Title

Stochasticity in biological networks : two sides of a golden coin

Permalink

https://escholarship.org/uc/item/43w6589k

Author

Lu, Ting

Publication Date

2007

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Stochasticity in Biological Networks: Two Sides of a Golden Coin

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Physics (Biophysics)

by

Ting Lu

Committee in charge:

Professor Peter G. Wolynes, Co-Chair
Professor Jeff Hasty, Co-Chair
Professor Massimiliano Di Ventra
Professor Terence Hwa
Professor J. Andrew McCammon

2007

The dissertation of Ting Lu is approved, and it is acceptable in quality and form for publication on microfilm.

Co-Chair

Co-Chair

University of California, San Diego

2007

To my parents

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to take this chance to acknowledge many people without whom this work would not have been possible. First, I thank my advisors Dr. Jeff Hasty and Dr. Peter Wolynes for their support, patience, and guidance throughout my entire doctoral study. The vision, thinking and physical sense of Dr. Wolynes always illuminate my research and guide my way of thinking. The inspiration, intuition, and caring of Dr. Hasty carry me along the whole graduate path. Drs. Hasty and Wolynes provide a free environment with trust and caring to allow me to study various interesting biological problems. I thank Dr. Hasty for bringing me into the wet lab for a chance to carry out bench work experiments. I have been extraordinarily lucky to have such superb advisors and the chance to work on both theory and experiment.

I thank all of the members in Systems Biodynamics Laboratory and Wolynes group who create a supportive and integrative environment. They are Lev Tsimring, Dmitri Volfson, Jesse Stricker, Matt Bennett, Scott Cookson, Tal Danino, Mike Ferry, Chris Grilly, Martin Kolnik, Diane Longo, Octavio Mondragon, Sujata Najak, Natalie Ostroff, Lee Pang, Jennifer Marcinik, Tongye Shen, Chenghang Zong and Mike Prentiss. Lev, Tongye, Dmitri, Chenghang, and Matt gave numerous advices and helps on theoretical and computational aspects. Jesse and Natalie are my experimental mentors who taught me every step of molecular biology techniques with kindness and patience. Mike and Chris are my experimental consultants that are always willing to help. I have been fortune to meet all these people. These two labs are my base of research.

I thank Tongye Shen and Chenghang Zong for thousands of afternoons spent together for scientific discussions, sometime even arguments, as well as chats about everything. These fruitful and enjoyable moments have really made my graduate research and graduate life much easier. I sincerely thank Dmitri Volfson for his numerous amount of help in job hunting and passing through some hard time that I've suffered along with all of his scientific help.

Many people have greatly helped me on the writing of this thesis: Lee shared his thesis template to save me a lot amount of time for editing. Tal, Martin and Mike edited and refined this thesis. Scott helped me in revising a couple of chapters when they were just papers. Ms. Darlene Palmer, my English tutor, meets with me every week to introduce American culture and improve my spoken English.

Most Friday evenings in my graduate study were spent together with army chess friends: Jiucang Hao, Junsheng Han, Quan Gu, Hongxing Guo, Xuerong Liu, Jing Chen, Qiao Xiao, and Mi Zhang. These were really laughing and relaxing time for us. I thank my friends, Haijiang Zhang, Kai Zhao, Liang Jin, Zhiqiang Li, Lei Zhu and Peng Yu for going through hardcore courses together as well as a lot of happy times.

Finally, I'm very grateful to Jiayan Hu for her support and encouragement. And I owe it so much to my parents, sister, brother-in-law, and my cousin Yunyun Wang for their love and support. This five-year-long oversea journey towards a Ph.D. is a great fortune to me.

Chapter 2 contains materials in Lu T, Hasty J, and Wolynes P, Effective temperature in stochastic kinetics and gene networks. *Biophys. J.* 91:84-94 (2006). Chapter 3 contains materials in Lu T, Volfson D, Tsimring L, and Hasty J, Cellular growth and division in the Gillespie algorithm, *Systems Biology* 1:121–128, (2007). Chapter 4 contains materials in Lu T, Shen T, Zong C, Hasty J, and Wolynes P, Statistics of cellular signal transduction as a race to the nucleus by multiple random walkers in compartment/phosphorylation space, *Proc. Natl. Acad. Sci.*, 103:16752-16757 (2006). Portions of Chapter 5 appear in part in Lu T, Shen T, Bennett M, Wolynes P, and Hasty J, Phenotypic variability of growing cellular populations, submitted. Portions of Chapter 6 appear in part of Lu T, Stricker J. and Hasty J, A synthetic switch with phenotypic and genotypic transitions, to be submitted.

Curriculum Vitæ

Education

2007
Doctor of Philosophy in Physics (Biophysics)
University of California at San Diego, La Jolla, CA

2003
Master of Science in Physics
University of California at San Diego, La Jolla, CA

2002
Bachelor of Science in Physics
Zhejiang University, Hangzhou, P. R. China.

Research Experience

2002–2007
Research Assistant, University of California at San Diego, CA
Co-advisor: Jeff Hasty, Department of Bioengineering
Co-advisor: Peter Wolynes, Department of Chemistry and Biochemistry
Quantitative study of biological networks, especially gene networks, using theoretical modeling and experimental molecular biology techniques

2001-2002
Research Assistant, Zhejiang University, Hangzhou, P. R. China
Advisor: You-quan Li, Department of Physics
Quantum effects in mesoscopic electric circuits

Teaching Experience

2006
Teaching Assistant, Department of Physics, University of California at San Diego
Laboratory of Physics for Engineers 2B

2002-2003
Teaching Assistant, Department of Physics, University of California at San Diego
Course website maintenance for the department

Selected Publications

T. Lu, J. Stricker, and J. Hasty. A synthetic switch with phenotypic and genotypic transitions. To be submitted.

N. Ostroff, T. Lu, W. Blake, C. Luke, L. Tsimring, and J. Hasty, An engineered eukaryotic clock. Submitted to *Cell*.

T. Lu, T. Shen, M. R. Bennett, P. G. Wolynes, and J. Hasty. Phenotypic variability of growing cellular populations. Submitted to *Proc. Natl. Acad. Sci.*.

T. Lu, T. Shen, C. Zong, J. Hasty, and P. G. Wolynes. Statistics of cellular signal transduction as a race to the nucleus by multiple random walkers in compartment/phosphorylation space, *Proc. Natl. Acad. Sci.*, 103:16752-16757 (2006).

T. Lu, J. Hasty, and P. G. Wolynes. Effective temperature in stochastic kinetics and gene networks. *Biophys. J.* 91:84-94 (2006).

C. Zong, T. Lu, T. Shen, and P. G. Wolynes. Nonequilibrium self-assembly of linear fibers: microscopic treatment of growth, decay, catastrophe and rescue. *Phys. Biol.* 3:83-92 (2006).

T. Lu*, D. Volfson*, L. Tsimring and J. Hasty (* equal contribution). Cellular growth and division in the Gillespie algorithm. *IEE Systems Biology* 1:121-127 (2004).

T. Lu and Y. Li, Mesoscopic circuit with linear dissipation. *Modern Physics Letters B*, 16:1-5 (2002).

ABSTRACT OF THE DISSERTATION

Stochasticity in Biological Networks: Two Sides of a Golden Coin

by

Ting Lu
Doctor of Philosophy in Physics (Biophysics)

University of California, San Diego, 2007

Professor Peter G. Wolynes, Co-chair

Professor Jeff Hasty, Co-chair

Cells live with fluctuations arising from various sources and occurring at a broad spectrum of scales. Stochasticity plays an amazing role in cellular functions and physiological behaviors through biological networks that compose a living cell. Like two faces of a coin, noise may be destructive in many biological systems but can also be constructive on the other hand.

In this work, I combined computational, theoretical and experimental approaches to explore stochasticity in biological networks. The origins, consequences and significance of stochasticity were investigated through the developments of methodological techniques as well as studies of specific yet important networks.

Effective temperature was proposed as a measurement of noise in gene networks. It serves as an alternative to noise classification by “intrinsic” and “extrinsic” contributions and is in some sense a more fundamental approach. A generalized Gillespie algorithm was derived for stochastic simulation of biochemical reactions that allows one to simulate biological systems with time-dependent reaction rates and system volumes. In addition to these developments, an important network topology abstracted from the multi-site phosphorylation networks of nuclear factors of activated T-cells was studied. Signal transduction of the network was mapped onto a random walker problem in nonequilibrium statistical mechanics and an optimal enzyme concentration was found that favors fast transduction. Noise at the cellular population level was also studied. A generalized variation index was proposed to measure variability and diversity of cellular populations. We found that cellular population variability may depend on its initial conditions and environments. Finally we turned to stochastic recombinant events in a gene circuit. A synthetic switch with both phenotypic and genotypic transitions was studied using combined experimental and theoretical approaches. This led to the result that there is always a bias of cellular population to one specific fate.

These studies show the two sides of stochasticity and help us to better understand noise in biological systems and to aid in better design strategy of genetic circuits.

1

Introduction

1.1 Quantitative Systems Biology

Studying biological systems is a long march filled with thorns and treasures. But the mission of it is never changed: To understand, predict and manipulate life. Biology moved forward slowly in its long history because of the extreme complexity of life and limitations of technology but the march has accelerated especially in the past decades. With the completion of the Human Genome Project and the genome sequencing for major model organisms, we now face an explosion of data as well as new chances for fundamental understanding.

One of the most significant challenges of this mission is to understand biological functions in terms of the interactions betweens proteins, genes and other small molecules [51]. To predict and design living machines, such an understanding needs to be quantitative rather than qualitative. A set of fundamental laws, like those in chemistry, physics, and engineering must be found, which are essential for predictions of biological systems. At the molecular level, biological players are individually quite complicated but are also strongly entangled with each other. Therefore going beyond traditional approaches focusing on individual genes or molecules is of great difficulty but is necessary. What is needed is an understanding at a systems level, rather than knowledge about individual components. These two factors define quantitative systems biology [120, 64, 118].

Fortunately, recently developed techniques provide a set of powerful tools for the biotechnological arsenal: Microarray experiments can efficiently generate huge amounts of information about the connecting interactions between genes on a whole genome scale [98]; fluorescent proteins, tagged genes and quantum dots are increasingly used as labels in various aspects of biological experiments [112, 79]; microfluidic techniques can be used to easily manipulate microenvironments of cells and organisms [119]; microscopes and FACS machines are used to monitor dynamics at single-cell level [105]. All these techniques bring the study of biological systems into a stage where a quantitative systems level understanding becomes not only possible but obligatory.

1.2 An Integrated Approach

Systems biology uses modeling and simulation, combined with experiment, to explore behaviors in biological systems, especially dynamic behaviors [64, 3]. This thesis represents an example of such a combined theoretical, computational and experimental approach.

1.2.1 Analytical Modeling

At the cellular level, many biological behaviors can be interpreted in terms of more elementary physical and chemical processes. Once the actors are established then the principles and laws of physical sciences can be employed for quantitative descriptions of biological phenomena. In this integrative approach, modeling provides the basis for quantitative understanding since it converts qualitative information into a quantitative language.

Biological systems can be modeled in several ways and at different levels: The most fundamental description is at the single-molecule level where individual atoms are the basic elements. At this level microscopic principles like those in quantum mechanics govern biological behavior. The second layer of modeling arises at the single-cell level, where the fundamental units are molecules and thermodynamics and kinetic coefficients may be used to model individual reaction events at the microscopic level. The highest level of description is concerned with overall phenomenological behavior for a biological system at a macroscopic level.

The studies in this thesis mainly focus on the second and the third levels where microscopic and macroscopic descriptions overlap and are both employed.

At these levels, there are stochastic and deterministic descriptions. Among deterministic descriptions, traditional chemical kinetics allows one to transform a qualitative description of molecular and genetic interactions into a quantitative one. Many kinetic schemes, such as Michaelis-Menten reaction and Hill-type kinetics, occur over and over again in biological systems [20]. Since deterministic chemical kinetics yields sets of ordinary or partial differential equations, these models can be solved and analyzed using the tools developed in nonlinear dynamics [107].

Deterministic modeling is simple and powerful. Nevertheless, it is not sufficient when studying single cells. Biological systems usually behave in a stochastic rather than deterministic manner because the numbers of many key biomolecules in a cell are only on the order of tens to hundreds. This means the number fluctuations of molecules are not negligible. Many descriptions in nonequilibrium statstical mechanics can be used for modeling such stochastic systems. Among these descriptions, the most widely used ones are: Master equations, Fokker-Planck equations, and Langevin equations [113, 92, 40]. Because of the complicity and nonlinearity of biological systems, these stochastic models are hard to solve analytically even for simple systems. A set of approximation approaches such as Ω-expansion, linear noise approximation, Eikonal approximation, and variational methods, have been developed and play important roles [113, 92, 40].

1.2.2 Simulation

Computer simulation has been used in biological studies for more than half a century and has proved successful in numerous examples. It provides an alternative language to analytical modeling for many biological systems. With the increase of computation power, simulation has become more and more important.

Much like modeling, simulation can be either deterministic or stochastic. Among different methods, the Monte Carlo simulation developed by Gillespie in 1970s is regarded as a gold standard for stochastic simulation of biological networks [45]. Later developments of the Gillespie algorithm also greatly increased the efficiency and applicability of simulation [88, 73].

1.2.3 Molecular biology

Molecular biology provides the cornerstone for studying quantitative systems biology. Standard molecular biology techniques as well as efficient and cheap DNA synthesis methods are required for quantitative laboratory work [95].

1.2.4 Integration of the three aspects

The three aspects, analytical modeling, simulation and molecular biology, are not isolated but interact strongly. Preliminary experiments provide qualitative information about the main players in a system and the interactions between them. Modeling gives a theoretical foundation that transforms the qualitative information from such experiments into quantitative descriptions. The results and predictions of modeling give insights into new experiments that should be be performed. Simulation serves as an alternative to theoretical modeling when analysis is difficult and also serves as an alternate to experiment as a means of simply and quickly checking one's model. Comparing modeling, simulation and experiment is a cycle which ultimately converges to concrete results.

1.3 Stochasticity in Biological Networks

Stochasticity is ubiquitously exhibited in biological systems [89, 91, 60, 29]. Stochastic phenomena appear in organisms ranging from prokaryotic to eukaryotic and multicellular cells. It also occurs at different scales: At the molecular level, protein conformational changes, transcription and translation, and genetic mutations are all stochastic [37, 60, 63]. At the level of whole cells, growth, division, transitions between phenotypes and cell differentiation are all noisy. At the tissue and organ level, stochasticity plays a role as well, notably in embryonal development and tumor growth [26].

This behavior resides largely in various complicated networks, such as gene networks, signal transduction networks, and metabolic networks. Understanding biological networks is therefore the basis for understanding living matter. Stochasticity in such networks is the key to decoding mysteries of biology. Network stochasticity has therefore received extensive attention in the past few years with examples including: Protein production occurring in "bursts" at random time intervals rather than in a continuous manner [122, 19], observations of gene expression in individual cells that have also established the stochastic nature of transcription and translation, noise which has been found to limit the precision of biological rhythms, and contributions to the history of development of *Caenorhabditis elegans* and *Xenopus laevis* [106, 116].

Stochasticity arises from multiple sources [109, 110, 90, 114]. At the molecular level, noise comes from the intrinsic nature of biochemical reactions underlying the biological dynamics, such as fluctuations due to discrete gene transcription. This is termed intrinsic noise. Noise that originates from random fluctuation of other factors such as copy numbers of RNA polymerase is termed extrinsic noise. Noise is also contributed by the fluctuating environments of a cell, which is termed external noise.

Stochasticity in biological networks can have important consequences [89, 91, 60]. Much like its role in wireless communications, noise can destroy signals traveling through biological networks, it can disperse distributions of numbers of molecules, perturb systems away from their needed steady states and even drive cells to dramatically different states, such as pathological ones in cancer. Noise is evil from this point of view. Inevitably disturbed by noise, biological systems must still robust enough to resist to this molecular noise and function well most of the time. On the other hand, stochasticity is not always bad. The studies of stochastic focusing and stochastic resonance directly demonstrate the positive effects of noise [38, 83]. Furthermore, from the viewpoint of a physicist, noise randomizes all sorts of biological information, e.g., numbers of molecules and types of species, and increases entropy of biological systems. The randomization of biological information is important to the diversification of species and allows evolution to be speeded. Noise thus also plays a constructive role. We see that stochasticity is just like a coin with two faces: One evil and the other angelic.

Two big questions arise from the two faces of stochasticity:

- How are biological networks designed by nature to guard against noise and function robustly?
- How does noise diversify a species' richness and thereby benefit its survival in evolution?

Clearly the study of stochasticity in biological systems will be vital to improving our understanding of biology as a whole. Answering the above questions will help us to understand architectures of biological networks that can function robustly and help us understand ways that noise benefits evolution of biological systems. Furthermore, these studies can directly aid our final goal to design biological systems, a goal now termed as "synthetic biology" [5].

1.4 Goal

My thesis employs an integrated approach to study stochastic properties of biological networks for better understanding of network functions and architectures and to aid in synthetic design strategy.

I approach this goal by developing methodological techniques as well as studying specific systems. The concept of effective temperature in stochastic kinetics and gene networks is introduced in Chapter 2. The effective temperature serves as a fundamental quantity for the measurement of genetic noise. In Chapter 3, a more general stochastic algorithm is developed that can be used in systems with time-dependent reaction rates, as is almost always the case in study of prokaryotes. A specific network topology that commonly exists in gene networks and signal transduction networks is studied in Chapter 4. Variability of cellular population is investigated in Chapter 5 by proposing a general variation index and studying a simple example. A synthetic switch with genotypic and phenotypic transitions is studied using both experimental and theoretical approaches in Chapter 6. Lastly, conclusions and future outlook are provided along with some closing remarks.

2

Methodological Developments I: Effective Temperature as a Measurement for Gene Networks

2.1 Introduction

Gene networks are inherently noisy systems, with fluctuations arising from the stochastic nature of the underlying biochemical reaction events [76, 59, 51]. Fluctuations become more important when the numbers of reactant molecules are small, as is often the case in gene regulatory networks. The role of noise in gene expression has attracted much attention over the past few years, and many approaches have been used to model these systems. The Gillespie algorithm [45, 88, 73] is often considered to be the gold standard for performing stochastic simulations of such stochastic biological processes.

The noise in gene networks can been classified as either intrinsic, related to the specific biochemical process under evaluation, or extrinsic, related to factors upstream of this process [109, 110]. Various groups have experimentally investigated the effect that each source has on gene expression, and these studies have greatly improved our understanding of stochastic gene networks. Recent papers have laid a foundation for comparing experimental data with quantitative models of simplified gene networks [85, 82] and have shown that noise may limit the sensitivity of gene networks [102].

Many such models have employed the fluctuation-dissipation theorem, the central theorem in equilibrium mechanics, to describe these networks [65]. However, gene networks are complex, far-from-equilibrium systems for which the fluctuation-dissipation theorem may break down. In this paper we explore how the fluctuation-dissipation theorem may apply or break down in the realm of stochastic biochemical processes.

One of the key topics in quantitative biology is the search for measurable quantities which can be used to characterize the properties of a gene network. Here we explore the potential that effective temperature may provide such a quantitative measure. In thermodynamics, temperature must be the same for two bodies when they are in thermal equilibrium [49]. At equilibrium, temperature is independent of which part of a system and the precise measurements made. Near equilibrium, temperature differences determine the direction of heat flow. When a system is far from equilibrium, these key principles of thermodynamics may fail. Nevertheless it has been possible to extend the concept of temperature in thermodynamics to effective temperature in non-equilibrium systems [66] that change sufficiently slowly, such as glassy systems [24, 23].

Since genetic systems often respond much more slowly than the rate of their individual reaction events, we are encouraged to examine whether effective temperature plays an important role in gene networks. In fact, a few studies have already employed effective temperature to investigate biological systems. For example, it has been used to study the stability of motorized particles in cytoplasm [99] and to reveal the underlying active process in hair bundles [74]. Here we examine effective temperature as determined by comparing fluctuation to response in stochastic kinetics with the application to gene networks in mind. Our first example focuses on the effective temperature of a simple birth-death process. We then generalize our argument to describe the kinetics of two-species interactions.

The paper is organized as follows. After introducing the definition of effective temperature, we investigate a birth-death process where exact expressions for the effective temperature are derived both in the cases where there are large and small numbers of particles. Monte Carlo simulations are compared to these exact results. We then study a general system of two interacting chemical species. For this system, analytical results are derived using Langevin method relevant for a relatively large numbers of particles. For the two species problem we can examine whether temperature gradients determine the direction of flows. Finally we use the effective temperature to study an unregulated gene where both intrinsic and extrinsic nose are quantified using effective temperature. We conclude with a general discussion.

2.2 Effective Temperature

Let us start with a classical exercise which interested Einstein one century ago [30, 65]. In the presence of a given potential field $V(x)$, particles flow with the drift velocity $J_F = -\mu c \frac{\partial V(x)}{\partial x}$, where μ is the mobility and c is the concentration of particles at the position x. On the other hand, particles diffuse randomly and obey the Ficks's first law of diffusion as $J_D = -D\frac{\partial c}{\partial x}$. In thermal equilibrium, particles are in the Boltzmann distribution, i.e. $c \sim exp[-V(x)/k_BT]$, and the net current $J_N = J_F - J_D$ is null, i.e. $-\mu c\frac{\partial V(x)}{\partial x} + D\frac{\partial c}{\partial x} = 0$. Therefore the mobility, due to the friction of particles, and the diffusion, due to the random motions of the particles, are directly related

$$\mu = \frac{1}{k_BT}D \tag{2.1}$$

This has become known as the Einstein relation. This relation between fluctuation(D) and response(μ), when manifested in a more general manner, is called fluctuation-dissipation theorem.

The fluctuation-dissipation theorem states a general relationship between the response of a given system to an external disturbance and the internal fluctuations of the system in equilibrium. This relationship contains the temperature and is central in thermodynamics. However, when a system is out of equilibrium, the theorem breaks down and an extension of the theorem must be made. So the concept of effective temperature is introduced.

In a near equilibrium system, it is customary to study mechanical fluctuations and response. Consider such a mechanical system with a Hamiltonian H. If O_1 and O_2 are two observables of the system, then the correlation between these two observables is described by the function

$$C_{12}(t', t) = \langle O_1(t')O_2(t)\rangle - \langle O_1(t')\rangle\langle O_2(t)\rangle \tag{2.2}$$

where the brackets indicate the ensemble average.

If the system is subjected to a time-dependent small perturbation $-h(t)O_2(t)$ where $h(t)$ is a small field and $O_2(t)$ is an observable, then the Hamiltonian becomes

$$H \rightarrow H - h(t)O_2(t) \tag{2.3}$$

The response of the average of the observable O_1 to the small perturbation $-h(t)O_2(t)$ is then given by

$$R_{12}(t', t) = \frac{\delta\langle O_2(t')\rangle}{\delta h(t)} \tag{2.4}$$

For systems with slow dynamics Cugliandolo et. al. [24] suggests the effective temperature in the Fourier space may be defined as

$$T_{eff}(\omega) \equiv \frac{\omega \tilde{C}'_{12}(\omega)}{\tilde{R}''_{12}(\omega)} \tag{2.5}$$

where $\tilde{C}'_{12}(\omega)$ is the real part of the Fourier transform of the correlation function Eq.(2.2) and $\tilde{R}''_o(\omega)$ is the imaginary part of the Fourier transform of the response Eq.(2.4), i.e., $\tilde{C}'_{12}(\omega) = \Im\{\int_0^\infty dt C_{12}(t)e^{i\omega t}\}$, $\tilde{R}''_{12}(\omega) = \Re\{\int_0^\infty dt\ R_{12}(t)e^{i\omega t}\}$.

Genetic networks are generally multi-timescale systems. The characteristic time of binding events is much shorter than the half-life of the messenger RNA, which in turn is one order of magnitude shorter than the half-life of the resulting protein. Cell-to-cell communications and transportation processes are even possibly slower. These multiple scale properties encourage us to investigate the role of defining effective temperatures in genetic networks.

2.3 A Birth-Death Process

Genetic networks are extremely complex, involving binding processes, synthesis processes and degradation processes. To understand the fundamental properties of noise in gene networks, the response of networks to small perturbations, and the effective temperatures, we begin with the simplest birth-death process, which describes synthesis and degradation. Studying the simple birth death process already helps us understand the lim-

Table 2.1: Birth-Death Process

reaction	rate	specified rate
$\phi \rightarrow X$	$g(x)$	k_g
$X \rightarrow \phi$	$d(x)$	k_d

its of the effective temperature concept. The reactions of a simple birth-death process are shown in Tab. 2.1. The Master Equation for this process reads

$$\frac{\partial}{\partial t}P(x,t) = g(x-1)P(x-1,t) - g(x)P(x,t) + d(x+1)P(x+1,t) - d(x)P(x,t) \quad (2.6)$$

2.3.1 Large-N case

When the numbers of particles are large, the Master Equation can be approximated by the Fokker-Planck Equation, which is equivalently to the Langevin equation:

$$\frac{d}{dt}x = g(x) - d(x) + \eta(t) \quad (2.7)$$

where $\eta(t)$ is a Gaussian white noise term, i.e. $\langle \eta(t) \rangle = 0$, $\langle \eta(t)\eta(t') \rangle = [g(x) + d(x)]\delta(t-t')$. Comparing this simple birth-death process with the Brownian motion of an overdamped particle [17], we have the effective temperature of this process

$$T_{eff} = \frac{1}{2}\{g(x) + d(x)\} \quad (2.8)$$

where T_{eff} can be further simplified as $T_{eff} = g(x)$ if the system is near the steady state where $g(x) - d(x) = 0$ holds. If we specify $g(x)$ and $d(x)$ as those listed in the third column of the Tab. 2.1, the effective temperature becomes $T_{eff} = k_g$.

The effective temperature of this birth-death process can also be derived from the effective fluctuation-dissipation relationship. There are two kinds of perturbations that can be used to calculate effective temperature according to the FD relation. We may use a perturbation of generation rate k_g or a perturbation of degradation rate k_d. Both of these perturbations yield the same effective temperature $T_{eff} = k_g$. This result is consistent with Eq.(2.8), which verifies the validity of the definition of effective temperature in the large N limit.

2.3.2 Small-N case

In the last subsection, the Eq.(2.7) is appropriate for large number limit. But we also wish to define the effective temperatures in the small number case to see if they the same. In order to answer these questions, we first use the operator formulism for Master equations [75] combined with the Eyink's variational method [34] to get the correlation and response and thus the effective temperatures for any value of the mean number.

In the operator formulation, the Master Equation Eq.(2.6) with specified reaction rates is written as

$$\frac{d}{dt}|\Psi(t)\rangle = \hat{L}|\Psi(t)\rangle \tag{2.9}$$

where the Liouvillian is $\hat{L} = k_g(\hat{a}^+ - 1) + k_d(\hat{a} - \hat{a}^+\hat{a})$. The Liouvillian is generally non-Hermitian, i.e. $\hat{L} \neq \hat{L}^+$. The state function $|\Psi(t)\rangle$ is defined as $|\Psi(t)\rangle \equiv \sum_0^\infty P(n,t)|n\rangle$ where $P(n,t)$ is the probability of having n number of X at time t and $|n\rangle$ is the state with n number of X. $\hat{a}$ and $\hat{a}^+$ are creation and annihilation operators respectively which have the relations $\hat{a}|n\rangle = n|n-1\rangle$, $\hat{a}^+|n\rangle = |n+1\rangle$ and $[\hat{a}, \hat{a}^+] = 1$.

To solve the Eq.(2.9), we use the coherent trial *Ansätze* $\langle\Psi^L| = \langle 0|e^{\hat{a}}(1+\alpha(\hat{a}^+\hat{a} - 1))$ and $|\Psi^R\rangle = e^{u(\hat{a}^+-1)}|0\rangle$. The action for the system is

$$\Gamma = \int_0^\infty dt\ \langle\Psi^L(t)|(\partial_t - \hat{L})|\Psi^R(t)\rangle \tag{2.10}$$

The variations of the action generate two equations for the parameters α and u

$$\frac{d}{dt}\alpha(t) = -k_d\alpha(t) \tag{2.11}$$

$$\frac{d}{dt}u(t) = k_g - k_d u(t) \tag{2.12}$$

The steady states of the system $\langle\Psi_s^L| = \langle 0|e^{\hat{a}}$ and $|\Psi_s^R\rangle = e^{u_s(\hat{a}^+-1)}|0\rangle$ are then found by taking $\alpha = 0$ and $u = \frac{k_g}{k_d} \equiv u_s$ in the above equations. This indicates that the steady state distribution is a Poisson distribution with the mean u_s. This is confirmed to be the exact distribution by solving the Master equation using the generating function method [77]. The operator formulation, while cumbersome for the distribution function, is advantageous for calculating correlation and response functions, as we see in the following description.

Recall that there are two types of perturbations of the system – a perturbation of generation rate and a perturbation of degradation rate. These correspond to the perturbations of the Liouvillian $\hat{L}$ by $-h(t)(\hat{a}^+ - 1)$ and $-h(t)(\hat{a} - \hat{a}^+\hat{a})$. Because the Liouvillian is non-Hermitian ($\hat{L} \neq \hat{L}^+$), the left and state functions are not conjugate, i.e. $|\Psi^R(t)\rangle \neq (\langle\Psi^L(t)|)^+$. Likewise the time-dependent expression for an operator in the Heisenberg representation of the stochastic process is somewhat different from that for quantum system with a Hermitian Hamiltonians.

In the non-Hermitian case, a time-dependent operator can be defined as [25]

$$\hat{A}(t) = \hat{\hat{U}}^+\hat{A}_s\hat{U} \tag{2.13}$$

where $\hat{U}$ and $\hat{\hat{U}}^+$ are the evolution operators for the forward and backward Kolmogorov processes respectively. Because of the non-Hermitian property of the Liouvillian, the forward evolution operator is not the conjugate of the backward, i.e., $(\hat{U})^+ \neq \hat{\hat{U}}^+$. However, $\hat{U}$ and $\hat{\hat{U}}^+$ should conserve the evolution, i.e. $\hat{\hat{U}}^+\hat{U} = \hat{U}\hat{\hat{U}}^+ = 1$.

Now we turn to the calculation of correlation and response functions. For a perturbation of Liouvillian corresponding to the generation rate with $\hat{L} - h(t)(\hat{a}^+ - 1)$, the corresponding correlation and response functions can be obtained after some algebra(see Appendix A). These are

$$C(t', t) = u_s^2 + u_s e^{-k_d(t'-t)} \tag{2.14}$$

$$R(t', t) = e^{-k_d(t'-t)} \tag{2.15}$$

from which the effective temperature may be derived

$$T_{eff}^g = k_g \tag{2.16}$$

This is exactly the same as what was obtained in the large number limit.

On the other hand, for a perturbation of the degradation rate of the Liouvillian with $\hat{L} - h(t)(\hat{a} - \hat{a}^+\hat{a})$, the corresponding correlation and response functions(see Appendix A) are

$$C(t', t) = \frac{1}{2}(u_s^3 + u_s^2 + 2u_s^2 e^{-k_d(t'-t)} + u_s e^{-k_d(t'-t)}) \tag{2.17}$$

$$R(t', t) = u_s e^{-k_d(t'-t)} \tag{2.18}$$

The effective temperature is thus found to be

$$T_{eff}^d = k_g + k_d/2 \tag{2.19}$$

This temperature is different from either the one found by perturbing the generation rate in small number regime Eq.(2.16) or what was found in large number limit Eq.(2.8).

For this birth-death process, the percentage of the relative difference of the two temperatures is $\frac{50k_d}{k_g}\%$. When the average number of a protein $\frac{k_g}{k_d}$ is 10, the difference of the two temperatures is 5%, which is already obvious. Further, when the average number is around 1, the difference will be up to 50%. In some natural and synthetic genetic systems, the numbers of molecules is actually small. One study shows that the freely available dimers of the repressors CI in the phage lambda is only around 10 [8, 9], in which case small number will be an important effect. This effect might be dominant in those cases where there are few copies or even sometimes a single copy of a gene in the natural genome or in the case of there being only a few copies of plasmids in a cell.

However, when the number of molecules is large, i.e. $k_g/k_d >> 1$, both T^g_{eff} and T^d_{eff} merge consistently to the large number limiting value k_g.

2.3.3 Comparison with simulations

To check the effective temperatures derived analytically, we carried out some explicit simulations. We used the Gillespie simulation [45] to generate data following the reaction rules in Tab.2.1. We use the rate coefficients shown in the Tab.2.1 to generate data for each run and made 10000 runs of the program to represent the ensemble of trajectories. We numerically calculated the correlations and use averaged response.

To calculate the response function, we actually found the frequency-dependent susceptibility of the system rather than response function just as one would do in real experiments [24]. The simulation first "equilibrated" after waiting for a long enough time to allow the system comes to a steady state and to fluctuate around that state. We then turned on the perturbations (for a perturbation of generation rate we use $k_g^{'} = k_g(1 + 5\%)$, for a perturbation of degradation rate we use $k_d^{'} = k_d(1 + 5\%)$). We recorded all of the data once the perturbation was added. 10000 independent runs provided an ensemble that could be numerically averaged to get the susceptibility.

