Design, Analysis, And Computational Methods For Engineering Synthetic Biological Networks

Thesis by
Ania-Ariadna Baetica

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
Defended May 18, 2018

Ania-Ariadna Baetica
ORCID: 0000-0003-0421-8181

ACKNOWLEDGEMENTS

First, I would like to acknowledge my committee members for their valuable feedback and helpful comments. I would like to thank my adviser Richard Murray for being endlessly supportive and for giving the best advice. I pretty much struck gold with Richard and I genuinely do not think I could have chosen a better adviser. He provided the correct combination of space, support, and advice to help me grow. He also exposed me to opportunities and challenges that felt outside of my reach at the time, but they served to propel me forward. Just being in the same room as Richard and watching him interact with people is an invaluable lesson in leadership. I would also like to thank him for pushing me to give every opportunity a shot because who knows you might just start that thesis or get that position. To summarize, I would like to thank Richard for teaching me how to be a leader and how to solve all of my problems in at least twelve different ways.

I would also like to thank John Doyle for being an incredibly inspiring scientist and for thinking so far ahead of the curve. John has a unique and unparalleled understanding of science that has influenced how we all think. He has reinvented and redefined what it means to be an engineer in science and, in particular, in biology. It probably took me about five years to begin to understand John's thinking, but I greatly benefited from his comments and feedback. I am also thankful to Brian Munsky for creating an environment of opportunities at the q-bio summer school and for being very patient and meticulous. I am thankful to Niles Pierce for carefully listening to my ideas and for giving me his opinions and feedback.

In addition to my committee members, I am grateful to Lea Goentoro and to Jae Hyoung Cho for exposing me to the world of experimental biology through some very handsome *Xenopus* frogs. Thanks to Michael Abrams for sharing the bench space with me and to Christopher Frick for simply lighting up the lab with his smile.

I am also grateful to my many collaborators that have worked so hard with me to advance scientific research. So many of you have become my genuine friends! Special thanks to Ye Yuan for being my first great collaborator in graduate school. Another set of special thanks to Huy Vo and Zach Fox for being not only great collaborators, but awesome friends. I would like to thank Vipul Singhal, Flora Meng, and Cody Dunn for an summer of fun and research together followed by a long time of paper writing. I would like to thank Noah Olsman for being an inspiring

collaborator and a great storyteller. I am grateful for the chance to have had mutual mentorship relationships with Cindy Ren, Reed McCardell, and Fangzhou Xiao. Like all members of the Murray lab, I am grateful to Enoch Yeung, Victoria Hsiao, Anu Thurbagere, and Shaobin Guo for being people to look up to.

My gratitude also goes out to my friends for being there for me during graduate school. I want to thank Qiaochu Yuan for laughing with me, often at me, and for always giving me her couch to crash on. I want to thank Armeen Taeb for being a great friend, Tegan Brennan for being lovely and chill, and Yoke Peng Leong for being patient and supportive. I have greatly enjoyed dinners and stories at Settebello with my friends Armeen and Ioannis Filippidis. I would also like to thank my friends Yorie Nakahira, Yong Sheng Soh, Corina Panda, Tony Bartolotta, Monica and Laura Nastasescu, Bogdan Stoica, Mickey Wang, Andrey Shur, Leo Green, Michaelle Mayalu, and Niangjun Chen. Many thanks go to Monica Nolasco, Maria Lopez, Nikki Fauntleroy, and Sydney Garstang for helping out so many times and for organizing the department tea. Many thanks to Carmen Nemer-Sirois for helping Thomas and I with our wedding. I cannot give enough thanks to Jenny Butler for always being there for me and for guiding me in the right direction. Jenny has changed the course of my life and I will always benefit from having met her.

I am thankful to my family for inspiring me to work hard for my dreams and for believing that I could achieve anything and everything. I would like to thank my grandfather Tudor for his unquestionable support and patience and my grandmother Ioana for her spirit, passion, and confidence in me. Additionally, I would like to thank my mother Georgeta for being so hard working and my father Cornel for being my first teacher.

Finally, I would like to thank my husband Thomas A. Catanach for being the most loving, generous, kind, and supportive person I have ever met. I could not have imagined that I would meet you and marry you during graduate school. You constantly amaze me with your kindness and strength. I love you very much and I am excited to spend our future together! This thesis is dedicated to you, my sweetie.

ABSTRACT

This thesis advances our understanding of three important aspects of biological systems engineering: analysis, design, and computational methods. First, biological circuit design is necessary to engineer biological systems that behave consistently and follow our design specifications. We contribute by formulating and solving novel problems in stochastic biological circuit design. Second, computational methods for solving biological systems are often limited by the nonlinearity and high dimensionality of the system's dynamics. This problem is particularly extreme for the parameter identification of stochastic, nonlinear systems. Thus, we develop a method for parameter identification that relies on data-driven stochastic model reduction. Finally, biological system analysis encompasses understanding the stability, performance, and robustness of these systems, which is critical for their implementation. We analyze a sequestration feedback motif for implementing biological control.

First, we discuss biological circuit design for the stationary and the transient distributional responses of stochastic biochemical systems. Noise is often indispensable to key cellular activities, such as gene expression, necessitating the use of stochastic models to capture their dynamics. The chemical master equation is a commonly used stochastic model that describes how the probability distribution of a chemically reacting system varies with time. Here we design the distributional response of these stochastic models by formulating and solving it as a constrained optimization problem.

Second, we analyze the stability and the performance of a biological controller implemented by a sequestration feedback network motif. Sequestration feedback networks have been implemented in synthetic biology using an array of biological parts. However, their properties of stability and performance are poorly understood. We provide insight into the stability and performance of sequestration feedback networks. Additionally, we provide guidelines for the implementation of sequestration feedback networks.

Third, we develop computational methods for the parameter identification of stochastic models of biochemical reaction networks. It is often not possible to find analytic solutions to problems where the dynamics of the underlying biological circuit are stochastic, nonlinear or both. Stochastic models are often challenging due to their high dimensionality and their nonlinearity, which further limits the availability of

analytical tools. To address these challenges, we develop a computational method for data-driven stochastic model reduction and we use it to perform parameter identification. Last, we provide concluding remarks and future research directions.

PUBLISHED CONTENT AND CONTRIBUTIONS

[Men+17] X Flora Meng, Ania-Ariadna Baetica, et al. "Recursively constructing analytic expressions for equilibrium distributions of stochastic biochemical reaction networks". In: *Journal of The Royal Society Interface* 14.130 (2017), p. 20170157. DOI: 10.1098/rsif.2017.0157.
AAB designed the stationary behavior of the two-component transcriptional network using its analytical solution. AAB conceived the project and participated in writing the manuscript.

[Ols+17] Noah Olsman, Ania-Ariadna Baetica, et al. "Hard limits and performance tradeoffs in a class of sequestration feedback systems". In: *bioRxiv* (2017), p. 222042. DOI: 10.1101/222042.
AAB determined the stability and the performance of the sequestration feedback controller with controller species degradation. AAB conceived the project and participated in writing the manuscript.

[Ren+17] Xinying Ren, Ania-Ariadna Baetica, et al. "Population regulation in microbial consortia using dual feedback control". In: *bioRxiv* (2017). DOI: 10.1101/120253.
AAB analyzed the biological controller implemented by the chemical reaction network. AAB participated in the writing of the manuscript.

[Bae+16] Ania-Ariadna Baetica, Thomas A Catanach, et al. "A Bayesian approach to inferring chemical signal timing and amplitude in a temporal logic gate using the cell population distributional response". In: *bioRxiv* (2016). DOI: 10.1101/087379.
AAB adapted and implemented the stochastic temporal logic gate model. AAB formulated the research question and participated in writing the manuscript.

[Bae+15] Ania-Ariadna Baetica, Ye Yuan, et al. "A stochastic framework for the design of transient and steady state behavior of biochemical reaction networks". In: *2015 IEEE 54th Annual Conference on Decision and Control (CDC)*. IEEE. 2015, pp. 3199–3205. DOI: 10.1109/CDC.2015.7402699.
AAB mathematically formulated and solved the distributional design problem as an optimization problem. AAB also wrote the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 A brief introduction to synthetic biology

The field of synthetic biology engineers novel organisms, devices and systems for the purposes of improving industrial processes [Nar+16; Hem+10; SWS08; CG10; DKM10], discovering the principles of biological systems [GCC00; EL00; GHM14], and performing computation and compact information storage [Thu17; Hsi+16]. In synthetic biology, we build systems inspired by natural biological systems, but we do not restrict ourselves to parts and designs already available. The numerous applications of synthetic biology include industrial fermentation [GM12], waste detection [Sim+08], biosensors development [Hsi+16], diagnostics detection [Par+14], materials production [Ngu+15; CG08], novel protein design [Arn98], information storage [Thu17], and biological computation [Moo+12]. Since synthetic biology is a relatively novel field, its applications and capabilities are still expanding.

However, synthetic systems have a unique set of challenges and limitations. A goal of synthetic biology has been to engineer reliable, robust circuits composed of standardized parts that can easily be combined together [CLE08; Ark08; Kwo10; KC10; QD17]. Nevertheless, it has been demonstrated that synthetic circuits depend on their biological implementation [DM15; Yeu+14; Yeu+17]. Slight differences in the circuit's tuning, implementation in a different model organism [KZH09], or different experimental conditions can all cause synthetic circuits to cease functioning [GMB16; PW09]. This limits the modularity of synthetic circuits and the engineering of larger circuits. Additionally, synthetic systems are often subjected to strict resource limits that result in limited functionality and in competition with the host organism [QMD17; GD14]. Furthermore, synthetic systems oftentimes rely on transcriptional parts, which results in slower response timescales than natural biological systems, particularly when the product is a protein [Ngu+15].

For a comprehensive introduction to synthetic biology, we suggest the review papers [CLE08; Ark08; Kwo10; KC10; PW09] and the textbooks [DM15] and [Alo06].

1.2 Biochemical kinetics

Notation

Let $\mathbb{Z}$ and $\mathbb{R}$ be the integer set and the real set, respectively.

Let $n \geq 1$ be an integer. Let $P \in [0,1]^n$ be the n-dimensional probability vector set. For $p = (p_1, \ldots, p_n) \in P$, it must be that $p_i \geq 0$ and $\sum_{i=1}^{n} p_i = 1$. We denote by A^T the transpose of the matrix A.

Let the symbols A−B and A : B represent the binding of two molecules A and B into a complex.

For our synthetic circuit diagrams, we adhere to the conventions of the Synthetic biology open language [Gal+12].

Deterministic chemical kinetics

Our assumptions for deterministic chemical kinetics are that they employ either unimolecular or bimolecular reactions and mass action kinetics, unless otherwise specified. The law of mass action states that the rate of a chemical reaction is proportional to the product of the concentrations of the reactants. We consider a chemically reacting network with N species $\{S_1, \ldots, S_N\}$ and M chemical reactions $\{R_1, \ldots, R_M\}$. Then we can express the mass action kinetics as

$$\frac{dx}{dt} = Wv(t), \tag{1.1}$$

where x is the concentration of the chemical species, $t \in \mathbb{R}_{\geq 0}$ is the time, $W \in \mathbb{R}^{N \times M}$ is the stoichiometry matrix, and $v(x) \in \mathbb{R}^{M \times 1}$ is the reaction flux vector. Each row of the flux vector $v(x)$ corresponds to the rate at which a given reaction occurs. The columns of the stoichiometry matrix denote the changes in concentration of the chemical species. We describe the stoichiometry matrix in more detail in Section 1.2.

Example 1. We derive the deterministic model of an example biochemical reaction network. Let the following chemically reacting system with species S_1 be described by the two reactions:

$$\mathrm{S}_1 \underset{k_1}{\overset{k_2}{\rightleftharpoons}} \emptyset. \tag{1.2}$$

Here species S_1 can represent mRNA or protein that is being created at rate k_1 and degraded at rate k_2. Then the stoichiometry matrix is $W = (1, -1)$ and the reaction

flux vector is $v(t) = (k_1, k_2 S_1(t))^T$. Therefore, the mass action kinetics are given by:

$$\frac{dS_1(t)}{dt} = k_1 - k_2 S_1(t). \tag{1.3}$$

In addition to mass action kinetics, we employ Hill kinetics to describe the cooperative binding of ligands to a macromolecule [Hil10]. Cooperative binding means that the binding of a ligand to a macromolecule is enhanced by the presence of other ligands that are already bound to the macromolecule. Moreover, the Hill coefficient quantifies the degree of cooperativity between ligand binding sites. In the famous example of the O_2 molecule binding haemoglobin in red blood cells to form oxyhaemoglobin and be transported to body tissues, the Hill equation models these saturated binding dynamics. At low oxygen concentrations, there is haemoglobin in the blood and almost no oxyhaemoglobin in body tissues. However, at high oxygen concentration, there is almost no haemoglobin in the blood and lots of oxyhaemoglobin in body tissues. Each haemoglobin molecule can bind up to four O_2 ligands, which would suggest a Hill coefficient of four. In practice, the Hill coefficient is 2.8 due to the limitations of the Hill modeling framework and due to the haemoglobin molecule existing in multiple states. The Hill equation can be represented as

$$\mathrm{Hb} + 4\,\mathrm{O_2} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{Hb{-}4\,O_2}, \tag{1.4}$$

where k_a and k_d are the association and disassociation constants, Hb is the haemoglobin molecule, and $\mathrm{Hb{-}4\,O_2}$ is the oxyhaemoglobin molecule. We let the dissociation constant K_d be the ratio of k_d and k_a. Then the concentration of oxyhaemoglobin, $[\mathrm{Hb{-}4\,O_2}]$, can be computed as:

$$f_{\mathrm{Hill}}(\mathrm{O_2}) = \frac{[\mathrm{O_2}]^4}{K_d + [\mathrm{O_2}]^4}. \tag{1.5}$$

In this thesis, we employ the Hill function to model the transcription of a gene product when its DNA is regulated by multiple transcription factors such as activators and repressors [Alo06]. It is assumed that the transcription factors bind DNA cooperatively. Let A and R be the activator and the repressor, respectively, with basal expression levels of A_0 and R_0 and dissociation constants K_A and K_R. Let n_{Hill} be the Hill coefficient and α be the maximal transcription rate of the gene. Then the

Hill functions associated with the activator, $f^a_{\text{Hill}}(A)$, and the repressor, $f^r_{\text{Hill}}(R)$, are as follows:

$$f^a_{\text{Hill}}(A) = \frac{\alpha\left(\frac{A}{K_A}\right)}{1+\left(\frac{A}{K_A}\right)^{n_{\text{Hill}}}} + A_0, \qquad f^r_{\text{Hill}}(R) = \frac{\alpha}{1+\left(\frac{R}{K_R}\right)^{n_{\text{Hill}}}} + R_0. \tag{1.6}$$

For more details, we refer the reader to [Alo06] and [DM15].

Stochastic modeling with the chemical master equation

Stochasticity in biochemical systems comes from the thermodynamics of molecular reactions [DM15]. We consider a chemically reacting network with N species $\{S_1, \ldots, S_N\}$ and M monomolecular or bimolecular chemical reactions $\{R_1, \ldots, R_M\}$, as in [Gil07]. The dynamical state of the system at time $t \geq 0$ is described by the state vector $x(t) = (x_1(t), \ldots, x_N(t))$, where $x_i(t)$ is the integer population of species S_i at time t for all $1 \leq i \leq N$. The M chemical reactions change the state of the system according to the propensity function associated with each reaction.

The chemical master equation (CME) describes how stochastically reacting chemical species behave in a well-stirred solution at thermal equilibrium in a fixed, finite volume [Gil07]. The chemical kinetics of the N reacting molecular species are modeled as a discrete-state, continuous-time Markov process. A state x of the Markov process is a combination of counts of the N reacting molecular species, while the distribution state vector $p(x,t)$ denotes the probability that the system will be in state x at time t. The CME gives the time evolution law for $p(x,t)$ as the ordinary differential equation

$$\frac{\partial p}{\partial t}(x,t) = \sum_{j=1}^{M}(a_j(x-\xi_j)p(x-\xi_j,t) - a_j(x)p(x,t)), \tag{1.7}$$

where ξ_j is the j^{th} column of the stoichiometry matrix and a_j is the j^{th} propensity function associated with the chemical reaction network [Gil00].

More compactly, the CME can be written as an ordinary differential equation

$$\frac{dp}{dt}(x,t) = H(c)p(x,t), \tag{1.8}$$

where $c = (c_1, \ldots, c_K)$ are the K rate reaction parameters of the M chemical reactions, $K \in \mathbb{Z}$.

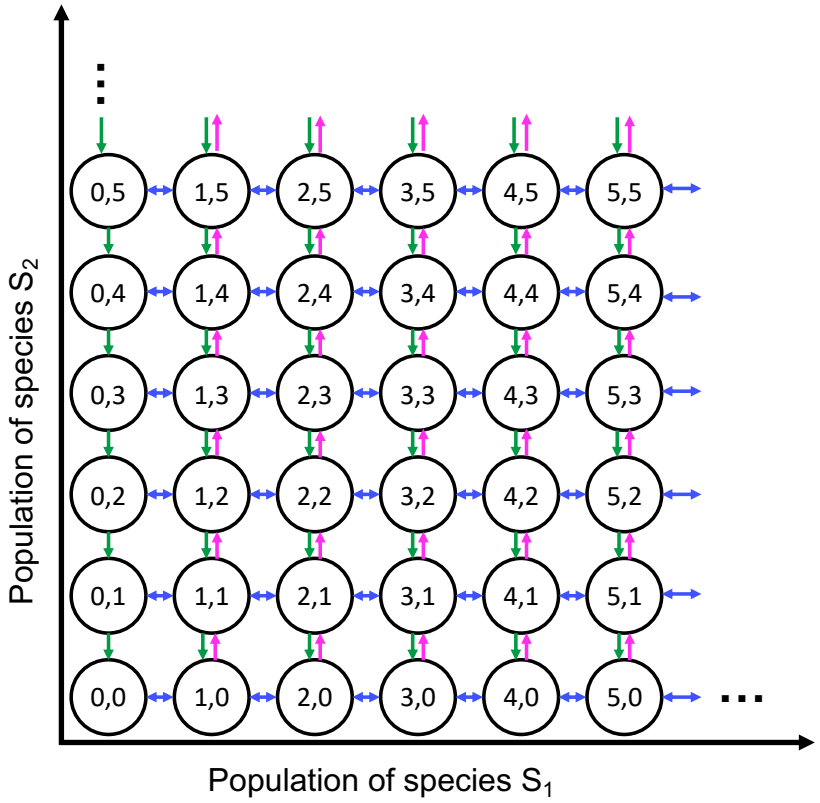

Figure 1.1: **The Markov state space associated with the chemical master equation.** We represent the two-dimensional, infinite Markov state space of a toy biochemical reaction network with species S_1 and S_2 ($N = 2$) and chemical reactions given in equations (1.9)-(1.11) ($M = 4$). The state x of the system is a two-dimensional vector whose values are all possible combinations of counts of species S_1 and S_2; therefore, x is infinite. The arrows represent the possible transitions between states, according to the four chemical reactions in equations (1.9)-(1.11) and under the assumption that state counts must be at least zero. The state transitions corresponding to the two chemical reactions in equation (1.9) are drawn as blue bidirectional arrows since species S_1 can either be created or degraded. The state transitions for the chemical reaction in equation (1.10) are drawn as green arrows pointing downwards to represent the degradation of species S_2. Finally, the state transitions for the chemical reaction in equation (1.11) are drawn as fuchsia arrows to represent the creation of species S_2; since species S_2 cannot be created in the absence of species S_1, the states with zero counts of S_1 do not have arrows. The probability of the transition between two states is given by the propensity function $a_j(x)$, associated to the j^{th} chemical reaction in the system and dependent on the value of state x. Therefore, transitions between some states are less likely to occur, an observation which is the basis of the finite state projection method.

Example 2. We derive the chemical master equation model for an example biochemical reaction network. Let the following chemically reacting system with species S_1 and S_2 be described by the set of reactions:

$$\mathrm{S}_1 \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \emptyset, \tag{1.9}$$

$$\mathrm{S}_2 \xrightarrow{k_3} \emptyset, \tag{1.10}$$

$$\mathrm{S}_1 \xrightarrow{k_4} \mathrm{S}_1 + \mathrm{S}_2. \tag{1.11}$$

Here species S_1 can represent mRNA and species S_2 can represent protein.

Then the propensity functions associated with the four chemical reactions are $a_1(x_1, x_2) = k_1 x_1$, $a_2(x_1, x_2) = k_2$, $a_3(x_1, x_2) = k_3 x_2$, $a_4(x_1, x_2) = k_4$ and the state-change vectors are $\xi_1 = (-1, 0)^T, \xi_2 = (1, 0)^T, \xi_3 = (0, -1)^T, \xi_4 = (0, 1)^T$, where x_1 is the number of molecules of species S_1 and x_2 is the number of molecules of species S_2.

Therefore the CME associated to this system is as follows:

$$\frac{\partial p}{\partial t}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, t\right) = k_1(x_1 + 1)p\left(\begin{bmatrix} x_1 + 1 \\ x_2 \end{bmatrix}, t\right) + k_2 p\left(\begin{bmatrix} x_1 - 1 \\ x_2 \end{bmatrix}, t\right) + \tag{1.12}$$

$$+ k_3(x_2 + 1)p\left(\begin{bmatrix} x_1 \\ x_2 + 1 \end{bmatrix}, t\right) + k_4 p\left(\begin{bmatrix} x_1 \\ x_2 - 1 \end{bmatrix}, t\right) - \tag{1.13}$$

$$- (k_1 x_1 + k_2 + k_3 x_2 + k_4)p\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, t\right), \tag{1.14}$$

for $x_1, x_2 \in \mathbb{Z}$, $x_1 > 0$, $x_2 > 0$.

A simulation of this system is illustrated in Figure 1.2.

The finite state projection algorithm

The finite state projection approach (FSP) [MK06] truncates the state space of chemical master equation and collects the probability mass that leaves the truncated region in a sink state, $g(t)$. The CME state space is split into two complete and disjoint sets, indexed by J and J', such that the CME becomes

$$\frac{d}{dt}\begin{bmatrix} p_J \\ p_{J'} \end{bmatrix} = \begin{bmatrix} H_{JJ} & H_{JJ'} \\ H_{J'J} & H_{J'J'} \end{bmatrix}\begin{bmatrix} p_J \\ p_{J'} \end{bmatrix} \tag{1.15}$$

The states in J' are then combined into a single sink state $g(t)$. Probability mass entering $g(t)$ is restricted to remain there for all time (i.e. probability may not leave

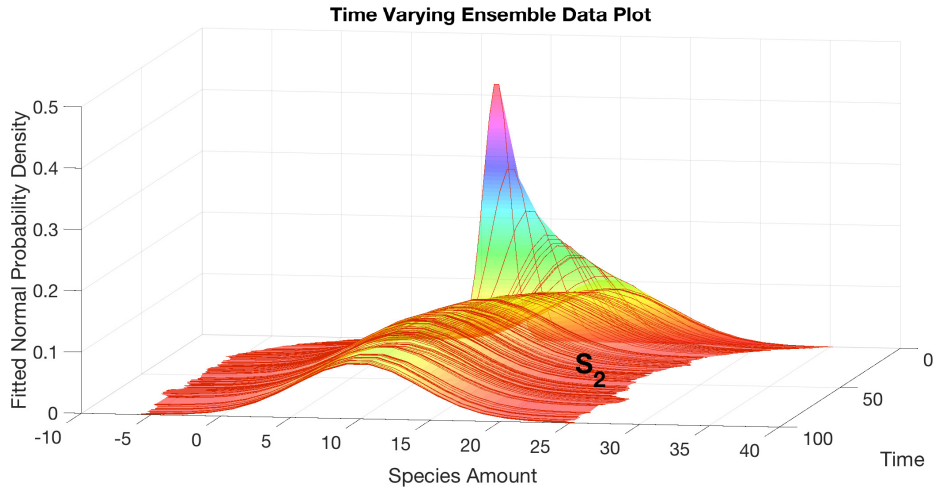

Figure 1.2: **The time-varying probability distributions of biochemical species S_2 computed using the chemical master equation.** For chemical reactions in equations (1.9)-(1.11) with reaction rate parameter values $k_1 = k_4 = 100$, $k_2 = 1$, $k_3 = 0.1$, we ran 1000 simulations for 100 seconds each to generate the set of probability distributions of the biochemical species S_2. We subsequently fit Gaussian distributions to the probability distributions from the CME for ease of interpretation.

the sink state after exiting J). This new master equation is known as the finite state projection, and can be written as

$$\frac{d}{dt}\begin{bmatrix} p_J^{FSP} \\ g(t) \end{bmatrix} = \begin{bmatrix} H_{JJ} & \mathbf{0} \\ -\mathbf{1}^{\mathbf{T}} H_{JJ} & 0 \end{bmatrix} \begin{bmatrix} p_J^{FSP} \\ g(t) \end{bmatrix} \tag{1.16}$$

The FSP approximation is guaranteed to be a lower bound on the true solution,

$$\begin{bmatrix} p_J^{FSP} \\ \mathbf{0} \end{bmatrix} \leq \begin{bmatrix} p_J \\ p_{J'} \end{bmatrix} \text{ for all } t > 0, \tag{1.17}$$

and it yields an exact error in the approximation,

$$\left| \begin{bmatrix} p_J^{FSP} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} p_J \\ p_{J'} \end{bmatrix} \right|_1 = g(t), \tag{1.18}$$

where $|\mathbf{v}|_1$ denotes the norm-1 of the vector. Proofs of these results are available in [MK06; MK07].

While the FSP approximation can significantly reduce the state space, the exponential increase in states with the number of species still limits the FSP approximation to be a practical choice for simulation of systems up to 10 species and for model identification of systems with up to 3 species and relatively low ($\approx 10^2$) molecule counts.

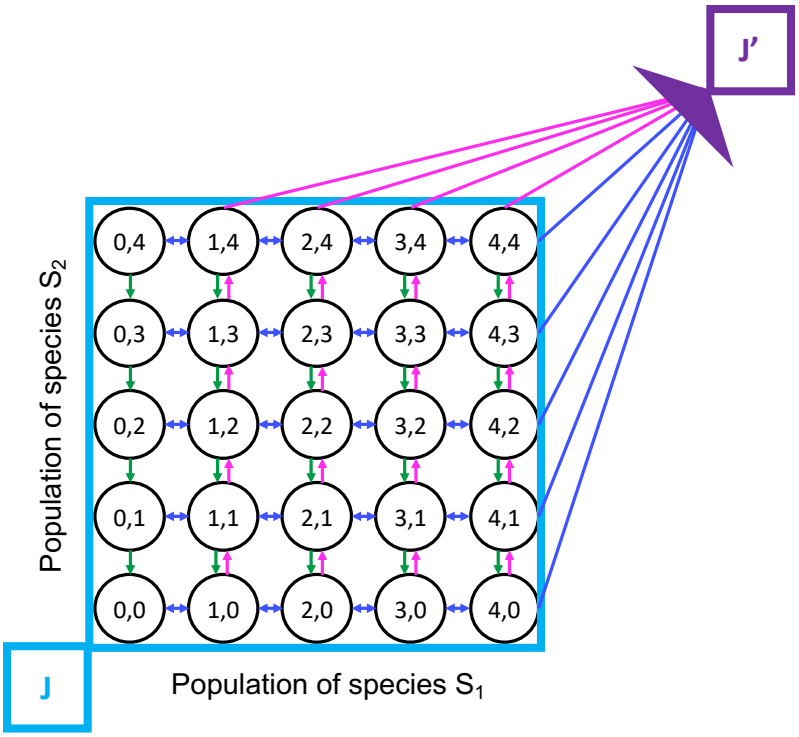

Figure 1.3: **The finite state projection method.** For chemical reactions in 1.9- 1.11, we truncate the infinite state space of the associated CME to a finite state space J (blue rectangle). The set J' (purple) collects all the remaining probability and once it enters it, probability does not escape it. The finite state space J is illustrated and is often chosen to be a hyperrectangle, but it can have any shape.

The matrix exponential solution to the chemical master equation

Using the finite state projection algorithm in [MK06], we truncate the Markov state space of the chemical master equation. Thus, we consider only a finite number of states S in each species of a biochemical reaction network.

In Chapters 2 and 4, we assume that the propensity functions a_j are linear in the reaction rate variables for all $1 \leq j \leq M$. Then the Markov transition matrix $H(c)$ is finite and we can represent it as the following sum:

$$H(c) = \Sigma_{j=1}^{M} c_j H_j, \tag{1.19}$$

where each matrix H_j corresponds to chemical reaction R_j for $1 \leq j \leq M$. The matrices H_j are sparse and of size $S^N \times S^N$.

Hence, equation (1.8) is equivalent to

$$\frac{dp}{dt}(x,t) = \Sigma_{j=1}^{M} c_j H_j p(x,t), \tag{1.20}$$

which implies linearity with respect to the reaction rates $c_1, \ldots, c_M$.

Therefore, the solution to equation (1.20) is given by

$$p(x,t) = e^{\sum_{j=1}^{M} c_j H_j t} p(x,0). \tag{1.21}$$

Remark 1. The exponential operator $e^{\sum_{j=1}^{M} c_j H_j t}$ is not separable into the product $\prod_{j=1}^{M} e^{c_j H_j t}$, unless the matrices $H_1, \ldots, H_M$ commute pair-wise. This is not usually the case, unless all the M reactions in the system are monomolecular and all the matrices are diagonal [JH07]. For bimolecular reactions, the corresponding matrices do not generally commute.

1.3 Control theoretical concepts for synthetic biology

In this section, we introduce the engineering perspective on the development of synthetic circuits. The engineering design cycle described in Figure 1.4 can be employed in synthetic biology; this approach has previously been used in mechanical and electrical engineering [AM08]. According to the engineering design cycle, synthetic circuits can be iteratively designed, built, tested, and improved until they achieve the desired performance standards. This thesis contributes to design through the performance specifications in Chapter 2, to testing through the performance properties in Chapter 3, and to improvement through the recommendations in Chapter 3. The computational methods in Chapter 4 contribute to both the design of synthetic circuits and to the evaluation of their performance.

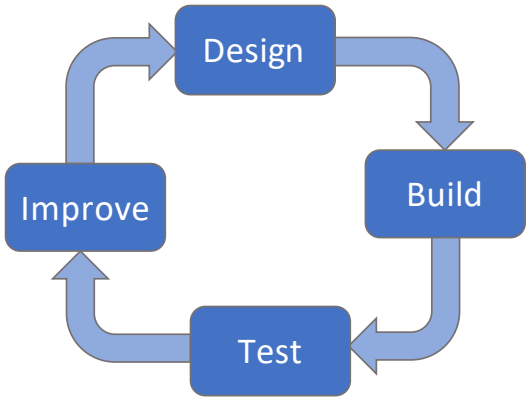

Figure 1.4: **The engineering design cycle as a methodical approach to problem solving.** The design cycle is a process by which an engineering (or biological) system is iteratively designed, built, tested, and refined until it becomes sufficiently improved in performance. Engineers follow this series of steps to solve a problem and to design a solution for it. The engineering design cycle and its repetitions represent an iterative process.

A common method to improve the performance of mechanical and electrical systems is feedback control. The principle of feedback control is to measure the error between the desired and the current performance of a circuit and to take corrective action as necessary [AM08]. Feedback is ubiquitous in endogenous biological systems, where it serves to regulate their behavior. Examples of feedback control found in endogenous systems include the regulation of body temperature [Wer10], of circadian rhythms [Rus+07], of calcium [EGK02], and of glycolysis [CBD11]. Analogously, we are developing feedback control to regulate the behavior of synthetic systems and to ensure that they behave robustly.

Feedback control provides undeniable benefits to biological, mechanical, and electrical systems. The foremost benefit is robustness to uncertainty; should the system undergo a change such an external disturbance, feedback corrects this change and ensures the system retains good performance properties. Additionally, feedback can stabilize and speedup an unstable or slow process. Yet, feedback control also comes with several drawbacks. Feedback can inadvertently amplify noise inside a system and it can also exacerbate instability if poorly designed. For a more detailed understanding of feedback control, particularly in the context of synthetic biology, see the textbooks [DM15; AM08].

In Figure 1.5, we introduce the reference tracking setup that we use for biological feedback control throughout Chapter 3. The goal of reference tracking is for a system's output to track a pre-determined reference signal. The error signal measures the difference between the output and the reference as a function of time. When the error converges to a real value e^{ss} as a function of time, the system is deemed stable and its steady state value is e^{ss}. Conversely, if there is no value $e^{ss} \in \mathbb{R}$ that the error converges to, then the system is unstable. The magnitude of the steady state error e^{ss} is a performance measure of the system in Chapter 3; a common goal is to ensure a small magnitude of e^{ss}.

Several common types of feedback control include proportional, integral, and derivative control. As explained in [DM15; AM08], proportional control responds proportionally to the error between the reference and the current output state for small errors. By itself, proportional control does not ensure a small steady state error and therefore it can result in poor performance. Integral control responds proportionally to the integral of the error signal over time. If a steady state of the closed loop system exists, then integral control guarantees zero steady state error ($e^{ss} = 0$). Derivative control responds proportionally to the derivative of the error signal and it

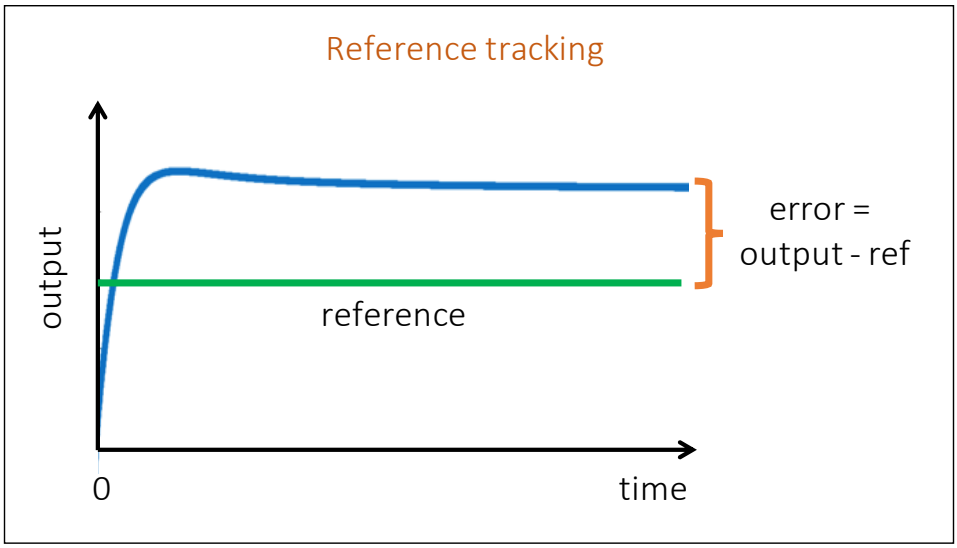

Figure 1.5: **Reference tracking.** The error measures the difference between the output of the process and the reference specified by the user. When the error converges to a real value as a function of time, we refer to it as the steady state error. Additionally, the closed loop system is deemed stable. If the error does not converge to a real value as a function of time, then the closed loop system is unstable. The magnitude of the steady state error will be used as a performance measure in Chapter 3 for the closed loop system. The reference can vary as a function of time.

aims to anticipate future errors; however, it can also amplify high frequency process noise. In practice, proportional and integral control are used ubiquitously to regulate mechanical and electrical systems [AM08].

There are multiple differences between implementing feedback control in engineered systems and in biological systems. One difference is that the reference signal may not be an explicit signal, but rather the result of other biochemical dynamics; additionally, when implemented by chemical reactions (such as the constitutive production of a biochemical species in Chapter 3), the reference can be dependent on temperature and pH, subjected to disturbances itself, or coupled to other dynamics external to the circuit. The differences between engineered and biological systems have created questions about the performance of synthetic circuits. We discuss this topic in Chapter 3 and we also further address the differences between engineered and biological feedback systems.

1.4 Contribution overview

In Chapter 2, we design the distributional response of stochastic biochemical networks whose dynamics are modeled by stochastic framework of the chemical master

equation. When available [JH07; GLO05; GSN12; SS08; GD12; ACK10], analytic solutions to the CME can greatly simplify the design of distributional responses of biochemical networks. Nonetheless, analytic solutions are challenging to find due to the high dimensionality of the Markov space underlying the CME model, which scales exponentially with the number of species in the network. In Meng *et al.* [Men+17], we have described recursive algorithms for gluing together simple Markov state spaces at one or two vertices to derive analytic stationary solutions to CME models with large Markov spaces [MHK14; MP15]. Using a recursive algorithm, we have derived the analytic stationary solution to the transcriptional network example in [GD12]. In Chapter 2, we illustrate how to use this analytic stationary solution to design the transcriptional network's stationary behavior. We employ this transcriptional network example to illustrate how designing the stochastic behavior of biochemical reaction networks is simplified by the availability of an analytic solution to the CME model.

Subsequently, we propose a general framework for the design of stochastic behaviors of biochemical reaction networks for which analytical solutions to their CME models may not be available [Bae+15]. Design specifications for distributions include specifying their modality, the locations of their modes, and their rate of convergence to stationarity [MG09]. We formulate these specifications as constraints in an optimization program that finds the reaction rate values that achieve in the desired distributional design. We apply our stochastic design framework to examples of biochemical reaction networks such as a protein production-degradation network [DM15], the Schlögl model [Gil91; Gun+05], and the genetic toggle switch model [GCC00] to illustrate its strengths and limitations.

The content of Chapter 2 has been published in [Bae+15] and [Men+17].

Contribution summary in [Men+17]:

- Designed the behavior of a two-component transcription network using the analytic form of its stationary distribution.

Contribution summary in [Bae+15]:

- Developed a framework for the design of stochastic behaviors of biochemical reaction networks.

- Mathematically described distributional design specifications of modality, location of modes, and convergence to stationarity.
- Formulated and solved the distributional design problem as an optimization problem constrained by the design specifications.

