Design and implementation of feedback control systems for synthetic biology



Harrison Steel
St John's College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Trinity 2019

Acknowledgements

My greatest appreciation goes to Professor Antonis Papachristodoulou, who was the most knowledgeable and encouraging supervisor I could have asked for. His faith in me sustained my pursuit of the diverse research topics covered in this thesis, and the time he took to guide and support my investigations was truly invaluable. I must thank him for being so understanding of my many extracurricular activities and travels which regularly distracted me. He was also most kind in introducing me to his broad network of research colleagues, and in providing me many opportunities to travel academically. My time as a student at Oxford wouldn't have been nearly as exciting without him!

I am very grateful to have had the opportunity to collaborate experimentally with Ciarán Kelly, Thomas Folliard, Robert Habgood, Xiaoyu Chen, and Nicolas Delalez during my studies; they greatly extended my understanding of laboratory techniques in synthetic biology, and provided a wealth of assistance and advice for the studies described herein. Similarly, I would like to thank Aivar Sootla, Andreas Harris, and Thomas Prescott for their advice on the modelling studies described herein, they all have fantastically collaborative attitudes, and a wealth of knowledge that they most generously share!

On a more general note I would like to thank all SySOS group members at Oxford, both past and present, who made working here so exciting. These are James Anderson, Edward Hancock, Axel Nyström, James Scott-Brown, Yang Zheng, Licio Romao, Carolin Schute, and Boon Lim.

During my DPhil I had the opportunity to visit and work with a range of academics outside of our research group, who added colour and variety to the experience. In particular, I would like to thank Domitilla Del Vecchio for hosting me at MIT for three months, Mustafa Khammash for hosting my visit to ETH Zurich, John Doyle for hosting my visit to Caltech, and Wei Huang for allowing me to use some of his laboratory space for testing of my Turbidostat device.

A special thank you goes to my family (Mum, Dad, Elliot!) and friends (I can't play favourites by only naming a few!) who offered support, and made living so far from Australia such an enjoyable experience. I'd particularly like to thank those from Australia and New Zealand that stopped in to Oxford during their travels to see me, as well as my extended family in the UK, for making this country feel like home.

Finally, I would like to thank the General Sir John Monash foundation for trusting me and funding my studies at Oxford. Without their scholarship I wouldn't have been able to undertake this degree, and without the vibrant Monash community in Oxford and around the world I wouldn't have had nearly as much fun!

Abstract

Synthetic biology is a growing field at the interface of the engineering and biological sciences which aims to apply rational engineering design principles in biological contexts. It has many applications ranging from healthcare and the environment to industrial chemical synthesis and computation. When researchers attempt to apply synthetic biological tools to tackle these problems, they often follow a similar design-build-test cycle to that employed in traditional engineering disciplines. However, there are major challenges at each step in this cycle due to the variability inherent in biological systems, and our limited ability to understand and measure their behaviour. This thesis addresses select problems in each step of the design-build-test cycle as applied to synthetic biology, from a control-theoretic perspective.

To aid the design process of synthetic biological circuits I begin by using traditional control-theoretic analysis to derive structural requirements that biological systems must satisfy if they are to achieve behaviours such as disturbance rejection. I demonstrate how these requirements can be used to inform the biological design process, and highlight the challenges that must be overcome for effective noise attenuation in such circuits.

To refine the build/creation process for synthetic biological systems, I describe the experimental implementation and subsequent analysis of new synthetic biological feedback controllers built from two classes of components; short RNA and Recombinase proteins. Our results demonstrate how these circuits can be used to tune (or create) feedback loops, and reduce the influence of cellular noise on the performance of our synthetic circuits. This work thus has the double benefit of demonstrating new approaches to the engineering of biological systems, as well as creating well-characterised biological circuits that can be applied directly to a range of applications.

Finally, to improve the testing/characterisation process for synthetic biological systems I develop a novel biological approach to measuring cell-to-cell variability, as well as a new automated experimental platform: The Turbidostat. This device provides many of the important measurement and actuation capabilities required for synthetic biological research in a cheap and easy-to-use robotic platform. It can thus be used to automate a diverse range of characterisation experiments that would otherwise be highly laborious and prone to variability/noise. I perform several validation experiments, demonstrating this device's broad range of capabilities.

Contents

Abstract **iii**

Contents **vi**

1 Introduction **1**
1.1 Motivation . . . 1
1.2 Thesis Outline . . . 6
1.3 Contribution to Literature . . . 7

2 Background **10**
2.1 Biological Background . . . 10
2.1.1 Dramatis Personæ: DNA, RNA, and Protein . . . 10
2.1.2 The Central Dogma of Molecular Biology . . . 12
2.1.3 A Synthetic Biologist's Toolbox . . . 14
2.1.4 Biological Noise . . . 18
2.2 Forward Design and Control in Synthetic Biology . . . 20
2.2.1 Biological Controllers and Forward Design . . . 20
2.2.2 Approaching Feedback Control . . . 23
2.3 Experimental Techniques in Synthetic Biology . . . 24
2.3.1 Creating New Circuits . . . 24
2.3.2 Synthetic Circuit Characterisation . . . 25
2.3.3 Automating Experimental Challenges . . . 26
2.4 Modelling Biological Systems . . . 27
2.4.1 Chemical Reaction Networks . . . 28
2.4.2 Systems of Ordinary Differential Equations . . . 29
2.4.3 Linearisation and Frequency Domain Analysis . . . 30
2.4.4 Deterministic Modelling . . . 32
2.4.5 Noise Modelling and Stochastic Simulation Algorithms . . . 33
2.4.6 Choosing and Fitting Parameters . . . 37
2.5 Conclusion . . . 38

3 Design Rules for Biological Control Systems **40**
3.1 Introduction . . . 40
3.2 Aim and Problem Outline . . . 43
3.2.1 Potential forms for $\mathbf{f}(\boldsymbol{x}, \boldsymbol{u})$ . . . 43
3.2.2 System Linearisation and Assumptions for Implementation . . . 45
3.2.3 Other General Constraints for Implementation . . . 47
3.3 General Linearised Network Structure . . . 48

3.4 Network Constraints for Adaptation . . . 49
3.5 Network Constraints for Disturbance Rejection . . . 53
3.6 Application To Biological System Design . . . 55
3.6.1 Implementation of parameter constraints . . . 55
3.6.2 Non-linear system selection and correspondence with linear system 56
3.6.3 Parameter selection and implementation . . . 58
3.7 Discussion . . . 60
3.8 Conclusion . . . 63

4 sRNA-Based Controllers 64
4.1 Introduction . . . 64
4.2 Experimental Summary . . . 67
4.3 Model Structures . . . 68
4.4 Parameter Selection . . . 70
4.5 Modelling Noise . . . 73
4.6 The Autorepressor . . . 74
4.6.1 Simple Model . . . 75
4.6.2 Detailed Model . . . 75
4.6.3 Parameter Fitting . . . 76
4.7 sRNA Circuit 1 . . . 77
4.7.1 Simple Model . . . 77
4.7.2 Detailed Model . . . 77
4.7.3 Parameter Fitting . . . 79
4.7.4 Dynamic Behaviour . . . 81
4.7.5 Noise Simulation . . . 84
4.8 sRNA Circuit 2 . . . 86
4.8.1 Simple Model . . . 87
4.8.2 Detailed Model . . . 88
4.8.3 Parameter Fitting . . . 89
4.8.4 Dynamic Behaviour . . . 90
4.8.5 Noise Simulation . . . 92
4.9 Discussion . . . 93
4.10 Conclusion . . . 95

5 DNA Recombinase Controller Implementations 96
5.1 Introduction . . . 96
5.2 Recombinase Circuit 1 . . . 98
5.2.1 Experimental Summary . . . 100
5.2.2 Modelling Circuit 1 . . . 101
5.2.3 Simulation and Parameter Values . . . 105
5.2.4 Results . . . 106
5.2.5 Discussion . . . 109
5.3 Recombinase Circuit 2 . . . 112
5.3.1 Approximating mRNA-mRNA Annihilation . . . 113
5.3.2 System Model . . . 114
5.3.3 Analysis . . . 118
5.3.4 Dynamic Simulation . . . 123
5.3.5 Stochastic Simulation . . . 124
5.3.6 Discussion . . . 125

5.4 Conclusion 126

6 Synthetic Biological Reporting Circuits 128
6.1 Motivation 128
6.2 Noise Distributions and Quantification 130
6.3 A Biological Implementation: FRET Reporter Pair 133
6.4 Parameter Selection 136
6.5 Deterministic Analysis 138
6.6 Parameter Dependence and Noise Sensitivity 140
6.7 Discussion 146
6.8 Conclusion 150

7 Characterisation Hardware: The Turbidostat 151
7.1 Introduction 151
7.1.1 Desired Capabilities and Features 153
7.1.2 Potential Applications 154
7.1.3 Comparison to Similar Devices 155
7.2 Turbidostat Design History 157
7.2.1 Version 1.0 158
7.2.2 Version 2.0 159
7.2.3 Version 3.0 159
7.3 Actuation Sub-Systems 160
7.3.1 Heating 160
7.3.2 Stirring 161
7.3.3 Pumping 162
7.3.4 OD Laser 163
7.3.5 Visible Light LED 164
7.3.6 Ultra-Violet Light LED 165
7.4 Measurement Sub-Systems 165
7.4.1 Thermometry 165
7.4.2 Spectrometry 166
7.5 Control Implementation 166
7.5.1 Temperature Regulation 166
7.5.2 OD Regulation 168
7.6 Electrical Implementation 168
7.7 Software 172
7.8 Experimental Testing and Validation 175
7.9 Conclusion 181

8 Conclusion 182
8.1 Future Outlook 184

References 189

Acronyms

aTc	Anhydrotetracycline
DNA	Deoxyribonucleic Acid
FP	Fluorescent Protein
FRET	Forster Resonance Energy Transfer
GFP	Green Fluorescent Protein
GUI	Graphical User Interface
HTML	Hypertext Markup Language
I^2C	Inter-Integrated Circuit
IO	Input/Output
IR	Infrared
LED	Light Emitting Diode
MOSFET	Metal Oxide Semiconductor Field Effect Transistors
mRNA	Messenger Ribonucleic Acid
OD	Optical Density
ODE	Ordinary Differential Equation
PCB	Printed Circuit Board
PDE	Partial Differential Equation
PID	Proportional Integral Derivative
PWM	Pulse Width Modulation
RBS	Ribosome Binding Site
RNA	Ribonucleic Acid
RNAP	Ribonucleic Acid Polymerase
sRNA	Small Ribonucleic Acid
TetR	Tetracycline Repressor
TF	Transcription Factor
tRNA	Transfer Ribonucleic Acid
UV	Ultraviolet

1

Introduction

1.1 Motivation

Synthetic biology is a developing field at the interface of the engineering and biological sciences which aims to apply rational design principles in biological contexts. Synthetic biological systems can be engineered to tackle real-world problems ranging from environmental protection to chemical synthesis [1], and from medical sensing to decision making and computation [2, 3]. Early breakthroughs in the field were achieved via the re-wiring of natural genetic components [4–6], followed by extensive trial-and-error experimentation to find appropriate parameter regimes for their operation. Though early efforts demonstrated the broad capabilities of synthetic biology, the field is far from reaching its potential [7]. Much must still be done to develop tools and design methodologies that account for the unreliable and often unpredictable behaviour of engineered biological systems [8, 9], which makes their application challenging in all but the most ideal conditions.

From the time that the first single-cell organisms emerged, life on earth has flourished and diversified, leaving few ecological niches unfilled [10]. Organisms have evolved to survive in the most demanding circumstances, ranging from sparse deserts to super-hot underwater volcanic vents [11]. In most places that life thrives we find an abundance of organisms, of a range of forms, sizes, and capabilities, interacting with one another to support complex ecosystems [12]. To achieve this, life has had to overcome many challenges that parallel those at the cutting edge of modern engineering disciplines. For example, nature has developed means for efficiently synthesising a huge range of molecules

from simple components and varying energy sources [1], sensory architectures that respond to the presence of tiny chemical concentrations [13], and methods for communication between organisms and cells over a range of scales and media [14]. Synthetic biology aims to re-purpose many of these functionalities found in nature to inspire and develop new biological tools that solve problems of human interest.

Over the past century many of the molecular mechanisms upon which life depends have been revealed. Genes (sequences of DNA) responsible for encoding particular functions or systems have been identified, which provide scientists with the fundamental information required to replicate or alter a biological process's behaviour in different contexts [15]. Simultaneously, significant advances have been made in the cost and speed of DNA sequencing and synthesis, which led to a flourishing of synthetic biological creations around the turn of the 21st century [16]. In many cases these constructs took natural DNA sequences and re-arranged them (often using a biological equivalent to cut-and-pasting) to define new functionalities [4, 5]. More recently approaches for *ad hoc* design of DNA sequences with a particular functionality have been developed, often using thermodynamic or machine-learning approaches informed by biological data [17, 18]. However, the creation of entirely synthetic proteins with complex functions generally remains beyond the state-of-the art (though in some cases substantial progress has been made [19, 20]). Thus, scientists are currently at a point where they re-purpose largely natural biological components (such as enzymes, regulatory proteins, or sensors) as building blocks to create novel biological systems.

As increasingly complex synthetic biological systems have been built it has become clear that naive re-arrangement of DNA sequences is rarely adequate for creating robust and predictable circuits [21, 22]. This is caused by a variety of complex interdependencies; the behaviour of a gene can vary greatly when it is transferred between organisms, or even different DNA contexts within a single organism [22]; the host organism's state, which is in turn acutely dependent upon its growth phase and environmental/growth conditions, can greatly affect a synthetic system's behaviour [23, 24]; the existence of a synthetic

system can significantly perturb the behaviour of its host cell (and in turn be influenced by changing processes around it) [25]. Because of these sources of variation synthetic biological components often require significant re-tuning (in terms of their expression levels, or constituent sequences) when employed in new contexts [26].

To improve the reliability (and predictability) of biological parts when they are employed in different contexts or environments researchers have begun to apply tools and techniques from control engineering. These techniques have been fundamental to progress in a wide range of traditional engineering disciplines (such as electrical and aeronautical systems), and so it is timely that they are now being applied to emerging biological challenges. Interestingly, study of natural biological systems has revealed that many employ control architectures similar to those from traditional control engineering [27, 28]. For example, autoregulatory architectures are ubiquitous in both prokaryotes [29] and eukaryotes [30], and have been studied in isolation via synthetic implementations [29, 31, 32]. Other frequently studied natural control systems include the osmoregulatory mechanism in yeast [33] and the chemotaxis pathway in bacteria [34], both of which have been demonstrated to rely on architectures that approximate integral feedback control [35].

Much recent research has focused on the development of synthetic biological control architectures, with a view to using them as regulatory building blocks in larger engineered systems. For example, synthetic biological systems which approximate the functionalities of control motifs including integrators [36], amplifiers [37], and summation gates [38, 39] have been demonstrated. These can then be used to build Lead-Lag [40], PID [41], or more advanced controllers, which (as known from traditional control engineering) can regulate noise, speed response times, and linearise system responses [32, 42]. In some cases such controllers have found practical application. For example, they have been used to improve the reliability of biosensors when they are deployed in differing situations [43], and to manage processes within host organisms to boost their capabilities as bio-factories [44]. However, the integration of control structures with larger synthetic biological designs remains difficult [45]. We can classify the challenges which impede this into three broad

categories, each of which must be overcome if Synthetic Biology is to make full use of the control engineering advances that have been instrumental to so many other fields of engineering.

Challenge 1 (design): Lack of systematic design tools and algorithms. Biological systems present a range of requirements (that are not present in traditional control engineering) that limit the range of control architectures that can be feasibly implemented. This necessitates the creation of new tools and algorithms for forward design that guide the development of control architectures that are specifically tailored to biological contexts.

Challenge 2 (build): Difficulty of Biological implementation: Once a control architecture has been designed the selection of particular biological parts, and their experimental implementation, remains difficult. This is a result of the context-dependent behaviour of biological components, the noise inherent to biological processes, and a range of secondary inter-dependencies that can arise when different systems are interfaced with one another [46].

Challenge 3 (test): Reliable characterisation of biological constructs: Once created in the laboratory, the performance of biological circuits must be experimentally characterised in such a way that results are reproducible in other laboratories, and useful for the forward design of new constructs. Achieving standardisation and reproducibility of this process remains challenging due to a lack of reliable experimental protocols and automation hardware.

This thesis contributes to addressing in part each of the above challenges, which together provide the components required for a classical engineering design-build-test cycle as applied to synthetic biology.

I approach **Challenge 1** from a traditional control theory / linear algebra perspective, in order to define sets of rules that must be satisfied by control architectures that achieve particular control goals (such as disturbance rejection), whilst also being feasible for biological implementation. This builds upon recent work which identified specific biological motifs that (in an idealised scenario) are able to reject disturbances to their constituent parameters [47–49], as well as studies of systematic design principles for biological systems [50]. My approach combines these two objectives (systematic design and disturbance rejection), and may therefore be used as a component in computer-aided-design (CAD) tools for biological systems, which have attracted much recent interest [51, 52].

Challenge 2 is addressed via a combination of biological experimentation and mathematical analysis, with which I study a number of different implementations of biological feedback systems. Many biological substrates have been employed for this purpose in the past: in this thesis I focus on using Recombinase enzymes, transcription factors, and sRNA, while other researchers have employed various regulatory proteins [36], inter-cell signalling architectures [53], or DNA itself [54]. In some cases multiple biological substrates have been combined to create architectures that operate over varying time-scales [55, 56] (done in this thesis by combining transcription factor and sRNA regulation) and species concentrations [41]. Different biological implementations offer differing strengths and weaknesses, and thus our contribution to this diverse toolbox of feedback controllers will be valuable as synthetic biology finds new applications.

Finally, to tackle **Challenge 3** I create an experimental hardware device, the *Turbidostat*, which automates and standardises typical Synthetic Biological characterisation experiments. Precursors to my device include bio-reactors for growth regulation [57] and optogenetics [58], which maintain cells at a constant density for prolonged periods whilst their behaviour is probed. These platforms have been widely adopted in synthetic biological research due to their ability to de-tangle the behaviour of synthetic circuits from their host cell's growth dynamics. However, until now the available Turbidostat-like devices have only

offered a subset of the capabilities required for complete characterisation of synthetic biological constructs, and have been difficult to use. My Turbidostat packages many required capabilities in a single low-cost platform, which performs many common characterisation experiments automatically (and can be easily programmed to execute others).

1.2 Thesis Outline

I now provide an outline of this thesis, briefly summarising the main contributions of each chapter.

Chapter 2: Synthetic Biology is a highly inter-disciplinary research area, which includes fundamental tools from biology, mathematics, computer science, control theory, and electrical engineering. This chapter begins with a brief overview of the fundamental biological processes and components that appear in the rest of the thesis, before describing a number of key experimental methods which will be refered to throughout. I then describe techniques for the mathematical modelling of biological systems, including initial model development, selection and fitting of parameters, and various algorithms for deterministic and stochastic simulation of differential equations. Also covered in this chapter are existing techniques for the forward design of biological systems, and past work on the implementation of synthetic biological control systems. These later aspects of biological engineering are reviewed in [59].

Chapter 3: This chapter presents a design framework for biological circuits that are capable of adaptation or disturbance rejection. I use traditional control theory to analyse a generalised system of linear ODEs, combining the structural requirements imposed by control theoretical analysis with constraints inherent in the biological implementations in question. This allows us to derive a set of rules for system structures and parameter regimes which must be satisfied for common control goals to be achieved. The results of this chapter were published in [60, 61].

Chapter 4: In this chapter I describe and analyse sRNA-based implementations feedback in synthetic biology. sRNA is used to tune autoregulatory architectures to reduce

noise and improve tunability, and to create novel feedback structures entirely. Results from this section were published in [42, 62].

Chapter 5: In this chapter I describe and analyse an implementation of feedback that employs integrase and excisionase proteins, which invert targeted regions of DNA, in order to reduce the cell-to-cell variability of a synthetic circuit's output. I then analyse and simulate a more complex feedback architecture, which additionally includes sRNA structures as described in Chapter 4. Some results from this chapter were published in [63, 64].

Chapter 6: In this chapter I design and simulate a synthetic biological system which aims to facilitate easy experimental measurement of cell-to-cell variability. Such measurements are required for the characterisation of noise-attenuating circuits like those described in Chapter 5. This chapter was motivated by the capabilities of the automation hardware described in Chapter 7. The results of this chapter were published in [65].

Chapter 7: In this chapter I describe the design, manufacturing, and testing of a novel turbidostat experimental device for characterisation of synthetic biological constructs. This device employs traditional (i.e. non-biological) control techniques to regulate many parameters that are important to cell growth and behaviour, including cell density, light intensity, temperature, and aeration. The results of this chapter are not yet published; we are currently investigating commercialisation of this new technology.

1.3 Contribution to Literature

The work described in this thesis has led to the following publications:

• H. Steel and A. Papachristodoulou, *Feedback Control and Synthetic Biology: Constraints on Design*, Proceedings of the 20th IFAC World Congress, 2017

• H. Steel, G. Lillacci, M. Khammash, and A. Papachristodoulou, *Challenges at the interface of control engineering and synthetic biology*, IEEE 56th Annual Conference on

Decision and Control, 2017

• H. Steel, A. Harris, E. Hancock, C. Kelly, and A. Papachristodoulou, *Frequency domain analysis of small non-coding RNAs shows summing junction-like behaviour*, IEEE 56th Annual Conference on Decision and Control, 2017

• A. Harris, C. Kelly, H. Steel, and A. Papachristodoulou, *The autorepressor: A case study of the importance of model selection*, IEEE 56th Annual Conference on Decision and Control, 2017

• T. Folliard, H. Steel, T.P. Prescott, G. Wadhams, L.J. Rothschild, and A. Papachristodoulou, *A synthetic recombinase-based feedback loop results in robust expression*, ACS Synthetic Biology, **6**(9) 2017

• H. Steel and A. Papachristodoulou, *Design constraints for biological systems that achieve adaptation and disturbance rejection*, IEEE Transactions on Control of Network Systems, **5**(2) 2018

• C. Kelly[1], A. Harris[1], H. Steel[1], E. Hancock, J. Heap, and A. Papachristodoulou, *Synthetic negative feedback circuits using engineered small RNAs*, Nucleic Acids Research, **46**(18), 2018
([1]joint first authors)

• H. Steel and A. Papachristodoulou, *Probing Inter-Cell Variability using Bulk Measurements*, ACS Synthetic Biology **7**, 2018

• H. Steel, and A. Papachristodoulou, *Low-burden biological feedback controllers for near-perfect adaptation*, ACS Synthetic Biology, 2019

In addition, the contributions of the following studies completed during this DPhil are not explicitly covered in this thesis:

• T. Folliard, B. Mertins, H. Steel, T.P. Prescott, T. Newport, C.W. Jones, G. Wadhams, T. Bayer, J.P. Armitage, A. Papachristodoulou, and L.J. Rothschild, *Ribo-attenuators: novel elements for reliable and modular riboswitch engineering*, Scientific Reports, **7**(1) 2017

• H. Steel, A. Sootla, B. Smart, N. Delalez, and A. Papachristodoulou, *Improving Orthogonality in Two-Component Biological Signalling Systems using Feedback Control*, IEEE Control Systems Letters, 2018

• X. Chen, H. Steel, Y. Wu, Y. Wang, J. Xu, C. Rampley, I. Thompson, A. Papachristodoulou, and W. Huang, *Characterisation and development of aspirin inducible biosensors in E. coli Nissle 1917 and SimCells*, Applied Environmental Microbiology, 2019.

• H. Steel, A. Sootla, N. Delalez, and A. Papachristodoulou, *Mitigating Biological Signalling Cross-talk with Feedback Control*, European Control Conference, 2019

• H. Steel, and A. Papachristodoulou, *The effect of spatiotemporal antibiotic inhomogeneities on the evolution of resistance*, Under review, 2019

2

Background

2.1 Biological Background

The aim of this section is to provide the reader with a introduction to the biological concepts that will be referred to throughout this thesis. Most of the biological processes and components in question are complex, and have many aspects that are still not fully understood. Consequently, our descriptions represent drastically simplified models of each process, providing only the level of detail required to enable understanding of later sections. Furthermore, each process will only be discussed in the form present in Bacteria, specifically *Escherichia coli* (or *E. coli* for short). *E. coli* is a common model organism used in laboratory research throughout the world (and all experiments described herein) due to its ease of manipulation, extensive past study, relevance to human applications (strains are found in the human intestine), and short generation time ($\sim$ 30 minutes for each asexual cell division). The information outlined below, if not cited, can be found in standard texts such as Molecular Biology of the Cell [66] and Molecular Genetics of Bacteria [67].

2.1.1 Dramatis Personæ: DNA, RNA, and Protein

Deoxyribonucleic acid (DNA) is the information storage molecule fundamental to all known life. It is a linear polymer comprised of four nucleotide bases, guanine (G), adenine (A), thymine (T), and cytosine (C). The order in which these appear is used to encode information, such as genes (sequences which encode a molecule that has a

function). In its natural conformation DNA forms a double stranded helix structure, with the two strands' sequences of bases generally mirroring one another (with C opposite G, and A opposite T). A range of tertiary structures and other modifications (such as supercoiling [68]) can impact its function (e.g. by making it difficult for certain genes to be read). When cells divide or reproduce, DNA is copied with a high degree of accuracy, though occasional mistakes occur (which provide one source of mutations that drive an organism's evolution). DNA (particularly the genome) has great longevity, aided by a number of cellular mechanisms that are dedicated to repairing any damage that may arise.

Ribonucleic acid (RNA) is a similar molecule to DNA, which is created during the transcription process (see Section 2.1.2). Unlike DNA it is only single stranded, and the base thymine is replaced by uracil (U). RNA is significantly less stable than DNA (it contains ribose, as opposed to deoxyribose of DNA), and is additionally targeted for degradation (and recycling) by a number of cellular mechanisms. RNA is the intermediate molecule in the central dogma of molecular biology (Section 2.1.2), which is translated to create proteins. That said, RNA molecules themselves can have useful capabilities. For example, they can bind to other RNA thereby affecting their function [42], or form complex secondary structures that sense the presence of other molecules [69].

Proteins are linear polymers comprised by a sequence of amino acids (of which there are 20 common naturally occurring kinds) which are translated (Section 2.1.2) from coding RNA. Proteins are the work-horses of the cell, responsible for a diverse range of functions such as catalysing reactions, controlling gene expression, transporting and recycling resources, and physically moving the cell. A protein's function depends upon its three dimensional structure, which is largely defined by its amino acid sequence (as each amino acid has differing chemical properties). The prediction of a protein's structure from its amino acid sequence, and of its function from its structure, are both difficult open problems in biology. These challenges make the creation of synthetic proteins difficult (though some progress has been made in recent years [70]), and thus the work in this

thesis will generally co-opt proteins found in nature. Proteins are recycled within the cell at varying rates, depending upon their structure and amino acid sequence.

2.1.2 The Central Dogma of Molecular Biology

The central dogma of molecular biology (Fig. 2.1) describes the fundamental process via which information stored in DNA is used to create functional proteins. This process is at the core of the synthetic biological control architectures described in this thesis, and so will be outlined in this section.

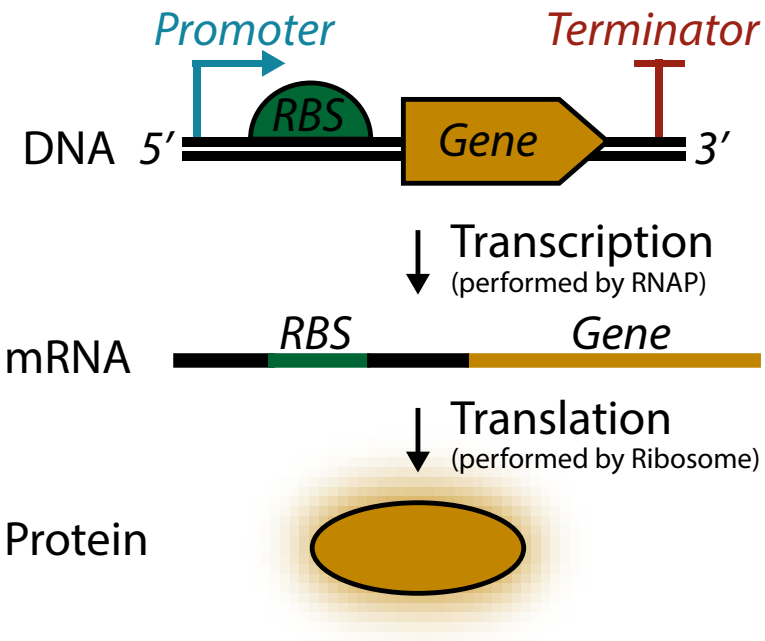


Figure 2.1: The central dogma of molecular biology. The double-stranded DNA segment (black parallel lines) contains a promoter (blue arrow), a Ribosome Binding Site (RBS, green), a gene to be expressed (orange) and a transcriptional terminator (red). In many diagrams the RBS and terminator will be omitted for simplicity; they can be assumed to be present on either side of the gene when not illustrated. The DNA is transcribed by RNA-Polymerase (RNAP) to produce messenger RNA (mRNA), which is in turn translated by a Ribosome to form a functional protein.

Transcription is the process via which DNA is read by an RNA Polymerase (RNAP, a complex molecular machine) enzyme to create an mRNA strand of complementary nucleic acid sequence. The process begins at a gene's upstream end (towards the DNA's 5' terminus) with the RNAP recognising and binding to a promoter region of the DNA. The binding of RNAP to DNA is regulated by a number of factors, including the strength

of the particular promoter sequence in question (which depends upon its nucleotide sequence), as well as the presence (or absence) of transcription factors (loosely defined as proteins which may bind to/near the promoter to encourage/block RNAP binding). Once bound, the RNAP separates the two DNA strands (locally pulling apart their natural double helix), and builds the mRNA molecule one nucleotide at a time. When the RNAP reaches a terminator region (defined by another DNA sequence, which can again be of varying strength) transcription ceases, and the RNAP and mRNA separate from the DNA.

Translation is the process via which an mRNA is read by a Ribosome (another molecular machine) to create a polypeptide (a chain of amino acids, also known as a protein). This process begins at a Ribosome Binding Site (RBS), a short sequence of nucleotides towards one end of the mRNA which is recognised by the Ribosome (the strength of this recognition is again dependent upon the particular sequence). The Ribosome then translates the nucleotide sequence of the mRNA (which is comprised of four kinds of nucleotides) into a sequence of amino acids (of which there are 20 naturally occurring possibilities). To overcome this numerical mis-match nucleotides are read in groups of three (referred to as a codons), of which there are 64 possibilities (i.e. 4^3), allowing many amino acids to be coded for by multiple codons. Translation continues until the Ribosome encounters a stop codon, at which point it separates from the mRNA and releases the amino acid sequence it has created.

Once separated from the Ribosome (or in some cases, even when still attached) the polypeptide chain can begin to fold into its final functional structure. The folding process involves complex thermodynamic interactions, which can be further regulated a number of environmental factors (such as temperature) or other cellular species (such as Chaperone proteins, which are specifically tasked with regulating the folding of other protein species). In some cases there may be additional post-transcriptional steps which must take place before a protein is able to perform its intended function, such as dimerisation (two proteins binding to form their functional complex).

Both mRNA and proteins are broken down into their individual components (nucleotides / amino acids) and are recycled by the cell, a process which is mediated by Ribonuclease and Protease enzymes respectively. In the case of proteins (which typically have a longer lifetime than mRNA) dilution due to cell growth can also play a significant role in regulating their intra-cellular concentration over time.

2.1.3 A Synthetic Biologist's Toolbox

In this section I describe a range of biological tools and components that are commonly used in synthetic biology, and touch on methods for tuning their behaviour. This list is by no means exhaustive; only components of central importance to this thesis are discussed. For an in-depth review of other synthetic biological "tuning dials" see [26].

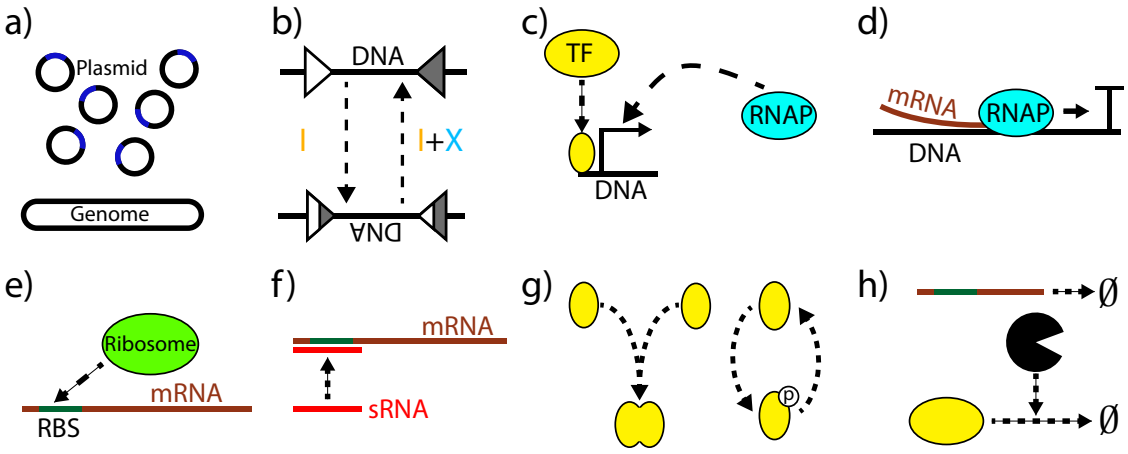


Figure 2.2: Common components and techniques found in a synthetic biologist's toolbox. **a)** Synthetic circuits can be inserted into the Genome or Plasmids, with potential differences in behaviour. **b)** Integrase (I) and Excisionase (X) proteins can flip DNA regions between recognition sites (triangles), thereby regulating circuit behaviour. **c)** Promoters vary in strength (rate of transcription initiation), and can be regulated by additional transcription factor (TF) proteins. **d)** Transcriptional terminators can be of varying efficiency, allowing RNAP to pass by and transcribe downstream genes. **e)** The strength of ribosome binding sites (RBS) can be varied to tune the rate at which RNAP is recruited and translation initiated. **f)** Short RNA (sRNA) can be designed to bind to an mRNA to obscure the RBS, thereby preventing translation initiation. **g)** Post-translational protein modifications such as dimerisation (left) or phosphorylation (right) can change a protein's behaviour. **h)** mRNA and Proteins degrade at varying (sequence dependent) rates, which can be actively regulated by specific enzymes (here a protease is pictured degrading the protein).

DNA Copy Number and Location: Synthetic Biological circuits are generally encoded within a cell as DNA in either Genomic or Plasmid form (Fig. 2.2a). Each cell has a single Genome (which is copied within the cell prior to division), which stores most

of its natural genetic code required for survival and self-replication. Synthetic genetic material can be inserted into the genome, however, this process can be experimentally intensive, and the transcriptional behaviour of the introduced code can be substantially context dependent (i.e. it matters *where* in the genome it is placed [68]). As an alternative to genome integration, synthetic biologists frequently introduce their constructs to a cell using Plasmids, which are short circular DNA strands (used by bacteria in nature to transfer genetic information) that can be easily manipulated in the laboratory. In contrast to the genome many copies of a given plasmid may be present within a cell (ranging from 2-3 to hundreds per cell), with the approximate number depending upon the plasmid's origin of replication (ORI). The regulation of the number of plasmid copies per cell is, however, imprecise, meaning that the absolute concentration of any synthetic protein that they may encode can vary substantially between cells.

DNA Inversion: Synthetic biological circuits can be regulated at the base DNA level via (among other methods) Recombinase enzymes [71], which are able to invert or excise regions of DNA (Fig. 2.2b). These regions are flanked by recognition sites which are altered in the DNA flipping operation (which is mediated by an Integrase protein), and the reverse flipping can only proceed in the presence of an additional Excisionase protein. DNA inversion is an effective means for regulating transcription (by inverting promoters or genes to change their behaviour), and can also be used for long-term information storage (since the state of the DNA is maintained when the genome/plasmid containing it is copied) [72] or logic operations [73].

Promoter Strength and Transcription Factor Regulation: Transcriptional promoters (Fig. 2.2c) can be broadly classified into two groups, constitutively active or regulated. The former category interact solely with the cell's native transcriptional machinery, and are thus always "ON" (able to initiate transcription). The rate at which transcription occurs (e.g. in terms of number of initiation events per unit time) depends upon the particular promoter sequence, and thus tuning of constituve promoter strength is a straight-forward method for controlling the output concentration of an RNA molecule

(and hence any proteins it encodes). Unlike constitutive promoters, the activity of regulated promoters (which are again defined by a particular sequence in DNA) can be influenced by interactions with additional proteins within the cell. For example, transcription factors (TFs) are a broad class of proteins which can recognise and bind to a specific DNA sequence in the promoter region, and in doing so potentially increase (*activating* TFs) or decrease (*repressing* TFs) the rate of transcriptional initiation. Engineered TF networks can be used to implement a wide range of useful regulatory architectures, and as such have been used as fundamental building blocks for significant synthetic biological circuits such as the toggle switch and repressilator [4, 5]. The behaviour of many transcription factors is additionally dependent upon the concentration of "inducer" molecules to which they may bind [74]. Varying the concentration of a specific inducer (which interacts with a transcription factor) is thus a common approach to introducing external control signals to a synthetic biological system.

Terminator Strength: Transcriptional terminators (Fig. 2.2d) provide another point of tuning for synthetic circuits: Their (DNA sequence-dependent) strength can be varied to allow some proportion of RNAP to pass by, such that any downstream genes can also be transcribed into mRNA (though at a lower rate than those upstream of the terminator) [75]. That said, for most synthetic circuits in this thesis transcriptional terminators with $\sim 100\%$ efficiency are employed.

RBS Strength and Riboswitches: Ribosome Binding Site (RBS) strength (rate of translation initiation) depends upon local mRNA sequence and, thanks to tools for sequence-based strength estimation [17], provides a (relatively) straight-forward method for tuning protein output levels of synthetic circuits (Fig. 2.2e). RBS and promoter strength combine multiplicatively to determine a protein's eventual concentration, though their individual strengths can significantly impact protein concentration variability [76]. It is also possible to regulate translation initiation using Riboswitches/Riboregulators: These are sequence-dependent mRNA structures that impact the Ribosome's ability to recognise and bind to the RBS region, and may potentially do so in a ligand dependent

manner (e.g. requiring of a secondary inducer molecule). They thus provide synthetic biologists with another approach for adding external control (i.e. via addition of chemical inducers) to their designs [77].

sRNA-mRNA Interactions: short RNA (sRNA) are a broad family of non-coding (i.e. they are not translated into proteins) RNAs which can be targeted (via a complementary nucleotide sequence) to interact with particular mRNA (Fig. 2.2f). The designs in this thesis typically employ sRNA that block the RBS region of an mRNA in order to prevent (or significantly decrease) its translation, as described in [18]. Such sRNAs are a valuable tool for synthetic biologists, since their sequence can be tuned to specify both their target (i.e. an arbitrary mRNA from which translation should be repressed) as well as their binding strength (which can be predicted to some degree using thermodynamic principles [18], and determines the degree of translational repression). In nature, non-coding RNAs perform a diverse range of roles in regulatory networks (for example, some sRNAs influence mRNA degradation rates rather than translation initiation), which are topics of substantial current research [78].

Post-Translational Protein Modification: Following translation, proteins can undergo a wide range of changes or modifications that impact their behaviour [26] (Fig. 2.2g). For example, many transcription factors are only active in their dimerised form (a complex of two identical proteins). In other cases the function of a protein can be changed by a phosphorylation reaction (addition of a phosphoryl group) which can be mediated by kinase and phosphatase enzymes [79]. Post-translational protein modification provides a powerful tool for synthetic biologists since these reactions can occur on fast timescales and do not consume a large quantity of cellular resources, however, in many cases they may be difficult to apply experimentally due to the complexities of protein engineering and potential for cross-talk [80].

Tuning Degradation: The concentration at which any cellular species is maintained depends upon a balance between its creation and degradation/dilution rates [81]. In

contrast with the above methods (which primarily focus on controlling the rate of production of an mRNA or protein), we can also build and tune synthetic biological circuits by changing the rate at which their constituent species are removed from the cell (Fig. 2.2h). The stability of mRNA molecules (whose average lifetime is on the order of minutes) is dependent upon several factors (such as their nucleotide sequence) and can be tuned by interactions with sRNA [82]. Proteins (with average lifetimes on the order of hours) show similar natural variation in degradation rate, and are diluted substantially as cells grow and divide. Furthermore, synthetic biological circuits can employ protease enzymes, which can be targeted toward specific proteins to regulate their rates of degradation (and hence intra-cellular concentration) [83].

2.1.4 Biological Noise

Noise and variability in biological systems is one of the factors that makes their reliable design so challenging. When measured experimentally (using techniques such as flow cytometry) even simple systems (such as the gene expression process described in Fig. 2.1) can present output levels (i.e. protein concentration) that vary over multiple orders of magnitude. In this section I briefly describe noise from the biological standpoint. This provides a background for future discussion of control systems that aim to mitigate its impact (Section 2.2.1), as well as methods for its mathematical modelling (Section 2.4.5).

In the biological context noise is classified into two broad categories which are defined relative to the particular sub-system that is being studied. These are intrinsic (due to stochasticity of the system's constituent biochemical processes) and extrinsic (due to fluctuations in the concentration and state of cellular components external to the system that nevertheless interact with it) noise types [84, 85]. Extrinsic noise sources include both short- and long-term phenomena: A circuit's behaviour can be influenced by temporary fluctuations in cellular machinery abundance [86] (such as ribosome sequestration, which impacts translation rate in synthetic circuits [87, 88]), or epigenetic differences in proteome and cell state that are passed on as cells divide [89–91]. Uneven

distribution of proteins or plasmids during cell division [92] can similarly introduce extrinsic fluctuations, as can environmental variability experienced between cells (for example, temperature heterogeneities). Total gene expression noise is thus a function of both cell-wide fluctuations, as well as variability in gene-specific regulation [93]. In general it can be challenging to identify the contribution that each noise type has to a system's total noise level [94] due to the interplay between different sources [95], though in many synthetic systems extrinsic noise is the greater contributor [65, 84, 95, 96].

In certain circumstances noise can be beneficial for a biological system or organism: In natural systems variability between cells provides an evolutionary advantage in changing environments [97, 98], and can enhance information transfer in genetic networks [99]. Indeed, on the most fundamental levels natural systems can benefit from its presence; the intrinsic variability in biochemical regulatory architectures can improve their sensitivity beyond that achievable by a purely deterministic system [100]. Motivated by this, synthetic systems have been developed which similarly benefit from the stochasticity of gene expression [48], and it is often necessary to include noise in their mathematical treatment to reproduce fundamental aspects of their behaviour [101, 102].

However, in many cases noise can negatively influence the function of synthetic biological circuits [103, 104]: fluctuations in the concentration of individual elements of a synthetic network can propagate throughout the system, impacting the behaviour of other components [86, 105]. This has motivated the design and implementation of synthetic systems that reduce variability [31], for example via inclusion of feedback control architectures [14, 32, 63] (discussed further in Section 2.2.1). So far biological engineers have largely attempted to carry across deterministic design approaches from traditional engineering fields, in many cases not accounting for the different limitations and constraints inherent to biology. Looking forward, scientists therefore aim to develop design philosophies which include noise at their most fundamental level, with a view to realising systems which (like many in nature) actually benefit from the presence of noise.

2.2 Forward Design and Control in Synthetic Biology

In the early days of synthetic biology it was hoped that, with sufficient characterisation and study, biological parts and components would become as reliable and modular as in traditional fields such as electrical engineering and computer science. This motivated the development of libraries of biological parts (such as the BioBrick library [106]), and tools for their automatic assembly to achieve user defined computational functions [51, 107]. However, when novel components are employed, or existing components are transplanted beyond highly regulated laboratory environments, plug-and-play design approaches prove to be unreliable. Significant variability in behaviour is observed due to a range of causes (many of which are still poorly understood), including the dependence of gene expression on surrounding DNA sequences, cross-talk and resource competition (which introduces unintended dependencies between processes occurring in the same cell), and the influence of a number of environmental factors on protein expression and function [22]. Gene expression noise, as outlined in Section 2.1.4, combines with these factors to make the behaviour of even the most simple synthetic constructs difficult to predict. To overcome these challenges, and in doing so improve the reliability and modularity of engineered biological systems, researchers have attempted to design and implement control architectures similar to those employed ubiquitously for this purpose in traditional engineering [28].

2.2.1 Biological Controllers and Forward Design

A range of methods have been proposed to aid the design of synthetic biological networks that achieve specific control goals. The first challenge in this process is the definition of an appropriate network topology, which can be approached via traditional control-theoretical methods whilst considering some biological limitations on component interactions (see Chapter 3). Constraints that must be satisfied for systems to achieve adaptation [45], disturbance rejection [49], or switch-like responses [108] have been explored, as have state-space approaches to finding design constraints for both linear [109] and non-linear [110] systems. Though these approaches can be used to find networks that are in some way minimal (for example, requiring the fewest reactions to realise), this does not guarantee that their implementation will be feasible, or that they will perform robustly. Particular

parameter value regimes can be selected during the design phase such that these constraints are satisfied (though they must then be achieved in implementation), using methods that parallel those commonly used in control and electrical engineering. For example, mathematical treatments of biological oscillators (such as the repressilator [5]) can guide parameter selection to meet specific frequency or amplitude goals [111].

The tuning of feedback systems via adjustment of their individual parameters and components is an ongoing challenge in synthetic biology. As is known from control engineering, certain fundamental constraints limit achievable controller performance (for example, the waterbed effect has been observed in yeast [112]), though in many cases experimentally realising desired parameter values poses a greater difficulty. As described in Section 2.1.3 this is often done by selecting RBSs or promoters of appropriate strengths [113], though their behaviour can be context-dependent [114]. Critical in this balance is the concept of *burden*, which is broadly used to describe the consumption of available cellular resources by synthetic circuits. Thus, if a synthetic (or natural) circuit is creating its components or products at a very high rate, the overall pool of free cellular resources (such as ribosomes and charged tRNAs) can be substantially reduced [87] (thereby "burdening" the cell). This may impact other cellular processes in a difficult-to-predict manner, potentially decreasing the rate at which certain proteins are produced and reducing cell growth rate [115, 116]. It is therefore desirable that synthetic control systems function at low component abundances, or build upon native cellular machinery to reduce the number of additional genes that must be expressed [48].

Recent work has demonstrated that natural systems can be built upon and re-purposed in this way by as little as the addition of a single appropriately acting protein. For example, inclusion of a new interaction in a phosphorylation cycle can re-purpose a system with a robust step-response to provide a user-defined output level, or even near-perfect adaptation [117]. This design approach is doubly beneficial; the introduced burden is small (since only a single additional protein is expressed), and due to its simplicity, experimental realisation of such a control system can be relatively straightforward. When considering

the burden imposed upon a cell by synthetic control systems, DNA- or RNA-based systems (such as strand displacement reactions) are similarly attractive, since they do not rely on resource-intensive protein production, whilst being able to perform the information processing necessary for effective control [115]. That said, the applicability of DNA-based frameworks can be limited as many can only be performed *in vitro*.

Synthetic biological systems that combine processes operating on widely varying timescales provide a range of benefits [118], including reducing the challenge presented by realisation of specific network structures in order to guarantee certain control behaviours. For example, systems that use a layer of fast-acting control components to regulate a layer of slower processes shift design challenges away from definition of precise network structures, instead relying on selection and tuning of parameter values in the fast layer [55]. Though this may in some cases be an inferior option, in others achieving a given network structure may prove problematic for reasons that include the introduction of secondary "virtual" network connections due to burden [88]. Similar secondary effects can be responsible for context-dependence and unforeseen circuit behaviours [22], though recently a number of approaches have been developed via which they may be overcome. One such approach is the utilisation of orthogonal transcription and translation machinery [119, 120], which are only used by the introduced synthetic systems and are not substantially disturbed by a cell's natural behavioural variations. Other undesirable secondary effects include the "load" that down-stream elements can apply to connected upstream components by consuming (or temporarily sequestering) them [121]. In electrical engineering designs often include parts that "insulate" up-stream components from those with which they interface, which has inspired the development of biological load drivers [122], which use fast-acting phosphotransfer processes to provide an insulating buffer between slower components [46]. Similar architectures can be used to implement amplifier-like circuits [123].

For many biological processes, robustness (a system's insensitivity to fluctuations in its components' properties and to its surrounding conditions) can be critical for reliable functioning [124]. Systems that are designed without consideration of how variation in

their individual parameters will impact their performance are thus at the mercy of noisy biochemical processes (described in Section 2.1.4), which can perturb or completely disrupt their desired behaviour [28]. Much research has been done to create control systems that can adapt to such disturbances [36], which will be described further in Chapter 3.

2.2.2 Approaching Feedback Control

The biological implementation of negative feedback has been an ongoing goal of synthetic biology, with early efforts focusing on its most elementary form, negative auto-regulation. These synthetic architectures demonstrated an ability to speed system response times, reduce noise, and linearise dose-responses [29, 31, 32], and have found practical implementation in metabolic engineering [125]. Recent work has developed negative feedback systems using a diversity of genetic components [41, 126], including sRNA-based transcriptional regulation (Chapter 4) and integrase/excisionase based stochastic feedback (Chapter 5). To further advance these designs it will be necessary to develop systems that approximate the fundamental components of traditional control systems, including integration, gain, and summation junctions [38, 39], which can be used to build Lead-Lag [40], PID [41], or more advanced controllers.

Like negative feedback, positive feedback architectures have been observed in a range of natural systems [127], and are now being included in synthetic biological designs. In their most fundamental implementations, synthetic positive feedback loops have been used to amplify biological circuits' responses to inducers [128], and when nested with negative feedback circuits can provide robustness and tunability for circuits such as genetic oscillators [129]. Positive feedback is also of use in the creation of bi-stable switches, as it encourages fast and permanent transition between switch states [130]. Feed-forward loops, which can be created by combining positive and negative regulatory actions [131], are well studied in natural processes [132], and provide topologies capable (in theory) of perfect adaptation [47].

Beyond biological contexts, a variety of feedback controllers that interface with synthetic biological systems have been implemented *in silico*. These take advantage of hardware setups that measure and control a cell colony's growth and gene expression [58, 133], which can minimise many major experimental noise sources (discussed further in Chapter 7).

2.3 Experimental Techniques in Synthetic Biology

In this section I provide a brief overview of select experimental techniques and procedures in synthetic biology that will be relevant to the remainder of the thesis.

2.3.1 Creating New Circuits

Synthetic biological circuits are defined in terms of a DNA coding sequence by combining genetic elements (e.g. nucleotide sequences for promoters or proteins) which can be found in a range of online libraries [106], or extracted from existing organisms and engineered circuits. A desired DNA coding sequence can then be physically created by cutting and re-combining existing DNA (e.g. using restriction enzymes), or by its direct synthesis (an increasingly cost-effective process performed by a number of commercial operators). As discussed in Section 2.1.3 this DNA will then be inserted into a host organism's genome, or a plasmid backbone (the method employed in the remainder of this thesis). Once created, plasmids (initially suspended as a solution of raw DNA) can be introduced to host cells (in our experiments, *E. coli*) via a procedure known as *transformation*. A number of techniques for this exist, of which a common example is heat-shock transformation (where cells are temporarily exposed to a high temperature which causes gaps to form in their outer membrane, allowing plasmids to enter). Plasmids used in synthetic biology will typically include an antibiotic resistance gene, meaning that following transformation cells can be grown in the presence of an antibiotic and only those which have taken up the plasmid (essential for the cell's proliferation) will survive. Appropriate antibiotics are included in subsequent characterisation experiments to force cells to retain their synthetic plasmid DNA (and whatever additional functions it may program).

2.3.2 Synthetic Circuit Characterisation

In order to characterise the behaviour of a synthetic biological construct it is necessary that it has some kind of experimentally measurable output or impact upon its host cell. Commonly this is achieved via use of fluorescent proteins, such as GFP (Green Fluorescent Protein), which when excited by a light of a particular wavelength will emit a red-shifted spectrum that can then be measured. The intensity of the emission spectrum can be assumed to be (approximately) proportional to the fluorescent protein's concentration, allowing this variable's value to be estimated. Typically a fluorescent protein will be included in a synthetic circuit such that it is expressed in tandem with the circuit's output (e.g. another protein), or in some cases it is directly fused to a protein of interest to allow direct measurement of that protein's concentration. Due to their ease of use, fluorescent protein reporters are employed widely in biological sciences and (in addition to estimating concentration) can be used to study the localisation of a range of molecules [134], as well as the dynamics of inter-protein interactions (e.g. using FRET [135], which is discussed in Chapter 6).

With a fluorescent protein as output, characterisation experiments are (at least in this thesis) then performed by growing cells in growth media (liquid "bacteria food" of various kinds) in containers (of various volumes), generally inside an incubator (to maintain an optimal temperature of $\sim 37°$C for *E. coli*) and shaking/stirring apparatus (which agitates the media to ensure oxygenation and homogeneity). Inducer chemicals (which can control synthetic circuits as described in Section 2.1.3) can then be added at various time intervals, and samples (of cells suspended in growth media) can be extracted for analysis. One common measurement tool is a plate reader, a bench-top robot which accepts "plates" consisting of a number (often 96 per plate) of small-volume chambers containing samples. Plate readers are able to measure the optical density of a sample (which can be used to estimate the number of cells per unit volume), and fluorescent protein emission/excitation spectra. Many plate readers have a range of additional measurement capabilities (omitted here for brevity), and may even be able to incubate and shake their plate (allowing

experiments to be automated by the robot).

Another common measurement technique is flow cytometry, which (using a different hardware device) can rapidly sort cells and measure their individual fluorescence levels. Such an experiment will often measure $10^4 \sim 10^5$ individual cells, allowing the distribution of fluorescence levels within a population of cells to be estimated. This contrasts with plate reader measurements which return only the mean fluorescence level of a population. A major challenge faced when using both flow cytometry and plate reader hardware tools is the need for calibration and standardisation: without this measurements can vary significantly even between (theoretically) identical devices from the same manufacturer, or in a single device over time [136]. A range of hardware calibration protocols have been developed to overcome these problems, thereby enabling accurate and reproducible measurements [137, 138]. Further calibration is then required if direct measurements of fluorescent reporter concentrations are required (e.g. in units of protein concentration), rather than the raw measurements of fluorescence output by measurment hardware (generally expressed in arbitrary units) [139].

2.3.3 Automating Experimental Challenges

For many synthetic circuits, characterisation experiments will ideally be run for long periods (> 1 day) in order to assess their steady-state behaviour, or to measure their response to time-varying inducer inputs. However, when cells are grown in a fixed concentration of media they will typically exit their exponential growth phase (during which an *E. coli* population can double in size every ~ 30 minutes) due to exhaustion of resources after $5 \sim 10$ hours. In order to maintain cells in exponential growth it is thus necessary to dilute them (into fresh media) at regular intervals to maintain an abundance of free resources. This can be done manually (usually each $2 \sim 3$ hours), though this is very tedious for researchers and introduces noise to the experiment (because of discontinuities in cell density). This process can be automated using liquid handling machines, though due to their high cost and complexity of setup they are impractical for

most synthetic biology research groups to operate. A superior alternative, in terms of both efficiency and data quality, is a Turbidostat/Chemostat device like that developed in Chapter 7 of this thesis. This device continually pumps in new (and pumps out old) growth media in order to maintain a constant cell density, and (in the case of our device) can also regulate the temperature, aeration, and measure fluorescence, over the course of a multiple-day experiment.

2.4 Modelling Biological Systems

In this section I outline methods for the mathematical modelling of biological systems which will be employed in the rest of this thesis. I explain each step in the modelling process using a simple example of constitutive gene expression (Fig. 2.3), as was presented in Fig. 2.1. To appreciate the utility and philosophy of mathematical modelling in the biological sciences it is worth recalling two famous quotes:

All models are wrong, but some are useful - George Box

and

Everything should be made as simple as possible, but no simpler - Albert Einstein

Consequently, our general approach to modelling will be to devise mathematical descriptions of a system that are sufficiently complex to capture important aspects of its behaviour, and to make useful predictions about how these may change under the influence of various parameters of interest. Ideally these models will be made as simple as possible in order to reduce the number of free parameters, and to make differentiation between similar model structures feasible given the limited diagnostic tools at our disposal. This will generally mean that very complex biological processes must be simplified drastically, often to the point of being described by a single rate parameter.

To account for the influence of surrounding cellular processes, mathematical biologists have attempted to create whole-cell models with some degree of success [140], however, the application of these models has a large computational cost, requires estimation

of many parameters, and is difficult to fit to experimental data. Therefore, in this thesis synthetic biological constructs are modelled as single sub-systems of a far more complex cell, which is assumed to be able to maintain a (temporally) constant supply of transcription/translation machinery and other resources/components employed by our circuit. In many cases I will assess the validity of this assumption by (for example) estimating the impact that variability of cellular processes (and hence parameters of our model) has on our sub-system's behaviour.

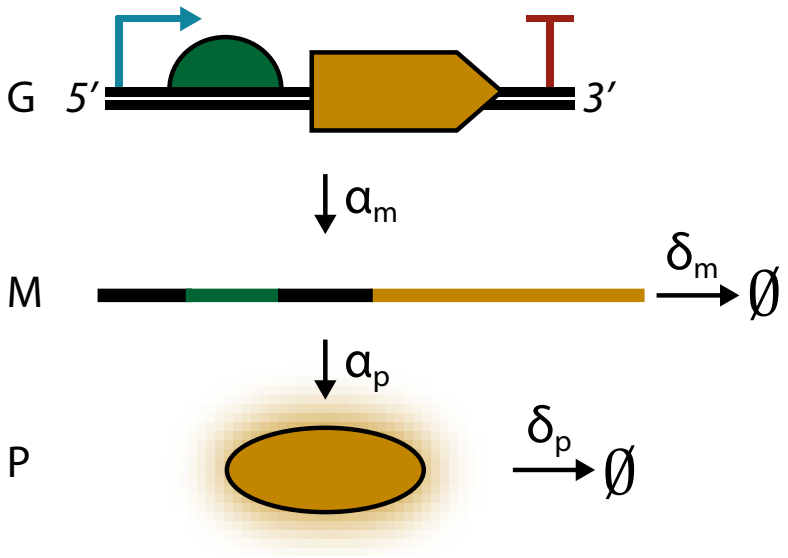


Figure 2.3: A simple model of gene expression. The DNA encoded Gene (G) is transcribed at rate α_m to create mRNA (M), which is then translated at rate α_p to create Protein (P). The mRNA and Protein are degraded/removed at rates δ_m and δ_p respectively to a null state $\emptyset$.

2.4.1 Chemical Reaction Networks

The first step in constructing a model of a biological process is to decide which biochemical interactions to account for. This may seem clear following a cursory glance at the case illustrated in Fig. 2.3, but the devil is in the details: each α and δ is itself a function of a number of dynamic biological processes happening throughout the cell. For example, α_p depends strongly upon the availability of ribosomes to mediate translation, and thus its value may change if these are being drawn upon by other processes [24, 120]. Ultimately approximations must be made, and so we assume that each of the processes depicted in our simple schematic of gene expressions proceeds at a constant rate. This system

can be described as a chemical reaction network of the form:

$$\begin{aligned} G &\xrightarrow{\alpha_m} G+M \\ M &\xrightarrow{\alpha_p} M+P \\ M &\xrightarrow{\delta_m} \emptyset \\ P &\xrightarrow{\delta_p} \emptyset \end{aligned} \tag{2.1}$$

The species on the left hand side of the reactions in (2.1) are referred to as the *reactants*, and those on the right as the *products*. For every reaction all reactants are required for it to occur, and when it does the outcome is the species listed on the right hand side. Thus, when a transcriptional event occurs the system goes from having one unit of G to one unit of G (the DNA it is not destroyed) and one of M. A more complex model of transcription might (for example) include intermediate states that account for binding of RNAP to the DNA, and then assembly of individual nucleotides, but in this case we combine the entire process into a single reaction occurring at rate α_m.

2.4.2 Systems of Ordinary Differential Equations

We can transform the chemical equations in (2.1) into a continuous description of each species' concentration by making two assumptions. First, that the numbers of each species are sufficiently high that they can be treated as continuous concentrations (rather than discrete values), and that these concentrations exist in a well mixed volume. It is then possible to employ the *Law of Mass Action* which states that the rate of a given reaction is proportional to the product of the concentrations of its reactants. For example, the transcriptional step in (2.1) proceeds at rate $\alpha_m g$ where I use lower-case g to represent the concentration of the gene component of our system. By considering whether each reaction creates or destroys its constituent species, we can thus express (2.1) as a system of differential equations. Each reaction contributes a term to the differential equation governing each species it consumes (in which case the term will be negative) or produces (the term will be positive), in this case giving:

$$\begin{aligned} \dot{m} &= \alpha_m G - \delta_m M \\ \dot{p} &= \alpha_p M - \delta_p P \end{aligned} \tag{2.2}$$

where the dot operator indicates a time derivative (i.e. $\dot{m} = \frac{dm}{dt}$). Note that no differential equation is required to govern g since (2.1) assumes the concentration of the gene to be constant (at no point is it created or destroyed). This assumption (which will generally be made in subsequent sections of this thesis) is based on a belief that DNA concentration is regulated by plasmid- or genome-copying machinery, which is assumed to be constant as it is not an active part of the system being studied.

2.4.3 Linearisation and Frequency Domain Analysis

Given a system of ODEs (e.g. as in (2.2)) a range of analysis techniques can be used to investigate various aspects of its behaviour. The particular system in (2.2) is linear (i.e. the right hand side is linear in each state variable) which makes its analysis easy, however, biological models will not always be so simple. A general first-order non-linear system describing the dynamics of n state variables $\boldsymbol{x}(t) = [x_1(t), \ldots, x_n(t)] \in \mathbb{R}^n$ can be expressed in the form:

$$\dot{\boldsymbol{x}}(t) = \mathbf{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)), \tag{2.3}$$

where $\dot{\boldsymbol{x}}(t) = \frac{d\boldsymbol{x}(t)}{dt} \in \mathbb{R}^n$ is a vector containing the time derivative of each state variable, $\boldsymbol{u}(t) \in \mathbb{R}^n$ is a vector containing the time-dependent inputs to the non-linear system's state variables, and $\mathbf{f}(\boldsymbol{x}(t), \boldsymbol{u}(t)) \in \mathbb{R}^n$ contains n general non-linear functions of the state variables and inputs.

