximate the action of the light inducer as a discrete set light on and light off $U = \{0, 1\}$. In classical control theory this type of control is often referred to as “bang bang”.

To create the training set we start from random initial conditions and the “light on” signals $u(t) = 1$ are applied with the probability 3% covering the interval of interest $[0, 400] \times [0, 400]$. The instantaneous reward function is defined in the following way: Two boundary conditions are used to specify the minimal and the maximal boundary for the instantaneous reward function. The first absorbing boundary condition is given by the minimal reward -1, which is assigned to the neighborhood of fixed point $B(p_1 \downarrow, p_2 \uparrow)$, which is the initial state we aim to reverse. By absorbing in this context we mean that if the system reaches that state after a series of random signal applications, we assign the according instantaneous reward to the last point and reset the simulation from random initial conditions. The second absorbing boundary with the maximal reward 1 is assigned to the neighborhood of the fixed point $(p_1 \uparrow, p_2 \downarrow)$, which we are trying to achieve. Each application of the control signal is associated with a cost for the cell to express the foreign protein and therefore the instantaneous reward is always diminished by “light punishment” $q_l = 0.005$. In summary the reward function is:

$$r(\mathbf{p}, u, t) = \begin{cases} -1 & \text{if } \mathbf{p} \in B(p_1 \downarrow, p_2 \uparrow) \\ 1 & \text{if } \mathbf{p} \in B(p_1 \uparrow, p_2 \downarrow) \\ r(\mathbf{p}, 0, t) - q_l & \text{if } u(t) = 1 \end{cases} \quad (5.10)$$

Algorithmic outcomes and interpretation. The intuitive outcome for a single switch is a threshold rule: if the concentration of the up gene is low then apply induction signal. However, the main disadvantage of this technique is that it does not take into account the system dynamics and the threshold value is not clear a priori. In the pioneering synthetic biology implementation by Gardner et al. (1) (revised in section 1.4) the chemical inducer IPTG has been applied during the whole switching time of 6 hours to the bacterial population. This technique can not be applied if not just one but several switches need to be reversed.

Illustrating our approach we have induced an optimized control strategy for the switch reversal, which should avoid over-expression of the first protein (see Fig. 5.7). Indeed if starting from initial conditions where the first gene is down and the second is up as shown in the picture only few kicks are required in order to reverse the switch. From the input-output relationships contained in the training time series the reinforcement learning technique deduced a control strategy, which takes advantage of the internal system dynamics.

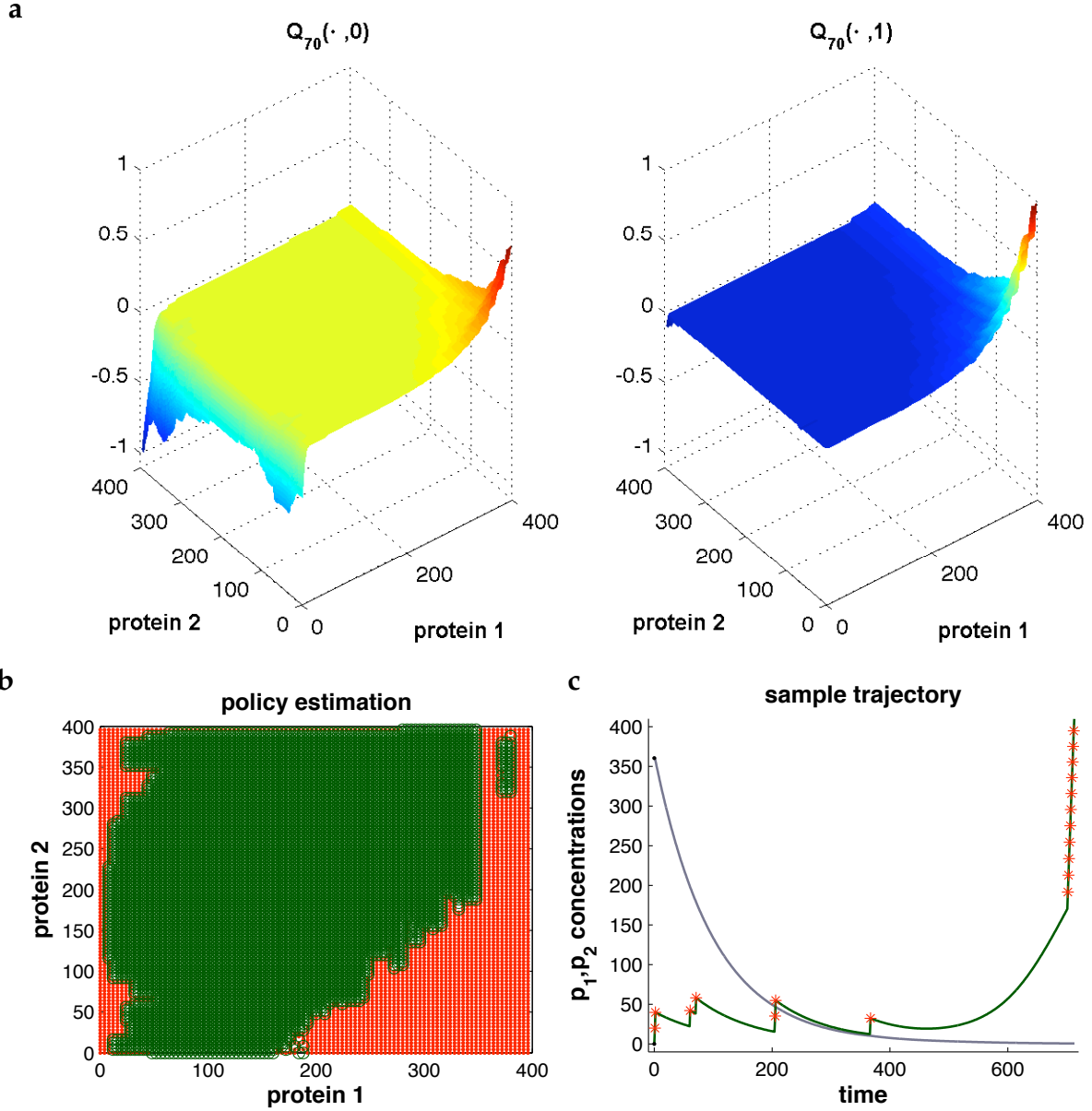
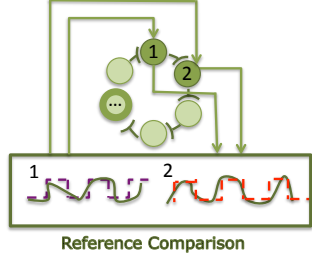


Figure 5.7: Optimized reversal of a genetic switch. **a)** Estimation of the Q - score function for the pair (p_1, p_2) protein concentrations obtained after 70 iterations ². **b)** Induced state dependent control policy (Eq. 5.5). The red markers indicate the set $\{Q(\cdot, 1) > Q(\cdot, 0)\}$, which means genetic network states, where the value for the action "apply control" $u = 1$ is larger than for "do not apply control" for $u = 0$. The complementary set within the boundary conditions is shown in green. **c)** Simulated trajectories for $p_1(t)$ (green) and $p_2(t)$ (gray) starting from $B(p_1 \downarrow, p_2 \uparrow)$ (black dots) is controlled using the algorithmically estimated policy shown in **(b)** towards the desired state $(p_1 \uparrow, p_2 \downarrow)$. System states for which the control action has been applied ($u = 1$), are marked by red stars. **Note:** Conducting the switch to the desired state with this algorithmic induced policy avoids over-expression. In comparison to the induced policy, naive or blind application of the control action on the first protein and waiting till the second protein will reach smaller concentration would result in 1000-fold over-expression and therefore lead to unnecessary cell burden.

