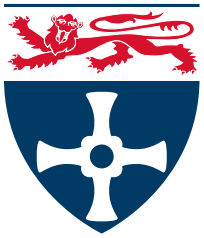
Exploiting contextual effects on synthetic biological constructs for the design of genetic and electrogenetic circuits.

Lewis Grozinger
Newcastle University
School of Computing
l.grozinger2@ncl.ac.uk



A thesis submitted for the degree of
Doctor of Philosophy
in September 2022

Abstract

Synthetic biologists build novel biological networks to perform predefined functions. The majority of work in this field has concentrated on engineering synthetic gene networks in bacteria, which coopt the host's genetic regulatory machinery for applications including biosensing, bioremediation and biocomputing.

The engineering of these synthetic gene networks is an expensive and time-consuming process that involves multiple iterations of a design-build-test-learn (DBTL) cycle. Often the results are impressive, and the synthetic gene network performs very well, but only in the environment in which they were engineered originally. The core problem is that the performance of a synthetic gene network relies fundamentally on factors external to the network itself — the context that the network is placed in. This includes for example the activity of the host's transcriptional machinery, or the genetic material close to the network, or the specific growth rate of the host.

Here I use a case study of a well-known synthetic gene network library to demonstrate that qualitative differences in network performance arise from changes in context and degrade the quality of the library. However I also show advantages to considering the context as a parameter of the library, and a cross-context library is of better quality than the original according to the metrics developed here.

At the population level it is often space that dominates context and spatial gradients may exist in any number of important factors. Bacteria such as *Escherichia coli* have developed chemotaxis systems to provide sensing of spatial gradients and motility to "solve" this problem for individuals. But it is not always desirable or possible for individuals to simply relocate. Natural systems that operate at the level of a populations often operate with space as a parameter, where individuals specialise and differentiate to contribute effectively to the overall function of the system.

In this work I take one such system, the electroactive biofilm, and show how two classical synthetic gene networks, the genetic toggle switch and the repressilator, can be designed not only to deal with spatial gradients, but also to fundamentally rely on them. The mathematical models developed are implemented in software and are modular, in the sense that other synthetic gene networks could easily be modeled using the same software. The results provide a foundation for moving toward more sophisticated work in synthetic electrogenetic networks.

Acknowledgment

I owe gratitude to my supervisory team, Ángel Goñi-Moreno, Elizabeth Heidrich, Harold Fellermann and Anil Wipat for their supervision and support.

I am also grateful for the support (often not academic) given by members of both the ICOS group at the Newcastle university school of computing, and the Biocomputation lab at the centre for biotechnology and plant genomics of the technical university of Madrid.

Thank you also to Natalio Krasnogor and Chris Barnes for their attention in assessing the work in this thesis.

Paolo Zuliani, Ruud Stoof, Fedor Shmarov, Nishant Singh, Yiming Huang, Matthew Crowther and Jack Jennings all warrant mentioning, even if they don't realise or recognise it.

I cannot give an exhaustive account of my debts of gratitude, but I encourage my creditors to come and collect.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This chapter outlines the content and an overall theme of the thesis.

A typical bacterial genome contains several thousand genes. Levels of expression of subsets of these genes can be strongly correlated and make up a gene network. These natural gene networks control the behaviour of bacteria, helping them to adapt to changes in their environment by sensing a wide variety of signals and adjusting expression of their genes.

Synthetic biology engineers novel (new to nature) gene networks which sense particular signals as input and adjust their gene expression in a predefined manner. A crucial aspect of engineering these networks is modeling of the processes that regulate gene expression to simulate the networks *in silico*. These simulations make predictions about performance of the network and are used to select the network topologies, genes and other genetic components that are likely to produce the best results.

1.1 Problem statement

In reality the expression of the genes of the network depends not only on the direct influence of a particular input signal, but also on the combined indirect influence and complex interactions of many other factors that define a 'context' for the synthetic gene network.

Therefore identical synthetic gene networks will perform differently in different contexts. Context can negatively affect the modularity and reusability of synthetic gene networks which are key aims for the field of synthetic biology.

Context cannot be disregarded because it is impossible (and indeed undesirable) to turn synthetic biological networks into completely closed systems. Context is therefore an aspect of synthetic biology that must be tackled.

1.2 Solution approach

This work uses modeling to look at how context affects the performance of synthetic gene networks and explores approaches for dealing with context in synthetic biology.

One such approach is to measure the effect of context by analysis of experimental data about identical synthetic gene networks in different contexts. Once the performance in different contexts is characterised, these characterisations can be used in existing synthetic

biological design workflows. In Chapter 3 this analysis is done with a library of well-known genetic logic gates. Although this approach relies on the collection of experimental data in each context, which can be an expensive process, it is shown here *in silico* how using cross-context libraries of genetic logic gates can lead to better libraries and potential added value in the design process.

Another approach is to model not only the synthetic gene network but also the contextual factors that affect its performance. Of course modeling all of the context is intractable, so instead these models should be focused on contextual factors that are considered most significant. In Chapter 4 and 5 two classical synthetic gene networks, the toggle switch and repressilator, are redesigned to have a significant dependence on the redox environment in an electroactive biofilm. The interaction between the biofilm and these gene networks is modeled in order to determine design principles that allow these systems to operate well in spite of differences in context. It is shown *in silico* how these models can predict how to modify synthetic biological designs to recover functionality that is lost due to context changes.