The development of synthetic biological controllers for microbes and fungi can help address problems in human health through scheduled oral probiotic delivery, in industrial fermentation through the improved commercial production of enzymes, and in waste recycling through the improved treatment of sewage effluent for drinking water. In Chapter 3, we use methods from control theory to determine the properties of stability and performance of biological controllers implemented by sequestration feedback. Using these metrics of stability and performance, we provide guidelines for the implementation of robust synthetic biological controllers.

First, we introduce biological control using sequestration feedback and we demonstrate that the controller species' sequestration binding strength, the process species' degradation rates, and the controller species' degradation rates affect their stability and performance properties. We then derive an analytical criterion for stability and we tune the stability margin by increasing either the process or the controller species degradation rates.

Second, we determine the performance of sequestration feedback networks when the controller species are degraded and diluted due to cell division or fungi budding. It has been demonstrated that a stable sequestration feedback controller with no controller species degradation implements perfect adaptation, which provides these biological controllers with robustness and zero steady state error [BGK16]. Moreover, we derive performance results that consider the controller species degradation and dilution and we provide robust implementation choices that guarantee a small steady state error.

Finally, we describe a tradeoff between the stability and the performance of sequestration feedback networks. Additionally, we give guidelines for building sequestration feedback networks with respect to this tradeoff.

The content of Chapter 3 has been published in [Ols+17] and [Ren+17].

Contribution summary to [Ols+17]:

- Incorporated the degradation of the controller species in the model for the sequestration feedback network.

- Determined the stability and the performance of the sequestration feedback controller with controller species degradation.
- Provided guidelines for the implementation of sequestration feedback networks.

Contribution summary to [Ren+17]:

- Analyzed the cell population controllers implemented using sequestration feedback networks.

In Chapter 4, we introduce a novel method for parameter identification of stochastic biochemical reaction networks. Probabilistic models are necessary to accurately represent the stochastic dynamics of biochemical reaction networks [Elo+02]. Since analytical solutions to stochastic models of biochemical reaction networks are rarely known [JH07; DS66; GLO05; GD12; MHK14; Men+17], these models are often solved computationally. However, stochastic models grow exponentially in complexity with the number of biochemical species considered [MNO12]. Thus, parameter identification methods such as the Metropolis-Hastings algorithm [Bro+11; Gel+14], which require computing the solution of the stochastic models for each parameter set, are very computationally intensive. Efficient and scalable computational methods for the parameter identification of stochastic biochemical methods are still being developed [KRS15].

Our method proposes to reduce the computational cost of solving stochastic models by two sequential projections of their dynamics. First, the finite state projection algorithm [MK06; MK07] reduces the state space of stochastic models by rendering the state space finite and by eliminating states with low probability. Subsequently, we project this reduced stochastic model onto a subspace spanned by radial basis functions [Fas07]. Other projections of stochastic models of biochemical reaction networks have been discussed in [MK08; JU10; TFM12; Zha+10; KRS15]; however, our method differs in that the radial basis functions are computed from simulated single-cell data of the stochastic model using the adaptive residual subsampling algorithm in [DH07]. We demonstrate that performing parameter identification on these reduced stochastic models results in a small loss in parameter accuracy, but in a large gain in computational efficiency.

The content of Chapter 4 has been developed in collaboration with Huy Vo, Zachary Fox, and Brian Munsky (Colorado State University).

Contribution summary to Chapter 4 by the author:

- Contributed to computing the finite state projections of the two chemical master equation models.
- Contributed to the code for the radial basis function projections.
- Performed parameter identification using the adaptive Metropolis-Hastings algorithm for the bursting gene example.

Chapter 2

STOCHASTIC BIOCHEMICAL SYSTEMS DESIGN

Stochasticity plays an essential role in the dynamics of biochemical systems. Stochastic behaviors of bimodality, excitability, and fluctuations have been observed in biochemical reaction networks at low molecular numbers. These stochastic behaviors can be described by modeling the biochemical system using the chemical master equation, a forward Kolmogorov equation in the biochemical literature. The chemical master equation describes the time evolution of probability distributions of biochemical species in the system. Analytic solutions to the chemical master can help expedite multi-scale simulations, identify system parameters, and design desirable stochastic behaviors. However, due to the large dimensionality of the state space of the chemical master equation, analytical solutions are rarely known.

In this chapter, we provide methods to design the behavior of stochastic biochemical systems by tuning the rates of the underlying biochemical reaction network model. We first demonstrate how to design the stationary stochastic behavior of a transcriptional network with a known analytical solution of its chemical master equation model. Then we introduce a method for the design of stochastic behaviors of biochemical systems when an analytical solution is not available. We focus on specifying the behaviors of the time evolving probability distributions that describe these stochastic behaviors. Our design specifications include the probability distributions' modality, the locations of their modes, and their rate of convergence to stationarity. We formulate these specifications as constraints in an optimization program that finds the optimal reaction rate parameters of the underlying biochemical network. We apply our stochastic design framework to examples of biochemical reaction networks to illustrate its strengths and limitations. We hope that our stochastic design method can contribute to the design-built-test engineering cycle for synthetic biology.

The results in this chapter were published in [Men+17] and [Bae+15]. With respect to [Men+17], the author contributed by designing the behavior of a transcriptional network using the analytic form of its stationary distribution. In [Bae+15], the author contributed by developing a framework for the design of stochastic behaviors of biochemical reaction networks. The author mathematically described design

specifications for the number of modes, location of modes, and convergence to stationarity of probability distributions. Then the author formulated and solved the stochastic design problem as an optimization problem constrained by the design specifications. The descriptions of the work contained in this chapter were written by the author.

2.1 Motivation

Biological behavior is commonly described using deterministic, nonlinear, continuous-time models [DM15; Van92]. Yet, the deterministic description of biochemical reaction network kinetics is not appropriate if the biochemical species are at low molecular numbers or if stochastic fluctuations are important in the time evolution of the system [Elo+02]. As such, biochemical reaction network kinetics inside living cells are better captured by discrete stochastic models since reactant molecules are often at low copy numbers and subject to random motion [Gil00; MA97; JH07]. Experimental evidence highlights stochastic effects in living cells by showing copy-number fluctuations in genetically identical cells and distinct cell fate decisions in populations of clonal cells [Elo+02; RO08; EE10; MRD07; Cho+08; ARM98].

To capture the observed discrete stochastic behaviors in biochemical systems, the network reactions can be modeled as a Markov jump process [Gil00]. The state of this process is a vector of the concentration of species in the reaction network. The state vector evolves in time with dynamics given by a forward Kolmogorov equation, known in the biochemical literature as the chemical master equation (CME). The distribution of states evolves in time according to an infinite-dimensional ordinary differential equation (ODE) specified by the CME. The coefficients in the ODE are determined by the rate constants and by the stoichiometry and propensity functions of the chemical reaction network. Analytical solutions to the CME are only available for a few examples of biochemical reaction networks (e.g. monomolecular reaction networks [JH07], first-order networks [GLO05], a genetic negative feedback system [GSN12], two- and three- stage models of gene expression [SS08], a transcriptional network in [GD12], deficiency zero chemical reaction networks [ACK10]); even fewer of these analytical solutions capture transient stochastic behaviors. Most commonly, no analytical solutions are available and Monte Carlo-based techniques are used to approximate the solutions [Gil07; MK06; GMK17]. Alternatively, analytical moment equations for the chemical master equation are sometimes used [Eng06].

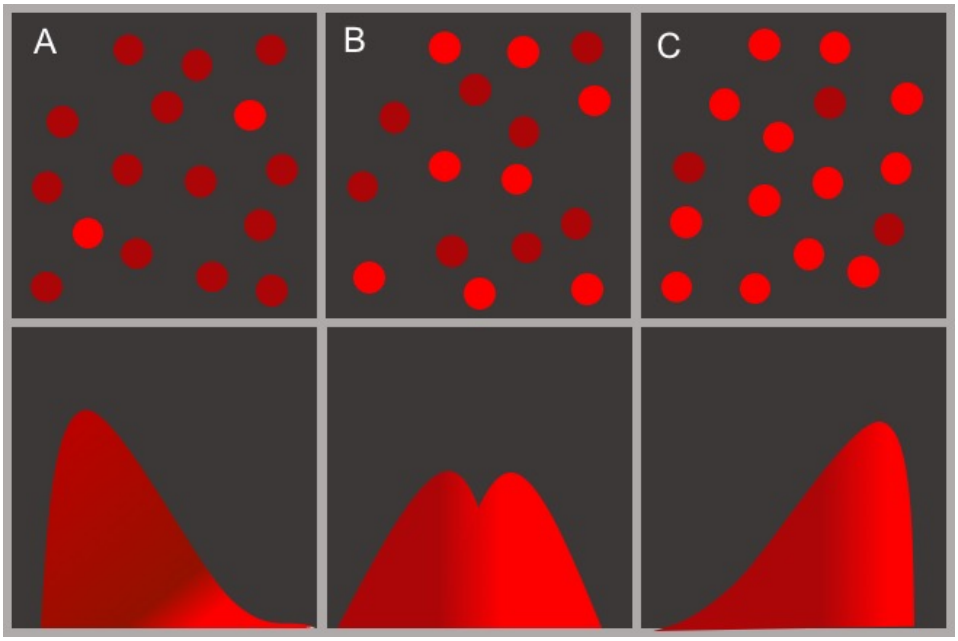

Figure 2.1: **Schematic representation of the fluorescence of an *E. coli* cell population that carries the genetic toggle switch.** The stochastic gene expression of red fluorescent reporter proteins in *E. coli* cells that carry the genetic toggle switch from [GCC00] creates a distribution of phenotypes. The reporter protein fluoresces red such as mCherry [Sha+04]. In panels A, the cell phenotypes form a unimodal distribution that peaks at a low (dark red) fluorescence level. A few cells have high (bright red) fluorescence levels and they represent the tail of the distribution of phenotypes. In panel B, the cell phenotypes form a bimodal distribution with peaks at both a low (dark red) and a high (bright red) fluorescence level. In panel C, the cell phenotypes form a unimodal distribution that peaks at a high (bright red) fluorescence level. Data measurements and discussion corresponding to this schematic are available in [GCC00; Por+07].

The shortage of analytical solutions to the CME model poses a challenge to the engineering design cycle in Chapter 1 for the field of synthetic biology. In this chapter, we aim to mitigate the lack of analytical solutions to the CME through our design framework, which does not make use of analytic solutions. We first design the stationary behavior of a transcriptional network for which we have computed an analytical solution to the CME model [GD12; Men+17]. Subsequently, we propose a framework for the design of stochastic behaviors of biochemical reaction networks, irrespective of the analytical solutions to their CME models.

In engineered systems, stochasticity is often an inconvenience and therefore, the engineering design goal is to minimize stochastic effects. Several examples of systems where stochastic effects are undesirable include semiconductors [Ian+94;

MA09] and power systems that use renewable energy sources such as wind or solar power [GA14]. Nonetheless, stochasticity can sometimes confer increased flexibility to uncertainty, as demonstrated by bet hedging strategies in bacterial persistence [VSK08; Bal+04] and in delayed seed germination in plants [Coh67]. Therefore, designing for stochastic behaviors is beneficial to some systems. Stochastic design specifications are encountered in applications where stochasticity is essential and stochastic models are ubiquitous, such as turbulent flow models for aerodynamics and for oceanography [GM13]. There, mathematical tools for stochastic PDEs inform the stochastic design constraints for turbulent flow models [Pop02].

The design features we propose for biological systems are inspired by unanswered questions in the design of genetic regulatory circuits. Our insight comes from the problem of designing the stochastic behavior of the genetic toggle switch in [GCC00]. Gene expression levels in cells of an *E. coli* population carrying the toggle switch form a distribution of phenotypes due to the heterogeneity of the cell population [Por+07], as illustrated in Figure 2.1. This distribution of cellular phenotypes follows either unimodal, bimodal, or trimodal stochastic transient and stationary behaviors [KRS15; Sch+10]. The phenotypic heterogeneity of the cell population is not typically designed for or specified. We can help control this heterogeneity by specifying the modality of the transient distributions: uni-/multi-modal, the protein expression levels, and the switching time. We mathematically formulate these design specifications similarly to [MG09] and we discuss how they result in remarkably different behaviors in a cell population.

Our design framework captures both stationary and transient distributional behaviors of biochemical networks such as uni-/multi-modality, the locations of the modes, and the rate of convergence to stationarity. These design features could not be captured by a deterministic framework; even the first moment, the single mode of a unimodal distribution, might be altered by stochastic effects [PBE00].

Even after selecting features relevant to the design of biochemical reaction networks, the stochastic design problem is difficult to solve. Our main challenge is that the exponential operator in the solution to the truncated CME from Chapter 1 has a dearth of exploitable mathematical properties [MV03] and a prohibitive computational cost. The exponential is not separable, which prevents us from leveraging a problem formulation in terms of relative entropy optimization as in [CS16]. We also considered its tensor projections as in [HG11; JH08; NHK12], but the orthogonal bases that we projected on were depleted of biological meaning; it was unclear how

to combine orthogonal basis polynomials in the space of projection so that they expressed the design features of uni-/multi-modality of distributions. This formulation would create overly elaborate problems that lose track of biological implementation. To avoid these issues, we simply consider the Taylor approximation to the exponential operator and we compute bounds on the error of this approximation.

If we use a first order Taylor approximation to the exponential operator, the design problem reduces to solving a linear program and a semi-definite optimization program [XB04; BV04]. There exist very efficient, scalable convex optimization tools, such as CVX [GBY08; GB08], that solve these programs. If the error of the first order Taylor approximation is large, we suggest using polynomial optimization methods as an alternative. Solving the design problem depends on the number of design features and the number of molecule counts of each biochemical species, particularly in the polynomially constrained case. Ultimately, we show that we can find accurate solutions for biochemical reaction networks with several species by using a first order Taylor approximation.

This chapter is organized as follows: In Section 2.2, we solve the design problem for a transcriptional network with a known analytical solution to the CME model. In Section 2.3, we set up the design problem and evaluate the error in the approximation of the exponential operator. In Section 2.4, we implement and solve design problems for classic examples of biochemical reaction networks: protein production-degradation, the Schlögl model, and the genetic toggle switch. Section 2.5 contains discussion of the applicability and limitations of our stochastic design framework, as well as an outline for future work.

2.2 The Design Of Stationary Stochastic Behaviors Of Biochemical Reaction Networks Using Analytical Solutions

In this section, we illustrate how the design of the stationary behavior of the transcriptional network in [GD12] is simplified by the availability of an analytical solution to its CME model. The analytical solution was first derived in [GD12] by applying the deficiency zero theorem from [ACK10] and then in [Men+17] by adapting the graph gluing technique proposed by Mélykúti *et al.* [MHK14; MP15]. No analytical transient solutions to the CME model of this transcriptional circuit are available, although an approximation was derived in [GD12] using singular perturbation theory [DM15].

Here, we use the analytical stationary solution derived in [GD12; Men+17] to de-

sign the equilibrium behavior of the transcriptional circuit simply by tuning the reaction rate parameters. This corresponds to altering the transcription factor's production or degradation rates, or its binding and unbinding rates to the downstream DNA. Tuning these rates can be implemented by choosing a transcription factor with the desired strength of binding and unbinding to DNA, by making more RNA polymerases or coactivators available to increase the transcription rate, by adding ubiquitin [DRH00], by tagging for phosphorylation [Gri+98] to increase the transcription factor's degradation rate, or even by changing the response element.

We first describe the two-component transcriptional system in Figure 2.2 [GD12]. The two transcriptional components are connected to each other through the action of the transcription factor Z. In the upstream component, Z is both produced and degraded. In the downstream component, transcription factor Z binds to DNA binding sites P and forms the complex C. The total amount of DNA, which is the sum of free binding sites P and of the complex C, is assumed to be conserved. We model the two-component transcriptional system stocastically using the chemical reactions in equation (2.1):

$$\begin{aligned} \mathrm{Z} &\underset{\delta}{\overset{k}{\rightleftharpoons}} \emptyset, \\ \mathrm{Z} + \mathrm{P} &\underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{C}. \end{aligned} \tag{2.1}$$

We let c be the number of molecules of the complex C and z be the number of molecules of transcription factor Z.

In [Men+17], we derived the analytical solution to the joint stationary distribution of transcription factors C and Z. The analytic solution of the stationary distribution was derived by recursively reconstructing its FSP-truncated finite state space from simpler shapes. First, simple "T" shapes were glued together sequentially at one vertex and then the proportionality condition in [MP15] was checked for the missing edges. Thus, the analytical solution over the finite state space was computed by reconstructing the state space from simple "T" shapes. Subsequently, the state space was allowed to be infinite and the analytic solution of the stationary distribution was obtained in that limit. We refer the reader to the derivation in [Men+17] for more information and we restate the analytical solution of the joint stationary distribution

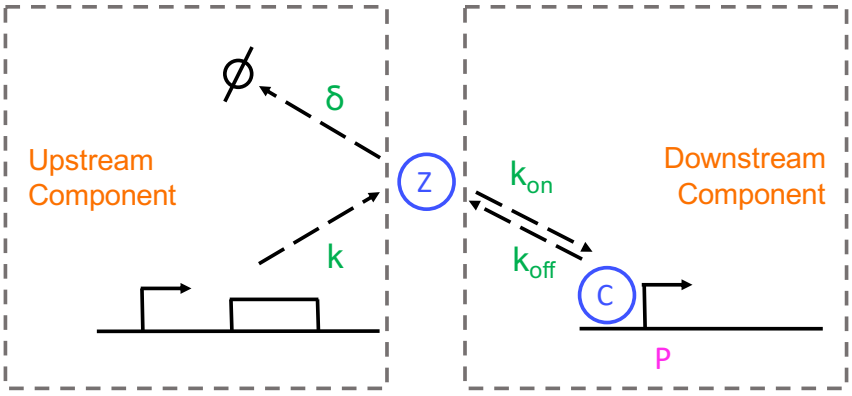

Figure 2.2: **A biological system of two interconnected transcriptional components [GD12].** In the upstream component, protein Z is transcribed from the DNA at rate k and degraded at rate δ. In the downstream component, protein Z acts as an activator that binds the DNA binding sites P and forms a complex C with it. Transcription factor Z and DNA sites P bind together with rate k_{on} and unbind with rate k_{off}. The total DNA is conserved, so the amount of DNA bound to transcription factor C and the amount of free DNA binding sites P is assumed to be constant.

of transcription factors in equation (2.2):

$$P(c,z) = \left(1 + \frac{\kappa\kappa_{\text{on}}}{\delta\kappa_{\text{off}}}\right)^{-N} \binom{N}{c} \left(\frac{\kappa\kappa_{\text{on}}}{\delta\kappa_{\text{off}}}\right)^{c} e^{-\kappa/\delta} \frac{(\kappa/\delta)^z}{z!} \tag{2.2}$$

for $c \in \{0,\dots,N\}$ and $z, N \in \mathbb{Z}_{\geq 0}$.

At stationarity, the upstream and downstream transcriptional systems are independent since the product of the stationary distributions of transcription factor Z and complex C equals their joint stationary distribution. This indicates that the system is not subjected to retroactivity at stationarity, meaning that the interconnection between the upstream and downstream modules does not slow down the dynamics of the upstream module [DM15; DDQ16]. Consequently, the expected value of their joint stationary distribution can be determined from the expected values of the transcription factor Z and of the complex C.

We formulate and find solutions for several design problems in Theorem 2.1, Proposition 2.2, and Corollary 2.3. We consider the following design features: the mean and the variance of the marginal probability distributions, and the location of the global maximum of the joint stationary distribution. For the latter, we prove that the stationary distribution has a unique global maximum if and only if the conditions in Proposition 2.2 are satisfied. Considering only stationary distributions with a unique

global maximum simplifies the subsequent design of its location, as illustrated in Figure 2.3. We present the proof of Theorem 2.1 in Appendix A.

Theorem 2.1. *Consider the system of two interconnected transcriptional components that are modeled by reactions as given in equation (2.2), where* $\kappa > 0$, $\delta > 0$, $\kappa_{on} > 0$, *and* $\kappa_{off} > 0$ *are the corresponding reaction rate constants. Let P, Z, and C be the numbers of promoters, transcription factors, and complexes, respectively. Let* $\alpha = \frac{\kappa\kappa_{on}}{\delta\kappa_{off}}$, $\beta = \frac{\kappa}{\delta}$, *and* $\gamma = \frac{N\alpha-1}{\alpha+1}$, *where N is a constant given by* $N = P + C$ *due to the conservation of DNA. In (i)–(iii), we set up and solve three design problems using the marginal stationary distributions of Z and C. Here,* α *and* β *are the design variables.*

(i) Since the marginal stationary distribution of Z is a Poisson distribution, its mean and variance are equal. The design problem of fixing the mean of Z at an objective value $\mu_z > 0$ *is feasible, and the solution is* $\beta = \mu_z$, *with N and the reaction rate constants being arbitrary otherwise.*

(ii) The design problem of setting the mean of C at an objective value $\mu_c \in (0, N)$ *is feasible, and the solution is* $\alpha = \frac{\mu_c}{N-\mu_c}$, *with N and the reaction rate constants being arbitrary otherwise.*

(iii) The design problem of choosing the variance of C to be an objective value $\sigma_c^2 > 0$ *is feasible if and only if* $\sigma_c^2 \leq \frac{N}{4}$, *and the solutions are* $\alpha = \frac{N-2\sigma_c^2 \pm \sqrt{N^2-4N\sigma_c^2}}{2\sigma_c^2}$, *with N and the reaction rate constants being arbitrary otherwise.*

We now consider designing the location of the global maximum of the joint stationary distribution in equation (2.2). In Proposition 2.2, we find necessary and sufficient conditions for the existence of a unique global maximum and provide its location. Corollary 2.3 shows how the uniqueness of the global maximum simplifies the design of its location. The notation $\lfloor\rfloor$ represents the integer part of a real number. The proofs of Proposition 2.2 and Corollary 2.3 are found in Appendix A.

Proposition 2.2. *Consider the system of two interconnected transcriptional components that are modeled by the reactions in equation (2.2). With the same notation as in Theorem 2.1, the stationary distribution in equation (2.2) has a unique global maximum if and only if* $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, *and* $\beta, \gamma \notin \mathbb{Z}$. *In this case, the maximum is at* $(c^*, z^*) = (\lfloor\gamma\rfloor + 1, \lfloor\beta\rfloor)$.

Corollary 2.3. *Under the constraints $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, and $\beta, \gamma \notin \mathbb{Z}$, designing the location of the unique global maximum of the two-component transcriptional system modeled by the reactions in equation (2.2) is equivalent to finding N, β, and γ such that $(\lfloor\gamma\rfloor + 1, \lfloor\beta\rfloor)$ is the objective location.*

When the global maximum of the two-component transcriptional system exists and is unique, there are infinitely many parameter values that can lead to the objective location of the global maximum. Since any α, β, γ, and N that satisfy the conditions in Corollary 2.3 lead to the desired global maximum design, this implies relative insensitivity with respect to experimental implementation.

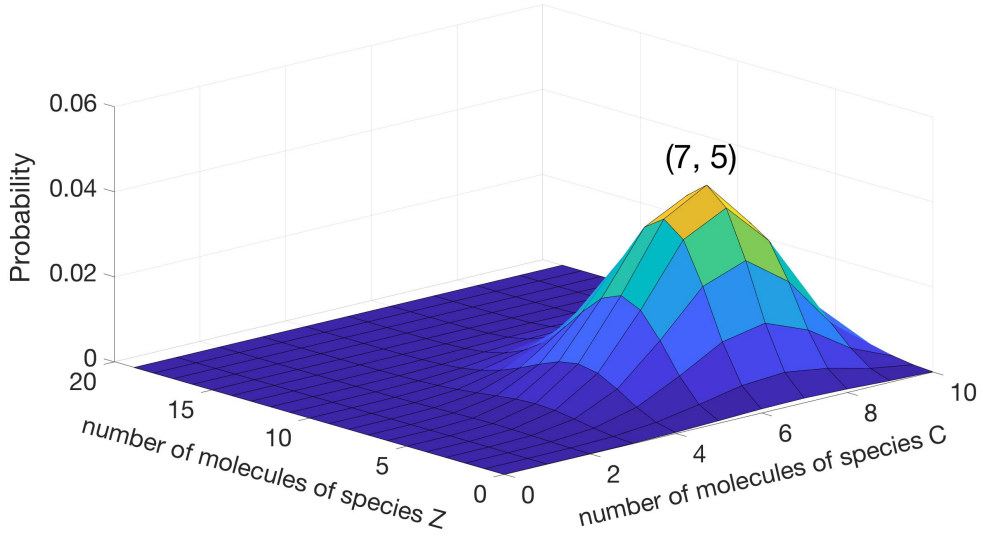

Figure 2.3: **Designing the global maximum of the joint stationary distribution of the complex species and the transcription factor species using its analytical form.** We consider the transcriptional circuit in [GD12] and we state the analytical form of the stationary distribution of the complex and the transcription factor species in equation (2.2). We then apply Proposition 2.2 and Corollary 2.3 to design the location of its global maximum. The parameter values we use are as follows: $N = 10$ molecules, $\alpha = 2$, $\beta = 5.5$, $\gamma = 6.33$. The joint probability distribution has a unique global maximum at 7 molecules of the complex C and 5 molecules of the transcription factor Z.

Theorem 2.1, Proposition 2.2, and Corollary 2.3 illustrate how the availability of analytical solutions to the CME greatly facilitates the design of stationary behaviors. Solving the four design problems in Theorem 2.1, Proposition 2.2, and Corollary 2.3 is equivalent to simply tuning the ratio of the production and decay rates of the transcription factor, the ratio of the binding and unbinding rates of the transcription

factor with promoters, and the total amount of DNA, as illustrated in the example in Figure 2.3. In Section 2.3, we consider designing similar features for distributions of reactant species in biochemical reaction networks when analytical solution to their CME models are unknown.

2.3 The Design Of Transient And Steady State Stochastic Behaviors Of Biochemical Reaction Networks

The design problem formulation

Our formulation of a stochastic design framework for biochemical reaction networks is a two-part contribution. We mathematically describe the desired transient and stationary behavior with design features and we find a solution for the design problem under these constraints. We then illustrate this stochastic design framework on several examples of biochemical reaction networks.

The design features

The design features we choose to constrain the stationary and transient probability distributions of biochemical species are as follows:

(i) uni- or multi-modality
(ii) the locations of the modes
(iii) the rate of convergence to stationarity

Our inclusion of design feature (i) is motivated by experimental evidence that demonstrates the presence of multi-modality in the genetic switching of the λ phage in the lactose operon [CLL10], in stochastic gene expression [SS08], and in cellular signal transduction pathways in mammalian cells [TB06]. The Gardner *et al.* genetic toggle switch [GCC00] is the first synthetic gene regulatory circuit to display multi-modality. An illustration of the genetic toggle switch's stochastic behavior is presented in Figure 2.1. Further discussion of the genetic toggle switch is available in Gardner *et al.* and Portle *et al.* [GCC00; Por+07]

Multi-modality is a purely stochastic behavior that cannot be reproduced or accounted for by deterministic modeling. Currently, there are no CME-based analytical tools to control its impact on the cell population heterogeneity. We aim for our mathematical formulation of multi-modality to fill this gap. Additionally, design features (ii) and (iii) control the expression levels across the cell population and the time to reach stationarity, respectively.

The design problem as an optimization program

We now consider the mathematical formulation of the design problem that incorporates the design features (i) - (iii). We assume a biochemical reaction network model with unknown reaction rate parameters subjected to these design constraints. We use the same notation for the biochemical reaction network and for the chemical master equation as in Chapter 1, but we replace the transition matrix $H(c)$ corresponding to the continuous CME with the matrix $D(c)$ that corresponds to the approximate discrete time dynamics:

$$D(c) = dtH(c) + \mathbb{I}_{M\times M} \tag{2.3}$$

In the design problem formulation we want to find the reaction rate vector $c = (c_1, \ldots, c_M)$ of the biochemical reaction network such that the transient (or stationary) probability distribution vector $p(x,t)$ is constrained according to our choice of design features at time points $t \in T = \{t_1, \ldots, t_k\}$, where $k \geq 1$ is the number of time points. Additionally, the dynamics of the probability distribution vector $p(x,t)$ are constrained by the CME. Lastly, we want to control the convergence rate of the transients to stationarity. These constraints and variables result in the following optimization program:

Find $c = (c_1, \ldots, c_M)$ such that

$$f_0 p_0 \leq \mu_0, f p^* \leq \mu_f, \tag{2.4}$$
$$f_i e^{H(c)t_i} p_0 \leq \mu_i, \tag{2.5}$$
$$(D(c) - p^* 1_M)^T (D(c) - p^* 1_M) \leq \mu^2 \mathbb{I}_{M\times M}, \tag{2.6}$$
$$D(c)p^* = p^*, \tag{2.7}$$
$$H(c) = \Sigma_{j=1}^{M} c_j H_j, \tag{2.8}$$
$$p_0 \in X_0, p^* \in X_f, p_{t_i} \in X_i, \tag{2.9}$$
$$X_0, X_i, X_f \subseteq P, \forall 1 \leq i \leq k, \tag{2.10}$$

where p_0 and p^* are the initial and stationary distributions, respectively; f_0, f_i, f are pre-selected projection operators that result in uni- or multi-modality of distributions; X_0, X_i, X_f are pre-selected subsets of P; μ, μ_i, μ_f are the tightness of the bounds, for all $1 \leq i \leq k$.

The inequalities in equations (2.4) and (2.5) impose design features (i) and (ii) at time points $\{t_1, \ldots, t_k\}$ under appropriate choices of operators. The initial, transient, and

stationary distributions are constrained by projection operators f_0, f_i, f, respectively for $1 \leq i \leq k$. An example of operator that imposes unimodality and the location m of the mode is the function $g : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$, $g(x) = (x-m)^2$ [MG09]. In Section 2.4, we give more examples of projection operator choices.

As shown in [XB04], the inequality in equation (2.6) uses the bound μ to tune the largest singular value of matrix $H(c)$. Thus, μ controls the rate of convergence to the stationary distribution. The inequality reduces to a semi-definite constraint by using the Schur complement formulation in [BV04].

The equality in equation (2.7) specifies that p^* is the stationary probability distribution vector of the Markov process, as explained in [XB04].

Remark 2. We clarify that design features (i) and (ii) apply to the marginal probability distributions of biochemical reactants in networks with more than just one species, $N > 1$. In order to marginalize the probability distributions, we multiply the operators f, f_0 and f_i, $1 \leq i \leq k$, by the appropriate marginalization matrices of sizes $M \times M^{N-1}$.

Remark 3. In practice, we choose to implement the design problem to minimize the linear objective function given by the sum of the bounds $\mu_0 + \mu_1 + \ldots + \mu_k + \mu_f + \mu$ with respect to the rate reaction rate vector $c = (c_1, \ldots, c_M)$ under the constraints in equations (2.4- 2.10). When the bounds $\mu_0, \mu_1, \ldots, \mu_k, \mu_f, \mu$ are pre-specified, the design problem reduces to finding a reaction rate vector c that satisfies equations (2.4- 2.10), if c exists. Thus, the optimization program simplifies to a feasibility problem.

Finding a solution to the design problem

Our main challenge in finding a solution to the design problem is the exponential operator present in equation (2.5). Our best approach has been to consider the Taylor approximation to the exponential operator and calculate the error of this approximation. Using the Taylor approximation of order $l \geq 1$ of the exponential operator, the inequality in equation (2.5) is replaced by

$$f_i \sum_{v=0}^{l} \frac{1}{v!} (H(c)t_i)^v p_0 \leq \mu_i, \forall 1 \leq i \leq k. \tag{2.11}$$

Subsequently, the design problem has linear constraints in equations (2.4) and (2.8), a semi-definite constraint in equation (2.6), and polynomial constraints in equations

(2.7) and (2.11). The problem is polynomial of degree $l+1$ in variables c, p_0, and p^* [BV04]. In Section 2.4, we find it useful to assume knowledge of p_0 and p^*, acquired either through experimental data or by computer simulations. This reduces the degree of the polynomial problem to l, eliminates equation (2.4), and makes equation (2.7) linear. The design problem in equations (2.4-2.11) is now a polynomial optimization program of degree l. In Section 2.4, we assume $l = 1$, so the equality in equation (2.11) is also linear. We then use CVX [GBY08; GB08] to solve the resulting optimization problems in Section 2.4.

Remark 4. There is a clear trade-off between choosing a larger truncation order l with the effect of decreasing the approximation error and keeping the degree of the polynomial inequalities in the design problem low.

2.4 Implementation of the Stochastic Design Framework

The protein production-degradation reaction network

We implement our design problem formulation to the gene regulatory network of protein production and degradation in [DM15]. Here, protein production and degradation are modeled stochastically as a birth-death Markov process. The chemical reaction network has only two reactions:

$$\mathrm{A} \underset{c_2}{\overset{c_1}{\rightleftharpoons}} \emptyset. \tag{2.12}$$

The reactions in equation (2.12) model the production and degradation of protein species A. The rates of the two reactions are c_1 and c_2. The birth occurs according to a Poison process with probability c_1 per unit time and the death occurs with probability per unit time proportional to $c_2 A(t)$. The analytic solution to its chemical master equation is $p(t) = \frac{c_2}{c_1}(1 - e^{-c_1 t})$ [GS01].

We choose the transient distribution to be unimodal with a single mode at 100 proteins using the operator $f(x) = (x - 100)^2$ and find the chemical reactions that satisfy this constraint. We assume that the stationary distribution is pre-determined by a Gaussian distribution with the same mean. The initial probability distribution is a Dirac delta function of height 1. Matrices H_1 and H_2 in equation (1.19) of Chapter 1 are the same as in [MG09].

Our stochastic design framework finds optimal reaction rates $c_1 = 3.9894$ and $c_2 = 0.0397$. The number of states in the FSP truncation is $S = 201$ and the convergence rate to the stationary distribution is $\mu = 0.1$. The approximation error

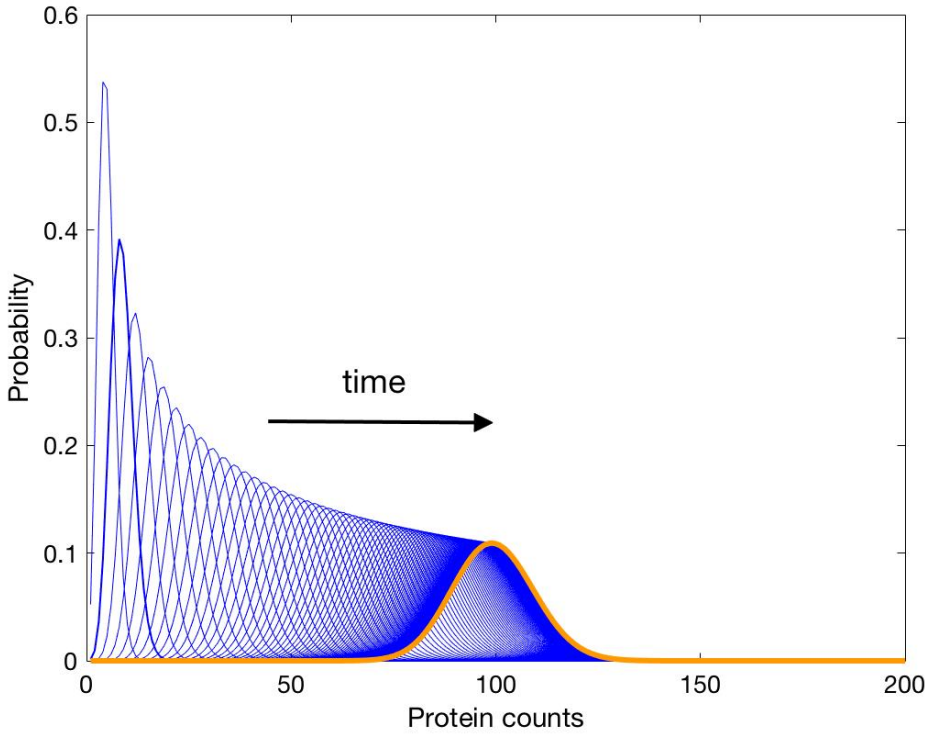

Figure 2.4: **Solution to the design problem for a protein production-degradation reaction network.** Time evolution of the unimodal transient distributions in the birth-death process in chemical reactions 2.12. The transient distribution after 100 seconds is displayed in orange and it has a single mode at 100 protein counts. The previous transient distributions are displayed in blue and they are indeed unimodal. The black arrow indicates the transients' progression towards stationarity as a function of time.

is $O(10^{-9})$. The results are illustrated in Figure 2.4. For a comparison of the last transient in Figure 2.4 and the stationary distribution, see Figure A.1 in Appendix A.

Remark 5. The solution to the optimization problem is not unique. The reaction rates c_1 and c_2 can take other values and they can certainly be adjusted by tuning the bounds $\mu_0, \mu_1, \ldots, \mu_k, \mu_f$ in equations (2.4-2.10).

The Schlögl chemical reaction network

The Schlögl chemical reaction network [Gil91] exhibits bistability in the deterministic model and bimodality in the CME model. See Gunawan *et al.* [Gun+05] for an in depth discussion of the Schlögl chemical reaction network. The set of chemical

reactions in the Schlögl network are as follows:

$$\begin{aligned} \mathrm{A} + 2\mathrm{X} &\underset{a_2}{\overset{a_1}{\rightleftharpoons}} 3\,\mathrm{X}, \\ \mathrm{B} &\underset{a_4}{\overset{a_3}{\rightleftharpoons}} \mathrm{X}. \end{aligned} \tag{2.13}$$

The concentrations of chemical species A and B are buffered and the four propensity functions are as follows:

$$\begin{aligned} a_1(X) &= k_1 A \frac{1}{2} X(X-1), \\ a_2(X) &= k_2 \frac{1}{6} X(X-1)(X-2), \\ a_3(X) &= k_3 B, \\ a_4(X) &= k_4 X, \end{aligned} \tag{2.14}$$

where k_1, k_2, k_3, and k_4 are the reaction rates.

We return to our previous notation by setting $c_1 = k_1 A$, $c_2 = k_2$, $c_3 = k_3 B$, and $c_4 = k_4$. The analysis of the deterministic model of the reaction network informs us that there exists a bifurcation with two equilibrium values of $s_1 = 84.79$ and $s_2 = 569.9$. We construct our projection operators to be centered around these values.

