

---

# Stochastische Modelle und Umgebungseffekte im Kontext der Synthetischen Biologie

---

**Stochastic Models and Environmental Effects in the Context of Synthetic Biology**

Zur Erlangung des Grades eines Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigte Dissertation von Johannes Falk aus Würzburg

Tag der Einreichung: 8.2.2019, Tag der Prüfung: 17.4.2019

Darmstadt 2019 – D 17

Referentin: Prof. Dr. Barbara Drossel

Korreferent: Prof. Dr. Harald Kolmar



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

Institut für Festkörperphysik  
Fachbereich Physik

**Stochastische Modelle und Umgebungseffekte im Kontext der Synthetischen Biologie**  
Stochastic Models and Environmental Effects in the Context of Synthetic Biology

Genehmigte Dissertation von Johannes Falk aus Würzburg

Referentin: Prof. Dr. Barbara Drossel

Korreferent: Prof. Dr. Harald Kolmar

Tag der Einreichung: 8.2.2019

Tag der Prüfung: 17.4.2019

Darmstadt – D 17

Bitte zitieren Sie dieses Dokument als:

URN: [urn:nbn:de:tuda-tuprints-86852](https://nbn-resolving.org/urn:nbn:de:tuda-tuprints-86852)

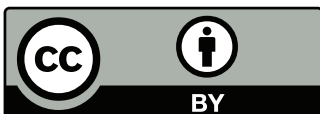
URL: <http://tuprints.ulb.tu-darmstadt.de/8685>

Dieses Dokument wird bereitgestellt von tuprints,

E-Publishing-Service der TU Darmstadt

<http://tuprints.ulb.tu-darmstadt.de>

[tuprints@ulb.tu-darmstadt.de](mailto:tuprints@ulb.tu-darmstadt.de)



Die Veröffentlichung steht unter folgender Creative Commons Lizenz:

Namensnennung 4.0 International

<http://creativecommons.org/licenses/by/4.0/>

---

---

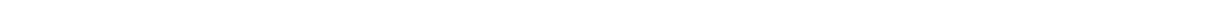
# Erklärung zur Dissertation

Hiermit versichere ich, die vorliegende Dissertation ohne Hilfe Dritter nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 11.06.2019

---

(Johannes Falk)



---

---

# Zusammenfassung

Die Synthetische Biologie versteht sich als ein interdisziplinäres Forschungsgebiet, das die Entwicklung komplexer künstlicher biologischer Systeme zum Ziel hat. Neben der erfolgreichen Integration von synthetischen Schaltungen in lebende Organismen, bedarf es hierzu passender Analyse- und Messtechnik sowie eines theoretischen Verständnisses der zugrunde liegenden komplexen biochemischen Wechselwirkungen. In der Synthetischen Biologie arbeiten folglich Biologie und Chemie eng mit Ingenieurwissenschaften, Physik und Informatik zusammen. Im Kontext dieser Interdisziplinarität werden in der vorliegenden Arbeit Methoden und Modelle der theoretischen Physik beschrieben, welche in der Synthetischen Biologie Anwendung finden.

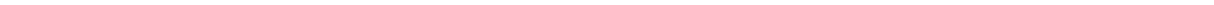
Im ersten Teil der Arbeit wird ein konkretes biologisches System, ein auf dem CRISPR/dCas9-System beruhender logischer Schalter, analysiert und modelliert. Experimentelle Daten von Messungen an Hefezellen zeigen hier, dass das System ein bimodales Schaltverhalten zeigt. Zur Erklärung dieser Beobachtung werden ein deterministisches sowie ein stochastisches Modell entwickelt. Mithilfe dieser Modelle lässt sich das Verhalten des CRISPR/dCas9-Systems nachbilden, womit ein Erklärungsansatz für die zugrunde liegenden biologischen Zusammenhänge gegeben wird.

Der zweite Teil dieser Arbeit geht auf ein generelles Problem der Synthetischen Biologie ein, nämlich die sogenannten Kontexteffekte. Hiermit werden allgemein alle Effekte bezeichnet, die durch die Umgebung, in welche ein synthetischer Schaltkreis eingebettet wird, verursacht werden. Zur besseren Analyse dieser oft nicht vernachlässigbaren Einwirkungen werden die durch verschiedene mögliche Umgebungstopologien erzeugten Memory-Effekte analytisch berechnet. Anhand von verschiedenen biologischen Beispielen zeigen wir dann, wie stark sich die Umgebung auf das Verhalten des Schaltkreises auswirken kann.

Das Phänomen, dass deterministische und probabilistische Beschreibung grundlegend verschiedene Ergebnisse liefern, betrachtet der dritte Teil der Arbeit. Die theoretische Physik ist bestrebt, beobachtbare Phänomene durch möglichst einfache und zugängliche Modelle darzustellen und damit auch zu erklären. Zu diesem Zweck wird im dritten Teil ein minimales Modell vorgestellt, welches ausreichend ist, um durch sogenanntes Burstrauschen verursachte stabile Zustände darzustellen. Mittels deterministischer Methoden kann das beobachtete Systemverhalten nicht korrekt vorhergesagt werden. Zur Analyse des Effektes nutzen wir daher eine Gleichung, welche die Extrema der Wahrscheinlichkeitsverteilung beschreibt.

Aufbauend auf den Ergebnissen des vorigen Kapitels wird im letzten Teil dieser Arbeit kurz die Rolle von minimalen Modellen in der Synthetischen Biologie diskutiert. Insbesondere vor dem Hintergrund der verschiedenen erkenntnistheoretischen Ziele von Ingenieur- und Naturwissenschaften beleuchten wir dabei, ob und wie theoretische Modelle Nutzen in der Synthetischen Biologie entfalten können.

Zusammenfassend wird in dieser Arbeit aufgezeigt, wie stochastische Modellierung und die analytische Behandlung von Umgebungseffekten in der Synthetischen Biologie eingesetzt werden können. Die verschiedenen vorgestellten Herangehensweisen lassen sich dabei als grundlegende Methoden verstehen, auf die Modellierer in Zukunft zurückgreifen können.



---

---

# Abstract

Synthetic Biology is an interdisciplinary field of research aiming at the development of complex artificial biological systems. In addition to the successful integration of synthetic circuits into living organisms, this requires appropriate analysis and measurement techniques as well as a theoretical understanding of the underlying biochemical interactions. Consequently, in Synthetic Biology biologists and chemists work closely together with engineers, physicists, and computer scientists. In the context of this interdisciplinarity, this thesis describes methods and models of theoretical physics that are applied in Synthetic Biology.

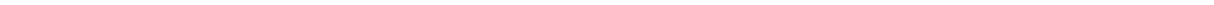
In the first part of the thesis, a concrete biological system, a logical switch based on the CRISPR/dCas9 system, is analyzed and modeled. Experimental data from measurements on yeast cells show that the system has a bimodal switching behavior. A deterministic and a stochastic model are developed to explain this observation. With the help of this model, the behavior of the CRISPR/dCas9 system can be simulated, thus providing an explanatory approach for the underlying biological mechanisms.

The second part of this thesis deals with a general problem of Synthetic Biology, namely the so-called context effects. This refers to all effects caused by the environment in which a synthetic circuit is embedded. For a better analysis of these often not negligible effects, the memory effects generated by different possible environment topologies are analytically calculated. Using various biological examples, we then show how strongly the environment can affect the behavior of the circuit.

The phenomenon that deterministic and probabilistic descriptions provide fundamentally different results is considered in the third part of this paper. Theoretical physics aims to represent and explain observable phenomena using models that are as simple and accessible as possible. For this purpose, the third part presents a minimal model, which is sufficient to describe bimodal probability distributions caused by burst noise. Using deterministic methods, this system behavior cannot be predicted correctly. For the analysis of the effect we therefore use an equation that gives the extreme of the probability distribution.

Based on the results of the previous chapter, the role of minimal models in Synthetic Biology is briefly discussed in the last part of this paper. Particularly against the background of the different epistemological goals of engineering and natural sciences, we will examine whether and how theoretical models can develop benefits in Synthetic Biology.

In summary, this thesis shows how stochastic modeling and the analytical treatment of environmental effects can be applied in Synthetic Biology. The different approaches presented can be understood as basic methods that modelers will be able to use in the future.



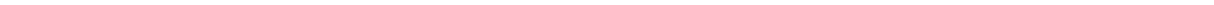


---

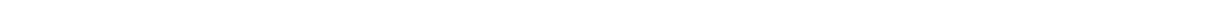
---

# Inhaltsverzeichnis

<b>Abkürzungsverzeichnis</b>	<b>xi</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Ziele und Nutzen . . . . .	2
1.2 Aufbau der Arbeit . . . . .	2
<b>2 Grundlagen</b>	<b>5</b>
2.1 Biologische Grundlagen . . . . .	5
2.1.1 Biologische Zellen . . . . .	5
2.1.2 DNA und Proteine . . . . .	6
2.1.3 Genregulationsnetzwerke . . . . .	7
2.2 Genregulationsnetzwerke im Modell . . . . .	8
2.2.1 Reaktionsgleichungen . . . . .	9
2.2.2 Reaktionsraten und Zustandsvektoren . . . . .	9
2.2.3 Simulation der Reaktionsgleichungen . . . . .	10
2.3 Theoretische Grundlagen . . . . .	13
2.3.1 Ergebnismenge und Zufallsvariablen . . . . .	13
2.3.2 Bedingte Wahrscheinlichkeiten . . . . .	14
2.3.3 Stochastischer Prozess . . . . .	14
2.3.4 Markov-Prozesse . . . . .	15
2.3.5 Von der Chapman-Kolmogorov-Gleichung zur Mastergleichung . . . . .	16
2.3.6 Grundannahmen der CME . . . . .	18
2.3.7 Von Teilchenzahlen zu Teilchenkonzentrationen . . . . .	19
2.3.8 Die Kramers-Moyal-Entwicklung und die Fokker-Planck-Gleichung . . . . .	20
2.3.9 Reaktionsratengleichungen – Dynamisches System . . . . .	22
2.3.10 Das Massenwirkungsgesetz und die Hill-Funktion . . . . .	23
<b>3 Modell für einen CRISPR/dCas9-basierten genetischen Schalter</b>	<b>27</b>
3.1 Übersicht: Genetische Schalter . . . . .	27
3.1.1 CRISPR/Cas9-System . . . . .	28
3.1.2 CRISPR-Interferenz . . . . .	28
3.1.3 CRISPR/dCas9-basierte Schalter in Hefe . . . . .	30
3.2 Ein CRISPR/dCas9-basiertes AND-Gate in Hefe . . . . .	30
3.2.1 Experimentelle Analyse des Schalters . . . . .	31
3.2.2 Charakterisierung des Schaltverhaltens . . . . .	33
3.2.2.1 Die ROC-Kurve . . . . .	33
3.2.3 Modellierung . . . . .	35
3.2.4 Zellschädigende Effekte . . . . .	39
3.3 Diskussion der Ergebnisse . . . . .	41
<b>4 Memory-Effekte monomolekularer Umgebungen</b>	<b>43</b>
4.1 Kontexteffekte in der Synthetischen Biologie . . . . .	43
4.1.1 Stand der Forschung . . . . .	44



4.2	Hintergrundwissen . . . . .	45
4.2.1	Definition: Lineare und monomolekulare Reaktionsnetzwerke . . . . .	45
4.2.2	Random Walk und absorbierende Zustände . . . . .	45
4.2.3	Absorptionswahrscheinlichkeit und Rückwärts-Mastergleichung . . . . .	46
4.2.4	Spektrale Leistungsdichte . . . . .	47
4.3	Subsystem und Umgebung . . . . .	47
4.3.1	Umgebung mit monomolekularen Reaktionen . . . . .	49
4.3.2	Zusammenhang zwischen linearer Kette und Random Walk . . . . .	50
4.3.3	Absorptionswahrscheinlichkeit . . . . .	52
4.4	Mögliche monomolekulare Umgebungen und deren Memory-Kerne . . . . .	53
4.4.1	Eine einzelne Spezies . . . . .	53
4.4.2	Kette mit beidseitiger Ankopplung an das Subsystem . . . . .	54
4.4.3	Kette mit einseitiger Ankopplung an das Subsystem . . . . .	54
4.5	Anwendungen . . . . .	57
4.5.1	Retroaktivität im Repressilator . . . . .	57
4.5.1.1	Der Repressilator . . . . .	59
4.5.2	Retroaktivität durch Kopplung an Promotorregion . . . . .	59
4.5.3	Kontexteffekte durch Promiskuität . . . . .	60
4.6	Diskussion . . . . .	64
<b>5</b>	<b>Ein minimales Modell für durch Burstrauschen induzierte Bimodalität</b>	<b>65</b>
5.1	Rauschen in biologischen Systemen . . . . .	65
5.2	Hintergrundwissen: Bifurkationen . . . . .	67
5.3	Das Schlögl-Modell . . . . .	68
5.3.1	Deterministische Beschreibung . . . . .	69
5.3.2	Probabilistische Beschreibung . . . . .	69
5.4	Analyse des unmodifizierten Schlögl-Modells . . . . .	70
5.5	Der Einfluss von Burstrauschen . . . . .	74
5.6	Diskussion . . . . .	77
<b>6</b>	<b>Minimalmodelle in der Synthetischen Biologie</b>	<b>81</b>
6.1	Der Nutzen von Modellen in Ingenieur- und Naturwissenschaften . . . . .	82
6.2	Minimale Spielzeugmodelle . . . . .	83
6.3	Das Schlögl-Modell als minimales Spielzeugmodell . . . . .	84
6.4	Das Schlögl-Modell in der Transformation . . . . .	85
<b>7</b>	<b>Fazit</b>	<b>87</b>
<b>8</b>	<b>Danksagung</b>	<b>89</b>
	<b>Literatur</b>	<b>91</b>



---

---

# Abkürzungsverzeichnis

Cas9 *CRISPR associated protein 9*

CKE *Chapman-Kolmogorov-Gleichung*

CME *Chemische Mastergleichung*

CRISPR *Clustered Regularly Interspaced Short Palindromic Repeats*

dCas9 *disfunctional Cas9*

DNA *Desoxyribonukleinsäure*

ES  *$\beta$ -Estradiol*

FPE *Fokker-Planck-Gleichung*

GFP *grün fluoreszierendes Protein*

gRNA *guide RNA*

ME *Mastergleichung*

mRNA *messenger RNA*

RBS *ribosomale Bindungsstelle*

RNA *Ribonukleinsäure*

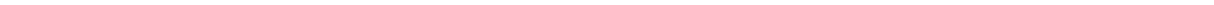
RNAP *RNA-Polymerase*

ROC *Receiver-Operating-Characteristic*

RRE *Reaktionsratengleichung*

scRNA *scaffold RNA*

SSA *Stochastischer Simulationsalgorithmus*



---

# 1 Einleitung

»What I cannot create, I do not understand.«

---

Richard Feynman [1]

Die Synthetische Biologie ist eine interdisziplinäre Wissenschaft, welche die künstliche Herstellung von einzelnen genetischen Schaltkreisen bis hin zu kompletten lebenden Organismen zum Ziel hat. Verglichen mit der verwandten Systembiologie treibt die Synthetische Biologie hierbei das »Erschaffen von Neuem« voran und hebt dabei den Erkenntnisgewinn des Konstruierens hervor [2]. Gut lässt sich dieses Verständnis in Feynmans berühmtem letzten Tafelanschrieb »What I cannot create, I do not understand.« [1] ausdrücken. Demgegenüber verfolgt die klassische Systembiologie eher die Methodik der Dekomposition, um die Funktionsweise einzelner Teile zu analysieren [3]. Ein bekanntes Beispiel für die Arbeit der Synthetischen Biologie stellt in diesem Zusammenhang der Repressilator, ein synthetisch erzeugter genetischer Oszillator, dar. Mit Bezug auf diesen oszillierenden Schaltkreis formulieren Sprinzak et al. in ihrer Veröffentlichung »Reconstruction of genetic circuits« die Motivation der Synthetischen Biologie dabei prägnant in zwei Sätzen: »By taking apart an old clock, you could probably come up with a pretty good guess at how it works. But a more concrete understanding of the clock mechanism might be obtained by designing and building one's own clock out of similar parts« [4].

Um möglichst diverse Anwendungsgebiete abzudecken sowie maximale Flexibilität zu erreichen, verfolgt die Synthetische Biologie in ihren Designansätzen das Konzept der Modularität, welches in anderen Wissenschaften wie etwa der Informatik bereits erfolgreich eingesetzt wird [5–8]. Anstatt ein großes Ganzes zu generieren, werden kleine, möglichst eigenständige Teilmodule mit speziellen Funktionen entwickelt, die später beliebig zu verschiedenen größeren Netzwerken zusammengesetzt werden können.

Die Synthetische Biologie wird allgemein als ein Forschungsgebiet eingestuft, das zwar aus der Biologie hervorgegangen ist, aber auch Ansichten und Methoden aus den Ingenieurwissenschaften und anderen klassischen Naturwissenschaften wie der Physik kombiniert [9]. Um beim Beispiel des Repressilators zu bleiben, war die Implementierung und Integration des genetischen Netzwerks in Hefezellen klassische Biologie, der Entwurf des Schaltkreises sowie die Analyse der Daten erfolgte aber auf Basis physikalischen und mathematischen Verständnisses. In den letzten Jahren wurde dabei immer offensichtlicher, dass die derzeit vorhandenen, vor allem deterministischen mathematischen Verfahren den Anforderungen der Synthetischen Biologie nicht gerecht werden. Da die betrachteten Systeme sowie die Teilchenzahlen oftmals sehr klein sind, bedarf es neuer und schnellerer Modellierungsmethoden, welche die stochastischen Effekte berücksichtigen [10, 11]. Hierbei hat sich insbesondere gezeigt, dass die in zellulären Systemen vorhandenen Fluktuationen – auch Rauschen genannt – oftmals ein zentraler Bestandteil des zellulären Verhaltens sind und daher zwingend in die Modelle einfließen müssen [12].

Vor allem aufgrund der bisherigen Ergebnisse der Systembiologie kam man in der Synthetischen Biologie zu der Erkenntnis, dass für eine erfolgreiche Modellierung und Analyse der beabsichtigten Schaltkreise eine holistische Herangehensweise vonnöten ist. Dies ist ein Analyseverständnis, welches in der theoretischen Physik – hier insbesondere im Bereich der komplexen dynamischen Systeme – bereits seit Längerem verfolgt wird. Komplexe Systeme werden dabei am treffendsten als Systeme definiert, die aus vielen miteinander interagierenden Bestandteilen bestehen. Hierbei gilt insbesondere, dass sich das Verhalten des Gesamtsystems nicht durch die isolierte Analyse seiner Teile vorhersagen lässt [13]. So hat sich etwa die im 20. Jahrhundert

---

noch vorherrschende Meinung, dass jedes Gen genau eine Funktion hat, mittlerweile als falsch herausgestellt. Es wurde klar, dass Funktion und Nutzen verschiedener Gene sich nur im Bild komplexer gegenseitiger Interaktionen verstehen lassen [14].

Vor dem Hintergrund der hier beschriebenen Entwicklungen und Probleme der Synthetischen Biologie wird eine mathematische Beschreibung der biologischen Prozesse immer wichtiger. So lassen sich durch mathematische Modelle zum Beispiel Fragen zum Verhalten eines experimentell noch nicht realisierten Systems beantworten oder durch Vergleich zwischen Modell und Experiment Hypothesen testen. Wie oben bereits ausgeführt, ist es hierzu allerdings notwendig, dass modellierte Systeme nicht isoliert, sondern mit Berücksichtigung der Umgebung analysiert werden. Außerdem müssen vorhandene Fluktuationen auch im Modell beachtet werden. Folglich bedarf es geeigneter theoretischer Methoden, welche die Bedürfnisse der jungen interdisziplinären Wissenschaft erfüllen. Diese Arbeit knüpft hier aus der Perspektive der theoretischen Physik an. Die explizit behandelten Fragestellungen werden dazu im nächsten Abschnitt beschrieben.

---

## 1.1 Ziele und Nutzen

---

Den bisherigen Ausführungen folgend werden in dieser Arbeit drei verschiedene Themen behandelt, welche – jedes für sich – auf bestimmte Fragestellungen der Synthetischen Biologie eingehen. Folgende Fragen stehen hierbei im Fokus:

- Wie lässt sich ein typisches System der Synthetischen Biologie mathematisch modellieren?
- Wie lässt sich die zelluläre Umgebung bei der Modellierung berücksichtigen?
- Welche besonderen Effekte kann schubartiges Rauschen (sogenanntes Burstrauschen) hervorrufen?
- Können abstrakte Modelle Nutzen in der Synthetischen Biologie entfalten?

Die behandelten Themen sind dabei beispielhaft für einen Bereich der Synthetischen Biologie, in dem Naturwissenschaften wie Physik und Mathematik wertvolle Beiträge leisten müssen, damit die Unternehmung Synthetische Biologie ihre gesteckten Ziele erreichen kann.

---

## 1.2 Aufbau der Arbeit

---

Nachdem wir in Kapitel 2 einige für diese Arbeit wesentliche Grundlagen erläutert haben, werden wir in Kapitel 3 einen konkreten biologischen Schaltkreis, welcher auf dem sogenannten CRISPR/dCas9-System beruht, kennenlernen. Bei CRISPR/dCas9 handelt sich um ein bisher vor allem in der Genmanipulation oft genutztes System, welches in dieser Arbeit derart erweitert wird, dass es als AND-Gatter einen genetischen logischen Schalter bereitstellt. Im Sinne der beabsichtigten Modularität ist dieses System so aufgebaut, dass es als eigenständiges Modul möglichst wenig externe Abhängigkeiten aufweist. Ein aktuelles Ziel der Biologie ist die Optimierung des Schaltverhaltens von CRISPR/dCas9-basierten Gattern. Hierzu ist allerdings ein tieferes Verständnis für die zugrunde liegenden Mechanismen vonnöten. Das von uns in dieser Arbeit entwickelte Modell schlägt einen Mechanismus vor, welcher den im Experiment erhaltenen Beobachtungen zugrunde liegen könnte. Anhand verschiedener Modellierungsmethoden zeigen wir, dass die Vorhersagen unseres Modells qualitativ und in vielen Parameterbereichen auch quantitativ gut mit den experimentellen Daten übereinstimmen.

In Kapitel 4 gehen wir dann näher auf die von der Synthetischen Biologie angestrebte, aber oft nur schwer erreichbare Unabhängigkeit der einzelnen Teilmodule ein. Die Wechselwirkungen, welche die Teilmodule mit der Umgebung haben, werden wir hierbei als Kontexteffekte definieren. Ausgehend von einer Einschränkung auf lineare Netzwerke berechnen wir dann analytisch, wie sich verschiedene mögliche Umgebungen auf das betrachtete genetische Modul auswirken, wobei



---

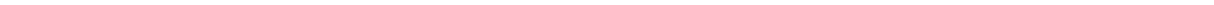
---

wir nicht nur spontane, sondern auch zeitlich verzögerte Wechselwirkungen in die Analyse mit aufnehmen. Verschiedene Beispiele werden hierbei verdeutlichen, wie wichtig die Einbeziehung der Umgebung in die Analyse von Schaltkreisen ist, selbst wenn die Wechselwirkungen mit der Umgebung nur klein sind. Unsere Arbeit gliedert sich damit in das weiter oben bereits ausgeführte Verständnis der Synthetischen Biologie ein, dass biologische Systeme, obwohl modularisiert, dennoch holistisch analysiert und simuliert werden müssen und liefert Methoden, dies effizient durchzuführen.

Dass das in biologischen Systemen vorhandene Rauschen, also zufällige Fluktuationen, starke Auswirkungen auf das Verhalten des betrachteten Systems haben kann, zeigt sich besonders gut in Kapitel 5. Aus verschiedenen Arbeiten ist bereits bekannt, dass Rauschen dazu führen kann, dass klassische Methoden wie deterministische Beschreibungen das Systemverhalten nicht korrekt vorhersagen. Zur Veranschaulichung und näheren Analyse einer speziellen Art des Rauschens, dem sogenannten Burstrauschen, definieren wir ein Reaktionssystem, bei dem die Stärke impulsartiger Fluktuationen über einen Parameter einstellbar ist, ohne dass dabei die deterministische Beschreibung verändert wird. Hierbei zeigt sich, dass die Fluktuationen zum einen klassisch vorhergesagte stabile Zustände zerstören, aber auch neue stabile Zustände erzeugen können. Insbesondere zeigt unsere Analyse, dass die impulsartige Produktion von Molekülen zu einem Systemverhalten führen kann, welches durch deterministische Methoden nicht richtig vorhergesagt wird. Unser Modell bietet dabei als minimales Modell einen Ausgangspunkt für weitere Untersuchungen von durch Rauschen induzierten Bifurkationen.

Das von uns in Kapitel 5 verwendete Modell zur Analyse von Burstrauschen kann nicht als konkretes Modell der Synthetischen Biologie verstanden werden, da es eine trimolekulare Reaktion enthält, welche so in der Natur nicht existiert. Nach biologischen und ingenieurwissenschaftlichen Gesichtspunkten hat das Modell also wenig, oder sogar keinen Nutzen. Aus physikalischer Sicht muss ein Modell allerdings nicht zwingend ein reales System abbilden, sondern kann, so wie in unserem Beispiel, auch nur einen interessanten Effekt zeigen. Somit besteht unweigerlich ein Konflikt in den erkenntnistheoretischen Zielen der verschiedenen wissenschaftlichen Disziplinen, welche in der Synthetischen Biologie vereint sind. Ob und wie unsere Arbeit und allgemein theoretische minimale Modelle dennoch Nutzen für die Synthetische Biologie haben können, werden wir in Kapitel 6 diskutieren.

Abschließend folgt in Kapitel 7 eine Zusammenfassung der Arbeit.



---

## 2 Grundlagen

---

### 2.1 Biologische Grundlagen

---

Diese Arbeit bewegt sich im Kontext der Synthetischen Biologie. Wie bereits im vorigen Kapitel beschrieben, handelt es sich hierbei um ein Teilgebiet der Biologie, das sich mit der künstlichen Erzeugung und Implementierung von sogenannten Genregulationsnetzwerken in natürliche Organismen, wie etwa Zellen, befasst. Im Folgenden sollen daher die wichtigsten biologischen Grundlagen und deren mathematische Formulierung aufgezeigt werden. Zu dem CRISPR-System, welches in Kapitel 3 behandelt wird, folgt in der dortigen Einleitung eine spezifische Erklärung.

---

#### 2.1.1 Biologische Zellen

---

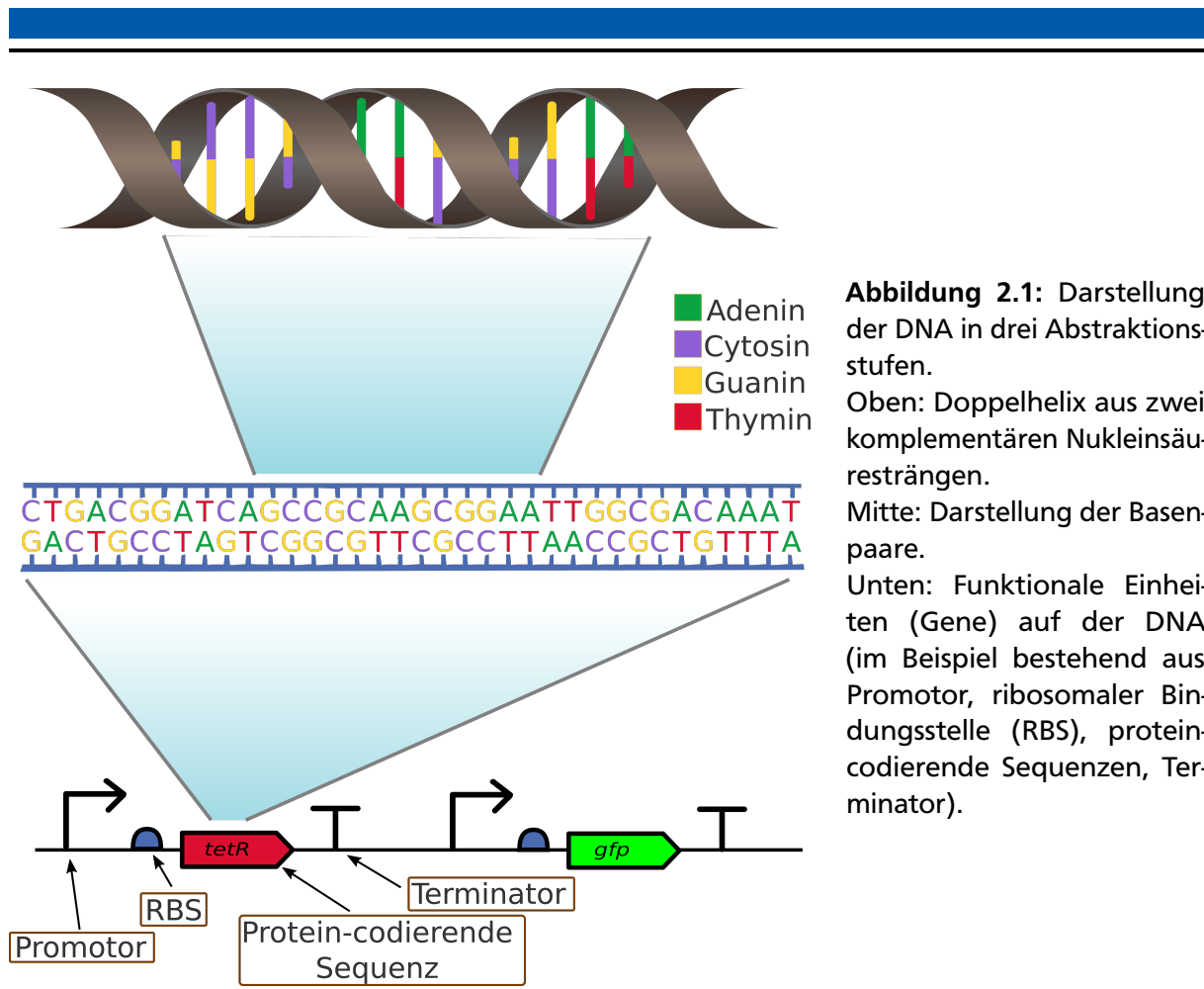
Die meisten Lebewesen, denen man im Alltag begegnet, sind in einer hierarchischen Struktur aufgebaut, welche aus kleineren Untereinheiten besteht. So haben z. B. alle Säugetiere verschiedene Organe, welche wiederum aus verschiedenen Gewebearten zusammengesetzt sind. Gewebe wiederum werden von bestimmten Zellen gebildet, wobei die Zelle hierbei die kleinste lebende Untereinheit dieser Hierarchie darstellt. All diese Lebewesen bestehen also aus einer großen Zahl von Zellen, welche zu funktionellen Einheiten verbunden sind. Man spricht deshalb von *Vielzellern* (ein erwachsener Mensch besteht zum Beispiel aus ca.  $10^{13}$  Zellen [15]) [16]. Obwohl es sich immer um Zellen handelt, haben jene, welche z. B. die menschliche Haut bilden, andere Eigenschaften und Funktionen als jene, welche z. B. einen Muskel bilden. Die Zellen haben verschiedene spezialisierte Zustände entwickelt, was man als *Differenzierung* bezeichnet [17]. Solche spezialisierten Zellen sind alleine nicht überlebensfähig, sondern abhängig von den umgebenden anderen hierarchischen Zellstrukturen.

Anders ist dies bei den sogenannten *Einzellern*, also jenen Lebewesen, welche nur aus einer einzigen Zelle bestehen. Einzeller finden sich beispielsweise unter den Pilzen und Bakterien, wobei der wahrscheinlich bekannteste Vertreter unter den Pilzen die Backhefe *Saccharomyces cerevisiae* (*S. cerevisiae*) und unter den Bakterien das Darmbakterium *Escherichia coli* (*E. coli*) ist.

Die Zellen aller Lebewesen haben unter anderem drei Dinge gemeinsam: Sie besitzen eine *Zellmembran*, welche die Zelle von der Umgebung abtrennt, sie sind gefüllt mit dem sogenannten *Cytoplasma* und sie enthalten *Desoxyribonukleinsäure* (DNA) [17]. Der Aufbau und die Bedeutung der DNA werden später in diesem Kapitel noch genauer erklärt.

Alle Lebewesen lassen sich abhängig vom internen Organisationsgrad ihrer Zellen in drei Kategorien einordnen, nämlich in *Eukaryoten*, *bakterielle Prokaryoten* und *Archaeen* [16]. Die eukaryotischen Zellen besitzen eine intrazelluläre Unterteilung in sogenannte *Organellen*, welche spezielle Aufgaben der Zelle übernehmen. Das wichtigste Organell der Eukaryoten ist der *Zellkern* (*Nukleus*), welcher über eine Kernhülle vom Rest der Zelle getrennt ist und die chromosomale DNA enthält. Aufgrund der höheren Organisationsstruktur können eukaryotische Zellen größer werden als Prokaryoten. Alle Tiere und Pflanzen sind Eukaryoten. Die Prokaryoten haben keinen Zellkern und auch keine sonstigen Organellen. Prokaryotische Zellen sind immer Einzeller und umfassen die Bakterien. Die weiterhin noch vorhandenen Archaeen stellen eine Form von Lebewesen dar, welche sich zwischen Eukaryoten und Prokaryoten befindet [18].

Allen Zellen, ob Teil eines Vielzelllers oder für sich als Einzeller, ist gemein, dass sie in der Lage sind Nährstoffe zu verarbeiten sowie auf Änderungen der Umwelt zu reagieren [19]. Die Backhefe *S. cerevisiae* ernährt sich beispielsweise vorzugsweise von Glucose (auch bekannt als Traubenzucker). Wenn in der Umgebung allerdings nur Galactose (Schleimzucker) vorhanden ist,

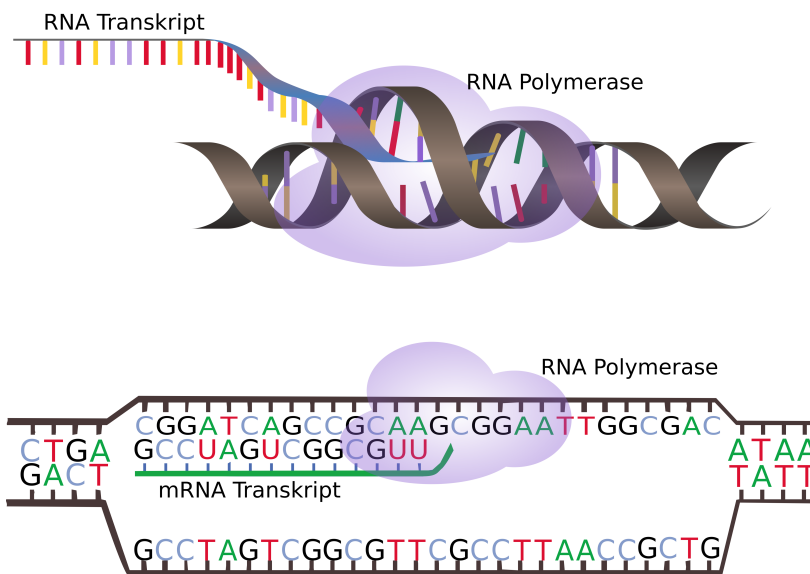


**Abbildung 2.1:** Darstellung der DNA in drei Abstraktionsstufen. Oben: Doppelhelix aus zwei komplementären Nukleinsäuresträngen. Mitte: Darstellung der Basenpaare. Unten: Funktionale Einheiten (Gene) auf der DNA (im Beispiel bestehend aus Promotor, ribosomaler Bindungsstelle (RBS), protein-codierende Sequenzen, Terminator).

kann die Hefe in kurzer Zeit ihren Stoffwechsel so umstellen, dass sie Galactose metabolisieren kann. Wie eine solche Anpassung abläuft, wollen wir im Folgenden beispielhaft erklären. Wir beziehen uns dabei auf den Vorgang, wie er in eukaryotischen Zellen auftritt. Die wesentlichen Abläufe sind allerdings übertragbar auf Prokaryoten und Archaeen. Wenn nicht gesondert angegeben, sind die beschriebenen Vorgänge den Quellen [16, 17, 20] entnommen.

### 2.1.2 DNA und Proteine

Die Moleküle, welche in einer Zelle zentrale Funktionen ausführen, sind die sogenannten *Proteine*. Wenn eine Hefezelle Funktionen wie etwa ihren Stoffwechsel umstellt, dann geht dies folglich einher mit der *Expression*, also der Herstellung von speziellen Proteinen. Im Falle der Umstellung von Glucose auf Galactose benötigt die Zelle beispielsweise Proteine, welche Galactose verarbeiten können. Alle Informationen, welche eine Zelle zur Expression von Proteinen benötigt, sind in der DNA enthalten. Hierbei handelt es sich, wie in Abbildung 2.1 in braun dargestellt, um zwei in einer Doppelhelix angeordnete Zucker-Phosphat-Stränge, welche oft als *Phosphatrückgrat* bezeichnet werden. An den Zucker-Molekülen hängt jeweils eine von vier möglichen Basen (Adenin und Thymin, sowie Cytosin und Guanin), welche über Wasserstoffbrückenbindungen mit der jeweils komplementären Base am anderen Strang verbunden ist. Diese DNA-Stränge lassen sich unterteilen in codierende und nichtcodierende Bereiche. Der Nutzen und die Funktion der nichtcodierenden Bereiche ist Gegenstand kontroverser Diskussionen und bisher in großen Teilen unverstanden. Wir wollen uns daher in der folgenden Arbeit auf die protein-codierenden Bereiche beschränken. Die Abfolge der verschiedenen Basen der codierenden Bereiche bildet den *genetischen Code*, welcher die Erbinformationen der Zelle enthält. Basierend auf Unterabschnitten der codierenden DNA, den sogenannten *Genen*, können verschiedenste Proteine hergestellt werden.



**Abbildung 2.2:** Darstellung der Transkription in zwei Abstraktionsstufen.

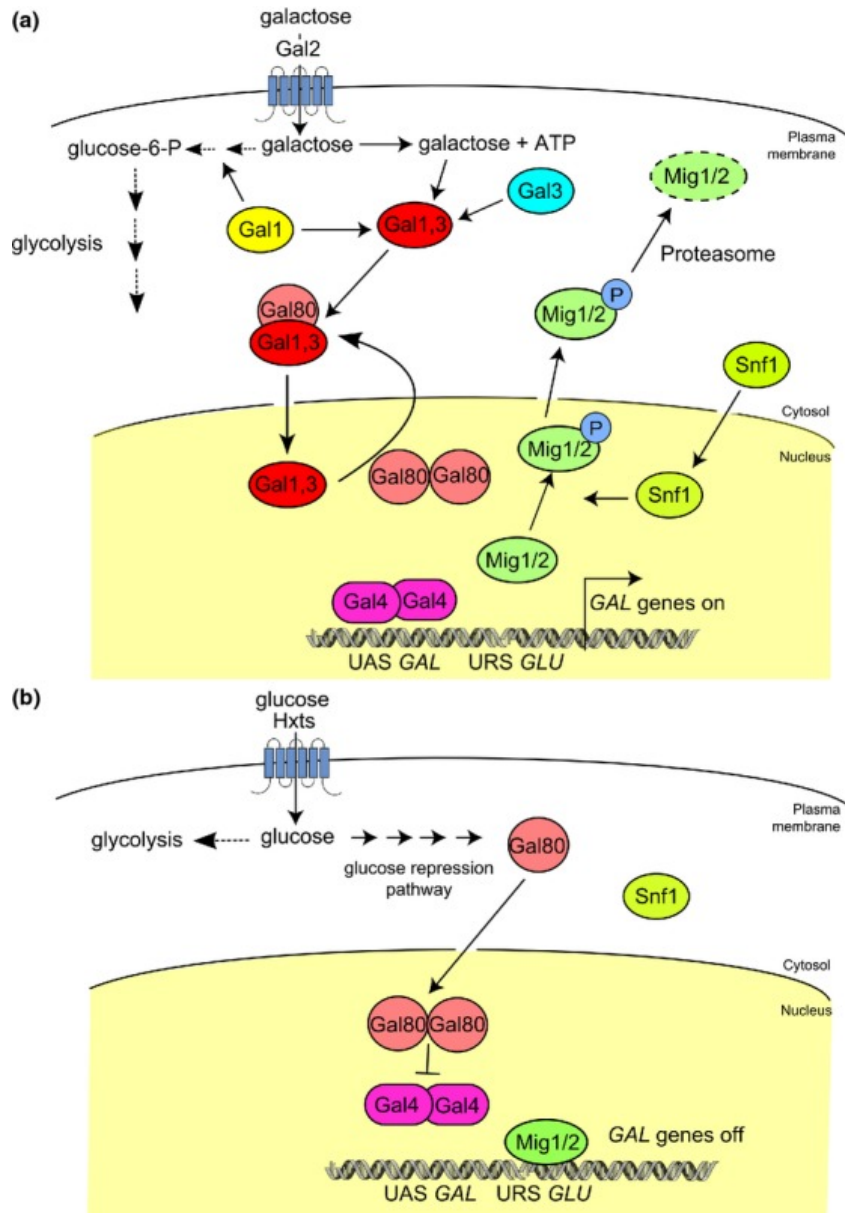
Oben: Die RNA Polymerase bricht die Doppelhelix auf und erstellt das mRNA-Transkript

Unten: Zweidimensionale Darstellung der Transkription

Hierunter gibt es beispielsweise auch einen Satz von sogenannten GAL-Genen [21]. Diese stellen den Code für jene Proteine bereit, die die Hefezelle zur Nutzung von Galactose benötigt. Die Synthese von Proteinen, ausgehend von der auf der DNA gespeicherten Information, geschieht in zwei wesentlichen aufeinander folgenden Schritten. Im ersten Schritt, der *Transkription*, wird die DNA von einem speziellen Protein, der *RNA-Polymerase* (RNAP) ausgelesen. Bei diesem, in Abbildung 2.2 dargestellten Vorgang, setzt sich die RNAP an eine am Beginn des Gens liegende DNA-Sequenz, den sogenannten *Promotor*. Während die RNAP dann bis zum *Terminator* über das Gen wandert, trennt sie die DNA-Stränge kurz auf und bildet eine einsträngige Kopie des DNA-Codes, die sogenannte *messenger RNA* (mRNA). Diese mRNA verlässt nach einer weiteren Prozessierung den Zellkern und gelangt ins Cytoplasma. Hier kommt es zur *Translation*. Dabei setzt ein Komplex aus Ribonukleinsäuren (RNAs) und Proteinen, das *Ribosom*, an der sogenannten *ribosomalen Bindungsstelle* (RBS) an die mRNA an, liest diese aus und verbindet gemäß des dort gespeicherten Codes verschiedene Aminosäuren miteinander. Die resultierende Kette von Aminosäuren unterliegt nach deren Herstellung meist noch einigen Modifikationen, wie etwa Faltungen, bis sie der Zelle als funktionsfähiges Protein zur Verfügung steht. Die so entstehenden Proteine können – abhängig von der zugrunde liegenden DNA-Sequenz – die verschiedensten Funktionen haben. So gibt es etwa Proteine, welche Strukturinformationen tragen und somit durch Anlagerung innerhalb der Zelle oder an deren Membran für bestimmte strukturelle Eigenschaften verantwortlich sind. Andere Proteine dienen zum Transport verschiedener Stoffe oder können Reaktionen katalysieren (man spricht dann von einem *Enzym*). Die in unserem Beispiel erwähnten GAL-Proteine katalysieren so etwa verschiedene Reaktionen, die Galactose in Glucose umsetzen. Nach deren Expression ist die Zelle also fähig, Galactose als Kohlenstoffquelle zu nutzen.

### 2.1.3 Genregulationsnetzwerke

Weiter oben wurde bereits erwähnt, dass Hefezellen Glucose bevorzugen. Um nicht unnötig Energie zu verbrauchen, wird die Expression der GAL-Gene daher unterdrückt, wenn Glucose vorhanden ist und aktiviert, wenn keine Glucose, dafür aber Galactose detektiert wird. Dies funktioniert, da bestimmte Proteine regulatorisch als sogenannte *Transkriptionsfaktoren* wirken können, d. h. sie aktivieren oder inhibieren spezifische Transkriptionen. Hierzu binden Transkriptionsfaktoren meist am oder in der Nähe des Promotors reversibel an die DNA. Diese Regulierung der Transkription durch Proteine bildet häufig sehr komplexe Schaltungen, in denen z. B. erst zwei Proteine unter bestimmten Voraussetzungen zu einem Dimer binden müssen und dann



**Abbildung 2.3:** Schematische Darstellung eines kleinen Genregulationsnetzwerkes, welches für die Aktivierung der GAL-Gene zuständig sein könnte. (a) Die Zelle befindet sich in Galactose-Medium. (b) Die Zelle befindet sich in Glucose-Medium.

Bildquelle: [22], Lizenz: CC BY-NC-ND

eine regulative Funktion übernehmen. Auch sind positive und negative Selbstbeeinflussung ein häufiges Phänomen. Aus diesem Grund spricht man von *Genregulationsnetzwerken*. Im Beispiel des gerade diskutierten GAL-Systems könnte das komplette Regulationsnetzwerk beispielsweise aussehen wie in Abbildung 2.3 dargestellt, wobei die beteiligten Proteine hier als bunte Blasen und Interaktionen durch Pfeile dargestellt sind.

## 2.2 Genregulationsnetzwerke im Modell

Diese Arbeit befasst sich mit der Modellierung der gerade beschriebenen Genregulationsnetzwerke. Es geht also darum, wie sich beobachtete oder vermutete Genregulationsnetzwerke durch ein mathematisches Modell beschreiben und basierend auf dem Modell Vorhersagen über deren

Eigenschaften machen lassen. Dieses Kapitel dient als Einführung in die formale Darstellung von biochemischen Reaktionen sowie deren Simulation. Im darauffolgenden Kapitel wird dann näher auf die nötigen mathematischen Grundlagen eingegangen.

### 2.2.1 Reaktionsgleichungen

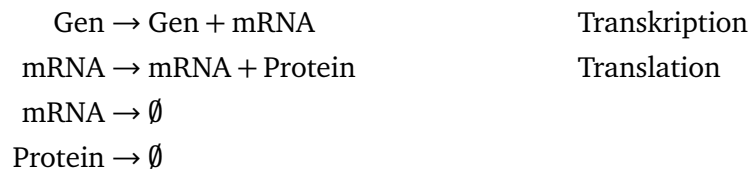
Um biochemische Wechselwirkungen, wie sie in der biologischen Zelle stattfinden, zu skizzieren, bietet sich die Schreibweise durch sogenannte *Reaktionsgleichungen* an [23]. Hierbei stehen links von dem sogenannten *Reaktionspfeil*  $\rightarrow$  die *Reaktanten* des Vorgangs und rechts davon die *Produkte* [24]. Einen Satz von  $M$  Reaktionen mit  $N$  beteiligten Substanzen  $\mathbb{X}_1, \mathbb{X}_2, \dots, \mathbb{X}_N$  schreiben wir im Folgenden als:



wobei  $j = 1, \dots, M$  die Reaktionen nummeriert. Die *stöchiometrische Konstante*  $s_{ij}$  gibt die Menge des Reaktanten  $\mathbb{X}_i$  in Reaktion  $j$  an und  $r_{ij}$  entsprechend die Anzahl der Produkte. An dieser Stelle definieren wir noch eine weitere Größe, welche im Laufe dieser Arbeit die Schreibweise vereinfachen wird: Die *stöchiometrische Matrix*  $\mathbf{S}_{ij} = r_{ij} - s_{ij}$  gibt für jede Reaktion  $j$  an, wie sich bei deren Ablauf die Mengen der verschiedenen Substanzen  $i$  ändern.

### Beispiel

Wenn wir die im vorausgegangenen Abschnitt beschriebene Transkription und Translation durch Reaktionsgleichungen angeben wollen, so erhalten wir:



wobei wir hier noch eine Degradation von mRNA und Protein eingefügt haben.

### 2.2.2 Reaktionsraten und Zustandsvektoren

Die Angabe der Reaktionsgleichungen gibt bereits einen Überblick über die möglichen Interaktionen zwischen den Bestandteilen des betrachteten Systems. Um die Dynamik, also das zeitliche Verhalten, eines Systems zu beschreiben, fehlen aber noch zwei wichtige Informationen: Zum einen benötigen wir eine Anfangsbedingung, also die Anzahl der verschiedenen Substanzen zum Zeitpunkt  $t = t_0$ , zum anderen benötigen wir Informationen darüber, wie oft und wann welche Reaktion in gegebenen Zeitabschnitten abläuft. Hierbei ist allerdings zu beachten, dass natürliche Systeme immer auch durch Wechselwirkungen beeinflusst werden, welche nicht exakt vorhersagbar sind. Daher können wir niemals genaue Zeitpunkte für bestimmte Reaktionen angeben, sondern nur deren Wahrscheinlichkeiten. Anstatt exakte Vorhersagen über zeitliche Abfolgen von Reaktionen zu tätigen, müssen wir also die vorhandene Zufälligkeit beachten. Zur generellen Vereinfachung der Modelle werden wir dabei eine Annahme treffen, welche von zentraler Bedeutung für alle weiteren Betrachtungen ist: Wir setzen voraus, dass das betrachtete System gut durchmischt ist und sich im thermischen Gleichgewicht befindet [25]. Jegliche Information über die Position von einzelnen Molekülen soll für uns also nicht von Interesse sein. Ein System mit  $N$  verschiedenen Spezies lässt sich dann vollständig durch den *Zustands-*

vektor  $\mathbf{n} = (n_1, \dots, n_N)^T$  beschreiben, wobei  $n_i$  die jeweilige Anzahl der Substanz  $\mathbb{X}_i$  angibt. Wenn die Reaktion  $R_j$  abläuft, dann verändern sich die einzelnen Elemente des Zustandsvektors entsprechend der zuvor definierten stöchiometrischen Matrix:  $\mathbf{n} := \mathbf{n} + \mathbf{S}_{\cdot j}$ , wobei wir mit  $\mathbf{S}_{\cdot j}$  die  $j$ -te Spalte der Matrix  $\mathbf{S}$  bezeichnen. Basierend auf der Annahme der guten Durchmischung kann man für jede Reaktion  $R_j$  eine Ratenfunktion  $f_j(\mathbf{n})$  angeben, für die gilt:

$$f_j(\mathbf{n}) dt = \text{Die Wahrscheinlichkeit dafür, dass} \quad (2.2)$$

bei gegebenem Zustand  $\mathbf{n}$  innerhalb des Zeitintervalls  $[t, t + dt)$  die Reaktion  $R_j$  abläuft.

Diese Ratenfunktion, die sogenannte *Propensität*  $f_j$ , ist dabei gegeben durch

$$f_j(\mathbf{n}) = c_j h_j(\mathbf{n}), \quad (2.3)$$

wobei  $c$  als *Reaktionskonstante* und  $h$  als *kombinatorischer Faktor* bekannt sind. Hierbei ist  $c_j dt$  die Wahrscheinlichkeit dafür, dass eine bestimmte Kombination der Reaktanten von  $R_j$  miteinander reagiert und  $h_j$  gibt, basierend auf  $\mathbf{n}$ , alle möglichen Kombinationen der Reaktanten an. Für die allgemeine Formulierung der Reaktionen wie in Gleichung (2.1) ergibt sich somit:

$$h_j(\mathbf{n}) = \binom{n_1}{s_{1j}} \cdots \binom{n_i}{s_{ij}} = \prod_{i=1}^N \frac{n_i!}{s_{ij}!(n_i - s_{ij})!}. \quad (2.4)$$

Je nach Konvention werden hierbei die  $s_{ij}!$  oftmals auch in den  $c_j$  absorbiert.