To simplify analysis it is often advantageous to examine non-linear systems in a regime where they can be approximated by a linear system. To do this we first find an equilibrium point $\boldsymbol{x}_e \in \mathbb{R}^n$ by setting $\mathbf{f}(\boldsymbol{x}_e, \mathbf{0}) = \mathbf{0}$ and solving this system of equations for each state variable. Sometimes this can be done analytically (as in simple cases like (2.2)), though for more complex systems (particularly when there are non-linear dependencies between parameters) this may require numerical approaches (algorithms for this are provided in most programming languages). The system in (2.3) can then be linearised via a Taylor series expansion of $\mathbf{f}(\boldsymbol{x}, \boldsymbol{u})$ about the equilibrium given by $\boldsymbol{x}_e$ to give:

$$\dot{\tilde{\boldsymbol{x}}} \approx \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_e, \mathbf{0}} \tilde{\boldsymbol{x}} + \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{u}}\right|_{\boldsymbol{x}_e, \mathbf{0}} \boldsymbol{u}, \tag{2.4}$$

where $\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\Big|_{\boldsymbol{x}_e,\mathbf{0}} \in \mathbb{R}^{n\times n}$ is the Jacobian of $\mathbf{f}$ evaluated at $(\boldsymbol{x}_e, \mathbf{0})$, $\frac{\partial \mathbf{f}}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{x}_e,\mathbf{0}} \in \mathbb{R}^{n\times n}$ is a matrix of all partial derivatives of $\mathbf{f}$ with respect to elements of $\boldsymbol{u}$ evaluated at $(\boldsymbol{x}_e, \mathbf{0})$, and $\tilde{\boldsymbol{x}} = (\boldsymbol{x} - \boldsymbol{x}_e)$ is the deviation in $\boldsymbol{x}$ from $\boldsymbol{x}_e$. We can express the linearisation (2.4) as a general linear system:

$$\begin{aligned} \dot{\boldsymbol{x}}(t) =& \mathbf{A}\boldsymbol{x}(t) + \mathbf{B}\boldsymbol{u}(t), \\ \boldsymbol{y}(t) =& \mathbf{C}\boldsymbol{x}(t) + \mathbf{D}\boldsymbol{u}(t), \end{aligned} \tag{2.5}$$

which is the standard form often employed in control engineering. In this thesis we will generally have $\boldsymbol{D} = 0$, and for single-output systems $\boldsymbol{C}$ will be an indicator vector (one element equal to 1 and 0 otherwise) that selects whichever state variable is considered as the system's output.

A range of approaches from control theory can be used to analyse systems in the form of (2.5). One common technique is to convert the system from its state space form (i.e. that in (2.5)) to a transfer function, which facilitates analysis of its response to input functions of varying frequency. The first step in this process is to take the Laplace transform of (2.5), which gives:

$$\begin{aligned} s\boldsymbol{X}(s) =& \mathbf{A}\boldsymbol{X}(s) + \mathbf{B}\boldsymbol{U}(s), \\ \boldsymbol{Y}(s) =& \mathbf{C}\boldsymbol{X}(s) + \mathbf{D}\boldsymbol{U}(s), \end{aligned} \tag{2.6}$$

where s is the frequency domain (Laplace) parameter. To assess the system output's dependence upon its input we want to solve for the ratio of $\boldsymbol{Y}(s)$ to $\boldsymbol{U}(s)$, which we can do by first solving for $\boldsymbol{X}(s)$:

$$\boldsymbol{X}(s) = (s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B}\boldsymbol{U}(s) \tag{2.7}$$

and then deriving the output/input transfer function $\boldsymbol{G}(s)$:

$$\boldsymbol{G}(s) = \frac{\boldsymbol{Y}(s)}{\boldsymbol{U}(s)} = \boldsymbol{C}(s\boldsymbol{I} - \boldsymbol{A})^{-1}\boldsymbol{B} + \boldsymbol{D} \tag{2.8}$$

There are many techniques for analysing transfer functions of this form (for an introduction see [141]) which I will introduce throughout the thesis as they are needed.

2.4.4 Deterministic Modelling

Typically, characterisation (as described in Section 2.3.2) of synthetic biological systems will include measurement of both their "steady state" output, as well as their dynamic (time-dependent) response following a change in their input. It is therefore useful to model the time-evolution of a system as it departs from a given starting state. Though in some cases this will be possible analytically (for example, the very simple system outlined in (2.2)), in most cases deriving a closed-form solution will be impossible and thus numerical approaches are required. A range of tools exist for numerically integrating a system of differential equations (as in (2.3)), such as the solvers built into common programming languages, or simple iterative algorithms such as the Runge-Kutta 4th order method [142]. Time-varying input signals can be supplied to these algorithms, allowing us to assess a model's behaviour following (for example) a step change in its inducer concentration.

Some synthetic biological systems are bistable [90] (sometimes intentionally, but sometimes providing an undesirable surprise for their designers), meaning that they may have multiple equilibriums/stationary points (the solutions $\boldsymbol{x}_e$ to $\mathbf{f}(\boldsymbol{x}_e, \mathbf{0}) = \mathbf{0}$ in (2.3)). A number of theoretical studies have analysed various conditions under which bistability may arise in biological systems [143]. The initial conditions (state variable values) used in deterministic modelling can have a significant impact when bistability is present since, depending upon which equilibrium's basin of attraction the initial point lies within, the numerical simulation may converge to a different point. The selection of initial state variables for numerical simulations is often challenging, as these states are generally not directly measurable, and concentrations (of proteins etc.) in question will depend upon the conditions experienced by the cells prior to the beginning of a characterisation experiment. If synthetic circuits that require addition of a chemical inducer (which is not present prior to the experiment) for their operation are being studied then I will often assume that state variables can be initialised with concentrations of zero, though in more complex cases additional consideration of this point will be required.

2.4.5 Noise Modelling and Stochastic Simulation Algorithms

In order to make reliable models of biological systems it is essential that we have methods to mathematically account for the noise/variability that is expected in their output [144]. Variability can be observed as variation in time of the state of a single cell, or as variability between cells within a population at a single time point. Observing the former kind (variation in time) is difficult experimentally, as it requires isolation and observation of single cells for extended periods. In this thesis I will predominantly focus on the latter case, as this reflects the experimental measurements that we are able to perform with flow cytometry (see Section 2.3.2). Such measurements allow us to estimate probability distributions for a synthetic circuit's measurable output, and to assess the effectiveness of control systems that aim to counteract the influence of unwanted noise. Different methods are required for modelling extrinsic and intrinsic noise (defined in Section 2.1.4), which are outlined in this section.

Extrinsic noise is defined as that arising due to factors external to a system under study, and can (for example) arise as fluctuations in Transcription/Translation due to variation in the available pool of RNAP/Ribosomes that mediate these processes. This will occur over time in a single cell (i.e. due to dynamic processes in the cell sequestering resources in response to environmental changes), and between cells in a population (due to phenotypic variation in cell state / proteome). As we are predominantly concerned with the latter form of variation we can model extrinsic noise by allowing parameter values to vary between cells within a simulated population. Typically this will involve simulating a large number of cells with one or many parameters sampled from a probability distribution. Regarding the choice of distribution, experimental studies have demonstrated that many biological variables of interest, such as concentrations of mRNA or protein, follow approximately log-normal distributions during steady growth [145] (though it may be better described by a normal distribution for cells in stationary phase [146]). Log-normal distributions have been shown to emerge in these cases due to the inherent complexity of the noisy biochemical processes involved [147]. With a distribution chosen, we are presented with the significant challenge of choosing its parameters (generally in

terms of a distribution's mean and standard deviation). Selecting mean parameter values is discussed in Section 2.4.6, and can generally be done with some degree of confidence. Selecting the width of the distribution in a given parameter (e.g. estimating how much translation rate will vary between cells within a population) is much more difficult as a parameter's effect would need to be isolated from the measured output of a circuit (which generally depends on many of its modelled parameters). Consequently, we typically aim to analyse the influence of extrinsic noise only from a qualitative perspective, often by introducing variability to different parameters in a model and seeing whether the observed patterns in output variation align with those found experimentally.

Intrinsic noise is defined as that arising within a biological system itself, primarily due to the stochasticity of the biochemical processes involved. The fact that biochemical species are present in integer quantities is significant when concentrations are small, because an integer change in the number of molecules per cell corresponds to a large percentage change in the overall quantity. In contrast, deterministic modelling techniques (as described in Section 2.4.4) treat state variables as being continuous, which can be acceptable when the concentrations (and hence number of molecules per cell) are large. To model stochasticity in a synthetic biological system we must therefore move beyond ODE descriptions (as in (2.3)), and instead treat its constituent reactions (as in (2.1)) as stochastic processes which can "fire" discretely, with the probability of this occurring depending upon the reaction rate [144]. A number of algorithms for stochastic simulations exist; the Gillespie Algorithm provides an exact method for simulating a chemical reaction network, but is computationally intensive, whilst the Tau Leaping Algorithm [148] approximates the Gillespie approach well in many cases, and is significantly more efficient. By running a large number of stochastic simulation trials using one of these algorithms we can estimate the degree of inter-cell variability (in terms of probability distributions of species' concentrations) that may arise in a biochemical system due to intrinsic noise.

Gillespie Simulation Algorithm

We can express a system of chemical reactions (as in (2.1)) in the general form:

$$\sum_i \boldsymbol{a}_{i,j}\boldsymbol{x}_i \xrightarrow{c_j} \sum_i \boldsymbol{b}_{i,j}\boldsymbol{x}_i \tag{2.9}$$

where j is the index of the chemical reaction (which would take values $j = 1, ..., 4$ in the case of (2.1)) and i is the index of the species involved. The $\boldsymbol{a}_{i,j}$ and $\boldsymbol{b}_{i,j}$ are the stoichiometric coefficients of each reactant/product (i.e. how many of each species are involved in the reaction), and c_j are the reaction rate constants. We will also define the quantities

$$\boldsymbol{d}_{i,j} = \boldsymbol{b}_{i,j} - \boldsymbol{a}_{i,j} \tag{2.10}$$

where $\boldsymbol{d}$ is referred to as a stoichiometry matrix, in which each element $\boldsymbol{d}_{i,j}$ reflects the net change in abundance of species i each time reaction j fires, and

$$e_j(\boldsymbol{x}) = c_j \prod_i \boldsymbol{x}_i^{\boldsymbol{a}_{i,j}} \tag{2.11}$$

which is the reaction propensity of reaction j according to the law of mass action. With these quantities defined, the Gillespie Simulation Algorithm proceeds as follows.

1. Set up the initial state variable values $\boldsymbol{x}$.
2. Calculate each reaction propensity $e_j(\boldsymbol{x})$
3. Assign probabilities to each reaction j which are proportional $e_j(\boldsymbol{x})$, randomly select a reaction j^* to occur according to this probability distribution.
4. Select a random timestep dt from an exponential distribution with parameter $\lambda = \sum_j e_j$.
5. Increment each species i by $\boldsymbol{d}_{i,j^*}$, and the simulation time by dt.
6. Repeat from step 2 until the simulation's maximum time t_{max} is reached.

To summarise, in each algorithm loop the Gillespie algorithm selects one reaction to "fire" based on the relative rates of all reactions in the system, and selects a random timestep from a distribution which depends on the sum of all reaction rates in the system.

The standard Gillespie algorithm outlined above is generally easy to implement, but is computational costly for systems that include a large number of reactions, or reactions occuring at widely varying timescales. This is because its fundamental procedure fires only one reaction during each algorithm loop, and as such when there are many reactions that can potentially fire the timestep dt will be very small (so it will take a long time to reach t_{max}). To address this shortcoming a range of approximations to the Gillespie Algorithm have been devised [148], one of which is the Tau Leaping algorithm which I now describe.

Tau Leaping Simulation Algorithm

The Tau Leaping simulation algorithm [148] differs from the standard Gillespie approach by using a fixed timestep, and allowing each reaction to fire (potentially multiple times) within this time interval. Using the quantities defined for the Gillespie algorithm above it can be outlined as follows:

1. Set up the initial state variable values $\boldsymbol{x}$, and select a timestep dt.
2. Calculate each reaction propensity $e_j(\boldsymbol{x})$
3. Sample a random number γ_j from a Poisson distribution with mean $e_j(\boldsymbol{x}) \times dt$, this is the number of times reaction j fires.
4. Increment each species i by $\gamma_j \times \boldsymbol{d}_{i,j^*}$, and the simulation time by dt.
5. Repeat from step 2 until the simulation's maximum time t_{max} is reached.

The fundamental assumption underlying the Tau Leaping approximation to the Gillespie algorithm is that we can assume each species concentration stays approximately constant during the interval dt, allowing us to update each species concentration simultaneously. The timestep is therefore independent of the number of reactions that are modelled in our system. This assumption can cause problems in some cases, particularly when species concentrations are very low: when this occurs it may be possible (for example) for a species initially present as a single molecule to be degraded/removed from the system twice within a single dt, leading to a negative abundance. This can be generally avoided by selecting a sufficiently small dt, or employing a range of approximation methods that can handle these cases of the Tau Leaping algorithm [149].

2.4.6 Choosing and Fitting Parameters

In some cases models which provide only qualitative insight into a synthetic biological construct's behaviour are sufficient. For example, they can reveal phenomena such as bistability, noise attenuation, and highlight particular parameter combinations that are important in defining a system's overall behaviour. However, in many cases we would like to devise models which quantitatively capture a system's behaviour, meaning that the parameter values chosen are important. Two broad approaches to selecting parameters will be employed in this thesis; the first is *a priori* estimation of their values based upon past studies in the literature or known properties of the system in question, and the second is *a posteriori* fitting of their values to experimental data. In this section I will briefly discuss each of these approaches.

Parameter estimation is frequently performed *a priori* via literature search; many biological studies have measured rates of production, degradation, binding, and other parameters of interest for a diverse array of biological systems [150]. In theory these parameters can be extracted from papers that present similar experimental situations to those anticipated in a synthetic biological circuit, however, in practice these parameter estimates are often inaccurate for a variety of reasons. For example, the context dependence and environmental sensitivity of biological circuits means that parameter values can vary substantially when they are interfaced with different components, or grown in different media or cell lines [114]. This leads to literature estimates of the same parameter often varying by many orders of magnitude [151]. In other cases it may be difficult to transfer parameter values from the literature to a model due to the diverse range of (non-standardised) measurement techniques that are employed in the field. Even assays as simple as fluorescence protein measurement (see Section 2.3.2) are rarely standardised, meaning experimental scaling factors (in this case between measured fluorescence and protein concentration) can vary greatly between different studies/laboratories. Some parameters (such as cellular growth rate) can be estimated with a good degree of accuracy (though as described in Section 2.3.3 they can change during the course of an experiment), but these are the exception rather than the norm. Overall, modelling using *a priori*

parameter estimations can be useful, particularly as a tool for examining the qualitative behaviour of a synthetic circuit, but will often result in models with poor quantitative alignment with experimental measurements.

To address the impracticalities of *a priori* parameter estimation, models will frequently be fitted to experimental data to provide better quantitative agreement in their behaviour. Literature abounds on the challenge of model fitting across a range of sciences [152]. The fundamental steps in this process are first defining an objective function (which in some way measures the discrepancy between the model prediction and experimental results), and then minimising this objective over a set of feasible parameter values to find the combination that results in best alignment between model and data. A range of model fitting tools exist in many programming languages (such as *fmincon* in MATLAB), and so we do not need to implement these ourselves. Implicit in the parameter fitting process are a range of challenges surrounding the feasibility of parameter identification: depending upon the model chosen experimental data may only allow fitting of a limited number of free parameters in a system (we observe this later in Section 4.7). Consequently I will often use *a priori* estimates for some parameter values, usually those which can best be estimated from the literature, such that a feasible subset remains to be fit to the experimental data. Many questions studied in the field of model identification [153] are also relevant to the parameter fitting process, as in some cases there can be models of significantly differing structure that nevertheless provide approximately equivalent fits to experimental data.

2.5 Conclusion

The design, implementation, and analysis of synthetic biological systems requires a broad skillset spanning the biological sciences and its experimental procedures, control engineering, and mathematics. In this chapter I have provided a high-level review of these fields in order to provide the reader with the background knowledge necessary to understand the remainder of the thesis. Ultimately, synthetic biology faces major

challenges at virtually all steps in the design-build-test cycle: its constituent components are highly variable and context-dependent, accurate characterisation and diagnostics of circuits in the laboratory is difficult, and the feasibility of modelling their behaviour is limited due to the great degree of simplification (of very complex biological processes) required to create tractable models. The remainder of this thesis aims to contribute solutions and insights into each of these problems.

3

Design Rules for Biological Control Systems

Though the widespread need for synthetic biological control systems has been recognised [27], and many architectures have been proposed, their implementation remains challenging [45]. This stems from a lack of systematic approaches to design, as well as methods for optimal (or even feasible) implementation [40]. To this end it has been recognised that constraint sets must be developed to limit the space of viable designs [50]. In this chapter I use tools from classical control theory to derive a set of constraints (based upon limitations inherent in biological systems) to aid the design of systems that achieve adaptation or disturbance rejection. The results outlined in this chapter were published in [60] and [61].

3.1 Introduction

Past work has established some constraints that must be satisfied for systems with properties such as switch-like responses [108], disturbance rejection (homeostasis) [49], and adaptation [45]. Such approaches can be used to determine structural dependencies of variables upon one another [154] (an approach originally popularised in the determination of ecological community matrices [155, 156]), and to reveal adaptive and oscillatory behaviours [157]. State-space approaches have been used in both linear [109] and non-linear [110] systems to provide necessary conditions for adaptation. Furthermore, "cofactor conditions," which require systems to have attractive steady-state functions for which changes in the input do not affect the output, have been employed to determine design principles for construction of homeostatic systems [158]. However, even though these

constraints can substantially narrow feasible design spaces, they are often mathematically abstract, and do not address many of the physical challenges posed by biological components that make the implementation of control structures of any substantial complexity challenging [45]. Adding to this difficulty, many challenges arising from the unpredictability and context dependence of biological systems persist (see Section 2.2.1).

Working toward addressing some of these challenges in implementation, we now derive design constraints on the structure of networks that provide two common control capabilities (Fig. 3.1): adaptation and disturbance rejection (homeostasis). We provide guidelines that can be used to simplify the design and implementation of systems with these capabilities, which due to our restrictive constraint set provide a greater level of specificity than previous work. We take a state space approach to designing the linearised control system, to which we apply restrictions that aid implementation, and then derive constraints on the network structure and parameters. Some of these results follow from the Internal Model Principle [159], which in the present context dictates that should a feedback system adapt to a class of input signals, the loop must contain a subsystem able to generate signals of that class [160]. Our design constraints are narrowed to consider the biological systems in question, and due to the restrictive assumption set proposed yield only the most readily implementable designs. Once an appropriate linear system has been designed, we demonstrate methods for its embedding in a non-linear model of biological relevance. The final stage of the synthetic biology design process, the selection of particular biological parts for implementation, is left to subsequent chapters.

The results in this chapter build upon the modelling background provided in Section 2.4, and in particular makes reference to the non-linear (2.3) and linear (2.4) ODE models from Section 2.4.3. Section 3.2 formally defines this chapter's aim, and provides examples of functional forms for biological interactions that may be used to implement the systems described. Section 3.2 also outlines two general constraints that are used in the rest of the chapter to exclude systems with some difficult-to-implement properties. In Section 3.3

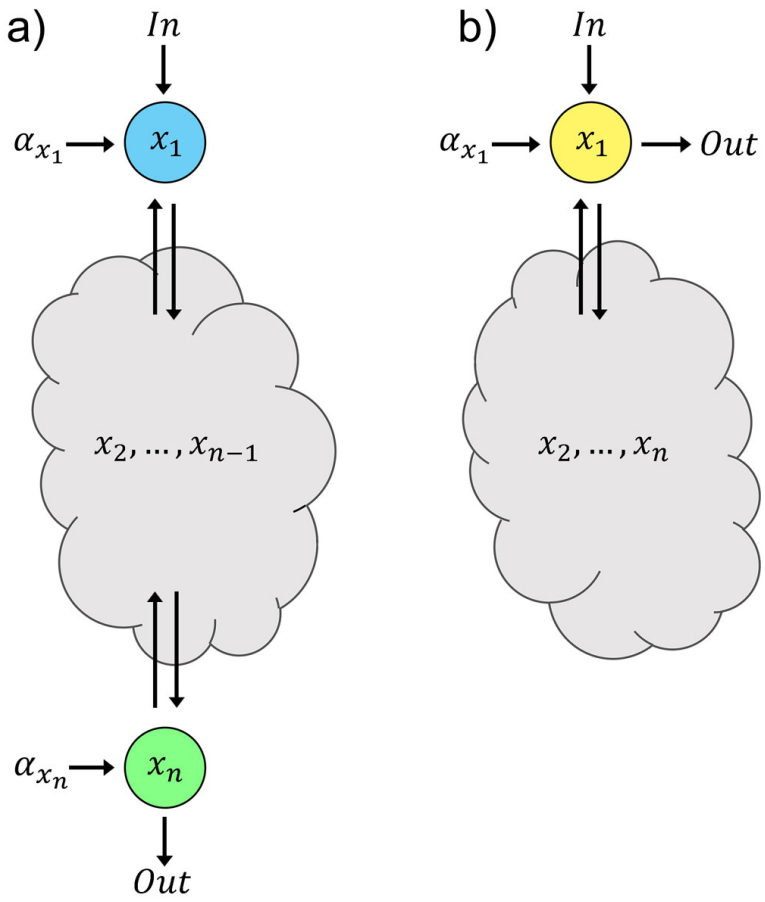


Figure 3.1: Schematic representation of systems that achieve adaptation and disturbance rejection. a) For an adaptive system the long-term behaviour of the output node (x_n, green) is independent of a time-dependent signal applied at the input node (x_1 , blue). **b)** The output variable of a disturbance-rejecting system (x_1, yellow) is in the long-term independent of a time-dependent input acting upon the same node. The α_{x_i} represent basal expression rates of the system's state variables x_i, and grey clouds represent a yet undefined network of interactions between the system's state variables.

results are derived that govern the equilibrium behaviour of linear systems, which are then used in Sections 3.4 and 3.5 to derive constraints for networks capable of adaptation and rejection of disturbances respectively. In Section 3.6 the results of Section 3.4 are employed to derive a linearised network that can adapt to step-type disturbances. This network is used to design a non-linear system with these capabilities, which is simulated using illustrative parameters. Section 3.7 discusses the broader implications of the assumptions made in our approach to design, and the chapter is concluded in Section 3.8.

3.2 Aim and Problem Outline

This chapter aims to provide constraints for the design of biological systems of minimal size that achieve *adaptation* and *disturbance rejection* (defined in Fig. 3.1) for inputs of the form $u(t) = t^{k-1}$, $t > 0$ (where $k = 1$ would indicate a step, $k = 2$ a ramp etc.). Rejection of step inputs is necessary in many biological systems for re-baselining (for example, in chemotaxis [161]), and adaptation to higher order inputs (i.e. $k \geq 2$) may be useful for distinguishing biological processes that create products at a constant rate from those that have higher order dynamics. Furthermore, disturbance rejecting systems can better function as modular components of larger synthetic circuits since they can be designed to have outputs which can (for example) maintain a constant concentration whilst being consumed by downstream processes. Designing non-linear systems of biological relevance that achieve these goals, and are also feasible for implementation, is challenging. Therefore, we approach this problem by designing a linear system, which is then used to determine the form of an appropriate non-linear system. As such our design approach considers the linearisation process outlined in Section 2.4.3 in the opposite direction: we first derive the necessary structure of $\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\big|_{\boldsymbol{x}_e,\mathbf{0}}$ from a set of constraints, and then attempt to find a non-linear biological system as in (2.3) possessing these linear dynamics about one of its equilibrium points. For a given linear system there will be many feasible forms of $\mathbf{f}(\boldsymbol{x}, \boldsymbol{u})$ depending on the physical components chosen, and so this choice will be narrowed down by selection of implementations that are most biologically tractable.

3.2.1 Potential forms for $\mathbf{f}(\boldsymbol{x}, \boldsymbol{u})$

Though there is a wide (and growing) range of synthetic parts from which biological circuits can be built (a selection is reviewed in Section 2.1.3), some commonly used elements (and their potential mathematical expressions that could appear in $\mathbf{f}(\boldsymbol{x}, \boldsymbol{u})$ from (2.3)) are outlined below. In these equations we have placed the state variables in square brackets to indicate that in this case they refer to concentrations.

- **Constant Expression**: An element of a network for which the concentration increases at a constant rate, α.

$$\frac{d[x]}{dt} \propto \alpha. \tag{3.1}$$

- **Activated / Repressed Expression**: An element for which the creation rate depends on the concentration of a secondary species $[y]$. Activation corresponds to the case in (3.2) for which large $[y]$ results in increased expression, whilst repression removes the factor of $[y]$ from the equation's numerator, meaning that $d[x]/dt$ becomes a decreasing function of $[y]$. The Hill coefficient (n) describes the cooperativity of the activation/repression process ($n = 1$ for a non-cooperative process).

$$\frac{d[x]}{dt} \propto \frac{\beta[y]^n}{k^n + [y]^n}. \tag{3.2}$$

- **Degradation / Dilution**: Reduction in an element's concentration due to its degradation over time, or dilution (due to cell growth), which is proportional (via a factor δ) to its own concentration.

$$\frac{d[x]}{dt} \propto -\delta[x]. \tag{3.3}$$

- **Michaelis-Menten Enzyme Kinetics**: Reactions in which an enzyme y is used to increase or decrease the concentration of a substrate x can be described using the Michaelis-Menten equation [144]. The rate of this process depends on the enzyme's concentration ($[y]$) and its catalytic rate (k_{cat}).

$$\frac{d[x]}{dt} \propto \frac{k_{cat}[y][x]}{K_M + [x]}. \tag{3.4}$$

- **Mass Action Kinetics**: Chemical reactions that transform one group of species into another can be described by mass action kinetics (introduced in Section 2.4.2). For example, for a reaction of the form:

$$\alpha x + \beta y \rightarrow \gamma c, \tag{3.5}$$

we have that

$$\frac{d[x]}{dt}, \frac{d[y]}{dt} \propto -k[x]^\alpha[y]^\beta. \tag{3.6}$$

Again, the above list of potential elements of $\mathbf{f}(\boldsymbol{x}, \boldsymbol{u})$ is by no means exhaustive; it aims to provide examples of simple mathematical approximations for the biological systems considered in this work.

3.2.2 System Linearisation and Assumptions for Implementation

Recalling (2.5), we can express the linearisation (2.4) as a general linear system of the form

$$\begin{aligned} \dot{\boldsymbol{x}} &= \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{u}, \\ \boldsymbol{y} &= \mathbf{C}\boldsymbol{x}, \end{aligned} \tag{3.7}$$

where $\boldsymbol{x}(t) \in \mathbb{R}^n$ represents the system's state variables and $\dot{\boldsymbol{x}}(t)$ their time derivatives (as in Section 2.4.3). $\boldsymbol{u}(t) \in \mathbb{R}^m$ is the system's input and $\mathbf{y}(t) \in \mathbb{R}^p$ the system output. For SISO (single input single output) systems as considered in Sections 3.3, 3.4 and 3.5 of this chapter, $m = p = 1$. $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are matrices of appropriate dimensions that define the system, with structure that will be constrained in terms of the sign or nullity of their elements. For practical implementation the elements of these matrices will then take values in $\mathbb{R}$ defined by the system's biological parameters. We focus on SISO systems due to their ease of analysis in the present context, though approaches to expanding this framework to incorporate systems in which inputs affect multiple state variables are discussed in Section 3.7.

In the remainder of this chapter sub-matrix notation will be as follows:

- $\mathbf{A}_{i,j}$ refers to the element in the ith row and jth column of $\mathbf{A}$
- $\mathbf{A}_{i:j,k:l}$ refers to a sub-matrix of $\mathbf{A}$ of size $j - i + 1$ by $l - k + 1$, which includes elements from the ith to jth row, and kth to lth column of $\mathbf{A}$.

To provide biologically tractable designs, we make a range of assumptions that limit the properties of designed linear systems. First, we require that $\mathbf{A}$ matrices are Hurwitz (stable in the control theory sense). In order to create systems for which the *structure*, rather than particular parameter values, guarantees performance goals are achieved, we make the assumption of *element independence*, defined in Assumption 3.2.2. Due to

the leaky nature of biological systems, particularly at the promoter level [162], we will constrain the constant term in the differential equation describing each species' time evolution to be non-negative. This will be referred to as a *basal* level of production (Constraint 3.2.3). Furthermore, we wish to avoid systems in which individual components impact other state variables both positively and negatively, and thus require that the partial derivatives with respect to $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ of the differential equation governing the dynamics of x_i have the same sign for all $\boldsymbol{x}$ (Constraint 3.2.4).

Definition 3.2.1. We define (and denote as $\mathbb{S}^{\mathbf{A}}$) a *System of Constraints* on a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, to be an $n \times n$ array of functions of elements of $\mathbf{A}$.

For example, we might have that $\mathbb{S}^{\mathbf{A}}_{i,j} = \mathbf{A}_{a,b} + \mathbf{A}_{c,d}\mathbf{A}_{e,f}$ if under a given constraint it was required that the element $\mathbf{A}_{i,j}$ be equal to the combination of other elements, $\mathbf{A}_{a,b} + \mathbf{A}_{c,d}\mathbf{A}_{e,f}$.

Assumption 3.2.2. *Element Independence*
We say that a matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ has the property of *element independence* if, for a given system of constraints $\mathbb{S}^{\mathbf{A}}$, the following holds:

$$\frac{\partial \mathbb{S}^{\mathbf{A}}_{i,j}}{\partial \mathbf{A}_{l,m}} = 0, \tag{3.8}$$

for all $\mathbf{A}_{l,m} \in \mathbf{A}$ subject to $l \neq i$ or $m \neq j$.

Element independence can be thought of as a criterion that eliminates constraint sets which require elements of the system matrix $\mathbf{A}$ to be equal to the sum, product, or any arbitrary non-trivial function, of other elements of $\mathbf{A}$. This criterion must be understood within the context of constraints applied to a matrix's elements, which impose functional relationships between different elements (i.e. the $\mathbb{S}^{\mathbf{A}}_{i,j}$) for families of systems that achieve given design goals. We desire element independence in biological contexts in order to aid the realisation of designed systems, since tuning parameter relationships for different

biological parts of a system is challenging in many cases. Similar requirements to element independence have previously been investigated [163], and are discussed further in Section 3.7.

3.2.3 Other General Constraints for Implementation

In addition to the mathematical constraints derived above, there are a number of desirable properties that our designed system should possess in order to ease its implementation. These will not be used to prove the results in Sections 3.3, 3.4 and 3.5 of this work, but are applied during the design of a specific biological implementation in Section 3.6.

Previously we assumed that in a biological system each state variable could have a non-negative basal expression level. If a given state variable x_i is to reach an equilibrium, there must be a balance between positive and negative rate contributions to its non-linear dynamics. Since we assume each state variable will have a non-negative basal expression rate α_{x_i}, there must be at least one negative element in each row of $\mathbf{A}$. This requirement yields the following constraint:

Constraint 3.2.3. $\mathbf{A}_{i,j} < 0$ for at least one $j = 1, \ldots, n$, for all $i = 1, \ldots, n$

To aid implementation we would like to avoid situations in which one element of a network has influences of differing signs on other elements. Either it should have a positive influence (e.g. a transcriptional activator), or a negative influence (e.g. a transcriptional repressor), but not both. Thus, the off diagonal terms in each column of $\mathbf{A}$ should all be strictly non-negative, or non-positive. This requirement yields the following constraint:

Constraint 3.2.4. Either $\mathbf{A}_{i,j} \leq 0 \; \forall \; i = 1, \ldots, j-1, j+1, \ldots, n, \; \forall \; j = 1, \ldots, n$ or $\mathbf{A}_{i,j} \geq 0 \; \forall \; i = 1, \ldots, j-1, j+1, \ldots, n, \; \forall \; j = 1, \ldots, n$

3.3 General Linearised Network Structure

Proposition 3.3.1. Consider a SISO system in the form of (3.7) with n state variables, and in which $\mathbf{A}$ is Hurwitz. The following statements are equivalent:

1. The system rejects inputs of the form $u(t) = t^{k-1}$ with $k \in \mathbb{N}^+$, i.e. $\lim_{t\to\infty} y(t)$ is constant.

2. $$\lim_{s\to 0} \det \begin{bmatrix} s\mathbf{I} - \mathbf{A} & -\mathbf{B} \\ \mathbf{C}/s^{k-1} & 0 \end{bmatrix} = 0, \tag{3.9}$$

   where s is the frequency domain (Laplace) parameter, $\mathbf{A} \in \mathbb{R}^{n\times n}$, $\mathbf{B} \in \mathbb{R}^{n\times 1}$, and $\mathbf{C} \in \mathbb{R}^{1\times n}$, and $\mathbf{I}$ is the $n \times n$ identity matrix.

Proof Under the stability assumptions of the proposition, from the final value theorem [141] we have that

$$y(\infty) = \lim_{s\to 0} sY(s). \tag{3.10}$$

For this system to be unperturbed (in the $t \to \infty$ limit) by the input/disturbance $u(t)$ we thus require $\lim_{t\to\infty} y(t) = 0$. For our system we can express the transfer function from input to output in terms of its Rosenbrock system matrix as [164, 165]:

$$G(s) = \frac{1}{\det(s\mathbf{I} - \mathbf{A})} \det \begin{bmatrix} s\mathbf{I} - \mathbf{A} & -\mathbf{B} \\ \mathbf{C} & 0 \end{bmatrix}, \tag{3.11}$$

which, taking $Y(s) = G(s)U(s)$ and a general input of the form $u(t) = t^{k-1}$, equivalent to $U(s) = (k-1)!/s^k$ in the Laplace domain, gives

$$\lim_{s\to 0} \frac{(k-1)!}{\det(s\mathbf{I} - \mathbf{A}) \times s^{k-1}} \det \begin{bmatrix} s\mathbf{I} - \mathbf{A} & \mathbf{-B} \\ \mathbf{C} & 0 \end{bmatrix} = 0. \tag{3.12}$$

Here $\det \begin{bmatrix} s\mathbf{I} - \mathbf{A} & -\mathbf{B} \\ \mathbf{C} & 0 \end{bmatrix}$ will yield a polynomial in s of maximal order $n-1$, whilst the lowest order s term in $\det(s\mathbf{I} - \mathbf{A}) \times s^{k-1}$ will be of order $k-1$, and equal to $\det(-\mathbf{A}) \times s^{k-1}$. By placing $k-1$ zeros at the origin this term causes cancelling of inputs of the form the system aims to reject, as required by the internal model principle. Thus, since we require this limit to approach zero, as $s \to 0$ the denominator's value

will be dominated by its lowest order term, and we can replace the condition in (3.12) with one which is equivalent in the $s \to 0$ limit

$$0 = \lim_{s\to 0} \frac{(k-1)!}{\det(-\mathbf{A}) \times s^{k-1}} \det \begin{bmatrix} s\mathbf{I} - \mathbf{A} & -\mathbf{B} \\ \mathbf{C} & 0 \end{bmatrix}. \tag{3.13}$$

Condensing the $1/s^{k-1}$ pre-factor into the bottom row of the determinant and eliminating constant terms gives the desired result. □

3.4 Network Constraints for Adaptation

For a system to achieve adaptation (Fig. 3.1a), the long-term behaviour of one of its state variables (the output) must be independent of a time-dependent input applied to a different state variable (the input). We will arbitrarily define the input state variable as the first element in the system, and the output state variable as the last, giving

$$\mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 0 & \cdots & 0 & 1 \end{bmatrix}. \tag{3.14}$$

Recall that the *relative degree* (D_g) of a system is defined as the number of times the system's output (y) must be differentiated before the input (u) appears explicitly in its expression. Equivalently, it is equal to the difference in the order of the polynomials (in s) found in the numerator and denominator of a system's state space transfer function ($G(s)$). Described in a network context, D_g is the shortest path (in the graph theory sense) between the system's input/output state variables. For a system with relative degree D_g this path will thus include D_g state variables, two of these being the input and output (1st and nth state variables respectively). If we arbitrarily label the state variables corresponding to the rows of $\mathbf{A}$ such that the first $1, \ldots, D_g - 1$ state variables sequentially form the shortest path to the output (nth) state variable, then requiring connectivity between the input and output elements of a network for a given $D_g = 1, \ldots, n$ is equivalent to the following

$$\begin{aligned} \mathbf{A}_{n,D_g-1} \prod_{j=2}^{D_g-1} \mathbf{A}_{j,j-1} &\neq 0, \\ \mathbf{A}_{n,1:D_g-2} &= 0. \end{aligned} \tag{3.15}$$

For example, if $D_g = 2$, the input state variable connects directly to the output ($\mathbf{A}_{n,1} \neq 0$), or if $D_g = 3$ the input state variable connects to the second state variable ($\mathbf{A}_{2,1} \neq 0$) which connects to the output ($\mathbf{A}_{n,2} \neq 0$).

Proposition 3.4.1. Consider a linear system (3.7) that satisfies the assumptions of Proposition 3.3.1, with $\mathbf{B}, \mathbf{C}$ as in (3.14). Suppose that for this system element independence (Assumption 3.2.2) of $\mathbf{A}$ holds, and that there exists a solution to (3.15) for some D_g, so that the input and output of the network are connected. If such a system has the minimal number of state variables (n), and is able to reject an input of the form $u(t) = t^{k-1}$, then it has the following properties:

1. $n = k + 2$
2. $\mathbf{A}_{i,i} = 0$, and the product $\mathbf{A}_{i,1}\mathbf{A}_{n,i} = 0$, for all $i = 2, \ldots, n-1$
3. $\mathbf{A}_{i,n} \neq 0$ for at least one $i = 2, \ldots, n$, and $\mathbf{A}_{1,j} \neq 0$ for at least one $j = 1, \ldots, n-1$.
4. $D_g = 2$, and thus $\mathbf{A}_{n,1} \neq 0$.

Proof From (3.9) this system must satisfy

$$\lim_{s\to 0} \det \begin{bmatrix} & & 1 \\ s\mathbf{I} - \mathbf{A}_{1:n,1:n} & & 0 \\ & & \vdots \\ \cdots & 0 \quad \frac{1}{s^{k-1}} & 0 \end{bmatrix} = 0. \tag{3.16}$$

Using Laplace's formula, this determinant can be expanded along the bottom row and up the right column to give

$$\lim_{s\to 0} \frac{1}{s^{k-1}} \times \det(\mathbf{A}_r) = 0, \tag{3.17}$$

where

$$\mathbf{A}_r = \begin{bmatrix} \mathbf{A}_{2:n-1,1} & \mathbf{A}_{2:n-1,2:n-1} - s\mathbf{I} \\ \mathbf{A}_{n,1} & \mathbf{A}_{n,2:n-1} \end{bmatrix}. \tag{3.18}$$

Let $\tilde{\lambda}_i(s)$, $i = 1, \ldots, n-1$ denote the eigenvalues of the matrix $\mathbf{A}_r$, then for (3.17) to be satisfied we require

$$\lim_{s\to 0} \frac{1}{s^{k-1}} \prod_{i=1}^{n-1} \tilde{\lambda}_i(s) = 0. \tag{3.19}$$

We will now examine two cases, the first where one of the $\tilde{\lambda}_i(s) = 0$ when s is arbitrary, and the second where this does not hold, thus requiring that a term of order s^k can be factorised out of the product of eigenvalues in (3.19) (so that (3.17) is satisfied). We will rule out the first case due to its inability to satisfy (3.15) for any value of D_g, and thus conclude the second case must be true.

Case 1: If one of the $\tilde{\lambda}_i(s) = 0$ when s is arbitrary then $\mathbf{A}_{2:n,1:n-1}$ is degenerate. Expanding $\det(\mathbf{A}_r)$ along the bottom row and separating out the first term we have:

$$\begin{aligned} &\mathbf{A}_{n,1} \times \det(\mathbf{A}_{2:n-1,2:n-1} - s\mathbf{I}) + \\ &\sum_{j=2}^{n-1} (-1)^{j+1} \mathbf{A}_{n,j} \mathbf{M}_{n,j} = 0, \end{aligned} \tag{3.20}$$

where $\mathbf{M}_{i,j}$ is the standard matrix minor equal to the determinant of the $(n-2) \times (n-2)$ matrix that results from removing the ith row and jth column from $\mathbf{A}_r$. The first term in (3.20) is the only one proportional to s^{n-2}, and therefore for this expression to be equal to zero for arbitrary s we require $\mathbf{A}_{n,1} = 0$, ruling out the $D_g = 2$ case of (3.15).

Expanding $\det(\mathbf{A}_r)$ down the left column, the first term is given by

$$\mathbf{A}_{2,1} \times \det \begin{bmatrix} \mathbf{A}_{3:n-1,2} & \mathbf{A}_{3:n-1,3:n-1} - s\mathbf{I} \\ \mathbf{A}_{n,2} & \mathbf{A}_{n,3:n-1} \end{bmatrix}. \tag{3.21}$$

Due to the assumption of element independence for $\mathbf{A}$, (3.21) must equal zero for all s. As the $D_g = 2$ case of (3.15) has already been excluded, $\mathbf{A}_{2,1} \neq 0$, and so expanding this determinant along the bottom row requires $\mathbf{A}_{n,2} = 0$, it being the coefficient of the only s^{n-3} term. Therefore, the $D_g = 3$ case of (3.15) is likewise excluded. Continuing to expand (3.21) in a similar manner (for example, in the next stage expanding the determinant in (3.21) about $\mathbf{A}_{3,2}$) rules out all possible values of D_g, thereby proving Case 1 to be impossible.

Case 2: With Case 1 excluded, it is necessary that a factor of s^k is shared between the $n-1$ eigenvalues in (3.19) for the limit to hold. This implies the matrix $\mathbf{A}_{2:n,1:n-1}$ is one-fold degenerate (that is, it contains one linearly dependent row/column), which

follows from the fact that $\mathbf{A}$ is not degenerate. This also requires that at least one of the values $\mathbf{A}_{2,n}, \ldots, \mathbf{A}_{n,n} \neq 0$, and one of $\mathbf{A}_{1,1}, \ldots, \mathbf{A}_{1,n-1} \neq 0$, to avoid row- and column-wise degeneracy of $\mathbf{A}$ respectively, which gives property 3 of Proposition 3.4.1.

In order for a factor of s^k to appear in the eigenvalues $\tilde{\lambda}_i$ it is necessary that there are at least k terms including s in $\mathbf{A}_r$. Since $\mathbf{A}_r$ contains $n-2$ terms linear in s we can constrain the minimal size of the complete network (n) as $n = k+2$, which gives property 1 of Proposition 3.4.1. If we express the determinant of $\mathbf{A}_r$ as a polynomial in s:

$$\beta_{n-2}s^{n-2} + \beta_{n-3}s^{n-3} + \ldots + \beta_1 s^1 + \beta_0. \tag{3.22}$$

Then $\beta_0, \ldots, \beta_{k-1} = 0$ in order for an s^k term to be factored out. By breaking down this determinant to calculate coefficients of the highest order s terms we find $\beta_{n-2} = \mathbf{A}_{n,1}$ and

$$\beta_{n-3} = \sum_{i=2}^{n-1} \mathbf{A}_{i,1}\mathbf{A}_{n,i} + \mathbf{A}_{n,1} \sum_{i=2}^{n-1} \mathbf{A}_{i,i.} \tag{3.23}$$

Since $n = k+2$ the only non-zero term in (3.22) is the one with coefficient $\beta_k = \mathbf{A}_{n,1}$, and thus the network's input must be connected to its output (i.e. $D_g = 2$, yielding property 4 of Proposition 3.4.1). The requirement that $\beta_{k-1} = 0$ in (3.23) dictates that $\mathbf{A}_{i,i} = 0$ and $\mathbf{A}_{i,1}\mathbf{A}_{n,i} = 0$ for all $i = 2, \ldots, n-1$, which gives property 2 of Proposition 3.4.1. □

From a biological standpoint, the first constraint in property 2 of Proposition 3.4.1 corresponds to a requirement that each non-input/output species has a constant rate of degradation/dilution (i.e. zeroth-order degradation), independent of its concentration. This requirement, which has previously been established for perfectly adapting systems [166], can be realised (for example) by saturation of proteolysis - forcing the protease responsible for degradation of a given species to work at saturation [167]. The second constraint in property 2 of Proposition 3.4.1 means that a species i (except for the input species) can be directly influenced by the input node, or it directly influences the output node, but both conditions are never simultaneously true. This is equivalent to requiring that there are no paths from input to output that go through precisely one other node. For the minimal relative degree ($D_g = 2$), the minimum size of a network that can reject

inputs of order $k-1$ can be equivalently calculated by recalling that rejecting disturbances of the form t^{k-1} requires the system's transfer function to have k zeros at the origin, and therefore by the definition of relative degree the system must be of minimal order $k+2$.

Although the minimally sized network for adaptation to a given input has $D_g = 2$ via Proposition 3.4.1, in some situations it may be advantageous to select a larger network in order to relax some constraints on the constituent component dynamics. With each increase in D_g we must include an additional term in (3.22), and thus the minimal size of an adaptive network for a given D_g is $n = D_g + k$.

3.5 Network Constraints for Disturbance Rejection

To formulate a network that can reject a disturbance in one of its state variables (Fig. 3.1b) we will consider a system in which the input and output are (arbitrarily) applied to the first state variable, giving

$$\mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}. \tag{3.24}$$

In a biological sense, such a network would be able to stabilise the concentration of an output species (whose equilibrium concentration may be a function of inputs elsewhere in the network) as it is being consumed by secondary processes. This system is therefore able to function as a *load-driver* [114]. Such systems have been implemented in a synthetic biological context in recent years, and have been shown to provide much-needed modularity for the interfacing of biological constructs [46].

Proposition 3.5.1. Consider a linear system (3.7) that satisfies the assumptions of Proposition 3.3.1, with $\mathbf{B}, \mathbf{C}$ as in (3.24). Suppose that for this system element independence (Assumption 3.2.2) of $\mathbf{A}$ holds. If such a system has the minimal number of state variables (n), and is able to reject an input of the form $u(t) = t^{k-1}$, then it has the following properties:

1. $n = k + 1$
2. $\mathbf{A}_{1,1} < 0$, $\mathbf{A}_{i,i} = 0$ for all $i = 2, \ldots, n$, and $\mathbf{A}_{j,i}\mathbf{A}_{i,j} = 0$ for all combinations of $i, j = 2, \ldots, n$.
3. $\mathbf{A}_{i,1} \neq 0$ and $\mathbf{A}_{1,i} \neq 0$ for at least one $i = 2, \ldots, n$.

Proof Analogously to the process following from (3.16), it is necessary that

$$\lim_{s \to 0} \frac{1}{s^{k-1}} \times \det(\mathbf{A}_t) = 0, \tag{3.25}$$

where

$$\mathbf{A}_t = \left[s\mathbf{I} - \mathbf{A}_{2:n,2:n} \right]. \tag{3.26}$$

If we express the determinant of $\mathbf{A}_t$ as a polynomial in s, we get

$$s^{n-1} + \beta_{n-2}s^{n-2} + \ldots + \beta_1 s + \beta_0 = 0. \tag{3.27}$$

To satisfy (3.25) an s^k term must be factored out of (3.27), which sets a minimal network size of $n = k + 1$ (property 1 of Proposition 3.5.1). Setting $n = k + 1$ in (3.27) we have

$$s^k + \beta_{k-1}s^{k-1} + \ldots + \beta_1 s + \beta_0 = 0, \tag{3.28}$$

which requires $\beta_{k-1}, \ldots, \beta_0 = 0$ in order to leave s^k as the lowest order term. Since $\beta_0 = \det(-\mathbf{A}_{2:n,2:n})$ we have that $\mathbf{A}_{2:n,2:n}$ is degenerate. However, since $\mathbf{A}$ is invertible, $\mathbf{A}_{2:n,2:n}$ has at most one degenerate row/column, and so at least one of $\mathbf{A}_{2,1}, \ldots, \mathbf{A}_{n,1} \neq 0$ and one of $\mathbf{A}_{1,2}, \ldots, \mathbf{A}_{1,n} \neq 0$ to avoid row- and column-wise degeneracy of $\mathbf{A}$ respectively (property 3 of Proposition 3.5.1). In (3.28) the term β_{k-1} is given by:

$$\beta_{k-1} = (-1)^{k-1} \sum_{i=2}^{k+1} \mathbf{A}_{i,i}, \tag{3.29}$$

which if $\beta_{k-1} = 0$, given the assumption of element independence for $\mathbf{A}$, gives the second part of property 2 in Proposition 3.5.1. Biologically, this constraint requires that all state variables bar the input/output must have zeroth-order degradation dynamics [166]. Since all eigenvalues of $\mathbf{A}$ have negative real part, $\text{Trace}(\mathbf{A}) < 0$, and thus $\mathbf{A}_{1,1} < 0$, giving the first part of property 2 in Proposition 3.5.1. Since $\mathbf{A}_{i,i} = 0$ for $i = 2, \ldots, n$, we have

$$\beta_{k-2} = (-1)^{k} \sum_{j=2}^{k+1} \sum_{i=j+1}^{k+1} \mathbf{A}_{j,i}\mathbf{A}_{i,j}, \tag{3.30}$$

which if $\beta_{k-2} = 0$, given again the assumption of element independence for $\mathbf{A}$, constrains every pair of non-output state variables to be joined only by a single link. This gives the final part of property 2 in Proposition 3.5.1. □

Considering a non-minimal network with $n = k + 2$, we still require s^k to be the lowest order term in $\det(\mathbf{A}_t)$, which is now given by

$$s^{k+1} + \beta_k s^k + \ldots + \beta_1 s + \beta_0 = 0. \tag{3.31}$$

The term β_k has the same form as in (3.29), but with the sum index going to $k + 2$. This sum can now be non-zero, and thus diagonal elements may be non-zero. The next coefficient in (3.31) is

$$\beta_{k-1} = (-1)^k \left(\sum_{j=2}^{k+2} \sum_{i=j+1}^{k+2} \mathbf{A}_{j,i}\mathbf{A}_{i,j} + \sum_{j=2}^{k+2} \sum_{i=2, i\neq j}^{k+2} \mathbf{A}_{i,i}\mathbf{A}_{j,j} \right), \tag{3.32}$$

which if $n = k + 2$, must equal zero. The right hand sum in (3.32) indicates that all pairwise products of diagonal terms are zero. Since the non-zero β_k allowed these to be individually non-zero, precisely one of the diagonal terms can be non-zero. The left hand sum in (3.32) constrains every pair of non-output nodes to be joined only by a single link, as was the case before.

3.6 Application To Biological System Design

3.6.1 Implementation of parameter constraints

We now apply the derived constraints to design a biological system capable of adapting to a step ($k = 1$) disturbance. For a minimal realisation of this system, following Proposition 3.4.1 we require a network of size $n = 3$, which in linearised form will be expressed as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{1,1} & \mathbf{A}_{1,2} & \mathbf{A}_{1,3} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \mathbf{A}_{2,3} \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} \end{bmatrix}. \tag{3.33}$$

From Proposition 3.4.1 we have $\mathbf{A}_{3,1} \neq 0$ and $\mathbf{A}_{2,2} = 0$. Since all eigenvalues of $\mathbf{A}$ have negative real part we know $\mathbf{A}_{1,1} + \mathbf{A}_{3,3} < 0$, and so we will set both $\mathbf{A}_{1,1}, \mathbf{A}_{3,3} < 0$ to avoid constraining the relative magnitudes of these parameters (if one were positive it would need to be lesser in magnitude than the other). In doing this we have also satisfied property 3 of Proposition 3.4.1. From property 2 of Proposition 3.4.1 we have

$\mathbf{A}_{2,1}\mathbf{A}_{3,2} = 0$, and will set $\mathbf{A}_{3,2} = 0$ as this choice allows us to set $\mathbf{A}_{2,1} \neq 0$ so that there is a non-zero element in the matrix's second row. Expanding the determinant of $\mathbf{A}$ down its middle column now gives $-\mathbf{A}_{1,2}(-|\mathbf{A}_{3,3}|\mathbf{A}_{2,1} - \mathbf{A}_{2,3}\mathbf{A}_{3,1})$, which must be less than zero for $\mathbf{A}$ to be stable (since this determinant is equal to the product of three negative eigenvalues), and so we set $\mathbf{A}_{2,3} = 0$ and $\mathbf{A}_{1,2}\mathbf{A}_{2,1} < 0$. At this point we can set $\mathbf{A}_{1,3} = \mathbf{A}_{2,3} = 0$ to simplify the system, since these elements are not required to satisfy other constraints. By Constraint 3.2.3 we then have that $\mathbf{A}_{2,1} < 0$, and therefore by Constraint 3.2.4 $\mathbf{A}_{3,1} < 0$. Thus we are left with

$$\mathbf{A} = \begin{bmatrix} -|\mathbf{A}_{1,1}| & |\mathbf{A}_{1,2}| & 0 \\ -|\mathbf{A}_{2,1}| & 0 & 0 \\ -|\mathbf{A}_{3,1}| & 0 & -|\mathbf{A}_{3,3}| \end{bmatrix}, \tag{3.34}$$

which forms part of a single-input multiple-output linear system as in (3.7) given by

$$\begin{aligned} \dot{\tilde{\boldsymbol{x}}} &= \mathbf{A}\tilde{\boldsymbol{x}} + \begin{bmatrix} u_{x_1}(t) \\ 0 \\ 0 \end{bmatrix}, \\ \boldsymbol{y} &= \tilde{\boldsymbol{x}}, \end{aligned} \tag{3.35}$$

where as before $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{x}_e$, and $u_{x_1}(t)$ is the time-dependent disturbance to species x_1 to which we wish x_3 to adapt. We now seek a non-linear system which can be approximated near to an equilibrium by the constrained linear system in (3.35).

3.6.2 Non-linear system selection and correspondence with linear system

To design a non-linear system (of the form of (2.3)) that can be approximated by (3.35) we begin by selecting biological mechanisms to account for each of the elements of (3.34). There are many ways in which this can be done depending on the biological components selected; we outline one such implementation.

The elements $-|\mathbf{A}_{1,1}|$ and $-|\mathbf{A}_{3,3}|$ can be implemented by attaching fast degradation tags to x_1 and x_3, or by introducing a protease external to this system which acts in saturation upon x_2, giving x_2 zeroth-order degradation [166]. The element $|\mathbf{A}_{1,2}|$ can be provided by having species x_2 non-cooperatively activate the expression of species x_1 in the regime in which the concentration of x_2 is small compared to the activator binding equilibrium constant. Elements $-|\mathbf{A}_{2,1}|$ and $-|\mathbf{A}_{3,1}|$ can be implemented by choosing

species x_1 to be a protease enzyme operating in the saturation regime (where enzyme concentration is small compared to substrate concentration).

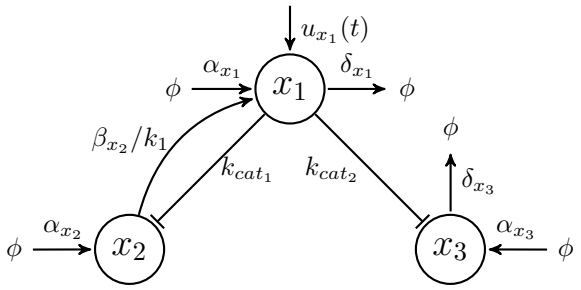


Figure 3.2: Network diagram for linear system in (3.35) with parameter values from the non-linear approximation in (3.38). Pointed (blunt) arrows indicate a positive (negative) affect of one species' concentration on that of another.

Using the standard Hill equation (for non-cooperative activation) and Michaels-Menten equation (for enzyme activity) [168] the system's non-linear dynamics can be expressed as

$$\mathbf{f}(\boldsymbol{x},\boldsymbol{u}) = \begin{bmatrix} \alpha_{x_1} - \delta_{x_1}[x_1] + \frac{\beta_{x_2}[x_2]/k_1}{1+[x_2]/k_1} + u_{x_1}(t) \\ \alpha_{x_2} - \frac{k_{cat_1}[x_1][x_2]}{K_{M_1}+[x_2]} \\ \alpha_{x_3} - \delta_{x_3}[x_3] - \frac{k_{cat_2}[x_1][x_3]}{K_{M_2}+[x_3]} \end{bmatrix}, \tag{3.36}$$

where we have placed the state variables in square brackets to indicate that they represent concentrations. The α_{x_i} represent basal expression rates of each species, β_{x_i} represent the maximal rate increase due to activation, and δ_{x_i} represent the first-order degradation rates. The K_M are Michaelis constants, and the k_{cat} the corresponding maximal enzyme catalytic rates. The structure of this system's Jacobian is given by

$$\left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_e,\mathbf{0}} = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ * & 0 & * \end{bmatrix}. \tag{3.37}$$

To set $\mathbf{A}_{2,2} = 0$ we need the activator x_2 to operate in the regime where its concentration is small compared to its binding equilibrium constant (so that $k_1 \gg [x_2]$), making the magnitude of this partial derivative small. Furthermore, forcing the protease enzyme x_1 to operate in its saturation regime gives $K_{M_1} \ll [x_2]$ and $K_{M_2} \ll [x_3]$. With these constraints enforced the Jacobian can be approximated by

$$\left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_e,\mathbf{0}} \approx \begin{bmatrix} -\delta_{x_1} & \beta_{x_2}/k_1 & 0 \\ -k_{cat_1} & 0 & 0 \\ -k_{cat_2} & 0 & -\delta_{x_3} \end{bmatrix}. \tag{3.38}$$

We are now able to use (2.4) to construct a system in the form of (3.35) by setting

$$\begin{aligned} \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{u}}\right|_{\boldsymbol{x}_e,\mathbf{0}} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_e,\mathbf{0}} = \mathbf{A}, \\ -\left.\frac{\partial \mathbf{f}}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_e,\mathbf{0}} \boldsymbol{x}_e &= \begin{bmatrix} \alpha_{x_1} \\ \alpha_{x_2} \\ \alpha_{x_2} \end{bmatrix}. \end{aligned} \tag{3.39}$$

3.6.3 Parameter selection and implementation

With the equivalences in (3.39) set, the non-linear system described by (3.36) is approximated by the linear system as in Fig. 3.2. To model this system, values for the parameters and ratios in (3.38) (which define the elements of the linear system (3.34)) are selected in order to ease implementation, subject to two constraints. First, that the eigenvalues of this matrix all lie in the open left half plane as per Proposition 3.4.1. Second, that there exists an equilibrium solution to $\mathbf{f}(\boldsymbol{x}_e, \mathbf{0}) = 0$ with $\mathbf{f}$ as in (3.36), with all elements of $\boldsymbol{x}_e$ and each α_{x_i} term strictly positive.

Based on the equilibrium concentrations $[\boldsymbol{x}_e]$ the activation binding constant is selected to satisfy $k_1 \gg [x_{e_2}]$, and the Michaelis-Menten constants to satisfy $K_{M_1} \ll [x_{e_2}]$ and $K_{M_2} \ll [x_{e_3}]$. Values for β_{x_2} and β_{x_3} are then fully defined by the ratios in (3.38).

The above process was followed to yield an illustrative set of satisfactory parameters, summarised in Table 3.1.

Table 3.1: Sample model parameters (unitless)

Parameter	Value	Parameter	Value
$[x_{e_1}]$	2.0	k_{cat_1}	1.5
$[x_{e_2}]$	2.5	k_{cat_2}	1.0
$[x_{e_3}]$	1.5	k_1	80
α_{x_1}	1.5	K_{M_1}	0.01
α_{x_2}	3.0	K_{M_2}	0.01
α_{x_3}	5.0	δ_{x_1}	2.0
β_{x_2}	80	δ_{x_3}	2.0

Using these values we simulated the linear (3.35) and non-linear (3.36) systems using MATLAB, for which results are presented in Fig. 3.3. With all state variable

concentrations $[\boldsymbol{x}]$ initially zero, designed equilibrium values (Table 3.1) are quickly reached. The time-dependent input $u_{x_1}(t)$ applies a negative unit step at $t = 10$. Though $[x_3]$ departs from equilibrium at this discontinuity in $u_{x_1}(t)$, it quickly returns to the desired value $[x_{e_3}]$, achieving the system's design purpose. Agreement between the linear and non-linear systems is good, particularly in terms of the controlled variable x_3, which was found to be robust to variation of model parameters, though these variations can shift the system's equilibrium position. Substantial (e.g. order of magnitude) changes in k_1, K_{M_1}, K_{M_2} can reduce the applicability of assumptions made when linearising our non-linear implementation. This results in departure between the linear and non-linear models, particularly in x_2, with x_1 and the controlled variable x_3 affected to a lesser extent. However, since the assumptions made in the derivation of (3.38) still loosely hold, the structure of the system is maintained and so adaptation by x_3 is still achieved.

Whilst the example illustrated in this section demonstrates the application of constraints to design a minimal system for step-disturbance rejection, the result is not unique. Rather than specifying particular parameter values, the constraints outlined in this chapter define the structure of $\mathbf{A}$ in terms of the sign (or existence) of its individual elements. When these constraints (and the assumptions that underlie them, such as the stability of $\mathbf{A}$) are satisfied, our results then guarantee that (due to its structural properties) the linear system is able to reject a given class of inputs. After all structural constraints are satisfied, unconstrained elements of $\mathbf{A}$ can then be set based on any other design goals; in our case, we set any uncertain elements to zero in order to minimise the number of required interactions within the system (thereby simplifying implementation). For networks that are designed to reject higher order disturbances ($k > 2$) the structure imposed by our design constraints will be increasingly sparse, leaving many more elements undetermined (though generally setting these to zero to reduce the complexity of the designed system will be desirable).