5.3.2 Controlling even gene rings around unstable periodic orbits



Another area, where model-free feedback control can be applied is the control around predefined trajectories. To showcase this we use a UPO due to the first Hopf bifurcation in the generalized repressilator model with 10 genes, which we have previously characterized numerically (see chapter 2, chapter 3) and analytically 3.2. Even rings act in the steady state as switches, but possess unstable periodic orbits (UPOs), which give rise to

long-lived oscillating transients. In section 2.4.4 we have introduced an intuitive control scheme in stochastic regime which uses a simple threshold rule to keep even rings in oscillation around these UPOs. From the intuitive implementation we realized that if we observe and control only 2 genes, the whole ring can be kept in oscillating state. We use this prior knowledge and formulate the control-theoretical problem - tracking a 10 gene ring around its UPO associated to the first Hopf bifurcation.

Formal definition of the control problem for phase tracing We simulate the generalized repressilator as an SDE using the standard Euler-Maryama, taking a small noise term $\sigma_i \approx 0.01$ and using otherwise the same set of constants as in chapter 2:

$$\begin{aligned}\dot{m}_j &= \frac{c_1}{1 + p_{j-1}^2} - c_2 m_j + \sigma_m d\omega_1 + \delta_{j1} u_1 q_{s_1} + \delta_{j2} u_2 q_{s_2} \\ \dot{p}_j &= c_3 m_j - c_4 p_j + \sigma_p d\omega_2.\end{aligned}\tag{5.11}$$

with the usual definition for δ_{ij} as Kronecker symbol (20). The control signals $\mathbf{u} \in \{u_1, u_2, 0\}$ are applied on the *mRNA* level as already indicated in the problem definition picture above only to the first u_1 and the second u_2 gene. The influence of the light signals on the mRNA production is denoted by q_{s_i} .

Unlike the switch example, the aim of the control strategy for the ring is not to reach a particular target state, but to keep oscillating pattern following a predefined path. This problem is computationally quite demanding and we divide it into two parts. The first part is starting from the stable up/down solution

direct the ring to the vicinity of the UPO, which can essentially be formulated as control towards the nearest point on the UPO and we have seen this type of control problems in the previous section. But here we want to demonstrate a solution of an operation around an unstable periodic orbit, and therefore we start in the vicinity of the UPO and the control aim will be to follow the oscillating trajectory. Since the periodic orbit is unstable we know that the ring will dephase and slowly converge to the stable up/down solution. Therefore, the objective of the control strategy will be to prevent the slow oscillation decay and keep the ring in the oscillating state.

We approximate the predefined oscillating pattern with a nonlinear curve and a phase. In practical applications this pattern can be induced after repeated single cell experiments and then averaging the time trajectories. The reward function can then be defined as a simple distance function between the observed state $\mathbf{p}(t) = (p_1(t), p_2(t))$ and the reference trajectory $\mathbf{p}_0(t)$:

$$r(\mathbf{p}, \mathbf{u}, t) = \exp(-|\mathbf{p}(t) - \mathbf{p}_0(t)|) - \sum_i q_i u_i(t) \quad (5.12)$$

As for the switch the cost for light application q_l is applied if the first (u_1) or the second (u_2) gene is induced via a light flash. **Note** that the read-out for the system contains only the state of the first two proteins, other variable such as *mRNA* and other proteins are not part of the algorithmic input.

Technical discussion of the algorithmic realization A solution for an operation around an unstable periodic orbit has not been tried before with such algorithms. Therefore we spend a couple of lines to discuss the technical details for control of these types of problems with the fitted-Q algorithm. Obviously the extension to general tracking problems from the control towards or around a fixed point is straight forward. A time coordinate for the trajectories has to be added to define the should trajectory or a phase as in our case and this information should be part of the trainings set. The reward function can be based on the distance between the perfect trajectory and the observation during the training phase. An appropriate metric for the distance measure must be bounded and can be adjusted to the noise and measurement uncertainties in the problem and an outlier robust alternative could be the metric in Eq. 5.12.

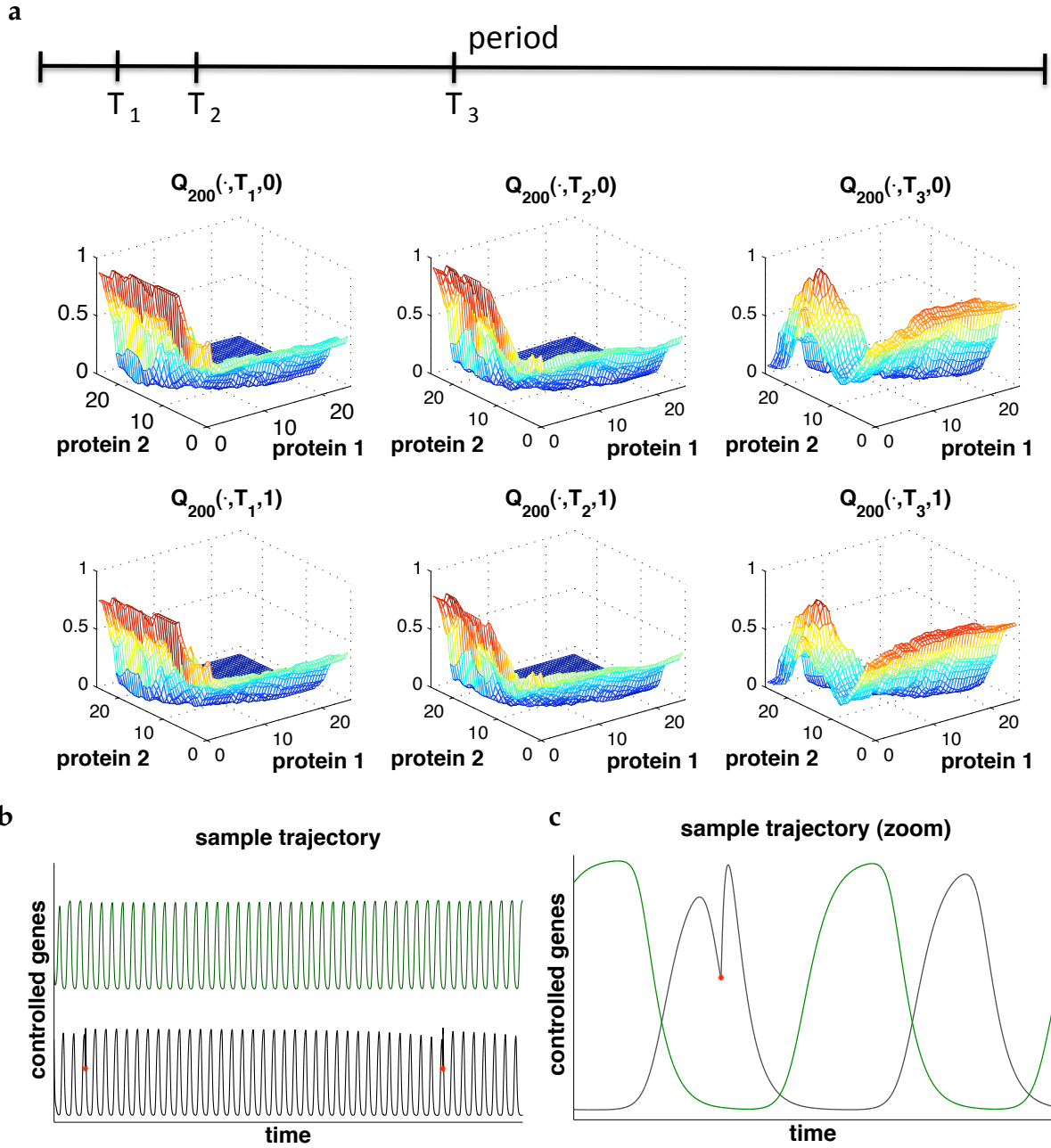


Figure 5.8: **Optimized tracking of an unstable periodic orbit in the generalized repressilator model** **a)** Estimation of the Q -score function for two proteins in the ring (p_1, p_2) obtained after 200 iterations. We show here the time snapshots of two points (T_1, T_2) lying closely in the periodic time space and the third (T_3) lying further apart. As expected the Q function for close points is similar in shape, but changes throughout the period to adjust for time specific control signals. **b)** A trajectory started from random initial conditions, for which the control signals have been applied twice (marked with red stars). We have defined the control constants to show that the control action is taken to prevent the ring from dephasing and approaching the stationary up/down solution (the dynamical features of even rings are discussed in chapters 3, 2.) **c)** zoom into the trajectory shown in b) around the second control signal.