Für gewöhnlich werden die Reaktionskonstanten über den entsprechenden Reaktionspfeil geschrieben [24], sodass die vollständige Angabe der Reaktionen eines Systems lautet:




---

### 2.2.3 Simulation der Reaktionsgleichungen

---

Wir haben nun einen Satz von Reaktionsgleichungen definiert sowie die zugehörigen Wahrscheinlichkeiten dafür, dass eine bestimmte Reaktion in einem Zeitfenster  $dt$  abläuft, eingeführt. Diese Informationen reichen aus, um den zeitlichen Verlauf eines betrachteten Systems zu simulieren. Die zwei zentralen Fragen hierzu sind: »Wann tritt die nächste Reaktion auf?« und »Welches ist die nächste Reaktion?«. Eine detaillierte Herleitung zur Beantwortung dieser Fragen wurde zuerst von Daniel Gillespie veröffentlicht, weshalb der im Folgenden kurz hergeleitete *Stochastische Simulationsalgorithmus* (SSA) oft auch *Gillespie-Algorithmus* genannt wird. Die im Weiteren beschriebene Herleitung orientiert sich dabei auch direkt an den ursprünglichen Arbeiten von Daniel Gillespie [25, 26].

Wir beginnen damit die Wahrscheinlichkeitsverteilung  $P(\mathbf{n}, \tau) d\tau$  herzuleiten, welche die Wahrscheinlichkeit dafür angibt, dass in einem Zeitraum  $(t, t + \tau)$  keine Reaktion stattfindet und im darauf folgenden Zeitintervall  $(t + \tau, t + \tau + d\tau)$  irgendeine der möglichen Reaktionen. Die Größe  $P(\mathbf{n}, \tau) d\tau$  lässt sich schreiben als:

$$P(\mathbf{n}, \tau) d\tau = P_0(\mathbf{n}, \tau) \cdot \sum_j f_j(\mathbf{n}) d\tau, \quad (2.6)$$



wobei  $P_0(\mathbf{n}, \tau)$  die Wahrscheinlichkeit für keine Reaktion im Zeitabschnitt  $\tau$  angibt und  $\sum_j f_j(\mathbf{n}) d\tau$  (gemäß Gleichung (2.2)) die Wahrscheinlichkeit für irgendeine Reaktion im Zeitabschnitt  $d\tau$  beschreibt.

Weiter oben haben wir  $c_j dt$  als die Wahrscheinlichkeit dafür eingeführt, dass bei gegebenem Zustand  $\mathbf{n}$  innerhalb des Zeitintervalls  $(t, t + dt)$  eine bestimmte der  $h_j$  möglichen Kombinationen der Reaktanten von  $R_j$  miteinander reagieren. Demnach können wir durch  $(1 - c_j dt)$  die Wahrscheinlichkeit dafür angeben, dass bei gleichem Zustand diese Reaktanten nicht miteinander reagieren. Folglich gilt für die Wahrscheinlichkeit, dass keine der möglichen Kombinationen miteinander reagiert:

$$(1 - c_j dt)^{h_j} = 1 - h_j c_j dt - \mathcal{O}(dt^2). \quad (2.7)$$

Die Wahrscheinlichkeit dafür, dass keine aller möglicher Reaktionen passiert, ist somit:

$$\prod_j (1 - h_j c_j dt + \mathcal{O}(dt^2)) = 1 - \sum_j h_j c_j dt - \mathcal{O}(dt^2) = 1 - \sum_j f_j(\mathbf{n}) dt - \mathcal{O}(dt^2). \quad (2.8)$$

Die Terme der quadratischen und höheren Ordnung in  $dt$  spiegeln hierbei die Möglichkeit wider, dass nicht nur eine, sondern mehrere Reaktionen im Zeitintervall  $dt$  ablaufen. Für kleines  $dt$  wird die Wahrscheinlichkeit für diesen Fall immer kleiner und kann vernachlässigt werden. Wenn nun bei gegebener Zeit  $t$  der Zustand  $\mathbf{n}$  die Wahrscheinlichkeit  $P(\mathbf{n}, t)$  besitzt, dann gilt für die Wahrscheinlichkeit, zu einem späteren Zeitpunkt immer noch im gleichen Zustand zu sein:

$$P(\mathbf{n}, t + dt) = P(\mathbf{n}, t) \left( 1 - \sum_j f_j dt \right) + \mathcal{O}(dt^2). \quad (2.9)$$

Nach Subtraktion von  $P(\mathbf{n}, t)$  und Durchführung des Übergangs  $\lim_{dt \rightarrow 0}$  folgt damit für die Wahrscheinlichkeit, dass zum Zeitpunkt  $t$  noch keine Reaktion stattgefunden hat:

$$P_0(\mathbf{n}, t) = e^{-\sum_j f_j t}. \quad (2.10)$$

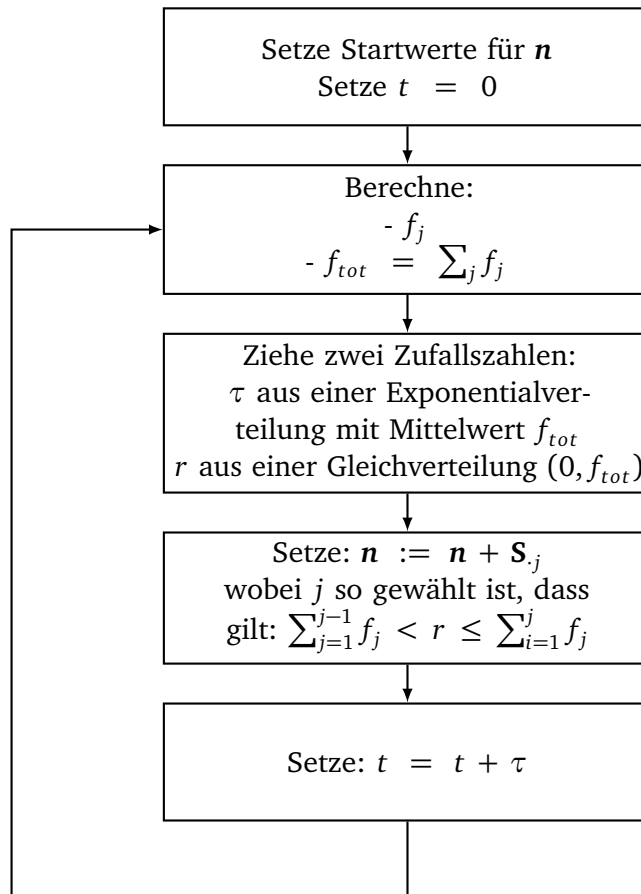
Eingesetzt in Gleichung (2.6) ergibt sich damit, dass die Zeiten zwischen zwei Reaktionen exponentialverteilt sind mit Mittelwert

$$f_{tot}^{-1} = \left( \sum_j f_j \right)^{-1}. \quad (2.11)$$

Nachdem wir nun wissen, wie die Zeiten zwischen den Reaktionen verteilt sind, stellt sich uns noch die Frage, welche der möglichen Reaktionen ausgeführt wird. Da die Wahrscheinlichkeiten der einzelnen Reaktionen proportional zu  $f_j$  sind, lässt sich die Wahrscheinlichkeit für jede Reaktion sofort angeben durch:

$$P_j = f_j / f_{tot}. \quad (2.12)$$

Basierend auf den gerade hergeleiteten Ergebnissen lässt sich nun für ein betrachtetes System der zeitliche Verlauf der Reaktionen und damit verbunden der vorhandenen Teilchenzahlen simulieren. Der SSA verläuft hierbei nach dem in Abbildung 2.4 gezeigten einfachen Schema ab. Unter der Annahme, dass das betrachtete System gut durchmischt ist, sich im thermodynamischen Gleichgewicht befindet und alle relevanten Wechselwirkungen durch die Reaktionsgleichungen abgebildet werden, lassen sich mit dem SSA somit mögliche Trajektorien, also zeitliche Verläufe eines gegebenen Reaktionssystems simulieren.



**Abbildung 2.4:** Ablaufdiagramm für den Gillespie-Algorithmus (SSA)

Obwohl die vorgestellte Methode im Rahmen ihrer Annahmen exakt ist, gibt es zahlreiche Nachteile dieser direkten Simulation:

- Der SSA erzeugt immer nur eine mögliche Trajektorie des Systems. Aufgrund der enthaltenen stochastischen Komponenten sind Trajektorien – wie auch in der Natur beobachtet – nie gleich, sondern hängen vom Zufall ab. Um Informationen über das mittlere Verhalten oder gar die Wahrscheinlichkeitsverteilung zu bestimmten Zeitpunkten zu erhalten, sind daher zahlreiche Läufe notwendig, was die Methode zeitaufwendig macht. Auch wenn mittlerweile verschiedene Algorithmen entwickelt wurden, die den SSA beschleunigen ([27–29]), kann die Simulation doch nur für kleine Systeme angewendet werden.
- Durch eine durchgeführte Simulation erhält man keinerlei Information über das Verhalten des Systems bei Änderung von Input-Parametern wie z. B. initiale Konzentrationen oder Raten. Für jede Änderung an diesen Parametern muss die Simulation erneut gestartet werden.
- Ein damit verbundenes Problem ergibt sich, wenn bestimmte Parameter eines beobachteten Systems abgeschätzt werden sollen, wenn also Messdaten gegeben sind und anhand der aufgenommenen Trajektorien die Parameter eines zugehörigen Modells bestimmt werden sollen. Hierzu wird normalerweise der zeitliche Verlauf des bekannten Modells durch Variation der Parameter an die Beobachtung angepasst. Da aber jede Simulation eines stochastischen Modells (auch bei gleichen Parametern) einen anderen Kurvenverlauf hervorbringt, scheitert diese Methode.

Um die oben genannten Nachteile der Simulation zu umgehen, werden wir in dieser Arbeit Methoden entwickeln und verwenden, welche auf analytische Weise Aussagen über stochastische Systeme ermöglichen. Dabei werden wir immer wieder auf den Gillespie-Algorithmus zurückkommen und diesen nutzen, um verschiedene Reaktionssysteme zu simulieren. Dies geschieht dann meist, um erhaltene theoretische Ergebnisse zu validieren.

Um den weiteren Betrachtungen ein theoretisches Fundament zu geben, werden wir im folgenden Kapitel zuerst die mathematischen Grundlagen der verwendeten Methoden einführen. Basierend auf diesen Grundlagen werden wir dann Gleichungen wie die Fokker-Planck-Gleichung oder die Ratengleichungen herleiten.

---

## 2.3 Theoretische Grundlagen

---

Im obigen Kapitel haben wir ausgeführt, warum und wie Zufallsprozesse das Verhalten eines Reaktionssystems beeinflussen. Basierend auf dem vorausgegangenem Beispiel werden im Folgenden grundlegende Begriffe der Wahrscheinlichkeitstheorie etwas genauer erläutert. Zugunsten eines intuitiven Zugangs wird aber auf eine detaillierte allgemeingültige mathematische Definition verzichtet. Die benötigten Begrifflichkeiten werden nur kurz eingeführt und direkt auf für uns relevante Fälle eingeschränkt. Wir werden uns hierbei insbesondere auf Systeme mit kontinuierlicher Zeit und diskreten Zuständen beschränken. Für eine allumfassende Erklärung sei auf jene Literatur verwiesen, welche auch hauptsächlich für dieses Kapitel herangezogen wurde [30–33].

---

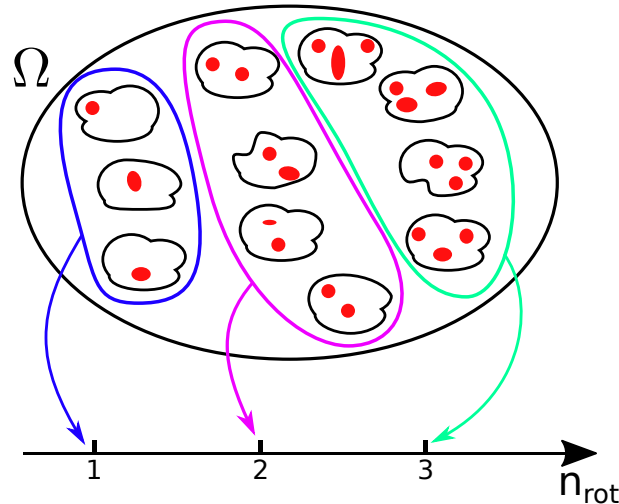
### 2.3.1 Ergebnismenge und Zufallsvariablen

---

Wenn wir uns ein System verschiedener Substanzen, wie etwa eine biologische Zelle, vorstellen, so ist es uns vielleicht möglich, mit geeignetem Instrument die Menge bestimmter, in der Zelle vorhandener Substanzen zu bestimmen. Durch verbesserte Methoden könnte unsere Messung aber auch zusätzlich deren Position, Geschwindigkeit oder detailliertere atomare Informationen hervorbringen. Jede Messung liefert uns als Ergebnis  $\omega$  gewisse Informationen über die beobachtete Zelle. Da die mikroskopischen Zustände der Zelle sowie das daraus resultierende Verhalten niemals exakt erfasst werden können, enthält das Messergebnis immer eine Zufallskomponente bzw. einen Grad der Unwissenheit. Aus diesem Grund muss man eine solche Messung als *Zufallsexperiment* auffassen. Alle im Experiment möglichen Ergebnisse bilden dabei die *Ergebnismenge*  $\Omega$  [34]. Die Ergebnismenge  $\Omega$  ist im Allgemeinen unendlich groß und besitzt beliebige Freiheitsgrade, welche durchaus abstrakte Repräsentanten darstellen können. Für eine mathematische Behandlung wünschen wir uns aber geordnete und vergleichbare Größen, am besten Zahlen oder Vektoren. Im vorigen Kapitel über den SSA haben wir beispielsweise begründet, warum wir für unsere Betrachtungen den Zustand des Reaktionssystems durch  $\mathbf{n}$  ausreichend genau angeben können. Da Stoffmengen nicht negativ werden, könnte unser Zustandsraum also durch  $\{\mathbf{n} | n_i \in \mathbb{N}_0\}$  gegeben sein. Wir benötigen demnach eine Funktion  $Z(\Omega)$ :

$$Z : \Omega \rightarrow Z(\Omega) : \omega \rightarrow \mathbf{n}, \quad (2.13)$$

welche die Ergebnismenge auf einen messbaren und für uns verwertbaren Raum abbildet. In Abbildung 2.5 ist eine mögliche solche Funktion grafisch dargestellt. Welchen exakten Zustand  $\omega$  annimmt und somit auch, welchen Funktionswert (auch genannt *Realisierung*)  $Z(\Omega)$  letztlich annimmt, hängt gemäß der Wahrscheinlichkeiten der einzelnen Ergebnisse vom Zufall ab, weshalb  $Z$  als *Zufallsvariable* bezeichnet wird. Wichtig ist hier zu bemerken, dass der komplette Zustand des betrachteten Systems mikroskopisch durch eine viel größere Zahl an Freiheitsgraden bestimmt ist und nur durch die nicht-injektive Zufallsvariable auf einen messbaren Raum abgebildet wird. Die Wahrscheinlichkeit  $P(Z(\omega) = z)$ , dass  $Z(\omega)$  den Wert  $z$  annimmt, wird



**Abbildung 2.5:** Beispiel für eine nicht-injektive Zufallsvariable. Die Ergebnismenge  $\Omega$  wird hierbei auf die Menge der natürlichen Zahlen abgebildet, welche die Anzahl an roten Substanzen in den jeweiligen Zellen repräsentiert.

durch eine Wahrscheinlichkeitsverteilung  $P$  gegeben. In Zukunft werden wir diese Schreibweise abkürzen und schreiben:

$$P(Z(\omega) = z) = P(z). \quad (2.14)$$

### 2.3.2 Bedingte Wahrscheinlichkeiten

An verschiedenen Stellen kommen wir in den folgenden Kapiteln mit sogenannten *bedingten Wahrscheinlichkeiten* in Berührung. Eine bedingte Wahrscheinlichkeit beschreibt dabei die Wahrscheinlichkeit für ein Ereignis  $A$  unter der Bedingung, dass das Vorliegen von Ereignis  $B$  bekannt ist. Geschrieben wird diese bedingte Wahrscheinlichkeit als  $P(A|B)$ . Ohne näher auf die Herleitung einzugehen, geben wir hier die Definition für bedingte Wahrscheinlichkeiten an [34]:

$$P(A|B) = \frac{P(A, B)}{P(B)}, \quad (2.15)$$

wobei  $P(A, B)$  die Verbundwahrscheinlichkeit dafür ist, dass die Ereignisse  $A$  und  $B$  gemeinsam auftreten.

### 2.3.3 Stochastischer Prozess

Ein *stochastischer Prozess* in kontinuierlicher Zeit wird im Folgenden als eine Abfolge von Zufallsvariablen  $Z_t$  mit  $t \in \mathbb{R}_+$  verstanden, welche nach aufsteigender Zeit  $t$  sortiert sind.

Die verschiedenen  $\Omega_t$  wie auch die zugrunde liegende Wahrscheinlichkeitsverteilung  $P_t$  können sich im Laufe des Prozesses ändern und die verschiedenen  $Z_t$  können beliebig voneinander abhängen.

## Beispiel

Ein einfaches Beispiel für eine gegenseitige Abhängigkeit der  $Z_t$  ist das Ziehen aus einer Urne ohne Zurücklegen. Wenn zu Beginn in der Urne 5 rote und 5 blaue Kugeln liegen, dann ist die Wahrscheinlichkeit für das Ziehen einer roten Kugel  $P(\text{rot}) = 50\%$ . Wenn nun zu den Zeitpunkten  $t_1 \dots t_5$  durch Zufall jeweils eine blaue Kugel gezogen wurde, so befinden

sich zum Zeitpunkt  $t_6$  natürlich nur noch rote Kugeln in der Urne. Die Wahrscheinlichkeit  $P(\text{rot}) = 100\%$  ist also davon beeinflusst, was zu den fünf vorigen Zeitpunkten passiert ist.

Diese Abhängigkeit der Zufallsvariablen von der Vergangenheit werden wir im folgenden Kapitel durch die Einführung der Markov-Prozesse drastisch einschränken.

## Anmerkung

Bei der Definition von stochastischen Prozessen kommt es immer wieder zu Verwechslungen. Grund dafür ist, dass zwei verschiedene Möglichkeiten existieren, wie man einen stochastischen Prozess verstehen kann. Ein stochastischer Prozess hängt von zwei Variablen, nämlich der Zeit  $t$  und dem Ergebnis  $\omega$  ab. Dementsprechend kann man ihn – so wie wir – als eine Abfolge von Zufallsvariablen verstehen, oder aber als eine Menge von Realisierungen, bei denen zu jedem Ergebnis  $\omega$  genau eine von der Zeit abhängige Funktion existiert, die – einmal gewählt – deterministisch abläuft. Während die zweite Anschauung in der Mathematik sehr beliebt ist, wird in der Physik eher erstere benutzt.

### 2.3.4 Markov-Prozesse

Wir werden nun die stochastischen Prozesse auf eine Unterklasse reduzieren, nämlich die der *Markov-Prozesse*. Hierbei handelt es sich um Prozesse, die kein »Gedächtnis« haben, deren nächster zukünftiger Zustand also nur von der Gegenwart und nicht von der Vergangenheit abhängt. Wir betrachten hierzu eine spezifische Realisierung eines stochastischen Prozesses  $(z_h, z_{h-1}, \dots, z_0)$  zu den Zeitpunkten  $(t_h > t_{h-1} > \dots > t_0)$ . Wenn nun gilt:

$$P(z_{h+1}|z_h, z_{h-1}, \dots, z_0) = P(z_{h+1}|z_h) \quad (2.16)$$

handelt es sich um einen Markov-Prozess. In Worten ausgedrückt bedeutet dies, dass die Wahrscheinlichkeit, zum nächsten beobachteten Zeitpunkt  $t_{n+1}$  in einem bestimmten Zustand  $z_{j+1}$  zu sein, ausschließlich vom aktuellen Zustand  $z_j$  abhängt. Jegliche Information über die Vergangenheit hilft nicht, die Aussage über die Zukunft zu präzisieren.

Die Grundannahmen und Implikationen von Markov-Prozessen werden sich durch die gesamte Arbeit ziehen. In Kapitel 4 werden wir aus dort aufgeführten Gründen von einem Markov-Prozess zu einem Nicht-Markov-Prozess übergehen. An einem kleinen Beispiel wollen wir daher verdeutlichen, wie Markov-Prozesse und Nicht-Markov-Prozesse miteinander zusammenhängen können.

## Beispiel

Wir stellen uns hierzu Studentin A vor, welche jede Woche 1 Euro geschenkt bekommt. Um wöchentlich über die Verwendung des Geldes zu entscheiden, überlegt sich A folgendes System: Sie legt jedes Mal nach Erhalt des Geldes vier Kugeln in eine Urne. Eine Kugel ist beschriftet mit E für Eis und drei Kugeln tragen den Buchstaben S für Sparen. Da A nach dem Befüllen den Inhalt ihrer Urne kennt, kann sie genau vorhersagen, dass sie bei ihrem Zug zu 25% ein E ziehen wird. Da die Urne wöchentlich neu befüllt wird, ist diese Wahrscheinlichkeit unabhängig davon, wie oft sie in den letzten Wochen bereits ein S gezogen hat. Information über die Vergangenheit, z. B. ein Blick auf den Kontoverlauf verbessert die Vorhersage der Wahrscheinlichkeiten also *nicht*. Es handelt sich um einen Markov-Prozess.

Nach einer längeren Phase ohne ein gezogenes E beschließt A nun, ihr System umzustellen. Anstatt die Urne jede Woche neu mit den vier Kugeln zu befüllen, befüllt sie diese nur einmal am Monatsanfang und zieht in den kommenden 4 Wochen jeweils eine Kugel heraus, ohne diese zurückzulegen. Auf diese Weise kann sie sich sicher sein, dass sie alle 4 Wochen genau

einmal ein Eis essen darf. Wir wollen für dieses Zug-System nun zwei Varianten unterscheiden: Angenommen A deponiert die bereits gezogenen Kugeln direkt neben der Urne. Wenn sie für ihren wöchentlichen Zug an die Urne herantritt, kann sie sofort erkennen, wie viele Kugeln und mit jeweils welchem Buchstaben sie bereits gezogen hat. Basierend auf denen ihr aktuell vorliegenden Informationen kann sie also sofort sagen, welche Kugeln sich noch in der Urne befinden und demnach die Wahrscheinlichkeiten für E oder S berechnen. Es handelt sich also auch hier um einen Markov-Prozess, da Ereignisse der Vergangenheit in den aktuellen Zustand aufgenommen wurden.

Anders wäre es, wenn A die Kugeln nach dem Zug nicht sichtbar in einer Schublade deponiert. Bei ihrem wöchentlichen Zug kann sie nun nicht mehr sofort erkennen, welche Kugeln sie bereits gezogen hat, sondern muss sich erinnern, ob sie in den letzten Wochen das Geld gespart oder ein Eis gegessen hat. Alternativ könnte Sie auch die Historie ihres Kontoverlaufs anschauen. Wichtig ist hier, dass A nun Wissen aus der Vergangenheit benötigt, um die aktuelle Konfiguration der Urne und damit die Wahrscheinlichkeit für ein baldiges Eis zu berechnen. Die Zukunft (Eis oder kein Eis) ist durch die Gegenwart und die Vergangenheit gegeben. Man sagt, es handelt sich um einen *Nicht-Markov-Prozess*. Wir werden in den folgenden Kapiteln basierend auf der Markov-Annahme Gleichungen herleiten und verschiedene Modelle diskutieren. In Kapitel 4 werden wir dann ein Modell bearbeiten, in dem einige Informationen des aktuellen Zustands nicht zugänglich sind. Genau wie bei A aktuelle Informationen über die Kugeln in der Schublade nicht vorhanden sind, wird in unserem Beispiel auch ein großer Teil der Informationen über den aktuellen Zustand fehlen und somit die Vergangenheit wichtig werden.

### 2.3.5 Von der Chapman-Kolmogorov-Gleichung zur Mastergleichung

Im vorigen Kapitel haben wir Markov-Prozesse als einen zentralen Prozess der Stochastik eingeführt. Basierend auf der Definition des Markov-Prozesses wollen wir nun eine wichtige Gleichung, nämlich die *Mastergleichung* (ME) herleiten. Wir bemerken hierzu, dass sich unter Verwendung von (2.15) und (2.16) schreiben lässt:

$$\begin{aligned} P(z_{h+1}, z_h, z_{h-1}, \dots, z_0) &= \\ P(z_{h+1}|z_h, z_{h-1}, \dots, z_0) \cdot P(z_h, z_{h-1}, \dots, z_0) &= \\ P(z_{h+1}|z_h) \cdot P(z_h, z_{h-1}, \dots, z_0) &= \\ P(z_{h+1}|z_h) \cdot P(z_h|z_{h-1}) \cdot \dots \cdot P(z_1|z_0) \cdot P(z_0). \end{aligned}$$

Dies impliziert:

$$\begin{aligned} P(z_3, z_2, z_1) &= P(z_3|z_2) \cdot P(z_2|z_1) \cdot P(z_1) && \text{und} \\ P(z_3, z_1) &= \sum_{z_2} P(z_3|z_2) \cdot P(z_2|z_1) \cdot P(z_1) \end{aligned}$$

womit folgt:

$$\Rightarrow P(z_3|z_1) = \sum_{z_2} P(z_3|z_2) \cdot P(z_2|z_1). \quad (2.17)$$

Die Wahrscheinlichkeit von Zustand  $z_3$  unter der Bedingung, dass zwei Schritte davor  $z_1$  vorgelegen hat, lässt sich somit berechnen, wenn man über alle möglichen Wege  $z_1 \rightarrow z_2 \rightarrow z_3$  summiert. Diese Gleichung ist bekannt unter dem Namen *Chapman-Kolmogorov-Gleichung* (CKE).

Wir führen nun eine neue Größe, die *Übergangsrate*  $a$ , ein und definieren:

$$a_{gh} = \lim_{\Delta t \rightarrow 0} \frac{P(z_g, t + \Delta t | z_h, t)}{\Delta t}. \quad (2.18)$$

Durch  $a_{gh}$  wird somit die Wahrscheinlichkeit dafür beschrieben, dass nach einer infinitesimalen Zeit  $\Delta t$  ein Zustand  $z_g$  vorliegt, wenn kurz davor der Zustand  $z_h$  vorlag. Für kleine Zeiten ist demnach die Wahrscheinlichkeit für einen Übergang von Zustand  $z_h$  zu Zustand  $z_g$  gegeben durch:

$$P(z_g, t + \Delta t | z_h, t) = \Delta t a_{gh} \quad (2.19)$$

und dementsprechend ist die Wahrscheinlichkeit dafür, dass in diesem Zeitbereich kein Übergang passiert:

$$P(z_h, t + \Delta t | z_h, t) = 1 - \Delta t \sum_{g|g \neq h} a_{gh} = \Delta t a_{hh}. \quad (2.20)$$

Unter Ausnutzung der CKE erhalten wir:

$$\begin{aligned} P(z_g, t + \Delta t | z_0, t_0) &= \sum_h P(z_g, t + \Delta t | z_h, t) P(z_h, t | z_0, t_0) \\ &= \sum_{h|g \neq h} (\Delta t a_{gh} P(z_h, t | z_0, t_0)) + \sum_{h|g=h} (\Delta t a_{gh} P(z_h, t | z_0, t_0)) \\ &= \Delta t \sum_{h|g \neq h} (a_{gh} P(z_h, t | z_0, t_0)) + a_{gg} P(z_g, t | z_0, t_0) \\ &= \Delta t \sum_{h|g \neq h} (a_{gh} P(z_h, t | z_0, t_0)) + \left(1 - \sum_{h|g \neq h} a_{hg}\right) P(z_g, t | z_0, t_0) \\ &= P(z_g, t | z_0, t_0) + \Delta t \sum_h (a_{gh} P(z_h, t | z_0, t_0) - a_{hg} P(z_g, t | z_0, t_0)), \end{aligned} \quad (2.21)$$

womit sich ergibt:

$$\frac{dP(z_g, t | z_0, t_0)}{dt} = \sum_h (a_{gh} P(z_h, t | z_0, t_0) - a_{hg} P(z_g, t | z_0, t_0)). \quad (2.22)$$

Nach Multiplikation mit  $P(z_0, t_0)$  und anschließender Summation über alle  $z_0$  erhalten wir schließlich die sogenannte *Vorwärts-Mastergleichung*<sup>1</sup>, oder oft einfach ME:

$$\frac{dP(z_g, t)}{dt} = \sum_h (a_{gh} P(z_h, t) - a_{hg} P(z_g, t)). \quad (2.23)$$

Die ME (bzw. eigentlich der Satz von MEs für alle  $z_g$ ) gibt uns für jeden der Zustände  $z_g$  die zeitliche Entwicklung dessen Wahrscheinlichkeit an. Da wir uns in dieser Arbeit ausschließlich mit chemischen Reaktionen befassen, werden wir die ME für unsere Zwecke anpassen. Zum einen ist uns aus Kapitel 2.2.2 bereits bekannt, dass in unserem System ein Zustand durch einen Vektor  $\mathbf{n}$  gegeben ist, der die Anzahl aller im betrachteten System enthaltenen Spezies enthält, weshalb wir substituieren:  $z \rightarrow \mathbf{n}$ . Weiterhin wissen wir, dass Übergänge zwischen verschiedenen Zuständen

<sup>1</sup> Der Name lässt bereits vermuten, dass es auch eine *Rückwärts-Mastergleichung* gibt. Diese werden wir aber nur in Kapitel 4 benötigen, weshalb sie dort erst vorgestellt wird.

ausschließlich über die  $M$  Reaktionen möglich sind. Jede Reaktion  $R_j$  löst hierbei gemäß der stöchiometrischen Matrix  $\mathbf{S}$  einen Übergang von  $\mathbf{n} \rightarrow \mathbf{n} + \mathbf{S}_j$  aus. Die Raten, mit denen diese Reaktionen auftreten, sind uns ebenfalls bereits als die reaktionspezifischen Propensitäten  $f(\mathbf{n})$  bekannt. Betrachtet man als Beispiel die Reaktionen  $\emptyset \rightarrow \mathbb{X}$  und  $\mathbb{X} \rightarrow 2\mathbb{X}$ , wird dabei deutlich, dass die Übergangsraten  $a_{gh}$  durchaus mehrere mögliche Reaktionen mit gleicher Wirkung beinhalten kann, also als Summe verschiedener  $f_j$  aufgefasst werden muss. Wir definieren daher:

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_j f_j(\mathbf{n} - \mathbf{S}_j)P(\mathbf{n} - \mathbf{S}_j, t) - \sum_j f_j(\mathbf{n})P(\mathbf{n}, t). \quad (2.24)$$

Wir summieren also nicht über alle möglichen Zustände des Systems, sondern über alle Reaktionen, welche auftreten können. Bei dieser für Reaktionssysteme zugänglicheren Form handelt es sich um die *Chemische Mastergleichung* (CME).

### 2.3.6 Grundannahmen der CME

Die CME wird uns in der gesamten Arbeit begleiten. Aus mathematischer Sicht war die Herleitung der ME dabei relativ gut zu verstehen. Mit Übergang zur CME gehen wir aber von einer mathematisch zu einer physikalisch motivierten Gleichung über. Wichtig ist daher, die physikalischen Implikationen der mathematischen Näherungen zu verstehen [35].

Die CME fußt auf der Näherung, dass die in der Gleichung betrachteten mikroskopischen Größen innerhalb des Zeitintervalls  $\Delta t$  einen Gleichgewichtszustand im betrachteten System einnehmen. Insbesondere wird hierbei aufgrund der Markov-Annahme (siehe Kapitel 2.3.4) angenommen, dass alle mikroskopischen Variablen zum Zeitpunkt  $t + \Delta t$  nur vom Zustand zum Zeitpunkt  $t$  abhängen. In jedem Zeitschritt  $\Delta t$  wird das System durch die jeweils ablaufende Reaktion gestört und relaxiert sofort in einen neuen Gleichgewichtszustand. Dies impliziert direkt, dass der Übergang  $\lim_{\Delta t \rightarrow 0}$  kein mathematisch rigoroser Übergang sein kann, sondern  $\Delta t$  immer groß genug sein muss, damit die Relaxationen ablaufen können [35].

Im Vergleich zu der Herleitung des SSA in Kapitel 2.2.3 wurden bisher keine weiteren Näherungen verwendet, weshalb die CME auch die gleichen Wahrscheinlichkeitsverteilungen vorhersagt, welche man mit dem SSA erhalten würde (man könnte auch sagen, dass der SSA die numerische Simulation der CME darstellt [27]). Das Problem ist allerdings, dass die CME nur für wenige und kleine Systeme lösbar ist. Bereits bei den meisten Systemen mit drei Reaktionen muss man sich auf numerische Methoden (wie z. B. den SSA) stützen. Somit bietet die CME zwar eine mathematische Beschreibung für die Zeitentwicklung der Wahrscheinlichkeiten der einzelnen Zustände, in den meisten Fällen ist sie aber nicht analytisch lösbar.

### Anmerkung

Den Zusammenhang zwischen der ME und dem SSA können wir leicht erkennen, wenn wir zurück zu Gleichung (2.9) schauen. Dort haben wir die Wahrscheinlichkeit dafür definiert, zu einem bestimmten Zeitpunkt  $t + dt$  noch im gleichen Zustand zu sein wie zum Zeitpunkt  $t$ , da im Zeitintervall  $dt$  keine Reaktion stattgefunden hat. Dies lässt sich leicht erweitern auf den Fall, in dem entweder keine oder genau eine Reaktion abgelaufen ist. Man erhält dann:

$$P(\mathbf{n}, t + dt) = P(\mathbf{n}, t) \left( 1 - \sum_j f_j(\mathbf{n}) dt \right) + \sum_j P(\mathbf{n} - \mathbf{S}_j) f_j(\mathbf{n} - \mathbf{S}_j) dt.$$



Durch Subtraktion von  $P(\mathbf{n}, t)$  und anschließenden Übergang  $\lim_{t \rightarrow 0}$  ergibt sich auch hier die ME.

Wir werden in den nächsten Kapiteln Methoden suchen, mit denen die Mastergleichung derart behandelt werden kann, dass einfache analytische Ansätze möglich sind. Wir wechseln dazu von der diskreten Beschreibung durch Teilchenzahlen zu kontinuierlichen Konzentrationen und nutzen dann die Kramers-Moyal-Entwicklung.

### 2.3.7 Von Teilchenzahlen zu Teilchenkonzentrationen

Für die weitere Analyse wird es sich als wichtig erweisen, das betrachtete System nicht über Teilchenzahlen, sondern über deren Konzentrationen zu beschreiben. Konzentrationen sind Gehaltsangaben, die auf ein Volumen bezogen sind. Anhand eines Beispiels werden wir in diesem Kapitel daher zuerst untersuchen, wie die Reaktionskonstanten von der Systemgröße  $\Omega$  abhängen, und dann die Propensitäten in Abhängigkeit von Konzentrationen formulieren.

Angenommen eine Reaktion beinhaltet zwei Reaktanten vernachlässigbarer Ausdehnung, welche sich im betrachteten System mit Größe  $\Omega$  an den Positionen  $\mathbf{x}_1$  und  $\mathbf{x}_2$  befinden. Die Wahrscheinlichkeit dafür, dass sich einer der Reaktanten innerhalb des Untervolumens  $\Delta\Omega$  befindet, ist dann gegeben durch:

$$P(\mathbf{x}_1 \in \Delta\Omega) = \frac{\Delta\Omega}{\Omega}. \quad (2.25)$$

Wir interessieren uns nun für die Wahrscheinlichkeit, dass beide Reaktanten sich näher als den Abstand  $d$  kommen, um dann miteinander reagieren zu können:

$$P(|\mathbf{x}_1 - \mathbf{x}_2| < d) = \frac{4\pi d^3}{3\Omega}. \quad (2.26)$$

Die Wahrscheinlichkeit ist somit invers proportional zur Systemgröße. Je größer das betrachtete System ist, desto unwahrscheinlicher wird es, dass sich die beiden Reaktanten treffen. Gemäß der Definition der Reaktionskonstanten  $c$  in Kapitel 2.2.2 muss also auch diese im Falle einer Reaktion mit zwei Reaktanten invers proportional zum Systemvolumen sein [36].

Anders verhält es sich bei einer Reaktion, welche nur einen Reaktanten beinhaltet. Dieser muss sich nicht an einer bestimmten Position im Raum befinden, um die Reaktion auslösen zu können. Die Reaktionskonstante ist demnach nicht abhängig von  $\Omega$ .

Um die Abhängigkeit der Reaktionskonstante von der Systemgröße (bei konstant gehaltener Teilchenzahl) explizit anzugeben, müssen wir diese daher umdefinieren. Eine allgemeingültige Herleitung hierzu findet sich in [25, 36]:

$$c \rightarrow c(\Omega) = \frac{c}{\Omega^{|s_j|-1}}, \quad (2.27)$$

wobei  $s_j = \sum_i s_{ij}$  die Anzahl der beteiligten Reaktanten angibt.

Der für die Angabe der Propensität außerdem benötigte kombinatorische Faktor  $h$  ist bei konstanter Teilchenzahl nicht von der Systemgröße abhängig.

Meist definiert man eine Vergrößerung der Systemgröße allerdings so, dass dabei die Konzentration konstant gehalten wird. Die Teilchenzahl ändert sich also proportional zur Änderung von  $\Omega$ . Aus diesem Grund führen wir die Teilchenkonzentration ein:

$$\tilde{n}_i(t) := n_i/\Omega. \quad (2.28)$$

Wir wechseln also vom diskreten und extensiven Wert der Speziesanzahl auf den kontinuierlichen intensiven Wert einer Konzentration. Neben dem Speziesvektor müssen wir auch den kombinatorischen Faktor umformulieren, welcher nun auf Basis von Konzentrationen arbeitet. Gemäß der Herleitung von  $h$  aus Gleichung (2.4) schreiben wir hier:

$$\begin{aligned} h_j(\mathbf{n}) &= \prod_{i=1}^N \frac{n_i!}{s_{ij}!(n_i - s_{ij})!} = \prod_{i=1}^N \frac{1}{s_{ij}!} \prod_{k=0}^{s_{ij}-1} (n_i - k) \\ &= \prod_{i=1}^N \frac{\Omega^{s_{ij}}}{s_{ij}!} \prod_{k=0}^{s_{ij}-1} \left( \tilde{n}_i - \frac{k}{\Omega} \right) = \tilde{h}_j(\tilde{\mathbf{n}}). \end{aligned} \quad (2.29)$$

Die konzentrationsabhängige Propensität lautet dann:

$$\begin{aligned} c_j(\Omega)\tilde{h}_j(\tilde{\mathbf{n}}) &= \frac{c_j}{\Omega^{|s_j|-1}} \cdot \prod_{i=1}^N \frac{\Omega^{s_{ij}}}{s_{ij}!} \prod_{k=0}^{s_{ij}-1} \left( \tilde{n}_i - \frac{k}{\Omega} \right) \\ &= \Omega c_j \prod_i \frac{1}{s_{ij}!} \prod_{k=0}^{s_{ij}-1} \left( \tilde{n}_i - \frac{k}{\Omega} \right) = \Omega c_j \left( \prod_i \frac{\tilde{n}_i}{s_{ij}!} + \mathcal{O}(\Omega^{-1}) \right) \\ &\approx \Omega c_j \prod_i \frac{\tilde{n}_i}{s_{ij}!} = \tilde{f}_j(\tilde{\mathbf{n}}). \end{aligned} \quad (2.30)$$

Im letzten Schritt haben wir hier vorausgesetzt, dass  $\tilde{n}_i \gg \frac{s_{ij}}{\Omega}$  gilt, dass also  $\Omega$  sehr groß ist und  $s_{ij}$  entsprechend klein.

Für den Zusammenhang zwischen den Propensitäten, welche von der Teilchenzahl abhängen  $f$  und denen, welche mit Konzentrationen arbeiten  $\tilde{f}$  gilt demnach [33]:

$$\frac{f(\mathbf{n})}{\Omega} - \tilde{f}\left(\frac{\mathbf{n}}{\Omega}\right) = \begin{cases} 0 & \text{falls } s_{ij} \in \{0, 1\} \\ \mathcal{O}(\Omega^{-2}) & \text{ansonsten} \end{cases} \quad (2.31)$$

Für den Fall  $s_{ij} \in \{0, 1\}$  nennt man das betrachtete Reaktionsnetzwerk linear. Mit solchen linearen Netzwerken werden wir uns in Kapitel 4 noch intensiv beschäftigen.

Aufgrund dieses Zusammenhangs wird die Rate  $\tilde{f}$  meist als makroskopische Rate bezeichnet, welche in großen, durch Konzentrationen charakterisierten Systemen Anwendung findet.

Nachdem wir nun die Systemgröße als zusätzlichen Parameter eingeführt haben, werden wir in den nächsten zwei Kapiteln herleiten, wie sich die Dynamik großer Systeme vereinfacht beschreiben lässt. Hierzu nutzen wir die Kramers-Moyal-Entwicklung.

### 2.3.8 Die Kramers-Moyal-Entwicklung und die Fokker-Planck-Gleichung

Die Kramers-Moyal-Entwicklung ist eine Entwicklung der ME bzw. CME für kleine Sprungweiten, also für Reaktionen, welche nur kleine Änderungen der Speziesanzahl bewirken. In exakter Form ergibt die Entwicklung eine unendliche Reihe, welche nicht leichter zu lösen ist, als die ME. Allerdings zeigt sich, dass bereits die ersten ein bzw. zwei Glieder dieser Entwicklung sinnvolle Ergebnisse liefern, weshalb die Reihe früh abgebrochen werden kann. Wir betrachten für diese Entwicklung die CME, welche wir in Abhängigkeit von Konzentrationen formulieren:

$$\frac{\partial P(\tilde{\mathbf{n}}, t)}{\partial t} = \sum_j \tilde{f}_j \left( \tilde{\mathbf{n}} - \frac{\mathbf{S}_{\cdot j}}{\Omega} \right) P \left( \tilde{\mathbf{n}} - \frac{\mathbf{S}_{\cdot j}}{\Omega}, t \right) - \sum_j \tilde{f}_j(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t). \quad (2.32)$$

Wenn wir voraussetzen, dass die Einträge von  $\mathbf{S}$  klein sind im Vergleich zu den Einträgen in  $\mathbf{n}$ , wenn also eine einzelne Reaktion die Gesamtkonzentration nur wenig verändert und zusätzlich  $\Omega$  groß genug ist, können wir die CME um kleines  $\mathbf{S}_{ij}/\Omega$  entwickeln und erhalten:

$$\begin{aligned} \frac{\partial P(\tilde{\mathbf{n}}, t)}{\partial t} &= \sum_j \tilde{f}_j(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t) \\ &\quad - \sum_j \sum_i \frac{\mathbf{S}_{ij}}{\Omega} \frac{\partial}{\partial \tilde{n}_i} (\tilde{f}_j(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t)) + \sum_j \sum_{i,k} \frac{\mathbf{S}_{ij} \mathbf{S}_{kj}}{2\Omega^2} \frac{\partial}{\partial \tilde{n}_i} \frac{\partial}{\partial \tilde{n}_k} (\tilde{f}_j(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t)) + \dots \\ &\quad - \sum_j \tilde{f}_j(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t). \end{aligned} \quad (2.33)$$

Wenn wir die Funktion nach dem zweiten Glied abrechnen, zeigt sich, dass die dadurch entstehende Funktion zur Beschreibung der Zeitentwicklung von  $P$  normierbar ist und keine negativen Werte (was unphysikalisch wäre) annehmen kann. Durch Absorption eines der  $\Omega$  in der Zeit, erhalten wir folgende Gleichung, welche allgemein als die *Fokker-Planck-Gleichung* (FPE) bekannt ist:

$$\frac{\partial P(\tilde{\mathbf{n}}, t)}{\partial t} = - \sum_i \frac{\partial}{\partial \tilde{n}_i} (\mathcal{A}_i(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t)) + \frac{1}{2\Omega} \sum_{i,k} \frac{\partial}{\partial \tilde{n}_i} \frac{\partial}{\partial \tilde{n}_k} (\mathcal{B}_{ik}(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t)) \quad (2.34)$$

mit:

$$\mathcal{A}_i(\tilde{\mathbf{n}}) = \sum_j \mathbf{S}_{ij} \tilde{f}_j(\tilde{\mathbf{n}}) \quad (2.35)$$

$$\mathcal{B}_{ik}(\tilde{\mathbf{n}}) = \sum_j \mathbf{S}_{ij} \mathbf{S}_{kj} \tilde{f}_j(\tilde{\mathbf{n}}) = (\mathbf{S} \text{diag}(\tilde{f}(\tilde{\mathbf{n}})) \mathbf{S}^T)_{ik} \quad (2.36)$$

Der Vektor  $\mathcal{A}$  wird hierbei als die deterministische Drift bezeichnet, da er ausschließlich von den Ratenkonstanten abhängt. Die Matrix  $\mathcal{B}$  nennt man die Diffusion des Systems. Da die Sprünge quadratisch eingehen, enthält sie keine Richtungsinformation der Reaktionen und verschiebt das System hauptsächlich in Richtung einer Gleichverteilung.

Mit Gleichung (2.34) erhalten wir für jede mögliche Kombination von Konzentrationen die zugehörige Wahrscheinlichkeitsdichte und deren zeitliche Entwicklung. Auch wenn die FPE für komplexe Systeme nur numerisch gelöst werden kann, werden wir später sehen, dass sie uns einen analytischen Zugang zu wichtigen Informationen des beschriebenen Systems verschaffen kann.

## Anmerkung

Naiv betrachtet könnte man meinen, dass ein Abbruch der Kramers-Moyal-Entwicklung nach dem dritten oder einem noch größeren Glied die Genauigkeit des Ergebnisses gegenüber der FPE verbessern müsste. Während dies auf eingeschränkten Definitionsbereichen korrekt sein kann, lässt sich beweisen, dass es im Allgemeinen kein weiteres Abbruch-Glied gibt, welches eine ebenfalls normierbare Funktion erzeugt, die gleichzeitig stets positiv bleibt [37].

### 2.3.9 Reaktionsratengleichungen – Dynamisches System

Bricht man die Kramers-Moyal-Entwicklung bereits nach dem ersten Term ab, erhält man:

$$\frac{\partial P(\tilde{\mathbf{n}}, t)}{\partial t} = - \sum_i \frac{\partial}{\partial \tilde{n}_i} \mathcal{A}_i(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t). \quad (2.37)$$

Hierbei handelt es sich um eine spezielle Form der Liouville-Gleichung, welche aus der klassischen Mechanik bekannt ist und dort oft als Bindeglied zwischen der deterministischen und probabilistischen Beschreibung dient. Sei  $\mathbf{y}(t)$  die Lösung folgender Differentialgleichung:

$$\frac{d\mathbf{y}(t)}{dt} = \mathcal{A}(\mathbf{y}(t)). \quad (2.38)$$

Dann können wir die Lösung auch über eine Wahrscheinlichkeitsdichte  $P(\tilde{\mathbf{n}}, t)$  angeben, für die gilt:

$$P(\tilde{\mathbf{n}}, t) = \delta(\tilde{\mathbf{n}} - \mathbf{y}(t)) = \prod_{k=1}^N \delta(\tilde{n}_k - y_k(t)) \quad (2.39)$$

und die zeitliche Ableitung ergibt sich zu:

$$\begin{aligned} \frac{\partial P(\tilde{\mathbf{n}}, t)}{\partial t} &= - \sum_{i=1}^N \prod_{\substack{k=1 \\ k \neq i}}^N \delta(\tilde{n}_k - y_k(t)) \dot{y}_i(t) \frac{\partial}{\partial \tilde{n}_i} \delta(\tilde{n}_i - y_i(t)) \\ &= - \sum_i \dot{y}_i \frac{\partial}{\partial \tilde{n}_i} P(\tilde{\mathbf{n}}, t) = - \sum_i \mathcal{A}_i(\mathbf{y}(t)) \frac{\partial}{\partial \tilde{n}_i} P(\tilde{\mathbf{n}}, t) \\ &= - \sum_i \frac{\partial}{\partial \tilde{n}_i} \mathcal{A}_i(\mathbf{y}(t)) P(\tilde{\mathbf{n}}, t) = - \sum_i \frac{\partial}{\partial \tilde{n}_i} \mathcal{A}_i(\tilde{\mathbf{n}}) P(\tilde{\mathbf{n}}, t). \end{aligned} \quad (2.40)$$

Wenn also  $P(\tilde{\mathbf{n}}, 0)$  als Delta-Distribution gegeben ist, dann verhält sich der zeitliche Verlauf analog zu der Lösung der Differentialgleichung (2.38) und bleibt weiterhin als Delta-Peak auf der durch die Differentialgleichung vorgegebenen Trajektorie [30, 38]. Ein solches Differentialgleichungssystem

$$\frac{d\tilde{\mathbf{n}}(t)}{dt} = \mathcal{A}(\tilde{\mathbf{n}}(t)), \quad (2.41)$$

wie wir es dann für die zeitliche Entwicklung von  $\tilde{\mathbf{n}}$  erhalten, bezeichnet man allgemein als *dynamisches System* und im Kontext von Reaktionen oftmals als *Reaktionsratengleichung* (RRE).

Da diese Gleichung durch Abbruch nach dem ersten Glied bzw. nach dem Grenzübergang  $\lim_{\Omega \rightarrow \infty}$  erhalten wurde, kann man sofort schließen, dass hier alle stochastischen Fluktuationen unterdrückt werden. Der zeitliche Verlauf der Konzentrationen ist mit Angabe des Anfangswertes komplett determiniert, weshalb wir hier von einer *deterministischen Beschreibung* sprechen. Da  $\mathcal{A}$  nur von der aktuellen Position  $\tilde{\mathbf{n}}$  abhängt, gilt außerdem, dass der zeitliche Verlauf von  $P$  nur lokale Informationen besitzt und aus diesem Grund auch anfällig für lokale Minima ist.