1.3 Thesis outline

This thesis is organised as follows:

- Chapter 2 provides background on synthetic biology with a focus on well studied gene networks such as genetic logic gates, toggle switches and oscillators. It introduces some modeling frameworks that are typically used to describe these systems and will be used later for the modeling in this work. What is meant by 'context' in synthetic biology is also more precisely defined. Exoelectrogenic bacteria and the electroactive biofilm they form are introduced in this chapter, to be used in Chapters 4 and 5 as an example of an environment where space produces significant contextual effects.

- Chapter 3 analyses the performance of a genetic NOT logic gate library using experimental data collected in seven different contexts. I did not generate this data, but it was shared with me by a collaborator. Two contextual parameters are considered, the host organism and the plasmid vector used to engineer the library. Models are used to quantify the difference in performance between contexts, and the relative importance of the two contextual parameters is discussed based on these

measures. The notion of library quality is defined and computed in various ways. These include the number of gates in the library whose individual performance is satisfactory up to a threshold tolerance, the number of pairs of gates from the library which can be connected together, and the maximum length of a chain of such pairs of gates. These measures of quality are used to evaluate the utility of mixing different contexts to create new libraries with improved quality. A previously described statistical model is adapted and fitted to data about constitutive gene expression in three of contexts considered, with good results. The parameters of the model have a physical interpretation which suggests a mechanistic explanation for the observed differences in expression between the contexts. Finally, an existing combinatorial genetic circuit design framework built for one context is used to engineer a simple logic circuit using the characterisation of experimental data from a different context.

- Chapter 4 develops a partial differential equation modeling framework for an electroactive biofilm into which synthetic gene networks can be embedded, so as to predict the performance of these networks in the context of the electroactive biofilm. The framework is used for the design of the electrogenetic toggle switch, an adaptation of the genetic toggle switch that is a hybrid electronic-genetic system. The electroactive biofilm is seen to be a heterogeneous environment with implications for the electrogenetic toggle switch. It is shown that designs which exhibit the required behaviour in homogeneous environments do not in general operate well in the presence of biofilm gradients. Inclusion of the contextual effect of space in the biofilm in this model is used to predict changes to the parameters of the genetic portion of the switch which make the system more robust to these gradients. Nevertheless, the model highlights further challenges that emerge as a result of spatial gradients that also change over time.

- Chapter 5 uses the same electroactive biofilm model for a case study where a genetic oscillator is designed for operation in the biofilm. It is shown how the biofilm might be used to address the problem of phase drift in the individual oscillators. This is an example of how context can be used to improve the performance of a synthetic biological network.

- Chapter 6 is a summary and general discussion of the presented results.

1.4 Publications arising from this thesis

Parts of the work in this thesis have previously been published and are listed here.

- Grozinger, L., Amos, M., Gorochowski, T.E. et al. Pathways to cellular supremacy in biocomputing. Nat Commun 10, 5250 (2019). https://doi.org/10.1038/s41467-019-13232-z *
- Tas, H., Grozinger, L., Stoof, R. et al. Contextual dependencies expand the re-usability of genetic inverters. Nat Commun 12, 355 (2021). https://doi.org/10.1038/s41467-020-20656-5 *
- Tas, H., Grozinger, L., Goñi-Moreno, A., de Lorenzo, V., Automated design and implementation of a NOR gate in Pseudomonas putida, Synthetic Biology, Volume 6, Issue 1, (2021), https://doi.org/10.1093/synbio/ysab024
- Grozinger, L., Heidrich, T., Goñi-Moreno, A., An electrogenetic toggle switch design, BioRxiv (2022), https://doi.org/10.1101/2022.05.19.492718 * †

A significant amount of software was written as part of this work and is made publicly available.

- For the analysis of genetic logic gate libraries performed in Chapter 3 — https://github.com/lgrozinger/Pyolin.jl
- For the modeling and bifurcation analysis of the electroactive biofilm systems in Chapters 4 and 5 — https://github.com/Biocomputation-CBGP/An-electrogenetic-toggle-switch-desig

Other publications I have authored or contributed to but do not form directly a part of this thesis.

- Crowther, M., Grozinger, L., Pocock, M. et al, ShortBOL: A Language for Scripting Designs for Engineered Biological Systems Using Synthetic Biology Open Language (SBOL), ACS Synthetic Biology (2020), https://doi.org/10.1021/acssynbio.9b00470

*I am the first author of this publication
†At the time of writing this article is accepted in Microbial Biotechnology and awaiting publication

- Stoof, R., Grozinger, L., Tas, H., Goñi-Moreno, A., FlowScatt: enabling volume-independent flow cytometry data by decoupling fluorescence from scattering, BioRxiv (2020), https://doi.org/10.1101/2020.07.23.217869

Chapter 2

Background

Parts of this chapter are adaptations of Grozinger et al [36], a review paper which presents promising approaches towards increasing complexity and utility in cellular biocomputing, of which I was the first author.