We have specifically defined the cost for the light application very high in this example, to demonstrate that the ring will slowly dephase due to its inherent dynamics, but the algorithm will sparsely take appropriate actions working against that. With a smaller cost for the light application the dephasing can be made unnoticeable and the oscillation pattern perfect.

The induced function Q is now explicitly dependent on time points in the period. For smooth trajectories and well behaved nonlinear problems (which we usually have in nature) we expect that the function Q will vary smoothly between the neighbouring points in the trajectory. This is also what we are observing for our induced Q scores. We show the function Q for two points lying close to each other and compare to a distant point. As expected the close trajectory points have a similarly shaped Q function, which shape differs significantly for the distant point Fig. 5.8.

5.4 Discussion

We have presented a concept for a model-free controller design, which can efficiently handle nonlinear and stochastic *in silico* gene regulatory networks. The objective of the formulated feedback control is to reduce the burdening of the cells, which arises if heterologous proteins are produced in larger amounts than necessary. Therefore we have formulated a control theoretical optimization problem inducing the gene expression just enough for switch reversal in the first problem and just enough to keep the ring oscillating.

Using the algorithm for the control of the gene ring around a UPO we show that the formalism is in principle able to handle hidden dependent variables and therefore of a potential interest in practical applications. In our test network the variables are dependent and performing control signals on upstream genes has a predictable effect on the overall network. In this case a ring of ten genes could be kept in the oscillating state by reading and influencing only two genes.

Even though our example is far from topologies found in nature, natural biological systems are thought to be organized in a similar manner as cascades and feedback loops (58). If for a cascade or a feedback loop we can control a key gene then we can deduce interference protocols with our algorithm. An exhaustive knowledge about the whole gene network is not required. The power of the

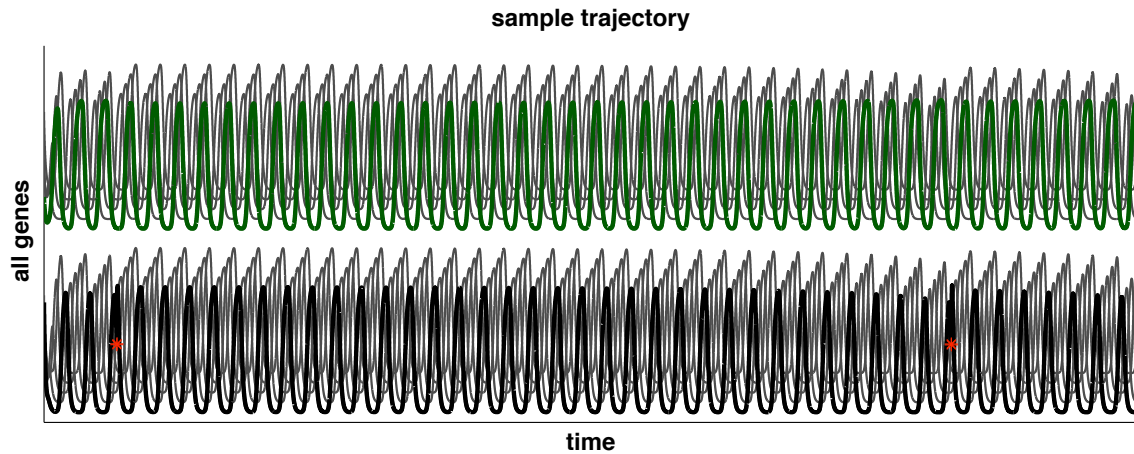


Figure 5.9: **Control of 2 genes keeps a 10 gene ring in the oscillating state around an unstable periodic orbit.** The control procedure has been designed for a gene ring with 10 genes, but we have used only the time traces of the first 2 genes for the policy induction. The genes which time evolution has been used in the trainings set are shown in thick colored lines, the time traces of the dependent genes inaccessible to the algorithm are shown in gray. The simplified example demonstrates that partially observable processes can be handled by the algorithm given that there is prior knowledge about hierarchical network organization.

algorithm drives the imagination that even a cell phenotype change in difficult problems such as for example stem cells specification (134) can be manipulated by imposing control actions on gene master regulators.

The algorithm was shown to deduce optimized control strategies for synthetic biology examples, which implementation is feasible due to progress in synthetic biology. In particular a construction of the switch and its controlled reversal can be a potentially interesting and feasible experiment. We would then compare switch reversals under constant induction as has been done by (1) with the optimized switch reversal deduced with our algorithm using the experimental time traces. The cell burden under both conditions can be assessed for instance via growth rate or other phenotypical changes of the cell morphology, such as the symmetry of the daughter cells, ribosomal usage etc. If experimentally a gain for the cells can be shown for a single synthetic switch, then a natural switch determining the cell phenotype, as for instance an antagonistic red-white blood cell (134) can be addressed opening a new perspective for tissue engineering and at the moment still rather imaginative path reversal for medical applications.

A potential interest of this algorithm in direct application to experiments is supported by the fact that the approach is model-free. Feedback control has

earlier been explored to some extent for Bayesian network genetic models (135–137). However, the earlier proposed approaches were formulated for specific model set ups, by (135, 136) as Bayesian networks implicitly assuming that only coarse information (microarray measurements) on the gene network is available. Our approach on the other hand is more flexible, since it would adapt to the training set accuracy and would handle information given by either coarse or more quantitative time traces in the best possible way.

Considering purely technical aspects of our fitted Q algorithm, we remark that it can be extended to the adaptive version. The supervised learning at each Q iteration uses the whole data set for the regression and only the output values (the current cumulative reward) change. It means that the successive calls to the supervised learning algorithm are independent of each other. Hence, at each step it is possible to replace small fractions of the trainings set with new real-time observations and extend the existing batch-mode algorithm to an adaptive version. Adaptivity for biological problems is of particular interest, because cell populations provide changing environments for genetic networks. Our formulation of the optimized control problem can account for changes occurring on slow, compared to the system dynamics, time scales. Then the information of the control strategy induced earlier would still be useful for the slightly altered environment. Therefore fitted Q is well suited to handle a slow biological adaptation process.

In summary, even though biological systems would never be as exactly controllable as their classical engineering counterparts as for instance flight control in jets (54), nuclear reaction shutdown (129) or a class of problems known as “games against nature” (129) with appropriate algorithmic techniques the major challenges of gene network uncertainty, hidden variables, and non-linearity can be tackled with these concepts. A potential application area, where we are planning to apply this concept is the tissue engineering in stem cells. And the toy example of the switch is expected to be helpful for reversals of antagonistic switches, which we are planning to use for direction of stem cells to cardiomyotic fate (61).