## Anmerkung

Um den Übergang von der stochastischen Beschreibung hin zur deterministischen Beschreibung darzustellen, haben wir den Weg über die Fokker-Planck-Gleichung gewählt, da diese im weiteren Verlauf der Arbeit noch Verwendung finden wird. Zur Veranschaulichung des Zusammenhangs zwischen stochastischer und deterministischer Beschreibung soll hier allerdings noch eine zweite Möglichkeit aufgezeigt werden [33, 39, 40]. Wir beginnen dazu mit der CME, welche wir mit  $\mathbf{n}$  multiplizieren und alsdann über alle Zustände summieren:

$$\begin{aligned}\sum_{\mathbf{n}} \mathbf{n} \frac{dP(\mathbf{n}, t)}{dt} &= \sum_{\mathbf{n}} \mathbf{n} \left( \sum_j f_j(\mathbf{n} - \mathbf{s}_{\cdot j}) P(\mathbf{n} - \mathbf{s}_{\cdot j}, t) - \sum_j f_j(\mathbf{n}) P(\mathbf{n}, t) \right) \\ &= \sum_j \mathbf{s}_{\cdot j} \sum_{\mathbf{n}} f_j(\mathbf{n}) P(\mathbf{n}, t) \\ &= \sum_j \mathbf{s}_{\cdot j} \mathbb{E}(f_j(\mathbf{n}))\end{aligned}\quad (2.42)$$

folglich gilt:

$$\frac{d}{dt} \mathbb{E}(\mathbf{n}(t)) = \sum_j \mathbf{s}_{\cdot j} \mathbb{E}(f_j(\mathbf{n}(t))).\quad (2.43)$$

Damit diese Gleichung lösbar ist, müssen wir nähern, dass der Erwartungswert der Propensität gleich der Propensität des Erwartungswertes ist:

$$\mathbb{E}(f_j(\mathbf{n})) \approx f_j(\mathbb{E}(\mathbf{n}))\quad (2.44)$$

womit wir letztlich erhalten:

$$\frac{d}{dt} \mathbb{E}(\mathbf{n}(t)) \approx \sum_j \mathbf{s}_{\cdot j} f_j(\mathbb{E}(\mathbf{n})).\quad (2.45)$$

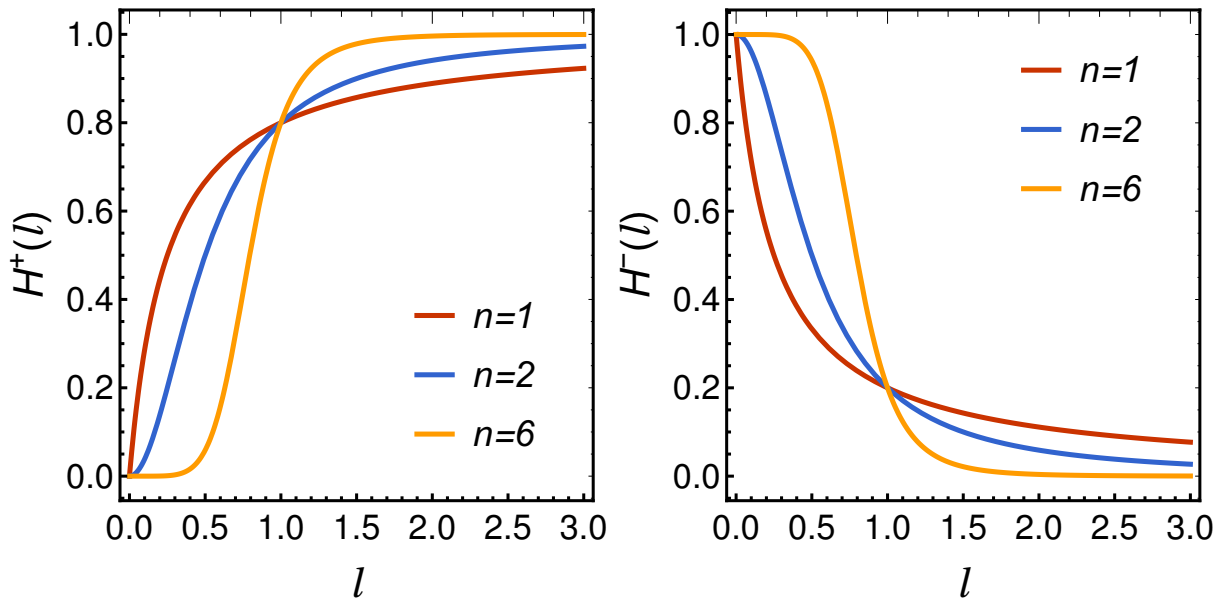
Die Näherung, welche wir in Gleichung (2.44) vollzogen haben, ist hierbei in linearen Reaktionsnetzwerken exakt, womit direkt der Zusammenhang mit Gleichung (2.31) ersichtlich ist.

### 2.3.10 Das Massenwirkungsgesetz und die Hill-Funktion

Da die Propensitäten der einzelnen Reaktionen in (2.41) proportional zum Produkt der Reaktantenkonzentrationen sind, spricht man in diesem Fall von einem dynamischen System, welches dem *Massenwirkungsgesetz* genügt. Gemäß unserer bisherigen Herleitung können dabei zur Beschreibung von Reaktionssystemen nur dynamische Systeme entstehen, welche auch das Massenwirkungsgesetz erfüllen. Basierend auf diesen dynamischen Systemen lassen sich aber weitere Gleichungen herleiten, für die dies nicht mehr gilt. Als Beispiel soll hier die Hill-Funktion dienen, welche später in den Kapiteln 3 und 4 wieder vorkommen wird.

Zur Herleitung der Hill-Funktion betrachten wir ein Protein  $P$ , an welches  $n$  Teile einer anderen Substanz  $L$  reversibel binden und dadurch einen funktionalen Komplex  $PL_n$  bilden können. Die zugehörige Reaktionsgleichung lautet dann:





**Abbildung 2.6:** Darstellung der Hill-Funktion mit  $L$  als Aktivator (links) und Repressor (rechts) mit  $K_d = 0.25$  und verschiedenen Werten von  $n$ .

wobei  $k_a$  und  $k_d$  die jeweiligen Reaktionskonstanten der Assoziation und Dissoziation sind. Aus Kapitel 2.3.9 wissen wir nun, dass sich deterministisch unter anderem folgende Differentialgleichung ergibt:

$$\frac{dp l_n}{dt} = k_a p \cdot l^n - k_d p l_n. \quad (2.47)$$

Hierbei bezeichnen  $p l_n$ ,  $p$  und  $l$  jeweils die Konzentrationen von  $PL_n$ ,  $P$  und  $L$ . Wenn wir hier nach der Lösung suchen, welche sich im stationären, also zeitlich konstanten Fall einstellt, setzen wir die Zeitableitung gleich Null und erhalten:

$$\frac{p \cdot l^n}{p l_n} = \frac{k_d}{k_a} = K_d. \quad (2.48)$$

Die Konstante  $K_d$  nennt man hierbei oft *Michaelis-Konstante*. Mithilfe dieses Ergebnisses können wir nun für den stationären Fall den Anteil der Proteine, welche sich in einem funktionalen Komplex befinden, in Abhängigkeit der Konzentration von  $L$  ausrechnen:

$$H^+(l) = \frac{p l_n}{p l_n + p} = \frac{l^n}{K_d + l^n}. \quad (2.49)$$

Die Funktion  $H^+(l)$  ist als *Hill-Funktion* mit  $L$  als Aktivator bekannt. Der Parameter  $n$  wird oft *Hill-Koeffizient* genannt.

Analog lässt sich auch der Anteil der Proteine ausrechnen, welcher sich *nicht* in einem gebundenen Komplex befindet. Man erhält dann die Hill-Funktion mit  $L$  als Repressor:

$$H^-(l) = \frac{p}{p l_n + p} = \frac{K_d}{K_d + l^n}. \quad (2.50)$$

In Abbildung 2.6 sind die beiden Hill-Funktionen beispielhaft für verschiedene Werte von  $n$  dargestellt.

---

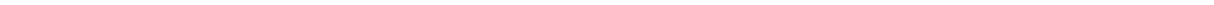
---

Wenn der Komplex  $PL_n$  nun zum Beispiel der aktivierende Transkriptionsfaktor für die Expression eines anderen Proteins  $Q$  ist, dann lässt sich die zeitliche Entwicklung der Konzentration von  $Q$  in Abhängigkeit von  $L$  angeben durch:

$$\frac{dq(l)}{dt} = aH^+(l), \quad (2.51)$$

wobei  $a$  in diesem Beispiel die maximale Expressionsrate darstellt. Offensichtlich genügt diese Gleichung nicht dem Massenwirkungsgesetz. In manchen Fällen, etwa bei der Berechnung höherer Momente, kann diese Verletzung des Massenwirkungsgesetzes zu ungenauen Ergebnissen führen [41, 42]. In diesen Fällen sollten Ergebnisse also immer z. B. mittels SSA validiert werden.

Aufgrund der guten analytischen Verwertbarkeit und der einfachen (numerischen) Berechnung wird die deterministische Beschreibung häufig als einzige Beschreibung eines biologischen Systems herangezogen. In den nächsten Kapiteln werden wir allerdings zwei Fälle kennen lernen, in denen die deterministische Beschreibung nicht genügt, um das System ausreichend zu charakterisieren. So ist in dem nun folgenden Kapitel 3 die deterministische Beschreibung beispielsweise nicht ausreichend, da Informationen über die genaue Verteilung der Wahrscheinlichkeiten vonnöten sind und in Kapitel 5 wird ein Fall behandelt, bei dem der Übergang  $\lim_{\Omega \rightarrow \infty}$  so starke Fehler induziert, dass die Aussage der deterministischen Beschreibung nicht mehr verlässlich ist.





---

## 3 Modell für einen CRISPR/dCas9-basierten genetischen Schalter

In diesem Kapitel wird ein synthetisches AND-Gatter betrachtet, das auf dem sogenannten CRISPR/dCas9-System beruht und in eine Hefezelle integriert wurde. Hierzu werden wir zuerst erklären, was genau das CRISPR/dCas9-System ist und wie dieses funktioniert. Nachdem kurz auf die zugrunde liegenden Experimente eingegangen wurde, beschreiben wir dann, wie die Modellierung und mathematische Analyse des Schaltkreises vonstattengeht und welche Ergebnisse sich daraus ergeben.

Das Projekt, in welchem die in diesem Kapitel erklärte Modellierung eingebettet ist, wurde dabei von Harald Kolmar konzipiert und geleitet. Das Design und die Durchführung der Experimente geschah durch Anja Hofmann (AG Kolmar). Alle im Folgenden genutzten experimentellen Daten sind folglich von ihr erhoben worden. Teilweise erhielt Anja dabei Unterstützung von Dominic Happel, Adrian Köber und Andreas Christmann (alle AG Kolmar). Von Anja stammen auch die Abbildungen 3.3 und 3.4. In dieser Arbeit nicht weiter eingehen werden wir auf Einzelmessungen, welche von Tim Prangemeier (AG Koepl) mit Unterstützung von Heinz Koepl durchgeführt wurden. Der Eigenbeitrag des Autors ist die Analyse der im Experiment erhobenen Daten und die darauf basierende Entwicklung des in diesem Kapitel beschriebenen theoretischen Modells.

Die Ausführungen in dieser Arbeit sind auf jene Punkte reduziert, die zum Verständnis des Modells nötig sind. Für eine genauere Beschreibung der verwendeten Konstrukte sowie des experimentellen Vorgehens sei hier auf die zugehörige Publikation [43] verwiesen, welche unter anderem aus der im Folgenden beschriebenen Forschung hervorgegangen ist. Insbesondere wird in der Veröffentlichung auch ein weiteres, leicht verändertes System analysiert.

---

### 3.1 Übersicht: Genetische Schalter

---

Wie in der Einleitung bereits angeklungen, ist das Ziel der Synthetischen Biologie die Entwicklung und Implementierung künstlicher genetischer Schaltkreise. Hierbei wird angestrebt, dass die einzelnen logischen Bauteile dieser genetischen Schalter beliebig miteinander kombiniert werden können und dennoch weiterhin zuverlässig und definiert arbeiten. Um die Expression bestimmter Gene zu regulieren, wurden in der Vergangenheit vor allem natürlich vorkommende Transkriptionsmechanismen verwendet. Ein Beispiel hierfür ist das in der Einleitung (Kapitel 2.1.3) beschriebene GAL-System. Durch Überführen der Hefekultur von Glucose auf Galactose werden im natürlichen System bestimmte Gene zur Verarbeitung von Galactose aktiviert. Durch Verwendung des gleichen Mechanismus kann man nun leicht auch andere Gene schalten [44]. Ein großer Nachteil vieler dieser natürlichen Transkriptionsfaktoren ist allerdings, dass sie mit dem natürlichen Metabolismus der Zelle interagieren und somit oftmals das Zellwachstum negativ beeinflussen [45–47]. Ein modernerer Weg ist daher der Einsatz von künstlichen Transkriptionsfaktoren wie etwa synthetischen Zinkfingerproteinen. Hierbei handelt es sich um Proteine, welche so designt werden können, dass sie spezifisch mit der DNA interagieren [48]. Leider lassen sich diese Zinkfingerproteine nur sehr umständlich und teuer herstellen [49]. Aktuell gibt es folglich große Bemühungen neue genetische Schalter zu finden, welche zuverlässig schalten und möglichst für die Zelle nicht beeinträchtigend oder sogar toxisch sind [50, 51].

Vor Kurzem wurde nun gezeigt, dass das sogenannte CRISPR/Cas9-System in einer eukaryotischen Zelle die oben genannten Einschränkungen überwindet und eine nützliche Plattform für

---

das Design verschiedener kontrollierbarer genetischer Schalter bietet [52–54]. Wie genau dieses CRISPR/Cas9 System funktioniert und welche Erweiterungen wir daran vorgenommen haben, erklären die nächsten Kapitel.

---

### 3.1.1 CRISPR/Cas9-System

---

Das Akronym CRISPR steht für *Clustered Regularly Interspaced Short Palindromic Repeats* und bezeichnet einen DNA-Bereich im bakteriellen Genom, in dem sich identische DNA-Sequenzen, die sogenannten Repeats, und dazwischenliegende Sequenzen, die sogenannten Spacer, abwechseln. Schon im Jahre 1987 berichteten Forschende von einer sich wiederholenden DNA-Sequenz im Genom des Bakteriums *E. coli*, konnten aber nicht erklären, welchen Nutzen diese haben [55]. Die Hypothese, dass diese Sequenzen ein Bestandteil des Immunsystems des Bakteriums sein könnten, kam erst 2005 auf [56, 57]. Die genaue Funktionsweise des CRISPR-DNA Strangs ist dabei bis heute nicht endgültig geklärt. Wahrscheinlich läuft der Vorgang aber wie folgt ab (siehe hierzu auch Abbildung 3.1)[58]: Wenn ein Bakterium durch einen Virus infiziert wird, fügt es einen Teil des viralen Erbguts als Spacer in die CRISPR-DNA ein. Wird die CRISPR-DNA nun transkribiert, so bilden sich nach einer definierten Prozessierung für jedes Spacer-Repeater Paar kleine *guide RNA* (gRNA) Stränge, die jeweils aus einem stabil gefalteten Repeater und einem freien Spacer bestehen. Zusammen mit dem DNA-schneidenden Enzym *CRISPR associated protein 9* (Cas9) bildet die gRNA einen funktionalen Komplex. Wenn nun die gleiche virale DNA wieder in das Bakterium eindringt, dann bindet der freie Spacer des Cas9 Komplexes an diese Sequenz und die Cas9 erzeugt einen Doppelstrangbruch an betreffender Stelle. Damit wird die virale DNA zerstört.

Den hier beschriebenen natürlichen Mechanismus macht man sich in der Synthetischen Biologie zunutze, indem man künstlich gRNA-Stränge in die Zelle einbringt, deren Bindungsstellen frei gestaltet werden können. So lässt sich der von der Cas9 erzeugte Doppelstrangbruch an beliebiger Stelle einfügen. An den entstehenden definierten Schnittstellen lassen sich dann eigene DNA Sequenzen in das Genom integrieren. Zusätzlich konnten Feng Zhang et al. den CRISPR/Cas9 Mechanismus erfolgreich von Bakterien auf Eukaryoten übertragen, womit CRISPR/Cas9 heute auch in Hefe anwendbar ist [59].

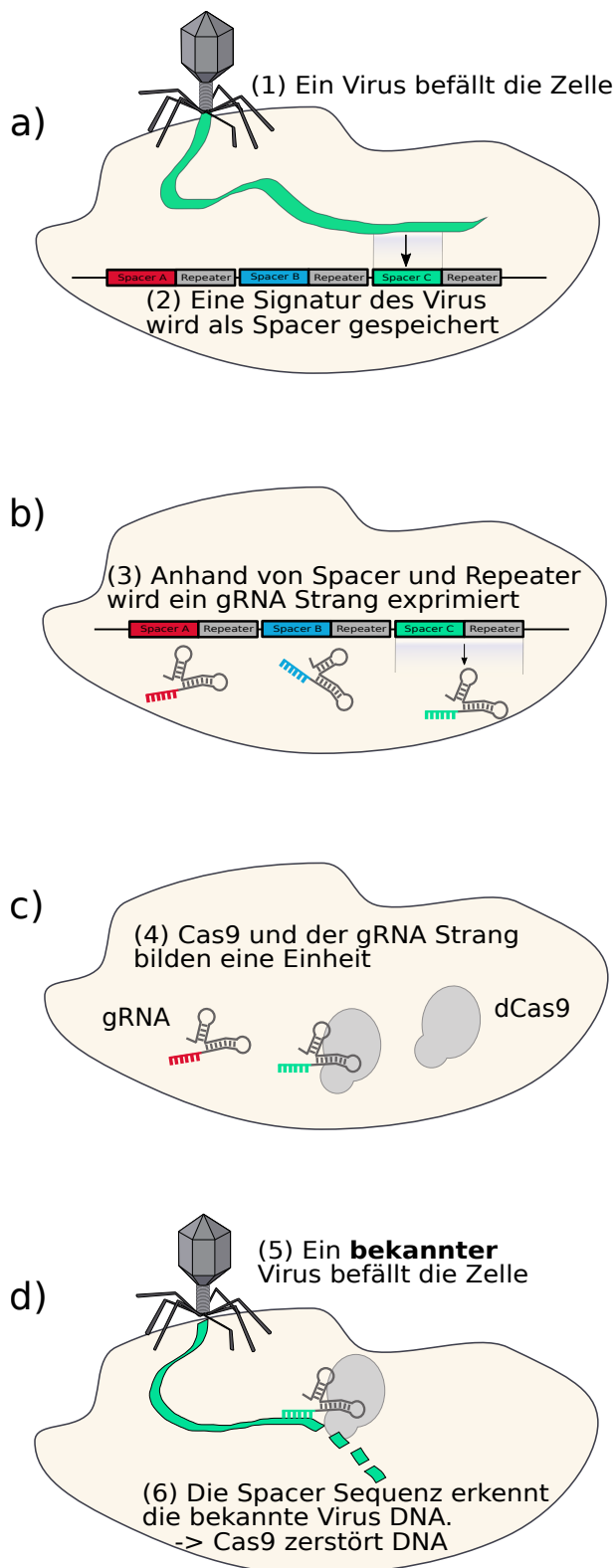
---

### 3.1.2 CRISPR-Interferenz

---

Die Funktionsweise des im Weiteren beschriebenen Schalters baut auf dem CRISPR/Cas9 System auf, verändert dieses aber an entscheidender Stelle und ermöglicht so die Expression eines bestimmten Gens zu aktivieren oder zu inhibieren. Für die im Folgenden genutzte CRISPR Interferenz wird dafür anstelle der Cas9, welche die DNA an definierter Stelle schneidet, eine sogenannte *disfunctional Cas9* (dCas9) verwendet, welche aufgrund zweier Punktmutationen keine DNA mehr schneiden kann, sich aber ansonsten wie funktionale Cas9 verhält. Im einfachsten Fall kann man, um die Expression eines bestimmten Gens zu blockieren, eine gRNA designen, welche an spezifischen Stellen der DNA, also beispielsweise in unmittelbarer Nähe des Promotors, bindet. Wie in Abbildung 3.2 dargestellt, blockiert dort dCas9 dann die Anlagerung von Transkriptionsfaktoren oder RNAP, wodurch die Transkription verhindert wird. Man spricht in diesem Fall auch von einer sterischen Hinderung.

Wie das nächste Kapitel zeigen wird, lassen sich basierend auf dem CRISPR/dCas9-System aber noch deutlich komplexere und damit auch für zukünftige Anwendungen interessantere Schaltungen bauen.

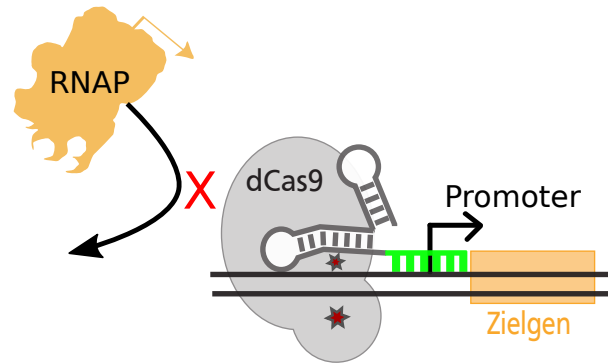


**Abbildung 3.1:** Beschreibung der wesentlichen Vorgänge des natürlichen CRISPR/Cas9-Systems:

a) Wenn ein Virus die Zelle befällt, schleust er eigenes Erbgut in diese Zelle ein (1). Über einen bisher nicht genau verstandenen Mechanismus kann die Zelle nun charakteristische Sequenzen des viralen Erbguts als sogenannte Spacer in die eigene DNA kopieren (2). Dies geschieht in einem bestimmten Pattern, in dem sich Sequenzen verschiedener Viren mit immer wieder gleichen Sequenzen, den Repeatern, abwechseln. Man spricht hier von den *Clustered Regularly Interspaced Short Palindromic Repeats* (CRISPR).

b) Die verschiedenen Spacer-Repeater Paare können zu sogenannter gRNA transkribiert werden (3). Diese kurzen RNA Sequenzen besitzen eine besondere Form, bei der sich die Repeater zu einer Art Schlaufe falten. c) Um die Schlaufe der gRNA kann sich ein besonderes Protein namens Cas9 legen, welches DNA zerschneiden kann. Cas9 und gRNA bilden zusammen eine funktionale Einheit.

d) Wenn die Zelle erneut von einem bekannten Virus befallen wird, so bindet der Spacer-Teil des gRNA/Cas9-Komplexes an die bekannte Sequenz des viralen Erbguts. Cas9 wird dadurch ebenfalls in die Nähe der viralen DNA geleitet und zerschneidet diese DNA. Dadurch wird die virale DNA unschädlich.



**Abbildung 3.2:** Beispielhafte Darstellung der CRISPR-Interferenz. An der DNA gebundene dCas9 verhindert die Anlagerung von RNAP und damit die Transkription des Zielgens. Die zwei Sterne sollen die zwei Punktmutationen verdeutlichen, welche dCas9 dysfunktional machen.

### 3.1.3 CRISPR/dCas9-basierte Schalter in Hefe

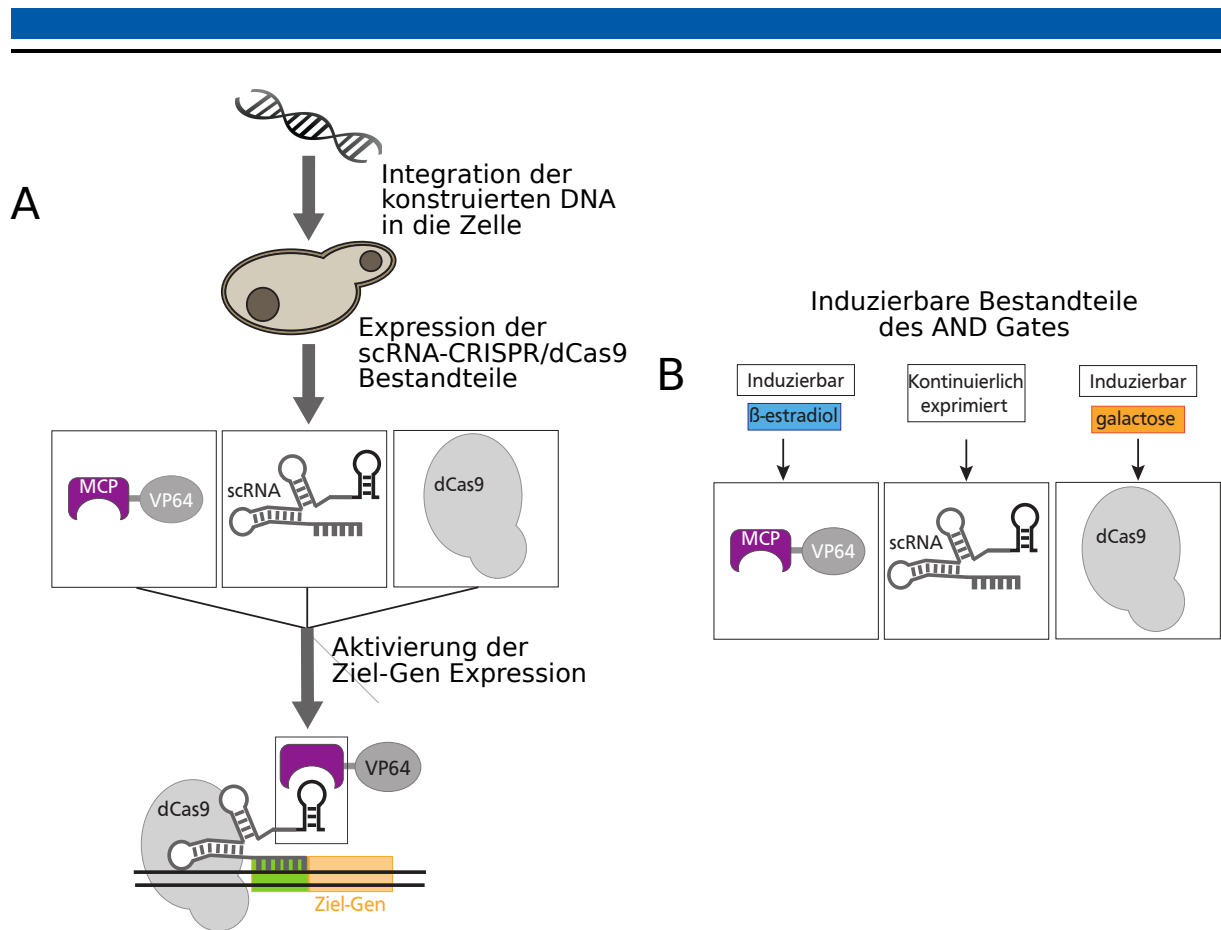
Um das CRISPR/dCas9-System als modularen Schalter zu verwenden, wurde die bisher beschriebene gRNA von Zalatan et al. durch eine sogenannte *scaffold RNA* (scRNA) getauscht [54]. Diese scRNA besitzt eine verlängerte Schleifenstruktur, sodass sie zum einen wie eine normale gRNA funktioniert und zusammen mit dCas9 einen funktionalen Komplex bildet, zum anderen jedoch über die zusätzliche Schleife einen Transkriptionsfaktor binden kann, der die Transkription aktiviert oder inhibiert. Eine schematische Darstellung dieses Systems ist in Abbildung 3.3(A) gezeigt, wobei der Transkriptionsfaktor in diesem Fall MCP-VP64 heißt. Je nach verwendetem Transkriptionsfaktor ist es mit diesem System also möglich, sehr modular und unterschiedlich stark verschiedene Gene zu aktivieren oder zu inhibieren [52–54, 60].

Bis jetzt ist das scRNA-CRISPR/dCas9-System aber noch in keiner Weise ein Schalter, denn es lässt sich, sobald es in die Zelle implementiert wurde, von außen nicht regulieren. Damit dies möglich ist, müssen Komponenten induzierbar, also durch äußere Einwirkung schaltbar gemacht werden.

Im Kapitel 2.1.3 haben wir die Funktionsweise von Genregulationsnetzwerken anhand des GAL-Systems erläutert und hierbei ausgeführt, wie sich der Stoffwechsel von Hefezellen umstellt, wenn sie statt Glucose nur Galactose vorfinden. Diese Umstellung und die damit verbundene Aktivierung der GAL-Gene haben Zalatan et al. genutzt, um die Expression von dCas9 zu aktivieren [47, 54]. Solange die Zellen sich in Glucose befinden, sind die GAL-Gene nicht aktiv und es wird keine dCas9 exprimiert, weshalb ein essenzieller Bestandteil des CRISPR/dCas9-Systems fehlt. Erst durch Umstellung auf Galactose werden die GAL-Gene aktiviert und dCas9 exprimiert. Dadurch kann nun über das modifizierte CRISPR/dCas9-System ein beliebiges Gen an- bzw. ausgeschaltet werden. Der GAL-Schalter ist gut erforscht und stellt ein Standardsystem der Synthetischen Biologie dar. Der große Nachteil dieses Systems ist allerdings, dass es aufgrund der umständlichen Umstellung von Glucose- auf Galactose-Medium nur schwer zu kontrollieren ist und ein langsames Schaltverhalten zeigt. Um diese Nachteile zu umgehen, haben wir den genetischen Schalter, wie im nächsten Kapitel erklärt, um eine weitere induzierbare Komponente erweitert und damit gleichzeitig ein AND-Gate konstruiert.

### 3.2 Ein CRISPR/dCas9-basiertes AND-Gate in Hefe

Wir werden nun auf ein auf dem CRISPR-System beruhendes AND-Gate eingehen, welches in eine Hefezelle (*S. cerevisiae*, Backhefe) integriert wurde. AND bedeutet in diesem Fall, dass zwei verschiedene Bedingungen erfüllt sein müssen, damit der Schalter seinen Zustand wechselt. In un-



**Abbildung 3.3:** A: Schematische Darstellung der Zusammenhänge im scRNA-CRISPR/dCas9-System. B: Induzierbare Bestandteile des AND-Gates.

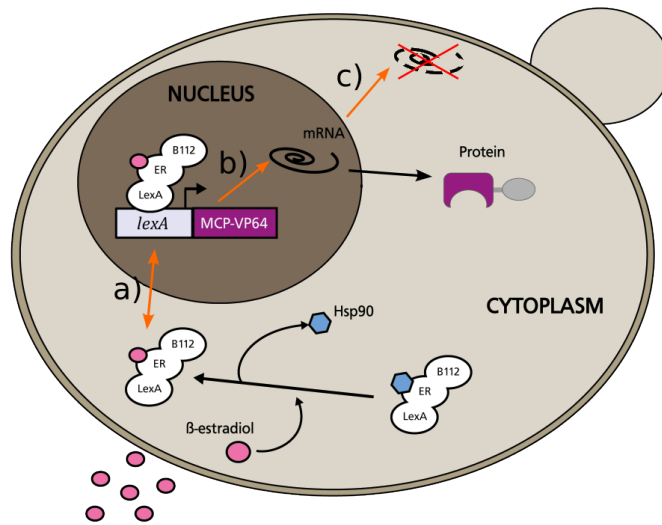
serem Fall werden diese zwei Bedingungen sein, dass sich die Zellen in einem Galactose-Medium befinden und einer ausreichenden Konzentration des Hormons  $\beta$ -Estradiol (ES) ausgesetzt sind.

Um die Nachteile des GAL-Systems zu umgehen, haben wir neben der dCas9 auch den an die scRNA bindenden Transkriptionsaktivator MCP-VP64 schaltbar gemacht. Seine Induktion erfolgt hierbei über das Hormon ES. Der zentrale Vorgang ist hierbei, dass die Expression von MCP-VP64 selbst wieder durch einen Transkriptionsaktivator gesteuert wird. Dieser kann aber nur in den Nukleus eindringen, wenn er vorher eine Bindung mit ES eingegangen ist [61–63]. Der genaue Vorgang ist in Abbildung 3.4 gezeigt. Eine schematische Übersicht über die beiden induzierbaren Komponenten, welche nun in unserem System ein AND-Gatter bilden, ist in Abbildung 3.3(B) dargestellt.

Die beiden verwendeten Induktionsarten kombinieren hierbei zwei verschiedene Vorteile. Wie oben bereits beschrieben ist die Induktion mit Galactose langsam und nur durch komplettes Überführen der Hefekultur in ein neues Medium möglich. Dafür ist der Schalter gut erforscht und nicht toxisch. Demgegenüber ist der ES-Schalter sehr dicht, funktioniert bei minimalen Konzentrationen durch einfache Zugabe von ES und reagiert sehr schnell sowie konzentrationsabhängig. Dafür ist ES in höheren Konzentrationen toxisch für die Hefezellen.

### 3.2.1 Experimentelle Analyse des Schalters

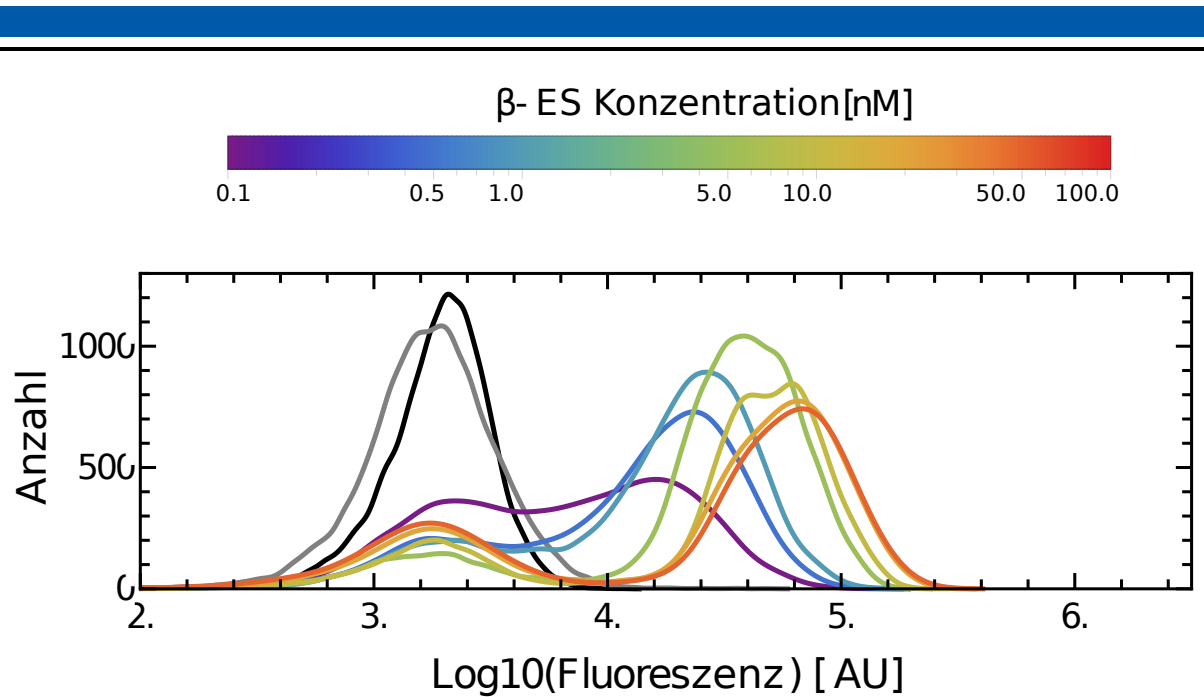
Nach Entwurf und biologischer Integration eines genetischen Schalters interessiert man sich meist dafür, wie gut das Schaltverhalten des Designs ist. Um schnell erkennen zu können, ob ein bestimmtes Zielprotein exprimiert wird, werden bei der Analyse oft fluoreszierende Proteine, wie etwa *grün fluoreszierendes Protein* (GFP) eingesetzt [64–66]. Durch Aktivierung mit einer definierten Wellenlänge lässt sich so anhand der detektierten Fluoreszenz quantifizieren, welche



**Abbildung 3.4:** Aktivierung der Transkription durch ES. Das Fusionsprotein, das eine sogenannte *lexA*-Bindungsdomäne, den menschlichen Östrogen-Rezeptor ER und den Transkriptionsaktivator B112 enthält, ist in Abwesenheit von ES an Hsp90 gebunden (blaues Sechseck). In diesem Zustand kann das Protein nicht in den Nucleus eintreten. Durch Zugabe von ES wird Hsp90 durch ES ersetzt, der Komplex kann in den Kern eintreten und an die *lexA*-Boxen vor dem MCP-VP64-Gen binden. Das Gen wird dann transkribiert, die mRNA geht über in das Zytoplasma und ein funktionales Protein kann exprimiert werden. In Orange sind die Reaktionen markiert, welche in Kapitel 3.2.3 explizit in das Modell aufgenommen werden.

Menge an GFP vorhanden ist. Die Messung der Fluoreszenz erfolgt hierbei meist in einem sogenannten Durchflusszytometer. Dabei handelt es sich um ein Messgerät, in welchem die Zellen einzeln an einem Laserstrahl vorbei fliegen, wodurch die Fluoreszenzproteine aktiviert werden. In einer nachfolgend stattfindenden Messung wird dann die Emission der Zelle gemessen, womit sich auf die relative Menge der enthaltenen Proteine zurückschließen lässt. Als Ergebnis der Zytometrie-Messung erhält man ein Histogramm der gemessenen Fluoreszenzwerte. Für das von uns betrachtete System ist diese Messung in Abbildung 3.5 für verschiedene Induktionsstufen dargestellt. Es ist gut zu erkennen, dass die Zellen bei Wachstum in Glucose keine Fluoreszenz zeigen. Bei Wachstum in Galactose ist die Fluoreszenz hingegen von der Konzentration von ES abhängig. Offensichtlich funktioniert der Schalter also wie angedacht als AND-Gate.

Auffallend an den Zytometrie-Daten ist, dass die Verteilung unter Zugabe von ES eine gewisse Bimodalität zeigt. Die Zellen befinden sich also entweder im AUS- oder im AN-Zustand. Eine häufige Beobachtung bei der Analyse von genetischen Schaltern ist allerdings ein anderes Schaltverhalten, nämlich, dass die Verteilung langsam und homogen in den AN Zustand wandert, also alle Zellen gleichmäßig schalten. Ein Vergleich der beiden Schaltverhalten ist in Abbildung 3.6 gezeigt. Was genau die Ursache für dieses bimodale Schaltverhalten ist, wollen wir im weiteren Verlauf dieser Arbeit durch ein Modell untersuchen. Zuerst soll aber analysiert werden, wie gut das betrachtete System als Schalter nutzbar ist.



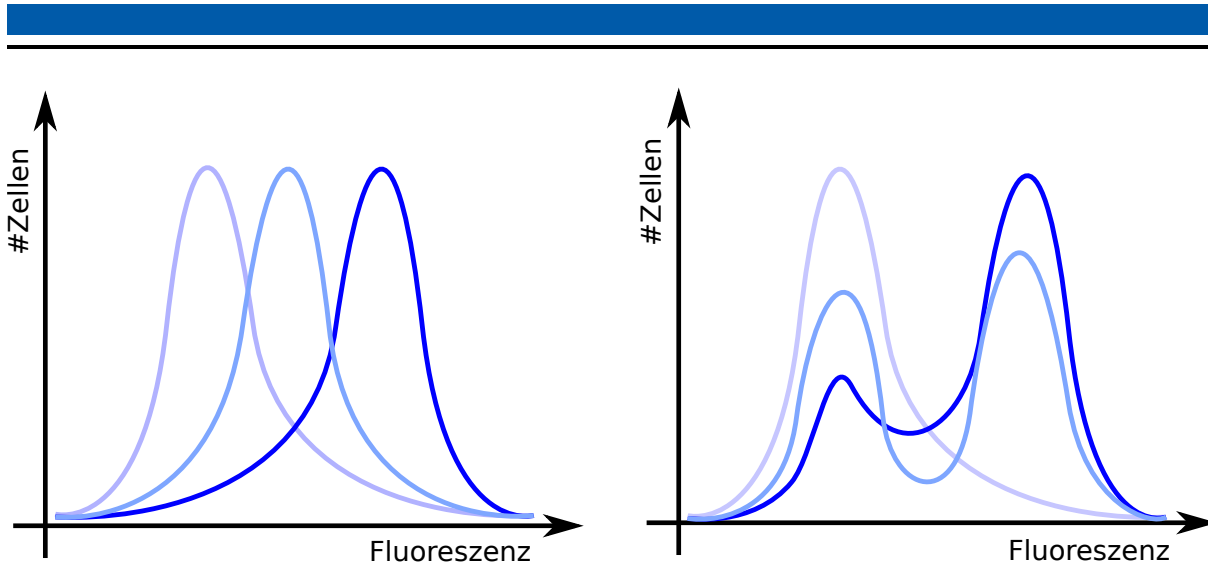
**Abbildung 3.5:** Dosis-Wirkungs-Histogramme des scRNA-CRISPR/dCas9-Systems. Das schwarze Histogramm zeigt den Versuch, bei dem Zellen in Glucose gewachsen sind, das graue Histogramm zeigt Galactose ohne  $\beta$ -Estradiol.

### 3.2.2 Charakterisierung des Schaltverhaltens

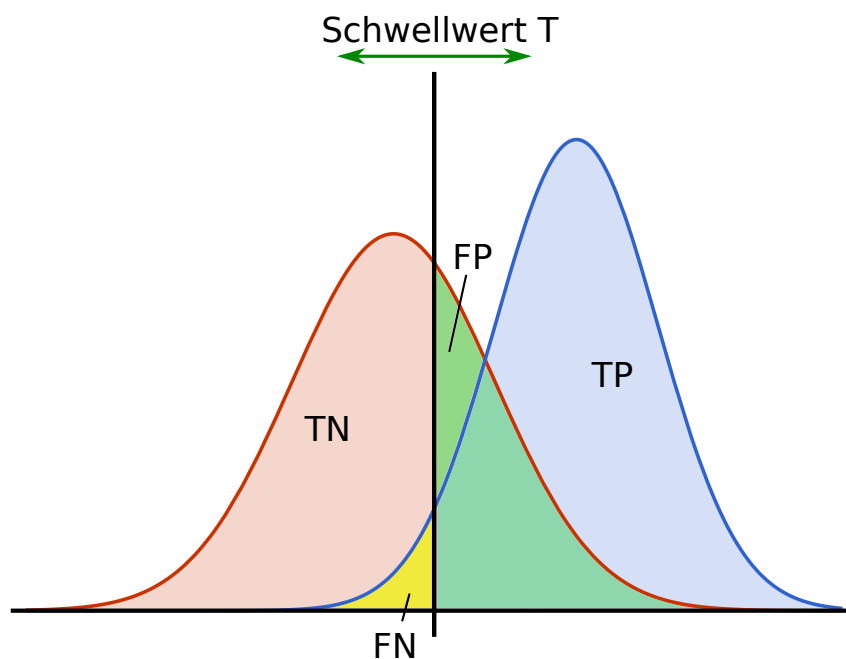
Die Qualität eines genetischen Schalters wird traditionell durch die sogenannte *Fold-Activation* angegeben, also den Unterschied zwischen den Expressionsleistungen im AN- bzw. AUS-Zustand. Eine gute Qualität haben demnach zum einen Schalter, die besonders dicht sind, also im AUS-Zustand nahezu keine Expression zeigen, sowie Schalter, die im AN-Zustand die Expression extrem steigern können. Die Angabe der Fold-Activation bezieht sich dabei immer auf den Mittelwert einer Verteilung von Zellen. Theoretisch ist also möglich, dass nur ein sehr kleiner Teil der Zellen eine sehr hohe Expressionsleistung zeigt und damit den Mittelwert für alle Zellen nach oben zieht, obwohl der Schalter eigentlich nicht gut funktioniert. Aus diesem Grund wurde vorgeschlagen, dass zur Bestimmung der Qualität eines Schalters lieber die sogenannte *Receiver-Operating-Characteristic* (ROC)-Kurve verwendet werden sollte [67]. Da unser Schalter aufgrund des bimodalen Verhaltens durch die Fold-Activation nur schlecht charakterisiert werden kann, werden wir demnach ROC-Kurven verwenden. Im Folgenden wird zuerst erklärt, wie genau diese Kurve berechnet wird und dann unser Schalter damit analysiert.

#### 3.2.2.1 Die ROC-Kurve

Das von uns betrachtete System eines genetischen Schalters kann man als ein logisches Gatter verstehen, welches optimalerweise einen binären Ausgang haben sollte, der entweder den Wert 0 (AUS) oder 1 (AN) annimmt. In der Signaltheorie werden ROC-Kurven genutzt, um die Performance solcher Gatter zu charakterisieren, weshalb diese Kurven auch für genetische Schalter geeignet erscheinen. Ein entscheidender Vorteil der ROC-Kurven ist dabei, dass mit ihnen die Schaltqualität des Gatters bewertet werden kann, ohne sich auf einen genauen Schwellwert  $T$  festzulegen, über dem ein Signal als AN bewertet wird. Dem Vorgehen in [67] folgend erhalten



**Abbildung 3.6:** Vergleich der zwei beobachtbaren Schaltverhalten von Hefezellen. Eine stärkere Induktion ist hierbei durch dunklere Farbe gekennzeichnet. **Links:** Häufig schalten alle Zellen gleichmäßig und proportional zu einer anliegenden Induktion. Das Histogramm wandert dann langsam in Richtung höherer Fluoreszenz. **Rechts:** Der von uns betrachtete Schalter zeigt ein bimodales Schaltverhalten. Die Zellen sind entweder AUS, oder AN. Es gibt kaum Zwischenwerte.



**Abbildung 3.7:** Interpretation der ROC-Kurve. Die Zellkolonie besteht aus zwei verschiedenen Gruppen, nämlich Zellen im AN-Zustand (blaue Kurve) und Zellen im AUS-Zustand (rote Kurve). Um diese Zellen mittels einer Messung zu unterscheiden, muss ein Schwellwert  $T$  eingeführt werden. Zellen, die heller fluoreszieren als der Schwellwert, gelten als AN, Zellen, die weniger stark fluoreszieren, gelten als AUS. Ein großer Teil der Zellen wird dadurch richtig klassifiziert. Bei dem gegebenen Schwellwert sind die Zellen im roten TN (true negative) Bereich tatsächlich AUS und werden auch als AUS klassifiziert. Ebenso werden die Zellen im blauen TP (true positive) Bereich als AN klassifiziert und sind dies auch. Gleichzeitig gibt es aufgrund der Streuung in den Messdaten aber auch Zellen, welche eigentlich AUS sind, durch den Schwellwert aber als AN klassifiziert werden (grüner Bereich, false positives) und Zellen, welche AN sind, aber so schwach leuchten, dass sie als AUS klassifiziert werden (gelber Bereich, false negatives). In der ROC-Kurve wird nun die Wahrscheinlichkeit  $P(TP)$  dafür, dass Zellen im AN Zustand true positives sind, über der Wahrscheinlichkeit  $P(FP)$  dafür, dass Zellen im AUS Zustand false positives sind, aufgetragen.



wir die Koordinaten  $(x(T), y(T))$  der ROC-Kurve für jeden möglichen Schwellwert  $T$  durch die folgenden zwei Gleichungen:

$$\begin{aligned} x(T) &= \frac{1}{N} \sum_{i=1}^N \mathbb{1}(\bar{f}_i > T) \\ y(T) &= \frac{1}{M} \sum_{i=1}^M \mathbb{1}(f_i > T) \end{aligned} \tag{3.1}$$

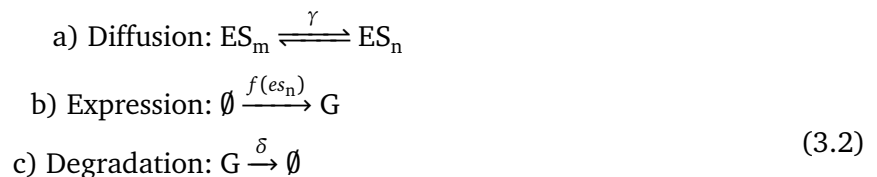
wobei  $N$  und  $M$  die Anzahl von Zellen im AUS- bzw. AN-Zustand darstellen und  $f_i$  sowie  $\bar{f}_i$  die zugehörigen Fluoreszenzwerte der  $i$ -ten Zelle angeben. Die Werte für  $x(T)$  und  $y(T)$  können dabei wie in Abbildung 3.7 dargestellt als Wahrscheinlichkeit für *false positives*  $P(\text{FP})$  bzw. *true positives*  $P(\text{TP})$  verstanden werden.

Ein Lichtschalter stellt einen perfekten Schalter dar, in dem Sinne, dass EIN und AUS komplett voneinander getrennt sind. Das Licht ist entweder vollständig an oder eben aus, ein Zwischenzustand ist nicht vorhanden. Bei einem solchen perfekten Schalter lässt sich also immer ein Schwellwert  $T$  finden, der klar eine Grenze zwischen den beiden Schaltzuständen darstellt. Für jeden möglichen Wert von  $T$  gibt die ROC-Kurve den Anteil der Zellen im AN Zustand an, die vom gewählten Schwellwert richtig klassifiziert werden (true positives) vs. den Anteil der Zellen, die falsch positiv klassifiziert werden (false positives). Ein dysfunktionaler Schalter erscheint in der ROC-Kurve folglich als eine diagonale Linie, da für jeden Schwellwert die selbe Menge an true positives und false positives erzeugt wird. Je besser der Schalter ist, desto mehr verformt sich die ROC Kurve in eine Stufenfunktion, welche vom Punkt  $(0, 0)$  zu  $(0, 1)$  springt und dort konstant bleibt.

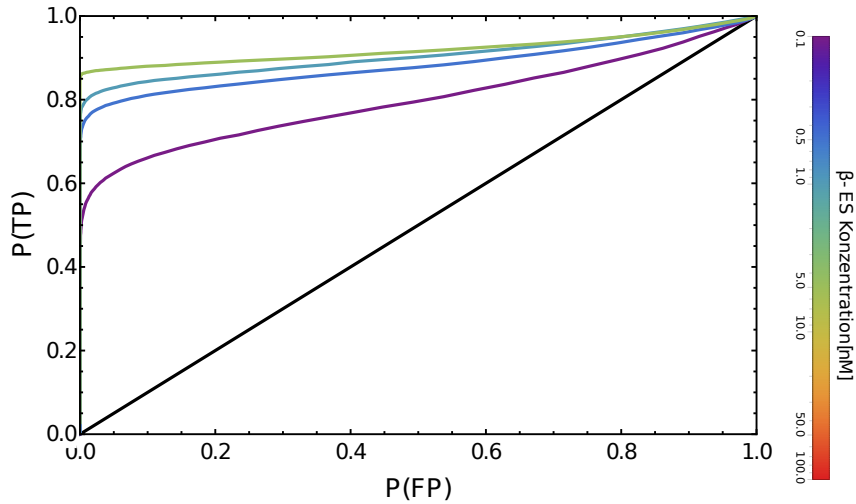
Für die Analyse unseres Schalters nehmen wir an, dass die Zellen, welche in Glucose gewachsen sind und keiner ES-Konzentration ausgesetzt waren, nahezu komplett im AUS-Zustand sind. Wir vergleichen diese dann mit einer jeweils anderen Gruppe von Zellen, welche in Galactose gewachsen sind und einer bestimmten ES Konzentration ausgesetzt waren. Die Analyse unseres Schalters ergibt dann die in Abbildung 3.8 gezeigten ROC-Kurven. Je höher die ES-Konzentration ist, desto besser separiert das System also zwischen AN- und AUS-Zustand.

### 3.2.3 Modellierung

Wie bereits weiter oben beobachtet, ist bei den Zytometrie-Daten auffallend, dass die Verteilung unter Zugabe von ES eine Bimodalität zeigt. Wie es zu diesem bimodalen Verhalten kommt, wollen wir im Folgenden anhand eines einfachen Modells versuchen zu erklären. Wir nehmen dabei an, dass die Zellen bereits in Galactose wachsen, dass also ein Teil des AND-Gates bereits aktiviert ist. Das von uns verwendete Modell beschreibt dann drei Prozesse, nämlich die Diffusion des MCP-VP64-ES-Komplexes in die Zelle und aus dieser heraus, die durch den Komplex aktivierte Expression der GFP-mRNA sowie deren Degradation:



$$\text{mit: } f(\text{es}_n) = \frac{\nu (\text{es}_n/k)^h}{1 + (\text{es}_n/k)^h}.$$



**Abbildung 3.8:** ROC-Kurven für den von uns betrachteten Schalter. Die schwarze Linie zeigt die diagonale ROC-Kurve, die sich aus einem dysfunktionalen Schalter ergeben würde. Die farbigen Linien entsprechen einem betrachteten System, das mit ES-Konzentrationen von 0.1 – 5 nM induziert wurde. Die Kurven sollten so gelesen werden: Wenn man für das 0.1 nM-System eine False-Positive-Rate von 60% anstrebt, würde dies zu einer True-Positive-Rate von etwa 80% führen. Durch Verschieben des Schwellenwerts kann man eine False-Positive-Rate von 5% erreichen, dies geht aber einher mit einem Rückgang der True-Positive-Rate auf 60%.