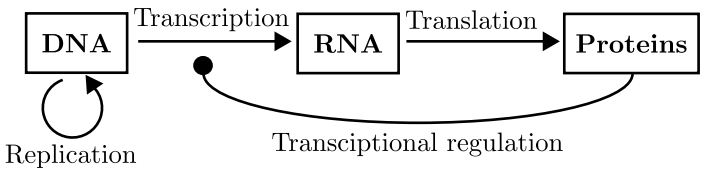


Figure 2.1: The central dogma of molecular biology. Genetic information is preserved by replication of DNA. Genetic information is expressed by transfer from DNA to RNA to proteins through the molecular processes of transcription and translation. Proteins can go on to control the expression of other genetic information (or their own information). One way they do this is by intervening in the transcription process.

Synthetic biology is the application engineering principles to rationally design novel biological systems [2]. These synthetic biological systems find application in bioproduction [17], bioremediation [21] and medicine [75] and biocomputation [1], as well as being of academic interest in the study of fundamental cellular processes.

The methodology used for building synthetic biological systems borrows from more traditional engineering disciplines and increasingly resembles a DBTL cycle [12]. In the design stage the synthetic biologist use their existing knowledge to conceive of a system *a priori* which might perform the desired function. At this stage the system is accompanied by a model which predicts how well the system will meet its objective. The build stage utilises technique from the biological and molecular sciences to realise the design, or often a library of the best design candidates. Testing characterises the actual performance of the system, the results of which are analysed in the learning stage, where new knowledge about the processes in the system is gained. This learning is used to adjust the model such that it's predictions are closer to the real performance of the tested system, and the design is modified so that the model's prediction for the new design more closely resembles the desired function. The cycle is started over again and is iterated until the performance of the synthetic biological system is satisfactory.

The focus of this thesis is on the design stage of bacterial synthetic biology, using gene networks, to perform biocomputations and to construct networks which display simple but interesting dynamical behaviours.

2.1 Gene networks

The flow of genetic information can be described using the central dogma of molecular biology shown in Figure 2.1. Information flows from DNA to RNA and to proteins via the

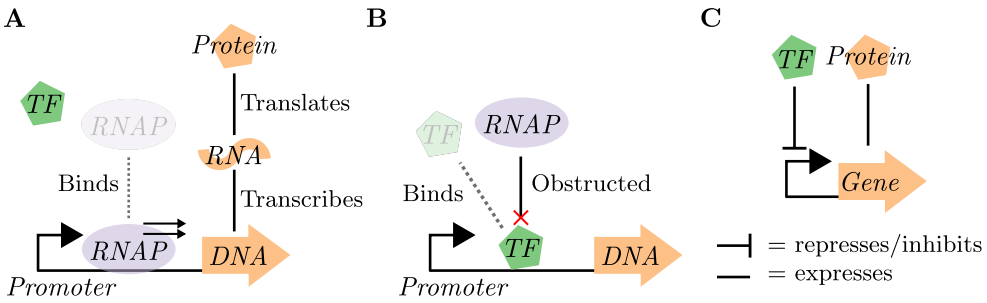


Figure 2.2: A schematic of a mechanism for transcriptional regulation by a repressive transcription factor. In **A** expression of the gene is on, since RNAP can bind to the promoter region to initiate transcription. In **B** the TF has bound to the promoter region and obstructs the recruitment of RNAP, stopping transcription and turning gene expression off. In **C** a more abstract representation of this repression interaction is shown.

processes of transcription and translation. Transcription transfers information from DNA molecules to RNA molecules, and translation transfers data from RNA molecules to newly synthesised proteins. The overall flow of information from DNA to protein is called the 'expression' of a gene.

The expression of one gene can effect the expression of another. There is more than one way this can happen, but in this thesis the focus is on the transcriptional regulation shown in Figure 2.1 as the feedback from protein to transcription. Genes can express proteins called TFs which bind to specific regions of DNA known as promoters. Promoters initiate the transcription of downstream DNA by recruiting RNAP, the enzyme that synthesises RNA from DNA. TF binding to promoters affects the affinity of RNAP for the promoter region and the recruitment of RNAP. Activators are TFs that encourage the recruitment of RNAP and so up regulate (activate) the expression of a gene. Repressors are TFs that obstruct the recruitment of RNAP, and so down regulate (repress) the level of expression of a gene.

The diagram of Figure 2.2 shows how repression works for some types of TF. Usually RNAP binds to the promoter region and initiates transcription of the DNA downstream, leading to the synthesis of RNA, which can be translated to express the protein for which the gene encodes. This is sketched out in Figure 2.2A. In the presence of TF with some affinity for DNA in the promoter region, TF may bind to the promoter region and prevent the binding of RNAP. Without RNAP transcription is not initiated and the gene is not expressed, as in Figure 2.2B. A more compact representation for this kind of interaction is shown in Figure 2.2C, and this sort of diagram will be used for representing gene networks throughout the rest of this work.

A collection of genes whose levels of expressions are connected in this way is a gene network. Naturally occurring gene networks are responsible for much of the 'decision making' of bacteria, for example in order to adapt to stimuli from the environment. Synthetic biology has developed sophisticated tools for engineering synthetic versions of gene networks using transcriptional regulation to connect genes in ways that are novel or 'new-to-nature'. These synthetic biological networks display a wide range of dynamic behaviours that can be useful for a variety of applications.

