Systems of Chemical Reactions in Biology: Dynamics, Stochasticity, Spatial Effects and Model Reduction

by

Carlos Alberto Gómez Uribe

Submitted to the Harvard-MIT Division of Health Sciences and Technology
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biomedical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

Author .....
Harvard-MIT Division of Health Sciences and Technology
May 16, 2008

Certified by .....
George C. Verghese, Ph.D.
Professor
Thesis Supervisor

Certified by .....
Leonid Mirny, Ph.D.
Associate Professor
Thesis Supervisor

Accepted by .....
Martha L. Gray, Ph.D.
Edward Hood Taplin Professor of Medical and Electrical Engineering
Director, Harvard-MIT Division of Health Sciences and Technology

Systems of Chemical Reactions in Biology: Dynamics, Stochasticity, Spatial Effects and Model Reduction

by

Carlos Alberto Gómez Uribe

Submitted to the Harvard-MIT Division of Health Sciences and Technology
on May 16, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biomedical Engineering

Abstract

Cells are continuously sensing and processing information from their environments and responding to it in sensible ways. The communication networks on which such information is handled often consist of systems of chemical reactions, such as signaling pathways or metabolic networks. This thesis studies the dynamics of systems of chemical reactions in the context of biological cells.

The first part of this thesis analyzes the osmo-regulation network in yeast, responsible for the regulation of internal osmolarity. We measure the system's step response in single cells, and find that the steady state is independent of the input, a property termed perfect adaptation that relies on integral feedback control. We then consider the signaling cycle, a pattern of chemical reactions that is often present in signaling pathways, in which a protein can be either active (e.g., phosphorylated) or inactive (e.g., unphosphorylated). We identify new regimes of static and dynamic operation, and find that these cycles can be tuned to transmit or digitize time-varying signals, while filtering input noise.

The second part of this thesis considers systems of chemical reactions where stochastic effects are relevant, and simplifies the standard models. We develop an approximate model for the time-evolution of the average concentrations and their variances and covariances in systems with and without spatial gradients. We also describe a framework to identify and derive approximate models for variables that evolve at different time scales in systems without spatial gradients. These tools can help study the impact of stochastic and spatial effects on system behavior.

Thesis Supervisor: George C. Verghese, Ph.D.
Title: Professor

Thesis Supervisor: Leonid Mirny, Ph.D.
Title: Associate Professor

Acknowledgments

This thesis has benefited greatly from my interaction with several people to whom I am grateful. Leonid Mirny, my thesis coadvisor, welcomed me into his biological research group knowing my background in biology was minimal. He was instrumental in getting me started in biological research in the field, and encouraged me to pursue interesting collaborations. Alexander van Oudenaarden similarly welcomed me into his experimental biology lab, and partnered me with Jerome Mettetal and Dale Muzzey, with whom I share similar research interests and with whom it was a pleasure to work. Dale taught me almost all that I know about experimental biology, and has become a great friend.

I am most grateful to George Verghese, who has been a remarkable mentor and a friend since I met him as an undergraduate student. He truly developed me as a researcher, and has provided me with very valuable guidance, often going out of his way to do it, and with a wisdom that extends beyond academics. I am also grateful to John Wyatt for teaching me about stochastic processes, and for several insightful discussions and feedback on the more mathematical topics in this thesis. I also benefited and greatly enjoyed my interaction with Rami Tazfiri, whose ideas for time-scale separation found their way into much of my work.

I want to thank the many wonderful teachers I have had, both here in Boston and elsewhere. John Bush's passion for science and *joie de vivre* have been very contagious. Venturing into the semiconductor industry under Geoff O'Donoghue's guidance has been exciting and rewarding. Darío Lerer and Raúl Fariñas, my high school teachers of mathematics and physics, were instrumental in my coming to MIT in the first place, and I owe them much.

I have been fortunate to have shared the last four delightful years with Aiste Narkeviciute, whom I dearly love. I am particularly grateful to her; to my parents, Estela Uribe Iniesta and Leopoldo Gómez Díaz; and to my brother Ricardo for their continuous love and support. I dedicate this thesis to the four of them, and to the baby that will be hiding in Aiste's belly for a few more months.

Contents

Chapter 1

Introduction

One of the main goals of quantitative or systems biology is to understand the cell — and ultimately, the whole organism — as a system whose inputs and outputs are well identified, and whose time-dependent behavior may be predicted from its inputs (see Figure 1-1). Such an understanding of the cell will enable one to predict the consequences of controlled or uncontrolled changes in the system structure (e.g., through mutations or acquired diseases) or inputs (e.g., through drugs or hormones). The resulting knowledge could be widely applied in a variety of fields, particularly to enhance human health through such endeavors as improved drug design, or to engineer biological systems that aid in the production of a wide range of useful compounds, including biofuels. However, the biological systems that regulate cellular processes are quite complex, typically comprising of multiple chemical species, such as proteins, that interact across many time scales and involve multiple levels of regulation through positive and negative feedbacks. Another complicating feature is the non-linearity of the resulting dynamics, which often also involve stochastic and spatial effects. Engineering disciplines have well-developed theories and tools to deal with similarly complex systems, and have the potential to significantly advance our understanding of biological systems.

Cells are the fundamental modules of life. They continuously sense information from their environment, process it, and respond in ways that presumably maximize their chances of reproducing, thereby continuing life. On the other hand, when ex-

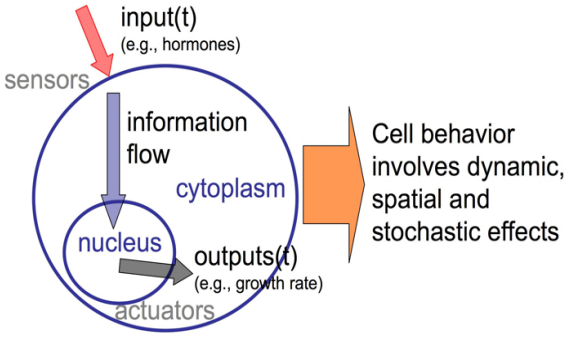

Figure 1-1: The cell as a dynamical system.

amined more closely a cell looks like a fluid-filled bag, perhaps with several internal compartments, where a mix of compounds diffuse happily with no apparent purpose or direction, occasionally reacting with one another. So how does the system-level cellular behavior emerge from the mix of compounds found within a cell? Finding a reasonable answer to this question is the main motivation behind the work in this thesis, which focuses on the development of modeling tools and frameworks for biochemical systems, and on the study of a particular biological system.

This thesis reports work on two main initiatives, both which often include ideas inspired by engineering systems theory: (1) the study of a specific biological system, the yeast osmo-regulation network, through a joint theoretical and experimental approach, and (2) the development and application of easy-to-use modeling tools and frameworks for biological systems.

The Yeast Osmo-sensing Network

For many years, most research of biological systems was either exclusively experimental or exclusively theoretical. Recently, however, several groundbreaking studies have intertwined theory and experiments, exploiting this interplay to broaden and deepen our understanding of biological phenomena. Such interdisciplinary studies are still relatively rare because of a considerable knowledge and communication gap in the theoretical/experimental spectrum. Some of the most exciting opportunities within systems biology focus on specific biological systems that require interdisciplinary re-

search teams and a strong interplay between theory and experimentation. I have been a part of one such study for the last two years.

The system of study is the osmo-regulation network in Saccharomyces cerevisiae, the organism more commonly known as budding yeast. This network adjusts the cell internal osmolarity in response to environmental changes to maintain an appropriate osmotic pressure across the cell membrane. Such homeostatic behavior is characteristic of biological systems, and is one of the properties of living organisms.

Together with a biologist and a physicist, I developed an assay to measure the dynamics of pathway activity in live single cells with a relatively high temporal resolution. We have shown that the system perfectly adapts to step inputs (i.e., has a steady-state that is independent of the inputs magnitude). Guided by ideas from control theory, we performed further experimentation to locate the part of the system providing a particular type of feedback (integral-feedback) that guarantees perfect adaptation. In the process, we have also shown that the system's dynamics are largely independent of gene expression. Additionally, we show that the enzymatic activity of a protein in the network is essential for the observed perfect adaptation, suggesting the existence of unknown protein-protein interactions in the system. This work is described in detail in Chapter 2.

Modeling Biochemical Systems

Biological networks often contain thousands of nodes and interactions among them. Even the simplest organisms have thousands of genes, and the protein product of each may exist in a range of different states (e.g., phosphorylated once, twice, or not at all). To better understand network function, recurrent motifs are identified and characterized in the hope that they behave similarly within different networks. In Chapter 3, we analyze the static and dynamic behaviors of the nearly ubiquitous phosphorylation cycle motif. The analysis there reveals the existence of four cycle regimes, two of which were previously known. It also shows that the cycle can be thought of as a first-order low-pass filter with gain and cut-off frequency that cells

can tune through changes in gene expression and/or in rate constants. This flexibility may allow cells to restrict the range of frequencies that stimulate a specific signaling pathway.

The mechanisms by which cells sense, transmit, process and react to external and internal information invariably involve systems of chemical reactions, such as signaling cascades, metabolic pathways, and gene networks. For many cellular processes, these seemingly different networks are interconnected with each other despite often operating on different time scales. As such, it is a remarkably difficult yet important problem to build a comprehensive, mathematical model that accurately describes a network's dynamics, even if the network is well characterized from a biological standpoint and has known parameter values.

To best reflect cellular conditions, such a comprehensive model would necessarily incorporate stochastic effects (e.g., due to low-molecule numbers), spatial effects (e.g., due to diffusion and compartmentalization), and nonlinear dynamics with multiple time scales. Currently, models that include these effects are available, but they are too complex for all but the simplest of systems. A fundamental goal of this thesis has been to apply model-reduction techniques (e.g., separation of time scales and moment truncation) to simplify available modeling frameworks, allowing the effects described above to be either simply modeled or neglected altogether.

To this end, this thesis develops a simple extension to the deterministic, spatially homogeneous model of mass-action kinetics to account for stochastic effects. The resulting model, referred to as mass fluctuation kinetics (MFK) and discussed in Chapter 5, is obtained through moment truncation from the fully probabilistic model of stochastic chemical kinetics introduced in Chapter 4. The MFK model tracks the expected concentrations of the species in the system, as well as their variances and covariances. Unlike other standard modeling approaches, the MFK model accurately describes the critical effect that fluctuations can have on the mean concentrations of a systems molecules. Additionally, it provides first and second-order statistics of the dynamic behavior of a biochemical network, while being significantly easier to build and analyze than the full model of stochastic chemical kinetics.

This thesis also develops in Chapter 6 a simple separation-of-time-scales method for stochastic chemical kinetics that identifies the slow variables in a system, facilitating creation of a simpler model in which the system state depends only on dynamics of the slow variables. Combining this approach with an MFK model yields an even simpler description of a stochastic biochemical network than either MFK or separation-of-time-scales alone could provide.

Incorporating both spatial and stochastic effects into a biochemical network model is an enormous but important challenge. Perhaps the most common current approach is to use the spatial chemical master equation (SCME), where a system's volume is partitioned into many sub-volumes, and diffusion is modeled as a first-order reaction between adjacent sub-volumes. The SCME describes the time evolution of the joint probability of the molecule numbers of all species in all sub-volumes. The resulting model is so complex that even statistical simulation of very simple systems is prohibitively difficult. This thesis describes an MFK model obtained from the SCME that directly tracks the averages, variances and covariances of all the concentrations in the system. This spatial MFK model is significantly simpler to construct and use than the SCME. The SCME and the related MFK model is described in detail in Chapter 7. Lastly, some concluding remarks and directions of future work are discussed in Chapter 8.

Chapter 2

The Osmo-regulation Network of *Saccharomyces cerevisiae*

Summary

This chapter describes a study of the osmo-regulation network in *Saccharomyces cerevisiae* (or budding yeast), a biological network comprised of signaling pathways, and metabolic and gene networks.[1] The osmo-regulation network, highly conserved across organisms, is responsible for regulating the internal osmolyte concentration in response to changes in environmental osmolarity. Negative feedback can serve many different cellular functions, including the reduction of noise in transcriptional networks and the creation of oscillations in circadian systems. However, only one special type of negative feedback (integral feedback) can ensure robust performance in homeostatic systems, yielding a system behavior termed perfect adaptation, where steady-state output is independent of steady-state input. Here we measure single-cell dynamics in the Saccharomyces cerevisiae hyperosmotic shock network, which regulates membrane turgor pressure. Importantly, we find that the nuclear enrichment of the MAP kinase Hog1 perfectly adapts. We use small-molecule inhibitors and dynamic measurements of cell volume, Hog1 nuclear enrichment, and glycerol accumulation to assess the network location of the mechanism responsible for perfect adaptation, and we build a

[1]The material in this chapter has been submitted for publication to *Cell*.

simple model of the system. Notably, Hog1 kinase activity, but not gene expression, is required for Hog1 perfect adaptation, suggesting that this networks homeostatic function may critically depend on protein-protein interactions.

2.1 Introduction

Cells are typically subjected to time-varying environments, yet must finely regulate the levels of many of their key physiological variables, which allows them to thrive in a wide range of experimental conditions. Often, this homeostatic behavior is accomplished by complex systems of chemical reactions known as biochemical networks. They typically consist of signaling pathways, where proteins interact and modify one another to process and transmit signals, metabolic networks that change the concentrations of key cellular metabolites, and gene networks that adjust protein levels. Measuring the system's dynamics is often a challenging task, particularly when measurements of signaling pathways or metabolic networks are needed.

Feedback is a ubiquitous regulatory feature of biological systems, characterized by the system output either directly or indirectly contributing also to the input. A system exhibits positive feedback when its output reinforces an input; for example, in the regulation of HIV-1 latency, the Tat protein is both an output of the system and a potent activator of many viral genes, including Tat itself ([151]). In systems with negative feedback, by contrast, the output opposes the input, as in the eukaryotic cell-cycle oscillatory system, where cyclin-B accumulation leads to formation of the MPF, which subsequently activates the APC, ultimately resulting in degradation of cyclin B ([115]).

Quantitative models are increasingly being used to study the function and dynamic properties of complicated, feedback-laden biological systems. These models can be broadly classified by the extent to which they represent specific molecular details of the network. At one extreme are the exhaustive models that dynamically track quantities of virtually all biomolecules in a system, using differential equations based either on known or assumed reaction stoichiometries and rates. This modeling

approach has produced exceptional results in the modeling of EGF receptor regulation ([127]) and crosstalk in yeast signal transduction ([125]), though in general it can suffer from overfitting or the omission of important reactions not yet biologically identified.

At the other end of the modeling spectrum is the minimalist approach, which aims to fit and predict a systems input-output dynamics with only a few key parameters, each potentially the distillation of a large group of reactions. Though such models frequently lack reaction- and network-specific details, they excel at providing intuitive and general insights into the dynamic properties of recurrent system architectures. For instance, two elegant studies of bacterial chemotaxis —a system said to perfectly adapt because abrupt changes in the amount of ligand only transiently affect the tumbling frequency, whereas steady-state tumbling is notably independent of the ligand concentration— highlighted a general feature of all perfectly adapting systems ([9, 156]).

Specifically, it was shown that a negative feedback loop implementing integral feedback is both necessary and sufficient for robust perfect adaptation in any biological system. Mathematically, a dynamic variable (e.g., x) whose rate of change is independent of the variable itself (i.e., the dx/dt equation contains no terms involving x) is an integrator, and integral feedback describes a negative-feedback loop that contains at least one integrator. Biologically, a biomolecule acts as an integrator if its rate equation is not a function of the biomolecule concentration itself; such a situation arises if, say, the synthesis and degradation reactions are saturated (see below). By providing specific mechanistic restrictions that apply to any perfectly adapting system, these two studies underscored the function and significance of perfect adaptation in homeostatic regulation and demonstrated the power of the minimalist modeling approach.

2.2 The Osmo-regulation Network of *Saccharomyces cerevisiae*

In order to activate the proper adaptive mechanism at the proper time, cells have evolved extensive signal-processing machineries, including signal-transduction cascades that initiate at the membrane and culminate in the nucleus. Mitogen-activated protein kinase (MAPK) cascades are among the most closely studied and highly conserved signal-transduction pathways. Their defining feature is a core module of three kinases —the MAPKKK, the MAPKK, and the MAPK— that sequentially phosphorylate each other. Activated by this phosphorylation, the MAPK itself often mediates a broad transcriptional response to the extracellular signal. Most eukaryotes contain several MAP kinases and several corresponding upstream pathways, each of which is tuned to respond to a specific stimulus. Remarkably, even though some pathways share common components, they are quite robust to cross-talk, precisely regulating fundamental processes including homeostasis, immunity, synaptic plasticity, and development.

S. cerevisiae is the yeast that is typically used for baking bread and brewing beer or wine. It is a unicellular eukaryotic (i.e., with a nucleus) organism that divides by budding, so it is typically called budding yeast. Budding yeast is an important model organism in cell biology, because it grows rapidly, it can be cultured easily, and can be genetically manipulated with relative ease. Budding yeast cells are separated from their environment by their cell membranes, which are semi-permeable, i.e., they let water and other molecules pass freely through them while stopping others. Those molecules stopped by the membrane are called osmolytes; their concentration in a solution is the solution's osmolarity.

When the osmolarities on the two sides of a semi-permeable membrane are different, and in the absence of other forces, osmosis results in water flow from the low osmolarity to the high osmolarity side of the membrane. The internal osmolarity of cells is typically higher than that of their environments. In the absence of other processes, osmosis would then drive water flow into the cell, resulting potentially in

cell rupture. However, as the internal volume of the cell increases, the cellular membrane and wall stretch and exert an opposing pressure on the incoming fluid resulting in a steady-state volume. Therefore, fluid within cells is typically pushing on the membrane and cell wall. This so-called *turgor pressure* is what drives bud growth in budding yeast, allowing them to divide.

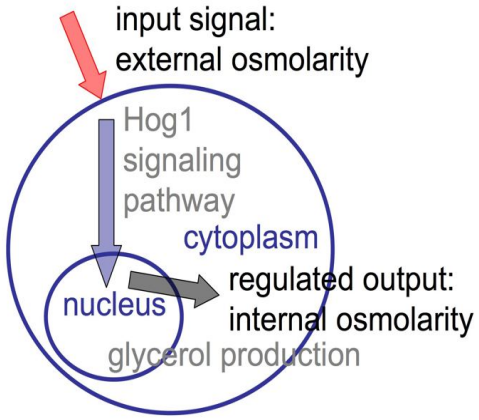

Figure 2-1: Budding yeast regulate their internal osmolarity in response to changes in external osmolarity by adjusting glycerol levels.

Maintaining an adequate difference between internal and external osmolarity is then of utmost importance for budding yeast. If the internal osmolarity is too high the cell membrane and wall can rupture, while if the internal osmolarity is too low then water flows out of the cell, leaving the cell unable to divide or, in the extreme case, completely dried up. Importantly, budding yeast are immobile creatures that live in environments with time-varying osmolarities. So how do they regulate their internal osmolarities in response to environmental osmolarity changes? It turns out they do so via a complex system of chemical reactions, the osmo-regulation network, that adjusts the internal concentration of a specific osmol, glycerol (see Figure 2-1). When the external osmolarity drops, this network lets glycerol leak out of the cell. When the external osmolarity increases, the osmo-regulation network dynamics lead to glycerol accumulation. We focus here on the system response to hyper-osmotic shocks, i.e., increases in the external osmolarity.

Minimalist modeling tactics have been successfully applied to the osmosensing network in the budding yeast *S. cerevisiae* ([86]). The core of this network is a

highly conserved mitogen-activated protein kinase (MAPK) cascade, one of several such cascades in yeast that regulate processes ranging from mating to invasive growth while being remarkably robust to cross-talk despite their many shared components [77, 76]. Yeast cells maintain an intracellular osmolarity in excess of the extracellular osmolarity, thereby creating positive turgor pressure across the cell wall and membrane that is required for many processes including budding itself. Sudden drops in turgor pressure, potentially caused by an upward spike in the external osmolyte concentration, are detected by membrane proteins such as Sln1 ([120]), which rapidly initiates a sequence of phosphorylation-dephosphorylation reactions that propagate down the Hog1 signaling pathway. This pathway is named after its key protein, the Hyper-Osmotic Glycerol protein 1, or Hog1. The last three proteins in the pathway constitute a MAPK cascade. Protein Hog1, the MAPK, is the human homologue of p38, which has been involved in cancer.

Activation of the Hog1 pathway results in the double phosphorylation of Hog1 (Figure 2-2a). Upon dual phosphorylation, the normally cytoplasmic and inactive Hog1 becomes activated and translocates to the nucleus ([45]), where it plays direct and indirect roles in a broad transcriptional response ([107]). *GPD1* and *GPP2* , which encode key glycerol-producing factors, are among the activated genes, and they facilitate osmoadaptation through the increase of intracellular osmolarity ([1, 121, 77]). In fact, glycerol accumulation, a process in which Gpd1, Gpp2 and other enzymes catalyze reactions that convert one glucose molecule into two glycerol molecues, has been shown to comprise 95% of the internal osmolarity recovery ([119]). The subsequent restoration of turgor pressure leads to nuclear export of Hog1, which is dephosphorylated by several nuclear and cytoplasmic phosphatases ([154, 79, 150]).

Non-transcriptional feedback mechanisms are also important for the hyperosmotic-shock response, and recently Hog1 has been implicated in these processes. For instance, Hog1 has been shown to phosphorylate the metal-ion antiporter Nha1, modulating its ion-efflux properties to accelerate osmoadaptation ([116]). Hog1 may also affect permeability of the Fps1 channel, which allows passive export of glycerol from the cell when at optimal turgor pressure ([96]). Upon hyperosmotic shock, however,

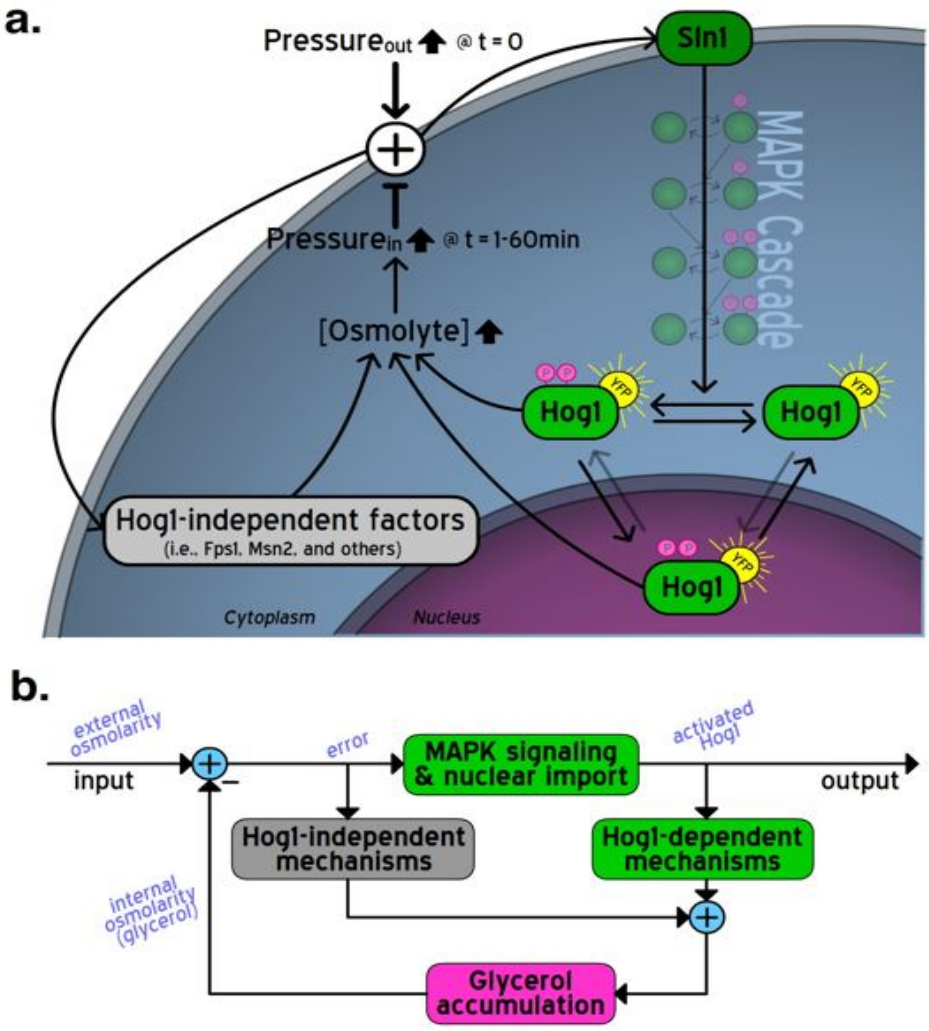

Figure 2-2: **a.** Diagram of the budding yeast osmo-regulation network, and **b.** corresponding system/engineering diagram.

the Fps1 channels close rapidly, leading to retention of intracellular glycerol ([137]). Under arsenite stress, Hog1 has been shown to phosphorylate Fps1 ([141]), thereby decreasing its permeability to arsenite, though it has not yet been shown whether Hog1 phosphorylates Fps1 during hyperosmotic shock. The Hog1 pathway has also been shown to upregulate glycerol synthesis via phosphorylation and potent activation of 6-phosphofructo-2-kinase (*PFK2*), a key glycolytic enzyme that increases production of glycerol precursors ([33]). Though direct phosphorylation of *PFK2* by Hog1 has not been investigated, *PFK2* has many consensus MAPK phosphorylation sites that become phosphorylated upon hyperosmotic shock ([33]).

Built upon the extensive genetic and biochemical characterization of hyperosmotic-shock regulation, the exhaustive dynamic model of the Hog1 system was successful at recapitulating many known system features with precise molecular detail ([86]). Furthermore, it accurately predicted previously untested behaviors, such as the dynamics of Hog1 activation, transcriptional activation, and glycerol accumulation in response to sequential osmotic shocks. Biochemical characterization of most systems is rarely so rich to be deemed exhaustive, nor so lacking to consider a system as a black box. Thus, models combining elements from both approaches can also be quite useful, such as a recent study that started with the exhaustive osmoadaptation model and abstracted several elements to yield a reduced representation ([52]). Here we take the reverse strategy, starting instead with the minimalist model (Figure 2-2b) and then using biological measurements and engineering principles to better understand systems-level dynamics and their relation with network topology.

I was involved in two studies of this system. In one of them, described in detail in the PhD thesis of Jerome Mettetal ([101]) and in [100], we viewed the osmo-sensing network as a "black box," and used sinusoidal inputs to build a minimal model of the system. In the other study, the one described next, we hyperosmotically stress yeast cells and monitor the dynamics of cell volume, Hog1 nuclear enrichment, and glycerol accumulation. In response to step inputs of osmolyte, we observe perfect adaptation of Hog1 nuclear enrichment, and this adaptation occurs with very low cell-to-cell variability. Perfect adaptation in this feedback system requires an integral-feedback mechanism, and we use molecular and modeling techniques to assess its location in the network. Specifically, we find that the perfect adaptation requires Hog1 kinase activity but does not require new protein production, suggesting that Hog1 may implement integral-feedback via a yet-unknown role in protein-protein interactions that increase the internal osmolyte concentration. Measurements of glycerol accumulation support this crucial role for Hog1 kinase activity and, as with the observation of perfect adaptation, impose further restrictions on the types of models that can describe the system. We present a simple quantitative model based on our various biological findings that captures the key features of our data.

The Network Diagram

Figure 2-2a summarizes the biological structure of the osmo-regulation network, and makes evident that the interactions amongst the different processes in the network result in several feedback loops that together determine the overall system behavior. Engineering disciplines, and particularly control theory, have developed tools and results that are specially well-suited for understanding such complex dynamical systems with multiple feedback loops (e.g., see [106, 81]). We later draw on some of these tools and results in developing a systems-level understanding of the osmo-regulation network, whose corresponding network diagram is shown in 2-2b. There the input to the network is the external osmolarity $u(t)$; the difference, or error, between $u(t)$ and the cell's internal osmolarity $g(t)$ defines the turgor pressure, which feeds into the cellular signaling subsystem that corresponds to the signaling pathway starting with Sln1 and resulting in phosphorylated Hog1. Phosphorylated Hog1 has an increased kinase activity, which is the input to Hog1-dependent processes that result in internal osmolyte (glycerol) accumulation (e.g., through Fps1 phosphorylation and gene expression). Other processes that are Hog1 independent also respond to the error and lead to glycerol accumulation.

2.3 Measuring the Dynamics of the Osmo-Regulation Network

The Yeast Strain

In order to study the osmo-regulation network, we developed an assay to measure pathway activity using Hog1's spatial localization as a proxy for its phosphorylation. We measure the nuclear accumulation of Hog1 and cellular volume in single cells over time in response to controlled changes in external osmolarity. We fused a yellow-fluorescent protein (YFP) to the C-terminus of endogenous Hog1 (Hog1-YFP) in haploid cells from which a protein called Sho1 was deleted. Sho1 deletion disables one of the two parallel branches that activate Hog1. We focus on the Sln1 branch here

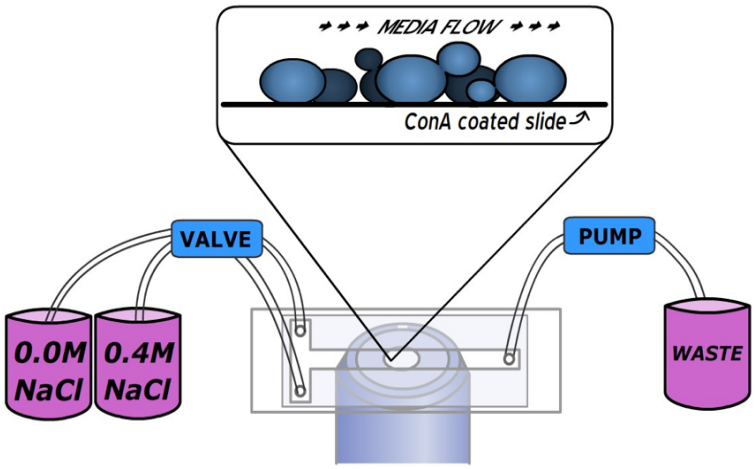

Figure 2-3: Apparatus with a chamber where yeast cells are stuck on its surface, while media from one of two inputs flows over them.

for several reasons: (1) Sln1 dominates the response at the shock levels we assayed (data not shown and old ref), and (2) whereas the stimuli affecting Sho1 are not well understood, turgor pressure is the well-characterized input to the Sln1 branch, thus aiding our modeling and data analysis. To identify the nucleus and thereby facilitate our computation of nuclear Hog1 enrichment, we fused a red fluorescent protein to Nrd1 (Nrd1-RFP), a strictly nuclear factor.

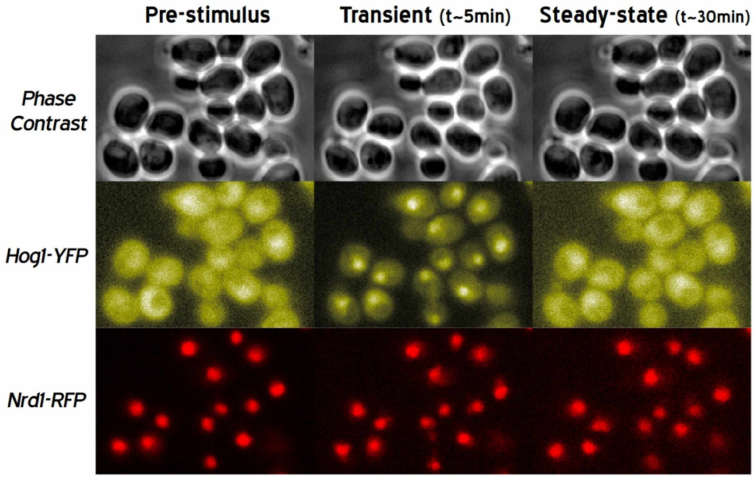

Figure 2-4: Phase-contrast, YFP and RFP images of the response to a step input sampled at three time points. Phase-contrast images allow us to find regions with and without cells, YFP images show the spatial distribution of Hog1 within each cell, and RFP images define nuclear regions within cells.

The Measuring Device

We developed a simple flow-chamber apparatus to permit the changing of media and acquisition of images simultaneously. Happily dividing cells (i.e., in log-growth phase) are loaded into the chamber (see Figure 2-3), where they adhere to a coverslip coated with concanavalin A. Media with or without excess osmolyte is then washed over the cells throughout each experiment, and in less than two seconds the media within the chamber can be swapped completely between the two types (data not shown). At each time point, we acquire phase-contrast, YFP, and RFP images. Figure 2-4 shows sample images taken at three different time points in response to a step input. These images show a transient accumulation of Hog1 in the nucleus shortly after a hyperosmotic shock, and ubiquitous localization shortly before and long after the shock.

Quantifying the System's Response

We developed custom image analysis algorithms to quantify the dynamic single-cell volume and Hog1 nuclear accumulation responses to changes in the extracellular osmolyte concentration, which we chose to be NaCl. Our algorithms segment the images at each time point to produce the image regions corresponding the each cell and to each cell's nucleus.

To quantify the Hog1 response of cell number i through its Hog1 nuclear accumulation, we first define the *raw* nuclear accumulation $r_i(t)$, defined as the ratio of the average nuclear YFP intensity to average whole-cell YFP intensity. That is, letting $\mathcal{N}_i(t)$ and $\mathcal{C}_i(t)$ respectively denote the pixels corresponding to the nuclear and whole cell regions of cell number i at time t, we have that

$$r_i(t) = \frac{\frac{1}{n_{n_i}(t)} \sum_{j \in \mathcal{N}_i(t)} y_j}{\frac{1}{n_{c_i}(t)} \sum_{j \in \mathcal{C}_i(t)} y_j} \ , \tag{2.1}$$

where y_j is the YFP intensity of pixel number j, and $n_{n_i}(t)$ and $n_{c_i}(t)$ are the number of pixels in $\mathcal{N}_i(t)$ and $\mathcal{C}_i(t)$, respectively. The raw nuclear enrichment $r_i(t)$ then

quantifies the relative concentration of Hog in the nucleus relative to that of the whole cell. Additionally, defining $r_i(t)$ as a ratio of YFP intensities automatically corrects for photo-bleaching, a process that results in decreasing YFP intensities in sequential images.

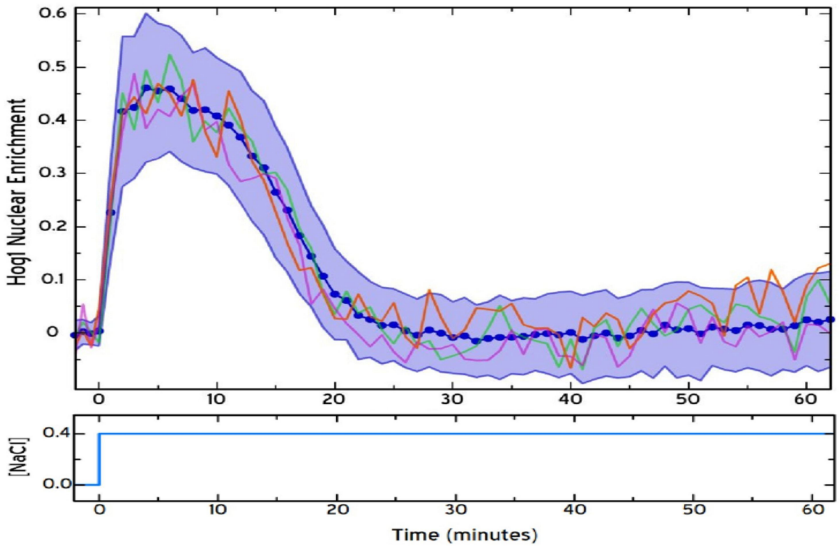

Figure 2-5: Step response of Hog1 nuclear enrichment $y(t)$. The jagged traces correspond to single cell traces, the blue line to the population average, and the shaded regions to the one-stadard-deviation bands about the population average.

We focus on what we call the Hog1 relative nuclear enrichment $h_i(t)$, or more simple just the Hog1 nuclear enrichment, which we define as

$$h_i(t) = \frac{r_i(t)}{r_i(t_o)} - 1 \ . \tag{2.2}$$

The nuclear enrichment $h_i(t)$ simply tracks relative changes of the initial ratio of nuclear to whole-cell YFP intensities.

Figure 2-5 shows the Hog1 nuclear enrichment in response to a step of 0.4M NaCl. The jagged traces there correspond to single-cell traces. The blue line is the population average, obtained as the sample mean of the nuclear enrichments of all cells in the sample. Similarly, the shaded region corresponds to the one-standard-deviation bands about the population average. Importantly, note that the steady-state nuclear enrichment is identical to its pre-stimulus level. We elaborate on this in the next

section.

Our algorithms also allow us to quantify relative volume changes from the collected images. Specifically, our analysis produces the number of pixels $n_{c_i}(t)$ that constitute each cell as a function of time; this number is proportional to the cell's area, so assuming that the cell is a sphere $n_{c_i}^{\frac{3}{2}}(t)$ is proportional to the cell's volume. Because we are interested in relative volume changes from baseline, we accordingly define the volume $e_i(t)$ of cell number i, referred to also as the error for reasons that will become clear in the next section, by

$$e_i(t) = \left(\frac{n_{c_i}(t)}{n_{c_i}(t_o)} \right)^{\frac{3}{2}} - 1 \ . \tag{2.3}$$

The error $e(t)$ thus tracks the relative volume changes.

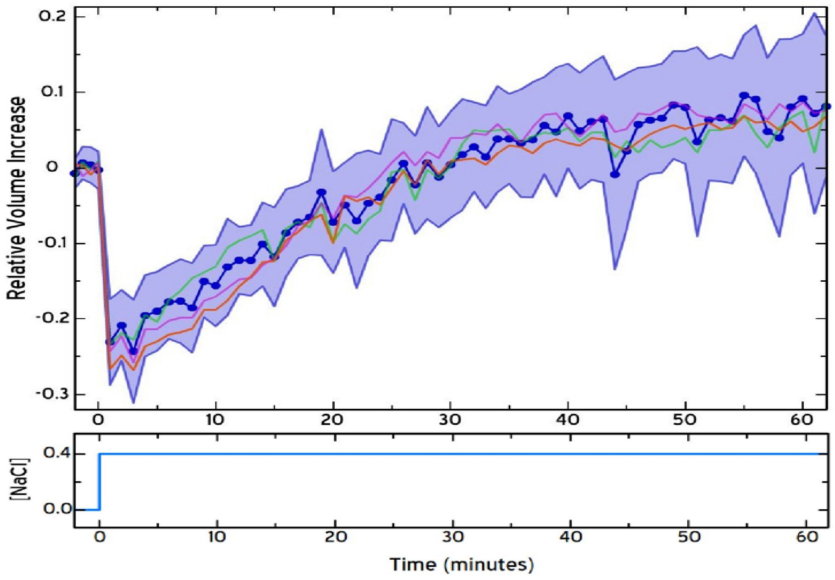

Figure 2-6: Step response of cell volume $e(t)$. The jagged traces correspond to single cell traces, the blue line to the population average, and the shaded regions to the one-stadard-deviation bands about the population average.

Figure 2-6 shows the cell volume $e(t)$ in response to a step of 0.4M NaCl. The jagged traces there correspond to single-cell traces. The blue line is the population average, obtained as the sample mean of the nuclear enrichments of all cells in the sample. Similarly, the shaded region corresponds to the one-standard-deviation bands about the population average. In response to the osmoshock, cell volume has a quick

decrease followed by a slower recovery. In contrast with the Hog1 nuclear enrichment, cell volume does not reach a steady-state, but instead continues increasing throughout the experiment. This indicates that at some point the turgor pressure, which decreases shortly after the osmoshock, has returned to a level that allows cells to grow and divide again.

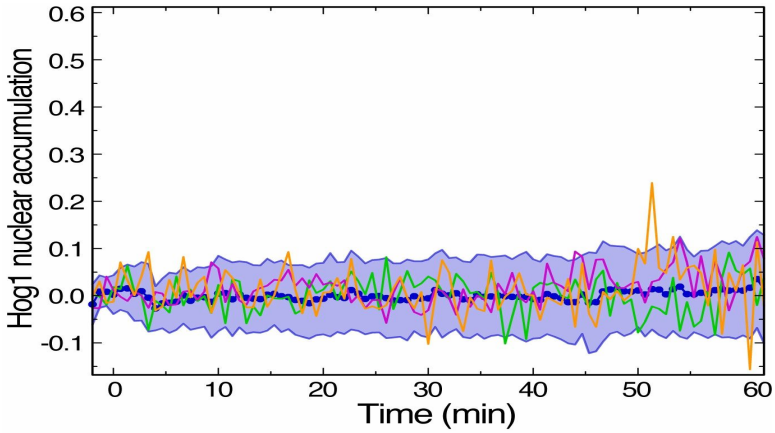

Figure 2-7: Single-cell traces and population statistics for Hog1 nuclear enrichment in unstimulated cells.

The previous figures also show that there is remarkably low cell-to-cell variability introduced by reactions in the MAPK cascade, since measurements both upstream (volume) and downstream (Hog1) were comparably noisy. In fact, fluctuations in unstimulated cells are of a similar magnitude (Figure 2-7), further suggesting that most of the cell-to-cell variability observed may come from the experimental setup itself. Taken together with studies investigating the noise generated from gene-expression-based systems, these data then suggest that the osmoadaptation signaling system generates output signals with very low noise, despite the fact that the system itself contains many proteins expressed at noisy levels.

2.4 Hog1 and its Perfect Adaptation

Quantitative measurement of Hog1 nuclear enrichment over a range of hyperosmotic shock strengths reveals that the kinetics of the response depend on shock-strength but steady-state nuclear enrichment does not (Figure 2-8). This behavior, having a steady

state that is independent of the input's steady state, is termed *perfect adaptation* in Biology.

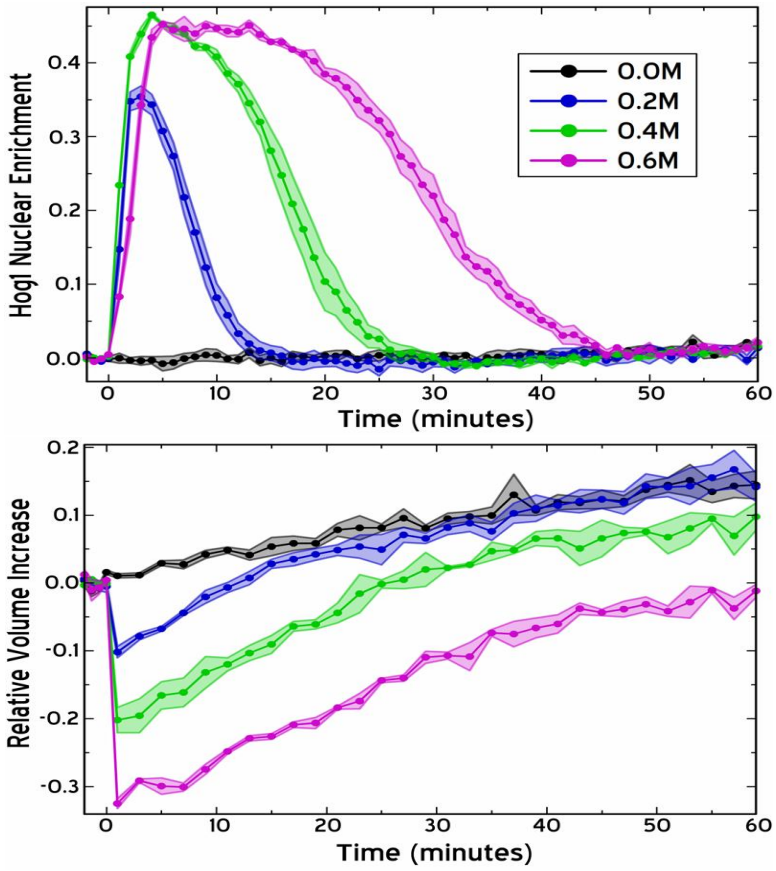

Figure 2-8: Hog1 nuclear enrichment (top) and cell volume (bottom) in response to step inputs of various magnitudes. The traces correspond to average behavior, obtained by averaging the population averages of at least four different experiments. The shaded regions around each trace correspond to day-to-day variability of the population averages; their width is twice the standard error of the population averages obtained from the different experiments.

A perfectly adapting system is particularly well suited for regulating homeostasis because it helps the cell respond to a deviation from a resting environment, but the response does not persist if the deviation itself defines a new resting environment. The best-characterized system that implements perfect adaptation is the bacterial chemotaxis network, in which the cell attempts to reach a maximum in the external chemoattractant concentration by reorienting less frequently as the level of chemoattractant rises (e.g., see [10, 9, 2, 156, 11]). Classic experiments measured the rate of

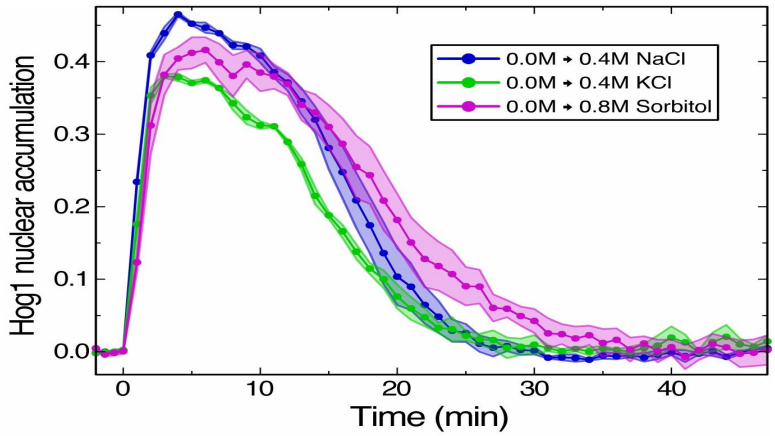

Figure 2-9: Hog1 nuclear enrichment in response to a step increase of external osmolarity, for a variety of osmols.

reorientation in response to an upward and persistent spike in chemoattractant and found that it perfectly adapts: despite a transient decrease in reorientation following the spike, the steady-state reorientation frequency is the same as its pre-spike level. Other biological systems that perfectly adapt include blood calcium levels in parturient cows ([34, 82]).

Figure 2-8 shows that within 5 minutes, Hog1 nuclear enrichment reaches a maximum, after which it slowly decreases. Crucially, the steady-state Hog1 nuclear enrichment of cells stimulated with 0.2M, 0.4M, and 0.6M NaCl is the same as that of cells not treated with NaCl. Furthermore, such behavior is not NaCl-specific, as it was observed with KCl and sorbitol treatment as well (Figure 2-9). The data on Figure 2-8 also demonstrate that the system is nonlinear; Hog1 nuclear enrichment saturates at about 0.45, and the dynamic response becomes wider with increasing input magnitudes.

Interestingly, in response to a 0.2M NaCl shock, nuclear enrichment returns to its steady-state within 15 minutes of the shock. This suggests that, at least at lower salts, the addition of new proteins through gene expression, which typically has slower dynamics, may play a minor role in the recovery. This is surprising in light of the current understanding in which enhanced expression of Gpd1 and Gpp2 is a key process of the system response.

Additionally, more frequent sampling of the Hog1 response shows that nuclear en-

richment starts within 10 seconds of the osmoshock (data not shown). This indicates that it takes less than 10 seconds for the phosphorylation signal to travel down the Hog1 signaling pathway.

The cell volume, shown in Figure 2-8, decreases within seconds of the osmoshock, and then slowly recovers. While Hog1 nuclear enrichement reaches a steady state (identical to its pre-stimulus level), cell volume continues increasing throughout the experiment, indicating that cells continue growing and dividing after a transient recovery to the osmoshock. This continued growth requires that the turgor pressure was restored during the transient response, and suggests turgor pressure also perfectly adapts to step inputs. Presumably, the main objective of the osmo-regulation network *is* to regulate turgor pressure so that it perfectly adapts, allowing the cell to continue growing and dividing despite step changes in external osmolarity; Hog1s perfect adaptation would then be a consequence of this regulation.

Interestingly, for stronger hyperosmotic shocks (e.g., [NaCl] = 0.6M), the Hog1 signal returned to its pre-stimulus level before the volume did. It is still possible, however, that pre-stimulus turgor pressure has indeed been restored at a smaller volume in these cells, since it has been shown that membrane invagination occurs in proportion to hyperosmotic shock strength. Because turgor pressure is a function of both membrane surface area and the osmotic difference across the membrane, this invagination would decrease the cell volume required to achieve the pre-stimulus turgor pressure.

2.4.1 Integral Feedback

Perfectly adaptation has been extensively studied in engineering and control. Control theoretic studies of perfect adaptation have shown that a system can perfectly adapt if and only if it contains an integral-feedback controller (which may only be made apparent after a change of coordinates, e.g., see [156]). Thus, the fact that we observe perfect adaptation in Hog1 nuclear accumulation requires the existence of integral-feedback control in the hyperosmotic shock response. We here show how integral feedback *can* give rise to perfect adaptation, and argue for necessity only later in the

chapter.

Consider the standard control theory system schematic in Figure 2-10. The system receives an external input $u(t)$, generates an observable output $y(t)$, and is comprised of two systems, called the plant and the controller. The output $y(t)$ of the plant *is* the input to the controller, which itself generates an output $g(t)$ that gets combined with the external input to generate the input to the plant $e(t) = y(t) - g(t)$, which we call the error.

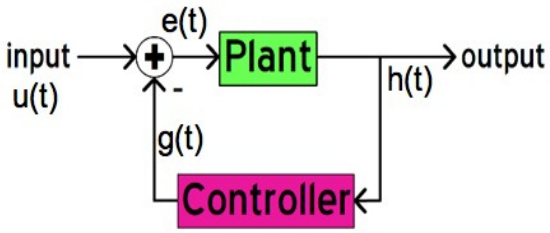

Figure 2-10: Simple control diagram network schematic.

Both the controller and the plant impact overall system behavior. For instance, suppose that the controller simply integrates its input over time to produce its output, i.e., that $\frac{dg}{dt} = ky(t)$, for some positive constant k. Further suppose that the overall system is stable, so that it settles to a steady-state in the presence of a sustained input $u(t) = u_o$. Then at steady-state, $\frac{dg}{dt}$ must be zero, which directly implies that $y(t)$ is zero at steady-state, so that the plant's output perfectly adapts. If the fact that the plant's steady state is zero implies that its steady-state input is zero, a not uncommon situation, then we also have that the error e is zero at steady state, so that $g = u_o$ at steady-state.

Could the perfect adaptation of the osmosensing network be understood in the context of the previous argument, taking the plant to be the system that takes in turgor pressure and outputs our measured Hog1, and the controller the system that takes measured Hog1 as its input and outputs internal osmolarity? Because $g = u_o$ at steady-state, then if we change the steady-state external osmolarity by Δu, the internal osmolarity g changes by the same amount $\Delta g = \Delta u$. However, since g is the controller's output we then have that $\Delta u = \int_{t_o}^{\infty} ky(\tau)d\tau$, namely that the

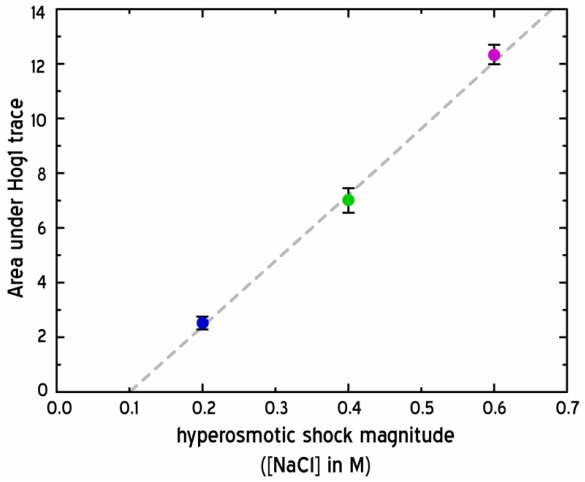

Figure 2-11: Area under the Hog1 response in Figure 2-8a versus the magnitude of the step input.

integral under the Hog1 response $y(t)$ is directly proportional to the change in external osmolarity.

Remarkably, despite the evident non-linearity of the Hog1 response (see Figure 2-8a) its integral does increase linearly with the magnitude of the input, as shown in Figure 2-11. However, the line does not cross the x-axis at zero, but at a positive shock. This is possibly due to the fact that Hog1 is not the single input into glycerol accumulation, so Hog1-independent processes potentially contribute part of the Δg, perhaps only up to a maximal amount after which they let Hog1 finish the adaptation. For now, we interpret the linearity of Figure 2-11 as an encouraging indicator that glycerol accumulation may behave like a simple integrator, responsible for the network's perfect adaptation.

Note that the previous argument did not require the plant to be a linear system, but only that the overall system reaches a steady state in the presence of a sustained input. The previous argument can be directly generalized to other types of controllers of the form

$$\frac{dg}{dt} = f(y(t)), \tag{2.4}$$

where $f(y(t))$ is a nicely behaved function that has the same sign as its argument h, but is otherwise arbitrary. The key observation here is that $f(y(t))$ is independent

of the output, so that $g(t)$ is simply the time-integral of $f(y(t))$. Such controllers are called *integral controllers.*

Biological Integrators

Saturated enzymatic reactions can result in biological Integrators. Consider a substrate (e.g., a protein) that can be in one of two chemical states: active (e.g., phosphorylated) and inactive (e.g., dephosphorylated). Inactive substrates become activated through an enzymatic reaction catalyzed by enzyme E_1, and active substrates get deactivated by another enzyme E_2. Under appropriate conditions (see Chapter 3 or [63]), the rate of change of the concentration of active substrate A is well approximated by

$$\frac{dA(t)}{dt} = k_1 \frac{\overline{E_1}\Big(\overline{S} - A(t)\Big)}{K_1 + \overline{E_1} + \overline{S} - A(t)} - k_2 \frac{\overline{E_2}\, A(t)}{K_2 + \overline{E_2} + A(t)},$$

where $\overline{E_1}$ and $\overline{E_2}$ are dynamic variables that denote the total concentrations of E_1 and E_2, respectively. The total concentration of substrate is $\overline{S}$. K_1 and K_2, and k_1 and k_2, are the Michaelis-Menten, and the catalytic rate constants of E_1 and E_2, respectively. The standard Michaelis-Menten rates are a special case of the functions above (which are themselves valid more generally), obtained when the enzyme concentrations E_1 and E_2 are small.

The equation above shows that in general, the rate of change of active substrate concentration depends on the concentration itself. However, when the activation and deactivation reactions are operating in saturating substrate excess (i.e., when $K_1 + \overline{E1} \ll (\overline{S} - A)$, and $K_2 + \overline{E_2} \ll A$) then the equation above simplifies to

$$\frac{dA(t)}{dt} = k_1 \overline{E_1} - k_2 \overline{E_2}.$$

In such a situation, the rate of change of active substrate is independent of the concentration itself, resulting in a biological integrator, where the dynamics of A(t) may be obtained simply by integrating the function $k_1\overline{E_1} - k_2\overline{E_2}$ over time. We speculate later in this chapter about the possibility of active *PFK2* , a protein involved

in glycerol synthesis that may be activated directly by Hog1 phosphorylation, being the key integrator in the system.

Perfect Adaptation For Linear Systems

We have argued above that integral feedback is sufficient for the plant's output to perfectly adapt. Showing necessity is much more elaborate mathematically (the interested reader is referred to [156] and [131]); here we satisfy ourselves with a weaker argument that applies to the osmosensing network diagram assuming it is a linear time-invariant (LTI) system (e.g., any driven system of linear ordinary differential equations with constant coefficients, which is at rest at the start of the experiment). All linear time-invariant systems are fully characterized by their *transfer function* $T(s)$, which relates the Laplace transforms[2] of the system's input and output to each other via:

$$Y(s) = T(s)U(s), \tag{2.6}$$

where $U(s)$ and $Y(s)$ are the Laplace transforms of the input and output, respectively. Importantly, an LTI system has one or more integrators in it if its transfer function explodes (i.e., its magnitude becomes infinite) when evaluated $s = 0$.

At steady-state, the output of the system is simply given by

$$y_{ss} = T(0)u_{ss}, \tag{2.7}$$

where u_{ss} and y_{ss} denote the steady-state input and output, respectively, and $T(0)$ is simply the transfer function evaluated at the frequency zero. For linear time-invariance systems, perfect adaptation then requires $T(0) = 0$, so that y_{ss} is independent of u_{ss}.

[2]The (uni-lateral) Laplace transform $X(s)$ of a time trace $x(t)$ is given by

$$X(s) = \int_t^\infty x(\tau)e^{-s\tau}d\tau, \tag{2.5}$$

where s is a complex number. When s is purely imaginary the Laplace transform is the Fourier transform. See [106] for a good overview of linear time-invariant systems, including Laplace and Fourier transforms.

Perfectly Adapting Signals Lie Downstream of Integrator

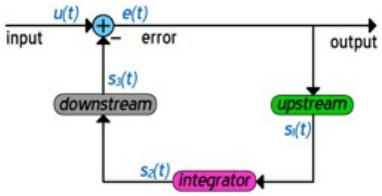

Figure 2-12: System diagram with an integrator flanked by two subsystems. The signals upstream of the integrator perfectly adapt, while those downstream do not.

Consider a system with the network diagram shown in Figure 2-12, where all the variables represent deviations from the systems rest level (i.e., all the internal variables and the system input are identically zero at rest). The system has a single input, $u(t)$, and a single negative feedback loop containing three subsystems: an integrator and two flanking subsystems called upstream and downstream based on their positions relative to the integrator. The error, $e(t)$, is the difference between the input $u(t)$ and $s_3(t)$, which is the output of the downstream subsystem and the last signal in the feedback loop. The error is the input to the upstream subsystem, which has output signal $s_1(t)$. The integrator subsystem integrates and potentially scales its input $s_1(t)$ to generate its output $s_2(t)$, which then serves as the input to the downstream subsystem.

Assume that the overall system is stable and that a persistent step-input, $u(t) = u_o$, is applied at time $t = 0$. At steady state, all the signals in the system are constant, and in particular, because s_2 is constant, s_1 must be zero. Therefore, the steady-state level of s_1 is independent of the system-input level, and the output of the upstream subsystem exhibits perfect adaptation. Now assume that when the steady-state output of the upstream subsystem is zero, its steady-state input is identically zero. This is certainly the case if the upstream subsystem is a stable linear system, but it is true more generally as well (e.g., for non-linear systems with a single steady state). Then, perfect adaptation of s_1 implies perfect adaptation of the error, e. This, in turn, directly implies that the last signal in the loop at steady state is identical to the input, i.e., $s_3 = u_o$. If the downstream subsystem is stable (and, in particular, has

no integrators in it), then, since its steady-state output is non-zero, its steady-state input s_2 must also be non-zero. This argument shows that the signals upstream of the last integrator in the negative feedback loop perfectly adapt, while those downstream of it do not.

2.4.2 The Integrator in the Osmo-sensing Network

These experimental data discussed so far already begin to restrict the possibilities of where in the osmo-sensing network the integrator can be. The potential locations we consider are four subsystems denoted H, I, D, and G (Figure 2-13). The H subsystem represents all relevant reactions that link an osmotic disturbance at the membrane with Hog1 nuclear enrichment; for example, the MAPK cascade and nuclear-import reactions are in this subsystem. In the D subsystem are Hog1-dependent mechanisms that promote glycerol accumulation, such as the transcriptional activation of genes that encode for glycerol-producing enzymes (e.g., Gpd1, Gpp2) and potential protein-protein interactions initiated by Hog1 that lead to glycerol accumulation. In subsystem I are the Hog1-independent mechanisms that contribute to osmolyte production, such as export-channel closure and gene-expression mediated by stress factors other than Hog1. Finally, subsystem G represents the metabolic reactions involved in glycerol synthesis and any other reactions that contribute to glycerol accumulation. Positing that each subsystem either contains an integrator or does not, we generated the 16 possible network configurations to guide our further analysis (Figure 2-13).