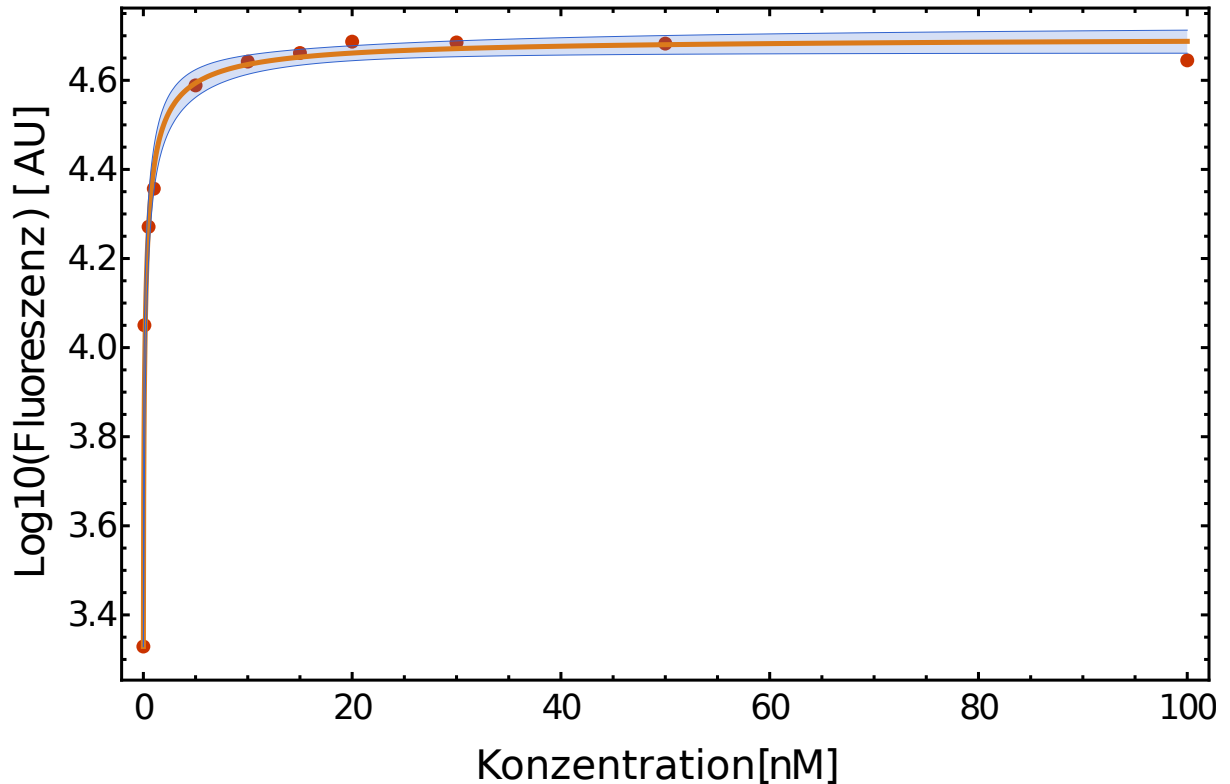
Hierbei stehen  $ES_n$  und  $ES_m$  für den ES-Komplex im Zellkern und im umgebenden Medium und  $G$  für die mRNA des fluoreszierenden Proteins. Die Reaktionen a), b) und c) korrespondieren mit den entsprechenden Kennzeichnungen in Abbildung 3.4. Um die regulierte Expression eines Gens zu modellieren, werden häufig Hill-Funktionen genutzt [20]. In unserem System aktiviert ES die Expression, weshalb wir für  $f(es_n)$  die in Kapitel 2.3.10 definierte Hill-Funktion nutzen. Die Hill-Funktion wird hierbei durch die Michaelis-Konstante  $k$ , den Hill-Koeffizienten  $h$  sowie den maximalen Expressionsfaktor  $\nu$  bestimmt und hängt von der ES-Konzentration im Nukleus ab. Der Parameter  $\delta$  repräsentiert die Degradationsrate der mRNA und  $\gamma$  steht für die Diffusionsrate von ES.

Basierend auf diesen Reaktionsgleichungen lässt sich analog dem Vorgehen in Kapitel 2.3.9 ein System von Differentialgleichungen erstellen, welches die zeitabhängige Konzentration von ES im Zellkern ( $es_n$ ) und die Konzentration der GFP-mRNA ( $g$ ) beschreibt:

$$\begin{aligned} \frac{\partial es_n(t)}{\partial t} &= -es_n(t)\gamma + es_m(t)\gamma \\ \frac{\partial g(t)}{\partial t} &= -g(t)\delta + f(es_n(t)) \end{aligned} \quad (3.3)$$

Die Konzentration von ES im Medium ist hierbei durch den jeweiligen Versuchsaufbau gegeben und kann für die Zeitdauer des Experiments als konstant angenommen werden.

Um die Daten der Zytometrie-Messung mit dem Ergebnis des Modells zu vergleichen, führen wir zusätzlich den Skalierungsfaktor  $\rho$  ein, welcher sowohl die nicht explizit modellierte Translation beinhaltet als auch die Anzahl der GFP-Proteine auf die gemessene Fluoreszenz abbildet. Im Experiment wird die Messung der Fluoreszenz erst nach einigen Stunden durchgeführt, weshalb



**Abbildung 3.9:** Dosis-Wirkungs-Kurve des Schalters. Die Punkte zeigen die Mittelwerte der jeweiligen Zytometrie-Messung an, die durchgezogene Linie zeigt die Erwartung nach unserem Modell. Das blaue Band markiert das 90%-Konfidenzintervall des Mittelwertes des Modells. Die Standardfehler der Versuchsdaten sind kleiner als die Marker, weshalb sie nicht eingezeichnet wurden.

wir von einem stationären Zustand ausgehen können. Folglich können wir in Gleichung (3.3) die Zeitableitungen Null setzen und erhalten:

$$es_n = es_m \quad (3.4)$$

$$g(es_n) = \frac{f(es_n)}{\delta} = \frac{(es_n/k)^h \phi}{1 + (es_n/k)^h} \quad (3.5)$$

wobei  $\phi = \frac{\nu}{\delta}$ . Wie bereits weiter oben beschrieben, müssen wir die Werte von  $g$  noch mit  $\rho$  skalieren, um sie mit den Fluoreszenzwerten vergleichen zu können. Die Funktion, welche wir an die Daten fitten, lautet also:

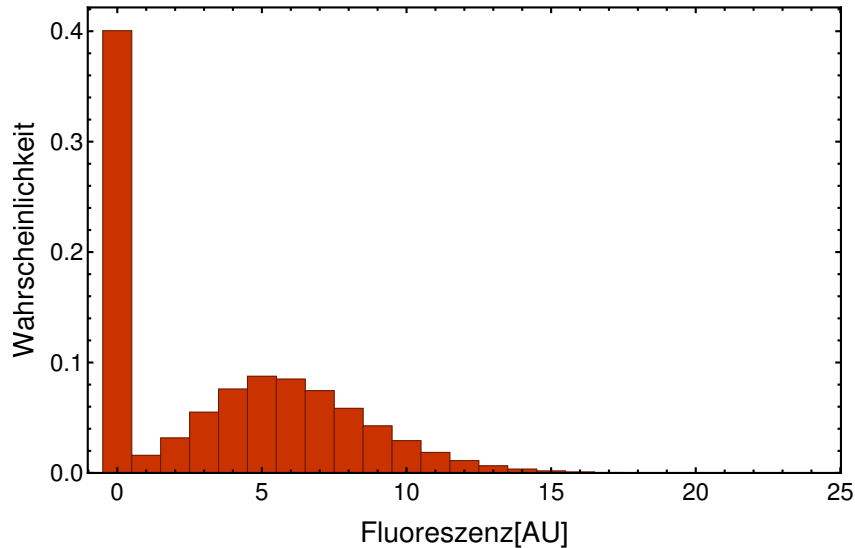
$$g(es_m) = \frac{(es_m/k)^h \Theta}{1 + (es_m/k)^h} \quad (3.6)$$

wobei  $\Theta = \rho \phi = \frac{\rho \nu}{\delta}$  hier ein generalisierter Skalierungsfaktor ist.

Durch den Fit von (3.6) an die Mittelwerte der jeweiligen Zytometrie-Messung erhalten wir dann den in Abbildung 3.9 dargestellten Kurvenverlauf und die in Tabelle 3.1 aufgelisteten Parameter. Das hier verwendete Modell zeigt gute Übereinstimmung mit dem experimentell bestimmten Verlauf der mittleren Fluoreszenz. Allerdings erhalten wir durch den deterministischen Ansatz keine Informationen über die Verteilung dieser Fluoreszenz, können also auch keine Aussagen darüber treffen, ob unser Modell die Bimodalität richtig vorhersagt. Wenn

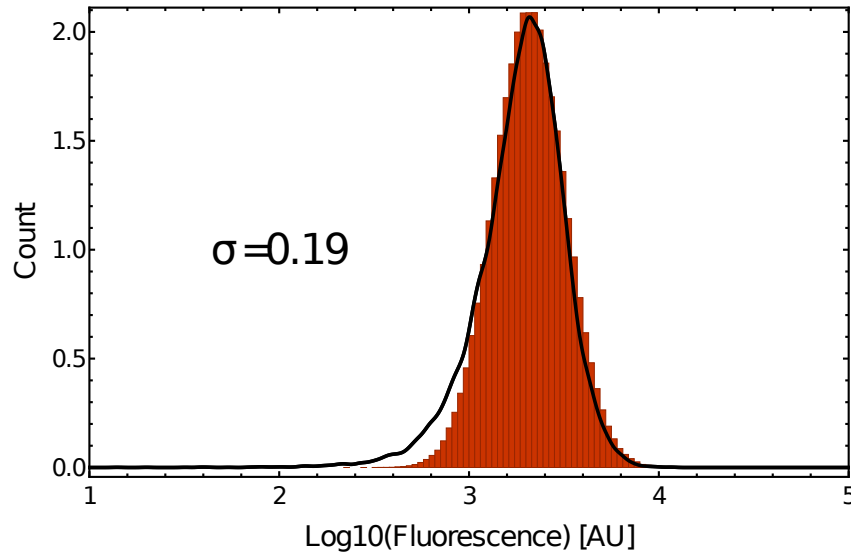
	Name	Schätzwert	Standardfehler	Dimension
h	Hill-Koeffizient	0.818458	0.139886	–
k	Michaelis-Konstante	1.07945	0.248332	nM
$\Theta$	Skalierungsfaktor	47708.6	2294.72	AU

**Tabelle 3.1:** Ergebnis des Fits von Gleichung (3.6) an die Daten



**Abbildung 3.10:** Histogramm der stochastischen Simulation. Für die Abbildung wurde eine ES-Konzentration von 0.5 nM gewählt. Die genutzten Parameter finden sich in Tabelle 3.1, zusammen mit:  $\delta = 0.8$ ,  $\nu = 11.6$ ,  $\gamma = 0.01$ .

man kurz überschlägt, wie groß eine Hefezelle ist und welche Konzentrationen von ES verwendet werden, erkennt man, dass der betrachtete genetische Schalter bei Konzentrationen von  $\approx 2 - 30 \frac{\text{ES-Teilchen}}{\text{Nukleus}}$  funktioniert (bei einer angenommenen Nukleus-Größe von  $\approx 3 \mu\text{m}^3$  [68]). Demnach ist es wahrscheinlich, dass stochastische Effekte eine Rolle spielen, was zu stark veränderten Wahrscheinlichkeitsverteilungen führen kann [69–71]. Wir ergänzen unser bisheriges deterministisches Modell daher um ein stochastisches. Das Modell enthält hierbei die in Gleichung (3.2) gegebenen drei chemischen Reaktionsgleichungen und kann als eine Modifikation des Genexpressionsmodells von Shahrezaei et al. [72] verstanden werden, wobei bei uns ein induzierbarer Promotor eingesetzt wird. Die Werte von  $k$  und  $h$  kennen wir bereits aus dem deterministischen Modell, die Konzentration  $es_m$  ist wieder durch das experimentelle Setup gegeben und die Parameter  $\rho$ ,  $\delta$  und  $\nu$  sind zwar nicht explizit gegeben, aber durch  $\Theta$  in ihrem Wertebereich eingeschränkt. Es bleibt einzig der Parameter  $\gamma$ , welchen wir manuell variieren müssen. Die Reaktionsgleichungen können dann mittels des SSA simuliert werden, wobei wir hierzu die Software *Dizzy* [73] nutzen. In Abbildung 3.10 ist für eine ES-Konzentration von 0.5 nM das Ergebnis einer solchen stochastischen Simulation als Histogramm dargestellt. Die im Experiment beobachtete Bimodalität ist dabei bereits gut zu erkennen. Ein direkter Vergleich zwischen der stochastischen Simulation und der realen Messung ist dadurch allerdings aus zwei Gründen noch nicht möglich. Zum einen registriert ein Zytometer – bedingt durch die Messmethode – auch dann ein Signal, wenn die betrachteten Zellen eigentlich keine Fluoreszenz zeigen. Zum anderen sind Messungen an biologischen Systemen wie Zellen immer einem Rauschen unterworfen, beispielsweise aufgrund der verschiedenen Aktivitätsphasen der einzelnen Zellen. Deutlich zu erkennen ist dies daran, dass auch die Negativkontrolle, in der kein ES vorhanden war und die Zellen in Glucose gewachsen sind, eine recht breite Verteilung um einen Wert  $\gg 0$  [AU] zeigt (schwarzes Histogramm in Abbildung 3.5). Um den Offset der Messung sowie



**Abbildung 3.11:** Analyse der Negativkontrolle zur Abschätzung des extrinsischen Rauschens. Die durchgezogene schwarze Linie stellt die experimentell erhaltenen Daten dar. Das Histogramm ist eine gefittete Gauß-Verteilung. Der Parameter  $\sigma$  gibt die gefittete Standardabweichung in linearer Skala an. Der Offset ist 1943[AU].

das Rauschen in das Modell zu integrieren, fitten wir daher, wie in Abbildung 3.11 dargestellt, eine Gauß-Verteilung mit Mittelwert  $O$  und Varianz  $\sigma^2$  an die Negativkontrolle und fassen diesen Fit als Maß für das Rauschen im System auf.

Letztlich erhalten wir die Daten des Modells dann durch folgende Operation:

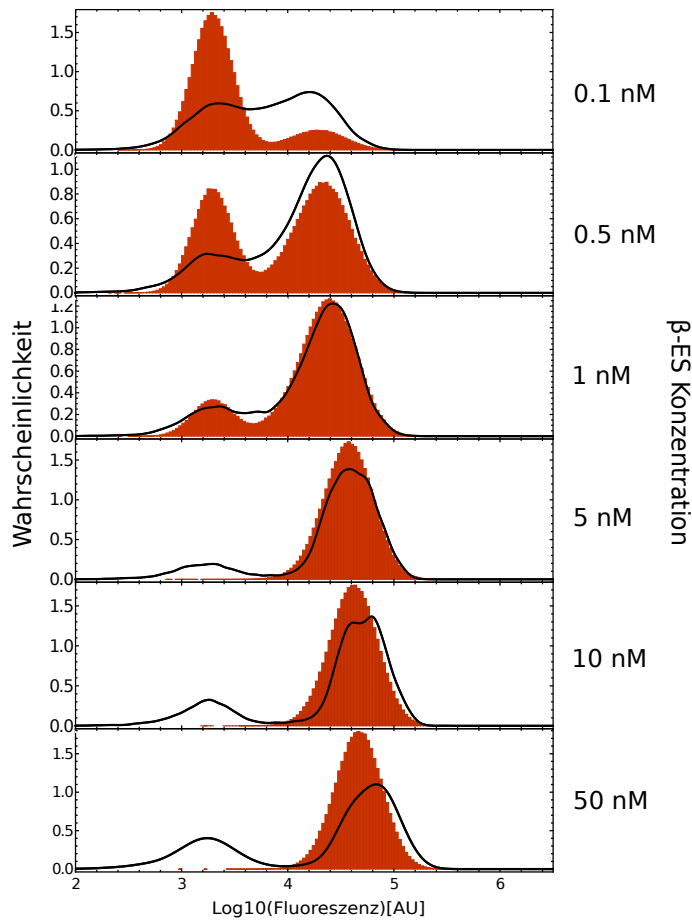
$$\log_{10}[S \cdot \rho + O] + N(0, \sigma^2), \quad (3.7)$$

wobei  $S$  Datenpunkte aus der stochastischen Simulation darstellt,  $\rho$  der Skalierungsfaktor, gegeben durch  $\Theta, \nu$  und  $\delta$  ist,  $O$  den Offset der Negativkontrolle darstellt und  $N(0, \sigma^2)$  eine Zufallsvariable mit Normalverteilung um 0 und Varianz  $\sigma^2$  angibt. Abbildung 3.12 zeigt einen Vergleich zwischen den experimentell erhaltenen Daten und den Ergebnissen der stochastischen Simulation.

Das Modell sagt also die Bimodalität zumindest qualitativ zufriedenstellend vorher. Es finden sich aber immer noch starke Unterschiede zwischen Modell und Experiment, welche sich vor allem bei höheren ES-Konzentrationen zeigen. Der Grund dafür liegt wahrscheinlich in den zwar geringen, aber doch vorhandenen zellschädigenden Effekten von ES. Im nächsten Kapitel werden diese daher noch zusätzlich in das Modell integriert.

### 3.2.4 Zellschädigende Effekte

Bei näherer Betrachtung von Abbildung 3.5 zeigt sich, dass bei einer erhöhten Konzentration von ES einige Zellen keine Fluoreszenz mehr zeigen. Wie in Kapitel 3.2 bereits angemerkt, ist ES in höheren Konzentrationen toxisch für Hefezellen, was in unserem Modell bisher nicht berücksichtigt ist. Nach Stand der aktuellen Experimente ist leider unklar, welcher Mechanismus diese zellschädigenden Effekte auslöst. Um dennoch zumindest eine Intuition dafür zu bekommen, ob und wie sich die Konzentration von ES möglicherweise negativ auf die Fluoreszenz auswirkt, führen wir eine zusätzliche generische Funktion  $\zeta(es)$  ein, welche von der ES-Konzentration abhängt und den Anteil der Zellen angibt, die keine Fluoreszenz zeigen. Den von  $\zeta(es)$  angegebenen Anteil von Datenpunkten setzen wir dann auf den Fluoreszenzwert 0 und reskalieren den

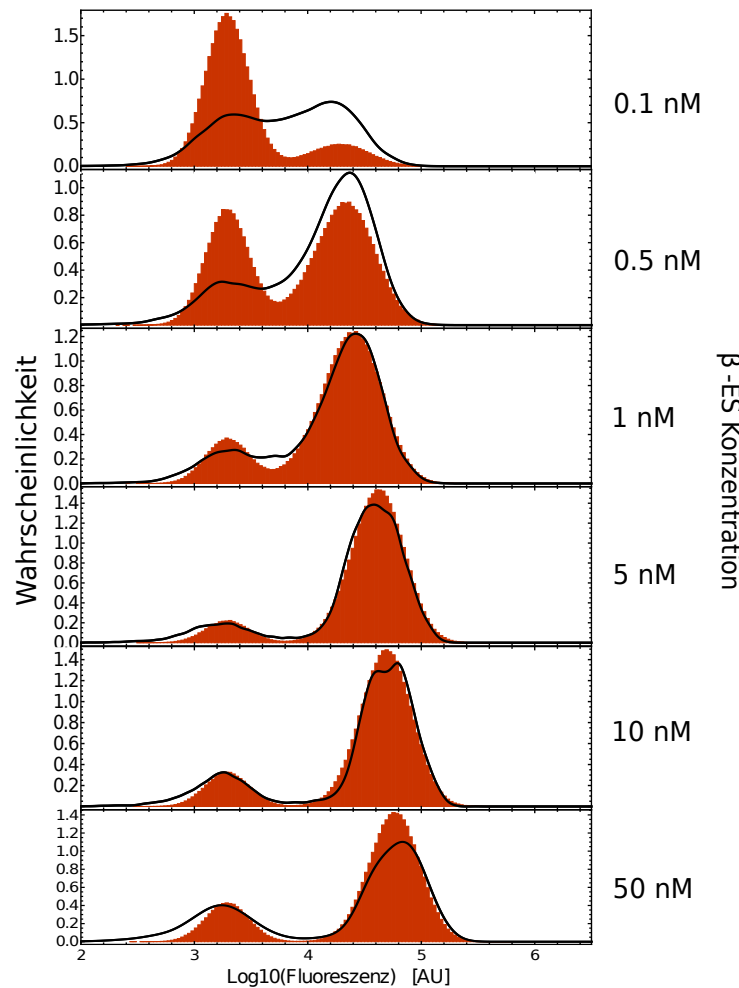


**Abbildung 3.12:** Vergleich der experimentellen Daten (schwarze Linie) und der stochastischen Simulation (Histogramm) für das beschriebene System ohne Beachtung der zellschädigenden Effekte. Die genutzten Parameter finden sich in Tabelle 3.1, zusammen mit:  $\delta = 0.8$ ,  $\nu = 11.6$ ,  $\gamma = 0.01$ .

Parameter  $\Theta$  um den Faktor  $1/\zeta$ , um den beobachteten Mittelwert konstant zu halten. Die Frage, welche sich durch dieses Vorgehen beantworten lässt, ist, ob und in welcher Weise  $\zeta(es)$  von ES abhängt. Leider ist die vorhandene Menge an experimentellen Daten nicht ausreichend, um hier ein belastbares Ergebnis anzugeben. Dennoch können wir erkennen, dass die zellschädigenden Effekte ebenfalls in Form einer Hill-Funktion von ES abhängen. In Abbildung 3.13 wurde hier beispielsweise  $\zeta(es) = 0.21 \frac{1}{1+(6/es)^{1.5}}$  mit  $[es] = \text{nM}$  gewählt.

Durch Inklusion der zellschädigenden Effekte wird die Übereinstimmung zwischen Daten und Modell demnach deutlich verbessert<sup>1</sup>. Die augenscheinlich ungenauen Vorhersagen für 0.1 nM resultieren dabei nicht unmittelbar aus einem Fehler des Modells, sondern zeigen vor allem, wie sensitiv das System in diesem Konzentrationsbereich arbeitet. Eine kleine Änderung von ES um weitere 0.1 nM resultiert bereits in einer starken Änderung der vorhergesagten Verteilung.

<sup>1</sup> Da durch die neue Funktion  $\zeta$  weitere freie Parameter eingeführt wurden, war zu erwarten, dass sich die Übereinstimmung verbessert. Der aktuelle Datenbestand reicht leider nicht aus, um belastbare Aussagen über die Signifikanz der Effekte zu erhalten.



**Abbildung 3.13:** Vergleich der experimentellen Daten (schwarze Linie) und der stochastischen Simulation (Histogramm) für das beschriebene System mit Beachtung der zellschädigenden Effekte. Die genutzten Parameter finden sich in Tabelle 3.1, zusammen mit:  $\delta = 0.8$ ,  $\nu = 11.6$ ,  $\gamma = 0.01$ .

### 3.3 Diskussion der Ergebnisse

In jüngster Zeit ist eine Vielzahl von Technologien zur Bearbeitung und Manipulation genetischer Schaltkreise entstanden [74–76]. Im Idealfall soll der Synthetischen Biologie einmal ein universeller, robuster und adaptiver Werkzeugkasten zur Erzeugung von Gattern und Schaltungen zur Verfügung stehen. Zu diesem Zweck haben wir ein logisches UND-Gatter in *S. cerevisiae* entworfen, das auf einem scRNA-CRISPR/dCas9-System aufbaut [54, 77, 78]. Transkriptionsaktivierung mit scRNAs hat sich hierbei bereits in den letzten Jahren als vielseitiges Werkzeug zur Induktion von Genexpression auf hohem Niveau erwiesen [50, 51]. Die bisherige alleinige Verwendung von Galactose als Induktor der Genexpression macht aber eine Feinabstimmung der Zielgenexpression unmöglich [79]. Um dieses Problem zu lösen, kombinierten wir die induzierbare Expression von dCas9 mit einer ES-abhängigen Expression des MCP-VP64 Transkriptionsaktivators.

Um zu zeigen, dass unsere experimentellen Ergebnisse mit dem Verständnis übereinstimmen, dass ES die Diffusion des Transkriptionsfaktors vermittelt, haben wir ein minimales mathematisches Modell entwickelt. Obwohl nur auf drei möglichen Interaktionen basierend, konnten wir die dosisabhängigen Fluoreszenzwerte korrekt wiedergeben. Eine interessante Beobachtung in unseren Experimenten ist die bimodale Fluoreszenzverteilung, die im deterministischen Modell

---

---

nicht nachgewiesen werden kann. Um die stochastischen Effekte korrekt zu berücksichtigen, haben wir daher zusätzlich ein stochastisches Modell erstellt. Basierend auf den Ergebnissen der deterministischen Anpassung konnten wir hiermit die vollständigen Fluoreszenzverteilungen bei unterschiedlichen ES-Konzentrationen qualitativ nachbilden. Ob die drei von uns in (3.2) angenommenen Reaktionen wirklich die wesentlichen Interaktionen sind, welche letztlich das beobachtete Schaltverhalten erzeugen, lässt sich auf Basis der bisherigen experimentellen Daten allerdings nicht verlässlich sagen. Möglich ist auch, dass ein anderer, im Vergleich mit der Expression sehr langsamer Prozess reversibel ES oder den Promotor bindet. Es sind daher weitere Untersuchungen nötig, um genauere Aussagen über die mikroskopischen Prozesse in Erfahrung zu bringen. Ein interessantes Vorgehen wäre zum Beispiel, ES mit einem Fluorophor zu markieren. Über eine Einzelzellmessung könnte man dann beobachten, wie schnell ES in die Zelle diffundiert und die Rate mit den Vorhersagen des Modells vergleichen. Mit genauerem Verständnis und exakteren Werten für die verschiedenen Raten ist es dann evtl. möglich, auch die Dynamik des Schalters zu modellieren.

Zusammenfassend lässt sich feststellen, dass das in dieser Arbeit entwickelte Modell reproduzierbar und verlässlich die im Experiment beobachteten Phänomene darstellen kann. Es ist dadurch nach Kalibrierung auch in der Lage, für bisher nicht getestete ES-Konzentrationen eine Vorhersage über entstehende Fluoreszenzverteilungen zu machen. Zusätzlich schlägt es einen biologisch sinnvollen Mechanismus vor, der die Form der entstehenden Verteilung erklären kann und eröffnet damit verschiedene Möglichkeiten der weiteren Diskussion und Analyse des scRNA-CRISPR/dCas9-Systems.



---

## 4 Memory-Effekte monomolekularer Umgebungen

In Kapitel 2.3.4 haben wir anhand eines Beispiels erklärt, wie in Markov-Prozessen die Information über den kompletten Zustand des Systems sowie das Wissen über die Vergangenheit voneinander abhängen. Im folgenden Kapitel werden wir annehmen, dass wir keine volle Information über das betrachtete System haben, sondern nur ein Subnetzwerk betrachten, welches in eine uns unbekannt Umgebung eingebettet ist. Ähnlich wie Studentin A im Beispiel aus Kapitel 2.3.4 werden wir dabei zu dem Ergebnis kommen, dass – aufgrund der fehlenden Information über den aktuellen Zustand – die Vergangenheit benötigt wird, um Aussagen über die Zukunft zu tätigen.

Das Kapitel basiert auf Teilen der Ergebnisse, welche zusammen mit Maleen Hanst (AG Koepl), Leo Bronstein (AG Koepl), Barbara Drossel und Heinz Koepl erarbeitet und veröffentlicht wurden [80]. Die Idee, Kontexteffekte näher zu analysieren, entstammt von Heinz Koepl. Maleen Hanst und Leo Bronstein haben sich dabei besonders mit der Anwendung des Mori-Zwanzig-Projektionsoperator-Formalismus auf lineare Umgebungen beschäftigt und damit ein theoretisches Fundament für die Gleichungen 4.11 geschaffen. Die darauf folgende Behandlung von Umgebungen mit monomolekularen Reaktionen sowie die Wahl und Analyse der Beispiele ist dann die Eigenleistung des Autors. Vor allem beim Übergang von der linearen Kette zum Random Walk hat er dabei sehr von Ratschlägen und Ideen von Barbara Drossel und bei der Wahl der Beispiele von Heinz Koepl profitiert.

---

### 4.1 Kontexteffekte in der Synthetischen Biologie

---

Im Kapitel 3 haben wir mehrfach ein zentrales Ziel der Synthetischen Biologie erwähnt, nämlich die Entwicklung modularer genetischer Bausteine, welche frei kombiniert werden können [81]. Wie in einem elektrischen Schaltkreis soll jedes der Bauteile im optimalen Fall nur eine bestimmte Aufgabe robust und zuverlässig erfüllen, ohne dabei die Funktion der anderen Bauteile zu beeinflussen. Die Entwicklung der einzelnen Module geschieht häufig isoliert, womit gemeint ist, dass ein einzelnes Bauteil für sich analysiert und optimiert wird. Basierend auf den dabei erhaltenen Ergebnissen werden verschiedene Komponenten dann zu einem größeren Schaltkreis zusammengesetzt. Aus elektrischen Schaltkreisen ist bekannt, dass es zu Problemen kommen kann, wenn zum Beispiel die Impedanz der Bauteile nicht angepasst wird. Ähnlich verhält es sich in der Synthetischen Biologie. Einzelne Bauteile oder Schaltkreise, welche vorher in Isolation getestet und entwickelt wurden, befinden sich nach Integration in einen größeren Schaltkreisverbund in einem fluktuierenden Umfeld, welches die eigene Funktion stark beeinträchtigen kann [82]. All die Auswirkungen dieser Umgebung werden in der Synthetischen Biologie als Kontexteffekte zusammengefasst. Ausgehend von ihrem jeweiligen Ursprung werden die Umgebungen in verschiedene Klassen eingeteilt [83, 84]:

**Wirtsvermittelter Kontext:** Die Umgebung, welche in der Zelle natürlich vorhanden ist

**Kompositioneller Kontext:** All jene Umgebung, welche erst durch die Verknüpfung verschiedener Bauteile entsteht

**Externer Kontext:** Fluktuationen und Störung von außerhalb, also zum Beispiel Temperaturschwankungen

---

Um verlässliche genetische Bausteine zu designen, muss man sicherstellen, dass die verschiedenen Kontexteffekte keinen großen Einfluss auf deren korrekte Funktion haben. Die Module müssen also robust gegenüber dem Kontext sein. Zahlreiche Forschungsprojekte haben sich damit beschäftigt, wie man zum einen die Kontexteffekte minimieren kann, aber auch, wie man deren Auswirkungen schon während des Designprozesses vorhersagen und die Module diesbezüglich robustifizieren kann.

---

#### 4.1.1 Stand der Forschung

---

Saez-Rodriguez et al. und später auch Del Vecchio et al. betrachteten beispielsweise kompositionelle Kontexteffekte, indem sie untersuchten, wie der Zusatz eines Downstream-Moduls (also ein Modul, welches später in der Signalkette auftritt) die Qualität des Upstream-Moduls (also eines, welches früher in der Signalkette auftritt) verändern kann. Dieser Effekt wird als *Retroaktivität* bezeichnet [5, 85]. Die Autoren schlugen hierbei einen allgemeinen Ansatz zur Quantifizierung der Auswirkungen der Retroaktivität vor. Darüber hinaus entwickelten sie einen genetischen Isolator, der die Retroaktivität abschwächt, und demonstrierten später erfolgreich seine Funktionalität. Eine Grundannahme von Del Vecchios Analyse ist, dass eine Trennung der Zeitskalen zwischen der Dynamik der beiden beschriebenen Module möglich ist. Während diese Annahme für kompositionelle Kontexteffekte oftmals gilt, kann sie nicht auf wirtsvermittelte Kontexteffekte angewendet werden, da die zelluläre Umgebung oft ähnliche Zeitskalen wie das entworfene Subnetz hat.

Ein weiterer wichtiger Beitrag im Bereich der Kontexteffekte wurde von Liebermeister et al. geliefert, welche einen allgemeinen Formalismus entwickelten, mit dessen Hilfe kompositionelle und wirtsvermittelte Kontexteffekte in die Ratengleichungen eines Teilnetzes einbezogen werden können, insbesondere unter Berücksichtigung der dynamischen Reaktion der Umgebung [86]. Zu diesem Zweck teilten sie das gesamte Netzwerk in ein betrachtetes Netzwerk, das so genannte Subnetz, und die verbleibende Umgebung auf. Nach Linearisierung der Umgebungsbeschreibung um einen stationären Zustand, reduzierten sie die Dimensionalität des Problems, indem sie die Dynamik auf den Raum der dominanten dynamischen Moden projizierten. Mit der Anwendung ihrer Methode auf ein metabolisches Netzwerk, das in eine Umgebung mit Parameterunsicherheit eingebettet ist, demonstrierten Liebermeister et al., wie die Berücksichtigung der Umgebung die Vorhersagen des Modells verbessert, auch wenn die Parameter nicht genau bekannt sind. Ihre Modellreduktion erwies sich gegenüber herkömmlichen Ansätzen in vielen Fällen als besser. Dennoch erwähnten Liebermeister et al. mehrere Einschränkungen ihrer Methode, nämlich die Erfordernis eines stationären Zustands in der Umgebung und die Nichterhaltung von Erhaltungsgrößen in Subnetz und Umgebung.

Die Ergebnisse von Liebermeister et al. wurden kürzlich von Rubin et al. erweitert, die den Mori-Zwanzig-Projektionsoperator-Formalismus nutzten, um eine Beschreibung für ein generisches Subnetz zu erhalten, das die kontextuellen Effekte über sogenannte Memory-Terme [87] beinhaltet. Durch die Verwendung von Memory-Termen ist es möglich, komplexe gegenseitige Wechselwirkungen zwischen Subnetz und Umgebung in die Beschreibung des Subnetzes aufzunehmen, z. B. wenn das Subnetz die Umgebung beeinflusst und diese Umgebungsänderungen später wieder in das Subnetz koppeln. Mit der erwähnten Projektionstechnik zeigten die Autoren, dass die Einbeziehung von Memory-Termen zu einer höheren Genauigkeit der vorhergesagten Subnetzwerkodynamik führt.

In diesem Teil der Arbeit werden wir einen weiteren wichtigen Beitrag zum Verständnis und der Analyse von Umgebungseffekten liefern. Hierbei beschränken wir uns auf Umgebungen, welche aus monomolekularen Reaktionsnetzwerken bestehen. Während Reaktionsnetzwerke normalerweise natürlich nichtlinear sind, können sie in vielen Fällen zu solchen monomolekularen, linearen Netzen vereinfacht werden, entweder durch Entwicklung um einen stationären Zustand [86] oder durch Projektion auf ein näherungsweise äquivalentes System [88].

Basierend auf dieser linearen Darstellung werden wir dann systematisch den Einfluss verschiedener generischer Umgebungen, in welche das betrachtete Subnetzwerk eingebettet ist, untersuchen. Im Unterschied zu bisherigen Arbeiten legen wir uns dabei nicht auf eine bestimmte vorgegebene Umgebung fest [87], sondern gehen auf die Situation ein, dass kein spezifisches Wissen über die Umgebung vorhanden ist. Hierzu betrachten wir verschiedene prototypische Umgebungen mit kettenartiger Topologie und leiten analytische Ausdrücke für die sogenannten Memory-Kerne dieser Umgebungen her.

Um schließlich die Auswirkungen der Kontexteffekte zu verdeutlichen, zeigen wir – ähnlich wie auch Del Vecchio et al. – anhand eines genetischen Oszillators, wie eine Ankopplung an die Umgebung dessen korrekte Funktion beeinträchtigt. Zusätzlich zeigen wir am Beispiel einer negativen Feedbackschleife, welche unerwarteten und dennoch weitreichenden Effekte die korrekte Inklusion der Umgebung erzeugen kann.

---

## 4.2 Hintergrundwissen

---

Im Verlauf dieses Kapitels werden wir die Umgebung durch eine Kette von Substanzen nähern, welche durch sogenannte monomolekulare Reaktionen miteinander verbunden sind. Alsdann werden wir diese Umgebung durch einen Random Walk mit absorbierenden Rändern beschreiben. Aus diesem Grund definieren wir hier zuerst lineare und monomolekulare Reaktionsnetzwerke und geben eine kurze Einführung in Random Walks sowie absorbierende Zustände.

Zum Vergleich von stochastischen und deterministischen Trajektorien werden wir außerdem später die sogenannte spektrale Leistungsdichte verwenden, welche wir daher hier auch kurz definieren.

---

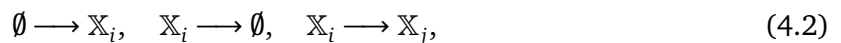
### 4.2.1 Definition: Lineare und monomolekulare Reaktionsnetzwerke

---

Eine Reaktion heißt *linear*, wenn die Propensität maximal linear von einer Teilchenzahl abhängt. Gemäß Definition (2.3) nehmen lineare Reaktionen also immer eine der folgenden beiden Formen an:



wobei die rechte Seite jeweils beliebig viele Produkte enthalten kann. Eine Reaktion heißt *monomolekular*, wenn sie linear ist und zusätzlich für die Gesamtzahl der Produkte  $r \leq 1$  gilt. Folglich gibt es nur die folgenden drei monomolekularen Reaktionen:



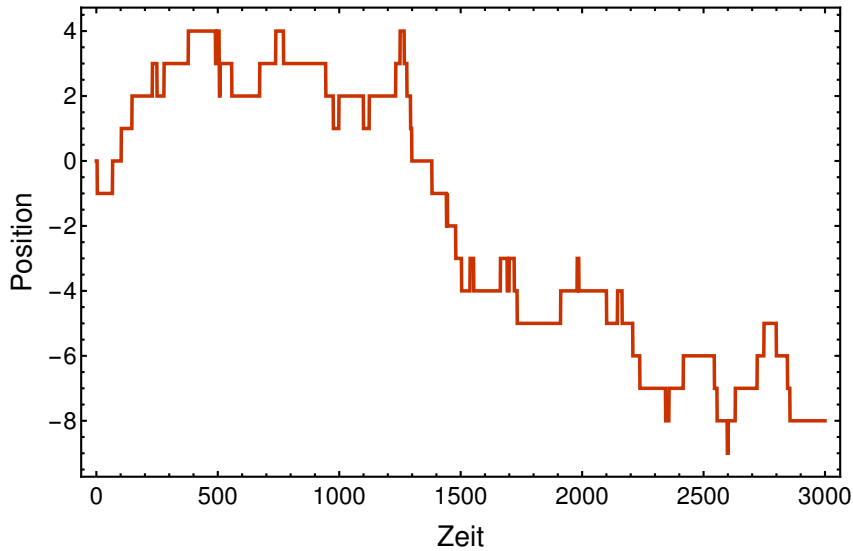
wobei die erste Reaktion eine Erzeugung, die zweite eine Zersetzung und die dritte eine Umwandlung einer bestimmten Substanz darstellt. Ein Reaktionsnetzwerk wird als linear (monomolekular) bezeichnet, wenn es ausschließlich aus linearen (monomolekularen) Reaktionen besteht. Im Weiteren werden wir dabei ein Netzwerk insbesondere dann als monomolekular definieren, wenn alle internen Reaktionen monomolekular sind.

---

### 4.2.2 Random Walk und absorbierende Zustände

---

Zur Motivation eines eindimensionalen *Random Walks* stellen wir uns einen Betrunkenen vor, der eine gerade Straße entlang läuft [89]. Aufgrund seines Zustands torkelt er in die zwei möglichen Richtungen. Wir gehen davon aus, dass jeder seiner Schritte die Länge  $\Delta x$  hat. Mit der Rate  $\alpha$  macht der Betrunkene einen Schritt nach links und mit der Rate  $\beta$  nach rechts. Die Bewegung



**Abbildung 4.1:** Möglicher Verlauf eines Random Walks. Parameter in diesem Beispiel:  $\alpha = \beta = 0.01$

des Betrunkenen ist dann ein Beispiel für einen Random Walk. Mittels der aus dem Kapitel 2 bekannten Grundlagen lässt sich ein Random Walk leicht durch eine ME beschreiben:

$$\frac{\partial P(x, t)}{\partial t} = \alpha P(x + \Delta x, t) + \beta P(x - \Delta x, t) - (\alpha + \beta)P(x, t). \quad (4.3)$$

Die mögliche Trajektorie eines einzelnen Betrunkenen lässt sich dann durch den SSA simulieren und könnte aussehen wie in Abbildung 4.1 gezeigt. Da es sich um einen Zufallsprozess handelt, sieht dieser Verlauf natürlich für jede neue Simulation anders aus. Wenn  $\alpha = \beta$  ist, hat die Bewegung keine Vorzugsrichtung, man spricht in diesem Fall von einem *ungerichteten Random Walk*. Für den Fall  $\alpha \neq \beta$  hat die Bewegung trotz des torkelnden Charakters eine Vorzugsrichtung und bewegt sich im Laufe der Zeit im Mittel in eine bestimmte Richtung. Man spricht dann von einem *gerichteten Random Walk*.

Wir stellen uns nun vor, dass der Betrunkene auf dem Weg zu seinem Bett ist. Sobald er das Bett erreicht, schläft er ein und bewegt sich nicht mehr. Das Bett ist dann ein sogenannter absorbierender Zustand, welcher dadurch definiert ist, dass er nur betreten, aber nicht verlassen werden kann. In der ME drückt sich dies dadurch aus, dass alle Raten heraus aus diesem Zustand null sind. In unserem eindimensionalen Fall kann der Random Walk in zwei Richtungen laufen, also nur zwei verschiedene absorbierende Zustände erreichen.

---

#### 4.2.3 Absorptionswahrscheinlichkeit und Rückwärts-Mastergleichung

---

Eine Größe, welche bei der Analyse von Random Walks mit absorbierenden Zuständen häufig von Interesse ist, ist die sogenannte *Überlebenswahrscheinlichkeit*  $G(x_a, x, t)$ . Sie gibt die Wahrscheinlichkeit dafür an, dass sich der Random Walker zum Zeitpunkt  $t$  noch nicht im absorbierenden Zustand  $x_a$  befindet, unter der Voraussetzung, dass er zum Zeitpunkt  $t = 0$  bei  $x$  gestartet ist. Dementsprechend ist *Absorptionswahrscheinlichkeit*, also die Wahrscheinlichkeit dafür, dass der Random Walker zum Zeitpunkt  $t$  bereits in  $x_a$  absorbiert wurde:  $1 - G(x_a, x, t)$ .

Bei der Herleitung der ME haben wir angemerkt, dass diese eigentlich Vorwärts-Mastergleichung genannt werden müsste. Dies deutete bereits darauf hin, dass es noch eine andere, nämlich die Rückwärts-Mastergleichung gibt. Die Vorwärts-Mastergleichung benötigt eine feste Anfangsbedingung, von der ausgehend die zeitliche Entwicklung des Systems berech-

net werden kann. Die Rückwärts-Mastergleichung dagegen setzt einen finalen Zustand voraus, von dem ausgehend rückwärts in der Zeit der Zustand berechnet wird, der zum finalen Zustand geführt hat. Da die Herleitung nahezu analog zu der der Vorwärts-Mastergleichung verläuft, geben wir hier nur das Ergebnis an:

$$-\frac{dP(z_g, t|z_0, t_0)}{dt_0} = \sum_h a_{h0} (P(z_g, t|z_h, t_0) - P(z_g, t|z_0, t_0)). \quad (4.4)$$

Die Rückwärts Mastergleichung stellt nicht die Frage »wohin geht die Wahrscheinlichkeit« sondern »woher ist sie gekommen«. Besonders bei der Berechnung von Absorptionswahrscheinlichkeiten spielt die Rückwärts-Mastergleichung eine wichtige Rolle. Anschaulich lässt sich der Vorteil der Rückwärts-Gleichung durch Betrachtung der Randbedingungen verstehen. Während in der Vorwärtsrichtung ein absorbierender Zustand mit einer bestimmten Rate betreten (und dann nicht mehr verlassen) werden kann, ist in der Rückwärtsrichtung die Rate dafür, von einem absorbierenden Zustand gekommen zu sein, einfach null.

---

#### 4.2.4 Spektrale Leistungsdichte

---

Im folgenden Kapitel werden wir eine deterministische Methode entwickeln, mit deren Hilfe sich die Dynamik von bestimmten Reaktionsnetzwerken approximiert berechnen lässt. Um die Qualität der Näherung zu bestimmen, vergleichen wir unsere Ergebnisse dann mit den Trajektorien aus dem SSA. Aufgrund der Stochastizität lassen sich allerdings deterministische und stochastische Trajektorien nicht direkt vergleichen. Vielmehr bedarf es geeigneter Größen, welche über den stochastischen Prozess mitteln. In Kapitel 3 haben wir hierzu beispielsweise die Wahrscheinlichkeitsverteilungen verglichen. Im Verlauf dieses Kapitels wollen wir allerdings die Dynamik von deterministischer und stochastischer Lösung vergleichen, wozu die statische Wahrscheinlichkeitsverteilung nicht geeignet ist. Stattdessen nutzen wir hierzu die sogenannte *spektrale Leistungsdichte*. Hierbei handelt es sich um eine Größe, welche die Leistung eines Signals pro infinitesimalem Frequenzband quantifiziert und meist in der Dimension dB/Hz angegeben wird. Zur Berechnung der spektralen Leistungsdichte eines Signals  $\xi(t)$  greifen wir zurück auf dessen Fourier-Transformation:

$$\hat{X}(\omega) = \int_0^T \xi(t) e^{-i\omega t} dt. \quad (4.5)$$

Die spektrale Leistungsdichte ist dann definiert als:

$$S(\omega) = \lim_{T \rightarrow \infty} \frac{1}{2\pi T} |\hat{X}(\omega)|^2 \quad (4.6)$$

---

### 4.3 Subsystem und Umgebung

---

Im Folgenden werden wir große Reaktionsnetzwerke betrachten, welche wir in zwei Unter-einheiten partitionieren. Das System, welches für uns von Interesse ist, weil es zum Beispiel hinsichtlich einer bestimmten Funktion entworfen und nun darauf getestet werden soll, nennen wir Subsystem (S). Das (meist) deutlich größere Netzwerk, welches dieses Subsystem umgibt und dessen genaue Topologie oft unbekannt ist, werden wir Umgebung (U) nennen. Das gesamte Reaktionsnetzwerk, welches gemäß Gleichung (2.5) beschrieben werden kann, fassen wir im

Folgendes als eine Kombination von Substanzen auf, welche entweder zum Subsystem gehören ( $\mathbb{X}^S$ ), oder zur Umgebung ( $\mathbb{X}^U$ ). Es lässt sich dann schreiben:

$$R_j : \sum_{i=1}^{N_S} s_{ij}^S \mathbb{X}_i^S + \sum_{i=1}^{N_U} s_{ij}^U \mathbb{X}_i^U \rightarrow \sum_{i=1}^{N_S} r_{ij}^S \mathbb{X}_i^S + \sum_{i=1}^{N_U} r_{ij}^U \mathbb{X}_i^U . \quad (4.7)$$

Für ein gut durchmisches System können wir dann den Zustandsvektor angeben, welchen wir aufteilen in:  $\mathbf{n} = (\mathbf{n}^S, \mathbf{n}^U)$ . Für die CME gilt:

$$\frac{dP(\mathbf{n}, t)}{dt} = \sum_j f_j(\mathbf{n} - \mathbf{s}_j) P(\mathbf{n} - \mathbf{s}_j, t) - \sum_j f_j(\mathbf{n}) P(\mathbf{n}, t), \quad (4.8)$$

wobei sich die stöchiometrische Matrix aus zwei Blockmatrizen für Subsystem und Umgebung zusammensetzt:

$$\mathbf{s} = \begin{bmatrix} \mathbf{s}^S \\ \mathbf{s}^U \end{bmatrix}. \quad (4.9)$$

Analog dem Vorgehen in Kapitel 2.3.9 lässt sich diese CME dann für große Systeme allgemein in ein System von RRE überführen:

$$\begin{aligned} \frac{d}{dt} \tilde{\mathbf{n}}^S &= \sum_j \mathbf{s}_{\cdot j}^S \tilde{f}_j(\tilde{\mathbf{n}}^S, \tilde{\mathbf{n}}^U) \\ \frac{d}{dt} \tilde{\mathbf{n}}^U &= \sum_j \mathbf{s}_{\cdot j}^U \tilde{f}_j(\tilde{\mathbf{n}}^S, \tilde{\mathbf{n}}^U). \end{aligned} \quad (4.10)$$

Wie weiter oben bereits beschrieben, interessieren wir uns insbesondere für das Verhalten des Subsystems, also die zeitliche Änderung der Konzentrationen  $\tilde{\mathbf{n}}^S$ . Die Umgebung stellt eine Art Störung dar, welche wir nicht – wie häufig gemacht – vernachlässigen wollen, sondern weiterhin mit in die Analyse einbeziehen. Im Allgemeinen liegen zur Topologie der Umgebung allerdings keine genauen Informationen vor, weshalb Annahmen über deren Aussehen getroffen werden müssen. Deutlich vereinfachen lässt sich die allgemeine Beschreibung beispielsweise, wenn wir annehmen, dass die Umgebung sowie die Kopplung zwischen Umgebung und Subsystem gemäß der Definition (4.2) aus linearen Reaktionen bestehen.

In diesem Fall sind die Raten der Umgebung und der Kopplung entweder konstant oder nur linear von einer Konzentration abhängig. Wir können daher das Reaktionssystem (4.10) vereinfachen zu:

$$\begin{aligned} \frac{d}{dt} \tilde{\mathbf{n}}^S &= q(\tilde{\mathbf{n}}^S) + \mathbf{T}^{SU} \tilde{\mathbf{n}}^U \\ \frac{d}{dt} \tilde{\mathbf{n}}^U &= \mathbf{T}^{UU} \tilde{\mathbf{n}}^U + \mathbf{T}^{US} \tilde{\mathbf{n}}^S + \mathbf{b}^U. \end{aligned} \quad (4.11)$$

Hier steht  $q(\tilde{\mathbf{n}}^S)$  für alle Subsystem-Subsystem Reaktionen, welche beliebiger Gestalt sein dürfen. Die Ratenmatrix  $\mathbf{T}^{UU}$  enthält als Einträge die Ratenkonstanten  $\{c_j\}$  der verschiedenen linear von  $\tilde{\mathbf{n}}$  abhängenden und der Vektor  $\mathbf{b}^U$  die der erzeugenden Umgebungsreaktionen. Die beiden Kopplungsmatrizen  $\mathbf{T}^{SU}$  und  $\mathbf{T}^{US}$  enthalten die Raten derjenigen Reaktionen, welche Umwandlungen von Umgebungssubstanzen in Subsystemsubstanzen und umgekehrt darstellen. Die gesamte Um-

gebungsgleichung in (4.11) ist nun linear und damit entscheidend vereinfacht. Durch formales Lösen der zweiten Gleichung und Einsetzen in die erste erhalten wir [90]:

$$\frac{d}{dt}\tilde{\mathbf{n}}^S = q(\tilde{\mathbf{n}}^S) + \int_0^t dt' \mathbf{T}^{\text{SU}} e^{(t-t')\mathbf{T}^{\text{UU}}} (\mathbf{T}^{\text{US}} \tilde{\mathbf{n}}^S(t') + \mathbf{b}^U) + \mathbf{T}^{\text{SU}} e^{t\mathbf{T}^{\text{UU}}} \tilde{\mathbf{n}}^U(0). \quad (4.12)$$

Bevor wir mit Hilfe dieser Gleichung die Effekte von linearen Umgebungsnetzwerken analysieren, wollen wir kurz die Bedeutung der einzelnen Terme diskutieren und für die weitere Verwendung eine konsistente, an [87] angelehnte Nomenklatur einführen:

- Der erste Term  $q(\tilde{\mathbf{n}}^S)$  beschreibt die lokale Dynamik innerhalb des betrachteten Subsystems, weshalb wir ihn im Folgenden auch als *Lokaldynamik* bezeichnen werden.
- Der Term  $\mathbf{T}^{\text{SU}} e^{t\mathbf{T}^{\text{UU}}} \tilde{\mathbf{n}}^U(0)$  beschreibt die Dynamik der Umgebung über die Zeit, sowie die Kopplung der Umgebung an das Subsystem. Die Wirkung dieses Terms wird determiniert von der Konzentration  $\tilde{\mathbf{n}}^U(0)$  zum Zeitpunkt  $t = 0$ . Da der Wert dieses Terms für jeden Zeitpunkt  $t$  unkorreliert mit den initialen Konzentrationen des betrachteten Subsystems ist, wird er als *Rauschterm* bezeichnet.
- Den Term:

$$\int_0^t dt' \mathbf{T}^{\text{SU}} e^{(t-t')\mathbf{T}^{\text{UU}}} \mathbf{T}^{\text{US}} \tilde{\mathbf{n}}^S(t') = \int_0^t dt' \mathbf{M}(t-t') \tilde{\mathbf{n}}^S(t') \quad (4.13)$$

nennen wir *Memory-Term*, wobei  $\mathbf{M}(t-t')$  der *Memory-Kern* ist. Der Memory-Term beschreibt, wie die Umgebung zum jetzigen Zeitpunkt auf das Subsystem wirkt, wenn es zum vergangenen Zeitpunkt  $t'$  eine Reaktion aus dem Subsystem in die Umgebung gab. Im Gegensatz zum Rauschterm lässt sich die Auswirkung des Memory-Terms also berechnen, wenn man die Vergangenheit des Subsystems sowie die Struktur der Umgebung kennt.