Synthetic biological genetic logic gates

Previous work has used synthetic networks of genes to perform biocomputations, notably to implement combinatorial logic [59, 92, 81, 34]. These gene networks are designed as modular units called genetic logic gates that are composed into larger networks that implement arbitrary combinatorial logic circuits.

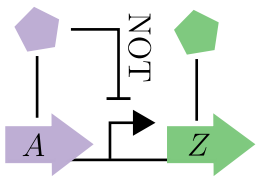
Two examples of genetic logic gates are shown in Figure 2.3. In Figure 2.3A a schematic of a genetic NOT gate is shown which implements the logical negation operation from Boolean algebra. This gate uses a single gene A and a single output gene Z. If A is highly expressed then expression of Z will be repressed. Conversely if A is not highly expressed then Z will be. If high expression is labeled **1** and low expression as **0**, the inputs and outputs can be tabulated as in Figure 2.3A.

The second example is of a NOR logic gate. As shown in Figure 2.3B this gate is composed from two NOT gates controlling the same output gene Z. The logical operation it performs is shown in the accompanying table, but in summary, expression of Z is only turned on if both inputs A and B are not expressed.

Connected genetic NOR gates should in principle be able to perform any combinatorial logic computation desired due to their universality. Therefore a key goal of synthetic biology is to make these kind of synthetic gene networks into modular devices so that they can be composed and reused to construct increasingly complex networks with predefined behaviours. In [59] it was shown that NOR gates could indeed be connected together with TFs in order to perform more complex combinatorial logic computations.

However combinatorial logic circuits require orthogonality of the logic gates in order to circuits to function as expected. Unwanted interactions (also known as cross-talk) between gates can lead to failure of the computation. This means that the same TFs cannot be

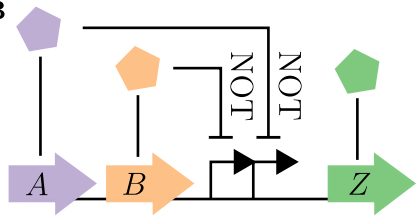
A



Formula: $Z = \neg A$

Input A	Transcription	Output Z
0	High	**1**
1	Low	**0**

B



Formula: $Z = \neg(A \vee B)$

Input A	Input B	Transcription	Output Z
0	**0**	Very High	**1**
0	**1**	High	**1**
1	**0**	High	**1**
1	**1**	Low	**0**

Figure 2.3: In **A** a gene network designed to use transcriptional repression to perform a logical negation operation $\neg$ on the expression of the input gene A to give expression of the output gene $Z = \neg A$. Therefore if A is highly expressed, Z is not, and vice versa, as shown in the table of **A**. As a modular device this is a genetic NOT gate. In **B** two of these NOT gates are combined with a common output Z, in order to construct another modular device, the genetic NOR gate. This network has two input genes A and B. The network is designed so that Z is highly expressed unless both A and B are highly expressed, performing the operation $Z = \neg(A \vee B)$, as shown in the table of **B**.

used in multiple gates in the same network, to prevent unwanted interactions between gates. This motivates the construction of libraries of genetic logic gates which perform the same operation, but which maintain their orthogonality through the use of different TFs [77].

A well known example of such a library was introduced by Nielsen et al as part of the *Cello* genetic design automation software [59]. The study characterised and optimised 20 genetic NOT gates for use in *E. coli* as well as developed software which could transform descriptions of combinatorial logic circuits into specifications of synthetic gene networks which composed the 20 gates to implement the circuits. This approach allowed the authors to automatically generate genetic designs for combinatorial logic circuits consisting of several gates [59].

Flow cytometry

The testing stage of the DBTL requires experimental data collected about the performance of the synthetic gene network. In most cases this amounts to measuring the levels of gene expression from the network. Flow cytometry is an experimental technique which is widely

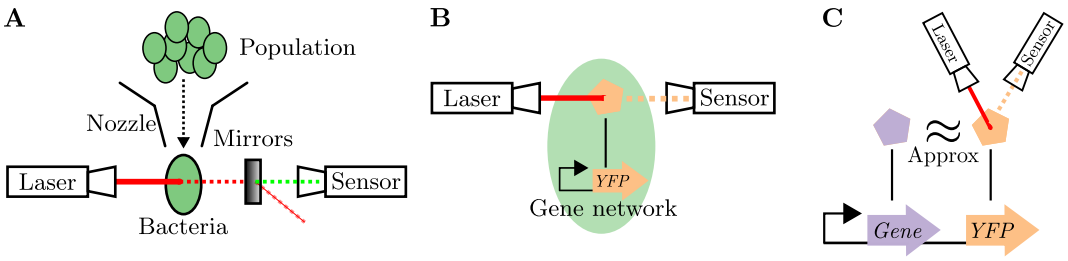


Figure 2.4: Flow cytometry is a fundamental experimental technique for measuring expression in synthetic gene networks. In **A** a schematic is shown of a flow cytometer. The population to be sampled is guided using microfluidics through a light beam of known wavelength. The light is scattered by the sample, filtered by a series of mirrors, and measured by a optical sensor. **B** illustrates how flow cytometers can be used to measure expression of fluorescent proteins in bacteria, and **C** shows how this technique is often used to measure expression of other genes, using a co-expressed fluorescent gene as a proxy.

used for this purpose in synthetic biology[91].