A critical aid in finding which of the subsystems contains the integrator(s) is the fact that with respect to the furthest-downstream integrator in a feedback loop, quantities upstream perfectly adapt, and those downstream do not. For instance, the observation that turgor pressure perfectly adapts —indeed, this is the primary function of regulating osmotic stress— stipulates that at least one integrator must exist in H, I, D, or G, since turgor pressure is upstream of each of these systems within the feedback loop. This allows us to reject network configuration (a), where none of the subsystems acts as an integrator. The previous result is easiest to see assuming that the subsystems are all LTI systems. Using techniques from control

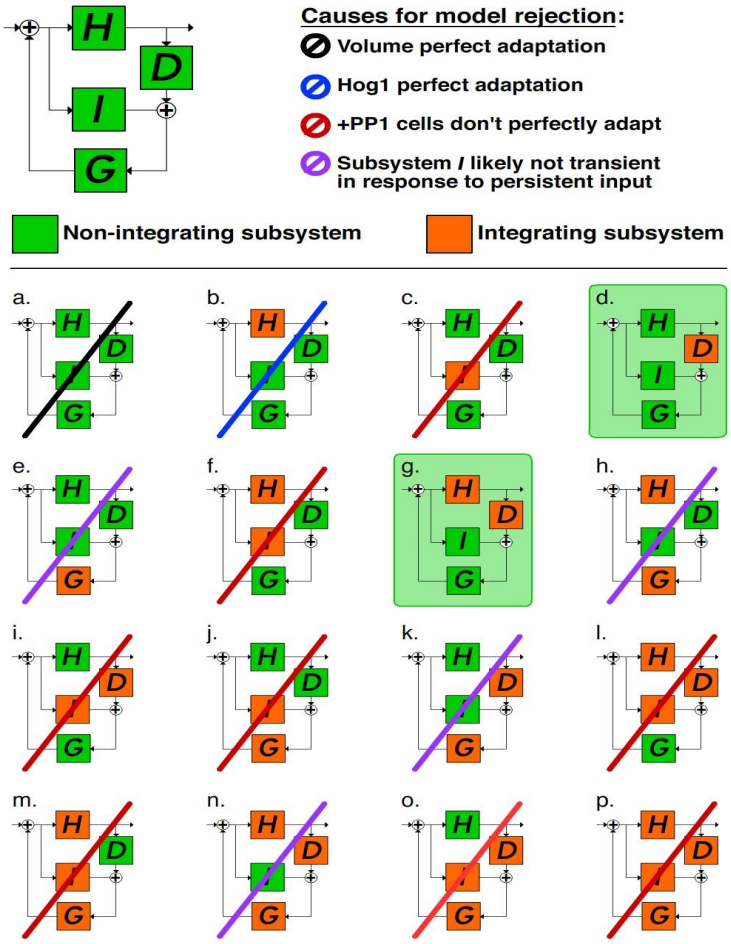

Figure 2-13: We consider all combinations for the presence (orange) or absence (green) of integrators in the subsystems H, D, I, and G, yielding the sixteen possibilities (a) through (m). For the reasons listed at top-right, which consider data from Figures 2-8 and 2-15, certain networks can be rejected. Only networks (d) and (g) satisfy all the constraints, and we implement (d) when later fitting the data to a model.

engineering, it is a straightforward task to obtain transfer functions from a given network diagram (given the transfer functions of the networks subsystems) or from the system of differential equations describing the system dynamics. The transfer function relating the system input to the error in the network of Fig 2-2b is given by

$$T_{ue}(s) = \frac{1}{1 + G(DH + I)},$$

where G, D, H and I are the transfer functions of the four subsystems in the network. Their dependence on the variable s has been omitted to avoid clutter (i.e., G(s)

is simply written as G). Perfect adaptation of the error requires that $T_{ue}(0) = 0$, which consequently implies that at least one of the four subsystems has a transfer function that goes to infinity (explodes) at $s = 0$. We can then conclude that perfect adaptation of the error requires that at least one of the four subsystems has at least one integrator.

Similarly, the transfer function between the system input and glycerol is given by

$$T_{ug}(s) = \frac{G(DH + I)}{1 + G(DH + I)},$$

which shows that as long as there is a single integrator in the feedback loop (i.e., as long as G(DH+I) explodes at s=0), then $T_{ug} = 1$, and the steady-state glycerol will equal the input, producing a steady-state error equal to zero.

The observation of Hog1 perfect adaptation imposes an additional constraint. Specifically, if the only integrator in the feedback loop were in subsystem H, then Hog1 nuclear enrichment would not perfectly adapt, because it is downstream of H. Equivalently, the transfer function of the input to Hog1 is given by

$$T_{uh}(s) = \frac{H}{1 + G(DH + I)},$$

so having integration only at the H subsystem does not result in $T_{uh}(0) = 0$. Thus, at least one of subsystems I, D, and G must contain an integrator, permitting the rejection of configuration (b).

Perfect Adaptation Requires Hog1 Kinase Activity, but not Gene Expression

We searched further for the location of the integrator by assessing whether biochemical inhibition of specific reactions ablated Hog1 perfect adaptation. Our first attempt blocked gene expression in the whole cell by pre-treating cells with the translation-elongation inhibitor cycloheximide, and then monitored cell volume and Hog1 nuclear enrichment in response to hyperosmotic shock. In the absence of salt shock, cyclo-

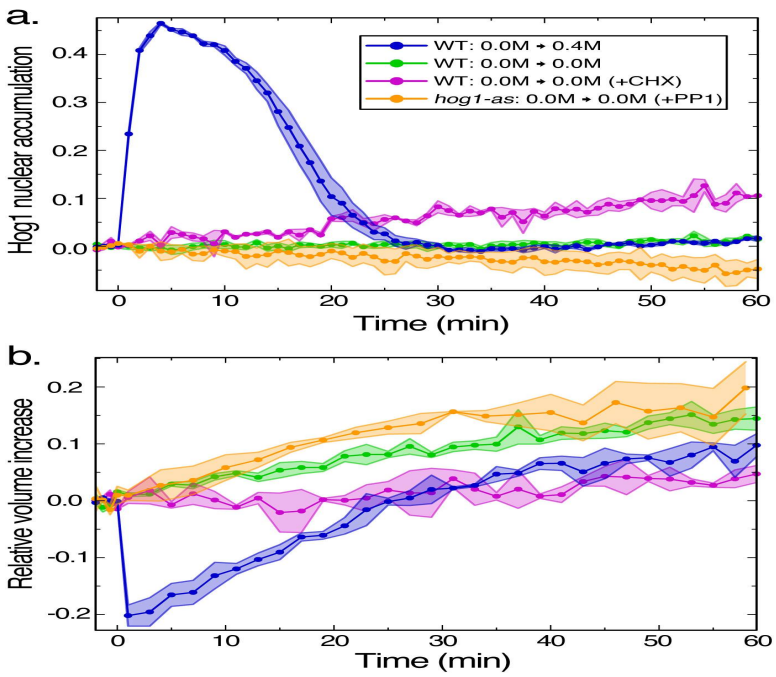

Figure 2-14: (A) Cultures from the cell line indicated (e.g., Wildtype or hog1-as) were treated with 100μg/ml cycloheximide (CHX) or 24μg/ml PP1, as specified in the legend. Only the dark-blue line corresponds to a culture that was osmoshocked, and this trace was included to provide a sense of scale. The +CHX and +PP1 traces shown here were subtracted from the response of osmostressed cells, to correct the raw traces and obtain a measurement of the osmostress-specific response. The traces of osmostressed cells corrected in this manner are shown in Figure 2-15. Dotted lines show the average response, obtained by averaging the population averages from three or four experiments, and the error boundaries depict the standard error of the average response. (B) Volume traces from the corresponding experiments in (A). As with the traces in (A), the traces here were substracted from those of osmostressed cells to correct raw volume measurements and obtain the osmostress-specific responses shown in Figure 2-15c.

heximide treatment decreased the growth in cell volume and led to a small increase in baseline Hog1 nuclear enrichment (Figure 2-14). Importantly, however, just as in wildtype cells, the steady-state level of both volume-growth and Hog1 nuclear enrichment in salt-stressed, cycloheximide-treated cells returned exactly to the baseline levels, indicative of perfect adaptation (Figure 2-15a). Despite the fact that gene expression is required for recovery from severe osmotic stress ([13]) and that hundreds of genes are activated or repressed in response to shocks ranging from mild to severe ([107]), these data indicate that gene expression is not required for Hog1 perfect

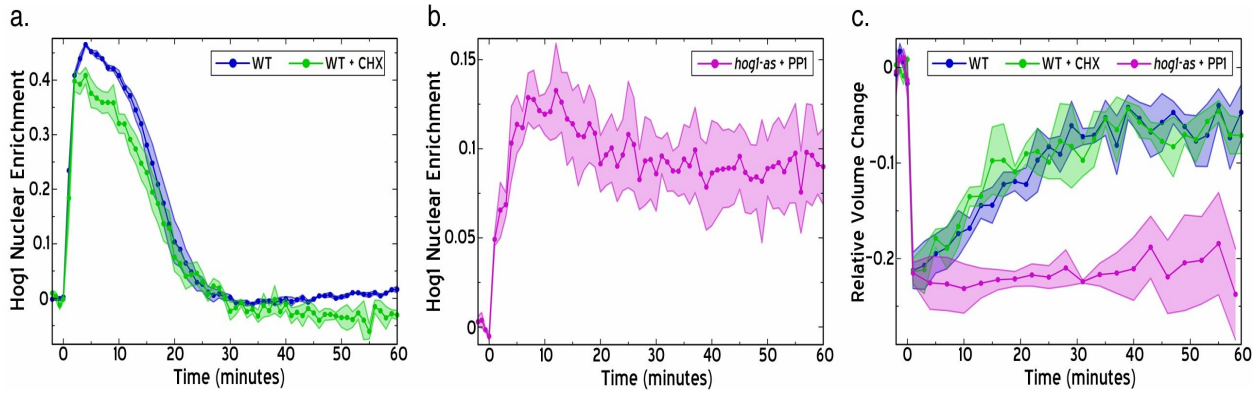

Figure 2-15: (A) Wildtype cells were loaded into a flow cell in minimal media. The green curve depicts the population mean and standard error (N=3) from experiments in which 100μg/ml cycloheximide was added to the cells five minutes before 0.4M NaCl entered the flow cell. The trace is corrected for drift in the Hog1 nuclear enrichment in cells not stressed with NaCl (see Figure 2-14). The blue curve illustrates cells that underwent the same osmo-shock but were not treated with cycloheximide. (B) Nuclear enrichment of (hog1-as)-YFP in cells treated with 24μM PP1 prior to hyperosmotic shock with 0.4M NaCl. As with the WT+CHX trace in (A), this trace is corrected to highlight the salt-specific response. The population mean and standard error (N=3) are plotted. (C) Volume traces from the corresponding experiments in (A) and (B), again corrected to highlight the salt-specific response.

adaptation to 0.4M NaCl, a moderate salt stress.

In contrast, Hog1 does require its kinase activity for perfect adaptation. To assay the effect Hog1 activity exerts on its own adaptation, we mutated the endogenous Hog1 gene such that the kinase activity of the mutant (hog1-as) could be specifically, rapidly, and inducibly ablated by the ATP-analog 1-NM-PP1 (PP1); see [152]. In the presence of PP1, perfect adaptation is lost (Figure 2-15b), as steady-state Hog1 accumulation does not return to its pre-stimulus level. As with the cycloheximide experiments, these results were corrected for the effect PP1 has on cells not shocked with salt (see Figure 2-14). This failure to adapt is not an artifact of the documented defect in nuclear import and export of hog1-as because nuclear enrichment returned to the pre-stimulus level upon removal of the stimulus (Figure 2-16). Furthermore, the effect is specific to Hog1-kinase activity, since cells expressing wildtype Hog1 perfectly adapted even in the presence of PP1 (Figure 2-17). Finally, PP1 treatment diminishes the post-shock volume recovery (Figure 2-15c), suggesting that kinase-dead Hog1

persists in the nucleus because turgor pressure is not restored. We conclude that Hog1 kinase activity is necessary for the proper functioning of the integral-feedback controller in the osmoadaptation network.

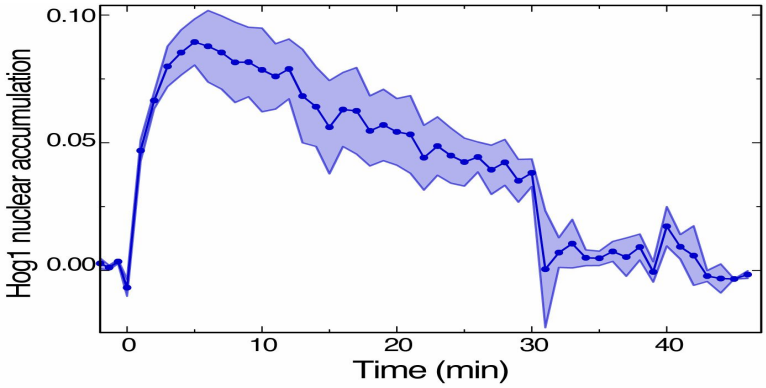

Figure 2-16: Hog1-as cells were grown and loaded into a flow-chamber using media lacking PP1. For five minutes prior to hyperosmotic shock, PP1 was introduced into the flow media. The +PP1 was supplemented with 0.4M NaCl to provide hyperosmotic shock. After 30 minutes, the +NaCl/+PP1 media with replaced by +PP1 media lacking NaCl. Hog1 nuclear enrichment was monitored throughout the experiment. Notably, upon removal of the hyperosmotic stress after 30 minutes, Hog1 exited the nucleus, returning to the pre-stimulus level, indicating the Hog1-as is capable of undergoing nuclear export if hyperosmotic stress is alleviated. Error bars represent the standard error of the population average obtained from three independent experiments.

In the presence of PP1, the connection between Hog1 and the D subsystem is broken, resulting in the modified transfer function

$$T_{uh}(s) = \frac{H}{1 + GI},$$

. Given that Hog1 does not perfectly adapt in this case, we know that the product $G(s)I(s)$ does not explode at $s = 0$, which implies that there is no integrator in the series connection of the G and I subsystems. This can be either because neither the G nor I subsystems posses integrator(s), or because one does but the other itself perfectly adapts. We argue later that the former is much more likely than the latter. The combination of all our findings points to D as the subsystem with the last integrator in the feedback loop, the key integrator responsible for perfect adaptation of the error and of Hog1.

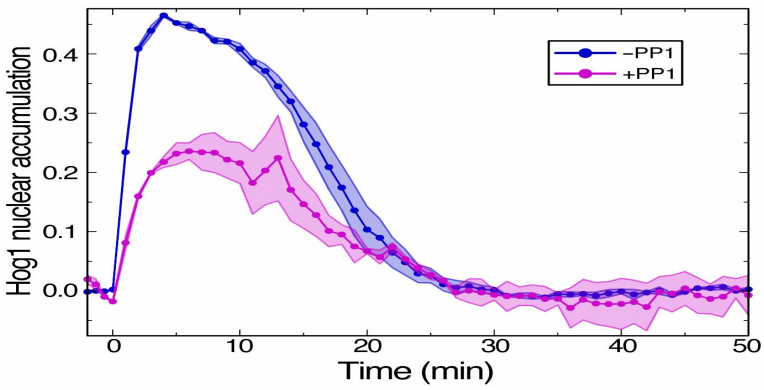

Figure 2-17: Wildtype cells grown in NaCl/PP1 media were loaded into the flow chamber, and then treated with PP1 for five minutes (e.g., NaCl/+PP1 media) before being hyperosmotically shocked by media with 0.4M NaCl (e.g., +NaCl/+PP1 media) at t = 0. Notably, nuclear enrichment throughout the experiment was diminished relative to PP1 cells, but perfect adaptation was still intact. Error bars represent the standard error of the population average obtained from three independent experiments.

Since PP1-treatment effectively severs the feedback loop between nuclear Hog1 and the Hog1-dependent mechanisms (i.e., the D subsystem), the loss of perfect adaptation further constrains the possible locations for the integrator(s). In particular, the Hog1-independent subsystem (the I subsystem) cannot contain the last integrator in the feedback loop. If it did, then the turgor pressure must perfectly adapt, and Hog1 likely would as well, yet we observe neither; thus, we reject scenarios (c), (f), (i), and (l) from Figure 2-13. Similarly, we can reject scenarios where both I and G (the glycerol-accumulation subsystem) act as integrators (i.e., scenarios (j), (m), (o), and (p), Figure 2-13), since this would also ensure perfect adaptation in the system.

We argue that the most likely network is one in which subsystem D has an integrator and G does not. If G did have an integrator, then the only way turgor pressure would fail to perfectly adapt in the presence of PP1 is if the output of the I subsystem itself perfectly adapted. We consider this scenario extremely improbable since it would require that all Hog1-independent osmotic-stress mechanisms be transient despite persistent loss of turgor pressure. The effect of each mechanism would have to cease within the 20-minute period before Hog1 reaches steady state in PP1-treated cells. On the basis of this argument, we reject scenarios (e), (h), (k), and (n) from

Figure 2-13, leaving only scenarios (d) and (g), where the D subsystem is the last integrator in the feedback loop.

Proper glycerol accumulation requires Hog1 kinase activity

We expected that the failure of PP1-treated cells to perfectly adapt corresponded to a glycerol-accumulation defect. Therefore, we measured internal glycerol accumulation in cells with intact or inhibited Hog1 kinase activity. In cells not treated with PP1 (PP1), intracellular glycerol rose rapidly after hyperosmotic shock but remained constant in unshocked cells (Figure 2-18a). In PP1-treated cells (+PP1), hyperosmotic shock caused a slight increase in internal glycerol, but the level achieved was significantly less than in PP1 cells, consistent with loss of turgor-pressure and Hog1 perfect adaptation. In accordance with our deduced network scenario, there was a striking similarity between the traces of Hog1 nuclear enrichment and the rate of internal glycerol accumulation in PP1 cells (compare Figure 2-8a and Figure 2-18b). If subsystem D is simply an integrator and most of the osmo-recovery is due to the Hog-dependent branch of the network, then the rate of glycerol accumulation would effectively be the derivative of the integral of Hog1. The rate of glycerol accumulation would therefore be simply a scaled version of the Hog1 curve. In +PP1 cells, where perfect adaptation is lost, this correspondence is also lost: the Hog1 and glycerol-accumulation-rate curves diverge at their post-stimulus steady-state levels.

Insufficient glycerol accumulation in +PP1 cells prompted us to investigate the effect of Hog1 kinase activity on glycerol synthesis and leakage. We measured cell-density-normalized levels of total glycerol and extracellular glycerol over time in the presence and absence of osmotic shock and PP1 (Figure 2-18c,d). First, we observed a role for Hog1 kinase activity in regulating glycerol leakage independent of osmotic stress. In particular, cells cultured overnight in PP1 media containing 0.0M, 0.1M, or 0.2M NaCl spanned a two-fold range in intracellular glycerol levels, and addition of PP1 caused significant glycerol leakage in each preparation (Figure 2-18e). Interestingly, upon stressing cells with salt, the transient decrease in glycerol leakage observed in PP1 cells was also apparent in +PP1 cells, consistent with certain glycerol-leakage

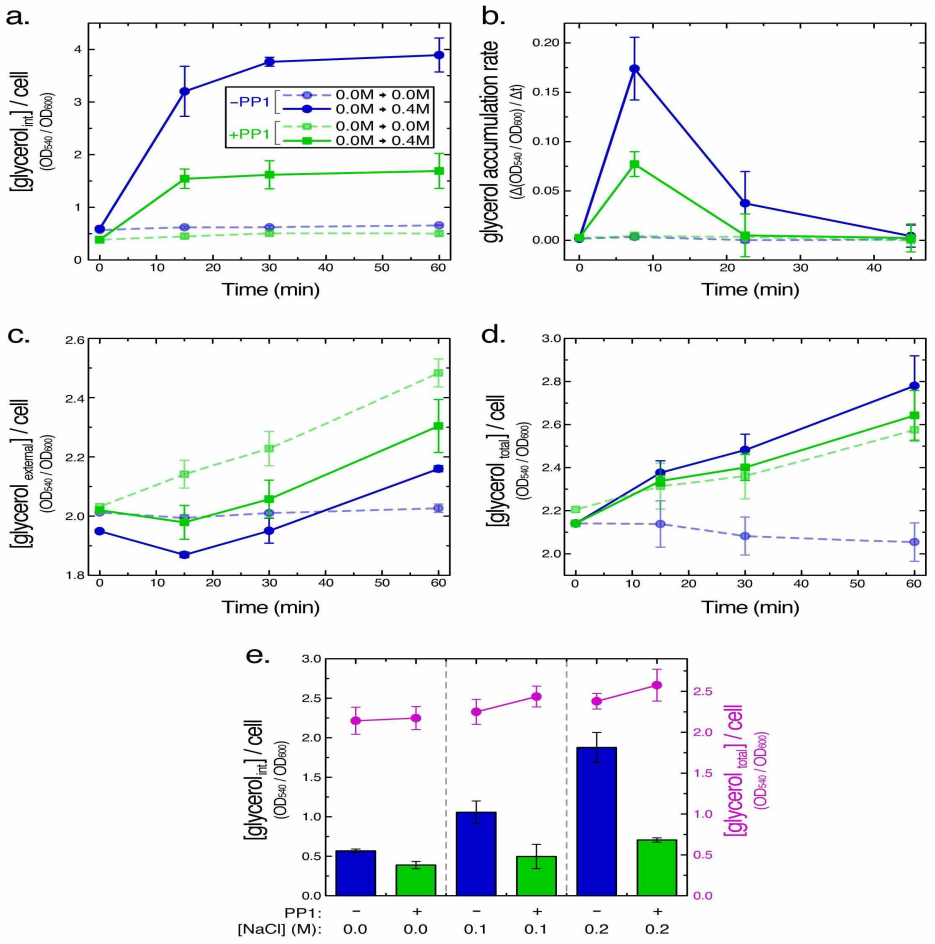

Figure 2-18: (A) Dynamics in the concentration of intracellular glycerol (OD540 measurement from glycerol kit; see Methods), corrected for cell-growth (OD600 measurement), were measured in hog1-as cells in the presence and absence of 24μM PP1 (administered 30 minutes before experiment) and hyperosmotic shock with 0.4M NaCl (administered at t = 0). Data points represent the mean of three independent experiments, with error bars showing the standard deviation. (B) The glycerol accumulation rate versus time. Rates were computed as the slope of the traces in (A), and the time of each plotted point corresponds to the midpoint of the two time points used in the slope computation (e.g., the slope between the 0-minute and 15-minute time points is plotted at 7.5 minutes). Error bars represent the combined standard deviation from data points in (A), assuming measurements were independent at different time points. (C) and (D): Dynamics in the extracellular (C) and total (D) glycerol concentration (see Methods) in experiments outlined in (A) and described in the main text. Data points and error bars are as described in (A). (E) PP1-treatment causes leakage of intracellular glycerol. hog1-as cells were grown overnight in media containing 0.0M, 0.1M, or 0.2M NaCl. Cultures were split, and one culture was treated with 24μM PP1. After 30 minutes, the intracellular glycerol concentration from both cultures was measured as in (A) and plotted as blue and green bars. From the same cultures, the total glycerol was measured as in (D; see Methods) and plotted in purple. The PP1-induced fall in intracellular glycerol is not coupled by a fall in total glycerol, indicating that glycerol is being leaked from the cell.

mechanisms operating independently of Hog1 kinase activity (Figure 2-18c).

As with glycerol leakage, the role for Hog1 in regulating glycerol synthesis is complicated. In the absence of osmotic stress, PP1 cells produce glycerol at a rate proportional to the cell-growth rate (Figure 2-18d). In contrast, synthesis grows at a much faster rate in +PP1 cells, presumably to counteract the increased glycerol leakage just discussed. Together, these data demonstrate that cells can upregulate glycerol synthesis even in the absence of Hog1 kinase activity. Upon hyperosmotic stress, however, PP1 cells rapidly and significantly upregulate glycerol synthesis, whereas +PP1 cells cannot further upregulate synthesis. We conclude that either this upregulated glycerol-synthesis rate is the maximum that can be achieved, or Hog1 kinase activity is required for the osmoshock-induced increase in glycerol synthesis that contributes to glycerol accumulation.

Simple model captures key data features

To test the validity of our deduced network scenario, we attempted to fit it to our Hog1 and glycerol data gathered in the presence and absence of PP1. For simplicity and to minimize parameters, we assume the system is linear and consider only data gathered for hyperosmotic shocks from 0.0M to 0.4M. This restriction neglects some aspects of system behavior (e.g., saturation of Hog1 nuclear enrichment for stronger osmotic shocks; Figure 3a) but still includes most important dynamic features.

We model the H and G subsystems as first-order linear systems, D as an integrator, and I simply as a constant. Letting $s_1(t)$ be the measured Hog1, $s_2(t)$ be the output of the D subsystem, and $s_3(t)$ be glycerol, the previous modeling assumptions result in the network diagram in Figure 2-19, and in the following differential equations:

$$\frac{d}{dt}\begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} = \begin{bmatrix} -\gamma_h & 0 & -k_h \\ \alpha_h & 0 & 0 \\ 0 & 1 & -(\alpha_i + \gamma_g) \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} k_h \\ 0 \\ \alpha_i \end{bmatrix} u. \tag{2.8}$$

The measured glycerol is simply a scaled copy of s_3, given by $g = c_g s_3$. We model addition of PP1 by setting α_h to zero, effectively breaking the connection

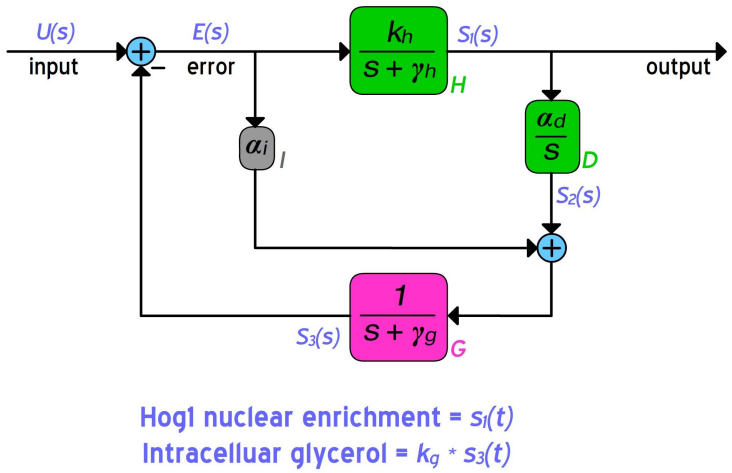

Figure 2-19: Model diagram showing the transfer functions we implemented for the network subsystems, resulting in the differential equations in (2.8).

between the H and D subsystems, and by letting the gain of Hog1 activation k_h take on a new value k_h to reflect the impact of loss of kinase activity on the nuclear import/export dynamics of Hog1. The resulting model is then completely specified by seven parameters.

Parameter	Value
k_h	4.96×10^{-1} min^{-1}
k'_h	1.47×10^{-1} min^{-1}
γ_h	3.69×10^{-1} min^{-1}
γ_g	1.19×10^{-1} min^{-1}
α_i	8.06×10^{-2} min^{-1}
c_g	7.76
α_h	1.06×10^{-2} min^{-1}

Table 2.1: Parameter values for the model fits shown in Figure 2-20a-d.

To test whether this model can explain the measured Hog1 and glycerol dynamics in the presence and absence of PP1, we fit the model simultaneously to four time traces: the measured Hog1 and glycerol dynamics both in the presence and absence of PP1 in response to a shock of 0.4M NaCl. We defined a fitting function that assigns a cost to every choice of parameter values. We computed a cost for each of the four time traces to which we fit the model, where the cost for each trace is the time-averaged squared difference between the model prediction and the data. The

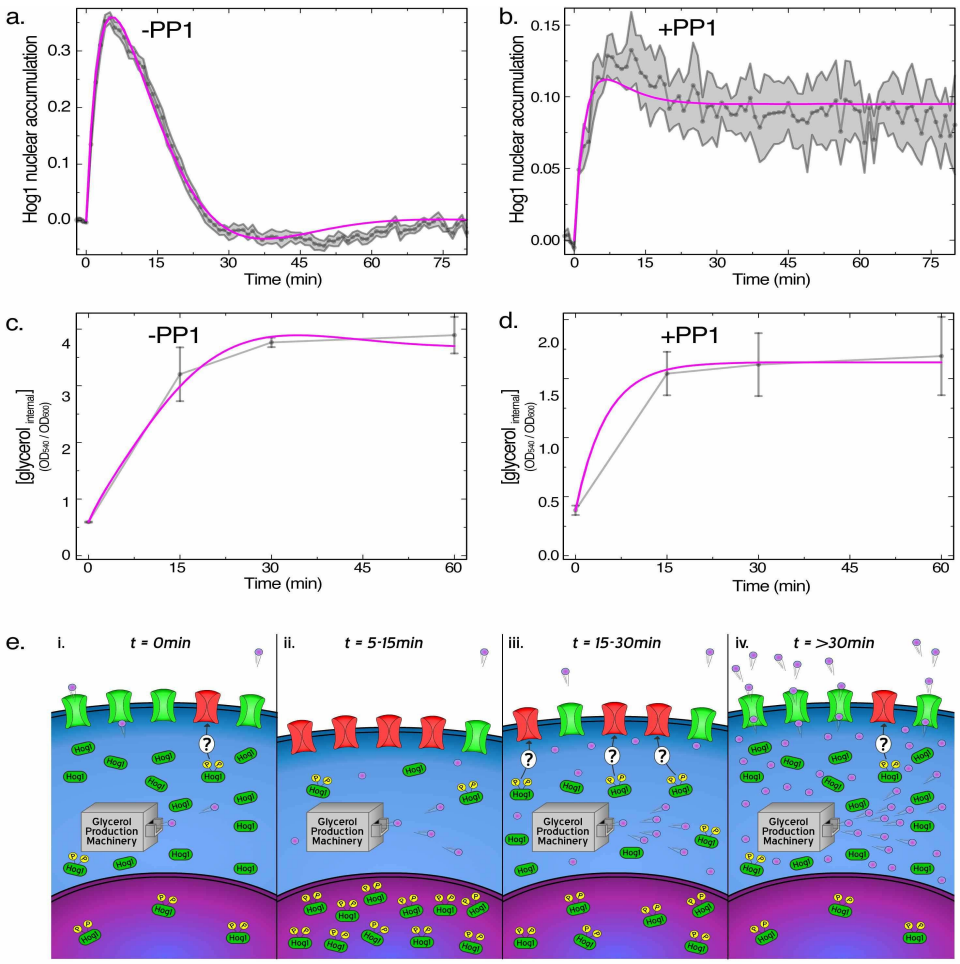

Figure 2-20: (A)-(D): Our seven-parameter model fits well with the Hog1-nuclear-enrichment and intracellular-glycerol data gathered from hog1-as cells stressed with 0.4M NaCl in the presence or absence of PP1 (model fit shown in purple, experimental data in gray). The model captures the perfect adaptation of Hog1 nuclear enrichment in the absence (A) of PP1, as well as the loss of perfect adaptation in the presence of PP1 (B). Similarly, the model fits well with the dynamics of internal glycerol accumulation (C and D). (E) Cartoon representation of key features of yeast hyperosmotic shock response, explained in detail in the Discussion. Hyperosmotic shock is applied at t = 0. Green membrane channels are open, allowing the transport of glycerol, shown as purple circles. Upon closure, the channel is depicted in red.

overall cost of a given set of parameter values is simply the sum of the costs from the four individual time traces. We then use the $fminsearch$ routine in Matlab (The Mathworks, Netwon, MA) for unconstrained nonlinear optimization to find a set of parameter values resulting in a minimum of the fitting cost function. The resulting parameters, listed in Table 2.4.2, were used to produce the model fits shown in Figure 2-20a-d. We find the model captures the key data features.

2.5 Discussion

We have measured the yeast hyperosmotic-shock response and shown that Hog1 nuclear enrichment perfectly adapts, which implies that the system must contain an integrator that ensures perfect adaptation. Despite a well-known role in mediating an osmostress-induced transcriptional response, Hog1 does not require gene expression to perfectly adapt. However, we were able to ablate perfect adaptation by inhibiting Hog1 kinase activity, suggesting that an interaction between Hog1 and a factor involved in glycerol accumulation may serve as the integrator. Taken together, these experiments greatly reduced the set of network scenarios that could describe the system dynamics. We chose one of the most-probable models and found that with only seven parameters, we could fit both Hog1 and glycerol dynamics in the presence and absence of Hog1 kinase activity. Our analysis suggests that the integrator is a Hog1-dependent mechanism (i.e., in the D subsystem). Most known reactions linking Hog1 with glycerol production rely on gene expression, such as the upregulation of *GPD1* and *GPP2* . However, less is known about reactions in which Hog1 phosphorylates a substrate directly involved in glycerol accumulation. Although future experiments will be required to determine exactly which reaction is an integrator, the search may be facilitated by considering reactions downstream of Hog1 that are most likely to operate at saturation, since saturated reactions are one way to biologically implement integration, as discussed above and in Chapter 3, and in [63].

A particularly enticing possibility is that active *PFK2* is the output of the integrator in the feedback loop. Active *PFK2* increases the production of glycerol precursors

from the glycolysis pathway, and it is activated by phosphorylation on Ser644 in response to hyperosmotic shock in a manner dependent on the Hog1 MAPK cascade. *PFK2* may be activated by Hog1 kinase activity since it has consensus sites for MAPK phosphorylation. Remarkably, estimates of protein abundance from a genome-wide fluorescent fusion-protein study ([53]) indicate that *PFK2* significantly outnumbers Hog1, approximately 90,000 to 7,000. If Hog1 indeed activates *PFK2* , then the rate of this reaction would likely be independent of the *PFK2* concentration because of active Hog1 saturation. Although the phosphatase of *PFK2* is not known, most characterized serine phosphatases have fewer than 10,000 copies in the cell, with the most-abundant having less than 20,000 copies. Thus, it is certainly possible that the reaction deactivating *PFK2* is also operating at saturation, with a rate independent of the *PFK2* concentration. If the reactions activating and deactivating *PFK2* are indeed saturated, then the rate of change of active *PFK2* will be independent of *PFK2* itself, making active *PFK2* the output of the integrator responsible for the networks perfect adaptation and proper homeostasis.

A dynamic perspective on hyperosmotic shock recovery

Our results yield the following dynamic portrayal of osmoregulation (Figure 2-20e). In the absence of hyperosmotic shock, cells maintain a constant internal glycerol concentration. Glycerol is synthesized at a rate slightly higher than that required to keep pace with cell-growth, and some glycerol is leaked. Hog1 kinase activity regulates this leakage of glycerol even in the absence of hyperosmotic shock as evidenced by the decrease in internal glycerol upon addition of PP1 to hog1-as cells (Figure 2-18e). The simultaneous and comparable increase in external glycerol suggests that the role for Hog1 is regulating not the degradation but rather the export of glycerol, presumably through a role at the membrane. The required Hog1 kinase activity in unstressed cells likely comes either from a few phosphorylated and active Hog1 molecules and/or from many unphosphorylated Hog1 molecules, each with low basal kinase activity. Our results cannot definitively distinguish between these two possibilities, but they nonetheless support the former: even in unstressed cells, it appears

that at least some Hog1 is activated since fluorescence in the nucleus and cytoplasm are comparable (Figure 2-4). Soon after a hyperosmotic shock (Figure 2-20e, panel ii), turgor pressure decreases, causing a concomitant fall in cell volume and rise in the activation and nuclear accumulation of Hog1. Within the first 15 minutes of shock with 0.4M NaCl, both the synthesis and leakage of glycerol are modulated to increase internal glycerol; the glycerol synthesis rate increases (Figure 2-18d) while the leakage rate falls in a manner independent of Hog1-kinase activity (Figure 2-18c). Beyond the first 15 minutes (Figure 2-20e, panel iii), cell volume rises (Figure 2-8b); nuclear Hog1 accumulation begins to fall (Figure 2-8a); the rate of glycerol synthesis continues to rise (Figure 2-18d), and some glycerol begins to leak to the exterior (Figure 2-18c). It is noteworthy that internal glycerol continues to increase beyond 15 minutes in PP1 cells (Figure 2-18a) but not in +PP1 cells, indicating that +PP1 cells prematurely leak as much glycerol as they synthesize. If leakage prevention were stronger and/or endured for longer, then these cells could also accumulate enough glycerol to perfectly adapt; these data suggest that Hog1 kinase activity plays a role in the maintenance of glycerol retention beyond 15 minutes, though its effect may be indirect.

Beyond 30 minutes (Figure 2-20e, panel iv), the Hog1 nuclear enrichment level is the same as the pre-stimulus level (Figure 2-8a), the hallmark of perfect adaptation. Although the pre- and post-stimulus rates of glycerol accumulation are identical (Figure 2-18b), glycerol synthesis and leakage are significantly higher after the stimulus than before it (Figure 2-18c,d). This has a nontrivial implication on system modeling. In particular, it invalidates models in which the integrator is immediately upstream of glycerol accumulation (i.e., the last step of the G subsystem). To see this, suppose that the rate equation describing glycerol accumulation were an integrator. By definition, the rate would then be independent of the current internal glycerol concentration, and thus both the synthesis and degradation rates would also be independent of internal glycerol. But, the inputs to the G subsystem, namely the activity of Hog1 and the Hog1-independent factors that respond to turgor pressure, are identical pre- and post-stimulus. Only the internal concentration of glycerol has changed, and yet the rates of synthesis and leakage change as well. Thus, even if the G subsystem

were to contain an integrator, which we argued against earlier, it would not be the last reaction in the subsystem. Because our deduced model has a first-order system downstream of the integrator, it can accommodate increased synthesis and leakage of glycerol despite perfect adaptation in turgor pressure and Hog1 activation.

A systematic approach for biological systems analysis

The Hog1 system is well characterized genetically and biochemically, and it has long served as a paradigm for studying eukaryotic MAPK signal transduction. More recently, the important dynamic properties of the network have been explored by several modeling efforts, including the exhaustive model in [86] and its simplified counterpart [52], as well as our minimalist model in [100]. Each has its own virtues; for instance, the exhaustive model makes predictions about the levels of each species in the model, such as GPD1 mRNA, phosphorylated Ssk2 , and glycerol-3-phosphate.

Our modeling efforts here and in [100] underscore the power of applying engineering principles to any biological system that has well-defined and quantifiable inputs, internal variables, and outputs. In analyzing such systems, first measuring the frequency response (as we do in [100]) provides an estimate of the number of relevant dynamic variables that a minimal model should have; additionally, these measurements may reveal the basic computation or function that the system performs. For systems potentially involving hundreds of reactions, each with unique kinetics, such analysis can appreciably reduce the complexity of a model without compromising, and potentially enhancing, the insight into systems-level behavior. Once the minimum number of dynamic variables has been estimated, the existing biochemical knowledge of the system can be leveraged to infer which biological quantities correspond to the relevant dynamic variables, and to create a basic network diagram such as the one in Figure 2-2b. Biological quantities whose importance has been established in previous studies, such as glycerol and cell volume for the yeast osmotic-shock response, and/or quantities with dynamics that are expected to change on an appropriate time-scale are all good candidates for relevant dynamic variables. By measuring the dynamics of the internal variables in the systems network diagram, the model can be further

constrained, and links can be drawn more confidently between model elements and biological mechanisms.

Perfect adaptation is one such dynamic feature that restricts potential models, where a basic result from control engineering provides important information about biological mechanism. We expect similar analyses to be particularly useful in the study of other homeostatic systems that are ubiquitous in biology (e.g., blood calcium levels in [34]), where perfect adaptation is presumably the paramount dynamic property of the network. Most of these systems will likely achieve robust perfect adaptation through a negative feedback loop with one or more integrators; thus, an important future endeavor is to better understand how biological systems implement integration at the molecular level. Since the simple loss of integral feedback can fundamentally transform the function of a systemfrom one with transient output into one with persistent and potentially deleterious outputidentifying and characterizing biological mechanisms providing integral feedback should be instrumental in the future study of homeostatic systems, the design of perfectly-adapting biosynthetic circuits, and the development of therapeutics to combat disease.

2.5.1 Methods

Strain background and construction

Our haploid wildtype strain here (DMY017) was derived from the DMY007 strain of Mettetal et al, the only difference being that the SHO1 ORF was excised via PCR, using methods described in Mettetal et al. Our PP1-sensitive strain (DMY034) was created by deleting the HOG1 ORF from DMY017 (DMY030), generating a single-base mutation of HOG1 on a plasmid using QuikChange (Stratagene), and then inserting the resulting hog1-as ORF into DMY030 using PCR integration. For more detail on this, please refer to the PhD thesis of Dale Muzzey.

Glycerol assays

All glycerol measurements were performed by Dale Muzzey. DMY034 cells were grown overnight in 100ml selective minimal media containing 0.0M, 0.1M, or 0.2M NaCl. Once cells entered log phase, cultures were supplemented either with PP1 (1-NM-PP1, EMD Biosciences) dissolved in DMSO or the same volume of DMSO alone. Thirty minutes later, 50ml of the culture was vortexed in a flask containing 1.17 grams NaCl crystals such that [NaCl] would rise by 0.4M when dissolved; this approach permitted us to change only the salt concentration, diluting neither the cells nor the media. At each time point (0, 15, 30, 60 minutes) relative to the addition of salt, three samples were collected: 100μl for cell density, 200μl for the total glycerol measurement, and 10ml that yields both the external and internal measurements. Between sample collections, flasks were shaking at 225rpm in the 30C incubator.

To account for changes in cell density throughout the experiment, the OD600 was measured on a sample containing cell culture and fresh water in the ratio 1:10; because salt-shock affects cell volume, we diluted cells in water to ensure that the volumes of salt-shocked and unshocked cells were roughly comparable.

The 200μl sample harvested for the total glycerol measurement was incubated at 95C for 10 minutes and then spun at 13000rpm for 3 minutes to pellet cell debris. A sample of the supernatant was added directly to the Free Glycerol Reagent Kit (Sigma) as directed and then the OD540 was meausured.

The 10ml sample was first spun at 2000rpm for 2 minutes. To acquire the external glycerol value, a sample of the supernatant was added to the glycerol kit, and the OD540 was measured. For the internal measurement, cells in the pellet were washed in 1ml fresh media that matched the cell media in [NaCl] in order to prevent hypoosmotic shock and internal glycerol leakage during washing. After another flash spin, cells were resuspended in 1ml fresh media and then incubated, spun, and sampled as described for the total glycerol measurement.

Chapter 3

Operating Regimes of Signaling Cycles: Statics, Dynamics, and Noise Filtering

3.1 Summary

A cell is subjected to constantly changing environments and time-varying stimuli. Signals sensed at the cell surface are transmitted inside the cell by signaling pathways. Such pathways can transform signals in diverse ways and perform some preliminary information processing. A ubiquitous building block of signaling pathways is a cycle of covalent modification (e.g., phosphorylation and dephosphorylation in MAPK cascades). What kind of information processing and filtering can be accomplished by this simple biochemical circuit?

Signaling cycles are particularly known for exhibiting a highly sigmoidal (ultrasensitive) input-output characteristic in a certain steady-state regime. Here we systematically study the cycle's steady-state behavior and its response to time-varying stimuli.[1] We demonstrate that the cycle can actually operate in four different regimes, each with its specific input-output characteristics. These results are obtained using

[1]The material in this chapter has been published in [63].

the total quasi-steady-state approximation, which is more generally valid than the typically used Michaelis-Menten approximation for enzymatic reactions. We invoke experimental data that suggests the possibility of signaling cycles operating in one of the new regimes.

We then consider the cycle's dynamic behavior, which has so far been relatively neglected. We demonstrate analytically that the intrinsic architecture of the cycles makes them act — in all four regimes — as tunable low-pass filters for small enough time-varying deviations of the input from baseline level, filtering out high-frequency fluctuations or noise in signals and environmental cues. Moreover, the cutoff frequency can be adjusted by the cell. Numerical simulations show that these analytical results continue to hold quite well even for bigger time-varying deviations from baseline level. We suggest that noise filtering and tunability make signaling cycles versatile components of more elaborate cell signaling pathways.

3.2 Introduction

Cells rely on chemical interactions to sense, transmit, and process time-varying signals originating in their environment. Because of the inherent stochasticity of chemical reactions, the signals transmitted in cell signaling pathways are buried in noise. How can cells then differentiate true signals from noise? We examine this in the context of a basic but ubiquitous module in signaling cascades: the signaling cycle. While an individual signaling cycle is simply an element of a large signaling network, understanding its response is an essential first step in characterizing the response of more elaborate signaling networks to an external stimulus ([117, 124]).

Each cycle consists of a substrate protein that can be in one of two states: active (e.g., phosphorylated) or inactive (e.g., dephosphorylated), see Fig.3-1. The protein is activated by a protein kinase that catalyzes a phosphorylation reaction. The protein gets inactivated by a second enzymatic reaction catalyzed by a phosphatase. The activity/concentration of the kinase can be considered as an input of the cycle. The response of the cycle is the level of phosphorylated substrate protein that is not bound

to the phosphatase and can thus interact with any downstream components of the signaling pathway.

Signaling cycles can also require multiple phosphorylations for activation. Furthermore, cycles of phosphorylation are frequently organized into cascades where the activated substrate protein serves as a kinase for the next cycle. Activation of the first kinase in a cascade can be triggered by a receptor that has received a specific stimulus (ligand, photon, dimerization, etc.). In addition, feedback processes may be present. The dynamics of signaling cascades have been the subject of active research using modeling and experiments. Theoretical and computational studies of eukaryotic signaling cascades span a broad range of questions such as those concerning the dynamics of the EGFR ([153]) or apoptosis signaling pathways ([94]), the propagation of noise and stochastic fluctuations ([139, 123, 12]), the role of feedbacks ([18, 89, 92, 84]) and scaffolding proteins ([91, 16]), the contribution of receptor trafficking ([147]) and spatial effects ([3, 85, 92]), the origin of bistability ([98, 109, 90]) and oscillations ([123, 26]), and the consequences of multiple phosphorylations ([123, 135, 66, 26, 108, 44, 43, 149]).

In this paper, our focus will be on the statics and dynamics of the basic singly modified signaling cycle. The seminal contribution Goldbeter and Koshland considered the steady-state response of this basic cycle and demonstrated that, under appropriate conditions, the response can be in a highly sigmoidal ultrasensitive regime, or in a hyperbolic regime ([61]; see below). Most modeling studies have assumed that all signaling cycles operate in the ultrasensitive regime; a few studies have also considered the hyperbolic regime ([32, 72]). Here we demonstrate that there are actually four major regimes, with the ultrasensitive and hyperbolic regimes being two of them.

Several previous studies that treat signaling cycles as modules have focused on the steady-state response to a constant input, largely ignoring responses to time-varying stimuli (see e.g., [61, 42, 66]). A study of Detwiler *et al.* ([32]) considered the dynamic response of the cycle in the hyperbolic regime (when both forward and backward reactions are first-order), and found low-pass filtering behavior. A recent study also examined the dynamic response of these two regimes and compared them

in their robustness to intrinsic and extrinsic noise ([93]).

Here we systematically consider both the steady-state response, and the dynamic response to time-varying stimuli. To model the enzymatic reactions in the signaling cycle, we use the total quasi-steady-state approximation (tQSSA; see [142]). The tQSSA is valid more generally than the Michaelis-Menten rate law, which assumes the enzyme to be present in much smaller concentration than its substrate, an assumption that is not generally valid in signaling pathways. We then use our model to examine possible regimes of the cycle, and identify two new steady-state regimes, for a total of four different behaviors, each being potentially useful in different signaling applications. Although these four regimes are defined at extreme parameter values, we numerically show that in fact together they cover almost the full parameter space. We obtain analytic approximations to the steady-state characteristics of each of the four regimes, and refine the conditions under which the two regimes identified by Goldbeter and Koshland are in fact achieved.

To obtain a fuller picture of the signaling cycle and its function, we then analyze its response to time-varying kinase activity. We demonstrate analytically that the intrinsic architecture of the cycles makes them act — in all four regimes — as tunable low-pass filters for small enough time-varying deviations of the kinase activity from baseline levels. Numerical simulations show that these analytical results continue to hold quite well even for bigger deviations from baseline level.

The four different regimes of the signaling cycle make it a versatile element, able to perform various signaling functions, while its low-pass filtering enables it to operate in noisy environments. These properties may help explain why signaling cycles are so ubiquitous in cell signaling.

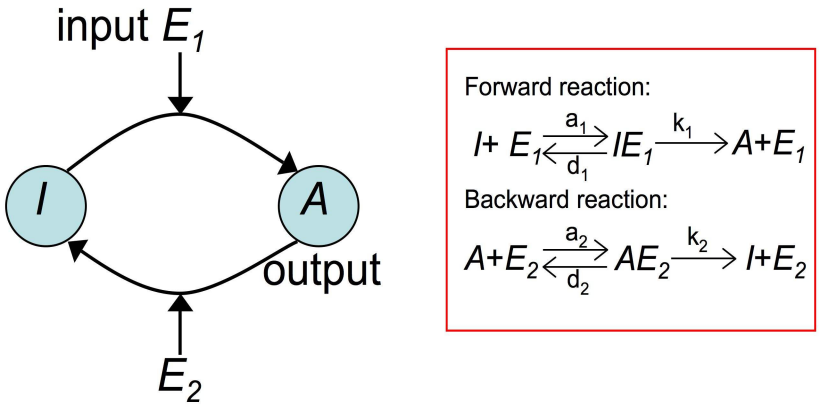

Figure 3-1: **Diagram of the signaling cycle**. The cycle consists of a protein that can be in an inactive (I) or active (A) form. It is activated and deactivated by two enzymatic species, termed kinase (E_1) and phosphatase (E_2), respectively. The reactions and reaction rates that describe the cycle are shown on the right.

3.3 Results

3.3.1 Model

The signaling cycle is modeled by two enzymatic reactions, as illustrated in Fig. 3-1: A forward enzymatic reaction catalyzed by kinases (enzyme 1, E_1) produces active proteins (A) from the inactive ones (I), and a backward reaction catalyzed by phosphatases (enzyme 2, E_2) de-activates active proteins:

$$I + E_1 \underset{d_1}{\overset{a_1}{\rightleftarrows}} IE_1 \xrightarrow{k_1} A + E_1 \tag{3.1}$$

$$A + E_2 \underset{d_2}{\overset{a_2}{\rightleftarrows}} AE_2 \xrightarrow{k_2} I + E_2. \tag{3.2}$$

Here a_1 (d_1) and a_2 (d_2) are substrate-enzyme association (dissociation) rates, and k_1 (k_2) is the catalytic rate of the forward (backward) enzymatic reaction. For notational convenience, we shall use the same symbol to denote a chemical species as well as its concentration. The input to the cycle is the total concentration of the active kinase, $\overline{E_1} = E_1 + IE_1$, while the output is the concentration of the free (i.e., not bound to phosphatase) active protein A.

While such systems are usually studied using Briggs-Haldane or Michaelis-Menten (MM) approximations (see [20, 102]), both can be inapplicable as they assume much lower concentration of the enzyme than of the substrate. In fact, substrates and enzymes of MAPK pathways are usually present at comparable concentrations in *S. cerevisiae* and *Xenopus* oocyte cells (as reported in [42] and consistent with data from the library of GFP-tagged proteins [53]).

Instead, we rely on the total quasi-steady-state approximation (tQSSA; see [142, 143, 144, 126] and Section 3.5 below) to obtain the following equation for the concentration of the total active protein, $\overline{A} = A + AE_2$:

$$\frac{d\overline{A}(t)}{dt} = k_1 \frac{\overline{E_1}\Big(\overline{S} - \overline{A}(t)\Big)}{K_1 + \overline{E_1} + \overline{S} - \overline{A}(t)} - k_2 \frac{\overline{E_2}\,\overline{A}(t)}{K_2 + \overline{E_2} + \overline{A}(t)}. \tag{3.3}$$

Here X denotes the concentration of an unbound chemical species and $\overline{X}$ denotes the total concentration of bound and unbound forms; $\overline{S}$ stands for the total concentration of substrate protein (in both active and inactive forms); and $K_1 = \frac{k_1+d_1}{a_1}$ and $K_2 = \frac{k_2+d_2}{a_2}$ are the MM constants for the kinase and the phosphatase, respectively. We have written $\overline{A}(t)$ explicitly with its time argument t to emphasize that it is a dynamic variable; however, for notational simplicity, we will omit the time argument in the rest of the paper and simply write $\overline{A}$. The quantities $\overline{E_1}, \overline{E_2}$ and $\overline{S}$ are constant here (although later in the paper we consider the dynamic response to small variations in $\overline{E_1}$). Even though the above equation is written in terms of $\overline{A}$, the free active protein concentration A, which is of primary interest, is simply recovered through the expression $A = \frac{K_2+\overline{A}}{K_2+\overline{E_2}+\overline{A}}\overline{A}$ (see Section 3.5.1).

Equation 3.3 shows the dependence of the rate of production of the active protein

on the number of kinases through the first term (phosphorylation), and on the number of phosphatases through the second term (dephosphorylation). In particular, when the total amounts of both kinase and phosphatase are small ($\overline{E_1} \ll K_1 + \overline{S} - \overline{A}$ and $\overline{E_2} \ll K_2 + \overline{A}$), the two terms in Equation 3.3 reduce to the standard MM rates for the forward and backward enzymatic reactions of the cycle. The tQSSA has also been recently proposed and applied by Ciliberto *et al.* in [29] to model networks of coupled enzymatic reactions, including interconnections of phosphorylation cycles.

Our key equation 3.3 simplifies for extreme combinations of parameter values (i.e., regimes) that are still of potential biological interest. This equation allows us to analytically examine (a) the possible cycle regimes of the system in steady state, and (b) the dynamic response of the system to time-varying inputs (time-varying activation of the kinase). The numerical results we present here are not constrained by the quality of the approximation since they are based on direct simulation of the mass action kinetics equation for the full system of reactions of Equations 3.1 and 3.2 (see Methods).

3.3.2 Four Regimes of the Signaling Cycle

Each enzymatic reaction can be in one of two qualitatively different regimes: a saturated one where almost all the enzyme is bound to its substrate, and an unsaturated one ([46, 148]). The regime of the reaction depends on the relative concentrations of a substrate and the enzyme (E), and on the MM constant (K) of the enzymatic reaction. The unsaturated (first-order) regime, where the rate of reaction is linearly proportional to the substrate concentration, occurs when the substrate is much less abundant than the sum of the MM constant of the reaction and the enzyme concentration (e.g., for the second reaction, $K_2 + \overline{E_2} \gg \overline{A}$). In the saturated (zero-order) regime, the rate of reaction is almost independent of the substrate concentration and is proportional to the enzyme concentration. This occurs when the substrate is much more abundant than the sum of enzyme concentration and its MM constant (e.g., for the second reaction, $K_2 + \overline{E_2} \ll \overline{A}$).

Since the signaling cycle is built of two enzymatic reactions, it can exhibit four

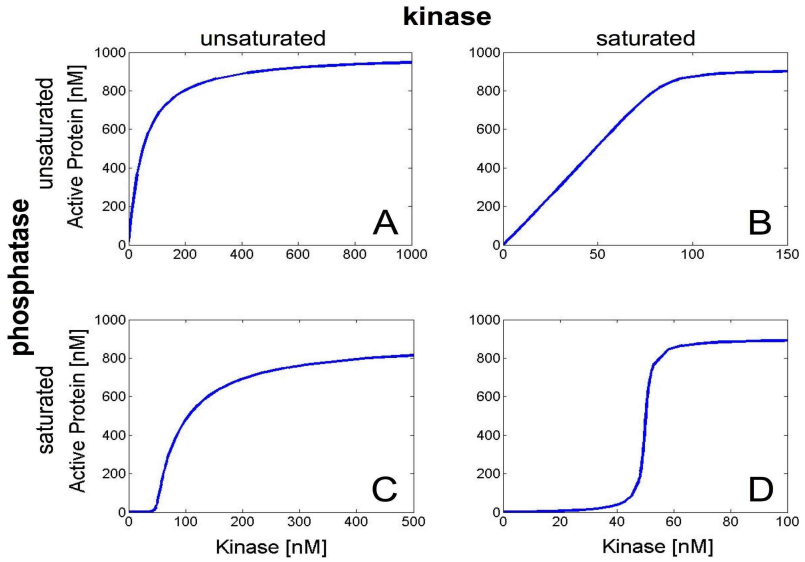

Figure 3-2: **Steady state behavior of the four cycle regimes. A** When both enzymes are unsaturated the steady state response is hyperbolic. The parameters used for this cycle are $\overline{S} = 1000$, $a_1 = 1$, $K_1 = 10000$, $a_2 = 1$, $\overline{E_2} = 50$, $K_2 = 10000$, $k_1 = 1$ and $k_2 = 1$ where all reaction rates are in units of $1/sec$, concentrations and Michaelis constants are in nM, and second order reaction rates (a_1 and a_2 are in $1/nMsec$. **B** When the kinase is saturated and the phosphatase unsaturated a linear response results. The parameters here are $\overline{S} = 1000$, $a_1 = 100$, $K_1 = 10$, $a_2 = 1$, $\overline{E_2} = 50$, $K_2 = 10000$, $k_1 = 500$ and $k_2 = 10000$. **C** When the kinase is unsaturated and the phosphatase saturated a threshold-hyperbolic response results. The parameters for this cycle are $\overline{S} = 1000$, $a_1 = 100$, $K_1 = 10000$, $a_2 = 100$, $\overline{E_2} = 100$, $K_2 = 1$, $k_1 = 25$ and $k_2 = 1$. **D** When both enzymes are saturated an ultrasensitive response results. The parameters used for this cycle are $\overline{S} = 1000$, $a_1 = 100$, $K_1 = 10$, $a_2 = 100$, $\overline{E_2} = 50$, $K_2 = 10$, $k_1 = 1$ and $k_2 = 1$. The parameters for the four cycles were chosen to be comparable in magnitude to values found in the literature (see [78, 84], for example).

regimes of signaling (see Fig. 3-2), corresponding to the two regimes of each reaction. The conditions for each of the four regimes are summarized in Table 3.1. The steady-state behavior of two of the four regimes (when the kinase and the phosphatase are either both saturated or both unsaturated, referred to as ultrasensitive and hyperbolic, respectively) has been characterized earlier by Goldbeter and Koshland ([61]). Using tQSSA, we are able to refine the range of parameter values for which these behaviors hold. The other two regimes have not been identified before, to the best of our knowledge.

Kinase Phosphatase	Unsaturated $K_1 + \overline{E}_1 \gg \overline{S} - \overline{A}$	Saturated $K_1 + \overline{E}_1 \ll \overline{S} - \overline{A}$
Unsaturated $K_2 + \overline{E}_2 \gg \overline{A}$	hyperbolic	signal transducing
Saturated $K_2 + \overline{E}_2 \ll \overline{A}$	threshold hyperbolic	ultrasensitive

Table 3.1: Conditions for the four cycle regimes.

3.3.3 Steady-State Response

Hyperbolic (unsaturated kinase, unsaturated phosphatase)

In this regime, the cycle exhibits a hyperbolic steady-state response that saturates at the value provided in Table 3.2 (see Fig. 3-2A). Using the tQSSA, we find that the hyperbolic regime requires weaker conditions than previously thought ($K_2 + \overline{E_2} \gg \overline{A}$ and $K_1 + \overline{E_1} \gg \overline{S} - \overline{A}$ instead of $K_2 \gg \overline{A}$ and $K_1 \gg \overline{S} - \overline{A}$).

Regime	Threshold for input (E_1)	Saturation level
hyperbolic	–	$\frac{k_1}{k_1+w_2}(1 - \frac{w_2}{k_2})\overline{S}$
signal-transducing	–	$(1 - \frac{\omega_2}{k_2})(\frac{k_1/\omega_2}{1+k_1/\omega_2})\overline{S}$
threshold-hyperbolic	$K_1/((k_1/k_2 - 1)(\overline{S}/\overline{E_2} - 1))$	$\overline{S} - \overline{E_2}(1 + \frac{k_2}{k_1})$
ultrasensitive	$\overline{E_2}k_2/k_1$	$\overline{S} - \overline{E_2}(1 + \frac{k_2}{k_1})$

Table 3.2: Expressions for threshold and saturation levels for steady-state regimes of the cycle. Here $\omega_2 = k_2 \frac{\overline{E_2}}{K_2+\overline{E_2}}$ is the characteristic frequency of the phosphatase.

The recent study in [93] suggests that the hyperbolic regime is much more robust to fluctuations and to cell-to-cell variability in kinase and phosphatase concentrations than the ultrasensitive regime, which requires fine-tuning of the threshold level. The hyperbolic regime transmits signals in a broad range of amplitudes, requiring no tuning of cycle parameters.

Signal-Transducing (saturated kinase, unsaturated phosphatase)

We refer to this new regime as signal-transducing because, as discussed below, it is ideal for transmitting time-varying signals without distortion, while attenuating

higher frequency noise. Here we only point out that its steady-state response is linear with a slope (gain) of k_1/ω_2, where ω_2 is referred to as the effective phosphatase frequency (see Table 3.2 and section on Dynamic Response below), until it reaches saturation (Fig. 3-2B, Table 3.2). Having a linear steady-state response, a property unique to this regime, is potentially desirable for signaling that involves graded stimuli. Available biochemical data and in vivo measurements argue for the possibility of this regime being present as a component in cell signaling cascades (see Discussion).