Using operator $f_{\text{unimodal}}(x) = (x - s_1)^2$, we impose unimodality on the transient distributions and we find optimal rate reaction values $c_1 = 1.0710 \times 10^{-5}$, $c_2 = 21.9939 \times 10^{-15}$, $c_3 = 0.3668$, and $c_4 = 0.0049$. We expect the reaction rate values to span several orders of magnitude [Gun+05]. The convergence rate in the application of our stochastic design framework is $\mu = 0.001$. This result is illustrated in Figure 2.5.

Alternatively, we can also impose a bimodal transient constraint as in [MG09] using projection operator

$$f_{\text{bimodal}}(x) = \begin{cases} \min((x - s_1)^2, 14920) & \text{if } x \geq 328 \\ \min((x - s_2)^2, 14920) & \text{otherwise.} \end{cases} \tag{2.15}$$

and, simultaneously, a unimodal stationary constraint $f^*(x) = (x - s_1)^2$. A plot of the bimodal projection operator is available in Figure A.2 in Appendix A.

The results of stochastic design framework are illustrated in Figure 2.6. We start from an initial distribution p_0 consisting of two Dirac delta functions with different weights and we move through a bimodal transient towards the unimodal steady state

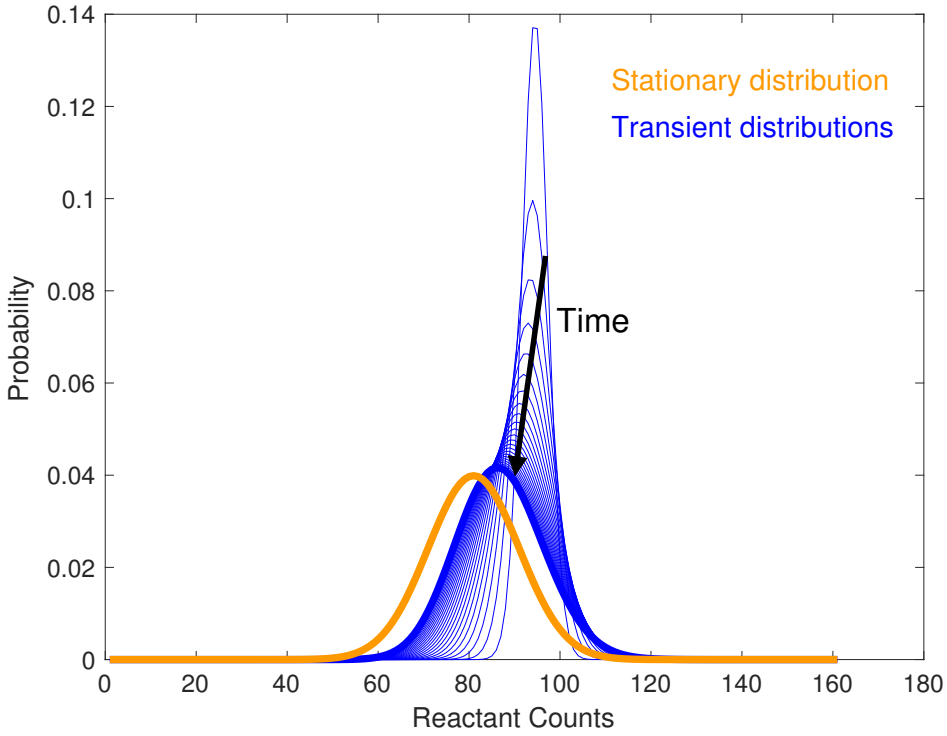

Figure 2.5: **Solution to the design problem for the Schlögl reaction network with unimodal transient constraints.** We plot the time evolution of the unimodal transients and compare it to the the stationary distribution. We find optimal rate reaction values $c_1 = 1.0710 \times 10^{-5}$, $c_2 = 21.9939 \times 10^{-15}$, $c_3 = 0.3668$, and $c_4 = 0.0049$ with a convergence rate to stationarity of $\mu = 0.001$. Not all transients are displayed.

distribution p^*. It is possible to find a solution to the problem almost irrespective of the placement and the heights of the Dirac delta functions. We demonstrate this in Figure 2.7 with a different unimodal stationary distribution choice of $f^*(x) = (x - s_2)^2$. It is also possible to define an initial distribution p_0 with Gaussian distributions replacing the Dirac delta functions. We can also replace the piecewise function with a sum of Gaussian distributions centered at s_1 and s_2. In all these cases, we are able to obtain solutions to the stochastic design problem.

Remark 6. When we impose a bimodal steady state distribution constraint, we can not find a satisfactory solution. The reason for this issue is that we implement equation (2.7) as the relaxation $\|D(c)p^* - p^*\| \leq \gamma$ for small γ. Hence, p^* is not forced to be an eigenvector of the matrix $D(c)$ and we do not guarantee that there are no other eigenvectors corresponding to eigenvalues closer to 0. In the biochemical reaction network examples in Figures 2.6, 2.7, and 2.9, the transient approaches

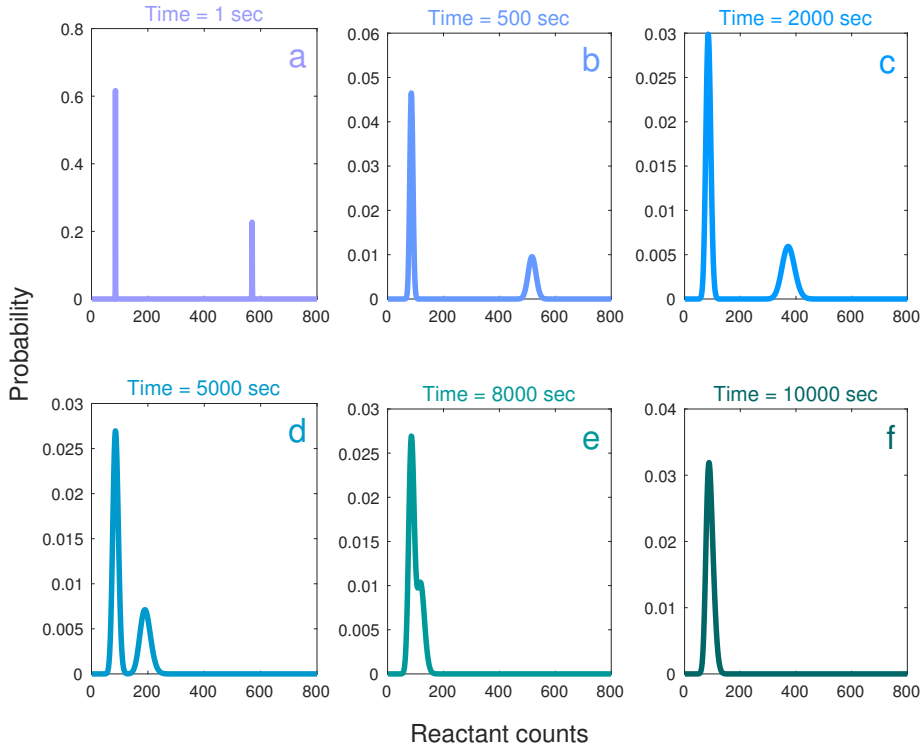

Figure 2.6: **Solution to the design problem for the Schlögl reaction network with bimodal transient constraints.** We plot the time evolution of the distributions. In part a, the initial distributions is pictured. We move through the bimodal transients in parts b-e. Part f has the stationary distribution. Not all transients are displayed.

the desired stationary distribution p^*, but it ultimately decays to a stationary distribution corresponding to the eigenvector with the largest eigenvalue. We choose not to implement equation (2.7) without the relaxation because the problem can be infeasible.

The genetic toggle switch

The genetic toggle switch is constructed from two promoters that mutually repress each other [GCC00]. The toggle switch flips between stable states using transient chemical or thermal induction. The bimodality property of the transients arises from the mutual inhibition of the two repressor genes, according to the diagram in Figure 2.8. Let A be the number of molecules of repressor A, B be the number of molecules of repressor B, I be the number of molecules of inducer I, according to the diagram in Figure 2.8. The mathematical model for the switch is adapted from the model in [GCC00] to consider only one inducer and to include the basal expression

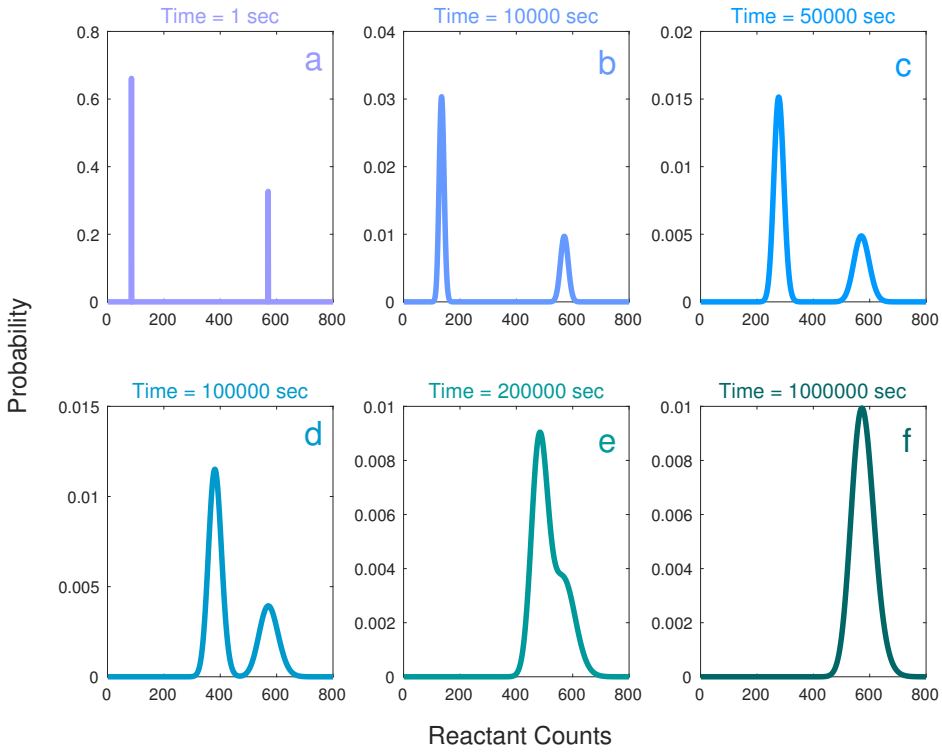

Figure 2.7: **Solution to the design problem for the Schlögl reaction network with bimodal transient constraints** We plot the time evolution of the distributions. In part a, the initial distributions is pictured. We move through the bimodal transients in parts b-e. Part f has the stationary distribution. Not all transients are displayed.

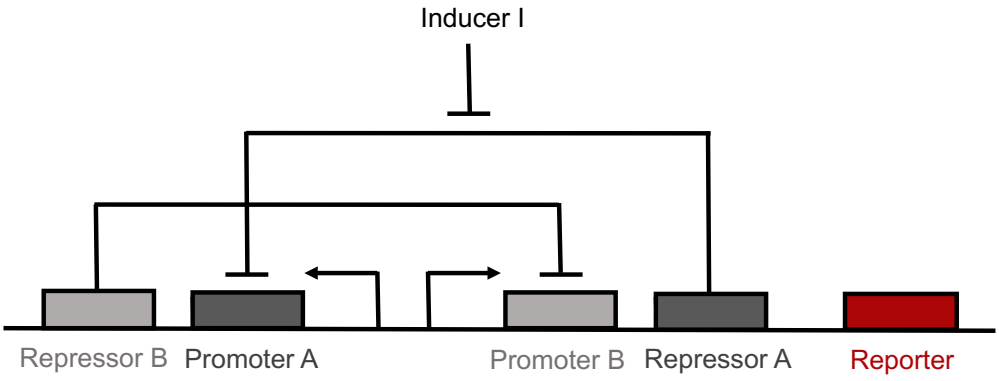

Figure 2.8: **The biological circuit diagram for the genetic toggle switch adapted from [GCC00].** Repressor A inhibits transcription from promoter A. Repressor B inhibits transcription from promoter B. Additionally, repressor A is inhibited by induction with inducer I. The reporter gene fluoresces red.

rates of the two repressor proteins. The propensity functions are as follows:

$$\begin{aligned} a_1(A,B) &= A_0 + c_1/(1+B)^{\beta}, \\ a_2(A,B) &= c_2 A, \\ a_3(A,B) &= B_0 + c_3/(1+(A/(1+I/K)^{\eta})^{\gamma}, \\ a_4(A,B) &= c_4 B. \end{aligned} \tag{2.16}$$

In this model, c_1 and c_2 are the rates of transcription and degradation of repressor A, respectively. Similarly, c_3 and c_4 are the rates of transcription and degradation of repressor B. β and γ are the cooperativity of repression of promoter B and promoter A, respectively. We use the parameter values from [GCC00]: $\beta = 2.5, \gamma = 1, \eta = 2.0015, K = 2.9618 \times 10^{-5}, I = 3.25 \times 10^{-5}$, along with the basal levels of protein expression $A_0 = 5, B_0 = 0.1$.

We constrain the transients to be bimodal and the stationary distribution to be unimodal. The results of our stochastic design framework are illustrated in Figure 2.9 for repressor A.

Reducing the error bound

If the error bound of the approximation is deemed too large, we can use a larger order approximation of the exponential operator to adjust it. In this case, the design problem becomes a polynomial optimization problem of order equal to that of the new Taylor approximation. Polynomial optimization problems (POPs) are computationally NP-hard [Par03]; but, in practice, solutions can usually be found for problems of small to moderate size [Wak+05b; Wak+05a]. Using our formulation, we expect the polynomial optimization problems to be solvable for biochemical reaction networks with several species. Our ability to obtain a solution to the POP will also depend on the number of design features we specify and on the number of molecule counts allowed for each species.

2.5 Conclusion and Future Work

In this chapter, we introduce methods for designing the stochastic behavior of biochemical reaction networks modeled by the chemical master equation. To perform biological design, we find reaction rate parameters to match desired stochastic distributions. Biological design is an underdeveloped area of research, even though the problem of choosing parameters that match measured stochastic distributions, known as parameter identification, has been well studied (Chapter 4).

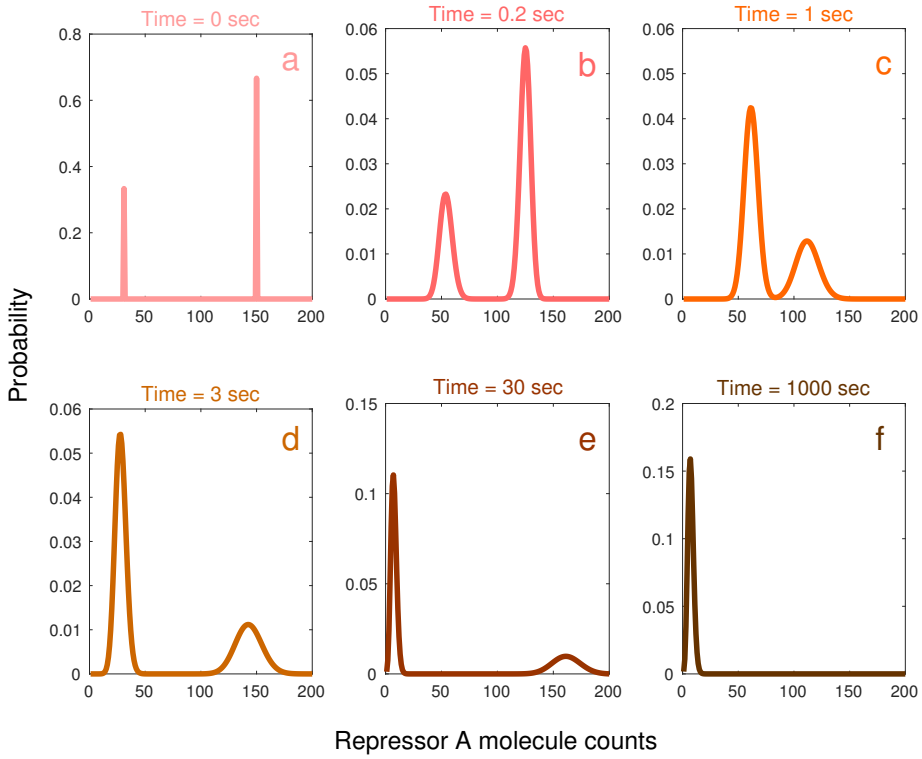

Figure 2.9: **Solution to the design problem for the genetic toggle switch with bimodal transient constraints.** We plot the time evolution of the distribution of repressor A. In part a, the initial distributions is pictured. We move through the bimodal transients in parts b-e. Part f has the stationary distribution. The rate reaction values we find are as follows: $c_1 = 150.7, c_2 = 0.97, c_3 = 1.57, c_4 = 0.12$. Not all transients are displayed.

First, we consider designing the stochastic behavior of a transcriptional network with a known analytical stationary solution to its chemical master equation model. We determine the reaction rates that resulted in the desired mean, variance, and global maximum of the joint stationary distribution of transcription factor and transcription factor bound to DNA. We demonstrate that these optimization problems are greatly simplified by the availability of an analytical solution to the chemical master equation model.

Additionally, we develop and implement a stochastic framework for the design of biochemical reaction networks when a solution to their chemical master equation may not be available. Our formulation of the stochastic design problem uses biologically meaningful design features to set up optimization problems and to find the optimal reaction rates of biochemical reaction networks. We illustrate this stochastic design framework with the two examples of the Schlögl chemical network and the

genetic toggle switch. We constrain the transient distributions of these two example biochemical networks to be either unimodal or bimodal and we determine the optimal reaction rates for their stochastic models.

Our stochastic design framework is particularly valuable for specifying the behaviors of biochemical reaction networks when analytical solutions to the chemical master equation model are unknown. This is often the case, though recently efficient approximations to stationary solutions [GMK17] and efficient algorithms for running the finite state projection algorithm have been proposed in [Vo17]. Nevertheless, transient analytical solutions to the chemical master equation model are almost never known and it has been difficult to even approximate them. Therefore, designing behaviors and finding solutions for specific transient stochastic behaviors of biochemical reaction networks is an open area of research.

The main limitation of our stochastic framework lies in the size of the problems we can accurately solve. A better approximation to the exponential operator might avoid the "curse of dimensionality", but none that we considered were viable. Hence, the polynomial optimization portion of the design problem formulation can only be accurately solved for problems with several biochemical species. Yet, this might be sufficient to offer insight into the behavior of larger gene regulatory circuits, when combined with results in reducing multiscale stochastic models [Bal+06] or when using quasi-steady-state and quasi-equilibrium approximations [MHK14]. In particular, we hope to use our framework to design multiscale genetic circuits with partial knowledge of rate reaction values.

Beyond the research in this chapter our stochastic design framework could offer insight into what is even biologically possible to build. For example, we might want to know if it is possible to build a genetic switch with a transient distribution with four modes. In particular, when designing gene regulatory circuits, it is challenging to predict what transient behavior might arise, how long the transient would the last for, or if the stationary behavior will even follow our specifications. Using the design feature language we have developed, we can test for these questions. Future work will include applying our stochastic design framework to the class of genetic switches and testing out what is possible to build. When combined with forward simulation techniques, theoretical design work can be done by iterating between the two, similar to the design process followed in engineering problems.

Chapter 3

IMPLEMENTING BIOLOGICAL CONTROL WITH SEQUESTRATION FEEDBACK

As discussed in Chapter 1, developing biological controllers using synthetic biology can help address problems in human health, industrial fermentation, and waste recycling. Since controller design tools have already been developed for engineered systems, we have considered applying them to the design of biochemical controllers. Nonetheless, biochemical and engineered systems differ in several important ways, such as biochemical systems' limited ability to record only a positive error signal when tracking a reference. Therefore, the controller design tools developed for engineering systems are not directly transferable to biochemical systems. In this chapter, we consider a sequestration controller for biochemical systems that relies on the sequestration reaction of two species, as illustrated in Figure 3.1. The unbound amount of controller species captures the error between the current and the desired state of the system and it is always a positive quantity. We develop guidelines for the biological implementation of sequestration feedback networks using properties of the controller and of the process species.

Several examples of possible parts for the sequestration controller species include transcriptional parts such as a mRNA and antisense RNA pair; protein parts such a sigma and anti-sigma pair or a toxin and antitoxin pair. Depending on the choice of parts, the sequestration reaction strength and the controller species degradation rates can vary over several orders of magnitude. Briat *et al.* [BGK16] studied stochastic sequestration feedback systems with no controller species degradation and demonstrated that they can achieve perfect adaptation when they are stable. Nevertheless, we have found the assumption of zero controller species degradation to be too restrictive for a realistic implementation of sequestration feedback networks [Ang+10; Ren+17; QD18; Ols+17]. In this chapter, we demonstrate that the controller species sequestration strength, as well as the process and controller species degradation rates affect the stability and the performance (steady state error) of sequestration feedback networks. We then use this analysis to provide guidelines for the biological implementation of sequestration feedback networks.

First, we find that the stability margin of the sequestration feedback network with

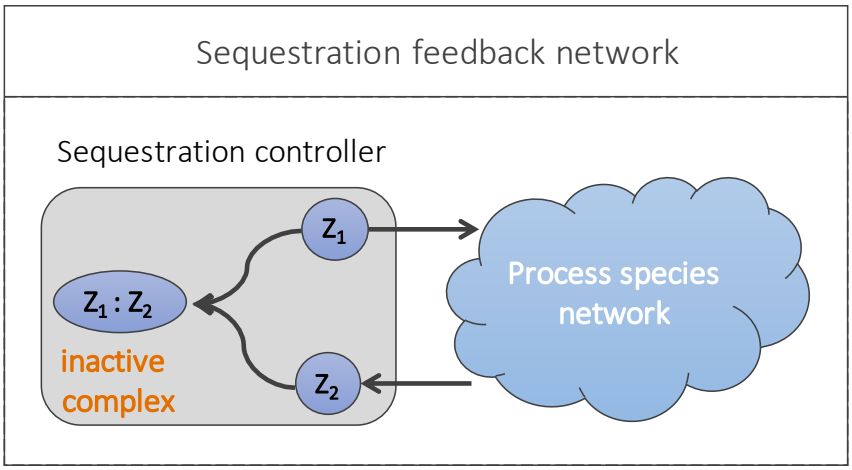

Figure 3.1: **Sequestration feedback network diagram.** The sequestration feedback network consists of a sequestration controller that controls the output of the process species network. The sequestration controller relies on the binding of two biochemical species Z_1 and Z_2 into the inactive complex $Z_1 : Z_2$. The rate of binding of species Z_1 and Z_2 is assumed to be much higher than their rate of unbinding. The sequestration controller and the process species network are connected by interactions that are further described in Figure 3.3.

controller degradation can be improved by increasing either the process species or the controller species degradation rates. We derive an analytical criterion for the stability of the sequestration feedback network by comparing the values of the process and the controller degradation rates under the assumption of strong sequestration feedback.

Second, we consider how the process and the controller species degradation rates influence the steady state error of the sequestration feedback system. A stable sequestration feedback system with no controller species degradation implements perfect adaptation, which implies zero steady state error. Additionally, we find another optimal controller degradation rate that ensures zero steady state error.

Last, we uncover a tradeoff between the stability margin and the steady state error in sequestration feedback systems by allowing both the process and the controller degradation rates to vary. Increasing either the process or the controller species degradation rate improves stability, but this can result in a large steady state error. We also find that the steady state error is more sensitive to changes in the process species degradation rate than the controller species degradation rate. Hence, we have more options to tune the controller species without negatively impacting the sequestration feedback system's performance.

Using our understanding of steady state error and stability margin, we provide general guidelines for the implementation of sequestration feedback networks. We suggest simple tuning options to improve the stability of sequestration networks. When the process is already specified by experimental constraints, we explain how to choose a controller for good stability and performance. We also suggest robust parts for the implementation of sequestration feedback networks.

The results in this chapter were published in [Ols+17] and [Ren+17]. With respect to [Ols+17], the author contributed by considering the implementation of sequestration feedback networks. The author incorporated the degradation of the controller species in the model for the sequestration feedback networks and then determined their properties of stability and performance. The author also developed guidelines for the implementation of sequestration feedback networks. In [Ren+17], the author contributed to the mathematical analysis of cell population controllers implemented by sequestration feedback networks. The description of the work contained in this chapter was written by the author.

3.1 Motivation

Biological control with sequestration feedback

Negative feedback is a ubiquitous motif in endogenous biological systems; the negative autoregulation motif is present in 40% of genes in *E. coli* [REA02]. Similarly to endogenous systems, synthetic biological feedback controllers have also typically relied on an implementation that uses negative feedback loops to achieve homeostasis. Examples of synthetic negative feedback circuits include the cell growth regulator in [You+04], the two-gene transcriptional network for RNA production matching in [Fra+14], the stochastic feedback implemented with integrases in [Fol+17], the co-culturing of two populations of microbes with different growth rates in [Sco+17], and the optogenetics-based real time genetic compensator in [HME18].

Implementing synthetic biological controllers using negative feedback has been appealing due to the properties it confers. Negative feedback is known to reduce noise in gene expression and to improve response speed [DM15; Alo06]. The strength of the negative feedback is usually proportional to the deviation from the desired set point [Fra+14; You+04], which means that it behaves as a proportional controller [AM08]. This implies that negative feedback cannot guarantee adaptation to stimuli and that it confers limited, parameter-dependent disturbance rejection properties. To maintain homeostasis, multiple feedback loops are usually employed,

such as in the temperature control of the human body [Wer10]. Accordingly, synthetic controllers that use negative feedback must be carefully tuned to provide good performance.

It has been natural for synthetic biology to consider alternative feedback strategies that result in better controller performance. Incoherent feedforward loops often display the property of perfect adaptation [Goe+09; Ma+09], which means that the system returns to the desired set point with no steady state error after responding to a stimulus. The incoherent feedforward motif has also been identified in the regulation of endogenous systems [Man+06; Alo06]. Yet, incoherent feedfoward loops are challenging to engineer in synthetic biology because they rely on an engineered combinatorial promoter [DKM10] or on an enzyme in the saturation regime [Ma+09]. A synthetic incoherent feedback loop based on an activator and a repressor binding to the same promoter was engineered in [GHM14].

Sequestration feedback, as introduced in Figure 3.1, has been another potential implementation of perfect adaptation through integral control. Examples of synthetic circuits that use sequestration feedback include the concentration tracker in [Hsi+14], the two bacterial growth controllers in [McC+17; Ann+17], and the gene expression controller in [Ann+17]. An endogenous sequestration feedback system relies on the binding of sigma factor σ^{70} to anti-sigma factor Rsd [JI99]. Other biochemical reaction network designs that implement robust perfect adaptation through integral control are studied in [XD18] and a proportional integral derivative controller design is studied in [Che+18].

Another advantage of using sequestration feedback to implement synthetic control has been that it emulates the ability of electrical and mechanical controllers to perform two-sided subtraction. Biochemical systems differ from classical engineering systems in that they can only perform comparison and control for non-negative concentrations of biochemical species. Therefore, biochemical systems can not exactly implement two-sided subtraction (of either positive or negative sign) and thus can not compute the two-sided error signal often required for control action. In particular, negative feedback in biological systems often only performs one-sided subtraction because it is active only once a quantity is above the set point. One approach to approximating two-sided subtraction with biochemistry is to use sequestration feedback to perform two one-sided subtractions, as illustrated in Figure 3.1. If there are more molecules of Z_1 than Z_2, then the sequestration feedback encodes the quantity $Z_1 - Z_2$, whereas if there are more molecules of Z_2 than Z_1, then it encodes the

quantity $Z_2 - Z_1$. To build controllers that can encode a two-sided error signal, two-sided subtraction must be approximated for biological control.

Chapter contribution

Previous literature has suggested implementing sequestration feedback networks with no controller species degradation to ensure that they achieve perfect adaptation [BGK16]. Two advantages of this strategy are good performance (zero steady state error if the closed loop system is stable) and a flexible implementation of the process network. However, it has become clear that the controller species are degraded [Ang+10; Ren+17; QMD17] and that the stability of the sequestration feedback network depends on both the process and the controller parameters, as we explain in Section 3.4. We suggest another possible implementation of sequestration feedback networks that accounts for the controller species' degradation and ensures zero steady state error (if the closed loop system is stable) in Section 3.5. Nevertheless, this implementation depends on an exact relationship between the controller and the process network parameters, which renders it inflexible and impractical.

Perfect adaptation (integral control) is a property that can be found in endogenous biological and engineering systems [Fer16; Goe+09; EGK02; Yi+00; AM08]. Nonetheless, current synthetic biological controllers may or may not be able to achieve perfect adaptation in a practical, flexible implementation. In Sections 3.6 and 3.7, we relax the requirement that the sequestration controller achieve perfect adaptation and we simulate a set of sequestration controllers with good performance, large stability margin, and good disturbance rejection properties. Furthermore, these controllers are advantageous because they do not require a precise implementation of their biological parameters.

In the previous literature for sequestration feedback networks, both the process and the controller networks have been simultaneously built [McC+17; Hsi+14; Fol+17; Lil+17]. Consequently, sequestration controllers have been built and optimized for a single process network, which could limit their generality for future use; a class of optogenetics-based real time controllers could be a more versatile option [HME18]. Moreover, these process networks have occasionally been complicated to ensure good performance of the sequestration feedback [Lil+17; McC+17]. This approach differs from the work flow of engineering systems [AM08], where the process network is often assumed to be fixed and unchangeable and the controller is built to stabilize the process and to achieve good performance. Thus, we also consider

the problem of building a sequestration controller for a fixed process network in Section 3.7.

This chapter is organized as follows: In Section 3.2, we introduce implementation considerations for the process and the controller species such as the strength of the controller species' sequestration reaction, as well as the degradation of both the process and the controller species. In Section 3.3, we introduce a model of the sequestration feedback controller and we use it to derive a stability criterion for the sequestration feedback networks based on comparing the process and the controller degradation rates in Section 3.4. In Section 3.5, we discuss the performance of sequestration feedback networks based on the value of the controller species degradation rate. We ascertain a tradeoff between the stability and the performance of sequestration feedback networks in Section 3.6. In Section 3.7, we give guidelines for the design of sequestration feedback networks. Finally, we conclude and propose future research directions in Section 3.8.

3.2 Implementation Considerations for Sequestration Feedback Networks

Sequestration feedback networks can be built using a variety of biological parts for the two sequestering controller species. Several examples of parts for the two controller species are illustrated in Figure 3.2. They include transcriptional parts such a mRNA and antisense RNA pair, sigma factors such a sigma and anti-sigma pair, or protein parts such as a toxin and antitoxin pair.

Transcriptional parts can be obtained from systems such as the hok-sok type I toxin-antitoxin system in *E. coli* or from parts already mined for synthetic biological systems. The hok gene product is a toxin that kills cells without its antidote, the antisense RNA sok that is complementary to the hok mRNA [Ger88]. A synthetic system of RNA and antisense RNA that originally performed translation initiation control is adapted to regulate transcriptional elongation in [Liu+12]. Protein parts can be obtained from type II toxin-antitoxin bacterial systems such as CcdA-CcdB. The toxin CcdB targets DNA gyrase and induces the breaking of DNA and subsequently cell death, while its antitoxin CcdA inhibits CcdB toxicity by sequestering it into a very stable CcdA-CcdB complex [De +09]. Protein parts can also be sigma and anti-sigma factors such as σ70 binding to bacteriophage T4 AsiA and inhibiting RNA polymerase, which slows transcription and inhibits *E. coli* growth [SC08].

Depending on the choice of parts for the controller species in sequestration feedback networks, the binding affinity of the sequestration reaction can vary over several

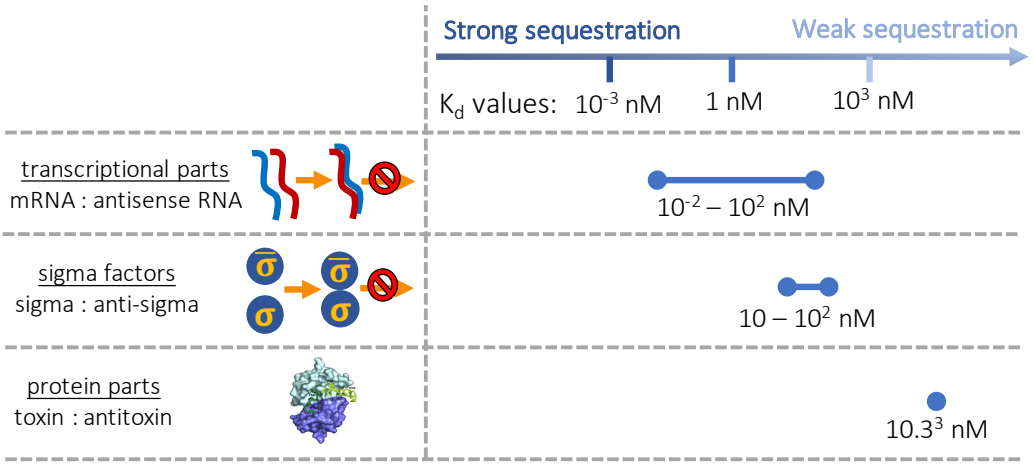

Figure 3.2: **The strength of the sequestration reaction.** The controller sequestration reaction can be implemented with a multitude of biological parts. Example transcriptional parts are mRNA and antisense RNA. Antisense RNA inhibits the translation of complementary mRNA by base pairing to it and physically obstructing the translation machinery of the cell. Anti-sigma factors bind sigma factors to inhibit transcriptional activity. Protein parts include the toxin-antitoxin module CcdA-CcdB in *E. coli*. When CcdB outlives CcdA, it kills the cell by poisoning DNA gyrase. The antitoxin CcdA blocks the activity of the toxin CcdB by binding together into a complex, thus allowing cells to grow normally. We include representative ranges of the binding constants (K_d values) for the transcriptional parts [Wal+02], the sigma factors [SC08], and the toxin parts [De +09]. For small K_d values, the association rate of the two molecules is high and the sequestration reaction is considered strong. In contrast, for large K_d values, the sequestration reaction is considered weak. This figure includes a partial reproduction of a figure in [De +09] with permission from the author.

orders of magnitude, as illustrated in Figure 3.2 and Table B.5. Yet, the binding affinity of the sequestration reaction influences the stability of the sequestration feedback network, as discussed in Section 3.4. Accordingly, we need to carefully consider whether our implementation choice for the controller parts results in stable closed loop control.

Additionally, depending on the choice of implementation of the controller species, their degradation rates can also vary, as illustrated in Table 3.1. Typically, the half-life of proteins inside the cell is long since they are slowly degraded, at a rate between 1 hr^{-1} and 2 hr^{-1}. However, the half-life of mRNA inside the cell is very brief (Table B.4), so we must include this degradation rate in our model. The degradation rate of the controller species influences the stability of the sequestration feedback network, as we discuss in Section 3.4.

mRNA	Hok	RpoS	CcdA
6.9 hr^{-1} - 20.8 hr^{-1}	2 hr^{-1}	1.39 hr^{-1}	1.39 hr^{-1}

Table 3.1: **The degradation rates of biological parts that could be used to build sequestration controllers.** We compute the degradation rates of mRNA, toxin Hok, sigma factor RpoS, and antitoxin CcdA. The degradation rate of mRNA is between 6.9 hr^{-1} and 20.8 hr^{-1} [Ber+04]. The degradation rate of toxin Hok in the type I toxin-antitoxin pair in *E. coli* is 2 hr^{-1} [SM12]. The degradation rate of the sigma factor protein RpoS is 1.39 hr^{-1} when the *E. coli* cells are in stationary phase at $37°C$ or under stress conditions [ZG98]. The antitoxin CcdA is degraded in wild-type cells with a rate of 1.39 hr^{-1} in the absence of toxin CcdB and a rate of 0.69 hr^{-1} when bound in a complex with toxin CcdB [De +09]. These degradation rates were computed from the half-lives in Table B.4.

The controller species degradation rate is particularly important when analyzing the stability and the performance of cell population growth controllers. Cell population growth control can be implemented using sequestration feedback with two proteins or RNAs as the controller species [Ren+17; McC+17]. In these implementations, the controller species are subject to both degradation and dilution. In addition to the degradation rate, the dilution rate equals the growth rate of the cells and thus can not be neglected in the controller's dynamics. Hence, cell population growth controllers lose the property of perfect adaptation due to the degradation and dilution of the controller species.

To recover the property of near-perfect adaptation in the presence of dilution of the controller species, Qian *et al.* propose a time-scale separation between the controller species reactions and their dilution [QD17]. In section 3.3, we introduce a model of sequestration feedback that includes the controller species degradation and dilution. We employ this model to perform a general analysis of the stability and the performance of sequestration feedback systems. Furthermore, we discuss how our biological implementation of the sequestration feedback network results in good stability and performance properties.