Flow cytomers have a fluidic device which can direct bacterial cells to pass through the beam of a laser. The bacterial cells scatter the light of this laser which is separated by a system of mirrors into wavebands whose intensity is measured by optical sensors. Flow cytometers are capable of taking many samples (measurements of individual bacteria) of large populations very quickly [91]. A rough schematic of this setup is shown in Figure 2.4A.

Bacteria can be genetically engineered to express fluorescent proteins such as YFP. The laser of the flow cytometer can also excite fluorescent proteins, and when excited these proteins emit light in specific wavebands that can detected by the optics of the flow cytometer (Figure 2.4B). As such the flow cytometer can be used to measure the expression of fluorescent proteins in bacterial cells [26].

In order to measure expression of a gene which is not fluorescent a synthetic gene network can be engineered to co-express a fluorescent gene with the gene of interest as in Figure 2.4C. The expression of the fluorescent protein can then be assumed to be proportional to the expression of the gene of interest, and fluorescence measurements can be used as a proxy or approximation for the expression of the gene of interest. This approach underpins much of the work characterising the performance of synthetic gene networks.

2.2 Modeling gene networks

A model of a gene network is an analog for the real network under study. Modeling is a fundamental part of synthetic biology, particularly in the design and learn stages of the DBTL cycle, where modeling can save the effort and expense associated with wet lab experiments [72]. In order to be useful the model must make indicative predictions about the performance of gene networks, and generate useful counterfactual predictions relevant to the design of the synthetic gene networks and testable through experiment.

It is for this reason that design choices available to the synthetic biologist inform the formulation of the model and in particular which features to include and at what level of abstraction. For example, if one of the choices to be made in the design is which promoter region to use, the model should be able to predict the consequences of that choice, that is it should answer the question "If I built my gene network with this promoter instead of this other one, what would happen?". However most synthetic biologists do not consider distance between the genetic components as a design parameter, so this detail is often (appropriately) left out of the model [72]. In contrast Stoof et al [79] do consider the distance between their genetic components as designable, and accordingly diffusion of TF molecules is included in their model.

Ordinary differential equations for gene network models

Concerning mathematical representations of gene network models, a variety of formalisms and frameworks can be used for be used to meet a range of model requirements. In this work I make extensive use of ordinary differential equations as a formalism to mathematically model gene expression and other reaction networks. ODEs are used for deterministic models of gene expression with continuous state variables. The state variables keep track of the concentrations of each of the species in the model, for example the proteins that are being expressed, or the RNA that is being transcribed. The equations describe the rate of change of the state variables as a function of their concentrations:

$$\partial_t \mathbf{u} = f(\mathbf{u}(t); \mathbf{p}) \tag{2.1}$$

where t is an independent variable for time, $\mathbf{u}(t)$ is the vector of the state variables at time t, $\mathbf{u}(t) = (u_1, ..., u_n)$ and $\mathbf{p}$ is a vector of parameters for the system. f is a function which

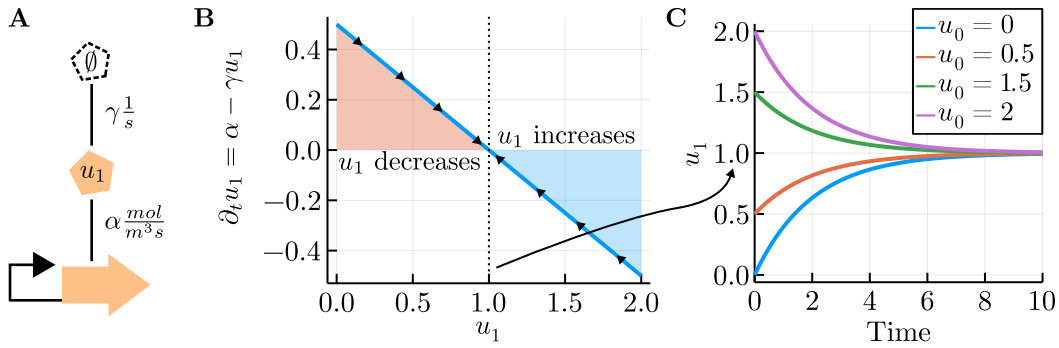


Figure 2.5: Modeling expression of a single gene u_1 with an ODE. The model (**A**) uses two parameters α and γ, the expression rate and degradation/dilution rate, in a single differential equation. **B** plots the equation against the expression level. This line intersects with zero at a fixed point of the equation. In **C** Time course simulation of the equation shows that trajectories that start from different initial conditions converge toward that fixed point.

describes the rate of change of $\mathbf{u}(t)$ with respect to time. Solving the ODE means finding the functions $\mathbf{u}(t)$ which satisfy the equation. Actually, we usually want one particular function, the one where $\mathbf{u}(t) = \mathbf{u_0}$ where $\mathbf{u_0}$ are the initial concentrations or the 'starting state' of the system.