Threshold-Hyperbolic (unsaturated kinase, saturated phosphatase)

In this new regime, the output below a given input threshold is zero, and then increases hyperbolically until it reaches its saturation level (approximated by the same expression as the saturation level of an ultrasensitive regime). Figure 3-2C shows the steady-state response of such a cycle.

Ultrasensitive (saturated kinase, saturated phosphatase)

The output in this regime is close to zero for inputs below a threshold, and increases rapidly to a saturation value, consistent with the results obtained in [61] using the MM approximation. Such highly sigmoidal behavior effectively quantizes the signal (see Fig. 3-2D). This regime was termed ultrasensitive because, when the input is close to the threshold, small input changes result in large changes of the steady-state output. Interestingly, cells may adjust the threshold of this cycle by changes in phosphatase level, $\overline{E_2}$.

The MM approximation fails, however, when the amount of enzyme becomes comparable to that of its substrate. Using the tQSSA we are able to refine the range of parameter values required for ultrasensitive signaling. The criteria for ultrasensitivity obtained from the MM model ([61]), namely $K_2 \ll \overline{A}$ and $K_1 \ll \overline{S} - \overline{A}$, are actually not sufficient conditions for the cycle to be ultrasensitive; instead we need $K_2 + \overline{E_2} \ll \overline{A}$ and $K_1 + \overline{E_1} \ll \overline{S} - \overline{A}$. When the enzyme concentrations become comparable to those of their substrates, there is no ultrasensitivity, as noted recently by Bluthgen *et al.* in [14] by more complicated arguments.

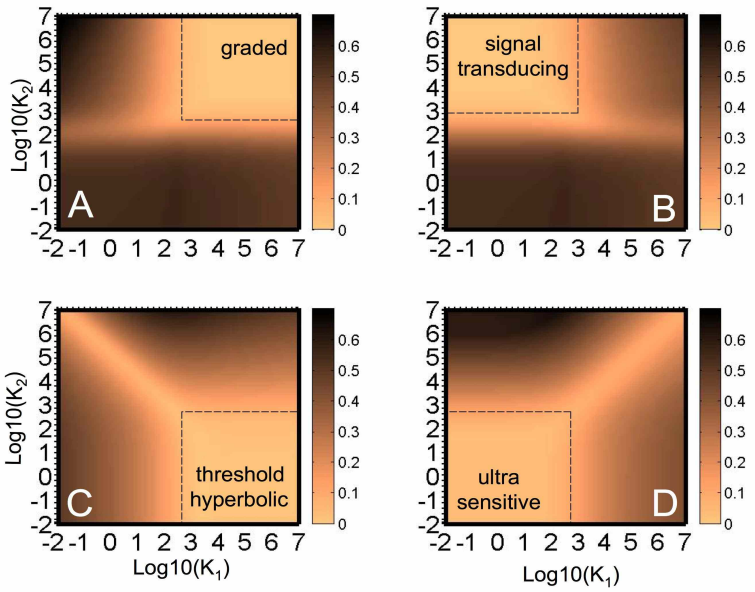

Figure 3-3: **Relative error.** Subfigures **A, B, C** and **D** respectively show the relative error between the steady-state characteristic of the hyperbolic, signal-transducing, threshold-hyperbolic and ultrasensitive regimes, and that of the tQSSA in Equation 3.3. To compute the error for a regime we first approximated the average squared difference between the regime's steady state and that of Equation 3.3, and then divided its square root by the total substrate S_t (see Section 3.5.5 for more details). A relative error of .1 then corresponds to an average absolute difference between the steady state characteristic of the regime and that of Equation 3.3 of $.1S_t$. The figures here show that the relative error for each regime is small for a wide region of the K_1 versus K_2 space, demonstrating that the four regimes cover almost the full space. The parameters used for this cycle are the same as those in Figure 3-2D, except K_1 and K_2 which were varied in the range of values shown in the x and y axes in this figure. The dashed lines enclose the regions where each regime is expected to describe the system well.

Incorporating Stochastic Effects

Considering stochastic effects due to low molecule numbers or small volumes, at least at steady state, still results in the same four regimes and their corresponding behaviors. Analogously to the mass action model introduced earlier, the signaling cycle can be described by a fully probabilistic model with a time-varying probability distribution governed by the master equation (see Chapters 4, 5 and 6) of the reactions

in Equations 3.1 and 3.2. Separation of time scales (see Chapter 6) in this probabilistic model, under appropriate conditions, results in the following approximate model for the probability α_i ($i \in \{1, \ldots, \overline{S}\Omega\}$) that the number of molecules of $\overline{A}$ equals i:

$$\frac{d}{dt}\alpha_i = \alpha_{i-1} f_{i-1} + \alpha_{i+1} b_{i+1} - \alpha_i (f_i + b_i). \tag{3.4}$$

Here f_i and g_i are functions that respectively correspond to the probability per unit time that a forward or a backward enzymatic reaction occur given that the number of $\overline{A}$ molecules is i. These functions can be shown to be approximately given by (see Chapter 6):

$$f_i = k_1 \Omega \frac{\overline{E_1}\left(\overline{S} - \overline{A}\right)}{K_1 + \overline{E_1} + \overline{S} - \overline{A}} \text{ and } b_i = k_2 \Omega \frac{\overline{E_2}\,\overline{A}}{K_2 + \overline{E_2} + \overline{A}},$$

where Ω equals the system volume times Avogadro's number, and $\overline{A} = i/\Omega$.

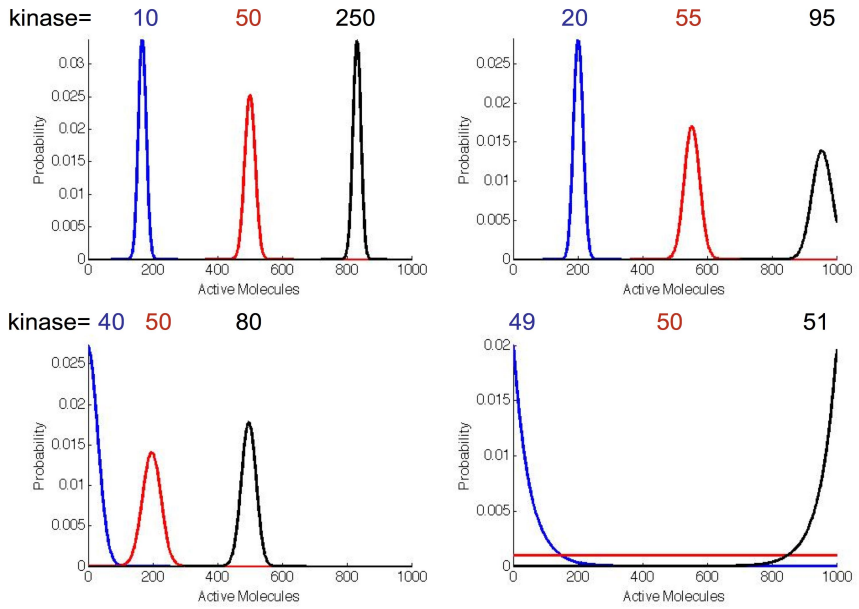

Figure 3-4: Steady-state distributions for the four cycles in Figure 3-2 for three steady-state kinase levels.

Equation 3.4 is the forward-Kolmogorov equation of a birth-death continuous-time discrete-state Markov chain (e.g., see [49]). At steady-state, the probability

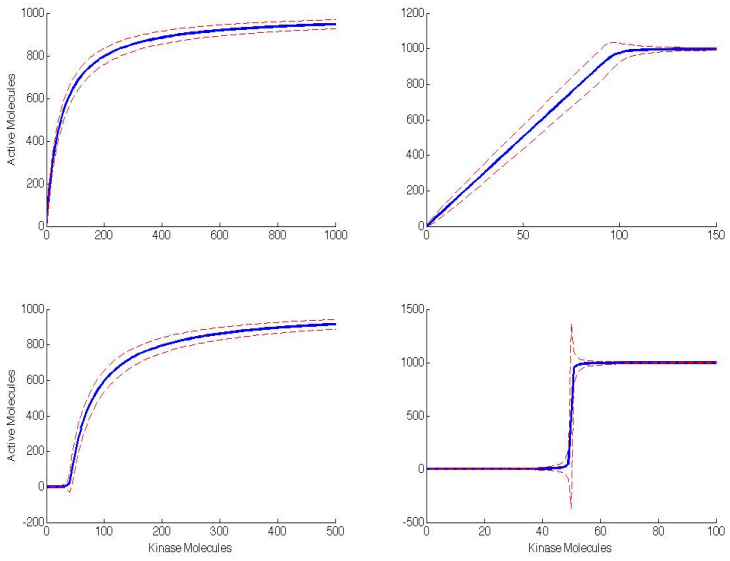

Figure 3-5: Steady state behavior of the four cycle regimes in the presence of stochasticity. The blue and red lines corresponds to the mean, and mean +/- three standard deviations of the steady-state distribution. The same parameters as in Figure 3-2 were used, along with $\Omega = 1$.

distribution for these Markov chains may be shown to be given by:

$$\alpha_i = \left(\prod_{j=0}^{i-1} \frac{f_j}{b_{j+1}}\right)\alpha_0, \text{ where } \sum_i \alpha_i = 1. \tag{3.5}$$

The functions f_i and g_i simplify in each of the four regimes depending on whether the enzymatic reactions are operating in the saturated or in the linear regime. The equation above can then be used to obtain the steady-state distribution of the four cycle regimes, from which the mean and standard deviations of active molecules can be computed. The resulting steady-state distributions for the four cycle regimes are specified in Table 3.3. Figure 3-4 shows the steady-state distributions for the four cycle regimes at three steady-state kinase levels. Figure 3-5 shows the mean number of active substrate molecules, along with the three-standard-deviation envelopes, for the same cycles shown in Figure 3-2. Chapter 5 shows that the fluctuations around the mean become proportionally greater as the number of molecules decreases. Here assumed the total number of substrate molecules is 1,000, and fluctuations do not seem

significant, except perhaps in the ultrasensitive regime. However, several proteins may be present at significantly smaller numbers and fluctuations can become quite significant.

Regime	$\alpha_i \propto$	Steady-State Distribution Type
hyperbolic	$\binom{N}{i} p^i (1-p)^{N-i}$	binomial
signal-transducing	$\left(\frac{\Omega k_1 \overline{E_1}}{b}\right)^i / i!$	truncated Poisson
threshold-hyperbolic	$\binom{N}{i} \left(\frac{a}{\Omega k_2 \overline{E_2}}\right)^i i!$	unknown
ultrasensitive	$\left(\frac{k_1 \overline{E_1}}{k_2 \overline{E_2}}\right)^i$	truncated geometric

Table 3.3: Steady-state distributions of active substrate, perhaps off by a normalization constant. N is the total number of substrate molecules $\overline{S}\Omega$, $a = \frac{k_1 \overline{E_1}}{K_1 + \overline{E_1}}$, $b = \frac{k_2 \overline{E_2}}{K_2 + \overline{E_2}}$ and $p = \frac{a}{a+b}$.

In summary, we have demonstrated that a signaling cycle can operate in four regimes that have qualitatively different steady-state responses to kinase activation. Of the newly identified regimes, the signal-transducing regime is a good candidate for sensing stimuli, when a graded and undistorted response is required. Depending on the slope of its response, which is controlled by parameters of the cycle and can be easily adjusted by the cell to a required level, the input signal may be amplified or diminished. We consider factors influencing the choice of the regime for natural signaling cycles in different cellular processes in the Discussion.

The four regimes we consider, although obtained only at extreme parameter values, are actually quite descriptive of the system for a wide range of parameters, and naturally partition the space of possible steady-state behaviors of the signaling cycle into quadrants, as shown Fig. 3-3. This figure shows the relative error between the steady-state characteristic of each of the four regimes and that of Eq.3.3 for a wide range of kinase and phosphatase MM constants (see Section 3.5.5). It reveals that

the regime approximations are quite good at a wide range of values of MM constant (for example, the region with a relative error of less than 10% for each regime covers almost a full quadrant in the plots), and not only when the MM constants take the very large or very small values required in the regime definitions. This demonstrates that these four regimes, though defined by extreme values of system parameters, actually encompass the full space of cycle behaviors.

Understanding the steady-state response of the cycle is informative, but is only part of the story; signaling cycles do not necessarily transmit steady inputs but rather deal with time-dependent signals that reflect changing environmental conditions.

3.3.4 Dynamic Response

Signaling cascades in the cell are activated by receptors, which in turn get activated by ligand binding and inactivated by internalization and other mechanisms. All of these mechanisms produce time-varying signals, and are subjected to noise (i.e., rapid and stochastic fluctuations) due to small numbers of molecules, diffusion, and other effects. How can a cell extract a time-varying signal from noisy stimuli?

Response to signals of various frequencies: low-pass filtering

To address this question, we first study the response of the four regimes to time-varying stimuli. A high-frequency signal is a proxy for the noise in the signal, so understanding how the cycle responds to high frequencies is essential for understanding its response to noise.

We studied the cycle's response to oscillating kinase levels at different frequencies and amplitudes: $\overline{E_1}(t) = E_0(1 + a \sin \omega t)$. This is not to say that sinusoidal inputs need be biologically relevant, but systematically understanding the response to such inputs gives one intuition about the response to more general inputs. Furthermore, for small enough input variations around some background baseline level, the cycle's behavior is, to a first-order approximation, linear and governed by time-independent parameters; in this situation the response to sinusoids determines the response to

arbitrary inputs. In fact, in the signal-transducing regime the dynamic response of the cycle (just as its static response) is linear for all non-saturating inputs, without the restriction to small variations.

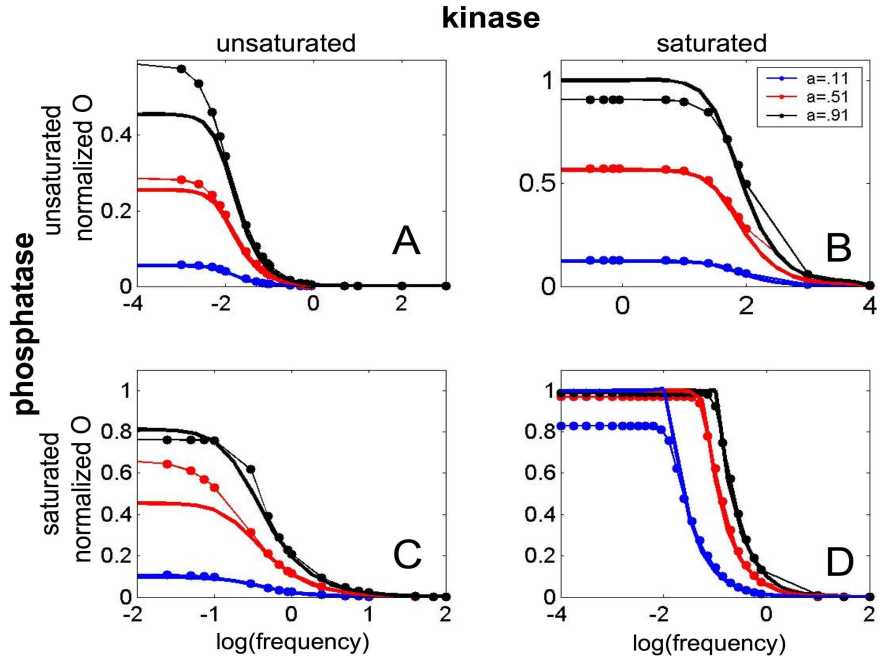

Figure 3-6: **Magnitude of the response of the cycle O (normalized by the steady-state saturation value) versus the input frequency ω, for three different input amplitudes a.** The traces in **A, B, C** and **D** show the response of the hyperbolic, signal transducing, threshold-hyperbolic and ultrasensitive switches, respectively, shown in Figure 3-2. The solid lines are the analytical approximation (Equation 3.6). The dotted lines are obtained from numerical simulation of the full system.

Figure 3-6 shows the amplitude O of the variations in the output (normalized by the steady-state saturation value of the cycle), obtained by numerical simulation for three values of a, and as a function of input frequency ω. Invariably, the response is flat and high at low frequencies, but starts to decrease after a particular frequency is reached. These results are very well described in the case of the smallest a (corresponding to 11% deviations) by the expression obtained analytically using small-signal approximations (see Section E of Supporting Information):

$$O = \left(\frac{g}{\sqrt{\omega^2 + \omega_c^2}} \right) E_0 a, \tag{3.6}$$

where E_0 is the background kinase level, and where the gain g and the cutoff frequency ω_c are functions of the cycle parameters that are different for the four regimes (see Table 3.4). The analytical approximation continues to hold quite well even for larger values of a, the deviation amplitude (up to 91% of the baseline for the results in Figure 3-6). For frequencies much smaller than the cutoff, the amplitude of the output variations is constant and proportional to the ratio of gain to cutoff frequency. For frequencies above the cutoff, the output variations have an amplitude that decays as $1/\omega$. Figure S1 presents more detailed results on the variation of O as a function of both a and ω, again obtained by numerical simulations.

Regime	Gain g	Cutoff Frequency ω_c
hyperbolic	$\frac{\overline{S}K_1}{E_0(K_1+E_0)}\frac{\omega_2}{\omega_2+\omega_1}(1-\frac{\omega_2}{k_2})\omega_1$	$\omega_1+\omega_2$
signal-transducing	$(1-\frac{\omega_2}{k_2})k_1$	ω_2
threshold-hyperbolic	$\frac{\overline{E_2}K_1}{E_0(K_1+E_0)}k_2$	ω_1
ultrasensitive	k_1	$2k_1a\frac{E_0}{\overline{S}-\overline{E_2}(1+k_2/k_1)}$

Table 3.4: Expressions for gain and cutoff frequency for four regimes of the cycle (in response to the input $\overline{E_1}(t) = E_0(1+a\sin\omega t)$. Here ω_2 is the characteristic frequency of the phosphatase, defined in Table 3.2; while $\omega_1 = k_1\frac{E_0}{K_1+E_0}$ is the characteristic frequency of the kinase.

An essential property of this signaling low-pass filter is that the cutoff frequency ω_c can be easily adjusted by varying enzymatic parameters and concentrations of the kinase and the phosphatase. Although all four regimes act as low-pass filters, their cut-off frequencies ω_c and gains g depend differently on the cycle parameters (see Table 3.4 and Fig. 3-6).

Importantly, for the two newly characterized regimes (the signal-transducing and threshold-hyperbolic), the gain and the cutoff frequency can be adjusted independently, thus allowing greater flexibility to the signaling requirements of individual signaling pathways (Table 3.4 and Fig. 3-6).The gain and the cut-off frequency for three of the regimes are independent of the input parameters a and ω; the exception is the cutoff frequency for the ultrasensitive regime, which depends on the input amplitude.

It is easy to understand the origin of the low-pass filtering behavior. First consider a cycle subjected to a slowly varying input (Fig.3-7): If the input changes so slowly that the cycle has enough time to reach its steady-state level before the kinase level changes by a significant amount, the cycle simply tracks the kinase level as a function of time through its steady-state response curve, characteristic for its operational regime. Now consider a rapidly changing input. Since the kinase level changes faster, the cycle has less time to adjust to its steady state corresponding to the new value of the input before the kinase level changes again. Thus the output will not be able to reach its full amplitude before the kinase levels change again in the opposite direction, and the amplitude of the output is thus decreased (see Fig.3-7). As the signal changes faster and faster, the amplitude of the output will decrease, until the kinase levels vary so fast that the cycle simply does not respond.

The response of the cycle thus depends on the two time-scales: the duration of the stimulus $\tau = 1/\omega$ and the intrinsic switching time of the cycle $\tau_c = 1/\omega_c$. If the stimulation is longer than the switching time, $\tau \gg \tau_c$, then the cycle will adjust its response by $2aE_0g/\omega_c$. On the other hand, a shorter, transient stimulus $\tau \ll \tau_c$ is not likely to activate the cascade.

Interestingly, ligands activate a kinase by binding to it. The results here imply that weak ligands binding for a time interval shorter than τ_c are unlikely to produce any down-stream activation of the pathway, while those that stay bound longer than τ_c activate the pathway. Low-pass filtering can thus perhaps make a signaling cascade more selective to higher-affinity ligands.

Response to a noisy signal

Importantly, low-frequency inputs are proxies for longer input activation, while high-frequency inputs are proxies for short, transient activations of the cascade and for high-frequency noise. Because of low-pass filtering, cycles respond to noise less than to signals, and as the noise shifts to higher frequencies, the cycle responds to it less. Figure 3-8 makes the point more precisely: it shows the response of the cycle to a slowly varying signal buried in noise, and demonstrates that the noise is filtered out

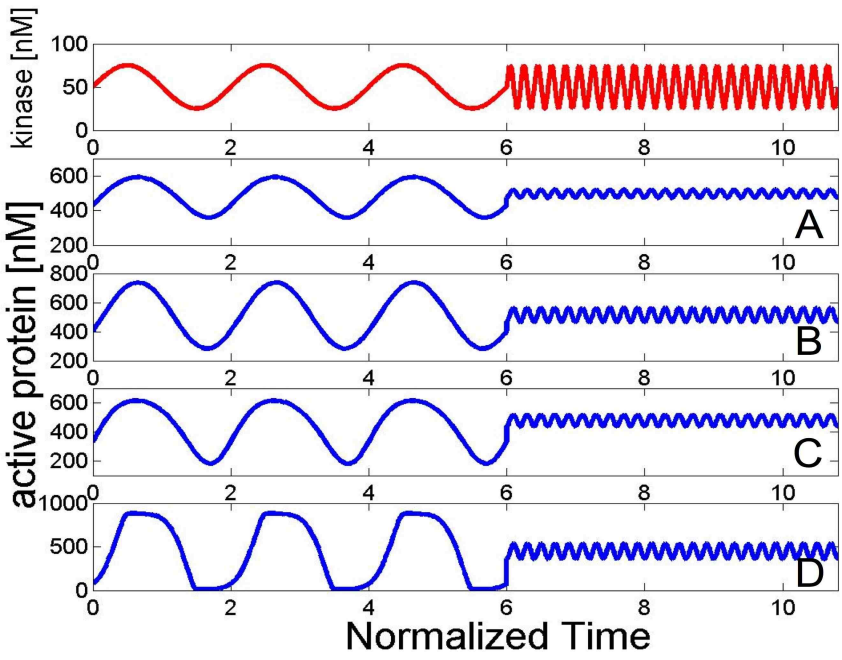

Figure 3-7: **Dynamic response of the cycles to fast and slow inputs.** The cycle has a characteristic response time τ_c that is a function of its parameters (see Section Dynamic Response), and which is different for all four regimes. This plot shows the response of all four regimes to (i) a slow input that has a period equal to twice the characteristic response time of the cycle followed by (ii) a fast input with a period equal to one fifth of the cycle's response time. For clarity, time was normalized by dividing by the characteristic time of each cycle. The signal in red represents the input kinase levels (for the threshold-hyperbolic switch the input used is actually twice the red signal) while the blue traces in **A, B, C** and **D** show the response of the hyperbolic, signal transducing, threshold-hyperbolic and ultrasensitive switches, respectively, shown in Figure 3-2.

and the signal is revealed.

In summary, analysis of dynamic response demonstrates that (i) the cycle acts as a low-pass filter in all four regimes; (ii) the cutoff frequency and the gain of signaling can be adjusted by the cell to achieve better performance (independently of each other in the case of the signal-transducing and the threshold-hyperbolic cycles); and (iii) low-pass filtering makes signaling cascades insensitive to noise and transient activations. Below we discuss some biological implications of these findings.

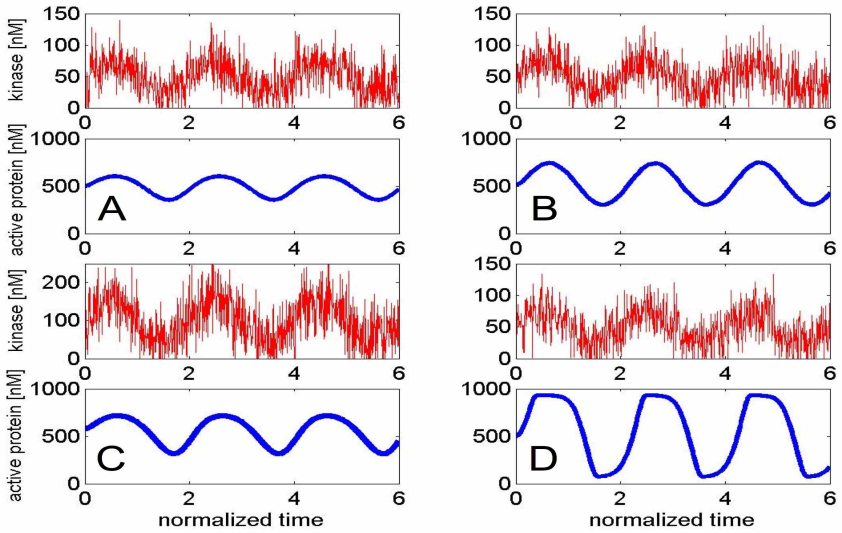

Figure 3-8: **Response of the four cycles to the input buried in noise.** The input is a sum of a slow signal (same as in Fig.3-7) and a Gaussian uncorrelated noise. The resulting input signals are shown in red. The blue traces in **A, B, C** and **D** show the response of the hyperbolic, signal transducing, threshold-hyperbolic and ultrasensitive switches, respectively, as shown in Figure 3-2. The response shows that the cycles respond to the signal only and ignore or filter out the noise in the input. Time was normalized by the characteristic time of each cycle to facilitate comparison amongst cycles.

3.3.5 Low-pass filtering

Figure 3-9 summarizes the low-pass filtering behavior for the four regimes. It shows O (color coded) versus a and ω (i.e., a horizontal cut through this plot would simply be an O versus ω curve such as those shown in Fig. 3-6). Some salient features are evident in this figure. (i) All regimes act as low-pass filters. (ii) Although Equation 3.6 is obtained using a small signal approximation and is expected to hold for small a, it provides a good guide for describing O for all values of a. Perhaps the biggest discrepancy is the fact that O does not increase linearly with a but saturates (see Fig. 3-9A,C,D). The signal-transducing regime, however, does seem to have a response that increases linearly with a so for a given frequency an input with twice the amplitude of another will result in twice the output O. (iii) Finally, for all regimes except the ultrasensitive regime (Fig. 3-9) the response starts decreasing at about the same frequency, independent of a, where as for the ultrasensitive cycle smaller a results in

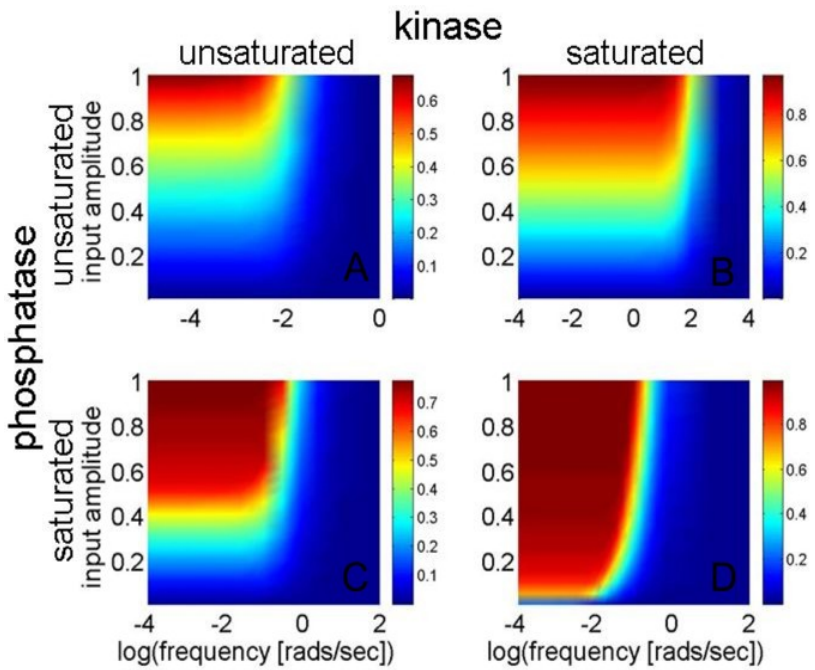

Figure 3-9: **Magnitude of the oscillations in the output as a response to oscillations in the input about a background kinase level.** Plots **A, B, C** and **D** show the output oscillations O of the hyperbolic, signal transducing, threshold-hyperbolic and ultrasensitive switches, respectively (normalized by the steady-state saturation value of each cycle), shown in Figure 3-2, in response to an input of the form $\overline{E_1} = E_0(1 + a \sin wt)$. The magnitude of O is color coded and shown as a function of the input amplitude a and frequency ω. Output oscillations increase with increasing a and decrease with increasing ω as expected. The four cycles, however, respond very differently to their inputs. The parameters used for the cycles are the same as those in Figure 3-2, and $\overline{E_2} = 50\ nM$ except for the threshold-hyperbolic switch, where $\overline{E_2} = 100\ nM$.

smaller cut-off frequency.

The ultrasensitive regime

The ultrasensitive cycle is the only cycle that oscillates between a level close to its saturation value and zero for a wide range of inputs (red region of Fig. 3-9). The cutoff frequency of this cycle is a function of both the total substrate protein of the cycle, and of the input's amplitude, a unique property of this cycle. For all other cycles, the cutoff frequency is independent of total substrate and of the input parameters.

The signal-transducing regime

The cutoff frequency depends only on the phosphatase level and phosphatase parameters, but the gain depends on both phosphatase and kinase parameters. Thus the cutoff frequency can be tuned by changing phosphatase parameters or level, and the gain can be independently adjusted by changing the catalytic rate of the kinase via evolution.

The threshold hyperbolic regime

The cutoff frequency depends only on the average kinase level and kinase parameters. Thus the cutoff frequency may be tuned by changing the kinase level and/or parameters, and the gain may be tuned independently by adjusting phosphatase levels or parameters.

The hyperbolic regime

This cycle has a cutoff frequency that depends on both the kinase and the phosphatase. Increasing either one increases the cutoff frequency. The gain also depends on both the kinase and the phosphatase, so adjusting their levels will modify both the gain and the cutoff frequency of the switch.

3.3.6 Incorporating Stochastic Effects

The effects of low molecule numbers can be incorporated into a model of the signaling cycle, resulting in the probabilistic description in (3.4), where α_i ($i \in \{0, 1, \ldots, T = \Omega\overline{A}\}$) is the probability of having i active substrate molecules. This equation can be solved at steady-state, as in (3.5), to yield the stationary distribution of active substrate molecules. For the signal-transducing regime we can obtain an approximate form of the distribution for the transient before steady state. The main result is that when the initial distribution is Poisson, then as long as the support of the distribution remains outside of the saturation region (i.e., as long as $\alpha_{S_t} = 0$), the time-varying

distribution is Poisson with a mean that satisfies the macroscopic equation. We briefly outline the derivation of this result below.

For the signal transducing regime, $f_i = k_1 \Omega \overline{E}_1 = f$ and $b_i = \overbrace{k_2 \frac{\overline{E}_2}{K_2 + \overline{E}_2}}^{=b} \overbrace{\Omega \overline{A}}^{=i} = bi$. We start by deriving the rate of change of the mean number of active total substrate molecules $\mu(t)$. We multiply (3.4) by i and sum over all $i \in \{0, \ldots, T\}$ to get

$$\frac{d}{dt}\mu = f(1 - \alpha_T) - b\mu \ , \tag{3.7}$$

which is almost like the macroscopic equation, except for the $(1 - \alpha_T)$ term.

Now suppose that at all times, including the initial time, the distribution α_i is a truncated Poisson with mean $\mu(t)$, i.e., that $\alpha_i(t) = \frac{\mu(t)^i}{i!} c(t)$ for $i \in \{0, \ldots, T\}$ and zero otherwise, where $c(t) = \left(\sum_{j=0}^{T} \frac{\mu(t)^j}{j!} \right)^{-1}$ is a normalizing constant. The time-derivative of $c(t)$ is then

$$\frac{d}{dt}c(t) = -c^2(t) \left(\sum_{j=1}^{T} \frac{\mu(t)^{j-1}}{(j-1)!)} \right) \frac{d}{dt}\mu(t). \tag{3.8}$$

Taking the time-derivative of the previous definition of α_i gives

$$\begin{aligned} \frac{d}{dt}\alpha_i &= \frac{\mu(t)^{i-1}}{(i-1)!} c(t) \frac{d}{dt}\mu(t) + \frac{\mu(t)^i}{i!} \frac{d}{dt} c(t) \\ &= f\alpha_{i-1} - f\alpha_i + \alpha_{i+1} b_{i+1} - \alpha_i b_i + f\alpha_T (2\alpha_i - \alpha_i \alpha_T - \alpha_{i-1}). \end{aligned} \tag{3.9}$$

This equation equals (3.4) when $\alpha_T = 0$, which shows that the truncated Poission distribution is a solution of (3.4) when the distribution at time $t = 0$ is a truncated Poisson, and when the probability that all substrate molecules are active remains at zero throughout the experiment. Equivalently, as the number of substrate molecules T goes to infinity then $c(t)$ becomes $e^{-\mu(t)}$, and the distribution for α_i becomes (an un-truncated) Poisson with mean $\mu(t)$ that satisfies (3.7) with α_T set to zero.

3.4 Discussion

Significant effort has been put in the elucidation and characterization of signaling cascades and pathways (see e.g., [140, 134, 124, 85] for reviews). When put together, these pathways form an intricate network of cell signaling, where each node in the network corresponds to a different chemical species. Because of the complexity of the network, it is natural to split it into interconnected modules (sets of nodes whose output depends only on its input and not on the network downstream of it) and analyze possible behaviors arising from different interconnections of modules (see e.g. [67, 72, 21]).

What constitutes a module in the network, however, is still hard to define, and significant efforts are directed at tackling this problem (e.g. [75, 132, 133, 114, 104]). What constitutes a good general representation for an arbitrary module in the network is also an open question. Other efforts have been aimed at understanding properties of the network as a whole, such as identifying the number of equilibrium states (e.g. [4, 25]).

Using a deterministic model, we have attempted to provide a systems-level input/output understanding of the signaling cycle, ubiquitous in signaling pathways. After identifying four parameter regimes (two of them not reported before, to our knowledge), their steady-state and dynamic behaviors were analyzed and numerically verified. The results indicate that cycles act as low-pass filters, and that each regime may be useful under different circumstances. Given the values for cycle parameters, one can use our results to determine the regime in which the cycle operates. Unfortunately, the scarcity of parameter values makes it hard to assess which of these regimes is more widely present in signaling pathways. The low-pass filtering behavior of the cycle demonstrates that inputs of the same magnitude but changing at different speeds may produce very different outputs, which argues in favor of studying the dynamical properties of signaling pathways.

All physical systems stop responding to fast enough inputs, but what makes the low-pass filtering behavior of the signaling cycle interesting is that it is first-order,

with a single cutoff frequency, and that the cutoff frequency can be adjusted by evolution (through changes in the enzymatic catalytic rates) and by the cell (through changes in gene expression). As such, the signaling cycle is a versatile module with simple dynamics that can be easily tuned for various noise filtering needs and used to construct signaling networks with more complicated functions and dynamics.

Of the two newly identified regimes, the signal transducing one is of particular interest because it appears ideal to transmit time-varying intracellular signals without distortion, while filtering out high-frequency noise in the input. Furthermore, because it is linear, it opens the possibility that at least parts of signaling pathways (those built of signal-transducing signaling cycles, or other yet unidentified linear signaling motifs) may be amenable to linear system analysis, a powerful set of tools to understand the properties of arbitrary network structures and motifs (for example, elucidating the roles of cascades, positive and negative feedbacks, etc.). If naturally occurring cycles operate in the signal-transducing regime, then analyzing networks built of these cycles becomes tractable as long as load effects can be neglected.

Can naturally occurring signaling cycles operate in this regime? While it was demonstrated that certain kinases in *S. cerevisiae* and *Xenopous* operate in saturation (with MM constant of $\sim$ 5nM and substrate concentrations of $\sim 30 - 100$nM for yeast; see [42, 8]), little is known about phosphatases. To explore the possibility that known signaling pathways operate in the signal-transducing regime, we manually collected values of MM constants from the biochemical literature. Then we used data for intracellular protein concentrations measured using GFP-tagged proteins ([53]). Phosphatases seem to have a broad specificity, with a relatively wide range of MM constants (e.g., 5 to 90μM for the PP2C phosphatases), and appear to be present in large concentrations (e.g., [Ptc1]$\approx$ 1520 molecules per cell, so $\overline{E_2} \approx 0.025\mu$M, while [Ptc2-3]$\approx$ 15000, so $\overline{E_2} \approx 0.25\mu$M, assuming a yeast cell volume of 0.1pl [42]). Data on singly-phosphorylated substrates is hard to find, but for a rough indication consider the doubly phosphorylated protein Pbs2 of *S. cerevisiae* as an example. Pbs2 is measured to have about 2000 molecules per cell so that $\overline{S} \approx 0.03\mu$M$=$ 30nM. If singly phosphorylated proteins were characterized by similar numbers, then their

phosphatases could potentially be unsaturated, since $\overline{A} < \overline{S} \ll K_2 + \overline{E}_2$. In contrast, kinases that act on Pbs2 are present at lower concentrations (e.g., [Ste11]= 736, [Ssk2]= 217, and [Ssk22]= 57 molecules per cell, or $\overline{E_1} \approx 1 - 3$nM). Such concentrations are consistent with kinases operating in saturation, since $\overline{E_1} + K_1 < \overline{S}$ (assuming K_1 is in the same range as those measured for Ste7, $K_1 \sim 5$nM). Taken together, these numbers suggest that the possibility of a signaling cycle operating in the signal-transducing regime.

Different signaling cycles, however, may be operating in different regimes, raising two questions: First, which regime has been chosen by evolution for a given signaling application? Second, what are the advantages and disadvantages of each regime for a cycle in a particular position for a particular signaling pathway? To answer the first question, we need to know *in vivo* concentrations and MM constants of involved enzymes. Unfortunately these data are either unavailable or scattered through publications in the biochemical literature. The applicability of MM constants measured *in vitro* is also questionable. An alternative experimental approach would be to obtain steady-state response curves for various cycles. The success of such measurements depends on and hence is limited by the availability of *in vivo* single cell probes for the phosphorylation state of a particular protein.

The second question, on advantages and disadvantages of each regime, can be addressed by systematic analysis of cycle properties: steady-state and dynamic response, robustness to fluctuations, etc. By matching these characteristics against the requirements of a particular signaling system, one can suggest the optimal regime for each signaling application. For example, one can think that signaling in retina cells shall be fast and graded, depending on the intensity of adsorbed light. Similarly, gradient sensing in motile cells has to provide graded responses on the time-scales required to change direction of motion. On the other hand, signaling of cell fate determining stimuli and signaling involved in various developmental processes may require an ultrasensitive ("on/off") response, while imposing much softer constraints on the time it takes to switch the system from off to on state (hours instead of the milliseconds needed in light-sensing). The performance of the signaling regimes in

the context of cascades and feedbacks is also important for understanding the rules that govern the choice of a regime for each cycle.

For cycles in signaling applications involving all-or-none decisions, such as differentiation, apoptosis, or the cell cycle, it has been argued that ultrasensitive cycles may be useful as they effectively generate a discrete output that is either high or low ([44]). When such a cycle is tuned appropriately (such that in the presence of the background input it is close to its threshold; see [93]), it is the best cycle at recovering time-dependent signals buried in noise, because its gain for low-frequency inputs is the highest among the regimes. Therefore an ultrasensitive cycle is desirable when the input signals are extremely noisy, and/or have to achieve binary level outputs.

A signal-transducing cycle, on the other hand, is the best choice to transmit time-dependent signals without distortion because its output is approximately a scaled but otherwise undistorted copy of low-frequency input signals, while noisy input components are filtered out. It is the only cycle that does not distort the input. What the other two regimes might be best at is not clear. The threshold-hyperbolic cycle, however, may prove useful in situations when no activation is desirable below a given input strength, and when a graded response is desired for inputs above this threshold.

We here considered the effect of temporal noise in kinase levels on the response of the signaling cycle. A more detailed model should also take into account the intrinsic noise coming from the cycle itself, since it consists of chemical reactions where the number of molecules per species is small, and thus a deterministic model based on mass action kinetics may be inadequate. For example, although the deterministic cycle is known to have a single steady-state solution, Samoilov *et al.* (see [123]) found that treating the cycle stochastically can give rise to a bimodal distribution for the phosphorylated protein. The "mass fluctuation kinetics" approach described in [62] may be useful in this regard; see also [65, 48]. Other sources of noise that should also be taken into account are fluctuations in molecule numbers from cell to cell, as has been well documented for gene levels (see [139, 110, 35], for example). Lastly, some of the species of the cycle may be found only in the cellular membrane rather than in the cytoplasm, or may be localized within specific cellular compartments, or

may move about the cell by diffusion or active transport in an activity-dependent manner (e.g., the yeast protein HOG1 that dwells in the cytoplasm unless doubly phosphorylated, when it translocates into the cell nucleus). The consequences of these spatial effects need to be understood (see [85] for a recent review).

Achieving a systems level understanding of signaling pathways is an important problem, and is highly limited by the lack of experimental data with enough temporal resolution to guide modeling efforts. Perhaps identifying and analyzing the relevant signaling modules as we have done here for a signaling cycle will shed some light on their behavior. Similar explorations could be done on other commonly occurring motifs in signaling pathways such as G-protein coupled receptors, or signaling cycles that require multiple phosphorylation events to become active.

The SwissProt ([136]) reference numbers for the genes Ptc1, Ptc2, Ptc3, Pbs2, Ste7, Ste11, Ssk2 and Ssk22, mentioned in this paper, are respectively P35182, P39966, P34221, P08018, P06784, P23561, P53599, P25390.

3.5 Methods

All analytical expressions were obtained starting from Equation 3.3, the tQSSA approximation of the cycle, the derivation of which is discussed in Section 3.5.1. The full mass action kinetics (MAK) description of the system (again, see Section 3.5.1) was analyzed numerically to obtain the data used in all the plots. Therefore, although the analytical expressions depend on the validity of the tQSSA, the general results do not as they have been numerically verified on the full system.

To analyze the steady state behavior, the cycle equation (Equation 3.3) corresponding to each regime was set to zero to obtain $\overline{A}$ at steady state. The quasi steady state expression for the phosphatase-protein complex, obtained in the process of deriving Equation 3.3 was then used to translate this into the steady state output A (see Section 3.5.3 for details).

To obtain the dynamic response O Eq.3.3 was linearized about a chosen steady state level, assuming that the deviations in the input from its steady state level are

small. The steady state level of the input for the four cycles was chosen such that the steady state output was about half-way to saturation, to allow the cycles to respond as much as possible. Choosing other steady state values where the slope of the steady state response curve is small would lead to little response. Particular care has to be placed in the ultrasensitive cycle, which has a very small range of inputs where its slope is non-zero, implying that this cycle needs to be finely tuned for it to transmit dynamic information.

All numerical analysis was done in Matlab, starting from the full MAK description of the cycle. The data in Figure 3-2 was obtained by setting the derivatives to zero and solving the resulting algebraic relations numerically. The data in Figures 3-7, 3-6, and S1 was obtained by numerically integrating the MAK equations for the given inputs using the $ODE23s$ Matlab function. Finally, the data in Figure 3-8 was obtained by integrating the MAK equations using the Runge-Kotta algorithm on inputs of the form $E_0(1 + a \sin \omega t_i + \eta(0, 1))$ where t_i is any time point in the numerical integration and $\eta(0, 1)$ is a normal random variable (with unit variance and zero mean).

3.5.1 Derivation of Equation 3.3

We start with the mass action kinetics description of the reactions specified by Equations 3.1 and 3.2. There are six chemical species and three conservation relations (the kinase, phosphatase and substrate protein are conserved), yielding a total of three variables. Letting C_1 (C_2) denote the concentration of the inactive (active) enzyme-substrate complex IE_1 (AE_1), we write down the mass action kinetics equations for the enzyme-substrate complexes and for the amount of active protein $\overline{A}$ yields

$$\begin{aligned}
\frac{d\overline{A}}{dt} &= k_1 C_1 - k_2 C_2 && (3.10)\\
\frac{dC_1}{dt} &= a_1\big[(\overline{S} - \overline{A} - C_1)(\overline{E_1} - C_1) - K_1 C_1\big] \\
&= a_1\big[C_1^2 - (K_1 + \overline{E_1} + \overline{S} - \overline{A})C_1 + \overline{E_1}(\overline{S} - \overline{A})\big] && (3.11)\\
\frac{dC_2}{dt} &= a_2\big[(\overline{A} - C_2)(\overline{E_2} - C_2) - K_2 C_2\big] \\
&= a_2\big[C_2^2 - (K_2 + \overline{E_2} + \overline{A})C_2 + \overline{E_2}\,\overline{A}\big], && (3.12)
\end{aligned}$$

where $\overline{S}$ denotes the total amount of substrate protein, a_1 and a_2 are the association rate constants of the two enzymatic reactions and K_1 and K_2 the Michaelis Menten constants.

To apply the tQSSA, we then hypothesize that the complexes have faster dynamics than the active protein and that they are always at equilibrium with respect to the active substrate protein. This allows us to substitute C_1 and C_2 in Equation 3.10 with the equilibrium values, which are in turn found by setting the left hand side of Equations 3.11 and 3.12 to zero and solving for the complexes.

So doing yields

$$C_1 = \frac{K_1 + \overline{E_1} + \overline{S} - \overline{A}}{2}\left(1 - \sqrt{1 - 4r_1}\right) \quad \text{and} \quad C_2 = \frac{K_2 + \overline{E_2} + \overline{A}}{2}\left(1 - \sqrt{1 - 4r_2}\right)$$

where

$$r_1 = \frac{\overline{E_1}(\overline{S} - \overline{A})}{\left(K_1 + \overline{E_1} + \overline{S} - \overline{A}\right)^2} \quad \text{and} \quad r_2 = \frac{\overline{E_2}\,\overline{A}}{\left(K_2 + \overline{E_2} + \overline{A}\right)^2}.$$

To further simplify the equilibrium expressions for the complexes, we approximate to first order in r_1 and r_2 (which is reasonable when $r_1 \ll 1$ and $r_2 \ll 1$), yielding the following expressions for C_1 and C_2:

$$C_1 = \frac{\overline{E_1}(\overline{S} - \overline{A})}{K_1 + \overline{E_1} + \overline{S} - \overline{A}} \quad \text{and} \quad C_2 = \frac{\overline{E_2}\,\overline{A}}{K_2 + \overline{E_2} + \overline{A}}. \tag{3.13}$$

These expressions are finally inserted into Equation 3.10 to yield the signaling cycle equation (Equation 3.3).

In [142] Rami Tzafriri describes in full detail how to obtain the same result for a single enzymatic reaction in a self-consistent manner. In particular, he finds conditions under which the complexes indeed reach equilibrium with respect to the substrate, and under which the first order approximation of the square root is valid. The same argument carries through for each enzymatic reaction in the signaling cycle. In particular, the tQSSA is expected to hold when

1. Either $K_1 + \overline{E_1} \gg \overline{S}$ or $K_1 + \overline{S} - \overline{A} \gg \overline{E_1}$, and

2. Either $K_2 + \overline{E_2} \gg \overline{S}$ or $K_2 + \overline{A} \gg \overline{E_2}$.

As these inequalities are better satisfied, the tQSSA describes the signaling cycle better. Similarly, whereas a Michaelis Menten approximation would be valid only at low enzyme concentrations, an inspection of the conditions above shows the tQSSA is also valid when the enzyme concentrations are high. The conditions we use to define the four signaling regimes of the cycle are consistent with the sufficient conditions for the validity of the tQSSA.

3.5.2 Equations for the Four Cycle Regimes

Regime 1: both kinase and phosphatase are saturated : ultrasensitive

This regime was first identified in [61], where its steady state behavior was analyzed. Equation 3.3 reduces to

$$\frac{d\overline{A}}{dt} = k_1\overline{E_1} - k_2\overline{E_2},$$

indicating that in this regime the signaling cycle effectively integrates the difference of its (scaled) input and a reference level specified by the (scaled) phosphatase level. When the difference maintains the same sign for long enough, it will become saturated at a low or a high output level, for a negative and a positive difference, respectively. This regime can be used in feedbacks requiring time integration (integral feedbacks), such as the one proposed to operate in bacterial chemotaxis ([156]).

Regime 2: kinase saturated, phosphatase unsaturated : signal-transducing

Of the two new regimes that we characterize in this study, the one with a saturated kinase and unsaturated phosphatase (Figure 3-2B) is of particular interest. We refer to this regime as signal-transducing because, as discussed below, it is ideal for transmitting noisy time-varying signals. Equation 3.3 for this regime becomes

$$\frac{d\overline{A}}{dt} = k_1\overline{E_1} - k_2\frac{\overline{E_2}}{K_2 + \overline{E_2}}\overline{A},$$

which is linear in $\overline{A}$. This has several interesting implications. In particular, it implies that for slow inputs (relative to the cut-off frequency $k_2\overline{E_2}/(K_2+\overline{E_2})$) the output $\overline{A}$ will simply be a scaled copy of the input. While the latter holds for slow inputs, quickly varying inputs (noise) are filtered out making this regime ideal for the transmission of signals. Furthermore, the fact that this cycle is a linear system implies that pathways (or part of pathways) built of cycles in this regime become highly tractable mathematically since all the well-developed signals and systems techniques would apply to them. Available biochemical data and in vivo measurements argue in favor of this regime to be present in cell signaling cascades (see Discussion).

Regime 3: kinase unsaturated, phosphatase saturated : threshold-hyperbolic

The second new regime has an unsaturated kinase and a saturated phosphatase, and Equation 3.3 becomes

$$\frac{d\overline{A}}{dt} = k_1\frac{\overline{E_1}(\overline{S}-\overline{A})}{K_1+\overline{E_1}} - k_2\overline{E_2}.$$

Its steady state output is zero for inputs below a threshold and then increases hyperbolically with increasing steady-state inputs.

Regime 4: both kinase and phosphatase unsaturated : hyperbolic

This regime was also first identified in [61] and exhibits a hyperbolic steady state response. Equation 3.3 becomes

$$\frac{d\overline{A}}{dt} = k_1\frac{\overline{E_1}(\overline{S}-\overline{A})}{K_1+\overline{E_1}} - k_2\frac{\overline{E_2}\,\overline{A}}{K_2+\overline{E_2}+\overline{A}}$$

for this regime.

3.5.3 Steady State

The output of the cycle is the amount of free active protein and may be found from the amount of active protein $\overline{A}$ and of active complex C_2, since $A = \overline{A} - C_2$. Analytic approximations to the steady state response of the signaling cycle may then be ob-

tained by finding expressions for $\overline{A}$ and C_2. The former can be found by setting the left hand side of Equation 3.3 to zero and solving for $\overline{A}$, while the latter is taken to be $\frac{\overline{E_2}\,\overline{A}}{K_2+\overline{E_2}+\overline{A}}$ as discussed in the appendix above.

So doing for the four signaling regimes results in analytic expressions for their steady state responses. The case of the ultrasensitive regime, however, involves a slightly different method.

Regime 1: ultrasensitive

Setting Equation 3.3 to zero for this regime results in $k_1\overline{E_1} = k_2\overline{E_2}$, and since for this regime $C_2 \approx E_2$ then this indicates that $C_2 \approx \frac{k_1}{k_2}\overline{E_1}$. Numerical simulation indicates that as long as $\frac{k_1}{k_2}\overline{E_1} \leq \overline{E_2}$ the previous relation is accurate and furthermore that the switch output is zero. As the input increases beyond this point, C_2 quickly increases to its maximal value $\overline{E_2}$ (i.e., the phosphatase becomes fully saturated, while the level of free inactive protein decreases to zero and the inactive complex $C_1 \approx \overline{E_2}\frac{k_2}{k_1}$. Together, these observations imply that for $\overline{E_1} \leq \frac{k_2}{k_1}\overline{E_2}$ the output of the cycle is zero and $\overline{A} \approx C_2 \approx \frac{k_1}{k_2}\overline{E_1}$, and that for inputs above this level $\overline{A}$ quickly saturates at $\overline{S} - \frac{k_2}{k_1}\overline{E_2}$, and the output level is given by $A = \overline{S} - \big(1 + \frac{k_2}{k_1}\big)\overline{E_2}$. This implies that no matter how high the input is, the output of the ultrasensitive cycle will never equal the total amount of substrate protein unless there is no phosphatase.

Regime 2: signal-transducing

Setting Equation 3.3 to zero for this regime results in $k_1\overline{E_1} - k_2\frac{\overline{E_2}}{K_2+\overline{E_2}}\overline{A} = 0$, so $\overline{A} = \frac{k_1}{k_2}\frac{K_2+\overline{E_2}}{\overline{E_2}}\overline{E_1}$. At the same time $C_2 \approx \frac{\overline{E_2}}{K_2+\overline{E_2}}\overline{A}$ so that $A = \frac{k_1}{k_2}\big(\frac{K_2+\overline{E_2}}{\overline{E_2}} - 1\big)\overline{E_1}$. This linear relationship between the output and the input can not hold for high inputs because the output must be less than the total amount of substrate. We therefore expect the output to saturate when there is not free inactive protein, i.e., when $\overline{A} + C_1 \approx \overline{S}$. Since $C_1 \approx \overline{E_1}$ in this regime, the previous expression implies that the switch will saturate when $\overline{E_1} \approx \frac{\overline{S}}{1+\frac{k_1}{\omega_2}}$, where $\omega_2 = \frac{k_2\overline{E_2}}{K_2+\overline{E_2}}$. Evaluating the output at this input level yields the saturation value of the switch in this regime:

$A = \left(1 - \frac{\omega_2}{k_2}\right)\left(\frac{\frac{k_1}{\omega_2}}{1+\frac{k_1}{\omega_2}}\right)\overline{S}$.

Regime 3: threshold-hyperbolic

Setting Equation 3.3 to zero for this regime results in $\omega_1(\overline{S} - \overline{A}) - k_2\overline{E_2} = 0$, where $\omega_1 = k_1\frac{\overline{E_1}}{K_1+\overline{E_1}}$. This implies that $\overline{A} \approx \overline{S} - \frac{k_2}{\omega_1}\overline{E_2}$ and since $C_2 \approx \overline{E_2}$, that $A \approx \overline{S} - (1 + \frac{k_2}{\omega_1})\overline{E_2}$. This approximation is not expected to hold at low inputs, where it blows up. Instead, at low inputs the free active protein is expected to be zero and $\overline{A} \approx C_2 \approx \frac{\omega_1}{k_2}(\overline{S} - \overline{A})$ from the first expression in this subsection. Solving for $\overline{A}$ gives $\overline{A} \approx \frac{\omega_1}{k_2+\omega_1}\overline{S} \approx A$ for low inputs. This expression is expected to break as the input level reaches a level $\overline{E_1}^*$ where the expression equals $\overline{E_2}$. Above that input the first expression for $\overline{A}$ is expected to hold. Therefore, for inputs below $\overline{E_1}^*$ the output is approximately zero, and then increases hyperbolically as $A \approx \overline{S} - (1 + \frac{k_2}{\omega_1})\overline{E_2}$.

Regime 4: hyperbolic

Setting Equation 3.3 to zero for this regime results in $\omega_1(\overline{S} - \overline{A}) - \omega_2\overline{A} = 0$, where ω_1 and ω_2 are as defined above. Therefore $\overline{A} \approx \frac{\omega_1}{\omega_1+\omega_2}\overline{S}$ and since $C_2 \approx \frac{\omega_2}{k_2}\overline{A}$ then $A = \left(1 - \frac{\omega_2}{k_2}\right)\frac{\omega_1}{\omega_1+\omega_2}\overline{S}$. The saturation level of this regime is obtained by evaluating the previous expression in the limit as $\overline{E_1}$ becomes infinite.

3.5.4 Dynamics

To find approximate analytic expressions for the response of the system to inputs of the form $\overline{E_1} = E_0(1 + a\sin\omega t)$, we use small signal analysis. This method consists of linearizing the system about its steady state level, and further assuming that the input deviates from its steady state level by small amounts. Any results thus obtained are expected to be valid for small E_0a, although numerically we have observed that the results so obtained describe the system better than they might have the right to when E_0a is not small. The method works as follows: First let the function $f(\overline{A}, \overline{E_1})$ (or just f for simplicity) denote the rate of change of $\overline{A}$ as described by Equation 3.3 (i.e., $\frac{d\overline{A}}{dt} = f(\overline{A}, \overline{E_1})$), and let $\overline{A}_{ss}$ be the steady state level of $\overline{A}$ when the input

is constant and equal to E_0, so that $f(\overline{A}_{ss}, E_0) = 0$. Then define the deviations from steady state levels $\delta\overline{A} = \overline{A} - \overline{A}_{ss}$ and $\delta\overline{E_1} = \overline{E_1} - E_0 = E_0 a \sin \omega t$. Assuming the deviations are always small and Taylor expanding $f(\overline{A}, \overline{E_1})$ about the steady state levels then yields

$$\frac{d\delta\overline{A}}{dt} = g\delta\overline{E_1} - \omega_c \delta\overline{A}, \tag{3.14}$$

where $g = \frac{\partial f}{\partial \overline{E_1}}\big|_{(\overline{A}_{ss}, E_0)}$ is referred to as the gain and $\omega_c = \frac{\partial f}{\partial \overline{A}}\big|_{(\overline{A}_{ss}, E_0)}$ as the cut-off frequency. This equation is linear and may be solved for arbitrary inputs $\delta\overline{E_1}$ by one of the many useful techniques to work with linear differential equation (i.e., by Laplace transforms). In particular, when $\delta\overline{E_1} = aE_0 \sin \omega t$ and the initial condition is zero

$$\delta\overline{A} = aE_0 \frac{g}{\sqrt{\omega^2 + \omega_c^2}} \cos\left(\omega t + \tan^{-1}(\frac{-\omega_c}{\omega})\right) + aE_0 g\omega e^{-\omega_c t},$$

where $\tan^{-1}$ denotes the inverse tangent. Here, we are only interested in twice the amplitude of the steady state oscillations in $\overline{A}$, from maximum to minima. These are evidently given by Equation 3.6, such that for frequencies smaller than the cut-off ω_c the oscillations are proportional to $\frac{g}{\omega_c}$ and oscillations for frequencies larger than ω_c decay as $1/\omega$.

Because the output of the system is $A = \overline{A} - C_2$, we need to translate these oscillations in $\overline{A}$ to oscillations in A. In the ultrasensitive and threshold-hyperbolic regimes, $C_2 \approx \overline{E_2}$ so the oscillations in A equal those in $\overline{A}$. In the hyperbolic and signal-transducing regimes, $C_2 \approx \frac{\omega_2}{k_2}\overline{A}$, so the amplitude of the oscillations in A is that amplitude of the oscillations in $\overline{A}$ multiplied by a factor of $1 - \frac{\omega_2}{k_2}$.

Regime 1: ultrasensitive

For the ultrasensitive regime we do not need to use the method above. This regime needs to be fine-tuned to transmit signals because, as evidenced by its steady state response curve, is only responsive to changes in the input close to its inflection point, at $\overline{E_1} = \frac{k_2}{k_1}\overline{E_2}$. Choosing E_0 at this level results in the cycle equation becoming $\frac{d\overline{A}}{dt} = k_1 a E_0 \sin \omega t$, which is identical to Equation 3.14 with a gain of k_1 and cut-off frequency of zero. The previous equation does not hold for small enough frequencies;

instead at some effective cut-off frequency the oscillations will cover the full range of values that the ultrasensitive cycle may take. That is, the effective cut-off frequency satisfies $2E_0 a\frac{k_1}{\omega_c} = \overline{S} - \left(1+\frac{k_2}{k_1}\right)\overline{E_2}$, where the right hand side is the saturation level of the cycle. Solving for ω_c in this expression yields the cut-off frequency in Table 3.4. The ultrasensitive regime is the only one that achieved oscillations that cover its full steady state response range, and where the (effective) cut-off frequency depends on the input amplitude a.

Regime 2: signal-transducing

Because Equation 3.3 for this regime is already linear in $\overline{A}$ and in $\overline{E_1}$, it already has the same form as Equation 3.14 with $g = k_1$ and $\omega_c = \omega_2$. Multiplying the gain by $1 - \frac{\omega_2}{k_2}$ to translate to oscillations in $\overline{A}$ gives the result in Table 3.4.

Regime 3: threshold-hyperbolic

Applying the method described above results in the expressions in Table 3.4 (These results are not expected to hold when the steady state input E_0 is below the regime's threshold and the output is zero). For simplicity though, we let $\omega_0 = \omega_1\big|_{(\overline{E_1}=E_0)} = \frac{k_1 E_0}{K_1+E_0}$, which turns out to be ω_c for this regime.