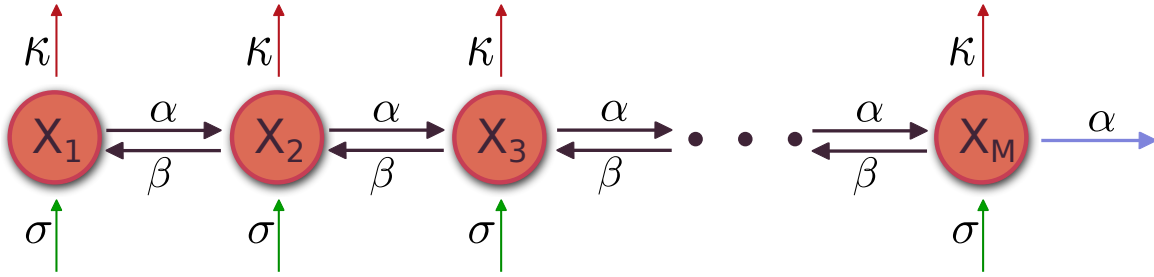
Gemäß Gleichung (4.12) müsste hier noch ein Term mit dem Vektor  $\mathbf{b}^U$  auftauchen. Im nächsten Unterkapitel werden wir aber begründen, warum dieser Term für unsere Umgebungstopologien im Rauschterm absorbiert wird.

---

### 4.3.1 Umgebung mit monomolekularen Reaktionen

---

Typische zelluläre Umgebungen sind meist hochgradig komplex und beinhalten verschiedenste bimolekulare Wechselwirkungen. Im Folgenden werden wir uns allerdings auf Umgebungen beschränken, welche durch monomolekulare Reaktionen beschrieben werden können. Aus Sicht der mathematischen Beschreibung wird durch diese Beschränkung ein einfacherer analytischer Zugang möglich. Gleichzeitig gibt es aber auch zahlreiche Interaktionen, welche gut durch monomolekulare Reaktionen abgebildet werden. Beispiele sind das reversible Binden von Transkriptionsfaktoren oder der konstante Zu- und Abfluss von Substanzen durch die Zellmembran. Weiterhin ist von einigen Proteinen bekannt, dass sie für kurze Zeiten ihre Struktur ändern und so eine neue Funktion ausführen können, was *Promiskuität* genannt wird. Die Übergänge zwischen den verschiedenen Strukturen können oft durch monomolekulare Reaktionen genähert werden [91]. Häufig sind Proteine auch an zyklischen Reaktionen beteiligt, bei denen sie zwischen verschiedenen funktionalen und dysfunktionalen Faltungszuständen wechseln können, was sich ebenfalls gut durch monomolekulare Übergänge approximieren lässt [91–93]. Ebenso kann auch RNA-Faltung zu Übergängen zwischen verschiedenen Strukturen führen, von denen nur eine aktiv ist [94]. All diese ungewollten Interaktionen zwischen dem Subsystem und der Umgebung und die damit einhergehenden Memory-Terme sollten in der Modellierung und der Analyse von genetischen Schaltkreisen beachtet werden. Es ist folglich wünschenswert, für



**Abbildung 4.2:** Lineare Kette mit monomolekularen Reaktionen, wie wir sie im Weiteren als generisches Modell für die Umgebung verwenden werden. Die Raten  $\alpha$  und  $\beta$  geben die Übergangsraten zwischen den verschiedenen Spezies  $X_i$  an. Die Raten  $\sigma$  und  $\kappa$  quantifizieren den konstanten Zu- und linearen Abfluss. Der blaue Pfeil steht für die Kopplung an das Subnetzwerk.

einen Satz von prototypischen Umgebungen Information über deren Wirkung zu haben. Wir werden daher im Weiteren analytische Ausdrücke für die Memory-Kerne der in Abbildung 4.2 gezeigten monomolekularen Umgebungen berechnen. Da in der Biologie nahezu alle Reaktionen reversibel sind, werden wir besonderes Augenmerk darauf legen, auch die inversen Reaktionen in die Beschreibung einzubauen. Die Parameter  $\alpha$  und  $\beta$  sind hierbei die Reaktionsraten für die Übergänge zwischen den einzelnen Zuständen,  $\sigma$  und  $\kappa$  sind Zu- und Abflüsse.

Bevor wir beginnen die zugehörigen Memory-Funktionen herzuleiten, gehen wir zunächst darauf ein, welche Auswirkungen ein ausschließlich konstanter Zufluss in die Kette ( $\sigma \neq 0$ ) auf das angekoppelte Subsystem hat. Nach einer anfänglichen transienten Phase, welche wir nicht weiter beachten wollen, befindet sich die Umgebung im stationären Zustand. Um die durch konstanten Zufluss erzeugte Ankopplung an das Subsystem zu berechnen, müssen wir dann folgendes Gleichungssystem lösen:

$$\begin{aligned}
 \tilde{n}_1 &= \frac{\sigma + \tilde{n}_2 \beta}{\alpha + \kappa} \\
 \tilde{n}_i &= \frac{\sigma + \alpha \tilde{n}_{i-1} + \beta \tilde{n}_{i+1}}{\alpha + \beta + \kappa} \\
 \tilde{n}_N &= \frac{\sigma + \alpha \tilde{n}_{N-1}}{\alpha + \beta + \kappa} \\
 r(t) &= r = \alpha \tilde{n}_N.
 \end{aligned} \tag{4.14}$$

Ein konstanter Zufluss in die zelluläre Umgebung lässt sich also durch Hinzufügen einer Konstanten im Rauschterm berücksichtigen und hat keinen Einfluss auf den Memory-Term.

### 4.3.2 Zusammenhang zwischen linearer Kette und Random Walk

Um die Memory-Kerne zu berechnen, werden wir auf das Wissen und wesentliche Ergebnisse der Random-Walk-Theorie zurückgreifen. Insbesondere werden wir nutzen, dass die Memory-Funktion für monomolekulare Systeme gleich der Rückkehrwahrscheinlichkeit eines Random Walkers ist. Zur Veranschaulichung dieser Zusammenhänge betrachten wir erneut als prototypische Umgebung eine monomolekulare Kette aus Abbildung 4.2. Das Subsystem koppelt an beliebiger Stelle  $x$  in die Umgebung und diese koppelt an beliebiger anderer Stelle zurück in das Subsystem. Die Kopplung erzeugt also beispielsweise zum Zeitpunkt  $t'$  ein Teilchen der Substanz  $X_x$ . Da die Kette nach Definition nur aus monomolekularen Reaktionen besteht, bewegt sich dieses Teilchen unabhängig von allen anderen Teilchen auf dieser Kette. Entsprechend der Reaktionsraten  $\alpha$  und  $\beta$  springt es zwischen verschiedenen Speziesarten (bzw. als Random Walk aufgefasst zwischen verschiedenen Position) hin und her, bevor es die Umgebung wieder verlässt.



Aufgrund der Unabhängigkeit der Teilchen gilt für den Mittelwert jeder Umgebungsspezies zum Zeitpunkt  $t$ :

$$\mathbb{E}(n_i^U(t)) = N^U P(n_i^U, t), \quad (4.15)$$

wobei  $N$  die Anzahl der in der Umgebung befindlichen Teilchen darstellt und  $P(n_i, t)$  die Wahrscheinlichkeit dafür angibt, dass ein einzelnes betrachtetes Teilchen sich zum Zeitpunkt  $t$  auf Position  $n_i$  befindet. Da wir in der Umgebung nur monomolekulare Reaktionen zugelassen haben, gilt analog (2.45) somit auch für die deterministische Konzentration:

$$\tilde{n}_i^U(t) = N^U P(n_i^U, t). \quad (4.16)$$

Für monomolekulare Reaktionsnetzwerke wurde weiterhin bereits gezeigt, dass eine Multinomialverteilung<sup>1</sup>  $M(\mathbf{n}, \hat{\mathbf{p}}(0))$  unter der Dynamik des Netzwerks multinomial bleibt [95]. Der Wahrscheinlichkeitsvektor  $\hat{\mathbf{p}}(t)$  evolviert dabei über die Zeit gemäß einer deterministischen Gleichung:

$$\hat{\mathbf{p}}(t) = e^{\mathbf{B}t} \hat{\mathbf{p}}(0), \quad (4.17)$$

wobei  $\mathbf{B}$  die deterministische Übergangsmatrix des Netzwerks darstellt. Zum Zeitpunkt  $t = 0$  ist die gesamte Wahrscheinlichkeit für die Position des Teilchens auf der Einkopplungsstelle konzentriert. Dieser Zustand lässt sich als Multinomialverteilung mit  $\hat{\mathbf{p}}(0) = (1, 0, 0, \dots, 0)$  auffassen. Demnach gilt obige Aussage und die Zeitentwicklung von  $\hat{\mathbf{p}}(t)$  ist durch die deterministische Gleichung (4.17) gegeben.

Ein Teilchen der Umgebung verlässt diese wieder, sobald es zurück in das Subsystem koppelt. Wir fassen diese Rückkopplung nun als absorbierenden Zustand  $n_{\text{abs}}$  der Umgebung auf, welcher nicht verlassen werden kann. Folglich enthält auch der Wahrscheinlichkeitsvektor der Multinomialverteilung eine Komponente  $\hat{p}_{\text{abs}}$ , welche die Wahrscheinlichkeit für den absorbierenden Zustand enthält. Die Wahrscheinlichkeit zum Zeitpunkt  $t$  absorbiert zu sein, ist demnach gegeben durch:

$$P_{\text{abs}}(t) = 1 - M(n_{\text{abs}} = 0, \hat{\mathbf{p}}(t)) = \hat{p}_{\text{abs}}(t), \quad (4.18)$$

womit sich die Wahrscheinlichkeit der Absorption zum Zeitpunkt  $t$  ergibt zu:

$$\frac{dP_{\text{abs}}}{dt} = \frac{d\hat{p}_{\text{abs}}}{dt} = \mathbf{B}^{\text{AN}} \hat{\mathbf{p}}(t). \quad (4.19)$$

Die Matrix  $\mathbf{B}^{\text{AN}}$  enthält dabei die Raten vom Netzwerk in den absorbierenden Zustand.

Mit diesen Ausführungen lässt sich nun der Memory-Term aus Gleichung (4.12) neu interpretieren.

- Der zweite Teil  $\mathbf{T}^{\text{US}} \tilde{\mathbf{n}}^{\text{S}}(\tau) d\tau =: P^{\text{US}}(\tau)$  entspricht der Konzentration einer Spezies, die im Zeitintervall  $d\tau$  aus dem Subsystem kommend in die Umgebung eintritt. Die dadurch entstehende Verteilung kann als Multinomialverteilung mit  $\hat{\mathbf{p}} = \mathbf{e}_i$  aufgefasst werden. Der erste Teil des Memory-Kernels  $\mathbf{T}^{\text{SU}}$  entspricht dabei (analog zu  $\mathbf{B}^{\text{AN}}$  in (4.19)) der Rate aus der Umgebung wieder in das Subsystem absorbiert zu werden.
- Der Term  $e^{\mathbf{T}^{\text{UU}}(t-\tau)}$  entspricht der deterministischen Lösung von Gleichung (4.17), wobei  $\mathbf{T}^{\text{UU}}$  die deterministische Ratenmatrix der isolierten Umgebung ist.

<sup>1</sup> Unter einer Multinomialverteilung versteht man die multivariate Verallgemeinerung der Binomialverteilung. Im Gegensatz zur Binomialverteilung gibt es also nicht nur jeweils 2, sondern  $k$  mögliche Versuchsausgänge. Die  $k$  Komponenten des Wahrscheinlichkeitsvektors  $\hat{\mathbf{p}}$  geben dabei für jeden der möglichen Ausgänge die Wahrscheinlichkeit an. Folglich gilt  $\|\hat{\mathbf{p}}\|_1 = 1$ .

- Demnach gibt  $\mathbf{T}^{\text{SU}} e^{\mathbf{T}^{\text{UU}}(t-\tau)} =: P^{\text{SU}}(t-\tau|\tau)$  die Wahrscheinlichkeit dafür an, die Umgebung nach einer Zeit  $t-\tau$  zu verlassen, wenn sie zum Zeitpunkt  $\tau$  betreten wurde.

Wie in Bild 4.2 bereits angedeutet, beinhaltet das Modell unserer Umgebung neben Zuflüssen auch Terme für einen Zerfall von Teilchen aus der Umgebung. Ein solches zerfallenes Teilchen kann nicht mehr in das Subsystem zurück koppeln. Da es keine Interaktion zwischen einzelnen Teilchen der Umgebung gibt, ergibt sich sofort, wie eine lineare Zerfallsrate die Wahrscheinlichkeit für eine Rückkopplung modifiziert:

$$P^{\text{SU}}(t-\tau|\tau) = \mathbf{T}^{\text{SU}} e^{\bar{\mathbf{T}}^{\text{UU}}(t-\tau)} e^{-\kappa(t-\tau)}, \quad (4.20)$$

wobei  $\bar{\mathbf{T}}^{\text{UU}}$  die Ratenmatrix der Umgebung ohne die Zerfallsreaktionen darstellt.

Wir fassen die verschiedenen Umgebungsspezies (in Abbildung 4.2 mit  $X_i$  bezeichnet) nun als mögliche Positionen eines Random Walkers auf, wodurch jede der linearen Reaktionsraten zu einer Übergangswahrscheinlichkeit zwischen den Zuständen wird. Mithilfe dieser Umdeutung lässt sich  $\mathbf{T}^{\text{SU}} e^{\mathbf{T}^{\text{UU}}(t-\tau)} = P^{\text{SU}}(t-\tau|\tau)$  dann als Absorptionswahrscheinlichkeit eines solchen Random Walks verstehen.

### 4.3.3 Absorptionswahrscheinlichkeit

Wir werden nun einen analytischen Ausdruck für  $P^{\text{SU}}(t-\tau|\tau)$  herleiten, mit dem sich jedes monomolekulare Reaktionsnetzwerk mit absorbierenden Rändern beschreiben lässt. Wir greifen dazu auf Ergebnisse der Random-Walk-Theorie zurück. Um die Notation zu vereinfachen, nummerieren wir die Zustände der Kette mit  $\omega \in (0, \dots, \Omega)$ . Wie sich später zeigen wird, besteht der allgemeinste Ansatz darin eine Kette zu betrachten, welche zwei absorbierende Zustände an beiden Enden besitzt, also bei  $\omega = 0$  und  $\omega = \Omega$ . Um die Wahrscheinlichkeit  $P_{\omega,\chi}$  dafür, dass ein Random Walker, der auf Position  $\omega$  startet, nach einer Zeitdauer  $t-\tau$  den absorbierenden Zustand bei  $\chi$  erreicht, zu beschreiben, nutzen wir die Rückwärts Mastergleichung. Für das gegebene System lautet diese:

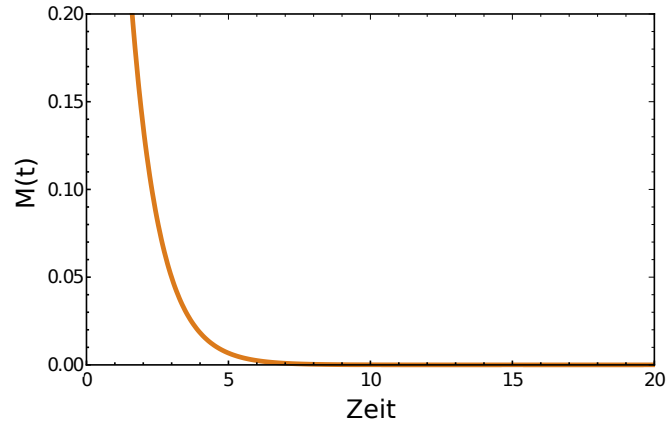
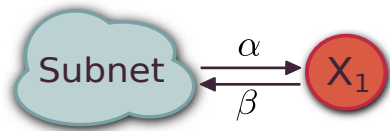
$$\frac{dP_{\omega,\chi}(t)}{dt} = -(\alpha + \beta)P_{\omega,\chi}(t) + \alpha P_{\omega+1,\chi}(t) + \beta P_{\omega-1,\chi}(t), \quad (4.21)$$

wobei  $\alpha$  und  $\beta$  die Übergangswahrscheinlichkeiten innerhalb der Kette für einen Schritt nach links bzw. rechts sind. Heathcote und Moyal haben für diese Gleichung und unter der Annahme von absorbierenden Zuständen bei 0 und  $\Omega$  analytische Lösungen berechnet [96]. Für die Wahrscheinlichkeit, zum Zeitpunkt  $t$  bereits absorbiert zu sein, erhielten sie:

$$\begin{aligned} \mathcal{P}_{\omega,0}(t) &= \nu^{-\omega} \sum_{j=0}^{\infty} \int_0^t \tau^{-1} e^{-\mu\tau} \left( \zeta(2j\Omega + \omega) - \zeta(2(j+1)\Omega - \omega) \right) d\tau, \\ \mathcal{P}_{\omega,\Omega}(t) &= \nu^{\Omega-\omega} \sum_{j=0}^{\infty} \int_0^t \tau^{-1} e^{-\mu\tau} \left( \zeta((2j+1)\Omega - \omega) - \zeta((2j+1)\Omega + \omega) \right) d\tau, \end{aligned} \quad (4.22)$$

wobei  $\zeta(x) = x I_x$ ,  $\mu = \alpha + \beta$ ,  $\nu = \sqrt{\alpha/\beta}$  und  $I_x$  die modifizierte Besselfunktion erster Art<sup>2</sup> mit Argument  $2\tau\sqrt{\alpha\beta}$  darstellt. Da es sich bei dieser Wahrscheinlichkeit um eine kumulative Verteilung handelt, haben wir hier die Bezeichnung  $\mathcal{P}$  eingeführt. Wenn wir uns nun für die

<sup>2</sup> Die modifizierte Besselfunktion erster Art:  $I_n(z)$  ist die Lösung folgender Differentialgleichung:  $z^2 y'' + z y' - (z^2 + n^2) y = 0$ .



**Abbildung 4.3:** Skizze der Topologie (links) und der entsprechenden Form der Memory-Funktion (rechts) des Beispiels mit nur einer einzelnen angekoppelten Spezies aus Abschnitt 4.4.1. Für dieses Beispiel gewählte Parameter:  $\alpha = \beta = 1$ .

Wahrscheinlichkeit dafür interessieren, genau zum Zeitpunkt  $t$  absorbiert zu werden, so müssen wir diese kumulative Verteilung nach der Zeit ableiten und erhalten:

$$P_{\omega,0}(t) = \nu^{-\omega} \sum_{j=0}^{\infty} t^{-1} e^{-\mu t} \left( \zeta(2j\Omega + \omega) - \zeta(2(j+1)\Omega - \omega) \right), \quad (4.23)$$

$$P_{\omega,\Omega}(t) = \nu^{\Omega-\omega} \sum_{j=0}^{\infty} t^{-1} e^{-\mu t} \left( \zeta((2j+1)\Omega - \omega) - \zeta((2j+1)\Omega + \omega) \right).$$

Da die Umgebung als endlich angenommen wird, gilt:

$$\int_0^{\infty} (P_{\omega,0}(t) + P_{\omega,\Omega}(t)) dt = 1. \quad (4.24)$$

In der Umgebung herrscht also Massenerhaltung.

#### 4.4 Mögliche monomolekulare Umgebungen und deren Memory-Kerne

Wir werden nun die Gleichungen (4.23) der Absorptionswahrscheinlichkeit eines Random Walks nutzen, um verschiedene Arten von generischen zellulären Umgebungen zu beschreiben sowie die zugehörigen Memory-Kerne bestimmen.

##### 4.4.1 Eine einzelne Spezies

Um ein Gefühl für die Auswirkungen der Umgebung zu bekommen, beginnen wir mit einer möglichst einfachen Beispielumgebung, nämlich der einer einzelnen Spezies. Wie in Abbildung 4.3 links gezeigt, ist diese mit den Raten  $\alpha$  und  $\beta$  an ein Subsystem angekoppelt. Beispiele für diese einfache Form der Umgebung finden sich z. B. bei der reversiblen Bindung einer Subsystemspezies an eine Bindungsstelle in der Umgebung. Für den Fall dieser einfachen Kopplung reduzieren sich in (4.13) die Matrizen  $\mathbf{T}^{\text{SU}} = -\mathbf{T}^{\text{UU}}$  zu der Rate  $\beta$  und  $\mathbf{T}^{\text{US}}$  zur Rate  $\alpha$ , wodurch sich der Memory-Kern vereinfacht zu:

$$M(\tau) = \beta e^{-\beta\tau} \alpha. \quad (4.25)$$

Dies entspricht (siehe hierzu auch Abbildung 4.3 rechts) der erwarteten Exponentialverteilung für ein Teilchen, welches mit linearer Rate zerfällt.

#### 4.4.2 Kette mit beidseitiger Ankopplung an das Subsystem

Wir befassen uns jetzt mit der allgemeinsten Variante der von uns eingeführten Umgebungs-  
topologien, nämlich einer Kette, welche an beiden Enden an das Subsystem koppelt. In Ab-  
bildung 4.4(a,b) sind zwei solche Beispiele dargestellt. Der Random Walker kann die Kette  
an beiden Enden verlassen. Beispiele für diese Art von Umgebungen lassen sich etwa in pro-  
miskuitären Konformationsänderungen von Proteinen finden. Proteine sind nur in bestimmten  
Faltungszuständen – welche in der Beschreibung des Subsystems vorausgesetzt werden – funk-  
tional. Durch verschiedene äußere Einflüsse können sich die Proteine aber im zeitlichen Verlauf  
in andere Faltungszustände begeben. Diese intermediären Zustände zeigen eine schleifenar-  
tige Reaktions-Topologie und können daher durch die verschiedenen Zustände der von uns  
eingeführten Umgebung repräsentiert werden [91]. Abbildung 4.5 zeigt anhand eines Proteins,  
welches einen Liganden bindet, wie eine solche Konformationsänderung verstanden werden  
kann. Um den Memory-Kern zu erhalten, können wir direkt die Summe der beiden Terme aus  
Gleichung (4.23) benutzen, welche dann eine endliche Kette mit absorbierenden Zuständen an  
beiden Enden beschreibt:

$$M_{\text{biChain}}(t) = P_{\omega,0}(t) + P_{\omega,\Omega}(t). \quad (4.26)$$

Im von uns angeführten Beispiel der Konformationsänderungen von Proteinen wäre die Kopplung  
des Subsystems an die Umgebung am ersten Glied der Kette und somit  $\omega = 1$ . Im Allgemeinen  
besitzt unsere Herleitung aber keine solche Einschränkung und  $\omega$  kann frei gewählt werden.

Enzymatische Reaktionskreisläufe haben oft eine ausgewiesene Vorzugsrichtung. Ein solcher  
Kreislauflauf ist in Abbildung 4.4(c) dargestellt. In diesem Fall können wir dann nähern:  $\lim_{\alpha \rightarrow 0}$ . Es  
gibt folglich keine Absorption am Eintrittspunkt ( $\omega = \Omega$ ), womit gilt:  $P_{\omega,\Omega}(t) = 0$ . Wir erhalten  
dann für den Memory-Kern:

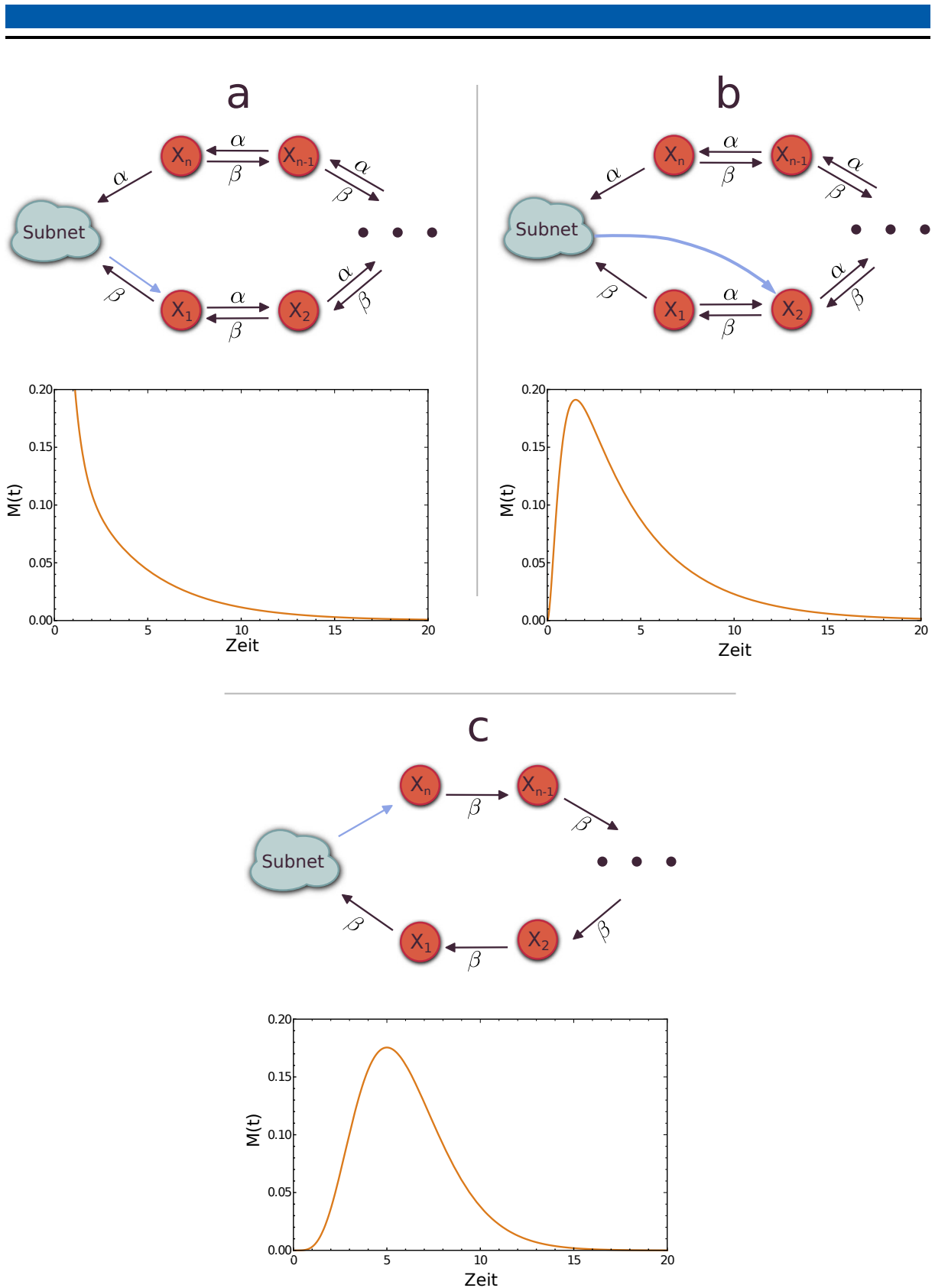
$$M(t) = \lim_{\alpha \rightarrow 0} P_{\omega,0}(t) = \frac{\beta^\omega e^{-t\beta} t^{\omega-1}}{\Gamma(\omega)}. \quad (4.27)$$

Dies ist eine in der Mathematik bekannte Funktion, nämlich die sogenannte Erlang-Verteilung.  
Sie stellt die Verteilung der Summe von  $\omega$  unabhängigen exponentialverteilten Zufallsvariablen  
dar. Für eine Kettenlänge von  $\omega = 1$  geht diese Verteilung wie erwartet über in das Ergebnis von  
Gleichung (4.25).

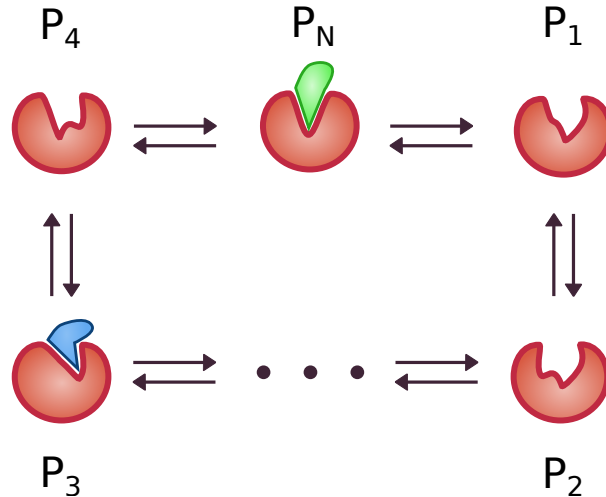
#### 4.4.3 Kette mit einseitiger Ankopplung an das Subsystem

Indem wir in Gleichung (4.26)  $\Omega \rightarrow \infty$  fordern, können wir den Einfluss des zweiten absorbie-  
renden Endes minimieren. Die Wahrscheinlichkeit dort absorbiert zu werden geht folglich gegen  
Null. Wir erhalten also eine Beschreibung für eine Umgebung, welche nur an einer einzelnen  
Stelle an das Subsystem koppelt. Die Wahrscheinlichkeit absorbiert zu werden vereinfacht sich  
dann zu [96]:

$$\begin{aligned} \lim_{\Omega \rightarrow \infty} P_{\omega,0}(t) &= \lim_{\Omega \rightarrow \infty} \nu^{-\omega} \sum_{j=0}^{\infty} t^{-1} e^{-\mu t} \left( \zeta(2j\Omega + \omega) - \zeta(2(j+1)\Omega - \omega) \right) \\ &= \omega \nu^{-\omega} t^{-1} e^{-\mu t} I_\omega. \end{aligned}$$



**Abbildung 4.4:** Exemplarische Skizze der Topologie (oben) und des Verlaufs des entsprechenden Memory-Kerns (unten) von in Abschnitt 4.4.2 beschriebenen zirkulären Umgebungen: (a,b) Bidirektionale lineare Kette mit Kopplung in die Umgebung an (a) der ersten Position (b) einer Zwischenposition (siehe (4.26)); (c) Gerichtete lineare Kette (siehe Gleichung (4.27)); Parameter sind:  $\alpha = \beta = 1$ ,  $\Omega = 6$ . Zusätzlich für (b):  $\omega = 3$ . Die drei Punkte stehen hierbei für eine beliebige lange Fortsetzung der Kette.



**Abbildung 4.5:** Beispielhafte Darstellung für ein Protein (orange), welches verschiedene Konformationen annehmen kann. Nur in der nativen Konformation  $P_N$  kann das Protein den passenden Ligand (grün) binden. In den alternativen Konformationen ist das Protein entweder ohne Funktion, oder zeigt promiskuitäre Eigenschaften (wie etwa  $P_3$ , welches mit einem ungewollten Ligand (blau) interagiert).

Die Abbildung basiert auf Fig.1 aus Science 324, Issue 5924, pp. 203-207 [91].

Reprinted with permission from the authors and AAAS.

Interessant ist hier ein Vergleich mit dem ersten Beispiel, in dem die Umgebung aus nur einer einzelnen Spezies bestanden hat. Da jetzt eine lange Kette von Spezies die Kopplung aus dem Subsystem puffert, fällt der Verlauf des Memory-Kerns im Vergleich langsamer ab (siehe Abbildungen 4.6(a,b)).

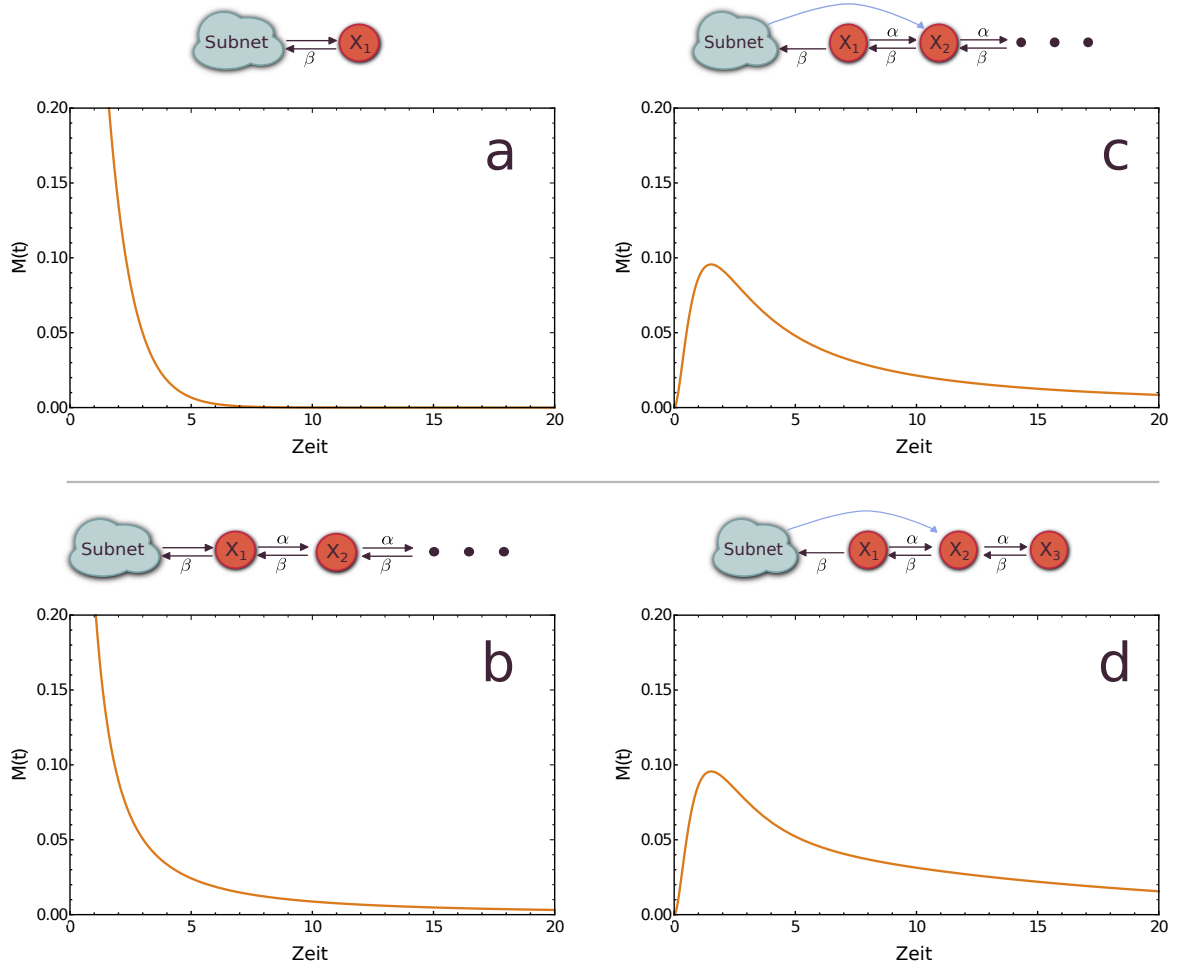
Analog zum beidseitigen Ende ist es auch in diesem Fall möglich, dass das Subsystem, wie in Abbildung 4.6(c) dargestellt, an einer frei wählbaren Position an die Umgebung koppelt.

Während unendliche Ketten, wie wir sie gerade behandelt haben, eine gute Näherung für ausgedehnte Reaktionskaskaden sein können, ist der Fall von endlich vielen Spezies im Allgemeinen relevanter. Daher werden wir im Folgenden noch Ketten endlicher Länge mit einem absorbierenden Zustand bei  $\omega = 0$  und einem reflektierenden Zustande bei  $\omega = \Omega$  betrachten. In Abbildung 4.6(d) ist ein Beispiel für eine solche Umgebung dargestellt.

An jeder Position innerhalb der Kette springt ein Teilchen mit der summierten Propensität  $\alpha + \beta$  entweder nach rechts oder links. An letzter Position (im Abbildung 4.6(d) bei  $X_3$ ) lässt sich allerdings nur mit der Propensität  $\beta$  nach links springen, da die Kette rechts endet. Ohne die Dynamik der Kette zu verändern, können wir eine neue Reaktion für  $X_3$  definieren, welche wieder auf  $X_3$  zurück führt (siehe hierzu Abbildung 4.7(2)). Jetzt ist die summierte Propensität an Stelle  $X_3$  wie bei den anderen Positionen  $\alpha + \beta$ , allerdings ändert nur die Reaktion mit Rate  $\beta$  tatsächlich die Position. Die zusätzliche Reaktion kann man aber auch als den Übergang von einer endlichen Kette auf das Ende einer zweiten endlichen Kette verstehen (Abbildung 4.7(3)). Dadurch können wir die endliche Kette mit einseitiger Ankopplung umdefinieren als eine endliche Kette mit beidseitiger Ankopplung, wie wir sie weiter oben bereits behandelt haben. Die Dynamik eines Teilchens auf einer endlichen Kette mit einseitiger Kopplung ist also äquivalent zu der Bewegung eines Teilchens auf einer Kette mit der Länge  $2\Omega + 1$ , bei der die Absorptionsrate gegeben ist durch:

$$P_{\text{finiteChain}\Omega}(t) = \hat{P}_{\omega,0}(t) + \hat{P}_{\omega,2\Omega+1}(t), \quad (4.28)$$

wobei  $\hat{P}$  die Absorptionswahrscheinlichkeit einer Umgebung der Länge  $2\Omega + 1$  angibt.

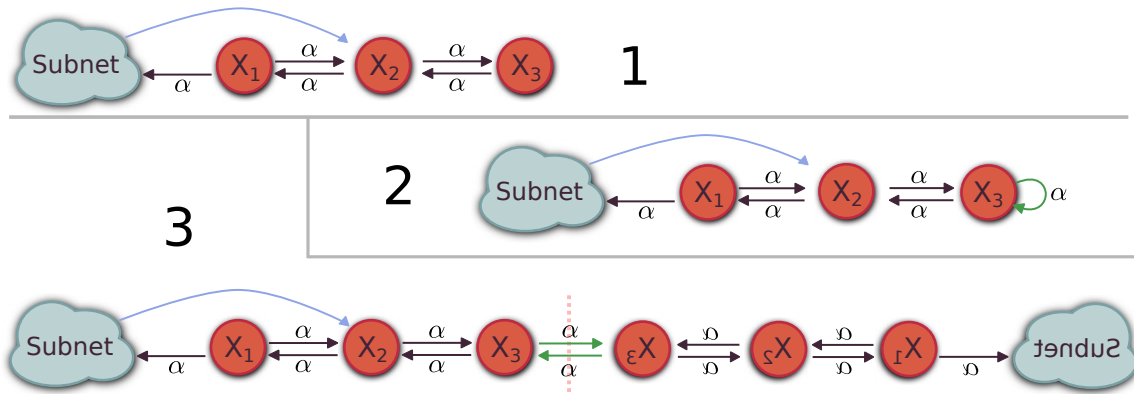


**Abbildung 4.6:** Skizze der Topologie (oben) und des entsprechenden Verlaufs des Memory-Kerns (unten) von linearen Umgebungen mit, wie in Abschnitt 4.4.3 beschrieben, nur einer Kopplung in die Umgebung: (a) Umgebung mit einer einzigen Spezies (siehe (4.25)); (b) Unendliche Kette von miteinander verbundenen Spezies, wobei sich die Kopplung zwischen Subsystem und Umgebung an der gleichen Stelle befindet (siehe (4.28) mit  $\omega = 1$ ); (c) Unendliche Kette von miteinander verbundenen Spezies, wobei sich die Kopplung zwischen Subsystem und Umgebung an verschiedenen Positionen befindet (siehe (4.28) mit  $\omega > 1$ ); (d) Endliche Kette von miteinander verbundenen Spezies (siehe (4.28)); Parameter sind:  $\alpha = \beta = 1$ . Zusätzlich für (c):  $\omega = 3$  und für (d):  $\Omega = 6, \omega = 3$ .

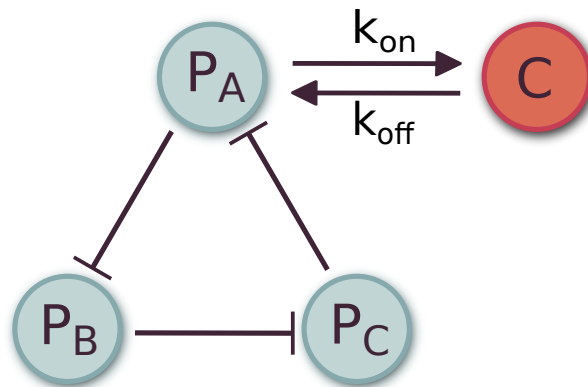
## 4.5 Anwendungen

### 4.5.1 Retroaktivität im Repressilator

Del Vecchio et al. führten Retroaktivität als ein grundlegendes Design-Problem ein, welches immer dann auftritt, wenn verschiedene Module (eigene Subsysteme) hintereinander geschaltet werden, um deren Funktionalitäten zu kaskadieren [5]. Retroaktivität ist hierbei ein Signal, welches sich rückwärts, also von einem später folgenden Modul hin zu einem der vorausgehenden Module, bewegt. Del Vecchio et al. leiteten eine Formel her, um diesen Effekt zu quantifizieren. Eine grundlegende Annahme dieser Herleitung ist, dass sich die Zeitskalen zwischen einem langsamen Modul und dessen schnellem nachfolgenden Modul stark unterscheiden. Unsere analytischen Ergebnisse eröffnen nun Möglichkeiten, die Auswirkungen der Retroaktivität zu



**Abbildung 4.7:** Skizze für die äquivalente Umwandlung einer endlichen Kette in eine zwei-endige Kette. (1) Die anfängliche endliche Kette, die wir beschreiben wollen, wird durch eine sich selbst beziehende Schleifenreaktion (2, grün) erweitert. Diese zusätzliche Reaktion kann als Übergangsreaktion zwischen zwei äquivalenten endlichen Ketten betrachtet werden, die nun eine zwei-endige Netzwerktopologie bilden (3). Die Spiegelachse ist als rosa gestrichelte Linie dargestellt.



**Abbildung 4.8:** Schematische Darstellung der Interaktionen der Proteine  $P_i$  des Repressilators sowie der später eingeführten Kopplung an die Umgebungsspezies C.

bestimmen, ohne auf eine Zeitskalentrennung angewiesen zu sein. Da wir weiter oben bereits Memory-Kerne für verschiedenste generische Umgebungen entworfen haben, können wir jetzt anhand einiger Beispiele zeigen, wie sich Kontexteffekte und damit auch Retroaktivität auf das Verhalten von genetischen Schaltkreisen auswirken können.

Del Vecchio et al. demonstrierten den Effekt der Retroaktivität anhand einer externen oszillierenden Proteinkonzentration  $k(t)$ , welche an eine Promotorregion bindet. Isoliert betrachtet, zeigt das Eingangssignal  $k(t)$  die erwartete und gewünschte Oszillation. Sobald  $k(t)$  aber an den deutlich schnelleren Promotor bindet, wird die Oszillation gedämpft. In Del Vecchios Studie wurde  $k(t)$  als generisches Output-Signal eines vorgeschalteten und perfekt funktionierenden Moduls angenommen. Im Allgemeinen wird aber die komplette Funktion des Oszillators durch die Kopplung an die Promotorregion stark gestört werden. Wir werden daher mithilfe des bekannten und im Folgenden kurz eingeführten Repressilators das oszillierende Modul explizit modellieren. Dann betrachten wir, wie die Oszillation des Repressilators durch Kopplung an die Promotorregion gestört wird und können somit den Einfluss der Retroaktivität untersuchen.



---

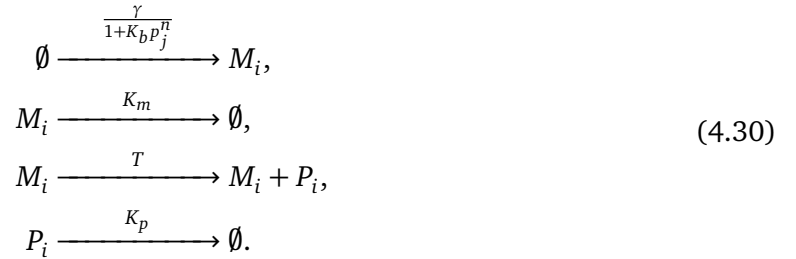
### 4.5.1.1 Der Repressilator

---

Der Repressilator ist ein von Elowitz und Leibler entworfener synthetischer Schaltkreis [97]. Er besteht aus drei Genen und deren zugehörigen Proteinen, wobei jedes Protein, wie in Abbildung 4.8 gezeigt, ein anderes Gen in zyklischer Abhängigkeit reprimiert. Die Wechselwirkungen des Systems können durch folgendes Gleichungssystem erfasst werden:

$$\begin{aligned}\frac{dm_i}{dt} &= -K_m m_i + \frac{\gamma}{1 + K_b p_j^n}, \\ \frac{dp_i}{dt} &= -K_p p_i + T m_i,\end{aligned}\tag{4.29}$$

wobei  $m_i$  die Konzentration der drei verschiedenen mRNA Stränge darstellt,  $p_i$  die der Proteine und  $i \in \{A, B, C\}$ . Die Degradierungsraten von mRNA und Proteinen sind gegeben durch  $K_m$  und  $K_p$ ,  $T$  gibt die Translationsrate an und  $\gamma$ ,  $n$ ,  $K_b$  sind die Parameter der Hill-Funktion (siehe hierzu auch Kapitel 2.3.10). Auf Reaktionsbasis ergibt sich damit:



Wie in Abbildung 4.9 dargestellt, lässt sich mit den von Elowitz und Leibler in [97] angegebenen Parametern sowohl in der deterministischen wie auch in der stochastischen Simulation eine Oszillation der Proteinkonzentrationen beobachten.

---

### 4.5.2 Retroaktivität durch Kopplung an Promotorregion

---

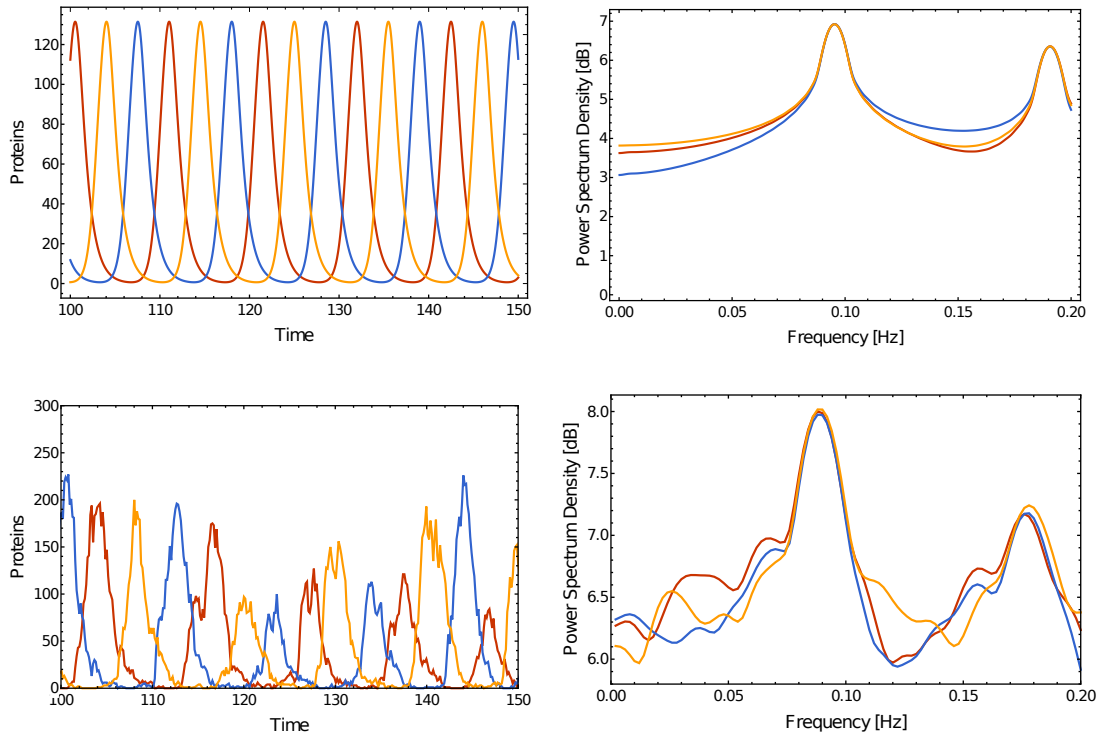
Basierend auf dem gerade eingeführten Repressilator können wir jetzt das von Del Vecchio analysierte System mit explizit modelliertem Eingangssignal nachbilden. Eines der oszillierenden Proteine soll hierbei ein in der Signalkaskade nachfolgendes Modul durch Bindung an dessen Promotor aktivieren. Um dieses Binden (und Entbinden) im Modell zu berücksichtigen, fügen wir für Protein A eine reversible Reaktion zu den Reaktionsgleichungen (4.30) hinzu:



wobei  $C$  hier den Komplex aus Protein und Promotor darstellt (siehe auch Abbildung 4.8). Basierend auf den Ergebnissen aus Kapitel 4.4.1 erhalten wir dadurch für das veränderte Gleichungssystem:

$$\dot{m}_i = -K_m m_i + \frac{\gamma}{1 + K_b p_j^n},\tag{4.32}$$

$$\begin{aligned}\dot{p}_i &= -K_p p_i + T m_i \\ &+ \delta_{Ai} \left( \int_0^t k_{\text{off}} e^{-k_{\text{off}}(t-\tau)} k_{\text{on}} p_i(\tau) d\tau - p_i k_{\text{on}} \right).\end{aligned}\tag{4.33}$$



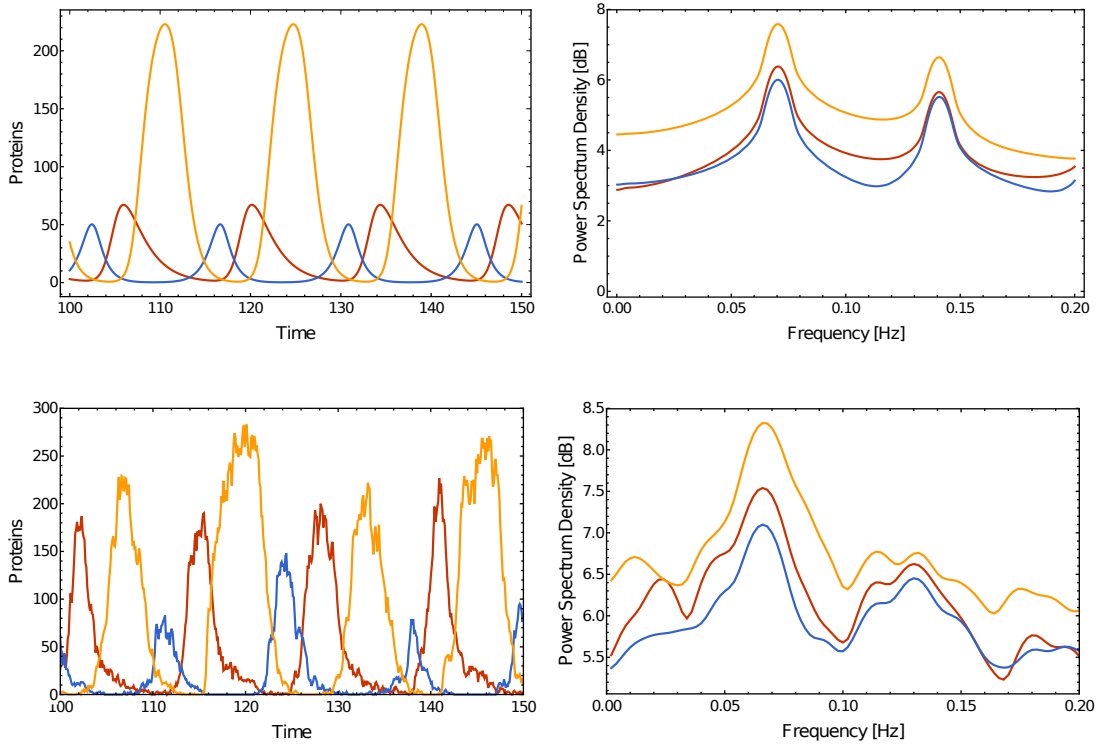
**Abbildung 4.9:** Vergleich der deterministischen Lösung (oben) und der stochastischen Simulation (unten) des unveränderten Repressilatorsystems (4.29), wobei rot, blau und gelb jeweils für  $P_A$ ,  $P_B$  und  $P_C$  stehen. Parameter aus Elowitz und Leibler [97]:  $K_p = T = 5$ ,  $K_m = 1$ ,  $n = 2.1$ ,  $K_b = 1$ ,  $\gamma = 250$ .