3.3 Modeling Sequestration Feedback Networks with Controller Species Degradation

In Figure 3.3, we introduce a mathematical model of a sequestration feedback network with nonzero controller species degradation and dilution rates. We only include the controller species equations in this section. For the full model, we refer

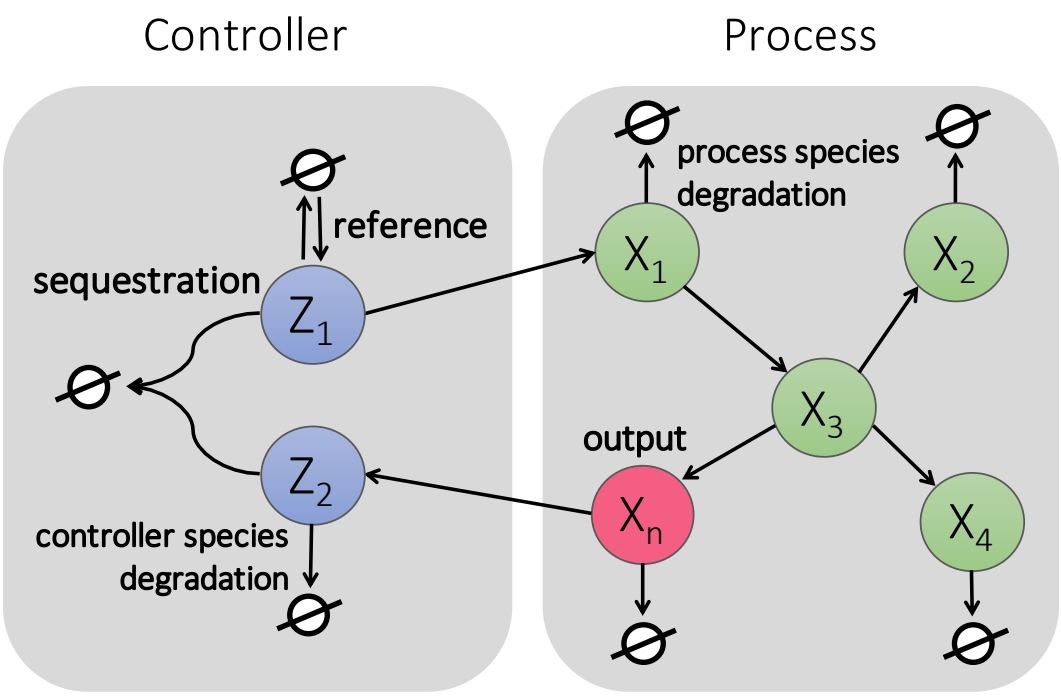

Figure 3.3: **The sequestration feedback network's controller and process networks.** The goal of the sequestration controller is for the process output species X_n to track the reference, as explained in Figure 1.5. The controller species Z_1 and Z_2 bind together in a sequestration reaction to form an inactive complex that is represented by an empty set. The reference acts on the sequestration feedback network as the constitutive production of controller species Z_1. We assume that the process species $X_1, \ldots, X_n$ are reactants in bimolecular chemical reactions. They interact with each other, as well as with the controller species. The controller species Z_1 acts on the process input species X_1. The process output species X_n acts on the controller species Z_2. The controller and process species are subjected to degradation and dilution, which is indicated by arrows pointing to empty sets. This diagram extends the setup in [BGK16].

the reader to Appendix B.

The two controller species Z_1 and Z_2 sequester each other into an inactive complex at rate η. The reference signal comes in through controller species Z_1's constitutive production at rate μ. Controller species Z_1 actuates the process network input species X_1 at rate θ_1, while the process network output species X_n acts on the second controller species Z_2 (the sensing species) at rate θ_2. For simplicity, we assume that the controller species are degraded at the same rate γ_c and that the process species are also degraded at the same rate γ_p. Under these assumptions and notation, the chemical reactions that describe the sequestration feedback network

are as follows:

$$\emptyset \xrightarrow{\mu} Z_1, \qquad Z_1 \xrightarrow{\theta_1} Z_1 + X_1, \qquad X_n \xrightarrow{\theta_2} X_n + Z_2,$$
$$X_i \xrightarrow{\gamma_p} \emptyset, \quad \forall 1 \leq i \leq n, \qquad Z_1 \xrightarrow{\gamma_c} \emptyset, \qquad Z_2 \xrightarrow{\gamma_c} \emptyset,$$
$$Z_1 + Z_2 \xrightarrow{\eta} \emptyset.$$

We let t denote time and $x_1, \ldots, x_n$ denote the concentrations of the process species $X_1, \ldots, X_n$. We let z_1 and z_2 denote the concentrations of the controller species Z_1 and Z_2, respectively. Thus, the sequestration controller can be modeled as follows:

$$\begin{aligned} \dot{z}_1 &= \mu - \eta z_1 z_2 - \gamma_c z_1, \\ \dot{z}_2 &= \theta_2 x_n - \eta z_1 z_2 - \gamma_c z_2. \end{aligned} \tag{3.1}$$

If we define the error signal as $e(t) = \frac{\mu}{\theta_2} - x_n(t)$, then the controller species have the following dynamics with respect to it:

$$\frac{d}{dt}\Big(z_1(t) - z_2(t)\Big) = \theta_2 e(t) - \gamma_c(z_1(t) - z_2(t)). \tag{3.2}$$

Therefore, the control action $z_1(t) - z_2(t)$ integrates the error signal $e(t)$ as follows:

$$z_1(t) - z_2(t) = \theta_2 \int_0^t e^{\gamma_c(s-t)} e(s) ds. \tag{3.3}$$

If the closed loop sequestration feedback network is stable, then equations (3.1) have a steady state and we can evaluate the magnitude of the steady state error. An in depth discussion of this topic is presented in Appendix B.

Assuming stability, when the controller species do not degrade, the sequestration feedback network exhibits the property of perfect adaptation since the error equals zero at steady state and since it is integrated by the controller. As explained in Chapter 1, perfect adaptation is a desirable property of biological systems because it guarantees zero steady state error and robustness to disturbances, irrespective of the process network (as long as the closed loop system remains stable). This allows for an imprecise implementation of the process network, although the zero controller degradation rate requires a very precise implementation of the sequestration controller.

With a nonzero degradation of the controller species, the sequestration controller is a lag compensator that integrates the error signal [AM08; Ren+17]. When the

controller species degradation rate is small, this sequestration controller approaches integral control, whereas when the controller species degradation rate is large, it approaches proportional control. It is not immediately apparent whether this controller retains the properties of zero steady state error or perfect adaptation. In Section 3.5, we show that while integral control guarantees zero steady state error under a variety of process network implementations, the performance of the lag compensator depends on the parameters of the sequestration feedback network. Consequently, we analyze the performance of sequestration feedback networks in Section 3.6. Furthermore, the stability criterion changes between sequestration feedback networks with zero and nonzero controller species degradation rate, as we discuss in Section 3.4.

Since both the process and the controller networks can only be imprecisely built with biological parts, we consider how this affects the properties of stability and performance of sequestration feedback networks in Section 3.6. In particular, we replace the stringent performance requirement of zero steady state error with a more flexible and practical performance requirement of small steady state error [AM08]. In Section 3.7, we formulate guidelines for implementing sequestration feedback networks with large stability margin and small steady state error.

3.4 Stability Analysis of Sequestration Feedback Networks

In this section, we derive an analytic criterion for the stability of sequestration feedback networks with nonzero controller species degradation. For simplicity, we assume strong sequestration binding of the controller species. We refer to this condition as the "strong sequestration feedback" regime, as defined in [Ols+17], and we restate it mathematically in the stability criterion.

First, we motivate the stability analysis by assessing the effect of varying the process species degradation rate, while keeping the controller species degradation rate constant (Figure 3.4). As we decrease the value of either the process species degradation rate (or the controller species degradation rate), the sequestration feedback network's stability margin decreases, as evidenced by the oscillations in the output species. For a small value of the process degradation rate, the sequestration feedback network becomes unstable, as evidenced by the sustained oscillations in panel B of Figure 3.4, and the controller becomes unable to track the reference signal. Hence, depending on the values of the controller and the process species degradation rates, the sequestration feedback network can be either stable or unstable.

A second insight we derive from Figure 3.4 is applicable to the implementation of sequestration feedback networks; we must measure the process degradation rate before choosing a sequestration controller to ensure that the controller can stabilize it. Intuitively, the process and the controller should not operate at vastly different speeds under the assumption of strong sequestration binding of the controller species; the opposite is true for weak sequestration feedback [Ols+17]. A controller with either a higher value of the degradation rate or under a "weak sequestration feedback" [Ols+17] might be able to stabilize the output species in panel B of Figure 3.4, as discussed in Section 3.6.

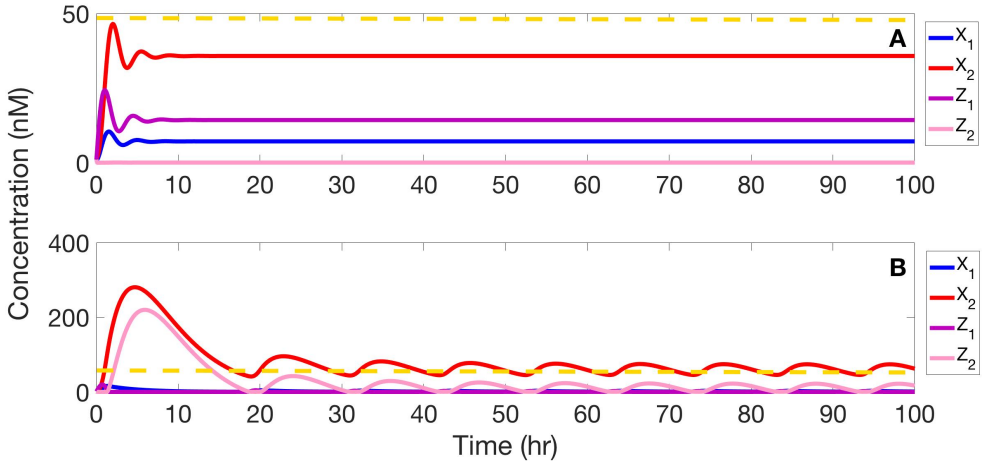

Figure 3.4: **The stability of the sequestration feedback network with varying process species degradation rate and strong sequestration feedback.** We illustrate the stability of sequestration feedback networks by assuming strong sequestration binding, indicative of a sigma factor implementation of the controller species. In panel A, we use a process degradation rate of 2 hr^{-1}, corresponding to a protein implementation of the process species. We note stability since the process output species X_2 (red) tracks the reference concentration of 50 nM (orange dashed line). In panel B, a process species degradation rate of only 0.25 hr^{-1} results in instability, as evidenced by the sustained oscillations in the process output species X_2. The parameters used for this simulation are given in Table B.7.

As the insights from Figure 3.4 suggest, the criterion for the stability of sequestration feedback networks with controller degradation is derived by comparing the process degradation rate and the controller degradation rate. The comparison between the process and the controller species degradation rates is also motivated by the ratio of the two degradation rates appearing as a term in the characteristic polynomial

associated with the stability of the linearized sequestration feedback network:

$$(s+1)^n\left(s+\frac{\gamma_c}{\gamma_p}\right)\left(s+\frac{\gamma_c}{\gamma_p}+\frac{\alpha+\frac{\beta}{\alpha}}{\gamma_p}\right)=-\frac{\beta}{\gamma_p^2}, \tag{3.4}$$

where $\alpha=\frac{\theta_1\theta_2\prod_{i=1}^{n-1}k_i}{\gamma_p^n}$ and $\beta=\eta\mu$ are constants and $s\in\mathbb{C}$ is the variable. The value of the ratio of the process and the controller degradation rates informs the location of the complex roots of the characteristic polynomial.

We formalize the stability criterion in Theorem 3.1 below and we include its derivation in Appendix B and in [Ols+17]. We consider three possible comparisons between the process degradation and the controller degradation rates: $\gamma_p \gg \gamma_c$, $\gamma_p \approx \gamma_c$, and $\gamma_p \ll \gamma_c$.

Theorem 3.1. *We consider the sequestration feedback network with controller degradation described in equations (3.1) under the assumption of strong sequestration feedback ($\beta \gg \alpha^2, \alpha\gamma_p$). The closed loop stability criterion depends on the relationship between the process degradation rate γ_p and the controller degradation rate γ_c. We consider the following three cases:*

Case I: *If the controller species degradation rate is much smaller than the process species degradation rate, the stability criterion is the same as the production-degradation inequality in [Ols+17], which we reproduce here as:*

$$\begin{aligned}&\gamma_p \gg \gamma_c,\\ &\gamma_p > \sqrt[n+1]{\frac{\theta_1\theta_2\prod_{i=1}^{n-1}k_i}{\tan\left(\frac{\pi}{2n}\right)\left(1+\tan\left(\frac{\pi}{2n}\right)^2\right)^{\frac{n}{2}}}}.\end{aligned} \tag{3.5}$$

Case II: *If the controller and process species degradation rates are approximately equal, then the closed loop system is stable if*

$$\begin{aligned}&\gamma_p \approx \gamma_c,\\ &\gamma_p > \frac{\sqrt[n+1]{\theta_1\theta_2\prod_{i=1}^{n-1}k_i}}{\sqrt{\tan\left(\frac{\pi}{n+1}\right)^2+1}}.\end{aligned} \tag{3.6}$$

Case III: *If the controller species degradation rate is much larger than the process species degradation rate, then the closed loop system is stable if*

$$\begin{aligned} &\gamma_c \gg \gamma_p, \\ &\gamma_p > \frac{\sqrt[n]{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}}{\sqrt[n]{\gamma_c}\sqrt{1 + \tan\left(\frac{\pi}{n}\right)^2}}. \end{aligned} \tag{3.7}$$

When the process degradation rate is much larger that the controller degradation rate, then the stability criterion is the same as the production-degradation inequality in [Ols+17]. Intuitively, the stability criterion requires the degradation of the process species to be faster than their production such that the process can respond effectively to the control action. When the process and the controller degradation rates are similar, the stability criterion only differs by a constant from the first case. Finally, when the process degradation rate is much smaller than the controller degradation rate, then the stability criterion relies on the controller degradation term to compensate for the slow process degradation rate. Since the process network is slow, the sequestration feedback network is challenging to stabilize and its performance can be very poor, as illustrated in Figure 3.6. In this case, stability may not be sufficient to ensure that the sequestration controller tracks the reference well. We study the performance of sequestration feedback networks in Sections 3.5 and 3.6.

3.5 The Performance of Sequestration Feedback Networks

When the controller species degradation is negligible, the sequestration feedback network exhibits the property of perfect adaptation, as discussed in Section 3.3 and in Appendix B. The controller implements integral control, which ensures that the steady state error equals zero, provided that the closed loop system of the sequestration feedback network is stable.

When the controller degradation rate is not negligible, we must include it in the model for the sequestration feedback network. In this case, the closed loop system has zero steady state error for zero controller species degradation, as well as for a unique value of the controller degradation rate, which we call the "critical" controller degradation rate. We illustrate the critical controller degradation rate in Figures 3.5 and B.1.

We illustrate the effect of the controller degradation rate for an example sequestration feedback network with two process species in Figure 3.5. When the controller

degradation rate has value zero, then the steady state error is also zero. Similarly, at the critical controller degradation rate, the steady state error also equals zero. Indeed, the degradation rate values of zero and of the critical degradation rate are the only ones for which the network tracks the reference with zero steady state error. We can use the theorems in this section to determine the conditions under which the critical degradation rate exists and to compute it.

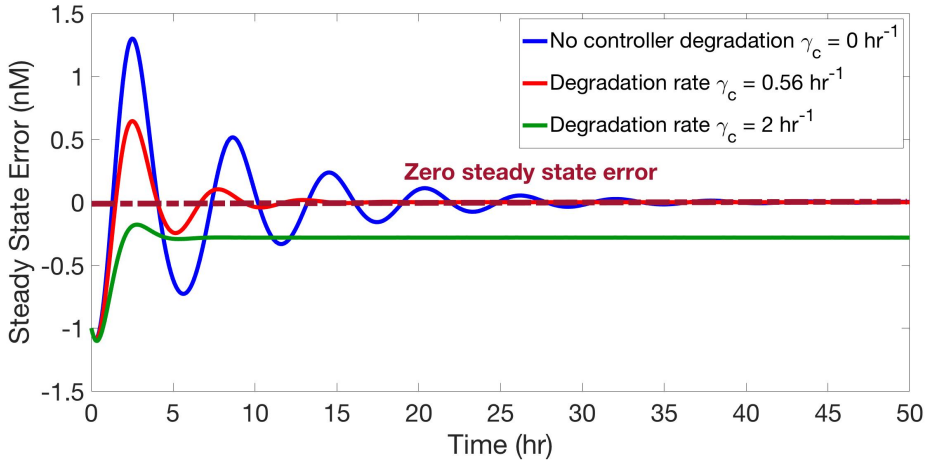

Figure 3.5: **The critical controller species degradation rate.** For a sequestration feedback network with only two process species, we vary the controller species degradation rate and measure the steady state error. With no controller species degradation (blue), the steady state error is zero (purple dashed line). When the controller species degradation rate is $\gamma_c = 2$ hr^{-1} (green), the steady state error has a value of 0.3 nM. However, at the critical controller species degradation rate, $\gamma_c = 0.56$ hr^{-1} (red), the steady state error has value zero. The parameters used for this simulation are given in Table B.1.

In Appendix B.3, we analytically derive the critical controller species degradation rate such that the steady state error of a sequestration feedback network with a general process network equals zero. We also derive conditions such that the critical controller degradation rate is achievable by the network parameters and we demonstrate that it is unique. Here, we only state the analytical result for a simplified process network.

Theorem 3.2. *The critical controller degradation rate for a simplified process network (i.e. each process species X_i is created by the previous process species X_{i-1}*

and creates the next process species X_{i+1}, $\forall 2 \leq i \leq n-1$, as in Figure B.2) is:

$$\gamma_c = \frac{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}{\gamma_p^n} - \frac{\eta \mu \gamma_p^n}{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}. \tag{3.8}$$

It can be achieved if and only if

$$\gamma_p < \sqrt[n]{\frac{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}{\sqrt{\eta \mu}}}. \tag{3.9}$$

The critical controller species degradation rate offers the possibility of implementing sequestration feedback networks that achieve zero steady state error even when the controller species degrade. Yet, as the mathematical expression in Theorem 3.2 suggests, the critical controller degradation rate is dependent on all the parameters of the sequestration feedback network. Therefore, the implementation of the controller could be sensitive and impractical. This motivates us to analyze the performance of an enlarged class of sequestration feedback networks with small, but nonzero steady state error.

3.6 The Tradeoff Between Stability and Performance

As discussed in Sections 3.4 and 3.5, the critical controller species degradation rate and the zero controller species degradation rate confer advantageous properties to sequestration feedback networks, but can be impractical to implement. In this section, we evaluate how sensitive the steady state error and the stability margin are to the process and to the controller degradation rates. We are interested in sequestration feedback network implementations with a small steady state error and a large stability margin.

First, we define the stability margin of a sequestration feedback network with two process species as:

$$S(\gamma_p, \gamma_c) = \frac{\gamma_p (\gamma_p + \gamma_c)^2}{\theta_1 k_1 \theta_2} \tag{3.10}$$

The stability criterion in [Ols+17] states that the sequestration feedback network is stable when $S \geq 0.5$. For simplicity, we assume that the process species production rates are equal: $\theta_1 = \theta_2 = k_1$. Therefore, we can define the normalized process and controller degradation rates as: $\gamma_p' = \frac{\gamma_p}{k_1}$ and $\gamma_c' = \frac{\gamma_c}{k_1}$, respectively. The normalized

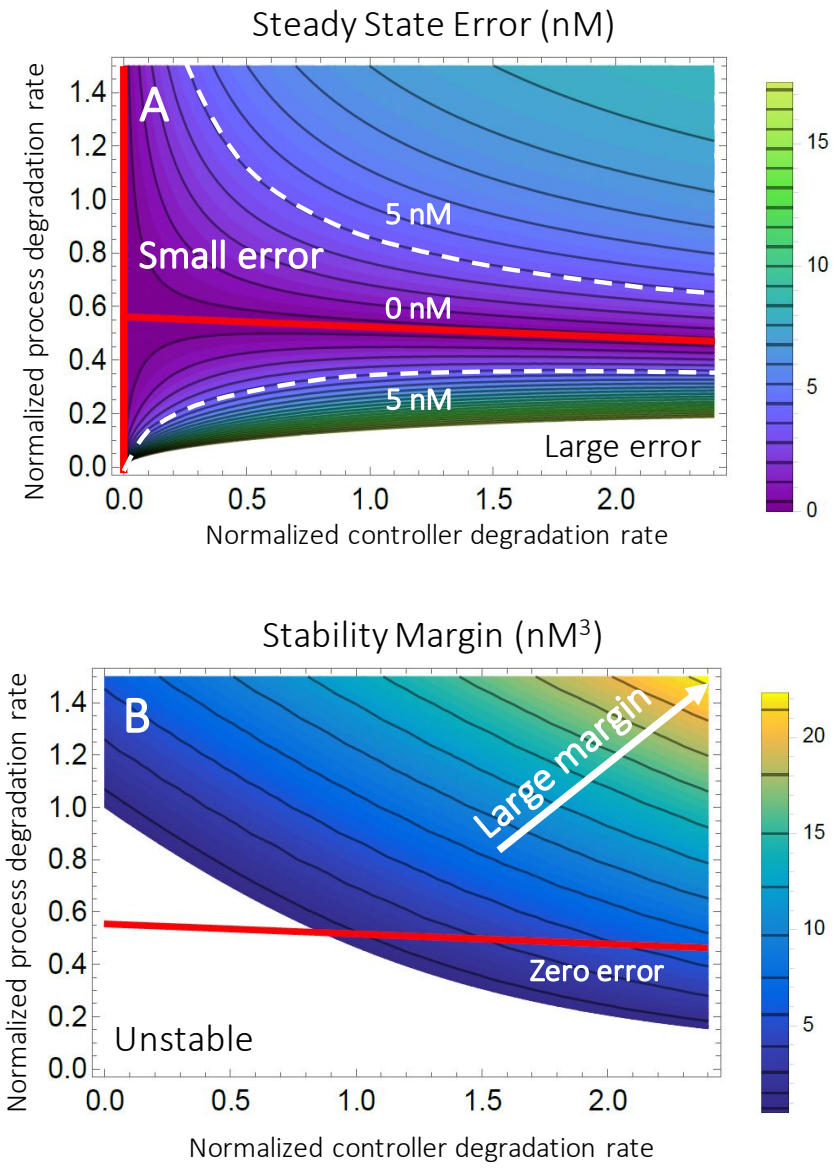

Figure 3.6: **The tradeoff between small steady state error and large stability margin.** We plot both the steady state error and the stability margin as functions of the normalized process species degradation rate and of the normalized controller species degradation rate for an example sequestration feedback network with two process species. A. The red lines indicate zero steady state error, corresponding to no controller degradation and to the critical controller degradation rate. We note that perturbing the process degradation rate by a small amount will deviate from the red line more than perturbing the controller degradation rate. In white, we draw contours for a steady state error of 5 nM. B. Increasing either the process degradation rate or the controller species degradation rate improves the stability margin. The white arrow in the top right corner indicates the increasing stability margin. The white region in the bottom left corner is unstable. This simulation was generated by Yoke Peng Leong.

process and controller rates, $\gamma_p^{'}$ and $\gamma_c^{'}$, are unitless, according to Table B.3. The steady state error can also be written in terms of the normalized process and controller degradation rates. We use the normalized process and controller degradation rates throughout Sections 3.6 and 3.7.

There is a tradeoff between achieving a large stability margin and a small steady state error in sequestration feedback networks, as illustrated in Figure 3.6. To obtain a small steady state error corresponding to good performance, the process and the controller degradation rates must be close to one of the two red lines that represent the zero controller species degradation rate and the (normalized) critical controller species degradation rate in Figure 3.6A. Simultaneously, to obtain a large stability margin, the process and the controller species degradation rates should be as large as possible, as in Figure 3.6B. This result matches the stability criterion in Appendix B.4; increasing either the process degradation rate or the controller degradation rate improves the stability margin. Since the two objectives cannot be simultaneously maximized, there is a tradeoff between achieving a large stability margin and a small steady state error in sequestration feedback networks.

Second, Figure 3.6A indicates that the steady state error of sequestration feedback networks is more sensitive to the process species degradation rate than to the controller species degradation rate. Accordingly, we have more freedom to tune the controller species degradation rate without negatively impacting the sequestration feedback network's performance.

In addition, we observe that slow process networks (small ratio between the process species' degradation and production rates) are difficult to stabilize for any controller, as indicated in Figure 3.6B. Increasing the controller degradation rate can improve stability, but can result in a large steady state error, as illustrated in Figure 3.6A. Hence, sequestration feedback networks with a small ratio between the process species' degradation and production rates are challenging to stabilize and their tracking performance can be very poor.

Last, we consider the problem of designing sequestration feedback controllers, as part of the engineering cycle introduced in Chapter 1. In the next section, we provide guidelines for the selection of a sequestration controller that achieves good stability and performance properties.

3.7 Implementation Guidelines for Sequestration Feedback Networks

In the previous sections of this chapter, we have determined the properties of sequestration feedback networks. Subsequently, we consider the design step of the engineering cycle and we determine sequestration feedback controllers for specified process networks. We use Figure 3.6 to provide guidelines for the implementation of sequestration feedback networks that have good stability and performance properties.

First, we note that there are process networks for which the tradeoff between steady state error and stability margin is severe. Hence, designing a sequestration controller that ensures stability and good performance may not be possible. We refer to the ratio between the process species' degradation and production rates as the "process speed". In the previous section, we have observed that process networks with a small process speed value such as 0.1 can only be stabilized by a sequestration controller with a high controller species degradation rate; however, even then, the resulting steady state error is very large. Therefore, sequestration feedback networks with small process speed are difficult to stabilize and their tracking performance is poor. Similarly, process networks with a large process speed value such as 1.4 require a small controller species degradation rate in order to keep the steady state error low, as illustrated in Figure 3.6A. However, this can result in a small stability margin. Thus, sequestration feedback networks with a large process speed value can be close to instability.

Subsequently, we consider a process network with a medium process speed value. Let us consider a process speed value of 0.6. This results in a small steady state error when the controller degradation rate is at most three times the process species production rate (Figure 3.7A). Accordingly, the design problem simplifies to finding a controller degradation rate that ensures stability. In Figure 3.7B, we observe that a controller degradation rate that is at least twice as fast as the process production rate results in a good stability margin. Depending on our implementation of the process species, achieving a controller degradation rate that is at least twice as fast as the process production rate could require a transcriptional implementation of the controller species (Table 3.1). Figure 3.7 suggests that a robust implementation of sequestration feedback networks uses a medium process speed since this allows for a range of controller species degradation rates. The corresponding controller species implementation can be determined using the degradation rates in Table 3.1.

Using Figure 3.7, we can first determine whether a specified process network is subject to a severe tradeoff between performance and stability. If the process speed

value is in the severe tradeoff region of Figure 3.7, building a sequestration controller for the specified process network might not be possible or might require a controller degradation rate outside the range of values given in Table 3.1. In this case, we might have to modify the process network to achieve a process speed that is subject to a less severe tradeoff between performance and stability. Additionally, Figure 3.7 and Table 3.1 can be used to decide the implementation of the sequestration controller when the process species' production and degradation rates have been measured. Depending on the ratio between the controller species degradation rate and the process species production rate, the implementation of the controller species is informed by the values in Table 3.1.

3.8 Conclusion and Future Work

The development of a first generation of synthetic biological controller for microbes is hopefully the beginning of an era when synthetic biology can help address new problems in human health, industrial fermentation, and waste recycling. In this chapter, we evaluate the stability and the performance of synthetic biological controllers implemented by a class of sequestration feedback networks. Using control theoretical methods, we develop conditions on the parameters of the sequestration feedback network that ensure stability and good performance. We tune the strength of the controller species' sequestration reaction, as well as the degradation of both the process and the controller species to obtain a large stability margin and a small steady state error. Moreover, we discuss practical considerations for the biological implementation of sequestration feedback networks and we suggest guidelines for a robust implementation.

The application of control theory to synthetic biological controllers aims to ensure that they function robustly, in different host organisms, despite perturbations in fluctuating environments. Nevertheless, almost all sequestration feedback controllers in the current synthetic biology literature have been concurrently built with the process networks they control [McC+17; Hsi+14; Fol+17; Lil+17]. This approach will likely limit the versatility of these biological controllers as they will be optimal to a single host organism (microbe or fungus) and to its environment; a potentially more general class of synthetic controllers are described in [HME18]. Therefore, we believe that it is important to continue assessing the constraints imposed by building sequestration feedback controllers for fixed process networks. By incorporating these constraints in our design of sequestration feedback controllers, we can guarantee that they function robustly, across a variety of biological host organisms.

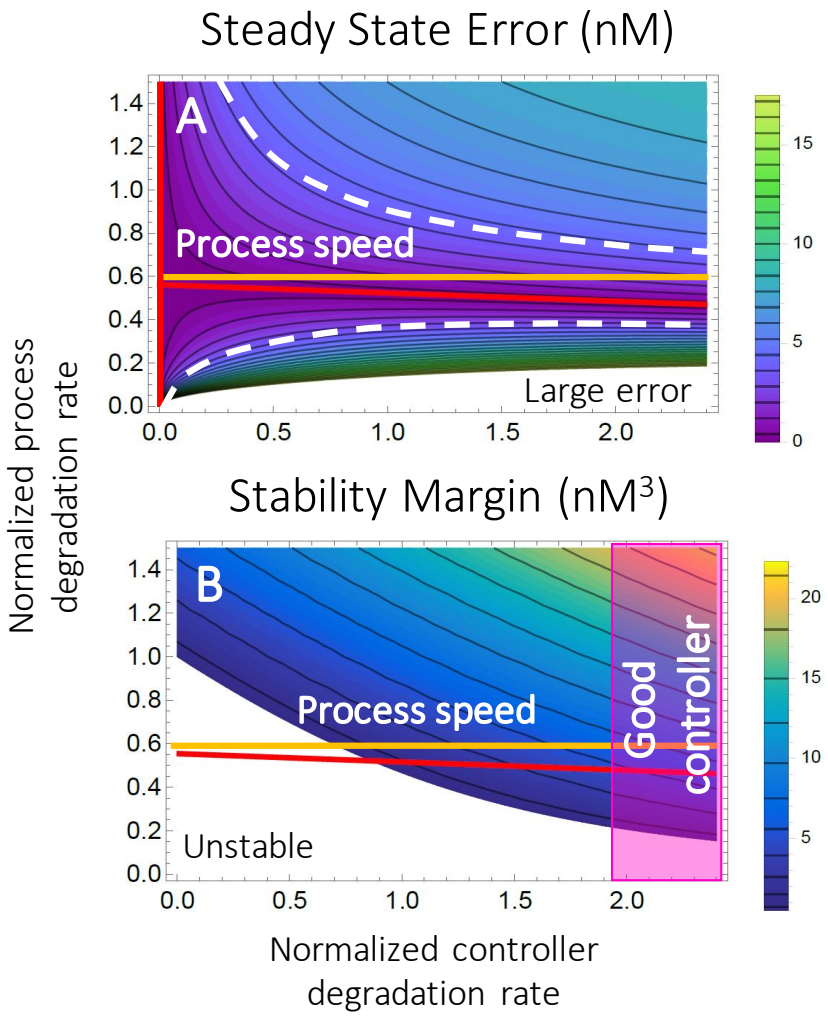

Figure 3.7: **Guidelines for designing sequestration controllers.** We design the controller of a sequestration feedback network for a specified process network. We first measure the process species' production and degradation rates and we render the process speed in the plots in Figure 3.6 (orange solid lines). In this example, we consider a process speed value of 0.6. This results in a small steady state error, as indicated in panel A. Accordingly, the design problem simplifies to finding a controller species degradation rate that ensures a good stability margin. In panel B, we observe that a controller species degradation rate that is at least twice the process species production rate results in a stability margin of at least 5 nM^3 (pink). We can match this controller species degradation rate with the values in Table 3.1 to inform our choice of controller species. We consider this to be a good controller implementation. If we want a higher stability margin value, we can increase the controller species degradation rate to match its value in panel B.

Moreover, we will concurrently expand to consider other mechanisms for biological control in addition to sequestration feedback. Several mechanisms for biological control that are currently being explored include: paradoxical extracellular signaling inspired by process regulation [Har+14] and post-translation mechanisms such as multi-protease regulation. Using control theoretical tools, we will develop models for these biological controllers and we will similarly assess their properties of stability and performance. Depending on the application of interest to synthetic biology, we will benefit from multiple mechanisms of feedback control to choose the most suitable implementation.

Chapter 4

COMPUTATIONAL METHODS

4.1 Introduction

The ability to model biochemical reaction networks has significant ramifications in the scientific fields of synthetic biology, molecular biology, and medicine. When biochemical species exist in large numbers, they can be treated as continuous quantities modeled by deterministic ordinary differential equations [Hig08]. In contrast, at the single cell level, biochemical species such as RNAs are produced at low copy numbers during gene expression, thus requiring a stochastic representation of their dynamics [Elo+02].

Probabilistic models such as the chemical master equation offer an accurate representation of stochastic biochemical networks, but they grow exponentially in complexity with the number of species considered [MNO12]. Therefore, it is challenging to study these stochastic, nonlinear models and to develop theoretical and computational methods for them. In particular, an important aspect of modeling stochastic biochemical reaction networks is the identification of parameter values that fit experimental data obtained by measuring the output species of that network. Most methods for parameter identification, such as the Metropolis-Hastings algorithm [Bro+11; Gel+14], require a very large number of solutions to the underlying model, one for each parameter evaluation. Hence, stochastic models of biochemical reaction networks are very computationally intensive to fit parameters for, even when we have data measurements of the networks they represent.

In this chapter, we demonstrate the benefits and the disadvantages of reducing the dimensionality of stochastic models of biochemical reaction networks prior to performing parameter identification for them. Our method proposes two sequential projections of these models; the finite state projection eliminates states with low probability and an additional interpolation using radial basis functions further simplifies the dynamics. We demonstrate that performing parameter identification on these two types of reduced models results in a small loss in parameter accuracy, but in a large gain in computational performance.

The results in this chapter were derived in collaboration with Zachary Fox, Huy Vo, and Brian Munsky (Colorado State University). The author contributed to the code

used by the projection methods. Additionally, the author performed parameter identification using the Metropolis-Hastings algorithm for the bursting gene example. The descriptions of the work contained in this chapter were written by the author based on a manuscript developed in collaboration with Huy Vo and Zachary Fox. For more details on the adaptive Metropolis-Hastings algorithm used for parameter identification, we refer the reader to [Bro+11; Bae+16].

4.2 Motivation

Biological processes in single cells can be modeled probabilistically using continuous-time, discrete-state Markov processes [MTK09; Neu+13; She+13]. Each state of the process is the integer vector with entries that are the number of molecules of all species. Finding the probability distribution over these states amounts to solving a first-order, linear, infinite system of ordinary differential equations known in biochemistry as the chemical master equation (CME [MG67; Gil92]). We have previously introduced the CME model in Chapter 1.

A critical task in understanding stochastic gene expression networks is the identification of parameters that fit experimental data. Parameter identification methods such as the Metropolis-Hastings algorithm [Has70; Bro+11] rely on maximizing the value of a likelihood function [MFN15]. For measurements obtained using RNA FISH [OJ16], the likelihood of a parameter combination given the data is computed by comparing the CME model predictions to the discrete data measurements. The evaluation of this likelihood function requires the solution of the CME for each parameter combination and a very large number of evaluations are required to fit parameters to the data.

Since analytical solutions to the CME are known only for special classes of biochemical reaction networks [JH07; DS66; GLO05; MHK14; Men+17], the CME has to be solved computationally. The most popular approach is to sample trajectories of the Markov process using Gillespie's stochastic simulation algorithm [Gil77] variants such as τ-leaping [CGP07], or to solve a continuous approximations to the CME instead; two such approximations are the chemical Langevin equation and the Fokker-Planck equation [Gil00]. These continuous state alternatives to solving the CME can sometimes be solved analytically or more easily computed, but they lose the advantage of directly matching discrete data.

The finite state projection (FSP) [MK06; MK07] approach computes the discrete probability distribution directly from the CME. The principle of FSP is to retain

only states with significant probabilities and to discard the rest of the state space, thus effectively truncating the CME into a finite problem. Nonetheless, this finite state problem is still too large to solve efficiently for a general biochemical reaction network. This hinders the effective use of the FSP for the parameter identification of these stochastic models.

In this chapter, we further reduce the computational cost of the FSP approximation to the CME model by interpolating the FSP approximation using a small number of radial basis functions. While similar ideas have been previously investigated in [MK08; JU10; TFM12; KRS15], our work differs in that the interpolation bases are derived from simulated single-cell data. This projection method is closely related to proper orthogonal decomposition (POD) [Ker+05], a popular method in other computational science and engineering fields. Yet, directly applying the POD projection to the data is difficult because single-cell data snapshots are often sparse and they are collected only at a few time points.

Instead, we implement a projection based on radial basis functions (RBFs) [Fas07] that we derive from the data. Previous work demonstrates that projecting the CME model onto RBFs can yield a significant reduction in computational time and resources [Zha+10; KRS15]. The RBFs are easily described since they are defined by only two parameters, the locations of their centers and the values of their shape parameters. They have also proven advantageous for the interpolation of high-dimensional data in the field of machine learning [BL88; CCG91; CS09].

Single-cell measurement data can be approximated by a linear combination of discrete RBFs using a modification of the adaptive residual subsampling algorithm proposed by Driscoll and Heryudono [DH07]. The RBFs we compute from single-cell data interpolate the CME model for every parameter combination explored by the Metropolis-Hastings algorithm. The reduced RBF-based model is evaluated to provide an approximation to the likelihood function. Since the reduced model requires less time to solve than the FSP model, the combination of the Metropolis-Hastings algorithm with the interpolation of the FSP model using RBFs can potentially decrease the computational cost of parameter identification.

This chapter is organized as follows: In Section 4.3, we describe the projection-based model reduction of stochastic biochemical reaction models. We also introduce the likelihood function that matches these models and single-cell data measurements. In Section 4.4, we discuss model reduction using radial basis functions computed from single-cell data. Our method is described and then applied to two stochastic

models of gene regulatory networks in Section 4.5. We conclude and discuss future directions in Section 4.6.

4.3 Background

Projection-based model reduction of the chemical master equation

In this section, we introduce a general framework for projection-based model reduction of the chemical master equation. First, we truncate the state space of the chemical master equation using the finite state projection algorithm; subsequently, we reduce it through another projection on a set of vectors. In our work, these vectors are RBFs derived from measured data.