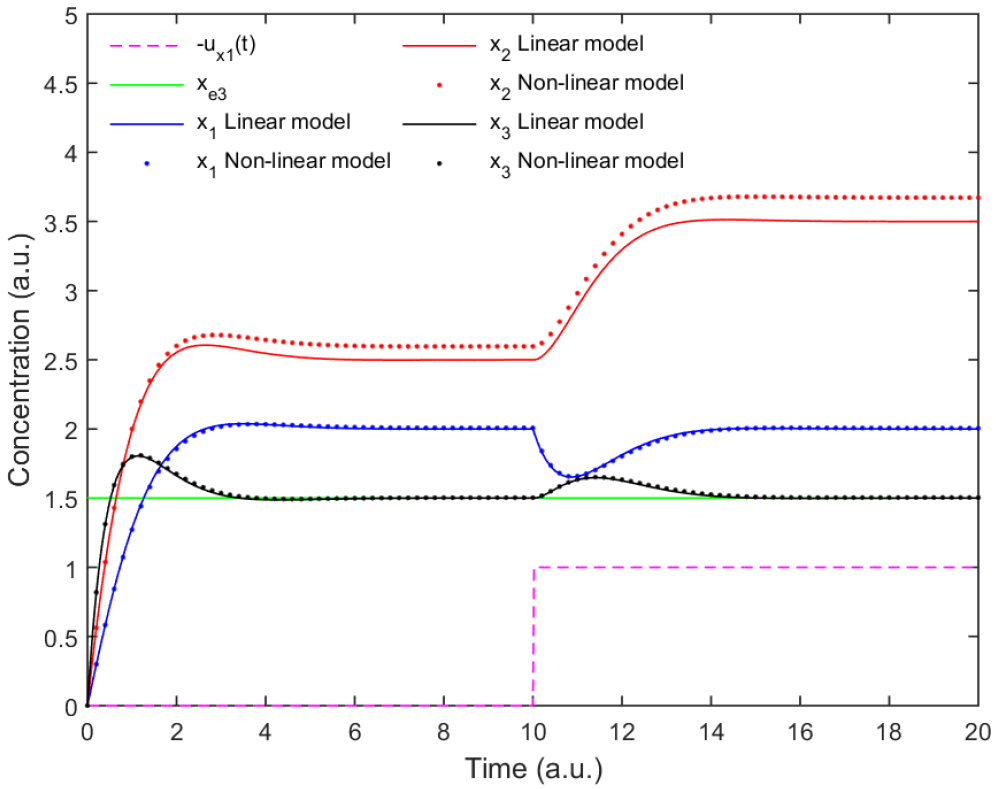


Figure 3.3: Comparison of linear (3.35) and non-linear (3.36) realisations of a system designed such that x_3 is able to adapt to step disturbances of x_1. Close agreement is found between the value of x_3 in both models and its desired equilibrium value x_{e3}. Model parameters used are as in Table 3.1. The negative of the time-dependent input $u_{x_1}(t)$ is plot as a rate, and state variables (x) are plot as concentrations. All values in arbitrary units (a.u.).

3.7 Discussion

The constraints derived in this chapter provide a general guideline for designing systems capable of adaptation and disturbance rejection, but do so under a limited set of assumptions which may, in practice, be violated. For example, the element independence requirement (that elements of **A** be independent of one-another) was investigated by Drengstig *et al.* [163], who classified adaptive networks as *robust* (those in which perfect adaptation is a property of the network's topology) or *non-robust* (those in which fine-tuning of rate constants is necessary to achieve adaptation). Our assumptions ruled out all such non-robust systems, however, their coarseness also eliminated a number of robust strategies. For example, the element independence assumption for **A** could be

violated in practice by reversible reactions such as

$$\mathrm{x}_1 \underset{\mathrm{k}_1}{\overset{\mathrm{k}_2}{\rightleftharpoons}} \mathrm{x}_2, \tag{3.40}$$

which when linearised gives a matrix of the form

$$\mathbf{A} = \begin{bmatrix} -k_1 & k_2 \\ k_1 & -k_2 \end{bmatrix}. \tag{3.41}$$

Despite this, we elected to avoid any form of parameter dependency since systems which violate these assumptions (in the non-robust case) only function within a limited parameter space, or (in the robust case) can be challenging to implement. For example, for reversible reactions as in (3.40) it may be difficult to tune the relative rate values k_1 and k_2, or to find a biological component with this property that is also able to interact with other species in the system as required. Other derived constraints can also be violated in practice (for example, the λ repressor protein CI can act as both an activator and repressor [169], violating Constraint 3.2.4), but were included since our goal is to limit the design space to include only the most readily implementable systems.

The present analysis can be extended to consider systems of more than one input by careful selection of node identities and layout. For example, feed-forward architectures often found in natural biological systems [170] (illustrated in Fig. 3.4) can be constructed by having the first node (x_1) act as proxy for the system input. By enforcing fast dynamics of this node it can track the input signal, though it will be subject to consumption/disturbance by downstream processes (as are inputs such as chemical inducers [151]).

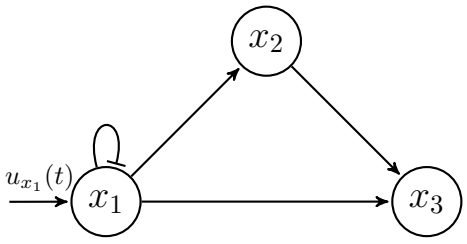


Figure 3.4: Simple feed-forward network structure, in which an inducer drives the steady-state concentration of x_1. By treating x_1 as a direct proxy for the input to this network it can be thought of as an input interacting with both x_2 and x_3.

The linearised approach to design described in this chapter simplifies the process of defining network structures and guiding implementation, but does so at the expense of

ruling out a class of non-linear systems for which the system matrix (as in (3.38)) would have non-constant terms. Such systems can be designed to adapt to a wide range of signals whilst requiring fewer components than those described by our linear design approach. For example, an $n = 2$ dimensional non-linear system can adapt to all sub-exponential inputs (including $u(t) = t^k$ for all k) [171, 172]. However, such systems can be challenging to implement using biological components [171], and in fact for many applications we desire adaptation or disturbance rejection for certain classes of signals but not others [173]. For example, many natural systems (including bacterial chemotaxis [174]) utilise negative feedback architectures which adapt to step inputs, but not to ramps, in order to sense nutrient gradients in their environment. This contrasts with other natural regulatory architectures, such as incoherent feed-forward loops, which adapt to a wider range of signals in order to reduce a system's sensitivity on environmental fluctuations [173].

As with all mathematical descriptions of biological systems, it is important to note that our approach involves making a range of approximations which in some circumstances may be violated. For example, our seemingly simple assumption that it is possible to build a network in which one species does *not* influence another (i.e. there is an off-diagonal zero in $\mathbf{A}$) may not hold: limited ribosomal resources mean that if one gene is translationally activated this may have a repressive influence on the translation of others [88]. Deliberate utilisation of these secondary effects may provide design approaches which could (for example) violate Constraint 3.2.4. Furthermore, zeroth-order degradation (zero terms on the diagonal of $\mathbf{A}$) is difficult to achieve in practice, since there will always be some rate of protein dilution and passive degradation [175]. Another assumption is made when linearising the more biologically-relevant non-linear system, as this approximation is only applicable within a given range of parameter values, thus constraining the regime in which our system can achieve its design specification. Combined, these and other second-order effects (which are prevalent throughout biological systems) mean that biological networks can almost never be *perfectly* adapting or rejecting of disturbances, though they can approximate these capabilities [158].

3.8 Conclusion

In this chapter I have described a set of constraints that can be employed to simplify the design and implementation of biological networks that achieve adaptation and disturbance rejection. Using this framework an engineer can quickly narrow the design space to a set of viable linear network topologies, which can then be embedded in a non-linear system of biological relevance. Though biologically feasible designs are possible outside of the assumption set outlined, these assumptions ensure that the design method provides only the most readily-implementable architectures. These can then be built in the laboratory from the ever-growing array of components available to synthetic biologists, which will be employed in subsequent chapters.

4

sRNA-Based Controllers

In this chapter I use small-RNA (sRNA) to implement and tune synthetic biological feedback loops. I develop a simple *a priori* model of this system to make qualitative predictions about its behaviour, it is implemented experimentally, and then a more complex model of its behaviour is derived and fit. This model is used to analyse variability in the system, and investigate how it can be tuned to reduce sensitivity to variation in each input. Discussion in this chapter focuses on the design, modelling, and analysis of these architectures as this was my primary contribution to the study, though the actual experimental procedures are outlined briefly in Section 4.2. Most of the results from this chapter are published in [42] and [62].

4.1 Introduction

sRNA (small-RNA) based regulatory networks have been a recent focus of synthetic biological research due to their adaptability and ease of implementation [18, 78]. At a fundamental level they utilise complementary base-pairing between RNA strands to cause structural changes in the RNA, thereby influencing the rate at which the genes it encodes are read (discussed in Section 2.1.3). Such systems can be used to build relatively fast-acting feedback loops, and provide a means for realising summing and difference junctions in biological contexts [62]. Systems based on the similar principle of DNA strand complementarity have been developed both theoretically and experimentally [176–178], and can be adapted for implementation of feedback control [39, 54]. Though the design of synthetic biological devices that utilise complementarity in such a way is conceptually

straight-forward, challenges remain in the accurate *a priori* prediction of their behaviour, due to factors including their tendency to form complex secondary structures which impact their function [179]. Nevertheless, in this section I aim to demonstrate the ability of sRNA to tune existing, and create new, synthetic biological feedback loops that may address some of the control challenges outlined in Section 2.2.1.

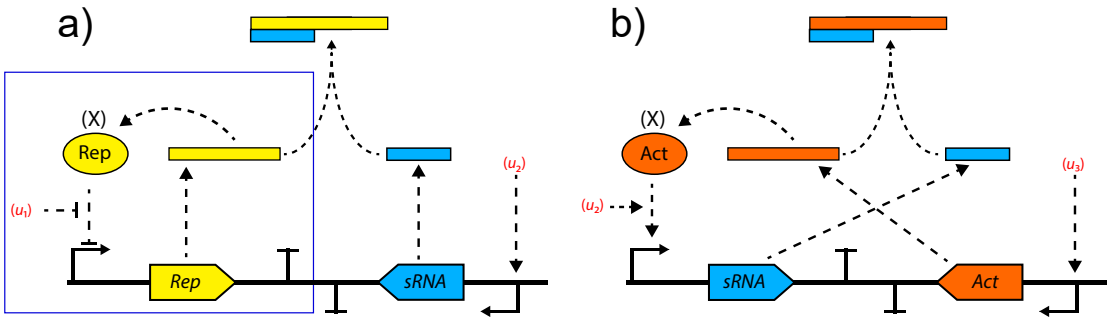


Figure 4.1: a) Circuit 1 - An sRNA-tuned autorepressor circuit consisting of: a gene encoding a repressor protein (Rep) whose activity is inhibited by u_1; a promoter upstream of the repressor-encoding gene (Rep), which the repressor protein can bind to and inhibit transcription from; an sRNA, whose expression can be induced through an external input (u_2) and which reduces the translation of the autorepressor mRNA, thus reducing the effective feedback strength of the autorepressor loop (blue box). **b)** Circuit 2 - A closed-loop sRNA feedback circuit consisting of a translation-inhibiting sRNA, whose expression can be induced through an external input (u_2), in closed-loop feedback with an mRNA encoding the transcriptional activator (Act) of the inducible promoter upstream of the sRNA-encoding gene. The steady-state output can primarily be set using a different external input (u_3) which binds to a constitutively-expressed transcriptional activator (not shown) inducing transcription of the Act gene.

The first system investigated (**Circuit 1**) is based on a TetR Autorepressor architecture (Fig. 4.1a): the autorepressor (or negative autoregulator) is a common genetic control motif, consisting of a protein TetR, which binds to (and prevents transcription) from its own promoter P_{tet}. When a chemical inducer u_1 (Tetracycline, or its common analogue anhydroTetracycline or aTc for short) binds to TetR it is no longer able to bind to P_{tet}, and hence transcription can proceed. This architecture has been well-studied in various contexts [31, 32, 103, 180], and is employed ubiquitously to control expression of repressors in both prokaryotes [29] and eukaryotes [30]. Since it represents a standard, minimal, regulatory unit, the autorepressor can be used as a model system for the comparison of experimental results with *in silico* modelling of simple gene regulatory networks (and

indeed this is done in our paper [151]). To the standard autorepressor (blue box in Fig. 4.1a) we introduce an sRNA which binds to (and prevents translation from) the mRNA encoding TetR. This sRNA is in turn under the regulation of an activating transcription factor RhaS, which responds to the chemical inducer u_2 (L-Rhamnose). By varying this inducer we can tune the strength of feedback in the autorepressor in real time.

In the second system investigated (**Circuit 2**) sRNA is used directly to "close the loop", creating a novel feedback structure with the activating transcription factor RhaS (Fig. 4.1b). In its normal state the promoter P_{rhaBAD} has a low transcription initiation rate, but when RhaS binds to its inducer L-Rhamose it can then bind to P_{rhaBAD} and drastically increase translation [181]. For this circuit we introduce a second tuning inducer, u_3 (m-toluic acid, which interacts with the XylS transcription factor) to govern the production rate of the regulatory protein RhaS. A similar architecture to ours (which was published after our work) has been proposed as a controller capable of adaptation/disturbance rejection, though direct comparison to our work is difficult due to substantially differing parameter regimes (it is designed to operate *in vitro* rather than *in vivo*) [182].

To compare our models of each circuit with experimental data we set all time derivatives to zero, and numerically solve each system of equations in MATLAB to find steady state concentrations of each species. These are then fitted to the experimental data to identify unknown parameters. This modelling procedure therefore approximates each circuit as being near steady-state when measurements are taken. This assumption is commonly made in the modelling of biological systems [144], though it may be violated in cases where synthetic circuits are particularly slow to respond / stabilise. In that case our model fitting will tend to over-estimate rate parameters, though by a similar factor for each reaction or inducer combination.

The experimental implementation and characterisation of these circuits is briefly summarised (as I only played a minor part) in Section 4.2. In Section 4.3 we describe modelling structures that will be used to describe each of the regulatory architectures

employed in this section. Parameter values (some of which are taken from the literature, and some fit in subsequent sections) are then outlined in Section 4.4. In Section 4.6 I describe a simple *a priori* model for the autorepressor component of Circuit 1, before deriving a more complex model and fitting this to experimental data. Section 4.7 extends this analysis to the sRNA-tuned autorepressor, and the noise and dynamic response of this model are investigated. In Section 4.8 a similar investigation is performed for Circuit 2. Finally, Section 4.9 discusses the results presented in this section and reflects on the future potential for sRNA-based control in synthetic biology, before Section 4.10 concludes the chapter.

4.2 Experimental Summary

The laboratory procedures used to create and characterise the circuits described in this chapter were primarily performed by Ciarán Kelly and Andreas Harris, so are only briefly summarised here. The genes required for our synthetic circuits were synthesized by Integrated DNA Technologies Inc. and inserted into standard plasmid backbones (summarised in the supplementary of [42]), and then transformed into *E. coli* strains MG1655 or JW3876 for characterisation. GFP is fused to the protein of interest in each circuit (TetR in Circuit 1 and RhaS in Circuit 2) to allow its concentration to be estimated. Cells were grown in either EZ rich defined media or M9 minimal media, to which appropriate antibiotics were added. Fluorescence characterisation experiments were performed by growing cell cultures overnight for 16 hours, and then diluting into subcultures with optical density of 0.05 in 96-well plates. Appropriate inducers were added at this time, and plates were incubated at 37°C with rapid shaking. The optical density of each sample was measured during growth by a plate reader. After 5 hours (when the cells were in late exponential growth phase) samples were taken and distributions of fluorescence per cell were measured via flow cytometry; this is the experimental data presented in this chapter. For an in-depth description of the experimental procedures see the published results in [42].

4.3 Model Structures

The inducible TetR transcription regulation system

We model the transcription rate from P_{tet} as a function of the concentration of the regulatory protein TetR (T) and its inducer anhydroTetracycline (aTc), u_1. This formulation is based upon past studies of this regulatory system [183, 184], and has been shown to accurately model the dynamic and steady-state behaviour of similar systems [185]. The activity of this promoter is given by:

$$\gamma_T = \frac{\alpha_t}{1 + \left(\frac{T}{K_T(1+(u_1/K_{u_1})^{n_{u_1}})}\right)^{n_T}} + \frac{\alpha_L\left(\frac{T}{K_T(1+(u_1/K_{u_1})^{n_{u_1}})}\right)^{n_T}}{1 + \left(\frac{T}{K_T(1+(u_1/K_{u_1})^{n_{u_1}})}\right)^{n_T}} \tag{4.1}$$

where α_t is the maximum translation rate from the promoter, K_T and K_{u_1} are equilibrium constants that define the system's saturation point in terms of its regulatory protein and inducer respectively, and n_T and n_{u_1} are exponents that define the cooperativity of the TetR and aTc binding processes respectively. Moreover, α_L parameterises the promoter's leakiness, and represents the transcription rate from promoters to which TetR is bound. Alternate approaches to modelling the *tet* promoter exist [5, 144, 151]; we select the form in (4.1) due to its ability to model autoregulatory systems (as it depends on both TetR and inducer concentration), and the availability of parameters in the literature [185].

The inducible RhaS transcription regulation system

We model the transcription rate from P_{rhaBAD} as a function of its regulatory protein RhaS (R) and its inducer L-Rhamnose (u_2) as follows:

$$\gamma_R^* = \alpha_r^* \frac{\frac{Ru_2}{K_{u_2}^* + u_2}}{K_R + \frac{Ru_2}{K_{u_2}^* + u_2}} \tag{4.2}$$

where α_r^* parameterises the absolute transcription rate, K_R is an equilibrium constant that defines the system's saturation point in terms of RhaS concentration, and $K_{u_2}^*$ is the cooperativity of the RhaS binding process. The form of (4.2) mirrors that used to model other activating transcription factors, where a Hill coefficient of 1 is used [185]. This formulation accounts for saturation of the promoter as the concentration of the regulatory protein increases, as well as saturation of regulatory protein activity as inducer

increases. In Circuit 1 the concentration of R is assumed to be constant (constitutive expression), and so in this case (4.2) simplifies to:

$$\gamma_R = \alpha_r \frac{u_2}{K_{u_2} + u_2} \tag{4.3}$$

In this case the parameter K_{u_2} depends upon the concentration of regulatory protein R and hence its value is different to that in (4.2).

The inducible XylS transcription regulation system

The XylS transcriptional regulatory system is only employed constitutively in our work (that is, we do not regulate the concentration of XylS), and so we model its behaviour using a saturating function of its inducer m-toluic acid (u_3):

$$\gamma_X = \alpha_x \frac{u_3}{K_{u_3} + u_3} \tag{4.4}$$

where α_x parameterises the absolute transcription rate, and K_{u_3} is an equilibrium constant that defines the system's saturation point in terms of its inducer. Here we again have a Hill coefficient of 1 for the inducer binding process. For Circuit 2 we replace α_x by α_x^* in (4.4) to give γ_X^* since the context of P_m, and the sRNA it regulates, changes.

Translational regulation by sRNA

In the experimental component of this study we employed the small-RNA scaffolds described by Na *et al.* [18], which depend upon Hfq chaperone proteins for mediation of sRNA-mRNA binding. Hfq mediation of sRNA-mRNA binding is a complex process for which a range of mathematical models exist [186], with the influence of Hfq varying greatly depending on the particular sRNA employed [187]. Past studies demonstrate that Hfq generally separates quickly from an sRNA-mRNA complex once complementarity is established [188]. This allows it to "actively cycle" between many RNAs targets [189], which facilitates its sharing between natural sRNA regulatory processes and those introduced in synthetic systems. For modelling purposes we have therefore assumed that our synthetic system does not saturate the natural Hfq pool.

To enable parameter fitting, and to simplify our model, we describe Hfq-mediated sRNA-mRNA binding as a first-order reaction of the form:

$$s + t \xrightarrow{K_1} c$$

where s is sRNA concentration, t is mRNA concentration, and c is the concentration of their bound complex. We assume that the lifetime of complex c is short compared to its un-binding half-life, which justifies modelling of this reaction as uni-directional (i.e. sRNA-mRNA binding permanently removes the mRNA from the system). The approximations we make in this description of the sRNA/mRNA/Hfq interaction will likely result in our model misrepresenting the actual concentration of sRNA, since the active binding and catalysis process mediated by Hfq would provide a higher "apparent" sRNA concentration than is actually present within the cell. This will influence the absolute values of parameters that govern sRNA levels (i.e. its transcription and binding rates), which could quantitatively impact simulations of the system's dynamic response.

4.4 Parameter Selection

The experimental measurements in this study are recorded in terms of fluorescence per cell in arbitrary units. However, because our models describe each species in terms of molecules per cell there is an unknown (ideally linear) relationship between a cell's copy number of GFP, and its total fluorescence. Protein copy numbers for similar circuits typically range from 10^1 to 10^4 per cell [103], with the upper end of this range corresponding to strong expression of stable proteins (such as those considered in our study). Consequently, we assume a one-to-one relation between measured fluorescence and protein copy number. This assumption will mean the value of certain parameters that we fit in our model (for example $\beta_{T,R}$, $K_{1,2}$, and $K_{T,R}$) is inclusive of a fluorescence to copy-number scaling factor, though this will not qualitatively impact the model's behaviour. Due to the number of free parameters in our model we use the literature to estimate some parameters' values, such that the rest are identifiable from our experimental data. The values and sources of each parameter are summarised in Table 4.1.

Table 4.1: Parameter values used in modelling each circuit with structures as illustrated in Figs. 4.2 and 4.10. The "Source" column describes the origin of each parameter value; they either derived from the literature (in which case references are given), or inferred from data in the indicated section.

	Description	Value	Unit	Source
α_t	Transcription rate of *tetR-GFP* gene from P_{tet} in sRNA-tuned Autorepressor	0.85	nM s^{-1}	[190, 191]
α_L	Transcription leak rate of *tetR-GFP* gene from P_{tet} in sRNA-tuned Autorepressor	0.090	nM s^{-1}	Section 4.7
α_r	Transcription rate of sRNA 1 from P_{rhaBAD} in sRNA-tuned Autorepressor	1.67×10^3	nM s^{-1}	Section 4.7
α_r^*	Transcription rate of sRNA 2 from P_{rhaBAD} in closed-loop sRNA feedback circuit	1.67×10^3	s^{-1}	Set to α_r
α_x	Transcription rate of sRNA 1 from P_m in sRNA-tuned Autorepressor	841	nM s^{-1}	Section 4.7
α_x^*	Transcription rate of *rhaS-GFP* gene from P_m in closed-loop sRNA feedback circuit	0.85	nM s^{-1}	[190, 191]
β_T	Translation rate of tetR-GFP mRNA	0.0154	s^{-1}	Section 4.7
β_T^x	Translation rate of tetR-GFP mRNA (XylS inducer)	0.0224	s^{-1}	Section 4.7
β_R	Maximum translation rate of rhaS-GFP mRNA	0.0243	s^{-1}	Section 4.8
K_1	sRNA-mRNA binding rate in sRNA-tuned Autorepressor	7.79×10^{-9}	nM^{-1} s^{-1}	Section 4.7
K_2	sRNA-mRNA binding rate in closed-loop sRNA feedback circuit	2.24×10^{-7}	nM^{-1} s^{-1}	Section 4.8
K_T	P_{tet}-TetR binding constant	0.0119	nM	Section 4.7
K_{u_1}	TetR-aTc binding constant	5.41×10^{-4}	nM	Section 4.7
K_R	P_{rhaBAD}-rhaS binding constant	1.036×10^6	nM	Section 4.8
K_{u_2}	RhaS-L-Rhamnose binding constant from (4.3)	2.6×10^3	nM	Section 4.7
$K_{u_2}^*$	RhaS-L-Rhamnose binding constant from (4.2)	1.4×10^3	nM	Section 4.8
K_{u_3}	XylS-*m*-toluic acid binding constant	1.78×10^5	nM	Section 4.8
n_T	P_{tet}-TetR binding cooperativity	3	none	[185]
n_{u_1}	TetR-aTc binding cooperativity	1.11	none	Section 4.7
δ_t	TetR-GFP mRNA degradation rate	4.1×10^{-3}	s^{-1}	[192]
δ_r	rhaS-GFP mRNA degradation rate	4.1×10^{-3}	s^{-1}	[192]
δ_c	mRNA-sRNA complex degradation rate	4.1×10^{-3}	s^{-1}	[192, 193]
δ_T	TetR-GFP protein degradation rate	3.9×10^{-4}	s^{-1}	[191]
δ_R	RhaS-GFP protein degradation rate	3.9×10^{-4}	s^{-1}	[191]
δ_s	sRNA degradation rate	8×10^{-4}	s^{-1}	[187, 194]

The maximum transcription initiation rates for the *tetR-GFP* and *rhaS-GFP* genes are set as $\alpha = 0.05\ s^{-1}\ gene^{-1}$ to correspond to typical levels for a mid-strength promoter in *E. coli* [190, 191]. The plasmids used in our circuits have typical copy numbers between 15 and 20 [42], which we approximate as ∼17 per cell and multiply by α to give $\alpha_t, \alpha_x^* = 0.85\ s^{-1}$. The mRNA degradation rate is set as $\delta_{t,r} = 4.1 \times 10^{-3}\ s^{-1}$ to represent a typical mRNA half-life of 2.8 minutes for *E. coli* growing in exponential phase [192], and we assume that degradation of the sRNA-mRNA complex (δ_c) occurs at this same rate (as has been found in sRNA regulatory systems [193]). The protein degradation rate is set as $\delta_{T,R} = 3.9 \times 10^{-4}\ s^{-1}$ to correspond to dilution during growth with a 30 minute doubling time [191]. For the sRNA we sum its dilution rate (equal to $\delta_{T,R}$) with its measured degradation rate of $\delta_{deg} = 4.1 \times 10^{-4}\ s^{-1}$ [194] (which is an order of magnitude slower than typical mRNA due to the stabilising effect of Hfq [187]) to give $\delta_{m_s} = 8 \times 10^{-4}\ s^{-1}$. The precise values of degradation rates are not critical for the fitting of our models to steady-state circuit outputs: each biochemical reaction is a trade-off between creation and degradation, and thus a changed value of δ can be countered by a change in creation rate (to yield an equivalent equilibrium species abundance). However, the dynamic behaviour of our system *is* a function of these assumed parameters, and hence any quantitative conclusions (e.g. estimations of response time) drawn from dynamic simulations will depend upon their accuracy. The assumption underlying the protein degradation rate (that it occurs due to dilution exponential growth) is only true for part of each experiment: examining the growth curves measured (available in [42]) reveals that cells are beginning to exit their exponential growth phase prior to the experiment's completion. Nevertheless, we make this assumption throughout as a significantly more complex model (with many additional unknown parameters) would be required to analyse a case with variable growth rate. Experimental setups that allow indefinite growth (i.e. using a Turbidostat as described in Chapter 7) will allow experiments where exponential growth is maintained in the future.

For the model of activity for the *tet* promoter (4.1) we set $n_T = 3$, as found in past studies of similar autorepressor systems [185]. Cooperativity of aTc binding to TetR (n_{u_1}) was fit to the experimental data, as there was a large degree of uncertainty in

its reported value [185]. To convert inducer units (typically ng/ml) into concentrations we use molecular weights of 426 g mol^{-1} for aTc, 164 g mol^{-1} for L-Rhamnose, and 136 g mol^{-1} for *m*-toluic acid. All other parameters (summarised in Table 4.1) were fit to experimental data as discussed in subsequent sections. Assuming a cell volume of 0.6 μm^3 (typical of *E. coli*) a 1 nM concentration corresponds to approximately 1 molecule per cell, and thus parameters values can easily be converted to units of molecules per cell for use in stochastic simulations.

4.5 Modelling Noise

To assess the impact of variability on our system we will simulate its behaviour in the presence of intrinsic (due to stochasticity of the system's biochemical processes) and extrinsic (due to fluctuations in the concentration and state of cellular components that it interacts with) noise types [84, 85]. The sources and interactions between these noise sources are discussed in Section 2.1.4. In this chapter we model different noise types independently and then demonstrate the anticipated outcome when combined.

We quantify noise in terms of the Coefficient of Variation (CoV) of the output (concentration of protein T) in a population, which is a metric commonly used in studies of gene expression noise [195]. CoV is a useful metric for describing the variability (relative to the mean) in data that is symmetrically distributed (e.g. normally distributed, or log-normally distributed and transformed on a log scale). For log-normally distributed data (as is often found in biological systems [145, 147]) CoV is given by:

$$CoV(T) = \sqrt{e^{s_{ln}^2} - 1} \tag{4.5}$$

where $s_{ln} = \sigma(ln(T))$ is the standard deviation of the natural logarithm of the variable T.

In the following sections intrinsic variability in each system is modelled using the tau-leaping method, based on the Gillespie algorithm [148]. Such an approach (described in detail in Section 2.4.5) treats each biochemical reaction as a stochastic process, probabilistically incrementing each species abundance by an integer amount in each

time step depending upon the calculated reaction rate. Stochastic simulations of intrinsic noise are complemented by a model of extrinsic noise, wherein we sample individual parameter values from a log-normal distribution, and then observe the distribution of output levels in a cell population. In reality extrinsic noise sources would have some impact upon every parameter's value, but we must select a few likely places for introduction of noise to make our simulations tractable. The particular parameters chosen, and the magnitude of the variability introduced at each, are ultimately somewhat arbitrary, and we will therefore be more interested in qualitative trends in our simulation results rather than the absolute levels predicted.

4.6 The Autorepressor

In this section the autorepressor component of sRNA Circuit 1 is analysed, for which a detailed schematic is provided in Fig. 4.2. Initial modelling of the autorepressor component of Circuit 1 (i.e. in the absence of sRNA, when $u_2 = 0$) allows us to fit some parameters and compare our results with the literature, before moving on to describe the circuit in its entirety in Section 4.7.

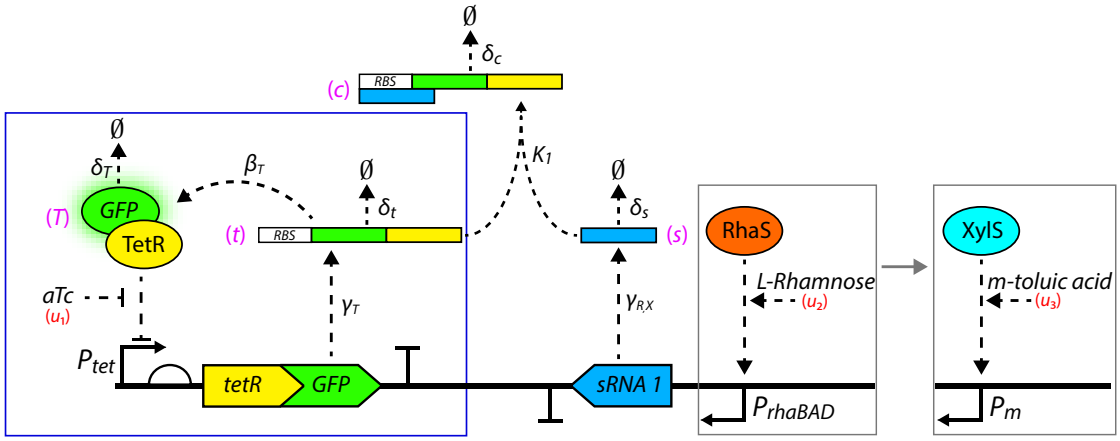


Figure 4.2: Layout for the Autorepressor (blue box only), and Circuit 1, the sRNA-tuned Autorepressor (whole figure). The alternate XylS inducer system was tested experimentally (Fig. 4.6), which corresponds to exchanging the grey boxes. Components are colour coded across their gene / mRNA / protein forms, and degrade to the null state Ø. Parameters are as defined in the text, state variables considered in modelling are highlighted in purple, and inputs (inducers) are in red.

4.6.1 Simple Model

In order to allow *a priori* investigation of our system's behaviour we build a simple model of the autorepressor, of the form:

$$\frac{dX}{dt} = \beta L + \frac{\beta(1-L)}{1+\left(\frac{X}{0.01+u_1}\right)^3} - X \tag{4.6}$$

where X is the system's protein concentration, L is a leakage term, and β is a noise term that will be used to introduce variability. This model reflects those used previously to describe the autorepressor [185], though it is largely de-parameterised as we only intend to investigate its qualitative behaviour. We find good qualitative agreement between experimental observations of the autorepressor's behaviour (Fig. 4.3) and calculated steady-state values of X as u_1 is varied in Fig. 4.4 (simulation where $u_2 = 0$). We crudely simulate the influence of extrinsic noise on this system by sampling the parameter β from a log-normal distribution (motivated in Section 4.5) with a mean of one, finding that the output achieves minimal variability for intermediate induction levels (as observed experimentally in past studies [103]).

4.6.2 Detailed Model

We construct a more detailed model for the Autorepressor (Fig. 4.2 - blue box) that depends on two states: the concentration of mRNA (t) and protein (T) for the TetR-GFP complex. The underlying biochemical reactions can be expressed as:

$$\begin{aligned} \emptyset &\xrightarrow{\gamma_{\mathrm{T}}} \mathrm{t} \xrightarrow{\delta_t} \emptyset, \\ \mathrm{t} &\xrightarrow{\beta_{\mathrm{T}}} \mathrm{t} + \mathrm{T}, \\ \mathrm{T} &\xrightarrow{\delta_{\mathrm{T}}} \emptyset, \end{aligned}$$

where γ_T is defined as in (4.1), and β_T is the maximal translation rate (Fig. 4.2). These reactions can be described using two coupled differential equations of the form:

$$\begin{aligned} \dot{t} &= \gamma_T - \delta_t t \\ \dot{T} &= \beta_T t - \delta_T T \end{aligned} \tag{4.7}$$

4.6.3 Parameter Fitting

We initially set $\alpha_L = 0$ in γ_T and use least-squares minimisation to fit the unknown parameters K_{u_1}, K_T, n_{u_1}, and β_T such that the steady-state output of the model (4.7) best matches the experimental data. For this we use the measurements of Circuit 1 with the *rhaBAD* promoter when sRNA is not induced (concentration of L-Rhamnose is zero). Experimental data and the model fit are presented in Fig. 4.3a. We observe that the model accurately captures the system's behaviour over the range of inducer concentrations measured, and that it correctly approximates the low- and high-inducer saturation levels anticipated from this promoter.

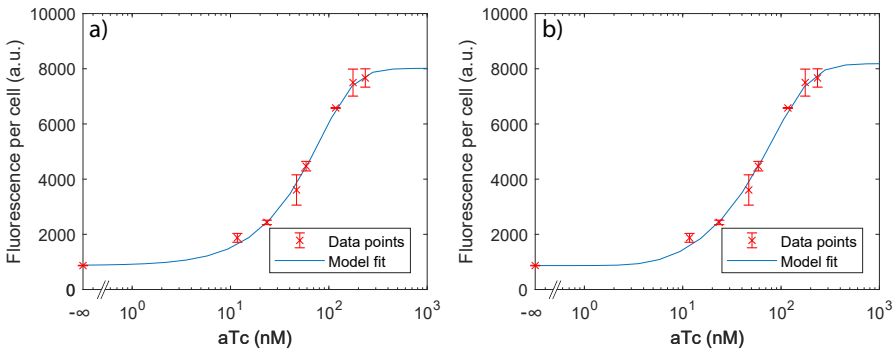


Figure 4.3: Autorepressor data and model fit. a) Model fit with $\alpha_L = 0$, yielding $K_{u_1} = 8.98$ nM, $K_T = 433$ nM, $n_{u_1} = 1.23$, and $\beta_T = 0.0151$ s^{-1}. Data points are measurements of the sRNA-tuned Autorepressor (pCK221) with the RhaS induction system but zero L-Rhamnose. **b)** Model fit from Section 4.7, yielding $K_{u_1} = 5.41 \times 10^{-4}$ nM, $K_T = 0.0119$ nM, $n_{u_1} = 1.11$, $\beta_T = 0.0154$ s^{-1}, and $\alpha_L = 0.090$ nM s^{-1}. These parameters are those summarised in Table 4.1. Error bars shown represent the standard deviation of three independent biological replicates.

From this data set alone it is not possible to uniquely identify the parameters K_{u_1}, K_T, n_{u_1}, and β_T in addition to α_L. This is because there are two possible contributing factors to the $u_1 \to 0$ limit of the promoter's output (4.1): first, if all promoters are bound to TetR at all times, then the $u_1 = 0$ output level is determined by α_L (this is considered in [183]). However, it is also possible for there to be a non-zero output when $u_1 = 0$ in the absence of leakage ($\alpha_L = 0$) if some promoters are not bound to TetR, as determined by the value of K_T. It is not possible to distinguish the relative contribution of these two mechanisms from experiments in which only aTc concentration is varied, and thus we must fit all parameters for this circuit simultaneously in Section 4.7, in which the apparent

value of α_t is varied due to sRNA induction. The result from this fit is presented in Fig. 4.3b, demonstrating close correspondence to Fig. 4.3a. When $\alpha_L \neq 0$, K_{u_1} takes a much smaller value, as this parameter is no longer required in determining the system's output level when $u_1 = 0$. Consequently, given the present data the absolute values of K_{u_1} and K_T are not identifiable in isolation. However, the quantity $(K_T)^{1/n_{u_1}}/K_{u_1}$ which arises in (4.1) for high inducer concentrations ($u_1 > K_{u_1}$) *is*: it is equal to 15.50 when $\alpha_L = 0$ and 34.12 when $\alpha_L = 0.09$, and thus the system's sensitivity to TetR concentration (T) is similar in both cases.

4.7 sRNA Circuit 1

In this section we move on to analyse the tuned autorepressor (Fig. 4.2), in which the inducible sRNA can regulate the strength of feedback.

4.7.1 Simple Model

To investigate the anticipated behaviour of this circuit prior to experimental implementation we extend the simple model of the autorepressor proposed in Section 4.6.1 to include the interaction of sRNA:

$$\frac{dX}{dt} = \beta L + \frac{\beta(1-L)}{1+\left(\frac{X}{0.01+u_1}\right)^3} - X - u_2 X \tag{4.8}$$

Here the final term ($u_2 X$) approximates the contribution of the inhibitory sRNA, which acts proportionally to the quantity of its targets (the mRNA that would be transcribed to produce X) and its induction level u_2. We calculate the steady-state value of X as u_1 is varied in Fig. 4.4a (for a range of sRNA levels determined by u_1), demonstrating the sRNA's ability to tune the system's response to u_1 (aTc). We also introduce variability into the parameter β, as described in Section 4.6.1, for which results are plot in Fig. 4.4b.

4.7.2 Detailed Model

We now extend the autorepressor model considered in Section 4.6 to include two additional states that account for sRNA inhibition of translation: the concentration of sRNA (s),

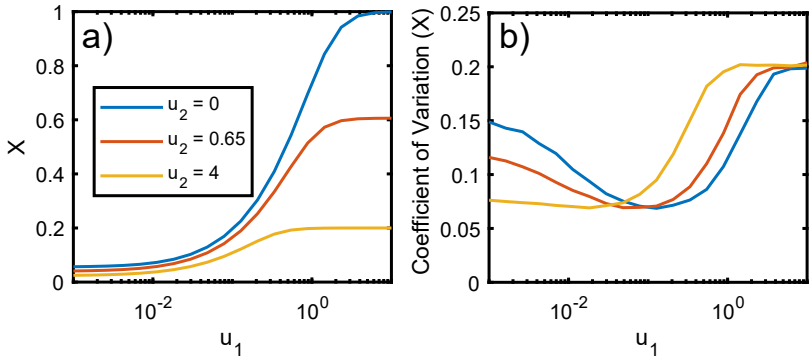


Figure 4.4: Simulations of the simple model of Circuit 1 outlined in Section 4.7.1. When $u_2 = 0$ this circuit is equivalent to the autorepressor as described in Section 4.6.1. **a)** The steady state output predicts that the addition of the sRNA would increase the fidelity of steady-state output (X) tuning when both external inputs are present. **b)** Noise simulations as described in Section 4.6.1, demonstrating that the use of an sRNA should not result in large increases in output noise.

and the concentration of the sRNA-mRNA complex (c). The underlying biochemical reactions can be expressed as:

$$\begin{aligned}
\emptyset &\xrightarrow{\gamma_{\mathrm{T}}} \mathrm{t} \xrightarrow{\delta_t} \emptyset,\\
\emptyset &\xrightarrow{\gamma_{\mathrm{R,X}}} \mathrm{s} \xrightarrow{\delta_s} \emptyset,\\
\mathrm{t} + \mathrm{s} &\xrightarrow{\mathrm{K_1}} \mathrm{c} \xrightarrow{\delta_c} \emptyset,\\
\mathrm{t} &\xrightarrow{\beta_{\mathrm{T}}} \mathrm{t} + \mathrm{T},\\
\mathrm{T} &\xrightarrow{\delta_{\mathrm{T}}} \emptyset,
\end{aligned}$$

where γ_R and γ_X are defined for the different inducer systems as in (4.3) and (4.4) respectively, K_1 is the binding rate of sRNA and mRNA, and β_{L_1} is the translation leak rate from the sRNA-mRNA complex. These reactions can be described using four coupled differential equations of the form:

$$\begin{aligned}
\dot{t} &= \gamma_T - \delta_t t - K_1 t s\\
\dot{s} &= \gamma_{R,X} - \delta_s s - K_1 t s\\
\dot{c} &= K_1 t s - \delta_c c\\
\dot{T} &= \beta_T t - \delta_T T
\end{aligned} \tag{4.9}$$

where the choice of γ_R or γ_X depends on the inducible promoter used (either P_{rhaBAD} or P_m) to control sRNA expression.

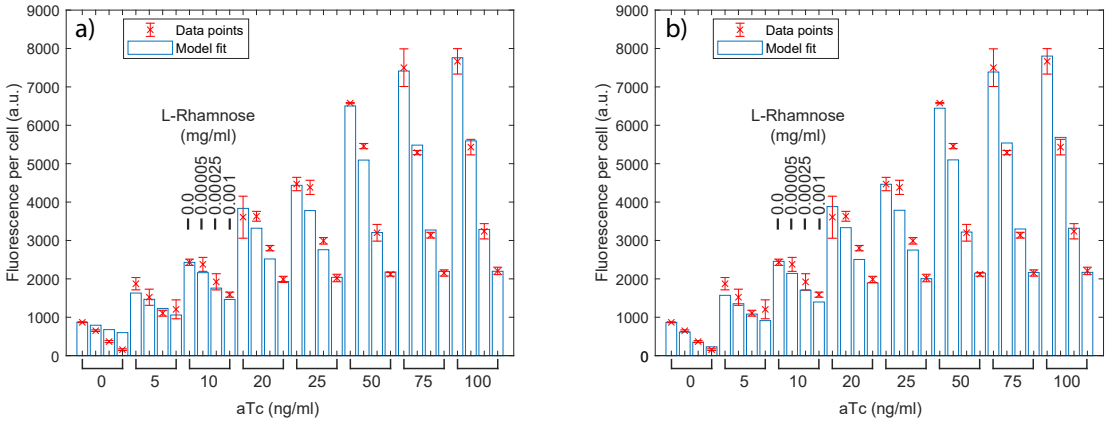


Figure 4.5: Circuit 1 with RhaS, data and model fit. a) Model fit for Circuit 1 in Fig. 4.2 with RhaS induction system. In this case we set $\alpha_L = 0$ and fit the three remaining parameters, yielding $K_1 = 1.080 \times 10^{-8}$ nM^{-1}, $K_{u_2} = 2.4 \times 10^3$ nM, and $\alpha_r = 1.12 \times 10^3$ s^{-1}. Other parameters are as in Fig. 4.3a. We observe good agreement between model and experiment when aTc ≥ 5 ng ml^{-1}, but when aTc $= 0$ there is substantial discrepancy. **b)** Model fit with α_L allowed to vary. Resulting parameter values are as in Table 4.1. We now find good agreement between model and data across all input combinations. Error bars shown represent the standard deviation of three independent biological replicates.

4.7.3 Parameter Fitting

We initially set $\alpha_L = 0$ and use least-squares minimisation to fit the unknown parameters K_1, K_{u_2}, and α_r in the model (4.9) such that its steady-state behaviour best matches the data presented in Fig. 4.5. Here we are using the parameters fit for the zero L-Rhamnose data in Fig. 4.3a when $\alpha_L = 0$. Experimental data and the model fit are presented in Fig. 4.5a. We observe that the model replicates the data well when aTc$\geq$ 5 ng ml^{-1}, but not so well as L-Rhamnose is increased in the zero aTc case. This discrepancy is a fundamental property of our model when $\alpha_L = 0$: it is impossible for the fold-change of output due to sRNA induction (that is, the output when L-Rhamnose equals zero divided by the output when L-Rhamnose is 0.001 ng/ml) to be greater at low aTc concentrations than at high aTc concentrations. Due to this phenomenon we conclude that when no aTc is present the system's output level is determined to a substantial degree by leakage from TetR-bound promoters (discussed in Section 4.3), thus requiring $\alpha_L \neq 0$.

We now allow α_L to vary and fit all free parameters (K_{u_1}, K_{u_2}, K_T, n_{u_1}, β_T, α_L, K_1, and α_r), recognising that only a ratio of the values of K_T and K_{u_1} will be identifiable as discussed in Section 4.6. The resulting parameters are listed in Table 4.1, and we observe

good agreement (Fig. 4.5b) across all inducer concentrations, with the model accurately capturing the system's behaviour for 0 aTc. Without aTc induction TetR concentration is predominantly a function of promoter leakiness, which means that a reduction in the concentration of T has little impact upon translation from P_{tet}. Therefore, the sRNA is able to achieve its maximal repression fold with the L-Rhamnose inducer saturated, reducing the system's output to ~ 0.2 times its value when L-Rhamnose is not present. This is not the case when aTc ≥ 5, since the dependence of (4.1) on both u_1 and T means that decreased translation of tetR mRNA due to sRNA-mRNA binding results in higher promoter activity, which feeds back to provide more free mRNA.

We now fit our models to the second variant of Circuit 1, in which the RhaS induction system is replaced by XylS (illustrated in grey box in Fig. 4.2). Since the autorepressor circuit and sRNA interaction remain unchanged we maintain most parameter values found for the RhaS circuit variant. We only fit β_T^x and α_x, and take the value of K_{u_3} found in Section 4.8. We first fit our model in the absence of m-toluic acid (Fig. 4.6a), yielding $\beta_T^x = 0.0224$ s^{-1}. Theoretically this value should be identical to that found in the RhaS version of Circuit 1 ($\beta_T = 0.0154$ s^{-1}): in neither case is the sRNA component of this circuit induced, so assuming minimal promoter leakage translation of identical mRNA should occur at the same rate. We hypothesise that their $\sim 30\%$ difference is likely due to the anticipated variability in inoculants between different experiments, as well as any differing burden effects that arise due to the constitutive expression of RhaS vs XylS within the cell. We then fit the value of α_x to the whole experimental data set, and we observe that the model is able to capture our system's behaviour (Fig. 4.6b), including at low aTc levels.

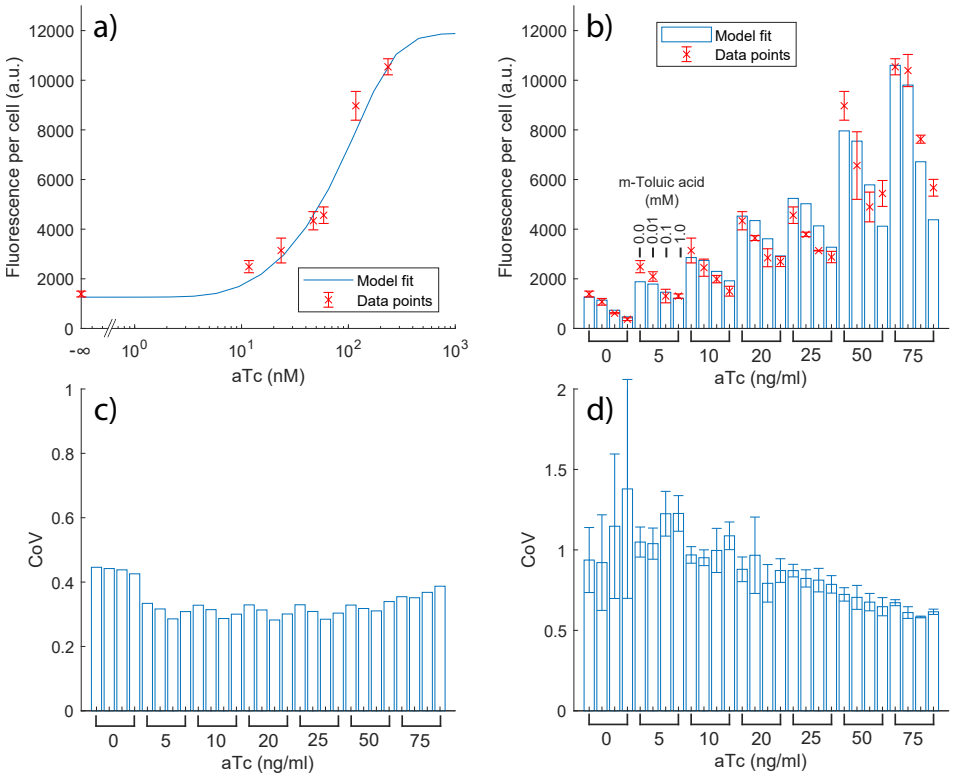


Figure 4.6: Circuit 1 with XylS, data and model fit. a) Validation of the autorepressor model (blue line) for the XylS circuit (when the concentration of m-Toluic acid is zero) with parameters as described in the text. **b)** Model fit for Circuit 1 in Fig. 4.2 with XylS induction system, demonstrating that a good fit is achieved despite many parameters being carried over from the RhaS variant of Circuit 1. **c)** Predicted coefficients of variation in the TetR-sfGFP concentration when tuned via both external inputs. Simulations include the influence of both extrinsic and intrinsic noise as described in Section 4.7.5. **d)** Experimentally-obtained coefficients of variation of fluorescence. Error bars shown represent the standard deviation of three independent biological replicates.

4.7.4 Dynamic Behaviour

To assess the dynamic behaviour of our feedback system we numerically integrate the system of equations (4.9) and apply a dynamic input to each inducer in turn. In doing so we are implicitly assuming that some reactions within our circuit (for example, inducer-transcription factor binding) occur quickly, since we do not explicitly model these. This assumption has been made in past studies which investigated the dynamic behaviour of similar transcription-factor systems [185] and was found to correspond well with experimental results. Other factors that introduce a time-delay, such as the uptake of inducer molecules into cells, are also not considered, and so simulated response times will under-estimate those achievable experimentally. We introduce a dynamic signal

to our system's inducers by setting:

$$u_i(t) = u_i^*(t) \times u_i^0 \tag{4.10}$$

where the time-dependent profile u_i^* determines the magnitude of variation for the inducer input i about its baseline value u_i^0. Results for these simulations for Circuit 1 with the RhaS induction system are presented Fig. 4.7. We observe that it is possible to reduce the system output's sensitivity to one input, whilst maintaining a given mean output, by using the other input as a tuning dial. For example, in Fig. 4.7 if $u_1^0 = 20$ ng ml^{-1} and $u_2 = 0$ ng ml^{-1} we see high sensitivity to fluctuations in u_1. However, if we set $u_1^0 = 75$ ng ml^{-1} and $u_2 = 1.45 \times 10^{-4}$ ng ml^{-1} the system's mean output remains the same, but equal percentage-change variations in u_1 have a substantially reduced impact. This is due to the sRNA (controlled by u_2) shifting the autorepressor's negative feedback equilibrium further towards saturation, reducing its sensitivity to changes in u_1. We also simulate the system's dynamic response to step inducer inputs (applied at $t = 0$), for which results are presented in Fig. 4.8. This simulation demonstrates an additional benefit of tuning the system using u_2 rather than u_1; the same output levels can be achieved without significant variation in response time.

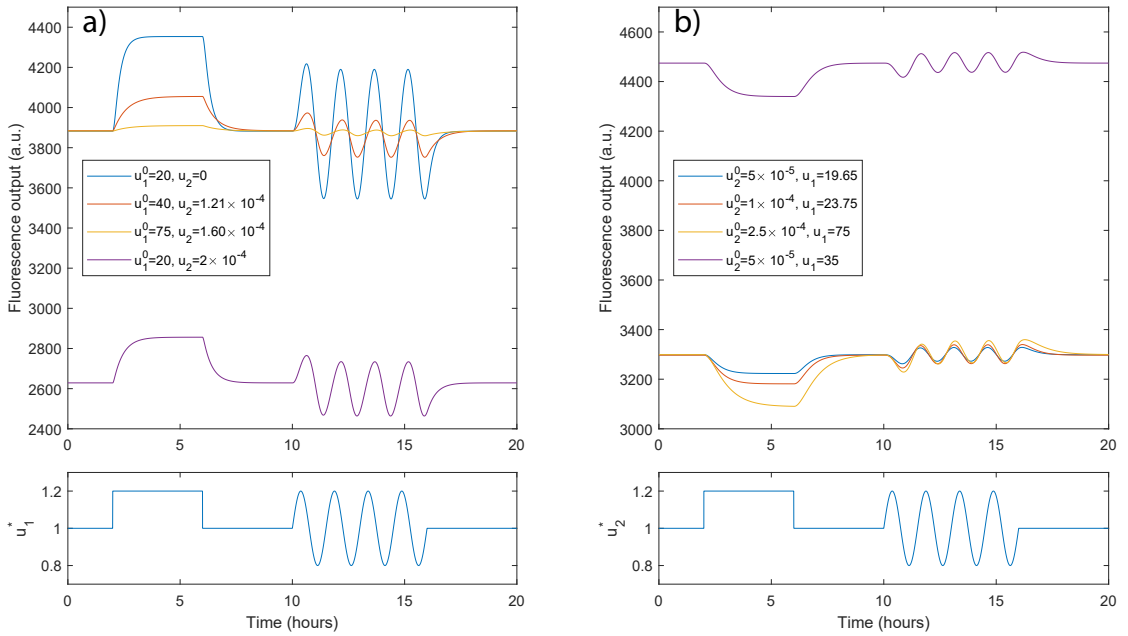


Figure 4.7: Dynamic simulations of Circuit 1. All inducer concentrations are in ng ml^{-1}. **a)** System output over time as u_1 (aTc) is varied according to (4.10) with the u_1^* profile illustrated. **b)** Similar to (a) but with u_2 (L-Rhamnose) varying over time. We observe a slower response to changes in u_2, and smaller output fluctuations for a given fold-change in this inducer.

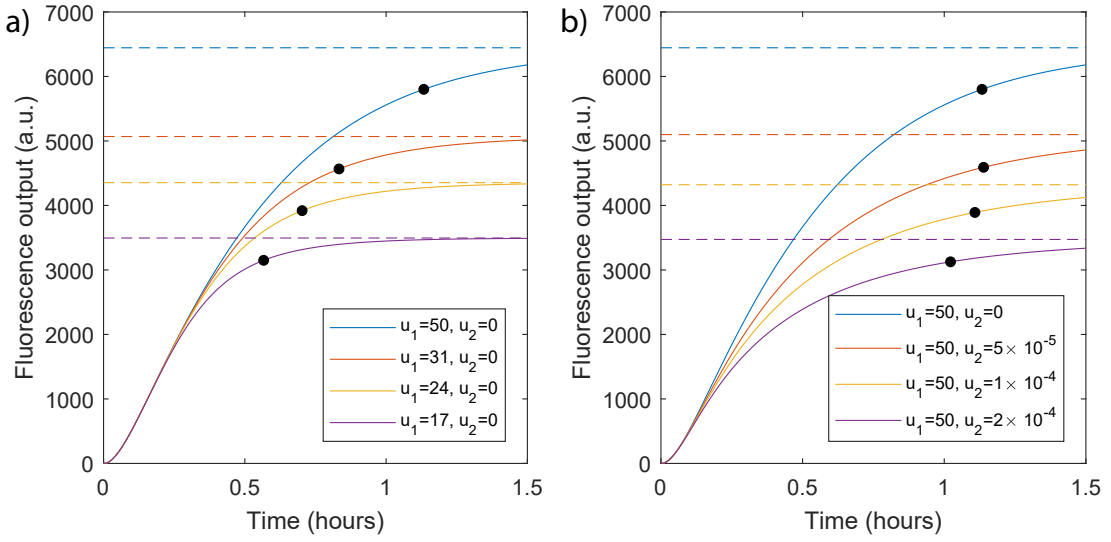


Figure 4.8: Step-response simulations of Circuit 1. **a)** Simulation without sRNA induction: The system's response time (black dots, defined as the point at which the output reaches 90% of its maximum value) increases as the system's input u1 is increased. **b)** Simulation of the autorepressor with sRNA induction, demonstrating that sRNA tuning allows the same absolute output levels to be achieved without substantially changing response time.

4.7.5 Noise Simulation

We now simulate the impact of intrinsic and extrinsic noise on our system, as described in Section 4.5. First we simulate intrinsic variability of the system in (4.9) with parameters as in Table 4.1 using the tau-leaping method. To account for stochastic variability in plasmid copy number we introduced a new equation to describe plasmid dynamics:

$$\dot{P} = \delta_P(P_0 - P) \tag{4.11}$$

where we set the plasmid degradation rate $\delta_P = \delta_T$ to correspond to dilution by cell growth, and P_0 is the mean plasmid copy number (set to be 17 for our plasmid backbone as described in Section 4.4). The RNA transcription terms in (4.9)a,b are then scaled by the factor $P(t)/P_0$ to account for variation in transcription due to stochastic variation in plasmid copy number.

It is likely that stochastic simulation of (4.9) will under-estimate noise intrinsic to our circuit's performance, since this system of equations lumps many complex processes into single parameters, and thus the influence of stochasticity at these steps is not modelled. For example, the Hfq-mediated binding of sRNA/mRNA is described by the single parameter K, which does not account for any stochasticity in intermediate steps (which occur at low species concentrations, where the influence of biochemical stochasticity is greatest [144]). We run stochastic simulations of a population of 10^5 individual cells until the population mean concentration of each species reaches an equilibrium, and then calculate the mean CoV of the equilibrated population over a further 1 hour of simulation. Results are presented in Fig. 4.9a, demonstrating that the magnitude of intrinsic noise is generally low ($CoV < 0.2$), and that for a given aTc (u_1) concentration noise decreases as sRNA concentration is increased (i.e. for larger values of u_2 (L-Rhamnose)).

We introduce extrinsic noise in this circuit by adding variability to gene expression at the translational step, and in the uptake/mixing of each inducer, by setting:

$$\beta_T^* = \Delta_1 \times \beta_T,\ u_{1,2}^* = \Delta_2 \times u_{1,2} \tag{4.12}$$

where β_T^* and $u_{1,2}^*$ are the translation rate and inducer concentration parameters with noise introduced, and $\Delta_{1,2} \sim \text{Log-normal}(\mu, \sigma^2)$ are independent random parameters sampled from a log-normal distribution for which we set $\mu = -\sigma^2/2$ such that the distribution's mean is unity. For the system in (4.9) we simulate a population of 10^5 cells with parameters sampled as described with $\sigma = 0.4$, for which results are presented in Fig. 4.9b. We observe qualitatively similar behaviour to that predicted by our simple model of this system (Fig. 4.4): a reduced level of variability is found for intermediate levels of aTc (u_1). The magnitude of variation introduced due to extrinsic noise here is substantially greater than that of intrinsic noise, though this comparison ultimately depends on the value chosen for σ.

In a real biological system it is difficult to separate measurements of noise introduced by intrinsic and extrinsic factors due to the influence of each factor depending on that of the other [94]. To investigate how the interplay of these sources would impact our model system we perform a combined simulation of both sources: we stochastically simulate our system (introducing intrinsic fluctuations) with parameters varied according to log-normal distributions as described above (to introduce extrinsic sources). Results from this simulation are presented in Fig. 4.9c, and are similar to those observed for simulations of extrinsic noise alone (Fig. 4.9b). Qualitatively, the predicted trends as aTc and L-Rhamnose concentration are varied match well with those measured experimentally (Fig. 4.9d).