Regime 4: hyperbolic

Applying the method described above results in the expressions in Table 3.4, where the cut-off turns out to be $\omega_c = \omega_0 + \omega_2$.

3.5.5 Quantifying the Quality of the Four Regime Approximations

Taking extreme values of the kinase and phosphatase MM constants allows us to obtain the four signaling regimes previously discussed. However, the results obtained from these approximations apply reasonably well to a wide range of MM constants,

and not only at the extreme. The quality of the approximation does increase, however, as the MM constants become more extreme. To demonstrate this we numerically solved for the steady-state characteristic of Equation 3.3 for a wide range of kinase and phosphatase MM constants and compared them to the characteristics of each of the four regimes. For each set of K_1 and K_2 values, we set the left hand side of Equation 3.3 to zero and solve for $\overline{A}$ and then subtract $\frac{\overline{E_2}\,\overline{A}}{K_2+\overline{E_2}+\overline{A}}$ as discussed in the section above to obtain A. We do so for a range of total kinase values K_t and for each, we compute the difference from the steady-state of each of the four regime steady states. We finally square these differences and compute their mean resulting in the mean squared error for each regime. In Figure 3-3 we show what we refer to a the relative error, the square root of the squared error normalized by the total substrate S_t. This figure again shows that the regime approximations are each approximately valid over a large part of a quadrant, covering almost the full K_1 versus K_2 space when combined.

Chapter 4

Systems of Chemical Reactions and Stochastic Chemical Kinetics

4.1 Summary

Systems of chemical reactions, such as gene and metabolic networks, and signaling pathways, are pervasive in biology, as well as in other fields of study. These systems are typically described by a system of ordinary differential equations that track the concentrations of each chemical species in the system as a function of time. Such a description assumes that the system is well-mixed, so that there are no significant spatial concentration gradients, and deterministic, typically because the number of molecules, or system volume, is large. When either of these two assumptions fails, a completely different model must be used. This chapter describes the most popular framework, termed Stochastic Chemical Kinetics, used to model well-mixed systems of chemical reactions where stochastic effects (e.g., due to low molecule numbers) become relevant and need to be incorporated into the model.

4.2 Introduction

Two important dimensions determine the nature of the model appropriate to describe a system of chemical reactions: the mixing rate and the system size (see Figure 4-1).

The mixing rate refers to the speed with which the molecules in the system become uniformly distributed in space. The mixing can be active, for example, by the use of a stir-bar, or passive, by diffusion. The system size refers to the system volume, or alternatively, to the number of molecules in the system (see Chapter 5).

When the system mixes very fast (e.g., faster than the time until the concentrations change significantly because of reactions), spatial effects can be ignored and the state of the system can be specified through the number of molecules, or concentration, of each chemical species in the whole system. If the system mixes more slowly, then spatial concentration gradients arise and need to be incorporated into the model. Similarly, when the system size is large, as in macroscopic systems like a test tube, the system behavior is practically deterministic. For small systems, however, the probabilistic nature of individual reactions dominates and a fully-probabilistic model must be used.

When the system mixes quickly and has a large size, the simplest model framework results. The system state is then specified by the concentrations of each of the chemical species, and their time-evolution is governed by a set of ordinary differential equations. This modeling framework, termed mass action kinetics, is often used in biology not necessarily because it is the most appropriate, but because it is the simplest one. When the system is large but mixing is slow, the appropriate framework becomes reaction-diffusion kinetics, where the system is now described by (deterministic) partial differential equations that track the concentrations as a function of space and time.

Systems with fast mixing but small system size require a fully probabilistic model. The standard and perhaps most popular framework to model such systems, termed stochastic chemical kinetics, tracks the time-evolution of the probabilities of all physically possible system states (i.e., all possible values for an array with the molecule numbers for each chemical species). The resulting model, at first glance, seems nothing like mass action kinetics (but see Chapter 5), and is significantly more complex. This chapter describes the mass action kinetics, and the stochastic chemical kinetics frameworks in some detail, and the next two chapters apply model reduction tools

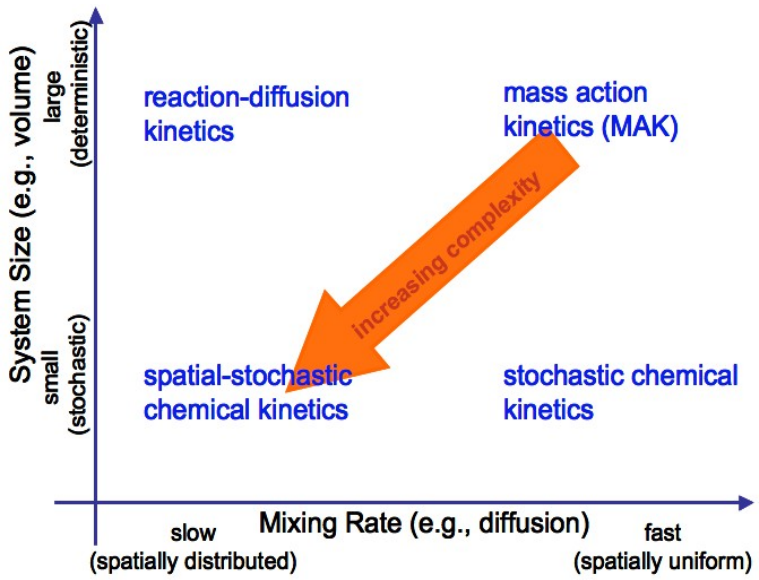

Figure 4-1: The nature of the model required to accurately describe a system of chemical reactions depends on two system variables: the mixing rate and the system size. The modeling framework the is often used in biology assumes, for simplicity, fast mixing and large system size. The other frameworks become much more complex. This chapter describes a standard modeling framework appropriate for systems with fast mixing but small system size.

to stochastic chemical kinetics in order to reduce the model complexity. Specifically, Chapter 5 obtains an approximate model for the evolution of the mean concentrations, and their variances and covariances, that closely resembles and is not much more complex than mass action kinetics. The hope is that this simpler although approximate model may be almost as easy to work with as mass action kinetics. Chapter 6, on the other hand, describes a separation of time scales method for stochastic chemical kinetics analogous to the one often applied to mass action kinetics models, which results in a model of reduced complexity, such as the Michaelis-Menten rate law.

When the system is small in size, and mixes slowly, the most complicated framework must be used (lower-left corner of Figure 4-1). The corresponding model is fully probabilistic, yet must account for spatial concentration gradients. Chapter 7 develops a simpler approximation for this modeling regime that tracks the mean concentrations, and their variances and covariances, as a function of space and time.

I have loosely used the words small and large to describe system sizes, and slow

and fast to describe system mixing rate, because it is not yet well known how large a system needs to be, or how fast the mixing rate needs to be, for stochastic and spatial effects, respectively, to be negligible. The hope is that the simplified approximations developed in this thesis will be useful to study when such effects become negligible.

The four modeling frameworks mentioned above and shown in Figure 4-1 should be compatible with one another, in the sense that the more complex ones should converge to the simpler ones in the appropriate limit, as denoted by the arrows in Figure 4-2. However, these convergences are not all well understood in a mathematical sense. The green arrows correspond to convergences that are more-or-less understood (for example, the green arrow on the right represents the limit as system volume goes to infinity while concentrations are kept constant; see Chapter 5), while the orange arrows correspond to less understood relationships. In Chapter 7, I propose a limit relating spatial-stochastic chemical kinetics to reaction-diffusion kinetics.

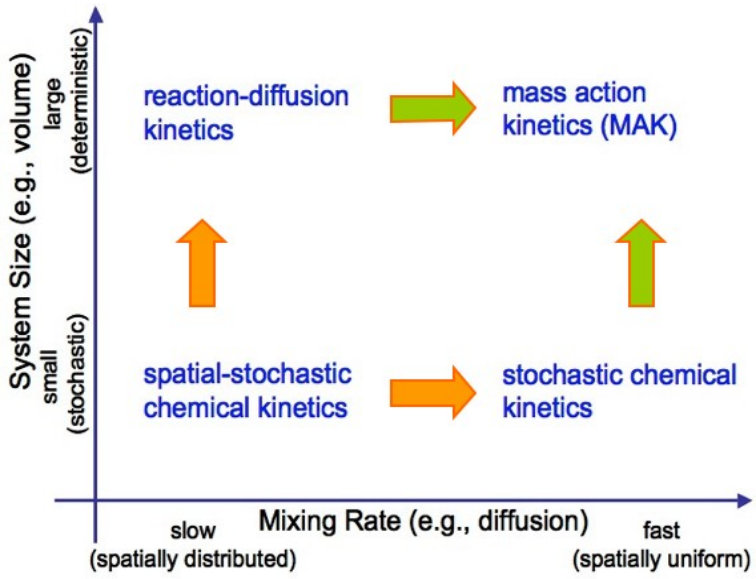

Figure 4-2: The four modeling frameworks and the convergences between them. The arrows denote the assumed convergences; green arrows correspond to limits that are more-or-less understood, while the orange ones to those that remain somewhat elusive.

4.3 Systems of Chemical Reactions

Systems of chemical reactions consist of n chemical species $\{X_1, ..., X_n\}$ interacting in a volume v through L reactions labeled $R_1, ..., R_L$.

Reaction R_ℓ corresponds to

$$\underline{s}_{\ell 1} X_1 + ... + \underline{s}_{\ell n} X_n \xrightarrow{k_\ell} \overline{s}_{\ell 1} X_1 + ... + \overline{s}_{\ell n} X_n \tag{4.1}$$

where $\underline{s}_{\ell i}$ and $\overline{s}_{\ell i}$ are the numbers of molecules of species X_i that are consumed and produced, respectively, on every *firing* of the reaction. The consumed species are referred to as *reactants*. The parameter k_ℓ in Equation 6.1 above is the *rate constant* of the reaction, and does not depend on Ω; its precise role in models of chemical reactions will be stated later.

Typically the system is assumed to be well-stirred, so that spatial concentration gradients do not arise within the system volume. This may be the case, for example, when diffusion is much faster than the chemical reactions. We make the well-stirred assumption in this and the next few chapters. However, Chapter 7 focuses on systems where this assumption does not hold, and describes frameworks to model these systems.

We let x_i and y_i be the number of molecules and the concentration, respectively, of species X_i. Molecule numbers are normalized by the system size, given by $\Omega = Av$ where A is Avogadro's number, to yield concentrations in moles per unit volume, so that $y_i = x_i/\Omega$. Letting $\mathbf{x}$ be the column vector of molecule numbers (with i-th component $\{\mathbf{x}\}_i = x_i$), a firing of reaction R_ℓ changes $\mathbf{x}$ instantaneously to $\mathbf{x} + \mathbf{s}_\ell$, where $\{\mathbf{s}_\ell\}_i = \overline{s}_{\ell i} - \underline{s}_{\ell i} = s_{\ell i}$ is the *stoichiometric coefficient* of species X_i in reaction R_ℓ, and $\mathbf{s}_\ell$ is referred to as the reaction's *stoichiometry vector*. The stoichiometry vectors are grouped into the $n \times L$ *stoichiometry matrix* $\mathbf{S}$ whose ℓ-th column equals $\mathbf{s}_\ell$.

4.3.1 Deterministic Models: Mass Action Kinetics

Traditionally, models of systems of chemical reactions describe the evolution of the concentrations of the different species through ordinary differential equations obtained from the mass action kinetics (MAK) rate laws. Reaction R_ℓ has a *macroscopic reaction rate*, denoted by $\overline{r}_\ell$, with $A\overline{r}_\ell$ being the average or expected number of times per unit volume per unit time that the reaction fires. Accounting for stoichiometry results in $s_{\ell i}\overline{r}_\ell$ being the expected rate of change in the concentration y_i caused by reaction R_ℓ. Letting $\mu_i = E[y_i]$ be the expected concentration of X_i and adding up the effects of all the reactions gives

$$\frac{d\mu_i}{dt} = \sum_{\ell=1}^{L} s_{\ell i}\overline{r}_\ell, \tag{4.2}$$

for each i, where we have omitted the time argument t from μ_i for notational simplicity. This set of equations may be expressed in matrix form by letting $\overline{\mathbf{r}}$ be the *macroscopic reaction rate* (column) *vector* (with $\{\overline{\mathbf{r}}\}_\ell = \overline{r}_\ell$), and similarly defining the expected concentration vector μ, so that

$$\frac{d\mu}{dt} = \mathbf{S}\overline{\mathbf{r}}. \tag{4.3}$$

The MAK rate laws are used to relate reaction rates to concentrations. Typically, as will be discussed in the next chapter, the respective macroscopic rates of 1) zero-order, 2) first-order, and 3) heterogeneous and 4) homogeneous second-order reactions are taken to be 1) independent of the concentrations, proportional to 2) the concentration of the reacting species, 3) the product of the concentrations of the two reacting species, 4) the squared concentration of the reacting species. Thus the rate $\overline{r}_\ell$ of reaction R_ℓ can, under these assumptions, always be expressed as a quadratic function of the concentration vector μ. The function is specified by the scalars k_ℓ and b_ℓ, an $n \times 1$ vector $\mathbf{c}_\ell$, and an $n \times n$ symmetric matrix $\mathbf{D}_\ell$, such that

$$\overline{r}_\ell(\mu) = k_\ell(\overline{b}_\ell + \overline{\mathbf{c}}_\ell{}'\mu + \mu'\overline{\mathbf{D}}_\ell\mu), \tag{4.4}$$

where we have made explicit the argument μ of $\overline{r}_\ell$, and where the superscript $'$ denotes the transpose of a vector (or matrix, later in the thesis). Collectively, $\overline{b}_\ell, \overline{\mathbf{c}}_\ell$ and $\overline{\mathbf{D}}_\ell$ are referred to as the *macroscopic rate parameters* of reaction R_ℓ. They are independent of the system size Ω and may be found by inspection from the specifications of the reactions, as illustrated in the examples of Section 6.9. In what follows, we let $c_{\ell i} = \{\overline{\mathbf{c}}_\ell\}_i$ and $d_{\ell ij} = \{\overline{\mathbf{D}}_\ell\}_{ij}$ denote the entries of $\overline{\mathbf{c}}_\ell$ and $\overline{\mathbf{D}}_\ell$.

Substituting for the rate $\overline{r}_\ell$ in Equation 6.37 using expressions of the form of Equation 4.4 results in the MAK equations, the conventional deterministic, macroscopic description of the system.

4.4 Stochastic Chemical Kinetics

To capture the inherent stochasticity of a system of chemical reactions, one can aim to track the evolution of the joint probability distribution (i.e., the joint probability mass function) of the molecule numbers that constitute the entries of the state vector $\mathbf{x}$. The standard probabilistic model is a continuous-time Markov process with time-dependent state given by $\mathbf{x}$; the time argument of $\mathbf{x}$ is omitted for notational simplicity. Each reaction R_ℓ fires in a small interval of time of length dt with a probability equal to $a_\ell(\mathbf{x})dt$, independently of the other time intervals and other reactions, and given that the state of the system is $\mathbf{x}$ at the start of the interval.

The quantity $a_\ell(\mathbf{x})$ is referred to as the *propensity* of reaction R_ℓ; its precise functional form, and all the assumptions in the preceding paragraph, can be derived from basic statistical mechanics considerations (see [56]). Propensities turn out to be proportional to the number of ways one can choose the reacting molecules.[1] Therefore, a reaction that has more than one molecule as reactant has a nonlinear propensity. The previous modeling assumptions lead to the Chemical Master Equation, and serve as the starting point for a family of algorithms to simulate system trajectories ([55,

[1]For example, if R_ℓ has one molecule of species X_i as its only reactant, then $a_\ell(\mathbf{x})$ is proportional to x_i. Similarly, if R_ℓ requires two molecules of species X_i as reactants, its propensity is then proportional to $x_i(x_i-1)/2$. Lastly, if R_ℓ requires one molecule of X_i and one molecule of X_j, its propensity is then proportional to $x_i x_j$.

24, 60]), such as Gillespie's widely used Stochastic Simulation Algorithm (SSA).

4.4.1 The Chemical Master Equation

Let $P(\mathbf{x}, t)$ be the probability distribution of the state $\mathbf{x}$ at time t, conditioned on an initial state $\mathbf{x}(t_o) = \mathbf{x}_o$, where we omit the conditioning on $\mathbf{x_o}$ for notational simplicity. Our goal here is to obtain an equation describing the evolution of this probability distribution. Suppose that the state of the system at time $t + dt$ is given by $\mathbf{x}$. If dt is small enough, then the probability that more than one reaction occurred in the interval dt is negligible. The state at time t must then have been at most one reaction away from $\mathbf{x}$, and since there are only L reactions, the state at time t could have only been one of $L + 1$ different states: $\mathbf{x}$, if no reaction occurred in the small time interval, or $\mathbf{x} - \mathbf{s}_\ell$ (with $\ell \in \{1, \ldots, L\}$) if reaction R_ℓ occurred.

This argument allows one to relate the probability of being in state $\mathbf{x}$ at time $t + dt$ to the the distribution of $\mathbf{x}$ at time t, resulting in

$$P(\mathbf{x}, t + dt) = P(\mathbf{x}, t) \times \big[1 - \sum_{\ell=1}^{L} a_\ell(\mathbf{x}) dt\big] + \sum_{\ell=1}^{L} P(\mathbf{x} - \mathbf{s}_\ell, t) a_\ell(\mathbf{x} - \mathbf{s}_\ell) dt. \qquad (4.5)$$

The previous equation obtains $P(\mathbf{x}, t + dt)$ by conditioning on the possible states at time t, and summing the product of the probability of each possible previous state with the probability of transitioning to state $\mathbf{x}$. The term between square brackets in the equation above denotes the probability that no reaction occurred in the time interval dt.

Subtracting $P(\mathbf{x}, t)$ on both sides of the previous equation, and then dividing by dt and taking the limit $dt \to 0$ results in the so-called Chemical Master Equation (CME), namely

$$\frac{dP(\mathbf{x})}{dt} = \sum_{\ell=1}^{L} P(\mathbf{x} - \mathbf{s}_\ell) a_\ell(\mathbf{x} - \mathbf{s}_\ell) - P(\mathbf{x}) a_\ell(\mathbf{x}). \qquad (4.6)$$

Here and from now on, I omit the argument t in $P(\mathbf{x}, t)$ for notational simplicity, so that $P(\mathbf{x})$ denotes the distribution of $\mathbf{x}$ at time t conditional on $\mathbf{x}_o$. The expectations

$E[\,\cdot\,]$ that we take in this and related chapters, unless explicitly stated, are all with respect to this distribution. The CME is actually the forward Kolmogorov equation (e.g., see [49]) for a continuous-time discrete-state Markov process model of the system of reactions. A careful derivation of the CME can be found in [56], for instance.

The CME is not a single equation, but rather a family of equations, one for each possible value of the state $\mathbf{x}$. Stacking the probabilities for all possible system states in a (possibly infinite) vector would result in a linear differential equation for $P(\mathbf{x})$. Unfortunately, the CME is typically hard to work with directly, mostly because the number of possible states is typically too prohibitive for all but the simplest of systems. For example, suppose that the reactions in a system are such that the number of molecules per species can never be more than M. Then the number of possible states is M^n, which grows exponentially with the number of species.

The CME also serves as the basis for approaches that reduce the complexity of the model (e.g., see [62, 57, 64, 97, 130, 103, 23]). For instance, Chapter 5 in this thesis develops approximate deterministic evolution equations for the means, variances and covariances of the species concentrations from the CME. In the process, it can be shown that the evolution of the mean concentrations converges to the deterministic MAK model as a limiting case when the effects of fluctuations on the means are neglgible. Similarly, Chapter 6 develops a reduced CME obtained by separation of time scales.

4.4.2 Stochastic Simulation

Traditionally, people have resorted to numerically simulating sample state trajectories of $\mathbf{x}$ rather than attempting to solve the CME. The key goal of this approach is to generate state trajectories stochastically, such that the probability of generating any given trajectory is consistent with the model statistics. This goal is achieved by exploiting a standard result for discrete-state continuous-time Markov processes (e.g., see [49]), which gives the probability distribution for the time until the next change in state and the actual next state, given the current state. Often many state trajectories are so simulated. The ensamble of states at a specific time may be used to

obtain an empirical probability distribution for the state at that time, and then used to estimate the expected value of arbitrary state functions. The resulting estimates will become more accurate as the number of trajectories is increased.

Between reactions, when the system state does not change, the model can simply be thought of as a merger of L *independent* Poisson processes (e.g., see [49]), with process number ℓ having a firing rate of $a_\ell(\mathbf{x})$. The standard results for these types of processes may be invoked to obtain the main results in this chapter. For example, the resulting combined process is also Poisson with rate $1/\mu_\tau = \sum_{\ell=1}^{L} a_\ell(\mathbf{x})$, so the time until the next reaction is exponentially distributed with mean μ_τ. Additionally, the next reaction will be of type R_ℓ (i.e., will correspond to process number ℓ) independently of the time until the next reaction and with probability $a_\ell / \sum_{j=1}^{L} a_\ell(\mathbf{x})$. After each reaction occurs, the state gets updated accordingly, and and a "new" process with updated rates starts. We describe next how to obtain the previous results without explicitly invoking Poisson processes, consistent with the literature in the field.

Suppose that at some time t the system is in state $\mathbf{x}$, and we want to generate a sample trajectory according to the model statistics. Let τ and ℓ respectively denote time until the next reaction, and the index of the next reaction, given the current state $\mathbf{x}$. If we knew the joint distribution $P(\tau, \ell | \mathbf{x}, t)$ of τ and ℓ, conditioned on the current state, we could sample from it, update our state at time $t+\tau$ to be $\mathbf{x}+\mathbf{s}_\ell$, and iterate this procedure until we have obtained a long-enough trajectory. This is exactly what the so-called stochastic simulation algorithm (SSA), more commonly referred to as the Gillespie algorithm, does (e.g., see [60]). All the numerical simulations of stochastic chemical systems in this thesis are obtained from this algorithm. The key random variables, τ and ℓ turn out to be independent, with τ exponentially distributed with mean

$$\mu_\tau = \frac{1}{\sum_{\ell=1}^{L} a_\ell(\mathbf{x})},$$

and with reaction R_ℓ being the next reaction with a probability of $a_\ell(\mathbf{x})\mu_\tau$. The following argument, adapted from [60] where it is described in more detail, shows

how to obtain the previous results.

The quantity $P(\tau, \ell|\mathbf{x}, t)dt$ is the probability that the next reaction in the system *is* R_ℓ, and will occur in the time interval $[t+\tau, t+\tau+dt)$. This probability can be expressed as the probability $F(\tau|\mathbf{x}, t)$ that *no* reaction occurs between t and $t+\tau$, times the probability that R_ℓ fires in $[t+\tau, t+\tau+dt)$. The latter is simply $a_\ell(\mathbf{x})dt$ by definition, so that

$$P(\tau, \ell|\mathbf{x}, t)dt = F(\tau|\mathbf{x}, t)a_\ell(\mathbf{x})dt.$$

Now consider $F(\tau + d\tau|\mathbf{x}, t)$. This probability must equal the probability that no reaction occurred between t ant $t+\tau$, namely $F(\tau|\mathbf{x}, t)$, times the probability that no reaction occurs in $[t+\tau, t+\tau+dt)$. This results in

$$F(\tau + d\tau|\mathbf{x}, t) = F(\tau|\mathbf{x}, t)\big[1 - \sum_{\ell=1}^{L} a_\ell(\mathbf{x})d\tau\big], \tag{4.7}$$

which after subtracting $F(\tau|\mathbf{x}, t)$ on both sides, dividing by $d\tau$, and taking the limit $dt \to 0$ becomes the ordinary differential equation

$$\frac{dF(\tau|\mathbf{x}, t)}{d\tau} = -\frac{1}{\mu_\tau}F(\tau|\mathbf{x}, t). \tag{4.8}$$

The initial condition of the previous equation is $F(0|\mathbf{x}, t) = 1$, since no reaction can occur when no time has elapsed, resulting in the solution $F(\tau|\mathbf{x}, t) = e^{-\frac{\tau}{\mu_\tau}}$. Substituting this result into our previous expression for $P(\tau, \ell|\mathbf{x}, t)$ gives the desired result:

$$P(\tau, \ell|\mathbf{x}, t) = a_\ell(\mathbf{x})e^{-\frac{\tau}{\mu_\tau}} = (a_\ell\mu_\tau)(\frac{1}{\mu_\tau}e^{-\frac{\tau}{\mu_\tau}}). \tag{4.9}$$

The last expression here shows that τ and ℓ are independent random variables, since their joint distribution is simply the product of their respective marginals. Furthermore, τ is exponentially distributed with mean μ_τ and R_ℓ is the next reaction with probability $a_\ell(\mathbf{x})\mu_\tau$.

To obtain sample state trajectories starting at some time t with state $\mathbf{x}$, one may simply evaluate μ_τ, which is a function of $\mathbf{x}$, and draw a sample from an exponential

distribution with mean μ_τ to obtain τ. Another random number is then drawn from the distribution for ℓ to choose the next reaction. Lastly, the state at time $t+\tau$ gets updated to $\mathbf{x}+\mathbf{s}_\ell$ before repeating the process. An alternate but equivalent approach samples L independent exponential distributions labeled by an index j, where the mean of distribution j is $1/a_j(\mathbf{x})$, and lets τ be the minimum of these L numbers, and ℓ the index j of the distribution that achieves this minimum.[2]

The previous algorithms generate sample paths that are statistically exactly consistent with the underlying model. Unfortunately, however, these simulations are often too computationally prohibitive, particularly when multiple trajectories are desired, for example, to compute the evolution of the expected concentrations, and their variances and covariances. Although certain improvements over the SSA can reduce computation (e.g., the Gibson-Bruck algorithm in [54] that only generates one random number per reaction by re-using previously obtained samples), exact algorithms must still simulate trajectories one reaction at a time. This is precisely why they are often too computationally costly: the expected time between reactions μ_τ decreases rapidly as the number of molecules increases, since propensities increase with molecule numbers (potentially faster than linearly), so simulating a trajectory of time-length T will require more and more iterations of the algorithm as the number of molecules increases.

Approximate Stochastic Simulation and Models

To circumvent the previous problem, some approximate approaches have been developed. Perhaps the best-known is the so-called *tau leaping* method (e.g., see [60]), which updates the system state only every τ time-units, where τ is a pre-specified deterministic quantity. The number of firings of the R_ℓ reaction during an interval of duration τ is approximated by a Poisson random variable with mean $a_\ell(\mathbf{x})\tau$, where $\mathbf{x}$ is the system state at the start of the interval. The number of firings of all reactions are assumed mutually independent, resulting in an algorithm that draws from L independent Poisson distributions at every iteration, updating the state accordingly

[2]It can be shown that τ and ℓ are still distributed according to Equation 4.9.

at the end of each"leap". The duration of the leap τ is chosen so that the expected change of state in each leap is sufficiently small that the propensity values change little, while having the expected number of reaction firing being much greater than one. Not all systems have values of τ that satisfy these requirements, though those with a large enough number of molecules tend to do so. Moreover, special care has to be taken with this approach to prevent molecule numbers from becoming negative.

A further approximation can be used to obtain a stochastic differential equation for that approximates the system state $\mathbf{x}$, which is known as the chemical Langevin equation. This description is obtained for systems where tau leaping is expected to work, *and* where the expected number of firings of *each* of the L reactions is much greater than one. The number of firings of reaction R_ℓ can then be approximated by a normal distribution with mean and variance equal to $a_\ell(\mathbf{x})\tau$, resulting in a differential equation driven by white noise called the Chemical Langevin Equation, after some re-arrangement and taking an appropriate limit. This argument is described in detail in [57].

The next chapter proposes a complement, and in some situations alternate approach to stochastic simulation. It derives a set of ordinary differential equations that approximately track the time-evolution of the mean concentrations, and their variances and covariances. The resulting equations, which are completely deterministic, can be numerically integrated to give the dynamics of the basic statistics of the system. Additionally, Chapter 6 develops a method to simplify the CME description for systems that have fast and slow reactions, a common situation.

Chapter 5

Mass Fluctuation Kinetics: Capturing Stochastic Effects in Systems of Chemical Reactions through Coupled Mean-Variance Computations

Summary

The intrinsic stochastic effects in chemical reactions, and particularly in biochemical networks, may result in behaviors significantly different from those predicted by deterministic mass action kinetics (MAK). Analyzing stochastic effects, however, is often computationally taxing and complex. We describe here the derivation and application of what we term the mass fluctuation kinetics (MFK), a set of deterministic equations to track the means, variances and covariances of the concentrations of the chemical species in the system. [1] These equations are obtained by approximating the dynamics of the first and second moments of the Chemical Master Equation. Apart from needing knowledge of the system volume, the MFK description requires only

[1] The material in this chapter has been published in [62].

the same information used to specify the MAK model, and is not significantly harder to write down or apply. When the effects of fluctuations are negligible, the MFK description typically reduces to MAK. The MFK equations are capable of describing the average behavior of the network substantially better than MAK, because they incorporate the effects of fluctuations on the evolution of the means. They also account for the effects of the means on the evolution of the variances and covariances, to produce quite accurate uncertainty bands around the average behavior. The MFK computations, although approximate, are significantly faster than Monte Carlo methods for computing first and second moments in systems of chemical reactions. They may therefore be used, perhaps along with a few Monte Carlo simulations of sample state trajectories, to efficiently provide a detailed picture of the behavior of a chemical system. In this chapter, we use the CME to obtain simple equations that describe the evolution of the expected concentrations, analogously to the MAK equations, but that also take into account the effects of stochastic fluctuations by tracking and using the evolution of the variances and covariances of the concentrations.

5.1 Introduction

Deterministic models based on mass action kinetics (MAK) are widely used to analyze and simulate systems of chemical reactions. However, at the volumes and concentrations that are found in, for example, gene networks and cellular signaling pathways, stochastic effects can dominate ([110, 112, 138, 30]). These effects originate in the intrinsically probabilistic nature of chemical reactions, and can lead to fluctuations about the deterministic behavior, to quantitative differences between the deterministic predictions and the average behavior, and even to qualitatively different behaviors (see Paulsson *et al.* [111] and Vilar *et al.* [146] for well-documented examples). Unfortunately, current methods for capturing the stochasticity of chemical reactions are significantly more complex and computationally taxing than deterministic models, despite recent advances ([24, 57, 103, 130, 64, 23]).

Reactions involving more than one reactant molecule are assumed to occur only

when the reacting molecules are found in physical contact with one another. As these encounters are governed by chance, reactions involving more than two reactant molecules are typically taken to be negligible ([60]). With this assumption, it suffices to consider systems where $\sum_{i=1}^{n} \underline{s}_{\ell i} \leq 2$ for every reaction R_ℓ. That is, we typically consider four types of reactions, depending on the reactants: (1) *zero-order* reactions that require no reactant, (2) *first-order* reactions that require a single reactant molecule, (3) *heterogeneous second-order* reactions that require two reactant molecules from different chemical species, and (4) *homogeneous second-order* reactions that require two reactant molecules from the same species.

We here show that the deterministic ordinary differential equations of MAK may be simply modified to account for stochastic effects, by starting from the Chemical Master Equation and using a moment-closure technique to approximate the evolution of the means, variances, and covariances of the concentrations of the chemical species. We refer to the resulting representation as the *mass fluctuation kinetics* (MFK) model. The model shows what the sources of noise are, and how they feed into the system. The use of the MFK equations is illustrated on several examples where MAK models fail in various ways, but where the MFK equations describe quite accurately the means and variances obtained by more complex and computer-intensive Monte Carlo simulations. The main results presented here have been published in [62].

5.2 Results

To proceed towards the goal stated at the end of the previous paragraph, let $\sigma_{ij} = E[(y_i - \mu_i)(y_j - \mu_j)]$ be the covariance (or *fluctuation*) between y_i and y_j, with $\sigma_{ii} = \sigma_i^2$ being the variance of y_i. Let the matrix $\mathbf{V}$ with $\{\mathbf{V}\}_{ij} = \sigma_{ij}$ be the *concentration covariance matrix*. Also, let the *microscopic reaction rate* ρ_ℓ be the propensity of the reaction R_ℓ, normalized by Ω and expressed as a function of $\mathbf{y}$ instead of $\mathbf{x}$:

$$\rho_\ell(\mathbf{y}) = \frac{a_\ell(\mathbf{x})}{\Omega}\Big|_{\mathbf{x}=\mathbf{y}\Omega}. \tag{5.1}$$

Its units are moles per unit volume per unit time, just as with the macroscopic reaction rate. The microscopic rate $\rho_\ell(\mathbf{y})$ is also quadratic in $\mathbf{y}$, with

$$\rho_\ell(\mathbf{y}) = k_\ell(b_\ell + \mathbf{c}_\ell'\mathbf{y} + \mathbf{y}'\mathbf{D}_\ell\mathbf{y}). \tag{5.2}$$

Here k_ℓ is the rate constant introduced earlier, while the *microscopic rate parameters* $b_\ell, \mathbf{c}_\ell$ and $\mathbf{D}_\ell$ are identical to the corresponding parameters in the expression for the macroscopic reaction rate in Equation 4.4, except in the case of homogeneous second-order reactions, which contribute extra entries to $\rho_\ell(\mathbf{y})$ that vary inversely with Ω (see Appendix 5.7). Thus, the microscopic and macroscopic reaction rates have the same functional form, i.e., $\rho_\ell(\mathbf{y}) = \overline{r}_\ell(\mathbf{y})$, for reactions other than homogeneous second-order ones, while for the latter we have $\lim_{\Omega\to\infty} \rho_\ell(\mathbf{y}) = \overline{r}_\ell(\mathbf{y})$. Microscopic rate parameters, like the macroscopic parameters, may be found by simple inspection from the specifications of the reactions, as illustrated in the examples of Section 6.9.

5.2.1 The Mass Fluctuation Kinetics (MFK) Equations

Starting from the CME, the evolution of μ and $\mathbf{V}$ are shown below to be well approximated by a closed system of equations when third-order moments about the mean are negligible. In these equations, each reaction R_ℓ has an associated *effective reaction rate* r_ℓ that equals the sum of two distinct components: the *average rate* $\rho_\ell(\mu)$, which depends only on the expected concentrations, and a *stochastic rate* $\xi_\ell(\mathbf{V})$, which depends only on the variances and covariances. The stochastic and effective rates of R_ℓ are given by

$$\xi_\ell(\mathbf{V}) = k_\ell \sum_{i,j=1}^{n} d_{\ell ij}\sigma_{ij} \tag{5.3}$$

$$r_\ell = \rho_\ell(\mu) + \xi_\ell(\mathbf{V}). \tag{5.4}$$

The stochastic rate ξ_ℓ is actually obtained by summing the products of the second-order partial derivatives of ρ_ℓ (evaluated at μ) and the corresponding fluctuations σ_{ij}, resulting in the expression in Equation 5.3 above.

To state the MFK equations, define vectors ρ, ξ and $\mathbf{r}$ such that $\{\rho\}_\ell = \rho_\ell(\mu)$, $\{\xi\}_\ell = \xi_\ell(\mathbf{V})$ and $\{\mathbf{r}\}_\ell = r_\ell$; these are the *average, stochastic* and *effective rate vectors,* respectively (all with dimensions $L \times 1$). Also define the $L \times L$ diagonal matrix $\mathbf{\Lambda}$ by $\{\mathbf{\Lambda}\}_{\ell\ell} = r_\ell$, with non-diagonal entries equal to zero. Lastly, let $\mathbf{M}$ denote the *fluctuation dynamics matrix* to be defined shortly. Then, when third-order moments about the mean are negligible, μ and $\mathbf{V}$ satisfy the coupled MFK equations

$$\frac{d\mu}{dt} = \mathbf{Sr} = \mathbf{S}\rho + \mathbf{S}\xi \tag{5.5}$$

$$\frac{d\mathbf{V}}{dt} = \mathbf{MV} + \mathbf{VM}' + \frac{1}{\Omega}\mathbf{S\Lambda S}'. \tag{5.6}$$

The j-th column of $\mathbf{M}$ is $\mathbf{S}\frac{\partial}{\partial \mu_j}\rho(\mu)$, so that in terms of the reaction parameters

$$\mathbf{M} = \mathbf{SC}' + 2\sum_{\ell=1}^{L} k_\ell \mathbf{s}_\ell \mu' \mathbf{D}_\ell, \tag{5.7}$$

where the $L \times n$ matrix $\mathbf{C}$ has ℓ-th row equal to $k_\ell \mathbf{c}_\ell$.

In the MFK equations the evolution of the mean concentration vector μ is driven by the variances and covariances through the column vector $\mathbf{S}\xi$, while the evolution of the covariance matrix $\mathbf{V}$ depends on the means, both through $\mathbf{M}$ and through the *dissipation matrix* $\frac{1}{\Omega}\mathbf{S\Lambda S}'$ (borrowing from the fluctuation-dissipation terminology of statistical thermodynamics [80]). The i-th diagonal entry in the dissipation matrix effectively drives the variance of species X_i and has the form $\frac{1}{\Omega}\sum_{\ell=1}^{L} s_{\ell i}^2 r_\ell$. The greater this term is, the greater the fluctuations in the concentration of X_i, so that low system size, high stoichiometric coefficients and high effective reaction rates produce increased fluctuations.

The MFK equations, though deterministic, correct the MAK equations for stochastic effects. While the MAK description consists of n ordinary differential equations, one for each species, the MFK description consists of $\frac{n(n+3)}{2}$ equations (not counting repeated equations, since $\mathbf{V}$ is symmetric by definition). This may seem like a high price to pay for including stochastic effects, but it is much simpler and much faster than obtaining means, variances and covariances by Monte Carlo methods, yet

fares very well at describing the true mean and fluctuations in a variety of chemical systems.

When $\xi_\ell(\mathbf{V}) = 0$ and $\rho_\ell(\mu) = \overline{r}_\ell(\mu)$ for all reactions R_ℓ in the system, the MFK equation for μ, Equation 7.13, reduces to MAK. This happens, for example, when there are no second-order reactions in the system.

Note from the MFK equations that when the mean concentrations reach steady-state levels, the effective rate vector $\mathbf{r}$ lies in the right null space of the stoichiometry matrix $\mathbf{S}$. This is analogous to the MAK equations, where $\overline{\mathbf{r}}$ is in the right null space of $\mathbf{S}$ at steady state.

Table 5.1: Notation.

Symbol	Variable
X_i	species number i
R_ℓ	reaction number ℓ
$s_{\ell i}$	stoichiometric coefficient of X_i in R_ℓ
$\mathbf{S}$	stoichiometry matrix
v	system volume
A	Avogadro's number
Ω	system size (Av)
x_i	population of X_i
y_i	concentration of X_i (x_i/Ω)
μ_i	expected concentration of X_i ($E[y_i]$)
μ	expected concentration vector
σ_{ij}	covariance of y_i and y_j
$\mathbf{V}$	covariance matrix
η_{ij}	noise-to-signal ratio of X_i and X_j ($\frac{\sigma_{ij}}{\mu_i \mu_j}$)
$\mathbf{\Gamma}$	noise-to-signal matrix
$\overline{r}_\ell$	macroscopic reaction rate of R_ℓ
r_ℓ	effective reaction rate of R_ℓ
$\rho_\ell(\mu)$	average rate of R_ℓ
$\xi_\ell(\mathbf{V})$	stochastic rate of R_ℓ

Table 5.1 summarizes for convenience the main notation introduced so far (as well as some symbols that are introduced later).

5.2.2 Outline of Derivation

Although a more detailed derivation of the MFK equations is presented in Appendix 5.7, it is useful here to briefly outline the key ideas.

To obtain the MFK equation for the mean, we first multiply the CME, Equation 4.6, by x_i and sum over all possible values of $\mathbf{x}$, and then normalize by Ω. The result, after algebraic simplification, turns out to be

$$\frac{d\mu_i}{dt} = \sum_{\ell=1}^{L} s_{\ell i} E[\rho_\ell(\mathbf{y})] = \sum_{\ell=1}^{L} s_{\ell i} r_\ell \tag{5.8}$$

as noted in Gillespie ([57]), for example; here $r_\ell = E[\rho_\ell(\mathbf{y})]$ is by definition the effective rate of reaction R_ℓ. This equation is similar to the MAK expression in Equation 6.37, but with r_ℓ instead of $\overline{r}_\ell$.

If the expectation in Equation 5.8 is approximated by $\rho_\ell(E[\mathbf{y}]) = \rho_\ell(\mu)$, one arrives directly at the MAK description (provided there are no homogeneous second-order reactions, or in the limit of $\Omega \to \infty$). Instead, for quadratic microscopic rate functions, we can evaluate the expectation in Equation 5.8 exactly and explicitly, obtaining

$$\frac{d\mu_i}{dt} = \sum_{\ell=1}^{L} s_{\ell i}\big(\rho_\ell(\mu) + \xi_\ell(\mathbf{V})\big). \tag{5.9}$$

The previous equation makes evident that r_ℓ is the sum of the average rate $\rho_\ell(\mu)$ and the stochastic rate $\xi_\ell(\mathbf{V})$, under the assumption of quadratic microscopic rate functions. As already noted, $\rho_\ell(\mu) = \overline{r}_\ell(\mu)$, except in the case of homogeneous second-order reactions. Even in this exceptional case we have equality in the limit $\Omega \to \infty$ for any fixed μ. The MFK reaction rate r_ℓ adds to $\rho_\ell(\mu)$ the rate $\xi_\ell(\mathbf{V})$, which represents a correction term due to stochastic effects.

The evolution equations for the covariances are found by first finding the evolution equation for $E[y_i y_j]$ by a method similar to that used to obtain Equation 5.8 above,

and then using the fact that $\sigma_{ij} = E[y_i y_j] - \mu_i \mu_j$. This yields the general equation

$$\begin{aligned} \frac{d\sigma_{ij}}{dt} &= \sum_{\ell=1}^{L}\Bigg(\frac{1}{\Omega}s_{\ell i}s_{\ell j}r_\ell \\ &\quad + s_{\ell i}\Big\{E[\rho_\ell(\mathbf{y})y_j] - r_\ell\mu_j\Big\} + s_{\ell j}\Big\{E[\rho_\ell(\mathbf{y})y_i] - r_\ell\mu_i\Big\}\Bigg), \end{aligned} \tag{5.10}$$

where the terms in braces are the covariances of the microscopic reaction rates and the concentrations. For quadratic microscopic rate functions, the equation simplifies to

$$\begin{aligned} \frac{d\sigma_{ij}}{dt} &= \sum_{\ell=1}^{L}\Bigg(\frac{1}{\Omega}s_{\ell i}s_{\ell j}\big(\rho_\ell(\mu) + \xi_\ell(\mathbf{V})\big) \\ &\quad + s_{\ell i}k_\ell\big[\sum_{p=1}^{n} c_{\ell p}\sigma_{pj} + 2\sum_{p,q=1}^{n} d_{\ell pq}(\mu_p\sigma_{qj} + \frac{\gamma_{pqj}}{2})\big] \\ &\quad + s_{\ell j}k_\ell\big[\sum_{p=1}^{n} c_{\ell p}\sigma_{pi} + 2\sum_{p,q=1}^{n} d_{\ell pq}(\mu_p\sigma_{qi} + \frac{\gamma_{pqi}}{2})\big]\Bigg). \end{aligned} \tag{5.11}$$

Here $\gamma_{pqj} = E[(y_p - \mu_p)(y_q - \mu_q)(y_j - \mu_j)]$ is the third-order moment about the mean. Unfortunately, although Equations 5.9 and 5.11 are exact consequences of the CME, they do not constitute a closed system of ordinary differential equations because Equation 5.11 depends on γ_{pqj}. Assuming that third-order moments about the mean are zero (as would be the case for a multivariate normal distribution) or negligible, i.e., setting $\gamma_{pqj} = 0$ for all p, q, j, produces the MFK equations, which may then be put in the matrix form of Equations 7.13 and 7.14.

Equations similar to MFK may be obtained for systems where the number of reactants per reaction is unrestricted, as well as for arbitrary propensity functions, as summarized in Section 5.4. However, for the rest of this paper we focus on the setting under which the MFK has been derived above.

The next section demonstrates the use of the MFK equations on four systems of increasing complexity, and compares the MFK predictions to those from the deterministic MAK solution, and to the behavior obtained from Monte Carlo simulations based on the SSA. Three of the examples are drawn or adapted from Khammash and

El Samad ([83]).

5.3 Examples

5.3.1 Complex Formation and Dissociation

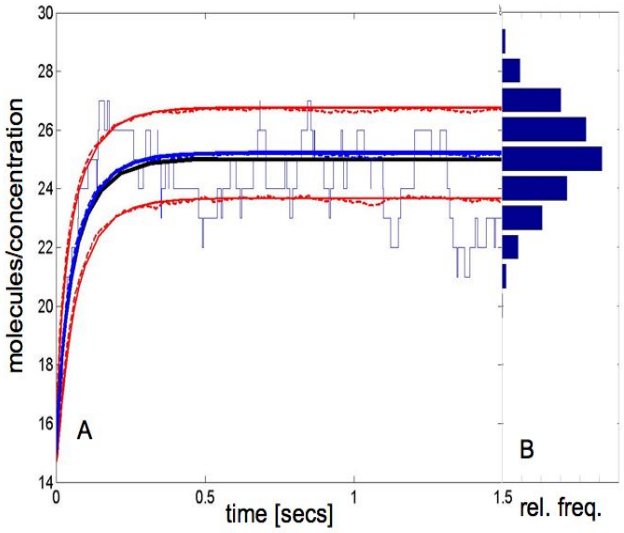

Figure 5-1: **(A)** Complex formation as a function of time for Example A. The (thick) blue and red solid lines are μ_3 and $\mu_3 \pm \sigma_3$, respectively, obtained from numerical integration of the MFK equations. The gray and red dotted lines respectively correspond to the "true" μ_3 and $\mu_3 \pm \sigma_3$, estimated from Monte Carlo simulation with 1000 runs, using the SSA. One such run is shown in the thin blue trace. The MAK solution (shown in black) does not follow the true mean as well as the MFK equations. **(B)** Histogram for the complex X_3 at time 1.5 secs, obtained from the 1000 runs of the simulation. The parameters used here are $\Omega = 1$ (so that molecule numbers and concentrations are the same), $k_1 = k_2 = 1$, initial conditions $\mathbf{x}(0) = [15\ 15\ 15]'$, and $\mathbf{V}(0) = \mathbf{0}$ (so that $\mu_3(0) = 15$, $\sigma_3^2(0) = 0$ and $\pi_1 = \pi_2 = 30$). In all the figures of this paper, rate constants for zero-, first- and second-order reactions are in units of $\frac{\text{mole}}{\text{sec}}$, $\frac{1}{\text{sec}}$ and $\frac{1}{\text{mole sec}}$, respectively. The system size is in units of liters.

Consider a system comprising the following two reactions:

$$X_1 + X_2 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X_3.$$

This represents the formation and dissociation of a complex (in this case X_3), such as

when a receptor and a ligand come together. The stoichiometry vectors for the two reactions are given by

$$\mathbf{s}_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, \quad \mathbf{s}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad \text{so} \quad \mathbf{S} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}. \tag{5.12}$$

For this example $\overline{r}_1 = k_1\mu_1\mu_2$ and $\overline{r}_2 = k_2\mu_3$. Thus $\overline{b}_1 = \overline{b}_2 = 0$, $\overline{\mathbf{c}}_\mathbf{1} = \mathbf{0}$, $\overline{\mathbf{c}}_\mathbf{2} = [0\ 0\ 1]'$,

$$\overline{\mathbf{D}}_\mathbf{1} = \begin{bmatrix} 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and} \quad \overline{\mathbf{D}}_\mathbf{2} = \mathbf{0},$$

where $\mathbf{0}$ indicates vectors or matrices of the appropriate dimensions with all their entries equal to zero.

At this point, expressing the rates in the quadratic form of Equation 4.4 may seem inefficient, as it appears that only one or two entries in the rate parameters will be non-zero. However, this turns out not to be true when conservation relations are used to reduce the number of state variables in the system, as will be discussed shortly for this example, and also in Section 5.5.1.

We could now write down the full deterministic MAK description of the system using Equation 6.37. The MAK equation for the concentration of the complex X_3 suffices for our purposes, and is

$$\frac{d\mu_3}{dt} = k_1\mu_1\mu_2 - k_2\mu_3. \tag{5.13}$$

Setting the derivative to zero yields the well known (macroscopic) equilibrium expression

$$\mu_3 = \frac{k_1}{k_2}\mu_1\mu_2. \tag{5.14}$$

To write down the MFK equations for this example, one may simply use Equations 7.13 and 7.14 directly. However, for our purposes here, it is helpful to first reduce

the number of variables used to track the state of the system. This may be done by noting that the network has two conservation relations: $y_1 + y_3 = \pi_1$ and $y_2 + y_3 = \pi_2$, where π_1 and π_2 are constants determined by the initial conditions. Using these to express y_1 and y_2 as functions of y_3, we may then represent the state of the system by tracking only species X_3. For this new state representation, the stoichiometric vectors and matrix simply reflect the changes in X_3 for each of the two reactions, and are respectively given by $\mathbf{s}_1 = 1$, $\mathbf{s}_2 = -1$ and $\mathbf{S} = [1 \; -1]$.

The microscopic rate function ρ_ℓ of zero-order, first-order, and heterogeneous second-order reactions have the exact same functional form and parameters as the macroscopic rate $\overline{r}_\ell$. It follows that in this example $\rho_1(\mu) = k_1\mu_1\mu_2 = k_1(\pi_1 - \mu_3)(\pi_2 - \mu_3)$ and $\rho_2(\mu) = k_2\mu_3$, so that $b_1 = \pi_1\pi_2$, $\mathbf{c_1} = -(\pi_1 + \pi_2)$, $\mathbf{D_1} = 1$, $b_2 = \mathbf{D_2} = 0$, and $\mathbf{c_2} = 1$. Then $\mathbf{C} = [\mathbf{c_1} \; \mathbf{c_2}]$, $\xi_1 = k_1\sigma_3^2$ and $\xi_2 = 0$. Lastly, $\mathbf{M} = \mathbf{c_1} - \mathbf{c_2} + 2k_1\mu_3 = k_1(2\mu_3 - \pi_1 - \pi_2) - k_2$. The MFK equations therefore become

$$\begin{aligned}
\frac{d\mu_3}{dt} &= r_1 - r_2 = \rho_1(\mu) + \xi_1 - (\rho_2(\mu) + \xi_2) \\
&= k_1(\pi_1 - \mu_3)(\pi_2 - \mu_3) - k_2\mu_3 + k_1\sigma_3^2 \qquad (5.15) \\
\frac{d\sigma_3^2}{dt} &= \frac{1}{\Omega}(r_1 + r_2) + 2[k_1(2\mu_3 - \pi_1 - \pi_2) - k_2]\sigma_3^2 \qquad (5.16) \\
&= \frac{1}{\Omega}(\rho_1 + \rho_2) + 2[k_1(\frac{1}{2\Omega} + 2\mu_3 - \pi_1 - \pi_2) - k_2]\sigma_3^2.
\end{aligned}$$

Numerical integration of these equations reveals (see Fig.1) that the MFK equations describe very well the mean and fluctuations of the system, as assessed by Monte Carlo simulations using repeated runs of the SSA. Note also that the MFK computations, complemented with just a single SSA run as in Fig.1, yield a much fuller picture of system behavior than either MFK or a run of the SSA alone, and certainly a more detailed picture than MAK provides.

For this simple network the effect of the fluctuations at equilibrium may be found analytically. Noting that $\mu_3 - \pi_1 = \mu_1$ and $\mu_3 - \pi_2 = \mu_2$, and setting the derivative in Equation 5.15 to zero, yields the stochastic equilibrium condition

$$\mu_3 = \frac{k_1}{k_2}(\mu_1\mu_2 + \sigma_3^2). \qquad (5.17)$$

Contrasting this expression with Equation 5.14 reveals that stochastic effects lead to a higher steady-state value of X_3 than predicted by MAK. This is consistent with the simulation results (see Fig.1).

As Ω goes to ∞ while keeping concentrations finite, the variance equation reduces to

$$\frac{d\sigma_3^2}{dt} = 2[k_1(2\mu_3 - \pi_1 - \pi_2) - k_2]\sigma_3^2. \tag{5.18}$$

With $k_2 > 0$ and noting that μ_3 can never be greater than π_1 or π_2 from the conservation relations, the previous equation describes a system that attains its unique stable equilibrium at $\sigma_3^2 = 0$ exponentially fast. The MFK equations for this example thus effectively become the MAK equations in this limit.

A Word of Caution

The MFK equations are approximate, and as such, one needs to verify that their behavior is indeed representative of the underlying system behavior. A good proxy for this may a single stochastic realization of a state trajectory of the system (e.g., obtained via the Gillespie algorithm), or even the MFK equations. Even in this simple example, some choices of initial conditions can result in unphysical predictions, including unbounded dynamics. The following argument, modified from [155], illustrates some such situations.

The MFK equation for the mean μ_3 in (5.15) *is* exact, but the one for the variance σ_3^2 is not. In its exact form, Equation (5.16) has an extra term equal to $+2k_1\gamma_3^3$ on the right of the equal sign, where $\gamma_3^3 = E[(y_3 - \mu_3)^3]$. Assuming this term is negligible, as we do in the MFK equations (5.15) and (5.16), can actually give rise to unphysical behavior for some physically-valid initial conditions.

To simplify things, suppose that $\pi_1 = \pi_2 = \pi$, and note that the complex concentration should always be bounded as follows $0 \leq \mu_3 \leq \pi$. The variance of the complex concentration σ_3^2 is bounded below by zero. Its upper bound, given a mean μ_3, is attained when the probability distribution for y_3 is concentrated at 0 and at π, with the probability that $y_3 = \pi$ being equal to μ_3/π. This results in $0 \leq \sigma_3^2 \leq \mu_3(\pi - \mu_3)$.

The physically-valid region of values for μ_3 and σ_3^2 , represented on a two-dimensional diagram with μ_3 on its x-axis and σ_3^2 on its y-axis, is then the area above $\sigma_3^2 = 0$ and below the parabola $\sigma_3^2 = \mu_3(\pi - \mu_3)$ that lies between $\mu_3 = 0$ and $\mu_3 = \pi$.

Ideally, for any initial condition within this region the solution of (5.15) and (5.16) should remain in this region. In particular, the velocity field of these equations should point into the physically-valid region along the limits of the region. It is easy to check this is indeed the case at the lower boundary where $\sigma_3^2 = 0$. To check whether this is true at the upper limit, we first define the column vector $\mathbf{N}$ that is normal to the parabola $\sigma_3^2 = \mu_3(\pi - \mu_3)$ and points into the physically-valid region. This vector is simply given by $\mathbf{N} = [(\pi - 2\mu_3) \quad -1]'$. To find out whether the velocity field of (5.15) and (5.16) points into the physically-valid region, we can simply test whether the dot product of $\mathbf{N}$ and $[\frac{d}{dt}\mu_3 \quad \frac{d}{dt}\sigma_3^2]'$ is positive for all physically-valid values of μ_3.

The computation, after some algebra, results in

$$\mathbf{N}' \begin{bmatrix} \frac{d}{dt}\mu_3 \\ \frac{d}{dt}\sigma_3^2 \end{bmatrix} = k_1\big[a^3 + 4\mu_3 a^2 - \mu_3^2 a - \frac{1}{\Omega}a(a+\mu_3)\big] + k_2\big[\mu_3^2 + \mu_3 a - \frac{1}{\Omega}\mu_3\big] \ , \tag{5.19}$$

where $a = \pi - \mu_3 \geq 0$. Unfortunately, as Ω becomes smaller the above number becomes more negative, and for any choice of parameters and value of μ_3 there is always an Ω such that for any smaller values of Ω the product above will be negative, and will thus point out of the physically-valid region.

John Wyatt suggests in [155] a way to modify the MFK equations in (5.15) and (5.16), for the situation when $k_2 \ll k_1$, so that the solution of the approximate equations will never leave the physically-valid region. Further work is required to understand more precisely when such situations arise, and how to correct for them, in general systems.

5.3.2 When MAK Fails to Quantitatively Describe the Mean Behavior: Stochastic Focusing

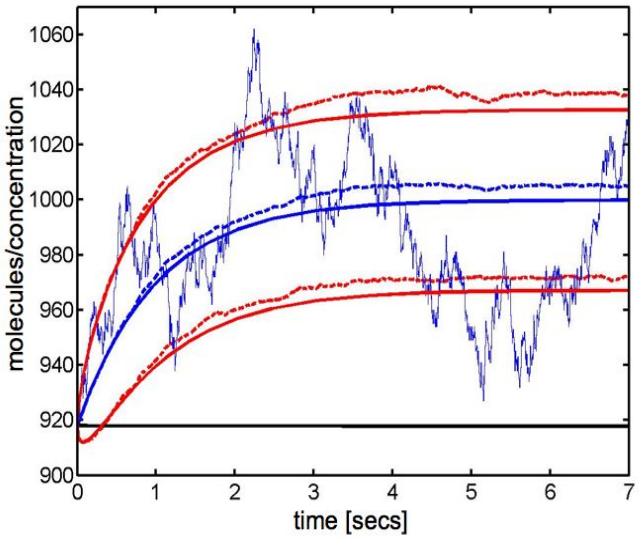

Figure 5-2: Product formation as a function of time for Example B. The blue and red solid lines are respectively μ_4 and $\mu_4 \pm \sigma_4$, obtained by numerical integration of the MFK equations. The blue and red dotted lines respectively correspond to the "true" μ_4 and $\mu_4 \pm \sigma_4$, estimated from Monte Carlo simulation with 1000 runs. One such run is shown in the thin blue line. The MAK solution (shown in black) significantly underestimates the true mean, while the MFK equations adequately capture the mean behavior and the uncertainty band. The parameters used are $\Omega = 1, k_1 = 10^4, k_2 = 9.9 \times 10^3, k_3 = 100, k_4 = 1.98 \times 10^5, k_5 = 10^4, k_6 = 1, k_7 = 10^4, k_8 = 10^3$, with initial conditions $\mathbf{x}(0) = [0\ 10\ 0\ 918]'$ and $\mathbf{V}(0) = \mathbf{0}$.

This example is based on one first described in Paulsson *et al.* [111], but modified to involve only the types of reactions considered in this paper. The system comprises the following 8 reactions:

$$\emptyset \xrightarrow{k_1} X_1, \quad X_1 + X_2 \underset{k_3}{\overset{k_2}{\rightleftharpoons}} X_3 \xrightarrow{k_4} X_2,$$
$$X_1 \xrightarrow{k_5} X_4 \xrightarrow{k_6} \emptyset \text{ and } \emptyset \underset{k_8}{\overset{k_7}{\rightleftharpoons}} X_2.$$

Here, X_4 is considered the product of the system. Its precursor X_1 binds to a signaling molecule X_2 to form the complex X_3, which then gets degraded. Because stochastic effects result in a higher concentration of X_4 than MAK predicts, the behavior was

termed *stochastic focusing*. (Observe a similar stochastic focusing effect in Equation 5.17 of the previous example.)

In the original example in [111], reactions 2, 3 and 4 are described by a single reaction, with a propensity function that is a rational function of the state (which may be obtained from a quasi-steady-state approximation). We here instead represent that reaction as a series of first- and second-order reactions.

The stoichiometry matrix $\mathbf{S}$ for our system is easily obtained, and the non-zero entries in the reaction parameters are $b_1 = 1$, $b_7 = 1$, $\{\mathbf{c_3}\}_3 = 1$, $\{\mathbf{c_4}\}_3 = 1$, $\{\mathbf{c_5}\}_1 = 1$, $\{\mathbf{c_6}\}_4 = 1$, $\{\mathbf{c_8}\}_2 = 1$, and $\{\mathbf{D_2}\}_{12} = \{\mathbf{D_2}\}_{21} = \frac{1}{2}$.

Figure 2 shows the mean and uncertainty bands (mean $\pm$ standard deviation) of X_4 obtained from the MFK and MAK equations, and from Monte Carlo simulation. The MAK model (in black) significantly underestimates the Monte Carlo mean. The MFK mean and uncertainty bands, on the other hand, fare much better at following the Monte Carlo results, the differences between them likely stemming from the fact that the MFK equations ignore third-moments about the mean. Although this system has only a single bimolecular reaction, that is enough to render the MAK description inadequate. Note again that the combination of the MFK results with a single run of the SSA, as in Fig.2, provides a satisfying picture of the stochastic behavior of the system.