The approach is illustrated by application to a simple model of expression of a single gene in Figure 2.5. The model keeps track of the concentration of a protein u_1 over time, which increases as it's gene is expressed at a rate of α, and decreases as it is diluted or degraded proportionally to its concentration at rate γ. The balance of rates of these two processes is the overall rate of change of u_1 and is written as:

$$\begin{aligned} \partial_t u_1(t) &= \text{production rate} - \text{dilution rate} \\ &= \alpha - \gamma u_1(t) \end{aligned} \tag{2.2}$$

This equation is plotted in Figure 2.5B for $\alpha = \gamma = 0.5$. For $u_1 < 1$ it can be seen that the rate of change of u_1 is positive, so that the concentration of u_1 is increasing. For $u_1 > 1$ it can be seen that the rate of change of u_1 is negative, so that the concentration of u_1 is decreasing. An important point is $u_1 = 1$ at which the rate of change is 0. The concentration of u_1 is neither decreasing or increasing, and will no longer change over time. $u_1 = 1$ is called a fixed point of the system.

The idea is generalisable to higher dimensions, where a fixed point is defined as a $\mathbf{u}$ for which $\partial_t \mathbf{u} = \mathbf{0}$, where $\mathbf{0}$ is the vector of zeros. At a fixed point, none of the state variables

of the system are changing anymore, so the ODEs predict their values will remain the same forever (in the absence of some external perturbation).

ODEs can be used to plot trajectories of the state variables over time. This is shown in Figure 2.5C where the trajectory of u_1 is plotted. In the case of this model it can be seen that all the trajectories approach the fixed point at $u_1 = 1$ as time goes on. The exact trajectory taken depends only on the initial conditions u_0, since ODEs are a deterministic mathematical model.

Modeling activation and repression with Hill equations

To model more complex gene networks with more than one gene, Hill equations can be useful to describe mathematically the effects of repression and activation on the expression of the target genes. The Hill equation is derived from mass action kinetics and describes the relationship between the concentrations of bound and unbound ligands to proteins as a function of the ligand concentration x.

$$h(x) = \frac{\text{Unbound protein}}{\text{Total protein}} = \frac{x^n}{K^n + x^n} \tag{2.3}$$

where K is the concentration of ligand for which half the protein is bound ($h(x) = \frac{1}{2}$) and n is a hill coefficient.

The justification for use of a Hill function for repression and activation of gene expression considers TFs as ligands and the promoter regions of DNA as the protein to be bound. The process of binding and unbinding of TF is often assumed to take place at a much faster rate than genetic processes such and reach equilibrium quickly. Under this assumption $h(x)$ is usually assumed proportional to the transcription rate from the promoter region and to gene expression [15].

For $n > 1$ the Hill equation is used for modeling activation of gene expression by TF. For $n < -1$ the Hill equation can be used for modeling repression of gene expression by TF. $n = 1$ is a special case where the Hill equation reduces to the Michaelis Menten equation which is a well-used model of enzyme kinetics. Figure 2.6A plots examples of these classes of Hill equations.

The Hill equation is incorporated into the ODE model by modifying Equation 2.2 as

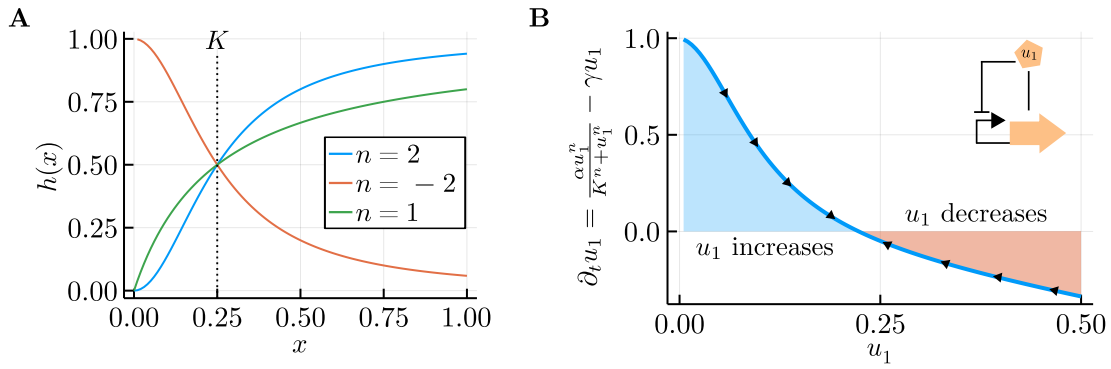


Figure 2.6: **A** plots examples of Hill equations for 3 different cases. For positive n (blue line), the Hill equation increases with x and so is useful for modeling activation of gene expression by a TF. For negative n (orange line) the Hill equation decreases with x and can model repression of gene expression by a TF. At $n = 1$ the Hill equation is the Michaelis Menten equation which is widely used in enzyme kinetics. **B** plots Equation 2.4 for $n = -2$, $\alpha = 1$, $K = 0.75$ and $\gamma = 0.1$. There is a single fixed point and it can be shown (by Descartes rule) that this is the only type of dynamical behaviour possible for this system.

follows:

$$\begin{aligned} \partial_t u_1(t) &= \alpha h(u_1(t)) - \gamma u_1(t) \\ &= \frac{\alpha u_1(t)^n}{K^n + u_1(t)^n} - \gamma u_1(t) \end{aligned} \tag{2.4}$$

Equation 2.4 is plotted in Figure 2.6B for $n = -2$, modeling negative feedback from u_1 to itself. This plot is similar to Figure 2.5B. There is still a single fixed point for the system, but now the rate of change is nonlinear in u_1. Nonlinearity can admit more complex (and interesting) dynamical behaviours such as bistability and oscillations, which are introduced in Sections 2.3 and 2.3.