We assume that the projection state space J of the FSP model includes all experimentally observed states from the data. For instance, J could be a hyper-rectangular domain delimited by the lower and upper bounds of the observed molecular copy numbers. Let positive integer $R \in \mathbb{Z}$ be such that $R \ll N$ represents the order of the interpolated CME model. Assuming that we have been given a set of N vectors $v_1, \ldots, v_R \in \mathbb{R}_{1\times N}$, we collect them into the matrix $\Phi = [v_1 \ldots v_N]$ of dimension $N \times R$. Then the FSP-reduced probability vector p_J^{FSP} can be projected onto Φ, yielding a reduced representation

$$q = \Phi^{-L} p_J^{\text{FSP}}, \tag{4.1}$$

where $\Phi^{-L} := (\Phi^T\Phi)^{-1}\Phi^T$ is the left inverse of Φ. Thus, we can define an infinitesimal generator for the reduced system as

$$B = \Phi^{-L} A \Phi. \tag{4.2}$$

The dynamics in the reduced basis are given by

$$\frac{d}{dt} q(t) = Bq(t), \quad q(0) = \Phi^{-L} p_J^{\text{FSP}}(0). \tag{4.3}$$

This projection reduces the FSP approximation to a $R \times R$ dimensional dynamical system. The efficiency and the accuracy of approximating the FSP model with the reduced system in equation (4.3) depends on the choice of the basis vectors $\{v_1, \ldots, v_R\}$, which we obtain from single-cell data measurements.

Parameter identification using single-cell data measurements

A variety of experimental methods are used to collect single-cell measurements, such as RNA FISH [OJ16], hybridization chain reactions [Cho+10], and flow cytometry [Sha05]. By using single-cell data measurements, we can perform parameter

identification for stochastic models of gene regulation. We introduce a likelihood function L to quantify the probability of the data D, depending on the parametrization Λ of the underlying CME model of the system. For details on deriving this likelihood function from RNA FISH data, we refer the reader to [MFN15]. The logarithm of the likelihood of observing single-cell data D given a model parameterized by the set Λ can be expressed as

$$\log L(D|\Lambda) = \sum_{i \in \mathcal{I}_D} d_i \log(P(x_i|\Lambda)), \tag{4.4}$$

where d_i is the number of measured cells in state x_i and $\mathcal{I}_D$ is the set of states observed experimentally. The quantities $P(x_i|\Lambda)$ represent the probabilities to observe the data in state x_i conditional on the parameter values Λ. To compute $P(x_i|\Lambda)$ for $i \in \mathcal{I}_D$, we solve the CME using the FSP approximation. We ensure that the set of observed states $\mathcal{I}_D \subset J$, where J is the state space of the FSP approximation. Then the log-likelihood function quantifies the quality of the fit for a parameter set Λ. Every evaluation of the log-likelihood function for a parameter set Λ requires solving the FSP model once.

The best parameter fit is given by the set Λ that maximizes the log-likelihood function. The optimal value of the likelihood function can be determined using Monte Carlo methods [Bro+11]. The most expensive component of the Metropolis-Hastings approach is the evaluation of the likelihood of each parameter set, which requires the CME model solution corresponding to each of those parameter combinations. Since the Metropolis-Hastings algorithm requires a large number of samples, speedup in the CME solver would have tremendous impact on the performance of parameter identification. Our method aims to implement this speedup by interpolating the FSP model using a linear basis derived from the data measurements.

4.4 Model Reduction Using Radial Basis Functions

There is a body of literature that uses projection-based model reduction of the FSP approximation to improve computational performance. Recently proposed methods include Krylov subspaces, wavelet bases, and polynomials bases [Bur+06; JU10; TFM12]. Our choice of projection uses Gaussian radial basis functions for model reduction. Radial basis functions are easily implemented and scale well with the dimensionality of the function they interpolate [BL88; Buh00; Zha+10; KRS15]. This makes them attractive for interpolating the multi-dimensional probability distributions that result from the FSP truncation of the CME model. This basis family has been recently used in the context of solving the CME model [Zha+10; KRS15].

An overview of Gaussian radial basis functions

In this section, we introduce the radial basis functions for our projection and we discuss their properties. A standard Gaussian radial basis function is defined as

$$\phi_{c,\epsilon}(x) = \exp\left(-\epsilon^2 \|x - c\|_2^2\right), \tag{4.5}$$

where $\|\|_2$ represents the Euclidean norm, c the interpolation center, ϵ the shape parameter, and x the point at which the RBF is being evaluated.

A vector p_{data} of dimensions $K \times 1$ collects the frequencies of states $x_1, \ldots, x_K$ observed from experimental data at a single time step. This vector can be decomposed into a linear combination of K Gaussian RBF functions and its entries can be recovered via the formula

$$p_{\text{data}}(i) = \sum_{j=1}^{K} \lambda_j \phi_{x_j,\epsilon_j}(x_i). \tag{4.6}$$

We let $\Lambda = \{\lambda_1 \ldots \lambda_K\}$ be a $K \times 1$ dimensional vector of expansion coefficients. Then Λ must satisfy the equation

$$\Phi\Lambda = p_{\text{data}}, \tag{4.7}$$

where the elements $\Phi(i, j) = \phi_{x_j,\epsilon_j}(x_i)$ and $1 \leq i, j \leq K$. To interpolate the data, the expansion coefficients Λ must be determined from solving equation (4.7).

We use a discrete variation of the adaptive residual subsampling algorithm in [DH07] to interpolate the data. This method requires successive steps of refinement and coarsening to choose RBF centers and shape parameters. As we iteratively interpolate the data using radial basis functions, the matrix Φ in equation (4.7) will be replaced with matrices Φ_k, where k denotes the iteration count.

Remark 7. The interpolation matrix Φ is not symmetric because it uses state-dependent shape parameters $\epsilon_j, 1 \leq j \leq K$. Moreover, the matrix Φ can become severely ill-conditioned, which means spectral convergence is difficult to achieve [Sch07]. As explained in [DH07], this is a linear algebra issue, where the solution accuracy and the dimensionality of matrix Φ trade-off each other. Choosing the locations of Gaussian RBF centers and their scaling is also a challenging problem, particularly for high-dimensional systems [For+02; PD05; Sch07].

Choosing the centers and the shape parameters of the radial basis functions

The choice of RBF centers and of scaling parameters is paramount for the interpolation, yet there is not a systematic method to determine them. We describe how the centers and the shape parameters can be iteratively computed. Our method is based on the residual subsampling algorithm in Driscoll *et al.* [DH07].

Let $x_1, \ldots, x_K$ be the states observed in the data and retained in the FSP state space, with p_{data} being the vector storing the fraction of cells occupying those states. Let $\times_{i=1}^{N}[\ell_i^0, u_i^0]$ be a box containing these data points. The center of this box will also be the center of our first RBF, while the box's dimensions give rise to the first shape parameter vector, chosen heuristically in our implementation as $\epsilon_0 = 10^{-3}\prod_{i=1}^{K}(u_i^0 - \ell_i^0)$. We can subdivide the initial box into 2^K sub-boxes (by halving each of its dimensions), which we will use for refinement later. The initial box will be classified as a *center box* and its sub-boxes as *check boxes*. We implement the hierarchy of boxes using a 2^K-tree where the check boxes are always the leaves of that tree and each check box has a center box as its parent.

Assume that at iteration k, we have the centers $c_1^{(k)}, \ldots, c_{m_k}^{(k)}$ and the shape parameters $\epsilon_1^{(k)}, \ldots, \epsilon_{m_k}^{(k)}$. We can then form the matrix Φ_k of size $K \times m_k$, where $\Phi_k(i,j) = \phi_{c_j^{(k)},\epsilon_j^{(k)}}(x_i)$, $1 \le i \le K$, $1 \le j \le m_k$. In other words, Φ_k contains the values of the RBFs evaluated at the data points. We then use a least square solver to solve the problem $\Phi_k\Lambda_k = p_{\text{data}}$. The residual $r_k := \Phi_k\Lambda_k - p_{\text{data}}$ will be used to decide the refinement and coarsening of the boxes.

Refinement and coarsening depend on two prespecified parameters θ_r and θ_c (chosen as $0.1\theta_r$ in our numerical tests below). Note that each entry r_i of the residual corresponds to a data point x_i for $1 \le i \le K$. If $r_i > \theta_r$, meaning that the residual at x_i is above the tolerance threshold, the check box containing x_i will change status into a center box and its sub-boxes will be added to the tree as the new check boxes. This is the refinement step. As in [DH07], we also look for opportunities for coarsening. Specifically, if the residual at all data points within a center box are below θ_c, we turn that box into a check box and its children are removed from the tree. The algorithm terminates and outputs the set of centers and shape parameters if the maximum number iterations is reached or if all residual values are below the tolerance threshold θ_r.

Our variant differs from the original approach of Driscoll *et al.* [DH07] in three aspects. First, we try to interpolate the data at all points in every iteration. Second, the residual is evaluated at all data points and not just at the centers of the check boxes.

Third, the original algorithm used multiquadric functions, while we use Gaussian radial basis functions instead. The original adaptive residual subsampling algorithm has no limits on the number of iterations since it was designed for interpolation on continuous domains, but that is not the case with the discrete-state data of interest here.

Choosing the right refinement and coarsening thresholds is crucial in decomposing the data using RBFs. If the thresholds are too strict, the interpolant will try to reproduce the noisy characteristics of the data and clutter the centers in a way that exacerbates numerical instability; if the thresholds are too relaxed, the interpolant will smooth away relevant features of the data such as multi-modality. A systematic method to decide these thresholds remains unknown.

4.5 Numerical Examples

For the numerical examples in this section, we begin by simulating the data using the Gillespie algorithm [Gil77]. We then apply our modified algorithm from [DH07] to a snapshot in time of the simulated single-cell data to find a suitable RBF basis representation. We use the RBF basis derived from single-cell data to interpolate the FSP model and to derive the reduced RBF-FSP model.

To identify the biochemical reaction rate parameters that best fit the single-cell data, we use the Metropolis-Hastings algorithm or we perform a parameter sweep. We compare the performance of the FSP model and of the RBF-FSP model. We find that the parameters we identify for the FSP model and for the RBF-FSP model are similar in accuracy. Nonetheless, using the RBF-FSP model significantly decreases the computational time of the parameter identification step.

The steps of our method are organized as follows:

1. Simulate single-cell data using the Gillespie algorithm or collect single-cell data experimentally.
2. Choose a snapshot of the single-cell data at a point in time to represent the system behavior.
3. Employ the modified residual subsampling algorithm from [DH07] to find an RBF basis representation of the data snapshot.
4. Interpolate the FSP model using the RBF basis obtained from the data to derive the RBF-FSP model.
5. Run the Metropolis-Hastings algorithm using both the FSP model and the RBF-FSP model.

6. For both the FSP and the RBF-FSP model, compute the posterior parameter distributions that can represent the single-cell data.
7. Compare the parameter identification accuracy and the computational time of the FSP and the RBF-FSP models.

When we substitute Metropolis-Hastings with a simple parameter sweep, we skip steps 5 and 6.

Bursting Gene Expression

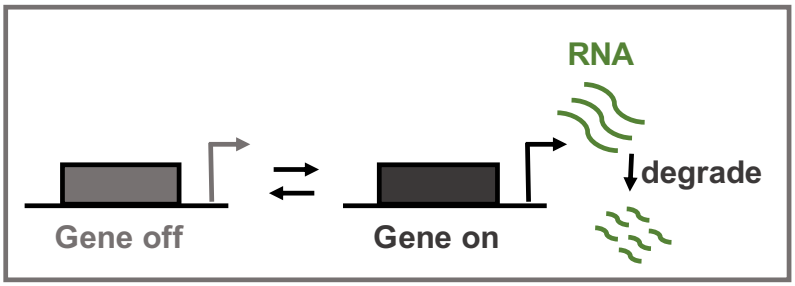

Figure 4.1: **Bursting gene expression diagram.** The two-state bursting gene model captures RNA transcription and degradation for a single gene that can switch between an active state (on) and an inactive state (off), depending on whether the transcription factor is bound or unbound to the gene's promoter. RNA is transcribed when the gene is in the on state; subsequently, it is degraded by ribonucleases.

The bursting gene expression model in Figure 4.1 describes the changes in the state of a gene's promoter corresponding to the binding and unbinding of a transcription factor. When a transcription factor binds to the gene, RNA is actively translated at rate k_r. For an RNA molecule R, this simple two-state view of gene expression can result in a variety of RNA dynamics, depending on the system's parameters [MNO12]. This gene regulatory network can be modeled by the following biochemical reactions that describe the state of the gene G and the dynamics of the RNA copy number R:

$$\mathrm{G_{off}} \underset{\mathrm{k_{off}}}{\overset{\mathrm{k_{on}}}{\rightleftharpoons}} \mathrm{G_{on}}, \tag{4.8}$$

$$\mathrm{G_{on}} \xrightarrow{\mathrm{k_r}} \mathrm{G_{on}} + \mathrm{R}, \tag{4.9}$$

$$\mathrm{R} \xrightarrow{\mathrm{k_{deg}}} \emptyset. \tag{4.10}$$

We simulate the bursting gene data at 30 linearly-spaced time points between 0 sec and 10 sec using the Gillespie algorithm. Figure 4.2(A) illustrates the RBF-based

representation (green line) of the simulated data (black line) with 5 RBF centers selected using the adaptive residual subsampling algorithm.

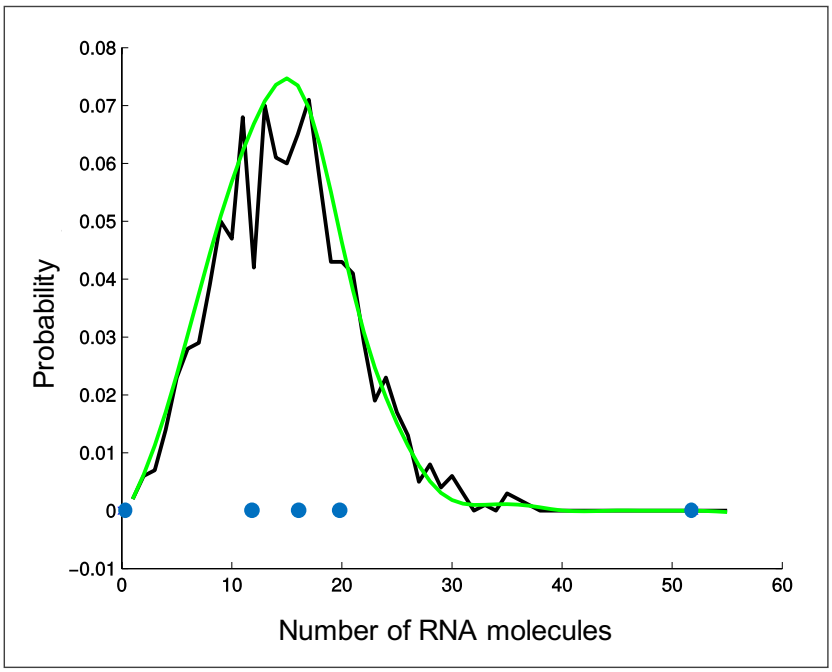

Figure 4.2: **Radial basis function interpolation of simulated single-cell data for the bursting gene model.** We interpolate the noisy data generated using the Gillespie algorithm (black) with the approximation provided by the radial basis functions (green) at time 10 sec. We use 5 RBF centers, positioned at the blue dots. The RBF interpolation provides a smooth approximation to the noisy data.

We then perform parameter identification of the FSP and the RBF-FSP projections of the busting gene model. We consider 2500 parameter combinations of the transcription rate k_r and of the gene off rate k_{off} and we search for the best fit to the simulated data. Our parameter combinations span one order of magnitude. The true values (i.e. the values that were used to generate the data) are at the center of the heat maps for both the FSP and the RBF-FSP models in Figures 4.3(A) and 4.3(B), respectively.

The parameter combinations that best fit the simulated data, their associated likelihood function values, and the computational times are given in Table 4.1. Only 5 RBF centers are required to represent the data in the RBF-FSP model, while the FSP model uses 80 states. Hence, the computational time required to identify the parameters of the RBF-FSP is only a tenth of the time required for the FSP model, while the accuracy of the identified parameter values is similar.

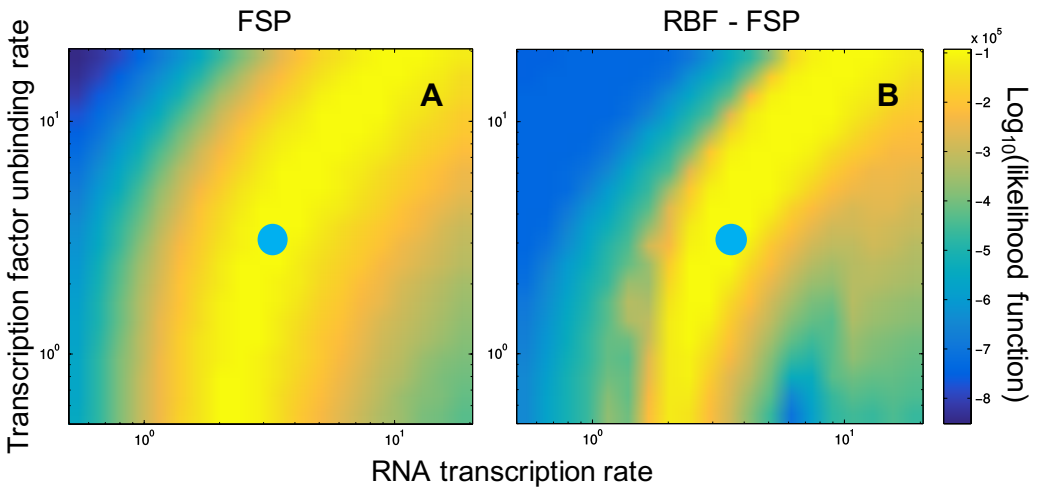

Figure 4.3: **Likelihood functions of the FSP and of the RBF-FSP for two parameter combinations of the bursting gene expression model.** We identify parameters of the gene bursting model using both the model reduced through the FSP projection and the model reduced through a subsequent RBF-FSP projection. We conduct a search for the gene bursting model parameters using 30 equally spaced time points from 0 sec to 10 sec. We compute and we plot the likelihood functions of parameter combinations of the RNA transcription rate k_r and of the transcription factor unbinding rate k_{off}. To find the maximum of the likelihood functions, we search a grid of potential parameter values centered around the true parameter values (bright blue dots). The size of the grid equals one order of magnitude. We find that the two likelihood functions are maximized along the bright yellow shapes and that the two shapes mostly match each other in panels A and B, although the shape in B is more restrictive.

Moreover, we perform a more in depth parameter search for the FSP approximation and the RBF-FSP projection of the gene bursting model. We plot the probability distributions of the four parameters of the bursting gene model in Figure 4.4, along with the parameter values used for the simulated data. We observe a loss in the accuracy of parameter identification, but a speedup in the simulation time of the RBF-FSP model. We include additional details in Appendix C.

Mutually-Repressing Toggle Switch

The genetic toggle switch circuit in Gardner *et al.* [GCC00] describes two mutually repressing promoter species *lacI* and λcI and their stochastic interactions. We have already provided an illustration of the genetic toggle switch in Figure 2.8; however, for an adaptation of the circuit diagram to the model in equations (4.11) and (4.12), see Figure C.3 in Appendix C.

	k_r	k_{off}	L(D|Λ)	time
FSP model	9.5410	1.3028	-9.2659e4	172.27
RBF-FSP model	9.5410	1.4311	-9.3031e4	16.03
True values	10.0	1.5	-9.2637e4	-

Table 4.1: **The maximum values of the RNA transcription rate and of the transcription factor unbinding rate for the FSP and the RBF-FSP reductions of the gene bursting model.** We performed a parameter sweep over a grid of 2500 parameter combinations for rates k_r and k_{off} with both the FSP model and the RBF-FSP model. We identified the best fitting parameters and we computed their associated likelihood function values. We obtain an order of magnitude decrease in runtime at a small cost in parameter identification accuracy.

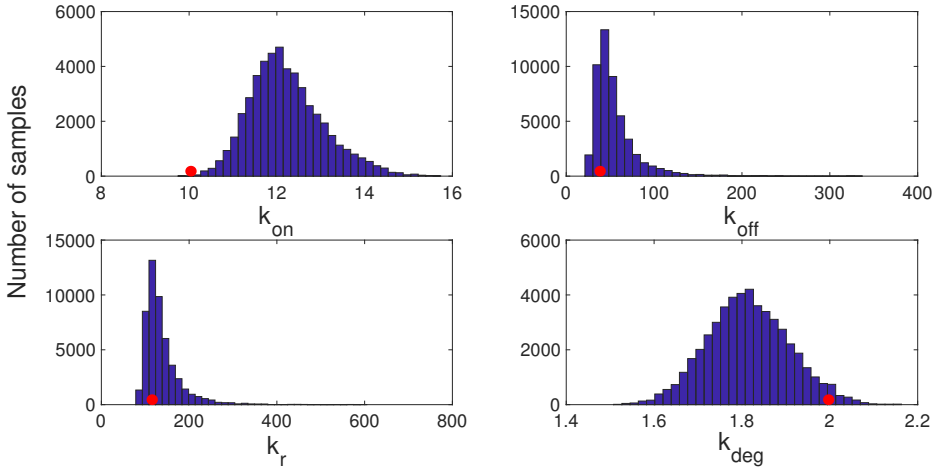

Figure 4.4: **The parameters of the bursting gene model, as identified using a version of the adaptive Metropolis-Hastings algorithm.** We run a chain of 100000 samples to identify the parameters of the gene bursting model using the model reduced through the RBF-FSP. We plot the probability distributions of the four parameters of the model. We report an acceptance rate of 76% of the proposed samples. The true parameter values are $\{10, 30, 100, 2\}$ and they are indicated in the histograms by red dots. The Metropolis-Hastings algorithm takes 5×10^3 sec (one and a half hours) and the initial search parameters are $\{5, 15, 10, 4\}$.

We assume that the reactions of this biochemical network are given by:

$$\emptyset \underset{w_2}{\overset{w_1}{\rightleftharpoons}} \lambda \mathrm{cI}, \tag{4.11}$$

$$\emptyset \underset{w_4}{\overset{w_3}{\rightleftharpoons}} \mathrm{LacI}, \tag{4.12}$$

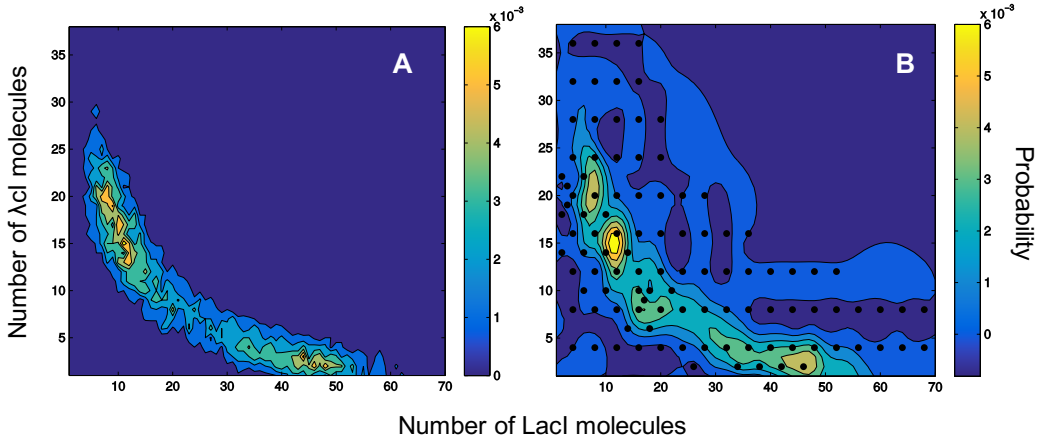

Figure 4.5: **Radial basis function interpolation of simulated single-cell data for the genetic toggle switch in [GCC00].** (A) We obtain the genetic toggle switch data from 1000 runs of the Gillespie algorithm. The genetic toggle switch is implemented using two repressor proteins LacI and λcI. We plot the joint probability mass of proteins LacI and λcI at 4 hrs. (B) The RBF interpolation of the simulated data for the genetic toggle switch at 4 hrs. We use 137 radial basis functions centered at the black dots to interpolate the data. To perform the interpolation, we use an adaptive mesh grid algorithm for two-dimensional distributions, as in [DH07].

where the propensity functions of the chemical reactions are as follows:

$$
\begin{aligned}
w &= \{w_1, w_2, w_4, w_4\}, \\
w_1 &= b_x + \frac{k_x}{1 + \text{LacI}^{n_{yx}}}, \\
w_2 &= \gamma_x \cdot \lambda\text{cI}, \\
w_3 &= b_y + \frac{k_y}{1 + \lambda\text{cI}^{n_{xy}}}, \\
w_4 &= \gamma_y \cdot \text{LacI},
\end{aligned}
$$

Here, b_x and γ_y are the basal production rates of proteins λcI and LacI, γ_x and γ_y are the degradation rates of proteins λcI and LacI, and k_x and k_y are the association constants. We limit our search to the 6 parameters $b_x, b_y, k_x, k_y, \gamma_x, \gamma_y$ that enter the CME model of the genetic toggle switch linearly, as explained in equation (1.20) in Chapter 1. The Hill coefficients are assumed to be approximately equal to 2.

Remark 8. Identifying parameters that appear linearly in the CME model is computationally advantageous. If the CME is linear in these parameters, then it can be

separated as:

$$\frac{dp(t)}{dt} = \Big(\sum_{i=1}^{M} \Lambda_i H_i \Big) p(t). \tag{4.13}$$

Then the infinitesimal generator of the reduced CME system is also linear in these parameters:

$$B = \Big(\sum_{i=1}^{M} \Lambda_i (\Phi^{-L} H_i \Phi) \Big). \tag{4.14}$$

The matrix products $\Phi^{-L} H_i \Phi$ can be precomputed and do not change with different parameter values. Thus, forming the projection of the CME model onto a linear basis is more computationally efficient. Additionally, the matrices $H_1, \ldots, H_M$ are even sparser than the original CME matrix H. The linear dependence on parameters in the reduced model significantly reduces our computation time; for the genetic toggle switch, it provides a five fold reduction.

For this numerical example, we give the true model parameters in Table 4.2. We then run the Metropolis-Hastings algorithm combined with the FSP model solution and the RBF-FSP model solution at each step to find the parameter values that maximize the likelihood function. We obtain the probability distributions of the parameters for the FSP model and the FSP-RBF model and we give the mean values in Table 4.2. In Figure 4.6, the 95% and 65% confidence intervals were computed for the second half of a 100000 iteration-long MCMC chain that has been thinned down to 10000 samples. The parameters are normalized by their true parameter values in Figure 4.6.

	b_y	k_x	k_y	γ_y	$L(D\|\Lambda)$	time(min)
FSP model	1.89e-3	1.46e-2	1.60e-2	3.46e-4	-3.0871e4	3000
RBF-FSP model	2.22e-3	1.65e-2	1.85e-2	4.31e-4	-3.114e4	200
True values	2.20e-3	1.60e-2	1.70e-2	3.8e-4	-3.022e4	-

Table 4.2: **The parameters of the genetic toggle switch identified by the Metropolis-Hastings algorithm.** Both the parameters identified using the FSP model and the RBF-FSP model match the values of the true parameters, as do their likelihood function values. There is a large speedup obtained by using the RBF-FSP model over the FSP approximation model. This speedup is much smaller (reduced only in half) if the Hill coefficients are added into the parameter search.

Parameter identification with the RBF-FSP model requires less than a tenth of the computational time of parameter identification with the FSP model. Nevertheless,

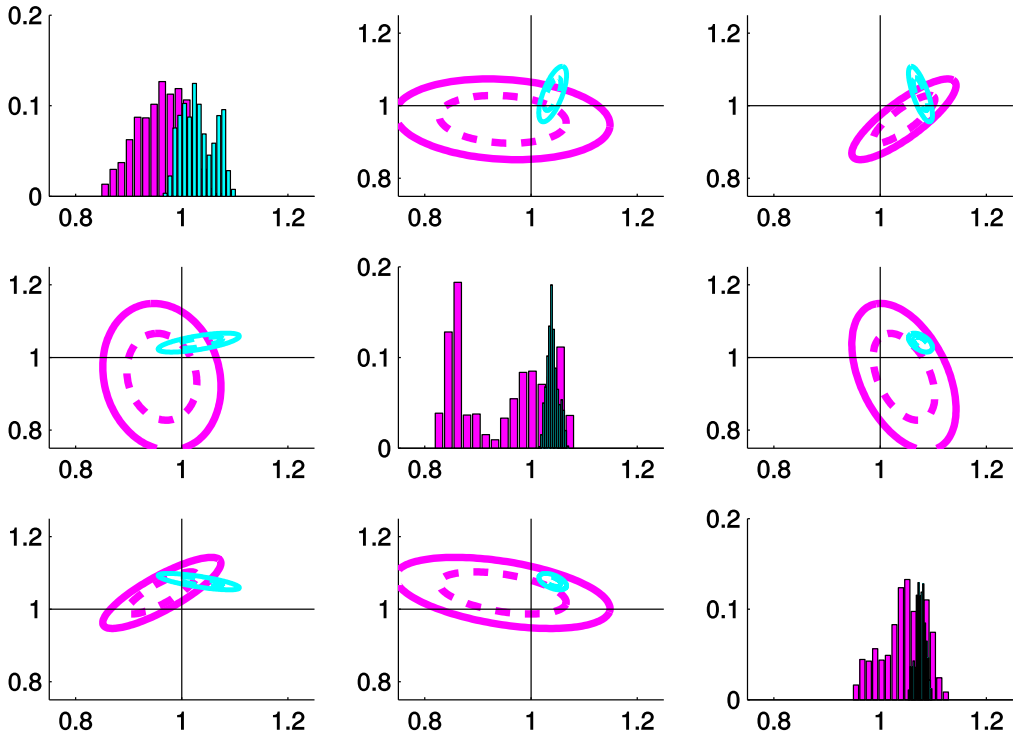

Figure 4.6: **The parameter distributions obtained from the Metropolis-Hastings algorithm for the genetic toggle switch with the FSP and the RBF-FSP models.** We demonstrate that the parameter values identified for both the RBF-FSP model and the FSP model are able to approach the true parameter values. The diagonal contains the histograms for the toggle switch parameters b_y, k_x, and k_y, normalized by their true values from Table 4.2, as identified for the FSP model (magenta) and for the RBF-FSP model (cyan). We observe that the three pairs of probability distributions are close to value one, as they should be following normalization to the true values, but that only the pair for b_y includes value one. This indicates a loss in parameter accuracy. Off the main diagonal, we illustrate the 95% (solid lines) and the 65% (dashed lines) regions of the parameter distributions sampled with Metropolis-Hastings for the FSP model (magenta) and for the RBF-FSP model (cyan). The true values are marked in the center, at the intersection of the vertical and horizontal lines. The parameter pairs we illustrate are (b_y, k_x), in the second from the top left plot, (b_y, k_y), in the third from the top left plot, and (k_x, k_y), in the second from the bottom right plot. The plot illustrations are symmetric across the diagonal.

the mean parameters values identified with both models are close to the true parameter values, as illustrated in Figure 4.6. Thus, RBF interpolation has provided us with improved computational time at a small cost in parameter accuracy.

4.6 Conclusion and Future Work

Stochastic models of biochemical reaction networks are subject to the curse of dimensionality [MNO12]. Their state spaces become extremely large, even for a small number of biochemical species. As a result, computationally solving these stochastic models is expensive. This becomes an extreme problem when fitting parameters to stochastic models using algorithms such as Metropolis-Hastings. These algorithms require a solution to the stochastic model for each parameter evaluation; hence, they can necessitate hundreds of thousands of solutions to find the best distribution of parameter values. Thus, parameter identification for stochastic models of biochemical reaction networks is a very computationally challenging task that the research in this chapter aims to address.

Frequently, despite its very large state space, the underlying dynamics of stochastic models such as the chemical master equation are less complex when projected onto a suitable interpolation basis. It is common to truncate the state space of the chemical master equation using the finite state projection algorithm (FSP). However, we propose that a subsequent projection-based model reduction, such as an interpolation via radial basis functions, can retain the accuracy of parameter identification, while reducing the computational burden of solving the chemical master equation.

In this chapter, we perform parameter identification for the stochastic models of two gene regulatory networks. First, we demonstrate the interpolation of simulated single-cell data generated through forward runs of the Gillespie algorithm [Gil77] using radial basis functions (RBFs). For the Gaussian RBF interpolation step, we use the residual subsampling algorithm in [DH07] to reduce the dimensionality of the stochastic models, after we truncate their state spaces using the FSP approximation. Then we employ the Metropolis-Hastings algorithm to perform parameter identification of the FSP and of the RBF-FSP models. We compare the results of parameter identification with the true parameter values; we observe a small loss in parameter accuracy and a ten-fold gain in computational time.

There are a number of open questions left about the interpolation of data using radial basis functions. The results in Figures 4.3 and 4.6 suggest that the RBF-FSP model constrains the parameter values more tightly than the FSP model. This could indicate that the RBF interpolation smooths out too much of the noise in the data; therefore, how smooth the data itself is can play an important role in identifying the underlying models. Yet, it is not yet understood how data sets with different levels

of noise affect the parameter identification results.

Moreover, the RBF interpolation matrix becomes severely ill-conditioned during successive iterations of refinement and coarsening, a possibility also indicated by [DH07]. Even though radial basis functions have been used extensively in solving PDEs [FS98], their practical implementation is well-known to be challenging [Sch07], particularly when they interpolate noisy data. The condition number of the interpolation matrix is often high and it is recommended to first regularize it through singular-value decomposition to eliminate the numerous eigenvalues that are close to zero; however, the computational expense of the singular-value decomposition increases with the size of the matrix and the truncation results in a loss in accuracy. Since several interpolation matrices are used during the residual subsampling algorithm, regularization could be expensive and lose the computational advantage afforded by the RBF interpolation. Furthermore, there are several well-known tradeoffs in RBF interpolation such as the tradeoff between error and stability when choosing smoothness; additionally, we observed a second common tradeoff induced by scaling: our wide-scaled RBFs improve the error, but induce instability.

Last, there is no theoretical method to choose the optimal thresholds for refinement and coarsening in the residual subsampling algorithm [DH07]. We have found that the RBF interpolation is very sensitive to these two threshold values since they determine the locations of their centers and their shape parameters. Moreover, other interpolation bases such as proper orthogonal decomposition and several wavelet bases have performed worse than RBFs in our numerical experiments to interpolate the genetic toggle switch data. Therefore, we believe that developing our understanding of RBF interpolation, of the numerical properties of their interpolation matrices, and of the optimal values of the two thresholds will be crucial to tackling problems in higher dimensions.

Chapter 5

CONCLUSION

This thesis contributed to three aspects of engineering biological systems: analysis, design, and computational methods. First, we designed the stochastic behaviors of biochemical reaction networks in Chapter 2. Second, we analyzed a biological controller implemented by sequestration feedback in Chapter 3. Last, we employed a model reduction method to efficiently perform parameter identification for stochastic biological systems in Chapter 4. In this section, we describe the contribution of this thesis and we suggest future research directions for each chapter.

In Chapter 2, we formulated and solved the problem of designing the stochastic behaviors of biochemical reaction networks. When analytic solutions for stochastic models are available, the design of stochastic behaviors is greatly simplified, as we demonstrated using the two-component transcription network from [Men+17]; however, solutions for these stochastic models are often unknown [MNO12]. Thus, we introduced a general framework for the optimization of the reaction rates of biochemical systems rates such that their transient and stationary probability distributions are constrained. Design constraints for these distributions included specifying their modality, the locations of their modes, and the rate of convergence to stationarity [MG09]. Under these constraints, we determined the optimal reaction rate values of the biochemical reaction networks by solving an approximation to a polynomial optimization problem. We demonstrated this framework on three examples of biochemical reaction networks: a protein production-degradation network [DM15], the Schlögl biochemical network [Gil91; Gun+05], and the genetic toggle switch [GCC00]. The content of Chapter 2 was published in [Bae+15] and in [Men+17].

Future research directions for this chapter include increasing the size of the networks we can solve design problems for, incorporating other distributional constraints in the optimization problem, and approximating the exponential operator using a polynomial series expansion. Polynomial optimization is an active research field [BPT12; Wak+05a] that employs sums of squares and semidefinite relaxations to find solutions to challenging optimization problems. These methods could be used in stochastic design problems after we replace the exponential operator with

a polynomial series approximation. This could ultimately help increase the size of the biochemical networks that we can solve design problems for. In particular, we anticipate that constraining the stochastic transients of biochemical reaction networks would be very helpful for design due to the almost complete lack of analytical solutions for them.

In Chapter 3, we investigated the properties of stability and performance of a synthetic biological controller implemented by a sequestration feedback motif. It was established in [BGK16] that sequestration feedback implements perfect adaptation under the two conditions of stability and of zero controller species degradation. First, we demonstrated that the controller species degradation is often nonzero and therefore, it must be included in the modeling of the sequestration feedback network. Then we derived an analytical criterion for the stability of sequestration feedback networks and determined their performance properties. We also found a trade-off between achieving a large stability margin and a small steady state error. Last, we provided guidelines for the implementation of sequestration feedback networks that fulfill desired specifications of stability and performance. The content of this chapter was published in [Ols+17] and [Ren+17].

Future research into synthetic biological controllers will include investigating other mechanisms for biological control, as well as testing the versatility of the current controllers. In addition to the sequestration feedback motif we analyzed, there are multiple other options for implementing biological control such as paradoxical extracellular signaling [Har+14] and post-translation mechanisms. The properties of stability and performance of these biological controllers will be assessed using the metrics introduced in Chapter 3. Furthermore, since the current synthetic biological controllers [Hsi+14; McC+17; Fol+17; Lil+17] have been developed alongside the processes they control, their versatility will have to be determined. It is likely that these controllers will require additional tuning before being able to control new processes. Future research includes the development of a more versatile class of synthetic biological controllers, perhaps inspired by the controller developed in [HME18].

In Chapter 4, we introduced a method for the parameter identification of stochastic biochemical reaction networks using simulated single-cell data. Stochastic models of biochemical reaction networks are often expensive to solve computationally; this problem is exacerbated when performing parameter identification for them using methods such as the Metropolis-Hastings algorithm that requires hundreds

of thousands of solutions to find the best parameter distribution. Therefore, we proposed to first truncate the state space of stochastic models and then to perform a projection-based model reduction step. Our projection step relies of Gaussian radial basis functions determined from the simulated data [DH07]. The model reduction steps decreased the computational time of parameter identification by an order of magnitude, although they introduced a small loss in accuracy. We demonstrated our method on two example networks of gene regulation: a bursting gene network and the genetic toggle switch in [GCC00]. The content of Chapter 4 was developed in collaboration with Huy Vo, Zachary Fox, and Brian Munsky (Colorado State University).