Das Kronecker-Delta  $\delta_{Ai}$  hat hierbei nur für  $i = A$  den Wert Eins und ist ansonsten Null. Ausgehend von der isolierten Betrachtung des Repressilators erwartet man, dass die Frequenz, mit der die Promotorregion an- und ausgeschaltet wird, vergleichbar mit der aus Abbildung 4.9 ist. Durch Lösen von (4.32) wie auch stochastische Simulation zeigt sich allerdings, dass dem mitnichten so ist (siehe hierzu Abbildung 4.10). Die durch reversible Bindung verursachte Kopplung des Repressilators an den Promotor sorgt zum einen für eine starke Reduzierung der Frequenz, zum anderen werden die Amplituden der drei Proteinkonzentrationen verschieden beeinflusst. Außer in den Zeitreihen lässt sich dies besonders gut in den spektralen Leistungsdichten erkennen, welche ebenfalls in Abbildung 4.9 und 4.10 gezeigt sind. Im unveränderten System sind die Spektren der drei verschiedenen Proteine sehr ähnlich. Durch Kopplung an die Umgebung werden die Spektren zu kleineren Frequenzen verschoben und insbesondere die Oszillationen von  $P_B$  deutlich gedämpft.

Dieses einfache Beispiel gibt uns bereits eine gute Intuition dafür, wie wichtig die Betrachtung von Umgebungseffekten beim Design von synthetischen Schaltkreisen ist.

#### 4.5.3 Kontexteffekte durch Promiskuität

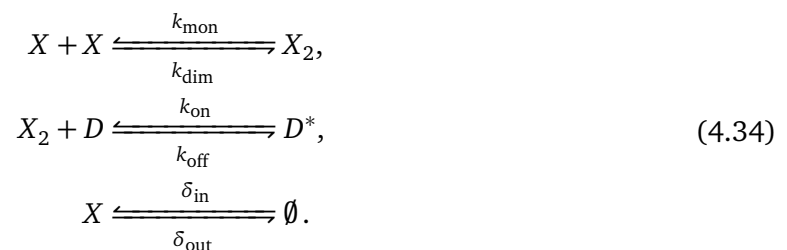
Wir wollen nun ein etwas komplexeres System betrachten, in welchem der Einbezug der Umgebung zu einem unerwarteten Effekt führt. Unser Modell baut hierbei auf einer Arbeit von Bratsun et al. auf [98]. Basierend auf der Beobachtung, dass in Genregulationsnetzwerken oftmals die Zeitskalen zwischen verschiedenen Reaktionstypen sehr unterschiedlich sind, entwickeln die Autoren in dieser Arbeit verschiedene synthetische Module, die durch Zeitverzögerung induzierte Oszillationen zeigen. Im Folgenden werden wir deren Beispiel einer negativen Feedback-Schleife mit Dimerisation nutzen und durch einen konstanten Zufluss von  $X$  erweitern. Folgend zeigen

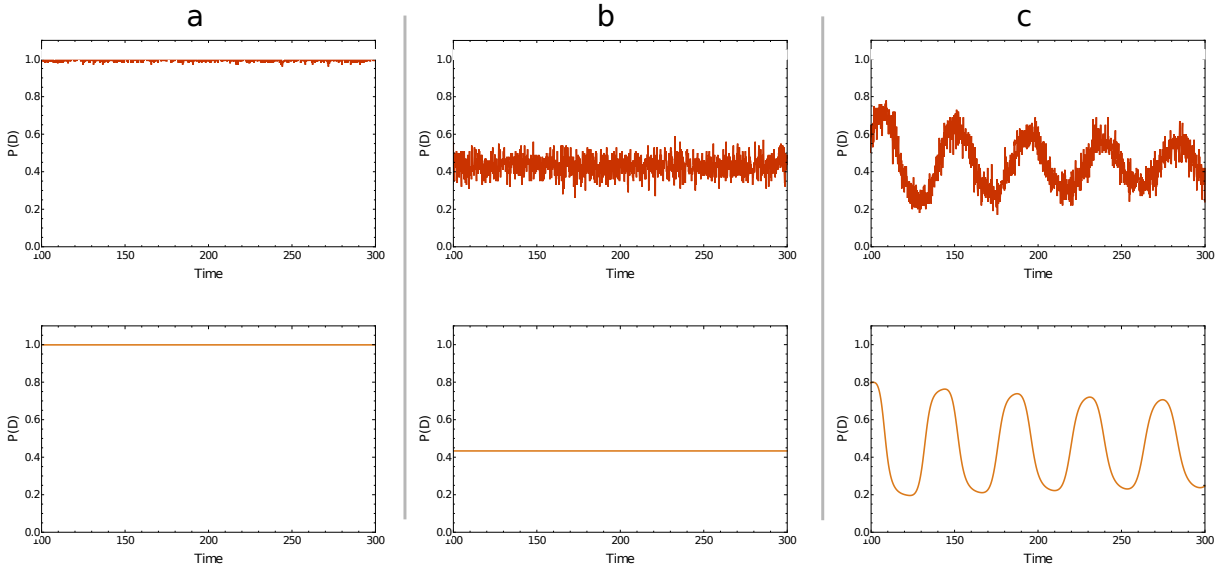


**Abbildung 4.10:** Vergleich der deterministischen Lösung (oben) und der stochastischen Simulation (unten) des Repressilatorsystems mit einem Protein, das an die Umgebung gekoppelt ist wie in (4.32) beschrieben, wobei rot, blau und gelb jeweils für  $P_A$ ,  $P_B$  und  $P_C$  stehen. Parameter aus Elowitz und Leibler [97]:  $K_p = T = 5$ ,  $K_m = 1$ ,  $n = 2.1$ ,  $K_b = 1$ ,  $\gamma = 250$ . Zusätzlich aus [5]:  $k_{\text{on}} = 1000$ ,  $k_{\text{off}} = 100$ .

wir, wie die Inklusion von nachfolgenden Modulen eine Zeitverzögerung auslösen und damit Oszillationen erzeugen kann.

Eine häufige Funktion synthetischer Schaltkreise ist die Aktivierung oder Deaktivierung nachfolgender Gene. Dies wird oft in zwei Schritten umgesetzt, wobei im ersten Schritt zwei Monomere zu einem Dimer reagieren müssen, bevor sie an eine Promotorregion binden können. Im Folgenden benennen wir die Monomere mit  $X$ , die Dimere mit  $X_2$  und die aktive bzw. inaktive Promotorregion mit  $D$  bzw.  $D^*$ . Zur Beschreibung einer Repression des Promotors  $D$  durch  $X_2$  lassen sich dann folgende Gleichungen angeben:





**Abbildung 4.11:** Darstellung der Wahrscheinlichkeit für einen aktiven Promotor im System aus Kapitel 4.5.3. Vergleich der deterministischen Lösung (unten) und stochastischen Simulation gemittelt über 100 Läufe (oben). (a) Das unmodifizierte Modul ohne Feedback, (b) mit direktem Feedback, (c) mit zeitverzögertem Feedback. Parameter:  $\delta_{\text{in}} = 2$ ,  $\delta_{\text{out}} = 4$ ,  $k_{\text{on}} = 1000$ ,  $k_{\text{off}} = 100$ ,  $k_{\text{mon}} = 1000$ ,  $k_{\text{dim}} = 200$ ,  $\gamma = 70$ ,  $\tau = 20$ .

Die für die weitere Betrachtung interessante Größe ist dabei die Wahrscheinlichkeit für einen aktivierten/deaktivierten Promotor. Das zugehörige System von Differentialgleichungen lautet:

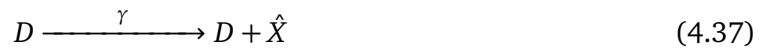
$$\begin{aligned}
 \dot{x}(t) &= \delta_{\text{in}} - \delta_{\text{out}} x(t) - 2k_{\text{dim}} x(t)^2 + 2k_{\text{mon}} x_2(t), \\
 \dot{x}_2(t) &= -k_{\text{off}} d(t) x_2(t) + k_{\text{on}} d^*(t) + k_{\text{dim}} x(t)^2 - k_{\text{mon}} x_2(t), \\
 \dot{d}(t) &= -k_{\text{off}} d(t) x_2(t) + k_{\text{on}} d^*(t), \\
 \dot{d}^*(t) &= k_{\text{off}} d(t) x_2(t) - k_{\text{on}} d^*(t).
 \end{aligned} \tag{4.35}$$

Im stationären Zustand erhält man für dieses System eine konstante Wahrscheinlichkeit für einen aktiven Promotor, welche von den gewählten Parametern der Übergangsraten abhängt:

$$P(d) = \frac{d}{d^* + d} = \frac{\delta_{\text{out}}^2 k_{\text{on}} k_{\text{mon}}}{\delta_{\text{out}}^2 k_{\text{on}} k_{\text{mon}} + \delta_{\text{in}}^2 k_{\text{dim}} k_{\text{off}}}. \tag{4.36}$$

Die Ergebnisse der stochastischen Simulation sowie die numerische Lösung des Gleichungssystems sind in Abbildung 4.11(a) gezeigt.

In einem kaskadierten Genregulationsnetzwerk wird der Promotor nun meist genutzt, um folgende Module zu schalten. Wir nehmen nun an, dass im weiteren Verlauf der folgenden Module ein Protein  $\hat{X}$  erzeugt wird, welches ähnliche Eigenschaften wie Protein  $X$  besitzt. Unter der Annahme, dass die Expression von  $\hat{X}$  keinen großen zeitlichen Versatz zur Aktivierung des Promotors  $D$  hat, könnte man dann das Reaktionssystem durch folgende Reaktion erweitern:

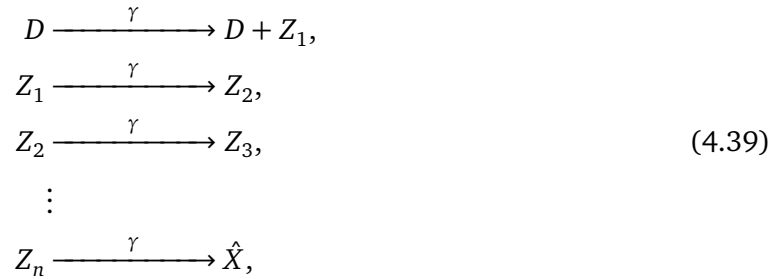


und folglich auch die betreffende Differentialgleichung für  $x$  verändern zu:

$$\dot{x}(t) = \delta_{\text{in}} - \delta_{\text{out}} x(t) - 2k_{\text{dim}} x(t)^2 + 2k_{\text{mon}} x_2(t) + \gamma d(t). \quad (4.38)$$

Da  $\hat{X}$  ähnlich zu  $X$  sein soll, haben wir hier  $x$  statt  $\hat{x}$  verwendet. Das resultierende Verhalten des neuen Netzwerks ist in Abbildung 4.11(b) dargestellt. Es zeigt sich, dass das Feedback des neu eingeführten Proteins  $\hat{X}$  dafür sorgt, dass sich die Wahrscheinlichkeit für einen aktiven Promotor verringert.

Wenn wir nun allerdings zusätzlich annehmen, dass die Expression von  $\hat{X}$  zeitlich verzögert zur Aktivierung des Promotors erfolgt, verändert sich das Resultat erheblich. Eine solche Zeitverzögerung wird beispielsweise auftreten, wenn die Anzahl von Modulen, welche zwischen der Dimerisation von  $X$  und der Expression von  $\hat{X}$  liegen, groß ist oder aber, wenn sehr langsame dazwischenliegende Prozesse ablaufen. Wir erfassen dieses Verhalten, indem wir als Reaktionssystem schreiben:



wobei die  $Z_i$  die Zwischenschritte hin zur Expression von  $\hat{X}$  darstellen. Die durch die  $Z_i$  erzeugte Kette ist dabei analog der gerichteten Kette, welche wir in 4.4.3 bereits untersucht haben. Von dort wissen wir, dass  $M(t - \tau)$  durch eine Erlang-Verteilung (siehe (4.27)) gegeben ist. Folglich könnten wir den hierdurch erhalten Memory-Kern in die Gleichung für  $\dot{x}$  einsetzen. Zuvor werden wir allerdings noch eine Vereinfachung treffen: Wenn wir in Gleichung (4.27) den Mittelwert  $\frac{\omega}{\alpha}$  konstant halten und die Kette vergrößern, verkleinert sich die Varianz mit  $\frac{1}{\omega}$ , wodurch folgt:

$$\lim_{\omega \rightarrow \infty} M(t - \tau) \Big|_{\frac{\omega}{\alpha} = \text{const}} = \lim_{\omega \rightarrow \infty} \frac{\beta^\omega e^{-(t-\tau)\beta} (t-\tau)^{\omega-1}}{\Gamma(\omega)} \Big|_{\frac{\omega}{\alpha} = \text{const}} = \delta\left(t - \tau - \frac{\omega}{\alpha}\right). \quad (4.40)$$

Als Näherung für eine lange Kette können wir daher den Memory-Kern der Umgebung in Form einer zeitverzögerten Delta-Distribution angeben. Diesen Memory-Kern können wir jetzt in die vorhandenen Gleichungen einsetzen und erhalten eine Näherung für das Verhalten des Subsystems bei korrekter Beachtung der Umgebung:

$$\begin{aligned} \dot{x}(t) &= \delta_{\text{in}} - \delta_{\text{out}} x(t) - 2k_{\text{dim}} x(t)^2 + 2k_{\text{mon}} x_2(t) + \int_0^t M(t - \tau) d(\tau) d\tau \\ &= \delta_{\text{in}} - \delta_{\text{out}} x(t) - 2k_{\text{dim}} x(t)^2 + 2k_{\text{mon}} x_2(t) + \gamma \delta(t - \tau), \end{aligned} \quad (4.41)$$

wobei wir  $M(t - \tau) = \delta(t - \tau)$  aus Gleichung (4.40) verwendet haben.

In Abbildung 4.11(c) ist die stochastische Simulation von (4.39) sowie die approximierte deterministische Lösung dieses Netzwerks dargestellt. Deutlich zu erkennen ist, dass die Wahrscheinlichkeit für einen aktivierten Promotor zu oszillieren beginnt. Ein Effekt, der ausschließlich durch die Berücksichtigung der zeitverzögerten Kopplung von  $\hat{X}$  entsteht.

---

## 4.6 Diskussion

---

In diesem Kapitel haben wir uns ausführlich mit den Kontexteffekten von genetischen Schaltkreisen beschäftigt, welche in eine zelluläre Umgebung eingebracht werden. Vorausgegangene Arbeiten setzen oft eingehendes Wissen über die Umgebung voraus. Um besser den Anforderungen der Synthetischen Biologie gerecht zu werden, haben wir verschiedene prototypische Umgebungstopologien analysiert und deren Memory-Kerne berechnet. Im Speziellen haben wir uns hierbei auf Umgebungen bestehend aus Ketten monomolekularer Reaktionen mit verschiedener Ankopplung an das Subnetzwerk beschränkt. Solche Reaktionsketten werden in verschiedensten Bereichen und Ausführungen in einer Zelle beobachtet. Indem wir einen Zusammenhang zwischen den Reaktionsketten und einem Random Walk herstellen konnten, ließen sich analytische Ausdrücke für die korrespondierenden Memory-Kerne finden.

Am Beispiel des Repressilators haben wir unsere Ergebnisse angewendet und veranschaulicht, wie die Beiträge der Umgebung die Funktion dieses Systems beeinträchtigen können. Durch die Kopplung einer Proteinart an einen Umgebungspromotor beobachteten wir dabei eine Veränderung der Oszillationsfrequenz. Zusätzlich haben wir mit einer einfachen negativen Feedbackschleife gezeigt, wie Memory-Effekte zu nicht intuitiven Phänomenen wie Oszillationen führen können, welche ohne Einbezug der Umgebung nicht erkannt werden.

Vor allem im Hinblick auf die immer komplexeren synthetischen Regulationsnetzwerke wird die robuste und verlässliche Funktion der einzelnen genetischen Module zusehends wichtiger werden. Die zahlreichen Vorarbeiten zur Analyse von verschiedenen Umgebungen konnten wir dabei durch die Berechnung von Memory-Termen um analytische Möglichkeiten erweitern.

Es steht außer Frage, dass eine Reduktion der Umgebung auf lineare Reaktionsketten in vielen Fällen eine Übereinfachung darstellt. Vor allem in der Nähe eines stationären Zustands oder in den Fällen, in denen einfache reversible Bindungen die vordergründigen Effekte ausmachen, dient es aber als gute Näherung. Zusätzlich wurde vor kurzem gezeigt, dass sich mit ausschließlich monomolekularen Reaktionsnetzwerken das Verhalten der meisten Umgebungen gut nähern lässt [99]. Unsere Ergebnisse bieten daher einen guten Ausgangspunkt für weitere Untersuchungen generischer Umgebungen im Kontext der Synthetischen Biologie.

---

## 5 Ein minimales Modell für durch Burstauschen induzierte Bimodalität

In diesem Kapitel werden wir uns näher mit einem Phänomen beschäftigen, welches deutlich mögliche Unterschiede zwischen stochastischer und deterministischer Modellierung aufzeigt. Hierzu untersuchen wir den Einfluss von Eigenrauschen auf die stabilen Zustände eines eindimensionalen Systems. Das betrachtete System, das sogenannte Schlögl-Modell, weist hierbei in seiner deterministischen Version eine Sattel-Knoten-Bifurkation zwischen monostabilem und bistabilem Verhalten auf. Indem wir sogenannte Burst-Reaktion einführen, variieren wir die Stärke des intrinsischen Rauschens, ohne dabei die deterministische Beschreibung zu verändern. Dabei werden wir in der probabilistischen Beschreibung Übergänge zwischen monostabilem und bistabilem Verhalten beobachten. Als Ergebnis wird sich dabei zeigen, dass Bistabilität durch Änderung der Burst-Größe zerstört, aber auch erzeugt werden kann.

Die Ergebnisse dieses Kapitels gehen aus einer Zusammenarbeit mit Marc Mendler und Barbara Drossel hervor [70]. Die Idee, mittels Gleichung (5.25) die Maxima der FPE zu berechnen, war dabei zentral von Barbara Drossel motiviert. Die Analyse der Verschiebung der Extrema mittels Gleichung (5.28) geschah in Zusammenarbeit mit Marc Mendler.

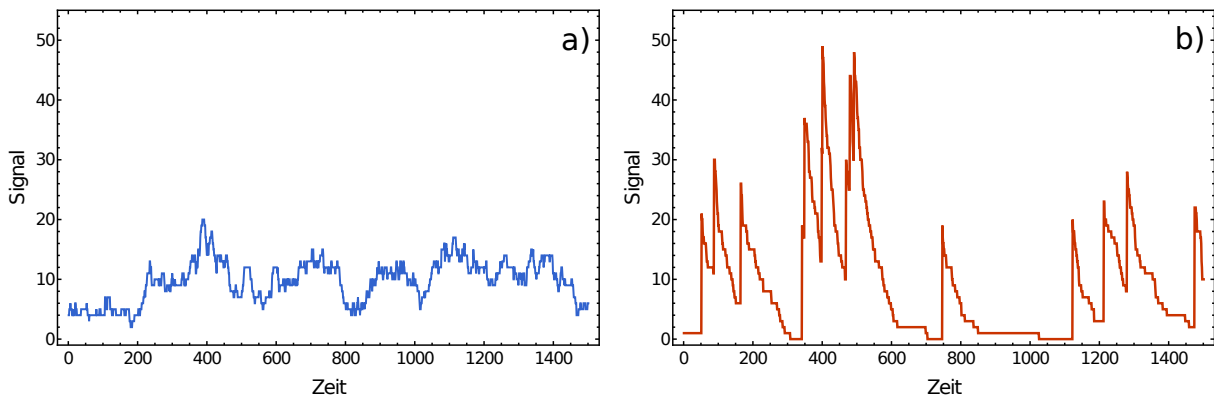
---

### 5.1 Rauschen in biologischen Systemen

---

Wir haben in den Grundlagen beschrieben, durch welche verschiedenen Näherungsformen chemische Reaktionssysteme beschrieben werden können. Als mikroskopische Variante haben wir hierbei die CME kennengelernt, welche die Diskretheit der chemischen Reaktionsgleichungen komplett widerspiegelt. Als mesoskopische Näherung haben wir dann die FPE erhalten und sind letztlich bei der makroskopischen RRE angelangt, welche eine deterministische Trajektorie erzeugt. Aus der Herleitung der Ratengleichung ging dabei bereits hervor, dass deren Beschreibung umso genauer wird, je unwichtiger der Einfluss des stochastischen Rauschens im ursprünglichen System ist. Da die Ratengleichungen sich meist leichter lösen lassen als die zugehörige FPE oder gar ME, werden diese häufig als erste und oft einzige Methode zur Beschreibung eines Systems herangezogen. Die Analyse der dabei entstehenden dynamischen Systeme, wie etwa die Untersuchung von stabilen Zuständen, ist folglich gut etabliert, weshalb hierfür zahlreiche analytische aber auch numerische Methoden existieren [100].

In Bezug auf mögliche ungewollte Einwirkungen galten die von außen auf das System wirkenden externen Fluktuationen lange als die wichtigste Quelle [101, 102]. Theoretische Berechnungen und Simulationen zeigen jedoch, dass für viele zellbiologische Systeme ein signifikanter Unterschied besteht zwischen dem vereinfachten deterministischen Verhalten und dem, das unter Verwendung der vollständigen Informationen über die stochastische Natur der mikroskopischen Dynamik des Systems erhalten wird. Dies gilt insbesondere, wenn die Anzahl von Molekülen oder Individuen klein ist. Der Ursprung der stochastischen Fluktuationen liegt hierbei in der Diskretheit der Reaktionen, so dass diese Art von Rauschen rein intrinsisch ist. Wie bereits in Kapitel 3 bemerkt, spielt gerade auf Einzelzellebene das stochastische Rauschen eine entscheidende Rolle in der Dynamik. Bis vor kurzem war es jedoch aufgrund fehlender geeigneter Messtechniken in der Mikro- und Systembiologie nicht möglich, die prognostizierten stochastischen Effekte in biologischen Zellen zu beobachten. Erst die Entwicklung neuer Verfahren ermöglichte es in den letzten Jahren, Wachstum und Stoffwechsel auf Einzelzellebene zu untersuchen. Beispiele



**Abbildung 5.1:** Vergleich zwischen (a) Herkömmlichem intrinsischem Rauschen mit einem stöchiometrischen Koeffizienten von  $\pm 1$  und (b) Burstrauschen mit einem stöchiometrischen Koeffizienten von  $+20$ . Beide Signalverläufe haben den gleichen Mittelwert.

sind Fluoreszenztechniken in Kombination mit z. B. Durchflusszytometrie oder mikrofluidischer Zellkulturanalyse [103, 104].

Die durch intrinsisches Rauschen hervorgerufenen Effekte sind hierbei verschiedenster Natur [105]. Während bislang zum Beispiel immer davon ausgegangen wurde, dass Rauschen ausschließlich die Breite der Verteilungen vergrößert und die bereits vorhandenen Fluktuationen verstärkt, haben neuere Ergebnisse gezeigt, dass gerade das intrinsische Rauschen in biologischen Systemen zu deren Stabilität beitragen kann und damit deren Leistungsfähigkeit verbessert [12, 106, 107].

Eine wichtige Form des intrinsischen Rauschens ist das sogenannte *Burstrauschen*, welches durch schubartige Sprünge gekennzeichnet ist. Während herkömmliche Reaktionsgleichungen lediglich stöchiometrische Koeffizienten von  $\pm 2$  aufweisen, entstehen beim Burstrauschen durchaus Werte von  $\gg \pm 2$ . Ein Vergleich der möglichen Signalverläufe von herkömmlichem intrinsischem Rauschen und Burstrauschen ist in Abbildung 5.1 dargestellt. Beispiele für Burstrauschen finden sich vor allem in der Zellbiologie, da Transkription und Translation oftmals in Schüben erfolgen [108, 109]. Grund hierfür ist zum Beispiel, dass die mRNA nach ihrer Entstehung schnell und durch viele Ribosome gleichzeitig abgelesen wird und damit Proteine synthetisiert werden. Im Vergleich zu der Synthetisierungsgeschwindigkeit ist die Lebenszeit der mRNA allerdings kurz. In der kurzen Zeit, in der die mRNA existiert, steigt die Konzentration der Proteine daher stark an und nach der Degradation der mRNA sinkt die Konzentration über längeren Zeitraum langsam wieder ab. Verschiedene Studien haben hierbei gezeigt, dass Bursts den grundsätzlich vorhandenen Einfluss von Rauschen noch verstärken und damit das Verhalten der betroffenen Systeme stark verändern können [110, 111].

Wie wir bereits in Kapitel 2.3 gelernt haben, lässt sich die Dynamik von Reaktionsnetzwerken durch die ME korrekt beschreiben und hierbei folglich auch das durch die möglichen Reaktionen ausgelöste intrinsische Rauschen behandeln. Da die ME allerdings oft nur schwer zu lösen ist, wurden verschiedene Methoden entwickelt, um die Auswirkungen des intrinsischen Rauschens auf einfachere Art analytisch zu erfassen. Eine wichtige Vertreterin dieser Methoden ist die Systemgrößenentwicklung von van Kampen [32], welche sich darauf stützt, dass für große Teilchenzahlen die Fluktuationen um den Mittelwert klein werden. Die ersten zwei Ordnungen der van Kampen-Entwicklung resultieren in einer auf dem deterministischen Mittelwert der betrachteten Größe zentrierten Gaußverteilung, deren Breite sich abhängig vom intrinsischen Rauschen anpasst. Durch Hinzunahme einer weiteren Ordnung erhält man dann Korrekturterme, welche die Gaußverteilung leicht verschieben [112]. Um die Stabilität von stabilen Zuständen verrauschter Systeme zu analysieren, nutzt M. Scott die Zeitskalen des Zerfalls der Autokorrela-



---

tionsfunktion [113]. Ein wichtiges Ergebnis dieser Arbeit ist, dass Rauschen – und hierbei vor allem Burstrauschen – stabile Punkte destabilisieren kann.

Während diese und ähnliche Methoden vor allem dazu dienen, stabile Punkte eines stochastischen System zu beschreiben, welche sich in der Nähe eines deterministischen Fixpunktes befinden, werden andere Techniken benötigt, um Bifurkationen und damit evtl. auftretende stabile Punkte zu beschreiben, welche keinen deterministischen Partner haben. Beispiele für Systeme, in denen genau dies passiert, gibt es zahlreich. So nutzen etwa Friedman et al. [109] eine kontinuierliche Form der ME, um zu zeigen, dass Rauschen zu bistabilem Verhalten in einem Genexpressionsnetzwerk führen kann, obwohl die deterministische Beschreibung stets Monostabilität vorhersagt. Ähnliche Ergebnisse wurden auch von zahlreichen anderen Autoren in verschiedenen komplexen mehrdimensionalen Systemen beobachtet [69, 71, 114–117].

In diesem Kapitel werden wir uns ebenfalls mit Bistabilität in stochastischen Netzwerken beschäftigen, dabei das System aber möglichst einfach wählen. Unser System hat nur eine Spezies und unterliegt dem Massenwirkungsgesetz (siehe Kapitel 2.3.10). Das System ist hierbei eine Version des *Schlögl-Modells* [118], welches ursprünglich als minimales Modell zur Untersuchung von Phasenübergängen eingeführt wurde [119–122]. Aufgrund seiner Einfachheit wird das Schlögl-Modell aber mittlerweile als generisches Template für die große Klasse von eindimensionalen bistabilen Systemen verwendet [123]. Die Einfachheit des Modells wird hierbei erkauft durch eine trimolekulare Reaktion, welche eigentlich als unphysikalisch gilt. Aus diesem Grund wurde das System lange Zeit auch nur als theoretisches Konzept gesehen. Vor kurzem hat Endres [124] allerdings ein biologisches System auf das Schlögl-Modell abbilden können.

Während das Schlögl-Modell bisher hauptsächlich in der deterministischen Beschreibung benutzt wurde, in der es eine sogenannte Sattel-Knoten-Bifurkation zeigt, werden wir uns mit der stochastischen Beschreibung beschäftigen. Hierbei werden wir die Extrema der stationären Lösung der Fokker-Planck-Gleichung benutzen, um die stabilen Zustände des Systems zu bestimmen. Auf diese Weise zeigen wir, wie intrinsisches Rauschen Bimodalität induzieren kann.

Immer wieder wird in diesem Kapitel von Sattel-Knoten-Bifurkationen die Rede sein. Aus diesem Grund beginnen wir den folgenden Abschnitt mit einer kurzen Erklärung zu diesen Bifurkation. Dann führen wir das von uns verwendete Schlögl-Modell in der ursprünglichen Form ein und analysieren die wichtigsten Eigenschaften. Alsdann werden wir eine Reaktion des Modells so modifizieren, dass Burstrauschen entsteht und zeigen, wie dieses Rauschen Bimodalität induzieren und zerstören kann. Im letzten Kapitel diskutieren wir unsere Ergebnisse.

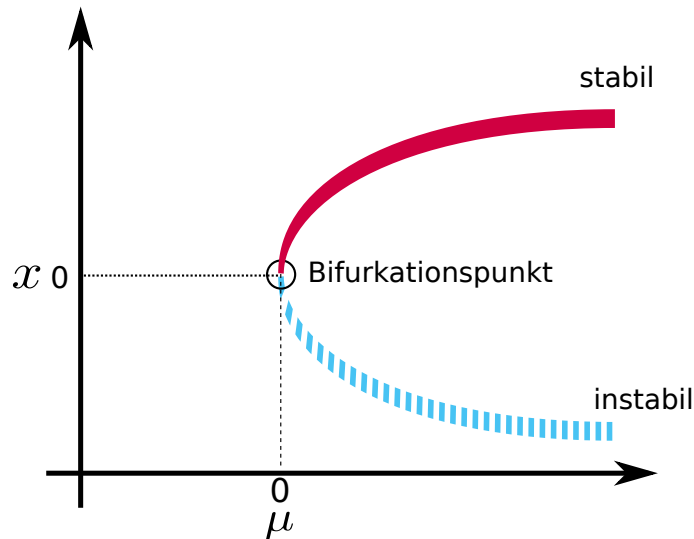
---

## 5.2 Hintergrundwissen: Bifurkationen

---

Bifurkationen beschreiben eine Änderung in der Anzahl oder der Stabilität von stationären Zuständen eines Systems. Die zur Erzeugung der Bifurkation veränderten Parameter heißen dabei *Bifurkationsparameter*. Ein genetischer Schalter könnte zum Beispiel zwei stabile Zustände, nämlich AN und AUS, haben. Durch Variation bestimmter Parameter (zum Beispiel Reaktionsraten) ist es aber evtl. möglich, dass einer der beiden Zustände nicht mehr stabil ist, dass also der Schalter beispielsweise immer in den AUS Zustand läuft. Genau diesen Übergang von zwei stabilen Zuständen auf einen würde man dann als Bifurkation bezeichnen. Der Bifurkationsparameter in diesem Fall wäre dann die Reaktionsrate, welche variiert wurde.

Es existiert eine große Zahl von verschiedenen Bifurkationen, welche basierend auf ihrer *Normalform* klassifiziert werden. Die Normalform ist dabei eine möglichst einfache Gleichung, die für eine bestimmte Bifurkation das Systemverhalten in der Nähe des *Bifurkationspunkts* beschreibt. Im folgenden Kapitel wird ausschließlich die Sattel-Knoten-Bifurkation auftauchen, weshalb wir uns hier auf diese beschränken.



**Abbildung 5.2:** Darstellung der Sattel-Knoten-Bifurkation. Am Bifurkationspunkt ( $\mu = 0, x = 0$ ) treffen sich ein stabiler (rot) und ein instabiler (blau) Fixpunkt.

Die (1D-)Sattel-Knoten-Bifurkation zeichnet sich dadurch aus, dass sich am Bifurkationspunkt ein instabiler (Sattel) und ein stabiler (Knoten) Fixpunkt treffen und beide verschwinden. Die Normalform dieser Bifurkation lautet dabei [125]:

$$\dot{x} = \mu - x^2, \quad (5.1)$$

wobei  $\mu$  der Bifurkationsparameter ist. Sucht man den stationären Zustand des System, in dem  $\dot{x} = 0$  gilt, erhält man die folgenden zwei Fixpunkte:

$$x_{1/2} = \pm\sqrt{\mu}. \quad (5.2)$$

Offensichtlich gibt es für  $\mu < 0$  also keinen Fixpunkt. Der Fall  $\mu = 0$  erzeugt genau einen Fixpunkt  $x = 0$  und stellt den Bifurkationspunkt dar. Für  $\mu > 0$  ergeben sich hingegen zwei Fixpunkte. Der Punkt  $x = \sqrt{\mu}$  ist dabei stabil, denn es gilt  $\dot{x} < 0$  für Werte größer  $x = \sqrt{\mu}$  und  $\dot{x} > 0$  für Werte kleiner  $x = \sqrt{\mu}$ . Das System bewegt sich für kleine Auslenkungen also immer wieder auf den Fixpunkt zurück. Analog gilt für den Punkt  $x = -\sqrt{\mu}$ , dass dieser instabil ist. In Abbildung 5.2 ist die Sattel-Knoten-Bifurkation graphisch dargestellt.

### 5.3 Das Schlögl-Modell

Das Schlögl-Modell ist ein Reaktionsmodell, welches aus zwei reversiblen Reaktionen besteht. Aufgestellt wurde es 1972 von dem deutschen Physiker Friedrich Schlögl. In der ursprünglichen Publikation hat Schlögl zwei Modelle vorgestellt, welche beide für bestimmte Parameterkombinationen einen Phasenübergang zeigen. Während das erste Modell – welches hier nicht weiter diskutiert wird – einen Phasenübergang erster Ordnung durchläuft, entsteht im zweiten Modell ein Phasenübergang zweiter Ordnung. Das Schlögl-Modell hat seit seiner Publikation zahlreiche Verwendungen erfahren und wurde in verschiedensten wissenschaftlichen Disziplinen angewendet. Wir wollen an dieser Stelle nur festhalten, dass das Schlögl-Modell das einfachste Re-

aktionssystem darstellt, welches Bistabilität erzeugen kann [123]. Die definierenden Reaktionen des Schlögl-Modells lauten:



$\mathbb{X}$  ist hierbei eine beliebige Spezies, deren Dynamik durch die vier Reaktionsgleichungen beschrieben wird. Um die Auswirkungen von Burstrauschen zu untersuchen, modifizieren wir dieses Modell allerdings. Dazu ersetzen wir die Reaktionsgleichung (5.5) durch:



Für  $r = 1$  ergibt sich damit das bekannte Schlögl-Modell. Durch Vergrößern von  $r$  lässt sich nun aber die Größe der sogenannten Bursts einstellen. Dabei werden beim einmaligen Ablaufen der Reaktion mehr  $\mathbb{X}$  erzeugt. Gleichzeitig haben wir aber die Rate durch  $r$  geteilt, wodurch die Reaktion seltener abläuft. Die mittlere Produktion von  $\mathbb{X}$  über die Zeit bleibt dadurch gleich.

---

### 5.3.1 Deterministische Beschreibung

---

Wie in den Grundlagen (Kapitel 2.3.9) erklärt, lässt sich anhand der Reaktionsgleichungen eine deterministische Differentialgleichung aufstellen, welche die Dynamik der Konzentration  $\tilde{n}$  von  $\mathbb{X}$  im zeitlichen Verlauf beschreibt. Diese lautet:

$$\dot{\tilde{n}} = k_1 - k_2\tilde{n} + k_3\tilde{n}^2 - k_4\tilde{n}^3, \quad (5.8)$$

was sich mit der stöchiometrischen Matrix  $\mathbf{S} = [1, -1, r, -1]$  und dem Propensitätsvektor  $\tilde{\mathbf{f}} = [k_1, k_2\tilde{n}, k_3/r \tilde{n}^2, k_4\tilde{n}^3]$  kompakter schreiben lässt als:

$$\dot{\tilde{n}} = \mathbf{S}\tilde{\mathbf{f}}. \quad (5.9)$$

Wichtig ist hierbei zu bemerken, dass die deterministische Beschreibung unabhängig von der Größe von  $r$  ist, was in gewisser Weise zu erwarten war, da in der deterministischen Beschreibung bekanntlich über die Fluktuationen gemittelt wird. Demnach ist es gleichwertig, ob wenige  $\mathbb{X}$  oft, oder viele  $\mathbb{X}$  entsprechend selten erzeugt werden, solange der Mittelwert gleich ist.

---

### 5.3.2 Probabilistische Beschreibung

---

Wie in den Grundlagen beschrieben, ist eine Grundannahme der deterministischen Gleichungen, dass das intrinsische Rauschen verschwindet, das System also groß genug gewählt wird. Für

kleinere Systeme ist es notwendig, diese auf Ebene der Teilchenzahl  $n$ , also mittels der ME zu beschreiben. Für obiges Reaktionssystem ergibt sich die CME zu:

$$\frac{\partial P(n, t)}{\partial t} = (\mathcal{E}^{-1} - 1)(a(n-1|n)P(n, t)) + \quad (5.10)$$

$$(\mathcal{E}^{-1} - 1)(a(n+1|n)P(n, t)) + \quad (5.11)$$

$$(\mathcal{E}^{-r} - 1)(a(n+r|n)P(n, t)),$$

wobei wir hier zur Vereinfachung der Notation den Operator  $\mathcal{E}$  eingeführt haben, für den gilt:

$$\mathcal{E}^a f(x) = f(x+a). \quad (5.12)$$

Die Übergangsraten sind dabei für unser Beispiel gegeben durch:

$$a(x+1|x) = k_1, \quad (5.13)$$

$$a(x+r|x) = \frac{k_3}{r}n^2, \quad (5.14)$$

$$a(x-1|x) = k_2n + k_4n^3.$$

Gemäß den Grundlagen aus Kapitel 2.3.8 lässt sich daraus eine FPE mit folgenden Werten für  $\mathcal{A}$  und  $\mathcal{B}$  nähern:

$$\begin{aligned} \mathcal{A}(\tilde{n}) &= k_1 - k_2\tilde{n} + k_3\tilde{n}^2 - k_4\tilde{n}^3 = \mathbf{S}\tilde{\mathbf{f}} \\ \mathcal{B}(\tilde{n}) &= (k_1 + k_2\tilde{n} + rk_3\tilde{n}^2 + k_4\tilde{n}^3) = \mathbf{S} \operatorname{diag}(\tilde{\mathbf{f}}) \mathbf{S}^T. \end{aligned} \quad (5.15)$$

Wie weiter oben bereits angemerkt, ist die Drift  $\mathcal{A}$  unabhängig von  $r$ . Die FPE enthält allerdings neben der Drift noch die Diffusion  $\mathcal{B}$ , welche ein Maß für das im System vorhandene Rauschen ist. Die Diffusion hängt sehr wohl von  $r$  ab, was für den weiteren Verlauf der Analyse noch wichtig werden wird.

Mithilfe der deterministischen sowie der probabilistischen Beschreibung werden wir nun das Verhalten des Schlögl-Modells analysieren.

## 5.4 Analyse des unmodifizierten Schlögl-Modells

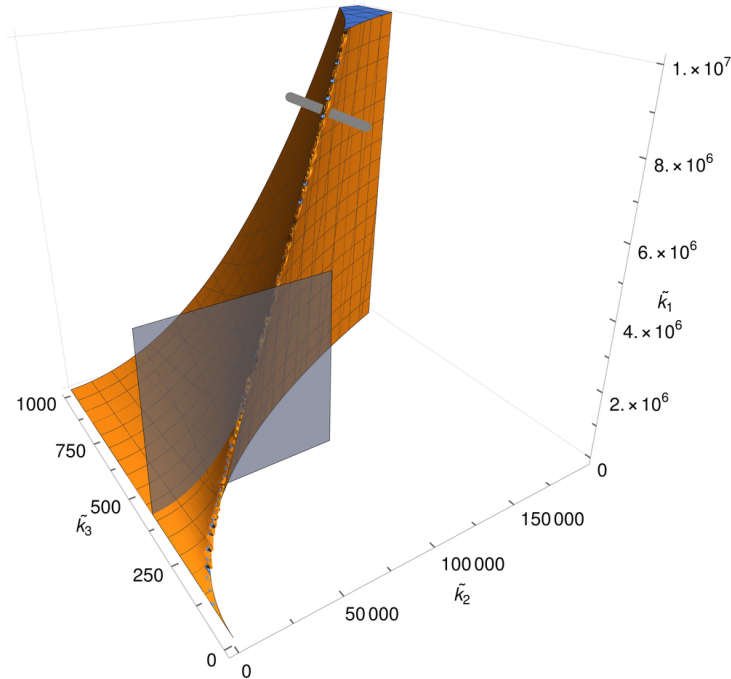
Bevor wir die Auswirkungen von Burstrauschen untersuchen, wollen wir uns hier zuerst das generelle Verhalten des unmodifizierten (ursprünglichen) Schlögl-Modells ( $r = 1$ ) veranschaulichen.

Für die stationäre Lösung des deterministischen Systems gilt:  $\dot{\tilde{n}} = 0$ . Hierdurch ergibt sich ein kubisches Gleichungssystem für  $\tilde{n}$ . Ob ein solches Gleichungssystem ein oder drei Fixpunkte besitzt, lässt sich durch Analyse der Diskriminante klären, welche lautet:

$$\Delta = -27\tilde{k}_1^2 + 18\tilde{k}_1\tilde{k}_2\tilde{k}_3 - 4\tilde{k}_1\tilde{k}_3^3 - 4\tilde{k}_2^3 + \tilde{k}_2^2\tilde{k}_3^2, \quad (5.16)$$

wobei wir hier  $\tilde{k}_i = k_i/k_4$  eingeführt haben.

Aus der Algebra ist für kubische Gleichungen bekannt, dass diese drei reale Nullstellen haben, wenn  $\Delta > 0$  ist, was in unserem Fall gleichbedeutend mit zwei stabilen und einem dazwischenliegenden instabilen Fixpunkt ist. Für den Fall  $\Delta < 0$  hat die Ratengleichung nur einen Fixpunkt, das System ist also monostabil [126]. Je nach Anfangsbedingung wird das deterministische System gegen einen der stabilen Punkte konvergieren oder auf dem instabilen Fixpunkt verweilen,



**Abbildung 5.3:** Darstellung der bistabilen (ausgefüllt) und monostabilen Bereiche der deterministischen Lösung des stationären unmodifizierten Schlögl-Modells. Die eingezeichnete graue Fläche sowie der graue Balken beziehen sich auf die Parameterregionen der Abbildungen 5.4 und 5.5.

falls die Anfangsbedingung bereits genau auf diesem Fixpunkt lag. Zur Veranschaulichung, für welche Parameterkombinationen das Schlögl-Modell zwei stabile Fixpunkte besitzt, ist dieser Bereich in Abbildung 5.3 in orange dargestellt.

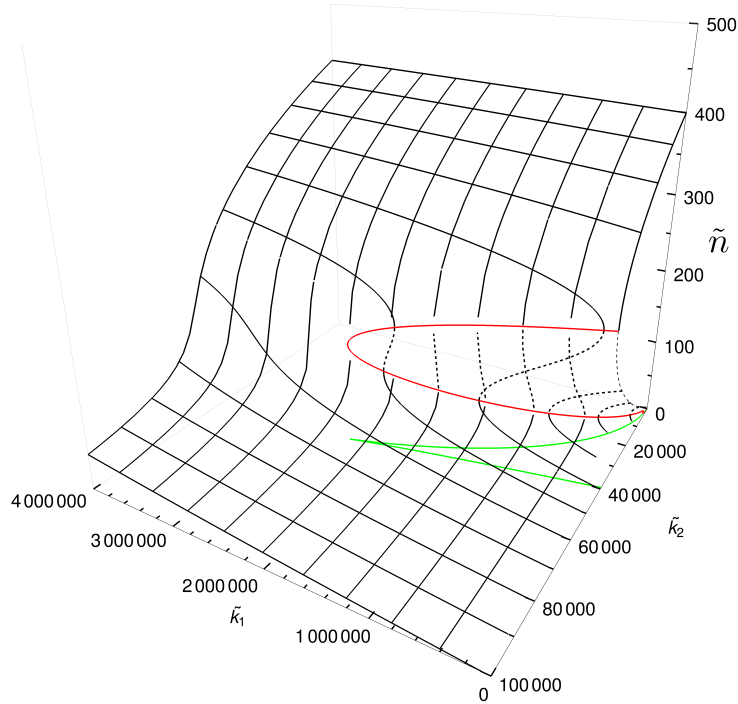
Durch Veränderung der Parameter  $k_i$  kann man das Schlögl-Modell also zwischen monostabil und bistabil transformieren. Die genaue Abhängigkeit der Fixpunkte von der Wahl der Parameter lässt sich übersichtlich durch ein sogenanntes Bifurkationsdiagramm darstellen, in dem die stationären Lösungen für verschiedene Parameterkombinationen aufgetragen sind. Ein solches Bifurkationsdiagramm für das Schlögl-Modell und Parameter  $\tilde{k}_1, \tilde{k}_2$  ist in Abbildung 5.4 gezeigt. Wie im einleitenden Kapitel 5.2 beschrieben, lassen sich Bifurkationen je nach Typ in verschiedene Klassen einteilen. Bei dem vorliegenden Fall handelt es sich um zwei Sattel-Knoten-Bifurkationen (jeweils mit Bifurkationsparameter  $\tilde{k}_1$  und  $\tilde{k}_2$ ), welche sich zu einer sogenannten Cusp-Bifurkation verbinden. Der Name rührt daher, dass die Projektion der Bifurkation auf die  $(\tilde{k}_1/\tilde{k}_2)$ -Ebene (in der Abbildung in grün eingezeichnet), eine spitze Form bildet.

Wenn wir nun von der deterministischen zur stochastischen Beschreibung wechseln, erhalten wir Wahrscheinlichkeitsverteilungen anstatt definierter Punkte. Das Pendant zu deterministischen Fixpunkten sind dann Extrema der Wahrscheinlichkeitsverteilung. Im stationären Zustand ergibt sich für die ME:

$$a^-(n)P_S(n+1) = a^+(n)P_S(n), \quad (5.17)$$

wobei wir als abkürzende Schreibweisen genutzt haben:

$$\begin{aligned} a^+(n) &= a(n+1|n) = k_1 + k_3 n^2, \\ a^-(n) &= a(n-1|n) = k_2 n + k_4 n^3. \end{aligned} \quad (5.18)$$



**Abbildung 5.4:** Stationäre Lösungen von  $\tilde{n}$  im zweidimensionalen Parameterraum. In der  $(k_1 - x)$ - sowie der  $(k_2 - x)$ -Ebene entstehen sogenannte Sattel-Knoten-Bifurkationen, welche sich zu einer Cusp-Bifurkation verbinden. Die rote Linie markiert den Übergang von stabil zu instabil. Die grüne Linie ist die Projektion der roten Linie auf die  $k_1 - k_2$  Ebene. Die dargestellte Parameterkombination ist in Abbildung 5.3 durch die graue Fläche dargestellt.

Demnach gilt die Rekursionsgleichung:

$$P_S(n+1) = \frac{a^+(n)}{a^-(n+1)} P_S(n), \quad (5.19)$$

woraus sich folgende allgemeine Lösung ergibt<sup>1</sup>:

$$P_S(n) = \frac{\prod_{i=0}^{n-1} a^+(i)}{\prod_{i=1}^n a^-(i)} P_S(0). \quad (5.20)$$

Das unmodifizierte Schlögl-Modell lässt sich also durchaus auf Basis der ME analytisch lösen. Da die Herleitung sich zunutze macht, dass der Zustand  $n$  nur von den Zuständen bei  $n \pm 1$  abhängt, wird der gerade vollzogene Weg sehr kompliziert, sobald wir unser System modifizieren und einen Burst mit einbauen. Wir werden daher im Folgenden die ME nicht weiter verwenden, sondern die Näherung durch die FPE betrachten, welche im stationären Zustand geschrieben werden kann als:

$$\frac{\partial}{\partial \tilde{n}} \mathcal{A}(\tilde{n}) P_s(\tilde{n}) = \frac{1}{2} \frac{\partial^2}{\partial \tilde{n}^2} \mathcal{B}(\tilde{n}) P_s(\tilde{n}). \quad (5.21)$$

<sup>1</sup> Vielen Dank an Bahram Houchmandzadeh für den Hinweis auf diesen Lösungsansatz

Unter der Voraussetzung, dass negative Konzentrationen nicht möglich sind<sup>2</sup>, lässt sich dies lösen durch [30]:

$$P_s(\tilde{n}) = \frac{\mathcal{N}}{\mathcal{B}(\tilde{n})} \exp \left[ \int_0^{\tilde{n}} \frac{2\mathcal{A}(\tilde{n}')}{\mathcal{B}(\tilde{n}')} d\tilde{n}' \right]. \quad (5.22)$$

Der Parameter  $\mathcal{N}$  ist hierbei eine Normierungskonstante.

Um das Integral in (5.22) zu lösen, bemerken wir, dass sich  $\mathcal{A}$  und  $\mathcal{B}$  für  $r = 1$  schreiben lassen als:

$$\begin{aligned} \mathcal{A}(\tilde{n}) &= -k_4(\tilde{n} - \tilde{n}_1)(\tilde{n} - \tilde{n}_2)(\tilde{n} - \tilde{n}_3) \\ \mathcal{B}(\tilde{n}) &= k_4(\tilde{n} + \tilde{n}_1)(\tilde{n} + \tilde{n}_2)(\tilde{n} + \tilde{n}_3), \end{aligned} \quad (5.23)$$

wobei  $\tilde{n}_x$  die verschiedenen (komplexen) Nullstellen von  $\mathcal{A}$  bezeichnen.

Partialbruchzerlegung führt dann auf folgende längliche, aber exakte Lösung:

$$\begin{aligned} P_s(\tilde{n}) &= \mathcal{N} e^{2\tilde{n}} (\tilde{n}_1 + \tilde{n})^{\frac{2\tilde{n}_1(\tilde{n}_1 + \tilde{n}_2)(\tilde{n}_1 + \tilde{n}_3)}{(\tilde{n}_1 - \tilde{n}_2)(\tilde{n}_1 - \tilde{n}_3)} - 1} \\ &\quad (\tilde{n}_2 + \tilde{n})^{\frac{2\tilde{n}_2(\tilde{n}_2 + \tilde{n}_1)(\tilde{n}_2 + \tilde{n}_3)}{(\tilde{n}_2 - \tilde{n}_1)(\tilde{n}_2 - \tilde{n}_3)} - 1} \\ &\quad (\tilde{n}_3 + \tilde{n})^{\frac{2\tilde{n}_3(\tilde{n}_3 + \tilde{n}_1)(\tilde{n}_3 + \tilde{n}_2)}{(\tilde{n}_3 - \tilde{n}_1)(\tilde{n}_3 - \tilde{n}_2)} - 1}. \end{aligned} \quad (5.24)$$

In Abbildung 5.5 werden die drei aufgeführten Methoden der Beschreibung für das unmodifizierte Schlögl-Modell miteinander verglichen.