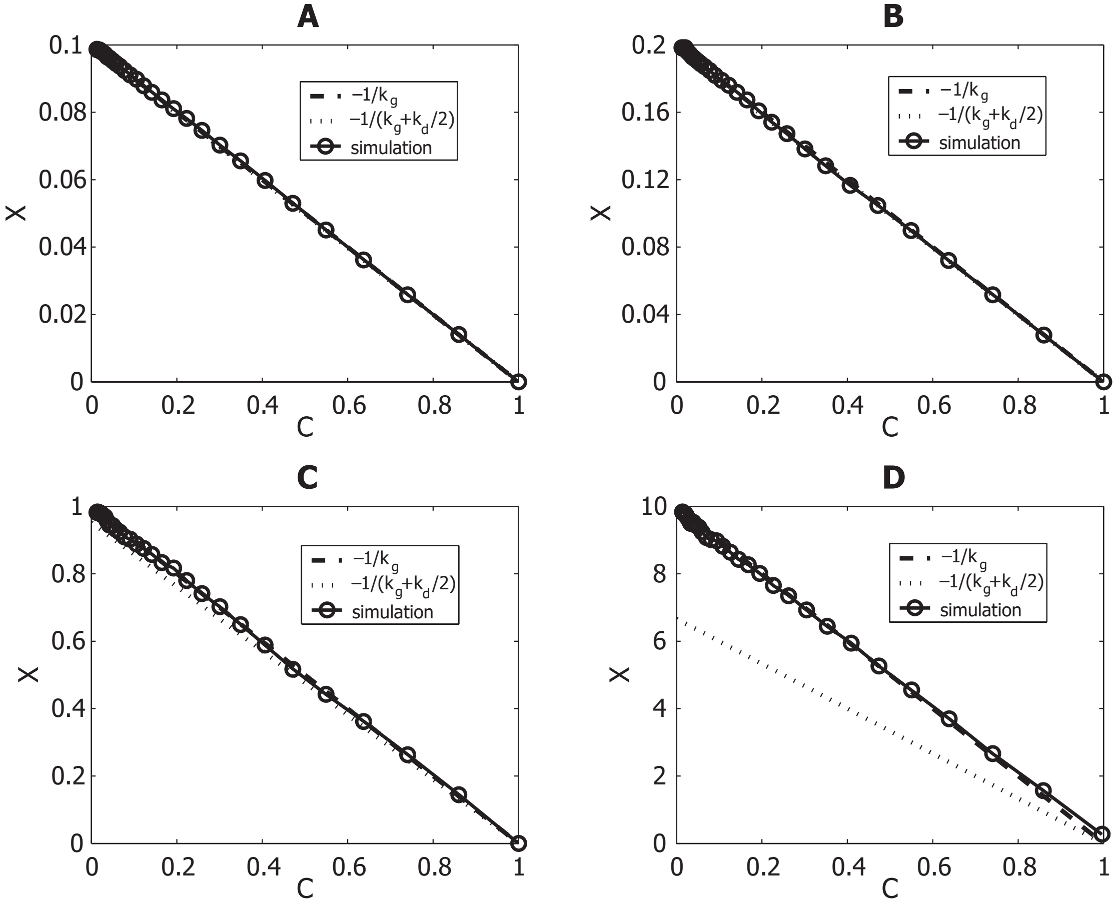

Figure 2.1: Effective temperatures corresponding to different mean value numbers k_g/k_d for a perturbation of $-h(t)x$. The parameters $\{k_g$=10.0, $k_d = 0.1\}$, $\{k_g$=5.0, $k_d = 0.1\}$, $\{k_g$=1.0, $k_d = 0.1\}$ and $\{k_g$=0.1, $k_d = 0.1\}$ are chosen in the panels A, B, C, and D respectively. Broken lines have a slope of $-\frac{1}{k_g}$, dotted lines have a slope of $-\frac{1}{k_g+k_d/2}$, and lines with circles are the simulation results with corresponding reaction rates. The X and Y axes are correlation and susceptibility respectively, and both are scaled so that the correlations range from 0 to 1 in all of the figures for comparison.

Following Cugliandolo and Kurchan we can draw parametric fluctuation-dissipation plots of the response versus the correlation, where the negative reciprocal of the slope of the curve indicates the effective temperature [24]. Fig.2.1 clearly indicates that for a perturbation of $-h(t)(\hat{a}^{+} - 1)$ the slope is always equal to $-1/k_g$. This result is found regardless of the mean values of the species number. This means that the corresponding effective temperature is always k_g. However, we can see from Fig.2.2 that for a perturbation of $-h(t)(\hat{a} - \hat{a}^{+}\hat{a})$ the slope of the simulation is equal to $-1/(k_g + k_d/2)$. This difference becomes more clear when the mean number is small in panel D. For this observable, we see that the corresponding effective temperature is $k_g + k_d/2$ rather than k_g. Although these two effective temperatures are not the same, they tend to become equal when the mean number becomes large (e.g. the left upper panel in Fig. 2.1 and Fig. 2.2). Both of the simulations agree very well with the analytical results in the above subsection.

2.3.4 Physical interpretation

Both the analytical and numerical results show that there is more than one effective temperature for a birth-death process. Effective temperature is thus not a unique quantity, but rather is observable-dependent. This result is surprising given the general properties of effective temperature for a far-from-equilibrium system[66]. In equilibrium systems all of the effective temperatures must converge to the same value regardless of the observable used in measurements [65] but this need not be the case for an out of equilibrium system. The fact that there are two different effective temperatures does not contradict our physical intuition, but simply emphasizes that the system in a steady-state is not a true state of equilibrium.

The two effective temperatures explore different aspects of the system dynamics in

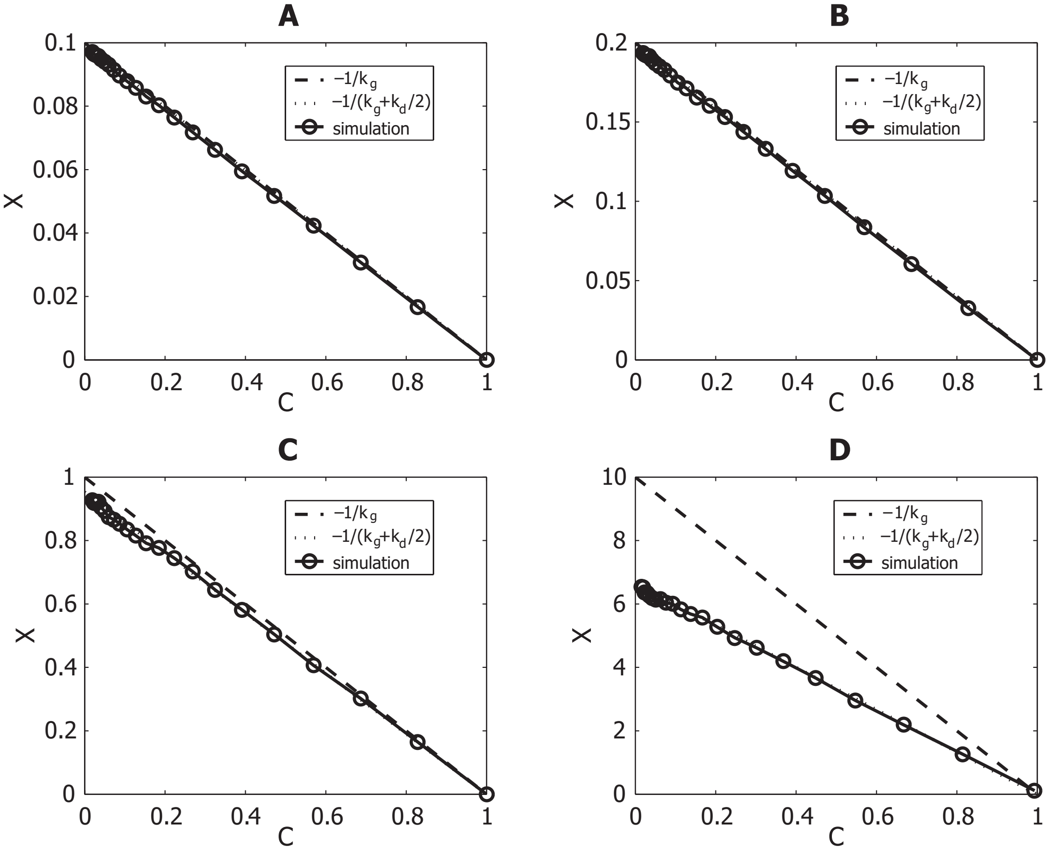

Figure 2.2: Effective temperatures corresponding to different mean value numbers k_g/k_d for the perturbations of $-h(t)x^2$. The parameters $\{k_g$=10.0, $k_d = 0.1\}$, $\{k_g$=5.0, $k_d = 0.1\}$, $\{k_g$=1.0, $k_d = 0.1\}$ and $\{k_g$=0.1, $k_d = 0.1\}$ are chosen in the panels A, B, C, and D respectively. Broken lines have a slope of $-\frac{1}{k_g}$, dotted lines have a slope of $-\frac{1}{k_g+k_d/2}$, and lines with circles are the simulation results with corresponding reaction rates. The X and Y axes are correlation and susceptibility respectively, both are scaled so that the correlations range from 0 to 1 in all of the figures for comparison.

Table 2.2: Two-Species Interacting Process

reaction	rate const.
$\phi \to A$	k_{g1}
$A \to \phi$	k_{d1}
$\phi \to B$	k_{g2}
$B \to \phi$	k_{d2}
$A \to m_1 A + n_1 B$	u
$B \to m_2 B + n_2 A$	v

the steady states corresponding to two different perturbations. The difference between the two temperatures is, however, a 'higher order' correction in some sense. As shown above, they converge to each other in the large number limit.

2.4 A Two-Species Interacting Process

Now we turn to a more complex example involving interactions between two species. The two species A and B may have their own generation and degradation processes, and if one species is consumed by the other, they become correlated due to this interaction (Table 2.2). When their numbers are high, the two species essentially interact as a predator-prey system and may be described by Langevin Equations [44]

$$\frac{\partial}{\partial t} A = k_{g1} - d_1 A + f_2 B + D_1 \xi_1(t) + D_3^a \xi_3(t) + D_4^a \xi_4(t) \tag{2.20}$$

$$\frac{\partial}{\partial t} B = k_{g2} - d_2 B + f_1 A + D_2 \xi_2(t) + D_3^b \xi_3(t) + D_4^b \xi_4(t) \tag{2.21}$$

where $d_1 = k_{d1} + (1 - m_1)u$, $d_2 = k_{d2} + (1 - m_2)v$, $f_1 = n_1 u$, $f_2 = n_2 v$. $\xi_i(t)$ is a Gaussian normal noise with the properties $\langle \xi_i(t) \rangle = 0$ and $\langle \xi_i(t) \xi_j(t') \rangle = \delta_{i,j} \cdot \delta(t - t')$. Their noise intensities are $D_1 = \sqrt{k_{g1} + k_{d1} A}$, $D_2 = \sqrt{k_{g2} + k_{d2} B}$, $D_3^a = (m_1 - 1)\sqrt{uA}$, $D_3^b = n_1 \sqrt{uA}$, $D_4^a = n_2 \sqrt{vB}$, $D_4^b = (m_2 - 1)\sqrt{vB}$.

The correlations arising from the interactions enter as common noise terms in Eqs.(2.20,2.21). These linear equations could arise from a nonlinear system by linearizing around the steady states. This is especially relevant in the current context since both the correlations and response are only calculated when the system is at a steady state. The numbers m_1, m_2, n_1 and n_2 can be arbitrary positive or negative integers, and represent different types of interactions between the two species. For example, $m_1 = m_2 = 1$, $n_1 = n_2 = -1$ implies that A consumes B and there is competition between A and B, while $m_1 = m_2 = 0$, $n_1 = n_2 = 1$ means A reacts to B and B reacts to A. All of the reaction rates are shown in column two of the Table 2.2. We note that for a system of molecular species the steady state values should be positive, i.e.

$$A^* = \frac{f_2 k_{g2} + d_2 k_{g1}}{d_1 d_2 - f_1 f_2} > 0 \tag{2.22}$$

$$B^* = \frac{f_1 k_{g1} + d_1 k_{g2}}{d_1 d_2 - f_1 f_2} > 0 \tag{2.23}$$

$$d_1 d_2 - f_1 f_2 \neq 0 \tag{2.24}$$

2.4.1 Effective temperature of the two-species system

As explored in the single birth-death process, generally there are multiple effective temperatures, with different temperatures corresponding to measurements using different perturbations. Instead of comparing all of the temperatures of the two species, we focus on the effective temperature corresponding to the perturbation of the generation rates and the autocorrelation of one species as a measurement for the interacting process. Other effective temperatures can be derived in straight-forward fashion. We will lose some information by using only one of the effective temperatures, corresponding to the autocorrelation, but we will still capture the essence of the situation.

The autocorrelations of A and B for the two-interacting species are derived using the fourier transform(see Appendix B)

$$C_{AA} = \frac{(\omega^2+d_2^2)D_1^2+f_2^2D_2^2+[(d_2D_3^a+f_2D_3^b)^2+\omega^2(D_3^a)^2]+[(d_2D_4^a+f_2D_4^b)^2+\omega^2(D_4^a)^2]}{(\omega^2+f_1f_2-d_1d_2)^2+(d_1+d_2)^2\omega^2} \tag{2.25}$$

$$C_{BB} = \frac{f_1^2D_1^2+(\omega^2+d_1^2)D_2^2+[(f_1D_3^a+d_1D_3^b)^2+\omega^2(D_3^b)^2]+[(f_1D_4^a+d_1D_4^b)^2+\omega^2(D_4^b)^2]}{(\omega^2+f_1f_2-d_1d_2)^2+(d_1+d_2)^2\omega^2} \tag{2.26}$$

The response functions can be derived similarly (see Appendix B)

$$R_{AA} = \frac{i\omega-d_2}{\omega^2+i\omega(d_1+d_2)+f_1f_2-d_1d_2} \tag{2.27}$$

$$R_{BB} = \frac{i\omega-d_1}{\omega^2+i\omega(d_1+d_2)+f_1f_2-d_1d_2} \tag{2.28}$$

Following the definition Eq.(2.5), we have the effective temperature for species A and B from Eqs.(2.25-2.28)

$$T_{eff}^{AA}(\omega) = \frac{(\omega^2+d_2^2)D_1^2+f_2^2D_2^2+[(d_2D_3^a+f_2D_3^b)^2+\omega^2(D_3^a)^2]+[(d_2D_4^a+f_2D_4^b)^2+\omega^2(D_4^a)^2]}{\omega^2+f_1f_2+d_2^2} \tag{2.29}$$

$$T_{eff}^{BB}(\omega) = \frac{f_1^2D_1^2+(\omega^2+d_1^2)D_2^2+[(f_1D_3^a+d_1D_3^b)^2+\omega^2(D_3^b)^2]+[(f_1D_4^a+d_1D_4^b)^2+\omega^2(D_4^b)^2]}{\omega^2+f_1f_2+d_1^2} \tag{2.30}$$

These are the effective temperatures of the predator-prey system. The expressions show that the effective temperature for each species depends not only on its own birth-death rate but also on the birth-death rate of the other species. They are correlated by interacting terms.

The temperatures depend on the time scale just as they do for other non-equilibrium systems, e.g. glasses [24], where the fluctuation-dissipation theorem in its simple form breaks down [65]. A convenient way to illustrate this breaking down is to draw correlation-susceptibility plots [24], since the negative reciprocal of the slope of the curves indicates the effective temperature of a system. For the simplicity, we set the birth-death noise terms D_1 and D_2 to be constants, and set the correlated noise terms D_3^a,D_3^b,D_4^a and D_4^b to be zero.

This is a simplified version of the general interacting process shown above. However, even this simple system suffices to demonstrate the breaking down of the fluctuation-dissipation relation. Here we employ the inverse Fourier transform to get the correlation and response functions in real space and then draw the plots.

$$C_{AA}(t) = \tfrac{[d_2^2-(x-y)^2]T_1+f_2^2T_2}{4xy(x-y)}e^{-(x-y)t} - \tfrac{[d_2^2-(x+y)^2]T_1+f_2^2T_2}{4xy(x+y)}e^{-(x+y)t} \tag{2.31}$$

$$C_{BB}(t) = \tfrac{[d_1^2-(x-y)^2]T_2+f_1^2T_1}{4xy(x-y)}e^{-(x-y)t} - \tfrac{[d_1^2-(x+y)^2]T_2+f_1^2T_2}{4xy(x+y)}e^{-(x+y)t} \tag{2.32}$$

$$R_{AA}(t) = \tfrac{(x+y)-d_2}{2y}e^{-(x+y)t} - \tfrac{(x-y)-d_2}{2y}e^{-(x-y)t} \tag{2.33}$$

$$R_{BB}(t) = \tfrac{(x+y)-d_2}{2y}e^{-(x+y)t} - \tfrac{(x-y)-d_1}{2y}e^{-(x-y)t} \tag{2.34}$$

where $x = (d_1 + d_2)/2$, $y = \sqrt{(d_1 - d_2)^2 + 4f_1f_2}/2$, $T_1 = D_1^2/2$ and $T_2 = D_2^2/2$.

Fig.2.3 illustrates the properties of the system on the different time scales. The parameters are chosen as follows: $m_1 = m_2 = n_1 = n_2 = 1$, $v = 0$ and $u = 0, 0.5, 1.0, 1.5, 2.0$. Because we set $v = 0$, the only interaction of the two species is $A \rightarrow A + B$, i.e. A can generate B. ¿From the diagram, it is clear that when there is no interaction, the effective temperature is a constant, which means that the fluctuation-dissipation theorem holds. However when there is an interaction, the curves are not straight lines, but instead bend to the left. The larger the coupling, the more the curve deviates from the straight line. These curves indicate that the fluctuation-dissipation theorem is breaking down in a frequency dependent way.

2.4.2 A specific two-species example

Here we study a special case to illustrate that the effective temperature in some ways can still play the same role as temperature does in thermodynamics. The interactions are chosen as $u = v$, $m_1 = m_2 = 0$ and $n_1 = n_2 = 1$, i.e. $A \rightarrow B$ and $B \rightarrow A$. This means

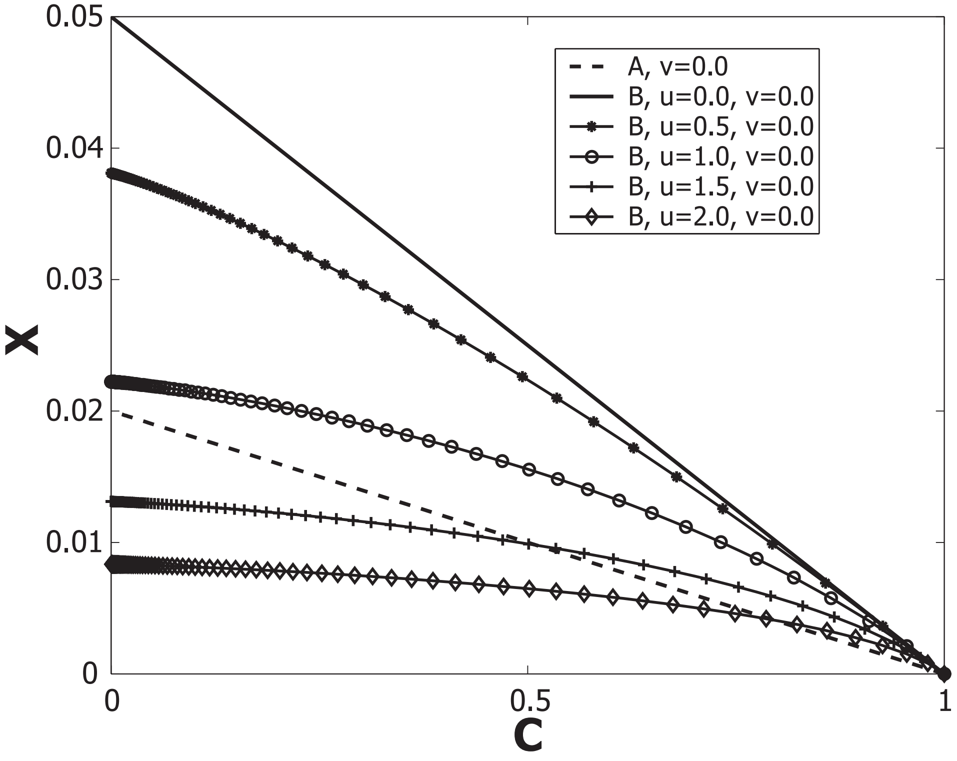

Figure 2.3: Illustration of multiple time scales properties of effective temperature. X axis is the scaled correlation, Y axis is the susceptibility. The dashed line represents A, the solid lines with symbols are for B. The solid lines with symbols from up to down correspond to the coupling strengths $u = 0, 0.5, 1.0, 1.5, 2.0$. The parameters are chosen as following: $k_{g1} = 50.0$, $k_{d1} = 1.0$, $k_{g2} = 20.0$, $k_{d2} = 1.0$, $m_1 = m_2 = n_1 = n_2 = 1$, $v = 0$.

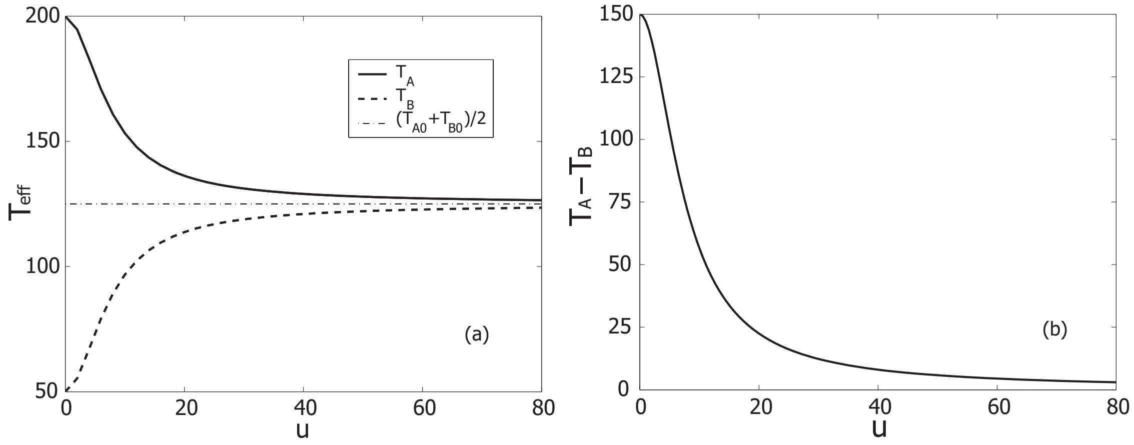

Figure 2.4: Effective temperature vs. interaction strength for the non-interacting case. Left Panel: The solid line, broken line and dash-dotted line stand for the effective temperature of the species A, the effective temperature of the species B, and the mean of the two effective temperature when there's no coupling. Right Panel: The difference of the two effective temperatures in the left panel versus the interaction strength. Here $k_{g1} = 200.0$, $k_{d1} = 1.0$, $k_{g2} = 50.0$, $k_{d2} = 1.0$, $m_1 = m_2 = 0$, $n_1 = n_2 = 1$, $u = v$ and $w = 10.0$.

that A reacts to become B and B reacts to become A.

To investigate the thermalization of the system, we focus on the difference of the two effective temperatures, i.e.

$$\Delta T_{eff}(\omega) = T_{eff}^{AA}(\omega) - T_{eff}^{BB}(\omega) \tag{2.35}$$

Together with the Eqs.(2.29, 2.30), the above equation tells us that as the coupling strength increases, the difference between the two effective temperatures decreases. That is, the two temperatures tend to equalize, as the 'hotter' one drops in temperature and the 'cooler' one increases in temperature.

Both Fig.(2.4) and Fig.(2.5) demonstrate this behavior. In Fig.(2.4) we assumed that the fluctuating terms are uncorrelated and their noise intensities are constant; that is, the terms D_3^a, D_3^b, D_4^a and D_4^b are zero and D_1 and D_2 are constants. The left panel shows that the 'hotter' species decreases in temperature and the 'cooler' species increases

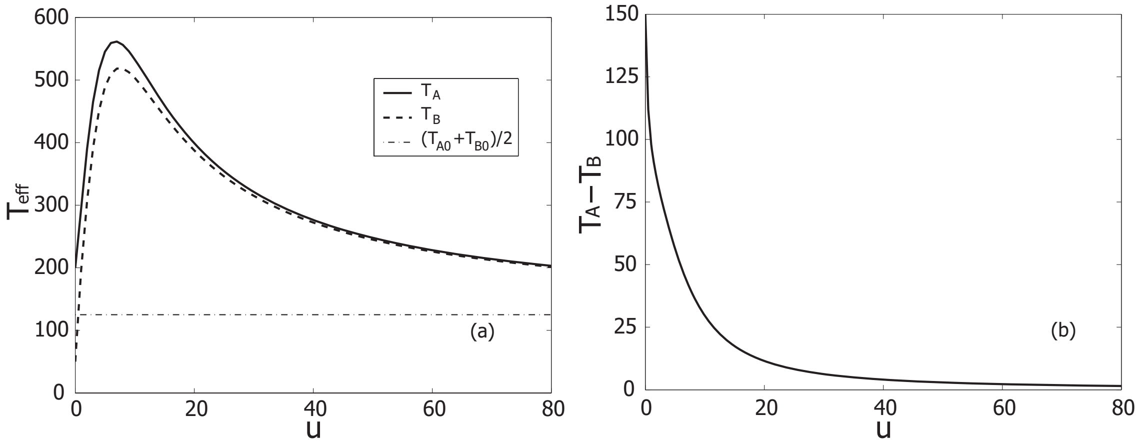

Figure 2.5: Effective temperature vs. interaction strength for the interacting case. Left Panel: The solid line, broken line and dash-dotted line stand for the effective temperature of the species A, the effective temperature of the species B, and the mean of the two effective temperature when there's no coupling. Right Panel: The difference of the two effective temperatures in the left panel versus the interaction strength. Here $k_{g1} = 200.0$, $k_{d1} = 1.0$, $k_{g2} = 50.0$, $k_{d2} = 1.0$, $m_1 = m_2 = 0$, $n_1 = n_2 = 1$, $u = v$ and $w = 10.0$.

in temperature monotonically with the increase of their interaction strength. When the coupling is strong enough, the two temperatures converge to the same value. The right panel clearly shows that the difference of the two temperatures decreases with increasing interaction strength.

Fig.(2.5) shows the result for the general correlated case. In the left panel we see that both of the temperatures increase at first and then drop after reaching a maximum, and ultimately they converge to the same value. The non-monotonic behavior, which is different from the uncorrelated case, comes from the correlation of the noise. The curves are similar to stochastic resonance curves[38]. Nevertheless, the difference of the two temperatures decreases monotonically, as seen in the right panel.

This example shows that the effective temperature in stochastic kinetics behaves much like ordinary temperature.

2.4.3 Remarks

The coupling between the two species as well as their own generation and degradation rates determine the effective temperatures. The effective temperatures are thus not always equal, but instead their discrepancy is system and interaction dependent. The effective temperature characterizes the properties and the state of that system. For example, with the physically meaningful coupling of the above example, the system holds the thermal properties: a) The flow goes from 'high' temperature to 'low' temperature, and b) if the coupling is strong enough, the temperatures of the two subsystems equalize.

The strategy for calculating effective temperature in solving Eqs.(2.20,2.21) is quite general, and it can be used to study more complex interaction networks involving multiple species, such as cascades.

2.5 Effective Temperature in Gene Networks

It has been suggested that the noise in gene networks can be broken down into intrinsic and extrinsic components[109]. If we consider the fluctuations of one particular species of a multiple species interacting network, then intrinsic noise originates from the stochastic nature of the reactions leading to expression of this species alone, whereas extrinsic variability arises from sources that effect the expression of all species. While this manner of noise classification is useful, it does have some limitations. For example, when there exists a correlation between the noise terms of the different species in a network, as there often is in gene networks, there is no simple way to separate the total noise into intrinsic and extrinsic components.

Fundamentally, the intrinsic-extrinsic classification is actually just the inverse

Fourier transform of the power spectrum of the correlation functions, because the noise intensity is just the autocorrelation function of the species. But the power spectrum is more general. For the two species case, Eq.(2.25) and Eq.(2.26) give the noise classification expressions. As seen from the expressions, there exist cross-correlation terms between intrinsic sources and extrinsic sources. The power spectrum expressions Eq.(2.25) and Eq.(2.26), beyond the intrinsic-extrinsic classification, include the correlated noise sources.

Based on what we know about effective temperature, it appears to be a good candidate for the quantitative analysis of gene networks. To explore this possibility, we will calculate the effective temperature of an unregulated gene present in many copies, and investigate what the effective temperature can tell us about this system.

We assume that there are N copies of a particular gene. Also, assuming that there are x number of bound operators, then the number of unbound operators is $N - x$. The operator sites are assumed to have no explicit regulation (i.e. the operator sites are fluctuating stochastically), and the state of each operator site (bound or unbound) will effect the rate of transcription. k_f and k_b are the binding and unbinding reaction rates for the operator, k_{g1} and k_{g2} are the protein generation rates of an bound operator and an unbound operator, and k_d is the degradation rate of the protein. Given these parameters, the Langevin equations describing the operators and the protein quantity, p, are:

$$\frac{dx}{dt} = -k_f x + k_b R_0 (N - x) - D_1 \eta_1(t) + D_2 \eta_2(t) \tag{2.36}$$

$$\frac{dp}{dt} = k_{g1} x + k_{g2}(N - x) - k_d p + D_3 \eta_3(t) + D_4 \eta_4(t) - D_5 \eta_5(t) \tag{2.37}$$

The noise intensity terms are $D_1 = \sqrt{k_f x}$, $D_2 = \sqrt{k_b R_0 (N - x)}$, $D_3 = \sqrt{k_{g1} x}$, $D_4 = \sqrt{k_{g2} x}$, and $D_5 = \sqrt{k_d p}$. The noise $\eta_i(t)$ (i=1,2,3,4,5) is uncorrelated normal Gaussian noise.

From the exact solutions Eqs.(2.29, 2.30) of the general equations Eqs.(2.20, 2.21), we find two different effective temperatures for the DNA and the protein number:

$$T_O(\omega) = T_1 \tag{2.38}$$

$$T_P(\omega) = T_2 + T_1 \frac{(k_{g1} - k_{g2})^2}{\omega^2 + (k_f + k_b)^2} \tag{2.39}$$

Here $T_1 = D_1^2 + D_2^2$ and $T_2 = D_3^2 + D_4^2 + D_5^2$ with the substitutions of x and p by the steady state values x^* and y^*. T_1 and T_2 are independent on the frequency.

These results show that the effective temperature of the DNA $T_O(\omega)$, T_1 is independent of the protein. The effective temperature of the protein $T_P(\omega)$, on the other hand, is composed of two parts, T_2 and $T_1 \frac{(k_{g1}-k_{g2})^2}{\omega^2+(k_f+k_b)^2}$, which means that the 'hotness' of the DNA actually effects the temperature of the protein. This is logical since the operator sites regulate the downstream production of the protein, but the protein does not regulate the production of the DNA.

To relate this to the notion of intrinsic and extrinsic noise, we can study the proteins as our system. Using this system, the fluctuation of the proteins due to the stochasticity of the chemical reaction Eq.(2.37) is the intrinsic noise, while the the fluctuation of the DNA operator is the extrinsic noise. Using a standard method to calculate the total noise [109, 102], we may write the total noise for this unregulated gene system in terms of intrinsic and extrinsic noises as:

$$\sigma_t^2 = \sigma_{in}^2 + \sigma_{ex}^2 \tag{2.40}$$

Following the same strategy that we used in the study of a two-species process, the equivalent power spectrum expression is:

$$\begin{aligned} C_p(\omega) &= \frac{D_3^2 + D_4^2 + D_5^2}{k_d^2 + \omega^2} + \frac{(k_{g1} - k_{g2})^2 (D_1^2 + D_1^2)}{[(k_f + k_b)^2 + \omega^2](k_d^2 + \omega^2)} \\ &= \frac{1}{k_d^2 + \omega^2} \cdot T_2 + \frac{(k_{g1} - k_{g2})^2}{[(k_f + k_b)^2 + \omega^2](k_d^2 + \omega^2)} \cdot T_1 \end{aligned} \tag{2.41}$$

Comparing the power spectrum expression with the effective temperature Eq.(2.39), we see that the effective temperature can function as an alternative means to quantify and analyze different sources of noise in gene networks. In this example, the temperature T_P divides noise into intrinsic and extrinsic components, but separates the noise source via time scale. In addition, because the effective temperature can measure the stability of a system [99], here T_{eff}^P tells how much of effective temperature (T_2) comes from the intrinsic noise (σ_{in}^2) and how much of the effective temperature ($T_1 \frac{(k_{g1}-k_{g2})^2}{\omega^2+(k_f+k_b)^2}$) comes from the extrinsic noise (σ_{ex}^2). This gives a measure of the contributions of intrinsic and extrinsic noises to the 'hotness' of the system, which is not merely a summation of intrinsic and extrinsic noise components. Moreover, the effective temperature explores the frequency selectivity of the system. From the expression of the effective temperature Eq.(2.39), we can see that the intrinsic noise always contributes to the system, but the extrinsic noise is frequency dependent. We see that the extrinsic noise plays an important role in the low frequency region, but it is filtered out in the high frequency region. This underlying property is masked by the power spectrum expression, demonstrating an advantage of using the effective temperature.