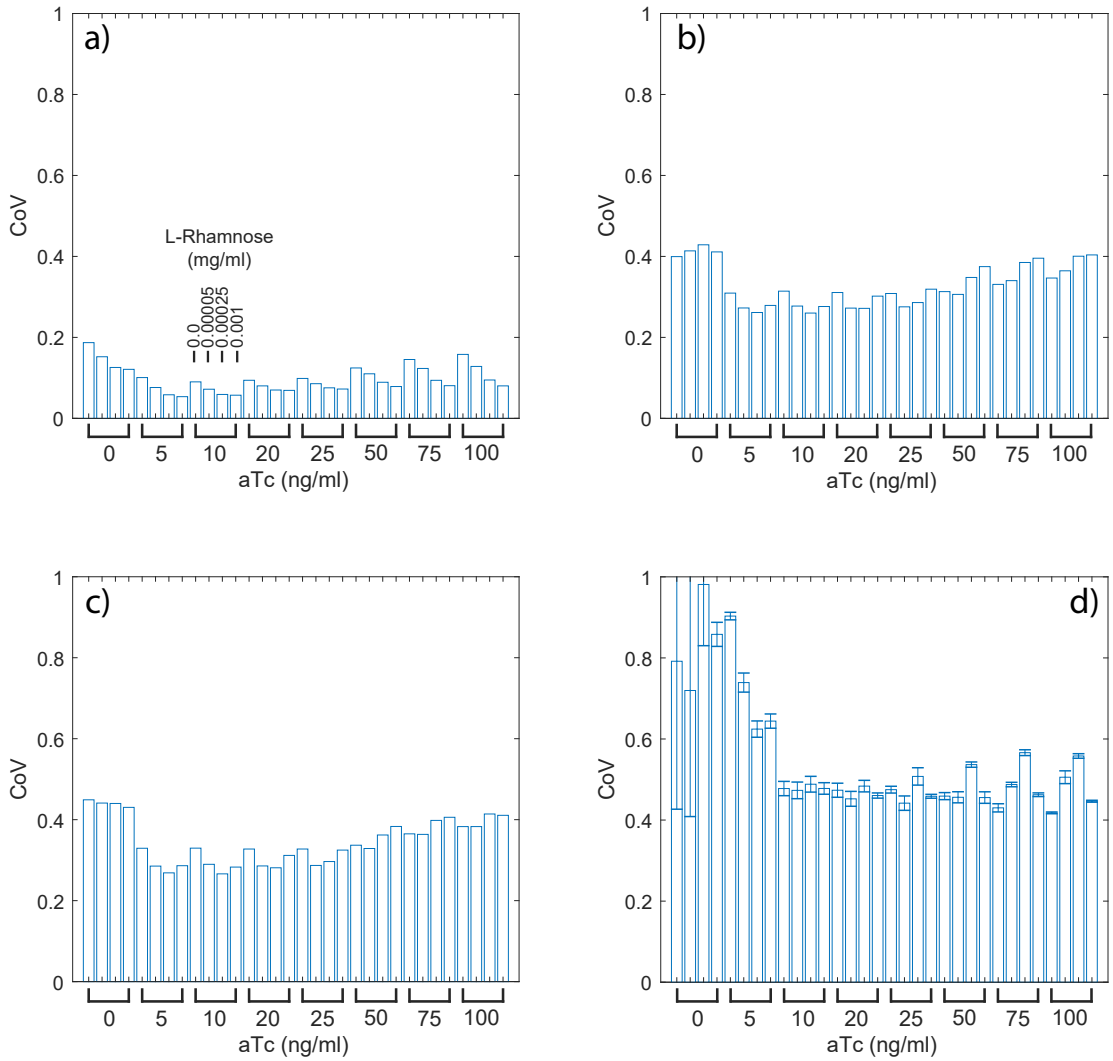


Figure 4.9: Circuit 1 noise simulations. Plots of coefficient of variation (CoV) calculated as in (4.5) for different noise sources, and experimental data. **a)** Simulation of intrinsic noise, due to stochasticity of cellular processes. **b)** Simulation of extrinsic noise, with parameters sampled from log-normal distributions as described by (4.12) with $\sigma = 0.4$. **c)** Combination of stochastic and extrinsic simulations. **d)** Experimentally-measured CoV.

4.8 sRNA Circuit 2

In this section we move on to analyse the second feedback circuit (Fig. 4.10), in which sRNA is used to "close the loop" with the activating RhaS transcription factor.

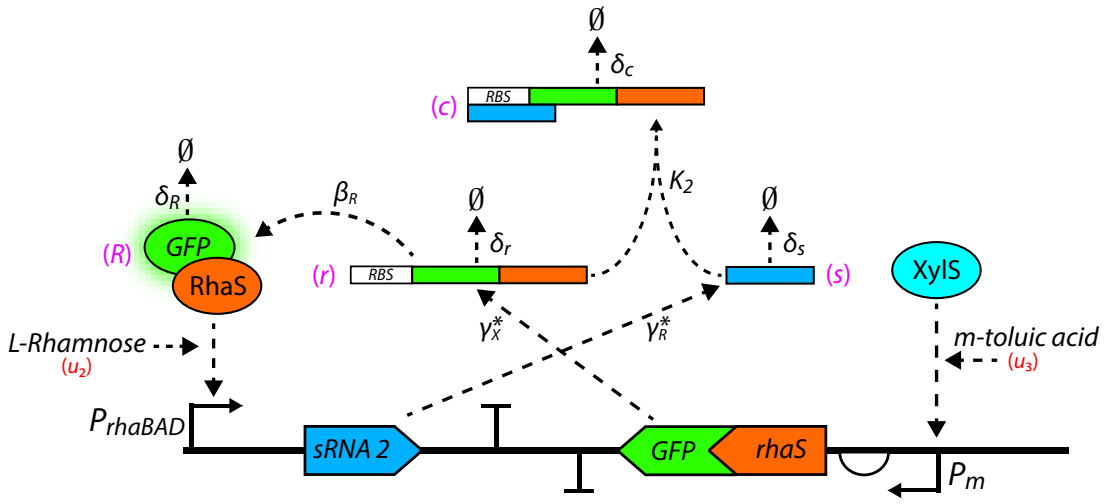


Figure 4.10: Layout for Circuit 2. The major difference between this and Circuit 1 in Fig. 4.2 is apparent; in this case the sRNA is included within the feedback loop. Components are colour coded across their gene / mRNA / protein forms, and degrade to the null state ∅. Parameters are as defined in the text, state variables considered in modelling are highlighted in purple, and inputs (inducers) are in red.

4.8.1 Simple Model

To investigate the anticipated behaviour of this circuit prior to experimental implementation we again designed a simple one-equation model of our system:

$$\frac{dX}{dt} = \frac{u_3}{u_3 + 1} \times \frac{\beta}{1 + \left(\frac{u_2 X}{1+X}\right)} - X \tag{4.13}$$

which combines the influence of u_2 (governing the activating transcription factor's activity and hence sRNA expression) and u_3 (which determines the translation rate of the transcription factor's mRNA) into a single equation. We calculate steady-state concentrations of this model in Fig. 4.11, and simulate noise using the parameter β as described in Section 4.6.1. We find that together the two inputs are able to tune the system's response over a wide range (Fig. 4.11b), and that CoV is anticipated to remain approximately constant (or decrease) as u_2 is increased (Fig. 4.12). To demonstrate the effect of feedback in this circuit we also investigate a model in which the activating transcription factor's concentration is decoupled from sRNA expression, given by:

$$\frac{dX}{dt} = \frac{u_3}{u_3 + 1} \times \frac{\beta}{1 + u_2} - X \tag{4.14}$$

for which results are presented in Fig. 4.11a. This simple simulation demonstrates qualitative differences in the steady-state behaviour of the open- and closed-loop systems as

inducer concentrations vary. Namely, the output of the closed-loop system is independent of u_2 when u_3 is small. We observe similar (closed-loop system) behaviour experimentally when u_3 is small: note that in Fig. 4.13b the system output is largely independent of L-Rhamnose when the m-toluic acid concentration is 0.01 mM.

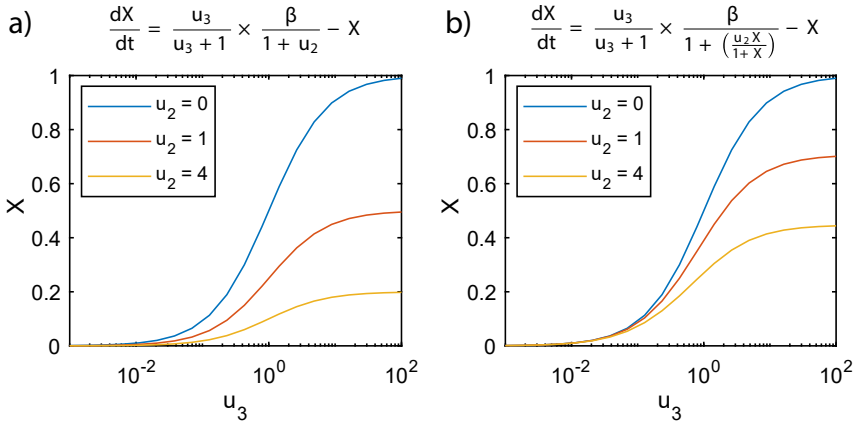


Figure 4.11: Simple model of the closed-loop sRNA feedback circuit. **a)** The open-loop model (4.14), in which the concentration of X does not impact its own rate of expression. **b)** Closed loop (feedback) model (4.13). We observe that in the closed-loop system the value of u_2 has no impact on the output when u_3 is small, whereas for the open-loop system the impact of u_2 remains similar as u_3 changes.

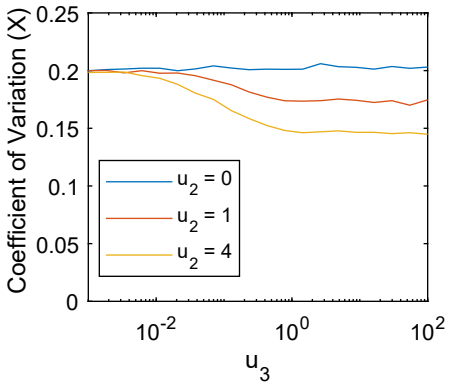


Figure 4.12: Noise simulations for the simple closed-loop (4.13) model of Circuit 2.

4.8.2 Detailed Model

We employ the structures outlined in Section 4.3 to create a model for the closed-loop sRNA feedback circuit (Fig. 4.10) that consists of four state variables: the concentration of mRNA (r) and protein (R) for the RhaS-GFP complex, the concentration of sRNA (s), and the concentration of the sRNA-mRNA complex (c). The underlying biochemical

reactions can be expressed as:

$$\begin{aligned}
\emptyset &\xrightarrow{\gamma_X{}^*} \mathrm{r} \xrightarrow{\delta_r} \emptyset, \\
\emptyset &\xrightarrow{\gamma_R{}^*} \mathrm{s} \xrightarrow{\delta_s} \emptyset, \\
\mathrm{t} + \mathrm{s} &\xrightarrow{\mathrm{K}_2} \mathrm{c} \xrightarrow{\delta_c} \emptyset, \\
\mathrm{r} &\xrightarrow{\beta_R} \mathrm{r} + \mathrm{R}, \\
\mathrm{R} &\xrightarrow{\delta_R} \emptyset,
\end{aligned}$$

These reactions can be described using four coupled differential equations of the form:

$$\begin{aligned}
\dot{r} &= \gamma_X^* - \delta_r r - K_2 r s \\
\dot{s} &= \gamma_R^* - \delta_s s - K_2 r s \\
\dot{c} &= K_2 r s - \delta_c c \\
\dot{R} &= \beta_R r - \delta_R R
\end{aligned} \tag{4.15}$$

4.8.3 Parameter Fitting

We begin by fitting parameters to our circuit's behaviour in open-loop when the sRNA is not induced (no L-Rhamnose is present), which allows (4.15) to be simplified by setting c and s to zero. The two free parameters are β_R and K_{u_3}, though we could have equivalently taken a value for β_R from the literature and fit α_x^*. The least-squares model fit for these parameters is illustrated in Fig. 4.13a, demonstrating that the promoter model in (4.4) is able to accurately capture the XylS induction system's open-loop behaviour.

Four parameters remain to be fit to experimental data, $K_{u_2}^*$, K_2, K_R, and α_r^*. There is a large amount of uncertainty when fitting the later three parameters: we find that values of K_R are generally large compared to the concentration of R (i.e. the model is far from saturation in rhaS concentration), and hence (4.2) has an approximately constant term of α_r^*/K_R. These two values can therefore vary together whilst maintaining a similar sRNA level, which is then constrained relative to r by K_2. To address this we fix α_r^* to the value of α_r found in Section 4.7, and least-squares fit $K_{u_2}^*$, K_2, and K_R to the experimental data. The results are presented in Fig. 4.13b, demonstrating good agreement between

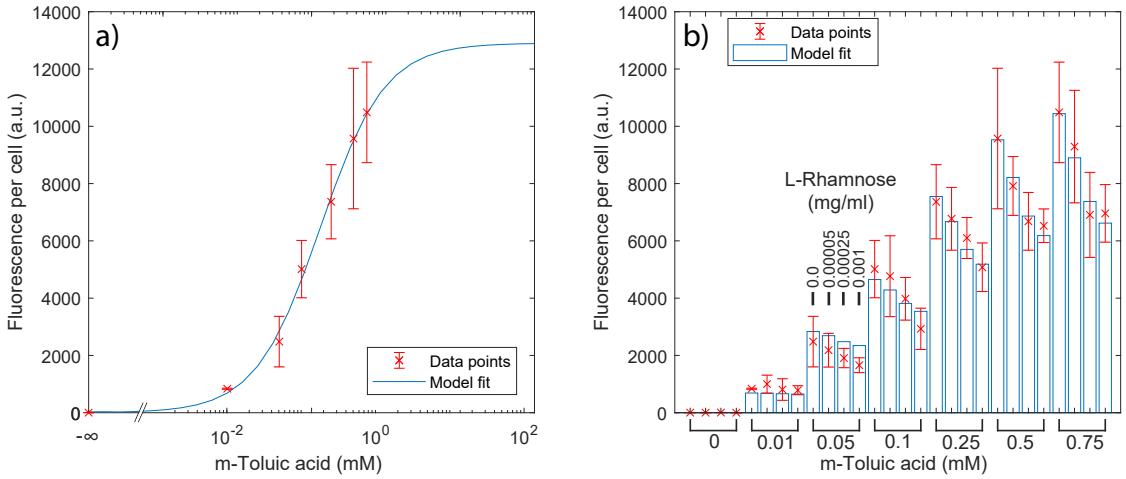


Figure 4.13: Circuit 2 model fitting. **a)** Fitting of the open-loop system (L-Rhamnose = 0) is used to estimate the parameters β_R and K_{u_3} in (4.4). We observe good agreement across the inducer range considered. **b)** Model fit for all combinations of inducer inputs. In this case we least-squares fit the values of K_2, $K^*_{u_2}$ and K_R, and set $\alpha^*_r = \alpha_r$. Resulting parameter values are summarised in Table 4.1.

our model and the experimental data.

4.8.4 Dynamic Behaviour

To assess the dynamic behaviour of the closed-loop sRNA feedback circuit we repeat a similar dynamic analysis to Section 4.7, again employing (4.10) to vary inducer concentrations over time. Results for these simulations are presented in Fig. 4.14. Similar to the tuned Autorepressor circuit we observe that it is possible to reduce the system output's sensitivity to one input, whilst maintaining a given mean output, by using the other input as a tuning dial. We observe that this circuit is much less sensitive to equivalent fold-changes in u_2 compared to u_3. This arises due to the smaller dynamic range of induction by u_2, which in Fig. 4.13 is observed to typically only reduce system output by $\sim 30\%$ between zero and maximal induction.

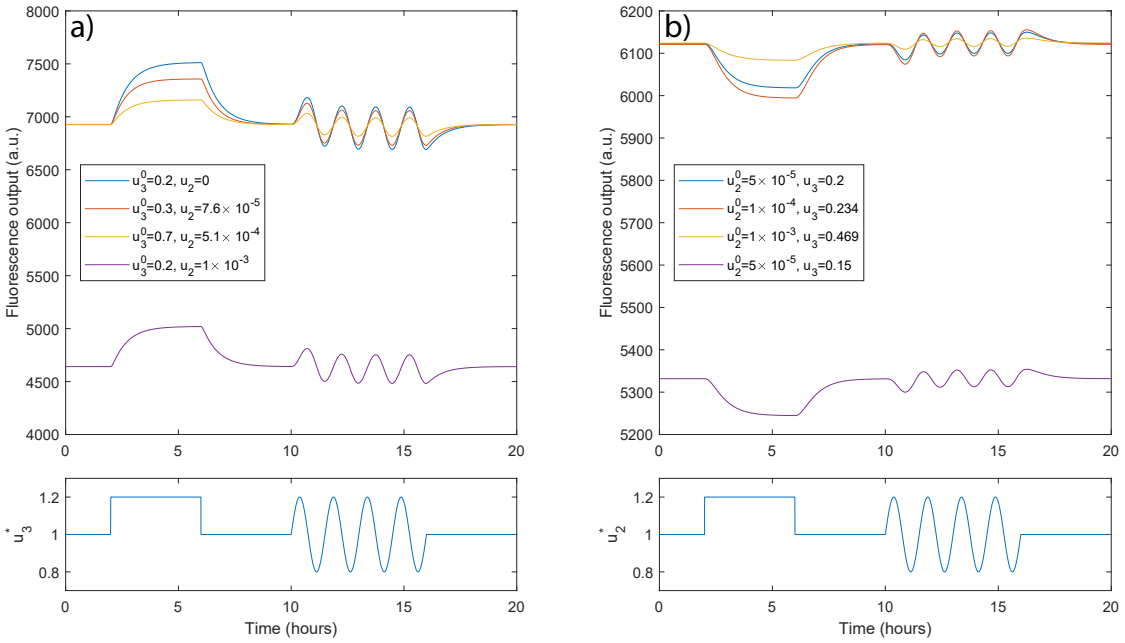


Figure 4.14: Closed-loop sRNA feedback circuit dynamic simulations. All inducer concentrations are in ng ml^{-1}. **a)** System output over time as u_3 (m-toluic acid) is varied according to (4.10) with the u_3^* profile illustrated. **b)** Similar to (a) but with u_2 (L-Rhamnose) varying over time. We observe a slower response to changes in u_2, and smaller output fluctuations for a given fold-change in this inducer.

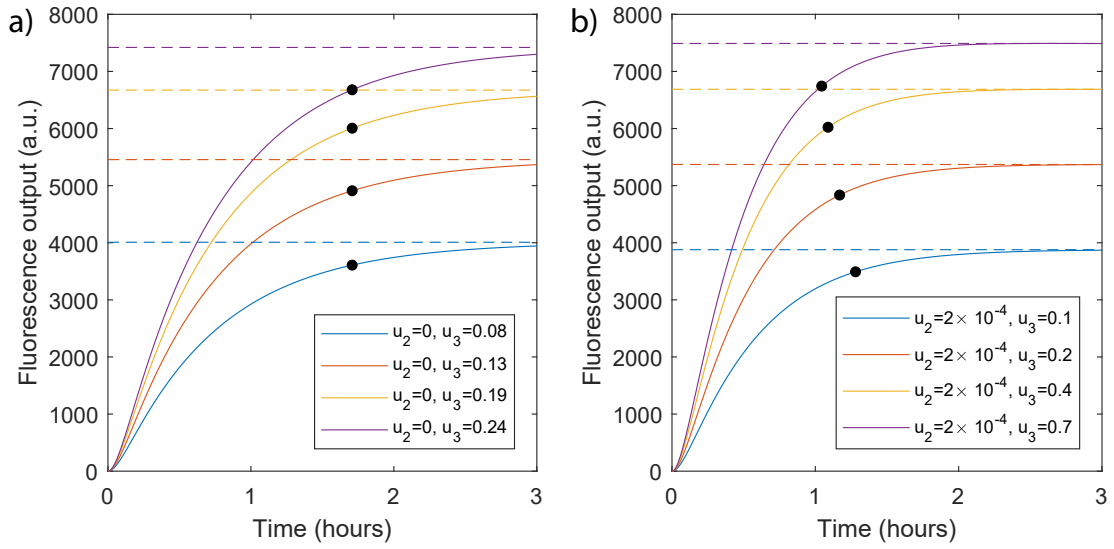


Figure 4.15: Step-response simulations of Circuit 2. **a)** Simulation of the closed-loop sRNA circuit without feedback ($u_2 = 0$). **b)** Simulation of the closed-loop sRNA circuit with feedback ($u_2 > 0$), demonstrating that feedback allows the system to achieve the same output levels with a shorter response time.

4.8.5 Noise Simulation

We perform simulations of the impact of both intrinsic and extrinsic noise on our system (4.15) similarly to as described in Section 4.5. In this case we set $\sigma = 1$, as a substantially greater magnitude of variability was measured for this circuit, and we again vary the translation rate and each inducer concentration:

$$\beta_R^* = \Delta_1 \times \beta_R,\ u_{2,3}^* = \Delta_2 \times u_{1,2} \tag{4.16}$$

Results are presented in Fig. 4.16. We find again that the contribution of intrinsic noise is low compared to that of extrinsic sources, and that the trends produced by extrinsic noise simulations qualitatively replicate those measured experimentally. Our model predicts zero output (and hence zero variation) when the concentration of *m*-toluic acid is zero, whereas the experimental data shows a non-zero CoV in this case likely due to cellular auto-fluorescence.

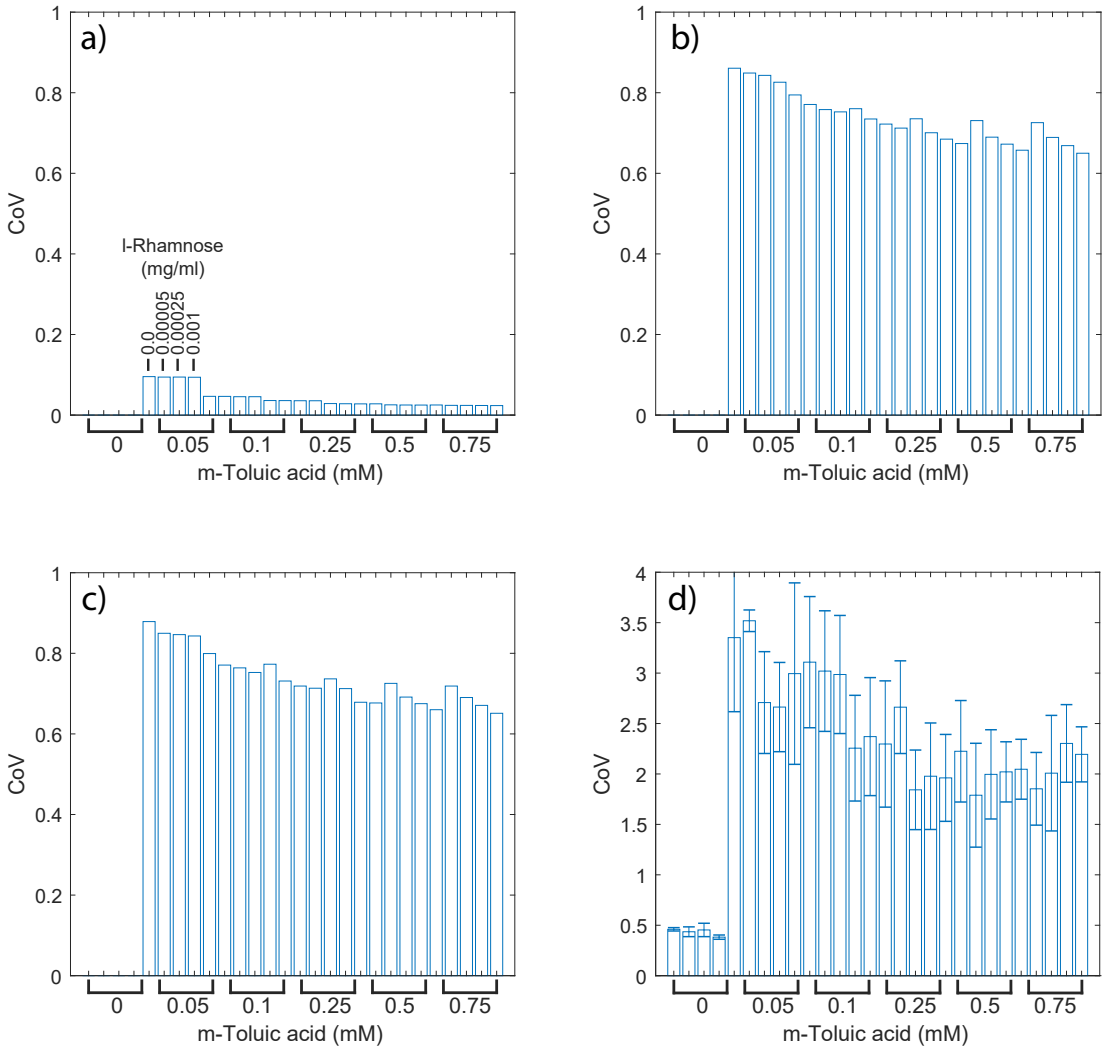


Figure 4.16: Closed-loop sRNA feedback circuit noise simulations. Plots of coefficient of variation (CoV) calculated as in (4.5) for different noise sources, and experimental data. **a)** Simulation of intrinsic noise, due to stochasticity of cellular processes. **b)** Simulation of extrinsic noise, with parameters sampled from log-normal distributions as described by (4.12) with $\sigma = 1.0$. **c)** Combination of stochastic and extrinsic simulations. **d)** Experimentally measured CoV.

4.9 Discussion

Our *in vivo* and *in silico* results demonstrated a range of benefits derived by tuning Circuit 1 (based on the classic TetR autorepressor) with an sRNA: by controlling the output using two inputs we can tune the circuit's sensitivity to each one, making it is easier to achieve intermediate output levels (which is usually difficult when tuning with aTc (u_1) alone due to the steep sigmoidal shape of the autorepressor's response). In addition, tuning via the second input allows us to largely de-couple response time from the steady-state output

level. This circuit can also be thought of as integrating multiple inputs to form a single output; such systems for signal combination have been a topic of significant research in synthetic biology, for example in the form of combinatorial promoters [196] or logic gates [197]. To demonstrate that our circuit could be used to combine a diverse range of signals we demonstrated its modularity by replacing the RhaS inducer with Xyls, finding that our model and parameters largely carried over between the two systems. Future investigations may further investigate this architecture's modularity by using different transcription factors (i.e. replacing TetR), though this is more difficult to do in practice (compared to replacing RhaS) as a new sRNA must be designed to target the changed mRNA sequence.

In Circuit 2 we demonstrated how an sRNA can be used as a tool to create novel feedback structures. Because this circuit uses an activating transcription factor (rather than the repressing TetR in Circuit 1, which must be expressed to block transcription from its promoter) it is able to achieve a near-zero output when its primary input (u_3) is zero. A system with architecture similar to Circuit 2 was analysed in another study (published after ours), which found it may be able to provide near-perfect adaptation [182]. In that case, simulations were performed assuming an *in vitro* implementation (i.e. in a cell-free media such as TX-TL), meaning that cell growth (and hence protein dilution/removal) is not present. This provides a significant advantage (in terms of accuracy as adaptation): theoretical studies [175] have demonstrated that dilution of control species provides a barrier to perfect adaptation (and indeed this was found in our theoretical studies of such systems in Chapter 3). Future experimental analysis of this circuit will include assessing its ability to reject disturbances by subjecting it to time-varying input signals (e.g. as in Fig. 4.14). This will likely require application of opto-genetic promoters [58], which allow steps both up and down (difficult with chemical inducers as they can't easily be removed) in the input, during growth in a Turbidostat device (Chapter 7).

4.10 Conclusion

Our results demonstrate the utility of sRNA as a tool for tuning feedback (Circuit 1) or implementing entirely novel feedback structures (Circuit 2). In the future it is likely that sRNA will become an increasingly commonplace component in synthetic biology as effective methods for the design of targeted sRNA are developed [18]. The modelling techniques we have developed will be valuable for application in these cases: as we demonstrated they can be used for both *a priori* prediction of circuit behaviours, as well as analysis and fitting of experimental results. This not only allows simulation of difficult-to-perform experiments (e.g. the dynamic simulations performed in this chapter), but it also provides a valuable conceptual framework for circuit analysis (e.g. the analysis of the leaky/non-leaky autorepressor variants in Circuit 1). Indeed, in subsequent chapters of this thesis I will apply these sRNA models in the design and analysis of other control systems.

5

DNA Recombinase Controller Implementations

The work in this chapter uses DNA recombinase enzymes (reviewed in Section 2.1.3) to implement synthetic biological controllers and feedback loops. I begin by modelling a simple feedback system which flips a region of DNA in order to attenuate cell-to-cell variability in transcription rate (as arises due to variable plasmid copy numbers). This system is implemented and characterised experimentally as described briefly in Section 5.2.1. I then analyse a more complex architecture, which employs both recombinase enzymes and mRNA-mRNA interactions (similar to those in Chapter 4), allowing it to better attenuate disturbances in transcription/translation. Many of the results from this chapter were published in [63] and [64].

5.1 Introduction

Integrase proteins, part of a broader enzyme family referred to as Recombinases, are used by viruses to integrate their genetic code with that of an invaded cell. By adjusting the orientation of their DNA recognition sites they can be re-engineered to flip (rather than excise) a sequence of DNA flanked by these sites [198]. The recognition sites are (predictably) changed by this process, and the reverse flipping operation (to regain the DNA segment's original orientation) can be catalysed by an integrase if a secondary protein (an excisionase) is present. This process is diagrammatically presented in Fig. 2.2b. The presence (and concentration) of both integrase and excisionase can be varied in order to control the rates of forward/reverse flipping. The ease with which this process can be engineered has led to the widespread application of recombinase systems in synthetic

biology. For example, they can be used to create biological "state machines" which permanently record the order in which inputs (i.e. inducer chemicals) are sensed [71, 199], and a broad array of logic gates [73, 200]. In this chapter I describe a new application of recombinase systems; the creation of dynamic feedback circuits that attenuate various sources of variability present in synthetic biological systems (reviewed in Section 2.1.4).

The first system we investigate (**Recombinase Circuit 1**, Fig. 5.1a) implements a simple variant of recombinase feedback: an integrase protein is expressed at a constant rate (under the control of the inducible Tet promoter), whilst excisionase (and the circuit output, GFP) are expressed from an invertible region of DNA. This region is initially oriented antisense to its promoter, meaning no GFP/excisionase is expressed. Integrase flips this DNA region, leading to excisionase expression which initiates the reverse flipping operation. The circuit thus eventually reaches a dynamic equilibrium with its forward/backward flipping rates balanced. Ideally this architecture reduces variability in the transcription rate of the output protein GFP: in cells with large transcription rates (i.e. due to a high plasmid copy number of the circuit), the expression of excisionase will be high and hence fewer DNA registers will be in their active ("ON") state. Though this circuit attenuates the effect of variable translation it does not, however, approximate "perfect adaptation" (i.e. as analysed in Chapter 3).

To realise an integrase/excisionase based circuit that gets closer to perfect adaptation we propose the architecture outlined in Fig. 5.1b (**Recombinase Circuit 2**, Fig. 5.1b), which additionally employs an mRNA-mRNA annihilation reaction to approximate a subtraction junction [62]. In this case we also include a new species D which is responsible for "measuring" cell-wide variability in protein translation rate: when the concentration of D changes (due to translational disturbance) a proportional change is felt in the transcription rate of the integrase protein. As demonstrated in Section 5.3 this architecture is able to adapt to changes in both transcription and translation rate of its components, and (thanks to the implementation using recombinase proteins) can do this whilst minimally burdening its host cell (i.e. consuming minimal resources).

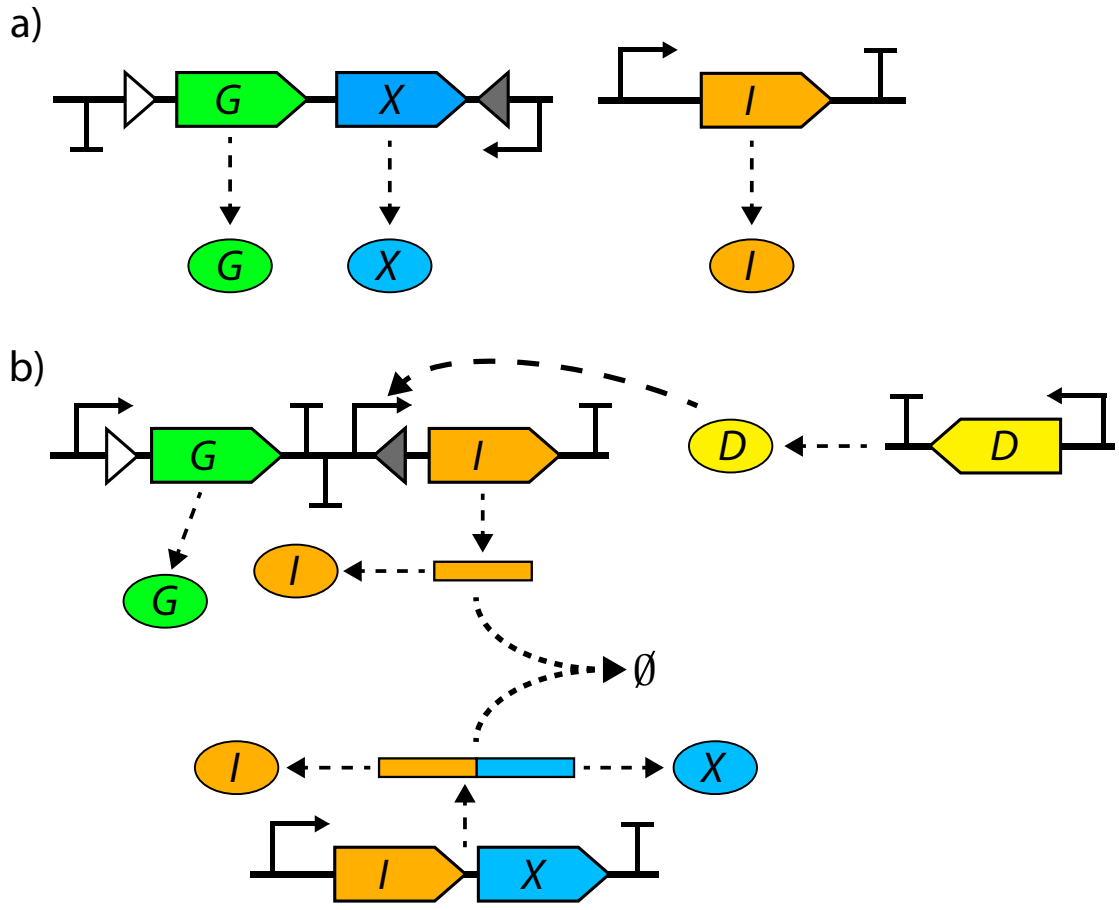


Figure 5.1: Recombinase-based feedback controllers considered in this chapter. a) Recombinase Circuit 1, which expresses integrase (I) continuously. A protein output (G) and excisionase (X) are expressed in an invertible DNA register, which is flipped back and forth to regulate the production rate of G. **b)** Recombinase Circuit 2, which additionally employs interacting RNA molecules (similar to those in Chapter 4), and a measurement species D. This circuit is designed with a secondary goal of reducing the burden it places upon its host cell by minimising expression rates of I and X when near to equilibrium.

5.2 Recombinase Circuit 1

Past studies have demonstrated the utility of flippable DNA registers in synthetic biological circuits with a range of purposes, such as event recording or decision making [71, 200]. However, when these systems are implemented they can be susceptible to noise arising from the variable number of DNA copies per cell (due to the stochasticity of plasmid replication and their uneven distribution during cell division). In this section we design and implement a simple feedback circuit that can partially attenuate this source of variability. To do this we build upon a previously reported DNA register [73] which

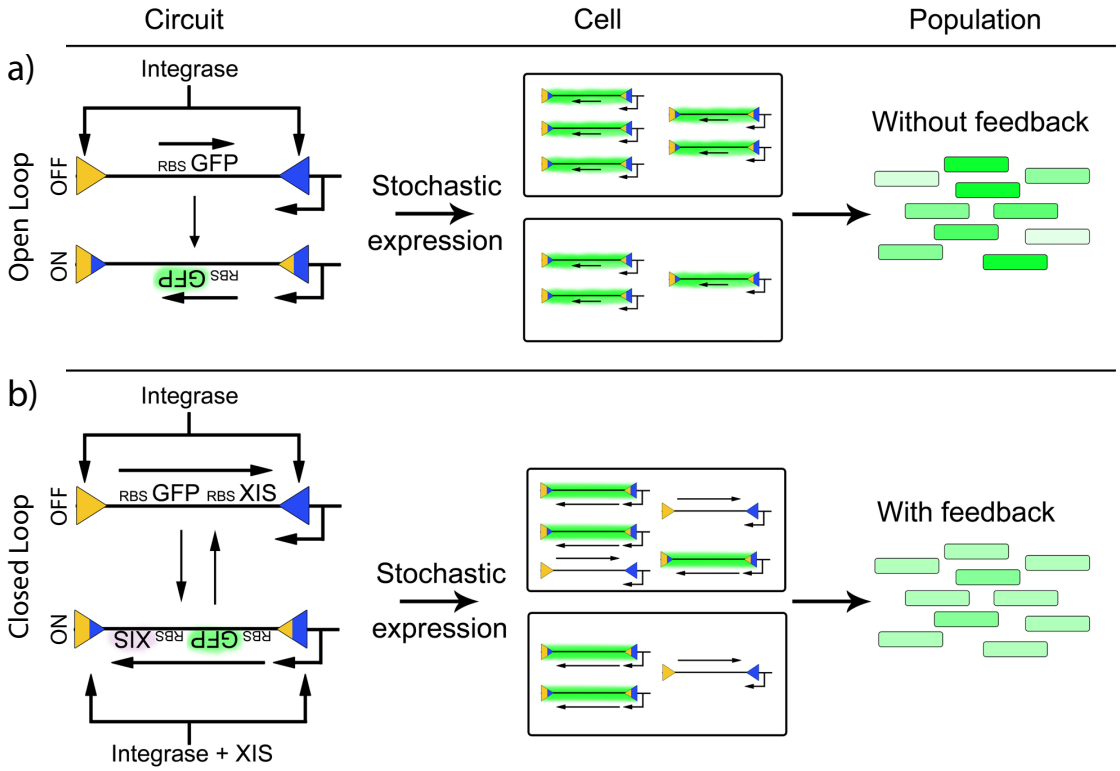


Figure 5.2: Schematics of the **a)** open-loop and **b)** closed-loop expression systems. The correlated expression of excisionase and GFP closes the loop around the integrase's switching process, allowing stochastic switching between ON and OFF states, thereby implementing negative feedback. Thus, while all DNA registers in the open-loop system eventually switch to the ON position, those in the closed-loop system are regulated by excisionase concentration leading to smaller variability in the number of ON registers per cell: if GFP concentration is high (low), excisionase concentration will be likewise, resulting in more OFF (ON) registers and down (up) regulation of GFP expression. Variability introduced by fluctuations in plasmid copy number, as illustrated in the central panel of this figure, is one of many noise sources for which negative feedback can compensate. The yellow and blue triangles represent attP and attB sequences which are recognised by the integrase, while half-coloured triangles represent attL and attR sequences that require the presence of excisionase for their recognition.

expresses GFP in response to DNA inversion. This circuit is referred to as the "open-loop" system (Fig. 5.2a), as there is no excisionase present to facilitate the backward flipping operation. Prior to the induction of the integrase protein (via the chemical inducer aTc) the circuit is in its OFF state (Fig. 5.2), with the GFP gene oriented antisense to its promoter. However, when integrase is induced it can flip this DNA segment, aligning GFP with its transcriptional promoter and leading to measurable fluorescence.

To add feedback we introduce an excisionase protein to the flippable DNA register

such that it is expressed in tandem with GFP, thus creating the "closed-loop" system (Fig. 5.2b). The excisionase closes the loop around the high gain integrase system by enabling registers in an ON state to stochastically and transiently switch to an OFF state. Over a sufficiently long time period the distribution of registers in either state stabilizes (illustrated in the middle column of Fig. 5.2), which due to the coupling of excisionase and GFP production results in a reduced variation in GFP expression between cells in the closed-loop system (both in absolute terms, and when normalized by mean expression levels) when compared to an open-loop system.

5.2.1 Experimental Summary

The laboratory procedures used to biologically construct and characterise the circuits described in this chapter were primarily performed by Thomas Folliard, so are only briefly summarised here. The backbone of the first plasmid (containing the GFP/excisionase genes on the left of Fig. 5.1a) was acquired from a past study [73] and modified (an excisionase gene with fast-degradation tag synthesised by IDT was added) using Gibson assembly. This plasmid has a low copy backbone ($\sim$ 15 per cell). The integrase enzyme is expressed under the control of the Tet promoter (i.e. as used in Section 4.6) on a second, higher copy plasmid ($\sim$ 60 per cell), which was co-transformed with the first plasmid for each experiment. All experiments were performed using *E. coli* strain DH5αZ1. Transformed cells were grown in standard LB media overnight, and then diluted into fresh media for characterisation experiments. End-point experiments were left to grow for 24 hours in 1.5ml opaque deep well plates at 37°C with rapid shaking, after which samples were taken and measured at the single-cell level in a flow cytometer. Time-course experiments were run for 12 hours with cells growing in 200 μL wells in a platereader at 37°C and with rapid shaking. The plate reader was programmed to take measurements of optical density (OD) and GFP fluorescence every 20 minutes over the course of the experiment. Each characterisation experiment was performed in triplicate (three biological replicates). For an in-depth description of the experimental procedures see the methods published in [63].

5.2.2 Modelling Circuit 1

To investigate the performance of our integrase/excisionase-based implementation of negative feedback we developed a mathematical model that combines both stochastic processes and deterministic differential equation models, for which a block diagram is presented in Figure 5.3. In our model we consider cell-to-cell variability in protein production arising from both intrinsic and extrinsic noise sources (for which biological sources and modelling techniques are discussed Sections 2.1.4 and 2.4.5 respectively). In particular, we allow certain system parameters to vary between cells (thereby introducing extrinsic noise), and treat the recombinase-mediated DNA flipping operation as a stochastic process. By utilising an approach that accounts for variability introduced in a range of cellular processes our model can replicate the impact of major noise sources (both gene-specific and cell-wide [93]) which cause cell-to-cell variability as well as inherited differences between triplicate populations.

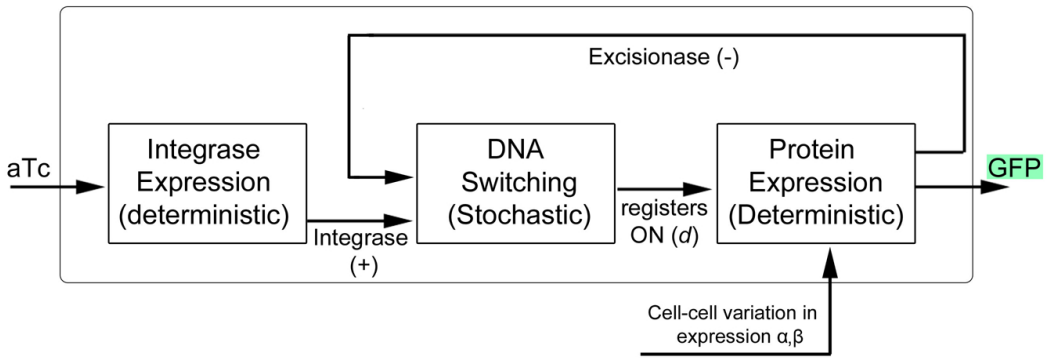


Figure 5.3: Block diagram for the modelled feedback system. The inducer (aTc) acts as an input to the deterministic expression of integrase via its interaction with the regulatory protein TetR. Integrase (and excisionase) concentration determines the transition probabilities of the random walk of the integer-valued state d representing the number of ON registers. Excisionase acts as negative feedback, because it biases the random walk toward transitions $p \to p-1$. The state d then acts as a random input to the deterministic protein expression dynamics, together with randomly selected constant scaling parameters α and β to represent variability between populations and cells.

Deterministic ODE Model

Considering our system at the single-cell level, we assume that each cell contains P copies of the plasmid containing the DNA register depicted in Figure 5.2, where the integrase

recognition sites are separated by either a reporter coding region (i.e. open-loop), or by both reporter and excisionase coding regions (i.e. closed-loop). Integrase is expressed on a different plasmid of copy number N, under the control of the Tetracycline transcription factor (TetR) which is constitutively expressed in the cell. We model the concentration of active (repressing) TetR ($[TetR]$) by assuming that (due to constitutive expression) there is a constant total TetR concentration per cell ($[TetR]_0$) that is reversibly converted to its inactive (non-repressing) form at a rate proportional to aTc concentration $[aTc]$. This is described by a deterministic differential equation of the form:

$$\frac{\mathrm{d}[TetR]}{\mathrm{d}t} = k_1(\alpha\beta_1[TetR]_0 - [TetR]) - k_{-1}[TetR][aTc] \tag{5.1}$$

where k_1 and k_{-1} are the forward and reverse reaction rates respectively. The combined parameter $\alpha\beta_1$ is a cell-specific scaling factor for expression rates which models the cell-to-cell uncertainty in protein expression (in this case the equilibrium TetR concentration per cell, $[TetR]_0$). In particular, α is used to simulate inherited differences in behaviour between populations grown from single cells, and β_1 introduces further cell-to-cell variation within each population. Thus, α accounts for some inherited noise sources (such as long-term inherited differences in proteome) as well as culture-wide variations (such as environmental differences between biological triplicates), whilst β_1 account for cell-to-cell variability (such as variations in plasmid copy number, which can strongly impact negative feedback systems [103]) and small scale extrinsic noise (such as local environmental inhomogeneities). We use these two parameters to simulate triplicates by splitting cells into three populations, each with a different value for α, and within each population values for β_1 are sampled from a probability distribution (as described in Section 5.2.3) for individual cells.

We model the dynamics of integrase concentration ($[I]$) as the net balance of its expression under the regulation of TetR with degradation and dilution. This results in the deterministic differential equation:

$$\frac{\mathrm{d}[I]}{\mathrm{d}t} = N\mu_I\alpha\beta_2\frac{K_I}{K_I + [TetR]} - \delta_I[I] \tag{5.2}$$

where μ_I and δ_I are parameters representing the maximal integrase expression per plasmid, and the degradation/dilution rates of integrase molecules respectively. The parameter K_I is the dissociation constant for TetR repressing integrase expression, and in this case $\alpha\beta_2$ models the cell-to-cell uncertainty in integrase expression around the nominal rate μ_I.

Similarly to integrase expression above, when p of the P plasmids have flipped to the ON state, we have expression of GFP and (in the closed-loop case) excisionase, and also degradation of each. We thus model the concentrations of GFP ($[GFP]$) and excisionase ($[X]$) using the differential equations

$$\begin{aligned} \frac{\mathrm{d}[X]}{\mathrm{d}t} &= p\alpha\beta_3\mu_X - \delta_X[X] \\ \frac{\mathrm{d}[GFP]}{\mathrm{d}t} &= p\alpha\beta_3\mu_G - \delta_G[GFP] \end{aligned} \tag{5.3}$$

where the scaling p of the expression rates μ_X and μ_G reflects an assumption that the expression of excisionase and GFP is linearly proportional to the number of ON registers. In this case the combined parameter $\alpha\beta_3$ models the cell-to-cell uncertainty in protein expression around the nominal rates μ_X and μ_G, which is set equal for $[X]$ and $[GFP]$ since they are expressed from the same operon on the same plasmid. Note that setting $\mu_X = 0$ models the open-loop behaviour, while $\mu_X > 0$ models closed-loop behaviour.

Stochastic DNA Flipping

In addition to the deterministic models described above we model the flipping of the P DNA registers as a stochastic process. The variable $p(t)$ representing the number of ON registers at time t takes a random walk on the set of integers $p \in \{0, 1, \ldots, P\}$. Given $p(t)$, the rate at which p jumps to $p+1$ (i.e. flipping from the OFF state to ON) is modelled as

$$R_{on}(I) = r_{on}(P - p)f_I(I) \tag{5.4a}$$

$$f_I(I) = \frac{I^4}{K_I^4 + I^4} \tag{5.4b}$$

This rate is constructed as a scalar parameter r_{on}, representing a switching rate, multiplying the expected number of OFF registers that have two integrase dimers bound,

thereby being in the configuration able to stochastically switch to the ON state. In particular, the integrase concentration-dependent fraction in this rate is a Hill function with parameter 4, representing the fact that a total of four integrase molecules need to bind to the DNA to be in the correct configuration to switch the register to the ON position. The parameter K_{on} is a combined parameter that is a function of the dissociation constant of integrase dimerization and the dissociation constant of integrase dimer binding to each DNA recognition site. More complex models of DNA recombination exist [201, 202] and may be used to expand our model in the future. However, at present we have utilised a simple Hill function to model recombination due to the small number of free parameters it introduces, and to reduce computational complexity.

In the closed-loop circuit the reverse flipping operation (from ON to OFF) is also possible: When some of the P registers have flipped ON, both GFP and excisionase are expressed, allowing some of the ON registers to flip back to OFF, thus slowing expression of GFP and excisionase. The rate at which p jumps to $p-1$ is modelled as

$$R_{off}(I,X) = r_{off} p f_X(X) f_I(I) \tag{5.5a}$$

$$f_X(X) = \frac{X^4}{K_X^4 + X^4} \tag{5.5b}$$

Similarly to (5.4), this rate is constructed as a scalar parameter r_{off}, representing a switching rate, multiplying the expected number of ON registers that have two integrase dimer/excisionase dimer complexes bound, thereby being in the configuration able to stochastically switch to the OFF state. In this equation both the integrase concentration-dependent fraction and the excisionase concentration-dependent fraction are Hill functions with parameter 4, since four integrase and four excisionase molecules are required for an ON register to flip to OFF. Similarly to the previous case, the parameter K_{off} represents the dissociation constant of excisionase with integrase.

5.2.3 Simulation and Parameter Values

Representative model parameters (presented in the Supplementary Material of [63]) were chosen in line with the assumptions made during the design of our system in order to define the relative rates of different processes. For example, excisionase degradation was made fast due to the presence of an ssrA tag, integrase/excisionase flipping rates were defined such that this process would occur over a multiple-hour time-scale, and plasmid copy numbers were defined based on known properties of the plasmid backbones used. Other free parameters in the model were subsequently tuned so that the model's behaviour would qualitatively capture that exhibited in the experimental data.

For the open- and closed-loop cases ($\mu_X > 0$ and $\mu_X = 0$ respectively) with input levels 0, 5, 15, 25, 50, 75, and 100 ng/mL, we simulated $5,000$ single cells using the mathematical model described above. This was done with values of $\alpha\beta_i$ sampled evenly (in probability) from the cumulative distribution function of a log-normal distribution, $\ln\mathcal{N}(\mu - \sigma_1^2/2, \sigma_1^2)$ with $\mu = 1$ and $\sigma_1 = 0.5$, which gave us a distribution of of cell behaviours from which we could artificially simulate triplicates. At every half-hour during the simulation the following process was repeated 2000 times to form and analyse an equivalent to 2000 sets of biological triplicates (triplicates are indexed by j): we sampled values $\alpha_j \sim \ln\mathcal{N}(-\sigma_2^2/2, \sigma_2^2)$ for $j = 1, 2, 3$ with $\sigma_2 = 0.25$, and then randomly selected 2000 cells with $\alpha_j\beta_i \sim \ln\mathcal{N}(ln(\alpha_j) - \sigma_2^2/2, \sigma_2^2)$ from the $5,000$ single cells simulations to form each of the three triplicate populations. Scaling factors for different processes $(\beta_1, \beta_2, \beta_3)$ are thus sampled independently from a log-normal distribution with mean α_j for each triplicate. This process of sampling cell populations from a constant pool of simulated cells with a pre-defined range of scaling factors was pursued instead of simulating unique populations to form triplicates as it allows us to compute reliable statistics much more efficiently.

For each simulated triplicate the mean fluorescence of each of the three populations was calculated, as was the standard deviation of these means. After the 2000 repetitions of this process the mean of the individual triplicate means was calculated, as was the

mean of the triplicate standard deviations, which is the data presented in (for example) Figure 5.6. To give the end-point data (as in Figure 5.4) a simulation of 5,000 cells was modelled for 24 hours, after which the process described above was performed to generate mean triplicate statistics.

5.2.4 Results

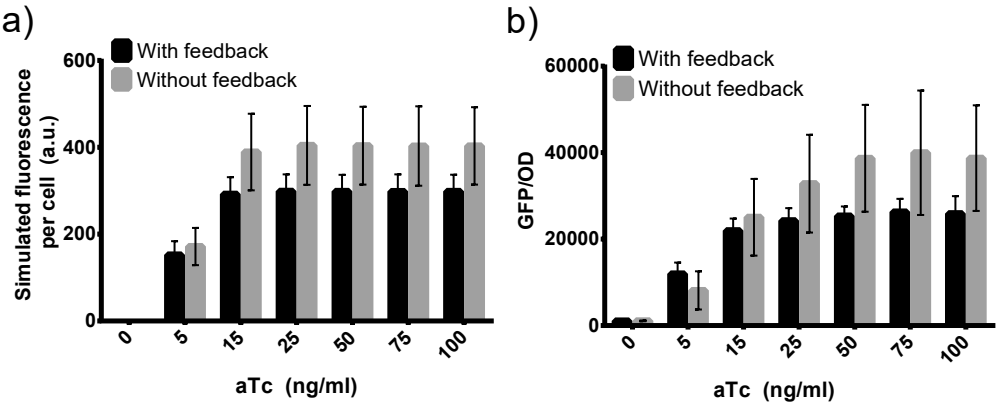


Figure 5.4: End-point responses for the open- and closed-loop systems over a range of aTc inputs. a) The simulated system run for 24 hours to generate triplicate statistics as described in Section 5.2.3. Data points here represent the mean end-point GFP concentration per cell (in arbitrary units) for 2000 sets of triplicates (each consisting of three populations of 2000 cells), and error bars correspond to the mean standard deviation in fluorescence within each triplicate set. **b)** Averaged end-point (taken to be 24 hours) GFP expression from triplicate biological repeats, where error bars indicate the standard deviation in fluorescence of the triplicates. The y-axis is GFP expression (measured via culture fluorescence) divided by optical density (OD), in arbitrary units. For both experimental and simulated data the feedback system demonstrates a reduction in steady state fluorescence, as well as a reduced variation (when normalized by mean expression level) in expression between triplicates.

Using our mathematical model we simulated the end-point response of our system (fluorescence output after 24 hours) to a range of aTc concentrations (Fig. 5.4a). We observe that the steady-state GFP expression per cell is uniformly decreased in the closed-loop system when compared to the open-loop variant. This trend, an anticipated consequence of negative feedback, is borne out experimentally (Fig. 5.4b). In both cases we observe that the mean intra-population variability across each experiment (represented by the error bars in Figs. 5.4) is decreased significantly in the feedback circuit, which is the intended noise-reducing function of our circuit. For the experimental results this corresponds to a reduction in coefficient of variation of fluorescence (the standard deviation of the triplicates' mean fluorescence divided by the mean fluorescence of each

triplicate set) for each triplicate set (with aTc $\geq$ 15 ng/mL) by a factor of $\sim$2.8. Simulated and experimentally-measured distributions of single-cell fluorescence at the experiment's end-point for induction with 50 ng/ml aTc are presented in Fig. 5.5. This makes the significant differences between the closed- and open-loop systems immediately apparent: in the closed-loop system the distributions of fluorescence for each triplicate are narrower, and have more similar mean values.

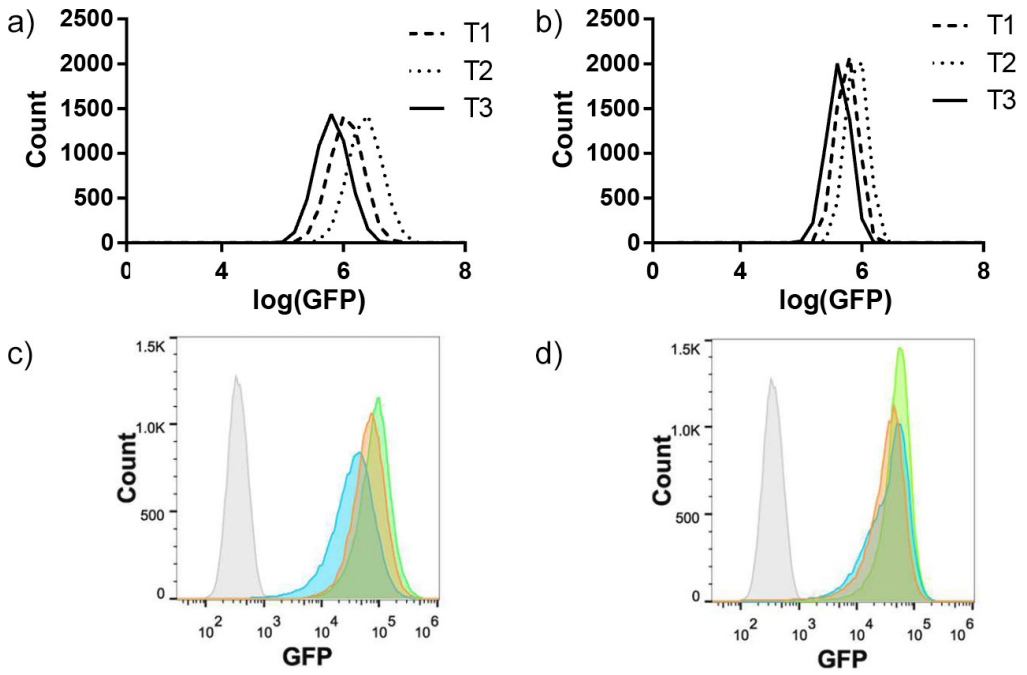


Figure 5.5: Simulated and experimentally-measured distributions of fluorescence. Simulated distributions of fluorescence for triplicates of the **a)** open- and **b)** closed-loop systems. Single-cell flow cytometry data for triplicates of the **c)** open-loop and **d)** closed loop systems. In all cases an induction level of 50 ng/ml of aTc is used. The x-axis is the level of GFP expression (measured via fluorescence) per cell, in arbitrary units. In the experimental data the gray distribution is a negative E. coli control. We observe that feedback results in higher reproducibility between replicates, both in terms of the location of the histogram's peak, as well as its overall shape. The system with feedback also results in a narrower histogram, indicating a reduction in cell-to-cell variability.

To investigate the impact of negative feedback on the system's transient response we performed time-course simulations and experiments, for which data at 50 ng/ml of aTc induction is presented in Fig. 5.6. As before, in both simulated and experimental data the closed-loop system provides a reduction in expression variability: the standard deviation of mean triplicate fluorescence is reduced in the closed- loop system, even early

in the response when the mean expression levels do not significantly differ. We next use our model to investigate the influence of the DNA register's copy number (P) on the system's performance (Fig. 5.7). We observe that the noise-reducing property of our negative feedback system is maintained over a wide range of copy numbers, except for the case in which the plasmid containing excisionase and GFP has very low copy (as would occur with genome integration). In this circumstance excisionase expression levels are too low to provide an adequate feedback signal, though this can be overcome (Fig. 5.7e) by increasing the rate of excisionase expression (for example, by using a stronger RBS).

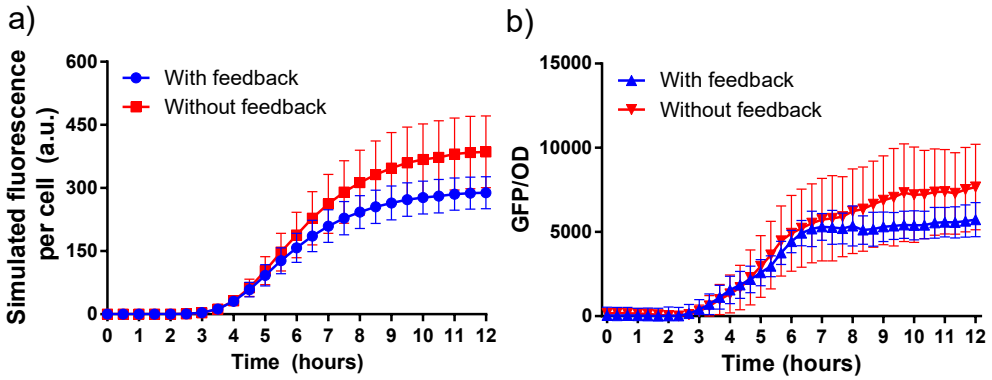


Figure 5.6: Time-course fluorescence data for the open- and closed-loop systems. All experiments are induced with 50 ng/mL aTc and grown for 12 hours. **a)** Simulated average expression level and mean standard deviation of expression for 2000 triplicate sets. **b)** Experimental results, which represent the average GFP expression for biological triplicates, and error bars denote the standard deviation of the individual triplicate means. The y-axis is GFP expression (measured as bulk culture fluorescence) divided by OD, in arbitrary units. Equivalent plots for the full range of inducer concentrations investigated in Fig. 5.4 are presented in [63]. In both the experimental and simulated data the feedback system yields reduced variation in output fluorescence levels.

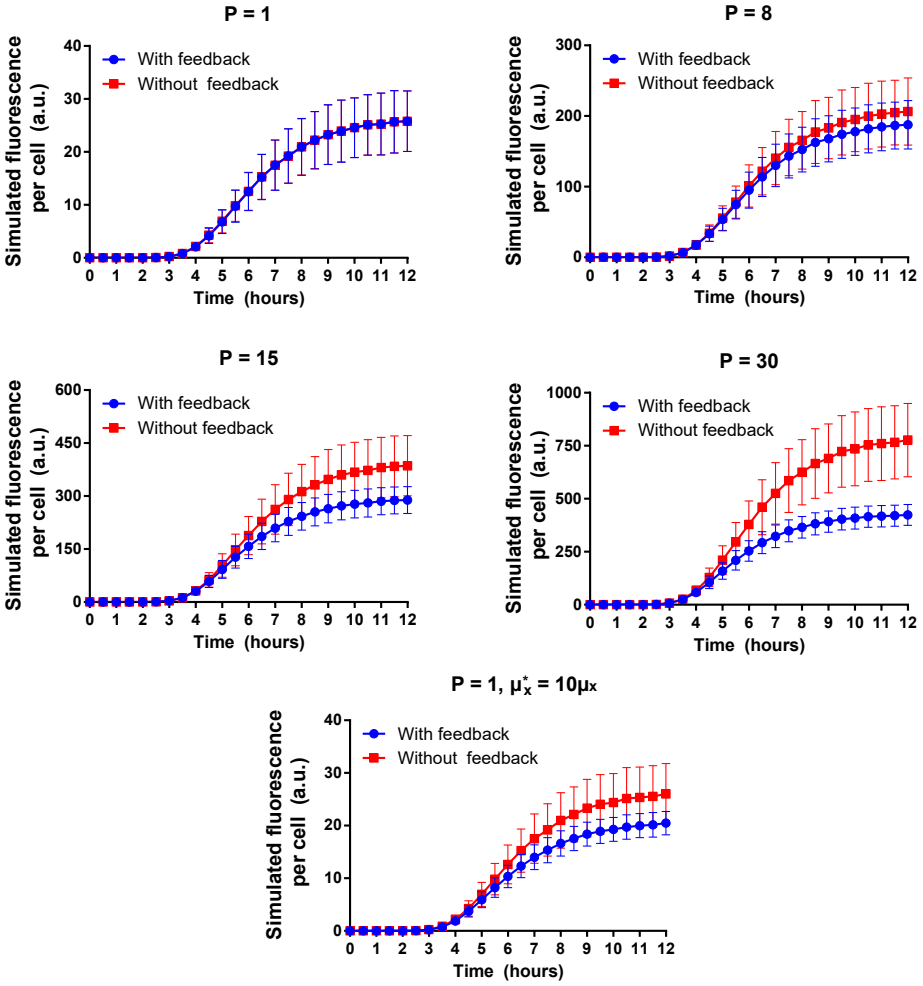


Figure 5.7: Influence of plasmid copy number on feedback circuit behaviour. Simulated time-course fluorescence data for a range of plasmid copy numbers (P) over 12 hours with 50 ng/ml aTc induction. We observe that as P increases the fluorescence output of the feedback case reduces compared to the non-feedback case, which arises due to the higher expression of excisionase resulting in a smaller proportion of plasmids being in the ON position. This demonstrates our negative feedback system's ability to compensate for fluctuating plasmid copy numbers, which is not possible in the non-feedback system (where end-point output is proportional to D). For very low values of P (i.e. $P = 1$, genome integration) our feedback system cannot function with the given parameter values as all plasmids are in the ON position since the low excisionase expression level provides insufficient feedback gain. This can be overcome by adjusting the expression level of excisionase (or reducing its degradation rate), which has been done in the lower panel (where μ_X^* is set to ten times its previous value). This increases the feedback gain of our system, shifting its functional range to be more effective at lower plasmid copy numbers.