Hill equations are routinely used as rates in all kinds of chemical reaction networks [58, 32], and are also a standard tool in synthetic biology for modeling transcriptional regulation in gene networks [65, 34, 86].

Partial differential equations for spatial models

ODEs described the rate of change of $\mathbf{u}(t)$, which only depended on a single independent variable t. In some models the state variables are dependent on several independent

variables. This often occurs in spatial models such as the biofilm model I develop in Chapter 4, where the concentrations of species in the model changes not only in time but also in space, for these models PDEs may be used as a mathematical description.

A common use case for PDEs is for the modeling of diffusion. For example if $\mathbf{u}(x,t)$ are state variables diffusing in a single dimension x, then Fick's second law is a PDE which describes their change over time:

$$\partial_t \mathbf{u}(x,t) = \mathbf{D}\partial_x^2 \mathbf{u}(x,t) \tag{2.5}$$

where $\mathbf{D}$ is the diffusion coefficient ('rate' of diffusion) and $\partial_x^2 \mathbf{u}(x,t)$ is the second order differential of $\mathbf{u}(x,t)$ with respect to x (the rate of change of the rate of change of $\mathbf{u}(x,t)$ with x). The PDEs used later in this thesis take this form.

Most PDEs cannot be easily solved analytically. Solving numerically involves discretisation of the spatial dimension using a finite difference scheme which approximates $\partial_x^2 \mathbf{u}(x,t)$. The approach I take for numerical simulation in this thesis is to discretise x into an evenly spaced grid of points Δx from each other as shown in Figure 2.7. I choose the second order central difference scheme:

$$\partial_x^2 \mathbf{u}(x,t) \approx \frac{\mathbf{u}(x-\Delta x,t) - 2\mathbf{u}(x,t) + \mathbf{u}(x+\Delta x,t)}{\Delta x^2} \tag{2.6}$$

Using this scheme the PDE can be transformed into a system of ODEs, one for each point on the grid, which approximates the original PDE.

Just as with systems of ODEs the initial conditions must be specified. In addition special care must also be taken at the boundaries of the spatial domain x. The boundary conditions used here are of two types, Neumann and Dirichlet. Neumann boundary conditions specify the derivative $\partial_x \mathbf{u}(x,t)$ at the boundary and can be used to describe flux across the boundary as shown in Figure 2.7. Dirichlet boundary specify the value of $\mathbf{u}(x,t)$ and are used modeling fixed conditions at the boundary.

PDEs are a common mathematical framework in synthetic biology when diffusion needs to be modeled. This often arises when modeling signaling between bacteria in multicellular systems[19], but also within bacterial cells when spatial effects are considered as important[79]. In this thesis they are used for mathematical modeling of the electroactive biofilm, following the example of previous studies in modeling bioelectrochemical systems

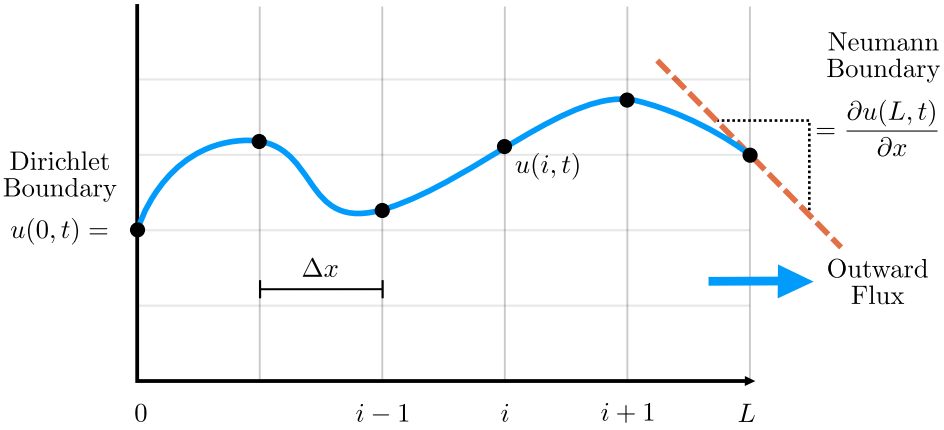


Figure 2.7: Finite difference discretisation of a spatial domain in order to solve a PDE numerically as a system of ODEs. The spatial dimensional x, which has boundaries at $x = 0$ and $x = L$ is split into a grid of points spaced Δx apart. An ODE describes the rate of change at each point i, using surrounding points to approximate the second derivative as in Equation 2.6. At the boundaries conditions are applied which can specify the value of $u(x,t)$, or its partial derivative with respect to x, depending on what happens at the boundary in the model.

[29, 67, 45].