2.6 Conclusions and Discussions

Temperature is a central notion of thermodynamics, and the fluctuation-dissipation theorem is important in near equilibrium statistical mechanics. However, the theorem breaks down for stochastic dynamical kinetics and gene networks since such systems are far from equilibrium. The FD plot introduced in glass theory nevertheless provides an understanding of the interacting stochastic processes.

The effective temperature in general is observable dependent and therefore is not unique, but when the systems are large it has some of the same valuable properties as found when describing glasses [66]. The observable-dependence shows that a birth-death process in a steady state is nevertheless not an equilibrium system. Although there are multiple effective temperatures corresponding to different types of correlations, the properly chosen effective temperature can be used to explore the dynamic properties of a system. The autocorrelation effective temperature in a simple situation controls the flow between two interacting biochemical species.

Effective temperature provides an alternative language for discussing the origin of noises in stochastic cell biology. It goes beyond the intrinsic-extrinsic classification that has already been introduced and works in cases where the noise of various species is correlated.

Moreover, using the general definition of the effective temperature for out-of-equilibrium systems, it should also be applicable to more complex genetic circuits that do not relax to fixed steady states, such as repressor oscillators [33]. When spatial heterogeneities in real cells are considered, diffusion and transportation of regulatory proteins also can have a great impact [78]. The slowness of the ordered dynamics of the regulatory networks in comparison to the molecular events suggest the idea of effective temperature could be generally useful. Further developments of the concept of effective temperature as a means to characterize more complex gene networks will be needed, however, to increase our understanding of multi-gene regulation in organism.

Acknowledgements: This chapter contains materials in Lu T, Hasty J, and Wolynes P, Effective temperature in stochastic kinetics and gene networks. *Biophys. J.* 91:84-94 (2006).

3

Methodological Developments II: Generalized Gillespie Algorithm

3.1 Introduction

Successful completion of the Human Genome Project has led to the realization that effective models for predicting cellular behavior must take into account the dynamic network interactions that mediate gene regulation. Since behavior arising from these complex interactions is difficult to predict without quantitative models, there is a need for experimentally validated computational modeling approaches that can be used to understand the complexities of gene regulation. Such model approaches will be invaluable in the generation of logically consistent hypotheses and will provide a framework for the systematic comparison of data across multiple experiments.

There is strong experimental evidence that the level of expression from the same gene varies significantly from one cell to another within a genetically-identical colony [41, 33, 14, 81, 13, 32, 58]. Such macroscopic variations are routinely observed in the cells of organisms ranging in complexity from bacteria to mammals. Theoretically, with mRNA numbers that are often less than ten, the intrinsic noise of the underlying biochemical reactions implies that large fluctuations originating from small molecular numbers simply must exist. While the importance of fluctuations in this context was stressed over thirty years ago [108], the recent experimental evidence has led to a resurgence of interest in the utilization of stochastic simulation techniques to model gene regulatory networks[6, 52, 109, 58, 89].

When fluctuations arise from the small number of reactant molecules, the stochastic simulation algorithm developed by Gillespie is considered the "gold standard" for modeling [43, 45]. The advantage of this algorithm is that it provides an exact numerical simulation of the underlying biochemistry. However, both the original algorithm and later developments have focused on volume-fixed systems, and have not systematically considered the effect of volume changes on the simulation routine [50, 88, 42]. Here we extend the Gillespie simulation routine to account for time-dependent rate constants arising from cellular growth and division. Specifically, we address the physical background for the stochastic simulation routine from the microscopic level, and reinvestigate the derivation of the Gillespie routine. Although our results are general for simulations involving time-dependent rate constants, we focus on the generation of an algorithm for cellular systems with changing volume. Examples are provided to illustrate the algorithm, and we propose a criterion that can be used to determine if a simple replacement of the traditional Gillespie algorithm is sufficient with changing volume.

3.2 Background

Generally, a biochemical reaction occurs when a combination of reactant molecules collides with relevant speeds above a certain threshold in a short interval of time. Because of the randomness of molecular position and speed, Brownian motion of nonreactive molecules, temperature fluctuations, and other influences, stochasticity is an inherent property of the underlying biochemistry. When the average cellular size is small (i.e. a bacterium), it is commonly assumed that the cell acts as a well-mixed bioreactor. While even in small cells there are specific gene regulatory processes where spatial compartmentalization is the first-order consideration, there is no evidence to date that such spatial effects dominate in a generic sense.

When spacial structure of the cell is not taken into account the current state of the system may be represented by the number of reactant molecules of each kind present in the cell along with the physical variables characterizing the state of the cell as a whole, such as volume, temperature, concentration of solute, etc. The sufficient condition for this description is the requirement that the chemical system inside cell is a well-stirred mixture, that is non-reactive collisions occur much more frequently than reactive collisions [45]. This assumption makes description of the system inherently *stochastic* in the sence that a given state of the system lacks positions and velocoties and therefore the evolution from this state can not be a deterministic process. Then the most complete description of stochastic system may be given in terms of the corresponding Master Equations.

It is easy to see that the probability $P_1(t,\mu)dt$ that a particular reaction R_μ among reactant molecules will occur in the next time interval $[t, t + dt]$ depends on the volume of the reactor

$$P(t, \mu)dt \quad = \quad \frac{a_\mu}{[V(t)]^{r_\mu - 1}} dt \tag{3.1}$$

The probability density $P(t, \mu)$ is usually called the propensity of the chemical reaction R_μ.

Living cells do not grow indefinitely: they reach a certain size and then divide. While this process itself is not deterministic, in order to illustrate our approach we will make the simplifying assumption that the cell division time is constant, and that all molecules are divided evenly among the daughter cells.

3.3 Modified Gillespie Algorithm for Reactions in Variable Volume

Let us now consider necessary modifications to Gillespie algorithm in the case of variable volume of the cell. Suppose the volume $V(t)$ contains a spatially homogeneous mixture of $X_i, i = 1..N$ species which may interact through the reaction channels $R_\mu, \mu = 1..M$. Next assume that a subset of these channels is characterized by the time-dependent propensities $a_s(t) = a'_s/V(t), s = 1..S$, while the propensities of the remaining channels do not depend on time, denote those $a_q, q = S + 1..M$. We normalize time so that the volume of cells doubles in a unit time interval, after which the cell divides.

Following Gillespie [45] we introduce the following probabilities,

- $P(\tau, \mu|Y, t)d\tau$ - probability that, given the state $Y = (X_1, \ldots, X_N)$ at time t, the next reaction in $V(t)$ will occur in the infinitesimal time interval $(t + \tau, t + \tau + d\tau)$, and it will be reaction R_μ

- $a_\mu(t)dt$ - probability that, given the state $Y = (X_1, \ldots, X_N)$ at time t, reaction R_μ will occur in V within the interval $(t, t+dt)$.

We compute $P(\tau, \mu|Y, t)d\tau$ as a product of the probabilities that no reaction will occur within $(t, t+\tau)$ times the probability that R_μ will occur within the subsequent interval $(t+\tau, t+\tau+d\tau)$

$$P(\tau, \mu|Y, t)d\tau = P_0(\tau|Y, t) \cdot a_\mu(t+\tau)d\tau \tag{3.2}$$

To find $P_0(\tau|Y, t)$ we note that $[1 - d\tau \sum_\mu a_\mu(t+\tau)]$ is the probability that no reaction will occur during $d\tau$, hence

$$P_0(\tau+d\tau|Y, t) = P_0(\tau|Y, t)[1 - d\tau \sum_\mu a_\mu(t+\tau)] = P_0(\tau|Y, t)[1 - d\tau \sum_s a_s(t+\tau) - d\tau \sum_q a_q] \tag{3.3}$$

Using the initial condition $P_0(\tau = 0|Y, t) = 1$, we solve the differential equation (3.3) to find

$$P_0(\tau|Y, t) = \exp[-\sum_s \int_t^{t+\tau} d\tau' a_s(t+\tau')] \cdot \exp[-\tau \sum_q a_q] \tag{3.4}$$

Next we combine Eqs.(3.2,3.4),

$$P(\tau, \mu|Y, t) = a_\mu(t+\tau) \cdot \exp[-\sum_s \int_t^{t+\tau} d\tau' a_s(t+\tau')] \cdot \exp[-\tau \sum_q a_q] \tag{3.5}$$

Now we specify the dependence of a_s on time explicitly, assuming that $V(t+\tau) = V(t)\exp[c\tau]$ with $c = \ln(2)$.

Performing the integration in Eq.(3.5) we find,

$$P(\tau, \mu|Y, t) = \begin{cases} a_s(t)\exp[-c\tau] \cdot \exp[-f_c(\tau)A_s - \tau A_q] & s = 1..S, \\ \\ a_q \cdot \exp[-f_c(\tau)A_s - \tau A_q] & q = S+1..Q, \end{cases} \tag{3.6}$$

with

$$A_s = \sum_s a_s(t), \quad A_q = \sum_q a_q, \quad f_c(\tau) = (1 - \exp[-c\tau])/c, \tag{3.7}$$

It is easy to compute the probability of any reaction to occur between time t and $t+T$. Integrating Eq. (3.6) over time and summing over all channels, we find,

$$\int_0^T \sum_{\mu} d\tau P(\tau,\mu|Y,t) = 1 - \exp[-A_s f_c(T) - TA_q] \tag{3.8}$$

The limit of this equation when $T \to \infty$ yields the probability that any reaction will occur after time t.

$$N_P = \begin{cases} 1, & A_q \neq 0, \\ \\ 1 - \exp[-A_s/c], & A_q = 0 \end{cases} \tag{3.9}$$

This probability is 1 when at least one time-independent channel is present ($A_q \neq 0$). However if $A_q = 0$, this probability is less than one because there is a finite probability $\exp[-A_s/c]$ that no reaction will ever occur due to the exponentially decaying propensity of all of the reactions.

When all channels are time-independent, $a_s = 0, \quad s = 1..S$ Eq.(3.6) is reduced to the standard formula derived by Gillespie [45],

$$P(\tau,\mu|Y,t) = a_\mu \cdot \exp[-\tau \sum_{\mu} a_\mu], \tag{3.10}$$

where the summation over μ now includes all channels. We note that the same result is recovered in the formal limit $c \to 0$ and $\lim_{c\to 0} f_c(\tau) = \tau$, which corresponds to time-independent volume.

We should note that in fact the cell volume grows exponentially only until it doubles from its original value V_0, after which the cell divides and the volume is reset to V_0. So we may only use formula (3.5) for times $t+\tau < t_n$, the next cell division time, (or equivalently, for $\mod(t+\tau) < 1$)

In order to implement Direct Gillespie algorithm we must address two questions: when the next reaction will occur and which reaction it will be.

Next reaction time. It is convenient to distinguish three possible scenarios, depending on the presence of time-dependent channels.

Case (1), $A_q \neq 0$, $A_s \neq 0$. Here we have both types of reactions and use Eq.(3.6). To find the time of the next reaction, $t+\tau$, we use the inversion method [46]. According to this method, in order to map the uniform random number (URN) to a number from a distribution with a given probability density function (PDF), one has to obtain a distribution from this PDF, draw a URN, set the distribution function equal to this number, and invert the equation. Using Eq.(3.6), and summing $P(\tau, \mu|Y, t)$ over all channels and integrating over time up to $\tau' = \tau$, we find the distribution function,

$$F(\tau, |Y, t) = \int_0^\tau d\tau' \sum_\mu P(\tau', \mu|Y, t) = 1 - \exp[-f_c(\tau)A_s - A_q\tau] \tag{3.11}$$

Let u_1 be a URN, then the time interval until the next reaction is given by the solution of the transcendental equation ($1 - u_1$ is also a URN),

$$\exp[-f_c(\tau)A_s - A_q\tau] = u_1 \tag{3.12}$$

which may be expressed using a Lambert function W [1] as

$$c\tau = W(\alpha \exp[\alpha + \beta]) - \alpha - \beta, \qquad \text{with} \quad \alpha = A_s/A_q, \quad \beta = c\ln(u_1)/A_q \tag{3.13}$$

If the time of the next reaction $t + \tau$ with τ obtained from Eq.(3.13) is less than the time of the next cell division t_n, the time is advanced to $t + \tau$ and one of the reactions is selected (see below). However, the time of the next reaction calculated using τ from Eq.(3.13) may exceed the time of the next cell division. In this case we simply advance time to the next cell division time t_n, divide the volume and the numbers of proteins by

[1] Lambert function, $W(x)$ is a solution of the equation $y\exp[y] = x$. Here we use the nonnegative part of the principal branch of W. It is of the same complexity as the log function and may be effectively evaluated for $x \geq 0$. For details see R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth, "On The Lambert W Function". Adv. Comp. Math. **5**, pp. 329-359, (1996)

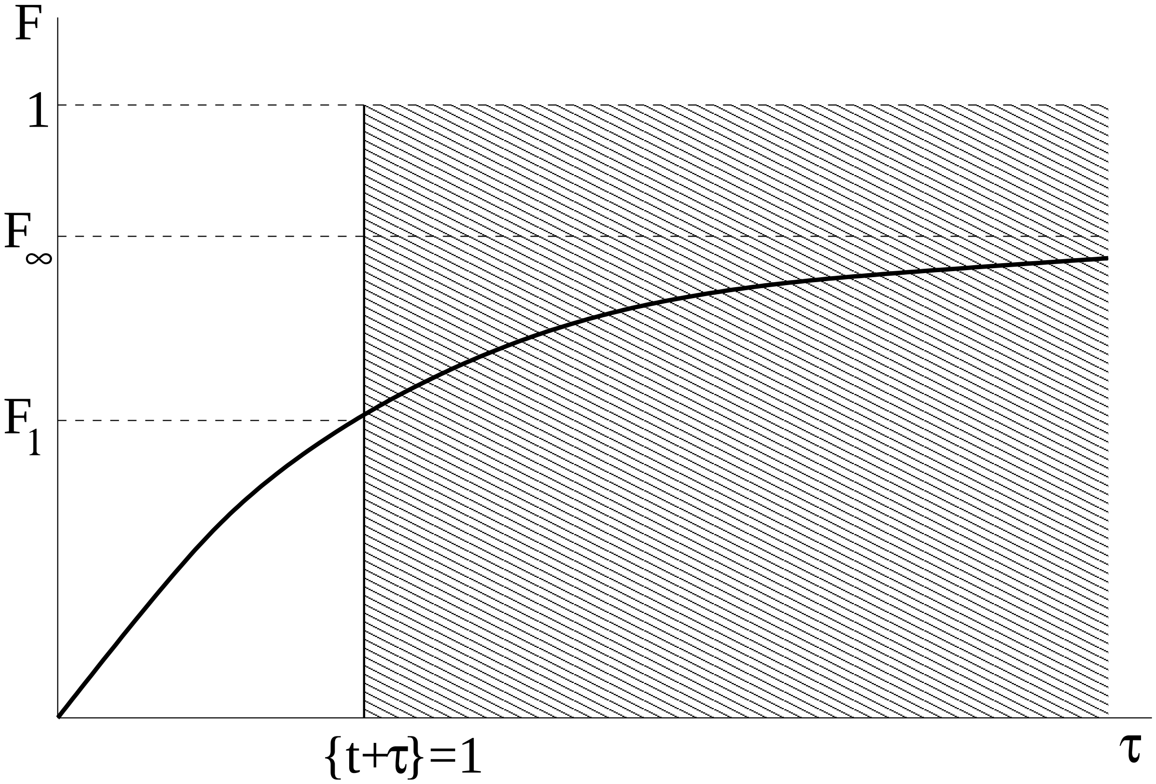

Figure 3.1: Distribution function of time until the next reaction for Case 2 when no time-independent channels are present. If URN u_1 is greater than F_1, instead of selecting the next reaction, the cell division at time $[t] + 1$ is performed.

two (more generally, this division may be non-even and chosen from a binomial distribution or partitioned in accordance with some empirically determined distribution), and select τ again.

Case (2), $A_q = 0$, $A_s \neq 0$. In this case only time-dependent channels are present and we use Eq.(3.6) with $a_q = 0, \quad q = S + 1..M$. As a result we obtain

$$c\tau = \ln\left[\frac{A_s}{A_s + c\ln(1 - u_1)}\right] \tag{3.14}$$

This case is special because there is a finite possibility that no reaction will happen in the interval $\tau = [0, \infty]$.

Formally, Eq.(3.14) has no solution τ for $u_1 > F_\infty = 1 - \exp[-A_s/c]$ (see Figure 3.1). In this case, as in the case when the solution τ exists but $\{t + \tau\} > 1$, no reaction is implemented but time is advanced to the time of the next cell division (to t_n), and the volume and number of proteins are reset.

Case (3), $A_q \neq 0$, $A_s = 0$. This is the standard situation covered by Gillespie algorithm, and the time to the next reaction is given by

$$\tau = -\frac{1}{A_q}\ln u_1 \tag{3.15}$$

Which reaction to choose. This step is similar to the standard DG algorithm. The only difference comes from the fact that in the interval $[t, t + \tau]$, the time-dependent propensities change due to cell growth, and thus one has to choose which reaction will occur based on the propensities at time $t + \tau$. Using a second URN u_2 we find the channel $\nu = \mu$

$$\sum_{\nu=1}^{\mu-1} a_\mu(t+\tau) < u_2(A_q + A_s(t+\tau)) \leq \sum_{\nu=1}^{\mu} a_\mu(t+\tau), \quad 1 \leq \nu \leq M \tag{3.16}$$

To summarize, the modified Direct Gillespie algorithm contains the following steps:

1. Input values for c_μ , $\mu = 1, \ldots, M$ and initial state $(x_1, \ldots, x_N)$, set $t = 0$.

2. Compute $a_\mu = h_\mu \cdot c_\mu$, along with $A_s = \sum_s a_s(t)$, $A_q = \sum_q a_q$, $s = 1..S$, $q = S+1..M$.

3. Generate uniform random numbers u_1, u_2.

4. Check if A_q is zero; if it is then use Eq.(3.14) to compute the time interval τ until the next reaction, otherwise compute τ according to Eq. (3.13)

5. Check whether $\mod (t + \tau) < 1$. If yes, go to the next step. If no, update time $t \to [t] + 1$, reset volume $V \to V_0$ and the number of proteins of each species in the cell $x_i \to x_i/2$, and return to step 2.

6. Find the channel of the next reaction μ using Eq. (3.16)

7. Update time $t \to t + \tau$, and adjust x_i in accordance with the particular reaction r_μ. Proceed to step 2.

Let us now discuss under what conditions our modified Time-dependent Direct Gillespie algorithm (TDG) is expected to yield results that differ from a naive application of the standard Direct Gillespie, which we will denote the Naive Direct Gillespie (NDG) approach. This naive approach arises from a natural assumption for generalizing the Gillespie algorithm, whereby the time-dependent rate constants are simply updated after each time step (i.e, the algorithm is not rederived). This corresponds to using the instantaneous values of the propensities associated with the current (at the moment of reaction) value of the cell volume. First consider the simple case with a single time-dependent channel.

Compare the distributions for the time interval until the next reaction,

$$P_1(\tau|t) = a(t)\exp[-f_c(\tau)a(t) - c\tau], \qquad (TDG) \tag{3.17}$$

$$P_0(\tau|t) = a(t)\exp[-a(t)\tau], \qquad (NDG) \tag{3.18}$$

In both cases the distribution functions decay exponentially at a rate proportional to $a(t)$. When $a(t) \gg 1$, in both cases the mean value of τ is much smaller than one. Additionally, $c\tau$ may be neglected as compared with the much larger term $f_c(\tau)a(t) = a(t)(\tau + O(\tau^2)) \approx a(t)\tau$. Therefore, for small values of τ, the distributions (3.17,3.18) are nearly the same. Thus, when the propensity a_s is large, the two algorithms should yield almost identical results. On the other hand, when $a(t)$ is small, the probability that a reaction will occur during the cell doubling time according to TDG algorithm approaches $a/(2\ln 2)$, whereas according to the naive NDG algorithm, it is a. Thus, the difference between these two algorithms for small a can be significant.

3.4 Example: Transcriptional Regulation without Feedback

In order to concretely illustrate the difference between the correct algorithm (TDG) and the naive approach (NDG), we now turn to a simple yet nontrivial example where analytic progress can be made. The example involves a single gene which fluctuates between two states S_0 and S_1. The transition $S_0 \to S_1$ occurs when a regulator protein (whose number is assumed to be constant throughout the cell cycle) is bound to the gene's promoter, and so this transition probability is inversely proportional to the cell volume v. Upon division of the cell, both its volume and the number of all proteins are halved. For simplicity, we assume that the number of regulator proteins quickly reaches a steady-state value. We neglect the fast relaxation of the number of regulator proteins, and assume that an effective propensity for the production of proteins is inversely proportional to the volume at any instance of time. The propensity of the reverse process $S_1 \rightarrow S_0$ is assumed independent of the cell volume. The gene S is producing protein X at a rate α_0 when it is in state S_0, and α_1 when it is in state S_1 (for definiteness, we assume $\alpha_0 < \alpha_1$). The biochemical reactions describing this model system are summarized in Table 1.

Table 3.1: Biochemical reactions for a simple system describing a constitutive promoter.

μ	reaction	a_μ
1	$S_0 \xrightarrow{k_1/v} S_1$	$k_1/v\, s_0$
2	$S_1 \xrightarrow{k_{-1}} S_0$	$k_{-1}\, s_1$
3	$S_0 \xrightarrow{\alpha_0} S_0 + X$	$\alpha_0\, s_0$
4	$S_1 \xrightarrow{\alpha_1} S_1 + X$	$\alpha_1\, s_1$
5	$X \xrightarrow{k_x}$	$k_x\, x$

Here $v = \exp[\ln 2\ t/T_0]$, and T_0 is the cell division time. At times $t = nT_0$, the volume v is halved $v \rightarrow v/2$, and the number of proteins is halved, $x \rightarrow x/2$. In the case of constant volume, this single-gene constitutive model was explored by Kepler and Elston [62] using a master equation for the time-dependent probability p_x^s of having both the promoter in the state $s = [0, 1]$ and x proteins. Using a similar approach for growing and dividing cells, we obtain equations for the dynamics of the partial moments $M_q^s \equiv \langle x^q \rangle_s = \sum_x x^q p_x^s$ (see Appendix). The zeroth moments M_0^s represent the probabilities for the promoter to be in the sth state. The sum of the first moments $M_1 = M_1^0 + M_1^1$ is the average number of proteins $\langle x \rangle$, and $Var = M_2^0 + M_2^1 - (M_1^0 + M_1^1)^2$ is the variance of the number of proteins. In the case of constant volume, these moments reach a steady-state [62], but when the volume is allowed to oscillate, the moments asymptotically approach an oscillatory regime as one might expect (see below Figure 3.2). The time-averaged probability of finding the promoter in an unbound (bound) state M_0^0 (M_0^1) can be accurately approximated by the formulas $M_0^0 = k_{-1}/(k_{-1} + 0.72k_1)$ and $M_0^1 = 1 - M_0^0$.

Let us now describe the results of the simulations of the reactions listed in Table 3.4. In this simple example, the fast reactions $3 - 5$ are computationally expensive in the direct Gillespie algorithm, since the average time step between these reactions is very small compared with the cell division time. In order to make progress on realistic problems involving many such fast reactions, a more natural approach is a hybrid simulation technique [1, 50], where the dynamics of the fast subset is modelled either deterministically or using Langevin equations, while the slow reactions are modelled with Gillespie. We adopt this approach and simulate the dynamics of the proteins X using a Langevin equation (B.6). This equation was integrated using the Euler-Murayama method [39].

The differences between the use of the standard and modified Gillespie algorithm for the slow reaction is evident in the distribution of residence times $p_0(t_r)$, and the distribution of phases $p(\theta)$ describing transitions from the S_0 to S_1 state within a unit cell-doubling interval, where $\theta = \mod(t, 1) \in [0, 1]$ (since the propensity of the inverse reaction is independent of volume, the residence time distribution for the bound state is simply described by the Poisson statistics).

For small k_1, most of the time step 2 of the algorithm yields negative result (randomly selected time $t + \tau$ exceeds the next cell division time), so the time will be advanced to the next cell division time and another draw is performed. Therefore, most of the transitions from S_0 to S_1 will be selected at the cell division time when $V = 1$. Then it is easy to see from (3.17), (3.18) that the phase distributions for the two versions of Gillespie algorithm are

$$p_0(\theta) \approx 1 \quad \text{(NDG)}, \quad p_1(\theta) \approx 2^{1-\theta} \ln 2 \quad \text{(TDG)} \tag{3.19}$$

Accordingly, the distribution of residence times at small k_1 is mostly determined by the number of unit time intervals during which the transition $S_0 \rightarrow S_1$ does *not* occur. The probability of the transition to not occur per unit time is e^{-k_1} for NDG algorithm and $\exp(-k_1/2\ln 2)$ for TDG algorithm, thus for long times the same distribution scales as

$$p_0(t_r) \propto e^{-k_1 t_r} \quad \text{(NDG)}, \quad p_1(t_r) \propto 2^{1-\theta} \ln 2 \quad \text{(TDG)} \tag{3.20}$$

In Figure 3.2 we show the time dependencies of ensemble-averaged promoter state s_1 (a,d) and protein concentration $\langle x \rangle / v$ (b,c,e,f) obtained numerically using the standard and modified Direct Gillespie algorithms for four different sets of parameters. In the same plots we put single realizations of s and x. As seen for the plots, modified Gillespie algorithm yields the results virtually indistinguishable from the theoretical curves. For comparison, Figure 3.2a,d also show the mean values of s_1 as a function of time obtained using standard Gillespie algorithm based on instantaneous volume. In this case visible deviations from theoretical dependencies are obtained.

Figure 3.3 shows the time-averaged characteristics $\overline{s_0}, \overline{\langle x \rangle}, \overline{Var}$ as functions of the forward propensity k_1 for $\alpha_0 = 100, \alpha_1 = 500, k_x = 10, k_{-1} = 1$. As seen for this Figure, TDG simulations are in excellent agreement with master equation analysis, whereas standard NDG simulations show systematic deviations.

Figure 3.4,a illustrates the distributions of phases θ of "forward" reaction $S_0 \rightarrow S_1$ within a single cell cycle and Figure 3.4,b shows the distribution of residence times t_r in S_0 state. These distributions obtained numerically with NDG and TDG algorithms are in good agreement with theoretical predictions (3.19), (3.20) for small k_1.

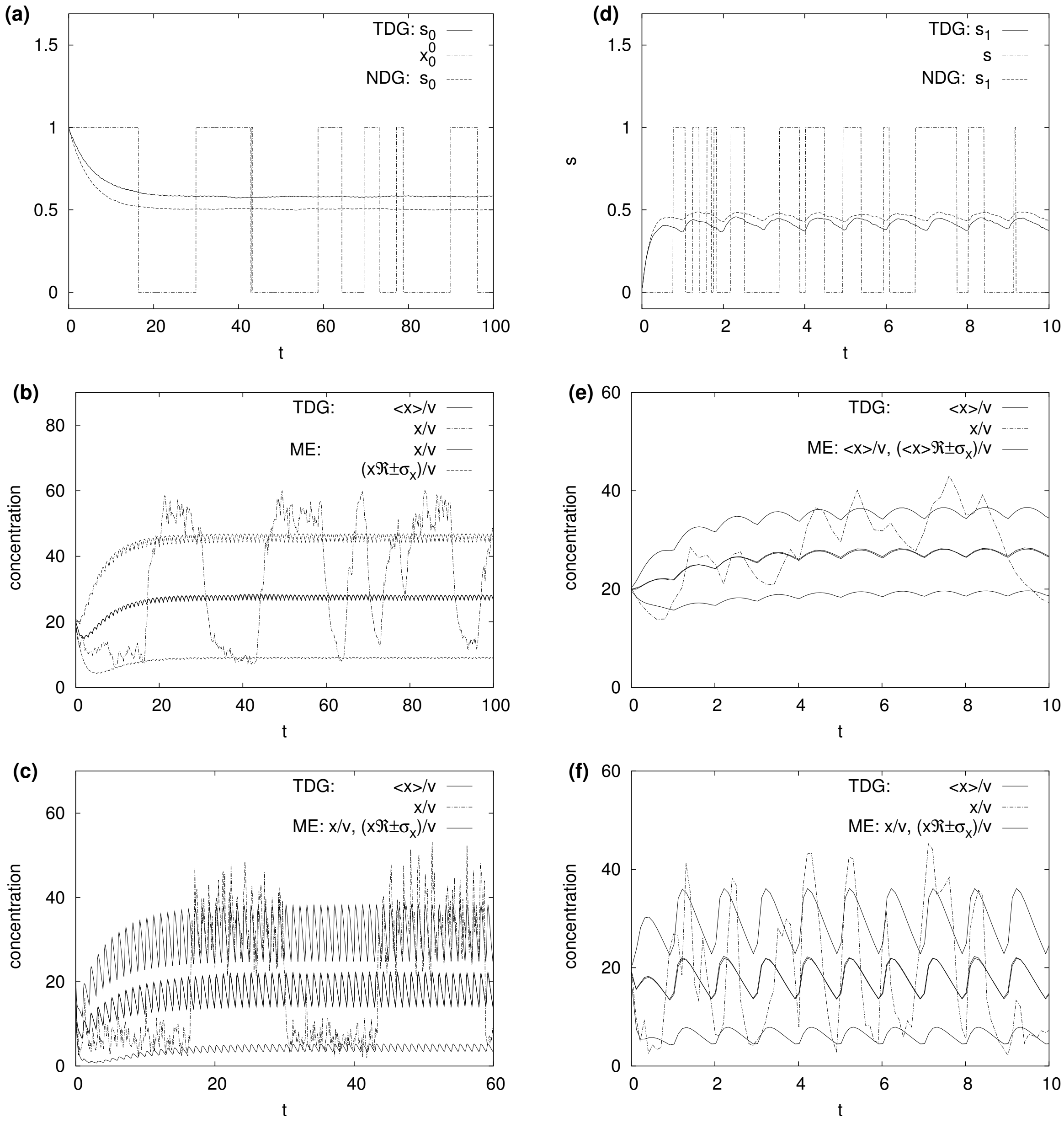

Figure 3.2: Time series of the probability of state S_1 of the promoter, s_1, (a,d) and the concentration of proteins $\langle x\rangle/v$ (b,c,e,f) obtained with NDG and TDG algorithms and theoretically using master equation approach (ME) for different parameter values. Left column: $k_1 = k_{-1} = 0.1$, right column $k_1 = k_{-1} = 10$, second row: $k_x = 0.01, \alpha_0 = 10, \alpha_1 = 50$, third row: $k_x = 10, \alpha_0 = 100, \alpha_1 = 500$. Dashed lines in panels (b,c,e,f) show the $(\langle x\rangle \pm \sigma_x)/v$ range of concentration fluctuations. Dash-dotted lines show a single typical trajectory of the stochastic system. The TDG time series virtually coincide with theoretical curves.

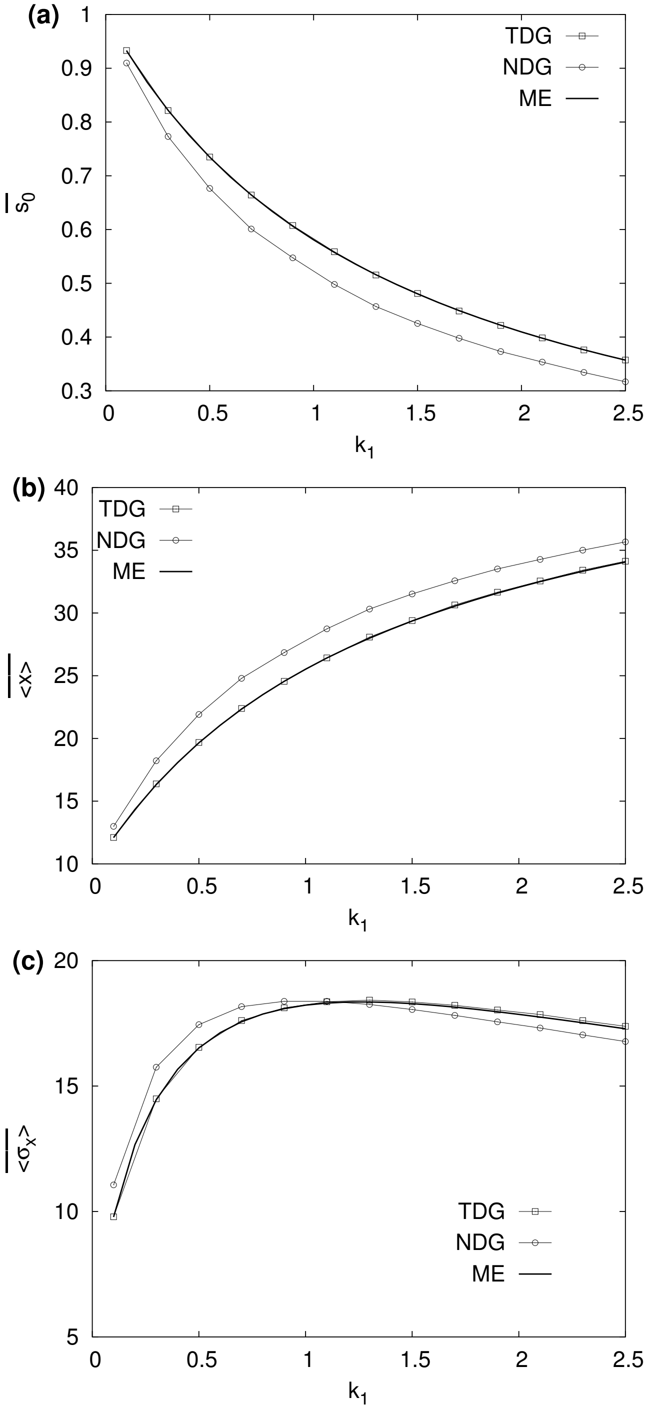

Figure 3.3: Time-averaged values of s_0, $\langle x \rangle$ (b) and standard deviation σ_x as a function of "forward" rate k_1 for $\alpha_0 = 100, \alpha_1 = 500, k_x = 10, k_{-1} = 1$.