Future research into the interpolation of discrete, noisy data using radial basis functions could expand the dimension of stochastic biochemical reaction networks that we can perform parameter identification for. The practical implementation of RBFs is challenging [Sch07; FS98], particularly when they interpolate noisy data; a good reproduction of noisy data requires handling several ill-conditioned matrices. Hence, the development of preconditioners and of regularization methods could improve the computational efficiency of RBF interpolation. Understanding how to choose optimal thresholds for refinement and coarsening in the residual subsampling algorithm [DH07] would also help the parameter identification of higher dimensional biochemical reaction networks.

BIBLIOGRAPHY

[ACK10] David F Anderson, Gheorghe Craciun, and Thomas G Kurtz. "Product-form Stationary Distributions For Deficiency Zero Chemical Reaction Networks". In: *Bulletin of Mathematical Biology* 72.8 (2010), pp. 1947–1970.

[Alo06] Uri Alon. *An Introduction To Systems Biology: Design Principles Of Biological Circuits*. CRC Press, 2006.

[AM08] Karl Johan Aström and Richard M Murray. *Feedback Systems: An Introduction For Scientists And Engineers*. Princeton University Press, 2008.

[Ang+10] Jordan Ang, Sangram Bagh, et al. "Considerations For Using Integral Feedback Control To Construct A Perfectly Adapting Synthetic Gene Network". In: *Journal of Theoretical Biology* 266.4 (2010), pp. 723–738.

[Ann+17] Fabio Annunziata, Antoni Matyjaszkiewicz, et al. "An Orthogonal Multi-input Integration System To Control Gene Expression In Escherichia coli". In: *ACS Synthetic Biology* 6.10 (2017), pp. 1816–1824.

[Ark08] Adam Arkin. "Setting The Standard In Synthetic Biology". In: *Nature Biotechnology* 26.7 (2008), p. 771.

[ARM98] Adam Arkin, John Ross, and Harley H McAdams. "Stochastic Kinetic Analysis Of Developmental Pathway Bifurcation In Phage λ-infected Escherichia coli Cells". In: *Genetics* 149.4 (1998), pp. 1633–1648.

[Arn98] Frances H Arnold. "Design By Directed Evolution". In: *Accounts of Chemical Research* 31.3 (1998), pp. 125–131.

[Bae+15] Ania-Ariadna Baetica, Ye Yuan, et al. "A stochastic framework for the design of transient and steady state behavior of biochemical reaction networks". In: *2015 IEEE 54th Annual Conference on Decision and Control (CDC)*. IEEE. 2015, pp. 3199–3205. DOI: 10.1109/CDC.2015.7402699.

[Bae+16] Ania-Ariadna Baetica, Thomas A Catanach, et al. "A Bayesian approach to inferring chemical signal timing and amplitude in a temporal logic gate using the cell population distributional response". In: *bioRxiv* (2016). DOI: 10.1101/087379.

[Bal+04] Nathalie Q Balaban, Jack Merrin, et al. "Bacterial Persistence As A Phenotypic Switch". In: *Science* 305.5690 (2004), pp. 1622–1625.

[Bal+06] Karen Ball, Thomas G Kurtz, et al. "Asymptotic Analysis Of Multiscale Approximations To Reaction Networks". In: *The Annals of Applied Probability* 16.4 (2006), pp. 1925–1961.

[Ber+04] Jonathan A Bernstein, Pei-Hsun Lin, et al. "Global Analysis Of Escherichia coli RNA Degradosome Function Using DNA Microarrays". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.9 (2004), pp. 2758–2763.

[BGK16] Corentin Briat, Ankit Gupta, and Mustafa Khammash. "Antithetic Integral Feedback Ensures Robust Perfect Adaptation In Noisy Biomolecular Networks". In: *Cell Systems* 2.1 (2016), pp. 15–26. ISSN: 2405-4712. DOI: 10.1016/j.cels.2016.01.004.

[BL88] D. S. Broomhead and D. Lowe. "Multivariable Functional Interpolation And Adaptive Networks". In: *Complex Systems* 2 (1988), pp. 321–355.

[BPT12] Grigoriy Blekherman, Pablo A Parrilo, and Rekha R Thomas. *Semidefinite Optimization And Convex Algebraic Geometry*. SIAM, 2012.

[Bro+11] Steve Brooks, Andrew Gelman, et al. *Handbook Of Markov Chain Monte Carlo*. CRC Press, 2011.

[Buh00] Martin Dietrich Buhmann. "Radial Basis Functions". In: *Acta Numerica 2000* 9 (2000), pp. 1–38.

[Bur+06] K. Burrage, M. Hegland, et al. "A Krylov-based Finite State Projection Algorithm For Solving The Chemical Master Equation Arising In The Discrete Modelling Of Biological Systems". In: 150^{th} *Markov Anniversary Meeting, Charleston, SC, USA*. Ed. by A.N. Langville and W.J. Stewart. Boson Books, 2006, pp. 21–38.

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[CBD11] Fiona A Chandra, Gentian Buzi, and John C Doyle. "Glycolytic Oscillations And Limits On Robust Efficiency". In: *Science* 333.6039 (2011), pp. 187–192.

[CCG91] S. Chen, C. F. N. Cowan, and P. M. Grant. "Orthogonal Least Squares Learning Algorithm For Radial Basis Function Networks". In: *IEEE Transactions on Neural Networks* 2.2 (Mar. 1991), pp. 302–309. ISSN: 1045-9227. DOI: 10.1109/72.80341.

[CG08] Izhack Cherny and Ehud Gazit. "Amyloids: Not Only Pathological Agents But Also Ordered Nanomaterials". In: *Angewandte Chemie International Edition* 47.22 (2008), pp. 4062–4069.

[CG10] James M Clomburg and Ramon Gonzalez. “Biofuel Production In Escherichia coli: The Role Of Metabolic Engineering And Synthetic Biology”. In: *Applied Microbiology and Biotechnology* 86.2 (2010), pp. 419–434.

[CGP07] Y. Cao, D.T. Gillespie, and L.R. Petzold. “Adaptive Explicit-implicit Tau-Leaping Method With Automatic Tau Selection”. In: *J. Chem. Phys* 126.22 (2007), p. 224101. ISSN: 0021-9606. DOI: 10.1063/1.2745299.

[Che+18] Michael Chevalier, Mariana Gomez-Schiavon, et al. “Design And Analysis Of A Proportional-Integral-Derivative Controller With Biological Molecules”. In: *bioRxiv* (2018). DOI: 10.1101/303545.

[Cho+08] Paul J Choi, Long Cai, et al. “A Stochastic Single-molecule Event Triggers Phenotype Switching Of A Bacterial Cell”. In: *Science* 322.5900 (2008), pp. 442–446.

[Cho+10] Harry MT Choi, Joann Y Chang, et al. “Programmable In Situ Amplification For Multiplexed Imaging Of mRNA Expression”. In: *Nature Biotechnology* 28.11 (2010), p. 1208.

[CLE08] Barry Canton, Anna Labno, and Drew Endy. “Refinement And Standardization Of Synthetic Biological Parts And Devices”. In: *Nature Biotechnology* 26.7 (2008), p. 787.

[CLL10] Youfang Cao, Hsiao-Mei Lu, and Jie Liang. “Probability Landscape Of Heritable And Robust Epigenetic State Of Lysogeny In Phage Lambda”. In: *Proceedings of the National Academy of Sciences* 107.43 (2010), pp. 18445–18450.

[Coh67] Dan Cohen. “Optimizing Reproduction In A Randomly Varying Environment When A Correlation May Exist Between The Conditions At The Time A Choice Has To Be Made And The Subsequent Outcome”. In: *Journal of Theoretical Biology* 16.1 (1967), pp. 1–14.

[CS09] Youngmin Cho and Lawrence K Saul. “Kernel Methods For Deep Learning”. In: *NIPS Conf.* 9 (2009), pp. 342–350.

[CS16] Venkat Chandrasekaran and Parikshit Shah. “Relative Entropy Relaxations For Signomial Optimization”. In: *SIAM Journal on Optimization* 26.2 (2016), pp. 1147–1173.

[DDQ16] Domitilla Del Vecchio, Aaron J Dy, and Yili Qian. “Control Theory Meets Synthetic Biology”. In: *Journal of The Royal Society Interface* 13.120 (2016), p. 20160380.

[De +09] Natalie De Jonge, Abel Garcia-Pino, et al. “Rejuvenation Of CcdB-poisoned Gyrase By An Intrinsically Disordered Protein Domain”. In: *Molecular Cell* 35.2 (2009), pp. 154–163.

[DH07] Tobin A. Driscoll and Alfa R.H. Heryudono. “Adaptive Residual Subsampling Methods For Radial Basis Function Interpolation And Collocation Problems”. In: *Comput. Math. with Appl.* 53.6 (2007), pp. 927–939. ISSN: 08981221. DOI: 10.1016/j.camwa.2006.06.005.

[DKM10] Mary J Dunlop, Jay D Keasling, and Aindrila Mukhopadhyay. “A Model For Improving Microbial Biofuel Production Using A Synthetic Feedback Loop”. In: *Systems and Synthetic Biology* 4.2 (2010), pp. 95–104.

[DM15] Domitilla Del Vecchio and Richard M Murray. *Biomolecular Feedback Systems*. Princeton University Press, 2015.

[DRH00] JMP Desterro, MS Rodriguez, and RT Hay. “Regulation Of Transcription Factors By Protein Degradation”. In: *Cellular and Molecular Life Sciences CMLS* 57.8-9 (2000), pp. 1207–1219.

[DS66] IG Darvey and PJ Staff. “Stochastic Approach To First-Order Chemical Reaction Kinetics”. In: *The Journal of Chemical Physics* 44.3 (1966), pp. 990–997.

[EE10] Avigdor Eldar and Michael B Elowitz. “Functional Roles For Noise In Genetic Circuits”. In: *Nature* 467.7312 (2010), pp. 167–173.

[EGK02] H El-Samad, JP Goff, and M Khammash. “Calcium Homeostasis And Parturient Hypocalcemia: An Integral Feedback Perspective”. In: *Journal of Theoretical Biology* 214.1 (2002), pp. 17–29.

[EL00] Michael B Elowitz and Stanislas Leibler. “A Synthetic Oscillatory Network Of Transcriptional Regulators”. In: *Nature* 403.6767 (2000), p. 335.

[Elo+02] Michael B Elowitz, Arnold J Levine, et al. “Stochastic Gene Expression In A Single Cell”. In: *Science* 297.5584 (2002), pp. 1183–1186.

[Eng06] Stefan Engblom. “Computing The Moments Of High Dimensional Solutions Of The Master Equation”. In: *Applied Mathematics and Computation* 180.2 (2006), pp. 498–515.

[Fas07] Gregory E. Fasshauer. *Meshfree Approximation Methods With MATLAB*. World Scientific, 2007. ISBN: 978-981-270-634-8.

[Fer16] James E Ferrell Jr. “Perfect And Near-perfect Adaptation In Cell Signaling”. In: *Cell Systems* 2.2 (2016), pp. 62–67.

[Fol+17] Thomas Folliard, Harrison Steel, et al. “A Synthetic Recombinase-based Feedback Loop Results In Robust Expression”. In: *ACS Synthetic Biology* (2017).

[For+02] Bengt Fornberg, Tobin A Driscoll, et al. “Observations On The Behavior Of Radial Basis Function Approximations Near Boundaries”. In: *Computers & Mathematics with Applications* 43.3-5 (2002), pp. 473–490.

[Fra+14] Elisa Franco, Giulia Giordano, et al. “Negative Autoregulation Matches Production And Demand In Synthetic Transcriptional Networks”. In: *ACS Synthetic Biology* 3.8 (2014), pp. 589–599.

[FS98] Carsten Franke and Robert Schaback. “Solving Partial Differential Equations By Collocation Using Radial Basis Functions”. In: *Applied Mathematics and Computation* 93.1 (1998), pp. 73–82.

[GA14] Mahmoud Ghofrani and Amirsaman Arabali. “A Stochastic Framework For Power System Operation With Wind Generation And Energy Storage Integration”. In: *Innovative Smart Grid Technologies Conference (ISGT), 2014 IEEE PES*. IEEE. 2014, pp. 1–5.

[Gal+12] Michal Galdzicki, Mandy Wilson, et al. *Synthetic Biology Open Language (SBOL) version 1.1.0*. Tech. rep. 2012.

[GB08] Michael Grant and Stephen Boyd. “Graph Implementations For Nonsmooth Convex Programs”. In: *Recent Advances in Learning and Control* (2008), pp. 95–110.

[GBY08] Michael Grant, Stephen Boyd, and Yinyu Ye. *CVX: Matlab Software For Disciplined Convex Programming*. 2008.

[GCC00] Timothy S Gardner, Charles R Cantor, and James J Collins. “Construction Of A Genetic Toggle Switch In Escherichia coli”. In: *Nature* 403.6767 (2000), pp. 339–342.

[GD12] Reza Ghaemi and Domitilla Del Vecchio. “Stochastic Analysis Of Retroactivity In Transcriptional Networks Through Singular Perturbation”. In: *American Control Conference (ACC), 2012*. IEEE. 2012, pp. 2731–2736.

[GD14] Andras Gyorgy and Domitilla Del Vecchio. “Limitations And Tradeoffs In Gene Expression Due To Competition For Shared Cellular Resources”. In: *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE. 2014, pp. 5431–5436.

[Gel+14] Andrew Gelman, John B Carlin, et al. *Bayesian Data Analysis*. Vol. 2. CRC Press Boca Raton, FL, 2014.

[Ger88] Kenn Gerdes. “The parB (hok/sok) Locus Of Plasmid R1: A General Purpose Plasmid Stabilization System”. In: *Nature Biotechnology* 6.12 (1988), pp. 1402–1405.

[GHM14] Shaobin Guo, Yutaka Hori, and RM Murray. “Systematic Design And Implementation Of A Novel Synthetic Fold-change Detector Biocircuit In Vivo”. In: *California Institute of Technology, Tech. Rep* (2014).

[Gil00] Daniel T Gillespie. “The Chemical Langevin Equation”. In: *The Journal of Chemical Physics* 113.1 (2000), pp. 297–306.

[Gil07] Daniel T Gillespie. “Stochastic Simulation Of Chemical Kinetics”. In: *Annu. Rev. Phys. Chem.* 58 (2007), pp. 35–55.

[Gil77] D.T. Gillespie. “Exact Stochastic Simulation Of Coupled Chemical Reactions”. In: *J. Phys. Chem.* 81.25 (1977), pp. 2340–2361.

[Gil91] Daniel T Gillespie. *Markov Processes: An Introduction For Physical Scientists*. Elsevier, 1991.

[Gil92] D.T. Gillespie. “A Rigorous Derivation Of The Chemical Master Equation”. In: *Physica A* 188.1-3 (Sept. 1992), pp. 404–425. ISSN: 03784371. DOI: 10.1016/0378-4371(92)90283-V.

[GLO05] Chetan Gadgil, Chang Hyeong Lee, and Hans G Othmer. “A Stochastic Analysis Of First-order Reaction Networks”. In: *Bulletin of Mathematical Biology* 67.5 (2005), pp. 901–946.

[GM12] D Ryan Georgianna and Stephen P Mayfield. “Exploiting Diversity And Synthetic Biology For The Production Of Algal Biofuels”. In: *Nature* 488.7411 (2012), p. 329.

[GM13] Ian Grooms and Andrew J Majda. “Efficient Stochastic Superparameterization For Geophysical Turbulence”. In: *Proceedings of the National Academy of Sciences* 110.12 (2013), pp. 4464–4469.

[GMB16] Marcella M Gomez, Richard M Murray, and Matthew R Bennett. “The Effects Of Time-varying Temperature On Delays In Genetic Networks”. In: *SIAM Journal on Applied Dynamical Systems* 15.3 (2016), pp. 1734–1752.

[GMK17] Ankit Gupta, Jan Mikelson, and Mustafa Khammash. “A Finite State Projection Algorithm For The Stationary Solution Of The Chemical Master Equation”. In: *arXiv* (2017).

[Goe+09] Lea Goentoro, Oren Shoval, et al. “The Incoherent Feedforward Loop Can Provide Fold-change Detection In Gene Regulation”. In: *Molecular Cell* 36.5 (2009), pp. 894–899.

[Gri+98] Regis Grimaud, Martin Kessel, et al. “Enzymatic And Structural Similarities Between The Escherichia coli ATP-dependent Proteases, ClpXP And ClpAP”. In: *Journal of Biological Chemistry* 273.20 (1998), pp. 12476–12481.

[GS01] Geoffrey Grimmett and David Stirzaker. *Probability And Random Processes*. Oxford University Press, 2001.

[GSN12] Ramon Grima, Deena R Schmidt, and Timothy J Newman. “Steady-state Fluctuations Of A Genetic Feedback Loop: An Exact Solution”. In: *The Journal of Chemical Physics* 137.3 (2012), p. 035104.

[Gun+05] Rudiyanto Gunawan, Yang Cao, et al. “Sensitivity Analysis Of Discrete Stochastic Systems”. In: *Biophysical Journal* 88.4 (2005), pp. 2530–2540.

[Har+14] Yuval Hart, Shlomit Reich-Zeliger, et al. “Paradoxical Signaling By A Secreted Molecule Leads To Homeostasis Of Cell Levels”. In: *Cell* 158.5 (2014), pp. 1022–1032.

[Has70] W. K. Hastings. “Monte Carlo Sampling Methods Using Markov Chains And Their Applications”. In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97.

[Hem+10] Christopher L Hemme, Housna Mouttaki, et al. “Sequencing Of Multiple Clostridial Genomes Related To Biomass Conversion And Biofuel Production”. In: *Journal of Bacteriology* 192.24 (2010), pp. 6494–6496.

[HG11] Markus Hegland and Jochen Garcke. “On The Numerical Solution Of The Chemical Master Equation With Sums Of Rank One Tensors”. In: *ANZIAM Journal* 52 (2011), pp. 628–643.

[Hig08] Desmond J. Higham. “Modeling And Simulating Chemical Reactions”. In: *SIAM Rev.* 50.2 (Jan. 2008), pp. 347–368. DOI: 10.1137/060666457.

[Hil10] Archibald Vivian Hill. “The Possible Effects Of The Aggregation Of The Molecules Of Haemoglobin On Its Dissociation Curves”. In: *J. Physiol.* 40 (1910), pp. 4–7.

[HME18] Patrick Harrigan, Hiten Madhani, and Hana El-Samad. “Real Time Genetic Compensation Operationally Defines The Dynamic Demands Of Feedback Control”. In: *bioRxiv* (2018), p. 244020.

[Hsi+14] Victoria Hsiao, Emmanuel LC de los Santos, et al. “Design And Implementation Of A Biomolecular Concentration Tracker”. In: *ACS Synthetic Biology* 4.2 (2014), pp. 150–161.

[Hsi+16] Victoria Hsiao, Yutaka Hori, et al. “A Population-based Temporal Logic Gate For Timing And Recording Chemical Events”. In: *Molecular Systems Biology* 12.5 (2016), p. 869.

[Ian+94] JM Iannelli, A Yariv, et al. “Stochastic Resonance In A Semiconductor Distributed Feedback Laser”. In: *Applied Physics Letters* 65.16 (1994), pp. 1983–1985.

[JH07] Tobias Jahnke and Wilhelm Huisinga. “Solving The Chemical Master Equation For Monomolecular Reaction Systems Analytically”. In: *Journal of Mathematical Biology* 54.1 (2007), pp. 1–26.

[JH08] Tobias Jahnke and Wilhelm Huisinga. “A Dynamical Low-rank Approach To The Chemical Master Equation”. In: *Bulletin of Mathematical Biology* 70.8 (2008), pp. 2283–2302.

[JI99] Miki Jishage and Akira Ishihama. “Transcriptional Organization And In Vivo Role Of The Escherichia coli Rsd Gene, Encoding The Regulator Of RNA Polymerase Sigma D”. In: *Journal of Bacteriology* 181.12 (1999), pp. 3768–3776.

[JU10] T. Jahnke and T. Udrescu. “Solving Chemical Master Equations By Adaptive Wavelet Compression”. In: *J. Comp. Phys.* 229.16 (2010), pp. 5724–5741. ISSN: 00219991. DOI: 10.1016/j.jcp.2010.04.015.

[KC10] Ahmad S Khalil and James J Collins. “Synthetic Biology: Applications Come Of Age”. In: *Nature Reviews Genetics* 11.5 (2010), p. 367.

[Ker+05] Gaetan Kerschen, Jean-claude Golinval, et al. “The Method Of Proper Orthogonal Decomposition For Dynamical Characterization And Order Reduction Of Mechanical Systems: An Overview”. In: *Nonlinear Dynamics* 41.1 (2005), pp. 147–169.

[KHM99] Sotirios C Kampranis, Alison J Howells, and Anthony Maxwell. “The Interaction Of DNA Gyrase With The Bacterial Toxin CcdB: Evidence For The Existence Of Two Gyrase-CcdB Complexes”. In: *Journal of Molecular Biology* 293.3 (1999), pp. 733–744.

[KRS15] Ivan Kryven, Susanna Röblitz, and Christof Schütte. “Solution Of The Chemical Master Equation By Radial Basis Functions Approximation With Interface Tracking”. In: *BMC Systems Biology* 9.1 (2015), p. 67.

[Kwo10] Roberta Kwok. “Five Hard Truths For Synthetic Biology”. In: *Nature News* 463.7279 (2010), pp. 288–290.

[KZH09] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. “Growth Rate-dependent Global Effects On Gene Expression In Bacteria”. In: *Cell* 139.7 (2009), pp. 1366–1375.

[Lil+17] Gabriele Lillacci, Stephanie K Aoki, et al. “A Synthetic Integral Feedback Controller For Robust Tunable Regulation In Bacteria”. In: *bioRxiv* (2017), p. 170951.

[Liu+12] Chang C Liu, Lei Qi, et al. “An Adaptor From Translational To Transcriptional Control Enables Predictable Assembly Of Complex Regulation”. In: *Nature Methods* 9.11 (2012), pp. 1088–1094.

[Ma+09] Wenzhe Ma, Ala Trusina, et al. “Defining Network Topologies That Can Achieve Biochemical Adaptation”. In: *Cell* 138.4 (2009), pp. 760–773.

[MA09] Mark D McDonnell and Derek Abbott. “What Is Stochastic Resonance? Definitions, Misconceptions, Debates, And Its Relevance To Biology”. In: *PLoS Computational Biology* 5.5 (2009), e1000348.

[MA97] Harley H McAdams and Adam Arkin. “Stochastic Mechanisms In Gene Expression”. In: *Proceedings of the National Academy of Sciences* 94.3 (1997), pp. 814–819.

[Man+06] Shmoolik Mangan, Shalev Itzkovitz, et al. “The Incoherent Feed-forward Loop Accelerates The Response-time Of The Gal System Of Escherichia coli”. In: *Journal of Molecular Biology* 356.5 (2006), pp. 1073–1081.

[McC+17] Reed D McCardell, Shan Huang, et al. “Control Of Bacterial Population Density With Population Feedback And Molecular Sequestration”. In: *bioRxiv* (2017), p. 225045.

[Men+17] X Flora Meng, Ania-Ariadna Baetica, et al. “Recursively constructing analytic expressions for equilibrium distributions of stochastic biochemical reaction networks”. In: *Journal of The Royal Society Interface* 14.130 (2017), p. 20170157. DOI: 10.1098/rsif.2017.0157.

[MFN15] B. Munsky, Z. Fox, and G. Neuert. “Integrating Single-molecule Experiments And Discrete Stochastic Models To Understand Heterogeneous Gene Transcription Dynamics”. In: *Methods* 85 (2015), pp. 12–21. ISSN: 10462023. DOI: 10.1016/j.ymeth.2015.06.009.

[MG09] Nuno C Martins and Jorge M Goncalves. “A Linear Programming Approach To Parameter Fitting For The Master Equation”. In: *IEEE Transactions on Automatic Control* 54.10 (2009), pp. 2451–2455.

[MG67] Donald A Mcquarrie and Daniel T Gillespie. “Stochastic Theory And Simulations Of Chemical Kinetics”. In: *J. Appl. Probab.* 478.1967 (1967), pp. 413–478. ISSN: 00219002. DOI: 10.2307/3212214.

[MHK14] Bence Mélykúti, Joao P Hespanha, and Mustafa Khammash. “Equilibrium Distributions Of Simple Biochemical Reaction Systems For Time-scale Separation In Stochastic Reaction Networks”. In: *Journal of The Royal Society Interface* 11.97 (2014), p. 20140054.

[MK06] Brian Munsky and Mustafa Khammash. “The Finite State Projection Algorithm For The Solution Of The Chemical Master Equation”. In: *The Journal of Chemical Physics* 124.4 (2006), p. 044104.

[MK07] Brian Munsky and Mustafa Khammash. “A Multiple Time Interval Finite State Projection Algorithm For The Solution To The Chemical Master Equation”. In: *J. Comput. Phys.* 226.1 (Sept. 2007), pp. 818–835. ISSN: 00219991. DOI: 10.1016/j.jcp.2007.05.016.

[MK08] B. Munsky and M. Khammash. “The Finite State Projection Approach For The Analysis Of Stochastic Noise In Gene Networks”. In: *IEEE Transactions on Automatic Control* 53.Special Issue (Jan. 2008), pp. 201–214. ISSN: 0018-9286. DOI: 10.1109/TAC.2007.911361.

[MNO12] Brian Munsky, Gregor Neuert, and Alexander van Oudenaarden. “Using Gene Expression Noise To Understand Gene Regulation”. In: *Science* 336.6078 (2012), pp. 183–187.

[Moo+12] Tae Seok Moon, Chunbo Lou, et al. “Genetic Programs Constructed From Layered Logic Gates In Single Cells”. In: *Nature* 491.7423 (2012), p. 249.

[MP15] Bence Mélykúti and Peter Pfaffelhuber. “The Stationary Distribution Of A Markov Jump Process Glued Together From Two State Spaces At Two Vertices”. In: *Stochastic Models* 31.4 (2015), pp. 525–553.

[MRD07] Hédia Maamar, Arjun Raj, and David Dubnau. “Noise In Gene Expression Determines Cell Fate In Bacillus subtilis”. In: *Science* 317.5837 (2007), pp. 526–529.

[MTK09] Brian Munsky, Brooke Trinh, and Mustafa Khammash. “Listening To The Noise: Random Fluctuations Reveal Gene Network Parameters”. In: *Mol. Syst. Biol.* 5.318 (2009), p. 318. ISSN: 1744-4292. DOI: 10.1038/msb.2009.75.

[MV03] Cleve Moler and Charles Van Loan. “Nineteen Dubious Ways To Compute The Exponential Of A Matrix, Twenty-five Years Later”. In: *SIAM review* 45.1 (2003), pp. 3–49.

[MVA00] Christopher K Mathews, Kensal E Van Holde, and Kevin G Ahern. *Biochemistry*. Add. Wesley Longman, San Francisco, 2000.

[Nar+16] Lauren Narcross, Elena Fossati, et al. “Microbial Factories For The Production Of Benzylisoquinoline Alkaloids”. In: *Trends in Biotechnology* 34.3 (2016), pp. 228–241.

[Neu+13] Gregor Neuert, Brian Munsky, et al. “Systematic Identification Of Signal-Activated Stochastic Gene Regulation”. In: *Science* 339.6119 (2013), pp. 584–587. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1231456.

[Ngu+15] Phuc HB Nguyen, Yong Wu, et al. “Design Space Exploration Of The Violacein Pathway In Escherichia coli Based Transcription Translation Cell-Free System (TX-TL)”. In: *bioRxiv* (2015), p. 027656.

[NHK12] Michael Nip, Joao P Hespanha, and Mustafa Khammash. “A Spectral Methods-based Solution Of The Chemical Master Equation For Gene Regulatory Networks”. In: *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*. IEEE. 2012, pp. 5354–5360.

[OJ16] Arturo V Orjalo and Hans E Johansson. “Stellaris® RNA Fluorescence In Situ Hybridization For The Simultaneous Detection Of Immature And Mature Long Noncoding RNAs In Adherent Cells”. In: *Long Non-Coding RNAs: Methods and Protocols* (2016), pp. 119–134.

[Ols+17] Noah Olsman, Ania-Ariadna Baetica, et al. “Hard limits and performance tradeoffs in a class of sequestration feedback systems”. In: *bioRxiv* (2017), p. 222042. DOI: 10.1101/222042.

[Par+14] Keith Pardee, Alexander A Green, et al. “Paper-based Synthetic Gene Networks”. In: *Cell* 159.4 (2014), pp. 940–954.

[Par03] Pablo A Parrilo. “Semidefinite Programming Relaxations For Semialgebraic Problems”. In: *Mathematical Programming* 96.2 (2003), pp. 293–320.

[PBE00] Johan Paulsson, Otto G Berg, and Måns Ehrenberg. “Stochastic Focusing: Fluctuation-enhanced Sensitivity Of Intracellular Regulation”. In: *Proceedings of the National Academy of Sciences* 97.13 (2000), pp. 7148–7153.

[PD05] Rodrigo B Platte and Tobin A Driscoll. “Polynomials And Potential Theory For Gaussian Radial Basis Function Interpolation”. In: *SIAM Journal on Numerical Analysis* 43.2 (2005), pp. 750–766.

[Pop02] Stephen B Pope. “Stochastic Lagrangian Models Of Velocity In Homogeneous Turbulent Shear Flow”. In: *Physics of Fluids* 14.5 (2002), pp. 1696–1702.

[Por+07] Stephanie Portle, Thomas B Causey, et al. “Cell Population Heterogeneity In Expression Of A Gene-switching Network With Fluorescent Markers Of Different Half-lives”. In: *Journal of Biotechnology* 128.2 (2007), pp. 362–375.

[PW09] Priscilla EM Purnick and Ron Weiss. “The Second Wave Of Synthetic Biology: From Modules To Systems”. In: *Nature Reviews Molecular Cell Biology* 10.6 (2009), p. 410.

[QD17] Yili Qian and Domitilla Del Vecchio. “Realizing “Integral Control” In Living Cells: How To Overcome Leaky Integration Due To Dilution?” In: *bioRxiv* (2017). DOI: 10.1101/141051.

[QD18] Yili Qian and Domitilla Del Vecchio. “Realizing ‘Integral Control’ In Living Cells: How To Overcome Leaky Integration Due To Dilution?” In: *Journal of The Royal Society Interface* 15.139 (2018). ISSN: 1742-5689. DOI: 10.1098/rsif.2017.0902.

[QMD17] Yili Qian, Cameron McBride, and Domitilla Del Vecchio. “Programming Cells To Work For Us”. In: *Annual Review of Control, Robotics, and Autonomous Systems* (2017).

[Raj+16] Karthik V Rajasekar, Konrad Zdanowski, et al. “The Anti-sigma Factor RsrA Responds To Oxidative Stress By Reburying Its Hydrophobic Core”. In: *Nature Communications* 7 (2016), p. 12194.

[REA02] Nitzan Rosenfeld, Michael B Elowitz, and Uri Alon. “Negative Autoregulation Speeds The Response Times Of Transcription Networks”. In: *Journal of Molecular Biology* 323.5 (2002), pp. 785–793.

[Ren+17] Xinying Ren, Ania-Ariadna Baetica, et al. “Population regulation in microbial consortia using dual feedback control”. In: *bioRxiv* (2017). DOI: 10.1101/120253.

[RO08] Arjun Raj and Alexander van Oudenaarden. “Nature, Nurture, Or Chance: Stochastic Gene Expression And Its Consequences”. In: *Cell* 135.2 (2008), pp. 216–226.

[Rus+07] Michael J Rust, Joseph S Markson, et al. “Ordered Phosphorylation Governs Oscillation Of A Three-protein Circadian Clock”. In: *Science* 318.5851 (2007), pp. 809–812.

[SC08] Umender K Sharma and Dipankar Chatterji. “Differential Mechanisms Of Binding Of Anti-sigma Factors Escherichia coli Rsd And Bacteriophage T4 AsiA To E. coli RNA Polymerase Lead To Diverse Physiological Consequences”. In: *Journal of Bacteriology* 190.10 (2008), pp. 3434–3443.

[Sch+10] D Schittler, J Hasenauer, et al. “Cell Differentiation Modeled via A Coupled Two-switch Regulatory Network”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20.4 (2010), p. 045121.

[Sch07] R Schaback. “A Practical Guide To Radial Basis Functions”. In: *Electronic Resource* 11 (2007).

[Sco+17] Spencer R Scott, M Omar Din, et al. “A Stabilized Microbial Ecosystem Of Self-limiting Bacteria Using Synthetic Quorum-regulated Lysis”. In: *Nature Microbiology* 2 (2017), nmicrobiol201783.

[Sha+04] Nathan C Shaner, Robert E Campbell, et al. “Improved Monomeric Red, Orange And Yellow Fluorescent Proteins Derived From Discosoma sp. Red Fluorescent Protein”. In: *Nature Biotechnology* 22.12 (2004), pp. 1567–1572.

[Sha05] Howard M Shapiro. *Practical Flow Cytometry*. John Wiley & Sons, 2005.

[She+13] Douglas P. Shepherd, Nan Li, et al. “Counting Small RNA In Pathogenic Bacteria”. In: *Anal. Chem.* 85 (2013), pp. 4938–4943. ISSN: 00032700. DOI: 10.1021/ac303792p.

[Sim+08] Michael L Simpson, Michael J Paulus, et al. *Microluminometer Chip And Method To Measure Bioluminescence*. US Patent 7,371,538. May 2008.

[SM12] Adi Steif and Irmtraud M Meyer. “The Hok mRNA Family”. In: *RNA Biology* 9.12 (2012), pp. 1399–1404.

[SS08] Vahid Shahrezaei and Peter S Swain. “Analytical Distributions For Stochastic Gene Expression”. In: *Proceedings of the National Academy of Sciences* 105.45 (2008), pp. 17256–17261.

[SWS08] David F Savage, Jeffrey Way, and Pamela A Silver. *Defossiling Fuel: How Synthetic Biology Can Transform Biofuel Production*. 2008.

[TB06] Tianhai Tian and Kevin Burrage. “Stochastic Models For Regulatory Networks Of The Genetic Toggle Switch”. In: *Proceedings of the National Academy of Sciences* 103.22 (2006), pp. 8372–8377.

[TFM12] Jose Juan Tapia, James R. Faeder, and Brian Munsky. “Adaptive Coarse-graining For Transient And Quasi-equilibrium Analyses Of Stochastic Gene Regulation”. In: *2012 IEEE 51st IEEE Conf. Decis. Control* 836 (2012), pp. 5361–5366. ISSN: 01912216. DOI: 10.1109/CDC.2012.6425828.

[Thu17] Anupama Thubagere Jagadeesh. “Programming Complex Behavior In DNA-based Molecular Circuits And Robots”. PhD thesis. California Institute of Technology, 2017.

[Van92] Nicolaas Godfried Van Kampen. *Stochastic Processes In Physics And Chemistry*. Vol. 1. Elsevier, 1992.

[Vo17] Huy D Vo. *Krylov Approximation And Model Reduction Methods For Solving The Chemical Master Equation*. The University of Alabama, 2017.

[VSK08] Jan-Willem Veening, Wiep Klaas Smits, and Oscar P Kuipers. “Bistability, Epigenetics, And Bet-hedging In Bacteria”. In: *Annu. Rev. Microbiol.* 62 (2008), pp. 193–210.

[Wak+05a] H Waki, S Kim, et al. “SparsePOP: A Sparse Semidefinite Programming Relaxation Of Polynomial Optimization Problems, Dept”. In: *Math. Comp. Sci., Tokyo Inst. Tech., Research Report B-414* (2005).

[Wak+05b] H Waki, S Kim, et al. “Sums Of Squares And Semidefinite Programming Relaxations For Polynomial Optimization Problems With Structured Sparsity”. In: *SIAM Journal on Optimization* (2005).

[Wal+02] S Patrick Walton, Gregory N Stephanopoulos, et al. “Thermodynamic And Kinetic Characterization Of Antisense Oligodeoxynucleotide Binding To A Structured mRNA”. In: *Biophysical Journal* 82.1 (2002), pp. 366–377.

[Wer10] Jürgen Werner. “System Properties, Feedback Control And Effector Coordination Of Human Temperature Regulation”. In: *European Journal of Applied Physiology* 109.1 (2010), pp. 13–25.

[XB04] Lin Xiao and Stephen Boyd. “Fast Linear Iterations For Distributed Averaging”. In: *Systems & Control Letters* 53.1 (2004), pp. 65–78.

[XD18] Fangzhou Xiao and John C Doyle. "Robust Perfect Adaptation In Biomolecular Reaction Networks". In: *bioRxiv* (2018). DOI: 10.1101/299057.

[Yeu+14] Enoch Yeung, Andrew Ng, et al. "Modeling The Effects Of Compositional Context On Promoter Activity In An E. coli Extract Based Transcription-translation System". In: *Decision and Control (CDC), 2014 IEEE 53rd Annual Conference on*. IEEE. 2014, pp. 5405–5412.

[Yeu+17] Enoch Yeung, Aaron J Dy, et al. "Biophysical Constraints Arising From Compositional Context In Synthetic Gene Networks". In: *Cell Systems* 5.1 (2017), pp. 11–24.

[Yi+00] Tau-Mu Yi, Yun Huang, et al. "Robust Perfect Adaptation In Bacterial Chemotaxis Through Integral Feedback Control". In: *Proceedings of the National Academy of Sciences* 97.9 (2000), pp. 4649–4653.

[You+04] Lingchong You, Robert Sidney Cox III, et al. "Programmed Population Control By Cell-cell Communication And Regulated Killing". In: *Nature* 428.6985 (2004), p. 868.

[ZG98] Yanning Zhou and Susan Gottesman. "Regulation Of Proteolysis Of The Stationary-phase Sigma Factor RpoS". In: *Journal of Bacteriology* 180.5 (1998), pp. 1154–1158.