5.2.5 Discussion

Our results demonstrate a good qualitative correspondence between the simulated and experimentally implemented systems, in both cases highlighting the useful properties of negative feedback architectures. There is a clear reduction in the mean steady-state

output of the closed-loop system, as a smaller fraction of DNA registers are in their ON position at any given time. A useful feature of many feedback systems is their ability to reduce rise time (as investigated for sRNA based controllers in Chapter 4), which has previously been demonstrated for biological systems [29]. However, based upon analysis of the time-course data from our experiments (using the method of Rosenfeld *et al.* [29]) we observed no statistically significant difference in response time. To further study these circuits' dynamic response a Turbidostat experimental setup would be useful for future experiments, which we develop in Chapter 7.

In all experiments it was found that there was a significantly reduced level of variability in the output of the closed-loop system. Considering this in the context of our model, if the random scaling factor $\alpha\beta$ is large (small), which might correspond to the GFP/excisionase plasmid having higher (lower) copy number in that cell, then for each ON register there will be a large (small) expression of both GFP and excisionase, which will reduce (increase) the number of ON registers. So, for cells with large (small) $\alpha\beta$ the level of expression of the target protein per ON register will be high (low), but there will be fewer (more) ON registers. Thus, the closed-loop circuit partially attenuates the influence of variability in $\alpha\beta$ on its output (GFP concentration). However, it is not perfectly adapting. This can be seen by expressing the DNA flipping process described by (5.4) and (5.5) in continuous form as:

$$\dot{p} = R_o ff(I, X) - R_o n(I) \tag{5.6}$$

which we can then solve at steady state ($\dot{p} = 0$) along with (5.3) to show that p must satisfy:

$$p = \frac{f_X(\frac{p\alpha\beta_3\mu_X}{\delta_X})}{\frac{r_{on}}{r_{off}} + f_X(\frac{p\alpha\beta_3\mu_X}{\delta_X})} \tag{5.7}$$

Similarly, we have for our system output:

$$[GFP]_{ss} = p\frac{\alpha\beta_3\mu_G}{\delta_G} \tag{5.8}$$

If $[GFP]_{ss}$ is to remain constant (i.e. perfect adaptation) following a change in $\alpha\beta_3$ then p must adjust accordingly to keep the product $p\alpha\beta_3$ constant. However, from (5.7) we observe that any change in p will require variation in $p\alpha\beta_3$, thus preventing perfect

adaptation. In Section 5.3 we propose and analyse a system that partially addresses this by changing the form of (5.7).

The variability between populations demonstrated by flow cytometry measurements in Figs. 5.5a,b (and their similarity to simulated results in Figs. 5.5c,d) supports the approach taken to introduce random variations between cells in our model: we observe (particularly in the open-loop case) substantial variation between triplicates in terms of the centre of their GFP fluorescence distribution, which corresponds to varying values for the random parameter α, as well as a wide distribution of fluorescence values within each population, corresponding to variation in the parameter β. Both of these parameters reflect natural differences in behaviour that arise between cells: α accounts for differences (for example, in the proteome) that are inherited from the small number of cells used to initiate each triplicate, while β reflects the distribution of cell states within each triplicate during growth and experiment (for example, due to inhomogeneities in environmental conditions or plasmid copy number fluctuations). Further work could be done to refine the probability distributions from which α and β are sampled, including experimental study to determine appropriate correlations between the β_i's that were used to introduce noise into different equations.

Though existing transcription factor-based feedback controllers provide one approach for regulating gene expression noise [32], the synthetic feedback system we developed in this section of particular interest due to its utilisation of integrase enzymes: due to their widespread use in Synthetic Biology, libraries of integrase proteins with orthogonal binding domains exist [203, 204], meaning our design could be easily scaled to allow multiple integrase/excisionase feedback circuits to function within the same cell. Recombinase proteins have been demonstrated to work well in systems with a wide variety of output proteins [205–207]. There should therefore be no major limitation to replacing our system's output (currently GFP) with alternate target proteins, making this framework widely applicable (and biologically tuneable, as demonstrated in the experimental supplementary of [63]) to a broad range of engineering challenges.

5.3 Recombinase Circuit 2

I now investigate a more complex recombinase-based control architecture, Circuit 2 from Fig. 5.1, which (in addition to re-arranging the proteins in Circuit 1) additionally includes a mRNA-mRNA interaction. As we will see this substantially changes the form of (5.7), allowing this circuit to achieve near-perfect rejection of disturbances to both cell-wide translation rates (frequently arising due to ribosome sequestration [116, 120]) and transcription rate (arising due to fluctuations in plasmid copy-number [103]). Furthermore, Circuit 2 is designed to operate in a "slack region" where (at equilibrium) the expression rate of its constituent proteins is minimised. This contrasts with Circuit 1, which at equilibrium expresses large concentrations of both integrase and excisionase in order to keep flipping rates balanced. In avoiding this, Circuit 2 reduces the burden (consumption of cellular resources) that it imparts upon its host cell, which could otherwise adversely impact cellular growth and the behaviour of other synthetic circuits [24, 25, 208].

The system proposed achieves these goals by using two biological motifs studied in previous sections: first, integrase/excisionase proteins are again used to create a biological feedback loop, but with a different architecture to that in Section 5.2. Circuit 2 thus demonstrates how recombinase enzymes can side-step the impact of dilution effects (hence minimising burden) experienced by controllers for which internal states are stored as protein concentrations; DNA registers avoid this as their state is maintained during replication. Second, an mRNA-mRNA annihilation reaction (an extension of the mRNA-sRNA interactions described in Chapter 4) provides functionality similar to a "subtraction" junction from traditional control engineering [62]. The interconnection of these two biological motifs is inspired by switching control architectures (commonly employed in classical control engineering [209]), which move between discrete operational modes depending on a reference signal (an interpretation of these modes in our system is outlined in Fig. 5.9).

In Section 5.3.1 a method was outlined for implementing a reaction system which approximates mRNA-mRNA annihilation. This is followed in Section 5.3.2 by a detailed

description of the proposed control architecture, and in Section 5.3.3 we analytically derive limits on the range of disturbances that it can effectively reject. We then use nominal parameter values to examine the system's steady-state behaviour, demonstrating that near-perfect adaptation can be achieved. In Section 5.3.4 our system's dynamics are simulated, demonstrating its time-varying response to various disturbance inputs. Section 5.3.5 further investigates the system's dynamics by treating the constituent reactions as stochastic processes, and analyses the impact of the discretised DNA flipping process.

5.3.1 Approximating mRNA-mRNA Annihilation

One structural motif which employed in this circuit is an mRNA-mRNA annihilation reaction (Fig. 5.8a), in which two mRNA irreversibly bind to each other to prevent translation of either strand. Similar behaviours have been observed in some natural systems [210], though are at present beyond the capabilities of commonly used RNA engineering tools (e.g. [18]). To overcome this experimental challenge here we briefly describe an equivalent RNA architecture (Fig. 5.8b) which additionally employs two sRNAs (as have been widely used in synthetic biological constructs [42, 211]). In this case we co-express an sRNA with each mRNA, which targets the opposing mRNA. We can model this architecture with a system of differential equations of the form:

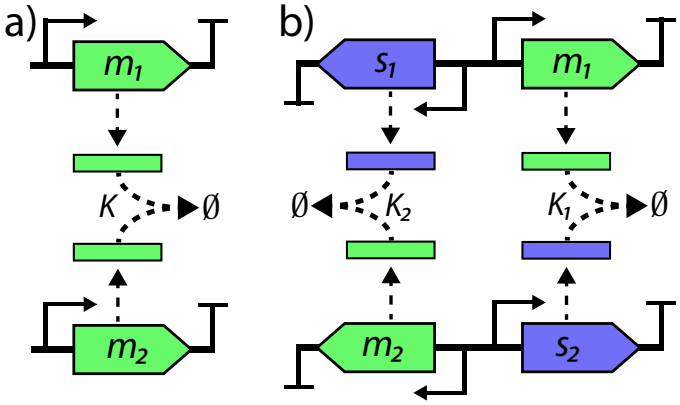


Figure 5.8: Circuit to approximate mRNA-mRNA Annihilation. a) An mRNA-mRNA annihilation interaction for which we propose **b)** an alternate architecture which employs sRNA to achieve equivalent functionality.

$$\dot{m_1} = \alpha_{1m} - \delta_{1m} m_1 - K_1 m_1 s_2 \tag{5.9a}$$

$$\dot{m_2} = \alpha_{2m} - \delta_{2m} m_2 - K_2 m_2 s_1 \tag{5.9b}$$

$$\dot{s_1} = \alpha_{1s} - \delta_{1s} s_1 - K_2 m_2 s_1 \tag{5.9c}$$

$$\dot{s_2} = \alpha_{2s} - \delta_{2s} s_2 - K_1 m_1 s_2 \tag{5.9d}$$

where α_i are constitutive expression rates, δ_i are degradation rates, and K_i are binding rates between each mRNA-sRNA pair. Examining (5.9) at steady state we find that m_1 must satisfy:

$$0 = \alpha_{1m} - \delta_{1m} m_1 - \frac{K_1 m_1 \alpha_{2s}}{\delta_{2s} + K_1 m_1} \tag{5.10}$$

Meanwhile, the equivalent mRNA-mRNA version of (5.9) would mirror (5.9a,b) with $K_1 = K_2$, and s_1 and s_2 replaced by m_1 and m_2 respectively. At steady state we would therefore have:

$$0 = \alpha_{1m} - \delta_{1m} m_1 - \frac{K_1 m_1 \alpha_{2m}}{\delta_{2m} + K_1 m_1} \tag{5.11}$$

which mirrors (5.10) up to the choice of α_{2m} and δ_{2m}. Indeed, if we set $\alpha_{im} = \alpha_{is}$ and $\delta_{im} = \delta_{is}$ (which is assumed in subsequent sections) then the two architectures in Fig. 5.8 have equivalent deterministic behaviour. That said, when variables are treated in a stochastic setting the architecture including sRNA will have differing noise properties due to the additional species involved.

5.3.2 System Model

The control architecture proposed in this section (Fig. 5.9 and Fig 5.10a) is comprised of a plasmid-encoded component (which includes a flippable DNA register and genes for the system output G and mRNA z), and a genome encoded component (which includes mRNA z^* and a gene for measurement species D pictured in Fig 5.10a). The system aims to regulate its output G to adapt to cell-wide disturbances Δ in translation rate (which could arise due to varying ribosome availability) and Θ in transcription rate of all plasmid encoded genes (which could arise due to varying plasmid copy number as in Section 5.2).

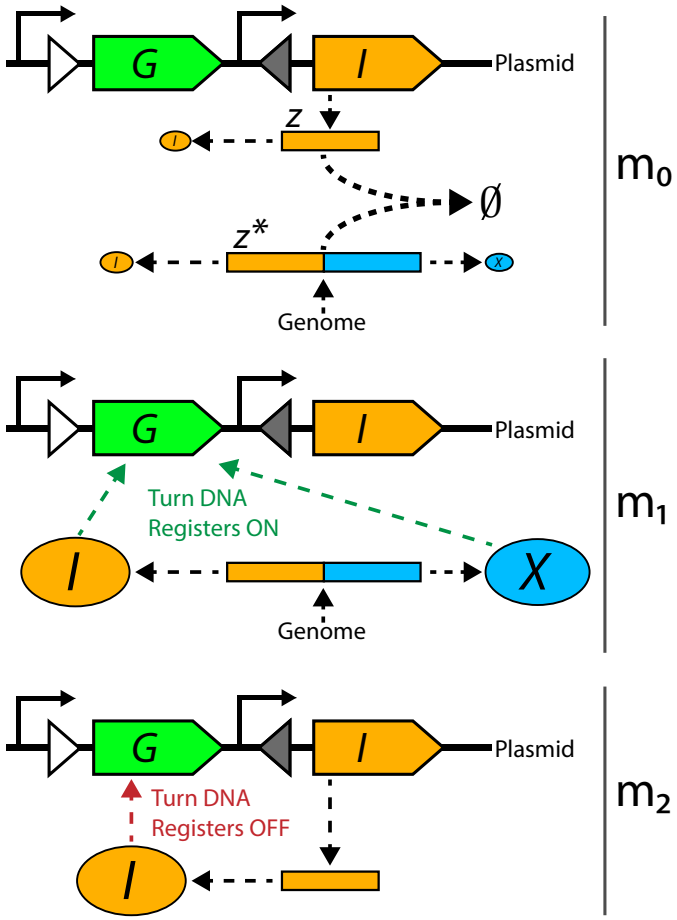


Figure 5.9: Illustration of switching control mechanism. Our proposed control architecture (full schematic presented in Fig. 5.10a) which employs an annihilation reaction between mRNA z and z^*, and a flippable DNA register (flanked by converging triangles, see Fig. 5.10b) containing the output gene (G) and a promoter. In mode m_0 production of mRNA species z and z^* is balanced, meaning most mRNA dimerises preventing translation, and thus DNA flipping is slow. When the system is disturbed (by a change in cell-wide translation rate or plasmid copy number) mRNA production is out of balance and the system enters either mode m_1 (to increase output G) or m_2 (to decrease G). In m_1 genome-expressed Integrase and Excisionase work together to flip DNA registers to their ON position, whilst in m_2 plasmid-expressed Integrase flips registers to their OFF position. Each mode is numerically illustrated in Fig. 5.10c. Measurement species D (see Fig. 5.10a) is omitted here for simplicity.

Rejection of disturbances in either of these variables is achieved by regulating the number of DNA registers in their ON position, which determines the transcription rate of G.

Control functionality is implemented using Integrase (I) and Excisionase (X) proteins that flip a DNA register flanked by recognition sites (Fig. 5.9), regulating the proportion of DNA registers that are in their ON position (p, which we initially treat as a continuous variable). Note that we have slightly re-defined p in this section (compared to Section

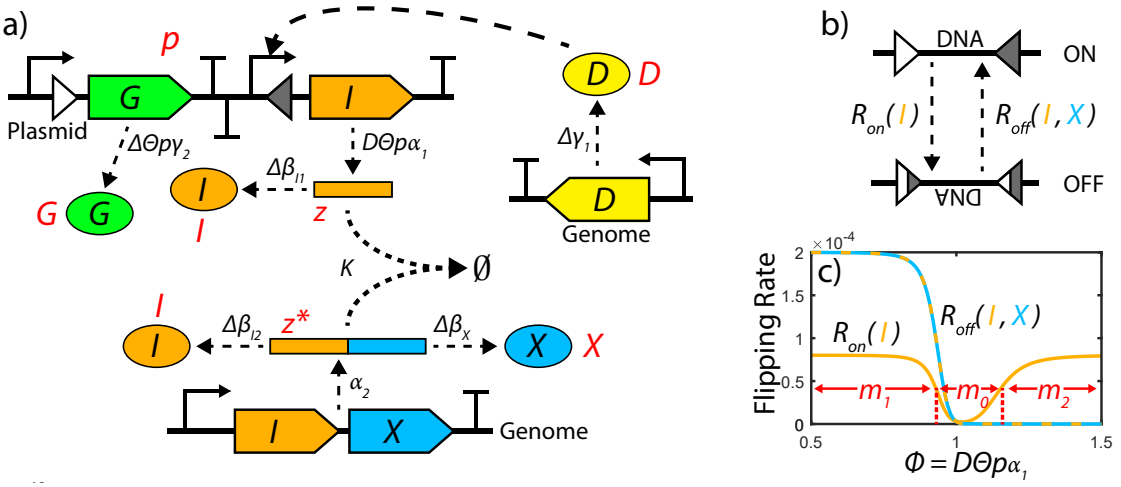


Figure 5.10: Circuit 2 control architecture. **a)** The Integrase/Excisionase control architecture analysed in this section. State variables from (5.12) are in red, other parameters are as described in the text. DNA is either plasmid- or genome-encoded as labelled. **b)** Illustration of Integrase/Excisionase mediated flipping of a DNA register between recognition sites (triangles) between its ON state (that in panel a) and OFF state (in which neither G nor I is transcribed). **c)** DNA flipping rates as a function of steady-state I and X values when $\Phi = D\Theta p\alpha_1$ is set to a fixed value, demonstrating a "slack region" (where $\Phi \approx 1$) near which both flipping rates are small. The controller's action approximates that of a switching controller, with modes m_0, m_1, and m_2.

5.2) to reflect the proportion (rather than absolute number) of DNA registers in their active form. From registers in their ON position the output G is transcribed, as is mRNA z (encoding I). Meanwhile, an mRNA z^* that encodes both I and Z is transcribed continuously from the genome, as is measurement species D. The mRNAs z and z^* irreversibly bind to each other in such a way that translation from each is supressed. This design can be conceptualised as an approximation of a switching controller which operates in three distinct modes, m_0, m_1, and m_2 (Fig. 5.9) defined as follows: when production rates of z and z^* are balanced little I or X is produced, meaning flipping is slow (the controller is in mode m_0) and p remains approximately constant. However, the concentration of z (and G) can be disturbed by a change in either transcription rate via Θ, or global translation rate via Δ (since the production of z is regulated by the measurement protein species D). When this imbalance in z and z^* arises either I (if $z \gg z^*$, mode m_2) or I and X (if $z \ll z^*$, mode m_1) will be produced, thereby flipping DNA registers (varying p) to return expression of z and G to its desired set-point.

The system can be described using a system of differential equations of the form:

$$\dot{D} = \Delta\gamma_1 - \delta_2 D \tag{5.12a}$$

$$\dot{z} = D\Theta p\alpha_1 - \delta_1 z - Kzz^* \tag{5.12b}$$

$$\dot{z^*} = \alpha_2 - \delta_1 z^* - Kzz^* \tag{5.12c}$$

$$\dot{I} = \Delta\beta_{I1} z + \Delta\beta_{I2} z^* - \delta_2 I \tag{5.12d}$$

$$\dot{X} = \Delta\beta_X z^* - \delta_2 X \tag{5.12e}$$

$$\dot{G} = \Delta\Theta p\gamma_2 - \delta_3 G \tag{5.12f}$$

$$\dot{p} = R_{off}(I, X)(1 - p) - R_{on}(I)(p) \tag{5.12g}$$

where state variables D, z, z^*, I, X, G are concentrations as defined above and $p \in [0, 1]$ is the proportion of registers in their "ON" state. The α's are transcription rates, β's are translation rates, γ's are lumped transcription/translation, δ's are degradation/dilution rates, and K is the binding rate between mRNAs. The parameter Δ is used to introduce disturbances in global (cell-wide) translation rate, and Θ introduces disturbances in transcription from the plasmid encoding G and z. Initially we have $\Delta, \Theta = 1$; variation from this value represents introduction of a disturbance. Integrase/excisionase mediated DNA flipping is described by:

$$R_{off}(I, X) = r_{off} f_X(X) f_I(I) \tag{5.13a}$$

$$R_{on}(I) = r_{on} f_I(I) \tag{5.13b}$$

$$f_X(X) = \frac{X^4}{K_X^4 + X^4} \tag{5.13c}$$

$$f_I(I) = \frac{I^4}{K_I^4 + I^4} \tag{5.13d}$$

where $r_{on,off}$ are scaling factors for the rates of DNA inversion, $K_{I,X}$ are equilibrium constants that determine the concentration of each protein required for the half-maximum flipping rate to be achieved, and the order four Hill functions arise due to four molecules of each protein being required (one dimer at each recognition site) to perform DNA inversion [63, 212]. The terms in (5.13) have been slightly re-defined from (5.4). In particular, the influence of the state of DNA registers (p) is now considered in (5.12g) instead. Values used for these parameter are presented in Table 5.1.

5.3.3 Analysis

We can solve (5.12) at steady-state to eliminate variables D, z^*, I, and X, finding that the equilibrium concentration of z must satisfy:

$$\frac{\alpha_1\gamma_1\Delta\Theta p}{\delta_2} = \delta_1 z + Kz\mu \tag{5.14a}$$

$$p = \frac{f_X(\Delta\beta_X\mu/\delta_2)}{\frac{r_{on}}{r_{off}} + f_X(\Delta\beta_X\mu/\delta_2)} \tag{5.14b}$$

$$\mu = \frac{\alpha_2}{\delta_1 + Kz} \tag{5.14c}$$

where we note that the factor $\Delta\Theta p$ (proportional to the production rate of G) appears on the left hand side of (5.14a). If we have $Kz \gg \delta_1$ (i.e. most mRNA binds to its complementary mRNA prior to degrading) then the right hand side of (5.14a) can be simplified to give $p\Delta\Theta \propto \delta_1 z + \alpha_2$, meaning that G will reject changes in Δ, Θ so long as $\delta_1 z$ does not vary significantly. This can be partially achieved by minimising δ_1 (as in other biological circuits targeting adaptation [175]), as well as by minimising the sensitivity of the steady-state value of z to changes in Δ and Θ. To do this we compute partial derivatives of (5.14a) with respect to Δ, setting $y = \Delta\beta_X\mu(z)/\delta_2$ to give:

$$\frac{\partial}{\partial\Delta}\frac{\alpha_1\gamma_1\Delta\Theta p}{\delta_2} = \frac{\partial}{\partial\Delta}(\delta_1 z + Kz\mu(z)) \tag{5.15}$$

$$\frac{\alpha_1\gamma_1\Theta p}{\delta_2} + \frac{\alpha_1\gamma_1\Theta\Delta}{\delta_2}\frac{\partial p}{\partial\Delta} = \frac{\partial z}{\partial\Delta}(\delta_1 + \frac{k\mu(z)\delta_1}{\delta_1 + Kz}) \tag{5.16}$$

$$\frac{\alpha_1\gamma_1\Theta}{\delta_2}(p + \Delta\frac{\partial p}{\partial\Delta}) = \frac{\partial z}{\partial\Delta}C \tag{5.17}$$

Where we have collected the term on the right hand side into C, then:

$$\frac{\alpha_1\gamma_1\Theta}{\delta_2}(p + \Delta\frac{\partial p}{\partial y}\frac{\partial y}{\partial\Delta}) = \frac{\partial z}{\partial\Delta}C \tag{5.18}$$

$$\frac{\alpha_1\gamma_1\Theta}{\delta_2}(p + \Delta\frac{\partial p}{\partial y}(\frac{\beta_X\mu(z)}{\delta_2} - \frac{\beta_X\mu(z)}{\delta_2}\frac{K\Delta}{(\delta_1 + Kz)}\frac{\partial z}{\partial\Delta})) = \frac{\partial z}{\partial\Delta}C \tag{5.19}$$

Which when the factor of $\frac{\partial z}{\partial\Delta}$ is isolated gives:

$$\frac{\partial z}{\partial\Delta} = \frac{\frac{\alpha_1\gamma_1\Theta}{\delta_2}(p + \Delta\frac{\partial p}{\partial y}\frac{\beta_X\mu(z)}{\delta_2})}{C + \frac{\alpha_1\gamma_1\Theta}{\delta_2}\frac{\partial p}{\partial y}\frac{\beta_X\mu(z)}{\delta_2}\frac{K\Delta^2}{(\delta_1+Kz)}} \tag{5.20}$$

We can follow a similar process to the above for parameter Θ:

$$\frac{\partial}{\partial\Theta}\frac{\alpha_1\gamma_1\Delta\Theta p}{\delta_2} = \frac{\partial}{\partial\Theta}(\delta_1 z + Kz\mu(z)) \tag{5.21}$$

$$\frac{\alpha_1\gamma_1\Delta p}{\delta_2} + \frac{\alpha_1\gamma_1\Theta\Delta}{\delta_2}\frac{\partial p}{\partial\Theta} = \frac{\partial z}{\partial\Theta}C \tag{5.22}$$

$$\frac{\alpha_1\gamma_1\Delta}{\delta_2}(p + \Theta\frac{\partial p}{\partial y}\frac{\partial y}{\partial\Theta}) = \frac{\partial z}{\partial\Theta}C \tag{5.23}$$

$$\frac{\alpha_1\gamma_1\Delta}{\delta_2}(p + \Theta\frac{\partial p}{\partial y}(-\frac{\beta_X\mu(z)}{\delta_2}\frac{K\Delta}{(\delta_1 + Kz)}\frac{\partial z}{\partial\Theta})) = \frac{\partial z}{\partial\Theta}C \tag{5.24}$$

Which when the factor of $\frac{\partial z}{\partial\Theta}$ is isolated gives:

$$\frac{\partial z}{\partial\Theta} = \frac{\frac{\alpha_1\gamma_1\Delta}{\delta_2}(p)}{C + \frac{\alpha_1\gamma_1\Theta^2}{\delta_2}\frac{\partial p}{\partial y}\frac{\beta_X\mu(z)}{\delta_2}\frac{K\Delta}{(\delta_1+Kz)}} \tag{5.25}$$

Since all parameters are positive and $\frac{\partial p}{\partial y} > 0$ (p is an increasing function of y) we observe from (5.20) and (5.25) that when $\Delta, \Theta \approx 1$ have $\frac{\partial z}{\partial\Delta} > \frac{\partial z}{\partial\Theta}$. It therefore follows from the above that $\frac{\partial(\Delta\Theta_p)}{\partial\Delta} > \frac{\partial(\Delta\Theta_p)}{\partial\Theta}$, which means the system's steady-state output will be more sensitive to changes in Δ than Θ.

This architecture's ability to reduce variation in the steady-state z value (and hence the term $\delta_1 z$) can be conceptually understood as arising from "ultrasensitive" behaviour due to z entering the first term in (5.14b) through the fourth-order Hill function f_X. Other biological feedback controllers have been proposed which utilise ultrasensitivity to provide similar adaptive behaviour [213], though the "ultrasensitive" component is typically implemented using a zeroth-order process (such as a balanced phosphorylation cycle) or high-cooperativity transcription-factor regulation [214]. We define our system's *target* as being the steady-state value G_0, determined by (5.12) when $\Delta, \Theta = 1$.

In Circuit 1 (Section 5.2) the proportion of registers in their active state was (at equilibrium) maintained by continuous flipping between states (i.e. I and X were both large and R_{on} and R_{off} were balanced). In contrast, Circuit 2 is designed to move between different switching modes (Fig. 5.9), which drive its state toward a "slack

Param.	Value	Unit	Param.	Value	Unit
α_2	1	nM s^{-1}	α_1	2×10^{-3}	s^{-1}
γ_1	1	nM s^{-1}	β_{I1}	6×10^{-4}	s^{-1}
γ_2	1	nM s^{-1}	β_{I2}	1.2×10^{-3}	s^{-1}
K	3×10^{-2}	nM^{-1} s^{-1}	β_X	2.4×10^{-3}	s^{-1}
K_I	30	nM	δ_1	4.1×10^{-3}	s^{-1}
K_X	20	nM	δ_2	7.8×10^{-4}	s^{-1}
r_{on}	8×10^{-5}	s^{-1}	δ_3	3.9×10^{-4}	s^{-1}
r_{off}	2×10^{-4}	s^{-1}			

Table 5.1: Nominal parameter values used in modelling Circuit 2.

region" (mode m_0) where both R_{on} and R_{off} are small (and hence so is $\dot{p}$). Within this region concentrations of Integrase/Excisionase are small, meaning that convergence to the steady-state dictated by (5.14) will be slow, though outside of the region it may be fast. This is illustrated in Fig. 5.10c, where we set the total production rate of z to a fixed value $\Phi = D\Theta p\alpha_1$ and then calculate corresponding steady-state I and X concentrations: when $\Phi \approx 1$ (mode m_0) we have $R_{on}, R_{off} \approx 0$ and so minimal flipping occurs in either direction. Thus, at (or near) its equilibrium our controller produces little of the control species I and X, minimising its burden upon the cell.

We can bound this slack region (i.e. approximate the boundaries of m_0) by maximal concentrations of I and X (denoted $\hat{K}_I$ and $\hat{K}_X$ respectively), which we define as the point at which flipping rates are below a fraction η of their maximum rates, i.e. $\eta = f_X(X = \hat{K}_X) = f_I(I = \hat{K}_I)$. Thus, if $\hat{K}_I < K_I$ and $\hat{K}_X < K_X$ we have that $\eta < 0.5$ and flipping occurs slowly. The system leaves this slack region (so either $I > \hat{K}_I$, or $X > \hat{K}_X$ and $I > \hat{K}_I$) for only short periods of time (immediately following a change in Δ or Θ), during which fast convergence (variation in p) occurs to bring the output G back toward G_0. To assess the range of Δ or Θ values for which the system's steady state lies in this slack region we solve (5.12c) and either (5.12d) or (5.12e) for z at steady state to gives:

$$z_X = \frac{\Delta\alpha_2\beta_X - K_X\delta_1\delta_2}{K_X K\delta_2} \tag{5.26a}$$

$$z_{I+,I-} = \frac{K\hat{K}_I\delta_2 - \Delta\beta_{I1}\delta_1 \pm \sqrt{s}}{2\Delta K\beta_{I1}} \tag{5.26b}$$

$$s = (K\hat{K}_I\delta_2 + \Delta\beta_{I1}\delta_1)^2 - 4\Delta^2\alpha_2\beta_{I1}\beta_{I2}K \tag{5.26c}$$

Which we then substitute into (5.14). We can solve the resulting equation for Δ (with $\Theta = 1$) or Θ (with $\Delta = 1$) to yield $\Delta_{I-,I+,X}$ and $\Theta_{I-,I+,X}$ corresponding to $z_{I-,I+,X}$ respectively. This gives an acceptable range of disturbances $\max[\Delta_{I-}, \Delta_X] < \Delta < \Delta_{I+}$ (and similar for Θ). Thus, when disturbances are in this range (considering variation to Δ and Θ separately) the system's equilibrium lies in a region within which both $I < \hat{K}_I$ and $X < \hat{K}_X$ at steady-state. This region is illustrated in Fig. 5.11a where steady state I and X concentrations are plot for varying disturbances; the various limiting values correspond to intersections between functions (e.g. $\Delta_{I-,+}$ are the Δ values at which I crosses $\hat{K}_I$). However, because we have not considered saturation of p we have as yet made no guarantee that these limits can actually be reached prior to controller saturation.

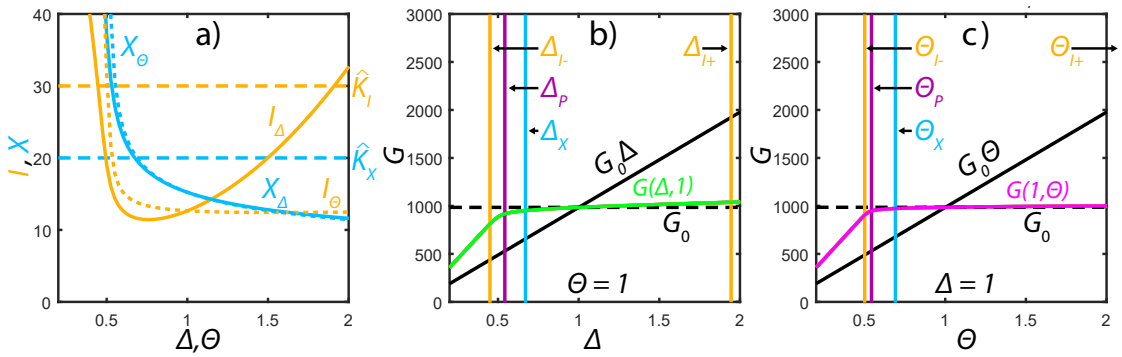


Figure 5.11: Circuit 2 analytic results. **a)** Steady-state I and X concentrations as a function of Δ (I_Δ and X_Δ, solid lines, $\Theta = 1$) or Θ (I_Θ and X_Θ, dotted lines, $\Delta = 1$). We set $\eta = 0.5$ here (and for subsequent figures), meaning $\hat{K}_I = K_I$ and $\hat{K}_X = K_X$. **b)** System steady-state output ($G(\Delta, 1)$) as a function of Δ with $\Theta = 1$, compared to its target value (G_0, black dashed line) and the result from an un-controlled case ($G_0\Delta$). Limiting values of Δ as described in the text are indicated by vertical lines. Control capability quickly declines for Δ values below the saturation point Δ_p. **c)** Similar to panel (b), but with $\Delta = 1$ and Θ varying. $\Theta_{I+} \approx 500$ is not pictured.

We can constrain the magnitude of disturbances that Circuit 2 can reject without saturating by examining its performance at limiting values of p. As Δ, Θ increase the value of the product $\Delta\Theta p$ in (5.12f) can be maintained by setting an appropriately small p. However, though the system does not run out of actuating potential in this case (i.e. we never reach $p = 0$), the steady-state error may increase. For the case where $\Delta, \Theta \to 0$ the value of $\Delta\Theta p$ can only be maintained up to the point where $z = 0$ and almost all registers are in their ON position. We can solve (5.14) to find the steady state value

$\hat{z}$ when $\Delta = \Theta = 1$, which we then substitute into:

$$\frac{f_X(\beta_X \hat{\mu}/\delta_2)}{\frac{r_{on}}{r_{off}} + f_X(\beta_X \hat{\mu}/\delta_2)} = \Theta\Delta \frac{f_X(\Delta\beta_X\alpha_2/\delta_1\delta_2)}{\frac{r_{on}}{r_{off}} + f_X(\Delta\beta_X\alpha_2/\delta_1\delta_2)} \tag{5.27a}$$

$$\hat{\mu} = \frac{\alpha_2}{\delta_1 + K\hat{z}} \tag{5.27b}$$

in which the right hand side of (5.27a) is the saturated signal $\Delta\Theta p$ when $z = 0$. Solving (5.27) for Δ (with $\Theta = 1$) or Θ (when $\Delta = 1$) gives minimal values Δ_p and Θ_p respectively that the system could compensate for without saturating, though steady-state error will again increase as this value is approached.

In Fig. 5.11b,c we analytically calculate the steady state outputs of (5.12) (denoted $G(\Delta, \Theta)$) over a range of Δ and Θ values. The various limiting values of Δ and Θ described previously are plot as vertical lines, as is the expected output ($G_0\Delta\Theta$) that would result without the controller's action (i.e. if G expression was independent of p). We observe that the system is able to reject a disturbance in Δ, and operates in the intended "slack region" for a wide range of Δ. Rejection of a disturbance to Θ is even better (i.e. the slope of $G(1, \Theta)$ is closer to 1), and the slack region extends to much higher values of Θ.

We can investigate the stability of the system in (5.12) by assuming that at equilibrium each individual term in the equation for $\dot{p}$ is zero (i.e. $R_{off}, R_{on} \approx 0$). Linearising about this point (in terms of D,z,z^*, I,X) gives the Jacobian:

$$\mathbf{J} = \begin{bmatrix} -\delta_2 & 0 & 0 & 0 & 0 \\ -\Theta p \alpha_1 & -\delta_1 - Kz^* & -Kz & 0 & 0 \\ 0 & -Kz^* & -\delta_1 - Kz & 0 & 0 \\ 0 & \Delta\beta_{I1} & \Delta\beta_{I2} & -\delta_2 & 0 \\ 0 & 0 & \Delta\beta_X & 0 & -\delta_2 \end{bmatrix} \tag{5.28}$$

which has eigenvalues $\lambda_{1,2,3} = -\delta_2$ and

$$\lambda_{4,5} = -\delta_1, -\delta_1 - K(z + z^*) \tag{5.29}$$

and thus all system poles (eigenvalues of (5.28)) are in the left half plane, meaning this equilibrium is stable. We can numerically evaluate the eigenvalues for the complete system's Jacobian (i.e. including p in (5.28)) about its equilibrium when $\Delta = \Theta = 1$,

which returns (approximately) the same eigenvalues as above, and an additional sixth value of $\lambda_6 = -4.4 \times 10^{-6}$, which is as expected much smaller than δ_2 (previously the eigenvalue of smallest magnitude).

5.3.4 Dynamic Simulation

In Fig. 5.12 we numerically integrate the system in (5.12), applying a time-varying disturbance profile $\tau(t)$ (Fig. 5.12a) to Δ (with $\Theta = 1$, $G(\tau, 1)$) and then to Θ (with $\Delta = 1$, $G(1, \tau)$). Following a step change in τ the system rapidly varies p (Fig. 5.12b), which is driven by fluctuating mRNA concentrations (Fig. 5.12c) that govern expression of I and X (Fig. 5.12d). We observe that the system is achieving the intended "slack zone" behaviour: I and X concentrations spike for short periods of time, rapidly driving p close to its new equilibrium, and then remain below $\hat{K}_I$ and $\hat{K}_X$ respectively. The system output (Fig. 5.12e) converges close to its pre-adaptation level following each disturbance, though this occurs over a longer time-scale since protein G is assumed to be stable (δ_3 is small).

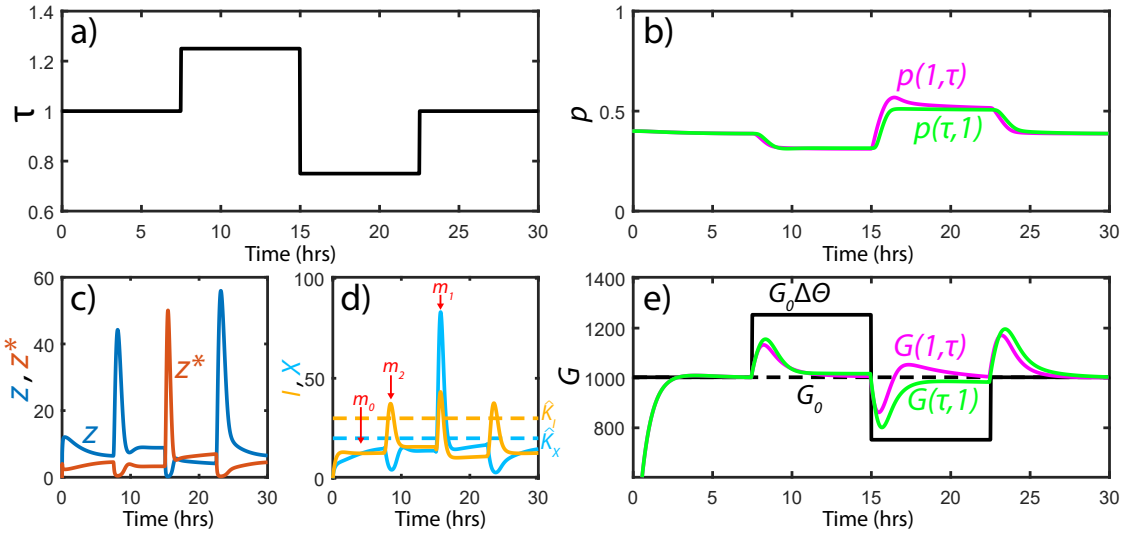


Figure 5.12: **a)** Disturbance profile τ substituted for Δ or Θ in subsequent sub-figures. **b)** The proportion of plasmids in their ON state (p) varies in response to disturbance in Δ ($G(\tau, 1)$) or Θ ($G(1, \tau)$). **c)** The concentration of each mRNA species and **d)** concentration of I and X for $\Delta = \tau$, $\Theta = 1$. Spikes in concentration appear immediately following application of a disturbance, which push the production rates of z and z^* out of balance. Example points at which the system is in each of its operating modes $m_{0,1,2}$ are labelled. **e)** System output G, demonstrating near-perfect rejection of disturbance to either Δ or Θ. $G_0\Delta\Theta$ indicates the system's expected steady-state output if no control system was in place.

5.3.5 Stochastic Simulation

Though in previous sections we have assumed that p is a continuous variable, in reality a finite number of DNA registers will be present within a given cell (with one, or perhaps a couple, per plasmid). We investigate the impact of discretising p by setting:

$$\hat{p} = \text{round}(p \times p_T) \tag{5.30}$$

and then substituting $p = \hat{p}/p_T$ in (5.12). Here p_T is the copy number (per cell) of the flippable DNA register, and the round function returns its argument rounded to the nearest integer. We solve (5.12) with this modification to find the steady-state value of G as Δ is varied in Fig. 5.13a, observing that as p_T is increased our system's output converges to the $p_T = \infty$ behaviour (i.e. Fig. 5.11b).

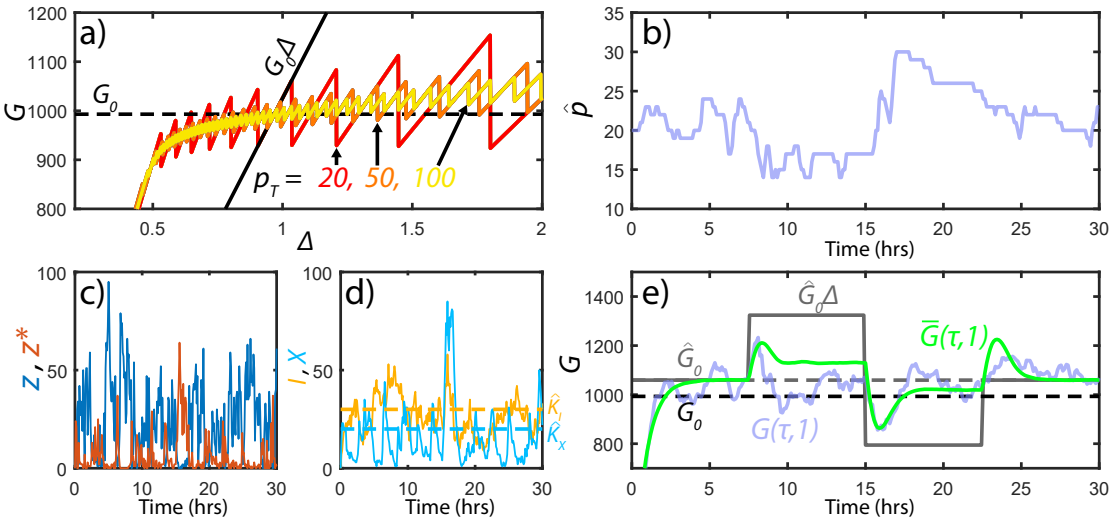


Figure 5.13: a) The dependence of our system's behaviour on the copy number of DNA registers (p_T) simulated as described in the text. Other quantities plot parallel those in Fig. 5.11b. **b-d)** Similar to Fig. 5.12b-d but for a single stochastic simulation run and with $\Theta = 1$. In panel b the total number of ON registers ($\hat{p}$) is plot, rather than the ON proportion of total registers (p). **e)** System output G for a single stochastic simulation ($G(\tau, 1)$) and the average output of 3000 stochastic simulations ($\bar{G}(\tau, 1)$). The target output for (mean steady-state G value when $\Delta = 1$) for the stochastic case ($\hat{G}_0$) differs to that for the deterministic case (G_0, see Fig. 5.12e) as described in the text.

As our system operates in a regime where concentrations of some system species (namely z and z^*) are low, the stochasticity of individual biochemical reactions may substantially influence its output. This may mean that intrinsic noise (discussed in

Section 2.1.4) may have a significant impact upon its behaviour. In Fig. 5.13b-e we perform tau-leaping Gillespie simulations of the system (algorithms for which are outlined in Section 2.4.5), setting $p_T = 50\ cell^{-1}$ and expressing each state variable in terms of integer copies per cell (rather than concentration). This follows from an average *E. coli* cell volume of 0.6 μm^3 meaning that a 1 nM concentration corresponds to approximately 1 molecule per cell. We observe substantial variability in z and z^* concentrations (Fig. 5.13c), which drives variability in I and X (Fig. 5.13d), and through that $\hat{p}$ (Fig. 5.13b) and ultimately G (Fig. 5.13e). In Fig. 5.13e we observe that stochastic treatment of this system increases its mean output level (from G_0 to $\hat{G}_0$), and (on average, $\bar{G}$) degrades the accuracy of adaptation to changing Δ. This changed behaviour arises due to the time-averaged value of Kzz^* in (5.12b,c) being reduced in the stochastic case. This intuitively follows from the fact that $\mathrm{E}[zz^*] = \mathrm{E}[z]\mathrm{E}[z^*] + \mathrm{cov}(z, z^*)$ (where E and cov are the expectation and covariance of a random variable respectively), and their covariance will be negative (due to negative inter-dependence introduced by $-Kzz^*$), thus reducing the expected magnitude of the product zz^*.

5.3.6 Discussion

Our results demonstrate that Circuit 2 is able to effectively reject a wide range disturbances to either translation or transcription rate. While Δ_p, Θ_p represent minimum disturbance limits that result in control saturation, the other limits described are solely constraints on the low-burden slack region in which the controller is designed to operate. These can be exceeded without degrading control capability, but come at the expense of increased protein production (and hence burden). Tuning of system parameters (in particular α_1) would allow the system's set-point to be placed closer to the middle of this range. As Δ, Θ grow the primary limit on the system's adaptative ability will thus be that imposed by the discretisation of DNA register state: p cannot be made arbitrarily small, and so large Δ, Θ cause a stepped response (Fig. 5.13a), particularly when the total number of DNA registers (p_T) is small. Furthermore, regulation of z by measurement protein species D may lose linearity (and saturate) for large Δ, Θ, providing an additional upper limit on disturbance size. The system's adaptive capability was degraded when constituent

reactions were simulated stochastically, though this may largely be addressed by parameter tuning. Our system's design is intended to minimise the burden it imposes on a host cell by operating in a state where I,X (and potentially D) protein concentrations are small. However, in doing so trade-offs are being made in terms of consumption of different cellular resources. For example, maintaining a large number of DNA registers (p_T) within a cell may also impact growth and other cellular processes.

Future work will include further stochastic analysis of the proposed control architecture's behaviour, as well as investigation of its parameter sensitivity. Experimental implementation would then likely require tuning of system parameters that are identified as being major determinants of performance, for example via Ribosome Binding Site variation or inclusion of protein degradation tags (i.e. to vary δ_2) as was done in past implementations of integrase/excisionase controllers (e.g. the fast degradation tag that was attached to the excisionase in Circuit 1). Another experimental challenge is posed by the design of mRNAs z and z^* which code for proteins whilst also being able to bind to each other, as discussed in Section 5.3.1.

5.4 Conclusion

In this chapter I have investigated two feedback control systems based upon DNA recombinase enzymes. The first (Circuit 1) provides a simple implementation of negative feedback, which was implemented *in vivo*. It can partially reject disturbances to the production rate of a protein output of interest, and may therefore find broad applications in situations where noisy species concentrations can impede the operation of synthetic biological constructs (described in detail in Section 2.1.4). Circuit 2 represents a significantly more complex feedback architecture, but takes the disturbance rejecting ability further to provide near-perfect adaptation to both transcriptional and translational disturbances. It also represents the first proposed biological implementation of a switching control architecture, which may prove popular for future applications requiring controllers that impart minimal cellular burden. Circuit 2 is, however, significantly more difficult to implement than Circuit 1 due to the RNA engineering required to achieve the desired

"subtraction junction" component of this design.

In the broader sense, control mechanisms that utilise Integrase/Excisionase mediated DNA flipping may present interesting opportunities for low-burden dynamic information storage and control: whilst information stored in the concentration of a protein (or phosphorylation state of a substrate) is diluted over time, the mean state of a DNA sequence is maintained (since it is copied in its present form). Future work may thus include development of different control schemes based upon the fundamental architectures described in this chapter, eventually leading to low-burden adaptive controllers with a diversity of synthetic biological applicatons [28].

6

Synthetic Biological Reporting Circuits

In Chapters 4 and 5 synthetic biological circuits were developed which (among other purposes) aimed to reduce variability or noise in gene expression. To assess the level of variability in these circuits we generally needed to use a flow cytometer to measure cell-to-cell variability in their output (quantified in terms of standard deviation of per cell fluorescence). The aim of the present chapter is to develop an alternate fluorescent reporter system which allows cell-to-cell variability to be estimated using only bulk measurements of fluorescence, thus avoiding the need for (time consuming and expensive) flow cytometry. In particular, the investigation presented in this chapter was motivated by a desire to provide the biological tools necessary for the Turbidostat device described in Chapter 7 to be used for estimation of cell-to-cell variability. Most of the results from this chapter were published in [65].

6.1 Motivation

As described in Section 2.1.4, a major challenge in the design of reliable synthetic biological systems is assessing and regulating the impact of variability and noise upon their behaviour [28]. When synthetic biological designs are realised experimentally, assessment of their noise properties and performance variability is therefore a critical part of their characterisation. Fluorescent reporter proteins are frequently used as an output for such systems due to the ease with which they are measured [215]. Experiments are often repeated many times (under theoretically identical conditions) using plate-reader hardware (e.g. as in Chapter 5) to observe variation in their behaviour, which is then quantified in terms of the variability in a mean experimental outcome (for example, a cell culture's bulk

fluorescence output). However, measurement of only a cell culture's mean fluorescence can disguise important but often unnoticed behaviours [216], such as bi-modal responses [217]. To accurately characterise many systems, measurement of intra-population cell-to-cell variability is thus required [216], for which techniques such as fluorescence microscopy and flow cytometry (which can measure individual fluorescence levels of a large number of cells) are employed extensively [218]. Recently, improvements in the capabilities of these single-cell measurement technologies have facilitated high-throughput studies of cell-to-cell variability [219]. However, single-cell measurement techniques remain time- and equipment-intensive, making their use unfavourable if large numbers of samples must be measured at regular time intervals, as is the case when parallelised experiments are used to test synthetic circuits over a range of input parameter combinations (often done using a plate reader).

This experimental challenge motivates the aim of the current chapter: to design a synthetic biological reporter circuit that allows characterisation of cell-to-cell variability using only measurements of a cell culture's mean behaviour. Such a reporter could be used in any synthetic biological system with a fluorescent protein output to give an estimate of cell-to-cell variability at little added cost. This would be of great utility for many high-throughput noise-measurement experiments (which currently require flow cytometry) as it would allow them to be performed efficiently and in parallel using a plate-reader device. We achieve noise estimation via a biological implementation of a multiplication function that can be used to measure a variable's mean value, as well as its mean squared value, at a population level. It is then possible to estimate the cell-to-cell standard deviation of a system's output, which previously required measurements at a single-cell level. Thus, though our system does not return the complete shape of the expression distribution (as does flow cytometry), for cases in which the distribution mean and width are the primary parameters of interest (as in many past studies [14, 31, 104] and Chapters 4 and 5) it can provide an efficient experimental alternative.

We begin in Section 6.2 with a description of the simple statistics behind this estimation procedure, showing how measurement of only bulk culture parameters can allow quantification of cell-to-cell variability. In the case of log-normally distributed behaviour (as is often found in biological systems [147]), we demonstrate that this provides a direct estimate of the width of the system's output distribution. In Section 6.3 we describe a fluorescent-protein FRET implementation for such a circuit (an alternate implementation using an sRNA-based circuit is described in the supplementary material of [65]), and discuss expected parameter values for its operation (Section 6.4). The FRET system's variance-estimating performance is then analysed (Section 6.5), and its operational parameter range and sensitivity to both intrinsic and extrinsic noise is assessed (Section 6.6). In Section 6.7 these results are discussed with reference to potential experimental conditions in which our reporter system might be employed, and potential sources for error (as well as alternate implementations which might minimise these) are outlined.

6.2 Noise Distributions and Quantification

We consider a measurable random biological variable of interest, X, which might (for example) correspond to the fluorescence output from a single cell due to production of a fluorescent reporter protein within that cell. The *expectation* of X, $\mathrm{E}[X]$, is its mean value over a large number of cells. In this context $\mathrm{E}[X]$ would (approximately) correspond to the total fluorescence of a cell colony divided by the number of cells of which it is comprised, hence giving the average fluorescence per cell. Many factors are not considered in this approximation, such as the attenuation of emitted fluorescence as it passes through experimental media (thereby making cells further from a detector device appear slightly dimmer), however these secondary effects will be ignored in our current analysis. The variance of a random variable X is given by [220]:

$$Var[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2, \tag{6.1}$$

where $\mathrm{E}[X^2]$ is the expected (arithmetic mean) value of the variable X squared, also referred to as the second *moment* of the random variable X. In the current context, X^2

would correspond to the squared fluorescence value of a single cell.

Values of X for individual cells will vary within a population, and when a large number of cells are measured individually (i.e. at a single-cell level) the approximate distribution of X can be ascertained. For many variables of interest in biological systems, such as the level of mRNA or protein produced, this distribution is observed to be approximately log-normal during steady growth [145] (though it may be better described by a normal distribution for cells in stationary phase [146]). It has been shown that a log-normal distribution can emerge in such systems due to the inherent complexity of the noisy biochemical processes involved [147]. A log-normal distribution for a given variable will be denoted Log-normal(μ,σ^2), where μ and σ are the mean and standard deviation respectively of the variable's natural logarithm. The nth moment of a log-normal distribution can be calculated analytically [221]:

$$\mathrm{E}[X^n] = e^{n\mu+\frac{1}{2}n^2\sigma^2}, \tag{6.2}$$

where n is a positive integer. In reality we will not be able to measure the absolute value of $\mathrm{E}[X^2]$, but rather a value "calculated" by our reporter system that is proportional to its value. We denote the measured second moment $\mathrm{E}_m[X^2] = \gamma_m \mathrm{E}[X^2]$, where γ_m reflects this proportionality and is termed the 2nd moment gain.

Experimentally, variation between cells in the value of a parameter X is often quantified in terms of the width of the distribution when the variable X is plotted on a log scale (i.e. the parameter σ). We can derive an expression for this variable for the log-normal distribution in terms of measurable quantities, giving:

$$\sigma = \sqrt{\ln\left(\frac{\mathrm{E}_m[X^2]}{\mathrm{E}_m[X]^2}\right) - \ln(\gamma_m)}, \tag{6.3}$$

where $\mathrm{E}_m[X]$ is the measured first moment (the mean). The square-root-log dependence of σ on $\mathrm{E}_m[X^2]$ is worth noting, as it means small changes in σ will be measurable as large changes in $\mathrm{E}_m[X^2]$. Even if we do not estimate/measure the parameter γ_m, we note that σ is an increasing function of $\mathrm{E}_m[X^2]/\mathrm{E}_m[X]^2$ (when the argument is real), meaning

that as the quantity $\mathrm{E}_m[X^2]/\mathrm{E}_m[X]^2$ grows, so does σ. If the underlying distribution is not log-normal and/or unknown we can estimate the population standard deviation using (6.1), though again estimation of γ_m is necessary.

Noise strength (fano factor) is a common parameter of interest when characterising the variability of biological systems, and can be used to discriminate between different mechanisms contributing to noise in protein abundance [194, 195]. This parameter, defined as the ratio of the expression variance to its mean is constant for a perfectly Poissonian process [222]. However, for most biological process (for which a log-normal distribution is anticipated) the coefficient of variation provides a better measure of gene-expression noise [195]. The (arithmethic) Coefficient of Variation (CoV) is defined as the ratio of the arithmetic standard-deviation to the mean ($SD[X]/\mathrm{E}[X]$ where $SD[X] = \sqrt{Var[X]}$) due to the linear relationship observed between these parameters [223]. We therefore have:

$$CoV = \sqrt{e^{\sigma^2} - 1} \approx \sigma \tag{6.4}$$

where the last approximation uses the truncated Taylor-series $e^x \approx 1 + x$ and is valid for small sigma ($\sigma \lesssim 0.8$) . This highlights the appropriateness of the coefficient of variation for assessing noise in biological systems exhibiting log-normal behaviour: it is proportional to the width of the expression distribution when plotted on a log scale.

We have thus outlined a method for determining the level of variability in a parameter X, as is often measured via flow cytometry, but here only using measurements of bulk parameters. To build an estimator of $\mathrm{E}_m[X^2]$ we must create a system with fundamental structure:

$$\mathrm{X} + \mathrm{X} \rightleftharpoons \mathrm{Y}$$

where we can both measure X, and the variable $Y \propto X^2$. We now describe and analyse a possible FRET implementation for such a variance estimating circuit. A second potential implementation, based upon an activating sRNA circuit, is described in the supplementary material of [65].

6.3 A Biological Implementation: FRET Reporter Pair

One approach to measuring a variable X in terms of its mean ($\mathrm{E}_m[X]$) and second moment ($\mathrm{E}_m[X^2]$) is to use a dimerising FP-FRET (Fluorescent Protein - Fluorescence Resonance Energy Transfer) protein pair that can approximate a multiplication function (Fig. 6.1a). In this case we have chosen the mClover3 (C) and mRuby3 (R) fluorescent proteins [134], but many other possibilities exist [135]. Both fluorescent fusion proteins are expressed in tandem from the promoter P_x, for which we aim to measure the variability in transcription. This variability may arise from any upstream processes (e.g. gene regulatory networks) interacting with P_x.

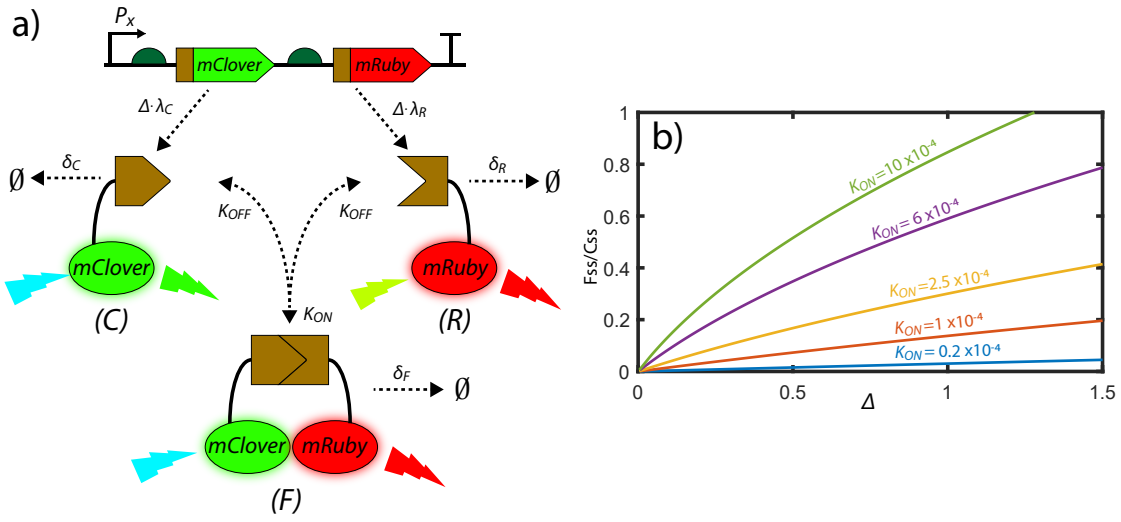


Figure 6.1: A FRET Reporter system for analysis of gene expression variability. a) Structure of the system, in which the concentration of a FRET-active dimer (F) is approximately proportional to the square of the concentration of its unbound components C and R. C and R each consist of a fusion of a fluorescent protein (mClover3 and mRuby3 respectively) with a protein binding domain (brown). Dashed arrows represent reactions and their directions (with rate parameters as defined in the text), and degradation/dilution reactions (δ) represent removal of species to the null state $\emptyset$. The DNA operon (top) consists of a promoter (P_x), Ribosome Binding Sites (RBS, dark-green), genes (labelled), and a transcriptional terminator. **b)** The dependence of the ratio F_{SS}/C_{SS} (as defined by (6.6) and (6.7)) on the value of Δ. Colours denote different values of K_{ON} in units of s^{-1}. For an ideal multiplication operation F_{SS}/C_{SS} would be linear in Δ, which is best satisfied for small values of K_{ON}.

When spatially separated, the fluorescent proteins have their usual excitation/emission behaviour. However, when they are brought into close proximity ($\sim$ 10nm or less [135]) it is possible to excite the complex at the lower excitation wavelength of mClover3 (termed

the *donor*), and measure a change in fluorescence in the emission spectrum of mRuby3 (the *acceptor*). This occurs via non-radiative energy transfer through long-range dipole to dipole interactions when the fluorophores are proximal [224]. A diversity of approaches to measuring FRET exist [135], though we will concentrate on ratiometric measurements of FRET intensity which may be performed using a plate reader (or fluorescence microscope or flow cytometer) [225].