5.3.3 When MAK Fails to Qualitatively Describe the Mean Behavior: A Genetic Oscillator

A system of chemical reactions with 9 species and 16 reactions is described in Vilar *et al.* ([146]) to model a genetic oscillator. In this network there are two DNA sequences (X_1 and X_3) and their respective mRNAs (X_5 and X_6) and gene products (X_7 and X_8). The sequence X_1 encodes an activator protein X_7 that increases transcription after binding to either X_1 or X_3 and creating the DNA-activator complexes X_2 and X_4, respectively. The sequence X_3, on the other hand, encodes the repressor protein X_8 that binds to activator proteins to produce the activator-repressor complex X_9, which

is degraded into a molecule of repressor, thereby reducing the number of activators. The reactions comprising the system are

$$X_1 + X_7 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X_2,\ X_2 \xrightarrow{k_3} X_2 + X_5,\ X_1 \xrightarrow{k_4} X_1 + X_5$$
$$X_3 + X_7 \underset{k_6}{\overset{k_5}{\rightleftharpoons}} X_4,\ X_4 \xrightarrow{k_7} X_4 + X_6,\ X_3 \xrightarrow{k_8} X_3 + X_6$$
$$\emptyset \xleftarrow{k_9} X_5 \xrightarrow{k_{10}} X_5 + X_7,\ \emptyset \xleftarrow{k_{11}} X_6 \xrightarrow{k_{12}} X_6 + X_8,$$
$$X_7 + X_8 \xrightarrow{k_{13}} X_9 \xrightarrow{k_{14}} X_8,\ X_7 \xrightarrow{k_{15}=k_{14}} \emptyset,\ \text{and}\ X_8 \xrightarrow{k_{16}} \emptyset.$$

The stoichiometry matrix $\mathbf{S}$ and the reaction parameters may be found by inspection of the above reactions.

The MAK description of such a network produces sustained oscillations for some combination of rate parameters. For another set of rate parameters, however, the MAK description fails to oscillate, while stochastic simulations reveal that the system does indeed oscillate ([146]), showing persistent bursts of activity with an irregular period (Fig. 3A shows 5 different stochastic simulations for the network). Thus, in this regime, MAK fails to even give a qualitatively accurate picture of the dynamic behavior of the system. Figure 3B shows the result from the MFK equations which, in contrast to the MAK solution, predict periodic oscillations. Monte Carlo simulation (see Fig. 3C) shows that the average levels and standard deviations do oscillate, but with an amplitude that decreases with time. As evident in Fig. 3A, the decrease in amplitude is due to the fact that different realizations of the processes do not exhibit synchronized bursts of activity, and that the level of de-synchronization increases with time (see the detailed analysis of a chemical system with a similar behavior in Gillespie and Mangel [58]). The differences between the MFK and Monte Carlo traces are likely due to the ignored third-order moments. Nevertheless, the MFK equations describe the behavior of the system much better than the MAK equations.

5.3.4 When both MAK and MFK Fail to Describe the Qualitative Mean Behavior: Bistability

When two different concentration vectors satisfy the MAK steady-state condition and are locally stable, the MAK equations describe a *bistable* system. In such a system, if the initial conditions lie in the basin of attraction of the first (second) steady state, the MAK equations settle to this steady state and remain there. When the system is treated stochastically, however, the behavior is more complex: starting from an arbitrary initial condition, either steady state may be reached in different realizations. If the fluctuations are small compared to the basins of attraction of the two steady states, then the system will fluctuate close to one of them. If the fluctuations are large enough, they will eventually kick the system out of whatever basin it is in and into the other one. This gives rise to a behavior typically referred to as *stochastic switching*, illustrated by the following network:

$$2X_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X_2 \xrightarrow{k_3} 3X_1, \quad X_1 + X_2 \underset{k_5}{\overset{k_4}{\rightleftharpoons}} X_3 \xrightarrow{k_7} \emptyset, \quad X_1 \underset{k_8}{\overset{k_6}{\rightleftharpoons}} \emptyset.$$

Species X_1 forms dimer X_2 which may turn into 3 molecules of X_1, or may bind one more molecule of X_1 to form the trimer X_3. Both X_1 and X_3 are respectively degraded through reactions 6 and 7, and X_1 is produced at some basal rate through reaction k_8.

For some parameter values, this network is bistable. Figure 4 shows the behavior of the network for one such set of parameters. The jagged blue curve shows a sample realization of the system intermittently *switching* between a high and a low level. The two black lines show the two equilibrium points obtained by numerical integration of the MAK equations, starting from different initial conditions, while the blue and red solid lines show the average concentration plus/minus a standard deviation obtained from the MFK equations. The MFK equations settle to a unique equilibrium for the two initial conditions that yielded different equilibria in the MAK description. The MFK equilibrium is between the two equilibria of the MAK description. The difference between the curves obtained from the MFK equations and the Monte Carlo

simulations (dotted blue and red lines) is attributed to having ignored third-order moments about the mean, which are significant, judging from the asymmetry of the empirical distribution (see Fig.4B).

Although the MFK description of this example fails to capture the system behavior, in that it does not indicate the two equilibria and the switching between them, the large standard deviation (comparable to the mean, rather than to its square root as in the other examples) is indicative of potential multi-stability and stochastic switching, suggesting the need for further exploration by stochastic simulation. For a more complex but biologically relevant example of stochastic switching, see Hasty *et al.* ([70]), where such behavior is found in a network of the bacteriophage λ.

5.4 General Propensity Functions

It is sometimes useful to consider chemical reactions with propensity functions that are not of the form in Equation 5.33. These result, for example, from a quasi-steady state approximation in systems with widely varying time scales. For example, Cao *et al.* ([22]) show that under appropriate conditions, the irreversible model of an enzymatic reaction

$$X_1 + X_2 \rightleftharpoons X_3 \rightarrow X_1 + X_4,$$

may well approximated by a system with a single reaction, namely

$$X_2 \rightarrow X_4.$$

Here, X_1, X_2, X_3 and X_4 respectively denote the enzyme, substrate, complex, and product species. The propensity function of the resulting reaction, however, is not quadratic in the molecule numbers but rather has a Michaelis-Menten form proportional to $\frac{x_{1t}x_2}{K+x_2}$, where K is a constant that depends on the rate constants of the original reactions, and x_{1t} is the total (free plus bound-to-substrate) enzyme concentration. Another situation where propensity functions are higher than quadratic in the concentrations occurs when reactions with more than two reactant molecules are

involved.

When propensities are smooth but otherwise arbitrary functions of the molecule numbers, MFK equations may still be obtained for the approximate evolution of the expected concentrations and their variances and covariances. The exact evolution equations for the expected concentrations and their variances and covariances, Equations 5.8 and 5.10, remain unchanged. However, now $E[\rho_\ell(\mathbf{y})]$ in Equation 5.8 generally involves higher moments and may not be expressed exactly as a function of μ and $\mathbf{V}$ as in Equation 5.4. A similar expression to Equation 5.4, though now approximate, may still be obtained by Taylor-expansion of $\rho_\ell(\mathbf{y})$ about the average concentrations μ, retaining only terms up to second order, and then taking the expectation. This results in

$$\begin{aligned} r_\ell &\approx \rho_\ell(\mu) + \frac{1}{2}\sum_{i,j=1}^{n}\left(\frac{\partial^2 \rho_\ell(\mathbf{y})}{\partial y_i \partial y_j}\Big|_{\mathbf{y}=\mu}\right)\sigma_{ij} \\ &= \rho_\ell(\mu) + \xi_\ell(\mu, \mathbf{V}). \end{aligned} \tag{5.20}$$

The effective rate is thus still a sum of the average rate and the stochastic rate, but now the stochastic rate $\xi(\mu, \mathbf{V})$ may be a function of the average concentrations as well. When the microscopic rate $\rho_\ell(\mathbf{y})$ has the form in Equation 5.2, the previous expressions for the effective and stochastic rates simplify to the expressions in Equations 5.3 and 5.4.

For the variance dynamics in Equation 5.10, the covariances between the microscopic rates and the concentrations (e.g., $E[\rho_\ell(\mathbf{y})y_j] - r_\ell\mu_j$) may similarly be approximated as a function of μ and $\mathbf{V}$ by first Taylor-expansion of terms of the form $\rho_\ell(\mathbf{y})y_j$ up to second order about μ, and then taking the expectation. Combining the resulting expression with Equation 6.6 yields

$$E[\rho_\ell(\mathbf{y})y_j] - r_\ell\mu_j \approx \sum_{p=1}^{n}\left(\frac{\partial \rho_\ell(\mathbf{y})}{\partial y_p}\Big|_{\mathbf{y}=\mu}\right)\sigma_{pj}. \tag{5.21}$$

Substituting Equations 6.6 and 5.21 into Equations 5.8 and 5.10 yields the MFK equations for systems of chemical reactions with arbitrary propensity functions. These

may be put in the same matrix form as Equations 7.13 and 7.14, except that now the effective rate r_ℓ is as defined in Equation 6.6, and the fluctuation dynamics matrix $\mathbf{M}$ is given by

$$\mathbf{M} = \mathbf{S}\frac{\partial \rho}{\partial \mu}, \tag{5.22}$$

rather than the expression in Equation 5.7. Here, $\frac{\partial \rho}{\partial \mu}$ is the Jacobian matrix of the effective rate vector ρ, with $\{\frac{\partial \rho}{\partial \mu}\}_{\ell p} = \frac{\partial \rho_\ell(\mu)}{\partial \mu_p}$. Therefore, in general $\mathbf{M}$ is now not a linear function of μ, though it is still independent of $\mathbf{V}$.

5.5 Additional Remarks

5.5.1 Conservation Relations

As the example in Section 5.3.1 illustrates, many networks have embedded conservation relations that may be used to eliminate some of the state variables, though this reduction comes at the expense of more non-zero rate parameters. In that example, neither of the two reactions changes the sums $x_1 + x_2$ and $x_1 + x_3$, so there are two conservation relations that express x_1 and x_2 as linear functions of x_3: (1) $x_1 = N_1 - x_3$ and (2) $x_2 = N_2 - x_3$, where $N_1 = \pi_1\Omega$ and $N_2 = \pi_2\Omega$ are determined by the initial numbers of molecules in the system. This allows the dynamics of the system to be described in terms of x_3 (and the parameters N_1 and N_2).

More generally, let $\mathbf{z}(t) = [z_1(t) \cdots z_M(t)]'$ where $z_j(t)$ is the number of times reaction R_j has fired from the start of the experiment at time t_o and up to time t. Then the state of the system can be expressed as $\mathbf{x}(t) = \mathbf{x}(t_o) + \mathbf{S}\mathbf{z}(t)$. Therefore, in order for the linear combination $\theta'\mathbf{x}$ to be conserved, i.e., to have $\theta'\mathbf{x}(t) = \theta'\mathbf{x}(t_o)$ for arbitrary $\mathbf{z}(t)$, it is necessary and sufficient that $\theta'\mathbf{S} = \mathbf{0}$. It follows that the number of conserved quantities equals the dimension ν of the left null space of $\mathbf{S}$. The dynamics of the system can thus be described in terms of just $n - \nu$ appropriately chosen state variables from the original n variables. The remaining ν variables can be expressed as linear combinations of the retained $n - \nu$ variables.

The MFK equations for the reduced set of $n - \nu$ variables have the same form

as for the original variables, since the quadratic microscopic rate functions remain quadratic in the retained variables. The stoichiometry matrix for the reduced description comprises just the appropriate $n - \nu$ rows of the original $\mathbf{S}$.

5.5.2 Steady State and the Large System Size Limit

The possible steady states of the MFK equations are found by setting the derivatives in Equations 7.13 and 7.14 to zero, resulting in the algebraic equations

$$\mathbf{0} = \mathbf{S}\rho + \mathbf{S}\xi \tag{5.23}$$

$$\mathbf{0} = \mathbf{M}\mathbf{V} + \mathbf{V}\mathbf{M}' + \frac{1}{\Omega}\mathbf{S}\mathbf{\Lambda}\mathbf{S}'. \tag{5.24}$$

These equations are still coupled, and are quadratic in the expected concentrations but linear in the variances. We assume for the rest of this subsection that all conservation relations have been taken into account as described in Section 5.5.1, so that $\mathbf{S}$ has no left null space.

The steady-state equations above may in principle be solved (numerically, if not analytically) to yield candidate steady-state solutions, μ_{ss} and $\mathbf{V}_{ss}$. The only solutions of interest have all entries of μ_{ss} non-negative and have $\mathbf{V}_{ss}$ being a positive semi-definite matrix.

Whether a particular (and physically meaningful) steady-state solution is attainable, i.e., whether the MFK equations converge to that steady state for initial conditions in the vicinity of the steady state, may generally be decided by studying the linear, time-invariant system of differential equations obtained by linearizing Equations 7.13 and 7.14 for small perturbations around the steady state. We do not pursue this here, but satisfy ourselves with some observations about the steady-state solutions themselves.

Because ξ is linear in the variances and covariances, so is $\mathbf{\Lambda}$. Therefore, given μ_{ss}, Equation 5.24 may be rearranged as a linear system of equations of the form

$$\big(\mathcal{M} + \frac{1}{\Omega}\mathcal{D}\big)\mathbf{v}_{ss} = \frac{1}{\Omega}\mathbf{p}, \tag{5.25}$$

where $\mathbf{v}_{ss}$ is the *covariance vector*, a column vector created by stacking the columns of the covariance matrix $\mathbf{V}_{ss}$, and the matrices $\mathcal{M}$ and $\mathcal{D}$ and the column vector $\mathbf{p}$ do not involve the entries of $\mathbf{v}_{ss}$. The matrix $\mathcal{M}$ can be explicitly written in terms of $\mathbf{M}$:

$$\mathcal{M} = \big(\mathbf{M} \otimes \mathbf{I}\big) + \big(\mathbf{I} \otimes \mathbf{M}\big), \tag{5.26}$$

where $\otimes$ denotes the Kronecker or tensor product, and $\mathbf{I}$ is the identity matrix. The entries of $\mathcal{D}$ and $\mathbf{p}$ are straightforward to determine as well. Those of $\mathcal{D}$ depend only on the stoichiometric coefficients, rate constants k_ℓ, and the rate parameter matrices $\mathbf{D}_\ell$; also, $\mathcal{D} = \mathbf{0}$ when there are no second-order reactions. Similarly, the entries of $\mathbf{p}$ depend only on the stoichiometric coefficients and the microscopic reaction rates at steady state, $\rho_\ell(\mu_{ss})$.

Equation 5.25 implies that the vector of variances and covariances at steady state is given by

$$\mathbf{v} = \big(\mathcal{M} + \frac{1}{\Omega}\mathcal{D}\big)^{-1}\frac{1}{\Omega}\mathbf{p}. \tag{5.27}$$

The indicated inverse exists if, for example, $\mathcal{M}$ is invertible (except in the unlikely circumstance that Ω is an eigenvalue of $-\mathcal{M}^{-1}\mathcal{D}$). In view of Equation 5.26, the invertibility of $\mathcal{M}$ is equivalent to the condition that no two eigenvalues of $\mathbf{M}$ sum to zero; this would be guaranteed if $\mathbf{M}$ is a Hurwitz matrix, i.e., if all its eigenvalues have strictly negative real parts. The matrix $\mathbf{M}$ turns out to govern small perturbations of μ away from μ_{ss} if one ignores the contribution of ξ to the evolution of μ in Equation 7.13, so the Hurwitz condition on $\mathbf{M}$ may be expected to hold for many systems of interest.

Assuming that $\mathcal{M}$ is invertible and Ω is large enough for all eigenvalues of $\frac{1}{\Omega}\mathcal{M}^{-1}\mathcal{D}$ to have magnitudes smaller than one, the expression in Equation 5.27 can be expanded in a convergent power series in $\frac{1}{\Omega}$:

$$\mathbf{v}_{ss} = \left(\mathbf{I} - \frac{1}{\Omega}\mathcal{M}^{-1}\mathcal{D} + ...\right)\frac{1}{\Omega}\mathcal{M}^{-1}\mathbf{p}. \tag{5.28}$$

For large enough Ω, therefore, the vector $\mathbf{v}_{ss}$ is well approximated by just $\frac{1}{\Omega}\mathcal{M}^{-1}\mathbf{p}$.

Equivalently, $\mathbf{V}_{ss}$ is well approximated by the Lyapunov equation obtained by setting $\xi = \mathbf{0}$ in the expression for $\mathbf{\Lambda}$ in Equation 5.24. Using this Lyapunov equation will yield a non-zero $\mathbf{V}_{ss}$ that causes ξ in Equation 5.23 to be non-zero, so in general the solution μ_{ss} will differ from the MAK solution. In the thermodynamic limit of $\Omega \to \infty$ with finite μ_{ss}, we get $\mathbf{V}_{ss} = \mathbf{0}$, and the MFK description (at least at steady state) becomes the MAK description.

Since the molecule number $x_i = \Omega y_i$, it follows that the variance of x_i is $\Omega^2 \sigma_i^2$. Hence, if σ_i^2 falls as $\frac{1}{\Omega}$ when $\Omega \to \infty$, the variance of the molecule number x_i will actually *grow* as Ω, i.e., the standard deviation of x_i will grow as $\sqrt{\Omega}$. The mean of x_i, on the other hand, is $\Omega \mu_i$, and thus grows as Ω for fixed μ_i.

The preceding results on the limiting behavior of y_i and x_i are consistent with those deduced in Gillespie ([60]) from the behavior of the Chemical Langevin Equation (which is the limiting form of the CME for large Ω), namely that relative fluctuations in concentration or molecule number scale as $1/\sqrt{\Omega}$ for increasing Ω.

5.5.3 Noise-to-Signal Ratio

Each diagonal element of $\mathbf{V}$ is the variance of the corresponding concentration. While its square root, the standard deviation, is a good indication of the magnitude of fluctuations about the mean concentration, a more calibrated measure is the standard deviation normalized by the mean, or alternatively, the variance normalized by the squared mean (we assumer in this subsection that all μ_i are positive). The latter may be termed the *noise-to-signal ratio* (NSR), though in the gene networks literature it is simply referred to as the *noise* ([110, 112, 138, 30]). We also define the cross-NSR for species X_i and X_j by

$$\eta_{ij} = \frac{\sigma_{ij}}{\mu_i \mu_j}, \tag{5.29}$$

and let $\{\mathbf{\Gamma}\}_{ij} = \eta_{ij}$ be the *NSR matrix*. It then follows from the previous section that for large Ω the NSR at steady state goes as $\frac{1}{\Omega}$, just like the variances and covariances.

Among the reasons why the NSR is a useful statistic, particularly in biological systems, is that concentrations can often be measured only indirectly, such as by

fluorescence microscopy. The measurable quantities are linearly proportional to the concentrations, but the proportionality constant is unknown. While the mean and fluctuations of the measurable quantities will depend on this constant, the NSR is independent of it. For this reason too, the NSR of y_i equals the NSR of x_i.

Letting $\mathbf{Q}$ be an $n \times n$ diagonal matrix with μ_i as its i-th diagonal entry, the NSR matrix is related to the covariance matrix by $\mathbf{\Gamma} = \mathbf{Q^{-1}VQ^{-1}}$, and the rate of change of $\mathbf{\Gamma}$ can thus be found from the MFK equations. The expression for the evolution of $\mathbf{\Gamma}$ can be derived directly in matrix notation, or entry-by-entry; both derivations are briefly described in Section 5.7. The expressions there make evident that the main driving term in the equation for $\frac{d\eta_{ii}}{dt}$ is

$$\frac{1}{\Omega}\frac{\sum_{\ell=1}^{L} s_{\ell i}^2 r_\ell}{\mu_i^2}. \tag{5.30}$$

This implies that lower system size, lower average concentration of X_i, high stoichiometric coefficients or high effective reaction rates all lead to higher NSR in X_i. Although these NSR sources have been identified before, the NSR equation provides a natural context that unifies them.

As an example, the NSR equation for the complex X_3 in Example A of Section IV is given by

$$\begin{aligned}\frac{d\eta_3^2}{dt} &= \frac{1}{\Omega}\left(k_1(\frac{\pi_1}{\mu_3}-1)(\frac{\pi_2}{\mu_3}-1)+\frac{k_2}{\mu_3}\right) \\ &\quad -2\eta_3^2 k_1\left(\frac{\pi_1\pi_2-\mu_3^2}{\mu_3}-\frac{1}{2\Omega}+\eta_3^2\mu_3\right),\end{aligned}$$

which in the limit of infinite system size becomes

$$\frac{d\eta_3^2}{dt} = -2\eta_3^2 k_1\left(\frac{\pi_1\pi_2-\mu_3^2}{\mu_3}+\eta_3^2\mu_3\right).$$

This equation describes a system with a stable equilibrium at $\eta_3^2 = 0$, analogous to the corresponding equation for the variance, but interestingly differs from the variance equation in that it depends only on k_1 and is thus independent of the second reaction.

5.6 Discussion

As noted earlier, the CME in Equation 4.6 is the forward Kolmogorov equation for a continuous-time discrete-state Markov process model of the system of reactions (e.g., see [49]). It constitutes a system of linear ordinary differential equations for the probabilities of each of the possible states of the system as a function of time. Unfortunately, for most systems of interest the state space of the system is combinatorially large.

Previous approaches for working with the Markov process model or the CME for general chemical systems (not restricting the number of reactant molecules in each reaction to be two or fewer) include: (1) simulating state trajectories using the SSA (or Gillespie algorithm) or some variant of it ([55, 24]), and this remains the most popular, albeit computer-intensive approach; (2) truncating the state space of the CME ([103]); (3) approximating the discrete-state Markov process by continuous-state Langevin or Ito differential equations, or correspondingly approximating the CME by a Fokker-Plank equation ([97, 57, 80, 36]), sometimes obtaining a fluctuation-dissipation relation ([80, 27]); and (4) tracing the evolution of the first m moments using various moment-closure techniques ([130, 129]). A survey of such approaches and the relation to deterministic models may be found in Érdi and Tóth ([38]).

All the above approaches are valuable; however, directly tracking the means, variances and covariances in the system generally provides more immediate insight into system behavior. Such an approach has been taken in Kaizer ([80]), Elf and Ehrenberg, ([36]) and Khammash and El Samad ([83]), but in the first two references the possibly significant effects of fluctuations on the means are ignored, while in the third reference only systems with zero-order and first-order reactions are treated. Notably, Goutsias also derived in [65] equations analogous to our MFK model described here, independently yet concurrent to our work.

The MFK equations presented in this paper require the volume Ω and the information used in the MAK description, but nothing further. They are completely specified by the same reaction diagrams and rate constants as MAK, which deter-

mine the stoichiometry and rate functions. The implementation is very direct, and amenable to numerical integration. The matrix formulation is particularly convenient in helping to structure the modeling process and organize computations.

The examples presented here demonstrate that the MFK equations are almost as simple to use as the MAK equations, while giving a much more accurate description of the system by tracking and correcting for stochastic effects. Furthermore, as illustrated by our examples, an approximation of the mean and the standard deviation bands via MFK, combined with a few simulations of sample state trajectories, often provides a much better picture of system behavior than either MFK or simulation by itself, and does this more efficiently than relying entirely on simulation.

The MFK equations may also be a useful tool when simulating chemical systems with widely different time scales (stiff systems). Methods to separate these systems into a slow and a fast process have been proposed recently ([64, 23]), where one obtains a CME for the slow process as a function of the steady-state moments of the fast process. Thus, an SSA study of the slow process requires estimation of the steady-state moments of the fast process, often a major challenge ([22]). The steady-state MFK equations (Equations 5.23 and 5.24) may then provide an efficient way to estimate the first two moments of the fast process at each iteration of the slow-scale SSA.

The form of the MFK equations resembles that of the fluctuation-dissipation relation of non-equilibrium statistical thermodynamics ([80]), which connects fluctuations to the dissipation of the system, for example associating higher reaction rates with larger fluctuations. The MFK equations differ in that they are obtained from the CME rather than from its Fokker-Plank approximation, and do not assume the means equal the macroscopic MAK solution. MFK thereby accounts for the effects of the fluctuations on the means.

Because the MFK equations are obtained by ignoring the contributions of third moments about the mean to the evolution of the variances and covariances, they only yield approximations to the first and second moments of the underlying distribution when these third moments about the mean are non-zero. It is thus expected that

the errors between the MFK results and the corresponding statistics of the true underlying distribution will grow as the third moments about the mean grow (i.e., when the underlying distribution becomes significantly asymmetrical). Rather than setting the third moments about the mean to zero, one could use other moment-closure techniques to yield approximations for these third moments as functions of the first and second moments. A particularly interesting approach may be to assume the underlying distribution is multivariate lognormal, as suggested by the results of Singh and Hespanha ([129, 130]). Other improvements may come from approximating the evolution of higher moments in the same fashion as done for second moments in MFK. As higher moments are taken into account, the error between the means, variances and covariances of the underlying distribution and their approximations should be reduced.

As the examples in Section 6.9 already demonstrate, the dynamic behavior of the MFK equations can be qualitatively different from that of the MAK equations on the one hand, and of the underlying Markov process model, on the other. A detailed study of the origins and characteristics of these differences would seem to be worthwhile. It is also of interest to better understand the dynamics of the MFK equations themselves, for instance extending chemical reaction network theory ([40, 39, 41, 38]) to this setting.

Extending the MFK equations to describe cell population behavior, where there is an ensemble of systems with the same set of chemical reactions but with different volumes and initial conditions, may also be worthwhile. Such explorations may help understand the origins and effects of the cell-to-cell variability that has recently been measured in gene networks ([110, 112, 138, 30]), and more generally, to determine whether stochastic effects need to be taken into account when describing signaling and metabolic pathways in cells. The same reasons make it worthwhile to study models of chemical reactions to systems that are not well-stirred, either because of compartmentalization (e.g., nucleus and cytoplasm) or diffusion. We start exploring this regime in Chapter 7.

5.7 Derivation of the MFK Equations

Because the CME in Equation 4.6 is formulated in terms of molecule numbers and not concentrations, we first find the evolution equations in molecule numbers and then normalize by Ω to obtain concentrations. For this purpose, let $\alpha_i = E[x_i]$ and $\beta_{ij} = E[(x_i - \alpha_i)(x_j - \alpha_j)]$ be the mean of x_i and covariance between x_i and x_j, respectively, and let α and β be the mean (column) vector and covariance matrix, respectively, with $\{\alpha\}_i = \alpha_i$ and $\{\beta\}_{ij} = \beta_{ij}$. To obtain the evolution equation for α_i, we multiply the CME by x_i and sum over all possible values of $\mathbf{x}$. The result, after algebraic simplification, turns out to be

$$\frac{d\alpha_i}{dt} = \sum_{\ell=1}^{L} s_{\ell i} E[a_\ell]. \tag{5.31}$$

The evolution equation of $E[x_i x_j]$ may be found similarly, and combined with the fact that

$$\frac{d\beta_{ij}}{dt} = \frac{dE[x_i x_j]}{dt} - \frac{d\alpha_i}{dt}\alpha_j - \alpha_i \frac{d\alpha_j}{dt}$$

to obtain the evolution equation for the covariance:

$$\begin{aligned} \frac{d\beta_{ij}}{dt} &= \sum_{\ell=1}^{L} s_{\ell i} s_{\ell j} E[a_\ell] + s_{\ell i}\big[E[a_\ell x_j] - E[a_\ell]\alpha_j\big] \\ &\quad + s_{\ell j}\big[E[a_\ell x_i] - E[a_\ell]\alpha_i\big]. \end{aligned} \tag{5.32}$$

The propensities a_ℓ for 1) zero-order, 2) first-order, 3) second-order heterogeneous and 4) second-order homogeneous reactions R_ℓ (see Gillespie[56, 60]) are respectively 1) independent of $\mathbf{x}$, proportional 2) to the number of molecules of its reactant species, 3) to the product of the molecule numbers of the two reactant species, 4) to the number of molecules of its reactant species times the same number minus one (reflecting the number of possible ways to choose any two reactant molecules from the reactant species). Accounting for their dependence on system size, propensities can then be

generally expressed as

$$a_\ell(\mathbf{x}) = k_\ell\big(\Omega b_\ell + \mathbf{c}_\ell{}'\mathbf{x} + \mathbf{x}'\frac{\mathbf{D}_\ell}{\Omega}\mathbf{x}\big). \tag{5.33}$$

where $\mathbf{c}_\ell$ may have some terms that depend on $\frac{1}{\Omega}$ (if R_ℓ describes a homogeneous second-order reaction). Using this expression, we expand the functions whose expectations we need in a Taylor series about the mean α, and find that

$$\begin{aligned} E[a_\ell] &= a_\ell(\alpha) + \varsigma_\ell(\beta) \quad \text{and} & (5.34)\\ E[a_\ell x_j] - E[a_\ell]\alpha_j &= k_\ell\Bigg(\sum_p c_{\ell p}\beta_{pj} & (5.35)\\ &\qquad +2\sum_{p,q}\frac{d_{\ell pq}}{\Omega}(\alpha_p\beta_{pq} + \frac{1}{2}\zeta_{pqj})\Bigg), \end{aligned}$$

where $c_{\ell p} = \{\mathbf{c}_\ell\}_p$, $d_{\ell pq} = \{\mathbf{D}_\ell\}_{pq}$, and where

$$\begin{aligned} \varsigma_\ell(\beta) &= k_\ell\sum_{p,q}\frac{d_{\ell pq}}{\Omega}\beta_{pq} \ \text{ and}\\ \zeta_{pqj} &= E[(x_p-\alpha_p)(x_q-\alpha_q)(x_j-\alpha_j)]. \end{aligned}$$

Substituting these expressions back into Equations 5.31 and 5.32 results in

$$\begin{aligned} \frac{d\alpha_i}{dt} &= \sum_{\ell=1}^{L} s_{\ell i}\big(a_\ell(\alpha) + \varsigma_\ell(\beta)\big) & (5.36)\\ \frac{d\beta_{ij}}{dt} &= \sum_{\ell=1}^{L}\Bigg(s_{\ell i}s_{\ell j}\big(a_\ell(\alpha) + \varsigma_\ell(\beta)\big) & (5.37)\\ &\quad +s_{\ell i}k_\ell\big[\sum_p c_{\ell p}\beta_{pj} + 2\sum_{p,q}\frac{d_{\ell pq}}{\Omega}(\alpha_p\beta_{qj} + \frac{\zeta_{pqj}}{2})\big]\\ &\quad +s_{\ell j}k_\ell\big[\sum_p c_{\ell p}\beta_{pi} + 2\sum_{p,q}\frac{d_{\ell pq}}{\Omega}(\alpha_p\beta_{qi} + \frac{\zeta_{pqi}}{2})\big]\Bigg). \end{aligned}$$

All that remains is normalizing by Ω and noting that $\mu_i = \frac{\alpha_i}{\Omega}$, $\sigma_{ij} = \frac{\beta_{ij}}{\Omega^2}$, $\rho_\ell(\mu) = \frac{1}{\Omega}a_\ell(\alpha)|_{\alpha=\mu\Omega}$, $\xi_{ij} = \frac{1}{\Omega^2}\varsigma_{ij}$ and $\gamma_{pqj} = \frac{\zeta_{pqj}}{\Omega^3}$. Doing so results in Equations 5.9 and 5.11, from which Equations 7.13 and 7.14 follow by setting $\gamma_{pqj} = 0$ for all p, q, j.

The evolution of the NSR is related to that of the mean concentrations and their variances and covariances by

$$\frac{d\eta_{ij}}{dt} = \frac{1}{\mu_i\mu_j}\frac{d\sigma_{ij}}{dt} - \eta_{ij}\big(\frac{1}{\mu_i}\frac{d\mu_i}{dt} + \frac{1}{\mu_j}\frac{d\mu_j}{dt}\big).$$

Substituting Equations 5.36 and 5.37 into the previous expression and setting third-moments to zero yields the evolution of the i,j-th entry of $\mathbf{\Gamma}$:

$$\begin{aligned}\frac{d\eta_{ij}}{dt} &= \sum_{\ell=1}^{L}\Bigg(\frac{1}{\Omega}\frac{s_{\ell i}}{\mu_i}\frac{s_{\ell j}}{\mu_j}r_\ell \\ &\quad + s_{\ell i}k_\ell\Bigg[-\eta_{ij}b_\ell + \sum_p(\eta_{pj}-\eta_{ij})c_{\ell p}\mu_p + \sum_{p,q} d_{\ell pq}\mu_p\mu_q(2\eta_{pj}-\eta_{ij}-\eta_{ij}\eta_{pq})\Bigg] \\ &\quad + s_{\ell j}k_\ell\Bigg[-\eta_{ij}b_\ell + \sum_p(\eta_{pi}-\eta_{ij})c_{\ell p}\mu_p + \sum_{p,q} d_{\ell pq}\mu_p\mu_q(2\eta_{pi}-\eta_{ij}-\eta_{ij}\eta_{pq})\Bigg]\Bigg).\end{aligned} \tag{5.38}$$

These may be arranged in a matrix form analogous to Equation 7.14. It is simpler, however, to directly derive the matrix form from the MFK equations. Letting $\mathbf{Q}$ be an $n \times n$ diagonal matrix with μ_i on its i-th diagonal, the NSR matrix is related to the covariance matrix by $\mathbf{\Gamma} = \mathbf{Q^{-1}VQ^{-1}}$, so that the rate of change of $\mathbf{\Gamma}$ is given by

$$\frac{d\mathbf{\Gamma}}{dt} = \mathbf{Q^{-1}}\frac{d\mathbf{V}}{dt}\mathbf{Q^{-1}} + \frac{d\mathbf{Q^{-1}}}{dt}\mathbf{VQ^{-1}} + \mathbf{Q^{-1}K}\frac{d\mathbf{Q^{-1}}}{dt}. \tag{5.39}$$

Substituting the rate of change of $\mathbf{V}$ from the MFK equation and using the relation $\frac{d\mathbf{Q^{-1}}}{dt} = -\mathbf{Q^{-1}}\frac{d\mathbf{Q}}{dt}\mathbf{Q^{-1}}$ results in

$$\begin{aligned}\frac{d\mathbf{\Gamma}}{dt} &= \mathbf{Q^{-1}}\big(\mathbf{MQ} - \frac{d\mathbf{Q}}{dt}\big)\mathbf{\Gamma} + \mathbf{\Gamma}\big(\mathbf{QM'} - \frac{d\mathbf{Q}}{dt}\big)\mathbf{Q^{-1}} \\ &\quad + \frac{1}{\Omega}\mathbf{Q^{-1}S\Lambda S'Q}\end{aligned} \tag{5.40}$$

where $\frac{1}{\Omega}\mathbf{Q^{-1}S\Lambda S'Q}$ is referred to as the *noise-driving* matrix. The term $\frac{d\mathbf{Q}}{dt}$ follows directly from the MFK equation for the evolution of the means, and is a function of the means and covariances. Rewriting the covariances as $\sigma_{ij} = \eta_{ij}\mu_i\mu_j$, Equation 5.40 may be written in terms of μ and the NSR, leading to an alternate closed system of

ordinary differential equations that describe μ and $\mathbf{\Gamma}$ rather than μ and $\mathbf{V}$.

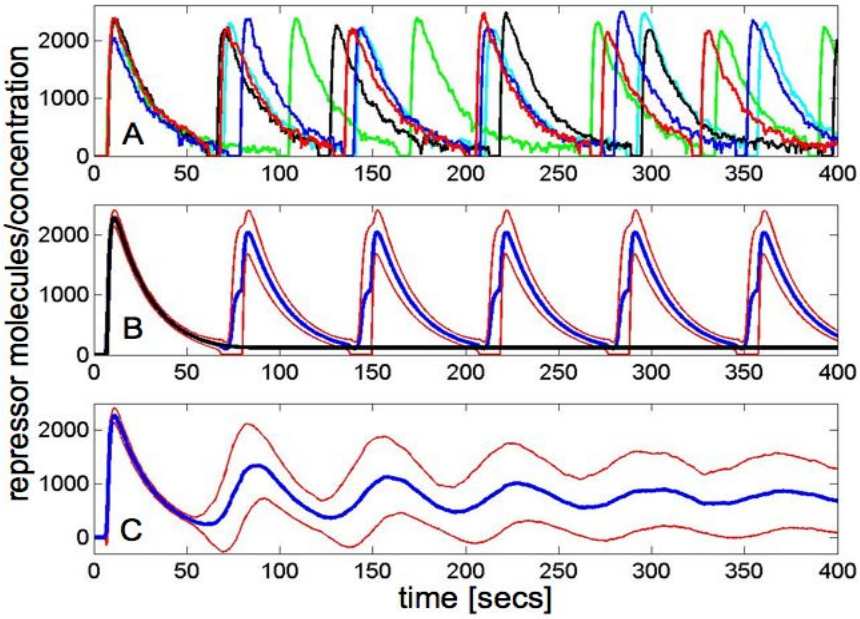

Figure 5-3: Repressor X_8 levels as a function of time for Example C. **(A)** Repressor levels for 5 different SSA simulations. All realizations show similarly persistent bursts of activity. However, these do not occur at fixed time intervals in each realization: the bursts across realizations are not synchronized, and the level of de-synchronization increases with time. **(B)** The MAK solution (shown in black) fails to predict persistent oscillations, instead settling to a fixed level after one activity burst. In contrast, the MFK equations predict persistent and periodic oscillations: the blue and red solid lines are respectively average repressor level μ_8 and the uncertainty bands $\mu_8 \pm \sigma_8$ obtained from numerical integration of the MFK equations. **(C)** The blue and red dotted lines respectively correspond to the "true" μ_8 and $\mu_8 \pm \sigma_8$, estimated from Monte Carlo simulation with 1000 runs. The average repressor level is found to oscillate with an amplitude that decreases with time, consistent with different realizations having de-synchronized activity bursts. The same parameters and initial conditions used for Fig. 5 in Vilar *et al.* [146] (although given there in the caption of Fig. 1) were used here. For the MFK equations we take $\mathbf{V}(0) = \mathbf{0}$.

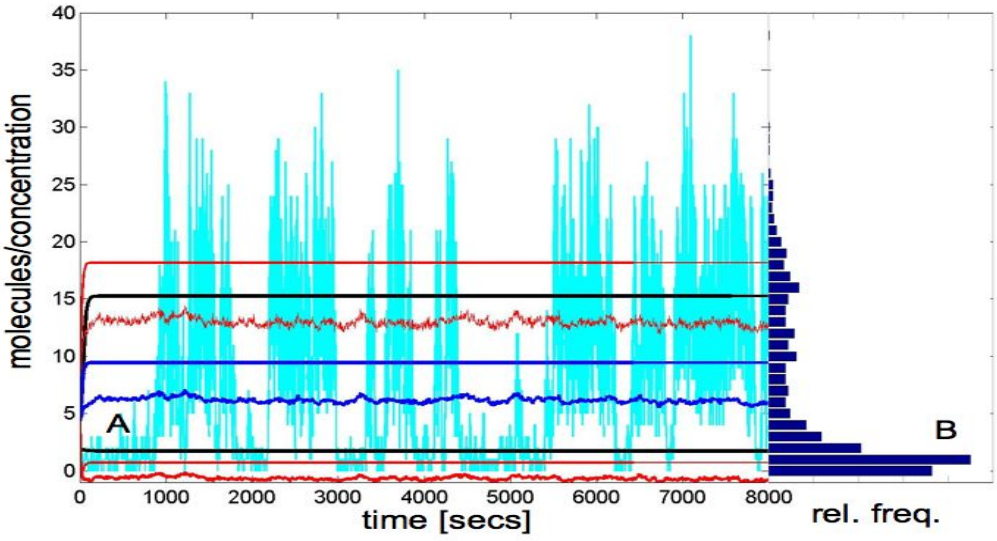

Figure 5-4: **(A)** Level of monomer X_1 as a function of time for Example D. The blue and red solid lines are respectively μ_1 and $\mu_1 \pm \sigma_1$ obtained from integrating the MFK equations. The blue and red dotted lines respectively correspond to the "true" μ_1 and $\mu_1 \pm \sigma_1$, estimated from a Monte Carlo simulation with 1000 runs, using the SSA. **(B)** Histogram of X_1 at the last time point obtained from the Monte Carlo simulation. The parameters used are $\Omega = 1, k_1 = 0.0275, k_2 = 0.1, k_3 = 15, k_4 = 0.4, k_5 = 0.1, k_6 = 0.1, k_7 = 0.145$ and $k_8 = 0.1$. The two initial conditions for both the MAK and MFK equations are $\mathbf{y}(0) = [2\ 0\ 0]'$ and $\mathbf{y}(0) = [5\ 0\ 0]'$. The first initial condition (along with $\mathbf{V}(0) = \mathbf{0}$) was used both for the Monte Carlo simulation and for the MFK curves shown here.

Chapter 6

Separation of Time Scales in Stochastic Chemical Kinetics

6.1 Summary

Systems of chemical reactions with widely different time scales are common. A classic example is the irreversible enzymatic reaction, for which a deterministic separation of time scales results in rate laws such as the Michaelis Menten. Recently, several stochastic chemical kinetics methods for separation of time scales have been developed that provide a Chemical Master Equation for the slow variables in the system. The first goal of this chapter is to present a revised method that extends the validity and improves the accuracy of existing methods by accounting for slow reactions when equilibrating the fast system. [1]This method is particularly powerful when applied to systems where the rates of the slow reactions are independent of the slow variables. The improvement we obtain for the irreversible enzymatic reaction system through our approach is analogous to the improvement in the deterministic case that one obtains by using the total quasi-steady-state approximation (tQSSA). We make a connection with the classical (deterministic) models of this system in this chapter. We also draw an analogy of the different stochastic approaches for time scale

[1]The material in this chapter has been submitted for publication in the *Journal of Chemical Physics*.

separation to a standard result of singular perturbation theory, which is widely used in deterministic models for separation of time scales. The reduced CMEs obtained by time scale separation are often used for stochastic simulation. The second main contribution of this chapter is to show how mass fluctuation kinetics (MFK) models, which give approximate evolution equations for the means, variances and covariances of the concentrations in a chemical system, can feed into time scale separation methods at a variety of stages. Particularly, we show how to obtain approximate evolution equations for the first and second moments of the slow variables, obtained either from the reduced CME or by time scale separation of the MFK model for the full system. Lastly, we show an alternative way to obtain the MFK model for the slow variables by first writing down the MFK model for the full system, followed by separation of time scales using standard results from singular perturbation theory.

6.2 Introduction

Systems of chemical reactions are often described by deterministic models that consist of ordinary differential equations that track the concentrations of the chemical species. At the small volumes and concentrations of many biological systems, however, models that capture the probabilistic nature of chemical reactions can be significantly more adequate descriptions. These models tend to be much more elaborate than their deterministic counterparts and can greatly benefit from model reduction.

In particular, systems of chemical reactions often comprise both fast and slow reactions and many state variables, yet their apparent dynamics can be described by a smaller number of slow state variables. Since experimental observations typically occur over slow time scales, resolution of the fast dynamics poses a serious computational challenge, without clear benefits. This issue is by no means unique to the stochastic limit and has received wide attention since the early days of chemical kinetics ([74, 102]). Yet while model reduction techniques through separation of time scales have played a pivotal role in the deterministic models of chemical reactions, it is only recently that such techniques have been developed for the stochastic formulation

([47, 118, 68, 23, 64, 99, 113, 122]).

The most widely used model for chemical kinetics that incorporates stochastic effects is the Chemical Master Equation (CME). Different authors ([118, 68, 23, 64]) have presented methods for obtaining reduced CMEs for the slow system variables on the basis of time scale separation, sparked by the papers of Rao and Arkin ([118]) and of Haseltine and Rawlings ([68]). The approach in Rao and Arkin ([118]) is based on assumptions that were subsequently justified in a more fundamental way in Cao ([23]) and in Goutsias ([64]) by ignoring the effects of slow reactions on the fast variables (defined as the complement of the slow variables). The resulting time scale separation methods in Cao ([23]) and in Goutsias ([64]), however, end up somewhat different from that of Rao and Arkin ([118]). Other authors have presented alternate approaches for separation of time scales, combining a truncation of the state space with time scale separation ([113]), obtaining the slow manifold for some CMEs ([122]), or applying singular perturbation analysis to the CMEs of certain systems ([99]).

The first main contribution of this chapter is to show that, for many biological systems of interest (those where the rates of the slow reactions are independent of the slow variables), a different assumption again yields a reduced order CME for the slow variables, but one that has a substantially greater range of validity than existing methods. Our reduced order CME is close to that of Rao and Arkin ([118]) but more tightly justified and more fully exploited. We illustrate this for the irreversible enzymatic reaction system. Drawing an analogy between the different stochastic approaches for time scale separation and singular perturbation theory clearly highlights the main differences between the various stochastic approaches, and suggests why our approach can be more accurate for a particular class of systems.

The reduced CMEs obtained by time scale separation are often used for stochastic simulation. Typically, however, a great deal of information is carried by the time evolution of just the means, variances and covariances of the state variables. Recent studies ([62, 65, 129]) have shown how to obtain approximate evolution equations, of complexity not much greater than that of mass action kinetics, for these first and second moments; we refer to these as models for mass fluctuation kinetics (MFK).

The second main contribution of this chapter is to show how the MFK approach can feed into time scale separation methods at a variety of stages. Specifically, it can be used to obtain:

1. the steady-state first and second moments of the fast variables needed for the reduced CME of the slow variables;

2. approximate evolution equations for the first and second moments of the slow variables, either from the reduced CME or by time scale separation of the MFK model for the full system.

6.3 Background

Systems of chemical reactions consist of n chemical species $\{X_1, ..., X_n\}$ interacting in a volume v through L reactions labeled $R_1, ..., R_L$. We let x_i and y_i be the number of molecules and the concentration, respectively, of species X_i. Molecule numbers are normalized by the system size, given by $\Omega = Av$ where A is Avogadro's number, to yield concentrations in moles per unit volume, so that $y_i = x_i/\Omega$. Letting $\mathbf{x}$ be the column vector of molecule numbers (with i-th component $\{\mathbf{x}\}_i = x_i$), referred to as the *species vector*, a firing of reaction R_ℓ changes $\mathbf{x}$ instantaneously to $\mathbf{x} + \mathbf{s}_\ell$, where $\{\mathbf{s}_\ell\}_i = s_{\ell i}$ is the *stoichiometric coefficient* of species X_i in reaction R_ℓ, and $\mathbf{s}_\ell$ is referred to as the reaction's *stoichiometry vector*.

Reaction R_ℓ may thus be specified by

$$\underline{s}_{\ell 1} X_1 + ... + \underline{s}_{\ell n} X_n \xrightarrow{k_\ell} \overline{s}_{\ell 1} X_1 + ... + \overline{s}_{\ell n} X_n, \tag{6.1}$$

where $\underline{s}_{\ell i}$ and $\overline{s}_{\ell i}$ are the numbers of molecules of species X_i that are consumed and produced, respectively, on every *firing* of the reaction, and where $\overline{s}_{\ell i} - \underline{s}_{\ell i} = s_{\ell i}$. The consumed species are referred to as *reactants*. The parameter k_ℓ in Equation 6.1 above is the *rate constant* of the reaction; its role is described later.

Under appropriate conditions, a continuous-time discrete-state Markov process ([49]) of the time evolution of $\mathbf{x}$ provides a good stochastic description of a system of

chemical reactions. Then the time evolution of $P(\mathbf{x})$, the probability distribution of the state $\mathbf{x}$ conditioned on an initial state $\mathbf{x}(t_o) = \mathbf{x}_o$, satisfies the Chemical Master Equation (see e.g., Chapter 4 or [56]), namely

$$\frac{dP(\mathbf{x})}{dt} = \sum_{\ell=1}^{L} P(\mathbf{x} - \mathbf{s}_\ell) a_\ell(\mathbf{x} - \mathbf{s}_\ell) - P(\mathbf{x}) a_\ell(\mathbf{x}). \tag{6.2}$$

The probability that reaction R_ℓ occurs in a small time interval with duration dt is approximately equal to $a_\ell(\mathbf{x})dt$; the quantity $a_\ell(\mathbf{x})$ is the *propensity* of reaction R_ℓ, which is assumed proportional to the rate constant k_ℓ. Let the *microscopic reaction rate* ρ_ℓ, with units of concentration per unit time, be the propensity of the reaction R_ℓ normalized by Ω and expressed as a function of $\mathbf{y}$ instead of $\mathbf{x}$:

$$\rho_\ell(\mathbf{y}) = \frac{a_\ell(\mathbf{x})}{\Omega}\Big|_{\mathbf{x}=\mathbf{y}\Omega}. \tag{6.3}$$

The mass fluctuation kinetics equations, which are derived from the CME in Equation 6.2, are an approximate model for the means, variances and covariances of the concentrations of all species in the system ([62, 65, 129]). The exact evolution of these moments is typically a function of higher-order moments; MFK models are approximations because they eliminate this dependence through various approaches, such as setting all higher-order central moments to zero ([62, 65]). In the latter approach, described in detail in the previous thesis chapter, and denoting by μ and $\mathbf{V}$ the mean (column) vector of species concentrations and the concentration covariance matrix, respectively, the MFK equations are given by

$$\frac{d\mu}{dt} = \mathbf{S}\mathbf{r} = \mathbf{S}\rho + \mathbf{S}\xi \tag{6.4}$$

$$\frac{d\mathbf{V}}{dt} = \mathbf{M}\mathbf{V} + \mathbf{V}\mathbf{M}' + \frac{1}{\Omega}\mathbf{S}\mathbf{\Lambda}\mathbf{S}'. \tag{6.5}$$

Here the symbol $'$ denotes vector or matrix transpose and $\mathbf{S}$ is the stoichiometry matrix; its ℓ-th column is the stoichiometry vector $\mathbf{s}_\ell$ of reaction R_ℓ.

The column vectors $\mathbf{r}, \rho, \xi$ denote the effective, average and stochastic reaction

rates, respectively. The effective rate r_ℓ of reaction R_ℓ is just a second-order approximation of the expected value of the reaction's microscopic rate in Equation 6.3; it is given by

$$\begin{aligned} r_\ell &\approx \rho_\ell(\mu) + \frac{1}{2}\sum_{i,j=1}^{n}\left(\frac{\partial^2 \rho_\ell(\mathbf{y})}{\partial y_i \partial y_j}\Big|_{\mathbf{y}=\mu}\right)\sigma_{ij} \\ &= \rho_\ell(\mu) + \xi(\mu, \mathbf{V}), \end{aligned} \tag{6.6}$$

where σ_{ij} is the covariance of y_i and y_j, the entry in row i and column j of $\mathbf{V}$. The effective rate is then equal to the microscopic rate evaluated at the mean concentrations—the average rate ρ_j—, plus a second order correction— the stochastic rate ξ_j,— which is a linear function of the concentration variances and covariances. Note, then, that when variances and covariances are negligible the evolution of the mean converges to mass action kinetics. Lastly, the j-th column of $\mathbf{M}$ is $\mathbf{S}\frac{\partial}{\partial \mu_j}\rho(\mu)$, and $\mathbf{\Lambda}$ is the fluctuation-dissipation matrix, an $L \times L$ diagonal matrix with ℓ-th diagonal entry equal to the effective rate r_ℓ of reaction R_ℓ. See Chapter 5 or [62, 65] for more on the derivation and application of the MFK equations.

6.3.1 Time Scales and Model Reduction

Reduced approximate models based on time-scale separation have been widely studied for deterministic systems of ordinary differential equations, in the context of singular perturbation theory (e.g., see [81]). Two key conditions enable the model reduction. The first requires a separation of state variables, perhaps made apparent only after a coordinate transformation, such that the initial rates of change of the variables in one set (the *slow variables*) are much smaller in magnitude than those of the variables in the other set (the *fast variables*). This condition results in a fast initial transient of the fast variables. The second condition requires the fast subsystem to be sufficiently exponentially stable throughout the experiment so that the fast transient settles to the neighborhood of a quasi-static equilibrium, and subsequently remains there as the slow system varies (over longer time scales).

The fast variables will thus quasi-equilibrate at the end of the initial transient, after which their rates of change are small, driven only by the slow variables. This allows one to construct a good reduced model on the time scale of the slow variables. If the fast transient is of interest, it too can be determined from a reduced model in which the slow variables are frozen at their initial values. (Section 6.9 briefly elaborates on this description.)

In the (deterministic or) stochastic chemical kinetics context, having slow and fast *reactions* (i.e., with small and large propensities, respectively) can similarly result in slow and fast subsystems for which reduced models can be obtained; this is the focus of the rest of the paper. The first step in such a process is to identify the slow and fast *variables* in the system. We develop a method that serves this purpose in Section 6.4, and describe how to obtain reduced models for the slow and fast subsystems in Section 6.5, finally illustrating the approach in detail in Section 6.6, on a widely used and important example. Before developing our general results, we introduce an example that shows the two time-scale behavior in its stochastic description.

Example: The Enzymatic Reaction System

Enzymatic reactions are pervasive in chemistry and biology. They are typically modeled through the reactions

$$E + S \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C \xrightarrow{k_3} E + P, \tag{6.7}$$

where E, S, C, and P respectively denote the enzyme, substrate, enzyme-substrate complex, and product; the symbols k_i denote the rate constants of the reactions. The standard setup assumes an initial non-zero concentration of enzyme and substrate, and zero concentrations for the complex and product.

The deterministic model of this system has been widely studied, where it has been found that two dynamic variables suffice to describe the state of the system. These variables are typically taken to be the complex C and the $t$otal susbtrate $S + C$, for reasons explained in Section 6.6, where this system is discussed in detail. Typically,

the complex has a fast transient that quickly reaches a quasi-steady-state after which it becomes enslaved to the total substrate, which has slow dynamics.

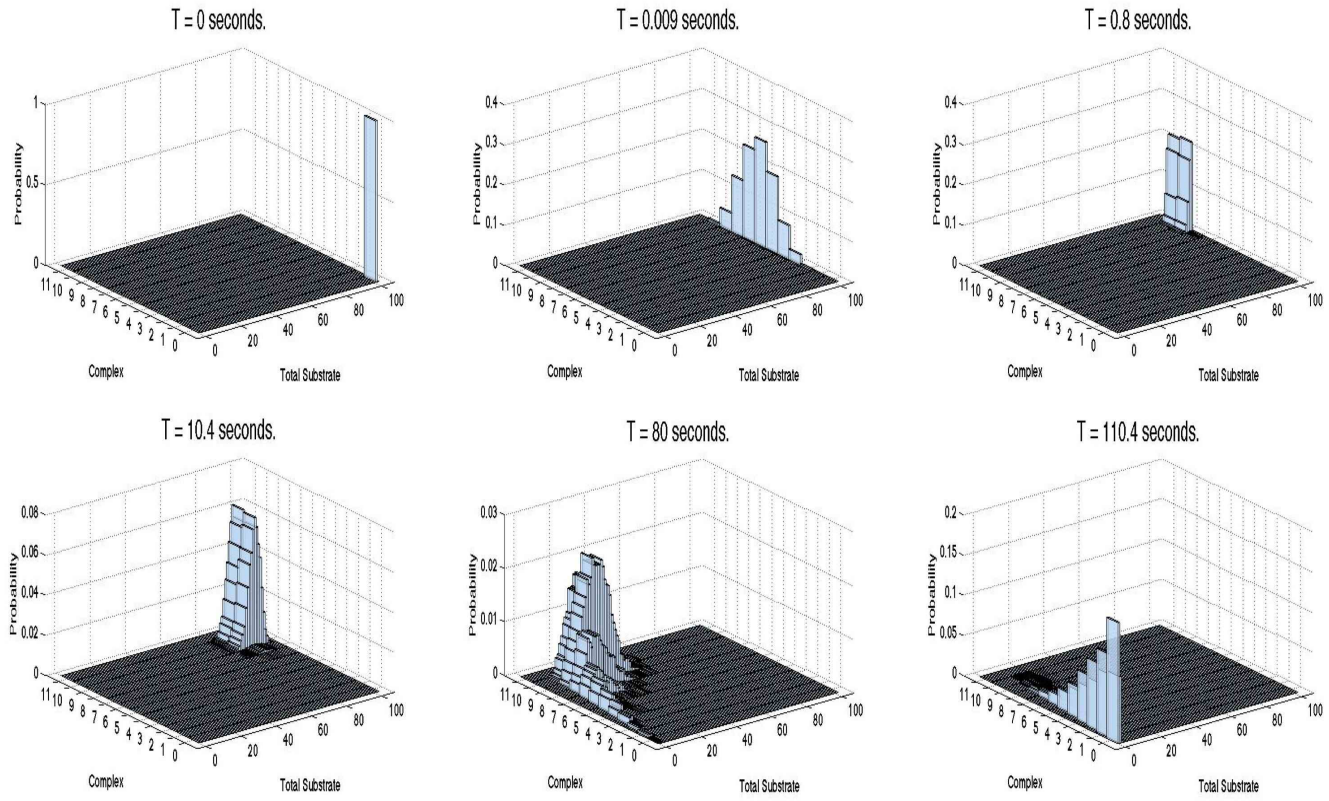

Figure 6-1: Joint probability distribution for the complex (fast variable) and total substrate (slow variable) at different time points, estimated from 20,000 samples obtained by Monte Carlo using the Gillespie algorithm. The top plots show the fast transient, where the distribution moves mostly in the dimension of the complex and quickly reaches quasi-steady state. The bottom plots correspond to the slow transient where the distribution exhibits only slow dynamics. The CME for the system is specified in Section 6.6; the parameter values used are the same as in Figure 6-3.

An analogous situation arises in the stochastic setting, where the system is described by the joint probability distribution of the complex and the total susbtrate. Figure 6-1 shows the time-evolution of this distribution for a certain choice of parameters. During the first 0.8 seconds (top plots in Figure 6-1) the distribution exhibits a fast transient where it moves mostly along the dimension of the fast variable (the complex). After this transient settles, the dynamics of the distribution becomes significantly slower (bottom plots of Figure 6-1). This separation of time scales can be

exploited as described in the rest of this chapter to significantly reduce the model complexity, and obtain approximate models for the distribution of the slow variable only, and of the fast variable conditioned on the slow variable.

6.4 Identifying Slow and Fast Variables

Given a set of slow and fast *reactions*, stoichiometry considerations alone suffice for the identification of slow and fast *variables* in a system of chemical reactions. The method discussed next can thus be applied to identify slow and fast variables in (deterministic) MAK models too, simply replacing the word *propensity* by *reaction rate*.

6.4.1 A Prototypical Example

The example we examine in this subsection has three types of variables, according to their dynamical time scales: conserved, slow, and fast. This example is discussed in Liu and Vanden-Eijnden ([95]) and in Gillespie *et al.* ([59]), and comprises the following reactions:

$$X_1 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} X_2 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} X_3 \underset{k_6}{\overset{k_5}{\rightleftharpoons}} X_4 \ . \tag{6.8}$$

Let $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]'$ denote the column vector of the species molecule numbers at some particular time t_o. Suppose we know that reactions R_1 (producing X_3 from X_2) and R_2 (producing X_2 from X_3) have propensities that are much smaller in magnitude than those of other reactions at time t_o, as assumed in Gillespie *et al* ([59]). Reactions R_1 and R_2 are accordingly termed slow, and the others fast. The corresponding stoichiometry matrix is given by

$$\mathbf{S} = \left[\begin{array}{cc|cccc} 0 & 0 & -1 & 1 & 0 & 0 \\ -1 & 1 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{array}\right] = \left[\begin{array}{cc} \mathbf{S}_s & \mathbf{S}_f \end{array}\right], \tag{6.9}$$

where the partitioning separates the columns $\mathbf{S}_s$ corresponding to slow reactions from the columns $\mathbf{S}_f$ corresponding to fast reactions.

We might expect that this partitioning of *reactions* into slow and fast gives rise to *variables* evolving at different time scales, depending on whether they get modified *only* by slow reactions, or by both fast and slow reactions. For the example system here, however, all species get modified by at least one fast reaction so they all have fast components to their dynamics.

It turns out that there actually *are* slow variables in this system, but identifying them requires considering linear transformations of the original coordinates. We describe the process in some more detail next, both to illuminate the discussion of this example in Liu and Vanden-Eijnden ([95]) and in Gillespie *et al.* ([59]), and because it serves to illustrate our proposed approach to identifying slow and fast variables in the general case.

Consider the variable $z = \alpha\mathbf{x}$, where α is a 4-component row vector. Every time reaction R_ℓ occurs, z immediately changes by $\alpha\mathbf{s}_\ell$. We now consider the stoichiometry of z for different choices of α.

First consider $z = z_1 = x_1 + x_2 + x_3 + x_4$, obtained when $\alpha = \alpha_1 = [1\ 1\ 1\ 1]$. The change of z_1 is exactly zero for all reactions, because α_1 lies in the left null space of $\mathbf{S}$ (i.e., $\alpha_1\mathbf{S} = 0$). The variable z_1 is thus a *conserved variable* that does not change throughout the experiment; the reason is that every reaction in this example involves creating one molecule of some species while destroying one molecule of another. Since the left null space of $\mathbf{S}$ has dimension equal to 1, there is no other vector α in this null space that is linearly independent of α_1, and hence no other independent conserved variable.