5.5 Appendix: constants for algorithmic realization

For the reference algorithmic constants for all examples are summarized in the table below:

	γ	K	$TS^\#$	N_{min}	# Trees	# It
Left - Right	0.75	2*	2000	10	25	10
Car on the Hill	0.95	3*	60000	2	25	100
switch	0.75	3	120000	3	25	70
gene ring	0.99	4	240000	3	25	200

Chapter 6

Summary and perspectives

The generalized repressilator model has been used throughout the thesis to exemplify design concepts and methods for synthetic biology. We started with the numerical bifurcation analysis and the characterization of transient dynamics originating from unstable periodic orbits (UPOs) in the generalized repressilator model. The importance of transient dynamics in genetic networks has been emphasized with the UPOs and a design of a controllable genetic oscillator around such orbits has been suggested. In the following we have used symmetry and topology arguments to analytically find the conditions for Hopf bifurcations leading to oscillatory behavior in the deterministic generalized repressilator model. Then the stochastic aspects of periodic orbits in the model have been explored. Finally, the accumulated knowledge about the system was used to showcase the formulation of biological problems as control-theoretical examples and algorithmic induction of optimized control policies.

6.1 Synthetic engineering and oscillator design

In the thesis chapter 2 a design for a controllable genetic oscillator was presented. The oscillator operates around an unstable periodic orbit with a feedback control loop, which switches between the oscillatory and fixed point solution. Operation around an unstable periodic orbit is a different mechanism than previously proposed autonomous oscillators (2, 3) operating around stable limit cycles. Autonomous devices are designed to robustly perform single functions, but our design presents a possibility to robustly switch between oscillations and fixed point

solution. Our design is also fundamentally different from the Atkinson et al. (74) proposal, which biochemically switches between oscillatory and switch-like stable states via modification of the gene network connectivity. The design excites and maintains the oscillations around an unstable periodic orbit using a light-based feedback control mechanism (Sec. 2.4.4) and does not change the network topology. Our mode of operation provides full control over the process secured with a feedback loop and therefore could be desirable for applications, where toxicity or virulence requires the full control over the process.

Further investigation of oscillatory modes in the generalized repressilator model can also lead to other insights in biological engineering. Longer rings can serve as dynamical test concept for orthogonality and modularity of repressors, which are required properties for synthetic biology engineering. Another possible direction for investigation of quasi-stable oscillators is their behavior in the population of cells and their synchronization characteristics if coupled. Finally, investigation of unstable periodic orbits can serve as an aid to collect knowledge about natural synchronized transient oscillators, for example such as observed in developmental biology.

In vivo implementation of larger synthetic networks requires hierarchical organization and scalable approaches. One possibility to achieve this goal, is to construct orthogonal and insulated modules as is common in classical engineering (4, 138, 139). *In vivo* repressor orthogonality and the degree of repressor insulation can be tested with larger rings of the generalized repressilator model, because any kind of cross-talk would immediately destroy the oscillation with the model-predicted frequency. Larger repressilators can therefore be used for an effective *in vivo* test of the orthogonality of several repressors at the same time.

Exploration of population dynamics for quasi-stable oscillators is potentially interesting, since such dynamics were shown to have nontrivial effects. For example, Simpsons paradox in bacterial engineered population by Chuang et al. (140) demonstrates that the information about the ensemble of cells is not contained in the single cell description, in this particular case the selective forces on a single cell are different from global selection on the population level. Similarly, the behavior of synthetic circuits in a population of cells can produce global oscillation pattern while the genetic circuits are not oscillators on the single cell level (48). The quoted examples show that population dynamics might have sometimes unintuitive ef-

fects and these effects are difficult to deduce from the single cell dynamics. It is for instance not clear, how our quasi-stable oscillators will behave in a population if coupled. An interesting hypothesis to test would be if the quasi-stable oscillations would have different synchronization properties compared to synchronization of inducible (48) or autonomous oscillators (141). Intuitively, we expect that better control over the oscillating phase can be achieved because quasi-stable oscillators are switchable and controllable (5).

Synthetic biology can also be used to uncover natural mechanisms for transient behavior, which dominates for instance the developmental biology. The mammalian segmentation clock has been fascinating several generation of scientists (88, 89, 142–144) and still the topology as well as the underpinning mechanisms are a target of recent research. One (rather unconventional) approach could be to synthetically recreate delay based oscillators, such as our even rings, with cell coupling and study the pattern formation of this population of synthetic oscillators. Effective hypothesis testing of mechanisms such as for instance the segmentation clock can be achieved with this technique, because the naturally occurring evolutionary burden will not be present in the synthetic system and the roles of feedback loops can effectively be assessed.

6.2 Transient dynamics and transferable concepts

In the chapter 3 of the thesis analytical bifurcation analysis has been presented for the deterministic version of the generalized repressilator model. The conditions on the model constants have been derived analytically pinpointing which modes of oscillations can occur for given parameter sets. Besides the modes originating from the first Hopf bifurcations in odd rings, all other emerging limit cycles are unstable. For this model however the unstable modes, originating from the first Hopf bifurcation in even rings feature in transient dynamics. The main reason why we are interested in the long-lasting unstable modes in biological models is that cells offer for the genetic circuits a temporarily limited horizon and therefore transient behaviour will be prevalent and even more important than the routinely studied steady states.

Transient states and UPOs in classical engineering are investigated less frequently than the steady states. This is because in the most classical engineering

cases transients occur on very fast time scales and the behaviour that matters is the steady state. With the analysis of UPOs performed here on the generalized repressilator model we showcased the importance of transient dynamics in nonlinear biologically inspired models. This example suggests that large scale synthetic networks with several genes and feedback loops are likely to have nontrivial transient behavior, which needs to be considered during the network design. A potential bioinformatics direction from this would be to extend the existing software as for instance (139), which is used for the construction of large scale circuits, to estimation and characterization of the network transient dynamics. In general,

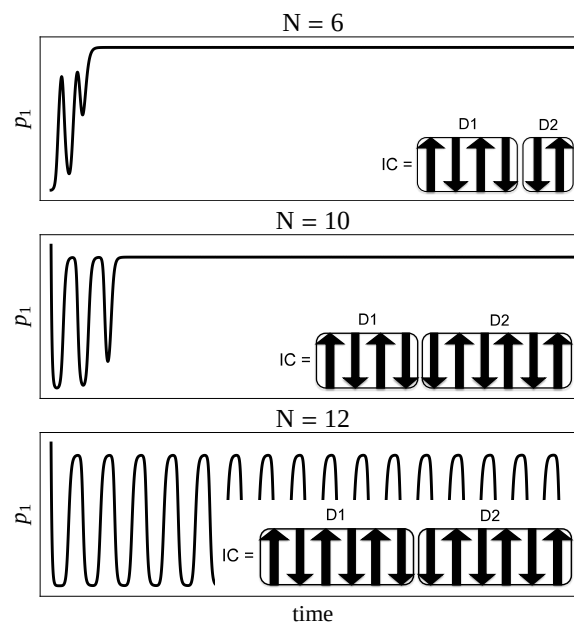


Figure 6.1: **Transient oscillations due to feedback loops.** The transient dynamics last longer in larger networks with feedback loops, which can be illustrated with the generalized repressilator model for different gene numbers in the ring. (The insets show the idealized initial conditions comparable between the topologies.)

nonlinear analysis and group theory as demonstrated here with ring symmetries help identify classical engineering or physics concepts, that are potentially transferable to part engineering in synthetic biology. Synthetic biology engineering can be considered to occur on two levels: the first level is the part design of single elements such as repressors and activators, i. e. the system specific level and the second level is the part interconnection. The symmetry arguments and topological concepts presented here for the rings operate on the second interconnection level. Topology based designs from engineering and physics can be transferred to

synthetic biological engineering if the single elements are effectively isolated from each other (138) and act as well characterized modules. Ring structures offer an impressive example that topology-based modular designs dynamically behave in the same way in different realizations: ring structures oscillate in natural biological systems (145), in engineered biological systems (2), and also in electronic devices (6). Hence, nonlinear mathematical analysis and symmetry arguments are a discipline unifying platform, which helps provide ideas for transferable designs in several scientific areas, ranging from synthetic biology to electrical engineering.