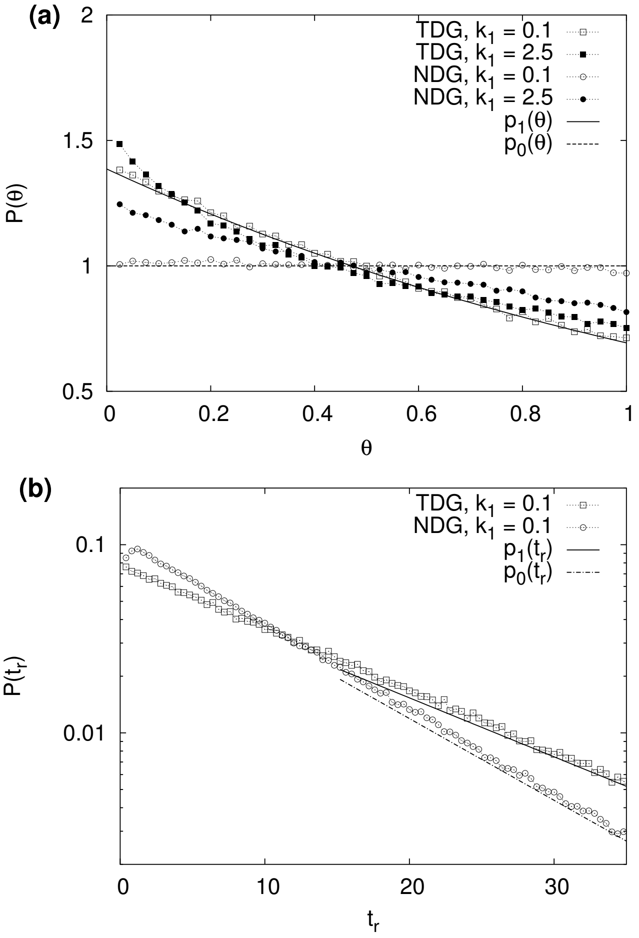

Figure 3.4: a - distribution of phases θ of "forward" reactions $S_0 \rightarrow S_1$ within a single cell cycle for $k_1 = k_{-1} = 0.1; 2.5$; b - distribution of residence times t_r in S_0 state for $k_1 = k_{-1} = 0.1$. Symbols show NDG and TDG simulations, and lines are plotted using theoretical formulas (3.19),(3.20)

3.5 Discussion

In this work, we have derived a generalisation of the Gillespie algorithm to account for cellular growth and division, and compared this algorithm with an adiabatic Gillespie routine where the volume is simply updated as time progresses. While the adiabatic routine was shown to be quite accurate if any of the chemical reactions are fast compared with the growth rate, it is typically not feasible in realistic settings to simulate all reactions with a pure Gillespie-type routine. We therefore focused our comparison on a hybrid simulation technique [17, 24], whereby the fast reactions were simulated with Langevin equations, and the slow reactions simulated with Gillespie method. In these simulations, the average time between random events may be significant as compared with the cell division time, and here we were able to demonstrate the necessity of using our generalisation of the Gillespie algorithm. Importantly, the generalised algorithm does not significantly increase the computational expense, so this derived method is preferred regardless of accuracy considerations. Our method specifies how a periodic deterministic event, namely cell division, can be incorporated into the Gillespie routine. While in order to illustrate our approach, we considered deterministic volume growth and division, future work could focus on additional sources of noise, such as a stochastic growth rate, fluctuations arising from unequal partitioning of molecules at cell division, or variations in the DNA replication time before division. As experiments begin to discriminate between the sources of noise in the cellular environment, simulation routines that correctly incorporate individual noise sources will be increasingly useful.

Acknowledgements: This chapter contains materials in Lu T, Volfson D, Tsimring L, and

Hasty J, Cellular growth and division in the Gillespie algorithm, *Systems Biology* 1:121–128, (2007).

4

Ladder Reaction Network

Achieving a quantitative understanding of the reaction networks that transduce cellular signals is one of the major challenges in biology. Signaling networks are found in a diverse set of organisms, ranging from prokaryotes to eukaryotes, and provide mechanisms for fundamental processes such as gene-regulatory control and cellular communication. Qualitative descriptions of the biomolecular components and mechanisms of cellular signaling have greatly improved our understanding of how cells function and have given insights into how to intervene therapeutically when such signals are miscommunicated. Experimental advances now allow quantitative studies of signal transduction and thereby inspire theoretical treatments. Many networks of nonlinear reactions exhibit interesting behavioral features as ultrasensitivity, adaption, robustness, and discrete "all-or-none" response which have been quantitatively explored [56, 18, 35, 36, 93, 97, 27, 16].

A commonly occurring network topology is the *reaction ladder network*. This network may be viewed as a generalization of multiple-site phosphorylation/dephosphorylation cascades, such as the pathway governing nuclear factor activation of T-cells (NFAT), which regulates the response of T-cells to antigen signaling [22, 54, 55, 100, 94]. To stimulate T-cells, NFAT must be transported to the nucleus. This transition occurs in response to a conformational change that exposes a nuclear localization sequence (NLS), which is normally buried in the protein interior in the inactive conformation and thus makes the NFAT inaccessible to transport by importin. The NLS becomes exposed in response to the progressive dephosphorylation of specific serine residues in its regulatory domain. This dephosphorylation occurs in response to an increase of intracellular calcium ions that activate calcineurin, which then dephosphorylates the masking residues. Once a sufficient number of sites have been dephosphorylated, conformational changes expose the NLS so that it can now be transported into the nucleus by the importin. This is not a one-way process. Inside the nucleus, the NFAT may be progressively re-phosphorylated by kinases and subsequently exported to the cytoplasm by the exportin Crm1 [54, 55]. To summarize this network generically, the NFAT can exist in a variety of phosphorylation states and at various locations within the cell. Transitions between these phosphorylation/compartmentalization states can be described as a network of reactions consisting of two groups of species C_i and U_i, where the C_i species reside in the cytoplasm and the U_i species reside in the nucleus (Fig. 4.1). On each side of the cytonuclear barrier there are $M + 1$ species having different levels of phosphorylation. This network topology can generally be interpreted as representing either processive or distributive mechanisms of phosphorylation [48, 7]. If the subscript i labels a specific order of phosphorylation, e.g., phosphorylation of residue A, followed by residue B, followed by residue C. . ., the network describes the processive (de)phosphorylation; if i represents for the number of (de)phosphorylated residues, the network describes the distributive phosphorylation mechanism, e.g., one residue is first phosphorylated, followed by two residues, then three residues.... Of course the rates connecting i and $i+1$ will be different for the two mechanisms, but the network topology remains the same. In the following, we study the processive phosphorylation of residues.

The rates of both phosphorylation (by kinases) and dephosphorlyation (by phosphatases) are naturally modeled as Michaelis-Menten reactions with rates that depend on the availability of enzymes. Although dephosphorylations dominantly occur in the cytoplasm and phosphorylations dominantly occur in the nucleus in the case of NFAT signaling [22, 54, 55], in our model we allow them both to occur in either environment with different rates. The protein phosphorylation state by affecting the conformation of that protein determines how easily it is translocated into or out of the nucleus. In our model k_+^i represents the reaction rate from C_i to U_i and k_-^i is the rate of going from U_i to C_i. An individual NFAT molecule thus makes random walks through its prosphorylation/location space according to these microscopic reaction rates.

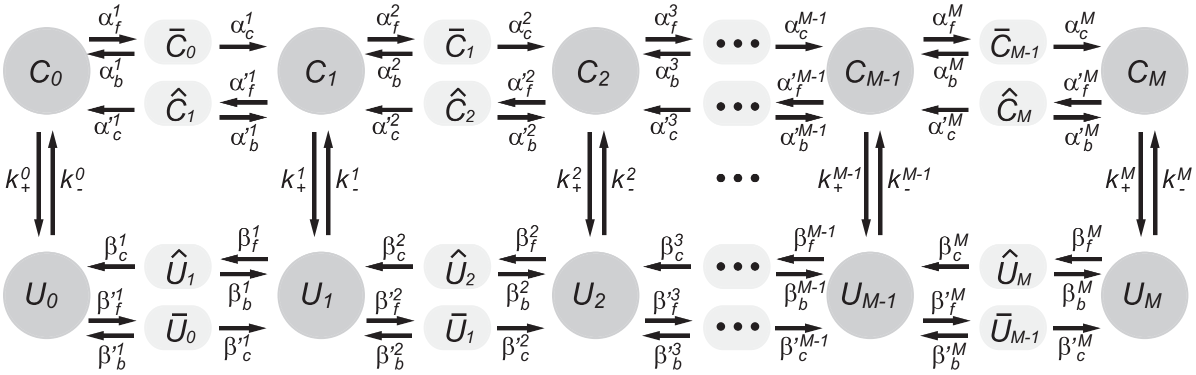
C0
C1
C2
CM-1
CM
C̄0
C̄1
C̄M-1
Ĉ1
Ĉ2
ĈM
U0
U1
U2
UM-1
UM
Û1
Û2
ÛM
Ū0
Ū1
ŪM-1
α1f α1b α1c α'1c α'1f α'1b
α2f α2b α2c α'2c α'2f α'2b
α3f α3b α'3c
αM-1c α'M-1f α'M-1b
αMf αMb αMc α'Mc α'Mf α'Mb
k0+ k0-
k1+ k1-
k2+ k2-
kM-1+ kM-1-
kM+ kM-
β1c β'1f β'1b β1f β1b β'1c
β2c β'2f β'2b β2f β2b β'2c
β3c β'3f β'3b
βM-1f βM-1b β'M-1c
βMc β'Mf β'Mb βMf βMb β'Mc

Figure 4.1: The reaction ladder network. The C and U represent two distinct compartmental locations of the signaling molecules, say, cytoplasmic and nuclear regions. The subscript i ($i = 0, 1, 2, ..., M$) indicates the dephosphorylation states. The $\bar{C}_i$ and $\hat{C}_i$ ($\bar{U}_i$ and $\hat{U}_i$) are the signaling protein-enzyme complex forms. The subscripts f, b and c represent forward, backward and catalyzed rates of each reaction. The k_+^i and k_-^i are the transport rates for transitions between C_i and U_i.

The NFAT signaling network shares its topology with many other networks in the cell. It simplifies to a typical two-state system if there is only a single phosphorylation state; the network resembles an enzymatic futile cycle when there are two phosphorylation states but no compartmental transport [96]; and the network is also similar to the Monod-Wyman-Changeux Model [80]. The reaction ladder network thus presents a paradigm for the interplay of spatial heterogeneity and post-transcriptional modification in the flow of biological information. The modification reactions, conformational changes and intercompartmental transports, are intrinsically stochastic events. On the ladder network each signaling molecule follows a path through the network, causing transcription to occur at random times. When the modifying enzymes are abundant, the network is effectively linear and each molecule walks through location/phophorylation space independently. When enzymes are limited in number, the network becomes nonlinear and the walks of different signaling molecules interact by competing for enzymes. Ultimately it takes only one NFAT to turn on its target gene. Thus we can say, somewhat anthropomorphically, that the individual NFAT's are competing in a race to the DNA. We thus have a problem of determining the statistics of mean first passage for multiple walkers.

The statistical problem of calculating the mean first-passage time for random walkers has a distinguished history [77, 113, 70, 15, 123]. Here we will find the first passage time and the survival probability in terms of a dynamic probability distribution. We will then show that the exact solution of this problem of multiple random walkers having the same goal does indeed agree with the results of Monte Carlo simulation of the network when enzymes are unlimited. The mean first passage time distribution is found to be asymmetric and has a long tail. The solution also shows there is an optimal forward reaction rate yielding the most rapid arrival of a viable signaling protein to the target.

4.1 The Distribution of First Passage Times and Survival Probability

Different phosphorylation and dephosphorylation processes are catalyzed by specific enzymes [87, 53]. For simplicity of treatment, we assume a universal kinase and a universal phosphatase, K_c and F_c, in the cytoplasm and another set, K_u and F_u, in the nucleus. Also for simplicity we assume only apo proteins can be transported between the cytoplasm and the nucleus. The enzymes (K_c, F_c, K_u and F_u) as well as the signaling protein-phosphatase complexes ($\bar{C}_i$ or $\bar{U}_i$) and the signaling protein-kinase complexes ($\hat{C}_i$ or $\hat{U}_i$) cannot be transported.

Suppose that there are c_i, $\bar{c}_i$ and $\hat{c}_i$ proteins in the C_i, $\bar{C}_i$ and $\hat{C}_i$ states and u_i, $\bar{u}_i$ and $\hat{u}_i$ proteins in the U_i, $\bar{U}_i$ and $\hat{U}_i$ states respectively ($i = 0, 1, 2, ..., M$). Then, the numbers of proteins in each state is described by a $6M + 6$ dimensional vector $\vec{\mathbf{n}} \equiv (\hat{c}_0, c_0, \bar{c}_0, \hat{u}_0, u_0, \bar{u}_0; \, ...; \, \hat{c}_M, c_M, \bar{u}_M, \hat{u}_M, u_M, \bar{u}_M)$, where $\hat{c}_0$, $\bar{u}_0$, $\hat{u}_M$ and $\bar{c}_M$ are zeros due to the boundaries of the network. The numbers of the enzymes in the system are defined by a four dimensional vector $\vec{\mathbf{E}} = (F_c, K_c, F_u, K_u)$. We define a state $|\Psi(t)\rangle$ as $|\Psi(t)\rangle \equiv \sum_{\vec{\mathbf{n}}, \vec{\mathbf{E}}} P(\vec{\mathbf{n}}, \vec{\mathbf{E}}, t)|\vec{\mathbf{n}}, \vec{\mathbf{E}}\rangle$, where $P(\vec{\mathbf{n}}, \vec{\mathbf{E}}, t)$ is the probability having $\vec{\mathbf{n}}$ and $\vec{\mathbf{E}}$ numbers of proteins and enzymes in the network. The Master equation describing the network can then be written as $\frac{\partial}{\partial t}|\Psi(t)\rangle = W|\Psi(t)\rangle$, where W is the transition rate matrix whose dimension depends on the total numbers of enzymes and substrates.

Generally solving this Master equation represents a challenging many-body problem. However, when the numbers of enzymes in the cytoplasm and the nucleus are very large compared to the total number of signaling proteins, as often happens in real biological systems, the phosphorylation and dephosphorylation processes which lead to the transitions of the signaling molecules are uncorrelated. Each protein can then be modeled as an independent random walker. Our assumption of an enzyme-saturated situation makes the mathematics of the network relative simple and the problem of multiple but independent random walkers can be solved exactly. This exact solution allows several interesting properties of the network to be explored.

A key aspect characterizing signaling pathways is the time to achieve a response after receiving an upstream signal, i.e., the typical delay time between a stimulus and the corresponding response. This is a stochastic quantity. The response occurs when one of the random walkers successfully binds to the DNA. To quantify this, we may consider the first passage time for a random walker starting from the initial position $\vec{r}_i$, arriving at the final position $\vec{r}_f$ for the first time. $F(\vec{r}_i, \vec{r}_f, t)$ is the probability distribution of such a random walker, initially in $\vec{r}_i$, whose first passage time of reaching the final position $\vec{r}_f$ is time t. $F(\vec{r}_i, \vec{r}_f, t)$ is related to the occupancy probability $P(\vec{r}_i, \vec{r}_f, t)$, which is the probability that a particle is found at the position $\vec{r}_f$ at time t irrespective of when it arrived. This relation is

$$P(\vec{r}_i, \vec{r}_f, t) = \int_0^t d\tau F(\vec{r}_i, \vec{r}_f, \tau) P^s(\vec{r}_f, \vec{r}_f, t-\tau) \tag{4.1}$$

Both the first passage time probability $F(\vec{r}_i, \vec{r}_f, \tau)$ and occupancy probability $P(\vec{r}_f, \vec{r}_f, t)$ are normalized through $\int dt F(\vec{r}_i, \vec{r}_f, t) = 1$ and $\int d\vec{r}_f P(\vec{r}_i, \vec{r}_f, t) = 1$. $P^s(\vec{r}, \vec{r}, t)$ is the occupancy probability with the identical initial and final position, i.e., the chance of a particle staying at and returning to the same position $\vec{r}$ after time t. In terms of Laplace transforms, the above equation can then be rewritten as

$$F(\vec{r}_i, \vec{r}_f, t) = \mathcal{L}^{-1}\left\{\frac{\mathcal{L}\{P(\vec{r}_i, \vec{r}_f, t)\}}{\mathcal{L}\{P^{(s)}(\vec{r}_f, \vec{r}_f, t-\tau)\}}\right\} \tag{4.2}$$

The survival probability is the probability that up to time t the random walker still has never reached the target position $\vec{r}_f$, which is represented as $S(\vec{r}_i, \vec{r}_f, t)$. By the definition, the survival probability S is

$$S(\vec{r}_i, \vec{r}_f, t) = 1 - \int_0^t d\tau F(\vec{r}_i, \vec{r}_f, \tau) \tag{4.3}$$

The above first passage time probability and survival probability are formulated for a single particle, it is, however, straightforward to expand this to the multi-particle case if there is no interaction between random walkers. In the case of large number of enzymes, multiple particles move independently and thus the probability for having all N particles can be obtained by multiplying the survival probabilities for each single particle, i.e.

$$S(\vec{r}_i, \vec{r}_f, t; N) = S^N(\vec{r}_i, \vec{r}_f, t; 1) = \Big(1 - \int_0^t d\tau F(\vec{r}_i, \vec{r}_f, \tau; 1)\Big)^N \tag{4.4}$$

The probability of having exactly z of total N particles in the position $\vec{r}_f$ at time t irrespective of their arrivals is

$$P(\vec{r}_i, \vec{r}_f, t; N, z) = \frac{N!}{z!(N-z)!} P^z(\vec{r}_i, \vec{r}_f, t) \cdot [1 - P(\vec{r}_i, \vec{r}_f, t)]^{N-z} \tag{4.5}$$

One defines the accumulated first passage time probability $F^{ac}(\vec{r}_i, \vec{r}_f, t; N, z)$ as the probability that at time t z of the total N particles have all arrived at the destination for the first time by time t. The expression is

$$F^{ac}(\vec{r}_i, \vec{r}_f, t; N, z) = \frac{zN!}{z!(N-z)!} S^{N-z}(\vec{r}_i, \vec{r}_f, t; 1) \cdot [1 - S(\vec{r}_i, \vec{r}_f, t; 1)]^{z-1} \cdot F(\vec{r}_i, \vec{r}_f, t; 1) \tag{4.6}$$

We may also defines the simultaneous first passage time probability $F^{si}(\vec{r}_i, \vec{r}_f, t; N, z)$, which is the probability that at time t z of the total N particles simultaneously arrived at the destination for the first time. The corresponding expression is the same as that for single particle case, i.e., $F^{si}(\vec{r}_i, \vec{r}_f, t; N, z) = \mathcal{L}^{-1}\left\{\frac{\mathcal{L}\{P(\vec{r}_i,\vec{r}_f,t;N,z)\}}{\mathcal{L}\{P^{(s)}(\vec{r}_f,\vec{r}_f,t-\tau;N,z)\}}\right\}$.

4.2 Results

It is easy to assign different rates to each step and carry out the calculations. For simplicity, we will first assume uniform forward reaction rates α_f as well as uniform backward rates α_b and catalyzed rates α_c for all phosphorylation and dephosphorylation events. We also assume that the transportation rates k_+^i increase evenly and k_-^i decrease evenly with the increase of the number of unphosphorylated sites i, i.e. $k_+^i = k_+^M(\gamma_+)^{i-M}$, $k_-^i = k_-^0(\gamma_-)^{-i}$ $(i = 0, 1, 2, \cdots, M)$. This assumption captures the empirical observation that fully dephospharylated NFAT is much easier to transport from the cytoplasm into the nucleus than phospharylated NFAT. In the nucleus, the fully phospharylated NFAT is most easily transported to the cytoplasm [87].

4.2.1 Comparison of the exact solution with simulation

Figure 4.2 illustrates three trajectories taken from a Monte Carlo simulation of a signaling protein traveling from the initial fully phospharylated state C_0 in the cytoplasm to the fully dephosphorylated state U_M in the nucleus [43, 73]. The red diamonds in Fig. 4.2 indicate that the protein is found in the cytoplasm regardless of its specific form (C_i, $\bar{C}_i$, or $\hat{C}_i$) while the green dots indicate that the protein is found in the nucleus. Transitions between red and green sites indicate a transversal across the cytoplasm-nucleus barrier while up and down transitions indicate phosphorylation and dephosphorylation events.

In Fig. 4.3, we compare the mean first passage time probability and survival prob-

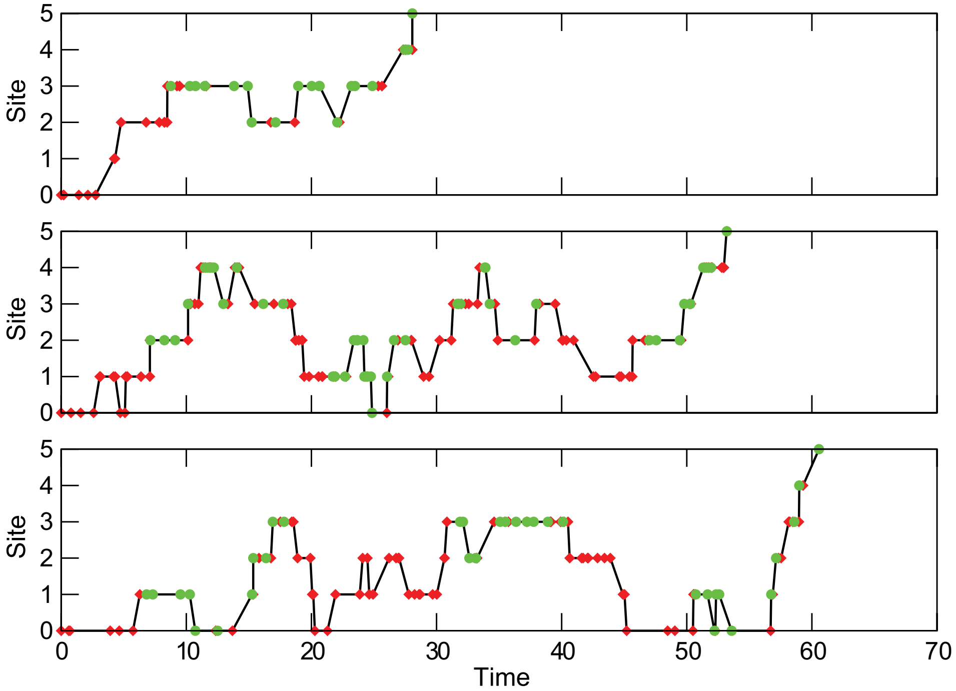

Figure 4.2: Three typical trajectories of a random walker traveling from C_0 to U_M is plotted as time vs. site number, i.e. the state of dephospharylation. The red diamonds label a protein in the cytoplasmic form (C_i, $\bar{C}_i$ or $\hat{C}_i$), while the green dots label the nuclear form (U_i, $\bar{U}_i$ or $\hat{U}_i$). The parameters are chosen as following: the phosphorylation site number $M = 5$, the forward reaction rates $\alpha'_f = \beta_f = \beta'_f = \alpha_f = 0.2$, the backward rates $\alpha'_b = \beta_b = \beta'_b = \alpha_b = 1.0$, the catalyzed rates $\alpha'_c = \beta_c = \beta'_c = \alpha_c = 1.0$, the transport rates $k_+^M = k_-^0 = 0.2$, and the ratio of transport rates $\gamma_+ = \gamma_- = 2.718$.

ability from the exact solution with those from the Monte Carlo simulations. The left panel shows several distributions of the first passage time probability computed from the exact solutions (solid lines) and those computed from the stochastic simulations (broken lines) with the same parameters, the right panel shows the corresponding survival probabilities from the exact solutions (red broken lines) and those from simulations (crosses). These simulation results agree very well with the exact solutions.

An exact solution can be found not only for the steady state distribution but also for the dynamics away from the steady states. Table 4.1 shows the occupancy probability

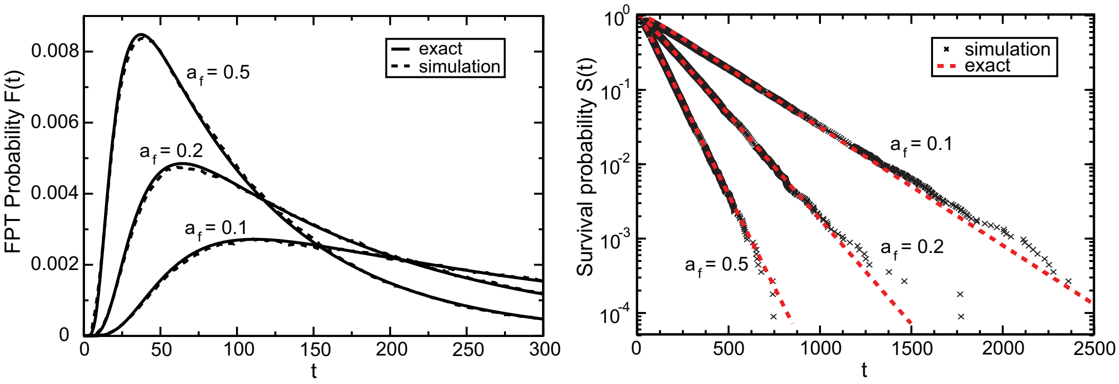

Figure 4.3: Comparisons of the probabilities of the first passage time and survival probabilities. Left panel: probability vs. the first passage time. Right panel: the survival probability vs. time. Parameters are the same as those in Fig. 4.2.

Table 4.1: Probability distribution of a random walker at time $t = 50$. For a given i, the three terms C_i, $\bar{C}_i$ and $\hat{C}_i$ with same phosphorylation states are collected together to represent the total probability of a protein at the i phosphorylation state in the C group. The three terms U_i, $\bar{U}_i$ and $\hat{U}_i$ are also collected together for the same reason. Parameters are chosen the same as those in Fig. 4.2.

$C_i + \bar{C}_i + \hat{C}_i$	$U_i + \bar{U}_i + \hat{U}_i$
0.2236	0.0294
0.2098	0.0358
0.1540	0.0410
0.0983	0.0424
0.0571	0.0404
0.0358	0.0325

distribution at time $t = 50$s when the network is far from the steady state. The difference between the exact solution and simulation is around 2%, which is essentially the sampling error. Figure 4.4 further explores the network dynamics. For a network with size M=5, the walker initially resides in the C_0 state ($t = 0$s) but propagates to other states by time $t = 10$s and $t = 50$s. Eventually the probability reaches a steady profile. The last time shown, $t = 1500$s, is much later than the mean first passage time 180s).

We can also compare the model's predications with laboratory experiments carried

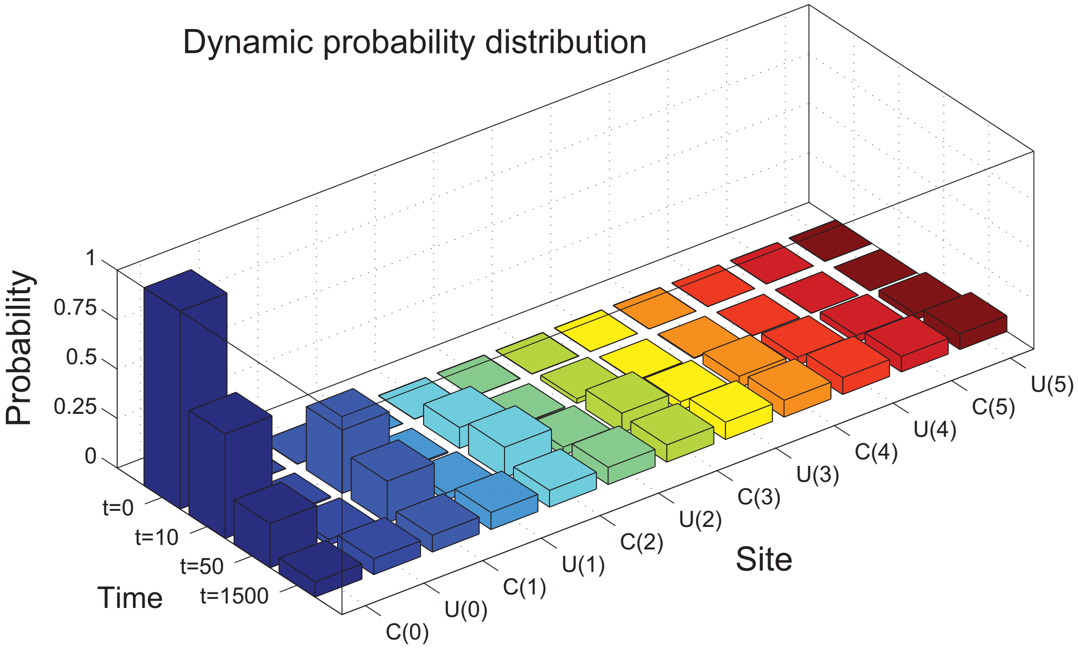

Figure 4.4: Dynamic propagation of the probability distribution. In the site axis, the $C(i)$ and $U(i)$ $(i = 0, 1, \cdots, 5)$, defined as $C(i) = C_i + \bar{C}_i + \hat{C}_i$ and $U(i) = U_i + \bar{U}_i + \hat{U}_i$, are collected terms representing the probabilities of a protein at the i phosphorylation state in the C group and the U group respectively. Four time slices $t = 0, 10, 50, 1500$ are chosen to show the dynamic propagation. The parameters are chosen the same as those in Fig. 4.2.

by Dolmetsch et al.. These experiments measure differential NFAT activation as a function of the amplitude and duration of a calcium stimulus [28]. Their work uncovered three different response patterns of the nuclear fraction of total NFATs that result from different stimulus (spike followed by plateau, a single spike and a low-level plateau). In Fig. 4.5, stimuli similar to experimentally used inputs (left panel) are entrained to our model and

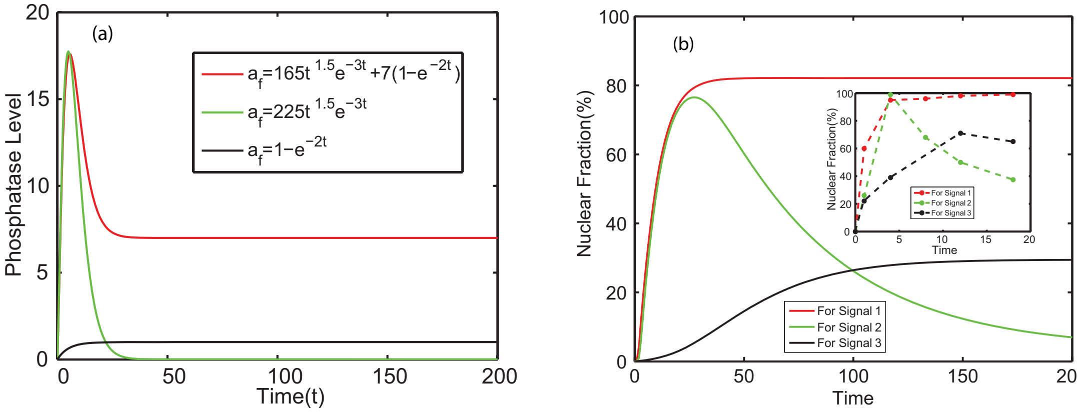

Figure 4.5: Comparisons of experiments with the model. (a) The concentration of phosphatases in cytoplasm is as the input with three different patterns: spike followed by plateau (red line), a single spike (green line), and a low-level plateau (black line), which are the same as experimental stimuli [28]. (b) Fractions of proteins in nuclear forms predicted from the model and those from experiments. Different colors correspond to different stimuli. The curves in the inset are experimental results from Dolmetsch et al. [28]. Parameters are chosen as: the phosphorylation site number $M = 5$, the forward reaction rates $\beta_f = 1.0$, $\alpha'_f = \beta'_f = 0.1$, the backward rates $\alpha_b = \beta_b = 1.0$, $\alpha'_b = \beta'_b = 0.1$, the catalyzed rates $\alpha_c = \beta_c = 10.0$, $\alpha'_c = \beta'_c = 10.0$, the transport rates $k^M_+ = 0.6, k^0_- = 0.1$, and the ratio of transport rates $\gamma_+ = \gamma_- = 2.718$.

are also shown to result in three response patterns (right panel). The predicted patterns agree well with those seen in the experimental studies (inset of right panel).