Conserved variables have stoichiometry coefficients that are zero regardless of what the propensities are, and can immediately be used to reduce the order of the system model, as is well known. In our example, the conservation relation implies, for instance, that x_4 can be written in terms of x_1, x_2 and x_3, so the latter three variables suffice to describe the dynamic state of the system.

Going beyond conserved variables, we are now interested in finding (combinations

of) variables with *slow* dynamics; these will correspond to quantities conserved on the fast time scale. We therefore focus only on the fast reactions, shutting off slow ones, which gives us

$$X_1 \underset{k_4}{\overset{k_3}{\rightleftharpoons}} X_2 \qquad X_3 \underset{k_6}{\overset{k_5}{\rightleftharpoons}} X_4 \ . \tag{6.10}$$

Now define slow variables via the quantities conserved in these fast reactions, i.e., via linear combinations of $\mathbf{x}$ whose stiochiometry coefficients are zero for all fast reactions. By inspection, it is evident that $x_1 + x_2$ and $x_3 + x_4$ are both conserved in the fast reactions, so these would be candidate slow variables. However, we can only select one of them for our transformed description, for example $z_2 = x_1 + x_2$, because the other would then be dependent on the selected one plus the previously obtained conservation relation:

$$x_3 + x_4 = (x_1 + x_2 + x_3 + x_4) - (x_1 + x_2) = z_1 - z_2 \ . \tag{6.11}$$

This shows that the system actually has exactly one slow variable.

There must now be exactly two fast variables to complete the system description. These can be any two linear combinations of the x_i's that are independent of those we have already picked, e.g., $z_3 = x_2$ and $z_4 = x_3$. The new system of coordinates $\mathbf{z} = [z_1 \ z_2 \ z_3 \ z_4]'$ is then related to the old via $\mathbf{z} = \mathbf{A}\mathbf{x}$, where

$$\mathbf{A} = \left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] , \tag{6.12}$$

so the new stoichiometry matrix governing changes in $\mathbf{z}$ is

$$\mathbf{B} = \mathbf{A}\mathbf{S} = \left[\begin{array}{cc|cccc} 0 & 0 & 0 & 0 & 0 & 0 \\ \hline -1 & 1 & 0 & 0 & 0 & 0 \\ \hline -1 & 1 & 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 & -1 & 1 \end{array}\right] , \tag{6.13}$$

where the first two columns correspond to the slow reactions. This transformed stoichiometry matrix $\mathbf{B}$ has a very particular structure. The row corresponding to the conserved quantity has only zeros as its entries. The row of the slow variable has only zeros in the columns corresponding to fast reactions, but is not all zero in the remaining columns (else it would correspond to a conserved variable). Similarly, the matrix block corresponding to the fast variables and the fast reactions has full row rank (else there would be additional conserved or slow variables).

What we have now accomplished is an identification of slow and fast *variables* based on the initial separation of *reactions* into slow and fast. For the approximations that will lead to our reduced-order models, we require the identity of the slow and fast variables to remain fixed over the course of the experiment. What this translates to is checking that slow variables are always only modified by slow reactions, and fast reactions are modified by at least one fast reaction throughout the experiment; such checking can be done in parallel with the main computation. One way to guarantee that separation of (slow and fast) variables is maintained is by the assuming the initial separation of reactions into slow and fast is maintained over the course of the experiment. However, it is possible that the identification of slow and fast variables can be maintained even with some switches (between slow and fast) in the classification of individual reactions over the experiment. This is because the same $\mathbf{A}$ matrix above could yield a $\mathbf{B}$ matrix with the same pattern and rank structure, even with the different assignment of the columns of $\mathbf{S}$ into those of $\mathbf{S_f}$ and $\mathbf{S_s}$ that occurs with the switched reaction(s).

It is precisely because this system has both slow and fast variables (although defined through linear combinations of the original species) that a meaningful separation-of-time-scales approximation, such as that in Gillespie *et al.* ([59]), can be developed. However, the same system may lack an appropriate partitioning into slow and fast variables if the slow and fast reactions are chosen differently. For example, consider now the case where the slow reactions in this system are given instead by R_1, R_3 and R_5. We still have the same conservation relation as before; but focusing now on only

fast reactions results in

$$X_1 \xleftarrow{k_4} X_2 \xleftarrow{k_2} X_3 \xleftarrow{k_6} X_4 \ , \tag{6.14}$$

which gives no further conserved quantities. The original system therefore has no slow variables, just one conserved quantity and three fast variables.

6.4.2 General Systems

The analysis of the preceding example can be generalized. In fact, Reinhart Heinrich had developed a very similar process to identify slow and fast variables for mass action models decades ago (e.g., see Chapter 4 in [73]). As in the example above, we assume that at the start of the experiment at time t_o, a subset of the reactions, termed slow reactions, have propensities that are much smaller in magnitude than those of the remaining reactions, which are termed fast reactions. Label the slow reactions by $R_1, ..., R_{L_s}$, with $L_s \leq L$; the remaining $L_f = L - L_s$ reactions are the fast reactions.

Given a set of slow and fast *reactions*, obtained by looking at the relative magnitudes of the reaction propensities at the start of the experiment, one must typically consider a linear change of coordinates to identify conserved, slow, and fast *variables* in the system. The following method can always be used to find a (non-unique) choice of coordinates that accomplishes this:

1. First identify a maximally independent set of linear combinations of the original variables that are conserved, no matter which reactions are running. The number n_c of such conserved variables is given by the dimension of the left null space of $\mathbf{S}$.

2. Note that the preceding combinations will all be conserved even if only the fast reactions are running. However, there may be *additional* linear combinations of these original variables, independent of those found in Step 1, that are conserved when *only the fast reactions* are running. Find a maximally independent set of such additional combinations; these are the *slow variables* in the transformed system. The number n_s of such slow variables will be the difference between

the dimensions of the left null space of $\mathbf{S}_f$ and the left null space of $\mathbf{S}$, where $\mathbf{S}_f$ denotes the columns of $\mathbf{S}$ corresponding to the fast reactions.

3. Finally, find an additional set of linear combinations of the original variables, independent of those found in Steps 1 and 2, that — taken together with the conserved variables and slow variables — provide an equivalent description to the original set of variables. These additional linear combinations will be the *fast variables* in the transformed description; more accurately, they are the non-slow variables, because they have fast components (possibly along with slow components). The number of such fast variables is given by $n - n_c - n_s$.

This identification of the conserved, slow, and fast variables is embodied in a transformation of the form $\widetilde{\mathbf{x}} = \mathbf{A}\mathbf{x}$, with $\mathbf{A}$ non-singular and of dimensions $n \times n$. Although we omit the details here, the construction of this matrix $\mathbf{A}$ can be carried out systematically with standard methods; see Appendix III. The original system can now be described as a function of only the new variables, using the relation $\mathbf{x} = \mathbf{A}^{-1}\widetilde{\mathbf{x}}$ to express the old variables as functions of the new ones, and using the matrix $\mathbf{B} = \mathbf{A}\mathbf{S}$ instead of the original stoichiometry matrix $\mathbf{S}$. By construction, $\mathbf{B}$ has the form shown in the following equation:

$$\mathbf{B} = \mathbf{A}\mathbf{S} \;=\; \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{B}_{ss} & \mathbf{0} \\ \mathbf{B}_{sf} & \mathbf{B}_{ff} \end{bmatrix} . \tag{6.15}$$

The matrices $\mathbf{B}_{sf}$ and $\mathbf{B}_{ff}$ have n_f rows, while $\mathbf{B}_{ss}$ has n_s rows. Importantly, our construction ensures that $\mathbf{B}_{ss}$ and $\mathbf{B}_{ff}$ have full row rank, which guarantees no further conserved or slow variables can be exposed by such linear transformation. This turns out to be essential for the model reduction in Section 6.5 to be accurate. The pattern of zeros in $\mathbf{B}$ makes evident that the state of the conserved variables is unmodified by any reaction, while the state of the slow variables can only be modified by slow reactions.

As already mentioned, the validity of our reduced-order models will require the

identity of the slow and fast variables to remain fixed over the course of the experiment, i.e., that the slow variables identified at the start of the experiment are always only modified by slow reactions, and fast variables are modified by at least one fast reaction, despite possible switches (between slow and fast) in the classification of individual reactions over the experiment. What this requires is that the initial $\mathbf{A}$ matrix continues to yield a $\mathbf{B}$ matrix with the same pattern and rank structure, even with the different assignment of the columns of $\mathbf{S}$ into those of $\mathbf{S_f}$ and $\mathbf{S_s}$ that occurs with the possible switched reaction(s).

Since the first n_c entries of $\widetilde{\mathbf{x}}$ are constant at their initial values, it suffices to keep the slow and fast variables as the dynamic state variables of the system. The system state therefore comprises the last $n_s + n_f$ components of $\widetilde{\mathbf{x}}$. To simplify the notation, we refer to the state of the system simply as $\mathbf{x}$ in what follows, and accordingly denote the slow and fast variables, which are typically linear combinations of the molecule numbers of the original species, by $\mathbf{x}_s$ and $\mathbf{x}_f$, respectively. Similarly, we denote simply by $\mathbf{S}$ the stoichiometry matrix for the transformed slow and fast variables, though the matrix actually comprises the last $n_s + n_f$ (block-triangular) rows of $\mathbf{B}$.

6.5 Obtaining Reduced Slow and Fast CME Models Through Time-Scale Separation

We assume as above that at the start of the experiment at time t_o, the first $L_s \leq L$ reactions are slow (i.e., have small propensities), and the remaining $L_f = L - L_s$ reactions are the fast (i.e., have large propensities). Then the CME in (4.6) can be written as

$$\begin{aligned} \frac{dP(\mathbf{x})}{dt} &= \sum_{\ell=1}^{L_s} P(\mathbf{x}-\mathbf{s}_\ell)a_\ell(\mathbf{x}-\mathbf{s}_\ell) - P(\mathbf{x})a_\ell(\mathbf{x}) \\ &\quad + \sum_{\ell=L_s+1}^{L} P(\mathbf{x}-\mathbf{s}_\ell)a_\ell(\mathbf{x}-\mathbf{s}_\ell) - P(\mathbf{x})a_\ell(\mathbf{x}) \ , \end{aligned} \tag{6.16}$$

where the first summation is over the slow reactions and the second over the fast reactions.

Further assume that the analysis in the previous section has been performed, resulting in a state vector with only slow and fast variables. To obtain the desired approximate system descriptions, partition the state vector $\mathbf{x}$ and the stoichiometry vector $\mathbf{s}_\ell$ of every reaction R_ℓ in the system into their slow and fast components:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_f \end{bmatrix}, \text{ and } \mathbf{s}_\ell = \begin{bmatrix} \mathbf{s}_{\ell s} \\ \mathbf{s}_{\ell f} \end{bmatrix}. \tag{6.17}$$

Here $\mathbf{s}_{\ell s}$ and $\mathbf{s}_{\ell f}$ are the changes in the slow and fast variables $\mathbf{x}_s$ and $\mathbf{x}_f$, respectively, with one occurrence of reaction R_ℓ. The goal of this section is to derive approximate models for the time-evolution of $P(\mathbf{x}_s)$ and $P(\mathbf{x}_f|\mathbf{x}_s)$, which jointly specify the full distribution $P(\mathbf{x})$ via

$$P(\mathbf{x}) = P(\mathbf{x}_s)P(\mathbf{x}_f|\mathbf{x}_s) \ . \tag{6.18}$$

Note from (6.18) that

$$\frac{dP(\mathbf{x})}{dt} = \frac{dP(\mathbf{x}_s)}{dt}P(\mathbf{x}_f|\mathbf{x}_s) + P(\mathbf{x}_s)\frac{dP(\mathbf{x}_f|\mathbf{x}_s)}{dt} \ , \tag{6.19}$$

an equation that will be useful in developing the reduced CMEs.

6.5.1 The Slow CME

We start by substituting (6.17) into the CME in (4.6), and then summing both sides of the equation over all possible values of the fast variables $\mathbf{x}_f$. The result, after some algebra, is the exact expression

$$\frac{dP(\mathbf{x}_s)}{dt} = \sum_{\ell=1}^{L} P(\mathbf{x}_s - \mathbf{s}_{\ell s})\overline{a}_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}) - P(\mathbf{x}_s)\overline{a}_\ell(\mathbf{x}_s) \ . \tag{6.20}$$

Equation (6.20) resembles the CME in (4.6) but with two important differences: it describes the probability distribution of just the slow variables $\mathbf{x}_s$; and it requires the

conditional propensity $\overline{a}_\ell$ of R_ℓ, which is the average of a_ℓ over the distribution of the fast variables $\mathbf{x}_f$ at time t, conditioned on the slow variables at time t :

$$\overline{a}_\ell(\mathbf{x}_s) = \sum_{\mathbf{x}_f} a_\ell(\mathbf{x}_s, \mathbf{x}_f) P(\mathbf{x}_f|\mathbf{x}_s) \ . \tag{6.21}$$

The summation in (6.20) is over *all* reactions, both slow and fast. However, note that $\mathbf{s}_{\ell s} = \mathbf{0}$ for all fast reactions, since by definition (and construction, using our procedure in Section 6.4) slow variables are not changed by fast reactions. The terms of the summation in (6.20) corresponding to the fast reactions therefore drop out, resulting in the following equation (that will become, after an additional approximation, the *slow CME*), which depends *only* on the slow reactions:

$$\frac{dP(\mathbf{x}_s)}{dt} = \sum_{\ell=1}^{L_s} P(\mathbf{x}_s - \mathbf{s}_{\ell s})\overline{a}_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}) - P(\mathbf{x}_s)\overline{a}_\ell(\mathbf{x}_s) \ . \tag{6.22}$$

Since the right side of this equation only involves the small propensities corresponding to the slow reactions, $P(\mathbf{x}_s)$ indeed varies slowly on the time-scale of the transients in the fast variables. Equations (6.21) and (6.22) were obtained by Frankovicz *et al.* ([47]) for systems with two variables. Rao and Arkin ([118]) later obtained these equations for general chemical systems as an approximate expression, though Goutsias ([64]) subsequently showed the identity is exact.

6.5.2 The Fast CME

We now develop a new reduced fast CME that approximately governs the conditional density of the fast variables $P(\mathbf{x}_f|\mathbf{x}_s)$. This CME is also needed to specify the slow CME, since the conditional propensities are a function of $P(\mathbf{x}_f|\mathbf{x}_s)$. As argued later, the fast CME we obtain is expected to be more accurate than previously published ones for systems where the slow propensities are independent of the slow variables. Importantly, many biological systems of interest, such as the irreversible enzymatic reaction discussed later, fall within this class. For other systems our approach is *a priori* not any worse than others.

As (6.22) shows, the distribution of the slow variables changes slowly within the time scale of interest for the transients in the fast variables, so we can write

$$\frac{dP(\mathbf{x}_s)}{dt} \approx 0 \tag{6.23}$$

over the duration of the fast transient. Using (6.18), (6.19), and (6.23) in the full CME in (4.6) then results in

$$\begin{aligned} &P(\mathbf{x}_s)\frac{dP(\mathbf{x}_f|\mathbf{x}_s)}{dt} = \\ &\sum_{\ell=1}^{L}\Big(P(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f - \mathbf{s}_{\ell f}) a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f - \mathbf{s}_{\ell f}) \\ &- P(\mathbf{x}_s)P(\mathbf{x}_f|\mathbf{x}_s) a_\ell(\mathbf{x}_s, \mathbf{x}_f)\Big) \ . \end{aligned} \tag{6.24}$$

In order to reduce (6.24) to a standard CME form (i.e., to the forward Kolmogorov equation for a Markov model of $\mathbf{x}_f$, e.g., see [49]), we further assume that for every reaction R_ℓ

$$\begin{aligned} &P(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f - \mathbf{s}_{\ell f})\, a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f - \mathbf{s}_{\ell f}) \\ &\approx P(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f})\, a_\ell(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f}) \ . \end{aligned} \tag{6.25}$$

This expression approximates the probability 'flow' out of any given state by that out of states one slow reaction away. For fast reactions (6.25) is exact, since $\mathbf{s}_{\ell s} = \mathbf{0}$ by definition, so the assumption only pertains to slow reactions.

For typical systems most of the entries in $\mathbf{s}_{\ell s}$ are zero except for one or two that are small integers, so the approximation in (6.25) is equivalent to requiring that the product $P(\mathbf{x}_s, \cdot)\, a_\ell(\mathbf{x}_s, \cdot)$ be relatively insensitive to small changes in the argument $\mathbf{x}_s$. Furthermore, there is a large class of relevant systems (e.g., the irreversible enzymatic reaction system discussed later in this paper) where the propensities of the slow reactions are independent of the slow variables. For these systems the approximation

in (6.25) simplifies to

$$P(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f - \mathbf{s}_{\ell f}) \;\approx\; P(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f}), \tag{6.26}$$

which only requires the more generally satisfied assumption that the probability distribution of the full state varies slowly in the slow variables (specifically, that any two states separated by a single slow reaction have similar probabilities). We assess the validity of (6.25) in a specific example in Section 6.6, where numerical simulation suggests that (6.25) is relatively valid when the support of the distribution covers several values of the slow variables for each value of the fast variable.

Substituting the expression (6.26) into (6.24) and dividing by $P(\mathbf{x}_s)$ results in the *fast CME*, given by

$$\begin{aligned}\frac{dP(\mathbf{x}_f|\mathbf{x}_s)}{dt} \;\approx\; \sum_{\ell=1}^{L}\Big(& P(\mathbf{x}_f - \mathbf{s}_{\ell f}|\mathbf{x}_s)a_\ell(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f}) \\ & - P(\mathbf{x}_f|\mathbf{x}_s)a_\ell(\mathbf{x}_s, \mathbf{x}_f)\Big),\end{aligned} \tag{6.27}$$

where the summation is over *all* reactions, and not only over the fast ones as in Cao *et al.* ([23]) and in Goutsias ([64]). Note that $\mathbf{x}_s$ here is a parameter of the distribution (i.e., a constant), and not a dynamic variable.

In Cao *et al.* ([23]), Goutsias ([64]), and in Haseltine and Rawlings ([69]), the authors obtain an approximate fast CME similar to (6.27), but dropping the terms corresponding to the slow reactions. We instead keep these terms, but have to assume (6.25). Importantly, (6.25) does not imply that slow reactions are negligible in the fast time scale. Thus, when assumption (6.25) holds our fast CME is more accurate. As already noted, assumption (6.25) is particularly likely to hold for systems where the slow reaction propensities are independent of the slow variables; for these systems (6.25) reduces to (6.26), a more plausible assumption. (A similar improvement over standard slow/fast reduction is obtained by our approach in an analogous class of singularly perturbed, linear, time invariant systems. Section 6.9 briefly develops the analogy, as it adds some insight and validation to our CME reduction approach.)

We demonstrate in Section 6.6 on a specific example that the our fast CME in (6.27) can indeed result, via its use in (6.21), in a much more accurate slow CME than alternative versions that ignore slow reactions. Rao and Arkin ([118]) also use (6.27) but assume it, rather than deriving it as we do here. Furthermore, they do not take advantage of an MFK model to obtain the conditional propensities, as we do in section 6.5.3 below, but instead obtain an assumed functional form for these propensities by invoking approximations obtained in the deterministic context.

6.5.3 Obtaining the Conditional Propensities

We can now complete the specification of the slow CME in (6.22) by using the fast CME to evaluate the conditional propensities in (6.21). The fast CME in (6.27) is guaranteed by our selection of fast variables in Section 6.4 to be driven by fast (and possibly slow) reaction propensities, so its dynamics are much faster than those of the slow CME in (6.22). Assuming the distribution of the fast variables, conditioned on the slow ones, essentially attains a stationary form on the fast time scale, and maintains it during the subsequent slow behavior, we can set

$$\frac{d}{dt}P(\mathbf{x}_f|\mathbf{x}_s) \approx 0 \tag{6.28}$$

in analyzing the slow CME.

The fast CME in (6.27) may accordingly be assumed at steady state, resulting in the expression

$$\sum_{\ell=1}^{L}\Big(P(\mathbf{x}_f - \mathbf{s}_{\ell f}|\mathbf{x}_s)a_\ell(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f}) - P(\mathbf{x}_f|\mathbf{x}_s)a_\ell(\mathbf{x}_s, \mathbf{x}_f)\Big) = 0. \tag{6.29}$$

This expression can be used to find the conditional propensities, which will now be functions of only the slow variables, not functions of time, because of the stationarity assumption in (6.28). **Equations (6.22), (6.21), and (6.29) together fully**

specify our approximate CME for the slow variables.

Using the MFK Model for the Fast CME

The MFK model for the fast CME in (6.27) can be used to obtain the conditional propensities. This model is specified by equations analogous to Equations 7.13 and 7.14 for the conditional mean concentration and covariance matrix of the fast variables; the slow variables are simply a parameter. Letting $\mathbf{y}_f = \mathbf{x}_f/\Omega$ and $\mathbf{y}_s = \mathbf{x}_s/\Omega$ denote the fast and slow concentrations, respecitvely, the resulting MFK equations are

$$\frac{d\mu_f}{dt} = \mathbf{S}_f\mathbf{r} = \mathbf{S}_f\rho + \mathbf{S}_f\xi \tag{6.30}$$

$$\frac{d\mathbf{V}_f}{dt} = \mathbf{M}_f\mathbf{V}_f + \mathbf{V}_f\mathbf{M}'_f + \frac{1}{\Omega}\mathbf{S_f}\mathbf{\Lambda}\mathbf{S}'_\mathbf{f}, \tag{6.31}$$

where $\mu_f = E[\mathbf{y}_f|\mathbf{y}_s]$ and $\mathbf{V}_f = E[(\mathbf{y}_f - \mu_f)(\mathbf{y}_f - \mu_f)'|\mathbf{y}_s]$ are the conditional mean and variance of the fast variables concentrations. The matrix $\mathbf{S}_f$ has L columns, one for each reaction; its ℓ-th column is $\mathbf{s}_{\ell f}$, the change in the fast variables from one reaction R_ℓ. The average and stochastic rates, ρ and ξ, are defined as in Equation 6.6, but keeping in mind that the slow variables now have a fixed concentration. That is, the average rate ρ_ℓ is now the microscopic rate function evaluated at $(\mu_f, \mathbf{y}_s)$ and

$$\xi_\ell = \frac{1}{2}\sum_{i,j=1}^{n_f}\left(\frac{\partial^2\rho_\ell}{\partial y_f(i)\partial y_f(j)}\Big|_{\mathbf{y}=(\mu_f,\mathbf{y}_s)}\right)\mathbf{V}_f(i,j),$$

where $y_f(i)$ denotes the i-th fast variable concentration and $\mathbf{V}_f(i,j)$ the entry in row i and column j of $\mathbf{V}_f$. The effective rate $\mathbf{r}$ is again just the sum of the average and stochastic rates. Similarly, $\mathbf{M}_f$ is just the Jacobian of $\mathbf{S}_f\rho$ with respect to the fast variables concentrations, evaluated at $(\mu_f, \mathbf{y}_s)$. The diagonal matrix Λ is the defined as before, with the effective rates as its diagonal entries.

The steady-state form of these MFK equations may be used to solve for the steady-state μ_f and $\mathbf{V}_f$, which can then be used to obtain the conditional propensities $\overline{a}_\ell$. For propensities that are at most quadratic in the fast variables, as is typically the

case, these steady-state statistics will fully specify the conditional propensities. In fact, MFK models of the fast CME may also be used to approximate conditional propensities that are non-linear non-quadratic twice-differentiable functions of the fast variables. By Taylor expanding such propensities around the means, keeping up to second-order terms and then taking expectations, the conditional propensities may always be approximated as a function of the means, variances and covariances of the fast variables.

The MFK Model For The Slow Variables

Approximate CMEs for the slow variables are typically used to obtain sample trajectories for the slow variables as a function of time through Monte Carlo simulation (e.g., via the Gillespie algorithm). Additionally, it is often desired to compute simple slow variable statistics, particularly the means, variances and covariances of the slow variables. While these statistics may be obtained through Monte Carlo simulation, the computation cost may still be prohibitive even for the reduced model. An alternate and often more efficient approach, when just the means, variances and covariances of the slow variables suffice, is to work with an MFK model of the slow variables.

Such a model may be obtained by two different approaches (see Figure 6-2). The first one works out the MFK model corresponding to the fully-specified slow CME in Equation 6.22 (see below). The second approach starts from the MFK model of the full unreduced CME (i.e., Equation 6.2), and carries out the separation of time scales in the full MFK model directly, for example, by using standard results of singular perturbation theory. Both approaches result in an MFK model for the slow variables, but while the former applies separation of time scales at the distribution level, and then uses moment truncation, the latter first applies moment truncation and then carries out the separation of time scales. The two stages involved in either approach, namely separation of time scales and moment truncation, need not be commutative, so the resulting MFK models obtained from the two approaches are slightly different. We here state the result of the first approach; the second approach is developed in

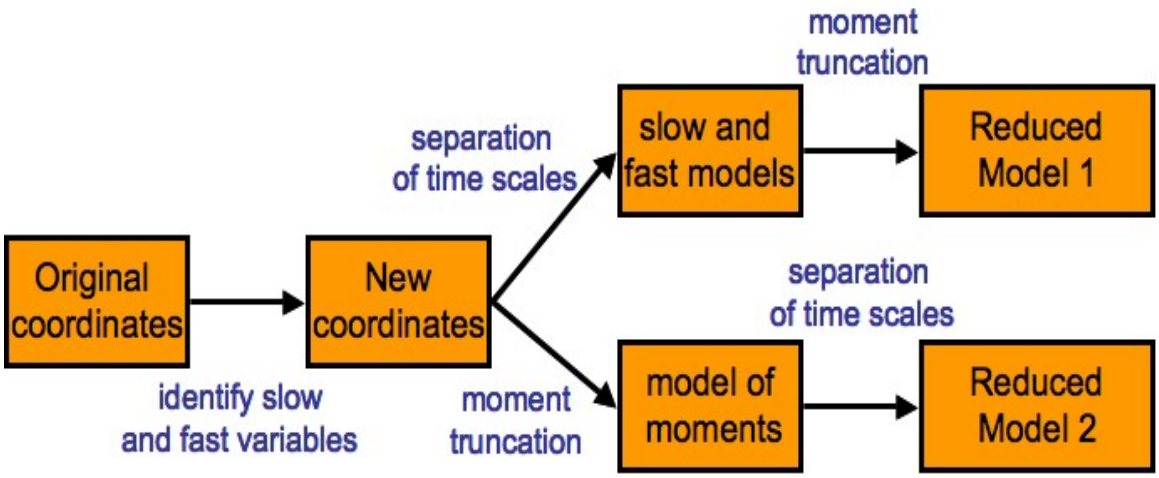

Figure 6-2: Diagram illustrating the process to obtain an MFK model for the slow variables in a chemical system. There are two complexity reduction steps: separation of time scales and moment truncation. Because these two steps are approximations that do not commute, performing these steps in a different order results in slightly different reduced models. This section describes the approach at the top of the diagram, where separation of time scales precedes moment truncation; Section 6.7 develops the alternative approach shown below.

Appendix 6.7.

The MFK model of the slow CME in Equation 6.22, like the slow CME itself, involves only the slow reactions and slow variables of the full system, with reaction propensities given by $\overline{a}_\ell$ rather a_ℓ. Letting μ_s and $\mathbf{V}_s$ respectively denote the mean vector and covariance matrix, respectively, of the slow variables concentrations, the MFK model for the slow variables is

$$\frac{d\mu_s}{dt} = \mathbf{S}_s\overline{\mathbf{r}} = \mathbf{S}_s\overline{\rho} + \mathbf{S}_s\overline{\xi} \tag{6.32}$$

$$\frac{d\mathbf{V}_s}{dt} = \mathbf{M}_s\mathbf{V}_s + \mathbf{V}_s\mathbf{M}'_s + \frac{1}{\Omega}\mathbf{S_s}\overline{\mathbf{\Lambda}}\mathbf{S}'_\mathbf{s}. \tag{6.33}$$

Here the matrix $\mathbf{S}_s$ has dimensions $n_s \times L_s$; its ℓ-th column equals $\mathbf{s}_{\ell s}$. The microscopic rate function for reaction R_ℓ is just $\overline{\rho}_\ell(\mathbf{y}_s) = \frac{\overline{a}_\ell(\mathbf{x}_s)}{\Omega}\big|_{\mathbf{x}_s=\mathbf{y}_s\Omega}$. The effective, average and stochastic rate vectors, respectively denoted by $\overline{\mathbf{r}}$, $\overline{\rho}$ and $\overline{\xi}$, now have L_s entries. The ℓ-th entry of the average rate vector is just $\overline{\rho}_\ell = \overline{\rho}_\ell(\mu_s)$, that of the stochastic rate

vector is

$$\overline{\xi}_\ell = \frac{1}{2}\sum_{i,j=1}^{n_s}\left(\frac{\partial^2\overline{\rho}_\ell}{\partial y_s(i)\partial y_s(j)}\Big|_{\mathbf{y}_s=\mu_s}\right)\mathbf{V}_s(i,j),$$

where $y_s(i)$ denotes the i-th slow variables concentration and $\mathbf{V}_s(i,j)$ the entry in row i and column j of $\mathbf{V}_s$. The matrix $\mathbf{M}_s$ is simply $\mathbf{S}_s\frac{\partial}{\partial\mu_s}\overline{\rho}(\mu_s)$, the Jacobian of $\mathbf{S}_s\overline{\rho}(\mu_s)$ with respect to the mean concentrations of the slow variables. Lastly, the $L_s \times L_s$ diagonal matrix $\overline{\Lambda}$ has $\overline{\mathbf{r}}$ in its diagonal.

These MFK equations can then be solved numerically to obtain approximations for the time-evolution of the means, variances and covariances of the slow variables.

6.6 A Slow-Time-Scale Model For The Irreversible Enzymatic Reaction

We now illustrate our proposed time-scale reduction approach on the irreversible enzymatic reaction system, specified by the reactions in (6.7). Since the same system was analyzed by Cao *et al.* ([22]), by Goutsias ([64]) and by Rao and Arkin ([118]), this example allows us to explicitly compare the different methods. Our analysis of this system is guided by concepts developed for its classical (or deterministic) counterpart, in particular the total quasi-steady-state approximation (tQSSA, see [142, 15]), which results in a rate law of wider validity than the Michaelis-Menten rate law.

6.6.1 Slow and Fast Variables

Simple inspection reveals the following two conservation relations in the system:

$$x_e + x_c = x_{e_o} \text{ and } x_s + x_c + x_p = x_{s_o}\ , \tag{6.34}$$

where a subscript denotes the species (e.g., x_c is the molecule number of species C), except for x_{e_o} and x_{s_o}, which are constants determined by the initial conditions, namely the initial total number of enzyme molecules, and of substrate molecules, re-

spectively. Further analysis of the stoichiometry matrix reveals that there are exactly two independent conserved quantities, which we take to be the ones listed in (6.34). This implies that only two dynamic variables are needed to describe the dynamics of this four-species system.

Let $x_t = x_s + x_c$ be the *total substrate*. The propensity functions for the reactions are then given by

$$\begin{aligned} a_1(x_t, x_c) &= \frac{k_1}{\Omega}(x_t - x_c)(x_{e_o} - x_c) \ , \\ a_2(x_t, x_c) &= k_2 x_c \ , \\ a_3(x_t, x_c) &= k_3 x_c \ . \end{aligned} \tag{6.35}$$

Now suppose that at the initial time $a_3 \ll \max(a_1, a_2)$, which in terms of the initial concentrations, denoted by y's, is equivalent to

$$\frac{k_3}{k_2} \ \ll \ \max\left(1, \frac{k_1}{k_2} y_c \left(\frac{y_t}{y_c} - 1\right)\left(\frac{y_{e_o}}{y_c} - 1\right)\right), \tag{6.36}$$

after some re-arrangement. Then R_3 is a slow reaction and R_1 or R_2 (or both) are fast reactions. To find the slow variables, we shut off the slow reactions (R_3 and perhaps R_1 or R_2) and note that the total substrate $x_s + x_c = x_t$ is the only additional conserved quantity independent of the two in (6.34). The total substrate is then a slow variable since it only gets modified by a slow reaction. (We could equivalently choose the product $x_p = x_{s_o} - x_t$ as a slow variable, and we do so in our discussion of the deterministic models, since this has often been the slow variable of choice in the deterministic mass action kinetics (MAK) context. In the stochastic context, however, we choose the total substrate x_t instead, as has been done in the previous literature on the total quasi-steady-state approximation.) The remaining variable, which must be linearly independent of the total substrate and of the conservation relations, is thus a fast variable. The number of enzyme-substrate complex molecules x_c satisfies the required independence condition, and is chosen as the fast variable. An analogous argument, using concentrations instead of molecule numbers and reaction

rates instead of propensities, results in the same identification of slow and fast system variables in the deterministic MAK model of the system.

6.6.2 Deterministic Models

The MAK model of the system consists of the following two coupled nonlinear ordinary differential equations for the mean concentrations of the fast and slow variables respectively:

$$\begin{aligned} \frac{d\mu_c}{dt} &= k_1\mu_e\mu_s - k_2\mu_c - k_3\mu_c \\ \frac{d\mu_p}{dt} &= k_3\mu_c \ , \end{aligned} \tag{6.37}$$

along with the conservation relations in (6.34), which can be written in terms of mean concentrations simply by replacing all x's in (6.34) by μ's. The conservation relations serve to determine μ_e and μ_s from the dynamic variables μ_c and μ_p in (6.37).

For a range of rate constants and initial conditions, the mean complex concentration μ_c in typical reactions of this type has much faster dynamics than the mean product concentration μ_p, resulting in an approximate description with the single slow dynamic variable μ_p, obtained through separation of time scales. Setting the rate of change of μ_c to zero (to capture the quasi-steady state of this fast variable that follows its rapid initial transient; see [142]) results in the approximation in

$$\frac{d\mu_p}{dt} = k_3\mu_c^{ss}, \tag{6.38}$$

where μ_c^{ss} is the steady-state concentration of the complex, which satisfies the relation

$$\mu_c^{ss} = \frac{(\mu_t + \mu_{e_o} + K_m)}{2}\left(1 \pm \sqrt{1-r}\right) , \tag{6.39}$$

where

$$r = \frac{4\mu_t\mu_{e_o}}{(\mu_t + \mu_{e_o} + K_m)^2} \ . \tag{6.40}$$

Here μ_t is the concentration of the *total substrate*, defined as $\mu_t = \mu_s + \mu_c = \mu_{s_o} - \mu_p$,

and $K_m = (k_2 + k_3)/k_1$ is the Michaelis-Menten constant.

When $r \ll 1$, which is the case when either the enzyme is scarce (i.e., $\mu_t + K_m \gg \mu_{e_o}$) or the substrate is scarce (i.e., $\mu_{e_o} + K_m \gg \mu_t$), the expressions above simplify to

$$\frac{d\mu_p}{dt} = k_3 \frac{\mu_{e_o}\mu_t}{\mu_t + \mu_{e_o} + K_m} = -\frac{d\mu_t}{dt} \ , \tag{6.41}$$

which is known as the first-order total quasi-steady-state approximation (tQSSA). The tQSSA has been quite successful at describing systems with enzymatic reactions, since it is more accurate and more generally valid than the Michaelis-Menten approximation ([63, 29]). Under conditions of substrate excess, (6.41) reduces to the Briggs-Haldane rate law ([20]), which is the Michaelis-Menten rate law when $k_3 \ll k_2$ ([74, 102]).

We will show next that applying the reduced CMEs presented in Section 6.5 to the stochastic model of the enzymatic reaction results in a more general approximation analogous to (6.38)-(6.40), where other methods ([118, 22, 64]) result in expressions related to the Briggs-Haldane or Michaelis-Menten rate laws.

6.6.3 The Stochastic Model and Slow Approximation

Assuming again that condition (6.36) holds, the total substrate is a slow variable since it is unchanged by the fast reaction(s). Similarly, since R_1, R_2, or both are fast reactions by assumption, the complex x_c is a fast variable since it is modified by at least one fast reaction.

Note that the condition in (6.36) is certainly true when $k_3 \ll k_2$, as has been often assumed in previous works. But the condition is also true more generally, for example, when $\mu_c(t_o) = 0$, or when the initial concentrations of total enzyme and total substrate are sufficiently different from one another (i.e., when either $\mu_{e_o} \ll \mu_t$ or $\mu_t \ll \mu_{e_o}$ at $t = t_o$, which is the case if $r \ll 1$ initially), since the complex concentration can never be more than $\min(\mu_{e_o}, \mu_t)$. Both these situations are commonly of interest. Additionally, the propensity of the slow reaction R_3 that modifies the slow variable is not a function of the slow variable itself, thereby falling within the class of systems

where our approach is expected to be more accurate than previous ones.

The CME for the full system is given by

$$\begin{aligned}\frac{dP(x_t,x_c)}{dt} &= P(x_t,x_c-1)a_1(x_t,x_c-1)+P(x_t,x_c+1)a_2(x_t,x_c+1) \\ &\quad +P(x_t+1,x_c+1)a_3(x_t+1,x_c+1) \\ &\quad -P(x_t,x_c)\Big(a_1(x_t,x_c)+a_2(x_t,x_c)+a_3(x_t,x_c)\Big).\end{aligned} \tag{6.42}$$

Similarly, the approximate CME for the fast variables, Equation 6.27, takes the form

$$\begin{aligned}\frac{d}{dt}P(x_c|x_t) &= P(x_c-1|x_t)a_1(x_t,x_c-1)-P(x_c|x_t)a_1(x_t,x_c) \\ &\quad +P(x_c+1|x_t)\big(a_2(x_t,x_c+1)+a_3(x_t,x_c+1)\big)-P(x_c|x_t)\big(a_2(x_t,x_c)+a_3(x_t,x_c)\big).\end{aligned} \tag{6.43}$$

Letting $\mathbf{x}_s = x_t$ and $\mathbf{x}_f = x_c$ denote the slow and fast variables of the system, (6.22) gives the approximate slow CME for the total substrate:

$$\frac{dP(x_t)}{dt} = P(x_t+1)\overline{a}_3(x_t+1)-P(x_t)\overline{a}_3(x_t)\ . \tag{6.44}$$

Equation (6.44) requires a single conditional propensity, namely $\overline{a}_3$, which is fully specified by $\overline{x}_c = E[x_c|x_t]$, the conditional mean of the complex, since by definition

$$\overline{a}_3(x_t) = k_3E[x_c|x_t] = k_3\overline{x}_c\ . \tag{6.45}$$

The MFK model corresponding to it results in the following evolution equations for $\overline{x}_c$ and $\sigma^2_{\overline{x}_c}$:

$$\frac{d\overline{x}_c}{dt} = \frac{k_1}{\Omega}\big[(x_{e_o}-\overline{x}_c)(x_t-\overline{x}_c)+\sigma^2_{\overline{x}_c}-\Omega K_m\overline{x}_c\big] \tag{6.46}$$

$$= \frac{k_1}{\Omega}\big[x_{e_o}x_t-(x_t+x_{e_o}+\Omega K_m)\overline{x}_c+\overline{x}_c^2+\sigma^2_{\overline{x}_c}\big]\ ,$$

$$\begin{aligned}\frac{d\sigma^2_{\overline{x}_c}}{dt} &= \frac{k_1}{\Omega}\big[2(3\overline{x}_c+\frac{1}{2}-x_t-x_{e_o}-\Omega K_m)\sigma^2_{\overline{x}_c} \\ &\quad +x_{e_o}x_t-(x_t+x_{e_o}-\Omega K_m)\overline{x}_c+\overline{x}_c^2\big]\ .\end{aligned} \tag{6.47}$$

These equations may then be solved (numerically) to obtain the steady-state values of $\overline{x}_c$ and $\sigma^2_{\overline{x}_c}$. Instead, we here pursue a simple approximation to draw a parallel to the classical approximations.

At the quasi-steady state following the fast transient, (6.46) gives a quadratic expression for the steady-state $\overline{x}_c^{ss}$ of $\overline{x}_c$, which can be solved to obtain

$$\overline{x}_c^{ss} = \frac{(x_t + x_{e_o} + \Omega K_m)}{2}\left(1 \pm \sqrt{1-r}\right) , \tag{6.48}$$

where

$$r = \frac{4(x_t x_{e_o} + \sigma^2_{\overline{x}_c})}{(x_t + x_{e_o} + \Omega K_m)^2} , \tag{6.49}$$

and $\sigma^2_{\overline{x}_c}$ is the steady-state variance of the complex C. Note the similarity of these expressions to (6.39) and (6.40), obtained for the deterministic case.

A sequence of further approximations helps connect the stochastic results presented here with the deterministic tQSSA results for the enzymatic system. As in the deterministic model, suppose that $r \ll 1$, and further assume that $\sigma^2_{\overline{x}_c} \ll x_t x_{e_o}$. Equation (6.48) then becomes

$$\overline{x}_c = \frac{x_t x_{e_o}}{x_t + x_{e_o} + \Omega K_m}. \tag{6.50}$$

Substituting this expression into (6.45) and using the resulting conditional propensity in (6.44) fully specifies a first approximation for the CME of the total substrate x_t.

Letting μ_t and $\sigma^2_{\mu_t}$ denote the mean and variance of the total substrate *concentration*, the MFK model for the reduced slow CME results in the evolution equations

$$\frac{d\mu_t}{dt} = -\frac{k_3 \mu_{e_o} \mu_t}{\mu_t + \mu_{e_o} + K_m} + \sigma^2_{\mu_t} k_3 \frac{\mu_{e_o}(\mu_{e_o} + K_m)}{(\mu_t + \mu_{e_o} + K_m)^3} , \tag{6.51}$$

$$\begin{aligned} \frac{d\sigma^2_{\mu_t}}{dt} &= \frac{1}{\Omega}\frac{k_3 \mu_{e_o} \mu_t}{\mu_t + \mu_{e_o} + K_m} \\ &\quad -\sigma^2_{\mu_t} k_3 \frac{\mu_{e_o}(\mu_{e_o} + K_m)}{(\mu_t + \mu_{e_o} + K_m)^3}[2(\mu_t + \mu_{e_o} + K_m) + \frac{1}{\Omega}]. \end{aligned} \tag{6.52}$$

These equations are the stochastic generalization of the deterministic tQSSA. Note

that in the absence of fluctuations (when $\sigma^2_{\mu_t} \approx 0$), the equations above reduce to the deterministic tQSSA in (6.41). The equations above may be used instead of the deterministic rate law to describe enzymatic reactions in small volumes where stochastic effects may be important.

The methods in Cao *et al.* ([22]), Goutsias ([64]) and Rao and Arkin ([118]) can all result in CME expressions identical to (6.44) and (6.45) for the total substrate in the system, but they all obtain different expressions for the complex $\overline{x}_c$. Cao *et al.* ([22]) and Goutsias ([64]) both ignore the catalytic reaction in the fast system. Cao *et al.* end up with an expression for $\overline{x}_c$ identical in form to (6.48), but with K instead of K_m, and with $\sigma^2_{\overline{x}_c}$ assumed negligible. Goutsias also ignores the slow reaction in the fast system, and after another assumption obtains the approximation $\overline{x}_c \approx \min(x_{e_o}, x_t)$; this approximation is equivalent to assuming that the minority species is in saturation. As further analysis reveals ([145]), this approximation is valid when $|x_t - x_{e_o}| \gg K_m\Omega$. Lastly, Rao and Arkin ([118]) do keep the slow reactions in the fast system, but instead of working with the fast part of the stochastic system to obtain $\overline{x}_c$, they rely on the macroscopic deterministic Michaelis-Menten approximation to set $\overline{x}_c = \frac{x_{e_o} x_t}{x_t + K_m\Omega}$, which has been shown to be valid ([128]) only when $x_t + K_m\Omega \gg x_{e_o}$.

6.6.4 Comparison of the Different Approximations

Our reduced model in (6.44), (6.45) and (6.48) results in better accuracy than the other models discussed above over a wider range of conditions. We demonstrate this next with numerical simulations. In this subsection, both the exact and approximate traces are obtained by Monte Carlo simulation using the Gillespie algorithm ([55]), the exact from the full system and the approximate from the approximation under evaluation. The quantities of interest (e.g., the average or standard deviation of the total substrate) are obtained from evenly spaced samples taken from the start of the experiment until the time when the average total substrate equals 1% of its initial value.

For certain parameter ranges, the assumptions of all the different approximations are satisfied: substrate excess, negligible catalytic reaction ($k_3 \ll k_2$), and saturation

of the minority species. In these cases all the different approximations are reasonably accurate. Figure 6-1 and 6-3 show one such example, where we used the same parameters as Figure 1(a) in Goutsias ([64]). Figure 6-1 shows the full distribution at various time points, and the top plot in Figure 6-3 shows the average total substrate as a function of time for the different approximations. Similarly, the lower plot in Figure 6-3 shows the standard deviation of the total substrate as a function of time for the various approximations. For this choice of parameters the different approximations result in adequate approximations for the evolution of the average and standard deviation of the total product.

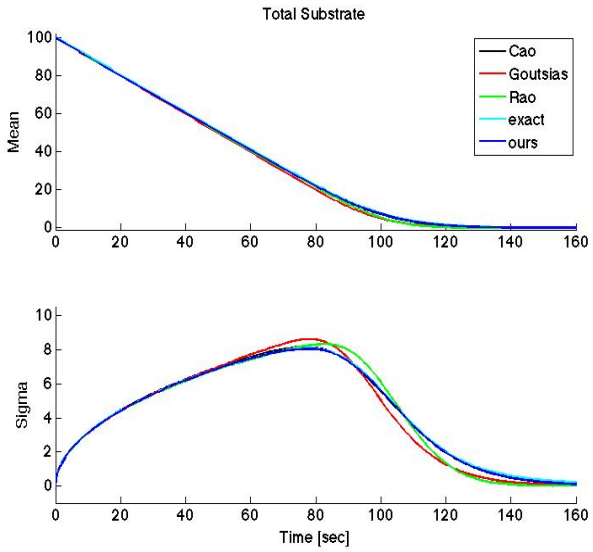

Figure 6-3: Mean and standard deviation of the total substrate x_t, obtained from $10,000$ evenly-sampled realizations for each of the stochastic descriptions considered. We chose the same parameters as those for Figure 1(a) in Goutsias ([64]), namely $\Omega = k_1 = k_2 = 1$, $k_3 = 0.1$, and initial enzyme and total substrate equal to 10 and 100 molecules, respectively. The realizations were all obtained from Monte Carlo simulation using the Gillespie algorithm.

To quantitatively assess the accuracy of an approximation, we define the error between an exact and an approximate time trace (of a mean or standard deviation), denoted by f and $\hat{f}$ respectively, by

$$\text{error} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(f_i - \hat{f}_i)^2} \quad , \tag{6.53}$$

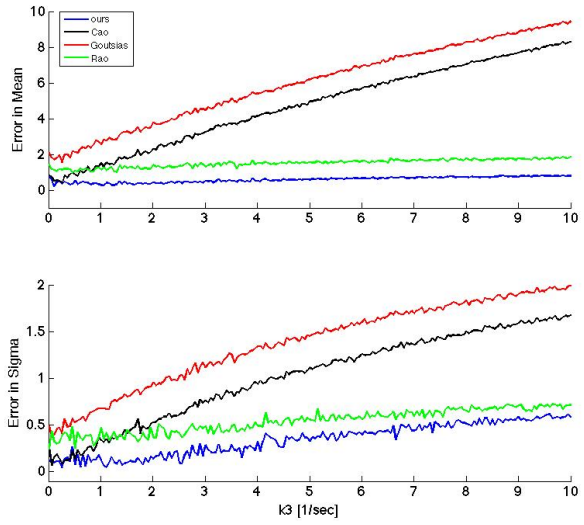

Figure 6-4: Errors of the mean and standard deviation of the total substrate x_t versus k_3 for the different approximations discussed here. Except for k_3, taken in the range shown in the plot, we used the same parameters as in Figure 6-3. We used 2,000 Monte Carlo realizations for the smallest value of k_3, and increased the number of realizations linearly, working with 20,000 of them for the largest value of k_3 shown. This was done in order to maximize accuracy, computing power and memory permitting.

where f_i and $\hat{f}_i$ respectively denote the i-th sample of the exact and approximate time trace, and m is the number of samples.

The errors for the mean total substrate in Figure 6-3, where initially there were 100 substrate molecules, can be shown to be less than 2 molecules for each of the approximations. Similarly, the errors for the total substrate standard deviation are less than 0.5 molecules. In Figure 6-4 we quantify the errors of the different approximations as a function of k_3 for the example of Figure 6-3. For each value of k_3 we first obtain time traces for the mean and standard deviation of the total substrate by Monte Carlo simulation (up to the time when the mean equaled 1% of the initial substrate), and then compute the error of the mean and of the standard deviation of each approximation as in (6.53).

As expected, the errors of the approximations proposed by Goutsias ([64]) and Cao *et al.* ([22]) grow quickly as k_3 increases, because they ignore R_3 in the fast subsystem. The approximation proposed by Rao and Arkin ([118]), on the other

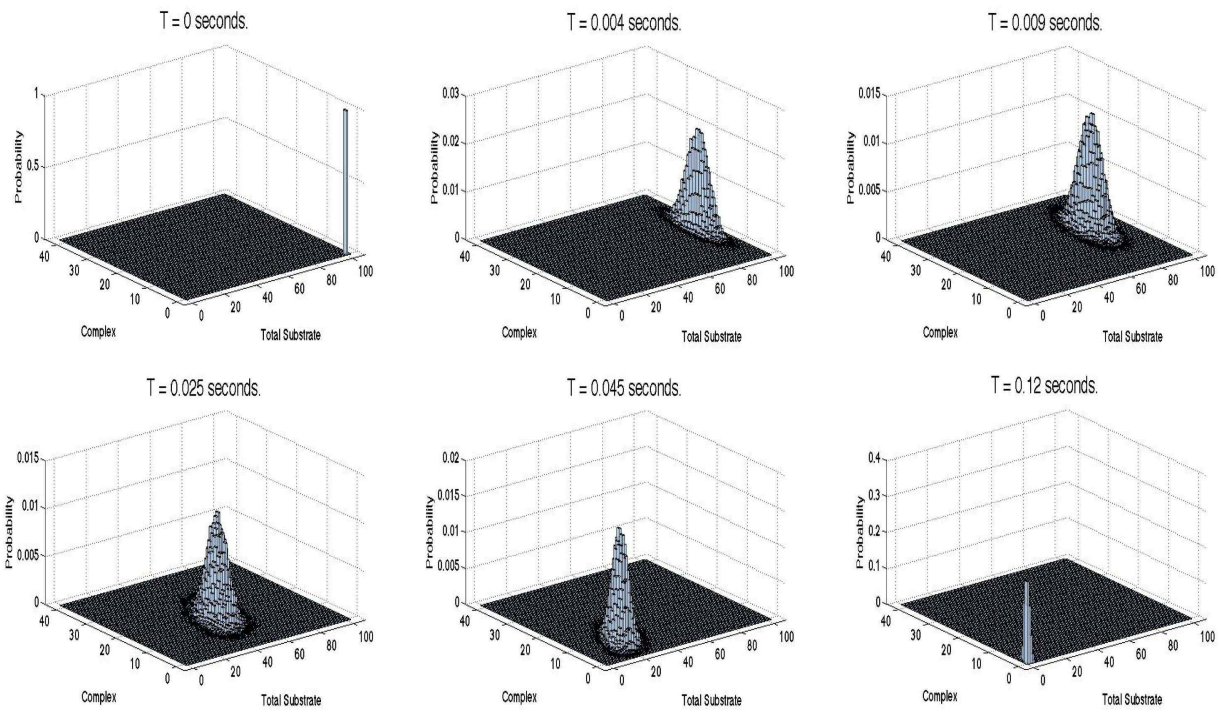

Figure 6-5: Joint probability distribution for the complex (fast variable) and total substrate (slow variable) at different time points, estimated from 60,000 samples obtained by Monte Carlo using the Gillespie algorithm. The top plots show the fast transient, where the distribution moves mostly in the dimension of the complex and quickly reaches quasi-steady state. The bottom plots correspond to the slow transient where the distribution exhibits only slow dynamics. The CME for the system is specified in Section 6.6; the parameter values used are the same as in Figure 6-6.

hand, does better than the previous ones as k_3 increases. This is again expected, since they do not ignore slow reactions in the fast system, and also because for this example the substrate is at excess, as they assume. However, for a wide range of k_3 values, the most accurate approximation is clearly the one proposed in this paper.

Our approximation remains valid where others fail. In particular, Figures 6-5 and 6-6 illustrate this is the case, for example, when $x_t = x_{e_o}$ and $k_2 = k_3$. Figure 6-5 shows the joint distribution at various time points, while Figure 6-6 shows the second order statistics of the total substrate for the exact model and the various approximations. Furthermore, Figure 6-7 shows that, in contrast to Cao *et al.* ([22]) and Goutsias ([64]), our approximation remains valid when k_3 significantly exceeds

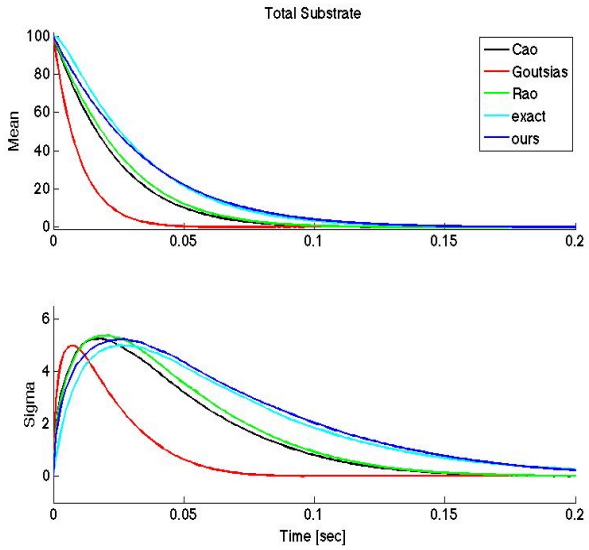

Figure 6-6: Mean and standard deviation of the total substrate x_t, obtained from $10,000$ evenly-sampled realizations for each of the stochastic descriptions considered. The system parameters are $\Omega = k_1 = 1$, $k_2 = k_3 = 100$, and initial enzyme and total substrate both equal to 100. The realizations were all obtained from Monte Carlo simulation using the Gillespie algorithm.

k_2. Interestingly, the error of the Rao and Arkin ([118]) approximation decreases as k_3 increases, although it is never smaller than the error of our approximation. The reason for this dependence is that their approximation assumes low enzyme numbers in the form of $x_{e_o} \ll x_t + K_m\Omega$. As k_3 increases, so does K_m, making their assumption more reasonable and their approximation more accurate.

6.6.5 Evaluating the Validity of Equation (6.25)

The fast CME relies on the validity of the approximation of Equation (6.25) for all (slow) reactions and possible states $\mathbf{x}$. This section discusses the validity of this approximation. We start by replacing the approximation in (6.25) by an exact identity through the incorporation of the multiplicative error $\delta_\ell(\mathbf{x})$, resulting in

$$P(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f)\, a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) \;=\; P(\mathbf{x}_s, \mathbf{x}_f)\, a_\ell(\mathbf{x}_s, \mathbf{x}_f)\big[1 + \delta_\ell(\mathbf{x}_s, \mathbf{x}_f)\big] \;. \quad (6.54)$$

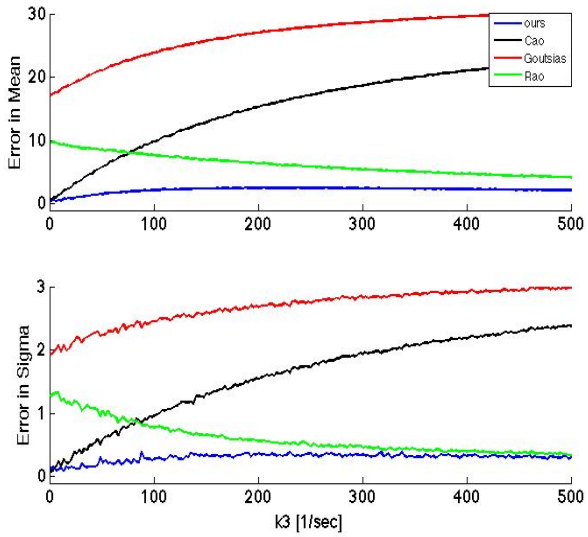

Figure 6-7: Errors of the mean and standard deviation of the total substrate x_t versus k_3 for the different approximations discussed here. Except for k_3, taken in the range shown in the plot, we used the same parameters as in Figure 6-6. We used 1,000 Monte Carlo realizations for the smallest value of k_3, and increased the number of realizations linearly, working with 10,000 of them for the largest value of k_3 shown. This was done in order to maximize accuracy, computing power and memory permitting.

For fast reactions, since $\mathbf{s}_{\ell s} = \mathbf{0}$, the error δ_ℓ is zero. For slow reactions, however, the error is generally not zero. We develop next some results regarding the value and computation of some simple statistics of the error δ_ℓ.

We think of $P(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) = p_2(\mathbf{x}_s, \mathbf{x}_f)$ and $P(\mathbf{x}_s, \mathbf{x}_f) = p_1(\mathbf{x}_s, \mathbf{x}_f)$ as two different distributions, where p_2 is obtained simply by shifting the $\mathbf{x}_s$ argument of the distribution p_1 by $\mathbf{s}_\ell$, and re-write (6.54) as

$$p_2(\mathbf{x}_s, \mathbf{x}_f)\, a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) - p_1(\mathbf{x}_s, \mathbf{x}_f)\, a_\ell(\mathbf{x}_s, \mathbf{x}_f) = p_1(\mathbf{x}_s, \mathbf{x}_f)\, a_\ell(\mathbf{x}_s, \mathbf{x}_f)\delta_\ell(\mathbf{x}_s, \mathbf{x}_f) \ . \tag{6.55}$$

Summing (6.55) over all values of $\mathbf{x}$ (recall that $\mathbf{x} = [\mathbf{x}_s' \ \mathbf{x}_f']'$) results after some algebra in the perhaps surprising statement

$$E[a_\ell(\mathbf{x})] - E[a_\ell(\mathbf{x})] = 0 = E[a_\ell(\mathbf{x})\delta_\ell(\mathbf{x})] \ , \tag{6.56}$$

so **the error is uncorrelated with the propensity, regardless of the functional form of the propensity.**

We now show how to compute the mean, variance, and mean absolute value of the error. Dividing both sides of (6.55) by $a_\ell(\mathbf{x}_s, \mathbf{x}_f)$ gives

$$p_1(\mathbf{x}_s, \mathbf{x}_f)\delta_\ell(\mathbf{x}_s, \mathbf{x}_f) = p_2(\mathbf{x}_s, \mathbf{x}_f)\frac{a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f)}{a_\ell(\mathbf{x}_s, \mathbf{x}_f)} - p_1(\mathbf{x}_s, \mathbf{x}_f) \ . \tag{6.57}$$

Summing (6.58) over $\mathbf{x}$ results in the mean error $E[\delta_\ell]$, i.e.,

$$E[\delta_\ell] = \sum_{\mathbf{x}_s, \mathbf{x}_f} \left(p_2(\mathbf{x}_s, \mathbf{x}_f)\frac{a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f)}{a_\ell(\mathbf{x}_s, \mathbf{x}_f)} - p_1(\mathbf{x}_s, \mathbf{x}_f) \right) . \tag{6.58}$$

Note that for a reaction R_ℓ where $a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) = a_\ell(\mathbf{x}_s, \mathbf{x}_f)$ for all values of $\mathbf{x}$ (e.g., if the propensity is independent of the slow variables), the expected error $E[\delta_\ell]$ is zero.

Perhaps a more informative statistic is the mean absolute value of the error $E[|\delta_\ell|]$. Taking the absolute value of (6.58) and summing the result over $\mathbf{x}$ gives

$$E[|\delta_\ell|] = \sum_{\mathbf{x}_s, \mathbf{x}_f} \left(\left| p_2(\mathbf{x}_s, \mathbf{x}_f)\frac{a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f)}{a_\ell(\mathbf{x}_s, \mathbf{x}_f)} - p_1(\mathbf{x}_s, \mathbf{x}_f) \right| \right) . \tag{6.59}$$

Note that for a reaction R_ℓ where $a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) = a_\ell(\mathbf{x}_s, \mathbf{x}_f)$ for all values of $\mathbf{x}$, $E[|\delta_\ell|]$ is bounded above by 2, which occurs when the distributions $p1$ and p_2 have a non-overlapping support.