To co-localise the two fluorescent proteins such that FRET can occur we fuse them to complementary protein binding domains, as has been done in a number of FRET-based protein localisation studies [226, 227]. Each fluorescent protein can therefore exist in isolation, or they can dimerise to form a complex F (Fig. 6.1a). The binding strength of the complementary domains represents the major parameter for tuning in our system, which could initially be investigated experimentally using a pair with ligand-dependent binding [227, 228]. Their binding properties will determine the forward and backward rates (K_{ON} and K_{OFF}) for the reaction:

$$C + R \underset{K_{OFF}}{\overset{K_{ON}}{\rightleftharpoons}} F$$

If we assume the concentrations of both C and R are proportional to the gene expression variability that we aim to analyse (the parameter Δ introduced at P_x), then this system provides multiplicative ability when the concentrations of C and R are greater than F. Following the terminology of the previous section the concentration (and hence fluorescence) of either mClover or mRuby is a proxy for the variable X which we aim to measure, and the concentration of F represents the variable Y. We can model this system using a system of differential equations of the form:

$$\dot{C} = \Delta\lambda_C - \delta_C C - K_{ON}CR + K_{OFF}F \tag{6.5a}$$

$$\dot{R} = \Delta\lambda_R - \delta_R R - K_{ON}CR + K_{OFF}F \tag{6.5b}$$

$$\dot{F} = K_{ON}CR - K_{OFF}F - \delta_F F \tag{6.5c}$$

where C, R are the concentrations of the mClover and mRuby fusion proteins respectively, and F is the concentration of the bound FRET complex. Δ is the random scaling factor

introduced by P_x which we aim to measure, $\lambda_{C,R}$ are lumped transcription/translation rates (we model these in a single step for simplicity), and $\delta_{C,R,F}$ are degradation rates for the corresponding species. By setting the time derivatives to zero it is possible to solve the system of equations (6.5) analytically to find the steady-state (ss) concentrations of each complex, giving:

$$C_{ss}, R_{ss} = \frac{\Delta\lambda_{C,R} - \delta_F F_{ss}}{\delta_{C,R}}. \tag{6.6}$$

and

$$F_{ss} = a(b + \Delta c - (\Delta^2 d + 2bc\Delta + b^2)^{0.5}) \tag{6.7}$$

where:

$$a = 1/(2K_{ON}\delta_F^2) \tag{6.8a}$$

$$b = \delta_C\delta_R(K_{OFF} + \delta_F) \tag{6.8b}$$

$$c = K_{ON}\delta_F(\lambda_C + \lambda_R) \tag{6.8c}$$

$$d = K_{ON}^2\delta_F^2(\lambda_C - \lambda_R)^2 \tag{6.8d}$$

For an ideal multiplication operation to occur, we aim to tune system parameters such that $F_{ss} \propto \Delta^2$, which would then mean (approximately) that $F_{ss} \propto C_{ss}^2, R_{ss}^2$. Observing that the constant term ($F_{ss}|_{\Delta=0}$) and first derivative ($\frac{\partial F_{ss}}{\partial \Delta}|_{\Delta=0}$) are zero, to achieve $F_{ss} \propto \Delta^2$ we thus desire:

$$\frac{\partial^2 F_{ss}}{\partial \Delta^2} = \frac{ab^2(c^2 - d)}{(\Delta^2 d + 2bc\Delta + b^2)^{1.5}} \sim Constant, \tag{6.9}$$

for all Δ. This criteria is best satisfied if b is large and c and d are small. Here b can be made large by setting the protein degradation terms to be fast (i.e. $\delta_{C,R,F}$ are large). However, tuning these parameters can be challenging (and may introduce noise): a more effective approach is to focus on c and d which can be made small by reducing K_{ON} (and in the case of d setting $\lambda_C = \lambda_R$). K_{ON} is the rate at which the fusion proteins C and R dimerise, which can be reduced by making the binding of their attached protein domains weak. By minimising F_{ss} this constraint also achieves $C, R \propto \Delta$ in (6.6). This follows from an intuitive analysis of the opposite scenario, in which a large F_{ss} means

most protein is in its dimerised form, and thus F_{ss} will be proportional to whichever of C_{ss} or R_{ss} is smaller.

These results can be rearranged to give:

$$F_{ss} = \frac{K_{ON} C_{ss}}{K_{OFF} + \delta_F} \cdot \frac{\Delta(\lambda_R - \lambda_C) + \delta_C C_{ss}}{\delta_R} \tag{6.10}$$

which if $\lambda_R = \lambda_C$ and $\delta_R = \delta_C$ (i.e. both fusion proteins are expressed and degraded at equal rates), simplifies to:

$$F_{ss} = \frac{K_{ON}}{K_{OFF} + \delta_F} C_{ss}^2 \approx \gamma_m C_{ss}^2 \tag{6.11}$$

giving a straight-forward estimate of γ_m for our simulated system. In the case where $\delta_F << K_{OFF}$ we have $\gamma_m \approx 1/K_d$, where $K_d = K_{OFF}/K_{ON}$ is the dissociation constant of the dimerisation process.

6.4 Parameter Selection

We express parameters values in terms of molecules per cell to aid stochastic simulations: assuming a cell volume of 0.6 μm^3 a 1 nM concentration corresponds to approximately 1 molecule per cell. The lumped transcription/translation rates ($\lambda_{C,R}$) can be expressed as a product of the constituent processes, $\lambda_{C,R} = \alpha_m \beta_{C,R}/\delta_m$, where α_m is the transcription rate of the (common) mRNA, $\beta_{C,R}$ are the translation rates of the two fusion proteins, and δ_m is the mRNA degradation rate. We set the transcription rate from the promoter as $\alpha_m = 0.05\ s^{-1}$ to correspond with typical levels from a mid-strength promoter in *E. coli* [190, 191], and the translation initiation rates are set as $\beta_{C,R} = 0.05\ s^{-1}\ mRNA^{-1}$ as is measured for mid-strength RBSs [190, 229]. The mRNA degradation rate is set as $\delta_m = 4.1 \times 10^{-3}\ s^{-1}$ to represent a typical mRNA half-life of 2.8 minutes for *E. coli* growing in exponential phase [192]. Combining these parameters we have $\lambda_{C,R} = 0.61\ s^{-1}$. The protein degradation rate is set as $\delta_{C,R,F} = 3.9 \times 10^{-4}\ s^{-1}$ to correspond to dilution during growth with a 30 minute doubling time [191]. These parameter values give a typical equilibrium protein abundance of $\lambda_{C,R}/\delta_{C,R,F} \sim 1500 \times \Delta\ cell^{-1}$, and correspond closely to those assumed in other studies of noise in *E. coli* [230].

As described in the previous section, the values of K_{ON} and K_{OFF} are the most important tuning dial for achieving the system's desired multiplicative behaviour. Generally the dimerisation timescales observed in FP-FRET experiments are substantially faster than cellular processes that they aim to measure, such as phosphorylation [231]. We therefore set $K_{OFF} = 1\ s^{-1}$ so that our system reflects the typical fast response times of ~ 10 s for FRET dimerisation [232]. To guide our selection of K_{ON} values we consider a typical dissociation constant for protein-protein binding in a FP-FRET system, $K_d = 4.4\ \mu$M [231]. This value is very similar to that measured for the same protein binding domains when they are not fused to fluorescent markers [231], which (assuming this consistency transfers to other systems) would simplify the task of selecting appropriate binding domains from the literature to implement our system. Using these values for K_{OFF} and K_d (and converting to units of molecules per cell) we have $K_{ON} = 2.4 \times 10^{-4}$ s^{-1}, which will be used to guide the range of K_{ON} values investigated. For typical experiments in which our variability reporter might be employed [51, 63, 69] we find values of σ in the range $0.5 \sim 1.5$, though we will focus analysis on the lower end of this range (as this is where disruption of our system's performance by noise and parameter variation will be greatest).

An important consideration in the FRET implementation's design is that its output remains measurable: if very small K_{ON} values are chosen (to best approximate a multiplication function), it may be challenging to measure the small amount of FRET signal against a high level of background fluorescent proteins that are not undergoing FRET [233]. In an *in vitro* study Gao *et al.* found that the FRET ratio was measurable with very high signal-to-noise for monomers at a concentration of $\sim 1\ \mu$M which (for their measured dimersiation K_d) corresponded to $\sim 15\%$ of the proteins being in their FRET-active dimerised state [231]. Further, Potzkei *et al.* demonstrated that accurate FRET measurements are possible in living *E. coli* when only a small fraction of a sensor was in its FRET-active state [234]. Due to overlap of fluorescent protein spectra, measurements include excitation and emission cross-talk which is exhibited as a baseline fluorescence level in the emission band of the acceptor fluorescent protein. Hoppe *et al.* demonstrated one

approach that uses measured fluorescence values to calculate the relative concentrations of monomers and a FRET-active dimer in the presence of this cross-talk [235] (and a range of other methods exist [233]). Furthermore, they demonstrated that FRET signals are distinguishable even when the concentration of the FRET-active dimer is <5% of that of one of its constituent components [235]. Therefore, in the following sections we analyse our circuit's behaviour in terms of the relative concentrations of its components ($\mathrm{E}_m[F]$, $\mathrm{E}_m[C]$), assuming that these can be inferred from accurate measurements of fluorescence (enabled by characterisation of the particular FRET pair's excitation/emission interaction [135, 235]).

6.5 Deterministic Analysis

In this section we analyse the system's steady-state output (given by (6.6) and (6.7)) using the parameter values discussed above, and with the value of K_{ON} varying between simulations. A similar analysis is performed for the alternate activating sRNA implementation in the supplementary material of [65]. In Fig. 6.1b the ratio F_{ss}/C_{ss} is plot for varying K_{ON}; for ideal multiplicative behaviour we desire a linear relation, which is best achieved for low values of K_{ON}. As discussed in the previous section this parameter's value represents an experimental trade-off: smaller K_{ON} values uniformly reduce the concentration of F (and hence the measured FRET dynamic range), but mean that the system better approximates a multiplication function.

We now examine the system's ability to retrieve parameters of the underlying probability distribution of Δ when only measurements of the circuit's mean outputs ($\mathrm{E}_m[F]$, $\mathrm{E}_m[C]$, and $\mathrm{E}_m[R]$) are available. The values for Δ are sampled from a log-normal distribution ($\Delta \sim \text{Log-normal}(\mu, \sigma^2)$). We vary the value of σ, which is proportional to the distribution's width when plot on log-scale, and $\mathrm{E}[\Delta]$, the expected value of Δ when sampled from this distribution. $\mathrm{E}[\Delta]$ therefore corresponds to the distribution's arithmetic mean, as would be measured using a plate reader. Fig. 6.2a provides examples of log-normal distributions with varying parameters: changing $\mathrm{E}[\Delta]$ with σ fixed shifts the

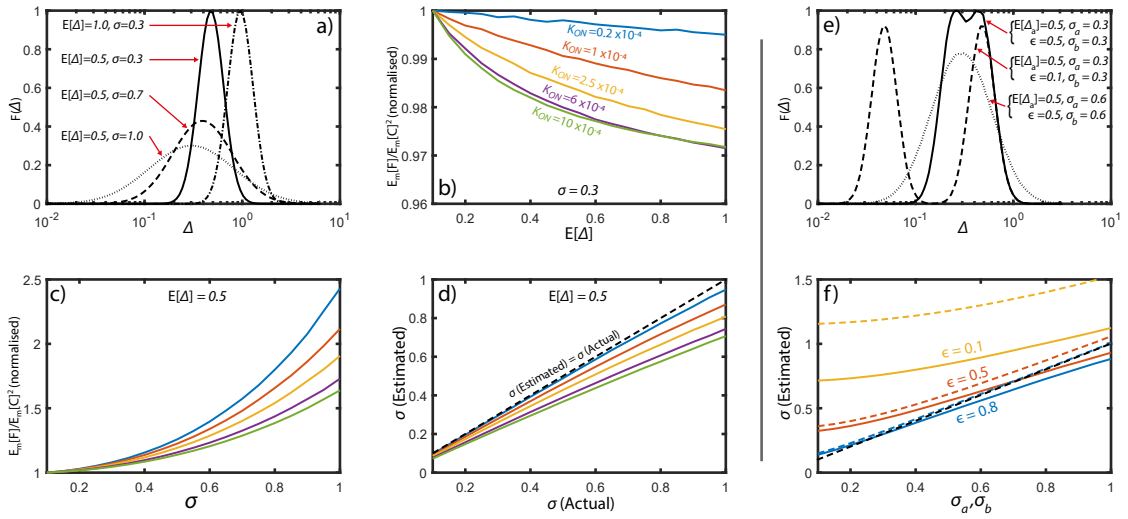


Figure 6.2: Deterministic Analysis of FRET reporter system. a) Log-normal distributions for varying $\mathrm{E}[\Delta]$ and σ. **b)** Normalised (by value at $\mathrm{E}[\Delta] = 0.1$) value of the system output $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$ (which is calculated as an average of system output for 10^6 values of log-normally distributed Δ) as the distribution's mean ($\mathrm{E}[\Delta]$) is varied for a fixed value of σ. An ideal system would be invariant under changes in $\mathrm{E}[\Delta]$, which is best achieved for small K_{ON} (colours denote different values of K_{ON} in units of s^{-1} and are maintained in panels c,d). **c)** System outputs simulated similarly to panel b, but with varying σ and fixed $\mathrm{E}[\Delta]$. An ideal system would show a strong dependence on σ, making experimental measurement of changes in this ratio easier, which is best achieved for small K_{ON}. **d)** Comparison of the value of σ estimated using (6.3) (and the approximation $\gamma_m = K_{ON}/K_{OFF}$ from (6.11)) with the underlying distribution's actual value. Best estimation accuracy is achieved for small K_{ON}. **e)** Bi-modal log-normal distributions described in (6.12) with varying σ and ϵ. **f)** Estimated value of σ (solid lines) and "true" value of σ (dashed lines, the coefficient of variation of the bi-modal distribution) for bi-modal distributions with $K_{ON} = 1 \times 10^{-4}$ s^{-1}. Black dashed line indicates $\sigma(\text{Estimated}) = \sigma_a, \sigma_b$.

distribution's peak, but maintains its width, whilst changing σ with $\mathrm{E}[\Delta]$ fixed flattens the distribution, and shifts its peak (since $\mu = \ln(\mathrm{E}[\Delta]) - \sigma^2/2$ depends on both $\mathrm{E}[\Delta]$ and σ).

In Fig. 6.2b,c we calculate the system's mean output over 10^6 log-normally distributed Δ values, and examine the experimentally measurable ratio of population means $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$ (which appears in (6.3)) as distribution parameters $(\mathrm{E}[\Delta], \sigma)$ vary. As $\mathrm{E}[\Delta]$ varies this ratio would ideally remain constant if $C \propto \Delta$ and $F \propto \Delta^2$, which is what is observed in Fig. 6.2b with the best agreement occurring for small K_{ON}. In Fig. 6.2c this ratio is plot for varying σ, demonstrating the system's sensitivity to increasing variability in Δ. At this point we have not assumed any fundamental properties of the log-normal distribution (though it has been used to generate random samples), and as such

the ratio $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$ can be used to assess relative changes in noise (the variability of Δ) in a given system, even if the noise distribution deviates from log-normal (though the shape of the curves in Fig. 6.2b,c may change). In Fig. 6.2d we have used this measurable ratio (and the approximation $\gamma_m = 1/K_d$ from (6.11)) to directly estimate the parameter σ using (6.3), demonstrating that the best agreement is found for low K_{ON}.

A useful application of our variance-estimating system would be to identify when biological processes exhibit a bi-modal output, which is obscured by traditional measurements of only a cell culture's mean fluorescence [216, 222]. We therefore analyse our system's behaviour when Δ is sampled from a sum (Δ_s) of two log-normal distributions (Δ_a and Δ_b) with equal weighting:

$$\Delta_s = 0.5 \cdot \Delta_a + 0.5 \cdot \Delta_b \sim 0.5 \cdot \text{Log-normal}(\mu_a, \sigma_a^2) + 0.5 \cdot \text{Log-normal}(\mu_b, \sigma_b^2) \quad (6.12)$$

Examples of such distributions are plot in Fig. 6.2e, where we have set $\sigma_a = \sigma_b$ and we define μ_b such that $\mathrm{E}[\Delta_b] = \epsilon \mathrm{E}[\Delta_a]$. In a standard plate-reader fluorescence measurement a decrease in ϵ (i.e. the distribution becoming increasingly bi-modal) would appear as a decrease in mean expression, which does not give clear indication of an increased variation due to bi-modality. However, in Fig. 6.2f bi-modality is exhibited as a substantial increase in the estimated value of σ, though our system's ability to estimate the "true" value for σ deteriorates as the underlying distribution deviates from uni-modal log-normal. Though bi-modality substantially shifts the estimated σ, it is not distinguishable from the case in which there is a very wide uni-modal distribution with equal σ.

6.6 Parameter Dependence and Noise Sensitivity

The extent to which the deterministic behaviour of the variance-estimating system will be achieved in an experimental realisation will depend upon its performance in a noisy cellular environment. In this section we further analyse the FRET reporter implementation, demonstrating that its functionality is maintained under variations in parameter values, and in the presence of both extrinsic and intrinsic noise.

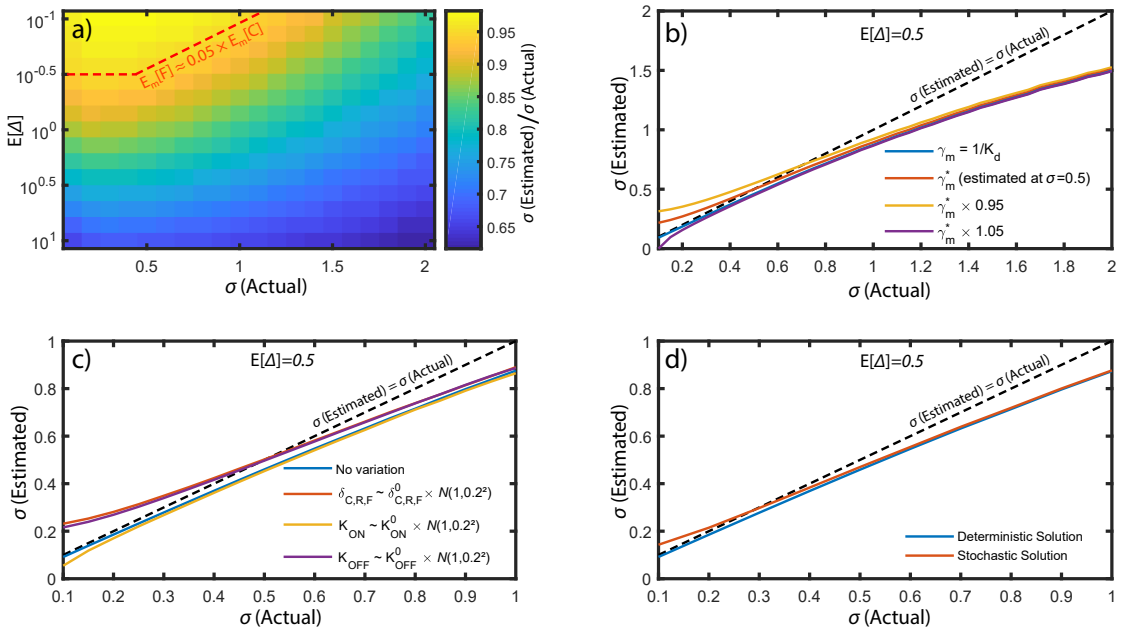


Figure 6.3: Parameter sensitivity and noise analysis with $K_{ON} = 1 \times 10^{-4}$ s^{-1}. a) Accuracy of σ estimation over an experimentally-relevant range of σ and E[Δ] values. The area above the dashed red line indicates the parameter regime in which the concentration of the FRET-active complex F is less than 5% of C. **b)** Dependence of calculated σ upon the estimated value of γ_m^* at $\sigma = 0.5$. **c)** Estimated values of σ when system parameters are sampled from a normal distribution. Note that we scale each of $\delta_{C,R,F}$ by the same randomly sampled value, since each is (predominantly) determined by the shared factor of cellular growth rate. **d)** Comparison of estimated values of σ for the deterministic steady state solution (e.g. (6.6) and (6.7)) and stochastic simulations (described Section 2.4.5).

To assess the range of gene expression levels over which a system with a given K_{ON} value can operate we calculated the accuracy of σ estimation as E[Δ] varied over two orders of magnitude (Fig. 6.3a). This range corresponds to mean protein abundances per cell varying between approximately 150 and 15,000 (see Section 6.4). Estimation accuracy declines (approximately) proportionally to log(E[Δ]), with the tradeoff between accuracy and measurability (i.e. the FRET signal being large compared to the background fluorescence) being governed by K_{ON}. Thus, by decreasing this value variance estimation will be more accurate at high protein abundances (high E[Δ]), but it will be increasingly challenging to measure the FRET signal when protein abundance is low.

Fig. 6.3b compares estimated values of σ as the parameter γ_m used in (6.3) varies. In practice a calibration experiment would be performed using flow-cytometry and a plate reader to give simultaneous measurements of σ, E$_m$[F], and E$_m$[C] from which γ_m can be

estimated using (6.3). The value of γ_m would then (ideally) carry over when different upstream circuits (i.e. inputs at P_x) are connected to the FRET reporter system. In Fig. 6.3b we simulate such a calibration experiment to estimate γ_m^* when $\sigma = 0.5$ (and hence for γ_m^* we have σ(Estimated)=σ(Actual) at this point). We observe that even when this parameter is over- or under-estimated, it does not have a major impact on the estimated value of σ over the range $\sigma = 0.5 \sim 1.5$.

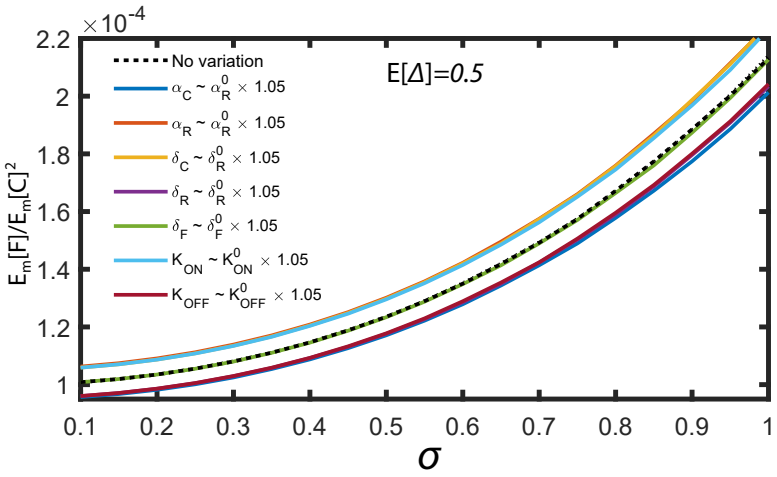


Figure 6.4: Dependence of system behaviour on individual parameter values. The value of the measurable parameter $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$ that can be used to estimate σ is plot as system parameters are varied individually by 5%. α_R, δ_C, and K_{ON} increase, and α_C, δ_R, and K_{OFF} decrease, the value of $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$ at a given σ, whilst δ_F has minimal impact. The superscript 0 (e.g. K_{ON}^0) indicates their previous values.

When system parameters are varied individually, we observe that for a given change in parameter value a similar magnitude of variation is observed in the quantity $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$ (Fig. 6.4). In practice isolated parameter variation is unlikely: it would require a coordinated shift in the mean value of a single parameter (independently of all others) across all cells in a population. More likely are consistent shifts across a group of parameters [230]; for example, λ_C and λ_R might decrease by a similar amount (e.g. due to ribosome sequestration), or $\delta_{C,R,F}$ might vary together (due to a change in mean cellular growth rate). In both these cases we find the opposing actions cancel each other out, returning the system to its previous output level.

In order to estimate the variability of an upstream process we ultimately require minimal noise to be introduced within the variance-estimating system itself. Since our system's purpose is to assess the noise in the upstream regulatory architecture / gene circuit that interacts with P_x, the variability we aim to measure is that which occurs "upstream" of the point at which transcription is initiated from P_x. We are thereby defining factors such as the varying copy number of the plasmid holding P_x (should our system be implemented on a plasmid) as being extrinsic to the system's internally introduced variability. Much of the variability in output protein levels arising within our system will be caused by variation at the transcription and translation steps [147]. We can model noise introduced due to extrinsic sources during this process by defining a new parameter Δ_*:

$$\Delta_* = \Delta \times \Delta_2 \tag{6.13}$$

where Δ is the variability at P_x as defined previously, and Δ_2 is the noise introduced during the gene expression process within our variance-estimating system itself. Due to the large number of secondary biological factors that combine to determine the rate of transcription and translation we anticipate that Δ_2 will also be distributed log-normally [147], and we denote its underlying parameters μ_2 and σ_2. Because Δ_* is the product of two log-normal distributions it will be similarly distributed, with parameters $\mu_* = \mu + \mu_2$ and $\sigma_*^2 = \sigma^2 + \sigma_2^2 + 2\sigma\sigma_2\rho$, where ρ is the correlation between the random variables Δ and Δ_2. For un-correlated noise ($\rho = 0$), $\sigma_* = \sigma\sqrt{1 + \sigma_2^2/\sigma^2}$, and thus the influence of a fixed σ_2 decreases rapidly as σ grows. For perfectly correlated parameters ($\rho = 1$) we have $\sigma_* = \sigma + \sigma_2$ which represents the "worst case" scenario: for any $|\rho| < 1$ we have an estimation error of $|\sigma_* - \sigma| < |\sigma_2|$.

Past studies have considered extrinisic noise to be the component of total noise that is correlated between expression of different proteins within a cell [84] (implying a value $\rho = 1$). However, with parameter Δ_2 we aim to model variability arising due to sources that are external to our circuit's stochastic biochemical reactions, but which may differ between the expression of the different proteins in our system [147]. For example, differing coding sequences for the two fusion proteins would mean that the (extrinsic) variability in

the concentration of charged tRNA species may impact their translation rates to different degrees. Thus, we expect a large (though not unity) value for ρ, since most of this noise arises due to shared parameters determining rates of transcription and translation [147]. That aside, we proceed with the worst case scenario in which the variability estimated by our system (σ_* measured from Δ_*) differs from the "true" variability (σ from the random parameter Δ) by an amount σ_2 due to the impact of extrinsic noise on the variance-estimating system itself.

The important consideration in this case is therefore the difference between σ_2 and σ. We can estimate this from past studies, in which extrinsic noise (measured by CoV, which by (6.4) is approximately equal to σ in the low σ region) was observed to increase by a factor of ~ 5 when a previously constitutively expressed fluorescence marker was put under the control of the inducible LacI transcription factor [84]. Therefore, considering that circuits that are substantially more complex and noisy than a LacI inducible promoter might generally be analysed by our proposed variance estimating system, extrinsic noise is only likely to have a minor influence on its estimating capability. Furthermore, by replacing the promoter P_x with a constitutive promoter known to provide low-noise transcription it may be possible to estimate σ_2 (though this parameter would show some level of context-dependence), and directly subtract its value from the measured σ_* to find σ.

Although the transcription and translation processes are likely to be a more substantial source of noise, we also analysed the impact of extrinsic variability in $\delta_{C,R,V}$ and $K_{ON,OFF}$ by sampling each of these values from a normal distribution (with relatively large standard deviation) and averaging the system's performance over a large number of trials (Fig. 6.3c). We observe that the influence of parameter noise decreases as the level of variability that we are attempting to measure grows, because in this case it is not correlated with variability in Δ.

When biomolecular systems operate at low concentration scales the fact that interactions occur in discrete increments may substantially alter a system's behaviour

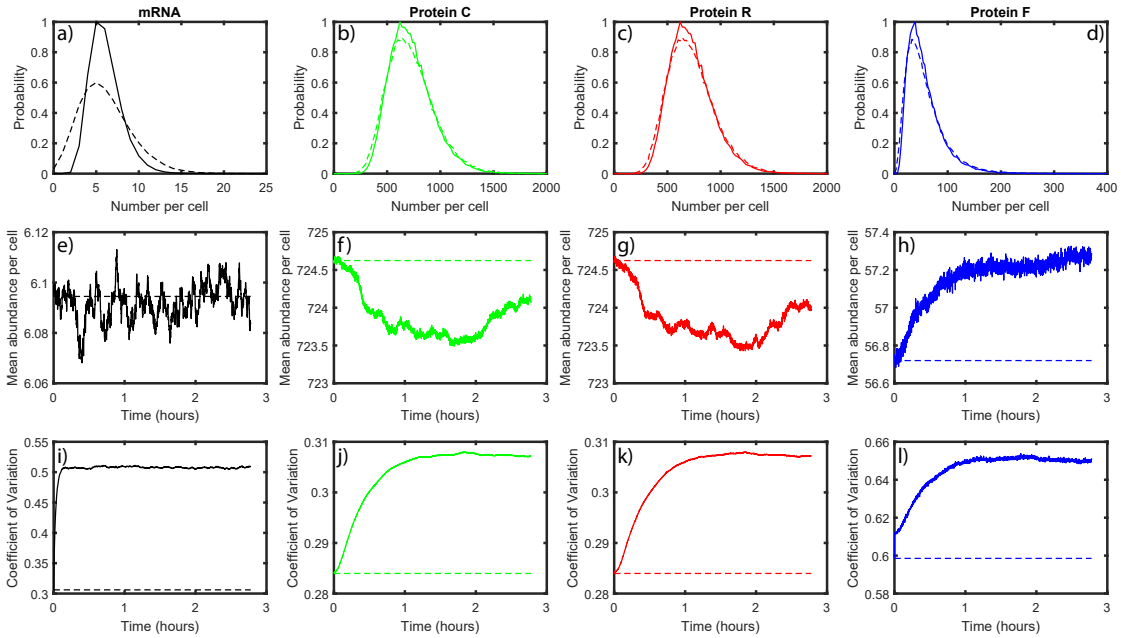


Figure 6.5: Stochastic simulation results with $\mathbf{E}[\Delta] = 0.5$, $\sigma = 0.3$ and 10^5 trials. a-d) Probability distributions of each system species. Solid lines indicate initial distributions (due to variability in Δ), and dashed lines indicate distributions after stochastic simulation has run for three hours. **e-h)** Mean abundance per cell of each species, with dashed lines indicating the value from deterministic calculations. Values for C and R are correlated because they are predominantly determined by fluctuations in mRNA levels. F increases over the first ~ 1 hour of simulation due to the introduction of stochasticity increasing variability within the variance-estimating system itself. **i-l)** Coefficient of variation for each species, with dashed lines indicating the output of the deterministic calculations.

[144]. To examine the influence of intrinsic noise introduced by stochasticity in our system we developed a stochastic model of its behaviour, following the methods outlined in Section 2.4.5 (and described in the supplementary of [65]). We separate the bulk transcription/translation process (characterised by rates $\lambda_{C,R}$) considered in the analytical analysis of our system into its individual steps, in order to explicitly simulate mRNA concentration dynamics. We anticipate large fluctuations in mRNA concentration due to its lower abundance, whereas for the higher concentration proteins the stochastic model results should converge towards the equilibrium levels predicted by analytical analysis of the differential equations [144]. The stochastic model was simulated for 10^5 trials over a range of values for σ. The estimated value of σ in the presence of stochasticity is plot in Fig. 6.3d and compared to that achieved by the analytical solution: we observe that the impact of intrinsic noise decreases rapidly as σ grows, which is anticipated since the influence of stochasticity within our system is not correlated with the variability in Δ.

Examples of the results from individual stochastic simulations are presented in Fig. 6.5 (and for other Δ distributions in the supplementary material of [65]), illustrating how the influence of stochastic noise on the probability distribution of each species' abundance decreases as σ grows.

6.7 Discussion

In this chapter we have demonstrated that a biological reporter system which returns a variable's mean value (X), as well as an output proportional to this value squared ($Y \propto X^2$), can provide an estimation of cell-to-cell variability from only bulk measurements of a cell culture's fluorescence. Though this does not return a complete distribution of reporter expression levels, for experiments in which distribution mean and width are of primary interest (such as those performed in Chapters 4 and 5) our system can eliminate the need for expensive and time-consuming single-cell measurements. It thereby prevents important behavioural properties (such as increased variability introduced by bi-modality) from being obscured as they are by traditional mean-only measurements [216, 222]. In this chapter we described one potential implementation for this system, based upon a FP-FRET reporter pair that reversibly dimerises to yield a measurable change in fluorescence. A second potential implementation, based upon an activating sRNA circuit architecture, is outlined in the supplementary material of [65]. We find that (for typical parameter values from the literature) the FRET system is able to accurately estimate the variability of gene expression from a promoter P_x from which transcription is controlled by an upstream genetic circuit. We also observe that both intrinsic and extrinsic noise have minimal influence on our system's functionality, except for the case in which the level of variability it aims to measure is very small. There are, however, a number of constraints that limit its operational range which will now be discussed.

The region for which accurate σ estimation is achieved is constrained to roughly two orders of magnitude of expression levels ($\mathrm{E}[\Delta]$) from P_x (Fig. 6.3a), a dynamic range which is likely sufficient for analysis of most synthetic biological systems. This range

can be shifted by tuning of K_{ON}; reduced K_{ON} shifts the system's ideal performance range to higher $\mathrm{E}[\Delta]$ values. For a given value of K_{ON} the feasible range is bounded below by the limitations of FRET fluorescence measurement sensitivity, as discussed in Section 6.4. In Fig. 6.3a we took this bound to be the point at which $\mathrm{E}_m[F]$ was 5% of $\mathrm{E}_m[C]$, though in practice this bound will depend on the fluorescent protein FRET pair chosen, media conditions (i.e. background fluorescence), and the particular measurement hardware employed. The upper bound on the system's operation is constrained by its departure from true multiplicative behaviour as $\mathrm{E}[\Delta]$ grows, which reduces the accuracy of σ estimation. The difference between the estimated and actual value of σ (as observed in Figs. 6.2d and 6.3) could be substantially reduced by performing calibration experiments over a range of σ values so that a more accurate expression linking σ, $\mathrm{E}_m[F]$, and $\mathrm{E}_m[C]$ could be derived and fit to this particular implementation; this would avoid the assumption that the system attains perfectly multiplicative behaviour in its output $\mathrm{E}_m[F] \propto \mathrm{E}_m[C]^2$ as is made implicitly when using (6.3).

Direct measurement of σ (which approximates CoV when $\sigma \lesssim 0.8$) requires the underlying distribution to be approximately log-normal, as is found ubiquitously in intracellular reaction dynamics [223]. This distribution arises due to the large number of noise sources which are applied multiplicatively to determine the outcome of biological processes, and thus only when one particular (non log-normal) noise source dominates a system's output is departure from log-normal behaviour anticipated [147]. We investigated this possibility for the case of bi-modality (Figs. 6.2e,f), demonstrating that though our design was not able to directly estimate the coefficient of variation for the bi-modal system, it was able to effectively rank them (i.e. gave higher outputs for distributions with greater variation). This highlights the usefulness of the measured parameter $\mathrm{E}_m[F]/\mathrm{E}_m[C]^2$, which increases in the presence of noise (even when not log-normal distributed) and allows the relative variability of two systems to be compared (if not estimated absolutely) without knowledge of the parameter γ_m.

In Fig. 6.2d γ_m was estimated using the approximation in (6.11), which holds for $\lambda_C = \lambda_R$. This equality is not necessary for our system to function, though it could be achieved in practice: both proteins are expressed from the same mRNA and are highly stable (so have approximately equal $\delta_{C,R}$, these being primarily due to dilution), and so their relative levels could be tuned by adjusting their RBSs (the RBS for C would likely need to be slightly stronger than that for R to account for transcriptional read-through of R). Regardless of whether $\lambda_C \approx \lambda_R$ is achieved, γ_m can be measured with flow cytometry and a plate-reader experiment, from which measured σ and mean fluorescence levels can be used to calculate its value using (6.3) (this process is simulated in Fig. 6.3b). Alternatively, if the approximation in (6.11) is satisfied then γ_m could be measured by expressing one fluorescent reporter from an inducible promoter and performing a titration experiment to determine K_d. Because γ_m is defined by parameters of the FRET reporter system itself, its value should remain fixed when different circuits are placed upstream of P_x, unless these were to actively interfere with the reporter binding processes (changing K_{ON} and K_{OFF}).

Stochastic simulation of our system demonstrated that its sensitivity to intrinsic noise is generally small, as anticipated for biological systems where concentrations of interacting components are large [230]. The influence of extrinsic noise was found to have substantial impact only when the variability introduced by P_x is of approximately the same level as that introduced within our system. Since our system expresses stable proteins, and noise introduced by transcriptional regulation of P_x is not considered as occurring "within" the system (this is included in the parameter Δ which we aim to measure), it represents what are the least noisy gene expression conditions obtainable without feedback [84]. Based on similar systems [84] we would anticipate the system's total internal noise level to be $\sigma \approx 0.1 \sim 0.2$, which is largely determined by transcriptional bursting [236]. The constraint (that a typical upstream regulatory system contributes significantly more noise than the expression of fluorescent reporters) is satisfied for many biological systems: Elowitz *et al.* [84] found that introducing an inducible transcription factor upstream of a fluorescent reporter increased extrinsic noise by a factor of ~ 5. Furthermore, studies of

synthetic biological circuits for which our reporter might be employed (e.g. the circuit in Section 5.2 or [51, 69]) typically yield values of σ in the range $0.5 \sim 1.5$, which coincides with the range considered in our simulations (Fig. 6.3a,b), and is above the σ levels at which intrinsic or extrinsic variability were found to impact our circuit (Figs. 6.3c,d). We thus conclude that our reporter system's functionality will be minimally disturbed by noise, with the exception of the case in which it is used to study a feedback regulatory system that is able to reduce variability of its output (i.e. whatever interacts with P_x) to below the variability internal to our reporter circuit.

As is the case for all synthetic biological systems, a range of secondary factors could disturb the behaviour of this (or any similar) variance estimating system. The dynamic range of the FRET complex is dependent upon the protein binding domains chosen and the linkers to their fluorescent markers [135], and can be improved by utilising "sticky" fluorescent protein FRET pairs [237]. For a stable cell-wide equilibrium (dependent on K_{ON} and K_{OFF}) to develop between proteins in their unbound (C, R) and FRET-active bound (F) state, it is important that each is well mixed through the cell. This could be violated if reporter proteins displayed a strong tendency to form inclusion bodies, or to be localised near a single transcription site [238] (though this would be unlikely if the system was introduced on a multi-copy plasmid). Transient interference with our reporter system by other cellular processes could also degrade its behaviour (for example, if the protein-protein binding domain used on C and R bound strongly to a native protein). Similarly, particular parameter values could be perturbed by changes in context (for example K_{ON} impacted by internal cellular crowding, or the efficiency of FRET emission varying with PH [235]), though such context-dependencies are challenges for all synthetic biological systems [22].

It is possible and likely that alternate implementations could be designed that are more simple, and provide better performance, than those proposed in this chapter. To estimate the variability of the parameter Δ the only fixed design requirement is that a system has states X and $Y \propto X^2$, both of which are measurable. Recently "sticky" FRET pairs have

been engineered by introducing weak hydrophobic interactions between the fluorescent proteins, thereby ensuring their proximity (hence boosting FRET efficiency) when they are brought together by the process that they aim to measure [237]. Similar protein engineering techniques could potentially be undertaken to further increase dimerisation affinity (but avoid homo-dimerisation), thereby eliminating the need for inclusion of protein binding domains. If variance estimation could be achieved using a single protein, then the noise introduced within the system would be reduced even further. Such a system could be conceivably built using a fluorescent reporter (X) that could form a homodimer (i.e. a complex with another copy of itself), with the dimerised state (Y) being distinguishable via bulk measurement (for example, using homo-FRET [239, 240]) from the monomeric state. Zhao *et al.* developed such a system, in which homodimerisation of a cpYFP construct results in a change in its excitation spectrum [241].

6.8 Conclusion

Mean-only measurements of biological systems frequently obscure complex behaviours, making single-cell measurements of variability critical to their accurate characterisation. In this chapter we have demonstrated that it is possible to estimate both the mean and cell-to-cell variability of a biological system's output using only measurements of a cell culture's bulk fluorescence, as might be provided by a plate reader or Turbidostat device like that described in Chapter 7. This is achieved by designing synthetic biological systems with measurable outputs that are proportional to the rate of transcription, and to this rate squared. We proposed a potential fluorescent protein FRET implementation of such a system, and demonstrated that it achieves this functionality over experimentally-relevant parameter ranges. We analysed the impact of both extrinsic and intrinsic noise, finding that they were only significant if the upstream system being measured had very low levels of variability. The results in this chapter thus support the feasibility of a variance-estimating reporter system and, along with further consideration of experimental factors, could be used to guide the *in vivo* implementation of such a system in future work.

7

Characterisation Hardware: The Turbidostat

In this chapter I outline the design and capabilities of a new "Turbidostat" experimental automation platform. An operational schematic is provided in Fig. 7.1, and a picture of the complete device is in Fig. 7.2. It can grow and measure cell cultures over long time periods, whilst regulating their temperature, aeration, and density in growth media, such that their response to time-varying inputs can be assessed. In doing so it automates many common characterisation experiments required in synthetic biology, such as those performed in Chapters 4 and 5. For the sake of brevity I provide a broad systems-level description of this device's components and capabilities, avoiding excessive detail regarding their electrical implementation. I also outline the software operating system and user-interface developed for the device, and present results from several biological characterisation experiments. The results from this chapter have not yet been published; we are currently pursuing commercial partnerships to facilitate production and dissemination of this technology to the research community.

7.1 Introduction

As described in Section 2.3.3 many synthetic circuits can take long periods (sometimes >1 day) to reach an equilibrium state. Furthermore, changes in resource availability and host cell state during growth can have a substantial impact upon their behaviour [24]. This makes the understanding of their behaviour (and its mathematical modelling) challenging. For characterisation experiments to be both reliable and reproducible it is therefore important to maintain cells for extended periods in their exponential growth phase. This can be achieved by growing cells in a bio-reactor known as a turbidostat:

such a device measures the density of cells in a growth vessel, and implements a control algorithm which pumps in fresh growth media / removes old media and cells, in order to maintain a constant optical density.

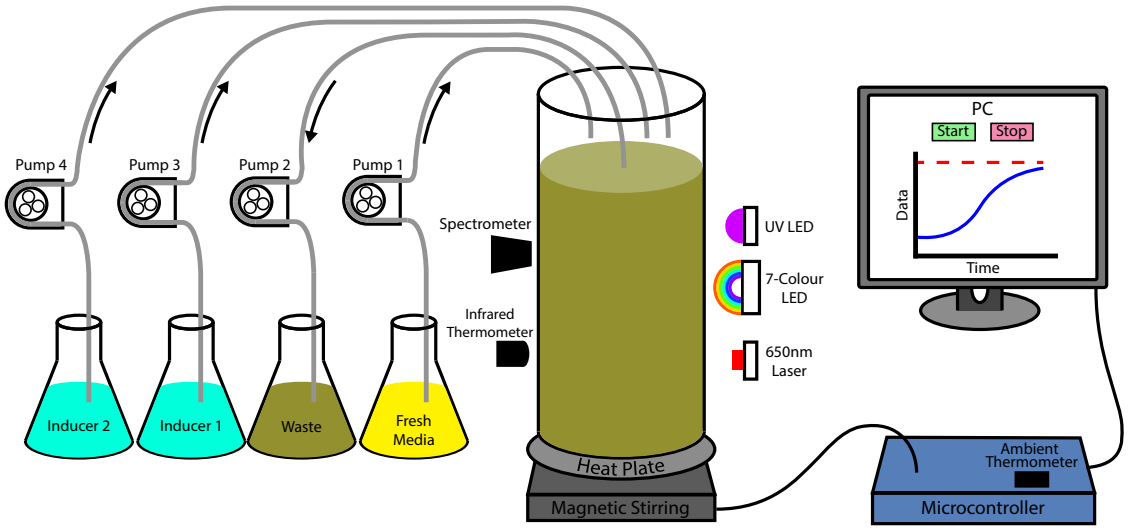


Figure 7.1: Schematic of the Turbidostat device. A PC interfaces with a microcontroller (a simple linux-based computer), which can control up to eight Turbidostat devices. Each Turbidostat consists of a central growth chamber of ~ 25 mL. Four pumps (for which pump rate and direction can be controlled independently) are included to input new / remove old growth media and cells (waste), as well as chemical inducers if required. The culture's optical density is measured by a laser and spectrometer, which can also measure fluorescence excited by a 7-colour LED. A UV LED is included to induce DNA damage to cells. A heat plate below the growth chamber is used in conjunction with an infrared thermometer to regulate the growing culture's temperature. Below this is a magnetic stirrer which can agitate the cells (via a magnetic stir bar within the growth chamber) to regulate aeration.

However, it is possible to include many additional features in such a device, which is the purpose of the current chapter: to design, manufacture, and demonstrate a Turbidostat that provides an all-in-one characterisation platform for synthetic biology. In the remainder of this section we list the capabilities and features that we desire such a device to have, both from the technical and consumer perspectives (Section 7.1.1). We also outline the broad array of research applications that these features provide (Section 7.1.2), and we compare our Turbidostat to similar devices in the literature (Section 7.1.3).

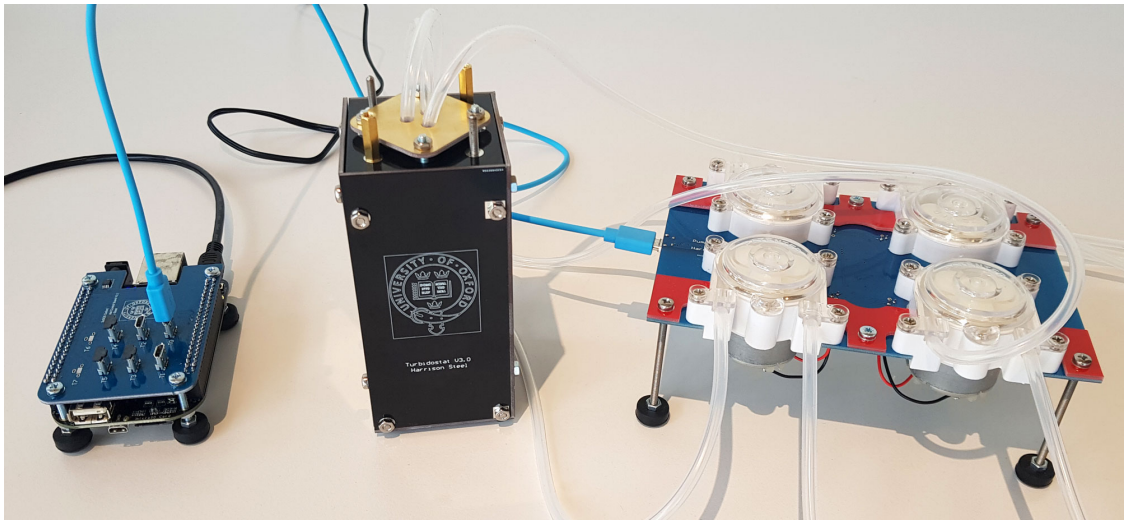


Figure 7.2: The completed Turbidostat device. The control computer (left) interfaces with the Turbidostat body (centre) which houses the culture chamber and measurement tools. The pump board (right) regulates pumping between the Turbidostat's growth chamber and waste/media reservoirs (not pictured).

7.1.1 Desired Capabilities and Features

- **Optical Density (OD) Regulation** - The core purpose of our Turbidostat is to regulate the optical density of a growing population of cells (which approximately reflects the concentration of cells in solution), which will be achieved by removing old cells and media whilst adding fresh growth media.

- **Temperature Regulation** - Cells should be maintained close to a temperature set point (generally $\sim 37°C$).

- **Provide Aeration** - Cells should be stirred, ideally at a variable speed in order to allow variation in their aeration.

- **Measure Fluorescence** - Many synthetic biological circuits use fluorescent proteins as their output, and thus our device should be able to measure the concentration of a variety of fluorescent protein targets.

- **Actuate Optogenetic Circuits** - Optogenetics provide a fast and effective method for providing inputs to synthetic biological constructs; our device should be able to provide light inputs of varying intensity and wavelength to actuate these circuits.

- **Chemical Actuation** - Additional pumps should be included to allow time-varying chemical inducer profiles to be added.

- **UV Irradiation** - An ultraviolet light source can be used to induce DNA damage in cells, thereby encouraging mutation / evolution.

- **Daylight Irradiation** - Ideally a light source with output spectrum close to that of day light would be included to regulate growth of photosynthetic species (e.g. Cyanobacteria).

7.1.2 Potential Applications

A broad range of scientific applications exist for a device with the features described above, including (but not limited to):

- **Synthetic Biological Characterisation Experiments** - This is our primary motivation for developing the Turbidostat; to provide a platform that can characterise the dynamic and steady-state behaviour of biological systems (such as those in Chapters 4 and 5).

- **Long-term Evolutionary Studies** - Our device can enable studies which grow cell populations over long periods whilst applying evolutionary pressure (for example, increasing antibiotic concentrations [242] or UV light) to encourage their adaptation.

- **Biosynthetic System Development** - Studies that develop biosynthesis systems (e.g. to produce biofuels) require tuning and optimisation of growth conditions to optimise yields and performance [243]; our platform provides an ideal bio-reactor for this.

- **Optogenetic System Characterisation** - Optogenetics have been used to actuate a range of synthetic biological systems [58], however many research groups develop individualised hardware for such experiments as there is no appropriate commercially available platform [244–246]. Our device provides an easy-to-use platform to perform these experiments.

- **Incubation/Stirring/Growth Measurement** - Our device can also be used as a simple growth measurement tool, incubator, and stirring platform. The pumps can be programmed individually to add inducers / media to a culture, allowing many simple experiments to be automated (even if the OD regulation capability is not required).

7.1.3 Comparison to Similar Devices

Due to the widespread usefulness of continuous-culture devices, a number have been presented in the literature. In Table 7.1 we review the capabilities of these, and compare them to our Turbidostat. We constrain our review to focus on devices with a working volume similar to ours (those considered operate with $5 \sim 50$ mL of media), thus excluding large industrial bioreactors (which are generally on the scale of 10s to 100s of litres). The devices with the most similarity to our Turbidostat are the inventions of Milias-Argeitis *et al* [58], Takahashi *et al* [57], and Wong *et al* [247] which we discuss briefly below.

Device	OD	Heat/Stir	Fluorescence	Optogenetics	Inducers	UV/Daylight	<£1000	GUI	Available
Ours	1	1	1	1	1	1	1	1	0.5
Milias-Argeitis [58]	1	0.5	0.5	1	0	0	0	0	0
Takahashi [57]	1	0.5	0.5	0	0	0	1	0	1
Wong [247]	1	1	0	0	0	0	0	1	0.5
Toprak [248]	1	0.5	0	0	1	0	1	0	1
Hoffmann [249]	1	0.5	0	0	0	0	1	1	1

Table 7.1: Comparison of our Turbidostat to similar devices in the literature. A 1 means the device has this capability, whilst 0 indicates its absence. A value of 0.5 means this requirement is partially satisfied. OD = Optical density regulation, Heat/Stir = In-built heating and stirring, Fluorescence = Fluorescent protein measurement, Optogenetic = Optogenetic actuation, Inducers = Capacity for automated addition of chemical inducers, UV/Daylight = Ultraviolet and Daylight light sources, <£1000 = Device can be purchased/assembled for less than £1000, GUI = An easy-to-use interface/operating system is available, Available = Device is available to buy, or schematics are available for self-build.

The device developed by Takahashi *et al* [57] is that used most widely in the synthetic biology community. It has been made available as a kit by the creators, which has led to its adoption by many laboratories [20, 211, 250]. Though in their original paper they demonstrated some ability to measure GFP fluorescence, their measurements were not particularly sensitive and hence the available version of the device does not come with this capability. Furthermore, this device does not include heating (an external incubator is required), and it uses a multiplexed pumping apparatus which is difficult to sterilise and prevents sampling from the device during operation (as positive pressure is used to expel waste media).

The device developed by Milias-Argeitis *et al* [58] represents that with capabilities closest to our device; it can heat and stir (using external hardware), regulate OD, and provide optogenetic inputs. It can also measure fluorescence, however this is done by transferring samples from the growth chamber to an external flow cytometer (hence measurements are not taken of the growth culture *in situ*). Due to the external hardware required this device is expensive and complex to construct, and consequently has not been made available or replicated outside of the creator's laboratory.

The device developed by Wong *et al* [247] represents a high-throughput parallelised variant of a basic Turbidostat platform, primarily for use in evolutionary studies. As such it is very expensive, complex, and difficult to set up, which combined make its adoption by synthetic biological researchers difficult. One advantage it has compared to our Turbidostat is a microfluidics "switchboard" that allows mixing between Turbidostat tubes; this could be crudely implemented in our device (by routing one pump between different Turbidostat tubes), but would be difficult to implement with the same degree of mixing configurability.

7.2 Turbidostat Design History

In this section we briefly review development of Versions 1.0 and 2.0 of our device, before outlining the additional capabilities that were included in Version 3.0.

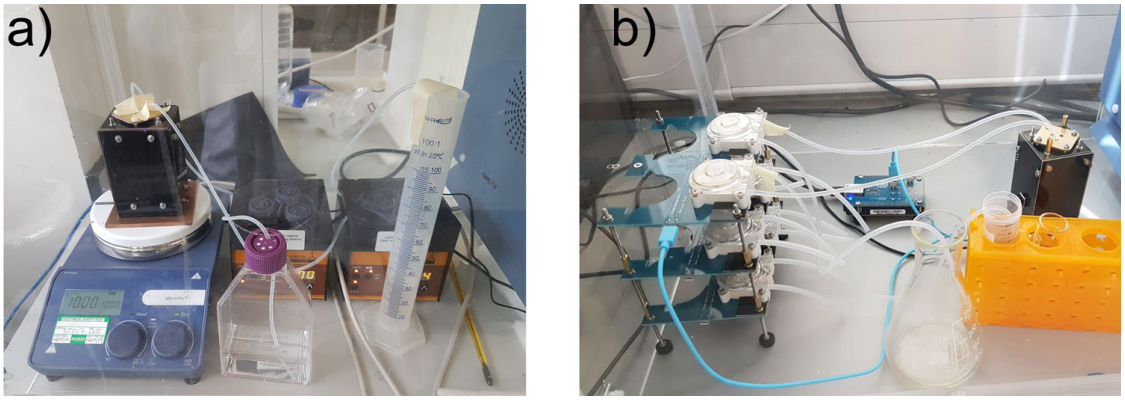


Figure 7.3: Previous versions of the Turbidostat device. a) Version 1.0 required an external incubator, stirrer, and expensive input pumps. **b)** Version 2.0 allowed devices to be multiplex from a single computer, used cheaper pumps, and included improved stirring and measurement capabilities.

7.2.1 Version 1.0

Version 1.0 (Fig. 7.3a) of our device was an initial proof-of-concept, which relied on many external instruments for operation. As such it was expensive, and had limited measurement capabilities compared to later versions. Its individual sub-systems are summarised as follows:

- **Heating:** Heating relies on either the external heat plate below the device (which interfaces with this device's control computer via a RS232 connection), or an external incubator. Temperature is regulated by control functionality either in the heat plate or incubator.
- **Pumping:** Pumping employed two expensive (∼1000 GBP) peristaltic pumps manufactured by Lambda Instruments. These interface with the control computer via an analogue voltage that determines pump rate.
- **Stirring:** Stirring is performed by the magnetic stirrer below the device.
- **Optical Outputs:** Version 1.0 included a laser, and a single low-power 450nm LED for excitation of Green Fluorescent Protein (GFP).
- **Optical Measurement:** Version 1.0 included the AS7262 6-channel spectrometer (Section 7.4.2).

7.2.2 Version 2.0

Version 2.0 (Fig. 7.3b) of our device improved upon Version 1.0 by including many additional features, allowing much of the external hardware to be removed. This version also included the ability to multiplex eight different devices, and shifted power regulation circuitry from the control board to the Turbidostat body itself.

- **Heating:** Heating was performed solely by the external incubator. Temperature is regulated such that the air temperature within the Turbidostat is maintained at 37°C.
- **Pumping:** Pumping is performed by up to four cheap (∼5 GBP) peristaltic pumps. These are controlled by a new Pump Board, which can be conveniently stacked when multiple devices are used in parallel.
- **Stirring:** Stirring is performed by a computer fan with attached magnets positioned within the device.
- **Optical Outputs:** Version 2.0 included a laser and the LZ7 7-colour LED (Section 7.3.5).
- **Optical Measurement:** Version 2.0 included the AS7262 6-channel spectrometer (Section 7.4.2).

7.2.3 Version 3.0

The final Turbidostat develops Version 2.0 further to include in-built temperature regulation, a UV light source, and more efficient power regulation circuitry. In subsequent sections we discuss each of these sub-systems in detail. In Section 7.3 the implementation of each actuating sub-system is described, before the measurement sub-systems are described in Section 7.4. Section 7.5 then describes the control algorithms that are used to tie together measurement and actuation devices to regulate various parameters of the cell culture.

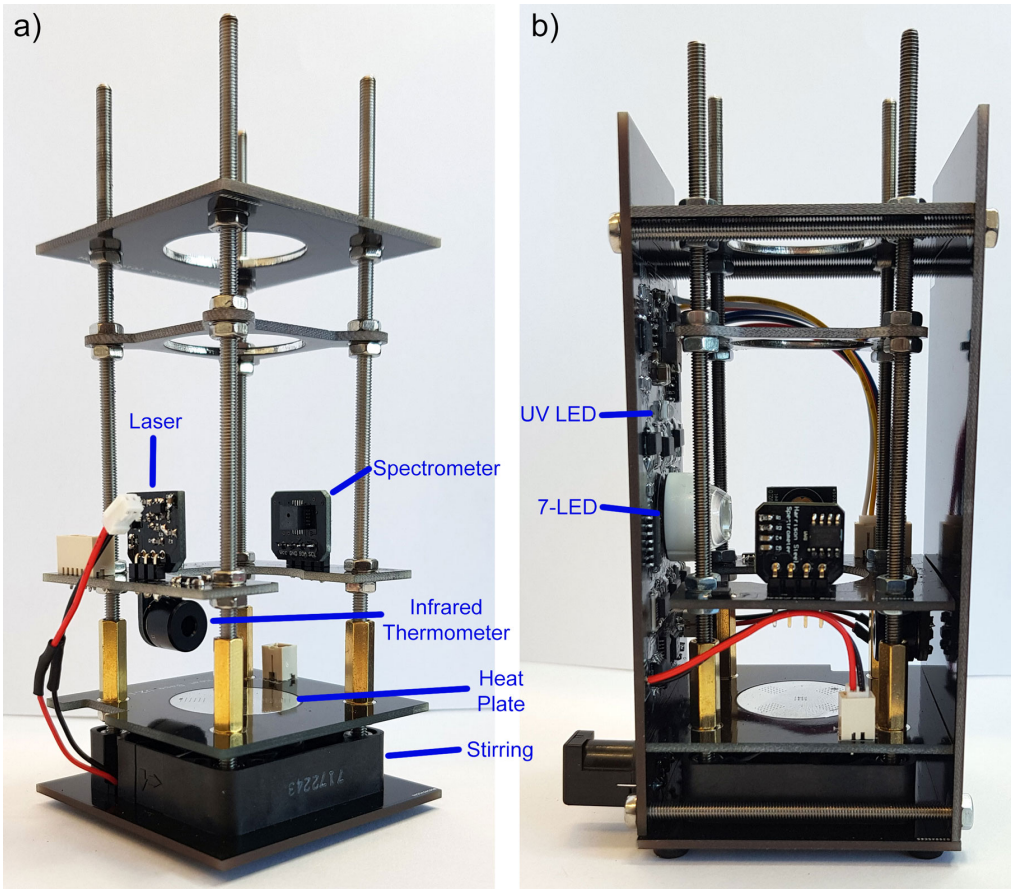


Figure 7.4: Internal components of Turbidostat Version 3.0 with **a)** all sides removed and **b)** two sides in place.

7.3 Actuation Sub-Systems

7.3.1 Heating

The simple heating element used in the Turbidostat is pictured in Fig. 7.5. Heat is generated by passing current through four 100 Ohm resistors arranged in parallel on the bottom of this PCB. There is an exposed copper pad on the upper side of the PCB, upon which the culture tube sits. Heat is transferred between the top and bottom sides of the PCBs using a large number of Vias (the copper plated holes which make electrical connection between opposite sides of a PCB) which are connected to the ground side of each resistor. Since the thermal conductivity of copper is much greater than that of fibreglass (the structural material of the PCB) most of the heat conducted away from the resistors is therefore transferred to the upper surface of the heat plate.

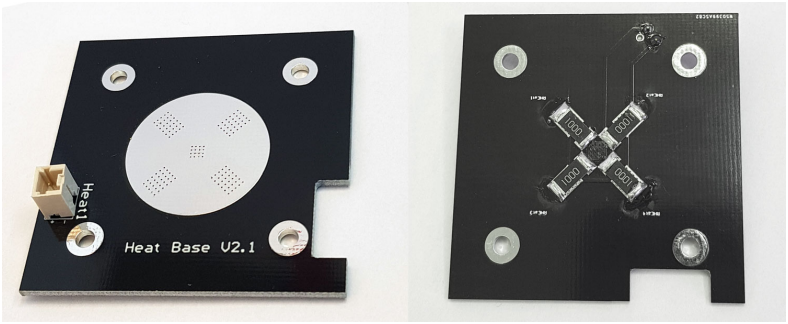


Figure 7.5: Top/bottom of the heat plate used to regulate temperature of the growth culture.

The heating element is connected to the 12V voltage rail of the Turbidostat, in series with a MOSFET transistor. This transistor is switched by a PWM (pulse width modulation) signal, which allows the rate of heat production to be varied according to the Turbidostat's needs.

7.3.2 Stirring

The magnetic stirring assembly used in the Turbidostat is pictured in Fig. 7.6. To minimise costs we employ an off-the-shelf 5V computer fan, to which we glue two neodymium magnets. When the fan is powered this assembly generates a rotating magnetic field in the bottom of the culture flask, which rotates a magnetic stir bar in the growth tube to provide aeration.

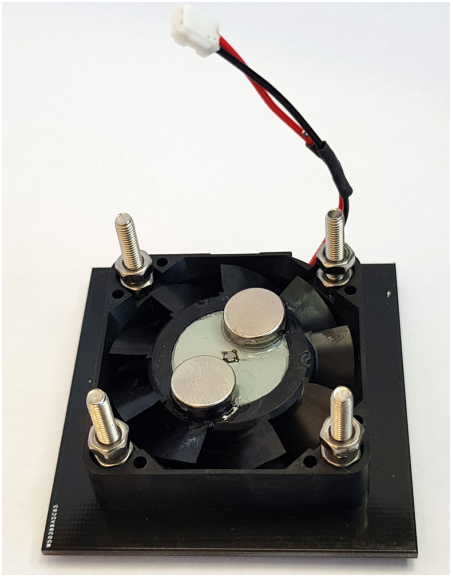


Figure 7.6: Magnetic stirring assembly used to control aeration of media.

The stirring fan is connected to the 6V power rail through a H-bridge motor driver, which is in turn controlled by a PWM signal. The duty cycle of the PWM can be used to control the fan's rotation rate.