4.2.2 An optimal forward reaction rate favors the passage

Many studies have hightlighted the efficiency, sensitivity, and robustness of signal transduction networks [56, 36, 69]. In this regard the ladder network exhibits an interesting property, the existence of an optimal value for the forward reaction rates. Figure 4.6 shows the mean, the most probable and the root-mean-square of the first passage time of a signaling protein from the C_0 state to the U_M. Clearly there is an optimal forward reaction rate for the passage: The optimum occurs at $\alpha_f \simeq 1$ for the mean first passage time, but the optimal values of $\alpha_f \simeq 2$ for the most probable and the root-mean-square passage times are around 2.

The existence of this optimum may seem to conflict with the intuition that the higher enzyme concentration is, the shorter the passage time is. When the forward reaction rates are very slow, the forward reactions do indeed constitute a bottleneck. But with increasing forward reaction rates, the walkers will be found more and more often in the signaling protein-enzyme complex forms. This helps signaling proteins move towards dephosphorylated states. Eventually when the forward reaction rate is too large compared to the transportation rates k_+ and k_-, the signaling proteins will then spend most of their time in the transport incompetent complex forms ($\bar{C}_i$, $\hat{C}_i$, $\bar{U}_i$ or $\hat{U}_i$) rather than in the apo forms required for transport. This ultimately leads to slower transport between the cytoplasm and the nucleus.

This intriguing phenomenon could also be qualitatively explained as follows. The passage time of a signaling protein from the fully phosphorylated state in the cytoplasm (C_0) to the fully dephosphorylated state in the nuclear (U_M) consists of two parts: the time for dephosphorylation (T_d) and the time for transport (T_t). The typical time from site A to its neighbor B can be estimated by adding the inverse rates using the "one-step-forward" approximation. Assuming there are total l neighbors C_i, $i = 1, ..., l$ (including $B = C_j$) that A can directly jump to with rate k_i, the typical time is $(\frac{k_j}{\sum k_i} \times k_j)^{-1}$. For each dephosphorylation step from state i to $i+1$, this estimate yields a time $(2\alpha_f + k_+^i)/\alpha_f^2 + (\alpha_b + \alpha_c)/\alpha_c^2$. The total dephosphorylation time is about $T_d = M \times [(2\alpha_f + \bar{k}_+)/\alpha_f^2 + (\alpha_b + \alpha_c)/\alpha_c^2]$, where $\bar{k}_+$ is the average of k_+^i's. The transport time from C_i to U_i can be approximated as $(2\alpha_f + k_+^i)/k_+^{i2}$ based on similar approximations. Since the transport could happen in any C_i state, the mean transportation time requires averaging over all i, which results in a time

$T_t = (2\alpha_f + \bar{k}_+)/\bar{k}_+^2$. The total mean passage time can therefore be approximated as

$$T = \Big\{ M[\frac{2\alpha_f + \bar{k}_+}{\alpha_f^2} + \frac{\alpha_b + \alpha_c}{\alpha_c^2}] + \frac{2\alpha_f + \bar{k}_+}{\bar{k}_+^2} \Big\} \tag{4.7}$$

The existence of an optimal forward reaction rate requires the existence of a solution to the equation $\partial T/\partial \alpha_f = 0$ has solution(s). As a result, we must have the equation $\alpha_f^3 - M\bar{k}_+^2\alpha_f - M\bar{k}_+^3 = 0$. With the parameters shown in Fig. 4.6, the optimal forward rate is estimated to be 0.49, which is comparable with the exact solution ($\simeq 1$) as shown in the figure.

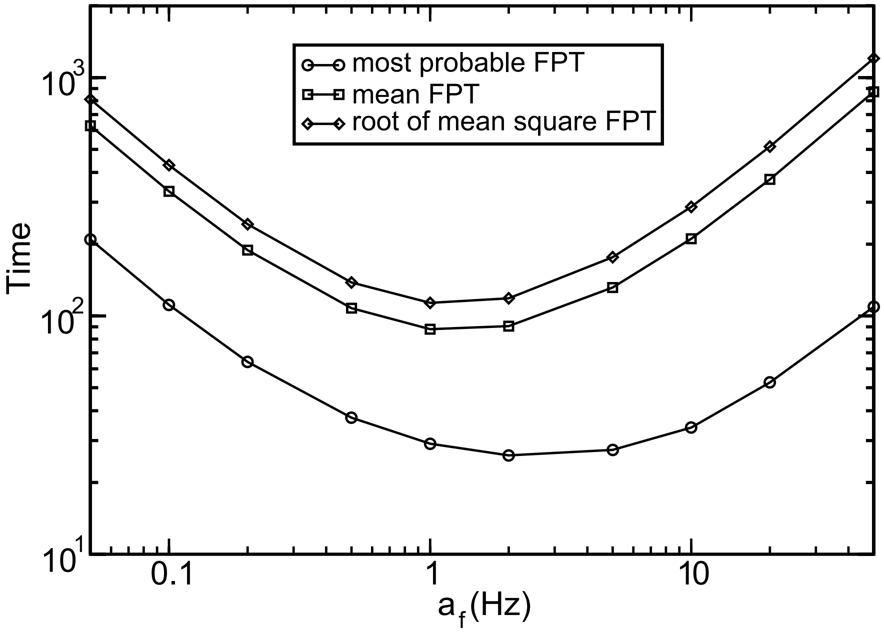

Figure 4.6: An optimal forward reaction rate α_f favors efficient passage. The horizontal axis is the forward reaction rate swept from 0.05 to 50. The three curves represent the mean first passage time, the root mean square first passage time and the most probable first passage time according to the forward rate α_f. Other parameters are fixed as the phosphorylation site number $M = 5$, the forward reaction rates $\alpha'_f = \beta_f = \beta'_f = \alpha_f$, the backward rates $\alpha'_b = \beta_b = \beta'_b = \alpha_b = 1.0$, the catalyzed rates $\alpha'_c = \beta_c = \beta'_c = \alpha_c = 1.0$, the transport rates $k_+^M = k_-^0 = 0.2$, and the ratio of transport rates $\gamma_+ = \gamma_- = 2.718$.

4.2.3 Asymmetry of the reaction network effects the passage

Although the ladder network is symmetrical in topology, the reaction rates in the cytoplasm and nucleus are quite different owing to different availabilities of appropriate enzymes. To illustrate the effect of the network's asymmetry on signal transduction, we employ two parameters to probe this: the phosphorylation asymmetry parameter and the transport asymmetry parameter. The phosphorylation asymmetry parameter is defined as $\theta = \frac{a'_f}{a_f} = \frac{a'_b}{a_b} = \frac{a'_c}{a_c} = \frac{b'_f}{b_f} = \frac{b'_b}{b_b} = \frac{b'_c}{b_c}$, where $\theta \leq 1$, which characterizes the preference for undergoing dephosphorylation compared to the phosphorylation processes in the different environments. The other parameter, transport asymmetry parameter ϕ, is defined as $\phi = \frac{a_f}{b_f} = \frac{a_b}{b_b} = \frac{a_c}{b_c} = \frac{a'_f}{b'_f} = \frac{a'_b}{b'_b} = \frac{a'_c}{b'_c}$, which characterizes the relative activities of reactions in the cytoplasm compared to the corresponding reactions in the nucleus.

Figure 4.7 illustrates the effects of these asymmetries on the network behavior. The left panel shows that increasing the phosphorylation asymmetry slows down signaling; the right panel shows that increasing the transport asymmetry speeds up signaling.

4.2.4 The first passage time distribution has a long tail

The probability distribution of the first passage time has a long tail due to the network's hierarchical structure. In the limits of either large or small forward reaction rates, the probability distribution is very flat and the tail can be extremely long.

With the long-tail distribution, the mean first passage time can be greatly different from the most probable first passage time. In Fig. 4.3, the most probable first passage times for $\alpha_f = 0.5, 0.2, 0.1$ are 40, 60 and 100s respectively. However, the ratio of the mean first passage time and the most probable first passage time ranges from on the order of 1 to

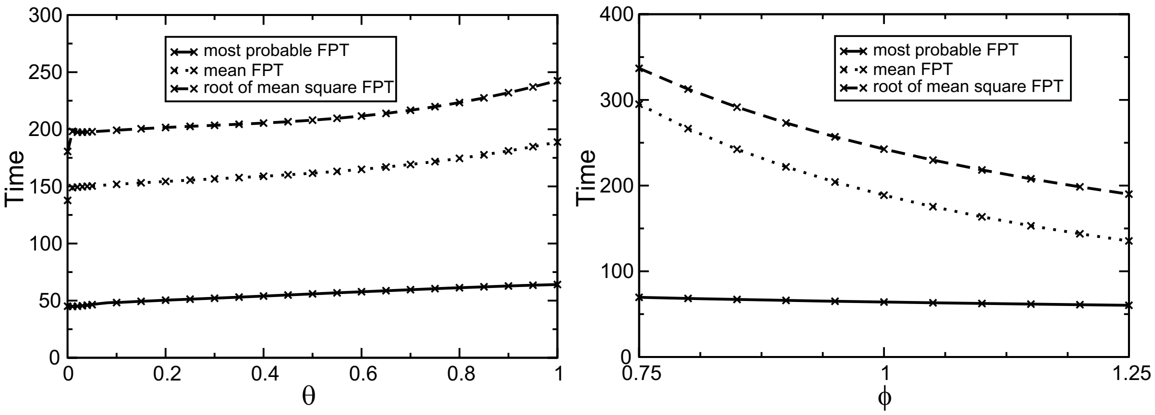

Figure 4.7: Effect of the asymmetry of the reaction networks to the passage rate. The left panel shows a plot of characteristic times as functions of the parallel asymmetry parameter θ. The other parameters are fixed at the phosphorylation site number $M = 5$, the forward reaction rates $\alpha_f = \beta_f = 0.2$, the backward rates $\alpha_b = \beta_b = 1.0$, the catalyzed rates $\alpha_c = \beta_c = 1.0$, the transport rates $k_+^M = k_-^0 = 0.2$, and the ratio of transport rates $\gamma_+ = \gamma_- = 2.718$. The right panel is the plot of characteristic times as functions of rung asymmetry parameter ϕ. The other parameters are fixed at $M = 5$, $\alpha'_f = \beta'_f = 0.2$, $\alpha'_b = \beta'_b = 1.0$, $\alpha'_c = \beta'_c = 1.0$, $k_+^M = k_-^0 = 0.2$, $\gamma_+ = \gamma_- = 2.718$.

the order of 3 with the chosen set of parameters. Fig. 4.6 shows that the ratio of the most probable first passage time and mean first passage time is large even on a logarithm scale.

4.3 Conclusions and Discussions

In this paper, we studied a general signal transduction network-reaction ladder network that models multiple—site phosphorylation and cytonuclear transport. As often happens for real networks, the enzymes are assumed to be abundant, and so each signaling protein independently wanders through its various states of phosphorylation and location in compartments. Except for the last binding step, the network is effectively linear and therefore can be solved exactly. This exact solution is confirmed using the Monte Carlo simulation. Even this simple network exhibits several interesting stochastic features. It exhibits a long tail of the probability distribution for the signaling time and there is an optimal forward reaction rate that favors speedy signaling. This optimum suggests there may be an optimal amounts of enzymes for efficient signal transduction.

When available enzymes are not abundant, the random walkers on the network become correlated because they must share the limited resources of available enzymes. Nonlinearity then starts to play an important role throughout the process not just in the acquiring last step of the final target. In this situation when an enzyme is bound, several phosphorylation or dephosphorylation events will occur before unbinding, similar to the processive aspect of transcription [84]. It will be interesting to study such scenarios in order to study the effects of oscillatory upstream signaling molecules on the network. These will help us further understanding the phenomenology and quantitative design criteria of effective signal transduction mechanisms.

Acknowledgements: This chapter contains materials in Lu T, Shen T, Zong C, Hasty J, and Wolynes P, Statistics of cellular signal transduction as a race to the nucleus by multiple random walkers in compartment/phosphorylation space, *Proc. Natl. Acad. Sci.*, 103:16752-16757 (2006).

5

Variability and Diversity of Cellular Populations

Stochasticity is an amazing property that exists ubiquitously in biological systems. Variation due to stochastic fluctuations occurs in various organisms and at different scales [89, 60, 104, 29]. Numerous studies show that noise is not always negligible but can play a significant role in many cellular functions and physiological behaviors. For example, noise seems to underly the emergence of neural precursor cells from an initially homogenous population during the development of *Drosophila melanogaster* [103]. Noise contributes to the traversal of start and progression of yeast cells into the cell cycle [12]. Noise can also limits the precision of circadian clocks [11]. A recent study illustrates that molecular random fluctuation influences the fates of cells infected with human immunodeficiency virus [117]. All these studies lead us to trace the origins of noise and looking for a corresponding mathematical characterization.

Noise occurring at the genetic and molecular level has been intensive studied in the past few years [89, 60, 104, 47, 91]. Multiple sources contribute to the observable variability [109]. Stochasticity inherent to biochemical reaction events of gene expression is classified as genetic intrinsic noise, such as fluctuations from transcription of a gene and translation of mRNA. This noise arises from the intrinsic nature of chemical reactions as molecular transformation. Variability arising from sources that are external to the biochemical process of gene expression under consideration is termed genetic extrinsic noise. Examples of extrinsic noise are fluctuations in the abundances of RNA polymerases and ribosomes and different copy numbers of a gene. Noise in gene expression is propagated through network cascades and the corresponding amplitude of the fluctuation can be either increased or damped [101, 85, 102, 72].

However, variability at molecular level is not the final product of noise. Genetic noise further triggers fluctuation and heterogeneity of entire cellular populations since different types of cells usually have distinct adaptations and fitness to environments, such as different growth rates and survival capabilities [111, 121, 67, 10, 61]. A random switch of an individual cell from one type to another can lead to dramatic changes at the level of an entire population [86]. The discrete nature of cells ensures that cellular growth, division, death and transitions between different types are ultimately and essentially non-deterministic. These must generate additional contributions to population variation. Diversity of a cellular population, representing the species richness of that population [68, 57], is therefore affected intrinsically by all such stochastic factors. Source-oriented observations and theoretical studies in ecology also have reported various population fluctuations of animals and plants, like plant species, animal populations, and all sorts of factors, like climate, age and sex, are studied and discussed [21, 4].

The existence of noise sources at multiple (molecular to cellular) levels and potential couplings between these levels encourage us to search a quantification of phenotypic cellular population variability.

5.1 Methods

Here is a simple example illustrating population variation and motivating the development for measurements. For a two-phenotype community as shown in Fig. 5.1a, we simulated the cellular population dynamics in a microfluidic chip using the Monte Carlo method [45, 88] multiple times. We then recorded the cell populations of each runs in the microfluidic chip and got the statistics of populations. Figure 5.1b shows the coefficients of variation (standard deviation divided by mean) for each phenotype versus trial number for different initial states. As can be seen clearly, two experiments, one starts with real low cell numbers and the other starts with high cell numbers which fully fill the chip. These have distinct difference of variation coefficients although these two sets of experiments have a same environment and same microscopic rate coefficients of the community.

5.1.1 Cellular Population Variation Index

Let us imagine an experiment starting with two genetically identical cells that belong to two different phenotypes, green and red, as shown in Fig. 5.2. Cells of each phenotype may grow, divide, and die in the culture as well as randomly switch from one phenotype to the other depending on the intrinsic mechanism and their environment. In this experiment, we take a couple of snapshots at different times and record the population of each phenotype. We then repeat the experiment many times: starting with the same two

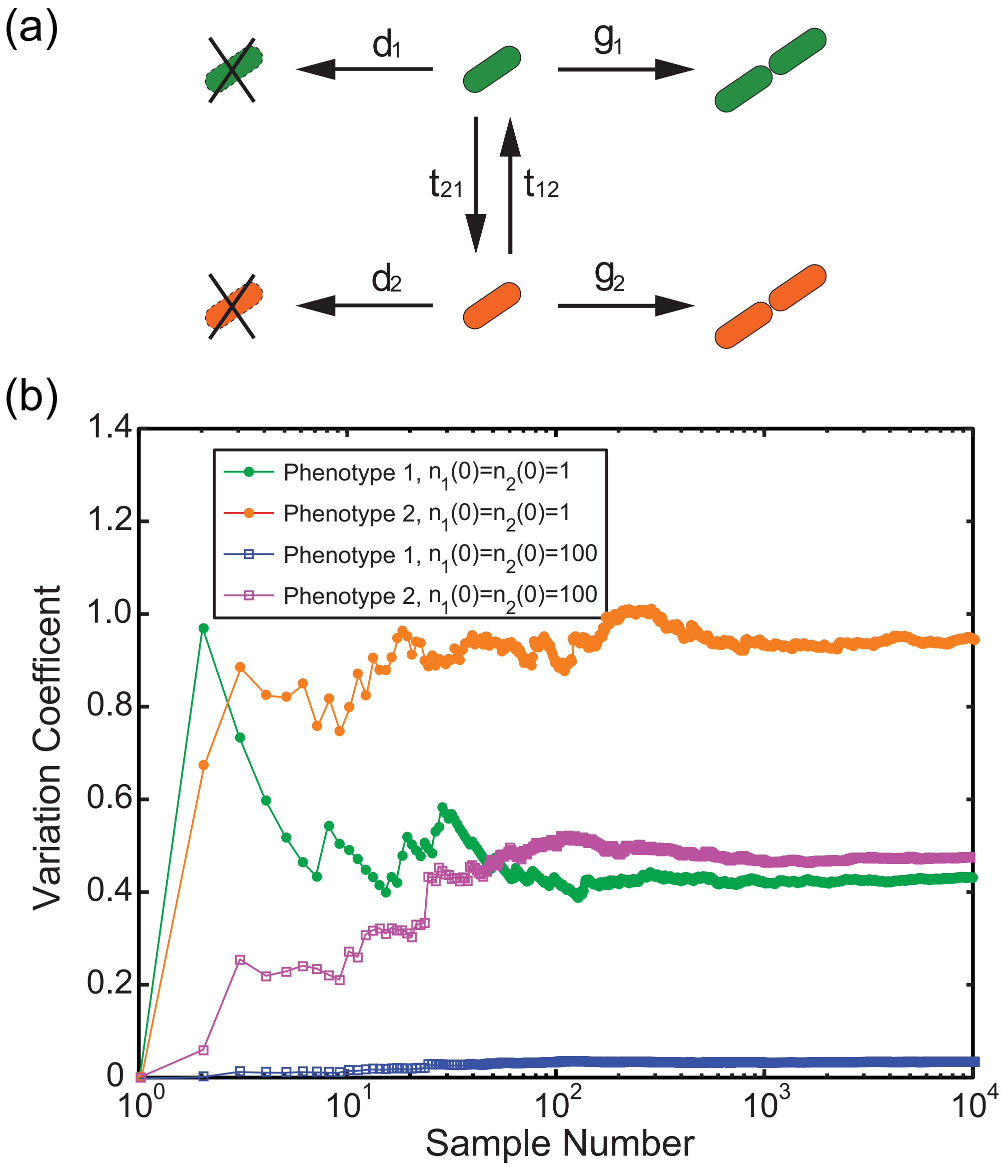

Figure 5.1: (a), A two-phenotype community. Each phenotype (green and red) has its own birth rate coefficient(g_1, g_2) and death rate coefficient(d_1, d_2). They are able to switch from one phenotype to the other with the transition rate coefficients (t_{12} and t_{21}). (b), Coefficient of variation (CV) versus sample number. A two-phenotype community (illustrated by (a)) grows in a chemostat environment which can contain a maximum of 200 cells and is taken to measure at $t = 6$ for all experiments. Blue and red curves are the CVs of phenotype 1 and 2 with the initial state $(1, 1)$ respectively, while green and magenta curves are the CVs of phenotype 1 and 2 for the initial state $(100, 100)$ respectively. It is clear to see that different initial states of a community in a same chemostat environment result in different variation coefficients. The parameters are the same as those in Fig. 5.3.

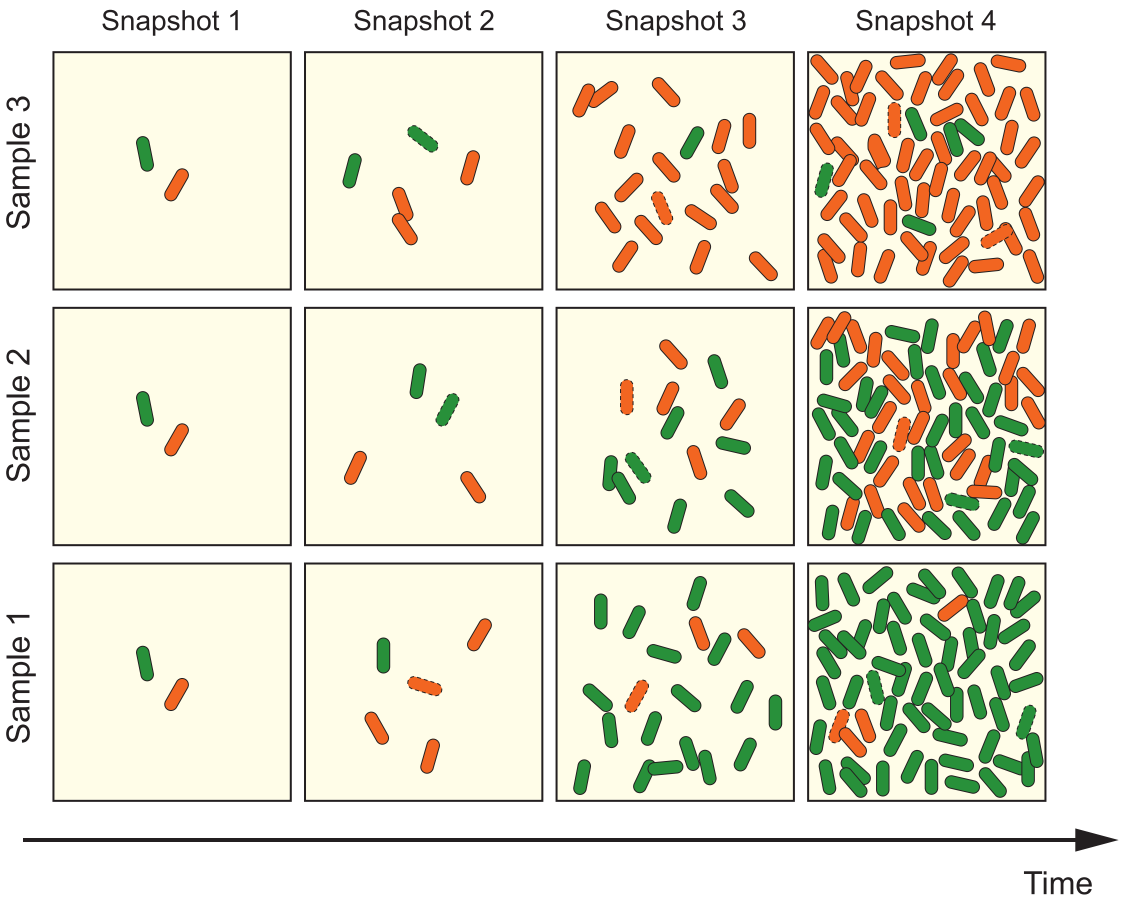

Figure 5.2: An illustration of the ensemble population diversity. The initial state, one phenotype each, is the same for three ensembles. Here colors represent distinct phenotypes. Cells grow, divide and die (the broken lines in the cartoon indicate that the circled cell is dead). Four snapshots are taken at specific times. Multiple runs of the experiment result in different population in each phenotype: Green cells turn out to be the dominant in the first sample, while red cells in the last sample, and both types are comparable in the middle sample.

cells, growing in the same culture and taking snapshots at the same time points. Finally, when we compare the snapshots at the same time from different runs, we will find that the percentage of each phenotype, as well as the overall population are different among runs, some of which might even be strikingly distinct.

Quantitative tools of measurements are needed to analyze these data. For a community with I species and n_i cells for species i $(i = 1, 2, ..., I)$, Simpson proposed a 'concentration' to measure the diversity, $\Theta = \frac{1}{N^2}\sum_{i=1}^{I} n_i^2$, where $N = \sum_{i=1}^{I} n_i$ is the overall population of the community [68]. Similar quantities were introduced in the study of spin glass dynamics and single molecule dynamics [115, 124]. This concentration describes the probability of any two randomly chosen individuals from a community belonging to the same species. The complement of the concentration, $1 - \Theta$, is termed the Simpson index and describes the probability that two randomly picked individuals belong to different species. Thus it is a measure of the population diversity of a particular community. However, the population for each species can vary from one run to another and the index proposed by Simpson does not account for the variation from this fluctuation. To investigate the population diversity with the incorporation of the cellular and intracellular stochasticity, we propose a generalized population variation index (detailed in Appendix) as

$$\mathfrak{D} \equiv \frac{\overline{\langle \mathbf{n}^2 \rangle} - \left(\overline{\langle \mathbf{n} \rangle}\right)^2}{\left(\overline{\langle \mathbf{n} \rangle}\right)^2} \tag{5.1}$$

where $\mathbf{n}$ is a I-dimensional vector of cellular population with each components n_i representing the population of species i, the bar operator, $\overline{(\cdot)}$, represents $\sum_{\mathbf{n}} P(\mathbf{n})(\cdot)$ the ensemble average over the different population distributions of a community, while the bracket operator, $\langle(\cdot)\rangle$, represents $\frac{1}{S}\sum_{j=1}^{S}(\cdot)$ the average over species (phenotypes and genotypes) with a given distribution. This index can be written as

$$\mathfrak{D} = \underbrace{\frac{\overline{\langle \mathbf{n}^2 \rangle - \langle \mathbf{n} \rangle^2}}{\left(\overline{\langle \mathbf{n} \rangle}\right)^2}}_{\equiv \mathfrak{D}_i} + \underbrace{\frac{\overline{\langle \mathbf{n} \rangle^2} - \left(\overline{\langle \mathbf{n} \rangle}\right)^2}{\left(\overline{\langle \mathbf{n} \rangle}\right)^2}}_{\equiv \mathfrak{D}_c} \tag{5.2}$$

Here the first part $\mathfrak{D}_i$, termed the intra-colony variation index, measures the population difference between species under the same average of population distribution. It is therefore the ensemble averaged Simpson concentration. This non-negative index plays a similar role as the Simpson concentration does: It is zero when the population is evenly distributed among species, increases with the unevenness of population distribution, and reaches its maximum when all of the population belongs to a single species. The second part, $\mathfrak{D}_c$, termed as the cross-colony variation index, measures the population variation among colonies. It goes to zero when the overall population is the same from colony to colony and becomes large when the variation in population size is large among colonies. These two variations are in some sense analogous to those of intrinsic and extrinsic contributions to the noise in gene expression[109]. Here 'intrinsic' refers to the intra-colony variation while 'extrinsic' refers to cross-colony variation. This generalized index reduces to Simpson's index when we neglect population dispersion, i.e., when the distribution of population size is a δ-function. In such cases, $\mathfrak{D}_c$ goes to zero and $\mathfrak{D}_i$ becomes $(\langle \mathbf{n}^2 \rangle - \langle \mathbf{n} \rangle^2)/\langle \mathbf{n} \rangle^2 = S\Theta - 1$, where S is the number of species and Θ is the Simpson concentration [68].

5.1.2 Cellular Population Dynamics

The population dynamics of a colony consists of cellular birth, death and transitions between types (phenotypes or genotypes), which can be fully characterized by the following Master equation

$$\begin{aligned} \tfrac{d}{dt}P(\mathbf{n},t) &= \sum_{\mathbf{r}} \big[\mathbf{G}(\mathbf{n}-\mathbf{r}) + \mathbf{D}(\mathbf{n}-\mathbf{r}) + \mathbf{T}(\mathbf{n}-\mathbf{r})\big] P(\mathbf{n}-\mathbf{r},t) \\ &- \sum_{\mathbf{r}} \big[\mathbf{G}(\mathbf{n}) + \mathbf{D}(\mathbf{n}) + \mathbf{T}(\mathbf{n})\big] P(\mathbf{n},t) \end{aligned} \tag{5.3}$$

where $P(\mathbf{n}, t)$ is the probability of a colony with its population size in state $\mathbf{n}$ ($\mathbf{n}$ is a vector $\mathbf{n} = \{n_1, n_2, ..., n_I\}$ that fully characterizes the state of the colony, i.e., there are n_1 cells for type 1, n_2 cells for type 2, etc.) at time t, $\mathbf{r}$ is a vector representing the numbers of cells changed in an corresponding event. $\mathbf{G}(\mathbf{n})$, $\mathbf{D}(\mathbf{n})$ and $\mathbf{T}(\mathbf{n})$ are the rates of cellular birth, death, and type transition processes. These rates are generally determined by intracellular events and environmental conditions. The proposed variation index as well as cellular population dynamics is also applicable to describe genotypic variations that root in DNA mutation and random drift and lead to molecular evolution [63].

5.2 Results

5.2.1 Free Growth Environment

The Variation Index

We first study the variation of cell populations growing in an ideal case where there is no constraints on nutrition or space. In these environments, all the reaction rates can be assumed for simplicity as linear functions of their arguments. We use a two-phenotype community as an example to explore the population variation as shown in Fig. 5.1. The population dynamics can be described by

$$\begin{aligned}\tfrac{d}{dt}P(n_1,n_2,t) &= g_1 n_1^- P(n_1^-,n_2,t) + d_1 n_1^+ P(n_1^+,n_2,t) + g_2 n_2^- P(n_1,n_2^-,t) + d_2 n_2^+ P(n_1,n_2^+,t) \\ &+ t_{12} n_2^+ P(n_1^-,n_2^+,t) + t_{21} n_1^+ P(n_1^+,n_2^-,t) - \big[(g_1+d_1+t_{21})n_1 + (g_2+d_2+t_{12})n_2\big] P(n_1,n_2,t) \qquad (5.4)\end{aligned}$$

where g_i and d_i are the rate coefficients of birth and death for phenotype i respectively, t_{ij} is the rate coefficient of the transition from type j to type i, and $n_i^{\pm} = n_i \pm 1$ $(i = 1, 2)$.

The analysis of the variation index and the two-phenotype system shows that we only need to track the dynamics of the first two moments of the probability distribution function to investigate the population variability and diversity of a community (detailed in Appendix). It is analytically unaccessible generally but, for this linear free growth case, is solvable. By constructing a vector consisting of the first and the second moment as

$\mathbf{M}(t) = \big(\bar{n}_1(t), \bar{n}_2(t), \overline{n_1^2}(t), \overline{n_2^2}(t), \overline{n_1 n_2}(t)\big)^T$, we have a simplified yet sufficient description

$$\frac{d}{dt}\mathbf{M}(t) = \mathbf{Q} \cdot \mathbf{M}(t) \tag{5.5}$$

where the matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \begin{bmatrix} s_1 & t_{12} & 0 & 0 & 0 \\ t_{21} & s_2 & 0 & 0 & 0 \\ r_1 & t_{12} & 2s_1 & 0 & 2t_{12} \\ t_{21} & r_2 & 0 & 2s_2 & 2t_{21} \\ -t_{21} & -t_{12} & t_{21} & t_{12} & s_1 + s_2 \end{bmatrix}$$

Here parameters $s_1 = g_1 - d_1 - t_{21}$, $s_2 = g_2 - d_2 - t_{12}$, $r_1 = g_1 + d_1 + t_{21}$ and $r_2 = g_2 + d_2 + t_{12}$.

Exact expressions of the first moment, $\mathbf{M}^{(1)}(t) = \big(\bar{n}_1(t), \bar{n}_2(t)\big)^T$, and second moments, $\mathbf{M}^{(2)}(t) = \big(\overline{n_1^2}(t), \overline{n_2^2}(t), \overline{n_1 n_2}(t)\big)^T$ (detailed in Appendix) allow us to study the temporal behavior of the population variation, which can be expressed as a function of $\mathbf{M}(t)$ as

$$\mathfrak{D}(t) = \frac{M_1^{(2)} + M_2^{(2)} - 2M_3^{(2)}}{(M_1^{(1)} + M_2^{(1)})^2} + \frac{(M_1^{(2)} + M_2^{(2)} + 2M_3^{(2)}) - (M_1^{(1)} + M_2^{(1)})^2}{(M_1^{(1)} + M_2^{(1)})^2} \tag{5.6}$$

where $M_j^{(i)}(t)$ is the jth element of the ith moment at time t. This variation expression is characterized by five exponents, $(\theta,\ \nu,\ \theta + \nu,\ 2\theta,\ 2\nu)$, where $\theta = \frac{1}{2}(s_1 + s_2 + \Delta)$, $\nu = \frac{1}{2}(s_1 + s_2 - \Delta)$ and $\Delta = \sqrt{(s_1 - s_2)^2 + 4t_{12}t_{21}}$. The exponent θ is the Lyapunov exponent for the deterministic exponential growth of the dynamic systems [67].