The last statistic we discuss is the second moment of the error. Dividing (6.58) by p_1, squaring the result, multiplying by p_1 and summing over all $\mathbf{x}$ results in

$$\begin{aligned} E[\delta_\ell^2] = E[|\delta_\ell^2|] &= \sum_{\mathbf{x}_s, \mathbf{x}_f} p_1(\mathbf{x}_s, \mathbf{x}_f)\delta_\ell^2(\mathbf{x}_s, \mathbf{x}_f) \\ &= \sum_{\mathbf{x}_s, \mathbf{x}_f} \left[\frac{1}{p_1(\mathbf{x}_s, \mathbf{x}_f)} \left(p_2(\mathbf{x}_s, \mathbf{x}_f)\frac{a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f)}{a_\ell(\mathbf{x}_s, \mathbf{x}_f)} - p_1(\mathbf{x}_s, \mathbf{x}_f) \right)^2 \right] . \end{aligned} \tag{6.60}$$

Interestingly, for a reactions R_ℓ where $a_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) = a_\ell(\mathbf{x}_s, \mathbf{x}_f)$ for all values of $\mathbf{x}$, and when p_1 and p_2 are close to one another, $E[\delta_\ell^2]$ is approximately equal to twice

the Kullback-Leibler distance between p_1 and p_2. [2]

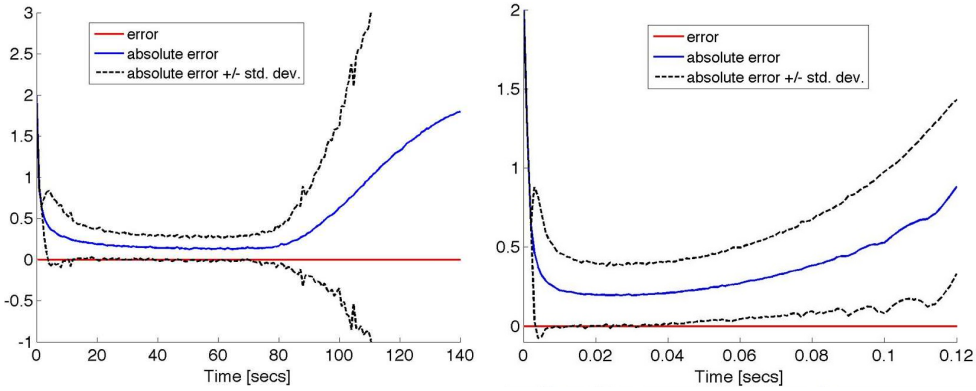

Figure 6-8: Expected error (red), and expected absolute error (blue) +/- the standard deviation (black) for the examples in Figures 6-1 (left plot) and 6-5 (right plot). Note that the expected absolute error starts at its maximum value of 2 and quickly decreases within the fast transient. Comparing these plots to Figures 6-1 and 6-5 shows that the expected absolute error is large when the support of the distribution is small.

The equations above immediately allow us to compute the discussed statistics of the error δ_ℓ, associated with reaction R_ℓ, as a function of time given the distribution $P(\mathbf{x}_s, \mathbf{x}_f) = p_1$. For simple enough systems like the enzymatic system discussed in detail in this chapter, we can estimate the distribution p_1 as a function of time from stochastic simulation, as shown in Figures 6-1 and 6-5, and use it to compute the statistics of the error δ discussed above. We focus on the error of the third reaction, which is assumed to be slow. Note that for this reaction $a_3(\mathbf{x}_s - \mathbf{s}_{\ell s}, \mathbf{x}_f) = a_3(\mathbf{x}_s, \mathbf{x}_f)$ since the reaction is independent of the slow variable. To compute the associated error statistics, we simply shift p_1 by $\mathbf{s}_\ell = 1$ to create p_2, and apply the formulas above for $E[\delta]$, $E[|\delta|]$ and combine them with $E[\delta^2]$ to obtain the standard deviation of the error. The results are shown in Figure 6-8. As derived above, the expected error is zero. Comparing the plots in Figure 6-8 to the distributions in Figures 6-1 and 6-5 shows that the expected absolute error $E[|\delta|]$ is maximum when the support

[2]The Kullback-Leibler distance between two distributions $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$, given by $D(p_1||p_2) = -\sum_{\mathbf{x}} p_1(\mathbf{x}) \log \frac{p_2(\mathbf{x})}{p_1(\mathbf{x})}$, provides a quantitative way to measure the distance between $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$. It is zero only when $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ are identical, and is positive otherwise (e.g., see [31]). Assuming $p_2(\mathbf{x})$ is close to $p_1(\mathbf{x})$ and Taylor expanding the log in the expression for $D(p_1||p_2)$ up to second order, results after some algebra in $D(p_1||p_2) \approx \frac{1}{2} \sum_{\mathbf{x}} \frac{1}{p_1(\mathbf{x})} (p_1(\mathbf{x}) - p_2(\mathbf{x}))^2$.

of the distribution is small, which in these examples happens when all the probability is concentrated at a single value of $\mathbf{x}$ (at time zero since all realizations were given the same initial condition, and at the end of the experiment when all substrate molecules have been converted to product).

We leave to further work to better assess how δ, the error in (6.25), propagates through our separation of time-scales approach and ultimately impacts the accuracy of the resulting approximate slow and fast models.

6.7 The MFK Equations for the Slow Species

In this section we present an alternate approach to obtain an MFK model for the means, variances and covariances of the slow variables; one that allows a direct application of the standard singular perturbation theory results ([81]). We start from the MFK model of the full system (i.e., corresponding to the unreduced CME in Equation 6.2), given by Equations 7.13 and 7.14. Now assume that all conservation relations have been taken into account (i.e., so that in the resulting coordinates $\mathbf{S}$ has a left null space of dimension zero). Further assume that the effective rate vector $\mathbf{r}$ at the initial time $t = t_o$ consists of slow rates of order $\epsilon \ll 1$, those corresponding to the slow reactions, and fast rates of order 1 corresponding to the fast reactions. Further assume that the slow reactions only change the state of a subset of the variables, the slow variables. These conditions lead to a partitioning of the state space into slow and fast variables, so that

$$\mathbf{x} = \left[\begin{array}{c} \mathbf{x}_s \\ \hline \mathbf{x}_f \end{array}\right], \text{ and } \mathbf{r} = \left[\begin{array}{c} \epsilon\mathbf{r}_s \\ \hline \mathbf{r}_f \end{array}\right] = \left[\begin{array}{c} \epsilon(\rho_s + \xi_s) \\ \hline \rho_f + \xi_f \end{array}\right],$$

where $\mathbf{r}, \rho, \xi$ denote the effective, average and stochastic reaction rates of the MFK equations of the full CME (Equation 6.2). The stoichiometry matrix may then be similarly partitioned as

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S_s} & \mathbf{S_{sf}} \\ \hline \mathbf{S_{fs}} & \mathbf{S_f} \end{array}\right],$$

where the i-th column of $\mathbf{S_s}$ ($\mathbf{S_{sf}}$) is the net change vector in the slow variables caused by a single occurrance of the i-th slow (fast) reaction. Similarly $\mathbf{S_f}$ ($\mathbf{S_{fs}}$) contain the net changes of the fast variables when a fast (slow) reaction occurs. The MFK equation for the average concentration vector, Equation 7.13, becomes

$$\begin{aligned} \frac{d}{dt}\mu &= \frac{d}{dt}\left[\begin{array}{c}\mu_s\\ \hline \mu_f\end{array}\right] = \mathbf{S}\mathbf{r} = \left[\begin{array}{c|c}\mathbf{S_s} & \mathbf{S_{sf}}\\ \hline \mathbf{S_{fs}} & \mathbf{S_f}\end{array}\right]\left[\begin{array}{c}\epsilon\mathbf{r}_s\\ \hline \mathbf{r}_f\end{array}\right] \\ &= \left[\begin{array}{c}\epsilon\mathbf{S_s}\mathbf{r_s} + \mathbf{S_{sf}}\mathbf{r_f}\\ \hline \epsilon\mathbf{S_{fs}}\mathbf{r_s} + \mathbf{S_f}\mathbf{r_f}\end{array}\right]. \end{aligned} \tag{6.61}$$

For the slow variables to have slow dynamics, their rate of change must be $O(\epsilon)$ which will be the case only if $\mathbf{S_{sf}} = \mathbf{0}$, i.e., if no fast reaction causes a net change in any of the slow variables, as required by our definition of slow variables. However, note that if the original variables do not satisfy this condition, there may still be a change of variables that will. From now on, we then assume that

$$\mathbf{S} = \left[\begin{array}{c|c}\mathbf{S_s} & \mathbf{0}\\ \hline \mathbf{S_{fs}} & \mathbf{S_f}\end{array}\right].$$

We similarly partition the covariance matrix $\mathbf{V}$ as follows:

$$\mathbf{V} = \left[\begin{array}{c|c}\mathbf{V_s} & \mathbf{V_{sf}}\\ \hline \mathbf{V'_{sf}} & \mathbf{V_f}\end{array}\right].$$

Here $\mathbf{V_s}$ and $\mathbf{V_f}$ are the covariance matrices of the fast and slow species, respecitvely, and $\mathbf{V_{sf}}$ is the mixed covariance matrix, with each entry being the covariance between

a slow and a fast species. The matrix $\mathbf{M}$ may be similarly partitioned as follows:

$$\mathbf{M} = \mathbf{S}\frac{\partial \rho}{\partial \mu} = \mathbf{S}\left[\begin{array}{c|c} \epsilon\frac{\partial \rho_s}{\partial \mu_s} & \epsilon\frac{\partial \rho_s}{\partial \mu_f} \\ \hline \frac{\partial \rho_f}{\partial \mu_s} & \frac{\partial \rho_f}{\partial \mu_f} \end{array}\right] \tag{6.62}$$

$$= \left[\begin{array}{c|c} \epsilon\mathbf{S_s}\frac{\partial \rho_s}{\partial \mu_s} & \epsilon\mathbf{S_s}\frac{\partial \rho_s}{\partial \mu_f} \\ \hline \epsilon\mathbf{S_{fs}}\frac{\partial \rho_s}{\partial \mu_s} + \mathbf{S_f}\frac{\partial \rho_f}{\partial \mu_s} & \epsilon\mathbf{S_{fs}}\frac{\partial \rho_s}{\partial \mu_f} + \mathbf{S_f}\frac{\partial \rho_f}{\partial \mu_f} \end{array}\right] \tag{6.63}$$

$$= \left[\begin{array}{c|c} \epsilon\mathbf{M_s} & \epsilon\mathbf{M_{sf}} \\ \hline \epsilon\mathbf{M_{fsss}} + \mathbf{M_{ffs}} & \epsilon\mathbf{M_{fssf}} + \mathbf{M_f} \end{array}\right]. \tag{6.64}$$

The fluctuation-dissipation matrix can also be partitioned, as in

$$\mathbf{\Lambda} = \left[\begin{array}{c|c} \epsilon\mathbf{\Lambda_s} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{\Lambda_f} \end{array}\right], \text{ resulting in} \tag{6.65}$$

$$\frac{1}{\Omega}\mathbf{S\Lambda S'} = \frac{1}{\Omega}\left[\begin{array}{c|c} \epsilon\mathbf{S_s\Lambda_s S'_s} & \epsilon\mathbf{S_s\Lambda_s S'_{fs}} \\ \hline \epsilon\mathbf{S_{fs}\Lambda_s S'_s} & \epsilon\mathbf{S_{fs}\Lambda_s S'_{fs}} + \mathbf{S_f\Lambda_f S'_f} \end{array}\right].$$

Combining all the previous expressions, the covariance MFK equation 7.14 the following expression for $\frac{d\mathbf{V}}{dt}$:

$$\left[\begin{array}{c|c} \epsilon\big(\mathbf{M_sV_s} + \mathbf{V_sM'_s} + \mathbf{M_{sf}V'_{sf}} + \mathbf{V_{sf}M'_{sf}}\big) & \epsilon\big(\mathbf{M_sV_{sf}} + \mathbf{M_{sf}V_f} + \mathbf{V_sM'_{fsss}} + \mathbf{V_{sf}M_{fssf}}\big) \\ & \quad + \mathbf{V_sM'_{ffs}} + \mathbf{V_{sf}M'_f} \\ \hline \epsilon\big(\mathbf{V'_{sf}M'_s} + \mathbf{V_fM'_{sf}} + \mathbf{M_{fsss}V_s} + \mathbf{M'_{fssf}V_{sf}}\big) & \epsilon(\mathbf{M_{fsss}V_{sf}} + \mathbf{V'_{sf}M_{fsss}} + \mathbf{M_{fssf}V_f} + \mathbf{V_fM'_{fssf}}) \\ \quad + \mathbf{M_{ffs}V_s} + \mathbf{M_fV'_{sf}} & \quad + \mathbf{M_{ffs}V_{sf}} + \mathbf{V'_{sf}M'_{ffs}} + \mathbf{M_fV_f} + \mathbf{V_fM'_f} \end{array}\right]$$
$$+\frac{1}{\Omega}\left[\begin{array}{c|c} \epsilon\mathbf{S_s\Lambda_s S'_s} & \epsilon\mathbf{S_s\Lambda_s S'_{fs}} \\ \hline \epsilon\mathbf{S_{fs}\Lambda_s S'_s} & \epsilon\mathbf{S_{fs}\Lambda_s S'_{fs}} + \mathbf{S_f\Lambda_f S'_f} \end{array}\right]. \tag{6.66}$$

Equations 6.61 (with $\mathbf{S}_{sf}$ set to $\mathbf{0}$) and 6.66 are already in a form where the standard result of singular perturbation theory, Equation 6.88, which yields a first-order approximation in ϵ, may be directly applied. Some analysis, described shortly, yields

the reduced model for the slow variables given by

$$\frac{d\mu_{\mathbf{s}}}{dt} = \epsilon \mathbf{S_s} \mathbf{r}_s \tag{6.67}$$

$$\frac{d\mathbf{V_s}}{dt} = \epsilon\big(\overline{\mathbf{M}}_{\mathbf{s}}\mathbf{V_s} + \mathbf{V_s}\overline{\mathbf{M}}_{\mathbf{s}}' + \frac{1}{\Omega}\mathbf{S_s}\mathbf{\Lambda_s}\mathbf{S_s'}\big), \tag{6.68}$$

where

$$\overline{\mathbf{M}}_{\mathbf{s}} = \mathbf{M_s} - \left(\mathbf{S_s}\frac{\partial \rho_{\mathbf{s}}}{\partial \mu_{\mathbf{f}}}\left[\mathbf{S_f}\frac{\partial \rho_{\mathbf{f}}}{\partial \mu_{\mathbf{f}}}\right]^{-\mathbf{1}}\mathbf{S_f}\frac{\partial \rho_{\mathbf{f}}}{\partial \mu_{\mathbf{s}}}\right), \tag{6.69}$$

assuming the inverse in the expression exists.

The previous equations have an MFK form but with two important differences. First, the $\mathbf{M}_s$ matrix of the original system becomes the matrix $\overline{\mathbf{M}}_s$ above. Second, the slow rates $\mathbf{r}_s$ and the matrix $\overline{\mathbf{M}}_s$ in Equations 6.67 and 6.68 generally depend on the concentrations, variances and covariances of both fast and slow species. The terms in these equations that involve the fast species may be obtained from the algebraic relations

$$\mathbf{0} = \mathbf{S_f}\mathbf{r}_f, \tag{6.70}$$

$$\mathbf{0} = \mathbf{V_s}\mathbf{M}_{\mathbf{ffs}}' + \mathbf{V_{sf}}\mathbf{M}_{\mathbf{f}}', \text{ and} \tag{6.71}$$

$$\mathbf{0} = \mathbf{M_{ffs}}\mathbf{V_{sf}} + \mathbf{V}_{\mathbf{sf}}'\mathbf{M}_{\mathbf{ffs}}' + \mathbf{M_f}\mathbf{V_f} + \mathbf{V_f}\mathbf{M}_{\mathbf{f}}' + \frac{1}{\Omega}\mathbf{S_f}\mathbf{\Lambda_f}\mathbf{S_f'}. \tag{6.72}$$

The previous equations are obtained from Equations 6.61 and 6.66 by setting the rate of change of the fast variables of the system (i.e., $\mu_f, \mathbf{V}_f$ and $\mathbf{V}_{sf}$) to zero, while setting ϵ to zero. Note that Equations 6.70, 6.71 and 6.72 are exactly analogous to Equation 6.88. Perhaps a better approximation would not set ϵ to zero on the right side of the evolution equations of the fast variables, resulting instead in expressions analogous to Equation 6.90. Systems where the slow reactions are independent of the slow species may particularly benefit from this approach. Simple inspection of Equation 6.66, however, reveals that these expressions can be significantly more com-

plicated than Equations 6.70, 6.71 and 6.72. The MFK method described here, which uses Equations 6.70, 6.71 and 6.72, should then be equivalent in terms of accuracy to the methods of Cao *et al* ([23]), Goutsias ([64]) and Rao and Arkin ([118]).

The reduced evolution equations for the slow variables, Equations 6.67 and 6.68, are obtained as well from the full MFK description in Equations 6.61 and 6.66, which directly gives Equation 6.67 for the mean concentrations and

$$\begin{aligned}\frac{d\mathbf{V_s}}{dt} &= \epsilon\big(\mathbf{M_s}\mathbf{V_s} + \mathbf{V_s}\mathbf{M_s'} + \mathbf{M_{sf}}\mathbf{V_{sf}'} + \mathbf{V_{sf}}\mathbf{M_{sf}'} \\ &\quad + \frac{1}{\Omega}\mathbf{S_s}\mathbf{\Lambda_s}\mathbf{S_s'}\big), \end{aligned} \tag{6.73}$$

for its variances and covariances. We then note that Equation 6.71 implies that the cross covariances between slow and fast species are given by

$$\mathbf{V_{sf}} = -\mathbf{V_s}\mathbf{M_{fs}'}\mathbf{M_f'^{-1}} = -\mathbf{V_s}\left(\frac{\partial \rho_\mathbf{f}}{\partial \mu_\mathbf{s}}\right)'\mathbf{S_f'}\left[\left(\frac{\partial \rho_\mathbf{f}}{\partial \mu_\mathbf{f}}\right)'\mathbf{S_f'}\right]^{-1},$$

assuming the inverse in the expression exists. This should often be the case since $\mathbf{M_f}$ is the Jacobian of the fast system, which must be stable for the singular perturbation result to hold, so $\mathbf{M_f}$ should be Hurwitz. The previous expression may then be substituted into Equation 6.73 to obtain Equation 6.68.

Connection to Distribution Based Approach

Is the reduced description of the slow variables in Equations 6.67 and 6.68 compatible with the MFK equations for the slow variables that one would obtain from the reduced CME in Equation 6.22? The former computes the MFK equations for the full system, and only then applies separation of time scales; the latter first applies separation of time scales at the CME level, and then one may write down the MFK equations for the reduced CME. Both approaches can then result in MFK-type of equations for the slow species. Also, while our distribution approach is analogous to not setting ϵ to zero in the evolution equations for the fast variables, the MFK-based approach described here does set ϵ to zero in the analogous equations. Therefore, the distribution approach

that keeps ϵ in the fast system should be expected to be more accurate. But are these two methods otherwise compatible? It turns out that they are, but some brief analysis is required to see their similarity.

The MFK equations obtained from the slow CME of Equation 6.22 result in an evolution equation for $\mathbf{V}_s$ where the matrix multiplying $\mathbf{V}_s$ is given by $\mathbf{S}_s \frac{d\rho_s}{d\mu_s}$, the slow part of the stoichiometry matrix times the total derivative of the average rates with respect to the slow variables. The average rate vector ρ_s is the same in both approaches, namely the vector of slow reaction propensities evaluated at the mean concentrations and normalized by the system size Ω. The previous matrix appears to be different from $\overline{\mathbf{M}}_\mathbf{s}$, the matrix playing the same role in the MFK-based approach. However, it turns out that when one takes into account the constraints imposed by Equation 6.70, the two matrices are equivalent. To make this relation explicit, we start with the following expression for the total derivative of $\mathbf{S_s}\rho_s$ with respect to μ_s:

$$\begin{aligned} \frac{d\mathbf{S_s}\rho_s}{d\mu_s} &= \frac{\partial \mathbf{S_s}\rho_s}{\partial \mu_s} + \frac{\partial \mathbf{S_s}\rho_s}{\partial \mu_f}\frac{d\mu_f}{d\mu_s} \\ &= \mathbf{M}_s + \mathbf{S_s}\frac{\partial \rho_s}{\partial \mu_f}\frac{d\mu_f}{d\mu_s}. \end{aligned} \tag{6.74}$$

Here d and ∂ denote total and partial differentials, respectively. Now, from Equation 6.70 we have that

$$\frac{d\mathbf{S_f}\rho_f}{d\mu_s} = \mathbf{S_f}\frac{\partial \rho_f}{\partial \mu_s} + \mathbf{S_f}\frac{\partial \rho_f}{\partial \mu_f}\frac{d\mu_f}{d\mu_s} = 0, \tag{6.75}$$

so that

$$\frac{d\mu_f}{d\mu_s} = -\left(\mathbf{S_f}\frac{\partial \rho_f}{\partial \mu_f}\right)^{-1}\mathbf{S_f}\frac{\partial \rho_f}{\partial \rho_s}. \tag{6.76}$$

Substituting the above expression into Equation 6.74 results in the expression for $\overline{\mathbf{M}}$ in Equation 6.69. This shows that $\overline{\mathbf{M}}$ is just $\mathbf{S}_s\frac{d\rho_s}{d\mu_s}$; both approaches (i.e., applying the separation of time scales at the distribution level and then obtaining the MFK equations, versus first obtaining the MFK equations and then applying the separations of time scales) are in this sense equivalent. They both arrive at a similar expression for the evolution of the covariance matrix of the slow species, even though in practice

the approximations required to obtain analytical results in each of the two methods may lead to actual different resulting expressions (as the next example illustrates).

6.7.1 Example: The Enzymatic Reaction

We demonstrate the MFK-based method in the irreversible enzymatic reaction discussed previously. Assume that the rate of the catalytic reaction is order ϵ, and consider the same dynamic variables as above, namely the total substrate x_t and the complex x_c. Then $\mathbf{S_s} = -1$, $\mathbf{S_f} = \begin{bmatrix} 1 & -1 \end{bmatrix}$, $\mathbf{S_{sf}} = [\ 0 \quad 0\ ]$ and $\mathbf{S_{fs}} = -1$. Let y_s denote the concentration of x_t, the slow species, and y_f that of x_c, the fast species. We then have

$$\rho_f = \begin{bmatrix} k_1(y_{e_o} - y_f)(y_s - y_f) \\ k_2 y_f \end{bmatrix}, \text{ and } \rho_s = k_3 y_f,$$

so that $\mathbf{M_s} = 0$, and

$$\overline{\mathbf{M_s}} = -k_3 \frac{y_{e_o} - \mu_f}{K + y_{e_o} + \mu_s - 2\mu_f}, \text{ where } K = \frac{k_2}{k_1}.$$

The slow MFK equations, Equations 6.67 and 6.68, then simply become

$$\frac{d\mu_s}{dt} = -k_3 \mu_f \tag{6.77}$$

$$\frac{d\sigma_s^2}{dt} = -2k_3 \frac{y_{e_o} - \mu_f}{K + y_{e_o} + \mu_s - 2\mu_f} \sigma_s^2 + \frac{1}{\Omega} k_3 \mu_f. \tag{6.78}$$

These equations, however, are a function of the fast variable μ_f. The equilibrium expressions of the fast system, Equations 6.70, 6.71 and 6.72 can now be used to obtain an expression for μ_f as a function of the slow variables μ_s and σ_s^2.

Equation 6.70 gives the following quadratic equation for μ_f after some re-arrangement:

$$0 = \mu_f^2 - (y_{e_o} + \mu_s + K)\mu_f + \mu_s y_{e_o} - \sigma_{sf} + \sigma_f^2. \tag{6.79}$$

Equations 6.71 in turn gives

$$\begin{aligned} 0 &= k_1(y_{e_o} - \mu_f)\sigma_s^2 + [k_1(2\mu_f - \mu_s - y_{e_o}) - k_2], \text{ so that} \\ \sigma_{sf} &= \frac{y_{e_o} - \mu_f}{\mu_s + y_{e_o} + K - 2\mu_f}\sigma_s^2. \end{aligned} \tag{6.80}$$

Similarly, Equation 6.72 becomes

$$\begin{aligned} 0 &= 2k_1(y_{e_o} - \mu_f)\sigma_{sf} + 2[k_1(2\mu_f - \mu_s - y_{e_o}) - k_2)\sigma_f^2 \\ &\quad + \frac{1}{\Omega}\left[k_1[(\mu_s - \mu_f)(y_{e_o} - \mu_f) - \sigma_{sf} + \sigma_f^2] + k_2\mu_f\right], \end{aligned}$$

which may be re-arranged to

$$\sigma_f^2 = \frac{[2(y_{e_o} - \mu_f) - \frac{1}{\Omega}]\sigma_{sf} + \frac{1}{\Omega}[(\mu_s - \mu_f)(y_{e_o} - \mu_f) + K\mu_f]}{2(\mu_s + y_{e_o} + K - 2\mu_f - \frac{1}{2\Omega})}. \tag{6.81}$$

Equations 6.80 and 6.81 can then be combined to find that

$$\begin{aligned} \sigma_f^2 - \sigma_{sf} &= -\frac{(\mu_s + K - \mu_f)(y_{e_o} - \mu_f)\sigma_s^2}{(\mu_s + y_{e_o} + K - \mu_f)(\mu_s + y_{e_o} + K - \mu_f - \frac{1}{2\Omega})} \\ &\quad + \frac{1}{2\Omega}\left[\frac{(\mu_s - \mu_f)(y_{e_o} - \mu_f) + K\mu_f}{\mu_s + y_{e_o} + K - 2\mu_f - \frac{1}{2\Omega}}\right]. \end{aligned} \tag{6.82}$$

The previous expression can already be substituted into Equation 6.79 to yield a quartic equation of μ_f as a function of μ_s and σ_s^2.

Instead of pursuing this approach, we instead make several approximations to find μ_f to draw a parallel to the classical case, and to the result obtained via the distribution approach. We start by pretending that $\sigma_f^2 - \sigma_{sf}$ is not a function of μ_f and solve for μ_f in Equation 6.79, resulting in

$$\mu_f = \frac{\mu_s + y_{e_o} + K}{2}\left[1 \pm \sqrt{1 - 4\frac{\mu_s y_{e_o} + \sigma_f^2 - \sigma_{sf}}{(\mu_s + y_{e_o} + K)^2}}\right]. \tag{6.83}$$

We then let $r = 4\frac{\mu_s y_{e_o} + \sigma_f^2 - \sigma_{sf}}{(\mu_s + y_{e_o} + K)^2}$, and assume $r \ll 1$, so that

$$\mu_f \approx \frac{\mu_s y_{e_o}}{\mu_s + y_{e_o} + K} + \frac{\sigma_f^2 - \sigma_{sf}}{\mu_s + y_{e_o} + K}. \tag{6.84}$$

Note that the first term in the right side of the previous equation is identical to the tQSSA approximation for the complex obtained in the classical case; we consider the second term to be a stochastic correction and approximate it as small. We then want to substitute Equation 6.82 into 6.84, but because Equation 6.84 is a function of μ_f, we first evaluate μ_f in Equation 6.82 as the first term on the right of Equation 6.84 before actually making the substitution. That is, we use $\mu_f = \frac{\mu_s y_{e_o}}{\mu_s + y_{e_o} + K}$ in the expression for $\sigma_f^2 - \sigma_{sf}$ before substituting it into Equation 6.84. Proceeding so results in our final expression for μ_f, which, letting $a = \frac{\mu_s y_{e_o}}{\mu_s + y_{e_o} + K}$ for simplicity of notation, becomes

$$\begin{aligned} \mu_f \approx\ & \frac{\mu_s y_{e_o}}{\mu_s + y_{e_o} + K} - \frac{(\mu_s + K - a)(y_{e_o} - a)}{(\mu_s + y_{e_o} + K)(\mu_s + y_{e_o} + K - 2a)(\mu_s + y_{e_o} + K - 2a - \frac{1}{2\Omega})}\sigma_s^2 \\ & + \frac{1}{2\Omega}\left[\frac{(\mu_s - a)(y_{e_o} - a) + Ka}{(\mu_s + y_{e_o} + K)(\mu_s + y_{e_o} + K - 2a - \frac{1}{2\Omega})}\right]. \end{aligned} \tag{6.85}$$

Substituting this expression into Equations 6.77 and 6.78 finally gives the slow MFK equations for the enzymatic reaction. The resulting equations are similar to the MFK equations obtained previously (in Equation 6.52), although with K in place of K_m because the MFK-based approach ignores the slow reactions in the fast system. For the same reason, the approximation of Equation 6.52 should be the more accurate.

6.8 Concluding Remarks

We have described here a systematic method to identify slow and fast variables in a system of slow and fast chemical reactions, and have developed reduced CMEs for these variables, starting from a CME for the full system. Our approach is distinct in its retention of the effects of slow reactions on the fast subsystem. The resulting model reduction method should be particularly useful in systems where the slow reactions

are independent of the slow variables, as occurs in many systems of interest. We have also highlighted the value of MFK models, with their tracking of means, variances and covariances, as adjuncts to this model reduction process.

Numerical experiments with our reduced CME model for the slow variables in a canonical example — the irreversible enzymatic reaction — have demonstrated consistently and predictably better results than those obtained with other models from the literature. Furthermore, our reduced MFK model for the slow system in this example specializes in the large-volume (low-fluctuation) limit to familiar deterministic total quasi-steady-state models that are more accurate and more broadly applicable than traditional Michaelis-Menten models.

It is worth noting that Goutsias ([64]) and Haseltine and Rawlings([68, 69]) work with a different coordinate system than that used here, specifying the state to be the degree of advancement of the reactions (i.e., the number of reactions of each kind that have occurred since the start of the experiment). This alternate state is particularly useful when reactions are not instantaneous, but rather change the molecule numbers in the system only after a specified amount of time since the beginning of the reaction has passed; see Goutsias ([64]), for instance, for a detailed explanation. However, the expressions here can be adapted quite straightforwardly for their model.

We end with the observation that considerable work remains to be done to bridge the more informally justified time-scale reduction methods developed here and in related literature, with more rigorous treatments such as those in Ball *et al.* ([6]), for models of interest in stochastic chemical kinetics.

6.9 Connection of Slow and Fast CME Models to Singular Perturbation Theory

Singular perturbation (SP) theory provides a very useful framework for separation of time scales in systems of ordinary differential equations (ODEs). Previous studies have successfully applied SP in the context of stochastic chemical kinetics ([113, 99]).

Furthermore, the deterministic approximations for the irreversible enzymatic reaction are understood to be first-order approximation obtained through SP analysis ([17, 71]). The standard SP model starts with the evolution equations of the fast and slow variables of a system, respectively denoted by $\mathbf{x}_f$ and $\mathbf{x}_s$; it is given by

$$\begin{aligned} \frac{d}{dt}\mathbf{x}_s &= \epsilon f(t, \mathbf{x}_s, \mathbf{x}_f, \epsilon) \\ \frac{d}{dt}\mathbf{x}_f &= g(t, \mathbf{x}_s, \mathbf{x}_f, \epsilon), \end{aligned} \tag{6.86}$$

where t denotes time, $\epsilon \ll 1$ is a small parameter, and $f(;)$ and $g(;)$ are functions of the indicated arguments, returning vectors of the appropriate dimensions.

The standard result is that, under appropriate conditions, it can be shown that approximating the slow variables by

$$\frac{d}{dt}\mathbf{x}_s = \epsilon f(t, \mathbf{x}_s, h(t, \mathbf{x}_s), \epsilon), \tag{6.87}$$

where $h(t, \mathbf{x}_s)$ is a solution of

$$g(t, \mathbf{x}_s, h(t, \mathbf{x}_s), 0) = 0, \tag{6.88}$$

results in an approximation with an error of order ϵ (see [81], for example). This approximation, which we refer to as the first-order approximation, requires finding the value of the fast variables as a function of the slow variables, $h(t, \mathbf{x}_s)$. This is achieved while assuming that the left side of their evolution equation is zero, which is equivalent to their being at steady state, and importantly, by setting ϵ to zero on the right side of the equation, resulting in the algebraic expression of Equation 6.88.

A simple re-scaling of time of Equations 6.22 and 6.27 shows high correspondence with the standard singular perturbation result. It also serves to highlight the difference between our approach and the approaches proposed by Cao *et al.* ([23]), Goutsias ([64]) and Haseltine and Rawlings ([68]). Suppose that τ_s and τ_f respectively denote the characteristic time scales of the slow and fast CMEs in Equations 6.22 and 6.27. These time scales could be thought of, for example, as the time con-

stant of the convergence of each CME to its respective stationary form (assuming ergodicity). The ratio of the slow and fast time scales, which we here call $\epsilon = \tau_f/\tau_s$, is expected to be small. We now look at the equations in the fast time scale by re-scaling time using $\tau = t/\tau_s$, the propensities of the slow reactions using $a'_\ell = a_\ell \tau_s$, and those of the fast reactions using $a'_\ell = a_\ell \tau_f$; τ and a'_ℓ denote the normalized time and propensities. These re-scalings, when substituted into Equations 6.22 and 6.27, result in the following CMEs:

$$\begin{aligned}
\frac{dP(\mathbf{x}_s)}{d\tau} &= \epsilon \sum_{\ell=1}^{L_s} P(\mathbf{x}_s - \mathbf{s}_{\ell s})\overline{a}'_\ell(\mathbf{x}_s - \mathbf{s}_{\ell s}) - P(\mathbf{x}_s)\overline{a}'_\ell(\mathbf{x}_s), \\
\frac{dP(\mathbf{x}_f|\mathbf{x}_s)}{d\tau} &= \epsilon \Bigg[\sum_{\ell=1}^{L_s} \Bigg(P(\mathbf{x}_f - \mathbf{s}_{\ell f}|\mathbf{x}_s) a'_\ell(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f}) \\
&\quad - P(\mathbf{x}_f|\mathbf{x}_s) a'_\ell(\mathbf{x}_s, \mathbf{x}_f) \Bigg) \Bigg] + \\
&\quad \sum_{\ell=L_s+1}^{L} \Bigg(P(\mathbf{x}_f - \mathbf{s}_{\ell f}|\mathbf{x}_s) a'_\ell(\mathbf{x}_s, \mathbf{x}_f - \mathbf{s}_{\ell f}) \\
&\quad - P(\mathbf{x}_f|\mathbf{x}_s) a'_\ell(\mathbf{x}_s, \mathbf{x}_f) \Bigg).
\end{aligned} \tag{6.89}$$

The normalized CMEs above are in the same form as the standard singular perturbation model in Equation 6.86. Furthermore, the method we suggest here is exactly the standard singular perturbation result cited above, except that we do not set ϵ to zero in the right side of the evolution equation for the fast system. That is, we solve

$$g(t, \mathbf{x}_s, h(t, \mathbf{x}_s), \epsilon) = 0 \tag{6.90}$$

rather than Equation 6.88, to obtain the steady-state values of the fast variables as a function of the slow ones.

As inspection of Equation 6.89 reveals, setting ϵ to zero in the equation for the fast variables of our CMEs directly results in ignoring the slow reactions in the fast system, this is exactly what Cao *et al.* ([23]), Goutsias ([64]) and Haseltine and Rawlings ([68]) do. Unfortunately, the methods of separation of time scales for stochastic systems are not equivalent to the standard singular perturbation problem. The full

description of the system is the original CME for all species, a system of equations (one for each possible state) where both fast and slow variables potentially appear in every equation. Putting it into the standard singular perturbation model already requires the series of assumptions leading to Equation 6.27. That said, analyzing a simple linear deterministic system using singular perturbation ideas does help clarify why keeping slow reactions in the fast system improves upon previous approximation schemes. This example is discussed next. Furthermore, a cleaner way to connect separation of time scales to singular perturbation is achieved by working not with the full CME, but with its moments, since each of these is already described by an ODE. We describe such an approach in Appendix 6.7, where we derive the first-order singular perturbation approximation to the means, variances and covariances of the slow variables' concentrations.

Example

We have shown that the method proposed here to obtain an approximate CME for the slow variables can result in significantly more accurate approximations for some systems. In fact, the example discussed next suggests our method is expected to do better for systems where the propensities of the slow reactions that modify any slow variables are independent of the slow variables themselves, while for other systems all methods are *a priori* similarly accurate.

Consider a linear time-invariant system with dynamics described by $\frac{d}{dt}\mathbf{x} = \mathbf{A}\mathbf{x}$, where $\mathbf{x}$ the (column) state vector and $\mathbf{A}$ a matrix of appropriate dimensions, with entries that are functions of a small parameter $\epsilon \ll 1$. Further assume that $\mathbf{A}$ is such that the first entries in $\mathbf{x}$— the slow variables, denoted by $\mathbf{x}_s$ — have rates of change that are $O(\epsilon)$, while the rest of the state variables— the fast ones, denotes by $\mathbf{x}_f$ — all have rates of change that are $O(1)$ in ϵ. More concretely, assume that

$$\frac{d}{dt}\left[\begin{array}{c}\mathbf{x}_s\\\hline\mathbf{x}_f\end{array}\right] = \left[\begin{array}{c|c}\mathbf{A}_1 & \mathbf{A}_2\\\hline\mathbf{A}_3 & \mathbf{A}_4\end{array}\right]\left[\begin{array}{c}\mathbf{x}_s\\\hline\mathbf{x}_f\end{array}\right] = \left[\begin{array}{c}\mathbf{A}_1\mathbf{x}_s + \mathbf{A}_2\mathbf{x}_f\\\hline\mathbf{A}_3\mathbf{x}_s + \mathbf{A}_4\mathbf{x}_f\end{array}\right], \tag{6.91}$$

where the matrices above have appropriate dimensions and are given by

$$\begin{aligned} \mathbf{A}_1 &= \epsilon \mathbf{A}_{11}, & \mathbf{A}_2 &= \epsilon \mathbf{A}_{21}, \\ \mathbf{A}_3 &= \mathbf{A}_{30} + \epsilon \mathbf{A}_{31}, \quad \text{and} & \mathbf{A}_4 &= \mathbf{A}_{40} + \epsilon \mathbf{A}_{41}. \end{aligned} \tag{6.92}$$

We would like to draw an analogy at this point to the stochastic chemical kinetics systems discussed above. The time-derivative of $\mathbf{x}_s$ is a sum of small terms, or reactions, that are all slow because they are proportional to ϵ; while that of $\mathbf{x}_f$ has terms independent of ϵ, due to the fast reactions, but also others proportional to ϵ from slow reactions.

Now suppose we wanted to obtain an approximate description of the slow variables in this system. The analogy to what we propose above for stochastic systems is to assume that the fast variables always equilibrate fast, so that $\frac{d}{dt}\mathbf{x}_s \approx 0$. This results in

$$\mathbf{x}_f \approx -\mathbf{A}_4^{-1}\mathbf{A}_3\mathbf{x}_s, \tag{6.93}$$

where we have assumed that $\mathbf{A}_4$ is invertible (and stable!). Substituting this expression into the rate equation for $\mathbf{x}_s$ then results in our slow approximation of the system, namely

$$\frac{d}{dt}\mathbf{x}_s \approx \big(\mathbf{A}_1 - \mathbf{A}_2\mathbf{A}_4^{-1}\mathbf{A}_3\big)\mathbf{x}_s. \tag{6.94}$$

A second possibility to obtain an approximation for the slow variables in the system results from applying the standard theorems of singular perturbation theory (cite Khalil). There, one Taylor-expands the exact solution for $\mathbf{x}$ as a function of ϵ and can put together approximate solutions of different accuracies in ϵ. The first approximation, known as the zero-order approximation, differs from the approach above in that it sets ϵ to zero in the equation for the rate of change in the fast variables. This results in

$$\mathbf{x}_f \approx -\mathbf{A}_{40}^{-1}\mathbf{A}_{30}\mathbf{x}_s \tag{6.95}$$

instead of Equation 6.93, and in

$$\frac{d}{dt}\mathbf{x}_s \approx \left(\mathbf{A}_1 - \mathbf{A}_2\mathbf{A}_{40}^{-1}\mathbf{A}_{30}\right)\mathbf{x}_s \tag{6.96}$$

in place of Equation 6.94. Note that this second approach is exactly analogous to what Goutsias and Cao propose in the stochastic case, namely to set the rate of slow reactions in the fast system to zero before equilibrating the fast system. But which approach is more accurate?

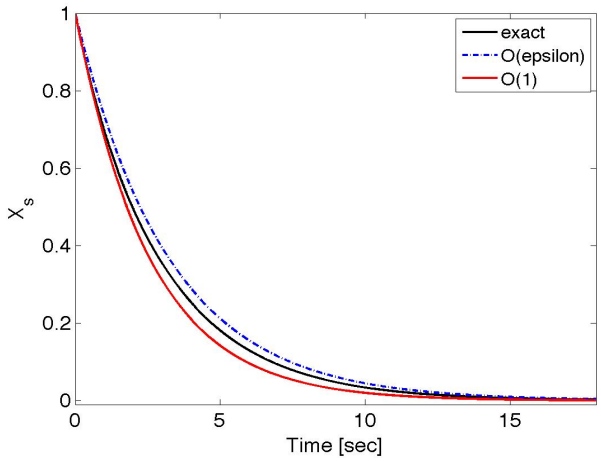

Figure 6-9: Exact solution to Equation 6.91 (black line), along with the two approximations (blue and red lines), for $a_{11} = -1.9$ and $\epsilon = .1$. The initial condition was taken to be $\mathbf{x}_0 = [1\ 1]'$.

It turns out, as demonstrated below first numerically and then analytically, that it is not *a priori* obvious which approach will be a better approximation for general systems of the form presented here. In fact, for very small ϵ both approximations do quite well. But when ϵ is larger, our approach does turn out to be the more accurate of the two for systems where the rate of change of the slow variables is independent of the slow variables themselves (i.e., when $\mathbf{A}_1 = \mathbf{0}$). Such systems turn out to be common in Biology (e.g., see the next example) and, in fact, the increase in accuracy that our method provides can be quite significant for systems of this class.

To simplify the argument we focus on the two-variable case, so all the matrices above, except for $\mathbf{A}$ itself, are now numbers and are denoted by a small a symbols

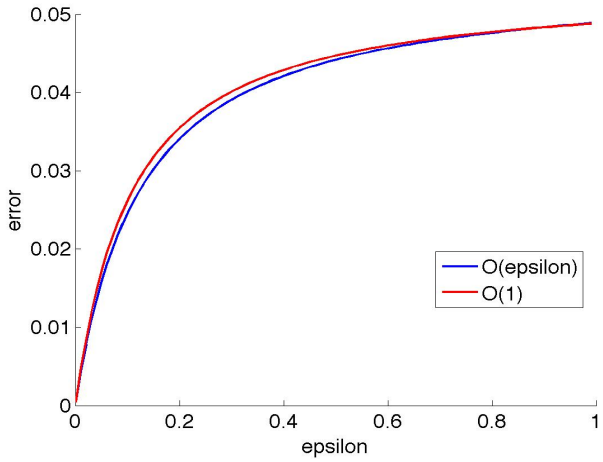

Figure 6-10: Error of the two approximations for the system in Figure 6-9 as a function of ϵ. We consider a range of values for ϵ, and for each we numerically obtain evenly-spaced samples of the solutions to Equations 6.91, 6.94 and 6.96, corresponding to a time-period equal to five times the slow time scale of the system. We then compute the error of each approximation according to Equation 6.53. For small ϵ both approximations have a small error, but as ϵ grows the error of both approximations increases. The initial condition used for every value of ϵ was $\mathbf{x}_0 = [1\ 1]'$.

(e.g., $\mathbf{A}_4 = \mathbf{A}_{40} + \epsilon\mathbf{A}_{41}$ is now $a_4 = a_{40} + \epsilon a_{41}$). Consider a system with $\mathbf{A}$ matrix given by

$$\begin{bmatrix} \epsilon a_{11} & -2\epsilon \\ 1+2\epsilon & -(1+10\epsilon) \end{bmatrix}. \tag{6.97}$$

When $a_{11} \neq 0$ either approximation can be better than the other one (depending on the actual value of a_{11}); keeping $O(\epsilon)$ terms in the equilibrium equation for the fast variables need not result in a better approximation (Figure 6-9). Figure 6-10 shows the error of both approximations for the example in Figure 6-9; it is small for small ϵ but grows as ϵ becomes larger. The accuracy of both approximations in this example (i.e., for the value of a_{11} chosen here) is comparable for a wide range of values of ϵ. Other values of a_{11} can result in either of the two approximations being significantly better than the other.

Now consider a system with the same parameter values above, but where a_{11} is now set to zero. Our approximation, Equation 6.94, is now clearly the better one; see Figure 6-11. Furthermore, the error of our approximation is now significantly smaller

than that of the other approximation for a wide range of ϵ, as shown in Figure 6-12. In fact, this figure shows that our approximation can remain a good approximation even when ϵ is not small. These results, taken together, argue for the use of our approximations when $a_{11} = 0$, which would correspond to slow reaction rates that are independent of slow variables.

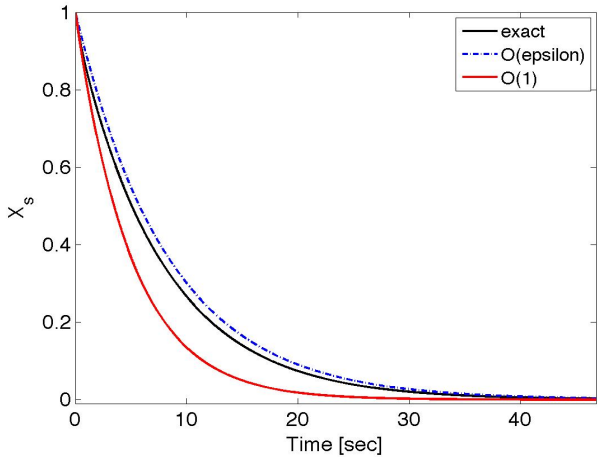

Figure 6-11: Exact solution to Equation 6.91 (black line), along with the two approximations (blue and red lines), for $a_{11} = 0$ and $\epsilon = .1$. The initial condition was taken to be $\mathbf{x}_0 = [1\ 1]'$.

We now compare the two approximations analytically. For this purpose, suppose that the goal of the approximation is to capture the time scale of the slow system accurately. The time scales of a linear system are simply the inverse of its eigenvalues, which may be found exactly for the two-variable case. They are given by

$$\begin{aligned} \lambda &= \left(\frac{a_1 + a_4}{2}\right)\left[1 \pm \sqrt{1 - 4\frac{a_1 a_4 - a_2 a_3}{(a_1 + a_4)^2}}\right] \\ &\approx \left(\frac{a_1 + a_4}{2}\right)\left[1 \pm \left(1 - 2\frac{a_1 a_4 - a_2 a_3}{(a_1 + a_4)^2}\right)\right], \end{aligned} \tag{6.98}$$

where we use the fact that $|a_4| \gg |a_1|, |a_2|$ to obtain the second equation, which is exact up to second order in ϵ. The eigenvalue corresponding to the slow time scale of

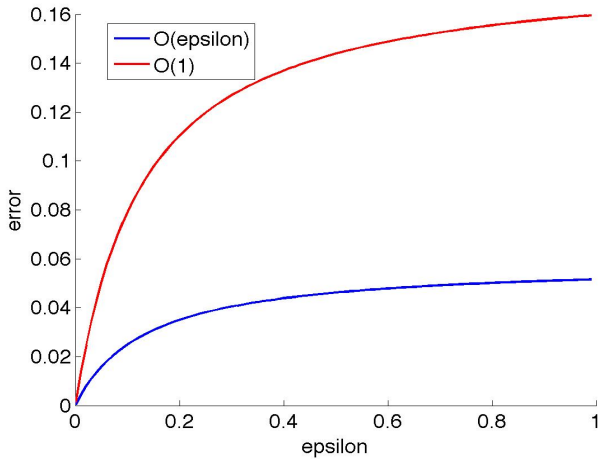

Figure 6-12: Error of the two approximations for the system in Figure 6-11 as a function of ϵ. For small ϵ both approximations have a small error, but as ϵ grows the error of both approximations increases. The initial condition used for every value of ϵ was $\mathbf{x}_0 = [1\ 1]'$.

the system is then

$$\begin{aligned}\lambda_s &= \frac{a_1a_4 - a_2a_3}{a_1 + a_4} = \epsilon \overbrace{\left[\frac{a_{11}a_{40} - a_{21}a_{30}}{a_{40}}\right]}^{\lambda_0} \\ &\quad +\epsilon^2\Bigg[\underbrace{\left(\frac{a_{11}a_{21}a_{30}}{a_{40}^2} + \frac{a_{11}^2}{a_{40}}\right)}_{\lambda_{10}} \underbrace{-\frac{a_{21}a_{30}}{a_{40}}\left(\frac{a_{31}}{a_{30}} - \frac{a_{41}}{a_{40}}\right)}_{\lambda_{11}}\Bigg] \\ &= \epsilon\lambda_0 + \epsilon^2(\lambda_{10} + \lambda_{11}),\end{aligned} \tag{6.99}$$

where we have ignored $O(\epsilon^3)$ terms.

The eigenvalue of the approximation resulting from our method may be read from Equation 6.94 directly; it is given by

$$\lambda_1 = \frac{a_1a_4 - a_2a_3}{a_4} = \epsilon\lambda_0 + \epsilon^2\lambda_{11} + O(\epsilon^3). \tag{6.100}$$

Similarly, the eigenvalue of the second approximation is simply

$$\lambda_2 = \frac{a_1a_{40} - a_2a_{30}}{a_{40}} = \epsilon\lambda_0. \tag{6.101}$$

The above results show that, in general, the zero-order singular perturbation approximation only recovers the $O(\epsilon)$ term of the slow system's eigenvalue. Our approximation, on the other hand, recovers the $O(\epsilon)$ term and (only) part of the $O(\epsilon^2)$ term. So for general systems, it is not *a priori* obvious which of the two approximations is better. It depends on the actual values of the system parameters whether Equation 6.100 is closer to Equation 6.99 than Equation 6.101.

However, our approximation does recover the full $O(\epsilon^2)$ of the eigenvalue for systems where a_{11} is zero (i.e., systems where a_1 is $O(\epsilon^2)$) since for these systems $\lambda_{10} = 0$. Note that this condition would be analogous to having the rates of all slow reactions be independent of the slow variables. For such systems, our approximation is always the more accurate, and the difference in accuracies is expected to grow with ϵ. These results are consistent with the numerical results described above.

Chapter 7

The Spatial Mass Fluctuation Kinetics Equations

7.1 Background

We consider systems of chemical reactions consisting of N chemical species $\{X_1, ..., X_N\}$ interacting in a volume V through L reactions labeled $R_1, ..., R_L$. The starting point is the reaction-diffusion master equation (RDME) proposed, for example, in [88, 50, 105, 7] since the early 1970s. In this context, the space is divided into n subvolumes of volume v, such that $V = nv$, and where the size v of each subvolume is carefully chosen (e.g., see [87]). The RDME is the forward Kolmogorov equation of a discrete state continuous time Markov chain model of the system, and describes the time evolution of the joint probability distribution for the number of molecules of each species within each subvolume (e.g., see [87] for a good introduction to the RDME).

We arbitrarily number the subvolumes and let $x_i(c)$ and $y_i(c)$ be the number of molecules and the concentration, respectively, of species X_i in subvolume c. Molecule numbers are normalized by $\Omega = Av$, where A is Avogadro's number, to yield concentrations in moles per unit volume so that $y_i = x_i/\Omega$. We also let $\mathbf{x}_c$ be the column vector of molecule numbers in subvolume c (with i-th component $\{\mathbf{x}_c\}_i = x_i(c)$). We further let $\mathbf{x} = [\mathbf{x}_1' \dots \mathbf{x}_n']'$ be the state vector of the system and x_i its i-th entry. Similarly, let $\mathbf{y}_c = \mathbf{x}_c/\Omega$ be the vector of concentrations in subvolume c, and $\mathbf{y} = [\mathbf{y}_1' \dots \mathbf{y}_n']'$

be the vector of concentrations in the whole system, with mean $\mu = E[\mathbf{y}] = [\mu'_1 \dots \mu'_n]'$ and covariance matrix $V = E[(\mathbf{y} - \mu)(\mathbf{y} - \mu)']$. A firing of reaction R_ℓ immediately changes $x_i(c)$ to $x_i(c) + s_{\ell i}(c)$, or in vector form, $\mathbf{x}$ changes to $\mathbf{x} + \mathbf{s}_\ell$, where $\mathbf{s}_\ell$ is the *stoichiometry vector* of R_ℓ and $s_{\ell i}(c)$ the *stoichiometry coefficient* of $x_i(c)$ in R_ℓ.

The time evolution of $P(\mathbf{x})$, the probability distribution of the state $\mathbf{x}$ conditioned on an initial state $\mathbf{x}(t_o) = \mathbf{x_o}$, satisfies the RDME, namely

$$\frac{dP(\mathbf{x})}{dt} = \sum_{\ell=1}^{L} P(\mathbf{x} - \mathbf{s}_\ell) a_\ell(\mathbf{x} - \mathbf{s}_\ell) - P(\mathbf{x}) a_\ell(\mathbf{x}). \tag{7.1}$$

The probability that reaction R_ℓ occurs in a small time interval with duration dt is approximately equal to $a_\ell(\mathbf{x})dt$; the quantity $a_\ell(\mathbf{x})$ is the *propensity* of reaction R_ℓ. Note that this equation has the same form as the CME of stochastic chemical kinetics, Equation (4.6). This is because we again have a discrete-state continuous-time Markov chain model of the system, and because we defined the system state as the single vector $\mathbf{x}$.

Two kinds of reactions are considered: those that only change the state of one subvolume (e.g., $\mathbf{x}_c$ for subvolume c), which we refer to as *local*; and those that change the state of more than one subvolume, which we refer to as *diffusive*. The propensities of local reactions are functions of the local state only (e.g., of $\mathbf{x}_c$ for a local reaction occurring in subvolume c). These propensities have the same functional form as their analogues in stochastic chemical kinetics. For example, the propensity of a local reaction that takes one molecule of X_i and one molecule of X_j in subvolume c as reactants is given by $kx_i(c)x_j(c)$, where k is the reaction rate constant. For a diffusive reaction, on the other hand, each firing removes one molecule from one subvolume and puts it in another subvolume. The propensity of a diffusive reaction that results in one less molecule of X_i in subvolume c is assumed to be given by $d_i x_i(c)$, where the rate constant d_i is a function of the diffusion constant D_i of X_i to be specified later.

Typically researchers resort to stochastic simulation as a means for getting to the system dynamics. The simulation algorithms available are analogous to those of (spa-

tially homogeneous) stochastic chemical kinetics. For example, a particularly efficient algorithm, the next-subvolume method (e.g., see [37]), is analogous to that proposed by Gibson and Bruck in [54] for stochastic chemical kinetics. However, as the number of subvolumes and the number of molecules in each subvolume increase, stochastic simulation quickly becomes computationally challenging. Additionally, the connection between the RDME and other modeling frameworks for systems of chemical reactions is not that clear, as discussed in some detail in Section 7.2.1.

To help address both of these issues, our interest in this chapter is to develop approximate equations for the (spatial) concentrations and their variances and covariances. The next two sections of this chapter can be read in any order. The next sections develops results for a simple cylindrical system consisting of cubic subvolumes stacked next to one another, to build intuition. We use this setup to discuss the convergence of this modeling framework to simpler ones in Section 7.2.1. We then develop in Section 7.3 the general case where subvolumes are spatially connected to one another in an arbitrary fashion.

7.2 Systems with one relevant spatial dimension

Consider a cylinder starting at position in space 0, square cross-sectional area l^2 and divide it into N_v subvolumes of length l (and with volume $v = l^3$). Number the subvolumes from left to right, so that subvolume number c is connected to the immediate neighboring subvolumes $c-1$ and $c+1$. Also let $\Omega = vA$ be the size of each subvolume, $y_i(c) = \frac{x_i(c)}{\Omega}$ the concentration of species X_i in subvolume c, $\mu_i(c)$ its expected concentration, and $\sigma_{ij}(c_1, c_2)$ be the covariance between $y_i(c_1)$ and $y_j(c_2)$, which potentially corresponds to species in different subvolumes (when $c_1 \neq c_2$). All expectations in this chapter are taken with respect to the distribution $P(\mathbf{x})$.

Suppose we are interested in the time-evolution of the expected value $E[f(\mathbf{x})]$ of a smooth but possibly non-linear function of the state $f(\mathbf{x})$. Multiplying (7.1) by $f(\mathbf{x})$

and then summing over all values of $\mathbf{x}$ results, after some algebra, in

$$\frac{d}{dt}E[f(\mathbf{x})] = \sum_{\ell=1}^{L}\Big(E[a_\ell(\mathbf{x})f(\mathbf{x}+\mathbf{s}_\ell)] - E[a_\ell(\mathbf{x})f(\mathbf{x})]\Big). \tag{7.2}$$

This equation, first derived in the context of the CME by Gillespie (cite 1981 paper), can be used to find the rate of change of the mean concentrations, and their variances and covariances.

We first let $f(\mathbf{x}) = \frac{x_i(c)}{\Omega}$, so that $E[f(\mathbf{x})] = \mu_i(c)$. Substituting this definition of $f(\mathbf{x})$ into (7.2) gives

$$\begin{aligned}\frac{d\mu_i(c)}{dt} &= \sum_{\ell=1}^{L}\Big(E\big[a_\ell(\mathbf{x})\frac{x_i(c)+s_{\ell i}(c)}{\Omega}\big] - E\big[a_\ell(\mathbf{x})\frac{x_i(c)}{\Omega}\big]\Big) \\ &= \sum_{\ell=1}^{L} s_{\ell i}(c)E[\frac{a_\ell(\mathbf{x})}{\Omega}] \;=\; \sum_{\ell=1}^{L} s_{\ell i}(c)E[r_\ell(\mathbf{y})]\;,\end{aligned} \tag{7.3}$$

where

$$r_\ell(\mathbf{y}) = \frac{a_\ell(\mathbf{x})}{\Omega}\Big|_{\mathbf{x}=\mathbf{y}\Omega} \tag{7.4}$$

is the *rate* of reaction R_ℓ, which has units of concentration per unit time. Note that for local reactions r_ℓ is a function of only the local concentrations, and not of the full concentration vector $\mathbf{y}$.

Although the summation in (7.4) is over all reactions in the system, only those with non-zero $s_{\ell i}(c)$ actually contribute to the evolution of $\mu_i(c)$. Such reactions are either local reactions in subvolume c, with reaction indices denoted by $\ell(c)$, or one of the four diffusive reactions that modify $x_i(c)$ (diffusion of X_i to and from subvolumes $c-1$ and $c+1$). The rate of a diffusive reaction involving species X_i originating in subvolume c is $d_i y_i(c)$, where d_i is related to the diffusion constant of X_i through (e.g., see [87])

$$d_i = \frac{D_i}{l^2}\;. \tag{7.5}$$

Then we can re-write (7.4) as

$$\begin{aligned}\frac{d\mu_i(c)}{dt} &= \overbrace{\sum_{\ell\in\ell(c)}\left(s_{\ell i}(c)E[r_\ell(\mathbf{y}_c)]\right)}^{\text{local reactions}} + \overbrace{d_i\left[(\mu_i(c+1)-\mu_i(c))-(\mu_i(c)-\mu_i(c-1))\right]}^{\text{diffusive reactions}} \\ &= \sum_{\ell\in\ell(c)}\left(s_{\ell i}(c)E[r_\ell(\mathbf{y}_c)]\right) + \frac{D_i}{l}\left[\frac{\mu_i(c+1)-\mu_i(c)}{l} - \frac{\mu_i(c)-\mu_i(c-1)}{l}\right] \quad (7.6)\end{aligned}$$

The first part on the right of (7.6) depends only on local reactions of subvolume c, which have rates that depend only on the local state $\mathbf{y}(c)$. Because these rates are potentially non-linear functions of $\mathbf{y}(c)$ (7.6) generally depends on higher moments (e.g., variances and covariances) of the local concentrations. The dynamics of $\mu_i(c)$ is coupled to that of other subvolumes through the diffusive reactions on the right of (7.6). The diffusive reactions in (7.6) are effectively computing the discrete version of the second spatial derivative of $y_i(c)$. It is tempting to take the limit of l to zero at this point to obtain the Laplacian operator from the diffusive reactions in (7.6). Unfortunately, it turns out such a limit results in exploding higher-moments. Alternate strategies to avoid this problem are discussed in Section 7.2.1.