6.3 Stochastic analysis of bifurcations and (Ω, c) -plane

Stochastic aspects and the influence of noise in the generalized repressilator model have been assessed in thesis chapter 4. We have explored numerically and analytically the influence of noise on the onset of stable and quasi-stable oscillations as well as oscillation quality depending on the level of noise. In the regime far away from the onset of oscillations, noise effectively prolongs the duration of transient oscillations in even rings. In odd rings near the onset of Hopf bifurcation point it leads to stochastic resonance effects and the oscillating behavior starts earlier than in the deterministic regime.

The investigation of low copy number regime might be interesting for synthetic biological engineering, because by small effective system size overall less heterologous proteins need to be produced by the cells, which results in smaller cell burdens. Previous synthetic constructs have been designed to operate in large number regimes mainly to showcase the proof of concept with clear signals, but it will become demanding for the cells to carry more sophisticated genetic networks if all nodes are operating in the full capacity deterministic regime. Methods and insights from the chapter 4, which take the effective system size into account can be applied if a low copy number regime is chosen for the design.

An interesting research direction is identified with the example of the generalized repressilator model namely further development of the Ω -expansion framework addressing the influence of biochemical parameter changes, which affect both the mean value as in the classical bifurcation analysis and the variations in the stochastic behavior. This effect is likely to be even more prominent in the eukaryote systems, where gene expression is associated with "burst"-like

stochastic effects (7) and stochastic memory (11). In these cases the engineering of gene copy number variation is expected to affect not only the mean, but also the stochastic properties of the system and the (system size, parameter)-plane view could be potentially useful.

Experimentally it would be interesting to find a way to move only along one of the directions, either only controlling the system size Ω or changing biochemical parameters. One possibility would be to manipulate two parameters at the same time, so that the overall number of the molecules in the system, i. e. the effective system size, does not change, but the mean is changed. There have been experiments in yeast (100), where introduced mutations in the promoter regions that had an effect on the noise measured in terms of the coefficient of variation, but almost no effect on the mean. This research is already going into the direction of noise control via effective system size triggering, but regulation only on the level of promoters still influences the mean even though not to the extent to which for example the variation of copy numbers would do it. So what we are currently assuming is that two parameters need to be tweaked to achieve the navigation on the (system size, parameter)-plane along one of the directions.

6.4 Readout-based control for gene transcription networks

In the last research chapter the accumulated knowledge of the gene repressilators was used to explore the possibility of optimized interference via external signals. We have taken a general read-out based approach and used Fitted-Q algorithm to numerically induce control strategies, which minimize the expression of heterologous proteins in cells.

The perspective, which readout-based control is adding to the standard learning from the experiment, is that it systematically improves the knowledge of the system and guides the next intervention step. In the standard engineering discipline this principle is known as "closing the loop" and it typically leads to accumulation of knowledge about a system in a systematic way. We have formulated two synthetic biology examples as control theoretical problems and applied a reinforcement learning algorithm for induction of optimized control protocols. The presented optimization for the synthetic biological parts is minimization of the cell burden through reduction of the expression of heterologous proteins. The

first example is the optimized reversal of a biological switch, where gene over-expression is avoided and the genetic switch is conducted from the first fixed point to the second via control signal applications. The second example is the tracking of an unstable periodic orbit and keeping an even gene ring in an oscillating state. The examples have been chosen to demonstrate the power of the algorithmic approach providing an intuitive example with the switch and a more complex example with the gene ring.

For the efficient application of this scheme in real-world experiments the main problem remaining to be overcome is that the input and readout in biological systems is still a non-negligible burden for the cells. The current state of the art allows to send control signals via monochromatic light to the surface receptors and then use fluorescence as the readout. The light activation is an indirect method and already requires the expression of several heterologous genes (28), which is associated with a metabolic burden for the cell. A faithful estimation of the cell burden through a synthetic circuit is very demanding and an active research field. The same argument applies for the fluorescent readouts. Potential alternatives for the readouts can become available in the near future as for example quantum dots or radio-frequency-controlled nanoparticles (146), which could be more quantitative and less burdening. The burden coming from the control realization should be small compared to the gain which will be obtained by the control technique, so that the overall cell burden is reduced. However, if these questions can be addressed in the near future, the presented type of model-free optimal control algorithms would open a new perspective for clinical and industrial applications.

By application of the algorithm on biological examples we realized that the efficiency of the control protocols would increase enormously if targeted protein degradation mechanisms would be available. While the current state of the art allows an effective and targeted activation of gene expression, a similar mechanism for protein degradation is not available for *E. coli*. This means that if we want to direct the genetic network to a state in which one of the proteins is absent, we need to wait till it degrades naturally and this time scale dictates the process. The experimental switch reversal (without the use of control protocols) by Gardner et al. (1) took several hours (ca 7h) to switch. Our simulation results for the optimal switching instructed the minimal required expression of the heterologous

proteins, but the duration of the procedure could not be shortened without an effective mechanism for the controlled degradation. Therefore the effectiveness in terms of the duration and minimal possible expression of heterologous proteins will be achieved in the future via both a mechanism for a targeted degradation and non-burdening control signal and readout realizations.

Nevertheless, the presented reinforcement learning has the potential to provide insights for quantitative biology and there are several ways in which the existing algorithm can be extended to better address the specific aspects of gene expression machinery. A potentially interesting direction to refine the algorithm is the systematic consideration of the stochastic delay between the application of the control signal and the observation of the result. In bacterial cells a delay on the order of minutes is expected between these events, due to protein production, which is abstracted as a single step in model equations. Since the gene expression is a stochastic process the delays are not constant, but their probability distributions in biochemical cascades can to some extent be estimated using lumped stochastic processes (11, 147). Based on these analytical results and the basic control theory we are planning to extend the algorithm taking these details into account.

A different way in which reinforcement learning technique could be adjusted for use with real biological experiments is the extension to adaptive control. Biological environments defined by the feeding media of the cell populations, temperature etc, can change over time and so should the control protocols be changing. Most of the changes will happen gradually and not abruptly and that means some of the information on the process, which has been obtained under previous conditions can be reused by the construction of new control protocols. The extension of reinforcement learning algorithm would need to effectively balance the previously gathered information and the new information obtained under the changed environment to optimally learn and control the process. We are currently following up this direction.

Bibliography

1. Gardner, T., C. R. Cantor, and J. J. Collins, 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–342.
2. Elowitz, M. B., and S. Leibler, 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338.
3. Stricker, J., S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, 2008. A fast, robust and tunable synthetic gene oscillator. *Nature* 456:516–519.
4. Andrianantoandro, E., S. Basu, D. K. Karig, and R. Weiss, 2006. Synthetic biology: new engineering rules for an emerging discipline. *Mol Sys Bio* 2:2006.0028.
5. Strelkowa, N., and M. Barahona, 2010. Switchable genetic oscillator operating in quasi-stable mode. *J R Soc Interface* 7:1071–1082.
6. Lindner, J. F., and A. R. Bulsara, 2006. One-way coupling enables noise-mediated spatiotemporal patterns in media of otherwise quiescent multistable elements. *Phys Rev E* 74:020105.
7. Raj, A., and A. van Oudenaarden, 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* 135:216–226.
8. Raser, J. M., and E. K. O’Shea, 2005. Noise in Gene Expression: Origins, Consequences, and Control. *Science* 309:2010–2013.
9. Swain, P. S., M. B. Elowitz, and E. D. Siggia, 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.
10. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain, 2002. Stochastic Gene Expression in a Single Cell. *Science* 297:1183–1186.
11. Pedraza, J. M., and J. Paulsson, 2008. Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* 319:339–343.