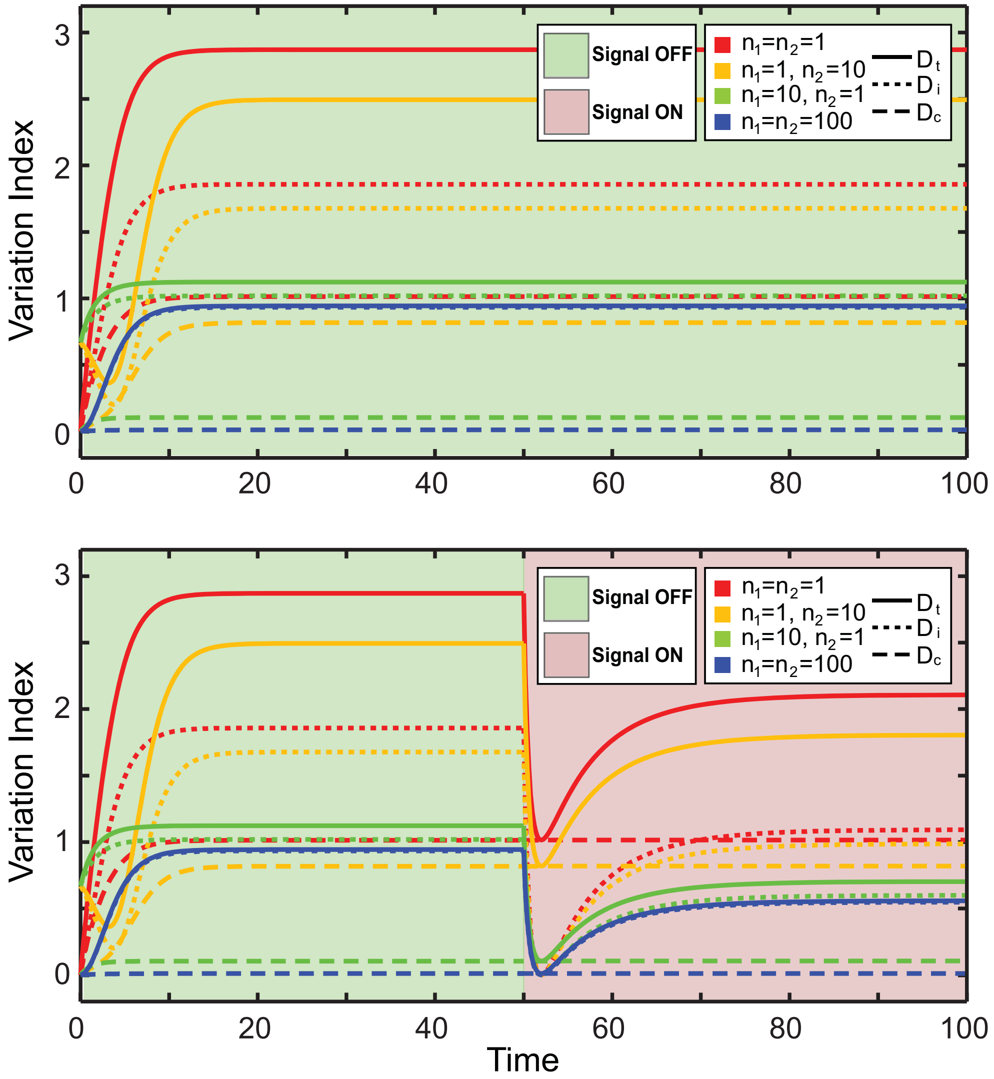

Figure 5.3: Propagation of the population variation over time. The dotted, dash and solid lines represent intra-colony, cross-colony, and overall population variation. The color of the line (red, orange, green, and blue) corresponds to the initial distribution (1,1), (1,10), (10,1), or (100,100) respectively. The change of the background color presents the onset of an external signal: pale green means no signals in the culture (t_{12}^{OFF}) while pale magenta means a signal is released into the culture (t_{12}^{ON}). Parameters are $g_1 = 1.0$, $g_2 = 0.5$, $d_1 = d_2 = 0.01$, $t_{12} = 0.01$, $t_{21}^{OFF} = 0.01$, $t_{21}^{ON} = 0.5$.

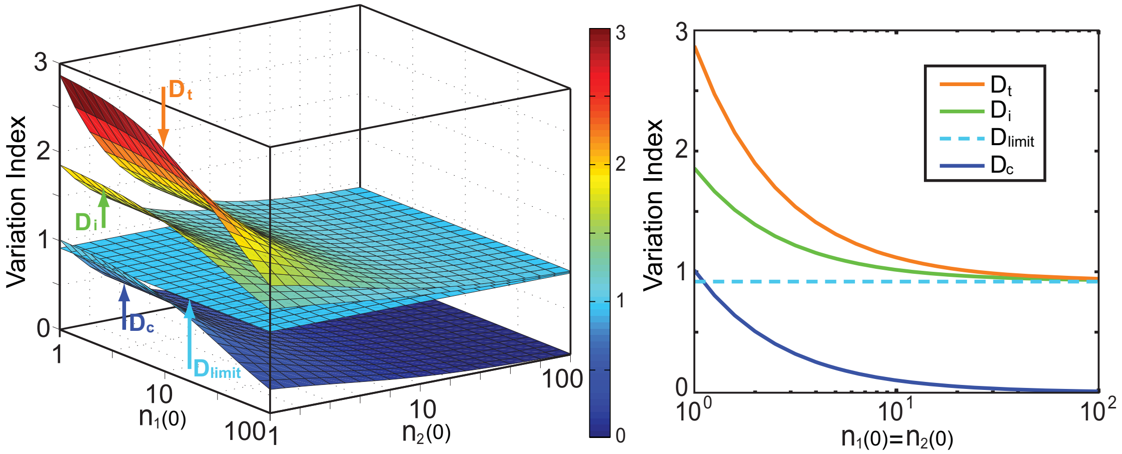

Figure 5.4: Asymptotic population variation with respect to different initial distributions. Left panel: The four surfaces correspond to cross-colony variation, deterministic variation, intra-colony variation and total population variation (summation of intra- and cross- colony variations). Right panel: A cut slice passing the vertical axis and the diagonal in horizontal plane in the left three-dimensional variation. The variation index is different from the deterministic variation but approaches it when the initial numbers of cells are large. Initial phenotypes are assumed to be delta-distributed and parameters are the same as those in Fig. 5.3.

Figure 5.3 illustrates the propagation of the population variation in different environmental conditions and with different initial states. Here the background colors represent different environments, the colors of the lines (red, orange, green, and blue) represent the initial states of the community ($(1, 1)$, $(1, 10)$, $(10, 1)$, and $(100, 100)$) and the types of those lines (dotted, dash, and solid) correspond to the contributions of the population variation (intra-colony, cross-colony, and overall population variation). All of the variation contributions approach constant values at the long time limit in a constant free growth environment (Upper panel). When the environment is changed, the population variations are out of the stationary values and adapted to new steady states according to the updated environment (Lower panel).

As can be seen in Fig. 5.3, different initial states result in different cellular population variabilities despite of the same rates for all events of the cellular population dynamics. For the long time and asymptotic behaviors, the largest exponent, 2θ, has the dominant contribution to population variation, where the intra-colony and cross-colony indices asymptotically become

$$\begin{aligned}\mathfrak{D}_i^\infty &= \frac{\big[C_{21} - F_1(0)\big]\big(K_{11} + K_{21} - 2K_{31}\big)}{\gamma_1^2\big[1 + \frac{s_2 - s_1 + \Delta}{2t_{12}}\big]^2} \\ \mathfrak{D}_c^\infty &= \frac{\big[C_{21} - F_1(0)\big]\big(K_{11} + K_{21} + 2K_{31}\big)}{\gamma_1^2\big[1 + \frac{s_2 - s_1 + \Delta}{2t_{12}}\big]^2} - 1 \end{aligned} \tag{5.7}$$

where the detailed expressions of C_{21}, $F_1(0)$ and K_{ij} are in Appendix.

Figure 5.4 explicitly shows the effects of the initial states of a cell population to the long time limit of the cellular population variability. The left panel illustrates the dependence of population variation on its initial cellular population. From bottom to up, the four surfaces correspond to cross-colony, deterministic, intra-colony, and overall population variation. Here the deterministic results are obtained for comparison from the corresponding mass action dynamics (mean-field dynamics) which was solved based on the first order moments alone and completely ignoring the effect of the second order. The right panel is a cross section passing the vertical axis and the diagonal line in the horizontal surface. As shown in both panels, smaller initial cell population results in larger final population variation. With the increasing of initial cell population, overall variation decreases and approaches the deterministic result which is independent of the initial state. This dependence of the population variation on the initial numbers of cells is somewhat analogous to the noise in gene networks, where smaller numbers of molecules bring larger randomness and gene noise becomes negligible when the mean numbers of molecules are sufficiently large [71].

A Numerical Experiment

We further perform a numerical experiment using the Monte Carlo method [45, 88]. The population dynamics of a two-phenotype community is simulated to illustrate the population variation and diversity. Following the processes in Fig. 5.1a, different phenotypes have their own rates of birth, death and transitions. In the experiment, the cellular populations grow from different initial states ($n_1(0)$, $n_2(0)$): (1, 1), (1, 10), (10, 1), and (100, 100) and for each set of given initial condition, we averaged over 100,000 simulation trajectories. The results show that both the ensemble averaged means and variances of each phenotypes agree well with those from analytical solutions (the differences are less than 0.5%).

Figure 5.5 is a direct illustration of celluar population variations with sample trajectories. The upper left, upper right, lower left and lower right panels correspond to different initial states (1, 1), (1, 10), (10, 1), and (100, 100). In each panel, the solid black line and broken black lines are the mean cell populations of phenotype 1 and 2 from analytical solutions. Each pair of a solid line and a broken line with a same color are the two trajectories of phenotypes 1 and 2 from a single run. There are 14 pairs of trajectories are shown in each panel. It is clear to see that a community with small initial numbers of cells has huge variability (e.g., Upper left panel) and the variability is remembered and does not diminish even when the population is extremely large. However, a community with large initial cellular population has less population variability. All these results nicely demonstrate the theoretical analysis.

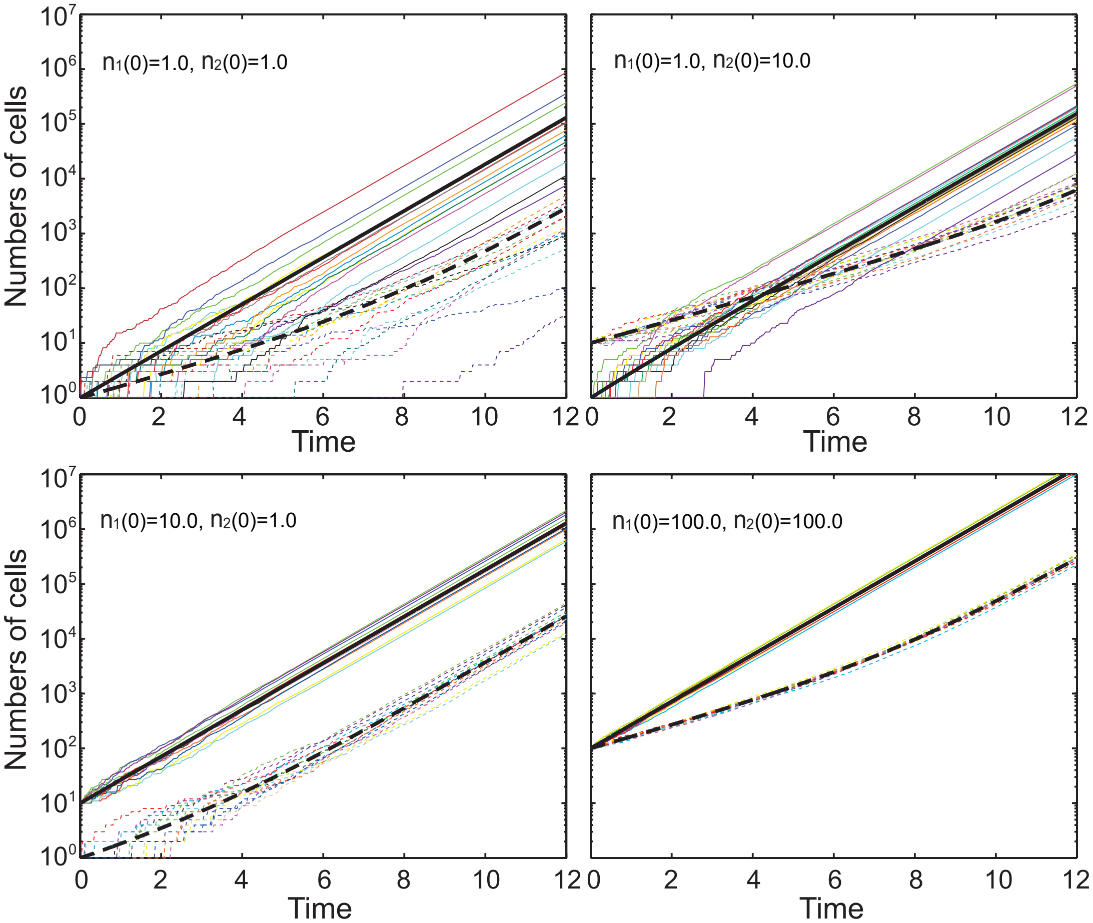

Figure 5.5: Simulation of cellular population dynamics in free growth environments. The upper left, upper right, lower left and lower right panels correspond to different initial numbers of cells ((1, 1), (1, 10), (10, 1), and (100, 100)). In each panel, the black bold solid and bold broken lines are the mean numbers of phenotype 1 and phenotype 2 from analytical solutions. Each pair of a solid line and a broken line with same color are the two trajectories of phenotype 1 and phenotype 2 respectively from a single simulation. There are 14 pairs of trajectories shown in each panel. Initial phenotypes are assumed to be delta-distributed and parameters are the same as those of Fig. 5.3.

5.2.2 Chemostat Environments: Eliminating the Cross-Colony Variation

Realistic environments of cellular dynamics usually have limited nutrition sources, like flask cultures and agarose plates [2], as well as space constraints, like microenvironments of microfluidic chips [31]. Here we turn to study the variability and diversity of cellular populations in these types of chemostat environments.

Cells grow in chemostat environments with sufficient nutrition supplies but restrict spatial limits, equivalently speaking, cells can grow exponentially as they are in free growth environments but their overall population is capped to a maximum value. After their overall population has reached the maximum, the dynamics changes: On average, whenever a new cell is born, one cell in the environment will be randomly washed out because of the space constraints.

We perform another numerical experiment by simulating the cellular population dynamics in such an ideal chemostat environment. Figures 5.6a and 5.6b show the typical trajectories from Gillespie simulations for initial states (1, 1) and (100, 100) respectively. Different colors of curves in the figures represent the trajectories from different runs. As can be seen in these figures, the variabilities of cell populations in different initial conditions are distinct when the overall populations are less than the maximum (= 1000 here), which is the same as that in free growth environments. After the cell populations reach the maximum, the population dynamics and the variations start to change. After an interesting transient period, it reach a new steady state.

Figure 5.6c shows the evolutions of intra-colony, cross-colony and overall variations in chemostat environments. All of the variation indices starting from different initial states converge to the same ones. Moreover, cross-colony variations go to zero. This is consistent

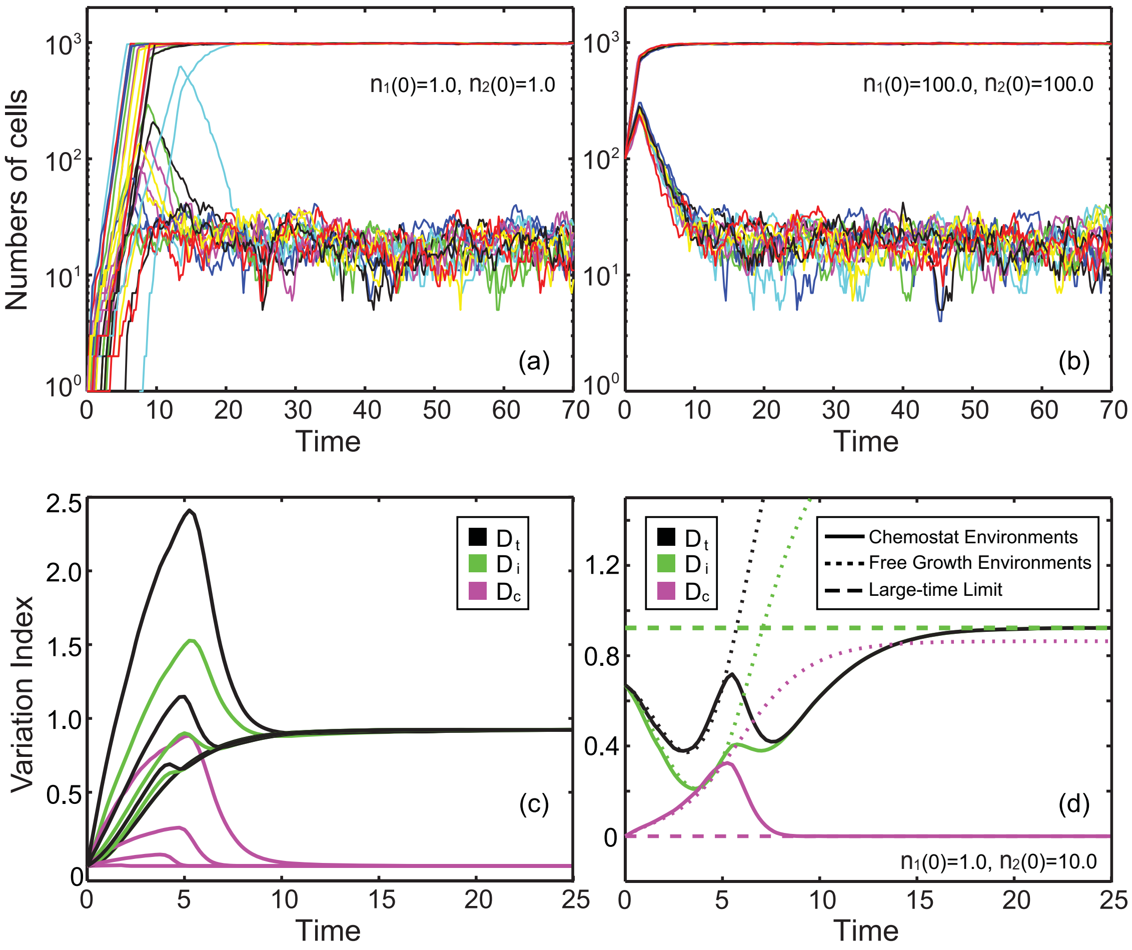

Figure 5.6: Cellular population variations in chemostat environments. (a) and (b), Typical trajectories of cell populations of the two phenotypes from the Gilespie simulations. (a) is for the initial state ($n_1(0) = n_2(0) = 1$) and (b) is for the initial state ($n_1(0) = n_2(0) = 100$). X-axis is time and Y-axis is cell population. Different colors of the curves represent the trajectories from different runs. The maximum population is chosen as 1000 and the rest parameters are the same as those in Fig. 5.3. (c), Variations collapse in chemostat environments. Green, magenta and black sets of curves correspond to intra-colony, cross-colony, and overall variations for different initial conditions (From up to down: (1,1), (3,3), (10,10) and (100,100)). After transit differences, all of the cross-colony go to zero and all of the intra-colony variations and overall variations come to a same one regardless of their different initial conditions. (d) Comparison of variation indices in chemostat environments from simulations with analytical limits. Green solid, magenta solid, and black solid curves are intra-colony, cross-colony, and overall variations from the simulations respectively; green dotted, magenta dotted, and black dotted curves are intra-colony, cross-colony, and overall variations in free growth environments, green dashed, magenta dash, and black dash curves are the long time limits of intra-colony, cross-colony, and overall variations from analytical calculations.

with the previous statement that cross-colony variation goes to zero when the population is the same from colony to colony since the overall populations here are constrained to be identical in chemostat environments.

The variation of cell populations cannot be calculated analytically in these chemostat environments, but the initial stage and the long-time limit of variation are accessible (detailed in Appendix) and are comparable to those from simulations. The steady state distribution of cellular populations in the long-time limit is expressed as

$$P(n_1) = \prod_{i=1}^{n_1} \left(\frac{t_{12}[N^* - (i-1)] + g_1(i-1)(1 - \frac{i-1}{N^*})}{t_{21} i + g_2 i (1 - \frac{i}{N^*})} \right) P(0) \tag{5.8}$$

where N^* is the population maximum, $P(0)$ is determined by normalization and from which variation indices are derived.

Figure 5.6d shows that the variation indices in chemostat environments are the same as those in free growth environments at the beginning when overall populations are away from the maximum. They start to deviate from those in free growth environments when overall populations approach the maximum, and finally reach new steady states regardless of initial conditions after a period of time. As shown in the figure, the transition time of variation indices from those in free growth environments to the new steady long-time limits could be long ($\sim$ 13 here), which tells that it needs a quite long time to have concrete statistics for experiments with different initial states after the overall populations have reached the maximum.

5.3 Conclusions and Discussions

Population diversity is one of the most important and fascinating characters of life designed by nature. It can benefit a community, e.g., reinforces the survival probability of species the under selective pressures and therefore offers better chances for life to evolve and to adapt to fluctuating and unpredictable environments [104, 91, 60, 121]. On the other hand, there are cases that population unification is highly demanded, for instance in the development of multicellular organisms [2].

A general population variation index was proposed to measure the diversity and variability of cellular populations. Using a two-phenotype community as an example, we examined the propagation of population variations in different environments and with initial conditions. One of the interesting results is that the initial state of a cell population is important to its variation in free growth environments and lasts even in the long time limit. However, this effect loses in chemostat environments upon the overall population has reached the maximum of the environments and all variations collapse to a same one regardless of their initial states.

This study provides a theoretical tool to study the variability and diversity of cellular populations. It also brings some suggestions for basic experimental protocols. Microfluidics techniques have revolutionary impacts on biological experiments in the past several years and are widely used in single cell experiments. Chambers in microfluidic chips are good chemostat environments with finite space to contain a certain numbers of cells. This study implies that it might be important to have a larger initial population or longer experimental time in such chemostat environments to decrease run-to-run variability and increase the accuracy of experiments. Another example is Polymerase Chain Reaction (PCR) [2]. With sufficient supply of DNA primers, Polymerases, Deoxynucleotide triphosphates, and buffer solution, a template DNA piece is amplified exponentially without any nutrition or space limitations. This amplification reaction is pretty much a free growth process. Since there is a certain probability that a template is not perfectly amplified, our study suggests that it is important to have a larger amount of initial DNA templates to obtain final DNAs with a higher purity.

Acknowledgements: Portions of this chapter appear in part in Lu T, Shen T, Bennett M, Wolynes P, and Hasty J, Phenotypic variability of growing cellular populations, submitted.

6

A Synthetic Switch with Phenotypic and Genotypic Transitions

In this thesis, we have studied stochasticity in different biological systems and at different levels, such as fluctuations of molecule numbers, variations of cellular populations and conformational changes of signaling molecules. However, this variability does not change the genetic content of the biological systems. Therefore, all of theses changes represent phenotypic variability. Here we present our work on variability in a gene circuit which contains genetic modifications as well as phenotypic variability.

6.1 A Bias of Cellular Populations

The circuit of our system is close to the toggle switch by Gardner et al. [41]. As shown in Figure 6.1(a), it consists of two constituitive promoters P_L and P_{trc-2} that regulate the expression of the $LacI$ and CI genes respectively. LacI is a repressor of the P_{trc-2} promoter and CI is a repressors of the P_L promoter. A ssra tagged-gene was fused to the CI gene to increase the degradation rate of the CI protein. The GFP gene is expressed from the P_{trc-2} promoter to serve as a reporter.

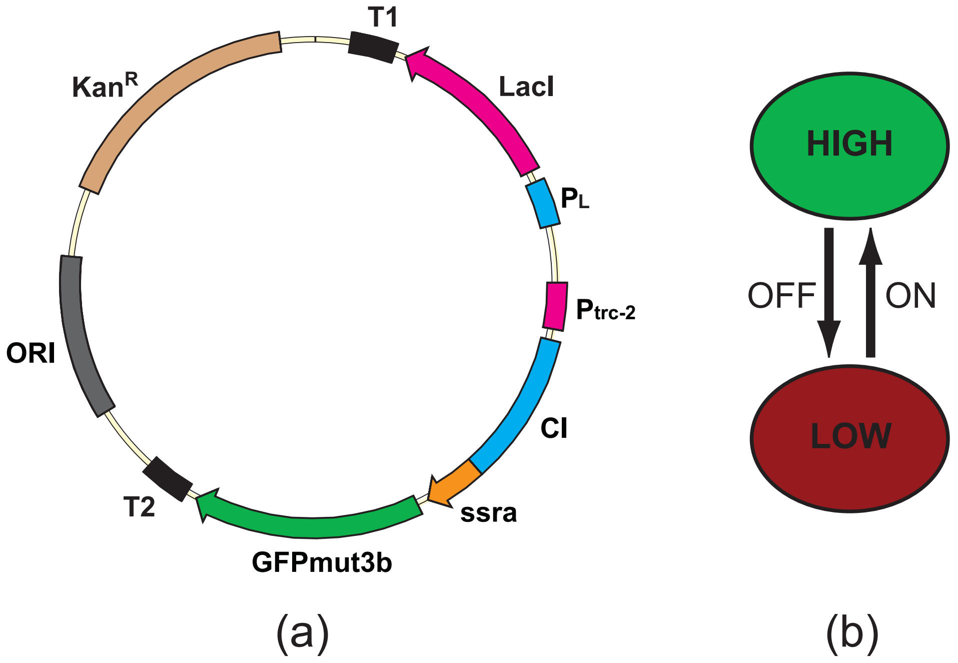

Figure 6.1: (a)Plasmid map of the genetic switch. The circuit is similar to the toggle switch by Gardner et al. [41]. It consists of two constituitive promoters P_L and P_{trc-2} that regulate the expression of $LacI$ and CI genes respectively. LacI is a repressor of the P_{trc-2} promoter and CI is a repressors of the P_L promoter. A ssra tagged-gene was fused to the CI gene to increase the degradation rate of the CI protein. The GFP gene is expressed from the P_{trc-2} promoter to serve as a reporter. (b)Two states of the switch according to its GFP expression level. Increasing temperature shifts the switch to a low GFP expression level while increasing IPTG concentration shifts the switch to a high GFP expression level.

Due to the design of the circuit, there are two states for the switch's GFP expression level – one high state and one low state, which are shown in Figure 6.1(b). This allele of the CI gene is temperature sensitive. It does not fold correctly when the environmental temperature is physiologically high ($> 42^o$C). Consequently this means that its binding affinity to the P_L promoter decreases quickly as a function of temperature, which results in a high expression level of the LacI gene and the switch is channeled into the low state at this temperature. LacI proteins bind to P_{trc-2} in their tetramer form. However, in the presence of Isopropyl β-D-1-thiogalactopyranoside (IPTG) the LacI protein undergoes a conformational change which causes the loss of its binding affinity to the promoter. In this situation, both the CI gene and the GFP gene are highly expressed and the switch is channeled into the high state. This switch is hence controllable by the environment: Increasing temperature shifts the switch to a low GFP expression level while increasing IPTG concentration shifts the switch to a high GFP expression level.

A set of flow cytometry experiments with different temperature and IPTG concentrations allow us to determine the optimal environment for cells in specific states. A culture at 42^oC and without IPTG (42^oC, 0M) is beneficial for the cells in the low state, which is termed a low state environment. Cells in a culture at 30^oC and with 10^{-3}M IPTG (30^oC, 10^{-3}M) are mainly in high GFP expression levels with few of them in the low state, we term this environment a high state environment. Cells in 30^oC and without IPTG (30^oC, 0M) can stay in either the low or the high states, which we call a bistable state environment. We can logically operate this switch by changing cellular environments.

Figure 6.2 illustrates the dynamics of a cell population containing this switch in a changing environment. The cellular environment starting with the high state is sequentially shifted to the bistable, low, bistable, high, and bistable environments over 28 hours. The middle panel illustrates the mean fluorescence of cell populations in the experiment. Background colors (orange, green, and sky blue) represents different environments (High, Bistable, and Low respectively). As can be seen clearly, mean fluorescence level is high (low) when cells are grown in the high (low) state environment and maintains the previous state when cells are shifted to the bistable state environment. However, the mean fluorescence level biases toward the low state after a long time period. It can be further seen in the single-cell fluorescence distributions surrounding the mean fluorescence plot each points in which corresponds to a fluorescence distribution panel (the last point is not shown).

To characterize the origin of the bias, we altered the experiment by changing procedures. A control experiment was done changing the environment of the over night culture from the high state to the low state. The result shows that mean fluorescence increases shortly to a high level when the environment is shifted to the high state. However the fluorescence starts to decrease gradually, which is the same as what happened in the previous experiment. Another control experiment was done by increasing the IPTG concentration to a higher level (10^{-2}M) which was supposed to help the switch stay in the high state. However, this did not reduce the bias (Data not shown). Therefore, neither changing the initial preparation or increasing the IPTG concentration helped to reduce the downward trend bias of mean fluorescence levels.

6.2 Phenotypic and Genotypic Transitions

All experiments in the above section encouraged us to speculate that there may be a third state for this switch in which a cell does not express fluorescence protein but can survive in kanamycin containing media. Furthermore, once a cell transits to this third state, it is rare or impossible to switch back. If this were not the case, a sufficiently high concentration of IPTG should eliminate or, at least largely reduce, the fluorescence bias.

This transition to the third state could be phenotypic, which means the genetic content of the circuit is kept the same as it was before the transition, but the expression of the fluorescent protein has ceased due to an environmental factor. A similar example is bacteria persistence. The transition could also be genotypic such as a random mutation including a recombinant event, in which case the genetic content of circuit is changed.

To determine which is the case, we perform a restriction enzyme digestion followed by gel electrophoresis to check the genetic structure of the plasmid in cells. A cell population initially prepared in the low state environment is split into two: One half of the cells still grow in the low state environment while the other half are changed to grow in the high state environment. At different time points, parts of the cell population were sampled for plasmid contents. To do this we digested the plasmids of the samples and measured the size of resulting DNA fragments by gel electrophoresis. Figure 6.3 shows the plasmid sizes of sampled cells. The left columns correspond to the plasmid sizes of cells in the low state environment all the time. The right columns columns correspond to the plasmid sizes of cells in the high state environment. As can been seen clearly, plasmid sizes of cells in the low state environment are constantly around 6kb over all time periods while those of the cells in the high state environment are initially around 6kb as well but appear to be 2kb also after a certain time. With time going on, the percentage of 2kb plasmids increases and later becomes dominant. This confirms that there is a genotypic transition which occurs for cells in the high state environment.

Upon examination of the plasmid map we found that 6kb is the size of the original plasmid while 2kb is approximately equivalent to the overall sizes of the replication region, antibiotic resistance region and one of the terminators. This suggests that the genotypic transition removed the whole genetic switch from the original plasmid while maintaining the original features of the plasmid.

To determine exactly the mutational mechanism, DNA sequencing of both types of plasmids was performed and verifies that the plasmids after the transition contain the replication and marker regions but do not contain the switch component. It is further verified that there are a identical 9 base pairs in both of the terminators which result in recombination events. This resolves the paradox that the third state of switch does not allow GFP expression but guarantees the survival of cells in the antibiotic environments.

Now we have a complete picture of this synthetic switch. The circuit has two distinct states according to its GFP expression level, which are termed HIGH state and LOW states respectively (Marked with green and red in Figure 6.4). These two states are switchable and the transitions between them are phenotypic. Besides these two states, there is a third state (Marked with gray in the figure) which is unidirectionally transited from both of the switchable states. Therefore this is a switch consisting of both genotypic and phenotypic transitions.

6.3 Modeling

With the finding of the third state of this switch, we are on the stage for quantitative understanding.

6.3.1 A cellular population model

A deterministic description can be employed to model the cellular population dynamics by following the states and transitions illustrated in Fig. 6.4.

$$\begin{aligned}
\dot{N}_h &= (g_h - d_h - w_h - k_h)N_h + k_l N_l \\
\dot{N}_l &= (g_l - d_l - w_l - k_l)N_l + k_h N_h \\
\dot{N}_n &= (g_n - d_n)N_n + w_h N_h + w_l N_l
\end{aligned} \tag{6.1}$$

where N_h, N_l, and N_n are the numbers of cells in the switchable high, the switchable low, and the non-switchable low states, g_h, g_l and g_n are the corresponding growth rates, d_h, d_l and d_n are the corresponding death rates, and k_h and k_l are phenotypic transition rates between the switchable states while w_h and w_l are the genotypic transition rates.

This simple model can be solved exactly with the solution as following

$$N_h = C_1 e^{\alpha t} + C_2 e^{\beta t}$$

$$N_l = C_1 (\tfrac{b_2 - a_1 + \Delta}{2b_1}) e^{\alpha t} + C_2 (\tfrac{b_2 - a_1 - \Delta}{2b_1}) e^{\beta t}$$

$$N_n = C_1 (\tfrac{2a_3 b_1 - a_1 b_3 + b_2 b_3 + b_3 \Delta}{2b_1(\alpha - c)}) e^{\alpha t} + C_2 (\tfrac{2a_3 b_1 - a_1 b_3 + b_2 b_3 - b_3 \Delta}{2b_1}) e^{\beta t} + C_3 e^{ct}$$

where $\alpha = \frac{a_1 + b_2 + \Delta}{2}$, $\beta = \frac{a_1 + b_2 - \Delta}{2}$, $\Delta = \sqrt{(a_1 - b_2)^2 + 4a_2 b_1}$, $a_1 = g_h - d_h - w_h - k_h$, $b_1 = k_l$, $a_2 = k_h$, $a_1 = g_l - d_l - w_l - k_l$, $a_3 = w_h$, $b_3 = w_l$, $c = g_n - d_o$. Coefficients C_1, C_2, and C_3 are determined by initial conditions.