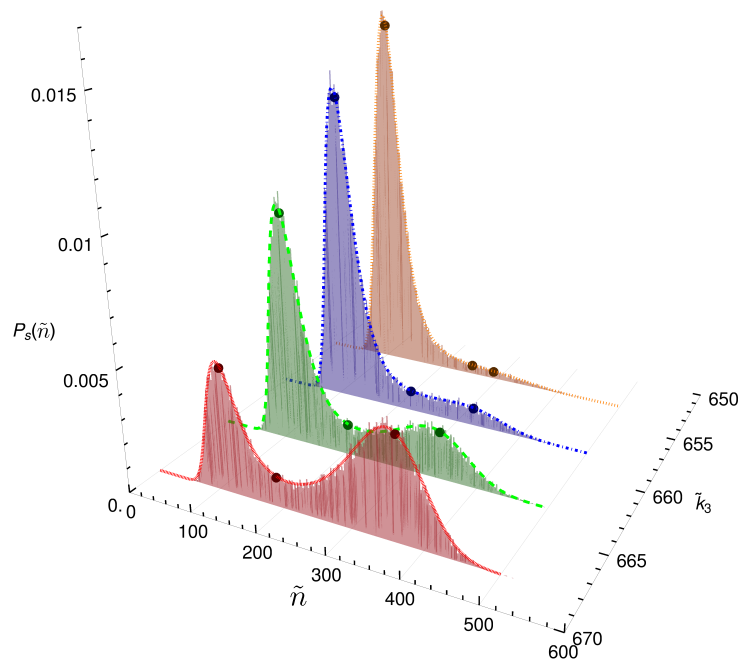
Wie bereits weiter oben angemerkt, lassen sich in der probabilistischen Beschreibung keine Fixpunkte mehr angeben, auf welche das System für  $t \rightarrow \infty$  zuläuft. Stattdessen kann man die Maxima der Wahrscheinlichkeitsverteilung heranziehen, also jene Punkte, in denen das System sich am häufigsten aufhält. In Abbildung 5.5 ist dabei zum einen zu erkennen, dass das Maximum der Wahrscheinlichkeitsverteilung sowie der deterministische Fixpunkt nicht immer aufeinander liegen, zum anderen wird aber auch deutlich, dass die stationäre Lösung der FPE und die stochastische Simulation gut übereinstimmen. Analytisch könnte man die Maxima und Minima der Wahrscheinlichkeitsverteilung also durch Ableitung der Gleichung (5.24) bestimmen. Einfacher und auch für das modifizierte System anwendbar ist aber ein anderer Weg. Ausgehend von Gleichung (5.21) erhalten wir durch einmalige Integration [127]:

$$\alpha(\tilde{n}) := \mathcal{A}(\tilde{n}) - \frac{1}{2} \mathcal{B}'(\tilde{n}) = 0. \quad (5.25)$$

Das Lösen dieser Gleichung für  $\tilde{n}$  ergibt ebenfalls die Extrempunkte der Wahrscheinlichkeitsverteilung [128].

Die Bifurkation, welche in Abbildung 5.4 dargestellt ist, entsteht durch Variation der Parameter  $k_i$  und somit durch Änderung der deterministischen Beschreibung des Modells. Im folgenden Kapitel werden wir zeigen, dass sich Bifurkationen auch ohne Änderung der deterministischen Beschreibung induzieren lassen. Dazu variieren wir den Burst-Parameter  $r$ , welcher bekanntlich die Drift unverändert lässt und nur auf die Diffusion der FPE wirkt.

<sup>2</sup> Die Aussage ist gleichbedeutend damit, dass an der Stelle  $\tilde{n} = 0$  eine reflektierende Randbedingung vorliegt.



**Abbildung 5.5:** Vergleich zwischen den probabilistischen Beschreibungen mit der Mastergleichung und der Fokker-Planck-Gleichung sowie der deterministischen Beschreibung für verschiedene Werte von  $\tilde{k}_3$ . Die stationären Lösungen, welche mittels stochastischer Simulation der Mastergleichung gewonnen wurden, sind als Histogramm dargestellt. Die Lösungen der Fokker-Planck-Gleichung sind durch die Kurven gegeben. Die schwarzen Punkte zeigen die Position der jeweiligen deterministischen Fixpunkte. (genutzte weitere Parameter:  $\tilde{k}_1 = 8 \cdot 10^6$ ,  $\tilde{k}_2 = 1.33 \cdot 10^5$ )

## 5.5 Der Einfluss von Burstrauschen

Wie bereits beschrieben, sind der Propensitätsvektor und die stöchiometrische Matrix für das Schlögl-Modell gegeben durch:

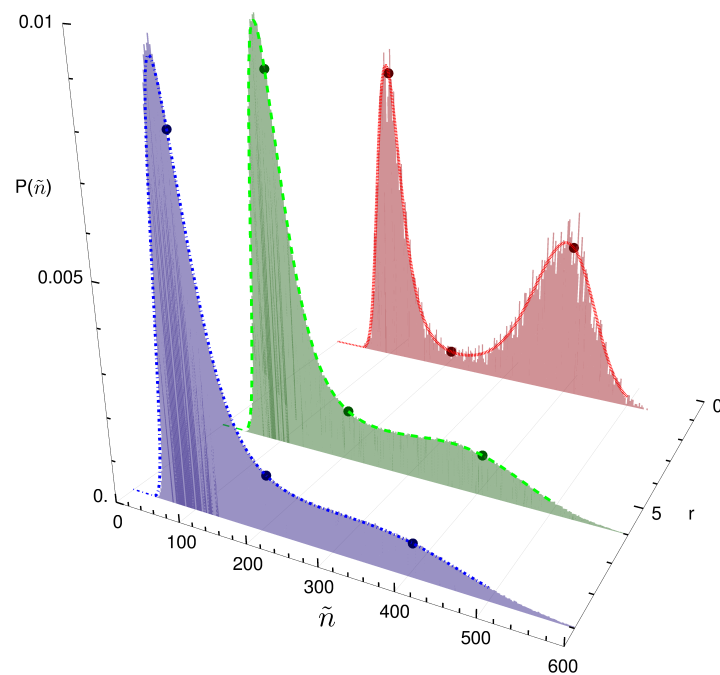
$$\mathbf{S} = [1, -1, r, -1] \quad (5.26)$$

$$\tilde{\mathbf{f}} = [k_1, k_2 \tilde{n}, k_3 / r \tilde{n}^2, k_4 \tilde{n}^3]^T. \quad (5.27)$$

Die deterministische Beschreibung bleibt bei Änderung von  $r$  unverändert, da Reaktion (5.5) zwar  $r$ -mal mehr  $\mathbb{X}$  erzeugt, aber dafür auch  $r$ -mal seltener abläuft. Dementsprechend ändert sich in der Fokker-Planck-Gleichung die deterministische Drift  $\mathcal{A}$  nicht. Wohl aber vergrößert sich die Diffusion  $\mathcal{B}$  mit zunehmendem  $r$ . Dies lässt sich leicht veranschaulichen, wenn man sich die Diffusion als eine Größe vorstellt, die die Stärke des Rauschens im System enthält. Wenn  $r$  vergrößert wird, steigt das Burstrauschen an, also auch die Diffusion. Da  $\mathcal{A}$  und  $\mathcal{B}$  bei  $r > 1$  verschiedene Nullstellen besitzen, ist eine Methode wie in Gleichung (5.23) und damit eine einfache analytische Berechnung der kompletten Verteilung nun nicht mehr möglich.

In Abbildung 5.6 ist die Auswirkung des Burst-Parameters  $r$  auf die stationäre Verteilung dargestellt. Die Raten  $k_i$  wurden hierbei konstant gehalten. Unter Einfluss der höheren Burst-Größe transformiert sich die Verteilung mit ursprünglich zwei Maxima in eine Verteilung mit nur noch einem Maximum. Anschaulich lässt sich das dadurch erklären, dass die Vergrößerung der Burst-Größe vor allem dazu führt, dass die Fluktuationen bei großen Werten von  $\mathbb{X}$  zunehmen. Die Verteilung an dieser Stelle verbreitert sich also bis zu dem Punkt, da sie vollkommen in der restlichen Verteilung aufgeht.

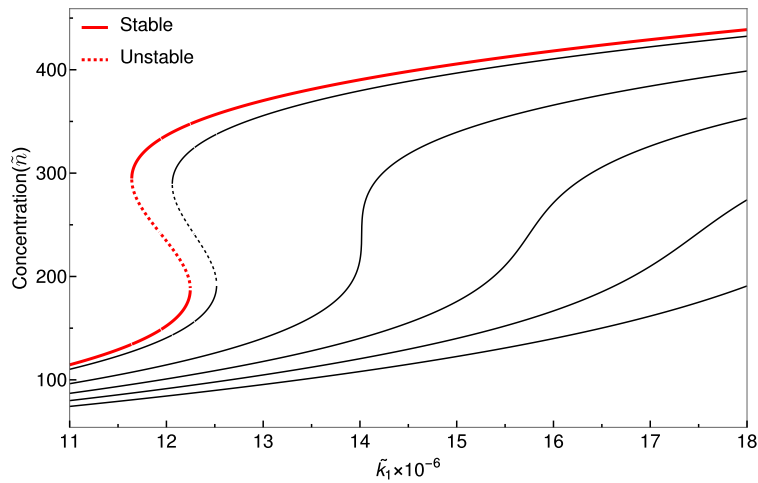




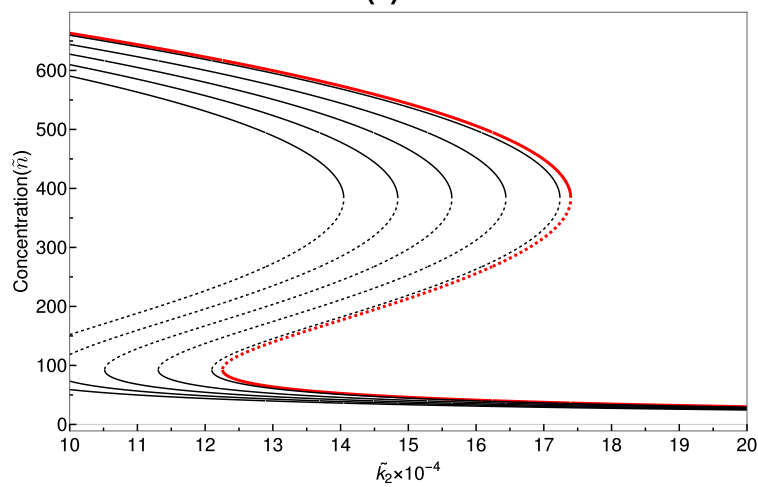
**Abbildung 5.6:** Vergleich zwischen den probabilistischen Beschreibungen mit der Mastergleichung und der Fokker-Planck-Gleichung, sowie der deterministischen Beschreibung für verschiedene Werte des Burst Parameters  $r$ . Die stationären Lösungen, welche mittels stochastischer Simulation der Mastergleichung gewonnen wurden, sind als Histogramm dargestellt. Die Lösungen der Fokker-Planck-Gleichung sind durch die Kurven gegeben. Die schwarzen Punkte zeigen die Position der jeweiligen deterministischen Fixpunkte, welche nicht von  $r$  abhängen. (genutzte weitere Parameter:  $k_1 = 9 \cdot 10^6, k_2 = 1.5 \cdot 10^5, k_3 = 720$ )

Zur näheren Betrachtung dieses Phänomens wenden wir uns nun wieder Bifurkationsdiagrammen zu. Um die Analyse von probabilistischen Systemen zu ermöglichen, definieren wir hierzu die Sattel-Knoten-Bifurkation verrauschter Systeme als eine Bifurkation, bei welcher zwei Extrema der Wahrscheinlichkeitsverteilung verschmelzen [115]. Gemäß der Theorie stochastischer Bifurkationen spricht man in diesem Fall auch von einer phänomenologischen Bifurkation (P-Bifurkation), im Gegensatz zur dynamischen Bifurkation (D-Bifurkation), welche sich nicht über Änderungen in der Verteilung, sondern in der internen Dynamik definiert [31]. Die jeweiligen Extrema der Verteilung lassen sich dabei mit Gleichung 5.25 schnell berechnen. In Abbildung 5.7 sind Bifurkationsdiagramme für verschiedene Stärken des Burstrauschens dargestellt. Es ist zu erkennen, dass eine Verstärkung des intrinsischen Rauschens Bifurkationen zerstören kann. Mit zunehmendem Rauschen ist dann also keine bimodale Verteilung mehr möglich. Interessanterweise lassen sich aber auch Systeme finden, welche erst durch das Rauschen bimodal werden (siehe hierzu Abbildung 5.7(c)).

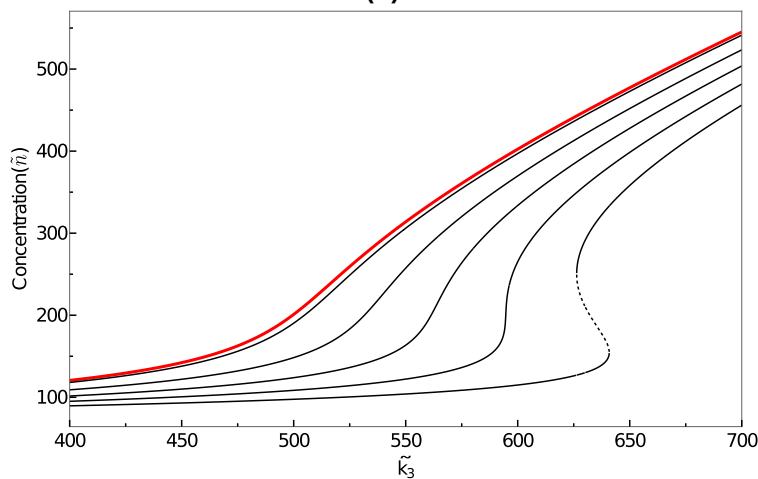
In Abbildung 5.7 ist weiterhin zu erkennen, dass sich das Bifurkationsdiagramm abhängig von dem aufgetragenen Parameter mit zunehmendem  $r$  entweder nach links oder nach rechts bewegt. Diese Richtung lässt sich auch mathematisch intuitiv verstehen. Mittels des Satzes der



(a)



(b)



(c)

**Abbildung 5.7:** Darstellung des Einflusses des Rauschens auf die Extrema der stationären Verteilung. Die rote Linie zeigt jeweils das Bifurkationsdiagramm der deterministischen Beschreibung. Die folgende erste schwarze Linie zeigt die Position der Extrema der probabilistischen Beschreibung für  $r = 1$ . Mit jeder weiteren schwarzen Linie erhöht sich  $r$  um 10. Deutlich zu erkennen ist, wie eine Erhöhung des Rauschens zum einen Bifurkationen zerstören (a), aber auch Bifurkationen induzieren kann (c). (weitere Parameter: (a)  $\tilde{k}_2 = 1.65 \cdot 10^5$ ,  $\tilde{k}_3 = 720$ , (b)  $1 = 5.33 \cdot 10^6$ ,  $\tilde{k}_3 = 800$ , (c)  $\tilde{k}_1 = 8 \cdot 10^6$ ,  $\tilde{k}_2 = 1 \cdot 10^5$ .)

impliziten Funktionen erhalten wir aus Gleichung (5.25) die folgenden Verschiebungen für die Extrema:

$$\begin{aligned} \frac{\partial k}{\partial r}(\tilde{n}, r) &= - \left( \frac{\partial \alpha(n, k, r)}{\partial k} \right)^{-1} \frac{\alpha(\tilde{n}, k, r)}{\partial r} \\ &= \frac{\frac{\partial^2 \mathcal{B}(\tilde{n}, k, r)}{\partial \tilde{n} \partial r}}{2 \frac{\partial \mathcal{A}}{\partial k} - \frac{\partial^2 \mathcal{B}}{\partial k \partial \tilde{n}}}, \end{aligned} \quad (5.28)$$

wobei wir hier ausgenutzt haben, dass  $\mathcal{A}$  nicht von  $r$  abhängt. Explizit ausgewertet für die verschiedenen  $k_i$  ergibt sich:

$$\frac{\partial \tilde{k}_1}{\partial r}(\tilde{n}, x) = \tilde{n} \tilde{k}_3 \quad (5.29a)$$

$$\frac{\partial \tilde{k}_2}{\partial r}(\tilde{n}, x) = - \frac{\tilde{n}}{\tilde{n} - 1} \tilde{k}_3 \quad (5.29b)$$

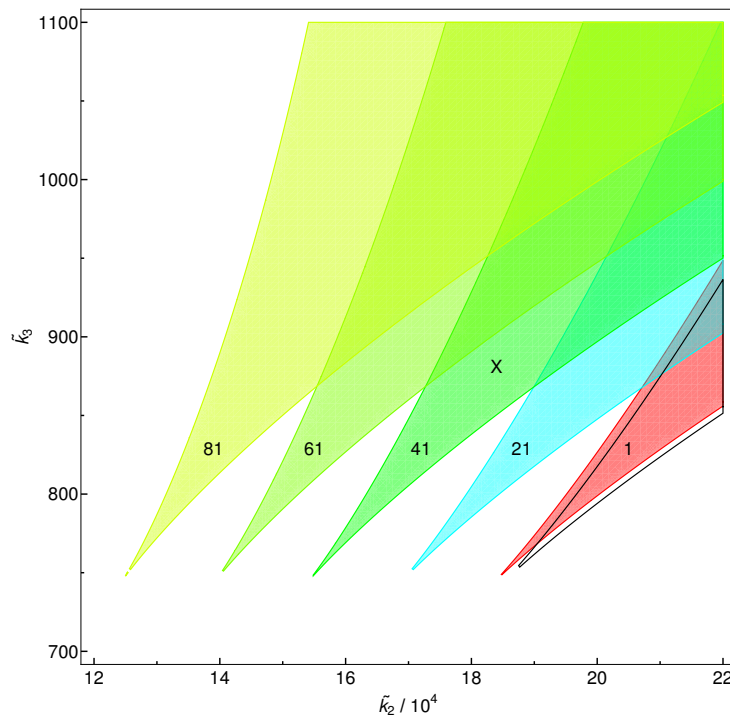
$$\frac{\partial \tilde{k}_3}{\partial r}(\tilde{n}, x) = \frac{1}{\tilde{n} - r} \tilde{k}_3. \quad (5.29c)$$

Wir erkennen, dass für  $\tilde{n} \gg 1$  Gleichung (5.29b) unabhängig von  $\tilde{n}$  ist, weshalb die Kurven in Abbildung 5.7(b) nahezu äquidistant sind und ein stärkeres Rauschen das qualitative Verhalten des Systems kaum beeinflusst. In Abbildung 5.8 ist hierzu noch einmal die Auswirkung verschiedener Stärken des intrinsischen Rauschens auf einen Querschnitt des Phasendiagramms gezeigt. Da hier ebenfalls gilt, dass  $\tilde{n} \gg 1$ , verschieben sich die Kurven äquidistant mit stärkerem Rauschen in  $\tilde{k}_2$  Richtung, wohingegen die Verschiebung in  $\tilde{k}_3$  Richtung (gemäß Gleichung (5.29c)) verschwindend klein ist. Wenn wir einen beispielhaften Satz von Parametern auswählen (hier durch ein schwarzes Kreuz markiert), können wir gut erkennen, wie ein ursprünglich unimodales System (das Kreuz liegt nicht in der farbigen Fläche) mit steigendem  $r$  bimodal wird (das Kreuz liegt in der farbigen Fläche) und dann wieder unimodal.

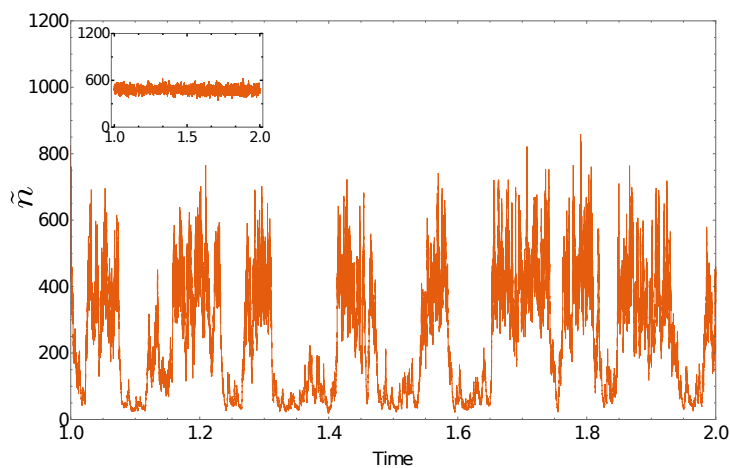
Bis jetzt haben wir nur stationäre Wahrscheinlichkeitsverteilungen betrachtet. Interessant ist aber auch die Frage, ob und wie der beobachtete Übergang zwischen unimodaler und bimodaler Verteilung in der Zeitreihe sichtbar ist. Hierzu wählen wir eine Parameterkombination, für welche wir bereits wissen, dass ein unimodales System mit steigender Rauschintensität bimodal wird und vergleichen die Zeitreihen der Simulation mittels des SSA für verschiedene Werte von  $r$ . Die Abbildung 5.9 zeigt im kleinen Inset die Zeitreihe für das unmodifizierte Schlögl-Modell, also für  $r = 1$ . Deutlich zu erkennen ist, dass das System schmalbandig um den deterministischen Fixpunkt bei  $n = 438$  fluktuiert. Demgegenüber zeigt die große Abbildung, wie das System bei sonst gleichen Parametern mit  $r = 25$  im zeitlichen Verlauf zwischen den beiden Zuständen bei  $\tilde{n} = 43$  und  $\tilde{n} = 376$  springt, also ein bistabiles Verhalten zeigt, welches in der deterministischen Beschreibung nicht erkennbar ist.

## 5.6 Diskussion

In diesem Kapitel haben wir die FPE verwendet, um die stabilen Punkte des Schlögl-Modells zu analysieren. Insbesondere haben wir dabei Burstrauschen so in das Modell eingebaut, dass dessen deterministische Beschreibung dadurch nicht verändert wird. Für die unmodifizierte Version des Schlögl-Modells liegen die Extrema der stationären Lösungen der FPE in der Nähe der Fixpunkte der deterministischen RRE. Wenn jedoch Burstrauschen hinzugefügt wird, kann das resultierende intrinsische Rauschen stabile Zustände erzeugen und zerstören. Die dabei



**Abbildung 5.8:** Querschnitt durch den Phasenraum in der ( $\tilde{k}_1 = 1.54$ )-Ebene. Für die rote Fläche wurde  $r = 1$  gewählt. Für jede weitere farbige Fläche wurde  $r$  um 20 vergrößert (wie auch durch die Zahlen in den jeweiligen Flächen angegeben). Die schwarze Linie zeigt die deterministische Lösung. Der hier dargestellte Parameterbereich entspricht der grauen Linie in Abbildung 5.3



**Abbildung 5.9:** Zeitreihe des modifizierten Schlögl-Modells mit Parametern:  $k_1 = 3 \cdot 10^6$ ,  $k_2 = 77518$ ,  $k_3 = 601$ ,  $r = 25$ . Die Abbildung zeigt eine bistabile Charakteristik mit zwei stabilen Punkten bei jeweils  $X = 43$  und  $X = 376$ . Die kleine Abbildung zeigt das selbe System mit  $r = 1$ . Hier besitzt das System augenscheinlich nur einen stabilen Punkt bei  $X = 438$ .

---

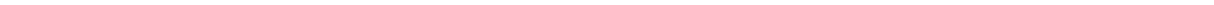
---

interessanteste Beobachtung ist, dass Burstrauschen in einem deterministisch monostabilen System eine Bimodalität hervorrufen kann.

Unser eindimensionales Modell ist hierbei das einfachste System, welches aufgrund von Eigenrauschen Bimodalität erzeugen kann. Die Ergebnisse sind dabei weit über die chemischen Reaktionssysteme hinaus relevant. So spielt Burstrauschen auch eine wichtige Rolle in der Populationsdynamik. Zum Beispiel wurde kürzlich gezeigt, dass die Hinzufügung von intrinsischem Rauschen zur dynamischen Beschreibung des Verhalten einer Spezies in Bezug auf Aussterbens- und Überlebenswahrscheinlichkeit [129] dramatisch verändern kann. Burstrauschen ist auch eine bekannte Art von Rauschen in Halbleitern wie z. B. NPN- und PNP-Transistoren [130], wo es im Niederfrequenzbereich dafür verantwortlich ist, dass das Rauschspektrum anders erscheint als man theoretisch erwarten würde [131].

Es ist zu erwarten, dass die Auswirkungen von intrinsischem Burstrauschen bei komplexeren Systemen noch einflussreicher und damit wichtiger werden. Allerdings wird die analytische Untersuchung eines größeren Reaktionsnetzes schnell undurchführbar. Während es mehrere Ansätze gibt, um die Simulation großer Reaktionsnetzwerke, die Burstrauschen beinhalten, zu beschleunigen [132, 133], dient unser minimales Modell als guter Ausgangspunkt für eine theoretische Analyse und hilft damit das grundlegende Verständnis des Phänomens zu verbessern.

Weiterhin wurde bereits gezeigt, dass intrinsisches Rauschen in zweidimensionalen Systemen Schwingungen induzieren [134, 135] und Hopf-Bifurkationen erzeugen kann [136, 137], die nicht durch die deterministische Beschreibung vorhergesagt werden. Für zukünftige Arbeiten erscheint es also sinnvoll, nach minimalen Modellen wie dem unsrigen zu suchen, welche auch die anderen Bifurkationen korrekt erfassen.



---

---

## 6 Minimalmodelle in der Synthetischen Biologie

»All models are wrong, but some are useful«

George Box [138]

In der Einleitung dieser Arbeit haben wir beschrieben, wie sich die verschiedenen Kapitel auf unterschiedliche Art und Weise in die Arbeitsweise und Ziele der Synthetischen Biologie eingliedern. Der Nutzen des in Kapitel 3 modellierten genetischen Schalters für die Synthetische Biologie lässt sich leicht erkennen. Auch die in Kapitel 4 beschriebenen Möglichkeiten zur Inklusion von Kontexteffekten sind direkt auf aktuelle Probleme der Synthetischen Biologie anwendbar. Anders verhält es sich mit dem im vorigen Kapitel beschriebenen minimalen Modell für durch Burstrauschen induzierte Bimodalität. Zwar ist Burstrauschen ein häufiges Phänomen in biochemischen Systemen, unser Modell enthält aber eine trimolekulare Reaktion. Hierfür müssten sich im dreidimensionalen Raum zeitgleich drei Substanzen an einer Stelle treffen, was nahezu unmöglich ist. Mit Recht kann man das Schlögl-Modell also als unbiologisch bezeichnen. Aus physikalischer Sicht ist die beschriebene, durch Burstrauschen induzierte Bimodalität aber ein spannendes Phänomen und verdient weitere Betrachtung. Dies ist hierbei nur eines von vielen möglichen Beispielen, welche aufzeigen, dass sich die Synthetische Biologie aufgrund ihrer Interdisziplinarität in einem Spannungsfeld verschiedener epistemologischer<sup>1</sup> Anschauungen und Ziele befindet. Auf die Ziele der Synthetischen Biologie bezogen wird dies besonders deutlich beim Vergleich der vorherrschenden »epistemischen Kulturen« in Naturwissenschaften und Ingenieurwissenschaften. Aus ingenieurwissenschaftlicher Sicht könnte ein langfristiges Ziel der Synthetischen Biologie beispielsweise die Entwicklung von Mikroorganismen sein, welche komplexe Substanzen produzieren, als Biosensoren fungieren oder in biomedizinischen Anwendungen wie der Krebstherapie eingesetzt werden können [139]. Im Gegensatz dazu zielt die Epistemik der klassischen Naturwissenschaften eher auf ein tieferes Verständnis der zugrunde liegenden Prozesse und Zusammenhänge in lebenden Organismen ab. Obwohl diese Ziele natürlich auch in der Systembiologie verfolgt werden konnten, ebnet erst die Synthetische Biologie die Bahn für die Herstellung kleinerer und gut isolierter Systeme, welche eine fokussierte Analyse einzelner Prozesse ermöglichen. Der Philosoph Alfred Nordmann vergleicht in seinem Aufsatz »Synthetic Biology at the Limits of Science« [9] die verschiedenen Interpretationen zentraler wissenschaftlicher Konzepte in Bezug auf die »epistemic communities«, welche in der Synthetischen Biologie zusammen arbeiten. Er schlussfolgert dabei, dass es nahezu unmöglich ist, die verschiedenen epistemischen Ideale zu vereinen, da sie in entgegengesetzte Richtungen arbeiten: »Here the reduction of complexity for the purposes of intellectual tractability, there the drawing together of scientific knowledge and technological capabilities. Here the identification of bottlenecks and needs-to-know for rational engineering, along with the demand for more and better theoretical knowledge in order to diminish ignorance, and there the attempt primarily to discover how much one can achieve even with how little we know, with considerable tolerance for ignorance of everything that can be black-boxed.« Abschließend erwähnt Nordmann allerdings, dass diese Spannung, oder sogar Antagonismus auf keinen Fall bedeutet, dass die Forschungsergebnisse der verschiedenen Disziplinen sich nicht gegenseitig nützen können. Eine These, die in Bezug auf solch unbiologischen Modelle wie das Schlögl-Modell fraglich erscheint. Inwieweit

---

<sup>1</sup> epistemologisch: Die Epistemologie, also das Zustandekommen von Wissen betreffend

---

dieser gegenseitige Nutzen sich bei der Verwendung von Modellen, und hierbei insbesondere bei dem von uns verwendeten Schlögl-Modell manifestiert, wollen wir daher diskutieren.

Im folgenden Abschnitt werden wir hierzu die teilweise widersprüchlichen Verwendungen und Rollen von Modellen in den Ingenieurwissenschaften und der Physik analysieren. Anschließend beziehen wir uns explizit auf das Schlögl-Modell, welches bekanntlich stark von physikalischen Idealen inspiriert ist, und zeigen, wie diese Art von Modell die Synthetische Biologie positiv beeinflusst hat und zu einem Modell mit biologischem Wert wurde. Wir kommen dann zu einem ähnlichen Schluss wie bereits Alfred Nordmann, nämlich, dass die verschiedenen epistemischen Kulturen in der Synthetischen Biologie nicht nur von gegenseitigem Nutzen sind, sondern dass die Synthetische Biologie auch Wege bietet, wie beide epistemischen Ideale verschmelzen können.

Die Diskussion philosophischer Fragestellungen wurde in enger Zusammenarbeit mit Thorsten Kohl (AG Nordmann) geführt. Thorsten Kohl hat sich dabei vor allem mit der Rolle von »epistemischen Dingen« im technowissenschaftlichen Experimentalsystem beschäftigt (Die zugehörige Veröffentlichung befindet sich in Revision). Der Fokus des Autors lag auf der im Folgenden ausgeführten Analyse von Minimalmodellen in der Synthetischen Biologie.

---

## 6.1 Der Nutzen von Modellen in Ingenieur- und Naturwissenschaften

---

Aus traditioneller Sicht werden Modelle in erster Linie dazu verwendet, zuverlässige Vorhersagen für das Verhalten von zu untersuchenden Systemen zu treffen, um ansonsten notwendige Experimente zu umgehen. Je mehr der relevanten Phänomene des betrachteten Systems das Modell korrekt abbilden oder darstellen kann, desto besser ist es [140]. Das Verhalten bei speziellen und im Normalfall nicht auftretenden Parameterkombinationen ist dabei nicht von Interesse. Die Systembiologie folgt diesem Verständnis von sinnvollen und nützlichen Modellen und versucht, alle verfügbaren Erkenntnisse in die Modelle zu integrieren, um eine Konvergenz zwischen Modell und Realität zu erreichen, was zu großen und komplexen Modellen führen kann. In der Synthetischen Biologie wird demgegenüber möglicherweise vorhandene Komplexität ignoriert, indem komplizierte Schaltungen auf eine Input-Output-Beziehung reduziert werden, die nur als sogenannte Black Boxes [141] erscheinen. Die Unwissenheit über die genauen Zusammenhänge in dieser Blackbox ist epistemisch kein Hindernis, da das Hauptziel die zuverlässige Beschreibung des beobachteten Phänomens ist. So lange dieses Ziel erreicht wird, ist die Korrektheit der Theorie nur sekundär. In beiden Fällen sind die Grundlagen, von denen ausgehend die technischen Modelle entwickelt werden, oft etablierte wissenschaftliche Gesetze, die dann aber durch viele Erweiterungen und Näherungen ihre Einfachheit verloren haben [142].

Auf der anderen Seite gibt es die idealisierten wissenschaftlichen Modelle, die auf viele Details verzichten und oft aus physikalischer Sicht an »Schönheit« und »Stärke« gewinnen, indem sie leicht verständlich und analytisch lösbar oder zumindest in Näherungen gut behandelbar sind [140]. Ihr Ziel ist es nicht, reale Phänomene genau zu beschreiben, sondern eine erklärende Funktion zu haben. Das Modell ist dabei oft nur eine abstrakte oder idealisierte Form eines Phänomens oder eines physikalischen Systems [143]<sup>2</sup>. Der Anspruch an die Modelle ist folglich deren Allgemeingültigkeit und Einfachheit, welche die Grundlage aller Erklärungskraft sind, und damit die Konzentration auf universelle Prinzipien. Die Vernachlässigung mikroskopischer Details und zunehmende Abstraktion sind dabei oft der einzige Weg, um Verständnis zu erlangen.

Für die Philosophin Nancy Cartwright ist klar, dass die Ziele der Modelle von Ingenieur- und Naturwissenschaften nicht gleichzeitig erreicht werden können. Sie stellt fest, dass es immer einen Kompromiss zwischen der Wahrheit und der Erklärungskraft gibt. Die grundlegendsten

---

<sup>2</sup> Die Philosophin Margaret Morrison unterscheidet an dieser Stelle zwischen Abstraktion als Prozess, in dem wir ein Phänomen in einer Weise beschreiben, die in der physikalischen Welt nicht realisierbar ist, und Idealisierung, die lediglich Näherungen enthält, die durch Korrekturfaktoren eliminiert werden können (siehe auch [142]). Beides erzeugt Aussagen über Phänomene, die nicht physikalisch real sind, aber letztere können dem realen System sehr nahe sein.



---

physikalischen Gesetze stellen selten reale Fakten dar, während die Grundbausteine der Ingenieurwissenschaften dies tun, aber – bedingt durch z. B. viele eingeführte Korrekturfaktoren und die Zusammenführung verschiedener Effekte in eine große Formel – ihre Aussagekraft verlieren und nur in dem Kontext gelten, für den sie kalibriert wurden. Sie kommt zu dem Schluss: »if the laws of physics are to explain how phenomena are brought about, they cannot state the facts« [142].

Im Folgenden wollen wir uns die Frage stellen, wie viel Realität ein Modell jemals repräsentieren kann. Die vorherrschende Überzeugung ist, dass Modelle zwar auf unterschiedliche Weise, aber letztlich immer fehlerhaft sind. Dieser Mangel darf jedoch nicht als Nachteil bewertet werden, sondern ist für die Rolle des Modells unerlässlich [142, 144].

William Wimsatt listet hierzu verschiedene Möglichkeiten auf, wie ein Modell falsch sein kann und kategorisiert deren Fehler nach Schwere. Die weiter oben charakterisierten Modelle, die in den Natur- oder Ingenieurwissenschaften verwendet werden, können hierbei leicht verschiedenen dieser Fehler zugeordnet werden. Dies zeigt einmal mehr die Vielfalt der Modelle und deren Nutzen. So können beispielsweise für das Ingenieurmodell häufig die nur lokale Anwendbarkeit und oft auch die Beschränkung auf eine phänomenologische Beschreibung als typische Fehler identifiziert werden. Im Gegensatz dazu sind die wissenschaftlichen Modelle meist Idealisierungen, deren Voraussetzungen für die Anwendbarkeit in der Natur nie gegeben sind/sein können<sup>3</sup>. An dieser Stelle sollte keine explizite Bewertung der verschiedenen Fehler und Modelle vorgenommen werden. Vielmehr muss die Eignung für unterschiedliche Anwendungsszenarien und epistemische Absichten jeweils neu geprüft werden.

George Box prägte die Aussage: »All models are wrong, but some are useful« [145]. Es ist interessant zu diskutieren, welche Art von Modell und damit welche Art von Fehlern welchen Einfluss auf den epistemischen Nutzen, d. h. das vom Modellen gewonnene Wissen, hat. Wimsatt kommt dabei zu dem Schluss, dass vor allem solche Modelle zu bevorzugen sind, die entweder idealisiert oder so unvollständig sind, dass ihnen kausal relevante Variablen fehlen. Nur diese Modelle sind im Allgemeinen so strukturiert, dass Fehler leicht lokalisiert und auf bestimmte im Modell enthaltene Mechanismen zurückgeführt werden können. Zu einem ähnlichen Ergebnis kommen Radde und Hütt in ihrem Vergleich der aktuellen Modellierungsmethoden der Biologie [146].

Im Folgenden gehen wir nun näher auf ein Modell ein, das nicht stärker von den Naturwissenschaften beeinflusst sein könnte, in gewisser Weise den Extremfall der oben vorgestellten naturwissenschaftlichen Modelle darstellt und damit auch die Fehler der Idealisierung und Unvollständigkeit in sich maximiert. Wir sprechen über das minimale Spielzeugmodell.

---

## 6.2 Minimale Spielzeugmodelle

---

Wir beginnen diesen Abschnitt mit einer Definition von Minimalmodellen. Der Theoretiker Nigel Goldenfeld definiert ein Minimalmodell als ein solches, welches »most economically caricatures the essential physics«. Basierend auf dieser Definition zeigen Batterman und Rice wie Modelle, die auf der Grundlage der Renormierungsgruppentheorie aus einer Vielzahl von zugrunde liegenden Modellen hervorgehen, als übergeordnetes Minimalmodell fungieren können [147]. Im Folgenden wollen wir auf deren Definition aufbauen, sie aber noch enger interpretieren. Als Minimalmodell wollen wir ein Modell bezeichnen, das ein bestimmtes Phänomen beschreibt, diese Fähigkeit aber mit jeder weiteren Reduzierung der Freiheitsgrade verliert. Im Gegensatz zur Nutzung durch Goldenfeld, Batterman und Rice muss unser Minimalmodell nicht mit einem realen System in Verbindung gebracht werden, oder aus der Analyse eines realen Systems hervorgehen.

---

<sup>3</sup> Wimsatt selbst stellt fest, dass die Fehler je nach Modell und Betrachtungskontext unterschiedlich zugeordnet werden können. Unsere Klassifizierung ist daher nicht als absolut zu verstehen. Eine globale Klassifizierung ist nicht möglich.

---

Vielmehr werden wir das Minimalmodell über Hartmanns Spielzeug- oder Sondierungsmodell definieren:

In seiner diachronen Untersuchung von Modellen beschäftigt sich Hartmann ausführlich mit dem sogenannten Spielzeugmodell [148]. Ihm folgend ist dies durch das Fehlen einer »direkt beabsichtigten physikalischen Anwendung« gekennzeichnet, d.h. es lehrt über nichts anderes als über sich selbst. Als Beispiel für das Spielzeugmodell wählt Hartmann die aus der Quantenfeldtheorie stammende  $\phi^4$ -Theorie, die ein selbstwechselwirkendes Skalarfeld  $\phi$  beschreibt. Es ist bemerkenswert, dass diese Theorie – obwohl basierend auf der zugrunde liegenden Quantenfeldtheorie wohl definiert – ein Teilchen beschreibt, das nicht existiert. Damit stellt sich sofort die Frage, warum eine Theorie, die wissentlich keinen Bezug zur Wirklichkeit hat und damit auf den ersten Blick auch keinen Erkenntnisgewinn bietet, überhaupt aufgestellt wird. An dieser Stelle ist aber wichtig zu erwähnen, dass die  $\phi^4$ -Theorie die einfachst mögliche Quantenfeldtheorie ist. Basierend auf dieser Beobachtung nennt Hartmann vier Gründe, warum das Studium dieser unrealistischen Theorie dennoch von epistemischem Nutzen sein kann:

- Man bekommt ein »Gefühl« für die Quantenfeldtheorie.
- Die  $\phi^4$ -Theorie kann genutzt werden, um neue mathematische Ansätze auszuprobieren.
- Die  $\phi^4$ -Theorie ermöglicht, generelle Eigenschaften zu extrahieren, welche möglicherweise jedes Modell der Quantenfeldtheorie besitzt.
- Eine Analyse der  $\phi^4$ -Theorie führt zu einem Satz von verschiedenen Mechanismen, welche auch auf komplexere Theorien angewendet werden können.

An dieser Stelle sind die Parallelen zwischen Hartmanns Spielzeugmodellen und Morrisons Abstraktionen zu beachten. Laut Morrison führt die Abstraktion dazu, dass Modelle zu »autonomen Agenten« werden. Häufig ersetzen diese Agenten dann das eigentliche System, das als »untersuchtes Objekt« nicht zugänglich ist. Unser Wissen über die reale Welt wird dann durch das Modell vermittelt, oder um es in einen Satz zu fassen: »Das mathematische Modell ersetzt das physikalische Phänomen als Untersuchungsgegenstand«. Im Gegensatz zu Hartmann basiert Morrisons Abstraktion jedoch auf einem zu beschreibenden realen System oder Phänomen, das durch mathematische Verzerrungen in ein Modell umgewandelt wird.

In Bezug auf Wimsatts Klassifizierung ist das Spielzeugmodell ein Modell, das den Fehler der Idealisierung (oder Abstraktion) bewusst auf die Spitze treibt, um eine maximale Vereinfachung eines wesentlichen Phänomens zu erreichen. Hartmanns vier Gründe, sich mit Spielzeugmodellen zu beschäftigen, finden sich daher auch in Wimsatts Liste der möglichen Vorteile fehlerhafter Modelle.

Gemäß dieser Definition eines minimalen Spielzeugmodells wollen wir nun das in Kapitel 5 schon behandelte Schlögl-Modell als ein solches analysieren. Darauf basierend werden wir zeigen, wie dieses Minimalmodell einen erkenntnistheoretischen Wert in der Synthetischen Biologie erlangt hat, d.h. die Barriere der beiden oben beschriebenen gegensätzlichen Modellansichten überschritten zu haben scheint.

---

### 6.3 Das Schlögl-Modell als minimales Spielzeugmodell

---

Wie in Kapitel 5 beschrieben, zeigt das Schlögl-Modell für bestimmte Parameterkombinationen ein bimodales Verhalten. Schlögl brachte dieses Verhalten in Zusammenhang mit der bereits bekannten Theorie der Phasenübergänge und zeigte phänomenologische Parallelen auf.

Im Hinblick auf Hartmanns Definition des Spielzeugmodells ist zu beachten, dass die Definition des Schlögl-Modells von einer trimolekularen Reaktion ausgeht. Damit dies in einem realen System geschieht, müssen sich drei Moleküle im kontinuierlichen Raum fast gleichzeitig treffen.

---

Trimolekulare Reaktionen sind daher äußerst unwahrscheinlich und werden meist als unrealistisch angesehen. Schlögl selbst war sich dieses Problems bewusst. Dennoch argumentierte er in Übereinstimmung mit Hartmann [118]: »Phasenübergänge in Nichtgleichgewichtszuständen sind immer mit nichtlinearen Transportgleichungen verbunden, deren Untersuchung im Allgemeinen kompliziert ist. Daher erscheint es sinnvoll, Modellsysteme zu diskutieren, die Phänomene dieser Art von Phasenübergängen zeigen und die in Bezug auf die mathematische Behandlung so einfach wie möglich sind«. Schon aus dieser Formulierung ist ersichtlich, dass das Schlögl-Modell ein Modell ist, das genau die Anforderungen erfüllt, die wir an ein Minimalmodell gestellt haben. Wie Schlögl schreibt, ist das Modell in Bezug auf die Darstellung des gewünschten Verhaltens »So einfach wie möglich«, so dass es sicherlich als minimal beschrieben werden kann. Durch das Vorhandensein trimolekularer Reaktionen werden im Schlögl-Modell Annahmen getroffen, die weit über die Approximation und sogar (im Sinne von Morrison) über die Idealisierung hinaus in Richtung Abstraktion gehen, denn selbst durch die Einführung eines einfachen Korrekturfaktors kann diese Unwirklichkeit nicht korrigiert werden.

An dieser Stelle ist zu beachten, dass Schlögl definitiv auch über mögliche Anwendungsbereiche nachdachte. So beschreibt er beispielsweise, dass es Analogien zum Laser und zur Tunnelodiode gibt. Diese Referenz ist jedoch rein phänomenologischer Natur und kann nicht mit einer Anwendungsreferenz verglichen werden, wie sie aus technisch-wissenschaftlicher Sicht verstanden wird.

---

#### 6.4 Das Schlögl-Modell in der Transformation

---

Analog zu Hartmanns Ausführungen über die Vorteile von Spielzeugmodellen wurde das Schlögl-Modell in den letzten Jahren in der wissenschaftlichen Forschung auf vielfältige Weise eingesetzt. So untersuchten Lubensky et al. beispielsweise mit dem Schlögl-Modell, wie sich die Systemgröße auf die Stabilität stochastischer bistabiler Systeme auswirkt, und identifizierten mit dem Modell allgemeine Merkmale der zugrunde liegenden Theorie stochastischer Systeme [149]. Hanggi et al. wiederum leiteten eine neue, verbesserte Variante der Fokker-Planck-Gleichung ab und verwenden das Schlögl-Modell als einfache Möglichkeit, die Qualität ihrer neuen Methode zu validieren. Sie nutzen also das Modell, um sich an neue formale Techniken zu gewöhnen [150]. Auch wir haben in Kapitel 5 auf den bekannten Mechanismen des Schlögl-Modells aufgebaut und haben daraus ein Minimalmodell für durch Burstrauschen induzierte Bimodalität entwickelt [70].

Während die genannten und viele hundert andere Arbeiten das Schlögl-Modell nach den von Hartmann für Spielzeugmodelle postulierten Punkten verwenden, zeigt die folgende Arbeit einen Bruch mit dieser Definition:

Robert Endres beschreibt in seinem Artikel »Bistability: Requirements on Cell-Volume, Protein Diffusion, and Thermodynamics«, wie er ein mögliches zellbiologisches Modell, nämlich einen selbstregulierenden Schalter, auf das »gut charakterisierte« Schlögl-Modell abbildet [124]. Um die Eigenschaften der Schaltung zu analysieren, verweist er auf zahlreiche bisherige Methoden, die in Bezug auf das Schlögl-Modell entwickelt wurden. Basierend auf der Abbildung eines möglichen biologischen Systems auf das Spielzeugmodell zeigt er auf, welche Bedingungen erfüllt sein müssen, damit ein zelluläres System bistabiles Schalten ermöglicht. Er vermeidet das Problem der trimolekularen Reaktionen, welche im Schlögl-Modell angenommen, aber biologisch nicht durchführbar sind, indem er das kooperative Verhalten zwischen zwei Proteinen approximiert. Endres kehrt also die epistemische Richtung zwischen Biologie und Physik um. Wo konventionell ein Modell nach einer biologischen Beobachtung erstellt wird, gestaltet er die biologische Anwendung nach dem Modell.

Was hier zu beobachten ist, lässt sich am besten mit Jane Calverts Worten ausdrücken: »A audacious example of how epistemic practices can come to influence biological materiality is given by synthetic biologists who say that if they discover that their models of biological phenomena do not work, then they will simply engineer the biological parts to fit the model

---

---

better, rather than changing the model to make it a more accurate representation of the biological system, which is how biological research would have proceeded in the past« [151].

Diese Beobachtung verleiht dem Nutzen von Spielzeugmodellen eine neue Dimension. Wie vorstehend ausgeführt, beschreiben die minimalen Spielzeugmodelle ein bestimmtes Phänomen und verlieren diese Fähigkeit mit jeder weiteren Reduzierung der Freiheitsgrade. Für den Ethos der Synthetischen Biologie Design der Dekonstruktion vorzuziehen, bieten die Spielzeugmodelle daher den perfekten Nährboden, um bestimmte Phänomene mit möglichst geringem Aufwand zu beobachten.

Hier schließen wir den Kreis dieser Arbeit und werfen erneut einen Blick auf die in der Einleitung angeführten berühmten letzten Worte von Richard Feynman: »What I cannot create, I do not understand.«. In Bezug auf die oben beschriebenen Minimalmodelle kann dieser Satz verstanden werden als [9]: Egal, wie gut meine Modelle sind und wie gut ich sie verstanden habe – wofür ein Minimalmodell besonders geeignet ist – all das reicht nicht aus, um ein echtes »Verständnis« zu erreichen. Dazu muss ich zuerst »gebaut« haben.

Die Schaffung und Konstruktion neuartiger Strukturen in der Synthetischen Biologie ermöglicht es, bestehende theoretische Ideen, insbesondere minimale Spielzeugmodelle, in die Realität umzusetzen. Man müsste also den Satz von Feynman eigentlich umkehren: »Was ich bauen kann, habe ich verstanden«. Die Klärung, ob es sich dabei um ein wissenschaftliches oder unwissenschaftliches Konzept handelt, ist Gegenstand eines laufenden philosophischen Diskurses [9]. An dieser Stelle ist jedoch klar, dass die Kombination von Minimalmodellen und Synthetischer Biologie dem Feynman-Prinzip des Verstehens sehr nahe kommt, da es nun möglich ist, ein Modell so einfach zu konstruieren, dass es im Sinne der konventionellen Naturwissenschaften zugänglich ist und damit nach entsprechender Charakterisierung im Sinne der Technikwissenschaften aufgebaut werden kann.

Die folgende Hypothese von Heinemann und Panke wird also womöglich bald Wirklichkeit werden: »Ultimately, mathematical models developed for research purposes (e.g. in systems biology) will be employed as design models in synthetic biology« [152].

---

## 7 Fazit

Die Synthetische Biologie ist ein junger interdisziplinärer Forschungsbereich, welcher unterschiedlichste Wissenschaften wie Biologie, Physik, Mathematik oder auch Informatik in sich vereint. Das gemeinsame Ziel ist die künstliche Erzeugung von genetischen Schaltkreisen, bis hin zu kompletten lebenden Organismen. Während die experimentellen Tätigkeiten hierbei vor allem von Forschenden aus Biologie und Chemie verrichtet werden, ist die Aufgabe der theoretischen Physik beispielsweise die Analyse von Daten und darauf basierend die Entwicklung neuer Modelle. Hierbei zeigte sich, dass die bisher verwendeten Methoden oftmals nicht den Bedürfnissen der Synthetischen Biologie genügen, sei es aufgrund der Nichtbeachtung von Kontexteffekten oder der Vernachlässigung vorhandener Fluktuationen.

Davon ausgehend wurden in der vorliegenden Arbeit drei verschiedene Themen der Synthetischen Biologie behandelt. Jedes für sich geht auf andere aktuelle Fragestellungen und Probleme der Synthetischen Biologie ein. Die erhaltenen Ergebnisse leisten dabei wertvolle Beiträge zur Analyse von Umgebungseffekten sowie dem Verständnis von stochastischen Modellen.

Im ersten Teil wurde ein genetischer Schalter, welcher auf dem CRISPR/dCas9-System beruht, analysiert. Basierend auf experimentellen Daten konnten ein deterministisches und ein stochastisches Modell entwickelt werden. Die Modelle umfassen hierbei lediglich drei Reaktionen, sind aber trotz ihrer Einfachheit in der Lage, das experimentell beobachtete bimodale Schaltverhalten gut nachzubilden. Unsere Ergebnisse stimmen dabei mit dem Verständnis überein, dass  $\beta$ -Estradiol (ES) die Diffusion der Transkriptionsfaktoren vermittelt. Die beobachtete Bimodalität entsteht in unserem Modell dadurch, dass die Diffusion von ES deutlich langsamer als die Transkription verläuft. Ob die von uns angenommenen drei Reaktionen aber tatsächlich die zentralen Mechanismen des Systems sind, lässt sich mit den bisherigen Experimenten nicht belegen. Ebenfalls vorstellbar wäre zum Beispiel, dass es einen anderen langsamen Prozess gibt, der den Transkriptionsfaktor oder ES reversibel bindet. Um belastbarere Aussagen machen zu können, gäbe es beispielsweise die Möglichkeit, ES mit einem Fluorophor zu markieren. Die Rate der ES-Diffusion könnte man dann in einer Einzelzell-Messung direkt bestimmen. Ergänzend würde eine quantitative Bestimmung der mRNA-Mengen eine direkte Analyse der Induktion mittels ES und Galactose ermöglichen. Unser Modell schlägt also einen biologisch sinnvollen Mechanismus vor, der die Form der entstehenden Verteilung erklären kann und eröffnet damit verschiedene Möglichkeiten der weiteren Diskussion und Analyse des scRNA-CRISPR/dCas9-Systems.