12. Bellman, R., 1957. Dynamic Programming. Princeton University Press.
13. Ernst, D., P. Geurts, and L. Wehenkel, 2005. Tree-Based Batch Mode Reinforcement Learning. *J Mach Learn Res* 6:503–556.
14. Heinemann, M., and S. Panke, 2006. Synthetic biology—putting engineering into biology. *Bioinformatics* 22:2790–2799.
15. Khalil, A. S., and J. J. Collins, 2010. Synthetic biology: applications come of age. *Nat Rev Genet* 11:367–379.
16. Tigges, M., T. T. Marquez-Lago, J. Stelling, and M. Fussenegger, 2009. A tunable synthetic mammalian oscillator. *Nature* 457:309–312.
17. Kobayashi, H., M. Kaern, M. Araki, C. K., T. S. Gardner, C. R. Cantor, and J. J. Collins, 2004. Programmable cells: interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA* 101:8414–8419.
18. Anderson, J., E. Clarke, A. Arkin, and C. Voigt, 2005. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* 355:619–627.
19. Karlebach, G., and R. Shamir, 2008. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* 9:770–780.
20. van Kampen, N. G., 2007. Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam, 3-rd edition edition.
21. Gillespie, D. T., 1977. Exact Stochastic Simulation of Coupled Chemical Reactions. *J Chem Phys* 8:2340–2361.
22. iGEM, 2010. MIT Registry. <http://partsregistry.org>. .
23. Batt, G., B. Yordanov, R. Weiss, and C. Belta, 2007. Robustness analysis and tuning of synthetic gene networks. *Bioinformatics* 23:2415–2422.
24. Tsai, T. Y.-C., Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. E. Ferrell, 2008. Robust, Tunable Biological Oscillations from Interlinked Positive and Negative Feedback Loops. *Science* 321:126–129.
25. Alberts, B., A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, 2007. Molecular Biology of the Cell. Garland Science, 5th edition edition.

26. El-Sharoud, W., 2008. *Bacterial Physiology: A Molecular Approach*. Springer, Berlin Heidelberg.
27. Margolin, A. A., K. Wang, W. K. Lim, M. Kustagi, I. Nemenman, and A. Califano, 2006. Reverse engineering cellular networks. *Nat Protocols* 1:662–671.
28. Shimizu-Sato, S., E. Huq, J. M. Tepperman, and P. H. Quail, 2002. A light-switchable gene promoter system. *Nat Biotech* 20:1041–1044.
29. Levskaya, A., A. A. Chevalier, J. J. Tabor, Z. B. Simpson, L. A. Lavery, M. Levy, E. A. Davidson, A. Scouras, A. D. Ellington, E. M. Marcotte, and C. A. Voigt, 2005. Synthetic biology: engineering *Escherichia coli* to see light. *Nature* 438:441–442.
30. Bennett, M. R., and J. Hasty, 2009. Microfluidic devices for measuring gene network dynamics in single cells. *Nat Rev Genet* 10:628–638.
31. Lässig, M., 2007. From biophysics to evolutionary genetics: Statistical aspects of gene regulation. *BMC Bioinformatics* .
32. Celesnik, H., A. Deana, and J. G. Belasco, 2007. Initiation of RNA Decay in *Escherichia coli* by 5' Pyrophosphate Removal. *Molecular Cell* 27:79–90.
33. Richards, J., T. Sundermeier, A. Svetlanov, and A. W. Karzai, 2008. Quality control of bacterial mRNA decoding and decay. *Biochimica et Biophysica Acta* 1779:574–582.
34. Belasco, J. G., and J. G. Brawerman, 1993. *Control of messenger RNA stability*. Academic Press, San Diego.
35. Mizusawa, S., and S. Gottesman, 1983. Protein degradation in *Escherichia coli*: the lon gene controls the stability of sulA protein. *Proc Natl Acad Sci USA* 80:358–362.
36. Rao, C. V., and A. P. Arkin, 2003. Stochastic Chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *J Chem Phys* 118:4999–5009.
37. Altland, A., and B. Simons, 2006. *Condensed Matter Field Theory*. Cambridge University Press.
38. Kierzek, A. M., 2002. STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. *Bioinformatics* 18:470–481.
39. Hill, A. V., 1910. The combinations of haemoglobin with oxygen and with carbon monoxide. *I J Physiol* 40:iv–vii.

40. Monod, J., J. Wyman, and J. Changeux, 1965. On the nature of allosteric transitions: a plausible model. *J Mol Biol* 12:88–118.
41. Weiss, J., 1997. The Hill equation revisited: uses and misuses. *FASEB J* 11:835–841.
42. Gadgil, C., C. H. Lee, and H. G. Othmer, 2005. A stochastic analysis of first-order reaction networks. *Bulletin of Mathematical Biology* 67:901–946.
43. Hemberg, M., and M. Barahona, 2008. A Dominated Coupling From The Past algorithm for the stochastic simulation of networks of biochemical reactions. *BMC Systems Biology* 2.
44. Wilkinson, D. J., 2006. Stochastic Modelling for Systems Biology. CRC Press, 1-st edition.
45. Øksendal, B. K., 2003. Stochastic Differential Equations. An Introduction with Applications. Springer-Verlag, Berlin, Heidelberg, New York, 6-th edition.
46. Ausubel, F. M., 1987. Current Protocols in Molecular Biology. Wiley, New York.
47. Müller, S., J. Hofbauer, L. Endler, C. Flamm, S. Widder, and P. Schuster, 2006. A generalized model of the repressilator. *J Math Biol* 53:905–937.
48. Danino, T., O. Mondragon-Palomino, L. Tsimring, and J. Hasty, 2010. A synchronized quorum of genetic clocks. *Nature* 463:326–330.
49. Strogatz, S. H., 1994. Nonlinear Dynamics and Chaos. Westview Press.
50. Altland, A., and B. Simons, 2006. Condensed Matter Field Theory. Cambridge University Press, 1st edition.
51. Bulsara, A. R., V. In, A. Kho, P. Longhini, A. Palacios, W.-J. Rappel, J. Acebron, S. Baglio, and B. Ando, 2004. Emergent oscillations in unidirectionally coupled overdamped bistable systems. *Phys Rev E* 70:036103.
52. Golubitsky, M., I. Stewart, and D. Schaeffer, 1988. Singularities and Groups in Bifurcation Theory, Volume II. Springer-Verlag, Berlin, applied mathematical sciences vol. 69 edition.
53. Perlikowski, P., S. Yanchuk, M. Wolfrum, A. Stefanski, P. Mosiolek, and T. Kapitaniak, 2010. Routes to complex dynamics in a ring of unidirectionally coupled systems. *Chaos* 20:013111.

54. Csete, M. E., and J. C. Doyle, 2002. Reverse Engineering of Biological Complexity. *Science* 295:1664–1669.
55. El-Samad, H., J. P. Goff, and M. Khammash, 2002. Calcium Homeostasis and Parturient Hypocalcemia: An Integral Feedback Perspective. *Journal of Theoretical Biology* 214:17 – 29.
56. Muzzey, D., C. A. Gómez-Urbe, J. T. Mettetal, and A. van Oudenaarden, 2009. A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation. *Cell* 138:160 – 171.
57. Yi, T.-M., Y. Huang, M. I. Simon, and J. Doyle, 2000. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci USA* 97:4649–4653.
58. El-Samad, H., and M. Khammash, 2010. Modelling and analysis of gene regulatory networks using feedback control theory. *International Journal of Systems Science* 41:17–33.
59. Purcell, O., N. J. Savery, C. S. Grierson, and M. di Bernardo, 2010. A comparative analysis of synthetic genetic oscillators. *J R Soc Interface* –.
60. Borneman, A. R., J. A. Leigh-Bell, H. Yu, P. Bertone, M. Gerstein, and M. Snyder, 2006. Target hub proteins serve as master regulators of development in yeast. *Genes Development* 20:435–448.
61. Liu, Y., M. Asakura, H. Inoue, T. Nakamura, M. Sano, Z. Niu, M. Chen, R. J. Schwartz, and M. D. Schneider, 2007. Sox17 is essential for the specification of cardiac mesoderm in embryonic stem cells. *Proc Natl Acad Sci USA* 104:3859–3864.
62. Bondue, A., G. Lapouge, C. Paulissen, C. Semeraro, M. Iacovino, M. Kyba, and C. Blanpain, 2008. Mesp1 Acts as a Master Regulator of Multipotent Cardiovascular Progenitor Specification. *Cell Stem Cell* 3:69–84.
63. Megerle, J. A., G. Fritz, U. Gerland, K. Jung, and J. O. Rädler, 2008. Timing and Dynamics of Single Cell Gene Expression in the Arabinose Utilization System. *Biophys J* 95:2103–2115.
64. Spencer, S. L., S. Gaudet, J. G. Albeck, J. M. Burke, and P. K. Sorger, 2009. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459:428–432.