6.3.2 Comparison of the model with experiments

This minimal model is simple, nevertheless, it can capture cellular population dynamics of the switch in changing environments.

To test the model, we start with cells prepared over night in the high state environment, then pass the cells to fresh media next day and keep them growing in the high state environment. The culture is diluted constantly to avoid overgrowth and the OD is kept around $0.1 \sim 0.4$. At time $t = 0$, 3 and 6, some cells are taken out from the culture in the high state environment and passed to low state environments. The results of this experiment is showed in Figure 6.5(a) which illustrates the dynamics of the percentage of cells in High state over time. The black curve corresponds to the population that grow in the high state environment all the time while blue, red, and green curves correspond to the populations that are shift to the low state environment at time $t = 0$, 3, and 6. Besides the percentage of cells in the high state, we have the dynamics of single-cell fluorescence distributions which are indicated in the first two rows of Figure 6.6: Each frame corresponds to the fluorescence distribution at a certain time. The background colors of sky blue and orange represents the high and low state environments. Clearly, the population in the high state environment is shrinking gradually in the high state environment (First row) and goes to zero much more quickly when it is shifted to the low state environment(Second row).

Figure 6.5(b) shows the results from the minimal population model. The colors of curves have the exact means as those corresponding ones in Fig. 6.5(a). Both qualitative and quantitative behaviors of population dynamics from the model and the experiment are consistent.

Similarly, cells prepared initially in the low state environment can grow in the same environment through out the experiment or can be shifted to the high state environment at different times ($t = 0, 3, 6$). The results are shown in Fig. 6.5(c) which are consistent with the corresponding modeling results Fig. 6.5(d). The corresponding single cell results are indicated in the third and fourth rows of Fig. 6.6.

6.4 Conclusions and Discussions

In this work, we studied a three-state synthetic switch in changing environments. Two of the three states have the same genetic content and are switchable depending on environmental conditions. However, transitions to the third state are irreversible and accompany the change of genetic content. This simple switch has rich content by exhibiting both phenotypic and genotypic transitions.

The genotypic transition events to the third state are rare and have disastrous consequences at the cellular population level. Because of a much heavier metabolic load for a cell in the switchable high state, it grows much slower than those in either the switchable low state or the non-switchable low state. This means that a single genotypic transition will cause the non-switchable state cells to take over a whole cell population finally.

One lesson we can learn from this study is to check possible DNA repeat in a designed circuit before we build it. It would help to at least reduce the chance for recombination. It is important to design criteria in synthetic biology in which we are aiming to build specific circuits.

This study also shows the evolution of biological systems. Because of heavier metabolic burdens for the switchable high state, cell population adapt to the non-switchable third state in which cells can survive but do not express fluorescence proteins. This is a great example demonstrating the Darwin's theory of evolution.

Acknowledgements: Portions of this chapter appear in part of Lu T, Stricker J. and Hasty

J, A synthetic switch with phenotypic and genotypic transitions, to be submitted.

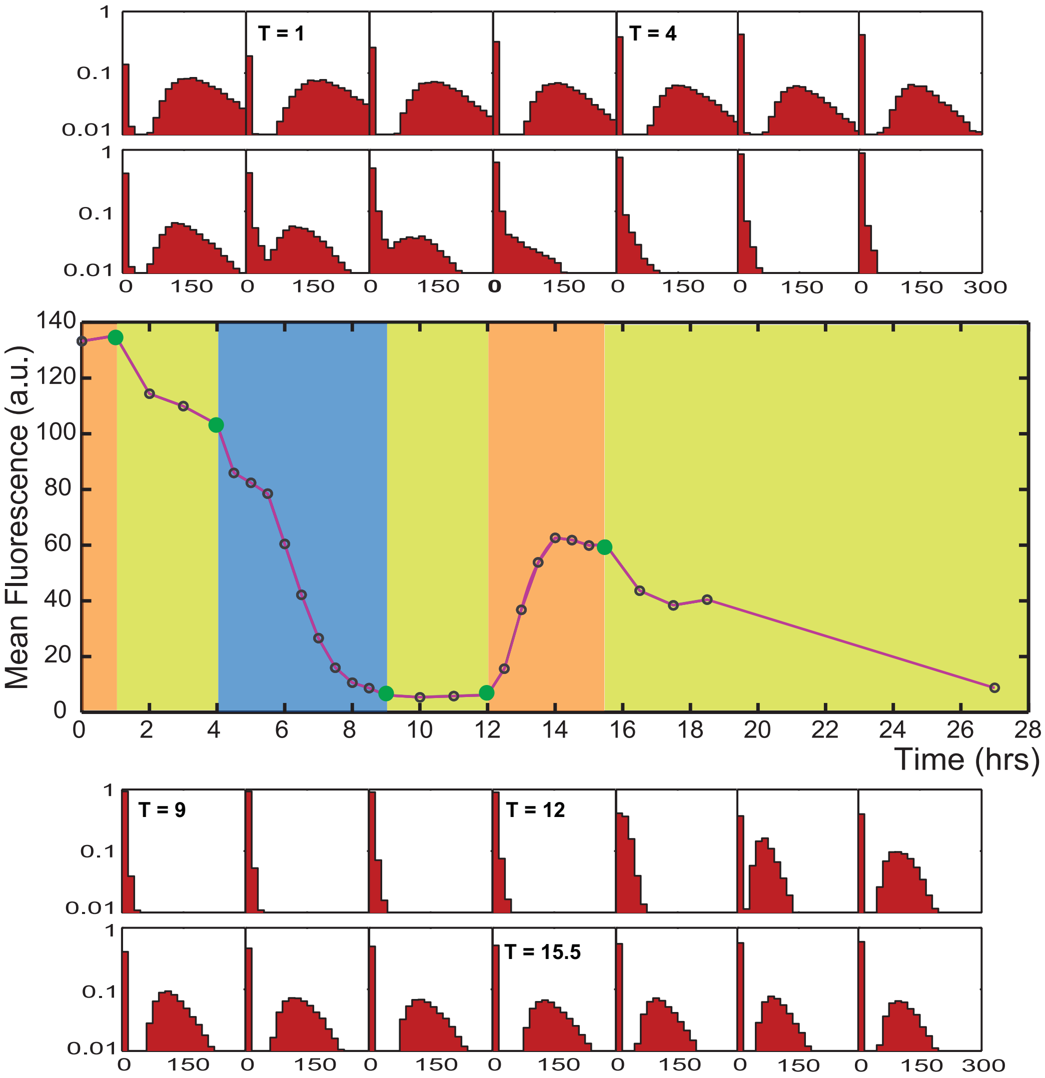

Figure 6.2: A long run of experiment shows the bias of GFP expression level. The middle plot is the time course of mean fluorescence of single cells in changing environments. Different background colors represent different environmental conditions where cells live: light blue, peridot and pumpkin correspond to Low state environment (42^oC, 0M), Bistable state environment(30^oC, 0M), and High state environment(30^oC, 10^{-3}M). The upper two and lower two rows of panels are single-cell fluorescence probability distributions at specific time points each of which corresponds to a data point in the middle plot. The panels marked with time($T = 1, 4, 9, 12, 15.5$) correspond to the moments for environmental changes that are denoted with green spots in the middle plot.

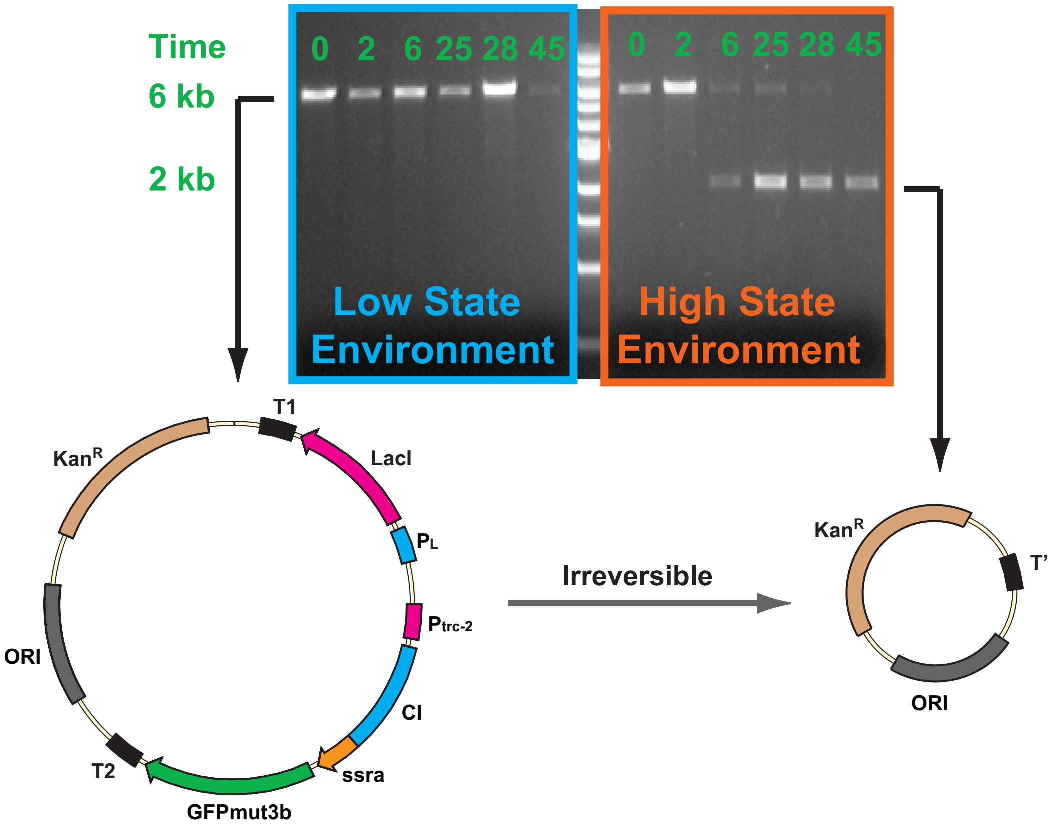

Figure 6.3: a restriction enzyme digestion followed by gel electrophoresis confirms genotypic transitions. The left column is the plasmid DNAs from cells grown in the low state environment while the right column is the plasmid DNAs from cells grown in the high state environment.

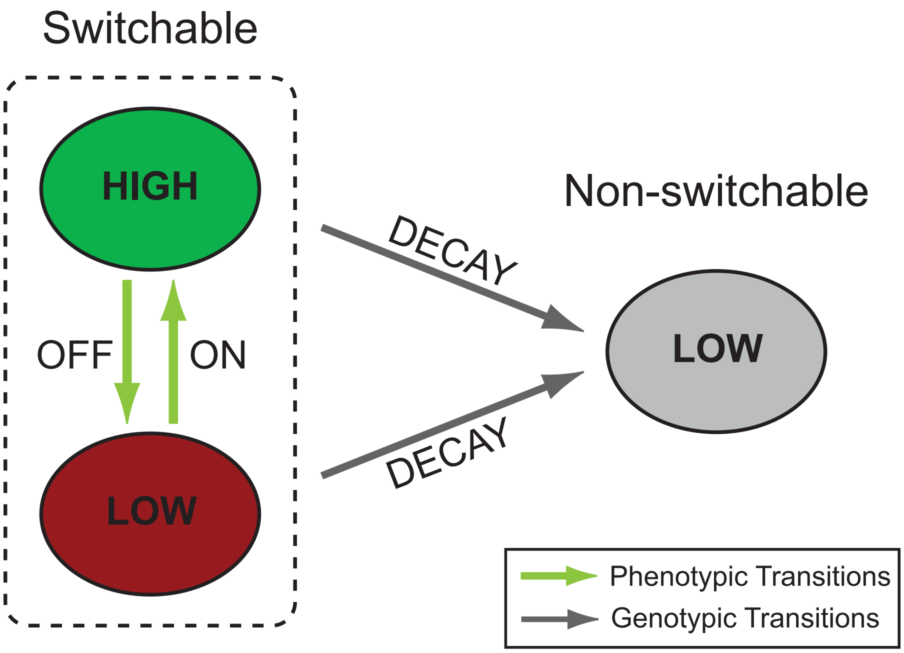

Figure 6.4: Phenotypic and genotypic transitions of the switch. The circuit has two distinct states corresponding to its GFP expression level. They are termed HIGH and LOW states (Marked with green and red in the plot). These two states are switchable and their transitions are phenotypic. Besides those, there is a third state (Marked with gray) which can be unidirectionally transited from both switchable states. Circuit's gene content is changed in these types of transitions.

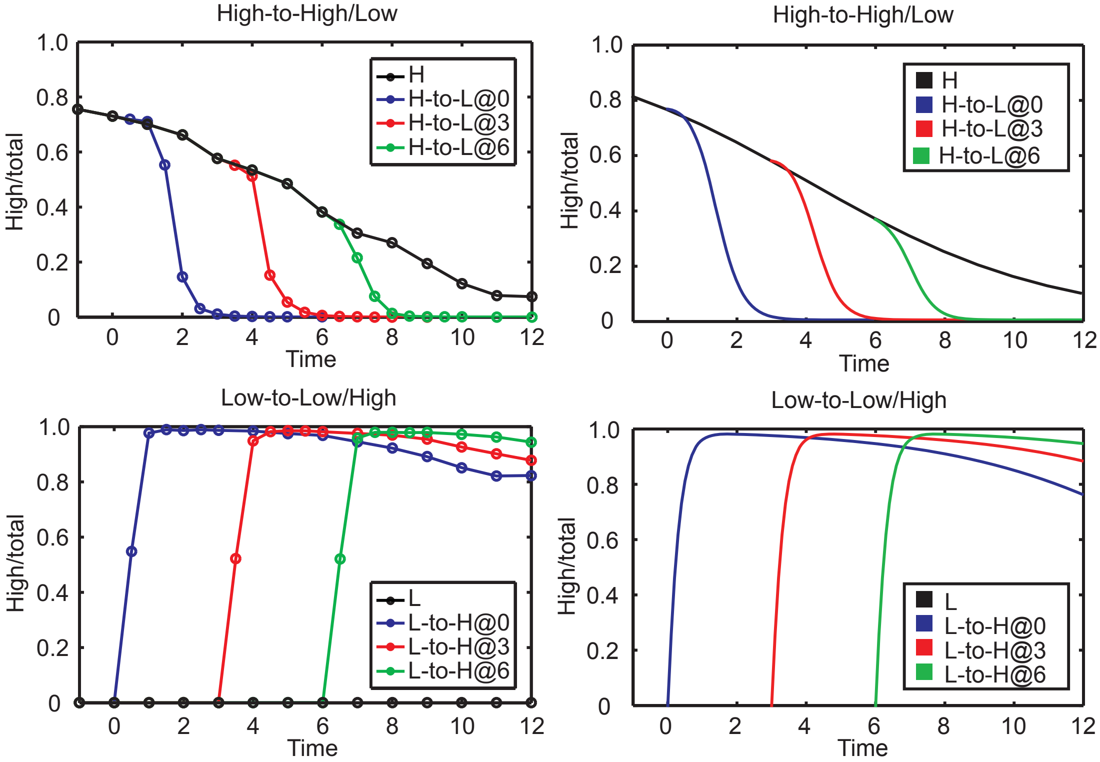

Figure 6.5: Cellular population dynamics in changing environments. (a)Percentage of cells in High state in different environments. Black curve: A cell population grows in High state environment. At $t = 0, 3, 6$ a portion of cells in High state environment(Black curve) are shifted to Low state environment. (b)Modeling results for cell population starting with the high state environment. (c)Percentage of cells in High state in different environments. Black curve: A cell population grows in Low state environment. At $t = 0, 3, 6$ a portion of cells in Low state environment(Black curve) are shifted to High state environment. (d)Modeling results for cell population starting with the low state environment.

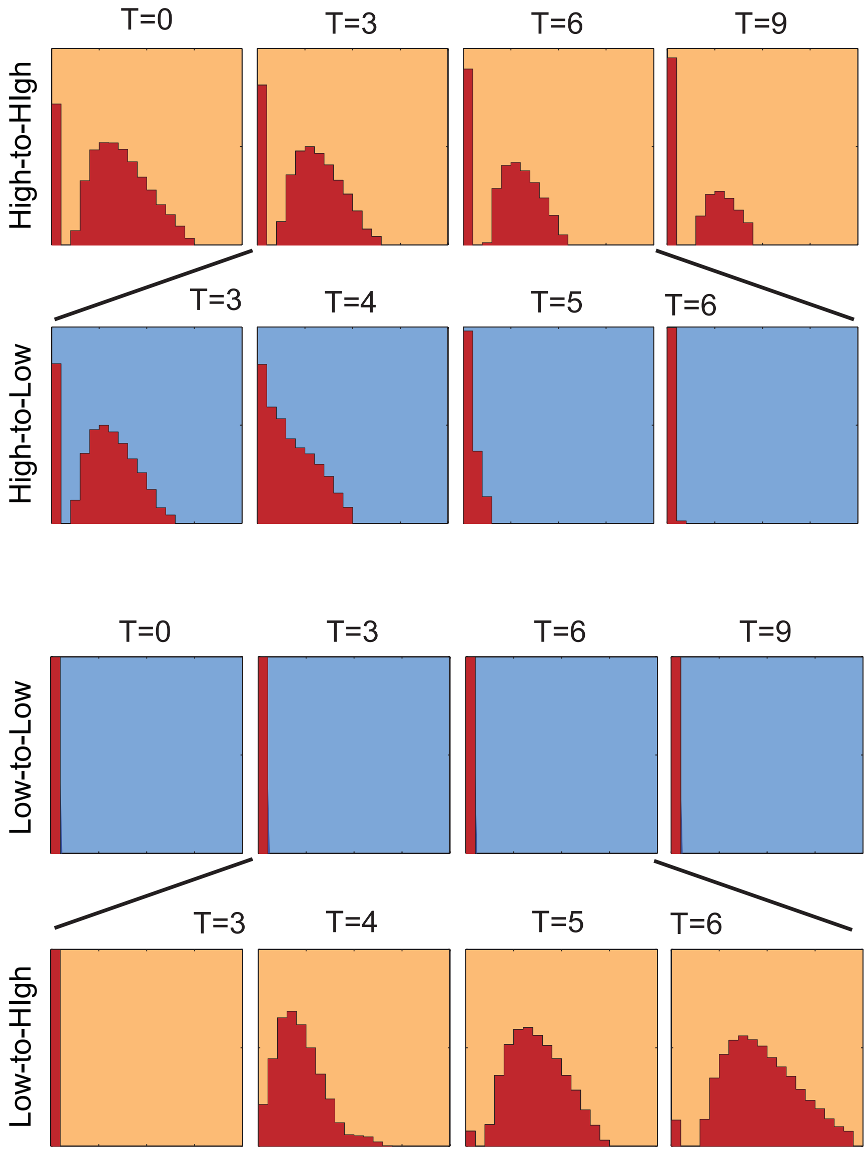

Figure 6.6: Propagation of single-cell fluorescence distribution in changing environments. The first and third rows correspond to the fluorescence distributions of cells in the high and the low state environments at time $t = 0, 3, 6, 9$ respectively. The second (fourth) corresponds to the fluorescence distributions of cells that are shifted from the high (low) state to the low (high) state environments starting with time $t = 3$. The background colors blue and orange indicate the low and the high state environments.

7

Concluding Remarks

7.1 Review

In this work, many aspects of stochasticity in biological networks were examined by both developing general methodological techniques and investigating several important network architectures. As a conceptual development, I proposed a measurable quantity, effective temperature, to quantify noise in stochastic kinetics and genetic networks. As a practical improvement to existing simulation approaches, I derived a generalized Gillespie simulation allowing for stochastic simulation of biological and chemical systems with time-dependent reaction rates or time-dependent system volumes. Besides these two developments, I studied a specific network topology, termed as ladder reaction network, that commonly occurs in signal transduction and gene regulation networks. I studied the signal transduction time of this architecture by mapping signal transduction to a random walker problem and found an optimal enzyme concentration that favors rapid signal transduction. Fluctuations at molecular level are not the only aspect of noise. By proposing a generalized variation index, I studied the consequences of stochasticity for cellular populations. I found that the variation of a cellular population may depend strongly on the population initial state and corresponding environments. A synthetic switch with phenotypic and genotypic transitions was also studied. A bias of cellular populations to their low state was puzzled out by the combined experimental and computational approaches.

7.2 Significance

The studies undertaken help us to better understand the designs of natural biological network architectures and to understand how noise can diversify species and aid in evolution. This work also gives insights to better design strategies for synthetic biology. Furthermore, the proposed concepts, such as effective temperature and variation index, derived simulation algorithm, as well as modeling approaches used in this work are broadly applicable to the study of many other biological systems.

A

Effective Temperature

A.1 The Calculation of Correlation and Response Functions for a Birth-Death Process

In order to calculate the correlation and response functions, we need an expression for the time-dependent observables in "Heisenberg" representation. From Eqs.(2.14,2.17), we find that the expressions of $\hat{a}^+(t)$ and $\hat{a}(t)$ are sufficient to describe the birth-death process.

The non-Hermitian time-dependent expression of $\hat{a}$ is

$$\hat{a}^+(t) \equiv e^{-\hat{L}t} \cdot \hat{a}^+ \cdot e^{\hat{L}t} \tag{A.1}$$

with the Lagrangian of the system $\hat{L} = k_g(\hat{a}^+ - 1) + k_d(\hat{a} - \hat{a}^+\hat{a})$.

To do this we use the following identity

$$e^{x\hat{A}}\hat{B}e^{-x\hat{A}} = \hat{B} + \frac{x}{1!}[\hat{A}, \hat{B}] + \frac{x^2}{2!}[\hat{A}, [\hat{A}, \hat{B}]] + \frac{x^3}{3!}[\hat{A}, [\hat{A}, [\hat{A}, \hat{B}]]] + \cdots \tag{A.2}$$

The algebra of the operators is given as

$$\begin{aligned}
\hat{a}^+ \quad & = \hat{a}^+ \\
[\hat{L}, \hat{a}^+] \quad & = k_d(1-\hat{a}^+) \\
[\hat{L}, [\hat{L}, \hat{a}^+]] \quad & = -k_d^2(1-\hat{a}^+) \\
[\hat{A}, [\hat{A}, [\hat{A}, \hat{B}]]] \quad & = k_d^3(1-\hat{a}^+) \\
& \cdots
\end{aligned}$$

Summarizing all of the above terms yields the expression for $\hat{a}^+(t)$:

$$\begin{aligned}
\hat{a}^+(t) \quad & = \hat{a}^+ + \frac{-1}{1!}k_d(1-\hat{a}^+) - \frac{1}{2!}k_d^2(1-\hat{a}^+) + \frac{-1}{3!}k_d^3(1-\hat{a}^+) + \cdots \\
& = (\hat{a}^+ - 1)e^{k_d t} + 1 \qquad (A.3)
\end{aligned}$$

Similarly, we have an expression for $\hat{a}(t)$

$$\hat{a}(t) \equiv e^{-\hat{L}t} \cdot \hat{a} \cdot e^{\hat{L}t} = (\hat{a} - \frac{k_g}{k_d})e^{-k_d t} + \frac{k_g}{k_d} \qquad (A.4)$$

Now we can calculate the correlation and response functions. For a perturbation of generation rate $k_g \to k_g + h(t)$, the Lagrangian $\hat{L}$ is perturbed by a small term $-h(t)(\hat{a}^+ - 1)$. Using the time-dependent expressions of $\hat{a}(t)$ and $\hat{a}(t)^+$, we have the correlation and response functions:

$$\begin{aligned}
C(t', t) \quad & = \langle \hat{a}^+(t')\hat{a}(t')\hat{a}^+(t)\hat{a}(t) \rangle \\
& = \langle 0 | e^{\hat{a}} \cdot \hat{a}^+(t')\hat{a}(t')\hat{a}^+(t)\hat{a}(t) \cdot e^{\frac{k_g}{k_d}(\hat{a}^+ - 1)} | 0 \rangle \\
& = u_s^2 + u_s e^{-k_d(t'-t)} \qquad (A.5) \\
R(t', t) \quad & \equiv \frac{\delta \langle \hat{a}^+(t')\hat{a}(t') \rangle}{\delta h(t)} \\
& = \langle 0 | e^{\hat{a}} \cdot [\hat{a}^+(t')\hat{a}(t')(\hat{a}^+(t) - 1) - (\hat{a}^+(t') - 1)\hat{a}^+(t)\hat{a}(t)] \cdot e^{\frac{k_g}{k_d}(\hat{a}^+ - 1)} | 0 \rangle \\
& = e^{-k_d(t'-t)} \qquad (A.6)
\end{aligned}$$

The corresponding effective temperature is derived from the definition Eq.(2.5):

$$T_{eff}(\omega) \equiv \frac{\omega \tilde{C}_{12}'(\omega)}{\tilde{R}_{12}''(\omega)} = \frac{-\partial_t C(t',t)}{R(t',t)} = k_g \tag{A.7}$$

For a perturbation of degradation rate, the Liouvillian is perturbed by a small term $-h(t)(\hat{a} - \hat{a}^+\hat{a})$. The correlation and response function are

$$\begin{aligned} C(t',t) &= \frac{1}{2}\langle \hat{a}^+(t')\hat{a}(t')\hat{a}^+(t)\hat{a}(t)\hat{a}^+(t)\hat{a}(t)\rangle \\ &= \frac{1}{2}\langle 0|e^{\hat{a}} \cdot \hat{a}^+(t')\hat{a}(t')\hat{a}^+(t)\hat{a}(t)\hat{a}^+(t)\hat{a}(t) \cdot e^{\frac{k_g}{k_d}(\hat{a}^+-1)}|0\rangle \\ &= \frac{1}{2}(u_s^3 + u_s^2 + 2u_s^2 e^{-k_d(t'-t)} + u_s e^{-k_d(t'-t)}) \qquad \text{(A.8)} \\ R(t',t) &\equiv \frac{\delta\langle \hat{a}^+(t')\hat{a}^(t')\rangle}{\delta h(t)} \\ &= \langle 0|e^{\hat{a}} \cdot \hat{a}^+(t')\hat{a}(t')[\hat{a}(t) - \hat{a}^+(t)\hat{a}(t)] - [\hat{a}(t') - \hat{a}^+(t')\hat{a}(t')]\hat{a}^+(t)\hat{a}(t) \cdot e^{\frac{k_g}{k_d}(\hat{a}^+-1)}|0\rangle \\ &= u_s e^{-k_d(t'-t)} \qquad \text{(A.9)} \end{aligned}$$

from which the effective temperature can be derived

$$T_{eff}(\omega) \equiv \frac{\omega \tilde{C}_{12}'(\omega)}{\tilde{R}_{12}''(\omega)} = \frac{-\partial_t C(t',t)}{R(t',t)} = k_g + \frac{1}{2}k_d \tag{A.10}$$

A.2 Solving the Eqs.(2.20, 2.21)

By taking the fourier transform of Eqs.(2.20, 2.21), we have

$$(d_1 - i\omega)\tilde{A}(\omega) - f_2\tilde{B}(\omega) = D_1\tilde{\xi}_1(\omega) + D_3^a\tilde{\xi}_3(\omega) + D_4^a\tilde{\xi}_4(\omega) + k_{g1}\delta(\omega) \tag{A.11}$$

$$-f_1\tilde{A}(\omega) + (d_2 - i\omega)\tilde{B}(\omega) = D_2\tilde{\xi}_2(\omega) + D_3^b\tilde{\xi}_3(\omega) + D_4^b\tilde{\xi}_4(\omega) + k_{g2}\delta(\omega) \tag{A.12}$$

where the noise intensities D_1, D_2, D_3^a, D_3^b, D_4^a and D_4^b depend only on the steady state values A^* and B^* of the two species.

These two equations can be solved as

$$\tilde{A}(\omega) = \frac{D_1(d_2-i\omega)\tilde{\xi}_1(\omega)+D_2f_2\tilde{\xi}_2(\omega)+[D_3^a(d_2-i\omega)+D_3^bf_2]\tilde{\xi}_3(\omega)+[D_4^a(d_2-i\omega)+D_4^bf_2]\tilde{\xi}_4(\omega)}{\omega^2-i\omega(d_1+d_2)+f_1f_2-d_1d_2} \tag{A.13}$$

$$\tilde{B}(\omega) = \frac{D_1f_1\tilde{\xi}_1(\omega)+D_2(d_1-i\omega)\tilde{\xi}_2(\omega)+[D_3^af_1+D_3^b(d_1-i\omega)]\tilde{\xi}_3(\omega)+[D_4^af_1+D_4^b(d_1-i\omega)]\tilde{\xi}_4(\omega)}{\omega^2-i\omega(d_1+d_2)+f_1f_2-d_1d_2} \tag{A.14}$$

where the delta functions are ignored since they have no contributions in the later calculations of correlation and response functions.

The autocorrelation functions for the species A and B are thus

$$C_{AA} = \frac{(\omega^2+d_2^2)D_1^2+f_2^2D_2^2+[(d_2D_3^a+f_2D_3^b)^2+\omega^2(D_3^a)^2]+[(d_2D_4^a+f_2D_4^b)^2+\omega^2(D_4^a)^2]}{(\omega^2+f_1f_2-d_1d_2)^2+(d_1+d_2)^2\omega^2} \tag{A.15}$$

$$C_{BB} = \frac{f_1^2D_1^2+(\omega^2+d_1^2)D_2^2+[(f_1D_3^a+d_1D_3^b)^2+\omega^2(D_3^b)^2]+[(f_1D_4^a+d_1D_4^b)^2+\omega^2(D_4^b)^2]}{(\omega^2+f_1f_2-d_1d_2)^2+(d_1+d_2)^2\omega^2} \tag{A.16}$$

In order to calculate the response functions, we introduce two small perturbations $\tilde{h}_1(\omega)$ and $\tilde{h}_1(\omega)$ to the system

$$(d_1-i\omega)\tilde{A}(\omega)-f_2\tilde{B}(\omega) = D_1\tilde{\xi}_1(\omega)+D_3^a\tilde{\xi}_3(\omega)+D_4^a\tilde{\xi}_4(\omega)+\tilde{h}_1(\omega)+k_{g1}\delta(\omega) \tag{A.17}$$

$$-f_1\tilde{A}(\omega)+(d_2-i\omega)\tilde{B}(\omega) = D_2\tilde{\xi}_2(\omega)+D_3^b\tilde{\xi}_3(\omega)+D_4^b\tilde{\xi}_4(\omega)+\tilde{h}_2(\omega)+k_{g2}\delta(\omega) \tag{A.18}$$

From which the response functions can be derived

$$R_{AA} = \frac{\delta\langle\tilde{A}(\omega)\rangle}{\delta\tilde{h}_1(\omega)} = \frac{i\omega-d_2}{\omega^2+i\omega(d_1+d_2)+f_1f_2-d_1d_2} \tag{A.19}$$

$$R_{BB} = \frac{\delta\langle\tilde{B}(\omega)\rangle}{\delta\tilde{h}_2(\omega)} = \frac{i\omega-d_1}{\omega^2+i\omega(d_1+d_2)+f_1f_2-d_1d_2} \tag{A.20}$$

B

Moment Equations for the Single Gene - No Feedback System

Similar to [62], we introduce the time-dependent probability p_x^s to have promoter in state $s = [0, 1]$ and x proteins. The evolution of this probability between cell division times is described by the two master equations

$$\dot{p}_x^0 \quad = \alpha_0(p_{x-1}^0 - p_x^0) + k_x[(x+1)p_{x+1}^0 - xp_x^0] + k_{-1}p_x^1 - k_1 v^{-1} p_x^0, \tag{B.1}$$

$$\dot{p}_x^1 \quad = \alpha_1(p_{x-1}^1 - p_x^1) + k_x[(x+1)p_{x+1}^1 - xp_x^1] + k_1 v^{-1} p_x^0 - k_{-1}p_x^1, \tag{B.2}$$

At cell division time t_n, the volume v is halved, and also the number of the proteins in the cell, so $p_x(t_n+) = p_{2x}(t_n-)$.

From Eqs.(B.1),(B.2) we can derive the equations for the partial moments of the distribution of the number of proteins, defined as

$$\langle x^q \rangle_{0,1} \equiv \sum_x x^q p_x^s \tag{B.3}$$

The zeroth moments $s_{0,1} = \langle x^0 \rangle_{0,1}$ give the marginal probabilities of the promoter to be in states $0, 1$, respectively. The equations for $s_{0,1}$ read

$$\dot{s}_0 = -\frac{k_1}{v(t)} s_0 + k_{-1} s_1 \tag{B.4}$$

$$\dot{s}_1 = \frac{k_1}{v(t)} s_0 - k_{-1} s_1 \tag{B.5}$$

The equations for the first moments read

$$\dot{\langle x\rangle}_0 = \alpha_0 s_0 - k_x \langle x\rangle_0 - \frac{k_1}{v(t)}\langle x\rangle_0 + k_{-1}\langle x\rangle_1 \tag{B.6}$$

$$\dot{\langle x\rangle}_1 = \alpha_1 s_1 - k_x \langle x\rangle_1 + \frac{k_1}{v(t)}\langle x\rangle_0 - k_{-1}\langle x\rangle_1 \tag{B.7}$$

and for the second moments,

$$\dot{\langle x^2\rangle}_0 = \alpha_0 s_0 + 2\alpha_0\langle x\rangle_0 + k_x(\langle x\rangle_0 - 2\langle x^2\rangle_0) - \frac{k_1}{v(t)}\langle x^2\rangle_0 + k_{-1}\langle x^2\rangle_1 \tag{B.8}$$

$$\dot{\langle x^2\rangle}_1 = \alpha_1 s_1 + 2\alpha_1\langle x\rangle_1 + k_x(\langle x\rangle_1 - 2\langle x^2\rangle_1) + \frac{k_1}{v(t)}\langle x^2\rangle_0 - k_{-1}\langle x^2\rangle_1 \tag{B.9}$$

Here at cell division times the values of $\langle x\rangle_{0,1}$ and $\langle x^2\rangle_{0,1}$ have to be reset, $\langle x\rangle_{0,1}(t_n+) = \langle x\rangle_{0,1}(t_n-)/2$ and $\langle x^2\rangle_{0,1}(t_n+) = \langle x^2\rangle_{0,1}(t_n-)/4$.