Im zweiten Teil der Arbeit wurde ein generelles Problem der Synthetischen Biologie, nämlich die sogenannten Kontexteffekte, behandelt. Immer wenn ein künstlich erzeugter genetischer Schaltkreis in eine vorhandene Zelle integriert wird, befindet sich der Schaltkreis in einer mit ihm wechselwirkenden Umgebung, dem sogenannten Kontext. Während diese Effekte bisher nicht zufriedenstellend in vorhandene Modelle integriert werden konnten, haben wir für lineare Umgebungen die spezifischen Memory-Terme exakt berechnet. Anhand zweier Beispiele konnten wir zeigen, welche unterschiedlichen Auswirkungen der Kontext auf die Funktionalität von Genregulationsnetzwerken haben kann. So verändert sich beispielsweise die Schwingungsfrequenz eines genetischen Oszillators, sobald ein beteiligtes Protein reversibel an einen Promotor binden kann. Im Gegensatz dazu beginnt ein System mit negativer Rückkopplung zu oszillieren, sobald eine bestimmte Reaktionskaskade der Umgebung mit berücksichtigt wird. Unsere Analyse beschränkte sich auf Umgebungen, welche durch monomolekulare Reaktionen dargestellt werden können, was als Einschränkung aufgefasst wurde. Mittlerweile haben Briat et al. aber gezeigt, dass das Verhalten der meisten Umgebungen durch monomolekulare Reaktionsnetzwerke genähert werden kann [99]. Im Hinblick auf die immer komplexeren synthetischen Regulationsnetzwerke

---

ermöglichen unsere analytischen Ergebnisse daher eine schnelle Analyse und Robustifizierung realistischer Schaltkreisentwürfe.

Sogenanntes Burstrauschen wurde im dritten Teil der Arbeit analysiert. Ein eindimensionales System, das sogenannte Schlögl-Modell, wurde dazu durch eine Burst-Reaktion derart erweitert, dass sich die Stärke des stoßartigen Rauschens variieren lässt, ohne die deterministische Beschreibung zu ändern. Wir konnten beobachten, dass Bistabilität durch Veränderung der Burst-Größe zerstört, aber auch erzeugt werden kann. Da Burstrauschen bei Transkriptions- wie auch Translationsprozessen ein häufiges Phänomen ist, leisten unsere Ergebnisse daher einen Beitrag zum besseren Verständnis von z.B. heterogenem Zellverhalten. Die von uns für die Analyse entwickelten Methoden und die erhaltenen Ergebnisse bieten dabei zahlreiche Anknüpfungspunkte für weitere Untersuchungen. So haben wir in unserer Analyse beispielsweise die FPE genutzt, um die Extrema von stationären Verteilungen zu berechnen. Mittlerweile wurde gezeigt, dass unsere Methode so erweitert werden kann, dass damit Phasenportraits stochastischer Systeme dargestellt und neuartige Bifurkationen analysiert werden können [128]. Obwohl unser Modell mit drei Reaktionen sehr einfach erscheint, zeigt es eine nicht intuitiv erwartete, durch Burstrauschen induzierte Bimodalität. Es dient daher auch dazu, für generelle Auswirkungen von stochastischem Rauschen zu sensibilisieren [153].

Motiviert durch das Schlögl-Modell wurde im abschließenden Teil der Arbeit noch diskutiert, ob – und wenn ja, wie – theoretische Minimalmodelle einen Nutzen in der Synthetischen Biologie entfalten können. Vergleicht man Modelle aus Ingenieur- und Naturwissenschaften miteinander, stellt man fest, dass diese oftmals auf Basis sehr unterschiedlicher epistemischer Überzeugungen entwickelt wurden. In der Synthetischen Biologie kooperieren nun Forschende aus genau diesen, in ihren wissenschaftlichen Konzepten oftmals konträren Wissenschaften. Theoretische Minimalmodelle beinhalten meist starke Näherungen und Abstraktionen, weshalb sie oft als unbiologisch und nicht relevant eingestuft werden. Häufig wird daher angezweifelt, dass theoretische Modelle wie das Schlögl-Modell in der Synthetischen Biologie einen Nutzen entfalten können. In der Synthetischen Biologie werden allerdings Methoden und Verfahren entwickelt, um neuartige biologische Systeme zu konstruieren. Minimalmodelle zeigen dabei auf, welche wenigen Komponenten zwingend notwendig sind, um eine bestimmte Funktionalität zu erreichen. Wir kommen in unserer Diskussion daher zu dem Schluss, dass gerade die Synthetische Biologie die idealen Voraussetzungen für eine Symbiose von theoretischen Naturwissenschaften und eher experimentell bezogenen Ingenieurwissenschaften bietet. Interdisziplinäre Forschung wird zunehmend wichtiger [154], verlangt von den Beteiligten aber Mut, wissenschaftliche Grenzen zu überschreiten [155]. Die Ergebnisse unserer Diskussion – wie auch der ganzen Arbeit – zeigen, dass die Synthetische Biologie sehr von den Methoden und Modellen der theoretischen Physik profitieren kann und motivieren zu interdisziplinärer Forschung.

---

## 8 Danksagung

Am Ende einer solchen Arbeit gilt es Danke zu sagen. Mein besonderer Dank gilt hierbei Barbara Drossel. Als Bewerber von einer anderen Universität hat sie mich herzlich in ihrer Arbeitsgruppe aufgenommen. Ich danke ihr sehr für die intensive Betreuung, die eine perfekte Balance zwischen angeleitetem und selbst-angeleitetem Forschen ermöglichte. Ich hätte mir keine bessere Betreuung vorstellen können. Außerdem bin ich natürlich dafür dankbar, dass sie mir die Arbeit im interdisziplinären CompuGene-Projekt ermöglichte, welches oftmals auch viel Zeit für organisatorische Arbeiten verschlang.

Weiterhin möchte ich Harald Kolmar danken, der bereitwillig das Zweitgutachten meiner Arbeit übernommen hat. Es war mir im letzten Jahr eine große Freude, zusammen mit ihm und Mitgliedern der AG Kolmar am CompuGene-Paper zum CRISPR/dCas9-Schalter zu arbeiten.

Ein wichtiger Faktor für produktives Arbeiten ist immer auch eine gute Arbeitsatmosphäre. Besser als in der AG Drossel und im CompuGene-Modelling-Lab kann es dabei nicht sein. Dementsprechend gilt mein Dank allen aktuellen und teilweise auch ehemaligen Mitgliedern der AG Drossel sowie den beteiligten Professor\*innen und Doktorand\*innen aus CompuGene. Einige Personen verdienen hierbei besondere Erwähnung (in mehr oder weniger zufälliger Reihenfolge):

**Marc Mandler** Als ich gerade mit meiner Doktorarbeit begonnen hatte, kam Marc als Masterand in die AG Drossel. Seither haben wir den Weg der stochastischen Modellierung gemeinsam bestritten. Durch viele gemeinsame Diskussionen hat Marc ganz entscheidend zu meinem Verständnis von stochastischen Systemen beigetragen. Dementsprechend hat er (implizit) auch großen Anteil an der Entstehung dieser Arbeit. Vielen Dank auch für das Korrekturlesen.

**Lara Becker** Während ihrer Bachelorarbeit noch auf »Irrwegen«, hat sie sich für die Master- und nun auch für ihre Doktorarbeit den stochastischen Effekten verschrieben. Lara hat ein ungefähr unendliches Wissen über aktuelle Literatur zu stochastischer Modellierung und vielen anderen Themen. Diskussionen mit ihr waren daher immer sehr hilfreich und konstruktiv. Außerdem hat Lara die Arbeit Korrektur gelesen.

**Christopher Priester** Während meines ersten Jahres in der AG Drossel teilte ich mein Büro mit »Mifo«. Durch seine Erfahrung und die Übersicht über die verschiedenen Themen, welche damals in der AG behandelt wurden, konnte ich im ersten Jahr viel von ihm lernen und später auf diesem Fundament aufbauen.

**Anja Hofmann** Als Biologin meines Vertrauens musste Anja all meine – gerade zu Beginn sehr dümmlichen – Fragen zu Hefezellen und genetischen Schaltern beantworten. Immer hat sie dies mit Geduld und auf physikerverständlichem Niveau getan. Außerdem hat sie Teile der Arbeit Probe gelesen. Vielen Dank dafür!

**Thomas Zoll und Niels Schlichting** Als Mitglieder des Modelling-Labs waren die beiden »Kabischler« meine direkten (und häufigen) Ansprechpartner, wenn ich mal wieder ein Verständnisproblem mit der Biologie hatte. Obwohl Thomas und Niels mit ihren Projekten eigentlich genug zu tun haben, habe ich dennoch immer eine ausführliche und durchdachte Antwort bekommen. Zusammen mit Eva Moldenhauer haben sie außerdem den Biologie-Teil dieser Arbeit Probe gelesen. Danke!

**Maleen Hanst** Zusammen mit Maleen durfte ich an den Kontexteffekten arbeiten. Als exzellente Mathematikerin hat Maleen mich mit ihren Gedankengängen oft abgehängt oder meine

---

---

»Beweise« entkräftet. Immer hat sie es dabei aber geschafft, mir anschaulich aufzuzeigen, wo meine mathematischen oder logischen Fehler lagen. Vielen Dank auch für das Probelesen des Theorie-Teils.

**Tim Prangemeier** Als studierter Maschinenbauer war Tim selten direkt in meine Forschung involviert. Auf seine fundierte Meinung und seinen »Blick von außen« konnte ich mich allerdings stets verlassen. Als Aborigine war Tim mir außerdem oft eine Hilfe, wenn es mit der englischen Sprache haperte.

**Thorsten Kohl** Zusammen mit Thorsten habe ich mich während meiner Zeit als Doktorand mit philosophischen Fragen beschäftigt. Thorsten war dabei allerdings nicht nur ein »Mitdenker«, sondern ein steter Ansprechpartner, auch für nichtphilosophische Themen. Gerne erinnere ich mich zum Beispiel an die Führung durch das Technikmuseum in Berlin. Außerdem hat Thorsten Teile der Arbeit Korrektur gelesen.

**Heinz Koepl** Als Mitinitiator von CompuGene war Heinz bei vielen Entscheidungen und Fragen zu meiner Forschung beteiligt. Seine Expertise im Bereich der Modellierung kam meiner Arbeit dabei auf vielerlei Weise zugute.

**Leo Bronstein** Leo war zusammen mit Maleen an der Analyse der Kontexteffekte beteiligt. Leos analytische Fähigkeiten sowie seine akribische Arbeitsweise haben dabei entscheidend zum finalen Paper beigetragen.

**Brigitte Held** Als wissenschaftliche Koordinatorin von CompuGene war Brigitte für mich Ansprechpartnerin für jegliche organisatorische Fragen. Ohne ihre Mithilfe wären viele Dinge (v. a. das Symposium) nicht zustande gekommen. Da sie ihre hilfreichen Dienste meist im Hintergrund verrichtet, verdient sie es, hier explizit genannt zu werden.

Außerdem gab es da noch einige motivierte, talentierte Bacheloranden, welche ich betreuen durfte und welche zum Entstehen dieser Arbeit ebenfalls den ein oder anderen Teil beigetragen haben: Christine Buck, Jan Heinz, Jonas Müller, Florian Swienty.

Abschließend möchte ich noch meinen Eltern und meinem Bruder danken, welche mich von klein auf geprägt und damit die Grundlage für all das gelegt haben, was ich nun bin, kann und mache. Außerdem geht ein herzlicher Dank an Tamara und Leonard. Gerade in den letzten Wochen war die Doktorarbeit ein zeitfressendes Monster, das oftmals Freizeitaktivitäten im Weg stand. Aber auch vorher schon schwirrte mir des Öfteren die Fokker-Planck-Gleichung im Kopf umher, wenn ich eigentlich die Monopoly-Figur weiter ziehen sollte. Beide haben sich dennoch nie beschwert und mir stattdessen in vielen Bereichen den Rücken frei gehalten.

Danke!



---

---

# Literatur

- [1] Stephen Hawking. *The Universe In A Nutshell*. Englisch. New York: Bantam Press, 2001. ISBN: 978-0-553-80202-3.
- [2] Maureen A. O'Malley. *Making Knowledge in Synthetic Biology: Design Meets Kludge*. In: *Biological Theory* 4.4 (Dez. 2009), S. 378–389. ISSN: 1555-5550. DOI: 10.1162/BIOT\_a\_00006.
- [3] Fred Boogerd u. a. *Systems Biology: Philosophical Foundations*. en. Elsevier, 2007. ISBN: 978-0-08-047527-1.
- [4] David Sprinzak und Michael B. Elowitz. *Reconstruction of genetic circuits*. en. In: *Nature* 438.7067 (2005), S. 443–448. ISSN: 1476-4687. DOI: 10.1038/nature04335.
- [5] Domitilla Del Vecchio, Alexander J Ninfa und Eduardo D Sontag. *Modular cell biology: retroactivity and insulation*. In: *Molecular Systems Biology* 4 (2008), S. 161. ISSN: 1744-4292. DOI: 10.1038/msb4100204.
- [6] Kyung Hyuk Kim, Deepak Chandran und Herbert M. Sauro. *Toward Modularity in Synthetic Biology: Design Patterns and Fan-out*. en. In: *Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology*. Hrsg. von Heinz Koepl u. a. New York, NY: Springer New York, 2011, S. 117–138. ISBN: 978-1-4419-6766-4. DOI: 10.1007/978-1-4419-6766-4\_6.
- [7] Herbert M Sauro. *Modularity defined*. In: *Molecular Systems Biology* 4 (2008), S. 166. ISSN: 1744-4292. DOI: 10.1038/msb.2008.3.
- [8] Donal Stewart und John Roger Wilson-Kanamori. *Modular Modelling in Synthetic Biology: Light-Based Communication in E. coli*. In: *Electronic Notes in Theoretical Computer Science*. The Second International Workshop on Interactions between Computer Science and Biology 277 (2011), S. 77–87. ISSN: 1571-0661. DOI: 10.1016/j.entcs.2011.09.037.
- [9] Alfred Nordmann. *Synthetic Biology at the Limits of Science*. en. In: *Synthetic Biology. Character and Impact*. Hrsg. von Bernd Giese u. a. Berlin: Springer, 2015, S. 31–58.
- [10] Tamás Székely und Kevin Burrage. *Stochastic simulation in systems biology*. In: *Computational and Structural Biotechnology Journal* 12.20 (2014), S. 14–25. ISSN: 2001-0370. DOI: 10.1016/j.csbj.2014.10.003.
- [11] Natalja Strelkowa. *Stochastic Complexity Analysis in Synthetic Biology*. en. In: *How Nature Works: Complexity in Interdisciplinary Research and Applications*. Hrsg. von Ivan Zelinka u. a. Emergence, Complexity and Computation. Heidelberg: Springer International Publishing, 2014, S. 161–194. ISBN: 978-3-319-00254-5. DOI: 10.1007/978-3-319-00254-5\_8.
- [12] Johan Paulsson, Otto G. Berg und Måns Ehrenberg. *Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation*. en. In: *Proceedings of the National Academy of Sciences* 97.13 (Juni 2000), S. 7148–7153. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.110057697.
- [13] Yaneer Bar-yam. *Dynamics Of Complex Systems*. Englisch. Reading, Mass: Westview Press, 1997. ISBN: 978-0-201-55748-0.
- [14] Jordi Garcia-Ojalvo. *Physical approaches to the dynamics of genetic circuits: a tutorial*. In: *Contemporary Physics* 52.5 (2011), S. 439–464. ISSN: 0010-7514. DOI: 10.1080/00107514.2011.588432.

- 
- [15] Eva Bianconi u. a. *An estimation of the number of cells in the human body*. eng. In: *Annals of Human Biology* 40.6 (2013), S. 463–471. ISSN: 1464-5033. DOI: 10.3109/03014460.2013.807878.
- [16] Domitilla Del Vecchio und E. D. Sontag. *Dynamics and Control of Synthetic Bio-molecular Networks*. In: *2007 American Control Conference* (2007), S. 1577–1588.
- [17] Sabine Schmitz und Christine Desel. *Der Experimentator Zellbiologie*. en. Experimentator. Springer Spektrum, 2018. ISBN: 978-3-662-56110-2.
- [18] C. R. Woese, O. Kandler und M. L. Wheelis. *Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya*. en. In: *Proceedings of the National Academy of Sciences* 87.12 (1990), S. 4576–4579. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.87.12.4576.
- [19] Ana Paula Oliveira, Michael C. Jewett und Jens Nielsen. *From Gene Expression to Metabolic Fluxes*. In: *Introduction to Systems Biology*. Hrsg. von Sangdun Choi. Totowa, NJ: Humana Press, 2007, S. 37–66. ISBN: 978-1-59745-531-2. DOI: 10.1007/978-1-59745-531-2\_3.
- [20] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Englisch. 1. Aufl. Boca Raton, FL: Taylor & Francis Inc, 2006. ISBN: 978-1-58488-642-6.
- [21] Ozlem Demir und Isil Aksan Kurnaz. *An integrated model of glucose and galactose metabolism regulated by the GAL genetic switch*. en. In: *Computational Biology and Chemistry* 30.3 (2006), S. 179–192. ISSN: 14769271. DOI: 10.1016/j.compbiolchem.2006.02.004.
- [22] Michaela Conrad u. a. *Nutrient sensing and signaling in the yeast Saccharomyces cerevisiae*. en. In: *FEMS Microbiology Reviews* 38.2 (2014), S. 254–299. ISSN: 1574-6976. DOI: 10.1111/1574-6976.12065.
- [23] M. P. Crosland M. Sc Ph.D. *The use of diagrams as chemical ‘equations’ in the lecture notes of William Cullen and Joseph Black*. In: *Annals of Science* 15.2 (1959), S. 75–90. ISSN: 0003-3790. DOI: 10.1080/00033795900200088.
- [24] A. D. McNaught und A. Wilkinson. *Compendium of Chemical Terminology*. Englisch. 2 edition. Oxford England ; Malden, MA, USA: Wiley, 1997. ISBN: 978-0-86542-684-9.
- [25] Daniel T. Gillespie. *Exact stochastic simulation of coupled chemical reactions*. In: *The Journal of Physical Chemistry* 81.25 (1977), S. 2340–2361. ISSN: 0022-3654. DOI: 10.1021/j100540a008.
- [26] Daniel T Gillespie. *A rigorous derivation of the chemical master equation*. In: *Physica A: Statistical Mechanics and its Applications* 188.1 (1992), S. 404–425.
- [27] Daniel T. Gillespie. *Stochastic Simulation of Chemical Kinetics*. en. In: *Annual Review of Physical Chemistry* 58.1 (2007), S. 35–55. ISSN: 0066-426X, 1545-1593. DOI: 10.1146/annurev.physchem.58.032806.104637.
- [28] Michael A. Gibson und Jehoshua Bruck. *Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels*. In: *The Journal of Physical Chemistry A* 104.9 (2000), S. 1876–1889. ISSN: 1089-5639. DOI: 10.1021/jp993732q.
- [29] Yang Cao, Hong Li und Linda Petzold. *Efficient formulation of the stochastic simulation algorithm for chemically reacting systems*. eng. In: *The Journal of Chemical Physics* 121.9 (2004), S. 4059–4067. ISSN: 0021-9606. DOI: 10.1063/1.1778376.
- [30] Crispin Gardiner. *Stochastic Methods - A Handbook for the Natural and Social*. Springer, 2009.
- [31] Ludwig Arnold. *Random Dynamical Systems*. en. Springer Monographs in Mathematics. Berlin Heidelberg: Springer-Verlag, 1998. ISBN: 978-3-540-63758-5.

- [32] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry (Third Edition)*. Hrsg. von N.G. VAN KAMPEN. Third Edition. North-Holland Personal Library. Amsterdam: Elsevier, 2007.
- [33] Tudor Udrescu. *Numerical Methods for the Chemical Master Equation*. Englisch. PhD Thesis. Karlsruhe, 2012.
- [34] Norbert Henze. *Stochastik für Einsteiger: Eine Einführung in die faszinierende Welt des Zufalls*. en. 10. Aufl. Springer Spektrum, 2013. ISBN: 978-3-658-03077-3.
- [35] Matthew Scott. *Applied Stochastic Processes*. Lecture Notes. 2013.
- [36] Paola Lecca. *Stochastic chemical kinetics*. In: *Biophysical Reviews* 5.4 (2013), S. 323–345. ISSN: 1867-2450. DOI: 10.1007/s12551-013-0122-2.
- [37] Professor Dr. Hannes Risken. *The Fokker-Planck Equation*. Springer Series in Synergetics. Springer Berlin Heidelberg, 1984.
- [38] C. Lange. Honerkamp, J., *Stochastische dynamische Systeme. Konzepte, numerische Methoden, Datenanalysen*. Weinheim, VCH Verlagsgesellschaft 1990. XIII, 340 S., DM 98.—. ISBN 3-527-27945-8. en. In: *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik* 73.1 (1993), S. 46–46. ISSN: 1521-4001. DOI: 10.1002/zamm.19930730113.
- [39] Tommaso Biancalani. *The Influence of Demographic Stochasticity on Population Dynamics: A Mathematical Study of Noise-Induced Bistable States and Stochastic Patterns*. en. Springer Theses. Springer International Publishing, 2014. ISBN: 978-3-319-07727-7.
- [40] Daniel T. Gillespie. *The chemical Langevin equation*. en. In: *The Journal of Chemical Physics* 113.1 (2000), S. 297–306. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.481811.
- [41] K. R. Sanft, D. T. Gillespie und L. R. Petzold. *Legitimacy of the stochastic Michaelis-Menten approximation*. eng. In: *IET systems biology* 5.1 (2011), S. 58. ISSN: 1751-8849. DOI: 10.1049/iet-syb.2009.0057.
- [42] Patrick Smadbeck und Yiannis Kaznessis. *Stochastic model reduction using a modified Hill-type kinetic rate law*. In: *The Journal of Chemical Physics* 137.23 (2012). ISSN: 0021-9606. DOI: 10.1063/1.4770273.
- [43] Anja Hofmann u. a. *A tightly regulated and adjustable CRISPR-dCas9 based AND gate in yeast*. In: *Nucleic Acids Research* 47.1 (2019), S. 509–520. DOI: 10.1093/nar/gky1191. eprint: /oup/backfile/content\_public/journal/nar/47/1/10.1093\_nar\_gky1191/1/gky1191.pdf.
- [44] Ana Traven, Branka Jelicic und Mary Sopta. *Yeast Gal4: a transcriptional paradigm revisited*. In: *EMBO Rep.* 7.5 (2006), S. 496–499. ISSN: 1469-221X. DOI: 10.1038/sj.embor.7400679.
- [45] Diana S M Ottoz, Fabian Rudolf und Joerg Stelling. *Inducible, tightly regulated and growth condition-independent transcription factor in Saccharomyces cerevisiae*. In: *Nucleic Acids Res.* 42.17 (2014), e130. DOI: 10.1093/nar/gku616.
- [46] Nancy A Da Silva und Sneha Srikrishnan. *Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae*. In: *FEMS Yeast Res.* 12.2 (2012), S. 197–214. ISSN: 1567-1364. DOI: 10.1111/j.1567-1364.2011.00769.x.
- [47] J. S. Flick und M. Johnston. *Two systems of glucose repression of the GAL1 promoter in Saccharomyces cerevisiae*. In: *Mol. Cell Biol.* 10.9 (1990), S. 4757–4769. ISSN: 0270-7306. DOI: 10.1128/MCB.10.9.4757.

- 
- [48] Gerald Thiel und Michael Lietz. *Regulator neuronaler Gene: Zinkfingerprotein REST*. de. In: *Biologie in unserer Zeit* 34.2 (2004), S. 96–101. ISSN: 1521-415X. DOI: 10.1002/biuz.200410244.
- [49] Thomas Gaj, Charles A. Gersbach und Carlos F. Barbas. *ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering*. English. In: *Trends Biotechnol.* 31.7 (2013), S. 397–405. ISSN: 0167-7799, 1879-3096. DOI: 10.1016/j.tibtech.2013.04.004.
- [50] Michael K. Jensen und Jay D. Keasling. *Recent applications of synthetic biology tools for yeast metabolic engineering*. In: *FEMS Yeast Res.* 15.1 (2015), S. 1–10. ISSN: 1567-1364. DOI: 10.1111/1567-1364.12185.
- [51] Elke Nevoigt. *Progress in metabolic engineering of Saccharomyces cerevisiae*. In: *Microbiol. Mol. Biol. Rev.* 72.3 (2008), S. 379–412. ISSN: 1098-5557. DOI: 10.1128/MMBR.00025-07.
- [52] Lei S. Qi u. a. *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression*. In: *Cell* 152.5 (2013), S. 1173–1183. ISSN: 1097-4172. DOI: 10.1016/j.cell.2013.02.022.
- [53] Prashant Mali u. a. *CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering*. In: *Nat. Biotechnol.* 31.9 (2013), S. 833–838. ISSN: 10870156. DOI: 10.1038/nbt.2675.
- [54] Jesse G. Zalatan u. a. *Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds*. In: *Cell* 160.1-2 (2015), S. 339–350. ISSN: 1097-4172. DOI: 10.1016/j.cell.2014.11.052.
- [55] Y. Ishino u. a. *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product*. eng. In: *Journal of Bacteriology* 169.12 (1987), S. 5429–5433. ISSN: 0021-9193.
- [56] C. Pourcel, G. Salvignol und G. Vergnaud. *CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies*. eng. In: *Microbiology (Reading, England)* 151.Pt 3 (2005), S. 653–663. ISSN: 1350-0872. DOI: 10.1099/mic.0.27437-0.
- [57] Alexander Bolotin u. a. *Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin*. eng. In: *Microbiology (Reading, England)* 151.Pt 8 (2005), S. 2551–2561. ISSN: 1350-0872. DOI: 10.1099/mic.0.28048-0.
- [58] Patrick D. Hsu, Eric S. Lander und Feng Zhang. *Development and Applications of CRISPR-Cas9 for Genome Engineering*. In: *Cell* 157.6 (2014), S. 1262–1278.
- [59] Le Cong u. a. *Multiplex Genome Engineering Using CRISPR/Cas Systems*. en. In: *Science* (2013), S. 1231143. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1231143.
- [60] John L. Rinn und Howard Y. Chang. *Genome regulation by long noncoding RNAs*. In: *Annu. Rev. Biochem.* 81 (2012), S. 145–166. ISSN: 1545-4509. DOI: 10.1146/annurev-biochem-051410-092902.
- [61] R. Scott McIsaac u. a. *Fast-acting and nearly gratuitous induction of gene expression and protein depletion in Saccharomyces cerevisiae*. In: *Mol. Biol. Cell* 22.22 (2011), S. 4447–4459. ISSN: 1939-4586. DOI: 10.1091/mbc.E11-05-0466.
- [62] Jean-François Louvion, Biserka Havaux-Copf und Didier Picard. *Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast*. In: *Gene* 131.1 (1993), S. 129–134. ISSN: 0378-1119. DOI: 10.1016/0378-1119(93)90681-R.

- 
- [63] D. M. Ruden u. a. *Generating yeast transcriptional activators containing no yeast protein sequences*. In: *Nature* 350.6315 (1991), S. 250–252. ISSN: 0028-0836. DOI: 10.1038/350250a0.
- [64] Victoria A. Doronina u. a. *Site-specific release of nascent chains from ribosomes at a sense codon*. In: *Mol. Cell Biol.* 28.13 (2008), S. 4227–4239. ISSN: 1098-5549. DOI: 10.1128/MCB.00421-08.
- [65] Vera S. Efimova u. a. *Expression of Cholesterol Hydroxylase/Lyase System Proteins in Yeast *S. cerevisiae* Cells as a Self-Processing Polyprotein*. In: *Mol. Biotechnol.* 59.9-10 (2017), S. 394–406. ISSN: 1559-0305. DOI: 10.1007/s12033-017-0028-5.
- [66] Julius Grzeschik u. a. *A simplified procedure for antibody engineering by yeast surface display: Coupling display levels and target binding by ribosomal skipping*. In: *Biotechnol. J.* 12.2 (2017). ISSN: 1860-7314. DOI: 10.1002/biot.201600454.
- [67] Christopher Schneider u. a. *ROC'n'Ribo: Characterizing a Riboswitching Expression System by Modeling Single-Cell Data*. In: *ACS Synth Biol.* 6.7 (2017), S. 1211–1224. ISSN: 2161-5063. DOI: 10.1021/acssynbio.6b00322.
- [68] Paul Jorgensen u. a. *The Size of the Nucleus Increases as Yeast Cells Grow*. In: *Mol. Biol. Cell* 18.9 (2007), S. 3523–3532. ISSN: 1059-1524. DOI: 10.1091/mbc.E06-10-0973.
- [69] Michael Samoilov, Sergey Plyasunov und Adam P. Arkin. *Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations*. In: *Proceedings of the National Academy of Sciences of the United States of America* 102.7 (2005), S. 2310–2315. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0406841102.
- [70] Johannes Falk, Marc Mendler und Barbara Drossel. *A minimal model of burst-noise induced bistability*. en. In: *PLOS ONE* 12.4 (Apr. 2017), e0176410. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0176410.
- [71] Andrew Duncan u. a. *Noise-induced multistability in chemical systems: Discrete versus continuum modeling*. In: *Phys. Rev. E* 91 (4 Apr. 2015), S. 042111. DOI: 10.1103/PhysRevE.91.042111.
- [72] Vahid Shahrezaei und Peter S. Swain. *Analytical distributions for stochastic gene expression*. en. In: *Proc. Natl. Acad. Sci. USA* 105.45 (2008), S. 17256–17261. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0803850105.
- [73] Stephen Ramsey, David Orrell und Hamid Bolouri. *Dizzy: stochastic simulation of large-scale genetic regulatory networks*. In: *Journal of Bioinformatics and Computational Biology* 3.2 (2005), S. 415–436. ISSN: 0219-7200.
- [74] Hidde de Jong. *Modeling and simulation of genetic regulatory systems: a literature review*. In: *J. Comput. Biol.* 9.1 (2002), S. 67–103. ISSN: 1066-5277. DOI: 10.1089/10665270252833208.
- [75] Allison Hoynes-O'Connor und Tae Seok Moon. *Programmable genetic circuits for pathway engineering*. In: *Curr. Opin. Biotechnol.* 36 (2015), S. 115–121. ISSN: 1879-0429. DOI: 10.1016/j.copbio.2015.08.007.
- [76] Yili Qian und Domitilla Del Vecchio. *Emerging Applications of Control and Systems Theory: The “Power Network” of Genetic Circuits*. In: *Emerging Applications of Control and Systems Theory: A Festschrift in Honor of Mathukumalli Vidyasagar*. Hrsg. von R. Tempo, S. Yurkovich und P. Misra. Springer, 2018, S. 109–121. ISBN: 978-3-319-67067-6. DOI: 10.1007/978-3-319-67068-3\_8.
- [77] Emil D. Jensen u. a. *Transcriptional reprogramming in yeast using dCas9 and combinatorial gRNA strategies*. In: *Microb. Cell Fact.* 16.1 (2017), S. 46. ISSN: 1475-2859. DOI: 10.1186/s12934-017-0664-2.

- 
- [78] Matthew Deaner, Julio Mejia und Hal S. Alper. *Enabling Graded and Large-Scale Multiplex of Desired Genes Using a Dual-Mode dCas9 Activator in Saccharomyces cerevisiae*. In: *ACS Synth. Biol.* 6.10 (2017), S. 1931–1943. ISSN: 2161-5063. DOI: 10.1021/acssynbio.7b00163.
- [79] P Hovland u. a. *Galactose as a gratuitous inducer of GAL gene expression in yeasts growing on glucose*. In: *Gene* 83.1 (1989), S. 57–64. ISSN: 0378-1119.
- [80] Johannes Falk u. a. *Context in Synthetic Biology: Memory Effects of Environments with Mono-molecular Reactions*. In: *arXiv:1809.09555 [cond-mat, physics:physics, q-bio]* (2018). arXiv: 1809.09555.
- [81] Reshma P Shetty, Drew Endy und Thomas F Knight. *Engineering BioBrick vectors from BioBrick parts*. In: *Journal of Biological Engineering* 2 (2008), S. 5. ISSN: 1754-1611. DOI: 10.1186/1754-1611-2-5.
- [82] Ernesto Andrianantoandro u. a. *Synthetic biology: new engineering rules for an emerging discipline*. en. In: *Molecular Systems Biology* 2.1 (2006), S. 2006.0028. ISSN: 1744-4292, 1744-4292. DOI: 10.1038/msb4100073.
- [83] Priscilla E. M. Purnick und Ron Weiss. *The second wave of synthetic biology: from modules to systems*. En. In: *Nature Reviews Molecular Cell Biology* 10.6 (2009), S. 410. ISSN: 1471-0080. DOI: 10.1038/nrm2698.
- [84] Stefano Cardinale und Adam Paul Arkin. *Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems*. In: *Biotechnology Journal* 7.7 (2012), S. 856–866. ISSN: 1860-6768. DOI: 10.1002/biot.201200085.
- [85] J. Saez-Rodriguez, A. Kremling und E. D. Gilles. *Dissecting the puzzle of life: modularization of signal transduction networks*. In: *Computers & Chemical Engineering. Computational Challenges in Biology* 29.3 (2005), S. 619–629. ISSN: 0098-1354. DOI: 10.1016/j.compchemeng.2004.08.035.
- [86] Wolfram Liebermeister, Ulrike Baur und Edda Klipp. *Biochemical network models simplified by balanced truncation*. eng. In: *The FEBS journal* 272.16 (2005), S. 4034–4043. ISSN: 1742-464X. DOI: 10.1111/j.1742-4658.2005.04780.x.
- [87] Katy J. Rubin u. a. *Memory effects in biochemical networks as the natural counterpart of extrinsic noise*. In: *Journal of Theoretical Biology* 357 (2014), S. 245–267. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2014.06.002.
- [88] Zhixing Cao und Ramon Grima. *Linear mapping approximation of gene regulatory networks with stochastic dynamics*. en. In: *Nature Communications* 9.1 (2018), S. 3305. ISSN: 2041-1723. DOI: 10.1038/s41467-018-05822-0.
- [89] Barbara Drossel. *Komplexe Dynamische Systeme*. Vorlesungsskript. Wintersemester 2009/10.
- [90] Harro Heuser. *Gewöhnliche Differentialgleichungen: Einführung in Lehre und Gebrauch*. Deutsch. 6. Aufl. Wiesbaden: Vieweg+Teubner Verlag, 2009. ISBN: 978-3-8348-0705-2.
- [91] Nobuhiko Tokuriki und Dan S. Tawfik. *Protein Dynamism and Evolvability*. en. In: *Science* 324.5924 (2009), S. 203–207. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1169375.
- [92] J. H. Hofmeyr, H. Kacser und K. J. van der Merwe. *Metabolic control analysis of moiety-conserved cycles*. eng. In: *European Journal of Biochemistry* 155.3 (1986), S. 631–641. ISSN: 0014-2956.

- [93] Thomas Schwander u. a. *A synthetic pathway for the fixation of carbon dioxide in vitro*. eng. In: *Science (New York, N.Y.)* 354.6314 (2016), S. 900–904. ISSN: 1095-9203. DOI: 10.1126/science.aah5237.
- [94] Katharina Semrad. *Proteins with RNA Chaperone Activity: A World of Diverse Proteins with a Common Task—Impediment of RNA Misfolding*. In: *Biochemistry Research International* 2011 (2011). ISSN: 2090-2247. DOI: 10.1155/2011/532908.
- [95] Tobias Jahnke und Wilhelm Huisinga. *Solving the chemical master equation for monomolecular reaction systems analytically*. eng. In: *Journal of Mathematical Biology* 54.1 (2007), S. 1–26. ISSN: 0303-6812. DOI: 10.1007/s00285-006-0034-x.
- [96] C. R. Heathcote und J. E. Moyal. *The Random Walk [In Continuous Time] and Its Application to the Theory of Queues*. In: *Biometrika* 46.3/4 (1959), S. 400–411. ISSN: 0006-3444. DOI: 10.2307/2333536.
- [97] Michael B Elowitz und Stanislas Leibler. *A synthetic oscillatory network of transcriptional regulators*. In: *Nature* 403.6767 (2000), S. 335.
- [98] Dmitri Bratsun u. a. *Delay-induced stochastic oscillations in gene regulation*. en. In: *Proceedings of the National Academy of Sciences* 102.41 (2005), S. 14593–14598. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0503858102.
- [99] Corentin Briat und Mustafa Khammash. *Ergodicity analysis and antithetic integral control of a class of stochastic reaction networks with delays*. In: *arXiv:1811.09188 [cs, math, q-bio]* (2018). arXiv: 1811.09188.
- [100] John Guckenheimer und P. J. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. New York: Springer, 2002. 453 S. ISBN: 978-0-387-90819-9.
- [101] Werner Horsthemke. *Noise Induced Transitions*. Hrsg. von Professeur Dr Christian Vidal und Professeur Dr Adolphe Pacault. Springer Series in Synergetics. DOI: 10.1007/978-3-642-70196-2\_23. Springer Berlin Heidelberg, 1984.
- [102] Chao Du und Wing Hong Wong. *Stochastic Modeling and Statistical Inference of Intrinsic Noise in Gene Regulation System via Chemical Master Equation*. In: *arXiv:1610.07213 [q-bio, stat]* (2016). arXiv: 1610.07213.
- [103] G. Nebe-von-Caron u. a. *Analysis of bacterial function by multi-colour fluorescence flow cytometry and single cell sorting*. English. In: *Journal of Microbiological Methods* 42.1 (Sep. 2000), S. 97–114. ISSN: 0167-7012. DOI: 10.1016/S0167-7012(00)00181-0.
- [104] David N. Breslauer, Philip J. Lee und Luke P. Lee. *Microfluidics-based systems biology*. en. In: *Mol. BioSyst.* 2.2 (Jan. 2006), S. 97–112. ISSN: 1742-2051. DOI: 10.1039/B515632G.
- [105] Tommaso Biancalani, Louise Dyson und Alan J. McKane. *Noise-Induced Bistable States and Their Mean Switching Time in Foraging Colonies*. In: *Phys. Rev. Lett.* 112 (3 Jan. 2014), S. 038101. DOI: 10.1103/PhysRevLett.112.038101.
- [106] J. Paulsson und M. Ehrenberg. *Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks*. In: *Physical Review Letters* 84.23 (2000), S. 5447–5450. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.84.5447.
- [107] K. Wiesenfeld und F. Moss. *Stochastic Resonance and the Benefits of Noise - from Ice Ages to Crayfish and Squids*. English. In: *Nature* 373.6509 (Jan. 1995), S. 33–36. ISSN: 0028-0836. DOI: 10.1038/373033a0.
- [108] Arjun Raj und Alexander van Oudenaarden. *Stochastic gene expression and its consequences*. In: *Cell* 135.2 (Okt. 2008), S. 216–226. ISSN: 0092-8674. DOI: 10.1016/j.cell.2008.09.050.

- 
- [109] Nir Friedman, Long Cai und X. Sunney Xie. *Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression*. In: *Physical Review Letters* 97.16 (2006), S. 168302. DOI: 10.1103/PhysRevLett.97.168302.
- [110] Pavol Bokes und Abhyudai Singh. *Gene expression noise is affected differentially by feedback in burst frequency and burst size*. In: *arXiv:1602.08568 [q-bio]* (2016). arXiv: 1602.08568.
- [111] Lev S. Tsimring. *Noise in Biology*. In: *Reports on progress in physics. Physical Society (Great Britain)* 77.2 (2014), S. 026601. ISSN: 0034-4885. DOI: 10.1088/0034-4885/77/2/026601.
- [112] R. Grima. *An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions*. eng. In: *The Journal of Chemical Physics* 133.3 (Juli 2010), S. 035101. ISSN: 1089-7690. DOI: 10.1063/1.3454685.
- [113] Matthew Scott, Terence Hwa und Brian Ingalls. *Deterministic characterization of stochastic genetic circuits*. In: *Proceedings of the National Academy of Sciences* 104.18 (2007), S. 7402–7407. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0610468104.
- [114] John K. McSweeney und Lea Popovic. *Stochastically-Induced Bistability in Chemical Reaction Systems*. In: *Annals of Applied Probability* 24.3 (2014), S. 1226–1268. ISSN: 1050-5164. DOI: 10.1214/13-AAP946.
- [115] Lisa M. Bishop und Hong Qian. *Stochastic Bistability and Bifurcation in a Mesoscopic Signaling System with Autocatalytic Kinase*. In: *Biophysical Journal* 98.1 (2010), S. 1–11. DOI: 10.1016/j.bpj.2009.09.055.
- [116] Rajesh Karmakar und Indrani Bose. *Positive feedback, stochasticity and genetic competence*. In: *Physical Biology* 4.1 (2007), S. 29. ISSN: 1478-3975. DOI: 10.1088/1478-3975/4/1/004.
- [117] Michael Assaf, Elijah Roberts und Zaida Luthey-Schulten. *Determining the Stability of Genetic Switches: Explicitly Accounting for mRNA Noise*. In: *Physical Review Letters* 106.24 (2011), S. 248102. DOI: 10.1103/PhysRevLett.106.248102.
- [118] F. Schlögl. *Chemical reaction models for non-equilibrium phase transitions*. In: *Zeitschrift für Physik* 253.2 (1972), S. 147–161.
- [119] P. Grassberger. *On phase transitions in Schlögl's second model*. In: *Zeitschrift für Physik B Condensed Matter* 47.4 (1982), S. 365–374.
- [120] Radek Erban u. a. *Analysis of a Stochastic Chemical System Close to a SNIPER Bifurcation of Its Mean-Field Model*. en. In: *SIAM Journal on Applied Mathematics* 70.3 (Jan. 2009), S. 984–1016. ISSN: 0036-1399, 1095-712X. DOI: 10.1137/080731360.
- [121] Melissa Vellela und Hong Qian. *Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited*. In: *Journal of The Royal Society Interface* 6.39 (2009), S. 925–940. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2008.0476.
- [122] Pierre Gaspard. *Fluctuation theorem for nonequilibrium reactions*. In: *The Journal of Chemical Physics* 120.19 (2004), S. 8898–8905. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.1688758.
- [123] Thomas Wilhelm. *The smallest chemical reaction system with bistability*. In: *BMC Systems Biology* 3 (2009), S. 90.
- [124] Robert G. Endres. *Bistability: Requirements on Cell-Volume, Protein Diffusion, and Thermodynamics*. In: *Plos One* 10.4 (2015), S. 1–22. ISSN: 1932-6203.



- 
- [125] Yuri A. Kuznetsov. *Elements of Applied Bifurcation Theory (2Nd Ed.)* Berlin, Heidelberg: Springer-Verlag, 1998. ISBN: 0-387-98382-1.
- [126] Ronald S. Irving. *Integers, Polynomials, and Rings: A Course in Algebra*. Springer Science & Business Media, 2003. 310 S. ISBN: 978-0-387-20172-6.
- [127] Daniel T Gillespie. *Markov processes: an introduction for physical scientists*. Elsevier, 1991.
- [128] Marc Mendler, Johannes Falk und Barbara Drossel. *Analysis of stochastic bifurcations with phase portraits*. en. In: *PLOS ONE* 13.4 (Apr. 2018), e0196126. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0196126.
- [129] Shay Be'er und Michael Assaf. *Rare events in stochastic populations under bursty reproduction*. In: *Journal of Statistical Mechanics: Theory and Experiment* 2016.11 (2016), S. 113501.
- [130] K. F. Knott. *Characteristics of burst noise intermittency*. In: *Solid-State Electronics* 21.8 (1978), S. 1039–1043. ISSN: 0038-1101. DOI: 10.1016/0038-1101(78)90182-X.
- [131] Andrzej M. Zaklikiewicz. *Influence of burst noise on noise spectra*. In: *Solid-State Electronics* 24.1 (1981), S. 1–3. ISSN: 0038-1101. DOI: 10.1016/0038-1101(81)90205-7.
- [132] Pavol Bokes u. a. *Transcriptional Bursting Diversifies the Behaviour of a Toggle Switch: Hybrid Simulation of Stochastic Gene Expression*. In: *Bulletin of Mathematical Biology* 75.2 (2013), S. 351–371. ISSN: 0092-8240, 1522-9602. DOI: 10.1007/s11538-013-9811-z.
- [133] Yen Ting Lin und Tobias Galla. *Bursting noise in gene expression dynamics: linking microscopic and mesoscopic models*. In: *Journal of The Royal Society Interface* 13.114 (2016), S. 20150772. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2015.0772.
- [134] Philipp Thomas u. a. *Signatures of nonlinearity in single cell noise-induced oscillations*. In: *Journal of Theoretical Biology* 335 (2013), S. 222–234. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2013.06.021.
- [135] Andrew J. Black und Alan J. McKane. *Stochastic formulation of ecological models and their applications*. In: *Trends in Ecology & Evolution* 27.6 (2012-06), S. 337–345. ISSN: 0169-5347. DOI: 10.1016/j.tree.2012.01.014.
- [136] Hong Qian, Saveez Saffarian und Elliot L. Elson. *Concentration fluctuations in a mesoscopic oscillating chemical reaction system*. In: *Proceedings of the National Academy of Sciences* 99.16 (2002), S. 10376–10381. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.152007599.
- [137] Rajesh Ramaswamy und Ivo F. Sbalzarini. *Intrinsic noise alters the frequency spectrum of mesoscopic oscillatory chemical reaction systems*. In: *Scientific Reports* 1 (2011), S. 154. ISSN: 2045-2322. DOI: 10.1038/srep00154.
- [138] George E. P. Box. *Science and Statistics*. In: *Journal of the American Statistical Association* 71.356 (1976), S. 791–799. ISSN: 0162-1459. DOI: 10.1080/01621459.1976.10480949.
- [139] Sara Green. *Philosophy of Systems and Synthetic Biology*. In: *The Stanford Encyclopedia of Philosophy*. Hrsg. von Edward N. Zalta. Fall 2018. Metaphysics Research Lab, Stanford University, 2018.
- [140] Robert W. Batterman. *Asymptotics and the Role of Minimal Models*. In: *The British Journal for the Philosophy of Science* 53.1 (2002), S. 21–38. ISSN: 0007-0882.
- [141] Eran Tal. *Enhancing Knowledge, Affording Ignorance*. Darmstadt, 2013.
- [142] Nancy Cartwright. *How the Laws of Physics Lie*. Englisch. Oxford : New York: Oxford University Press, USA, 1983. ISBN: 978-0-19-824704-3.

- 
- [143] Margaret Morrison. *Reconstructing Reality: Models, Mathematics, and Simulations*. Oxford Studies in the Philosophy of Science. Oxford, New York: Oxford University Press, 2015. ISBN: 978-0-19-938027-5.
- [144] William C. Wimsatt. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. English. Cambridge, Mass: Harvard University Press, 2007. ISBN: 978-0-674-01545-6.
- [145] George E. P. Box und Norman R. Draper. *Empirical Model-Building and Response Surfaces*. English. 1 edition. New York: Wiley, 1987. ISBN: 978-0-471-81033-9.
- [146] Nicole E. Radde und Marc-Thorsten Hütt. *The Physics behind Systems Biology*. In: *EPJ Nonlinear Biomedical Physics* 4.1 (2016), S. 7. ISSN: 2195-0008. DOI: 10.1140/epjnbp/s40366-016-0034-8.
- [147] Robert W. Batterman und Collin C. Rice. *Minimal Model Explanations*. In: *Philosophy of Science* 81.3 (2014), S. 349–376. ISSN: 0031-8248. DOI: 10.1086/676677.
- [148] Stephan Hartmann. *Models as a Tool for Theory Construction: Some Strategies of Preliminary Physics*. en. Preprint. 2005. DOI: <http://philsci-archive.pitt.edu/2410/1/Models.pdf>.
- [149] Sorin Tănase-Nicola und David K. Lubensky. *Exchange of stability as a function of system size in a nonequilibrium system*. eng. In: *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics* 86.4 Pt 1 (2012), S. 040103. ISSN: 1550-2376. DOI: 10.1103/PhysRevE.86.040103.
- [150] Peter Hanggi u. a. *Bistable systems: Master equation versus Fokker-Planck modeling*. In: *Phys. Rev. A* 29 (1 Jan. 1984), S. 371–378. DOI: 10.1103/PhysRevA.29.371.
- [151] Jane Calvert. *Synthetic biology: constructing nature?* In: *The Sociological Review* 58.1\_suppl (2010), S. 95–112.
- [152] Matthias Heinemann und Sven Panke. *Synthetic biology—putting engineering into biology*. en. In: *Bioinformatics* 22.22 (2006), S. 2790–2799. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btl469.
- [153] Vivek Kohar und Mingyang Lu. *Role of noise and parametric variation in the dynamics of gene regulatory circuits*. In: *npj Systems Biology and Applications* 4 (Nov. 2018), S. 40. ISSN: 2056-7189. DOI: 10.1038/s41540-018-0076-x.
- [154] Heidi Ledford. *How to solve the world’s biggest problems*. en. In: *Nature News* 525.7569 (2015), S. 308. DOI: 10.1038/525308a.
- [155] Harald Welzer. *Wissenschaften: Nur nicht über Sinn reden!* de-DE. In: *Die Zeit* (2006). ISSN: 0044-2070.

---

---

## Lebenslauf

---

- seit 01/2015 Promotion in Physik an der TU Darmstadt, AG Drossel
- 2013 – 2014 Studium an der Julius-Maximilians-Universität Würzburg mit Abschluss Master of Science, Physik (1,1).  
Master-Thesis: *Shared Information in Percolation Processes & The Abelian Sand-pile Model*
- 2012 – 2013 Studium an der Heriot Watt University Edinburgh mit Abschluss Master of Physics with Honours of the First Class (Grade A)
- 2009 – 2012 Studium an der Julius-Maximilians-Universität Würzburg mit Abschluss Bachelor of Science, Physik (1,4).  
Bachelor-Thesis: *Suche nach Leptoquarks der zweiten Generation mit Hilfe eines neuronalen Netzwerks*
- 2000 – 2009 Hermann-Staudinger-Gymnasium Erlenbach am Main, Abschluss Abitur (1,6)

---

## Publikationen

---

Die folgenden Publikationen sind während meiner Zeit als Doktorand entstanden und sind aus der in den Kapiteln 3–6 beschriebenen Forschung hervorgegangen:

- Johannes Falk, Marc Mendler und Barbara Drossel.  
»A Minimal Model of Burst-Noise Induced Bistability«.  
PLOS ONE 12, Nr. 4 (27. April 2017): e0176410.
- Marc Mendler, Johannes Falk und Barbara Drossel.  
»Analysis of Stochastic Bifurcations with Phase Portraits«.  
PLOS ONE 13, Nr. 4 (24. April 2018): e0196126.
- Anja Hofmann, Johannes Falk, Tim Prangemeier, Dominic Happel, Adrian Köber, Andreas Christmann, Heinz Koepl und Harald Kolmar.  
»A Tightly Regulated and Adjustable CRISPR-DCas9 Based AND Gate in Yeast«.  
Nucleic Acids Research 47, Nr. 1 (10. Januar 2019): 509–20.
- Johannes Falk, Leo Bronstein, Maleen Hanst, Barbara Drossel und Heinz Koepl.  
»Context in synthetic biology: Memory effects of environments with mono-molecular reactions«.  
The Journal of Chemical Physics 150, Nr. 2 (10. Januar 2019): 024106
- Thorsten Kohl und Johannes Falk,  
»The Schlögl Model: An Instrument in Synthetic biology«.  
In Revision (Journal for General Philosophy of Science)