65. Glick, B. R., 1995. Metabolic load and heterologous gene expression. *Biotechnology Advances* 13:247 – 261.
66. Glieder, A., 2009. Technische Universität Graz. Private Communication. *Privat Communication*.
67. Trefethen, L. N., and M. Embree, 2005. Spectra and pseudospectra. Princeton University Press, Princeton, NJ, USA.
68. Süel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440:545–550.
69. Turcotte, M., J. Garcia-Ojalvo, and G. M. Süel., 2008. A genetic timer through noise-induced stabilization of an unstable state. *Proc Natl Acad Sci USA* 105:15732–15737.
70. Smith, H., 1987. Oscillations and multiple steady states in a cyclic gene model with repression. *J Math Biol* 25:169–190.
71. Franklin, G. F., A. Emami-Naeini, and J. D. Powell, 1993. Feedback Control of Dynamic Systems. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
72. Ahuja, S., and C. W. Rowley, 2010. Feedback control of unstable steady states of flow past a flat plate using reduced-order estimators. *J Fluid Mechanics* 645:447–478.
73. McRuer, D., and D. Graham, 2004. Flight Control Century: Triumphs of the Systems Approach. *J of Guidance, Control, and Dynamics* 27:161–173.
74. Atkinson, M. R., M. A. Savageau, J. T. Myers, and A. J. Ninfa, 2003. Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in *Escherichia coli*. *Cell* 113:597–607.
75. Doedel, E., 2007. AUTO 07. Software for Continuation and Bifurcation Problems in Ordinary Differential Equations. download - <http://indy.cs.concordia.ca/auto/>.
76. Soos, Z., 2007. Identification of dimerization phase transitions driven by Peierls and other mechanisms. *Chemical Physics Letters* 440:87–91.
77. Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1992. Numerical recipes in C: the art of scientific computing. Cambridge University Press, 2 edition.
78. Zi, Z., Y. Zheng, A. E. Rundell, and E. Klipp, 2008. SBML-SAT: a systems biology markup language (SBML) based sensitivity analysis tool. *BMC Bioinformatics* 9.

79. Galassi, M., J. Davies, J. Theiler, B. Gough, G. Jungman, P. Alken, M. Booth, and F. Ross, 2009. GNU Scientific Library Reference Manual. Network Theory Ltd, 3 (v1.12) edition.
80. Halton, J., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer Math* 2 84–90.
81. Vandewoestyne, B., and R. Cools, 2006. Good permutations for deterministic scrambled Halton sequences in terms of L2-discrepancy. *J Comput Appl Math* 189:341–361.
82. Braun, O., and Y. Kivshar, 2004. The Frenkel-Kontorova Model. Concepts, Methods, and Applications. Springer, illustrated edition.
83. Hemberg, M., and M. Barahona, 2007. Perfect Sampling of the Master Equation for Gene Regulatory Networks. *Biophys J* 93:401–410.
84. Karin, M., and A. Lin, 2002. NF- κ B at the crossroads of life and death. *Nat Immunol* 3:221–227.
85. Naugler, W. E., and M. Karin, 2008. NF- κ B and cancer - identifying targets and mechanisms. *Current Opinion in Genetics & Development* 18:19–26.
86. Ashall, L., C. A. Horton, D. E. Nelson, P. Paszek, C. V. Harper, K. Sillitoe, S. Ryan, D. G. Spiller, J. F. Unitt, D. S. Broomhead, D. B. Kell, D. A. Rand, V. See, and M. R. H. White, 2009. Pulsatile Stimulation Determines Timing and Specificity of NF-kappaB-Dependent Transcription. *Science* 324:242–246.
87. Holtzendorff, J., D. Hung, P. Brende, A. Reisenauer, P. H. Viollier, H. H. McAdams, and L. Shapiro, 2004. Oscillating Global Regulators Control the Genetic Circuit Driving a Bacterial Cell Cycle. *Science* 304:983–987.
88. Yun-Jin, B. L. Aerne, L. Smithers, C. Haddon, D. Ish-Horowicz, and J. Lewis, 2000. Notch signalling and the synchronization of the somite segmentation clock. *Nature* 408:475–479.
89. Pourquie, O., 2003. The Segmentation Clock: Converting Embryonic Time into Spatial Pattern. *Science* 301:328–330.
90. Yanchuk, S., and M. Wolfrum, 2008. Destabilization patterns in chains of coupled oscillators. *Phys Rev E* 77:026212.

91. Landsman, A. S., and I. B. Schwartz, 2006. Predictions of ultraharmonic oscillations in coupled arrays of limit cycle oscillators. *Phys Rev E* 74:036204.
92. Strogatz, S. H., 2001. Exploring complex networks. *Nature* 410:268–276.
93. In, V., A. R. Bulsara, A. Palacios, P. Longhini, A. Kho, and J. D. Neff, 2003. Coupling-induced oscillations in overdamped bistable systems. *Phys Rev E* 68:045102.
94. Bulsara, A. R., C. Seberino, L. Gammaitoni, M. F. Karlsson, B. Lundqvist, and J. W. C. Robinson, 2003. Signal detection via residence-time asymmetry in noisy bistable devices. *Phys Rev E* 67:016120.
95. Hardy, G. H., and T. W. Körner, 2008. A course of pure mathematics. Cambridge University Press, New York, 10th edition.
96. Dhooge, A., W. Govaerts, and Y. A. Kuznetsov, 2003. MATCONT: A MATLAB package for numerical bifurcation analysis of ODEs. *ACM Trans Math Softw* 29:141–164.
97. Peierls, R., 1955. Quantum theory of solids. Oxford Univ. Press, Oxford.
98. Braun, O. M., and Y. S. Kivshar, 2004. The Frenkel-Kontorova Model: Concepts, Methods, and Applications (Theoretical and Mathematical Physics). Springer Verlag, Berlin Heidelberg, 1st edition.
99. Maamar, H., A. Raj, and D. Dubnau, 2007. Noise in Gene Expression Determines Cell Fate in *Bacillus subtilis*. *Science* 317:526–529.
100. Murphy, K. F., R. M. Adams, X. Wang, G. Balazsi, and J. J. Collins, 2010. Tuning and controlling gene expression noise in synthetic gene networks. *Nucl Acids Res* 38:2712–2726.
101. Eldar, A., and M. B. Elowitz, 2010. Functional roles for noise in genetic circuits. *Nature* 467:167–173.
102. Johnston, R. J., J. Desplan, and C. Desplan, 2010. Stochastic Mechanisms of Cell Fate Specification that Yield Random or Robust Outcomes. *Annu Rev Cell Dev Biol* 26:16.1–16.31.
103. Acar, M., J. T. Mettetal, and A. van Oudenaarden, 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* 40:471 – 475.