[Zha+10] Jingwei Zhang, Layne T. Watson, et al. "Radial Basis Function Collocation For The Chemical Master Equation". In: *Int. J. Comput. Met.* 07.03 (2010), pp. 477–498. DOI: 10.1142/S0219876210002234.

Appendix A

THEOREM PROOFS FOR THE DESIGN OF STOCHASTIC BIOCHEMICAL REACTION NETWORKS

Theorem 2.1. *Consider the system of two interconnected transcriptional components that are modeled by reactions as given in equation (2.2), where* $\kappa > 0$, $\delta > 0$, $\kappa_{on} > 0$, *and* $\kappa_{off} > 0$ *are the corresponding reaction rate constants. Let P, Z, and C be the numbers of promoters, transcription factors, and complexes, respectively. Let* $\alpha = \frac{\kappa\kappa_{on}}{\delta\kappa_{off}}$, $\beta = \frac{\kappa}{\delta}$, *and* $\gamma = \frac{N\alpha-1}{\alpha+1}$, *where N is a constant given by* $N = P + C$ *due to the conservation of DNA. In (i)–(iii), we set up and solve three design problems using the marginal stationary distributions of Z and C. Here,* α *and* β *are the design variables.*

(i) Since the marginal stationary distribution of Z is a Poisson distribution, its mean and variance are equal. The design problem of fixing the mean of Z at an objective value $\mu_z > 0$ *is feasible, and the solution is* $\beta = \mu_z$, *with N and the reaction rate constants being arbitrary otherwise.*

(ii) The design problem of setting the mean of C at an objective value $\mu_c \in (0, N)$ *is feasible, and the solution is* $\alpha = \frac{\mu_c}{N-\mu_c}$, *with N and the reaction rate constants being arbitrary otherwise.*

(iii) The design problem of choosing the variance of C to be an objective value $\sigma_c^2 > 0$ *is feasible if and only if* $\sigma_c^2 \leq \frac{N}{4}$, *and the solutions are* $\alpha = \frac{N-2\sigma_c^2 \pm \sqrt{N^2-4N\sigma_c^2}}{2\sigma_c^2}$, *with N and the reaction rate constants being arbitrary otherwise.*

Proof. Using the analytical solution of the joint probability distribution of species Z and C in equation (2.2), we prove that the marginal stationary distribution of species Z is a Poisson distribution with mean and variance equal to β and that the marginal stationary distribution of species C is a binomial distribution with the number of trials and the success probability in each trial being N and $\frac{\alpha}{1+\alpha}$, respectively.

First, we use the new notation to restate the joint probability distribution as

$$P(c, z) = (1 + \alpha)^{-N} \alpha^c \binom{N}{c} e^{-\beta} \frac{\beta^z}{z!} \tag{A.1}$$

for $c \in \{0, \ldots, N\}$ and $z, N \in \mathbb{Z}_{\geq 0}$.

We can compute the marginal stationary distribution P_1 of species Z as

$$P_1(z) = \sum_{c=0}^{N} (1+\alpha)^{-N} \alpha^c \binom{N}{c} e^{-\beta} \frac{\beta^z}{z!} \tag{A.2}$$

for $z, N \in \mathbb{Z}_{\geq 0}$.

By rearranging the terms, we rewrite the marginal distribution P_1 as

$$P_1(z) = e^{-\beta} \frac{\beta^z}{z!} \sum_{c=0}^{N} \binom{N}{c} \left(\frac{\alpha}{1+\alpha}\right)^c \left(\frac{1}{1+\alpha}\right)^{N-c} \tag{A.3}$$

for $z, N \in \mathbb{Z}_{\geq 0}$.

Since we are summing over the probability of getting exactly $c \in \{0, \ldots, N\}$ successes in N trials of probability $\frac{\alpha}{1+\alpha}$, this is equivalent to computing the probability mass of the binomial distribution, which equals 1. Thus, the marginal distribution of species Z is

$$P_1(z) = e^{-\beta} \frac{\beta^z}{z!}, \tag{A.4}$$

which represents the Poisson distribution of mean and variance equal to β.

We then prove that the marginal stationary distribution P_2 of species C is a binomial distribution with the number of trials and the success probability in each trial being N and $\frac{\alpha}{1+\alpha}$, respectively. By definition, the marginal distribution of species C is

$$P_2(c) = (1+\alpha)^{-N} \alpha^c \binom{N}{c} \sum_{z=0}^{\infty} e^{-\beta} \frac{\beta^z}{z!} \tag{A.5}$$

for $c \in \{0, \ldots, N\}$.

Since we are summing over the Poisson distribution of mean β, this is equivalent to computing its mass, which equals 1. Thus, the marginal stationary distribution P_2 is the binomial distribution

$$P_2(c) = \binom{N}{c} \left(\frac{\alpha}{1+\alpha}\right)^c \left(\frac{1}{1+\alpha}\right)^{N-c} \tag{A.6}$$

for $c \in \{0, \ldots, N\}$.

In design problem (i), setting the mean and variance of Z at $\mu_z > 0$ is equivalent to specifying $\beta = \mu_z$, which is always feasible.

Design problem (ii) corresponds to setting $\frac{N\alpha}{1+\alpha} = \mu_c$. A solution exists if and only if $0 < \mu_c < N$, in which case the solution is $\alpha = \frac{\mu_c}{N-\mu_c}$.

Design problem (iii) is equivalent to solving $\frac{N\alpha}{(1+\alpha)^2} = \sigma_c^2$, leading to solution $\alpha = \frac{N-2\sigma_c^2 \pm \sqrt{N^2-4N\sigma_c^2}}{2\sigma_c^2}$ when $0 < \sigma_c^2 \leq \frac{N}{4}$. □

Proposition 2.2. *Consider the system of two interconnected transcriptional components that are modeled by the reactions in equation (2.2). With the same notation as in Theorem 2.1, the stationary distribution in equation (2.2) has a unique global maximum if and only if $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, and $\beta, \gamma \notin \mathbb{Z}$. In this case, the maximum is at $(c^*, z^*) = (\lfloor \gamma \rfloor + 1, \lfloor \beta \rfloor)$.*

Proof. Let $\|\cdot\|$ denote the l^1-norm on $\mathbb{R}^2$. For $x \geq 0$, let $\lfloor x \rfloor$ denote the integer part of x. Since N is the total number of promoters P and complex molecules C, it is reasonable to assume that $N > 1$. The sample space of the probability mass function in equation (2.2) is $\Omega = \{(c, z) \mid c = 0, 1, \ldots, N \text{ and } z \in \mathbb{Z}_{\geq 0}\}$. Let Ω_∂ denote the boundary of Ω, namely, $\Omega_\partial = (\{0, N\} \times \mathbb{Z}_{\geq 0}) \cup (\{0, 1, \ldots, N\} \times 0)$. The probability mass in equation (2.2) has a strict local maximum at $(c^*, z^*) \in \Omega \setminus \Omega_\partial$ if and only if, for all $(c, z) \in \Omega$ with $\|(c - c^*, z - z^*)\| \leq 1$, we have $\Pr(c, z) < \Pr(c^*, z^*)$. Solving these inequalities simultaneously gives $\gamma < c^* < \gamma + 1$ and $\beta - 1 < z^* < \beta$. Since $(c^*, z^*) \in \Omega \setminus \Omega_\partial$, a unique solution exists if and only if $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, and $\beta, \gamma \notin \mathbb{Z}$. When these conditions hold, the unique strict local maximum on $\Omega \setminus \Omega_\partial$ is at $(c^*, z^*) = (\lfloor \gamma \rfloor + 1, \lfloor \beta \rfloor)$. Moreover, basic algebra shows that, for all $c \in \{0, 1, \ldots, N\}$, we have $\Pr(c, z) < \Pr(c, z + 1)$ for all $z \in \{0, 1, \ldots, z^* - 1\}$ and $\Pr(c, z) > \Pr(c, z + 1)$ for all integers $z \geq z^*$. It is also straightforward to verify that $\Pr(c, z^*) < \Pr(c + 1, z^*)$ for all integers $c \in \{0, 1, \ldots, c^* - 1\}$, and $\Pr(c, z^*) > \Pr(c + 1, z^*)$ for all integers $c \in \{c^*, c^* + 1, \ldots, N\}$. Therefore, the stationary distribution of the two-component transcriptional system has a unique global maximum at (c^*, z^*) if and only if $N > 1$, $\beta > 1$, $0 < \gamma < N - 1$, and $\beta, \gamma \notin \mathbb{Z}$. □

We include additional figures to complement the chemical reaction network results obtained using our stochastic design framework in Section 2.4. In Figure A.1, we

demonstrate that the transients approach the pre-specified stationary distribution for the protein production-degradation reaction network with the reaction rate values found by our framework. In Figure A.2, we plot an example of a projection operator that constraints a corresponding transient distribution to be bimodal through the inequality in equation (2.11).

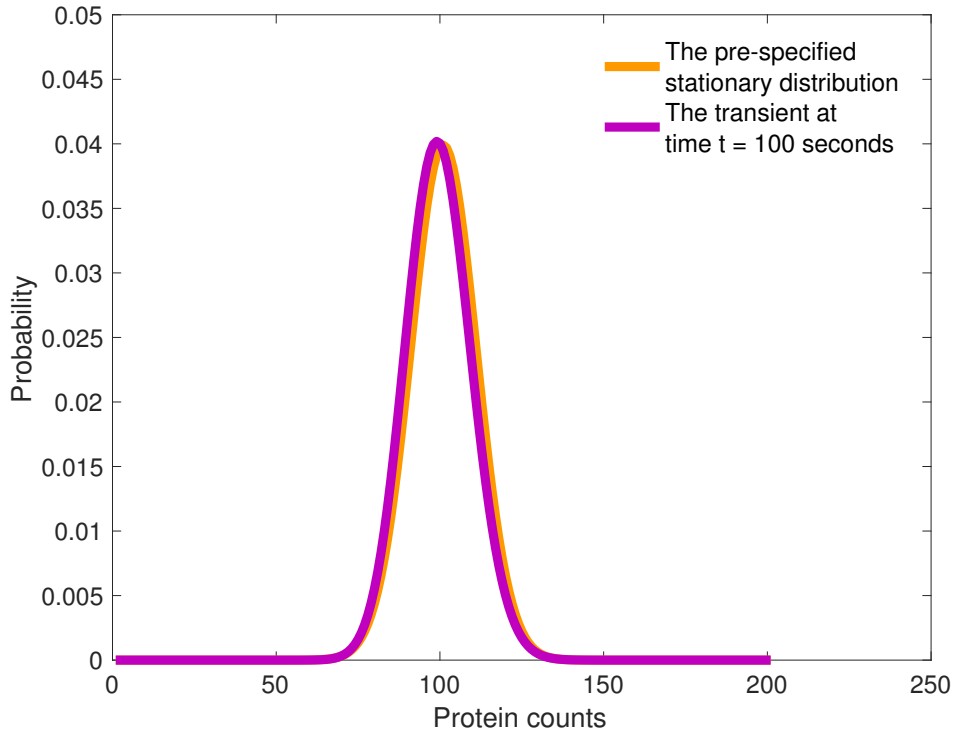

Figure A.1: **Solution to the design problem for a protein production-degradation reaction network.** We compare the transient distribution of the birth-death process in chemical reactions 2.12 after 100 seconds (orange) to the pre-specified stationary distribution (mauve). We find a very good match between the two probability distributions.

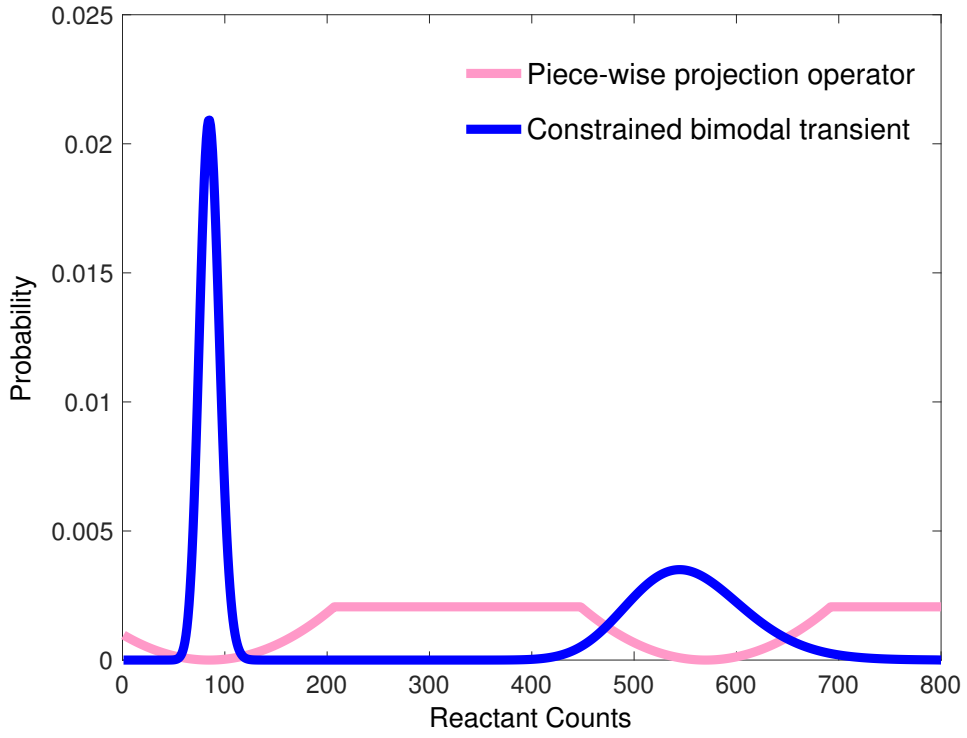

Figure A.2: **A piece-wise projection operator that constrains a corresponding transient distribution to be bimodal.** We plot an example of a piece-wise projection operator with two local minima (pink). This operator constrains a corresponding distribution to be bimodal (blue), according to the inequality in equation (2.11). There are many other choices of projection operators that result in corresponding bimodal transient distributions.

Appendix B

THEOREM PROOFS FOR SEQUESTRATION FEEDBACK NETWORKS

B.1 Sequestration feedback networks with no controller species degradation

We provide the following model for the two controller species, according to Figure 3.3:

$$\begin{aligned} \dot{z}_1 &= \mu - \eta z_1 z_2, \\ \dot{z}_2 &= \theta_2 x_2 - \eta z_1 z_2. \end{aligned} \tag{B.1}$$

For a complete model of the sequestration feedback network, we refer the reader to the system of equations (B.9).

We set the reference of the tracking problem at value $\frac{\mu}{\theta_2}$ and we define the error signal as $e(t) = \frac{\mu}{\theta_2} - x_n(t)$.

Then the controller species implement integral control since

$$\frac{d}{dt}(z_1(t) - z_2(t)) = \theta_2 e(t). \tag{B.2}$$

The control action $z_1(t) - z_2(t)$ integrates the error signal $e(t)$ as follows:

$$z_1(t) - z_2(t) = \theta_2 \int_0^t (e(s))\, ds. \tag{B.3}$$

If the sequestration feedback network is stable, then the model has a steady state and the integral controller ensures that the property of perfect adaptation holds since

$$\frac{dz_1(t)}{dt} = \frac{dz_2(t)}{dt} = 0 \implies x_n^{ss} = \frac{\mu}{\theta_2}. \tag{B.4}$$

Perfect adaptation is a desirable property of the sequestration feedback system because it allows for a variety of process network dynamics. We investigate whether this property is retained by including the controller species degradation in the model.

B.2 Sequestration feedback networks with controller species degradation

Incorporating the degradation of the controller species in the model description changes the controller equations (B.1) to the following:

$$\begin{aligned} \dot{z}_1 &= \mu - \eta z_1 z_2 - \gamma_c z_1, \\ \dot{z}_2 &= \theta_2 x_2 - \eta z_1 z_2 - \gamma_c z_2. \end{aligned} \tag{B.5}$$

The resulting controller is a lag compensator that integrates the error signal weighed by an exponential of the controller degradation rate:

$$\frac{d}{dt}(z_1(t) - z_2(t)) = \mu - \theta_2 x_n(t) - \gamma_c(z_1(t) - z_2(t)), \tag{B.6}$$

$$z_1(t) - z_2(t) = \theta_2 \int_0^t e^{\gamma_c(s-t)} \left(\frac{\mu}{\theta_2} - x_n(t) \right) ds. \tag{B.7}$$

With the same notation for the error signal, we obtain that:

$$z_1(t) - z_2(t) = \theta_2 \int_0^t e^{\gamma_c(s-t)} e(s) ds. \tag{B.8}$$

The exponential of the degradation rate biases the error measurement towards recent past over the distant past since $0 \leq s \leq t$.

Integral control has the property of perfect adaptation, provided that the closed loop system in stable. This implies that the closed loop system has zero steady state error. The lag compensator can also exhibit zero steady state error for a second controller degradation rate, as demonstrated in Section B.3.

B.3 The critical controller species degradation rate

The closed loop system can have zero steady state error for no controller species degradation, as well as for a value of the controller degradation rate, which we refer to as "critical". Assuming stability, zero controller species degradation guarantees perfect adaptation of the closed loop system, but can be challenging to implement. In this section, we derive a value of the critical controller species degradation rate such that the steady state error of a general stable sequestration feedback network with n process species equals zero. Depending on the parameters of the sequestration feedback system, the critical value of the degradation rate may or may not be achievable.

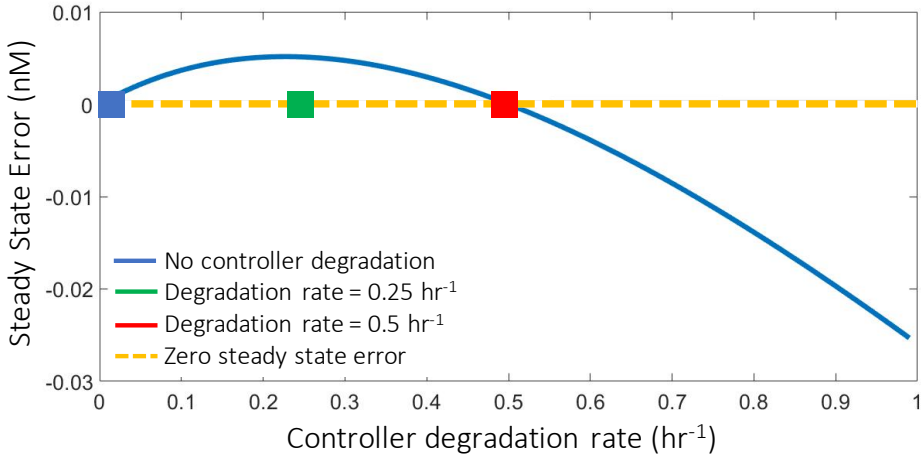

Figure B.1: **The steady state error as a function of the controller degradation rate.** For an example sequestration feedback network with only two process species, we plot the steady state error (blue line) as a function of the controller degradation rate, while keeping the other parameters of the network fixed. We obtain perfect adaptation for no controller degradation (blue square) and zero steady state error at the critical controller degradation rate (red square). The steady state error is nonzero at other controller species degradation rates (green square). Zero and the critical controller species degradation rate are the only two degradation rate values for which the steady state error equals zero (orange dashed line). This property holds for a general class of sequestration feedback networks, as discussed in Theorem B.1. For a given sequestration feedback network, we can analytically compute the steady state error as a function of the controller species degradation rate from equation (B.15).

We plot the steady state error as a function of the controller species degradation rate for an example sequestration feedback network with two process species in Figure B.1. When the controller degradation rate has value zero, then the steady state error is also zero. Similarly, at the critical controller degradation rate, the steady state error equals zero. Indeed, the degradation rate values of zero and of the critical degradation rate are the only ones for which the network tracks the reference with zero steady state error. In this section, we determine the conditions under which the critical degradation rate exists and how to compute it.

We consider the general deterministic sequestration feedback network with n process species, as illustrated in Figure 3.3 (i.e. the n process species can be reactants in any bimolecular reactions within the process network). Then the model of its dynamics

η	γ_p	μ	k_1	θ_1	θ_2
1 nM^{-1}	1.6 hr^{-1}	8 nM hr^{-1}	1 hr^{-1}	2 hr^{-1}	4 hr^{-1}

Table B.1: **The parameters used for the simulation in Figure 3.5**. The critical controller degradation rate is $\gamma_c = 0.56$ hr^{-1}. The reference concentration is 2 nM. The initial concentrations of the four species in the sequestration feedback network are 1 nM.

is given by the following system of equations:

$$\begin{aligned}
\frac{dx_1}{dt} &= \theta_1 z_1 + \alpha_{1,1} x_1 + \cdots + \alpha_{1,n} x_n, \\
\frac{dx_2}{dt} &= \alpha_{2,1} x_1 + \cdots + \alpha_{2,n} x_n, \\
&\vdots \\
\frac{dx_n}{dt} &= \alpha_{n,1} x_1 + \cdots + \alpha_{n,n} x_n, \\
\frac{dz_1}{dt} &= \mu - \eta z_1 z_2 - \gamma_c z_1, \\
\frac{dz_2}{dt} &= \theta_2 x_n - \eta z_1 z_2 - \gamma_c z_2.
\end{aligned} \tag{B.9}$$

We define the following notation:

$$A = \begin{pmatrix} \alpha_{2,1} & \alpha_{2,2} & \dots & \alpha_{2,n-1} \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n,1} & \alpha_{n,2} & \dots & \alpha_{n,n-1} \end{pmatrix},$$

$\alpha_1 = (\alpha_{1,1}, \dots, \alpha_{1,n-1})$, $\alpha_n = (\alpha_{2,n}, \dots, \alpha_{n,n})^T$, and $\Gamma = \theta_1^{-1}(\alpha_1 A^{-1} \alpha_n - \alpha_{1,n})$.

Theorem B.1. *The critical controller degradation rate of a general sequestration feedback network with n process species is given by*

$$\gamma_c = \frac{\theta_2}{\Gamma} - \frac{\Gamma \eta \mu}{\theta_2} \tag{B.10}$$

and it only exists if and only if the closed loop system is stable and $\Gamma < \frac{\theta_2}{\sqrt{\eta\mu}}$, *where* $\Gamma = \theta_1^{-1}(\alpha_1 A^{-1} \alpha_n - \alpha_{1,n})$.

Proof. In a 1-node sequestration feedback system, the degradation rate

$$\gamma_c = \frac{\theta_1 \theta_2}{\gamma_p} - \frac{\gamma_p \eta \mu}{\theta_1 \theta_2}$$

results in zero steady state error. This degradation rate value can only be achieved when $\gamma_p < \frac{\theta_1\theta_2}{\sqrt{\mu\eta}}$.

Assuming that the closed loop system is stable, at equilibrium, equation (B.9) reduces to

$$\begin{aligned} 0 &= \theta_1 z_1 + \alpha_{1,1}x_1 + \cdots + \alpha_{1,n}x_n \\ 0 &= \alpha_{2,1}x_1 + \cdots + \alpha_{2,n}x_n \\ &\vdots \\ 0 &= \alpha_{n,1}x_1 + \cdots + \alpha_{n,n}x_n \\ \mu &= \eta z_1 z_2 + \gamma_c z_1, \\ \theta_2 x_n &= \eta z_1 z_2 + \gamma_c z_2. \end{aligned} \tag{B.11}$$

Our system of equations in (B.11) reduces to

$$\begin{aligned} z_1 &= \Gamma x_n, \\ \mu &= \eta z_1 z_2 + \gamma_c z_1, \\ \theta x_n &= \eta z_1 z_2 + \gamma_c z_2. \end{aligned} \tag{B.12}$$

First, it must be the case that constant $\Gamma > 0$ and matrix A is invertible. Otherwise, the system cannot have a positive steady state. This is equivalent to $\alpha_1 A^{-1}\alpha_n > \alpha_{1,n}$. The input species should not be depleted to create the output species. The system in equation (B.12) simplifies to a single equation

$$x_n^2(\Gamma\eta\theta_2 + \Gamma^2\eta\gamma_c) + x_n(\Gamma\gamma_c^2 - \Gamma\mu\eta) - \gamma_c\mu = 0. \tag{B.13}$$

Equation (B.13) always has a positive solution

$$x_n = \frac{(\Gamma\mu\eta - \Gamma\gamma_c^2) + \sqrt{(\Gamma\gamma_c^2 - \Gamma\mu\eta)^2 + 4\gamma_c\mu(\Gamma\eta\theta_2 + \Gamma^2\gamma_c\eta)}}{2(\Gamma\eta\theta_2 + \Gamma^2\gamma_c\eta)}. \tag{B.14}$$

Thus, the steady state error signal is

$$e = \frac{-(\Gamma\gamma_c^2 + 2\Gamma^2\gamma_c\frac{\eta\mu}{\theta_2} + \Gamma\mu\eta) + \sqrt{(\Gamma\gamma_c^2 - \Gamma\mu\eta)^2 + 4\gamma_c\mu(\Gamma\eta\theta_2 + \Gamma^2\gamma_c\eta)}}{2(\Gamma^2\gamma_c\eta + \Gamma\eta\theta_2)}. \tag{B.15}$$

If we want the output of the dynamical system to follow the reference signal $\frac{\mu}{\theta_2}$, then it must be that

$$\gamma_c = \frac{\theta_2}{\Gamma} - \frac{\Gamma\eta\mu}{\theta_2}, \tag{B.16}$$

which can only be achieved if and only if $\Gamma < \frac{\theta_2}{\sqrt{\eta\mu}}$. □

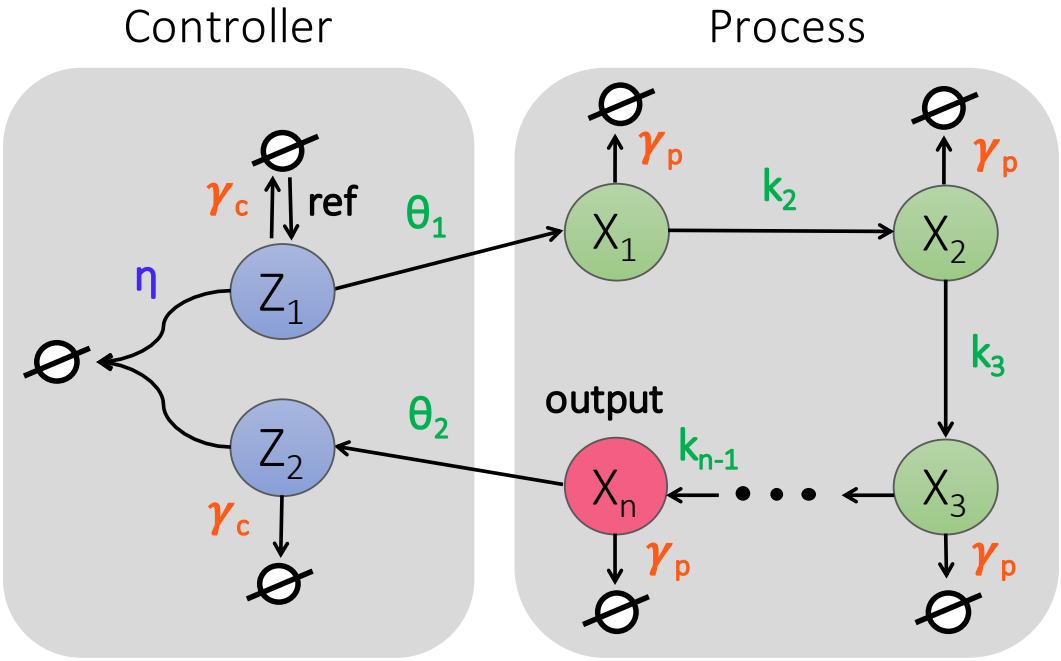

Figure B.2: **A sequestration feedback network with a simplified process.** The diagram matches Figure 3.3, but the process network is simplified to only allow bimolecular chemical reactions between species with consecutive numbering; each process species X_i is created by the previous process species X_{i-1} for $2 \leq i \leq n-1$. For simplicity, the process species degradation rates are assumed to be equal and represented as γ_p (orange). Similarly, the controller species degradation rates are assumed to be equal and represented as γ_c (orange). The sequestration reaction rate η is illustrated in blue and the process species production rates $\theta_1, \theta_2, k_1, \ldots, k_{n-1}$ are illustrated in green. A process network that follows these assumptions implements the process species as proteins subjected to dilution inside a bacterial cell.

Theorem 3.2. *The critical controller degradation rate for a simplified process network (i.e. each process species X_i is created by the previous process species X_{i-1} and creates the next process species X_{i+1}, $\forall 2 \leq i \leq n-1$, as in Figure B.2) is:*

$$\gamma_c = \frac{\theta_1\theta_2 \prod_{i=1}^{n-1} k_i}{\gamma_p^n} - \frac{\eta\mu\gamma_p^n}{\theta_1\theta_2 \prod_{i=1}^{n-1} k_i}. \tag{3.8}$$

It can be achieved if and only if

$$\gamma_p < \sqrt[n]{\frac{\theta_1\theta_2 \prod_{i=1}^{n-1} k_i}{\sqrt{\eta\mu}}}. \tag{3.9}$$

Proof. In this particular case, the matrix

$$A^{-1} = \begin{pmatrix} \frac{1}{k_2} & \frac{\gamma_p}{k_2 k_3} & \cdots & \frac{\gamma_p^{n-2}}{k_2 \ldots k_{n-1} \theta_2} \\ 0 & \frac{1}{k_3} & \cdots & \frac{\gamma_p^{n-3}}{k_3 \ldots k_{n-1} \theta_2} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\theta_2} \end{pmatrix},$$

the vectors $\beta_1 = (-\gamma_p, 0, \ldots, 0)$, $\beta_n = (0, \ldots, 0, -\gamma_p)^T$, $\beta_{1,n} = 0$, and the expression $\Gamma = \frac{\gamma_p^n}{\prod_{i=1}^{n-1} k_i \theta_2}$. Hence, the critical controller degradation rate is

$$\gamma_c = \frac{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}{\gamma_p^n} - \frac{\eta \mu \gamma_p^n}{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}. \tag{B.17}$$

The critical controller degradation rate can be achieved if and only if

$$\gamma_p < \sqrt[n]{\frac{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}{\sqrt{\eta \mu}}}. \tag{B.18}$$

□

B.4 Stability analysis of the sequestration feedback network with controller species degradation

In this section, we derive the stability criterion for the sequestration feedback network with controller degradation. The stability criterion depends on comparing the process and the controller degradation rates. To assess the stability of the sequestration feedback network, we derive the characteristic polynomial associated with its linearization.

First, we clarify our model assumptions. We assume that all process species degrade at the same rate and that the two controller species also degrade at the same rate. Additionally, we also assume a simplified process structure, as in Figure B.2 because otherwise it may be difficult to analytically derive the characteristic polynomial and its roots. For a more detailed discussion, see [Ols+17]. Thus, our model equations

are as follows:

$$\begin{aligned}
\frac{dx_1}{dt} &= \theta_1 z_1 - \gamma_p x_1 \\
\frac{dx_2}{dt} &= k_1 x_1 - \gamma_p x_2 \\
&\vdots \\
\frac{dx_n}{dt} &= k_{n-1} x_{n-1} - \gamma_p x_n \\
\frac{dz_1}{dt} &= \mu - \eta z_1 z_2 - \gamma_c z_1 \\
\frac{dz_2}{dt} &= \theta_2 x_n - \eta z_1 z_2 - \gamma_c z_2.
\end{aligned} \tag{B.19}$$

We introduce the following notation:

$$\alpha = \frac{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}{\gamma_p^n},$$

$$\beta = \eta \mu.$$

Then the linearized dynamics of the sequestration feedback network can be expressed according to the block matrix $M \in \mathbb{R}_{(n+2)\times(n+2)}$, where $A \in \mathbb{R}_{n\times n}$, $B \in \mathbb{R}_{n\times 2}$, $C \in \mathbb{R}_{2\times n}$, and $D \in \mathbb{R}_{2\times 2}$. Here,

$$M = \left[\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right],$$

$$A = \begin{bmatrix} -\gamma_p & 0 & \cdots & 0 \\ k_1 & -\gamma_p & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & \cdots & k_{n-1} & -\gamma_p \end{bmatrix}, B = \begin{bmatrix} \theta_1 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix},$$

$$C = \begin{bmatrix} 0 & \cdots & 0 \\ 0 & \cdots & \theta_2 \end{bmatrix}, D = \begin{bmatrix} -\eta z_2^* - \gamma_c & -\eta z_1^* \\ -\eta z_2^* & -\eta z_1^* - \gamma_c \end{bmatrix}.$$

where z_1^* and z_2^* are the steady state values of controller species z_1 and z_2. The steady state values of the controller species satisfy the equations:

$$\begin{aligned}
\mu &= \eta z_1^* z_2^* + \gamma_c z_1^*, \\
\alpha z_1^* &= \eta z_1^* z_2^* + \gamma_c z_2^*.
\end{aligned} \tag{B.20}$$

We can solve this system of equations and obtain that

$$z_1^* = \frac{\beta - \gamma_c^2 + \sqrt{(\beta - \gamma_c^2)^2 + 4\beta\gamma_c(\alpha + \gamma_c)}}{2\eta(\alpha + \gamma_c)}, z_2^* = \frac{\mu - \gamma_c z_1^*}{\eta z_1^*}. \tag{B.21}$$

By performing block matrix manipulations, we can compute the characteristic polynomial of matrix M as follows:

$$\begin{aligned} p(s) =& \det(s\mathbb{I} - M) = \det(s\mathbb{I} - A)\det[(s\mathbb{I} - D) - C(s\mathbb{I} - A)^{-1}B] \\ =&(s + \gamma_p)^n \left[(s + \eta z_1^* + \gamma_c)(s + \eta z_2^* + \gamma_c) - \eta^2 z_1^* z_2^*\right] + \alpha\eta z_1^* \gamma_p^n, \end{aligned} \tag{B.22}$$

where $\mathbb{I}$ is the identity matrix of dimensions $(n + 2) \times (n + 2)$.

By substituting in the values of z_1^* and z_2^* from equation (B.21), we obtain that

$$\begin{aligned} p(s) =&(s + \gamma_p)^n \left[s^2 + s\left(\frac{2\alpha\beta}{\beta - \gamma_c^2 + \sqrt{\Delta_n}} + \frac{\alpha\gamma_c + \beta}{\alpha + \gamma_c}\right) + \left(\frac{2\alpha\beta\gamma_c}{\beta - \gamma_c^2 + \sqrt{\Delta_n}} - \frac{\gamma_c(\gamma_c^2 - \beta)}{\alpha + \gamma_c}\right)\right] \\ &+\frac{\alpha(\beta - \gamma_c^2 + \sqrt{\Delta_n})}{2(\alpha + \gamma_c)}\gamma_p^n, \end{aligned} \tag{B.23}$$

where $\Delta_n = (\beta - \gamma_c^2)^2 + 4\beta\gamma_c(\alpha + \gamma_c) = (\beta + \gamma_c^2)^2 + 2\beta\gamma_c(2\alpha + \gamma_c)$.

To simplify the form of the variable Δ_n in the characteristic polynomial, we make the new "strong feedback" assumption that $\beta \gg \alpha\gamma_c$. Consequently, $\sqrt{\Delta_n} \approx \beta + \gamma_c^2$ since $\beta^2 \gg 4\alpha\beta\gamma_c$.

Therefore, the characteristic polynomial simplifies to

$$p(s) = (s + \gamma_p)^n \left[s^2 + s\left(\alpha + \frac{\beta}{\alpha + \gamma_c}\right) + \left(\alpha\gamma_c + \frac{\gamma_c(\beta - \gamma_c^2)}{\alpha + \gamma_c}\right)\right] + \frac{\alpha\beta}{(\alpha + \gamma_c)}\gamma_p^n. \tag{B.24}$$

Thus, the roots $s \in \mathbb{C}$ of the characteristic polynomial satisfy the following equation:

$$(s + \gamma_p)^n \left[s^2 + s\left(\alpha + \frac{\beta}{\alpha + \gamma_c}\right) + \left(\alpha\gamma_c + \frac{\gamma_c(\beta - \gamma_c^2)}{\alpha + \gamma_c}\right)\right] + \frac{\alpha\beta}{(\alpha + \gamma_c)}\gamma_p^n = 0.$$

We study the characteristic polynomial's roots to determine the stability of the sequestration feedback network with controller species degradation.

First, we make the change of variable $s = \gamma_p z$ and obtain a new formulation of the characteristic polynomial in equation (B.35) as the following:

$$(z+1)^n \left[z^2 + z\left(\frac{\alpha}{\gamma_p} + \frac{\beta}{\gamma_p(\alpha+\gamma_c)}\right) + \left(\frac{\alpha\gamma_c}{\gamma_p^2} + \frac{\gamma_c(\beta-\gamma_c^2)}{\gamma_p^2(\alpha+\gamma_c)}\right)\right] = -\frac{\alpha\beta}{\gamma_p^2(\alpha+\gamma_c)}. \tag{B.25}$$

By using the strong feedback assumption, we know that $\beta \gg \alpha^2, \alpha\gamma_c$, therefore the characteristic polynomial simplifies to

Let z be a complex root of the characteristic polynomial in equation (B.26). We use the characteristic polynomial in equation (B.26) to derive the stability criterion for the sequestration feedback system. We consider the following three cases: $\gamma_c \ll \gamma_p$, γ_c and γ_p the same order of magnitude, and $\gamma_c \gg \gamma_p$.