The asymptotic solution for s_0 at large time t can be written in the form

$$s_0(t) = k_{-1}\int_{-\infty}^{t} e^{-\int_{t'}^{t}\left(\frac{k_1}{v(y)}+k_{-1}\right)dy} dt' \tag{B.10}$$

The mean values of $s_{0,1}$ with good accuracy are approximated by the formulas

$$\overline{s_0(t)} = \frac{k_{-1}}{k_{-1} + k_1\overline{v^{-1}}} \tag{B.11}$$

$$\overline{s_1(t)} = \frac{k_1\overline{v^{-1}}}{k_{-1} + k_1\overline{v^{-1}}} \tag{B.12}$$

For small decay rate $k_x \ll 1$, the mean value of the number of proteins $\overline{\langle x\rangle} = \overline{\langle x\rangle_0} + \overline{\langle x\rangle_1}$ can be found from (B.6) assuming that the number of proteins doubles during the cell division time. For small decay rates is simply leads to

$$\overline{\langle x\rangle} = \frac{3}{2}\frac{k_{-1}\alpha_1 + k_1\overline{v(t)^{-1}}\alpha_0}{k_{-1} + k_1\overline{v(t)^{-1}}} \tag{B.13}$$

Similarly, we can obtain the mean variance of the number of proteins at large t, $\overline{\langle x^2 \rangle - \langle x \rangle^2}$ (it has to increase 4 times between consecutive cell divisions).

C

Variational Index

C.1 Variational index

Here, we introduce a general variation index as

$$D \equiv \frac{\int d\mathbf{n}P(\mathbf{n})\frac{1}{S}\sum\limits_{i=1}^{S}{}_{\mathbf{n}}n_i^2 - \left(\int d\mathbf{n}P(\mathbf{n})\frac{1}{S}\sum\limits_{i=1}^{S}{}_{\mathbf{n}}n_i\right)^2}{\left(\int d\mathbf{n}P(\mathbf{n})\frac{1}{S}\sum\limits_{i=1}^{S}{}_{\mathbf{n}}n_i\right)^2} \tag{C.1}$$

where $\int d\mathbf{n}P(\mathbf{n})(\cdot)$ is the average of different population distribution of a community, i.e., ensemble average, while $\frac{1}{S}\sum\limits_{i=1}^{S}{}_{\mathbf{n}}(\cdot)$ is the average of phenotypes. By denoting $\overline{(\cdot)}$ as $\int d\mathbf{n}P(\mathbf{n})(\cdot)$ and $\langle(\cdot)\rangle$ as $\frac{1}{S}\sum\limits_{i=1}^{S}{}_{\mathbf{n}}(\cdot)$, the variational index can be expressed as $D \equiv \frac{\overline{\langle\mathbf{n}^2\rangle}-\left(\overline{\langle\mathbf{n}\rangle}\right)^2}{\left(\overline{\langle\mathbf{n}\rangle}\right)^2}$, which could be further written as

$$\begin{aligned} D &= \frac{\overline{\langle\mathbf{n}^2\rangle}-\overline{\langle\mathbf{n}\rangle^2}}{\left(\overline{\langle\mathbf{n}\rangle}\right)^2} + \frac{\overline{\langle\mathbf{n}\rangle^2}-\left(\overline{\langle\mathbf{n}\rangle}\right)^2}{\left(\overline{\langle\mathbf{n}\rangle}\right)^2} \\ &\equiv D_i + D_c \end{aligned} \tag{C.2}$$

where D_i infers to intra-colony index and D_c infers to cross-colony variation.

By exchanging the average sequence (summation and integral), we may rewrite the variational index Eq.(C.2) as

$$\begin{aligned}D_t &= \frac{\frac{1}{S}\sum\limits_i\left(\overline{n_i^2}-\overline{n}_i\right)-\frac{1}{S^2}\sum\limits_{i,j}\left(\overline{n_in_j}-\overline{n}_i\right)}{\frac{1}{S^2}\left(\sum\limits_i\overline{n}_i\right)^2}+\frac{\frac{1}{S^2}\sum\limits_{i,j}\left(\overline{n_in_j}-\overline{n}_i\right)-\frac{1}{S^2}\left(\sum\limits_i\overline{n}_i\right)^2}{\frac{1}{S^2}\left(\sum\limits_i\overline{n}_i\right)^2}\\ &= \frac{S\sum\limits_i\overline{n_i^2}-\sum\limits_{i,j}\overline{n_in_j}}{\left(\sum\limits_i\overline{n}_i\right)^2}+\frac{\sum\limits_{i,j}\overline{n_in_j}-\left(\sum\limits_i\overline{n}_i\right)^2}{\left(\sum\limits_i\overline{n}_i\right)^2}\end{aligned} \tag{C.3}$$

From which we find that we need only calculate the first and the second moments rather than the whole probability distribution to have the variation index.

C.2 Calculation of the First and Second Moments

The moment equation Eq.(5.5) could be decomposed to two sets of equations, one set for the first moments and the other set for the second moments.

The first moments $\mathbf{M}^{(1)}(t)=\left(\bar{n}_1(t),\bar{n}_2(t)\right)^T$, the mean population of the dynamics, are given by

$$\mathbf{M}^{(1)}(t)=\begin{bmatrix}\bar{n}_1(t)\\ \bar{n}_2(t)\end{bmatrix}=\begin{bmatrix}1 & 1\\ \frac{s_2-s_1+\Delta}{2t_{12}} & \frac{s_2-s_1-\Delta}{2t_{12}}\end{bmatrix}\begin{bmatrix}\gamma_1e^{\theta t}\\ \gamma_2e^{\nu t}\end{bmatrix} \tag{C.4}$$

where $\Delta=\sqrt{(s_1-s_2)^2+4t_{12}t_{21}}$, $\theta=\frac{1}{2}(s_1+s_2+\Delta)$ and $\nu=\frac{1}{2}(s_1+s_2-\Delta)$. The coefficients are given by $\gamma_1=\frac{(s_1-s_2+\Delta)}{2\Delta}\bar{n}_1(0)+\frac{t_{12}}{\Delta}\bar{n}_2(0)$, $\gamma_2=\frac{(s_2-s_1+\Delta)}{2\Delta}\bar{n}_1(0)-\frac{t_{12}}{\Delta}\bar{n}_2(0)$.

We further can obtain the second moments, $\mathbf{M}^{(2)}(t)=\left(\overline{n_1^2}(t),\overline{n_2^2}(t),\overline{n_1n_2}(t)\right)^T$, which are given by

$$\mathbf{M}^{(2)}(t) = \mathbf{Y}(t)\mathbf{C}_2+\mathbf{Y}(t)\int_0^t\mathbf{Y}^{-1}(t')\mathbf{f}(t')dt'=\mathbf{Y}(t)[\mathbf{C}_2+\mathbf{F}(t)-\mathbf{F}(0)] \tag{C.5}$$

where $\mathbf{Y} = \mathbf{K}\mathbf{U}(t)$, $\mathbf{C}_2 = \mathbf{K}^{-1}\mathbf{M}^{(2)}(0)$, $\mathbf{L} = \mathbf{K}^{-1}\mathbf{f}$, and

$$\mathbf{K} = \begin{bmatrix} \frac{(s_1-s_2+\Delta)}{2t_{21}} & \frac{(s_1-s_2-\Delta)}{2t_{21}} & \frac{2t_{12}}{s_2-s_1} \\ \\ \frac{(s_2-s_1+\Delta)}{2t_{12}} & \frac{(s_2-s_1-\Delta)}{2t_{12}} & \frac{2t_{21}}{s_1-s_2} \\ \\ 1 & 1 & 1 \end{bmatrix}, \mathbf{U}(t) = \begin{bmatrix} e^{2\theta t} & 0 & 0 \\ \\ 0 & e^{2\nu t} & 0 \\ \\ 0 & 0 & e^{(\theta+\nu)t} \end{bmatrix}$$

$$\mathbf{f} = \begin{bmatrix} r_1 + \frac{s_2-s_1+\Delta}{2} & r_1 + \frac{s_2-s_1-\Delta}{2} \\ \\ t_{21} + \frac{r_2(s_2-s_1+\Delta)}{2t_{12}} & t_{21} + \frac{r_2(s_2-s_1-\Delta)}{2t_{12}} \\ \\ -(t_{21} + \frac{s_2-s_1+\Delta}{2}) & -(t_{21} + \frac{s_2-s_1-\Delta}{2}) \end{bmatrix}, \mathbf{F}(t) = \begin{bmatrix} \frac{\mathbf{L}_{11}\gamma_1}{-\theta} e^{-\theta t} + \frac{\mathbf{L}_{12}\gamma_2}{\nu-2\theta} e^{(\nu-2\theta)t} \\ \\ \frac{\mathbf{L}_{21}\gamma_1}{\theta-2\nu} e^{(\theta-2\nu)t} + \frac{\mathbf{L}_{22}\gamma_2}{-\nu} e^{-\nu t} \\ \\ \frac{\mathbf{L}_{31}\gamma_1}{-\nu} e^{-\nu t} + \frac{\mathbf{L}_{32}\gamma_2}{-\theta} e^{-\theta t} \end{bmatrix}$$

With the above exact expressions of the first and second moments, we can now study the whole temporal behavior, which can be expressed as a function of $\mathbf{M}(t)$

$$D_t(t) = \frac{M_1^{(2)} + M_2^{(2)} - 2M_3^{(2)}}{(M_1^{(1)} + M_2^{(1)})^2} + \frac{(M_1^{(2)} + M_2^{(2)} + 2M_3^{(2)}) - (M_1^{(1)} + M_2^{(1)})^2}{(M_1^{(1)} + M_2^{(1)})^2} \tag{C.6}$$

For the long-time, asymptotic behaviors of the population dynamics, the exponents are most useful quantities for the characterization, just as the Lyapunov exponents for general exponentially growth or relaxational dynamic systems. Up to the second moments, there are five exponents for the two-phenotype community, (θ, ν, $\theta + \nu$, 2θ, 2ν), where the largest exponent is 2θ. The long-time dynamics is mainly determined by the largest exponent, therefore the two indices asymptotically become

$$\begin{aligned} D_i^{\infty} &= \frac{\big[C_{21} - F_1(0)\big]\big(K_{11} + K_{21} - 2K_{31}\big)}{\gamma_1^2\big[1 + \frac{s_2-s_1+\Delta}{2t_{12}}\big]^2} \\ D_c^{\infty} &= \frac{\big[C_{21} - F_1(0)\big]\big(K_{11} + K_{21} + 2K_{31}\big)}{\gamma_1^2\big[1 + \frac{s_2-s_1+\Delta}{2t_{12}}\big]^2} - 1 \end{aligned} \tag{C.7}$$

C.3 Cellular Populations in Chemostat Environments

The cellular population in a chemostat environment is constrained by the corresponding container or chamber, where the dynamics is another story. It can be described by the following Master equation

$$
\begin{aligned}
\tfrac{d}{dt}P(n_1,n_2,t) &= g_1 n_1^- P(n_1^-,n_2,t)[1-\Theta(n_1^-+n_2-N^*)] + d_1 n_1^+ P(n_1^+,n_2,t) \\
&+ g_2 n_2^- P(n_1,n_2^-,t)[1-\Theta(n_1+n_2^- -N^*)] + d_2 n_2^+ P(n_1,n_2^+,t) + t_{21} n_1^+ P(n_1^+,n_2^-,t) \\
&+ g_1 n_1^- \big(\tfrac{n_2^+}{n_1^- + n_2^+}\big) P(n_1^-,n_2^+,t)\Theta(n_1^- + n_2^+ - N^*) \\
&+ g_2 n_2^- \big(\tfrac{n_1^+}{n_1^- + n_2^+}\big) P(n_1^+,n_2^-,t)\Theta(n_1^+ + n_2^- - N^*) \\
&- \Big[(d_1+t_{21})n_1 + (d_2+t_{12})n_2 + (g_1 n_1 + g_2 n_2)[1-\Theta(n_1+n_2-N^*)] \\
&+ \tfrac{(g_1+g_2)n_1 n_2}{n_1+n_2}\Theta(n_1^- + n_2^+ - N^*)\Big] P(n_1,n_2,t) \qquad \text{(C.8)}
\end{aligned}
$$

where $\Theta(x)$ is a step function with $\Theta(x) = 0 (x < 0)$ or $1 (x >= 0)$, N^* is the maximum population that the environment could contain.

The above equation is hard to solve analytically but could be computed numerically. Moreover, there are two limits that could be achieved analytically. One limit is the initial stage where the cell population is still under maximum N^*, the equation is the same as Eq.(C.12) and has been solved exactly. The other end is the long time limit where the overall population is constrained around the maximum value.

At the maximum population stage, the overall population is always around N^*, i.e. $n_1 + n_2 \simeq N^*$. Therefore the step function $\Theta(n_1^\pm + n_2^\pm - N^*) = 1$. The Eq.(C.8) could be nicely approximated by the following equation

$$
\begin{aligned}
\frac{d}{dt}P(n_1,t) &= \big(t_{12}(N^* - n_1^-) + g_1 n_1^-(1 - \frac{n_1^-}{N^*})\big)P(n_1^-,t) + \big(t_{21}n_1^+ + g_2 n_1^+ \\
&(1 - \frac{n_1^-}{N^*})\big)P(n_1^+,t) + \big(t_{12}(N^* - n_1) + t_{21}n_1 + (g_1+g_2)n_1(1 - \frac{n_1}{N^*})\big)P(n_1,t) \qquad \text{(C.9)}
\end{aligned}
$$

The steady state distribution of the above equation (detailed balance) brings

$$\frac{P(n_1)}{P(n_1-1)} = \frac{t_{12}[N^* - (n_1-1)] + g_1(n_1-1)(1-\frac{n_1-1}{N^*})}{t_{21}n_1 + g_2n_1(1-\frac{n_1}{N^*})} \tag{C.10}$$

From which we have the steady state distribution as

$$P(n_1) = \prod_{i=1}^{n_1}(\frac{t_{12}[N^* - (i-1)] + g_1(i-1)(1-\frac{i-1}{N^*})}{t_{21}i + g_2i(1-\frac{i}{N^*})})P(0) \tag{C.11}$$

where the probability having no cells is

$$P(0) = \Big(1 + \sum_{n_1=1}^{N^*}\prod_{i=1}^{n_1}(\frac{t_{12}[N^* - (i-1)] + g_1(i-1)(1-\frac{i-1}{N^*})}{t_{21}i + g_2i(1-\frac{i}{N^*})})\Big)^{-1}$$

C.4 Cellular Populations in an Environment with Finite Nutrition Supply Rates

The cellular population dynamics in an environment with finite nutrition supply rate is governed by

$$\begin{aligned}
&\frac{d}{dt}P(n_1,n_2,t) = g_1n_1^-(1-\frac{\alpha_1n_1^- + \alpha_2n_2}{N_n})P(n_1^-,n_2,t) + d_1n_1^+P(n_1^+,n_2,t)\\
&+g_2n_2^-(1-\frac{\alpha_1n_1 + \alpha_2^-n_2}{N_n})P(n_1,n_2^-,t) + d_2n_2^+P(n_1,n_2^+,t)\\
&+t_{12}n_2^+P(n_1^-,n_2^+,t) + t_{21}n_1^+P(n_1^+,n_2^-,t) - \big[(d_1+t_{21})n_1\\
&+(d_2+t_{12})n_2 + (g_1n_1+g_2n_2)(1-\frac{\sum_i\alpha_in_i}{N_n})\big]P(n_1,n_2,t)
\end{aligned} \tag{C.12}$$

where the terms $(1-\frac{\Sigma_i\alpha_in_i}{N_n})$ indicate the availability of left nutrition source, α_i is the nutrition cost rate for phenotype i and N_n is the overall nutrition flow.

Bibliography

[1] D. Adalsteinsson, D. McMillen, and T. C. Elston. Biochemical network stochastic simulator (bionets): software for stochastic modeling of biochemical networks. *BMC Bioinformatics*, 5:24, 2004.

[2] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York, USA, fourth edition, 2002.

[3] U. Alon. *An introduction to systems biology: Design principles of biological circuits*. Chapman and Hall/CRC, first edition, 2006.

[4] R. M. Anderson, D. M. Gordon, M. J. Crawley, and M. P. Hassell. Variability in abundance of animal and plant species. *Nature*, 296:245–248, 1982.

[5] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Mol. Syst. Biol.*, 2:44 – 51, 2006.

[6] A. Arkin, J. Ross, and H. H. McAdams. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *escherichia coli* cells. *Genetics*, 149:1633, 1998.

[7] B. E. Aubol, S. Chakrabarti, J. Ngo, J. Shaffer, B. Nolen, X. Fu, G. Ghosh, and J. A. Adams. Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proc. Natl. Acad. Sci. USA*, 100:12601, 2003.

[8] A. Bakk and R. Metzler. Nonspecifi binding of the o_R repressors ci and cro of bacteriophage λ. *J. Theor. Biol.*, 231:525–533, 2004.

[9] A. Bakk, R. Metzler, and K. Sneppen. Sensitivity of phage lambda upon variations of the gibbs free energy. *Israel Journal of Chemistry*, 44:309–315, 2004.

[10] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler. Bacterial persistence as a phenotypic switch. *Science*, 305:1622–1625, 2004.

[11] N. Barkai and S. Leibler. Biological rhythms: Circadian clocks limited by noise. *Nature*, 403:267–268, 1999.

[12] J. M. Bean, E. D. Siggia, and F. R. Cross. Coherence and timing of cell cycle start examined at single-cell resolution. *Mol. Cell*, 21:3–14, 2006.

[13] A. Becskei, B. Séraphin, and L. Serrano. Positive feedback in eukaryotic gene networks: Cell differentiation by graded to binary response conversion. *EMBO J.*, 20:2528, 2001.

[14] A. Becskei and L. Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405:590, 2000.

[15] D. Bedeaux, K. Lindenberg, and K. Shuler. On the relation between master equations and random walks and their solutions. *J. Math. Phys.*, 12:2116–2123, 1971.

[16] U. S. Bhalla and R. Lyengar. Emergent properties of networks of biological signaling pathways. *Science*, 283:381–387, 1999.

[17] W. Bialek. Stability and noise in biochemical switches. In T. K. Leen, T. G. Dietterich, and V. Tresp, editors, *Advances in Neural Information Processing*, volume 13, pages 159–165. MIT Press, Cambridge, MA, 2001.

[18] J. Changeux and S. Edelstein. Allosteric mechanism of signal transduction. *Science*, 308:1424–1428, 2005.

[19] C. Chapon. Expression of malt, the regulator gene of the maltose region in escherichia coli, is limited both at transcription and translation. *EMBO J.*, 1:369–374, 1982.

[20] R. G. Compton and G. Hancock. *Applications of Kinetic Modelling.* Elsevier Science, first edition, 1999.

[21] T. Coulson, E. A. Catchpole, S. D. Albon, B. J. T. Morgan, J. M. Pemberton, T. H. Clutton-Brock, M. J. Crawley, and B. T. Grenfell. Age, sex, density, winter weather, and population crashes in soay sheep. *Nature*, 292:1528–1531, 2001.

[22] G. R. Crabtree and E. N. Olson. Nfat signaling choreographing the social lives of cells. *Cell*, 109:S67–79, 2002.

[23] L. Cugliandolo and J. Kurchan. Thermal properties of slow dynamics. *Physica A.*, 263:242–251, 1999.

[24] L. Cugliandolo, J. Kurchan, and L. Peliti. Energy flow, partial equilibration and effect temp in systems with slow dynamics. *Phys. Rev. E.*, 55:3898–3914, 1997.

[25] G. Dattoli, A. Torre, and R. Mignani. Non-hermitian evolution of two-level quantum systems. *Phys. Rev. A.*, 42:1467–1475, 1990.

[26] R. Demicheli, M. W. Retsky, D. E. Swartzendruber, and G. Bonadonna. Proposal for a new model of breast cancer metastatic development. *Ann Oncol.*, 8:1075–1080, 1997.

[27] T. Doan, A. Mendez, P. Detwiler, J. Chen, and F. Rieke. Muliple phosphorylation sites confer reproducibility of the rod's single-photon responses. *Science*, 313:530–533, 2006.

[28] R. E. Dolmetsch, R. S. Lewis, C. C. Goodnow, and J. I. Healy. Differential activation of transcription factors induced by ca^{2+} response amplitude and duration. *Nature*, 386:855–858, 1997.

[29] J. K. Douglass, L. Wilkens, E. Pantazelou, and F. Moss. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. *Nature*, 365:337–340, 1993.

[30] A. Einstein. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Ann. Phys., Lpz.*, 17:549–560, 1905.

[31] J. El-Ali, P. K. Sorger, and K. F. Jensen. Cells on chips. *Nature*, 442:403–441, 2006.

[32] M. Elowitz, A. Levine, E. Siggia, and P. Swain. Stochastic gene expression in a single cell. *Science*, 297:1183–1186, 2002.

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

[34] G. Eyink. Action principle in nonequilibrium statistical dynamics. *Phys. Rev. E.*, 54:3419–3435, 1996.

[35] J. Ferrel. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *TIBS*, 21:460–466, 1996.

[36] J. Ferrel and E. Machleder. The biochemical basis of an all-or-none cell fate switch in xenopus oocytes. *Science*, 280:895–898, 1998.

[37] H. Frauenfelder, S. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.

[38] L. Gammaitoni, P. Hangi, P. Jung, and F. Marchesoni. Stochastic resonance. *Rev. Mod. Phys.*, 70:223–287, 1998.

[39] T. C. Gard. *Introduction to Stochastic Differential Equations.* Marcel Dekker, New York.

[40] C. W. Gardner. *Handbook of stochastic methods for physics, chemistry and the natural sciences.* Springer, second edition, 1996.

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

[42] M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem.*, 104:1876–1889, 2000.

[43] D. Gillespie. General method for numerically simulating stochastic time evolution of coupled chemical-reactions. *J. Comput. Phys.*, 22:403–434, 1976.

[44] D. Gillespie. The chemical langevin equation. *J. Chem. Phys.*, 113:297–306, 2000.

[45] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[46] D. T. Gillespie. *Markov Processes: An Introduction for Physical Scientists.* Academic Press, 1992.

[47] P. L. Graumann. Different genetic programmes within identical bacteria under identical conditions: the phenomenon of bistability greatly modifie our view on bacterial populations. *Mol. Microbiol.*, 61:560–563, 2006.

[48] J. Gunawardena. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc. Natl. Acad. Sci. USA*, 102:14617–14622, 2005.

[49] D. Halliday, D. Resnick, and J. Walker. *Fundamentals of Physics.* John Wiley & Sons, fifth edition, 1997.

[50] E. L. Haseltine and J. B. Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.*, 117(15):6959–6969, 2002.

[51] J. Hasty, D. McMillen, and J. J. Collins. Engineered gene circuits. *Nature*, 420:224–230, 2002.

[52] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci.*, 97:2075, 2000.

[53] G. Henkelman, M. LaBute, C. Tung, P. Fenimore, and B. H. McMahon. Conformational dependence of a protein kinase phosphate transfer reaction. *Proc. Natl. Acad. Sci. USA*, 102:15347–15351, 2004.

[54] G. D. Hogan, L. Chen, J. Nardone, and A. Rao. Transcriptional regulation by calcium, calcineurin, and nfat. *GENES & DEVELOPMENT*, 17:2205–2232, 2003.

[55] V. Horsley and G. K. Pavlath. Nfat: ubiquitous regulator of cell differentiation and adaptation. *J. Chem. Biol.*, 156:771–774, 2002.

[56] C. Huang and J. Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. USA*, 93:10078–10083, 1996.

[57] M. Jnr. Hunter. *Fundamentals of Conservation Biology.* Blackwell Publishers, Massachusetts, USA, second edition, 2002.

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

[59] M. Kaern, W. J. Blake, and J. J. Collins. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng.*, 5:179–206, 2003.

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

[61] D. B. Kearns and R. Losick. Cell population heterogeneity during growth of bacillus subtilis. *Genes Dev.*, 19:3083–3094, 2006.

[62] T. B. Kepler and T. C. Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.*, 81(6):3116–3136, Dec 2001.

[63] M. Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, London, 1983.

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

[65] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29:255–284, 1966.

[66] J. Kurchan. In and out of equilibrium. *Nature*, 433:222–225, 2005.

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

[68] R. Lande, S. Engen, and B. Sæther. *Stochastic Population Dynamics in Ecology and Conservation.* Oxford University Press, Oxford, UK, first edition, 2003.

[69] G. Li and H. Qian. Sensitivity and specificity amplification in signal transduction. *Cell Biochemistry and Biophysics*, 39:45–59, 2003.

[70] K. Lindenberg, V. Seshadri, K. Shuler, and G. Weiss. Lattice random walks for sets of random walkers. first passage times. *J. Stat. Phys.*, 23:11–25, 1980.

[71] T. Lu, J. Hasty, and P. G. Wolynes. Effective temperature in stochastic kinetics and gene networks. *Biophys. J.*, 91:84 – 94, 2006.

[72] T. Lu, T. Shen, C. Zong, J. Hasty, and P. G. Wolynes. Statistics of cellular signal transduction as a race to the nucleus by multiple random walkers in compartment/phosphorylation space. *Proc. Natl. Acad. Sci. USA*, 103:16752 – 16757, 2006.

[73] T. Lu, D. Volfson, L. Tsimring, and J. Hasty. Cellular growth and division in the gillespie algorithm. *IEE Proc. Sys. Biol.*, 1:121–127, 2004.

[74] P. Martin, A. Hudspeth, and F. Jülicher. Comparison of a hair bundle's spontaneous oscillations with its response to mechanical stimulation reveals the underlying active process. *Proc. Natl. Acad. Sci. USA*, 98:14380–14385, 2001.

[75] D. Mattis and M. Glasser. The uses of quantum field theory in diffusion-limited reactions. *Phys. Rev. E.*, 70:979–1001, 1998.

[76] H. McAdams and A. P. Arkin. It's a noisy business! genetic regulations at nonomolar scale. *Trends. Genets.*, 15(2):65–69, 1999.

[77] D. McQuarrie. *Stochastic Approach to Chemical Kinetics.* Methuen & Co Ltd, London, first edition, 1967.

[78] R. Metzler and J. Klafter. The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *J. Phys. A.*, 37:R161–R208, 2004.

[79] X. Michalet, F. F. Pinaud, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, A. M. Wu, S. S. Gambhir, and S. Weiss. Quantum dots for live cells, in vivo imaging, and diagnostics. *Science*, 307:538–544, 2005.

[80] J. Monod, J. Wyman, and J. P. Changeux. *J. Mol. Biol.*, 12:88–118, 1965.

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

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

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

[84] J. Paulsson and M. Ehrenberg. Noise in a minimal regulatory network: plasmid copy number control. *Q. Rev. Biophys.*, 34:1–59, 2001.

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

[86] M. Ptashne. *A Genetic Switch: Phage Lambda and Higher Organisms.* Blackwell Publishers, Massachusetts, USA, second edition, 1992.

[87] A. Rao, C. Luo, and P.G. Hogan. Transcription factors of the nfat family: Regulation and function. *Annu. Rev. Immunol.*, 15:707–747, 1997.

[88] C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasisteady-state assumption: Application to the gillespie algorithm. *J. Chem. Phys.*, 118:4999 – 5010, 2003.

[89] C. V. Rao, D. M. Wolf, and A. P. Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420:231 –237, 2002.

[90] J. M. Raser and E. K. O'shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304:1183–, 2004.

[91] J. M. Raser and E. K. O'shea. Noise in gene expression: Origins, consequences, and control. *Science*, 309:2010–2013, 2005.

[92] H. Risken. *The Fokker-Planck equation: Methods of solutions and applications.* Springer, second edition, 1996.

[93] P. P. Roux and J. Blenis. Erk and p38 mapk-activated protein kinases: a family of protein kinases with diverse biological functions. *Microbiol. Mol. Biol. Rev.*, 68:320–344, 2004.

[94] C. Salazar and T. Höfer. Allosteric regulation of the transcription factor nfat1 by multiple phosphorylation sites: A mathematical analysis. *J. Mol. Biol.*, 327:31–45, 2003.

[95] J. Sambrook and D. W. Russell. *Molecular cloning: A Laboratory Manual(3-Volume Set).* Cold Spring Harbor Laboratory Press, third edition, 2001.

[96] M. Samoilov, S. Plyasunov, and A. P. Arkin. Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. *Proc Natl Acad Sci USA*, 102:2310–2315, 2004.

[97] H. J. Schaeffer and M. J. Weber. Mitogen-activated protein kinases: Specific messages from ubiquitous messengers. *Mol. Cell. Biol.*, 19:2435–2444, 1999.

[98] M. Schena, D. Shalon, R. W. Davis, and P. O. Brown. Quantitative monitoring of gene expression patterns with a complementary dna microarray. *Science*, 270:467–470, 1995.

[99] T. Shen and P. G. Wolynes. Stability and dynamics of crystals and glasses of motorized particles. *Proc. Natl. Acad. Sci. USA*, 101:8547–8550, 2004.

[100] T. Shen, C. Zong, D. Hamelberg, J. A. McCammon, and P. G. Wolynes. The folding energy landscape and phosphorylation: modeling the conformational switch of the nfat regulatory domain. *FASEB*, 19:1389–1395, 2005.

[101] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.*, 31:64–68, 2002.

[102] T. Shibata and K. Fujimoto. Noisy signal amplification in ultrasensitive signal transduction. *Proc. Natl. Acad. Sci. USA*, 102:331–336, 2005.

[103] P. Simpson. Notch signalling in developments: on equivalence groups and asymmetric developmental potential. *Curr. Opin. Genet. Dev.*, 7:537–542, 1997.

[104] W. K. Smits, O. P. Kuipers, and J. W. Veening. Phenotypic variation in bacteria: the role of feedback regulation. *Nat Rev Microbiol*, 4:259–271, 2006.

[105] D. J. Stephens and V. J. Allan. Light microscopy techniques for live cell imaging. *Science*, 300:82–86, 2003.

[106] P. W. Sternberg and M. A. Felix. Evolution of cell lineage. *Curr. Opin. Genet. Dev.*, 7:543550, 1997.

[107] S. H. Strogatz. *Nonlinear dynamics and chaos: With application to physics, biology, chemistry, and engineering.* Perseus Books Publishing, first edition, 1994.

[108] R. N. Stuart and E. W. Branscomb. Quantitative theory of *in vivo lac* regulation: Significance of repressor packaging. *J. Theor. Biol.*, 31:313–329, 1971.

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

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

[111] M. Thattai and A. van Oudennarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167:523–530, 2004.

[112] R. Tsien. The green fluorescent protein. *Annu Rev Biochem*, 67:509–544, 1998.

[113] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry.* Elsevier Ltd., Oxford, fourth edition, 2003.

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

[115] J. Wang and P. G. Wolynes. Intermittency of single molecule reaction dynamics in fluctuating environments. *Phys. Rev. Lett.*, 74:4317–4320, 1995.

[116] F. C. Wardle and J. C. Smith. Refinement of gene expression patterns in the early xenopus embryo. *Development*, 131:46874696, 2004.

[117] L. Weinberger, J. C. Burnett, J. E. Toettcher, A. P. Arkin, and D. V. Schaffer. Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuations drive phenotypic diversity. *Cell*, 122:169–182, 2005.

[118] E. Werner. Meeting report: The future and limits of systems biology. *Sci. STKE*, 278:pe16, 2005.

[119] G. M. Whitesides. The origins and the future of microfluidics. *Nature*, 442:368–373, 2006.

[120] N. Wingreen and D. Botstein. Back to the future: education for systems-level biologists. *Nat Rev Mol Cell Biol*, 7:829–932, 2006.

[121] D. M. Wolf, V. V. Vazirani, and A. P. Arkin. Diversity in times of adversity: probabilistic strategies in microbial survival games. *J. Theor. Biol.*, 234:227–253, 2005.

[122] O. Yarchuk, N. Jacques, J. Guillerez, and M. Dreyfus. Interdependence of translation, transcription and mrna degradation in the lacz gene. *J. Mol. Biol.*, 226:581–596, 1992.

[123] S. B. Yuste and K. Lindenberg. Order statistics for first passage times in one-dimensional diffusion processes. *J. Stat. Phys.*, 85:501–512, 1996.

[124] Y. B. Zeldovich, A. A. Ruzmaikin, and D. D. Sokoloff. *The Almighty Chance*. World Scientific, Singapore, 1990.