104. Perkins, T. J., and P. S. Swain, 2009. Strategies for cellular decision-making. *Mol Syst Biol* 5:326.
105. El-Samad, H., and M. Khammash, 2006. Coherence Resonance: A mechanism for noise induced stable oscillations in gene regulatory networks. *Proceedings of the 45th IEEE Conference on Decision and Control* 1–4244–0171–2/06.
106. Chabot, J. R., J. M. Pedraza, P. Luitel, and A. van Oudenaarden, 2007. Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature* 450:1249 – 1252.
107. Marlovits, G., C. J. Tyson, B. Novak, and J. J. Tyson, 1998. Modeling M-phase control in *Xenopus* oocyte extracts: the surveillance mechanism for unreplicated DNA. *Biophysical Chemistry* 72:169 – 184.
108. Boland, R. P., T. Galla, and A. J. McKane, 2008. How limit cycles and quasi-cycles are related in systems with intrinsic noise. *Journal of Statistical Mechanics: Theory and Experiment* 2008:P09001.
109. Kuske, R., L. F. Gordillo, and P. Greenwood, 2007. Sustained oscillations via coherence resonance in SIR. *Journal of Theoretical Biology* 245:459 – 469.
110. McKane, A. J., and T. J. Newman, 2005. Predator-Prey Cycles from Resonant Amplification of Demographic Stochasticity. *Phys Rev Lett* 94:218102.
111. Goel, N., and N. Richter-Dyn, 1974. *Stochastic Models in Biology*. Academic press New York, San Francisco, London.
112. Kepler, T. B., and T. C. Elston, 2001. Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. *Biophys J* 81:3116–3136.
113. Shahrezaei, V., and P. S. Swain, 2008. Analytical distributions for stochastic gene expression. *Proc Natl Acad of Sci USA* 105:17256–17261.
114. Cai, L., N. Friedman, and X. S. Xie, 2006. Stochastic protein expression in individual cells at the single molecule level. *Nature* 440:358–362.
115. Nattermann, T., 2005. A scaling approach to ideal quantum gases. *American Journal of Physics* 73:349–356.
116. Guantes, R., and J. F. Poyatos, 2006. Dynamical Principles of Two-Component Genetic Oscillators. *PLoS Comput Biol* 2:e30.

117. Gardiner, C. W., 2004. Handbook of Stochastic Methods. Springer, 3-rd edition.
118. Risau-Gusman, S., and G. Abramson, 2007. Bounding the quality of stochastic oscillations in population models. *The European Physical Journal B - Condensed Matter and Complex Systems* 60:515–520.
119. Lin, F., K. Muthuraman, and M. Lawley, 2010. An optimal control theory approach to non-pharmaceutical interventions. *BMC Infectious Diseases* 10:32.
120. Stan, G.-B., F. Belmudes, R. Fonteneau, F. Zeggwagh, M.-A. Lefebvre, C. Michelet, and D. Ernst, 2008. Modelling the influence of activation-induced apoptosis of CD4⁺ and CD8⁺ T-cells on the immune system response of a HIV-infected patient. *IET Systems Biology* 2:94–102.
121. Schiff, S. J., and T. Sauer, 2008. Kalman filter control of a model of spatiotemporal cortical dynamics. *J Neural Eng* 5:1–8.
122. Ivanova, N., R. Dobrin, R. Lu, I. Kutenko, J. Levorse, C. DeCoste, X. Schafer, Y. Lun, and I. R. Lemischka, 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* 442:533–538.
123. Mettetal, J. T., D. Muzzey, C. Gomez-Urbe, and A. van Oudenaarden, 2008. The Frequency Dependence of Osmo-Adaptation in *Saccharomyces cerevisiae*. *Science* 319:482–484.
124. Levskaya, A., O. D. Weiner, W. A. Lim, and C. A. Voigt, 2009. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* 461:997–1001.
125. Prill, R. J., D. Marbach, J. Saez-Rodriguez, P. K. Sorger, L. G. Alexopoulos, X. Xue, N. D. Clarke, G. Altan-Bonnet, and G. Stolovitzky, 2010. Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges. *PLoS ONE* 5.
126. Ernst, D., P. Geurts, and L. Wehenkel, 2005. Tree-based batch mode reinforcement learning. *Journal of Machine Learning Research* 6:503–556.
127. Geurts, P., D. Ernst, and L. Wehenkel, 2006. Extremely randomized trees. *Machine Learning* 63:3–42.
128. Bertsekas, D., 2000. Dynamic Programming and Optimal Control, volume I. Athena Scientific, Belmont, MA, 2nd edition.

129. Bellman, R., and S. Dreyfus, 1957. *Applied Dynamic Programming*. Princeton University Press.
130. Breiman, L., 1996. Bagging Predictors. *Machine Learning* 24:123–140.
131. Breiman, L., 2001. Random forests. *Machine Learning* 45:5–32.
132. Breiman, L., J. Friedman, R. Olsen, and C. Stone, 1984. *Classification and regression trees*. Chapman and Hall/CRC, 1st edition.
133. Waddington, C. H., 1940. *Organisers and Genes*. Cambridge Univ. Press, Cambridge, UK.
134. Graf, T., and T. Enver, 2009. Forcing cells to change lineages. *Nature* 462:587–594.
135. Faryabi, B., A. Datta, and E. Dougherty, 2007. On approximate stochastic control in genetic regulatory networks. *IET Systems Biology* 1:361–368.
136. Datta, A., A. Choudhary, M. L. Bittner, and E. R. Dougherty, 2003. External Control in Markovian Genetic Regulatory Networks. *Machine Learning* 52:169–191.
137. Edwards, R., S. Kim, and P. van den Driessche, 2010. Control design for sustained oscillation in a two-gene regulatory network. *J Math Biol* online before print:27 Apr.
138. Vecchio, D., A. J. Ninfa, and E. D. Sontag, 2008. Modular cell biology: retroactivity and insulation. *Mol Syst Biol* 4:161.
139. Marchisio, M., and J. Stelling, 2008. Computational design of synthetic gene circuits with composable parts. *Bioinformatics* 24:1903–1910.
140. Chuang, J. S., O. Rivoire, and S. Leibler, 2009. Simpson’s Paradox in a Synthetic Microbial System. *Science* 323:272–275.
141. Garcia-Ojalvo, J., M. B. Elowitz, and S. H. Strogatz, 2004. Modeling a synthetic multicellular clock: Repressilators coupled by quorum sensing. *Proc Natl Acad Sci USA* 101:10955–10960.
142. Giudicelli, F., E. M. Özbudak, G. J. Wright, and J. Lewis, 2007. Setting the Tempo in Development: An Investigation of the Zebrafish Somite Clock Mechanism. *PLoS Biol* 5:e150.
143. Lewis, J., 2008. From Signals to Patterns: Space, Time, and Mathematics in Developmental Biology. *Science* 322:399–403.

144. Lacalli, T., D. Wilkinson, and L. Harrison, 1988. Theoretical aspects of stripe formation in relation to *Drosophila* segmentation. *Development* 104:105–113.
145. Takamatsu, A., R. Tanaka, H. Yamada, T. Nakagaki, T. Fujii, and I. Endo, 2001. Spatiotemporal Symmetry in Rings of Coupled Biological Oscillators of *Physarum* Plasmodial Slime Mold. *Phys Rev Lett* 87:078102.
146. Lu, T. K., A. S. Khalil, and J. J. Collins, 2009. Next-generation synthetic gene networks. *Nat Biotech* 27:1139–1150.
147. Anchang, B., M. J. Sadeh, J. Jacob, A. Tresch, M. O. Vlad, P. J. Oefner, and R. Spang, 2009. Modeling the temporal interplay of molecular signaling and gene expression by using dynamic nested effects models. *Proc Natl Acad Sci USA* 106:6447–6452.