7.3.3 Pumping

The pumping assembly used by the Turbidostat is pictured in Fig. 7.7. Each pump assembly includes four low-cost 6V peristaltic pumps (Fig. 7.7b) which employ 4.5mm/2.5mm outer/inner diameter silicone tubing. At maximum speed the flow rate through each pump is approximately 10 mL/minute, which is much faster than required by our Turbidostat. Consequently, pumps are pulsed for a short time period each 30 seconds in order to achieve much lower pump rates. One weakness of these low-cost peristaltic pumps is that pump rates are unpredictable, with different pumps potentially rotating at different speeds when the same input voltage is applied. Another source of variability arises from the fact that occasionally a pump will stop rotating in a location where the tubing is not completely pinched off. This allows any liquid within the tubing to slowly drain out under gravity. This source of error can be overcome by our control algorithm, as described in Section 7.5.

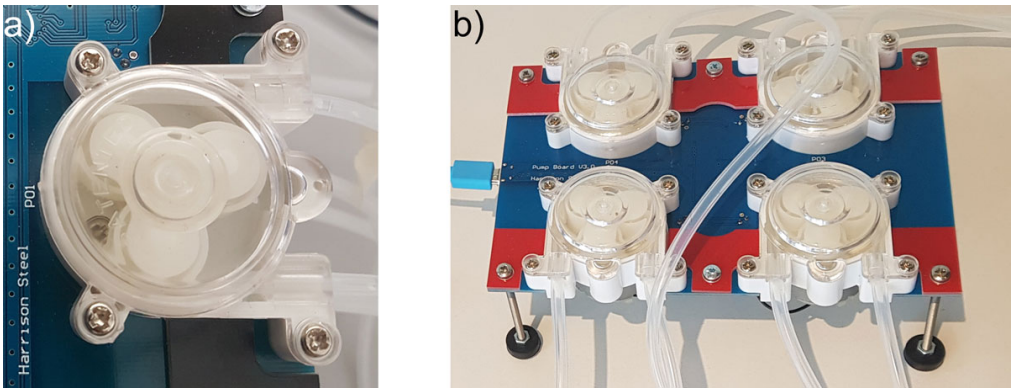


Figure 7.7: Pumps used in Turbidostat V3.0 a) A single, low-cost peristaltic pump. **b)** A pump assembly, which provides speed and directional control for four pumps.

Each pump is driven by a similar circuit to the magnetic stirring fan: power is provided by the 6V rail through a H-bridge motor driver, which is in turn controlled by a PWM signal. The duty cycle of the PWM can be used to control the pump's rotation rate.

7.3.4 OD Laser

The laser used by the Turbidostat to measure a culture's optical density is pictured in Fig. 7.8. The particular laser diode emits at 650nm, and includes a photodiode within its housing for use in power output regulation. No focusing optics are used with the laser (meaning it has a large beam divergence of $> 10°$ in each direction) as a diffuse beam results in reduced variability at the detector. The OD laser is oriented at 180° to the spectrometer (Fig. 7.4).

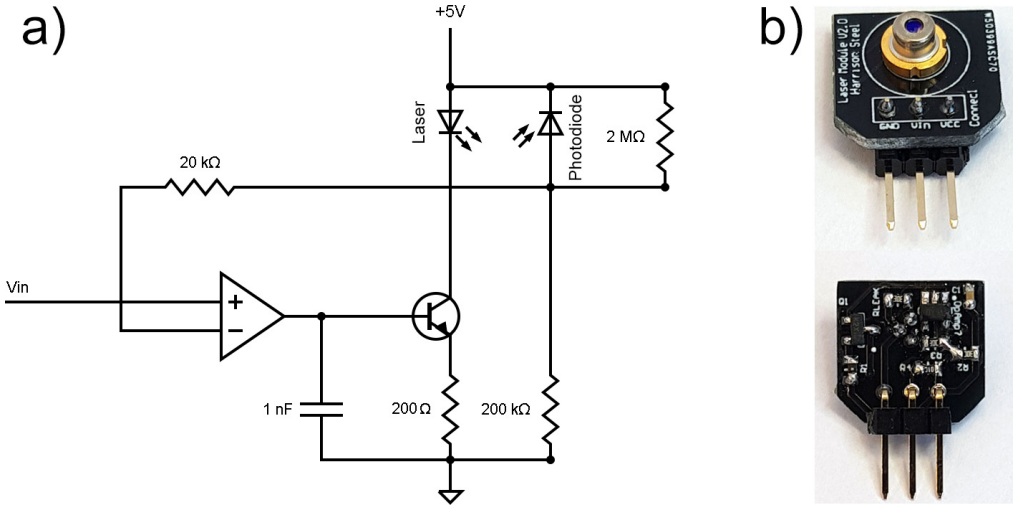


Figure 7.8: OD laser and electrical implementation. a) Diagram of laser driver circuit, which regulates emitted optical power to minimise output variation. **b)** Implementation of laser and driver circuitry in panel a.

The laser is driven by an optical power regulation circuit (Fig. 7.8a) which uses an OpAmp to set the voltage across a switching transistor such that a fixed current is generated by the laser diode's inbuilt photodiode. A 200 Ω resistor is used to limit the maximum current through the laser diode, and a 2 MΩ bypass resistor is placed around the photodiode to account for the OpAmp's voltage offset. We use this circuit (rather than regulating the current through the laser diode as in Fig. 7.9a) in order to minimise the dependence of the laser's output power on temperature. Similar regulatory architectures were implemented external to a light source (i.e. they did not use a laser's inbuilt photodiode) in past OD measurement tools [57, 58].

7.3.5 Visible Light LED

The visible light LED used by the Turbidostat to measure fluorescence is pictured in Fig. 7.9. The particular diode employed is the LZ7 emitter developed by LED Engin [251]. This LED has seven outputs of different colours, for which the wavelengths are summarised in Table 7.2. A plastic lens (present in Fig. 7.4b, not pictured in Fig. 7.9b) is glued to the front of the LED to focus its output. The LED is oriented at 90° to the spectrometer (Fig. 7.4b) to improve the ratio of emission signal (light emitted by fluorescent proteins, which is uniform in direction) to excitation signal (emitted by the LED, which is minimal at 90° to the beam).

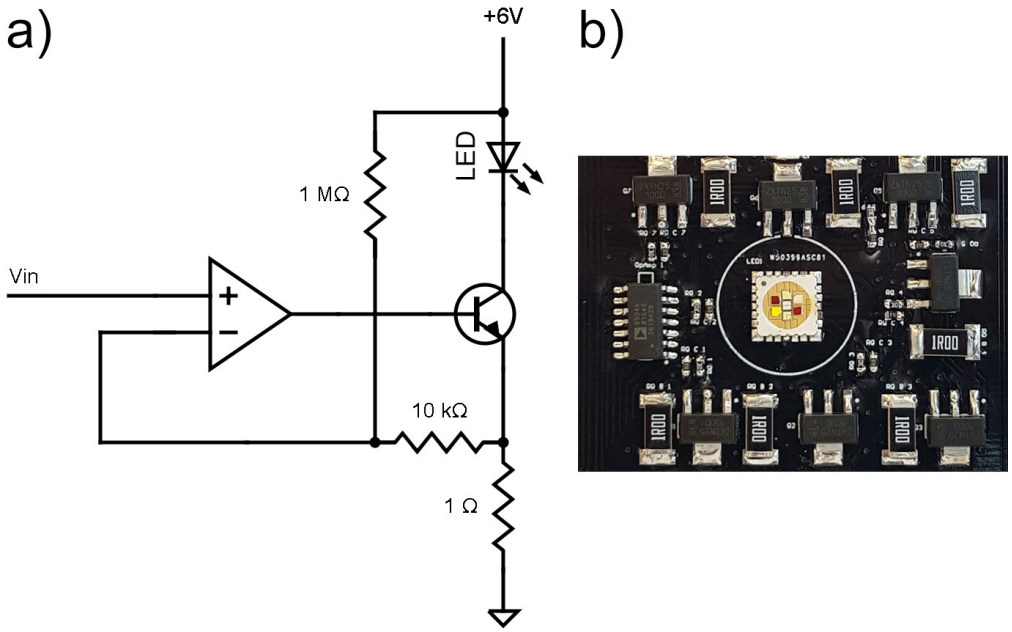


Figure 7.9: Visible LEDs and electrical implementation. a) Diagram of driver circuit used to regulate current through each LED. **b)** Implementation of LED and surrounding driver circuitry.

Each LED is driven by a current regulation circuit (Fig. 7.9a). This circuit uses an OpAmp and transistor to regulate the current through the LED similarly to Fig. 7.8a, but in this case the feedback signal (which enters the negative terminal of the OpAmp) is the current through the LED, rather than a photodiode output.

Wavelength (nm)	FWHM (nm)
395	30
457	35
497	55
525	70
595	25
625	30
6500K	N/A

Table 7.2: LED Emission peak wavelengths and spectral width (Full Width at Half Maximum, FWHM).

7.3.6 Ultra-Violet Light LED

The UV LED emits at peak wavelength of 280nm with FWHM of 12nm, which is optimised to maximise bacterial killing potential. The emitter has a viewing angle of approximately 120°; we do not employ any additional focusing optics as we desire uniform illumination of the growth media. The particular LED used is the XR UV Emitter developed by RayVio [252]. The UV LED employs a similar control circuitry to Fig. 7.9, though the device is powered from the 12V rail such that its typical forward voltage of $\sim$ 7V can be satisfied.

7.4 Measurement Sub-Systems

7.4.1 Thermometry

Temperature of the cell culture is measured by a medical grade infrared thermometer (Fig. 7.10) developed by Melexis [253]. With a stable voltage supply this device is able to remotely estimate temperatures near 37°C with an accuracy of 0.1°C. The device includes in-built thermal calibration (such that changes in ambient temperature do not affect its reading), and its restricted field of view ensures that the measured temperature reflects that of the media (and not its surroundings).

In addition to the infrared thermometer, there are two air temperature thermometers to monitor the device; one inside the Turbidostat and one on the control computer (which measures room temperature).

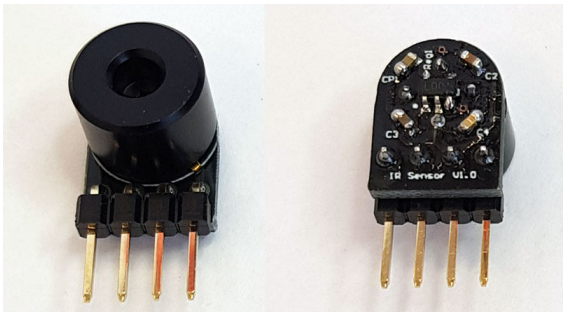


Figure 7.10: Infrared Thermometer assembly, including voltage regulation circuitry.

7.4.2 Spectrometry

Measurement of light intensity at various wavelengths is performed by a the AS7262 chip spectrometer developed by AMS [254] (Fig. 7.11). This device consists of six photodiodes with varying optical filters (the properties of which are summarised in Table 7.3), as well as amplification, filtering, and temperature correction circuitry. During production each device is calibrated to standardise measurements across each wavelength channel. An operating system is stored on an external memory chip, which can be programmed to take measurements with variable gain and integration time. In general we integrate each measurement for 0.5 seconds, after which the device returns a 16 bit floating point value that reflects the mean light intensity during that period.

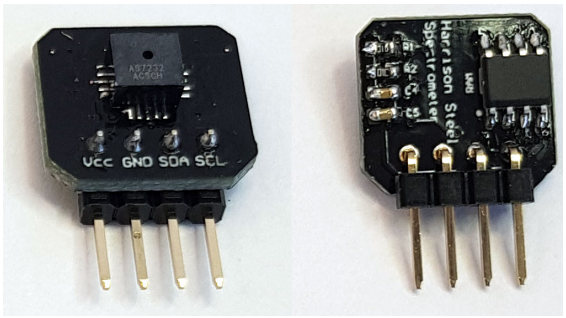


Figure 7.11: Spectrometer assembly, including flash memory chip.

7.5 Control Implementation

7.5.1 Temperature Regulation

Temperature regulation is implemented by a control algorithm which measures the culture's temperature using the infrared thermometer (Section 7.4.1), and actuates temperature

Wavelength (nm)	FWHM (nm)
450	45
500	45
550	45
570	45
600	45
650	45

Table 7.3: Spectrometer measurement band peak wavelengths and width (Full Width Half Maximum, FWHM) .

change using the heat plate (Section 7.3.1). Actuation is uni-directional; whilst we can actively increase system temperature, if the temperature is above the set-point we must wait for it to cool via heat transfer with the ambient environment. The control algorithm is implemented in Python, from which it is able to digitally communicate with each of these two sub-systems.

We employ a two-mode controller, depending on the difference (e) between the measured (T) and set-point (T_0) temperatures (so $e = T - T_0$). The two modes are as follows:

- **Mode 1:** $|e| > 1.5$ **°C.** We employ a proportional control law when $e < 0$, and set the heating output to zero when $e > 0$.

- **Mode 2:** $|e| \leq 1.5$ **°C.** We employ a proportional and integral (PI) controller, in which the proportional gain is much lower than in Mode 1. Furthermore, the integral term is reset to zero whenever Mode 1 is entered.

This two-mode controller is motivated by our strong aversion to temperature overshoot (which could kill cells) when the culture is being heated from below. Avoiding overshoot in this way comes at the expense of an increased response time during initial culture heating, but this is not of great significance as we often will need to wait significantly longer for the biological system to settle before commencing an experiment. We do not include a derivative term in Mode 2 due to potential disturbance by measurement noise.

7.5.2 OD Regulation

Cell density/OD regulation is implemented by a control algorithm which measures the culture's optical density using the laser and spectrometer, and actuates OD change (dilution) using the pumping system. Actuation is again uni-directional; whilst we can actively decrease the OD (dilution by pumping in fresh media), if OD is below the set-point we must wait for cells to grow for it to increase. The control algorithm is implemented in Python, from which it is able to digitally communicate with each of the hardware subsystems.

To determine the input (fresh media) pump rate we employ a PII (Proportional Integral Integral) controller which acts upon the difference (e) between the measured (OD) and target (OD_0) optical densities (so $e = OD - OD_0$). The PII controller terms have gains of G_p, G_{I1}, and G_{I2} respectively, where $G_{I2} \gg G_{I1}$. The P and first I term are primarily responsible for OD regulation near to OD_0. The second integrator is set to zero any time that $OD < OD_0$; its purpose is to account for rare situations (described in Section 7.3.3) in which the input peristaltic pump forms a poor seal on its tubing, which can lead to multiple pump cycles in which no media is delivered. The output (waste) pump rate is generally set to twice the input pump rate to guarantee a much greater rate of media removal, ensuring the volume of media in the growth chamber does not rise.

7.6 Electrical Implementation

In this section we briefly summarise some important aspects of the electrical implementation of the functions described in previous sections. Fig. 7.12 provides a schematic overview of the different PCBs in the device (their assembled form is pictured in Fig. 7.2). We now discuss each individually.

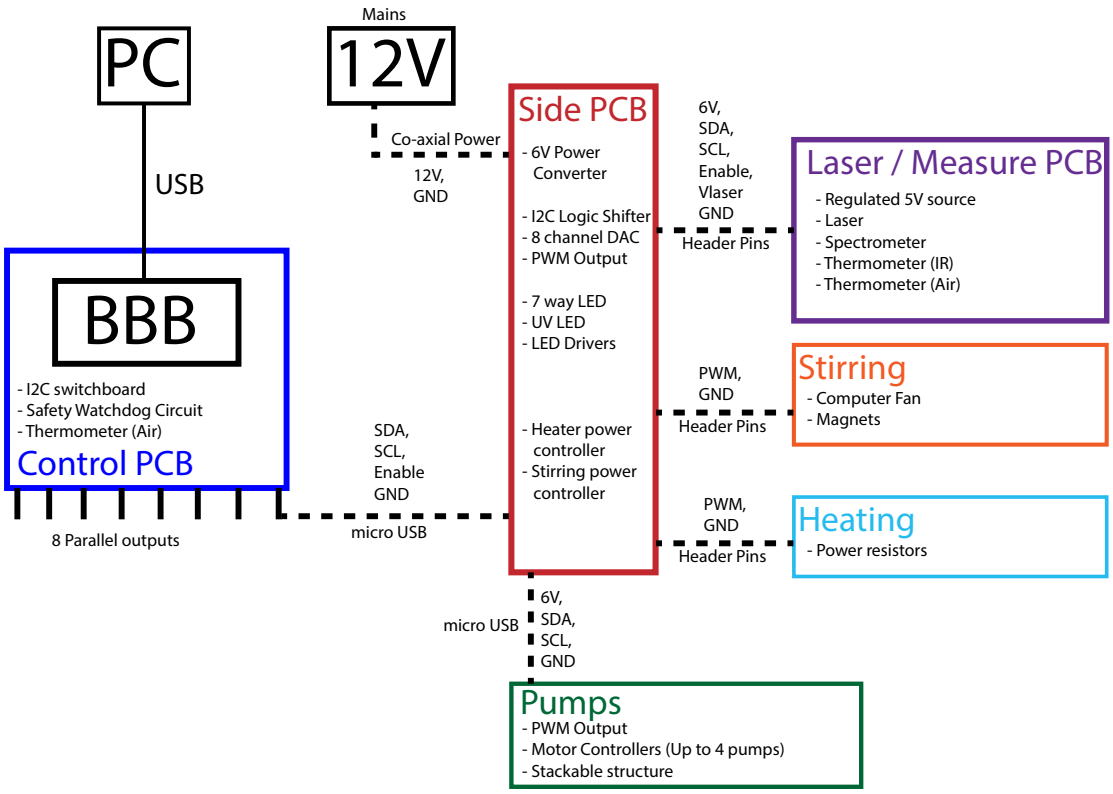


Figure 7.12: Electrical schematic of the complete Turbidostat device. Dashed lines indicate connections between different PCBs (Printed Circuit Boards). Each connection is labelled with the physical form of the connection (e.g. micro USB cable), as well as the electrical signals that it conveys.

Control Computer

The control computer is made of two components, an off-the-shelf Beaglebone Black computer, and the Control PCB (of my own design and construction).

The Beaglebone Black [255] is a small, low-cost linux computer that is widely used for rapid prototyping of electrical systems. In our implementation it is powered via a USB cable from a companion PC; the USB connection is also used to transfer data to the PC during an experiment. The Beaglebone provides digital IO ports, PWM outputs, and analogue voltage measurement. That said, in our case we only use a single digital output pin, and one of its I^2C busses. I^2C is a two-wire digital communication protocol, commonly used in embedded electrical systems, which is responsible for transferring information between our control computer and each of the digital components within the

Turbidostat (Thermometers, Spectrometer, etc).

Connected to the Beaglebone Black is a Control PCB of our design, which is responsible for multiplexing digital I^2C signals between each of the connected devices. This is required because each identical component (e.g. each Thermometer) has the same I^2C address, and so we must connect and communicate with one device's bus at a time. Also on this PCB is a digital thermometer to measure the room air temperature, and a safety watchdog circuit which is responsible for ensuring all connected devices switch off if there is a malfunction or bug in the operating software. This functionality is implemented using a digital timer and Q-latch on the PCB; simply put, if a digital IO pin on the Beaglebone Black isn't flipped each 0.5 seconds (from off-on or vice versa) then the "Enable" voltage will go low, which disables the power regulation circuitry on all connected PCBs. Since this digital pin will only flip when the Turbidostat's software is running, this mechanism allows all connected electronics to be disabled if a hardware error is detected or a software malfunction occurs.

Side PCB

The Side PCB (Fig. 7.13a) hosts many of the Turbidostat body's important regulatory circuits, and ties together the other components of the device with USB connections to the Control PCB, Pumps, and power input from the 12V mains converter.

A 12V 36W mains source is chosen as a convenient intermediate power level for input to our device. The Side PCB contains a switching power converter to 6V which can provide up to 4A, though during normal operation our device will not draw more than 1A. The digital electronics require a mixture of 3.3V and 5V logic, and thus we also include a logic level shifter (and low-power 5V and 3.3V sources) on this PCB. Also on this PCB is an 8 channel 12-bit DAC, used to control the output intensity of the 7 colour LEDs and the Laser diode (on the Laser/Measure PCB), and a PWM source which is used to

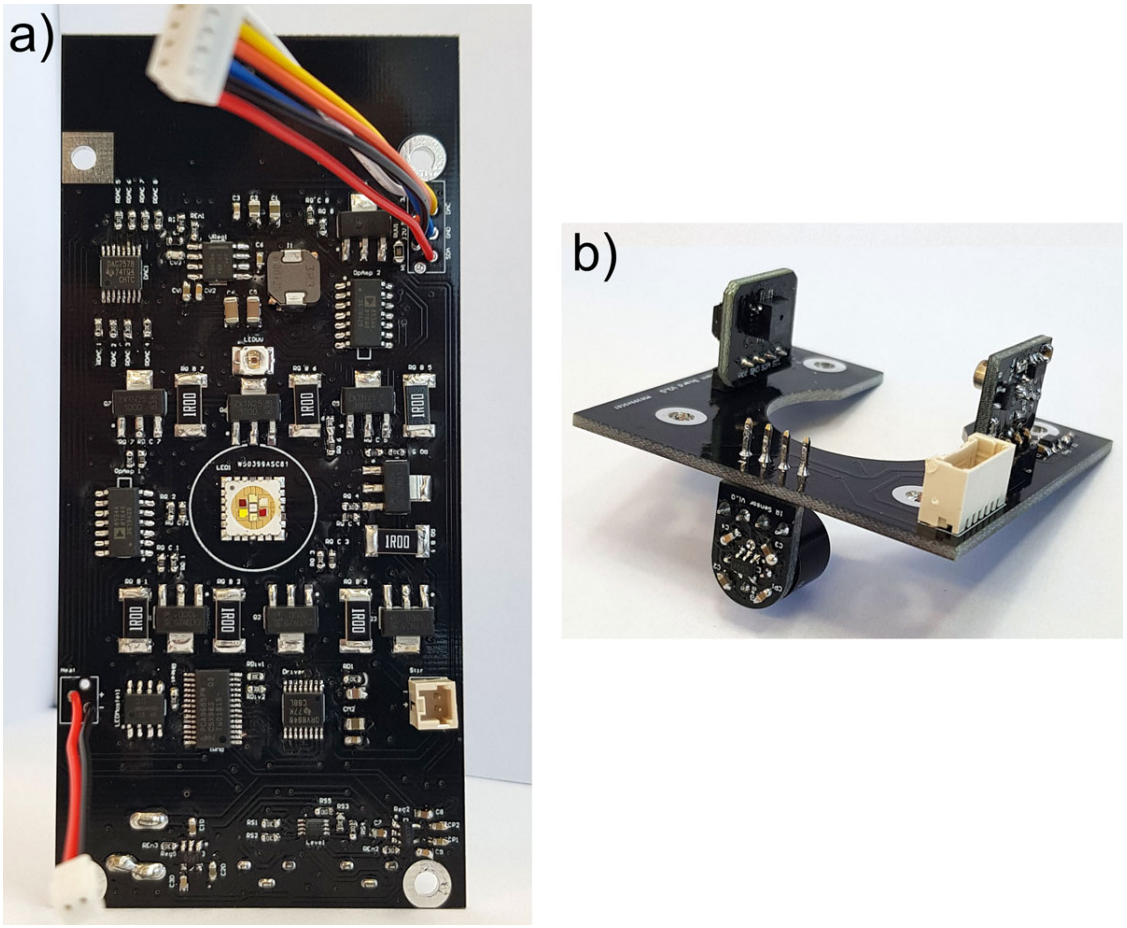


Figure 7.13: Internal Turbidostat circuitry. a) The Side PCB, without LED lens attached. **b)** The Laser/Measure PCB, with Laser, Spectrometer, and IR Thermometer attached.

regulate rates of stirring, heating, and UV LED intensity.

The 7-Colour LED is positioned in the middle of this PCB (Fig. 7.9), with sufficient free space around it for a lens to be glued to its front. The much smaller UV LED is then positioned above this (Fig. 7.2).

Laser/Measure PCB

The Laser/Measure PCB (Fig. 7.13b) is positioned horizontally within the Turbidostat's body (Fig. 7.2), and is primarily responsible for providing a mount for the Laser, Spectrometer, and Infrared Thermometer. Additionally, the Laser/Measure board contains an air temperature sensor, and a dedicated low-noise 5V source from which the laser is powered.

Pump Board

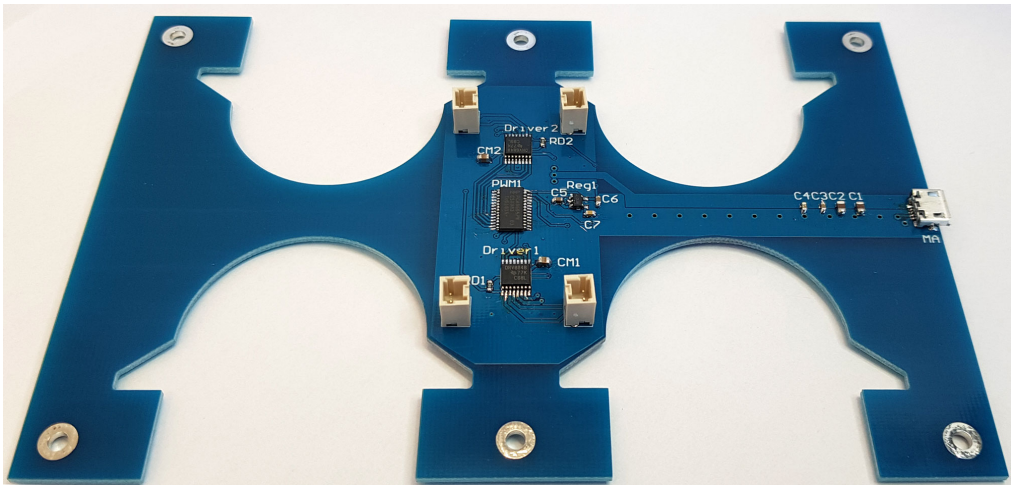


Figure 7.14: Control circuitry on the reverse of the Pump PCB.

The PCB dedicated to pump control is pictured dis-assembled in Fig. 7.14, and with pumps attached in Fig. 7.7. It is powered from the 6V source on the Side PCB, and contains power filters, a digital PWM output circuit, and two motor drivers. This allows up to four pumps to be driven, with a reversible voltage of up to 5.8V and 1.0A provided to each. In practice different pumps (peristaltic or otherwise) with similar supply voltage/power requirements can be attached to the outputs on this board.

7.7 Software

Developing a capable and easy-to-use operating system for our device is of great importance, not just for our own experiments, but also to encourage its adoption by other researchers. To achieve this we developed a web-based operating system, which provides the user with a convenient and intuitive GUI accessible via a web browser from a PC (connected via USB to the Beaglebone Black). Many control functions are automated by this operating system, which we now describe in detail.

Web Interface

The Turbidostat's web-based interface (Fig. 7.15) is implemented in HTML and Javascript, and is run as a web server from the Beaglebone Black using Gunicorn and Flask [256]. The web interface allows each connected device to be observed and controlled one at a

time (using the switchboard in the top-left of Fig. 7.15). The various system parameters (such as LED intensities and pump rates) can be set manually in this software: a value is typed into the "New" box corresponding to the particular function, and then the "Set" button will set the output level to this new value. The particular function or component can then be activated by toggling the corresponding "Switch" button, which will leave it enabled until pressed again (or until it is disabled in an automated software function). Temperature and spectrometer measurements can be performed manually if desired by clicking their corresponding "Measure" button. To measure a fluorescent protein the excitation and measurement wavelength bands must be set (bottom of Fig. 7.15), and then the particular measurement toggled into its "Active" state. When the Turbidostat is running, measured data is plot in real-time in the lower half of this web interface.

Back-end Implementation

Running behind the web interface is a Python script which is responsible for communicating with each of the Turbidostat sub-systems, implementing the control functionality described in previous sections, and passing data to/from the web interface. To achieve this a function library was written which translates user-given commands (such as output power/speed) into digital signals directed toward the appropriate component on the I^2C bus. When a command is to be sent to a particular device, a global flag is set to true (indicating that the I^2C bus is in use), and the I^2C multiplexer chip is directed to connect the control computer to the appropriate device's bus. Once the digital communication is complete this flag is released, allowing the next command to be sent (potentially to a different connected Turbidostat). Most of the Python operating system is run as separate threads, such that control processes can run simultaneously for up to eight devices in parallel. The operating system includes a background process which is responsible for pulsing one of the Beaglebone Black's digital output pins each 0.5 seconds; if this ceases then the watchdog circuitry connected to the Control Computer cuts power to all Turbidostats as a safety precaution.

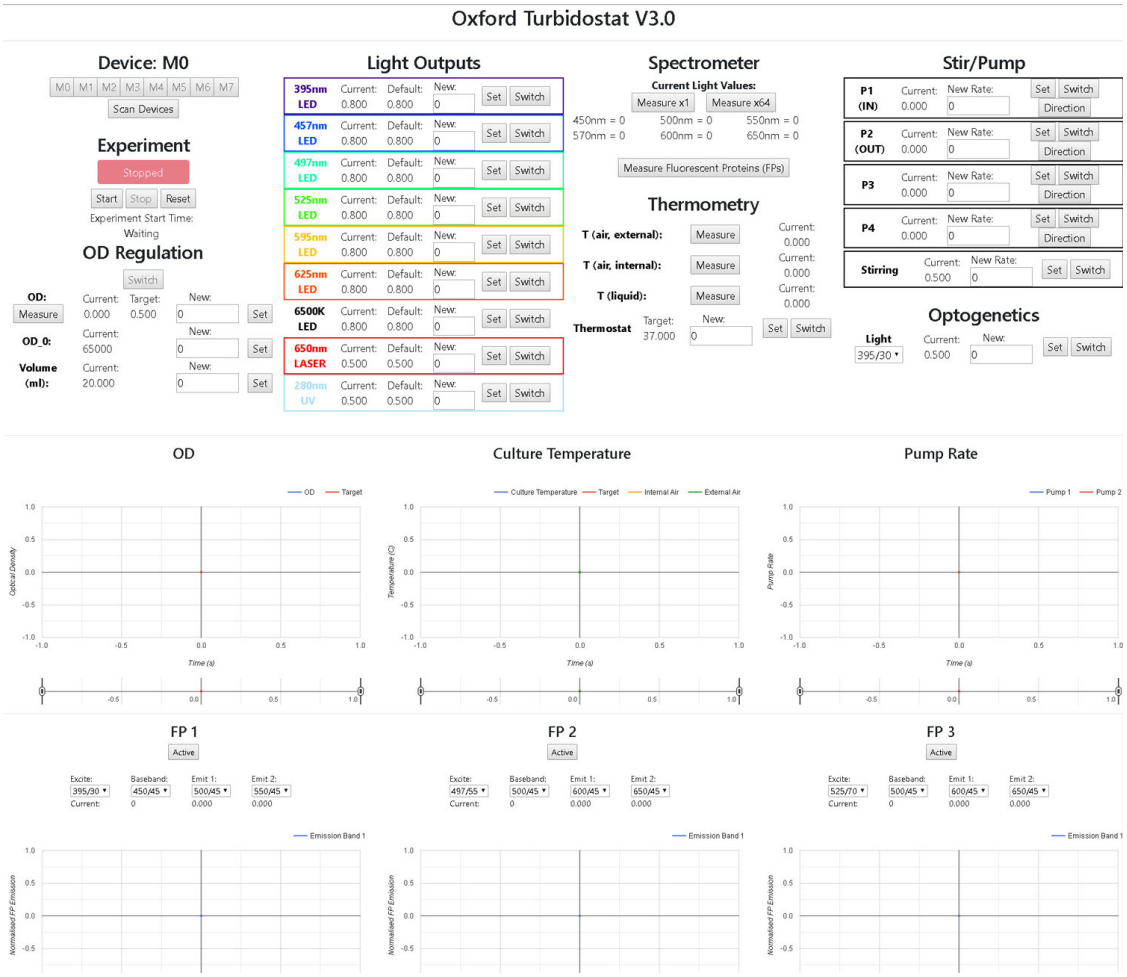


Figure 7.15: Web interface for Turbidostat control, which is accessed from a web browser on a connected PC.

Typical Operation Procedure

In this section we outline the typical procedure followed by our device when run in its automated experiment mode. An "Experiment" runs in a 1 minute-long cycle during which the culture is heated, stirred, measured, and data is processed and written to a file. This does not necessarily include OD regulation (i.e. pumping of fresh media), which must be enabled separately if desired. This procedure can be summarised as follows:

1. Click Start Experiment; this commences a new experiment which is named according to the time, date, and device upon which it is run.

2. All light sources, pumps, and stirring are disabled to allow the cells to settle. Wait 15 seconds.

3. Measure OD, any fluorescent protein measurements that are set "Active", and temperatures.

4. Calculate new pump rates required for OD regulation (if OD regulation function is enabled), or leave pumps off.

5. Switch stirring and pumps on at desired speeds.

6. Enable UV/Optogentic outputs if they have been set.

7. Write all measurement data, pump/stirring rates, and system parameters to a data file.

8. Wait until one minute has passed since Step 2. Repeat from Step 2.

An experiment continues until the Stop Experiment button is pressed. Following this an experiment can be re-started by clicking Start, or the Reset button can be used to clear all data (for this particular Turbidostat) to begin a fresh experiment.

7.8 Experimental Testing and Validation

In this section we present a selection of the validation experiments performed thus far in the Turbidostat, demonstrating a range of its capabilities. The biological constructs characterised were developed by collaborators, and so they will not be described in detail.

OD and Temperature Regulation

For this experiment (Fig. 7.16) we grew *E. coli* in rich EZ media to assess the Turbidostat's ability to regulate temperature, provide aeration, and maintain OD. We start the culture at an OD of ~ 0.1 and observe rapid growth until the target OD of 0.4 is reached in $\sim$ 1 hour. This OD is then maintained for ~ 13 hours. The temperature is maintained within 0.5°C of its target value (37°C), even when fresh media (which is at the room temperature, equal to the "External Air" temperature) is being added to dilute cells.

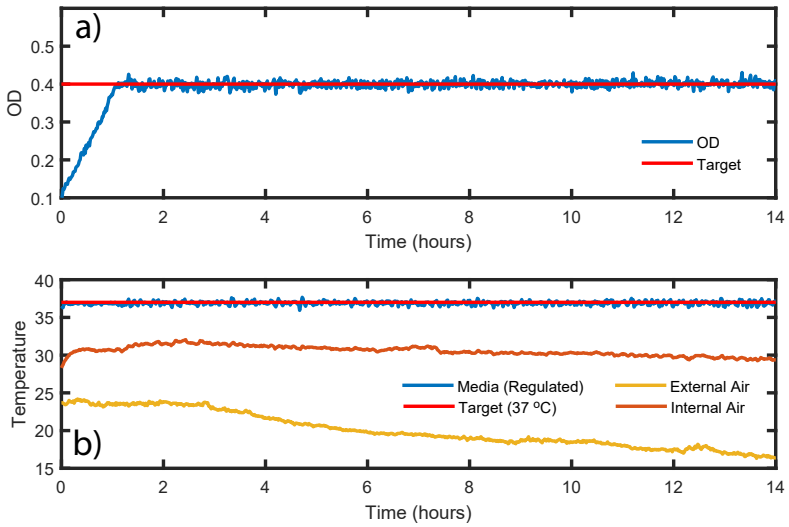


Figure 7.16: Example of OD (cell density) and media temperature regulation. *E. coli* is grown in rich EZ media and **a)** Optical density is maintained at 0.4 and **b)** Media temperature is regulated to maintain 37°C. Also plot is the external air temperature (which varies due to air conditioning/sunlight in the room) and internal device air temperature (higher than the external temperature due to heat generated by the Turbidostat's electronics).

Chemical Induction

For this experiment (Fig. 7.17) we grow standard *E. coli* in rich EZ media with the sRNA circuit described in Section 4.7 to investigate its response time following chemical induction. We grow the cells to an OD of 0.45, which we then regulate for the remainder of the experiment. After $\sim$ 4.5 hours we manually induce the circuit, and observe fluorescence begin to rise within approximately $\sim$ 30 minutes.

Characterising a RecA Promoter

For this experiment we characterise *E. coli* in minimal M9 media with genetic circuits which consist of GFP under the control of the RecA promoter (this circuit was created by our collaborator Xiaoyu Chen). RecA is a transcription factor that forms part of the cell's SOS response, which becomes active in response to DNA damage. In these experiments we induce DNA damage using our Turbidostat's UV light source, which we switch on for varying time periods.

In Fig. 7.18 we provide experimental data for one such construct. Initially we dilute the OD down to $\sim$ 0.6, and then allow cells five hours to exit stationary phase (Fig.

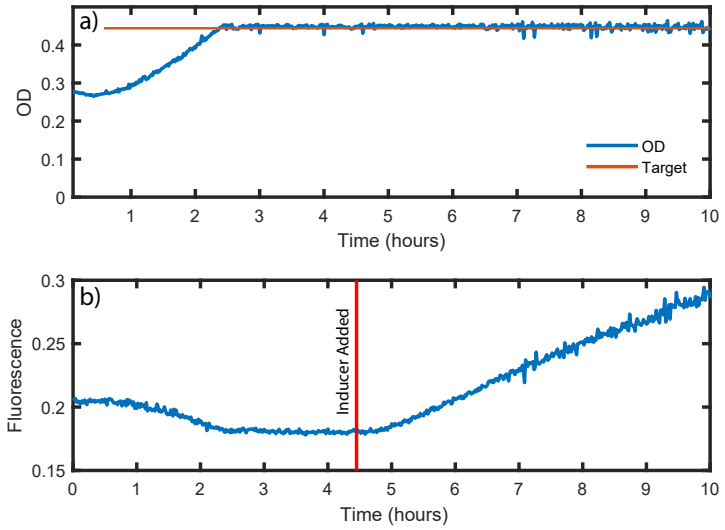


Figure 7.17: Test measurement of the sRNA circuit described in Section 4.7. **a)** Optical density, which is regulated to 0.45 throughout the experiment. **b)** Fluorescence, defined as measured light intensity at 500nm (emission wavelength) divided by light intensity at 450nm (excitation wavelength) when the cells are excited by the 395nm LED. We manually induce the circuit after $\sim$ 4.5 hours with 0.01 mg/mL of L-rhamnose and 0.1 mM m-toluic acid.

7.18b). During this period we observe the fluorescence per cell remains constant (Fig. 7.18d). After $\sim$ 5 hours have passed we give the cell a 1X dose of UV (which is equivalent to activating the UV LED at 50% power for 30 seconds), and begin regulating OD to a target of 0.7. After $\sim$ 20 minutes we observe that the fluorescence begins to rise because the SOS response (and in particular its transcription factor RecA) has been activated in response to DNA damage. After $\sim$ 7 hours we dose the cells with a UV dose 10X larger than the first. This initially leads to a rapid rise in fluorescence. However, as observed in Fig. 7.18c it also leads to a cessation of growth, and no increase in GFP expression (or OD rise) is observed beyond hour $\sim$ 12 of the experiment.

In Fig. 7.19 we present fluorescence data for two other genetic constructs that were characterised simultaneously in this experiment. No fluorescence is observed in the "Amplifier" construct. Sequencing revealed that there was a mutation in the plasmid, and thus this measurement serves as a convenient control baseline. This data highlights the utility of our instrument as a characterisation device; we can perform identical experiments on differing constructs and precisely assess their differences (in this case, the Amplifier + TetR construct has a greater degree of leakiness).

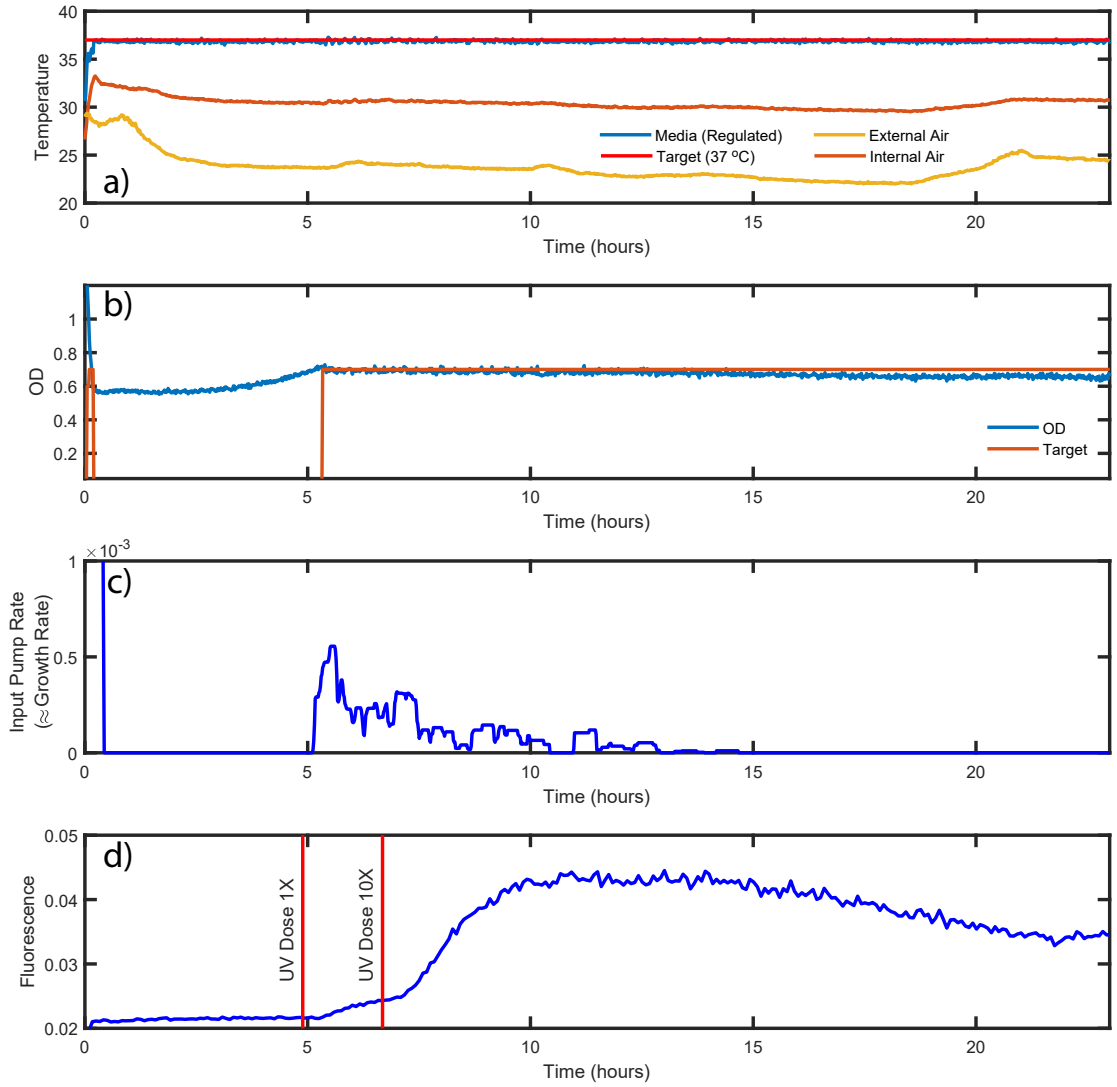


Figure 7.18: Turbidostat experiment performed in minimal M9 media with a synthetic circuit which produces GFP in response to UV-induced DNA damage. At ~ 5 hours the UV source is activated at 50% intensity for 30 seconds. At ~ 7 hours it is activated at 50% intensity for 5 minutes. **a)** The media temperature is regulated to 37°C throughout. **b)** OD of cells over the course of the experiment. We dilute the starting culture from an OD of ~ 1.2 to 0.6, and then allow them to begin growth. From ~ 5 hours onward we regulate OD to 0.7. **c)** The input pump rate required to maintain OD (which is hence zero prior to activation of OD regulation). This can be used as an approximate measure of cellular growth rate; we observe that growth slows following each UV dose, and eventually stops. **d)** Culture fluorescence, defined as measured light intensity at 550nm (emission wavelength) divided by light intensity at 450nm (approximate excitation wavelength) when the cells are excited by the 457nm LED.

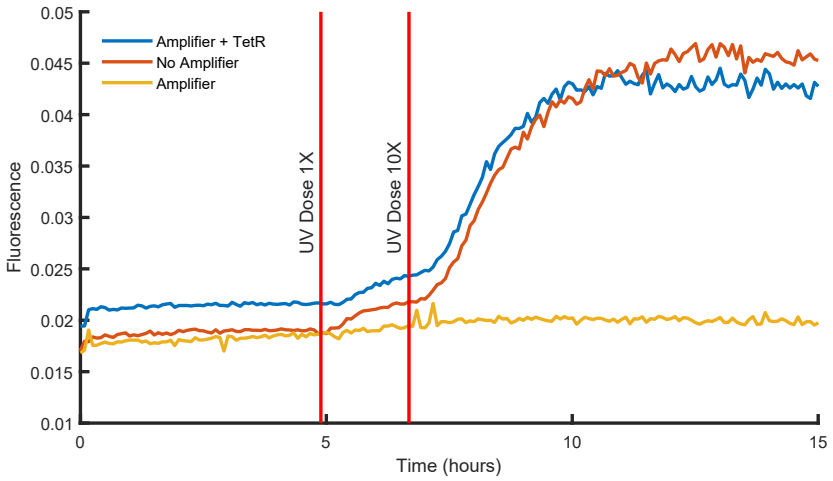


Figure 7.19: Measurements of fluorescence for experiments performed as in Fig. 7.18 for three constructs. The "No Amplifier" case is the RecA promoter immediately in front of GFP. The "Amplifier' case positions a transcription-factor based amplifier [37] downstream of the RecA promoter (though in this experiment the plasmid was found to be incorrect). The "Amplifier + TetR" case (that in Fig. 7.18) additionally includes a negative autoregulatory TetR structure.

Characterising an Optogenetic Transcriptional Regulator

For this experiment we characterised the CcaS-CcaR optogenetic gene regulatory system developed by Olson *et al* [246] (and also employed by Milias-Argeitis *et al* for characterisation of their Turbidostat [58]). This genetic circuit is sensitive to green ($\sim$ 535 nm) and red ($\sim$ 650 nm) light: green light activates gene expression (in this case the gene under its control is GFP), whilst red light represses expression. We use the 525 nm and 625 nm LEDs in our device for this purpose; we turn on one or the other for 30 seconds during each 60 second Turbidostat cycle to control gene expression as desired.

Fig. 7.20 presents the results from two different optogenetic characterisation experiments we performed with this construct. In both cases cells were grown in EZ media, and maintained at an OD of 0.4 throughout. In Fig. 7.20a we maintain the system in its off/on state for sufficient time for it to (approximately) reach steady state. We observe that there is a tight "OFF" expression level (i.e. following a long period of red light), and that the system responds quickly when subsequently exposed to green (within approximately $\sim$ 30 minutes, close to the maturation time of GFP). In Fig. 7.20b we demonstrate how optogenetics can be used to input quickly-varying signals to a biological system by switching the system between off and on states every two hours. We observe that within this period a two-fold change in protein expression can be achieved (i.e. the output varies from a fluorescence of $\sim$0.015 to $\sim$0.03), though the output more closely resembles a sinusoid (rather than the input square wave) due to the dynamics of gene expression and protein maturation.

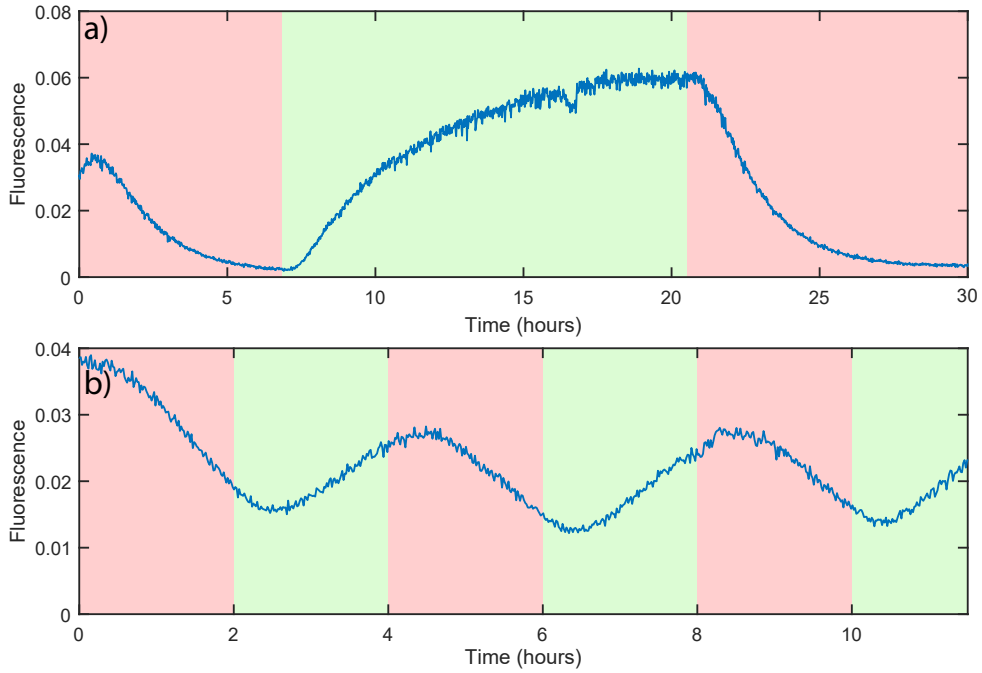


Figure 7.20: Measurements of fluorescence for the CcaS-CcaR transcriptional regulator [246] controlling the expression of GFP. For both experiments OD is maintained at 0.4 and temperature at 37°C throughout. A red background indicates the red (repressing) LED is active (with peak emission of 625 nm in our device), whilst a green background corresponds to the green (activating) LED input (525 nm in our device). Fluorescence is measured as the light intensity at 550nm (emission wavelength) divided by light intensity at 450nm (approximate excitation wavelength) when the cells are excited by the 457nm LED (as in Fig. 7.19), and a baseline fluorescence level (due to emission-excitation filter pass-through) of 0.016 is subtracted from the data.

7.9 Conclusion

In this chapter I have briefly described the design and construction of a Turbidostat device for the automated characterisation of biological systems. Its capabilities were experimentally validated, demonstrating that its design goals were achieved. We have built eight such Turbidostats by hand, and are now using them for a range of biological research projects beyond the scope of this thesis.

The device I created packages many functions desired by researchers into a cheap, compact, and easy to use platform. We have had expressions of interests from dozens of labs who would like to buy one (or many) for their own research. Consequently, as described in Chapter 8 we are now pursuing open-source commercialisation of our technology, such that it can be shared with the community and applied broadly for biological research.

8

Conclusion

The overarching aim of this thesis was to improve the engineering design-build-test cycle as it applies to synthetic biology. In Chapter 1 this process was broken down into three constituent challenges, which were addressed in turn throughout the thesis.

The **design** challenge was approached from a theoretical perspective in Chapter 3, where mathematical tools from traditional control engineering were refined by consideration of limitations inherent in biological systems. In this case we only concerned ourselves with creation of systems with particular control goals; adaptation and disturbance rejection. Though these goals represent only a small subset of functionalities required in biological engineering (i.e. a very different design approach was employed in Chapter 6), they are nevertheless useful for providing robustness for a broad range of applications. The constraints I derived in this chapter highlighted one of the major limitations of biological control systems, which is the challenge posed by achieving "perfect" adaptation in the face of dilution/degradation of a biological controller's constituent species [175]. This was observed in Chapter 5, where our recombinase-based control systems were only able to provide "near-perfect" adaptation to a range of disturbances.

To address the **build** challenge, Chapters 4 and 5 modelled and experimentally analysed synthetic biological circuits made from two primary components; recombinase (integrase and excisionase) proteins that can excise and rotate sections of DNA, and small-RNA (sRNA) designed to block translation from targeted mRNA. The two implementations of feedback have differing strengths and weaknesses. On one hand, sRNA

are easily programmable (in terms of both target sequence and repression strength), making them ideal for creating complex synthetic regulatory networks involving many species. In contrast, our recombinase implementation requires fixed recognition sequences (though different recombinase pairs that recognise different sequences can be used in parallel). However, their ability to affect permanent change to DNA (and not require constant production of the regulatory species as in the sRNA case) may be beneficial in some circumstances. Our research applied these biological "parts" in novel settings, demonstrating their utility for a broad range of applications in synthetic biology. In both chapters mathematical modelling was an important guide, both for *a priori* design and parameter tuning of each circuit, as well as *a posteriori* analysis of experimental results. The modelling approach developed will therefore be of great utility to future studies that employ these biological parts in a number of applications. Chapter 6 also partially addressed the challenge of experimentally building synthetic biological systems, by combining models of biological noise with the results of a broad range of experimental studies in the literature. The models in this chapter (and in particular, the performance trade-offs and parameter dependencies they reveal) provide a starting point for future implementation of these attractive variance estimating circuits.

The **test** (or characterisation) challenge was primarily addressed in Chapter 7 by the development of a new kind of experimental automation device for synthetic biology. This robotic platform (the Turbidostat) combines recently-developed electronic components, such as an on-chip spectrometer and multiple-output LEDs, to provide many attractive capabilities for biological researchers that were previously unavailable in a single platform. Reliable and repeatable characterisation of biological systems has long been a major difficulty in the field, which confounds the design and build steps of the engineering process (since they are based on uncertain data). Our Turbidostat addresses this by automating experimental characterisation procedures (thereby taking humans and their potential errors out of the loop), whilst standardising the actuation and measurement hardware. Chapter 6 also provides a valuable tool for experimental characterisation of biological systems which could be combined with the Turbidostat device to extend its

measurement capabilities.

Several obstacles remain in the path of synthetic biological researchers as the field develops into a mature discipline. Many of these problems stem from an incomplete understanding of the biological processes and systems involved, which will likely only be overcome by long-term scientific study. Synthetic biologists can contribute to this process by building, probing, and breaking natural systems: as Richard Feynman said, *What I cannot create, I do not understand.* Whenever a synthetic biological system is created, successfully or otherwise, we challenge the underlying assumptions held about the behaviour and properties of the biological systems in question. Thus, only once we are able to intelligently design a biological circuit *a priori* and have it work (approximately) as expect will we be able to claim we possess a good understanding of the biology underpinning its operation. Bringing researchers (slightly) closer to this ideal is what the research described herein has endeavoured to achieve.

8.1 Future Outlook

A number of the research directions described in this thesis have natural follow-on projects and studies, some of which are outlined below.

The challenge of creating tools and algorithms for forward-design in biology is a topic of much current research. The contribution herein (Chapter 3) develops constraints that must be satisfied for a particular design goal (disturbance rejection) to be achieved. These results could provide a partial basis for computer aided design (CAD) tools for biology (which have recently been developed for tasks such as logic circuit creation [51]). CAD packages that have been developed for traditional engineering disciplines (i.e. electrical or mechanical engineering) are built from a diversity of design algorithms that come in to play when particular requirements are specified. To develop more widely applicable CAD tools synthetic biologists will therefore need to explore design algorithms for a broader range of biological functions. To this end, we are currently working on design approaches for more general feedback architectures (such as lag and lead controllers). We

are also investigating stochastic approaches to biological design problems, because (as demonstrated in nature) working with stochasticity (rather than fighting against it) can in many cases yield superior performance [100].

Another study that is already underway is the experimental analysis of the dynamics of the circuits described in Chapters 4 and 5. To do so we will employ the Turbidostat devices (which were only developed toward the end of my DPhil) to probe each system with inputs of varying magnitude and frequency, which was not possible in previous plate reader and flow cytometer experiments. To improve the fidelity of input signalling we plan to replace some chemical inducers in these circuits with optogenetic regulation, which were demonstrated to worked well with the Turbidostat. Optogenetic actuation can be used to input complex time-dependent waveforms to our circuits, allowing us to (for example) characterise their response to sine-waves of varying frequency. It also facilitates measurement of each circuit's response time (in the presence and absence of feedback), which was difficult to separate from the influence of host growth dynamics in our past experiments. Finally, we intend to assess the disturbance-attenuating capabilities of these systems, which is of particular interest for the second sRNA circuit (described in Section 4.8) as it parallels similar architectures with this goal in the literature [182] (though experimental conditions differ).

The capabilities provided by the variance reporting circuit described in Chapter 6 are of great utility to researchers across both the academic and industrial biological sciences: it can eliminate the need for time-consuming flow-cytometry experiments when cell-to-cell variance is the parameter of experimental interest, which is particularly useful for techniques that require high-throughput screening (such as experimental evolution). At present we are discussing different implementations of this circuit with collaborators that employ FP-FRET in their research. An optimal implementation will likely employ homo-FRET, such that only a single fluorescent protein need be expressed (we focused on a two-fluorescent protein implementation in Chapter 6 as more data is available in the literature for that case). There are a number of FP-FRET systems in the literature

that would be a good starting point for this. For example, a cpYFP construct has been reported which homodimerises, resulting in a change in its excitation spectrum [241].

Development and commercialisation of our Turbidostat experimental platform (Chapter 7) is another direction of significant current and future work. At present, we are primarily focusing on further refinement of its hardware and operating system. For example, a new chip-based spectrometer (similar to the AS7262 [254]) was released recently by AMS which offers a greater number of optical filters covering the visual spectrum (each of which admits a narrower wavelength band), which we are currently integrating into the Turbidostat. This will allow the device to measure a greater variety of fluorescent proteins, and with reduced measurement noise. We are also in the process of implementing a wide range of additional software capabilities into the device's operating system. One feature of significance will be a standardised API (Application Programming Interface) to help users implement new control functionalities, and easily load past setups, without editing the device's operating system itself. This will be particularly attractive when the device is adopted by other labs for their own experiments, as functions can then be easily shared and refined between users. To this end we have recently lent Turbidostats to collaborating groups at Microsoft Research Cambridge and Imperial College London. Both institutions are testing our device in their own experiments, and have refined its operating system for their own needs.

As demand grows it will not be tenable for us to build devices by hand. In addition, biological research groups generally do not have the time or electrical engineering experience to assemble the device from scratch themselves. Consequently, we have recently entered into discussion with the university about the possibility of producing and distributing the Turbidostat on a larger scale. One approach to this that we are currently investigating is partnership with a company that specialises in the production of open-source scientific instruments. Ideally they will build the device in large batches and sell it to researchers, whilst allowing us to run the project as open-source (such that device schematics and software can remain freely available). We hope that we will be

able to get the device down to a price point of $\sim$ 500 USD, which will accelerate its adoption in both the research and education sectors. Our long term plan is to push for our hardware's adoption as a useful tool for characterisation of many biological systems: we hope that this will streamline the sharing of knowledge between researchers, standardise experimental procedures, and greatly improve the replicability of published results in the literature.

Over the course of the Turbidostat's development I have talked to many researchers about additional capabilities they might want such a device to have. Once we have begun to distribute a minimal viable product (i.e. that outlined in this Thesis) we will the move toward providing add-ons and additional capabilities. One common request is the capability to attach a pH probes and electrodes, which could be inserted through the device's lid to measure other parameters of the growth media. This was not previously built into our device as it was not necessary for our own experiments, and wished to avoid making direct contact with the growth media (hence why we used an infrared thermometer to measure its temperature). Other requests include a range of differing pumping capabilities: some researchers would like accurate syringe pumps for precise addition of media, and others methods for inter-Turbidostat mixing (i.e. as provided by the setup developed by Wong *et al.* [257]). Providing these capabilities should not be difficult with our platform (thanks to the modularity of the external pump board), and thus will be another direction for future development of this project.

In the longer term, feedback control architectures for synthetic biology (like those described herein) will be instrumental in helping the field reach its potential: *in vivo* control architectures will allow synthetic constructs to mitigate the impact of the noise and variability inherent in biological systems, which otherwise makes their behaviour unpredictable and highly variable. Meanwhile, *in silico* controllers and new hardware platforms will facilitate accurate and repeatable characterisation of biological constructs, improving the design-build-test cycle of biological engineering. Together, these advances will lead to breakthroughs in fundamental biological research, and facilitate the

development of synthetic biology to provide solutions to many of the 21st century's most pressing challenges.

References

[1] M. Eisenstein. "Living Factories of the Future". In: *Nature* 531.7594 (2016), pp. 401–403. DOI: 10.1038/531401a.

[2] J. J. Collins and A. S. Khalil. "Synthetic biology: applications come of age". In: *Nature reviews Genetics* 11.5 (2010), pp. 367–379. DOI: 10.1038/nrg2775.

[3] W. C. Ruder, T. Lu, and J. J. Collins. "Synthetic biology moving into the clinic." In: *Science (New York, N.Y.)* 333.6047 (2011), pp. 1248–1252. DOI: 10.1126/science.1206843.

[4] T. S. Gardner, C. R. Cantor, and J. J. Collins. "Construction of a genetic toggle switch in Escherichia coli". In: *Nature* 403.6767 (2000), pp. 339–342.

[5] M. B. Elowitz and S. Leibler. "A synthetic oscillatory network of transcriptional regulators". In: *Nature* 403.6767 (2000), pp. 335–338.

[6] N. Nandagopal and M. B. Elowitz. "Synthetic Biology: Integrated Gene Circuits". In: *Science* 333.6047 (2011), pp. 1244–1248.

[7] G. M. Church, M. B. Elowitz, C. D. Smolke, C. a. Voigt, and R. Weiss. "Realizing the potential of synthetic biology." In: *Nature reviews. Molecular cell biology* 15.4 (2014), pp. 289–294. DOI: 10.1038/nrm3767.

[8] M. J. Dunlop, R. S. Cox, J. H. Levine, R. M. Murray, and M. B. Elowitz. "Regulatory activity revealed by dynamic correlations in gene expression noise." In: *Nature genetics* 40.12 (2008), pp. 1493–1498. DOI: 10.1038/ng.281.

[9] A. A. K. Nielsen, T. H. Segall-Shapiro, and C. A. Voigt. "Advances in genetic circuit design: Novel biochemistries, deep part mining, and precision gene expression". In: *Current Opinion in Chemical Biology* 17.6 (2013), pp. 878–892.

[10] N. Merino, H. S. Aronson, D. Bojanova, J. Feyhl-buska, L Michael, S. Zhang, and D. Giovannelli. "Living at the Extremes : Extremophiles and the Limits of Life in a Planetary Context". In: *Frontiers in Microbiology* 10.April (2019), pp. 1–39. DOI: 10.3389/fmicb.2019.00780.

[11] P. Rampelotto. "Extremophiles and Extreme Environments". In: *Life* 3.3 (2013), pp. 482–485. DOI: 10.3390/life3030482.