Theorem 3.1. *We consider the sequestration feedback network with controller degradation described in equations (3.1) under the assumption of strong sequestration feedback ($\beta \gg \alpha^2, \alpha\gamma_p$). The closed loop stability criterion depends on the relationship between the process degradation rate γ_p and the controller degradation rate γ_c. We consider the following three cases:*

Case I: *If the controller species degradation rate is much smaller than the process species degradation rate, the stability criterion is the same as the production-degradation inequality in [Ols+17], which we reproduce here as:*

$$\begin{aligned} &\gamma_p \gg \gamma_c, \\ &\gamma_p > \sqrt[n+1]{\frac{\theta_1\theta_2\prod_{i=1}^{n-1} k_i}{\tan\left(\frac{\pi}{2n}\right)\left(1+\tan\left(\frac{\pi}{2n}\right)^2\right)^{\frac{n}{2}}}}. \end{aligned} \tag{3.5}$$

Case II: *If the controller and process species degradation rates are approximately equal, then the closed loop system is stable if*

$$\begin{aligned} &\gamma_p \approx \gamma_c, \\ &\gamma_p > \frac{\sqrt[n+1]{\theta_1\theta_2\prod_{i=1}^{n-1} k_i}}{\sqrt{\tan\left(\frac{\pi}{n+1}\right)^2+1}}. \end{aligned} \tag{3.6}$$

Case III: *If the controller species degradation rate is much larger than the process species degradation rate, then the closed loop system is stable if*

$$\begin{aligned} &\gamma_c \gg \gamma_p, \\ &\gamma_p > \frac{\sqrt[n]{\theta_1\theta_2\prod_{i=1}^{n-1} k_i}}{\sqrt[n]{\gamma_c}\sqrt{1+\tan\left(\frac{\pi}{n}\right)^2}}. \end{aligned} \tag{3.7}$$

Proof. **Case I:** $\gamma_p \gg \gamma_c$

In this case, the controller species degradation rate is much smaller than the process species degradation rate and therefore it does not influence the stability of the closed loop sequestration feedback system.

In particular, the characteristic polynomial in equation (B.26) reduces to the following characteristic polynomial:

$$(z+1)^n\left[z^2 + z\left(\frac{\alpha}{\gamma_p} + \frac{\beta}{\gamma_p(\alpha+\gamma_c)}\right) + \left(\frac{\alpha\gamma_c}{\gamma_p^2} + \frac{\gamma_c(\beta-\gamma_c^2)}{\gamma_p^2(\alpha+\gamma_c)}\right)\right] = -\frac{\alpha\beta}{\gamma_p^2(\alpha+\gamma_c)}. \tag{B.26}$$

$$(z+1)^n\left(z + \frac{\alpha+\frac{\beta}{\alpha}}{\gamma_p}\right) = -\frac{\beta}{\gamma_p^2}. \tag{B.27}$$

We know from [Ols+17] that the solution of the stability problem is provided by the production-degradation inequality. Consequently, in order for the closed loop sequestration feedback system to be stable, we must have that:

$$\begin{aligned} &\gamma_p \gg \gamma_c, \\ &\gamma_p > \sqrt[n+1]{\frac{\theta_1\theta_2\prod_{i=1}^{n-1} k_i}{\Delta_n}}. \end{aligned} \tag{B.28}$$

Case II: $\gamma_p \approx \gamma_c$

Let us assume that $\beta = \gamma_c^2$. Equivalently, the characteristic polynomial for z must satisfy the following equation:

$$(z+1)^n\left[\left(z\frac{\gamma_p}{\gamma_c}+1\right)\left(z\frac{\gamma_p(\alpha+\gamma_c)}{\beta-\gamma_c^2}+1\right)\right] = -\frac{\alpha\beta}{\gamma_c(\beta-\gamma_c^2)}. \tag{B.29}$$

Since $\gamma_p \approx \gamma_c$, then we must have that

$$(z+1)^{n+1}\left(z\frac{\gamma_c(\alpha+\gamma_c)}{\beta-\gamma_c^2}+1\right)=-\frac{\alpha\beta}{\gamma_c(\beta-\gamma_c^2)}. \tag{B.30}$$

Since we are in the strong feedback limit of $\beta >> \alpha^2, \alpha\gamma_p$. We multiply both side of the characteristic polynomial by factor $\frac{\gamma_p\alpha}{\beta}$.

Hence we obtain that

$$(z+1)^{n+1}\left(z\frac{\gamma_p\alpha}{\beta}+\frac{\gamma_p\alpha}{\beta}+\frac{\alpha^2}{\beta}+1\right)=-\frac{\alpha}{\gamma_p}, \tag{B.31}$$

which, under the strong feedback assumption, simplifies to

$$(z+1)^{n+1}\left(z\frac{\gamma_p\alpha}{\beta}+1\right)=-\frac{\alpha}{\gamma_p}, \tag{B.32}$$

Thus, the largest real root is at value $\approx -1-\frac{\beta}{\gamma_p\alpha}$.

If n is even, then the characteristic polynomial in (B.32) has two negative real roots and the remaining roots are conjugate complex pairs.

To determine the stability criterion, we analyze the bifurcation point at which the system goes from stable to unstable. Accordingly, we assume that there is a complex root $z = i\delta$. Then the magnitude and phase of the characteristic polynomial are given by

$$\begin{aligned}(\delta^2+1)^{\frac{n+1}{2}} &= \frac{\alpha}{\gamma_p},\\ \tan^{-1}(\delta) &= \frac{\pi(2k+1)}{n+1},\end{aligned} \tag{B.33}$$

where $0 \le k \le \frac{n}{2}$.

The second real root corresponds to index $k = \frac{n}{2}$.

We follow the proof in [Ols+17] and we obtain that stability is guaranteed if

$$\begin{aligned} &\text{n is even,} \\ &\gamma_p \approx \gamma_c, \\ &\gamma_p > \frac{\sqrt[n+1]{\theta_1\theta_2\prod_{i=1}^{n-1} k_i}}{\sqrt{\tan(\frac{\pi}{n+1})^2+1}}. \end{aligned} \tag{B.34}$$

If n is odd, then the characteristic polynomial in (B.32) has only one negative real root, located at value $\approx -1-\frac{\alpha+\frac{\beta}{\alpha}}{\gamma_p}$. The other roots are conjugate complex pairs. The rest of the argument is similar and stability is guaranteed for an analogous criterion.

Case III: $\gamma_p \ll \gamma_c$

$$(s+\gamma_p)^n\left[s^2+s\left(\alpha+\frac{\beta}{\alpha+\gamma_c}\right)+\left(\alpha\gamma_c+\frac{\gamma_c(\beta-\gamma_c^2)}{\alpha+\gamma_c}\right)\right]+\frac{\alpha\beta}{(\alpha+\gamma_c)}\gamma_p^n=0. \tag{B.35}$$

We can equivalently write the characteristic polynomial in equation (B.26) as the following:

$$(s+1)^n\left(s\frac{\gamma_p}{\gamma_c}+1\right)\left(s\frac{\gamma_p(\alpha+\gamma_c)}{\beta-\gamma_c^2}+1\right)=-\frac{\alpha}{\gamma_c} \tag{B.36}$$

Since $\frac{\gamma_p}{\gamma_c}$ and $\frac{\gamma_p(\alpha+\gamma_c)}{\beta-\gamma_c^2}$ are very small, we can infer that the characteristic polynomial (B.36) has two large negative real roots at values $\approx -\frac{\gamma_c}{\gamma_p}$ and $\approx -\frac{\beta-\gamma_c^2}{\gamma_p(\alpha+\gamma_c)}$. Here we have assumed that $\beta \gg \gamma_c^2$ in addition to the strong feedback assumption of $\beta \gg \alpha\gamma_c$.

We only consider the case n even. The case n odd has a similar proof with an additional real root.

Then the magnitude and the phase of the characteristic polynomial are given by:

$$\begin{aligned} (\delta^2+1)^{\frac{n}{2}} &= \frac{\alpha}{\gamma_c}, \\ \tan^{-1}(\delta) &= \frac{\pi(2k+1)}{n}, \end{aligned} \tag{B.37}$$

where $0 \le k \le \frac{n}{2}-1$.

All the roots are conjugate complex and they occur in pairs.

If the following conditions

$$\begin{aligned} &\gamma_p \ll \gamma_c, \\ &\gamma_p > \frac{\sqrt[n]{\theta_1 \theta_2 \prod_{i=1}^{n-1} k_i}}{\sqrt[n]{\gamma_c}\sqrt{1+\tan\left(\frac{\pi}{n}\right)^2}}, \end{aligned} \tag{B.38}$$

hold, then the stability of the closed loop sequestration feedback system is guaranteed.

□

Remark 9. *The sequestration feedback network in Figure B.2 can not simultaneously fulfill the strong feedback assumption and have a critical controller degradation rate since the inequalities in equation (B.3) and in $\beta \gg \alpha^2$ contradict each other for $n \geq 3$.*

B.5 Additional details about the numerical experiments

Scaling the sequestration feedback network model

We first rescale the sequestration feedback network to produce the simulations in Figures 3.4, B.1, and **??**, as well as throughout Section B.6. For simplicity, we only describe the rescaling step for the process network with two species (Figure B.3). However, this rescaling step is applicable for a general linear process network. The unscaled model of the sequestration feedback network is:

$$\begin{aligned} \dot{x}_1 &= \theta_1 z_1 - \gamma_p x_1, \\ \dot{x}_2 &= k_1 x_1 - \gamma_p x_2, \\ \dot{z}_1 &= \mu - \eta z_1 z_2 - \gamma_c z_1, \\ \dot{z}_2 &= \theta_2 x_2 - \eta z_1 z_2 - \gamma_c z_2. \end{aligned} \tag{B.39}$$

The units of these biochemical species and rates are given in Table B.2.

To rescale, we let $x_1 = x_1^{'} \cdot 10^9$, $x_2 = x_2^{'} \cdot 10^9$, $z_1 = z_1^{'} \cdot 10^9$, $z_2 = z_2^{'} \cdot 10^9$, $t = \frac{t^{'}}{3600}$, $\theta_1 = \theta_1^{'} \cdot 3600$, $\theta_2 = \theta_2^{'} \cdot 3600$, $\gamma_p = \gamma_p^{'} \cdot 3600$, $\gamma_c = \gamma_c \cdot 3600$, $k_1 = k_1^{'} \cdot 3600$, $\eta = \eta^{'} \cdot 10^{-9} \cdot 3600$, and $\mu = \mu^{'} \cdot 10^9 \cdot 3600$.

Then the units of the rescaled biochemical species and rates are given in Table B.3.

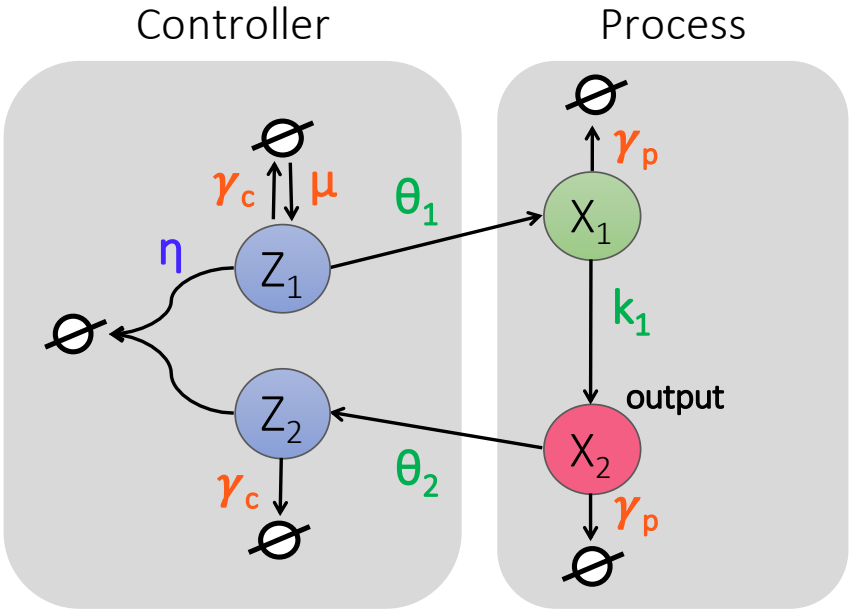

Figure B.3: **A sequestration feedback network with two process species.** For the numerical simulations in this section, we consider this sequestration feedback network with only two process species X_1 and X_2. We assume that process species X_1 and X_2 are degraded at the same rate γ_p and that the controller species Z_1 and Z_2 are degraded at the same rate γ_c. We give the model of this sequestration feedback network in equation (B.39).

x_1	x_2	z_1	z_2	t	θ_1	θ_2	γ_p	μ	η	γ_c	k_1
M	M	M	M	s	s^{-1}	s^{-1}	s^{-1}	M s^{-1}	M^{-1} s^{-1}	s^{-1}	s^{-1}

Table B.2: **The units of the biochemical species and rates in the sequestration feedback network.** Species X_1, X_2, Z_1, and Z_2 have units of molar (M) and time t has units of seconds (s). Thus, the degradation and the production rates θ_1, θ_2, γ_p, γ_c, and k_1 have units of s^{-1} since they correspond to first order reactions. Rate μ has units of M s^{-1} since it corresponds to a zero order reaction. Rate η has units of M^{-1} s^{-1} since it corresponds to a second order reaction.

$x_1^{'}$	$x_2^{'}$	$z_1^{'}$	$z_2^{'}$	t$^{'}$	$\theta_1^{'}$	$\theta_2^{'}$	$\gamma_p^{'}$	$\mu^{'}$	$\eta^{'}$	$\gamma_c^{'}$	$k_1^{'}$
nM	nM	nM	nM	hr	hr^{-1}	hr^{-1}	hr^{-1}	nM hr^{-1}	nM^{-1} hr^{-1}	hr^{-1}	hr^{-1}

Table B.3: **The units of the biochemical species and rates in the scaled sequestration feedback network.** Biochemical species $X_1^{'}$, $X_2^{'}$, $Z_1^{'}$, and $Z_2^{'}$ have units of nanomolar (nM) and time t$^{'}$ has units of hours (hr). Thus, the degradation and the production rates $\theta_1^{'}$, $\theta_2^{'}$, $\gamma_p^{'}$, $\gamma_c^{'}$, and $k_1^{'}$ have units of hr^{-1} since they correspond to first order reactions. Rate $\mu^{'}$ has units of nM hr^{-1} since it corresponds to a zero order reaction. Rate $\eta^{'}$ has units of nM^{-1} hr^{-1} since it corresponds to a second order reaction.

Following rescaling, the model of the sequestration feedback network can be described as:

$$\begin{aligned}
\dot{x}_1^{'} &= \theta_1^{'} z_1^{'} - \gamma_p^{'} x_1^{'}, \\
\dot{x}_2^{'} &= k_1^{'} x_1^{'} - \gamma_p^{'} x_2^{'}, \\
\dot{z}_1^{'} &= \mu^{'} - \eta^{'} z_1^{'} z_2^{'} - \gamma_c^{'} z_1^{'}, \\
\dot{z}_2^{'} &= \theta_2^{'} x_2^{'} - \eta^{'} z_1^{'} z_2^{'} - \gamma_c^{'} z_2^{'}.
\end{aligned} \tag{B.40}$$

We compute representative values of the controller's and the process' degradation rates from the half-life values in Table B.4. We compute the degradation rates in Table 3.1 in units of hr^{-1} by taking the inverses of the half-life values and multiplying them by the constant log(2). We gave the representative values of the controller's and the process' degradation rates in Table 3.1.

mRNA	Hok	RpoS	CcdA
0.5 – 6 min	20 min	30 min	30 min

Table B.4: **The half-lives of biological parts that could be used to build sequestration controllers.** We present the half-lives of mRNA, toxin Hok, sigma factor RpoS, and antitoxin CcdA. The median mRNA half-life is measured as 0.5-6 min in [Ber+04]. The half-life of toxin Hok in the type I toxin-antitoxin pair in *E. coli* is 20 minutes [SM12]. The half-life of the sigma factor protein RpoS is 30 min when the *E. coli* cells are in stationary phase at 37°C or under stress conditions. When sigma factor proteins RpoS are actively degraded by protease ClpXP during the exponential phase, their half-life is only 2 minute [ZG98]. The antitoxin CcdA is degraded in wild-type cells with a half-life of 30 min in the absence of toxin CcdB and a half-life of 60 min when bound in a complex with toxin CcdB [De +09].

Lastly, we present representative values of the sequestration binding on-rates in Ta-

ble B.5. The rate η' in the rescaled sequestration feedback model (equation B.40) represents the on-rate of the sequestration reaction between the two controller species. The off-rate of the sequestration reaction is much smaller than the on-rate. The values of the dissociation constants in Figure 3.2 can be computed as the ratios between the off-rates and the on-rates of the sequestration reaction.

mRNA : antisense RNA	sigma factor : anti-sigma factor
0.005 nM^{-1} hr^{-1} - 1.62 nM^{-1} hr^{-1}	18 nM^{-1} hr^{-1} - 72 nM^{-1} hr^{-1}

Table B.5: **The on-rate of the sequestration reaction.** To match Table B.3, we give the on-rates of the sequestration reactions in units of nM^{-1} hr^{-1}. From [Wal+02], we compute representative on-rate values of mRNA binding to antisense RNA. The values of the on-rates are between 0.005 nM^{-1} hr^{-1} and 1.62 nM^{-1} hr^{-1}. From [Raj+16], we compute representative on-rate values of sigma factors binding anti-sigma factors. The values of the on-rates are between 18 nM^{-1} hr^{-1} and 72 nM^{-1} hr^{-1}, depending on the temperature and the presence of zinc. To the best of our knowledge, the on-rate of the toxin CcdB binding the antitoxin CcdA has not been accurately determined [KHM99].

Details of numerical experiments

For the two-species process network in Figure B.3, both quantities $\alpha = \frac{\theta_1\theta_2 k_1}{\gamma_p^2}$ and $\beta = \eta\mu$ have units of hr^{-1}. For a transcriptional implementation of the controller species and a protein implementation of the process species, this network satisfies the assumption of strong sequestration feedback for the parameter values in Table B.6.

γ_c'	η'	γ_p'	μ'	k_1'	θ_1'	θ_2'	α'	β'
10 hr^{-1}	1 nM^{-1}	1 hr^{-1}	50 nM hr^{-1}	1 hr^{-1}	1 hr^{-1}	10 hr^{-1}	10 hr^{-1}	50 hr^{-1}

Table B.6: **The parameters used for a simulation with controller species implemented by transcriptional parts and process species implemented by protein parts.** The values of parameters γ_p, γ_c, and η are found in Tables 3.1 and B.5 for both the transcriptional and the protein parts. The reference concentration is 5 nM.

B.6 Other modeling assumptions

In this section, we consider different modeling assumptions than the setup introduced in Figure 3.3.

Firstly, we assume that the process species in the sequestration feedback network in Figure 3.3 degrade at different rates. If the process species are degraded at different

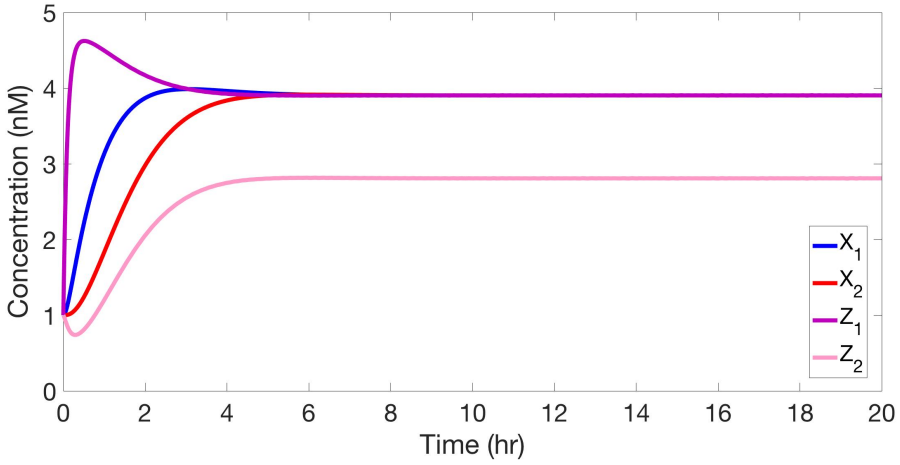

Figure B.4: **The sequestration feedback network with the controller species implemented with transcriptional parts and the process species implemented with protein parts.** The sequestration feedback network simulated using the parameters in Table B.6 is stable and has oscillations that settle within 5 hours. The reference concentration is 5 nM and the steady state error is 1.1 nM.

γ_c'	η'	μ'	k_1'	θ_1'	θ_2'
1 hr^{-1}	50 nM^{-1}	50 nM hr^{-1}	10 hr^{-1}	1 hr^{-1}	1 hr^{-1}

Table B.7: **The parameters used for the simulation in Figure 3.4.** In panel A, the process degradation rate is $\gamma_p = 2$ hr^{-1}. In panel B, the process degradation rate is $\gamma_p = 0.25$ hr^{-1}. The reference concentration is 50 nM.

rates, the characteristic polynomial in equation (B.26) does not necessarily factor the term $(s+1)^n$. Therefore, finding an analytical approximation to the roots of the characteristic polynomial can be challenging and finding an analytical criterion for the stability of sequestration feedback networks may not be possible. However, we consider the model with different process degradation rates:

$$\begin{aligned}
\dot{x}_1 &= \theta_1 z_1 - \gamma_{p_1} x_1, \\
\dot{x}_2 &= k_1 x_1 - \gamma_{p_2} x_2, \\
\dot{z}_1 &= \mu - \eta z_1 z_2 - \gamma_c z_1, \\
\dot{z}_2 &= \theta_2 x_2 - \eta z_1 z_2 - \gamma_c z_2.
\end{aligned} \tag{B.41}$$

As indicated by Figure B.5, the smaller of the two process species degradation rates results in instability. Although the larger process species degradation rate dampens the oscillations (compare Figures B.5B and 3.4B), it is not able to fully compensate

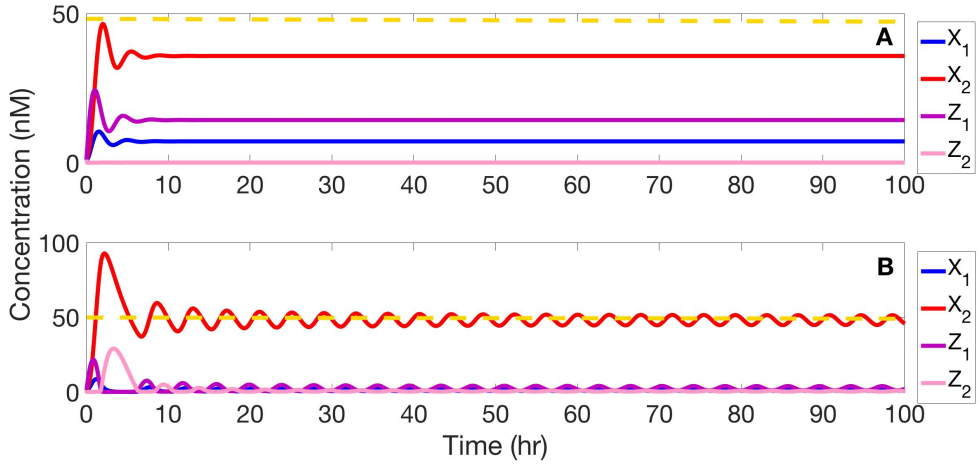

Figure B.5: **The stability of the sequestration feedback network with different process species degradation rate and strong sequestration feedback.** We illustrate the stability of sequestration feedback networks by assuming strong sequestration binding, indicative of a sigma factor implementation of the controller species. In panel A, we use large process degradation rates of 2 hr^{-1}. We note stability since the process output species X_2 (red) tracks the reference concentration of 50 nM (orange dashed line), albeit with a large steady state error. In panel B, we employ a process species with a small degradation rate (0.25 hr^{-1}) and a process species with a larger degradation rate (2 hr^{-1}). This results in instability, as evidenced by the sustained oscillations in the process output species X_2. The parameters used for these simulations are given in Table B.8.

$\gamma_c^{'}$	$\eta^{'}$	$\mu^{'}$	$k_1^{'}$	$\theta_1^{'}$	$\theta_2^{'}$
1 hr^{-1}	50 nM^{-1}	50 nM hr^{-1}	10 hr^{-1}	1 hr^{-1}	1 hr^{-1}

Table B.8: **The parameters used for the simulation in Figure B.5.** In panel A, the process degradation rates are $\gamma_{p_1}^{'} = \gamma_{p_2}^{'} = 2$ hr^{-1}. In panel B, the process degradation rates are $\gamma_{p_1}^{'} = 2$ hr^{-1}, $\gamma_{p_2}^{'} = 0.25$ hr^{-1}. The reference concentration is 50 nM.

for the small degradation rate of the other process species. Therefore, the smallest process species degradation rate must be carefully tuned to ensure stability of the sequestration feedback network.

Secondly, we consider the reactions between the process and the controller networks to be enzymatic. We note that the relationship between the process and the controller networks in Figure 3.3 can also be represented by the following chemical reactions:

$$\begin{aligned} \mathrm{Z}_1 &\xrightarrow{\theta_1} \mathrm{Z}_1 + \mathrm{X}_1, \\ \mathrm{X}_\mathrm{n} &\xrightarrow{\theta_2} \mathrm{X}_\mathrm{n} + \mathrm{Z}_2. \end{aligned} \tag{B.42}$$

Alternatively, we can assume that the transformation of controller species Z_1 into the process input species X_1 is catalyzed by the enzyme E_1 and that the transformation of the process output species X_n into the controller species Z_2 is catalyzed by the enzyme E_2. Then the dynamics of these catalytic reactions can be modeled using the Michaelis–Menten kinetics as follows:

$$\begin{aligned} \mathrm{E}_1 + \mathrm{Z}_1 &\underset{d_1}{\overset{a_1}{\rightleftharpoons}} \mathrm{C}_1 \xrightarrow{k_1} \mathrm{E}_1 + \mathrm{X}_1, \\ \mathrm{E}_2 + \mathrm{X}_\mathrm{n} &\underset{d_2}{\overset{a_2}{\rightleftharpoons}} \mathrm{C}_2 \xrightarrow{k_2} \mathrm{E}_2 + \mathrm{Z}_2, \end{aligned} \tag{B.43}$$

where the rates satisfy the inequalities $k_1 \gg a_1, d_1$ and $k_2 \gg a_2, d_2$.

We make the simplifying assumptions that the concentration of enzymes E_1 and E_2 are much smaller than the concentrations of species Z_1 and X_1 [MVA00; DM15]. We let the Michaelis constants be

$$\begin{aligned} K_m^1 &= \frac{d_1 + k_1}{a_1}, \\ K_m^2 &= \frac{d_2 + k_2}{a_2}. \end{aligned} \tag{B.44}$$

The Michaelis constants K_m^1 and K_m^2 that correspond to biological enzymes range between 10^{-6} M and 10^{-2} M [MVA00], which renders them to be much larger than the concentrations of species Z_1 and X_n. Therefore, the propensity functions associated with these kinetics are linear.

We let $\theta_1 = \frac{k_1 E_1^0}{K_m^1}$ and $\theta_2 = \frac{k_2 E_2^0}{K_m^2}$, where E_1^0 and E_2^0 are the initial concentrations of enzymes E_1 and E_2, respectively. Using this notation, the sequestration feedback

model remains unchanged from Figure 3.3 due to the linearity of the first-order kinetics.

Thirdly, we consider additional reactions between the controller and the process species.

The updated model of the system is:

$$\begin{aligned}
\dot{x}_1 &= \theta_1 z_1 - \gamma_p x_1 - k_1 x_1, \\
\dot{x}_2 &= k_1 x_1 - \gamma_p x_2 - \theta_2 z_2, \\
\dot{z}_1 &= \mu - \eta z_1 z_2 - \gamma_c z_1 - \theta_1 z_1, \\
\dot{z}_2 &= \theta_2 x_2 - \eta z_1 z_2 - \gamma_c z_2.
\end{aligned} \tag{B.45}$$

Appendix C

ADDITIONAL COMPUTATIONAL RESULTS FOR PARAMETER IDENTIFICATION

C.1 Results for the bursting gene model

The propensity functions of the bursting gene model's reactions in (4.8)-(4.10) are: k_{on}, k_{off}, k_r and $k_{deg}R$. Their stoichiometry matrix is:

$$\begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix}$$

For the bursting gene model, we use the parameters identified from the FSP and the RBF-FSP projections using a parameter sweep to compare the resulting distributions with the original data, as illustrated in Figure C.1. Since the original data is discrete and the radial basis functions continuous, we can only hope to approximate the original data with these two probability distributions.

We then use a version of adaptive Metropolis Hastings for the RBF-FSP model with the covariance matrix $0.7 \times V$, where

$$V = \begin{pmatrix} 0.001 & 0.005 & 0.003 & 0.000 \\ 0.005 & 0.031 & 0.022 & -4.293 \times 10^{-5} \\ 0.003 & 0.022 & 0.016 & 0.000 \\ 0.000 & -4.293 \times 10^{-5} & 0.000 & 0.000 \end{pmatrix}.$$

The joint distributions of the parameter values are illustrated in Figure C.2

C.2 Results for the genetic toggle switch model

In Figure C.3, we describe the setup of the genetic toggle switch that reflects the model in equations (4.11) and (4.12) in Section 4.5.

In Figure C.4(A) and (B), we compare the toggle switch data with distributions generated by the parameters that maximized the likelihood function for the FSP model (top panel) and the RBF-FSP model (bottom panel) of the genetic toggle switch. The results are qualitatively similar with Figure 4.5(B).

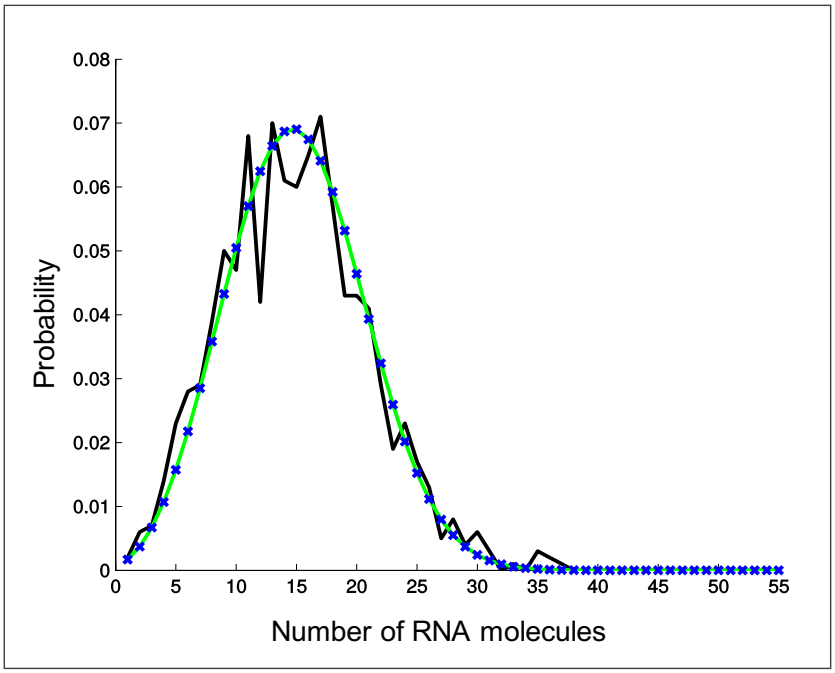

Figure C.1: **The probability distributions generated by the parameters identified using the two projections and the simulated data for the bursting gene model.** We compare the probability distribution of the simulated data (black) to the probability distributions obtained from the gene bursting model by using the parameters that maximized the likelihood of observing the data $L(D|\Lambda)$ in the RBF-FSP projection (green) and in the FSP projection (blue). We observe that the three probability distributions match.

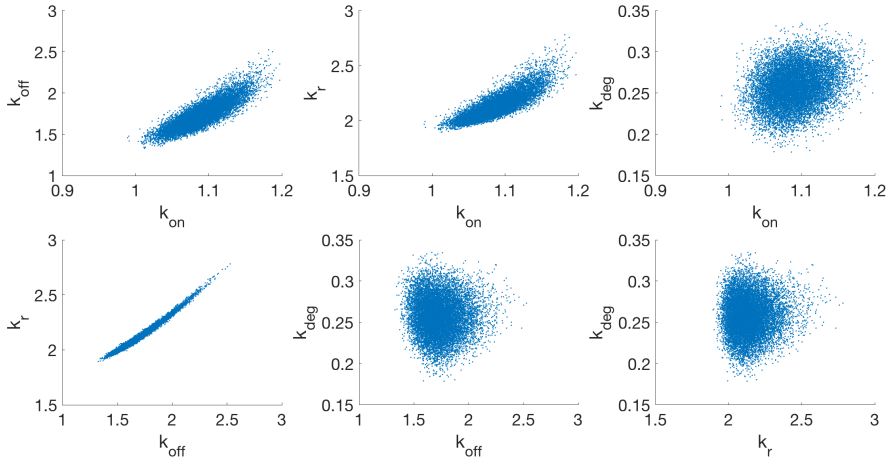

Figure C.2: **The joint distributions of the parameters identified using the Metropolis Hastings algorithm for the bursting gene model.** We identify probability distributions of the parameters of the bursting gene model using an adaptation of the Metropolis Hastings algorithm. By plotting the joint probability distributions of pairs of parameters, we can evaluate their correlations. We expect that the correlations are represented by compact, not diffuse shapes; this indicates that the chain has a sufficient number of accepted samples. We also look for odd shapes in the two-dimensional pairwise projections of the four-dimensional joint distribution; we observe that parameters k_r and k_{off} are anti-correlated. This result matches our expectation of the bursting gene model.

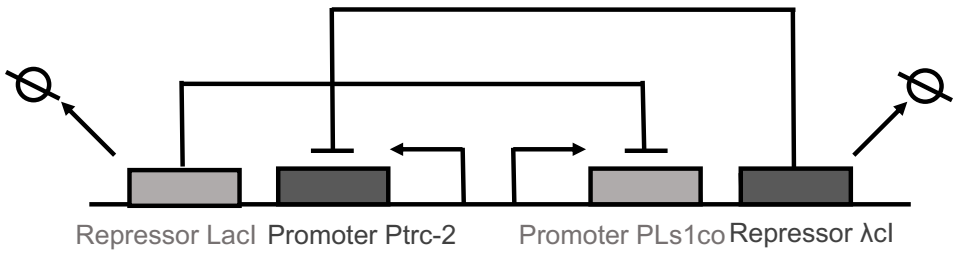

Figure C.3: **Illustration of the genetic toggle switch, adapted to the example in Section 4.5.** The two repressor proteins LacI and λcI mutually repress each other's promoters; LacI represses promoter PLs1co that drives the expression of λcI and λcI represses promoter Ptrc-2 that drives the expression of LacI. Additionally, both repressor proteins are subjected to degradation (represented using the empty set symbol).

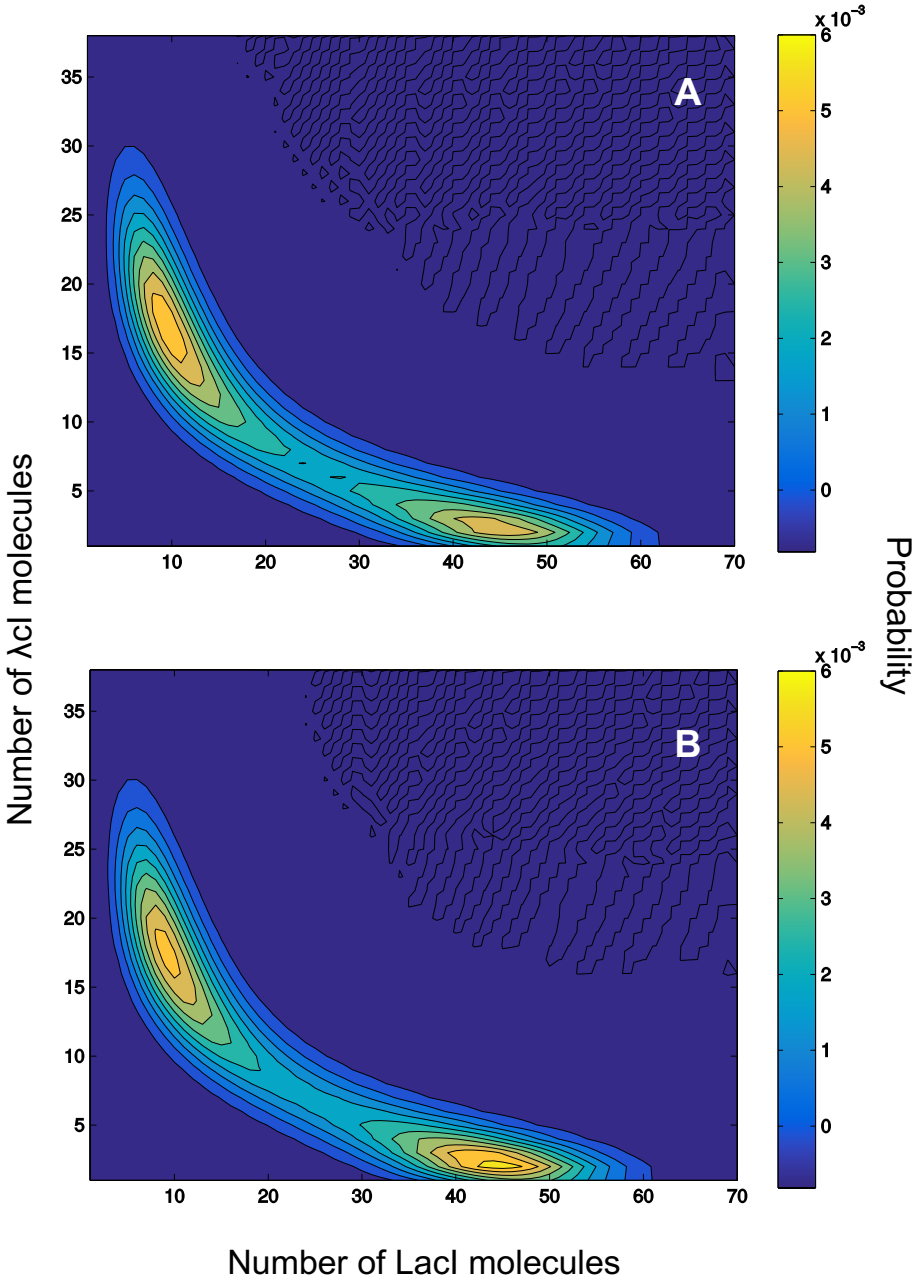

Figure C.4: **The probability distributions generated by the parameters identified using the two projections for the genetic toggle switch model.** We compare the simulated data for the genetic toggle switch in Figure 4.5(B) to the probability distributions for the parameters that maximized the likelihood of observing the data $L(D|\Lambda)$ in the FSP model (A) and in the RBF-FSP model (B). We use parameters identified after 100000 runs of the Metropolis-Hastings algorithm.