To obtain the rates of change of the variances and covariances $\sigma_{ij}(c1, c2)$, we now set $f(\mathbf{x}) = (\frac{x_i(c_1)}{\Omega} - \mu_i(c_1))(\frac{x_j(c_2)}{\Omega} - \mu_j(c_2))$, so that $E[f(\mathbf{x})] = \sigma_{ij}(c1, c2)$. Substituting this definition of $f(\mathbf{x})$ into $f(\mathbf{x})$ into (7.2), after some algebra, now results in

$$\begin{aligned}\frac{d\sigma_{ij}(c_1,c_2)}{dt} &= \sum_{\ell=1}^{L}\Bigg(\frac{1}{\Omega}s_{\ell i}(c_1)s_{\ell j}(c_2)E[r_\ell(\mathbf{y})] \quad (7.7)\\ &\quad + s_{\ell i}(c_1)E[r_\ell(\mathbf{y})\{y_j(c_2)-\mu_j(c_2)\}] + s_{\ell j}(c_2)E[r_\ell(\mathbf{y})\{y_i(c_1)-\mu_i(c_1)\}]\Bigg).\end{aligned}$$

Each reaction in the system can then contribute the three terms on the right of (7.8) to the rate of change of $\sigma_{ij}(c_1, c_2)$. The first term will only be non-zero when both $s_{\ell i}(c_1)$ and $s_{\ell j}(c_2)$ are non-zero, which for $i \neq j$ will only be the case when $c_1 = c_2$ for some local reactions (and not for any diffusive reaction), and for $i = j$ only for some diffusive reactions if $|c_1 - c_2| < 2$. The second and third term on the right of (7.8) will be non-zero for reactions that change $x_i(c_1)$ and $x_j(c_2)$, respectively. The

expectations (between square brackets) in the second and third terms in (7.8) are actually the covariances between the reaction rate $r_\ell(\mathbf{y})$ and the concentration of $X_j(c_2)$ and $X_i(c_1)$, respectively.

For nonlinear rate functions $r_\ell(\mathbf{y})$ these covariances involve central moments of $P(\mathbf{x})$ greater than two. The same is true of the expected rates $E[r_\ell(\mathbf{y})]$ that appear in (7.8) and in (7.6) when the rate functions are non-linear non-quadratic functions of the concentrations. Thus, the rate of change of the means, variances and covariances of the concentrations depends on higher-moments. However, in the spirit of the mass fluctuation equations in Chapter 5 we obtain approximate the rate of change of these statistics by Taylor-expansion of each of the functions whose expectations are needed about the average concentrations μ, retaining only terms up to second order, and then taking the expectation.

As shown in Chapter 5, the expected rate of a reactions is then

$$E[r_\ell(\mathbf{y})] \approx r_\ell(\mu) + \frac{1}{2}\sum_{p,q=1}^{n}\left(\frac{\partial^2 r_\ell(\mathbf{y})}{\partial y_p \partial y_q}\Big|_{\mathbf{y}=\mu}\right)\sigma_{pq} \tag{7.8}$$

$$= r_\ell(\mu) + \xi_\ell(\mu, \mathbf{V}), \tag{7.9}$$

where σ_{pq} corresponds to the entry in row p and column q of the covariance matrix $\mathbf{V}$. The stochastic rate ξ_ℓ is generally non-zero for reactions with non-linear rate functions (e.g., bimolecular reactions). Note that for a local reaction in subvolume c, the only non-zero terms in the summation in (7.9) will correspond to concentrations in that same subvolume. Similarly, because diffusive reactions are assumed to have rate functions that are linear in the concentrations, their stochastic rates are actually zero.

The covariances between the reaction rates $r_\ell(\mathbf{y})$ and the concentrations (e.g., $E[r_\ell(\mathbf{y})(y_j - \mu_j)]$) may similarly be approximated as a function of μ and $\mathbf{V}$ by first Taylor-expansion of terms of the form $r_\ell(\mathbf{y})y_j$ up to second order about μ, and then

taking the expectation. Combining the resulting expression with (7.9) yields

$$E[r_\ell(\mathbf{y})(y_j - \mu_j)] \approx \sum_{p=1}^{n} \left(\frac{\partial r_\ell(\mathbf{y})}{\partial y_p}\Big|_{\mathbf{y}=\mu} \right) \sigma_{pj}. \tag{7.10}$$

Substituting (7.9) into (7.6) results in the MFK equation for the mean

$$\begin{aligned} \frac{d\mu_i(c)}{dt} &= \sum_{\ell \in \ell(c)} s_{\ell i}(c) \big(r_\ell(\mathbf{y}_c) + \xi_\ell(\mathbf{y}_c, \mathbf{V}_c) \big) \\ &\quad + \frac{D_i}{l} \left[\frac{\mu_i(c+1) - \mu_i(c)}{l} - \frac{\mu_i(c) - \mu_i(c-1)}{l} \right], \end{aligned} \tag{7.11}$$

where $\mathbf{V}_c$ is the $N \times N$ covariance matrix of the concentrations in subvolume c. Note that only local fluctuations feed into the evolution of the mean concentrations.

As was the case for the mean equation (7.4), although the summation on the right of (7.8) is over all reactions, only a subset of the reactions actually have non-zero contributions. Specifically, of all local reactions in the system only those in subvolume c_1 or c_2 have non-zero $s_{\ell i}(c_1)$ or $s_{\ell j}(c_2)$ and contribute to (7.8). Similarly, only diffusive reactions that add or remove molecules from subvolumes c_1 and c_2 actually contribute to (7.8). This includes all diffusive reactions originating in subvolumes c_1 and c_2 but also those originating one subvolume away from c_1 or c_2. Because of the spatial structure of our setup, which is simply a cylinder, at most 8 diffusive reactions contribute to (7.8), with the actual number depending on whether $|c_1 - c_2|$ is zero, one, or greater than one.

Considering all reactions that contribute non-zero terms to (7.8), and using (7.9) and (7.10) to approximate the necessary expectations results in the MFK equation for the covariance $\sigma_{ij}(c_1, c_2)$. Letting $\delta_{p,q}$ be the Kronecker-delta function, equal to one when $p = q$ and to zero otherwise, this MFK equation becomes:

$$
\begin{aligned}
\frac{d\sigma_{ij}(c_1,c_2)}{dt} &= \sum_{\ell\in\ell(c_1)}\left(\delta_{c_1,c_2}\frac{1}{\Omega}s_{\ell i}(c)s_{\ell j}(c)(r_\ell+\xi_\ell)+s_{\ell i}(c_1)\left[\sum_{k=1}^{N}\frac{\partial}{\partial\mu_k(c_1)}r_\ell(\mu_{c_1})\sigma_{kj}(c_1,c_2)\right]\right)\\
&+\sum_{\ell\in\ell(c_2)}\left(s_{\ell j}(c_2)\left[\sum_{k=1}^{N}\frac{\partial}{\partial\mu_k(c_2)}r_\ell(\mu_{c_2})\sigma_{ik}(c_1,c_2)\right]\right)\\
&+\frac{D_i}{l}\left[\frac{\sigma_{ij}(c_1+1,c_2)-\sigma_{ij}(c_1,c_2)}{l}-\frac{\sigma_{ij}(c_1,c_2)-\sigma_{ij}(c_1-1,c_2)}{l}\right]\\
&+\frac{D_j}{l}\left[\frac{\sigma_{ij}(c_1,c_2+1)-\sigma_{ij}(c_1,c_2)}{l}-\frac{\sigma_{ij}(c_1,c_2)-\sigma_{ij}(c_1,c_2-1)}{l}\right]\\
&+\delta_{i,j}\frac{D_i}{l^2}\frac{1}{\Omega}\Big[\delta_{c_1,c_2}\big(\mu_i(c_1+1)+2\mu_i(c_1)+\mu_i(c_1-1)\big)\\
&-\delta_{c_1+1,c_2}\big(\mu_i(c_1+1)+\mu_i(c_1)\big)-\delta_{c_1-1,c_2}\big(\mu_i(c_1)+\mu_i(c_1-1)\big)\Big]. \qquad (7.12)
\end{aligned}
$$

The first two lines of this equation correspond to the contributions of local reactions. The third (fourth) line resembles the discrete Laplacian with respect to the first (second) spatial dimension c_1 (c_2), and the last two lines are dissipation terms arising because of diffusive reactions when $|c_1 - c_2| < 2$. **Equations (7.11) and (7.12) are the spatial MFK equations for this one spatial dimension system.**

7.2.1 The Continuum Limit and Concluding Remarks

It has been argued that the RDME converges in the limit of fast diffusion (where the non-zero entries in matrix **K** above become very large) to the Chemical Master Equation of Chapter 4 (e.g., see [51, 88]). I has been also proposed in [5] that as the number of reaction partners within the diffusion range of each molecule increases, the mean concentrations converge to those of the standard reaction-diffusion equations (a system of deterministic partial differential equations for the concentrations as a function of space and time, e.g., see [87]). It would be of interest to see if the spatial MFK equations above converge to the standard MFK equations in the former limit, and to a partial differential equation model in the latter limit. Further work is needed in this regard, but below we discuss some issues in obtaining the partial differential equation model from the spatial MFK equations in the continuous space limit. For

simplicity, we here discuss some of the issues that come up in the context of the one dimensional system of Section 7.2, with spatial MFK equations (7.11) and (7.12).

Let a denote the spatial coordinate of the relevant spatial dimension of the cylinder, so that subvolume c has location $a = cl$. Perhaps a first attempt at obtaining a continuum approximation would take the limit of l going to zero, so that the volume of each subvolume goes to zero. So doing certainly turns the diffusive part of the mean equation (7.11) into the second derivative of $\mu_i(a)$ with respect to a times the diffusion constant. Unfortunately, (7.11) generally depends on the variances and covariances, which have the rate of change in (7.12) that goes to infinity as l becomes small (since Ω goes to zero in this limit).

7.3 The Spatial MFK Equations

We now consider systems where the subvolumes are connected in an arbitrary fashion to one another, to allow, for example, geometries other than the cylinder considered in the previous section, or compartmentalization. We start by arbitrarily enumerating the n subvolumes in the system, and thinking of them as nodes of the larger network (the whole system). To describe the spatial structure of the network, we draw an edge between any pair of adjacent subvolumes. If subvolume c_1 is adjacent to volume c_2, we then imagine a directional edge out of subvolume c_1 and into subvolume c_2. The direction of the edge is arbitrary (we could have imagined the edge leaves from c_2 and leads onto c_1) and has no bearing on the results we develop.

We represent this edge through an $n \times 1$ column vector $\mathbf{e} = \mathbf{e}^{\text{out}} + \mathbf{e}^{\text{in}}$, where $\mathbf{e}^{\text{out}}$ has a -1 in its c_1-th entry and is zero elsewhere. Analogously, $\mathbf{e}^{\text{in}}$ has a +1 in its c_2-th entry and is zero elsewhere. We number all the E edges in the network (with $\mathbf{e}_k$ denoting edge number k) and store them in the $n \times E$ *node-edge incidence matrix* $\mathbf{E}$, whose k-th column is $\mathbf{e}_k$. The node-edge incidence matrix summarizes the spatial structure of the system (e.g., see [28]). We will also need two other related matrices $\mathbf{E}^{\text{out}}$ and $\mathbf{E}^{\text{in}}$ whose k-th columns respectively equal $\mathbf{e}_k^{\text{out}}$ and $\mathbf{e}_k^{\text{out}}$. Note that $\mathbf{E} = \mathbf{E}^{\text{out}} + \mathbf{E}^{\text{in}}$.

Recall that $\mathbf{y}_c$ is the vector of concentrations in subvolume c, and $\mathbf{y} = [\mathbf{y}_1' \dots \mathbf{y}_n']'$ is the vector of concentrations in the whole system, with mean $\mu = E[\mathbf{y}]$ and covariance matrix $\mathbf{V} = E[(\mathbf{y}-\mu)(\mathbf{y}-\mu)']$. Because the RDME has the same mathematical form as the CME of Chapter 5, we can actually directly write down the MFK equations for μ and $\mathbf{V}$, namely

$$\frac{d\mu}{dt} = \mathbf{S}\mathbf{r}(\mu) + \mathbf{S}\xi \tag{7.13}$$

$$\frac{d\mathbf{V}}{dt} = \mathbf{M}\mathbf{V} + \mathbf{V}\mathbf{M}' + \frac{1}{\Omega}\mathbf{S}\mathrm{diag}\big(\mathbf{r}(\mu) + \xi\big)\mathbf{S}', \tag{7.14}$$

where $\mathbf{r}$ is the vector with the rate functions of all (local and diffusive) reactions in the system, $\mathbf{M} = \mathbf{S}\frac{d\mathbf{r}(\mu)}{d\mu}$, and ξ the stochastic rate vector associated with $\mathbf{r}$, and $\mathrm{diag}\big(\mathbf{r}(\mu) + \xi\big)$ is a diagonal matrix with $\mathbf{r}(\mu) + \xi$ in its diagonal.

It turns out that an appropriate ordering of the reactions in the system, combined with the ordering we have chosen for the entries of $\mathbf{y}$, result in a very pleasant structure for the MFK equations above that is quite amenable for computations. For this purpose suppose that reactions are numbered such that the first $L_{\ell 1}$ correspond to the local reactions in subvolume 1, the next $L_{\ell 2}$ correspond to the local reactions of subvolume 2, and so on, until all local reactions have been taken into account. The remaining reactions are all the diffusive reactions in the system. The column vector of rate functions is then

$$\mathbf{r}(;) = \left[\begin{array}{c} \mathbf{r}_{\mathrm{local}}(;) \\ \hline \mathbf{r}_{\mathrm{diff}}(;) \end{array}\right] = \left[\begin{array}{c} \mathbf{r}_1(;) \\ \vdots \\ \mathbf{r}_n(;) \\ \hline \mathbf{r}_{\mathrm{diff}}(;) \end{array}\right],$$

where $\mathbf{r}_c(;)$ (for $c \in \{1, \dots, n\}$) denotes the rate vector of the local reactions in subvolume c, which depend only on the local concentrations $\mathbf{y}_c$.

The Kronecker product will turn out to be quite useful for our purposes here. It is denoted by $\mathbf{A} \otimes \mathbf{B}$ for $p \times q$ matrix $\mathbf{A}$ and $m \times n$ matrix $\mathbf{B}$, and results in the

$pm \times qn$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \dots & a_{1q}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \dots & a_{2q}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1}\mathbf{B} & a_{p2}\mathbf{B} & \dots & a_{pq}\mathbf{B} \end{bmatrix}.$$

See e.g., [19], for an introduction to the Kronecker product and a summary of the main identities involving such products.

Letting $\mathbf{I}_N$ denote the $N \times N$ identity matrix, the stoichiometry matrix $\mathbf{S}$ can then actually be written as

$$\mathbf{S} = \left[\begin{array}{c|c} \mathbf{S}_\ell & \mathbf{S}_d \end{array} \right] = \left[\begin{array}{ccc|c} \mathbf{S_1} & \mathbf{0} & \mathbf{0} & \\ \mathbf{0} & \ddots & \mathbf{0} & \left[\begin{array}{cc} \mathbf{E} & -\mathbf{E} \end{array} \right] \otimes \mathbf{I}_N \\ \mathbf{0} & \mathbf{0} & \mathbf{S_n} & \end{array} \right], \tag{7.15}$$

where the left part of the matrix is block diagonal and depends only on the local stoichiometry matrices of individual subvolumes. In most cases (except for compartmentalization) the local stoichiometry blocks are the same (e.g., $\mathbf{S}_{c_1} = \mathbf{S}_{c_2} = \mathbf{S}_{\text{local}}$ for any c_1 and c_2), so then $\mathbf{S}_\ell = \big(\mathbf{I}_n \otimes \mathbf{S}_{\text{local}}\big)$.

The structure of $\mathbf{S}_d$ is most easily seen by first considering a situation with a single species reacting and diffusing in the system. In such a scenario there are twice as many diffusive reactions as there are edges in $\mathbf{E}$, two for every edge (since edges were arbitrarily assigned a direction) to account for both directions of the diffusive process. The diffusive reaction removing one molecule from the subvolume corresponding to the non-zero entry in $\mathbf{e}_k^{\text{out}}$ and adding one molecule to the subvolume corresponding to the non-zero entry in $\mathbf{e}_k^{\text{in}}$ has a stoichiometry vector equal to $\mathbf{e}_k$, while the same reaction in the opposite direction has stoichiometry vector $-\mathbf{e}_k$. For this single species case, we can then order the diffusive reactions so that the first E reactions correspond to those in the same directions as the edges in $\mathbf{E}$, and the next E reactions correspond to those in the opposite direction as the edges in $\mathbf{E}$. Then the stoichiometry of the diffusive reactions would then be $[\mathbf{E} - \mathbf{E}]$. The Kronecker

product of the previous matrix with $\mathbf{I}_N$ generalizes this structure to the situation with multiple species, where the diffusive reactions have now been ordered so that the first N correspond to diffusion along $\mathbf{e}_1$ for the N species in the system, the next N to those along $\mathbf{e}_2$ and so on; once all edges in $\mathbf{E}$ have been so accounted for we start to simularly account for the N reactions in the opposite direction of $\mathbf{e}_1$, $\mathbf{e}_2$, etc.

The average and stochastic rate vectors above are similarly given by

$$\mathbf{r}(\mu) = \left[\begin{array}{c} \mathbf{r}_{\text{local}}(\mu) \\ \hline \mathbf{r}_{\text{diff}}(\mu) \end{array}\right] = \left[\begin{array}{c} \mathbf{r}_1(\mu_1) \\ \vdots \\ \mathbf{r}_n(\mu_n) \\ \hline \mathbf{r}_{\text{diff}}(\mu) \end{array}\right] \text{ and } \xi = \left[\begin{array}{c} \xi_{\text{local}} \\ \hline \mathbf{0} \end{array}\right],$$

where $\xi_{\text{local}} = [\xi_1' \ldots \xi_n']'$ is the vector of local stochastic rates, with

$$\{\xi_c\}_\ell = \frac{1}{2}\sum_{k,l=1}^{N} \left.\frac{\partial^2 r_\ell(c)}{\partial y_k(c)\partial y_l(c)}\right|_{\mathbf{y}(c)=\mu(c)}.$$

Importantly, the rates of the diffusive reactions can be shown to be given by

$$\mathbf{r}_{\text{diff}} = \big(\left[\begin{array}{cc} -\mathbf{E}^{\text{out}} & \mathbf{E}^{\text{in}} \end{array}\right]' \otimes \mathbf{K}\big)\mu,$$

where $\mathbf{K}$ is an $N \times N$ diagonal matrix with D_i/l^2 in its i-th diagonal.

This results in the following MFK equation for the mean:

$$\begin{aligned} \frac{d\mu}{dt} &= \mathbf{S}_\ell\big(\mathbf{r}_{\text{local}} + \xi_{\text{local}}\big) - \big(\mathbf{E}\mathbf{E}' \otimes \mathbf{K}\big)\mu \\ &= \big(\mathbf{I}_n \otimes \mathbf{S}_{\text{local}}\big)\big(\mathbf{r}_{\text{local}} + \xi_{\text{local}}\big) - \big(\mathbf{E}\mathbf{E}' \otimes \mathbf{K}\big)\mu\ . \end{aligned} \tag{7.16}$$

Because $\mathbf{S}_\ell$ is a block-diagonal matrix, the first of the two terms on the right, which depends only on local reactions and local quantities, has no coupling between different subvolumes. The second term on the right, however, incorporates the coupling across subvolumes due to diffusion. Interestingly, $-\mathbf{E}\mathbf{E}'$ is the discrete-space Laplacian operator (e.g., see [28]).

Similarly, the MFK equation for the variance $\mathbf{V}$ in (7.14) can be expressed as a sum of local decoupled effects, plus diffusion-based coupling of covariances in different subvolumes. To see this, note that

$$\begin{aligned}
\mathbf{M} &= \mathbf{S}\frac{d}{d\mu}\mathbf{r}(\mu) = \overbrace{\mathbf{S}_\ell \frac{d}{d\mu}\mathbf{r}_{\text{local}}(\mu)}^{\mathbf{M}_\ell} + \overbrace{\mathbf{S}_d \frac{d}{d\mu}\mathbf{r}_{\text{diff}}(\mu)}^{\mathbf{M}_d} \\
&= \begin{bmatrix} \mathbf{S_1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S_n} \end{bmatrix} \begin{bmatrix} \frac{d}{d\mu_1}\mathbf{r}_1(\mu_1) \\ \dots \\ \frac{d}{d\mu_n}\mathbf{r}_n(\mu_n) \end{bmatrix} + \overbrace{\left(\begin{bmatrix} \mathbf{E} & -\mathbf{E} \end{bmatrix} \otimes \mathbf{I}_N \right)}^{\mathbf{S}_d} \frac{d}{d\mu} \overbrace{\left(\begin{bmatrix} -\mathbf{E}^{\text{out}} & \mathbf{E}^{\text{in}} \end{bmatrix}' \otimes \mathbf{K} \right)\mu}^{\mathbf{r}_{\text{diff}}(\mu)} \\
&= \begin{bmatrix} \mathbf{S_1}\frac{d}{d\mu_1}\mathbf{r}_1(\mu_1) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S_n}\frac{d}{d\mu_n}\mathbf{r}_n(\mu_n) \end{bmatrix} + \left(\begin{bmatrix} \mathbf{E} & -\mathbf{E} \end{bmatrix} \begin{bmatrix} -\mathbf{E}^{\text{out}} & \mathbf{E}^{\text{in}} \end{bmatrix}' \otimes \mathbf{I}_N \mathbf{K} \right) \\
&= \big(\mathbf{I}_n \otimes \mathbf{S}_{\text{local}}\big)\frac{d}{d\mu}\mathbf{r}_{\text{local}}(\mu) + \big[\big(-\mathbf{E}\mathbf{E}'\big) \otimes \mathbf{K}\big] \,, \qquad (7.17)
\end{aligned}$$

where we used the identity $(\mathbf{A}_1 \otimes \mathbf{A}_2)(\mathbf{A}_3 \otimes \mathbf{A}_4) = (\mathbf{A}_1\mathbf{A}_3) \otimes (\mathbf{A}_2\mathbf{A}_4)$ for Kronecker products, which is valid as long the the dimensions of the matrices allow the indicated products (e.g., see [19]), and assumed in the last line that $\mathbf{S}_{c_1} = \mathbf{S}_{c_2} = \mathbf{S}_{\text{local}}$. This last assumption simplifies notation, but is by no means necessary. Note that $\mathbf{M}_\ell$ is block-diagonal, but $\mathbf{M}_d$, which is a function of the discrete Laplacian operator, is generally not. Note also that $\mathbf{M}_d = \big[\big(-\mathbf{E}\mathbf{E}'\big) \otimes \mathbf{K}\big]$ is symmetric, since $\mathbf{E}\mathbf{E}'$ and $\mathbf{K}$ are symmetric and $(\mathbf{A}_1 \otimes \mathbf{A}_2)' = (\mathbf{A}_1' \otimes \mathbf{A}_2')$ for any pair of matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ (e.g., see [19]).

The MFK equation for the variance then becomes

$$\begin{aligned}
\frac{d}{dt}\mathbf{V} &= \overbrace{\mathbf{M}_\ell\mathbf{V} + \mathbf{V}\mathbf{M}_\ell' + \frac{1}{\Omega}\mathbf{S}_\ell \text{diag}\big(\mathbf{r}_{\text{local}}(\mu) + \xi_{\text{local}}\big)\mathbf{S}_\ell'}^{\text{local decoupled computation}} \qquad (7.18) \\
&\quad \underbrace{+ \big[\big(-\mathbf{E}\mathbf{E}'\big) \otimes \mathbf{K}\big]\mathbf{V} + \mathbf{V}\big[\big(-\mathbf{E}\mathbf{E}'\big) \otimes \mathbf{K}\big] + \frac{1}{\Omega}\mathbf{S}_d \text{diag}\big(\mathbf{r}_{\text{diff}}(\mu)\big)\mathbf{S}_d'}_{\text{coupled (diffusive) computation}} \,.
\end{aligned}$$

The spatial MFK equations above (7.16) and (7.18) are consistent with the corre-

sponding equations, (7.11) and (7.12), for the 1-d case in the previous section. Note again the appearance of $-\mathbf{E}\mathbf{E}'$, the discrete Laplacian operator, which seems to act both column-wise and row-wise in $\mathbf{V}$.

We leave to further work an exploration of the convergence, under appropriate limits, of the more general spatial MFK equations, (7.16) and (7.18), to the continuous space limit and to the reaction-diffusion equations. Presumably the discrete Laplacian operators in (7.16) and (7.18) will converge to the standard Laplacian operator in the appropriate limit. Additionally, the spatial MFK equations should be compared against stochastic simulation of the RDME for specific examples.

Chapter 8

Concluding Remarks and Future Work

8.1 Concluding Remarks

The study of biology from a systems perspective is not new. However, the rate at which experimental data on biological systems is being generated, for example, through microarrays or DNA sequencing, is fairly recent. Decades of reductionistic studies, combined with the more recent technological breakthroughs, have resulted in a vast amount of detailed information, often at the molecular level, about many biological systems. The complete DNA sequences for several organisms, including humans, are now completely known.

Yet, it is still difficult to see how system function arises from this detailed (molecular) picture, or to generate quantitative or even qualitative predictions about system behavior under specific environmental conditions. It is even harder still to accurately predict the collective system behavior (e.g., at the cell population, tissue, organ, or organism level) from our detailed system knowledge. This situation has led to a renewed interest in the quantitative study of biological systems, but now with an integrative approach under the so-called field of systems biology.

The potential impact of achieving a quantitative systems understanding of biology is tremendous. It can significantly enhance drug design, for example, by the

use of accurate computational models for target identification, choice of mechanism of action, and even the molecular design of the drug itself. Furthermore, a solid systems-understanding of biology, combined with molecular biology, genetics and engineering, will allow (and is starting already) the engineering of biological systems for applications as varied as medical therapies, or the efficient production of biofuels.

Fortunately for us researchers, there is much work still to be done to realize the promise of systems biology. Better and higher-throughput techniques that enable dynamic measurements at the appropriate temporal resolution of biological systems will undoubtedly play a key role in the success of systems biology. Such technologies should allow, for example, the in-vivo measurement of biomolecule concentrations, reaction rate constants, and protein chemical states (e.g., phosphorylation).

However, the promise of systems biology will be difficult to realize solely on the availability of such measurement techniques. I believe that quantitative tools to model biological systems and to interpret biological measurements (e.g., to infer dynamic phosphorylation states from fluorescence resonance energy transfer data) are also necessary, and will play an increasingly more important role in systems biology, particularly as the complexity of the systems of study increases. These quantitative methods can become an integral part of the study of biological systems, for example, through the generation of concrete quantitative hypothesis amenable for experimental testing (as in Chapter 2), or through the interpretation of experimental data. Furthermore, from a biological engineering perspective, they can quite naturally provide an efficient framework to optimize the performance of a given biological system.

Some quantitative tools and results from engineering systems theory, or more generally, from the mathematical theory of dynamical systems, can be directly applied to the study of biological systems, as exemplified in Chapter 2. However, new quantitative tools specifically suited for biological systems will have to be developed. In particular, a systematic framework to model biological systems should be at the core of the quantitative tools in systems biology. At the cellular level, a good starting point could be a framework that yields minimal, accurate, and accessible models of systems of chemical reactions when provided with the structure (e.g., stoichiometry) of a bio-

chemical network and the relative magnitudes of its parameters (e.g., rate constants). The tools and frameworks developed could prove useful not only to biology, but also to other fields that deal with chemical reactions in small volumes (e.g., microreactors). Furthermore, these tools and frameworks, and the knowledge resulting from the study of biological systems, can also inspire the engineering of non-biological systems. For example, cells seem to robustly transmit information from their environment despite the fact that their communication network (e.g., signaling pathways) is built on inherently noisy components (e.g., diffusing proteins). The mechanisms by which this is achieved could perhaps be applied to engineered communication networks.

In this thesis we have tried to contribute to the larger effort of the systems biology community in two main ways. First, in Chapter 2 we have demonstrated how system engineering techniques and results can yield significant insight, and be an integral part of experimental studies of biological systems. Key to this approach is ability to measure system dynamics in response to carefully chosen time-dependent inputs. Second, in the remaining chapters we have applied model reduction techniques to characterize the system behavior of a motif that is pervasive in signaling pathways (Chapter 3), and to simplify the models for spatially-homogeneous (Chapters 5 and 6) and spatially-inhomogeneous (Chapter 7) stochastic systems of chemical reactions.

Chapter 3 applies separation of time scales to the MAK model of the signaling cycle, a network motif pervasive in signaling pathways, and allows the characterization of its steady-state and dynamic behavior. We show that this motif can be in one of four total regimes, each with their particular behavior characteristics, and two of which had not been identified before. One of the newly-identified regimes, the signal-transducing regime, seems particularly well suited for the robust transmission of information in the presence of noise. Remarkably, the system response in this regime is largely independent of protein concentrations (it depends only on protein chemical states instead), suggesting the transmission of information in different cells can be largely identical despite significant fluctuations in molecule numbers. Another regime, the ultrasensitive one, is mathematically an integrator (until saturation is reached). When embedded in negative feedback loops, ultrasensitive cycles would

provide integral feedback and result in the perfect adaptation of system behavior (see also Chapter 2).

Chapter 5 develops the mass fluctuation kinetics (MFK) equations, an approximate model for the means, variances and covariances for the concentrations of a spatially-homogeneous chemical system. The MFK model is not much more complex than the deterministic mass action kinetic (MAK) models that many researchers use because of their simplicity, and allows for a more direct comparison of system dynamics in the presence and absence of noise. In this context, the MFK model provides a clear way to see how system size directly affects the impact stochasticity has on system behavior. For example, it is easy to show in the MFK model that, as the system volume goes to infinity while keeping all concentrations constant (which is equivalent to letting molecule numbers go to infinity), the MFK equations often converge to MAK.

Even MAK models can be quite complex and generate a rich set of system behaviors, particularly because of their potential nonlinearities. A very useful tool that reduces the complexity of MAK models has been time-scale separation, where one aims to obtain a system description for a subset of the system variables, the slow variables. In Chapter 6 we present a method to systematically identify the slow and fast variables in spatially-homogeneous chemical systems, deterministic or stochastic, given a classification of reactions into slow and fast. This identification of slow and fast system variables turns out to depend only on stoichiometric considerations. In this chapter, we then develop separation of time-scale methods for stochastic chemical kinetics that are often more accurate and generally valid than previously proposed methods.

Chapter 7 is analogous to Chapter 5, but obtains an approximate model for the means, variances and covariances for the concentrations of a *s*patially-inhomogeneous chemical system. The resulting approximation, the spatial MFK equations, are much simpler to work with than the fully stochastic model. Together with the MFK equations, they open up a variety of possibilities for further work.

8.2 Further Work

The work described in this thesis suggests several directions of future work.

The Osmo-regulation Network of *S. cerevisiae*

Chapter 2 shows that the osmo-regulation network of *S. cerevisiae* perfectly adapts to steps of extracellular osmolarity, and that this adaptation depends on the kinase activity of protein Hog1. Further research related to this system could be:

- Studying the relationship between signaling and gene expression. Hog1 is known to mediate a wide gene expression response. Using the setup we developed to measure Hog1 signaling in single cells, and tracking gene expression in response to this signaling (e.g., by fluorescently tagging appropriate gene(s)) could enable the characterization of the cell-signaling-to-gene-expression system. A simple pole-zero model of the process could be used as a starting point. It will also be of interest if, as it seems will be the case, the cell-to-cell variability in gene expression levels will be much greater than the cell-to-cell variability in the cell signaling.

- Studying signaling in more detail and at a higher temporal resolution. The MAPK cascade reaches peak nuclear enrichment within minutes; this activation period reflects the dynamics of the signaling pathway and nuclear transport of Hog1 as cleanly as possible. It would be interesting to explore these initial dynamics in more detail to study MAPK signaling at the single cell level. Several experiments that would be of interest include tuning up and down the levels of each of the gene levels of the proteins in the signaling pathway, understanding the differences in signaling between the Sln1, Sho1, and combined branches of the signaling pathways. These experiments can be performed with a significantly higher sampling frequency, and on a newer microscope than the one we used; one with laser-autofocus to minimize the noise introduced by the experimental setup. It will be very interesting to see whether the observed behavior

can be understood in the context of what would be standard models of the chemical reactions involved in the signaling pathway.

- Determining the number of integrators in the feedback loop. If other inputs can be generated, such as ramps, the number of integrators in the feedback loop of the network can be determined through the application of the internal model principle of control theory. Our measurements of the amplitude of the response to sinusoidal inputs suggest that the system transfer function has a single zero at frequency zero, and thus a single integrator in the feedback loop. This suggests, for example, that the network will not adapt perfectly to ramp inputs.

- Identifying the biomolecules performing the integration. As mentioned in the discussion of Chapter 2, proteins involved in the breakdown of glucose into glycerol that are activated and deactivated by saturated reactions, such as PFK2, are good candidates. Inserting a genetic construct to externally manipulate PFK2 concentration could suffice to move the reactions away from saturation and, if indeed PFK2 is the integrator in the network, result in the loss of integration and of perfect adaptation.

More generally, biology is rife with systems involved in homeostasis, where a biophysical quantity is finely regulated to lie within a small range of values. Integral feedback may be the common mechanism to achieve such behavior, and it would be of interest to understand the different mechanisms employed by biology to implement integral feedback through the study of other homeostatic systems.

Quantitative Tools and Frameworks to Model Biology

The list of possible research projects is fairly large but the following can be mentioned here:

- Study the precise relationships between the four modeling regimes (as a function of system size and mixing rate) for systems of chemical reactions. In particular,

articulate the appropriate mathematical limits that connect the more detailed to the less detailed regimes. The spatial MFK equations of Chapter 7 provide a natural context for this endeavor.

- Develop a systematic procedure to determine which modeling regime is the most appropriate for a given system. Such a procedure could result from a better understanding of the error between the (spatial and standard) MFK equations and their (large system) deterministic counterparts, as a function of both mixing rate (e.g., diffusion constants) and system size.

- Study the general consequences of stochasticity and spatial effects on system behavior. For example, how far from one another are the steady-states corresponding to stochastic and non-stochastic models (e.g., as a function of system size)? How does stochasticity impact the stability of a steady-state? Presumably stochasticity is often destabilizing, but by how much? Another idea is to explore how (Turing) pattern formation conditions change in the presence of stochasticity.

- Generalizing results applicable to MAK models to other modeling regimes. There are various theories providing very general results for a large class of problems in the context of MAK. For example, Metabolic Control Analysis (MCA) describes how steady-state rates and concentrations depend on system parameters; Chemical Reaction Theory (CRT) yields the number and stability of steady-states for the MAK models of a large class of systems. It would be of interest to see whether these results apply to, or can be generalized, for the other modeling regimes. Here again the MFK models are natural frameworks to explore this.

- Develop model reduction tools for spatial MFK (e.g., separation of time-scales).

- Use techniques developed to study the system behavior of typical network motifs (e.g., the signaling cycle of Chapter 3 in the presence of spatial and stochastic effects).

Bibliography

[1] J Albertyn, S Hohmann, J M Thevelein, and B A Prior. Gpd1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway. *Mol Cell Biol*, 14(6):4135–44, Jun 1994.

[2] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. Robustness in bacterial chemotaxis. *Nature*, 397(6715):168–171, 1999.

[3] M Ander, P Beltrao, B Di Ventura, J Ferkinghoff-Borg, M Foglierini, A Kaplan, C Lemerle, I Tomas-Oliveira, and L Serrano. Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. *Syst Biol (Stevenage)*, 1(1):129–138, 2004.

[4] D. Angeli, J. E. Ferrell Jr., and E. D. Sontag. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc Natl Acad Sci U S A*, 101(7):1822–1827, 2004.

[5] L Arnold and M Theodosopulu. Deterministic limit of the stochastic model of chemical reactions with diffusion. *Adv Applied Prob*, 12:367–379, 1980.

[6] K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala. Asymptotic analysis of multiscale approximations to reaction networks. *Ann Appl Probab*, 16(1):1925–1961, 2006.

[7] F Baras and M M Mansour. Microscopic simulation of chemical instabilities. *Adv Chem Phys*, 100:393–475, 1997.

[8] L. Bardwell, J. G. Cook, E. C. Chang, B. R. Cairns, and J. Thorner. Signaling in the yeast pheromone response pathway: specific and high-affinity interaction of the mitogen-activated protein (map) kinases kss1 and fus3 with the upstream map kinase kinase ste7. *Mol Cell Biol*, 16(7):3637–3650, 1996.

[9] N. Barkai and S. Leibler. Robustness in simple biochemical networks. *Nature*, 387(6636):913–917, 1997.

[10] H. C. Berg. *Random Walks in Biology*. Princeton University Press, 1993.

[11] H. C. Berg. *E. coli in Motion*. Springer, 1st edition, 2003.

[12] William Bialek and Sima Setayeshgar. Physical limits to biochemical signaling. *Proc Natl Acad Sci U S A*, 102(29):10040–10045, 2005.

[13] A Blomberg and L Adler. Roles of glycerol and glycerol-3-phosphate dehydrogenase (nad+) in acquired osmotolerance of saccharomyces cerevisiae. *J Bacteriol*, 171(2):1087–92, Feb 1989.

[14] Nils Bluthgen, Frank J Bruggeman, Stefan Legewie, Hanspeter Herzel, Hans V Westerhoff, and Boris N Kholodenko. Effects of sequestration on signal transduction cascades. *FEBS J*, 273(5):895–906, 2006.

[15] J. A. Borghans, R. J. de Boer, and L. A. Segel. Extending the quasi-steady state approximation by changing variables. *Bull Math Biol*, 58(1):43–63, 1996.

[16] Nikolay M Borisov, Nick I Markevich, Jan B Hoek, and Boris N Kholodenko. Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys J*, 89(2):951–966, 2005.

[17] J. R. Bowen, A. Acrivos, and A. K. Oppenheim. *Chem Eng Sci*, 18(177), 1963.

[18] Onn Brandman, James E Jr Ferrell, Rong Li, and Tobias Meyer. Interlinked fast and slow positive feedback loops drive reliable cell decisions. *Science*, 310(5747):496–498, 2005.

[19] John W Brewer. Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, 25(9):772–781, Sept 1978.

[20] G. E. Briggs and J. B. Haldane. A Note on the Kinetics of Enzyme Action. *The Biochemical journal*, 19(2):338–339, 1925.

[21] F. J. Bruggeman, H. V. Westerhoff, J. B. Hoek, and B. N. Kholodenko. Modular response analysis of cellular regulatory networks. *Journal of theoretical biology*, 218(4):507–510, 2002.

[22] Y. Cao, D. T. Gillespie, and L. R. Petzold. Accelerated stochastic simulation of the stiff enzyme-substrate reaction. *J Chem Phys*, 123(14):144917, 2005.

[23] Y. Cao, D. T. Gillespie, and L. R. Petzold. The slow-scale stochastic simulation algorithm. *J Chem Phys*, 122(1):14116, 2005.

[24] Y. Cao, D. T. Gillespie, and L. R. Petzold. Efficient step size selection for the tau-leaping simulation method. *J Chem Phys*, 124(4):044109, 2006.

[25] M. Chaves, E. D. Sontag, and R. J. Dinerstein. Steady-states of receptor-ligand dynamics: a theoretical framework. *Journal of theoretical biology*, 227(3):413–418, 2004.

[26] Vijay Chickarmane, Boris N Kholodenko, and Herbert M Sauro. Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation. *J Theor Biol*, 244(1):68–76, 2007.

[27] A. Chorin and O. Hald. *Stochastic Tools in Mathematics and Science*. Springer, 2006.

[28] Fan R K Chung. *Spectral graph theory*. American Mathematical Society, 1997.

[29] A. Ciliberto, F. Capuani, and J. J. Tyson. Modeling networks of coupled enzymatic reactions using the total quasi-steady state approximation. *PLoS Comput Biol*, 3(3):e45, 2007.

[30] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437(7059):699–706, 2005.

[31] Thomas M Cover and Joy A Thomas. *Elements of information theory*. Wiley, 2nd edition, 2006.

[32] P. B. Detwiler, S. Ramanathan, A. Sengupta, and B. I. Shraiman. Engineering aspects of enzymatic signal transduction: photoreceptors in the retina. *Biophys J*, 79(6):2801–2817, 2000.

[33] Hassan Dihazi, Renate Kessler, and Klaus Eschrich. High osmolarity glycerol (hog) pathway-induced phosphorylation and activation of 6-phosphofructo-2-kinase are essential for glycerol accumulation and yeast cell proliferation under hyperosmotic stress. *J Biol Chem*, 279(23):23961–8, Jun 2004.

[34] H. El-Samad, J. P. Goff, and M. Khammash. Calcium homeostasis and parturient hypocalcemia: an integral feedback perspective. *J Theor Biol*, 214(1):17–29, 2002.

[35] H. El-Samad and M. Khammash. Regulated degradation is a mechanism for suppressing stochastic fluctuations in gene regulatory networks. *Biophys J*, 90(10):3749–3761, 2006.

[36] J. Elf and M. Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research*, 13(11):2475–2484, 2003.

[37] J Elf and M Ehrenberg. Spontaneous separation of bi-stable biochemical systems into spatial domains of opposite phases. *Systems Biol*, 2:230–236, 2004.

[38] P. Érdi and J. Tóth. *Mathematical Models of Chemical Reactions. Theory and Applications of Deterministic and Stochastic Models*. Manchester University Press, Princeton, NJ, 1989.

[39] M. Feinberg. *Chemical Reactor Theory: A Review*. Prentice-Hall, 1977.

[40] M. Feinberg and F. Horn. Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chem. Eng. Sci.*, 29(1):775, 1974.

[41] M. Feinberg and F. Horn. Chemical mechanism structure and the coincidence of the stoichiometric and kinetic subspaces. *Arch. Rational Mech. Anal.*, 66(1):83, 1977.

[42] J. E. Ferrell Jr. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends in biochemical sciences*, 21(12):460–466, 1996.

[43] J. E. Ferrell Jr. Building a cellular switch: more lessons from a good egg. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 21(10):866–870, 1999.

[44] J. E. Ferrell Jr. and E. M. Machleder. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science*, 280(5365):895–898, 1998.

[45] P Ferrigno, F Posas, D Koepp, H Saito, and P A Silver. Regulated nucleo/cytoplasmic exchange of hog1 mapk requires the importin beta homologs nmd5 and xpo1. *EMBO J*, 17(19):5606–14, Oct 1998.

[46] A. Fersht. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*. W. H. Freeman, NY, 1998.

[47] M. Frankowicz, M. Moreau, P.P. Szczęsny, J. Tòth, and L. Vicente. Fast variables elimination in stochastic kinetics. *J Phys Chem*, 97(1):1891–1895, 1993.

[48] C. Gadgil, C. H. Lee, and H. G. Othmer. A stochastic analysis of first-order reaction networks. *Bull Math Biol*, 67(5):901–946, 2005.

[49] Robert Gallager. *Discrete Stochastic Processes*. Springer, 2001.

[50] C Gardiner, K McNeil, D Walla, and I Matheson. Correlations in stochastic theories of chemical reactions. *J Stat Phys*, 14:307–331, 1976.

[51] C Gardiner and M Steyn-Ross. Adiabatic elimination in stochastic systems. ii. application to reaction diffusion and hydrodynamic-like systems. *Phys Rev A*, 29:2823–2833, 1984.

[52] Peter Gennemark, Bodil Nordlander, Stefan Hohmann, and Dag Wedelin. A simple mathematical model of adaptation to high osmolarity in yeast. *In Silico Biol (Gedrukt)*, 6(3):193–214, Jan 2006.

[53] Sina Ghaemmaghami, Won-Ki Huh, Kiowa Bower, Russell W Howson, Archana Belle, Noah Dephoure, Erin K O'Shea, and Jonathan S Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–41, Oct 2003.

[54] M. A. Gibson and J. Bruck. Exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A*, 105(1):1876–1889, 2000.

[55] D Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, Jan 1977.

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

[57] D. T. Gillespie. *J Chem Phys*, 113:297, 2000.

[58] D. T. Gillespie and M. Mangel. Conditioned averages in chemical kinetics. *J Chem Phys*, 75(2):704, 1981.

[59] D. T. Gillespie, L. R. Petzold, and Y. Cao. Comment on "nested stochastic simulation algorithm for chemical kinetic systems with disparate rates" [j. chem. phys. 123, 194107 (2005)]. *J Chem Phys*, 126(13):137101; discussion 137102, 2007.

[60] D.T. Gillespie. Stochastic chemical kinetics. In S. Yip, editor, *Handbook of Materials Modeling*, pages 1735–1752. Springer, 2005.

[61] A. Goldbeter and D. E. Koshland Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc Natl Acad Sci U S A*, 78(11):6840–6844, 1981.

[62] C. A. Gómez-Uribe and G. C. Verghese. Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *J Chem Phys*, 126(2):024109, Jan 2007.

[63] C. A. Gómez-Uribe, G. C. Verghese, and L. Mirny. Operating regimes of signaling cycles: Statics, dynamics, and noise filtering. *PLoS Comput Biol*, 3(13):e45, 2007.

[64] J. Goutsias. Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J Chem Phys*, 122(18):184102, 2005.

[65] J. Goutsias. Classical versus stochastic kinetics modeling of biochemical reaction systems. *Biophys J*, 92(7):2350–65, 2007.

[66] J. Gunawardena. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc Natl Acad Sci U S A*, 102(41):14617–14622, 2005.

[67] L H Hartwell, J J Hopfield, S Leibler, and A W Murray. From molecular to modular cell biology. *Nature*, 402(6761 Suppl):47–52, Dec 1999.

[68] Eric L Haseltine and James B Rawlings. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J Chem Phys*, 117(15), 2002.

[69] Eric L Haseltine and James B Rawlings. On the origins of approximations for stochastic chemical kinetics. *J Chem Phys*, 123(16), 2005.

[70] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins. Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci U S A*, 97(5):2075–2080, 2000.

[71] F. G. Heineken, H. M. Tsuchiya, and R. Aris. *Math Biosci*, 1(95), 1967.

[72] R. Heinrich, B. G. Neel, and T. A. Rapoport. Mathematical models of protein kinase signal transduction. *Molecular cell*, 9(5):957–960, 2002.

[73] Reinhart Heinrich and Stefan Schuster. *The regulation of cellular systems*, pages 112–134. Chapman & Hall, first edition, 1996.

[74] V. Henri. Lois générales de l'action des diastases. *Paris, Hermann*, 1903.

[75] K. P. Hofmann, C. M. Spahn, R. Heinrich, and U. Heinemann. Building functional modules from molecular interactions. *Trends in biochemical sciences*, 31(9):497–508, September 2006.

[76] S. Hohmann. Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev*, 66(2):300–372, 2002.

[77] S. Hohmann, M. Krantz, and B. Nordlander. Yeast osmoregulation. *Methods Enzymol*, 428:29–45, 2007.

[78] C. Y. Huang and J. E. Ferrell Jr. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A*, 93(19):10078–83, 1996.

[79] T Jacoby, H Flanagan, A Faykin, A G Seto, C Mattison, and I Ota. Two protein-tyrosine phosphatases inactivate the osmotic stress response pathway in yeast by targeting the mitogen-activated protein kinase, hog1. *J Biol Chem*, 272(28):17749–55, Jul 1997.

[80] Joel Kaizer. *Statistical Thermodynamics of Nonequilibrium Processes*. Springer-Verlag, NY, 1987.

[81] H. Khalil. *Nonlinear Systems*. Prentice-Hall, 2002.

[82] M. Khammash and H. El-Samad. Systems biology: From physiology to gene regulation. *IEEE Control Systems Magazine*, 24(4):62–76, 2004.

[83] Mustafa Khammash and Hana El Samad. Stochastic modeling and analysis of genetic networks. In *Proceedings of 44th IEEE Conference on Decision and Control, and the European Control Conference*, dic 2005.

[84] B. N. Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *European journal of biochemistry / FEBS*, 267(6):1583–1588, 2000.

[85] B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nature reviews. Molecular cell biology*, 7(3):165–166, 2006.

[86] E. Klipp, B. Nordlander, R. Kruger, P. Gennemark, and S. Hohmann. Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol*, 23(8):975–982, 2005.

[87] Karsten Kruse and Johan Elf. *System Modeling in Cellular Biology*, pages 177–198. The MIT Press, first edition, 2006.

[88] Yoshiki Kuramoto. Effects of diffusion on the fluctuations of open systems. *Prog Theor Phys*, 52:711–713, 1974.

[89] Stefan Legewie, Nils Bluthgen, and Hanspeter Herzel. Mathematical modeling identifies inhibitors of apoptosis as mediators of positive feedback and bistability. *PLoS Comput Biol*, 2(9):e120, 2006.

[90] Stefan Legewie, Nils Bluthgen, Reinhold Schafer, and Hanspeter Herzel. Ultrasensitization: switch-like regulation of cellular signaling by transcriptional induction. *PLoS Comput Biol*, 1(5):e54, 2005.

[91] A Levchenko, J Bruck, and P W Sternberg. Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc Natl Acad Sci U S A*, 97(11):5818–5823, 2000.

[92] Andre Levchenko and Pablo A Iglesias. Models of eukaryotic gradient sensing: application to chemotaxis of amoebae and neutrophils. *Biophys J*, 82(1 Pt 1):50–63, 2002.

[93] J. Levine, H. Y. Kueh, and L. A. Mirny. Intrinsic Fluctuations, Robustness and Tunability in Signaling. *Biophys J (submitted)*, 2006.

[94] Chen Li, Qi-Wei Ge, Mitsuru Nakata, Hiroshi Matsuno, and Satoru Miyano. Modelling and simulation of signal transductions in an apoptosis pathway by using timed petri nets. *J Biosci*, 32(1):113–127, 2007.

[95] D. Liu and E. Vanden-Eijnden. Nested stochastic simulation algorithm for chemical kinetic systems with disparate rates. *J Chem Phys*, 123(19):194107, 2005.

[96] K Luyten, J Albertyn, W F Skibbe, B A Prior, J Ramos, J M Thevelein, and S Hohmann. Fps1, a yeast member of the mip family of channel proteins, is a facilitator for glycerol uptake and efflux and is inactive under osmotic stress. *EMBO J*, 14(7):1360–71, Apr 1995.

[97] T. Manninen, M. L. Linne, and K. Ruohonen. Developing Ito stochastic differential equation models for neuronal signal transduction pathways. *Computational Biology and Chemistry*, 30(4):280–281, 2006.

[98] Nick I Markevich, Jan B Hoek, and Boris N Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J Cell Biol*, 164(3):353–359, 2004.

[99] E. A. Mastny, E. L. Haseltine, and J. B. Rawlings. Two classes of quasi-steady-state model reductions for stochastic kinetics. *J Chem Phys*, 127(9):094106, 2007.

[100] J. T. Mettetal, D. Muzzey, C. Gomez-Uribe, and A. van Oudenaarden. The frequency dependence of osmo-adaptation in saccharomyces cerevisiae. *Science*, 319(5862):482–484, 2008.

[101] JT Mettetal. PhD in Physics, Massachusetts Institute of Technology, 2007.

[102] L. Michaelis and M. Menten. Die kinetik der invertinwirkung. *Biochem Z*, 49(1):333–369, 1913.

[103] B. Munsky and M. Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys*, 124:044104, 2006.

[104] M E J Newman. Modularity and community structure in networks. *Proc Natl Acad Sci U S A*, 103(23):8577–8582, Jun 2006.

[105] G Nicolis and I Prigogine. *Self-organization in nonequilibrium systems*. John Wiley & Sons, New York, 1977.

[106] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals and Sytems*. Prentice-Hall, 2nd edition, 1996.

[107] Sean M O'Rourke and Ira Herskowitz. Unique and redundant roles for hog mapk pathway components as revealed by whole-genome expression analysis. *Mol Biol Cell*, 15(2):532–42, Feb 2004.

[108] Fernando Ortega, Jose L Garces, Francesc Mas, Boris N Kholodenko, and Marta Cascante. Bistability from double phosphorylation in signal transduction. kinetic and structural requirements. *FEBS J*, 273(17):3915–3926, 2006.

[109] Saurabh Paliwal, Pablo A Iglesias, Kyle Campbell, Zoe Hilioti, Alex Groisman, and Andre Levchenko. Mapk-mediated bimodal gene expression and adaptive gradient sensing in yeast. *Nature*, 446(7131):46–51, 2007.

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

[111] J. Paulsson, O. G. Berg, and M. Ehrenberg. Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci U S A*, 97(13):7148–7153, 2000.

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

[113] S. Peleš, B. Munsky, and M. Khammash. Reduction and solution of the chemical master equation using time scale separation and finite state projection. *J Chem Phys*, 125(20):204104, 2006.

[114] Jose B Pereira-Leal, Anton J Enright, and Christos A Ouzounis. Detection of functional modules from protein interaction networks. *Proteins*, 54(1):49–57, Jan 2004. Evaluation Studies.

[115] Joseph R Pomerening, Sun Young Kim, and James E Ferrell. Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. *Cell*, 122(4):565–78, Aug 2005.

[116] M. Proft and K. Struhl. Map kinase-mediated stress relief that precedes and regulates the timing of transcriptional induction. *Cell*, 118(3):351–361, 2004.

[117] C V Rao and A P Arkin. Control motifs for intracellular regulatory networks. *Annu Rev Biomed Eng*, 3:391–419, 2001.

[118] C.V. Rao and A.P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption:application to the gillespie algorithm. *J Chem Phys*, 118(11), 2003.

[119] R H Reed, J A Chudek, R Foster, and G M Gadd. Osmotic significance of glycerol accumulation in exponentially growing yeasts. *Appl Environ Microbiol*, 53(9):2119–23, Sep 1987.

[120] VladimIr Reiser, Desmond C Raitt, and Haruo Saito. Yeast osmosensor sln1 and plant cytokinin receptor cre1 respond to changes in turgor pressure. *J Cell Biol*, 161(6):1035–40, Jun 2003.

[121] M Rep, M Krantz, J M Thevelein, and S Hohmann. The transcriptional response of saccharomyces cerevisiae to osmotic shock. hot1p and msn2p/msn4p are required for the induction of subsets of high osmolarity glycerol pathway-dependent genes. *J Biol Chem*, 275(12):8290–300, Mar 2000.

[122] M. R. Roussel and R. Zhu. Reducing a chemical master equation by invariant manifold methods. *J Chem Phys*, 121(18):8716–8730, 2004.

[123] 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 U S A*, 102(7):2310–2315, 2005.

[124] H. M. Sauro and B. N. Kholodenko. Quantitative analysis of signaling networks. *Progress in biophysics and molecular biology*, 86(1):5–43, 2004.

[125] Jörg Schaber, Bente Kofahl, Axel Kowald, and Edda Klipp. A modelling approach to quantify dynamic crosstalk between the pheromone and the starvation pathway in baker's yeast. *FEBS J*, 273(15):3520–33, Aug 2006.

[126] S. Schnell and P. K. Maini. Enzyme kinetics at high enzyme concentration. *Bull Math Biol*, 62(3):483–489, 2000.

[127] Birgit Schoeberl, Claudia Eichler-Jonsson, Ernst Dieter Gilles, and Gertraud Müller. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. *Nat Biotechnol*, 20(4):370–5, Apr 2002.

[128] L. A. Segel. On the validity of the steady state assumption of enzyme kinetics. *Bull Math Biol*, 50(6):579–593, 1988.

[129] Abhyudai Singh and Joao Hespanha. Lognormal moment closures for biochemical reactions. In *Proceedings of the 45th IEEE Conference on Decision and Control*, dec 2006.

[130] Abhyudai Singh and Joao Hespanha. Moment closure techniques for stochastic models in population biology. In *Proceedings of the 2006 American Control Conference*, jun 2006.

[131] Eduardo Sontag. Adaptation and regulation with signal detection implies internal model. *Systems & Control Letters*, 50(2):119–126, Oct 2003.

[132] Victor Spirin, Mikhail S Gelfand, Andrey A Mironov, and Leonid A Mirny. A metabolic network in the evolutionary context: multiscale structure and modularity. *Proc Natl Acad Sci U S A*, 103(23):8774–8779, Jun 2006.

[133] Victor Spirin and Leonid A Mirny. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci U S A*, 100(21):12123–12128, Oct 2003.

[134] M. Steffen, A. Petti, J. Aach, P. D'haeseleer, and G. Church. Automated modelling of signal transduction networks. *BMC bioinformatics*, 3:34, 2002.

[135] Peter S Swain and Eric D Siggia. The role of proofreading in signal transduction specificity. *Biophys J*, 82(6):2928–2933, 2002.

[136] Swiss-Prot. http://expasy.org/sprot/.

[137] M J Tamás, M Rep, J M Thevelein, and S Hohmann. Stimulation of the yeast high osmolarity glycerol (hog) pathway: evidence for a signal generated by a change in turgor rather than by water stress. *FEBS Lett*, 472(1):159–65, Apr 2000.

[138] M. Thattai and A. van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A*, 98(15):8614–8619, 2001.

[139] M. Thattai and A. van Oudenaarden. Stochastic gene expression in fluctuating environments. *Genetics*, 167(1):523–530, 2004.

[140] The Alliance for Cellular Signaling. http://www.afcs.org, 2002.

[141] M. Thorsen, Y. Di, C. Tangemo, M. Morillas, D. Ahmadpour, C. Van der Does, A. Wagner, E. Johansson, J. Boman, F. Posas, R. Wysocki, and M. J. Tamas. The mapk hog1p modulates fps1p-dependent arsenite uptake and tolerance in yeast. *Mol Biol Cell*, 17(10):4400–4410, 2006.

[142] A. R. Tzafriri. Michaelis-Menten kinetics at high enzyme concentrations. *Bull Math Biol*, 65(6):1111–1119, 2003.

[143] A. R. Tzafriri and E. R. Edelman. The total quasi-steady-state approximation is valid for reversible enzyme kinetics. *Journal of theoretical biology*, 226(3):303–303, 2004.

[144] A. R. Tzafriri and E. R. Edelman. On the validity of the quasi-steady state approximation of bimolecular reactions in solution. *Journal of theoretical biology*, 233(3):343–350, 2005.

[145] A. R. Tzafriri and E. R. Edelman. Quasi-steady state kinetics at enzyme and substrate concentrations in excess of the michaelis-menten constant. *J Theo Biol*, 245(4):737–748, 2007.

[146] J. M. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proc Natl Acad Sci U S A*, 99(9):5988–5992, 2002.

[147] Jose M G Vilar, Ronald Jansen, and Chris Sander. Signal processing in the tgf-beta superfamily ligand-receptor network. *PLoS Comput Biol*, 2(1):e3, 2006.

[148] D. Voet and J. Voet. *Biochemistry*. Wiley, 1995.

[149] Liming Wang and Eduardo D. Sontag. A remark on the number of steady states in a multiple futile cycle. *Q-bio*, 2007.

[150] J Warmka, J Hanneman, J Lee, D Amin, and I Ota. Ptc1, a type 2c ser/thr phosphatase, inactivates the hog pathway by dephosphorylating the mitogen-activated protein kinase hog1. *Mol Cell Biol*, 21(1):51–60, Jan 2001.

[151] Leor S Weinberger, John C Burnett, Jared E Toettcher, Adam P Arkin, and David V Schaffer. Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuations drive phenotypic diversity. *Cell*, 122(2):169–82, Jul 2005.

[152] Patrick J Westfall and Jeremy Thorner. Analysis of mitogen-activated protein kinase signaling specificity in response to hyperosmotic stress: use of an analog-sensitive hog1 allele. *Eukaryotic Cell*, 5(8):1215–28, Aug 2006.

[153] H Steven Wiley, Stanislav Y Shvartsman, and Douglas A Lauffenburger. Computational modeling of the egf-receptor system: a paradigm for systems biology. *Trends Cell Biol*, 13(1):43–50, 2003.

[154] S M Wurgler-Murphy, T Maeda, E A Witten, and H Saito. Regulation of the saccharomyces cerevisiae hog1 mitogen-activated protein kinase by the ptp2 and ptp3 protein tyrosine phosphatases. *Mol Cell Biol*, 17(3):1289–97, Mar 1997.

[155] John L Wyatt. Stochastic representation of bimolecular chemical reaction systems. unpublished, Oct 2006.

[156] T M Yi, Y Huang, M I Simon, and J Doyle. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci U S A*, 97(9):4649–4653, Apr 2000.