[12] E. Rosenberg, G. Sharon, I. Atad, and I. Zilber-Rosenberg. "The evolution of animals and plants via symbiosis with microorganisms". In: *Environmental Microbiology Reports* 2.4 (2010), pp. 500–506. DOI: 10.1111/j.1758-2229.2010.00177.x.

[13] O. Dushek, A. C. Lellouch, D. J. Vaux, and V. Shahrezaei. "Biosensor architectures for high-fidelity reporting of cellular signaling". In: *Biophysical Journal* 107.3 (2014), pp. 773–782. DOI: 10.1016/j.bpj.2014.06.021.

[14] Y. Boada, A. Vignoni, and J. Picó. "Engineered Control of Genetic Variability Reveals Interplay among Quorum Sensing, Feedback Regulation, and Biochemical Noise". In: *ACS Synthetic Biology* 6.10 (2017), pp. 1903–1912. DOI: 10.1021/acssynbio.7b00087.

[15] R. Schleif. *Genetics and Molecular Biology*. 1993. DOI: 10.1146/annurev.pp.41.060190.000545.

[16] R. A. Hughes and A. D. Ellington. "Synthetic DNA Synthesis and Assembly : Putting the Synthetic in Synthetic Biology". In: *Cold Spring Harbor Perspectives in Biology* 9 (2017).

[17] H. M. Salis, E. A. Mirsky, and C. A. Voigt. "Automated design of synthetic ribosome binding sites to control protein expression". In: *Nature Biotechnology* 27.10 (2009), pp. 946–950. DOI: 10.1038/nbt.1568.

[18] D. Na, S. M. Yoo, H. Chung, H. Park, J. H. Park, and S. Y. Lee. "Metabolic engineering of Escherichia coli using synthetic small regulatory RNAs." In: *Nature biotechnology* 31.2 (2013), pp. 170–4. DOI: 10.1038/nbt.2461.

[19] L. Doyle, J. Hallinan, J. Bolduc, F. Parmeggiani, D. Baker, B. L. Stoddard, and P. Bradley. "Rational design of α-helical tandem repeat proteins with closed architectures". In: *Nature* 528.7583 (2015), pp. 585–588. DOI: 10.1038/nature16191. URL: http://dx.doi.org/10.1038/nature16191.

[20] S. D'Oelsnitz and A. Ellington. "Continuous directed evolution for strain and protein engineering". In: *Current Opinion in Biotechnology* 53 (2018), pp. 158–163. DOI: 10.1016/j.copbio.2017.12.020.

[21] B. Canton, A. Labno, and D. Endy. "Refinement and standardization of synthetic biological parts and devices". In: *Nature Biotechnology* 26.7 (2008), pp. 787–793. DOI: 10.1038/nbt1413.

[22] S. Cardinale and A. P. Arkin. "Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems". In: *Biotechnology Journal* 7.7 (2012), pp. 856–866. DOI: 10.1002/biot.201200085.

[23] X. Dai, M. Zhu, M. Warren, R. Balakrishnan, V. Patsalo, H. Okano, J. R. Williamson, K. Fredrick, Y. P. Wang, and T. Hwa. "Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth". In: *Nature Microbiology* 2.2 (2016), pp. 1–9. DOI: 10.1038/nmicrobiol.2016.231. arXiv: 15334406. URL: http://dx.doi.org/10.1038/nmicrobiol.2016.231.

[24] A. Nystrom, A. Papachristodoulou, and A. Angel. "A Dynamic Model of Resource Allocation in Response to the Presence of a Synthetic Construct". In: *ACS Synthetic Biology* 7 (2018), pp. 1201–1210. DOI: 10.1021/acssynbio.8b00015.

[25] F. Ceroni, R. Algar, G. B. Stan, and T. Ellis. "Quantifying cellular capacity identifies gene expression designs with reduced burden". In: *Nature Methods* 12.5 (2015), pp. 415–418. DOI: 10.1038/nmeth.3339.

[26] J. A. J. Arpino, E. J. Hancock, J. Anderson, M. Barahona, G. B. V. Stan, A. Papachristodoulou, and K. Polizzi. "Tuning the dials of synthetic biology". In: *Microbiology (United Kingdom)* 159.PART7 (2013), pp. 1236–1253. DOI: 10.1099/mic.0.067975-0.

[27] D. Del Vecchio, A. J. Dy, and Y. Qian. "Control theory meets synthetic biology." In: *J. R. Soc., Interface* 13.120 (2016), pp. 3–43.

[28] M. Khammash. "An engineering viewpoint on biological robustness." In: *BMC biology* 14.1 (2016), p. 22. DOI: 10.1186/s12915-016-0241-x.

[29] N. Rosenfeld, M. B. Elowitz, and U. Alon. "Negative autoregulation speeds the response times of transcription networks". In: *Journal of Molecular Biology* 323.5 (2002), pp. 785–793.

[30] T. I. Lee, N. J. Rinaldi, F. Robert, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitunger, E. G. Jennings, H. L. Murray, D Benjamin, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. "Transcriptional Regulatory Networks in Saccharomyces cerevisiae". In: *Science* 298.October (2002), pp. 799–804.

[31] A Becskei and L Serrano. "Engineering stability in gene networks by autoregulation." In: *Nature* 405.June (2000), pp. 590–593. DOI: 10.1038/35014651. arXiv: 9810036 [physics].

[32] D. Nevozhay, R. M. Adams, K. F. Murphy, K. Josic, and G. Balázsi. "Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression." In: *PNAS* 106.13 (2009), pp. 5123–8. arXiv: arXiv:1408.1149.

[33] D. Muzzey, C. A. Gómez-Uribe, J. T. Mettetal, and A. van Oudenaarden. "A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation". In: *Cell* 138.1 (2009), pp. 160–171. DOI: 10.1016/j.cell.2009.04.047. arXiv: NIHMS150003.

[34] H. C. Berg and D. A. Brown. "Chemotaxis in Escherichia coli analysed by three-dimensional tracking." In: *Nature* 239.5374 (1972), pp. 500–504. DOI: 10.1038/239500a0.

[35] T. M. Yi, Y Huang, M. I. Simon, and J Doyle. "Robust perfect adaptation in bacterial chemotaxis through integral feedback control." In: *Proc. Natl. Acad. Sci. U.* 97.9 (2000), pp. 4649–53.

[36] G. Lillacci, S. Aoki, D. Schweingruber, and M. Khammash. "A synthetic integral feedback controller for robust tunable regulation in bacteria". In: *bioRxiv* (2017), pp. 1–7. DOI: 10.1101/170951.

[37] B. Wang, M. Barahona, and M. Buck. "Engineering modular and tunable genetic amplifiers for scaling transcriptional signals in cascaded gene networks". In: *Nucleic Acids Research* 42.14 (2014), pp. 9484–9492. DOI: 10.1093/nar/gku593. arXiv: arXiv:1011.1669v3.

[38] R. Daniel, J. R. Rubens, R. Sarpeshkar, and T. K. Lu. "Synthetic analog computation in living cells". In: *Nature* 497.7451 (2013), pp. 619–623.

[39] K. Oishi and E. Klavins. "Biomolecular implementation of linear I/O systems". In: *IET Systems Biology* 5.4 (2011), pp. 252–260.

[40] A. W. K. Harris, J. A. Dolan, C. L. Kelly, J. Anderson, and A. Papachristodoulou. "Designing Genetic Feedback Controllers". In: *IEEE Transactions on Biomedical Circuits and Systems* 9.4 (2015), pp. 475–484.

[41] C. Briat, C. Zechner, and M. Khammash. "Design of a Synthetic Integral Feedback Circuit: Dynamic Analysis and DNA Implementation". In: *ACS Synthetic Biology* 5.10 (2016), pp. 1108–1116.

[42] C. L. Kelly, A. W. K. Harris, H. Steel, E. J. Hancock, J. T. Heap, and A. Papachristodoulou. "Synthetic negative feedback circuits using engineered small RNAs". In: *Nucleic Acids Research* (2018), pp. 1–15. DOI: 10.1101/184473.

[43] J. X. Chen, H. Steel, Y.-h. Wu, Y. Wang, J. Xu, C. P. N. Rampley, I. P. Thompson, A. Papachristodoulou, and W. E. Huang. "Development of Aspirin-Inducible Biosensors in Escherichia coli and SimCells". In: *Applied and Environmental Microbiology* 85.6 (2019).

[44] P. P. Peralta-Yahya, F. Zhang, S. B. del Cardayre, and J. D. Keasling. "Microbial engineering for the production of advanced biofuels". In: *Nature* 488.7411 (2012), pp. 320–328. DOI: 10.1038/nature11478. arXiv: 1201.0749v1.

[45] J. Ang and D. R. McMillen. "Physical constraints on biological integral control design for homeostasis and sensory adaptation". In: *Biophysical Journal* 104.2 (2013), pp. 505–515. DOI: 10.1016/j.bpj.2012.12.015.

[46] D. Mishra, P. M. Rivera, A. Lin, D. Del Vecchio, and R. Weiss. "A load driver device for engineering modularity in biological networks". In: *Nature Biotechnology* 32.12 (2014), pp. 1268–1275. arXiv: NIHMS150003.

[47] W. Ma, A. Trusina, H. El-Samad, W. A. Lim, and C. Tang. "Defining Network Topologies that Can Achieve Biochemical Adaptation". In: *Cell* 138.4 (2009), pp. 760–773. arXiv: NIHMS150003.

[48] C. Briat, A. Gupta, and M. Khammash. "Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks". In: *Cell Systems* 2.1 (2016), pp. 15–26.

[49] T. Drengstig, I. W. Jolma, X. Y. Ni, K. Thorsen, X. M. Xu, and P. Ruoff. "A basic set of homeostatic controller motifs". In: *Biophysical Journal* 103.9 (2012), pp. 2000–2010. DOI: 10.1016/j.bpj.2012.09.033.

[50] S. Green. "Revisiting Generality in Biology : Systems Biology and the Quest for Design Principles". In: *Biology & Philosophy* 30.5 (2014), pp. 1–25. DOI: 10.1007/s10539-.

[51] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt. "Genetic circuit design automation." In: *Science* 352.6281 (2016), aac7341.

[52] V. Libis, B. Delépine, and J.-L. Faulon. "Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways". In: *ACS Synthetic Biology* (2016), acssynbio.5b00225. DOI: 10.1021/acssynbio.5b00225. URL: http://pubs.acs.org/doi/abs/10.1021/acssynbio.5b00225.

[53] G. Fiore, A. Matyjaszkiewicz, F. Annunziata, C. Grierson, N. Savery, L. Marucci, and M. di Bernardo. "In-silico analysis and implementation of a multicellular feedback control strategy in a synthetic bacterial consortium". In: *Proc. of IEEE Conference on Decision and Control* Cdc (2016), acssynbio.6b00220. DOI: 10.1021/acssynbio.6b00220. URL: http://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00220.

[54] R. Sawlekar, F. Montefusco, V. V. Kulkarni, and D. G. Bates. "Implementing Nonlinear Feedback Controllers using DNA Strand Displacement Reactions". In: *IEEE Transactions on NanoBioscience* (2016).

[55] T. P. Prescott and A. Papachristodoulou. "Designing Conservation Relations in Layered Synthetic Biomolecular Networks". In: *IEEE Transactions on Biomedical Circuits and Systems* 9.4 (2015), pp. 572–580.

[56] P. Rivera-Ortiz and D. Del Vecchio. "Integral action with time scale separation: A mechanism for modularity in biological systems". In: *Proceedings of the IEEE Conference on Decision and Control* (2015), pp. 49–55.

[57] C. N. Takahashi, A. W. Miller, F. Ekness, M. J. Dunham, and E. Klavins. "A low cost, customizable turbidostat for use in synthetic circuit characterization". In: *ACS Synthetic Biology* 4.1 (2015), pp. 32–38. DOI: 10.1021/sb500165g. arXiv: arXiv:1408.1149.

[58] A. Milias-Argeitis, M. Rullan, S. K. Aoki, P. Buchmann, and M. Khammash. "Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth". In: *Nature Communications* 7.May (2016), p. 12546. DOI: 10.1038/ncomms12546.

[59] H. Steel, G. Lillacci, M. Khammash, and A. Papachristodoulou. "Challenges at the Interface of Control Engineering and Synthetic Biology". In: *Proc. IEEE Conf. Decision Control* 1014-1023 (2017).

[60] H. Steel and A. Papachristodoulou. "Design Constraints for Biological Systems that Achieve Adaptation and Disturbance Rejection". In: *IEEE Trans. Control Network Sys.* 5.2 (2018), pp. 807–817. DOI: 10.1109/TCNS.2018.2790039.

[61] H. Steel and A. Papachristodoulou. "Feedback Control and Synthetic Biology: Constraints on Design". In: *Proceedings of the 20th IFAC World Congress* (2017), pp. 10932–10937.

[62] H. Steel, A. W. K. Harris, E. J. Hancock, C. L. Kelly, and A. Papachristodoulou. "Frequency domain analysis of small non-coding RNAs shows summing junction-like behaviour". In: *Proceedings of the 56th IEEE Conference on Decision and Control.* (2017), pp. 5328–5333.

[63] T. Folliard, H. Steel, T. P. Prescott, G. Wadhams, L. J. Rothschild, and A. Papachristodoulou. "A synthetic recombinase-based feedback loop results in robust expression". In: *ACS Synthetic Biology* (2017).

[64] H. Steel and A. Papachristodoulou. "Low-Burden Biological Feedback Controllers for Near-Perfect Adaptation". In: *ACS Synthetic Biology* (2019). DOI: 10.1021/acssynbio.9b00125.

[65] H. Steel and A. Papachristodoulou. "Probing Intercell Variability using Bulk Measurements". In: *ACS Synthetic Biology* 7 (2018), pp. 1528–1537. DOI: 10.1021/acssynbio.8b00014.

[66] B Alberts, A Johnson, J Lewis, D Morgan, M Raff, K Roberts, and P Walter. *Molecular Biology of the Cell*. 6th. Garland Science, 2014.

[67] M Snyder, J Peters, T Henkin, and W Champness. *Molecular Genetics of Bacteria*. 4th. ASM Press, 2013.

[68] J. A. Bryant, L. E. Sellars, S. J. W. Busby, and D. J. Lee. "Chromosome position effects on gene expression in Escherichia coli K-12". In: *Nucleic Acids Research* 42.18 (2014), pp. 11383–11392. DOI: 10.1093/nar/gku828.

[69] T. Folliard, B. Mertins, H. Steel, T. Prescott, T. Newport, C. Jones, G. Wadhams, T. Bayer, J. Armitage, A. Papachristodoulou, and L. Rothschild. "Ribo-attenuators: novel elements for reliable and modular riboswitch engineering". In: *Scientific Reports* (2017).

[70] C. Lu. "Imaginary Proteins Made Real". In: *Cell* 164.1-2 (2016), p. 7. DOI: 10.1016/j.cell.2016.01.001.

[71] N. Roquet, A. P. Soleimany, A. C. Ferris, S. Aaronson, and T. K. Lu. "Synthetic recombinasebased state machines in living cells". In: *Science* (2016). DOI: 10.1126/science.aad8559.

[72] C. Merrick, J. Zhao, and S. Rosser. "Serine Integrases: Advancing Synthetic Biology". In: *ACS Synthetic Biology* (2018), acssynbio.7b00308. DOI: 10.1021/acssynbio.7b00308. URL: http://pubs.acs.org/doi/10.1021/acssynbio.7b00308.

[73] J. Bonnet, P. Yin, M. E. Ortiz, P. Subsoontorn, and D. Endy. "Amplifying genetic logic gates." In: *Science (New York, N.Y.)* 340.6132 (2013), pp. 599–603. DOI: 10.1126/science.1232758.

[74] A. A. Mannan, D. Liu, F. Zhang, and D. A. Oyarzún. "Fundamental Design Principles for Transcription-Factor-Based Metabolite Biosensors". In: *ACS Synthetic Biology* (2017), acssynbio.7b00172. DOI: 10.1021/acssynbio.7b00172. URL: http://pubs.acs.org/doi/abs/10.1021/acssynbio.7b00172.

[75] M.-T. Lin, C.-Y. Wang, H.-J. Xie, C. H. Y. Cheung, C.-H. Hsieh, H.-F. Juan, B.-S. Chen, and C. Lin. "Novel Utilization of Terminators in the Design of Biologically Adjustable Synthetic Filters". In: *ACS Synthetic Biology* (2016), acssynbio.5b00174.

[76] A. Raj and A. van Oudenaarden. "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences". In: *Cell* 135.2 (2008), pp. 216–226. DOI: 10.1016/j.cell.2008.09.050. arXiv: NIHMS150003.

[77] A. Serganov and E. Nudler. "A decade of riboswitches". In: *Cell* 152.1-2 (2013), pp. 17–24. DOI: 10.1016/j.cell.2012.12.024. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.cell.2012.12.024.

[78] S. Gottesman and G. Storz. "Bacterial small RNA regulators: Versatile roles and rapidly evolving variations". In: *Cold Spring Harbor Perspectives in Biology* 3.12 (2011), pp. 1–16. DOI: 10.1101/cshperspect.a003798.

[79] E. S. Groban, E. J. Clarke, H. M. Salis, S. M. Miller, and C. A. Voigt. "Kinetic Buffering of Cross Talk between Bacterial Two-Component Sensors". In: *Journal of Molecular Biology* 390.3 (2009), pp. 380–393. DOI: 10.1016/j.jmb.2009.05.007.

[80] M. R. Junttila, S.-P. Li, and J. Westermarck. "Phosphatase-mediated crosstalk between MAPK signaling pathways in the regulation of cell survival". In: *The FASEB Journal* 22.4 (2007), pp. 954–965. DOI: 10.1096/fj.06-7859rev.

[81] J. Hausser, A. Mayo, L. Keren, and U. Alon. "Central dogma rates and the trade-off between precision and economy in gene expression". In: *Nature Communications* 10.1 (2019). DOI: 10.1038/s41467-018-07391-8. URL: http://dx.doi.org/10.1038/s41467-018-07391-8.

[82] G. Storz, J. A. Opdyke, and A. Zhang. "Controlling mRNA stability and translation with small, noncoding RNAs". In: *Current Opinion in Microbiology* 7.2 (2004), pp. 140–144. DOI: 10.1016/j.mib.2004.02.015.

[83] N. C. Butzin and W. H. Mather. "Crosstalk between diverse synthetic protein degradation tags in Escherichia coli". In: *ACS Synthetic Biology* (2017). DOI: 10.1021/acssynbio.7b00122.

[84] M Elowitz, A Levine, E Siggia, and P Swain. "Stochastic gene expression in a single cell". In: *Science* 297.5584 (2002), pp. 1183–6. DOI: 10.1126/science.1070919.

[85] P. S. Swain, M. B. Elowitz, and E. D. Siggia. "Intrinsic and extrinsic contributions to stochasticity in gene expression." In: *Proceedings of the National Academy of Sciences of the United States of America* 99.20 (2002), pp. 12795–800. DOI: 10.1073/pnas.162041399.

[86] D. A. Oyarzún, J. B. Lugagne, and G. B. V. Stan. "Noise propagation in synthetic gene circuits for metabolic control". In: *ACS Synthetic Biology* 4.2 (2015), pp. 116–125. DOI: 10.1021/sb400126a.

[87] J. L. Snoep, L. P. Yomano, H. V. Westerhoff, and L. O. Ingram. "Protein burden in Zymomonas mobilis: Negative flux and growth control due to overproduction of glycolytic enzymes". In: *Microbiology* 141.9 (1995), pp. 2329–2337. DOI: 10.1099/13500872-141-9-2329.

[88] Y. Qian and D. Del Vecchio. "Mitigation of ribosome competition through distributed sRNA feedback". In: *Proc. of IEEE Conference on Decision and Control* (2016), pp. 1–30.

[89] J.-W. Veening, W. K. Smits, and O. P. Kuipers. "Bistability, Epigenetics, and Bet-Hedging in Bacteria". In: *Annual Review of Microbiology* 62.1 (2008), pp. 193–210. DOI: 10.1146/annurev.micro.62.081307.163002.

[90] L. Robert, G. Paul, Y. Chen, F. Taddei, D. Baigl, and A. B. Lindner. "Pre-dispositions and epigenetic inheritance in the Escherichia coli lactose operon bistable switch". In: *Molecular Systems Biology* 6.357 (2010). DOI: 10.1038/msb.2010.12.

[91] B. Snijder and L. Pelkmans. "Origins of regulated cell-to-cell variability". In: *Nature Reviews Molecular Cell Biology* 12.2 (2011), pp. 119–125. DOI: 10.1038/nrm3044.

[92] M. Soltani, C. A. Vargas-Garcia, D. Antunes, and A. Singh. "Intercellular Variability in Protein Levels from Stochastic Expression and Noisy Cell Cycle Processes". In: *PLoS Computational Biology* 12.8 (2016), pp. 1–23. DOI: 10.1371/journal.pcbi.1004972.

[93] A. Sanchez and I. Golding. "Genetic Determinants and Cellular Constraints in Noisy Gene Expression". In: *Science* December (2013), pp. 1188–1193. DOI: 0.1126/science.1242975. arXiv: NIHMS150003.

[94] A. Hilfinger and J. Paulsson. "Separating intrinsic from extrinsic fluctuations in dynamic biological systems". In: *Proceedings of the National Academy of Sciences* 108.29 (2011), pp. 12167–12172. DOI: 10.1073/pnas.1018832108. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.1018832108.

[95] A. Singh and M. Soltani. "Quantifying intrinsic and extrinsic variability in stochastic gene expression models". In: *PLoS ONE* 8.12 (2013). DOI: 10.1371/journal.pone.0084301.

[96] N. Rosenfeld, J. Young, U. Alon, P. Swain, and M. Elowitz. "Gene regulation at the single-cell level". In: *Science* 307.5717 (2005), pp. 1962–1965. DOI: citeulike-article-id:141524. URL: http://www.sciencemag.org/cgi/content/abstract/307/5717/1962.

[97] G. Balazsi, A. Van Oudenaarden, and J. J. Collins. "Cellular decision making and biological noise: From microbes to mammals". In: *Cell* 144.6 (2011), pp. 910–925.

[98] L. T. Macneil and A. J. M. Walhout. "Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression". In: *Genome Research* (2011), pp. 645–657. DOI: 10.1101/gr.097378.109.21.

[99] G. Rodrigo and J. F. Poyatos. "Genetic Redundancies Enhance Information Transfer in Noisy Regulatory Circuits". In: *PLoS Computational Biology* 12.10 (2016), pp. 1–20. DOI: 10.1371/journal.pcbi.1005156. arXiv: 1602.07082.

[100] J. Paulsson, O. G. Berg, and M. Ehrenberg. "Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation". In: *Proceedings of the National Academy of Sciences* 97.13 (2000).

[101] T. Tian and K. Burrage. "Stochastic models for regulatory networks of the genetic toggle switch." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.22 (2006), pp. 8372–8377. DOI: 10.1073/pnas.0507818103.

[102] N. Politi, L. Pasotti, S. Zucca, and P. Magni. "Modelling the effects of cell-to-cell variability on the output of interconnected gene networks in bacterial populations". In: *BMC Systems Biology* 9.Suppl 3 (2015), S6. DOI: 10.1186/1752-0509-9-S3-S6.

[103] Y Dublanche, K Michalodimitrakis, N Kummerer, M Foglierini, and L Serrano. "Noise in transcription negative feedback loops: simulation and experimental analysis". In: *Mol Syst Biol* 2 (2006), p. 41. DOI: msb4100081[pii]\r10.1038/msb4100081.

[104] K. F. Murphy, R. M. Adams, X. Wang, and J. J. Collins. "Tuning and controlling gene expression noise in synthetic gene networks". In: *Nucleic Acids Research* 38.8 (2010), pp. 2712–2726. DOI: 10.1093/nar/gkq091.

[105] M. Thattai and A. van Oudenaarden. "Intrinsic noise in gene regulatory networks". In: *Proceedings of the National Academy of Sciences* 98.15 (2001), pp. 8614–8619. DOI: 10.1073/pnas.151588598.

[106] B. Schulz. *BioBricks Foundation*. URL: http://biobricks.org/.

[107] E. Appleton, C. Madsen, N. Roehner, and D. Densmore. "Design Automation in Synthetic Biology". In: *Cold Spring Harbor Perspectives in Biology* (2017). DOI: 10.1101/cshperspect.a023978.

[108] N. A. Shah and C. A. Sarkar. "Robust network topologies for generating switch-like cellular responses". In: *PLoS Computational Biology* 7.6 (2011). DOI: 10.1371/journal.pcbi.1002085.

[109] S. U. Wei. "Topologies Needed for Perfect Adaptation". In: *Proceedings of the 30th Chinese Control Conference* (2011), pp. 5714–5717.

[110] S. U. Wei. "Perfect adaptation of general nonlinear systems". In: *Journal of Systems Science and Complexity* 29.1 (2016), pp. 61–73. DOI: 10.1007/s11424-015-4053-9.

[111] E. Lambert, E. J. Hancock, and A. Papachristodoulou. "Engineering a Genetic Oscillator Using Delayed Feedback". In: *Delay Systems, Advances in Delays and Dynamics Volume 1*. 2014, pp. 389–402.

[112] A. Mitchell, P. Wei, and W. A. Lim. "Oscillatory stress stimulation uncovers an Achilles heel of the yeast MAPK signaling network". In: *Science* 350.6266 (2015), pp. 1379–1383. DOI: 10.1126/science.aab0892.

[113] V. K. Mutalik, J. C. Guimaraes, G. Cambray, C. Lam, M. J. Christoffersen, Q.-A. Mai, A. B. Tran, M. Paull, J. D. Keasling, A. P. Arkin, and D. Endy. "Precise and reliable gene expression via standard transcription and translation initiation elements." In: *Nature methods* 10.4 (2013), pp. 354–60. DOI: 10.1038/nmeth.2404.

[114] D. Del Vecchio. "Modularity, context-dependence, and insulation in engineered biological circuits". In: *Trends in Biotechnology* 33.2 (2015), pp. 111–119.

[115] N. Roquet and T. K. Lu. *Digital and analog gene circuits for biotechnology*. 2014.

[116] Y. Qian, H. H. Huang, J. I. Jiménez, and D. Del Vecchio. "Resource Competition Shapes the Response of Genetic Circuits". In: *ACS Synthetic Biology* 6.7 (2017), pp. 1263–1272. DOI: 10.1021/acssynbio.6b00361.

[117] Y. C. Chang, J. P. Armitage, A. Papachristodoulou, and G. H. Wadhams. "A single phosphatase can convert a robust step response into a graded, tunable or adaptive response". In: *Microbiology (United Kingdom)* 159.PART7 (2013), pp. 1276–1285. DOI: 10.1099/mic.0.066324-0.

[118] R. M. Gordley, R. E. Williams, C. J. Bashor, J. E. Toettcher, S. Yan, and W. A. Lim. "Engineering dynamical control of cell fate switching using synthetic phospho-regulons." In: *Proceedings of the National Academy of Sciences of the United States of America* 113.47 (2016), pp. 13528–13533. DOI: 10.1073/pnas.1610973113.

[119] W. An and J. W. Chin. "Synthesis of orthogonal transcription-translation networks." In: *Proceedings of the National Academy of Sciences of the United States of America* 106.21 (2009), pp. 8477–82. DOI: 10.1073/pnas.0900267106.

[120] A. P. Darlington, J. Kim, J. I. Jiménez, and D. G. Bates. "Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes". In: *Nature Communications* 9.1 (2018). DOI: 10.1038/s41467-018-02898-6.

[121] A. Gyorgy and D. Del Vecchio. "Modular Composition of Gene Transcription Networks". In: *PLoS Computational Biology* 10.3 (2014). DOI: 10.1371/journal.pcbi.1003486.

[122] D. Del Vecchio, A. J. Ninfa, and E. D. Sontag. "Modular cell biology: retroactivity and insulation". In: *Molecular systems biology* 4.161 (2008). DOI: 10.1038/msb4100204.

[123] K. S. Nilgiriwala, J. Jimenez, P. M. Rivera, and D. Del Vecchio. "Synthetic Tunable Amplifying Buffer Circuit in E. coli". In: *ACS Synthetic Biology* 4 (2015), pp. 577–584.

[124] H. Kitano. "Biological robustness". In: *Nature Reviews Genetics* 5.November (2004), pp. 826–837. DOI: 10.1007/978-3-7643-7567-6_10.

[125] F. He, E. Murabito, and H. V. Westerhoff. "Synthetic biology and regulatory networks: where metabolic systems biology meets control engineering". In: *Journal of the Royal Society, Interface* 13 (2016), p. 20151046. DOI: 10.1098/rsif.2015.1046.

[126] J.-C. Lin and D Thirumalai. "Gene regulation by riboswitches with and without negative feedback loop." In: *Biophysical Journal* 103.11 (2012), pp. 2320–30. DOI: 10.1016/j.bpj.2012.10.026. arXiv: 1210.6998.

[127] J. L. Radzikowski, S. Vedelaar, D. Siegel, Á. D. Ortega, A. Schmidt, and M. Heinemann. "Bacterial persistence is an active r S stress response to metabolic flux limitation". In: *Molecular systems biology* 12.882 (2016), pp. 1–18.

[128] G. J. Nistala, K. Wu, C. V. Rao, and K. D. Bhalerao. "A modular positive feedback-based gene amplifier." In: *Journal of biological engineering* 4 (2010), p. 4. DOI: 10.1186/1754-1611-4-4.

[129] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty. "A fast, robust and tunable synthetic gene oscillator." In: *Nature* 456.7221 (2008), pp. 516–519. DOI: 10.1038/nature07389. arXiv: arXiv:1011.1669v3.

[130] T. Shopera, W. R. Henson, A. Ng, Y. J. Lee, K. Ng, and T. S. Moon. "Robust, tunable genetic memory from protein sequestration combined with positive feedback". In: *Nucleic Acids Research* 43.18 (2015), pp. 9086–9094. DOI: 10.1093/nar/gkv936.

[131] S Mangan and U Alon. "Structure and function of the feed-forward loop network motif." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.21 (2003), pp. 11980–11985. DOI: 10.1073/pnas.2133841100.

[132] K. Papenfort, E. Espinosa, J. Casadesús, and J. Vogel. "Small RNA-based feedforward loop with AND-gate logic regulates extrachromosomal DNA transfer in Salmonella." In: *Proceedings of the National Academy of Sciences of the United States of America* 112.34 (2015), E4772–81. DOI: 10.1073/pnas.1507825112.

[133] A. Milias-Argeitis, S. Summers, J. Stewart-Ornstein, I. Zuleta, D. Pincus, H. El-Samad, M. Khammash, and J. Lygeros. "In silico feedback for in vivo regulation of a gene expression circuit." In: *Nature biotechnology* 29.12 (2011), pp. 1114–6.

[134] B. T. Bajar, E. S. Wang, A. J. Lam, B. B. Kim, C. L. Jacobs, E. S. Howe, M. W. Davidson, M. Z. Lin, and J. Chu. "Improving brightness and photostability of green and red fluorescent proteins for live cell imaging and FRET reporting". In: *Scientific Reports* 6.1 (2016), p. 20889. DOI: 10.1038/srep20889.

[135] B. T. Bajar, E. S. Wang, S. Zhang, M. Z. Lin, and J. Chu. "A guide to fluorescent protein FRET pairs". In: *Sensors* 16.9 (2016), pp. 1–24. DOI: 10.3390/s16091488.

[136] J. Beal, T. Haddock-Angelli, M. Gershater, K. de Mora, M. Lizarazo, J. Hollenhorst, R. Rettberg, and iGEM Interlab Study Contributors. "Reproducibility of Fluorescent Expression from Engineered Biological Constructs in E . coli". In: *PLoS ONE* 11.3 (2016), pp. 1–22. DOI: 10.1371/journal.pone.0150182.

[137] J Kraan, J. W. Gratama, M Keeney, and J. L. D. Hautcourt. "Setting up and calibration of a flow cytometer for multicolor immunophenotyping". In: *Journal of Biological Regulators and Homeostatic Agents* (2003), pp. 223–233.

[138] L. Wang and R. A. Hoffman. "Standardization, Calibration, and Control in Flow Cytometry". In: *Current Protocols in Cytometry* January (2017), pp. 1–27. DOI: 10.1002/cpcy.14.

[139] J. Beal, T. Haddock-angelli, G. Baldwin, M. Gershater, A. Dwijayanti, M. Storch, K. D. Mora, M. Lizarazo, and R. Rettberg. "Quantification of bacterial fluorescence using independent calibrants". In: *PLoS ONE* (2018), pp. 1–15.

[140] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-garcia, J. I. Glass, and M. W. Covert. "Theory A Whole-Cell Computational Model Predicts Phenotype from Genotype". In: *Cell* 150 (2012), pp. 389–401. DOI: 10.1016/j.cell.2012.05.044.

[141] N. S. Nise. *Control Systems Engineering, 6th Ed.* Wiley, 2010, p. 944.

[142] Y. Zheng and G. Sriram. "Mathematical modeling: Bridging the gap between concept and realization in synthetic biology". In: *Journal of Biomedicine and Biotechnology* 2010.Figure 1 (2010). DOI: 10.1155/2010/541609.

[143] C. Mcbride and D. D. Vecchio. "The Number of Equilibrium Points of Perturbed Nonlinear Positive Dynamical Systems (Extended Version)". In: 1521925 (2019), pp. 1–22.

[144] B. P. Ingalls. *Mathematical Modelling in Systems Biology : An Introduction.* MIT Press, 2014. arXiv: arXiv:1011.1669v3.

[145] M. Bengtsson, A. Ståhlberg, P. Rorsman, and M. Kubista. "Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels". In: *Genome Research* 15 (2005), pp. 1388–1392. DOI: 10.1101/gr.3820805.4.

[146] B Banerjee, S Balasubramanian, G Ananthakrishna, T. V. Ramakrishnan, and G. V. Shivashankar. "Tracking Operator State Fluctuations in Gene Expression in Single Cells". In: *Biophysical Journal* 86.5 (2004), pp. 3052–3059. DOI: 10.1016/S0006-3495(04)74354-4.

[147] J. Beal. "Biochemical complexity drives log-normal variation in genetic expression". In: *Engineering Biology* 1.1 (2017), pp. 55–60. DOI: 10.1049/enb.2017.0004.

[148] D. T. Gillespie. "Approximate accelerated stochastic simulation of chemically reacting systems". In: *Journal of Chemical Physics* 115.4 (2001), pp. 1716–1733.

[149] D. T. Gillespie and L. R. Petzold. "Improved leap-size selection for accelerated stochastic simulation". In: *Journal of Chemical Physics* 119.16 (2003), pp. 8229–8234. DOI: 10.1063/1.1613254.

[150] L. Huynh and I. Tagkopoulos. "A Parts Database with Consensus Parameter Estimation for Synthetic Circuit Design". In: *ACS Synthetic Biology* (2016), acssynbio.5b00205. DOI: 10.1021/acssynbio.5b00205. URL: http://pubs.acs.org/doi/abs/10.1021/acssynbio.5b00205.

[151] A. W. K. Harris, H. Steel, C. L. Kelly, and A. Papachristodoulou. "The Autorepressor: a Case Study of the Importance of Model Selection". In: *Proceedings of the 56th IEEE Conference on Decision and Control.* (2017), pp. 1622–1627.

[152] G. Lillacci and M. Khammash. "Parameter estimation and model selection in computational biology". In: *PLoS Computational Biology* 6.3 (2010). DOI: 10.1371/journal.pcbi.1000696.

[153] A. F. Villaverde and J. R. Banga. "Reverse engineering and identification in systems biology: strategies, perspectives and challenges". In: *Journal of The Royal Society Interface* 11.91 (2013), pp. 20130505–20130505. DOI: 10.1098/rsif.2013.0505. URL: http://rsif.royalsocietypublishing.org/cgi/doi/10.1098/rsif.2013.0505.

[154] G. Giordano, C. Cuba Samaniego, E. Franco, and F. Blanchini. "Computing the structural influence matrix for biological systems". In: *Journal of Mathematical Biology* 72.7 (2016), pp. 1927–1958. DOI: 10.1007/s00285-015-0933-9.

[155] R. Levins. *Evolution in changing environments: some theoretical explorations*. Princeton: Princeton University Press, 1968.

[156] J. M. Dambacher, H. W. Li, and P. A. Rossignol. "Relevance of Community Structure in Assessing Indeterminacy of Ecological Predictions". In: *Ecology* 83.5 (2002), pp. 1372–1385.

[157] F. Blanchini, E. Franco, and G. Giordano. "Determining the Structural Properties of a Class of Biological Models". In: *Proceedings of the 51st IEEE Conference on Decision and Control (CDC)* (2012), pp. 5505–5510. DOI: 10.1109/CDC.2012.6427037.

[158] Z. F. Tang and D. R. McMillen. "Design principles for the analysis and construction of robustly homeostatic biological networks". In: *Journal of Theoretical Biology* 408 (2016), pp. 274–289. DOI: 10.1016/j.jtbi.2016.06.036.

[159] B. Francis and W. M. Wonham. "The Internal Model Principle of Control Theory The Internal Model Principle of Control Theory". In: *Automatica* 12.5 (1976), pp. 457–465.

[160] E. D. Sontag. "Adaptation and regulation with signal detection implies internal model". In: *Systems and Control Letters* 50.2 (2003), pp. 119–126. DOI: 10.1016/S0167-6911(03)00136-1. arXiv: 0309003v1 [arXiv:q-bio].

[161] U. Alon, M. G. Surette, N. Barkai, and S. Leibler. "Robustness in bacterial chemotaxis". In: *Nature* 397.6715 (1999), pp. 168–171.

[162] L. Huang, Z. Yuan, P. Liu, T. Zhou, H. P. Sørensen, W. H. Brondyk, G. J. Gopal, A. Kumar, Graslund, K. Ukkonen, J. Veijola, A. Vasala, and P. Neubauer. "Effects of promoter leakage on dynamics of gene expression." In: *BMC systems biology* 9.16 (2015). DOI: 10.1186/s12918-015-0157-z.

[163] T. Drengstig, H. R. Ueda, and P. Ruoff. "Predicting perfect adaptation motifs in reaction kinetic networks". In: *Journal of Physical Chemistry B* 112.51 (2008), pp. 16752–16758. DOI: 10.1021/jp806818c.

[164] H. H. Rosenbrock. "Transformation of linear constant system equations." In: *Proceedings of the Institution of Electrical Engineers* 114.4 (1967).

[165] H. H. Rosenbrock. *State-Space and Multivariable Theory*. Nelson, 1970.

[166] J. Ang, S. Bagh, B. P. Ingalls, and D. R. McMillen. "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. DOI: 10.1016/j.jtbi.2010.07.034.

[167] W. W. Wong, T. Y. Tsai, and J. C. Liao. "Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator." In: *Molecular systems biology* 3.130 (2007), p. 130. DOI: 10.1038/msb4100172.

[168] R. M. Murray and D. Del Vecchio. *Biomolecular Feedback Systems*. 2014. arXiv: arXiv:1309.7798.

[169] D. Lewis, P. Le, C. Zurla, L. Finzi, and S. Adhya. "Multilevel autoregulation of lambda repressor protein CI by DNA looping in vitro." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.36 (2011), pp. 14807–14812. DOI: 10.1073/pnas.1111221108.

[170] J Krishnan. "Effects of saturation and enzyme limitation in feedforward adaptive signal transduction". In: *IET Systems Biology* 5.3 (2011), pp. 208–219.

[171] E. D. Sontag. "A Dynamic Model of Immune Responses to Antigen Presentation Predicts Different Regions of Tumor or Pathogen Elimination". In: *Cell Systems* 4 (2017), pp. 231–241.

[172] M. Lang and E. D. Sontag. "Zeros of nonlinear systems with input invariances". In: *Automatica* 81 (2017), pp. 46–55. DOI: 10.1016/j.automatica.2017.03.030.

[173] H Chang and A Levchenko. "Adaptive molecular networks controlling chemotactic migration: dynamic inputs and selection of the network architecture." In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 368.1629 (2013), p. 20130117. DOI: 10.2307/42569254.

[174] C. J. Wang, A. Bergmann, B. Lin, K. Kim, and A. Levchenko. "Diverse Sensitivity Thresholds in Dynamic Signaling Responses by Social Amoebae". In: *Science Signaling* 5.213 (2012). DOI: 10.1126/scisignal.2002449.

[175] Y. Qian and D. 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), p. 20170902. DOI: 10.1098/rsif.2017.0902.

[176] D. Soloveichik, G. Seelig, and E. Winfree. "DNA as a universal substrate for chemical kinetics". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5347 LNCS.12 (2009), pp. 57–69. DOI: 10.1007/978-3-642-03076-5_6.

[177] T. Song, S. Garg, R. Mokhtar, H. Bui, and J. Reif. "Analog Computation by DNA Strand Displacement Circuits". In: *ACS Synthetic Biology* (2016), acssynbio.6b00144. DOI: 10.1021/acssynbio.6b00144.

[178] D. Y. Zhang and G. Seelig. "Dynamic DNA nanotechnology using strand-displacement reactions." In: *Nature chemistry* 3.2 (2011), pp. 103–113. DOI: 10.1038/nchem.957.

[179] D. Y. Zhang. "Towards Domain-Based Sequence Design for DNA Strand Displacement Reactions". In: *Internatonal Workshop on DNA-Based Computers*. Springer Berlin Heidelberg, 2010, pp. 162–175. DOI: 10.1007/978-3-642-18305-8.

[180] U Alon. "Network motifs: theory and experimental approaches". In: *Nat Rev Genet* 8.6 (2007), pp. 450–461. DOI: nrg2102[pii]\r10.1038/nrg2102.

[181] C. L. Kelly, G. M. Taylor, A. Hitchcock, A. Torres-me, and J. T. Heap. "A Rhamnose-Inducible System for Precise and Temporal Control of Gene Expression in Cyanobacteria". In: *ACS Synthetic Biology* (2018). DOI: 10.1021/acssynbio.7b00435.

[182] D. K. Agrawal, X. Tang, A. Westbrook, R. Marshall, C. S. Maxwell, J. Lucks, V. Noireaux, C. L. Beisel, M. J. Dunlop, and E. Franco. "Mathematical Modeling of RNA-Based Architectures for Closed Loop Control of Gene Expression". In: *ACS Synthetic Biology* 7 (2018), pp. 1219–1228. DOI: 10.1021/acssynbio.8b00040.

[183] T. Ellis, X. Wang, and J. J. Collins. "Diversity-based, model-guided construction of synthetic gene networks with predicted functions". In: *Nature Biotechnology* 27.5 (2009), pp. 465–471. DOI: 10.1038/nbt.1536.

[184] C. H. Wu, H. C. Lee, and B. S. Chen. "Robust synthetic gene network design via library-based search method". In: *Bioinformatics* 27.19 (2011), pp. 2700–2706. DOI: 10.1093/bioinformatics/btr465.

[185] D. Braun, S. Basu, and R. Weiss. "Parameter estimation for two synthetic gene networks: A case study". In: *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005* 5 (2005), pp. 769–772. DOI: 10.1109/ICASSP.2005.1416417.

[186] D. N. Adamson and H. N. Lim. "Essential requirements for robust signaling in Hfq dependent small RNA networks". In: *PLoS Computational Biology* 7.8 (2011). DOI: 10.1371/journal.pcbi.1002138.

[187] T. Soper, P. Mandin, N. Majdalani, S. Gottesman, and S. a. Woodson. "Positive regulation by small RNAs and the role of Hfq". In: *Pnas* 107.21 (2010), pp. 2–7. DOI: 10.1073/pnas.1004435107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1004435107.

[188] J. F. Hopkins, S. Panja, and S. A. Woodson. "Rapid binding and release of Hfq from ternary complexes during RNA annealing". In: *Nucleic Acids Research* 39.12 (2011), pp. 5193–5202. DOI: 10.1093/nar/gkr062.

[189] E. G. H. Wagner. "Cycling of RNAs on Hfq". In: *RNA Biology* 10.4 (2013), pp. 619–626. DOI: 10.4161/rna.24044.

[190] D. Kennell and H. Riezman. "Transcription and translation initiation frequencies of the Escherichia coli lac operon". In: *Journal of Molecular Biology* 114.1 (1977), pp. 1–21. DOI: 10.1016/0022-2836(77)90279-0.

[191] S. T. Liang, M. Bipatnath, Y. C. Xu, S. L. Chen, P. Dennis, M. Ehrenberg, and H. Bremer. "Activities of constitutive promoters in Escherichia coli". In: *Journal of Molecular Biology* 292.1 (1999), pp. 19–37. DOI: 10.1006/jmbi.1999.3056.

[192] H. Chen, K. Shiroguchi, H. Ge, and X. S. Xie. "Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli". In: *Molecular Systems Biology* 11.1 (2015), pp. 781–781. DOI: 10.15252/msb.20145794.

[193] C. A. McCullen, J. N. Benhammou, N. Majdalani, and S. Gottesman. "Mechanism of positive regulation by DsrA and RprA small noncoding RNAs: Pairing increases translation and protects rpoS mRNA from degradation". In: *Journal of Bacteriology* 192.21 (2010), pp. 5559–5571. DOI: 10.1128/JB.00464-10.

[194] R. Guantes, B. Cayrol, F. Busi, and V. Arluison. "Positive regulatory dynamics by a small noncoding RNA: speeding up responses under temperature stress". In: *Molecular BioSystems* 8.6 (2012), p. 1707. DOI: 10.1039/c2mb05479e.

[195] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. "Stochasticity in gene expression: From theories to phenotypes". In: *Nature Reviews Genetics* 6.6 (2005), pp. 451–464. DOI: 10.1038/nrg1615.

[196] N. E. Buchler, U. Gerland, and T. Hwa. "On schemes of combinatorial transcription logic." In: *Proceedings of the National Academy of Sciences of the United States of America* 100.9 (2003), pp. 5136–41. DOI: 10.1073/pnas.0930314100.

[197] T. S. Moon, C. Lou, A. Tamsir, B. C. Stanton, and C. a. Voigt. "Genetic programs constructed from layered logic gates in single cells." In: *Nature* 491.7423 (2012), pp. 249–53. DOI: 10.1038/nature11516. arXiv: NIHMS150003. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3904217{\&}tool=pmcentrez{\&}rendertype=abstract{\%}5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/23041931.

[198] N. D. Grindley, K. L. Whiteson, and P. A. Rice. "Mechanisms of Site-Specific Recombination". In: *Annu. Rev. Biochem.* 75 (2006), pp. 567–605. DOI: 10.1146/annurev.biochem.73.011303.073908.

[199] V. Hsiao, Y. Hori, P. W. K. Rothemund, and R. M. Murray. "A population-based temporal logic gate for timing and recording chemical events". In: *Molecular systems biology* 557 (2016), pp. 1–17. DOI: 10.1101/029967.

[200] J. Bonnet, P. Subsoontorn, and D. Endy. "Rewritable digital data storage in live cells via engineered control of recombination directionality". In: *Proceedings of the National Academy of Sciences* 109.23 (2012), pp. 8884–8889. DOI: 10.1073/pnas.1202344109.

[201] J. Bowyer, J. Zhao, P. Subsoontorn, W. Wong, S. Rosser, and D. Bates. "Mechanistic Modeling of a Rewritable Recombinase Addressable Data Module". In: *IEEE Transactions on Biomedical Circuits and Systems* 10.6 (2016), pp. 1161–1170. DOI: 10.1109/TBCAS.2016.2526668.

[202] J. E. Bowyer, V. Hsiao, W. W. Wong, and D. G. Bates. "Mechanistic modelling of a recombinase-based two-input temporal logic gate". In: *IET Engineering Biology* 1 (2017), pp. 40–50. DOI: 10.1049/enb.2017.0006.

[203] S. D. Colloms, C. A. Merrick, F. J. Olorunniji, and W. M. Stark. "Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination". In: *Nucleic Acids Research* 42.4 (2014), e23. DOI: 10.1093/nar/gkt1101.

[204] P. C. M. Fogg, S. Colloms, S. Rosser, M. Stark, and M. C. M. Smith. "New applications for phage integrases". In: *Journal of Molecular Biology* 426.15 (2014), pp. 2703–2716. DOI: 10.1016/j.jmb.2014.05.014. URL: http://dx.doi.org/10.1016/j.jmb.2014.05.014.

[205] A. Camilli and J. Mekalanos. "Cholerae Genes Induced During Infection". In: *Molecular Microbiology* 18 (1995), pp. 671–683.

[206] A. R. Castillo, A. J. Woodruff, L. E. Connolly, W. E. Sause, and K. M. Ottemann. "Recombination-based in vivo expression technology identifies Helicobacter pylori genes important for host colonization". In: *Infection and Immunity* 76.12 (2008), pp. 5632–5644. DOI: 10.1128/IAI.00627-08.

[207] J. P. Russell, D. W. Chang, A. Tretiakova, and M. Padidam. *Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells*. 2006. DOI: 10.2144/000112150.

[208] C. Mcbride and D. D. Vecchio. "Resource Sensor Design for Quantifying Resource Competition in Genetic Circuits". In: *Proceedings of the 58th IEEE Conference on Decision and Control*. (2018).

[209] D. Liberzon. *Switching in Systems and Control*. Springer, 2003.

[210] M. K. Thomason and G. Storz. "Bacterial Antisense RNAs: How Many Are There, and What Are They Doing?" In: *Annual Review of Genetics* 44.1 (2010), pp. 167–188. DOI: 10.1146/annurev-genet-102209-163523.

[211] H.-H. Huang, Y. Qian, and D. D. Vecchio. "A quasi-integral controller for adaptation of genetic modules to variable ribosome demand". In: *Nature Communications* 9.5415 (2018). DOI: 10.1101/336271.

[212] J. E. Bowyer, V. Hsiao, W. W. Wong, and D. G. Bates. "Mechanistic modelling of a recombinase-based two-input temporal logic gate". In: *Engineering Biology* 1 (2017), pp. 40–50. DOI: 10.1049/enb.2017.0006.

[213] C. C. Samaniego and E. Franco. "An ultrasensitive motif for robust closed loop control of biomolecular systems". In: *2017 IEEE 56th Annual Conference on Decision and Control, CDC 2017* (2017), pp. 5334–5340. DOI: 10.1109/CDC.2017.8264449.

[214] J. E. Ferrell and S. H. Ha. "Ultrasensitivity part I: Michaelian responses and zero-order ultrasensitivity". In: *Trends in Biochemical Sciences* 39.10 (2014), pp. 496–503. DOI: 10.1016/j.tibs.2014.08.003. arXiv: NIHMS150003.

[215] F. Delvigne, H. Pecheux, and C. Tarayre. "Fluorescent Reporter Libraries as Useful Tools for Optimizing Microbial Cell Factories: A Review of the Current Methods and Applications". In: *Frontiers in Bioengineering and Biotechnology* 3.September (2015), pp. 1–8. DOI: 10.3389/fbioe.2015.00147.

[216] B. F. Brehm-stecher and E. A. Johnson. "Single-Cell Microbiology: Tools, Technologies, and Applications Single-". In: *Microbiology and Molecular Biology Reviews* 68.3 (2004), pp. 538–559. DOI: 10.1128/MMBR.68.3.538.

[217] D. A. Siegele and J. C. Hu. "Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations". In: *Proceedings of the National Academy of Sciences* 94.15 (1997), pp. 8168–8172. DOI: 10.1073/pnas.94.15.8168.

[218] B. P. Tracy, S. M. Gaida, and E. T. Papoutsakis. "Flow cytometry for bacteria: Enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes". In: *Current Opinion in Biotechnology* 21.1 (2010), pp. 85–99. DOI: 10.1016/j.copbio.2010.02.006.

[219] O. K. Silander, N. Nikolic, A. Zaslaver, A. Bren, I. Kikoin, U. Alon, and M. Ackermann. "A genome-wide analysis of promoter-mediated phenotypic noise in Escherichia coli". In: *PLoS Genetics* 8.1 (2012). DOI: 10.1371/journal.pgen.1002443.

[220] P. Billingsley. *Probability and Measure - Third Edition*. Wiley, 1995, p. 608. DOI: 10.1007/s13398-014-0173-7.2. arXiv: arXiv:1011.1669v3.

[221] N. L. Johnson, S. Kotz, and N Balakrishnan. *Continuous univariate distributions. Vol. 1. Second edition. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics*. Wiley, 1994.

[222] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden. "Regulation of noise in the expression of a single gene". In: *Nature Genetics* 31.1 (2002), pp. 69–73. DOI: 10.1038/ng869. arXiv: NIHMS150003.

[223] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko. "Ubiquity of log-normal distributions in intra-cellular reaction dynamics". In: *Biophysics* 1 (2005), pp. 25–31. DOI: 10.2142/biophysics.1.25.

[224] T. Forster. "Zwischenmolekulare energiewanderung und fluoreszenz". In: *Annalen der physik* 437.1-2 (1948), pp. 55–75. DOI: 10.1007/BF00585226. arXiv: arXiv:1208.5721.

[225] B. Hochreiter, A. P. Garcia, and J. A. Schmid. "Fluorescent Proteins as Genetically Encoded FRET Biosensors in Life Sciences". In: *Sensors* 15 (2015), pp. 26281–26314. DOI: 10.3390/s151026281.

[226] C. Zhang, Z. H. Wei, and B. C. Ye. "Imaging and tracing of intracellular metabolites utilizing genetically encoded fluorescent biosensors". In: *Biotechnology Journal* 8.11 (2013), pp. 1280–1291. DOI: 10.1002/biot.201300001.

[227] V. Stein and K. Alexandrov. "Synthetic protein switches: Design principles and applications". In: *Trends in Biotechnology* 33.2 (2015), pp. 101–110. DOI: 10.1016/j.tibtech.2014.11.010. arXiv: arXiv:1011.1669v3.

[228] E. M. W. M. Van Dongen, L. M. Dekkers, K. Spijker, E. W. Meijer, L. W. J. Klomp, and M. Merkx. "Ratiometric fluorescent sensor proteins with subnanomolar affinity for Zn(II) based on copper chaperone domains". In: *Journal of the American Chemical Society* 128.33 (2006), pp. 10754–10762. DOI: 10.1021/ja0610030.

[229] A. Pai and L. You. "Optimal tuning of bacterial sensing potential". In: *Molecular Systems Biology* 5.286 (2009), pp. 1–11. DOI: 10.1038/msb.2009.43.

[230] V. Shahrezaei, J. F. Ollivier, and P. S. Swain. "Colored extrinsic fluctuations and stochastic gene expression". In: *Molecular Systems Biology* 4.196 (2008), pp. 1–9. DOI: 10.1038/msb.2008.31. arXiv: 0809.2973.

[231] R. Gao, Y. Tao, and A. M. Stock. "System-level mapping of Escherichia coli response regulator dimerization with FRET hybrids". In: *Molecular Microbiology* 69.6 (2008), pp. 1358–1372. DOI: 10.1111/j.1365-2958.2008.06355.x.

[232] L. Sanford and A. Palmer. *Recent Advances in Development of Genetically Encoded Fluorescent Sensors*. 1st ed. Vol. 589. Elsevier Inc., 2017, pp. 1–49. DOI: 10.1016/bs.mie.2017.01.019.

[233] D. W. Piston and G. J. Kremers. "Fluorescent protein FRET: the good, the bad and the ugly". In: *Trends in Biochemical Sciences* 32.9 (2007), pp. 407–414. DOI: 10.1016/j.tibs.2007.08.003.

[234] J. Potzkei, M. Kunze, T. Drepper, T. Gensch, K.-e. Jaeger, and J. Büchs. "Real-time determination of intracellular oxygen in bacteria using a genetically encoded FRET-based biosensor". In: *BMC Biology* 10.1 (2012), p. 28. DOI: 10.1186/1741-7007-10-28.

[235] A. Hoppe, K. Christensen, and J. A. Swanson. "Fluorescence Resonance Energy Transfer-Based Stoichiometry in Living Cells". In: *Biophysical Journal* 83.6 (2002), pp. 3652–3664. DOI: 10.1016/S0006-3495(02)75365-4.

[236] R. D. Dar, B. S. Razooky, L. S. Weinberger, C. D. Cox, and M. L. Simpson. "The low noise limit in gene expression". In: *PLoS ONE* 10.10 (2015), pp. 1–21. DOI: 10.1371/journal.pone.0140969.

[237] L. H. Lindenburg, M. Malisauskas, T. Sips, L. V. Oppen, S. P. W. Wijnands, S. F. J. V. D. Graaf, and M. Merkx. "Quantifying Stickiness: Thermodynamic Characterization of Intramolecular Domain Interactions To Guide the Design of Forster Resonance Energy Transfer Sensors". In: *Biochemistry* 53 (2014), pp. 6370–6381. DOI: 10.1021/bi500433j.

[238] A. Goni-Moreno, I. Benedetti, J. Kim, and V. De Lorenzo. "Deconvolution of Gene Expression Noise into Spatial Dynamics of Transcription Factor-Promoter Interplay". In: *ACS Synthetic Biology* 6.7 (2017), pp. 1359–1369. DOI: 10.1021/acssynbio.6b00397.

[239] P. Sharma, R. Varma, R. C. Sarasij, K. Gousset, G Krishnamoorthy, M. Rao, S. Mayor, Ira, K. Gousset, G Krishnamoorthy, M. Rao, and S. Mayor. "Nanoscale organization of multiple GPI-anchored proteins in living cell membranes." In: *Cell* 116.4 (2004), pp. 577–89. DOI: 10.1016/S0092-8674(04)00167-9.

[240] A. N. Bader, E. G. Hofman, J. Voortman, P. M. Van Bergen En Henegouwen, and H. C. Gerritsen. "Homo-FRET imaging enables quantification of protein cluster sizes with subcellular resolution". In: *Biophysical Journal* 97.9 (2009), pp. 2613–2622. DOI: 10.1016/j.bpj.2009.07.059.

[241] Y. Zhao, Q. Hu, F. Cheng, N. Su, A. Wang, Y. Zou, H. Hu, X. Chen, H. M. Zhou, X. Huang, K. Yang, Q. Zhu, X. Wang, J. Yi, L. Zhu, X. Qian, L. Chen, Y. Tang, J. Loscalzo, and Y. Yang. "SoNar, a Highly Responsive NAD+/NADH Sensor, Allows High-Throughput Metabolic Screening of Anti-tumor Agents". In: *Cell Metabolism* 21.5 (2015), pp. 777–789. DOI: 10.1016/j.cmet.2015.04.009.

[242] E. Toprak, A. Veres, J. B. Michel, R. Chait, D. L. Hartl, and R. Kishony. "Evolutionary paths to antibiotic resistance under dynamically sustained drug selection". In: *Nature Genetics* 44.1 (2012), pp. 101–105. DOI: 10.1038/ng.1034.

[243] E. M. Zhao, Y. Zhang, J. Mehl, H. Park, M. A. Lalwani, J. E. Toettcher, and J. L. Avalos. "Optogenetic regulation of engineered cellular metabolism for microbial chemical production". In: *Nature Publishing Group* 555.7698 (2018), pp. 683–687. DOI: 10.1038/nature26141. URL: http://dx.doi.org/10.1038/nature26141.

[244] A. Maia Chagas, L. L. Prieto-Godino, A. B. Arrenberg, and T. Baden. "The €100 lab: A 3D-printable open-source platform for fluorescence microscopy, optogenetics, and accurate temperature control during behaviour of zebrafish, Drosophila, and Caenorhabditis elegans". In: *PLoS Biology* 15.7 (2017), pp. 1–21. DOI: 10.1371/journal.pbio.2002702.

[245] H. Wang and Y. T. Yang. "Mini Photobioreactors for in Vivo Real-Time Characterization and Evolutionary Tuning of Bacterial Optogenetic Circuit". In: *ACS Synthetic Biology* 6.9 (2017), pp. 1793–1796. DOI: 10.1021/acssynbio.7b00091.

[246] E. J. Olson, L. A. Hartsough, B. P. Landry, R. Shroff, and J. J. Tabor. "Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals." In: *Nature methods* 11.4 (2014), pp. 449–55. DOI: 10.1038/nmeth.2884.

[247] B. G. Wong, C. P. Mancuso, S. Kiriakov, C. J. Bashor, and A. S. Khalil. "Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER". In: *Nature Biotechnology* 36.7 (2018), pp. 614–623. DOI: 10.1038/nbt.4151.

[248] E. Toprak, A. Veres, S. Yildiz, J. M. Pedraza, R. Chait, J. Paulsson, and R. Kishony. "Building a morbidostat: An automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition". In: *Nature Protocols* 8.3 (2013), pp. 555–567. DOI: 10.1038/nprot.2013.021.

[249] S. A. Hoffmann, C. Wohltat, K. M. Müller, and K. M. Arndt. "A user-friendly, low-cost turbidostat with versatile growth rate estimation based on an extended Kalman filter". In: *PLoS ONE* 12.7 (2017), pp. 1–15. DOI: 10.1371/journal.pone.0181923.

[250] B Van den Bergh, T Swings, M Fauvart, and J Michiels. "Experimental Design , Population Dynamics , and Diversity in Microbial Experimental Evolution". In: *Microbiology and Molecular Biology Reviews* 82.3 (2018), pp. 1–54. DOI: 10.1128/MMBR.00008-18CE.

[251] LEDEngin. *LZ7 LED Emitter*. 2019. URL: http://www.ledengin.com/products/emitters{\#}LZ7.

[252] RayVio. *XR Series UV Emitter*. 2019. URL: https://rayvio.com/xrseries/.

[253] Melexis. *Digital plug & play infrared thermometer in a TO-can*. 2019. URL: https://www.melexis.com/en/product/MLX90614/Digital-Plug-Play-Infrared-Thermometer-TO-Can.

[254] AMS. *AS7341 11-Channel Spectral Color Sensor*. 2019. URL: https://ams.com/as7341.

[255] BeagleBoard-Foundation. *Beaglebone Black*. 2019. URL: https://beagleboard.org/black.

[256] A. Ronacher. *Flask (A Python Microframework)*. 2019. URL: http://flask.pocoo.org/.

[257] B. G. Wong, C. P. Mancuso, S. Kiriakov, C. J. Bashor, and A Khalil. "A generalizable experimental framework for automated cell growth and laboratory evolution". In: *BioRxiv* (2018).