Other approaches to modeling in synthetic biology

ODEs and PDEs are staples in modeling all kinds of physical systems including synthetic gene networks. ODEs in particular are extremely common, and it is relatively straightforward to formulate a system of ODEs for a set of chemical reactions from a gene network or other biochemical system. In addition an enormous variety of *in silico* tools are available for working with and solving ODEs numerically or which include ODEs as part of a suite of modeling tools(for example [70, 41, 93, 16]). Furthermore, ODEs were used to design and engineer many of the canonical synthetic gene networks such as the toggle switch [27] and the repressilator [23]. [37].

However, other approaches have been developed including some that address one of the primary limitations of ODE modeling. That is, in the limit of small copy numbers of RNA and protein, the stochasticity of the processes that contribute to genetic expression can have significant impact on the dynamical behaviours of gene networks. Discrete stochastic models are often used in these cases in order to predict the effect of noise on the time course of gene expression [72]. The family of Gillespie simulation algorithms are the most often used for simulating these kinds of models [33]. Simulation with these algorithms give

an ensemble of stochastic trajectories which can be used to predict statistical properties of gene expression.

Another aspect of discretisation becomes important when considering that synthetic biological systems often operate at the population level. Individual cells have their own internal dynamics as well as those through which they interact with their environment. In this case, heterogeneity, space, and complex interactions between individuals can shape population level behaviour and require multi-scale hierarchical descriptions of synthetic biological systems[]. Agent-based simulations are a natural approach to investigating these kinds of effects and platforms have been developed capable of mixing both deterministic and stochastic models at multiple scales[58, 10].

There are also examples where the choice of modeling framework can elucidate new kinds of dynamics for gene networks that go undetected when modeled using ODEs. For example, negative autoregulation in a single gene network can only produce a system with a single fixed point when modeling deterministically with ODEs. However it has been shown, using a modified version of the Gillespie algorithm, that the addition of stochastic delays to negative autoregulatory networks can produce oscillations [9]. The choice of mathematical framework is therefore of fundamental importance when studying the dynamics of gene networks.

2.3 Dynamics of gene networks

The dynamics of a system is the way that its state changes over time. In the previous section the idea of a fixed point of a system was introduced by saying that when a system arrives at a fixed point, its state no longer changes over time. This is a very simple kind of dynamics.

In gene networks the dynamics is the way in which the level of expression genes change over time. Generally there are two classes of dynamical behaviour of interest in synthetic gene networks, (multi)stability and oscillation, and these are the dynamics available to the synthetic biologist.

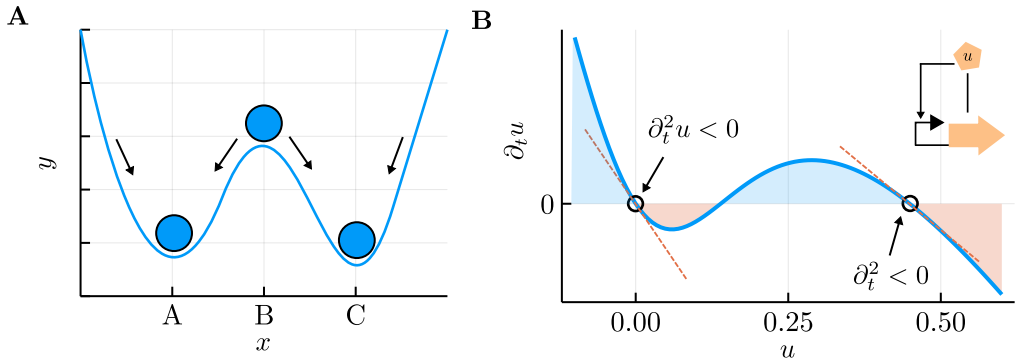


Figure 2.8: An illustration of bistability in dynamical systems. In **A** a one dimensional terrain is shown with 1 peak (B) and 2 valleys(A and C). A ball rolling along the terrain may be balanced at A, B or C, but will be unstable at B — it is easily pushed away. In contrast resting at A and B are stable — the ball will return to rest in the same position if perturbed. **B** shows how the same idea applies to positive feedback in gene expression. If the rate of change in expression is plotted there are 2 stable and 1 unstable fixed point. The nature of the stability of each is determined by the slope of the curve as it crosses the horizontal axis.

Bistability

Fixed points are classified as stable or unstable depending on the local behaviour of the system around the fixed point. The system tends to move toward fixed points that are stable away from those that are unstable. A good analogy is the ball moving along the one dimensional terrain shown in Figure 2.8A. The ball can be balanced in three positions A, B and C. It is clear that if the ball is balanced B, any small perturbation will push the ball out of balance, and the ball will not return spontaneously to B. In contrast, small perturbation to balls at A or C will be damped, and the ball will return to rest at its initial position and that these positions are stable.

This system is bistable, meaning that there are 2 stable fixed points of the system (and one unstable fixed point). Bistability in gene networks is achieved using positive feedback [50], as is illustrated in Figure 2.8B. In this example the rate of change of gene expression with positive feedback is plotted against the level of gene expression.As discussed in the previous section, fixed points are those where the curve intersects with the x-axis — where rate of change $\partial_t u(t) = 0$. Stable fixed points are those where the rate of change of the rate of change is negative at the fixed point — the curve crosses the x-axis with downwards slope. This condition can be written mathematically as $\partial_t^2 u(t) < 0$ at the fixed point.

Bistability is a fundamental dynamical behaviour both in natural and synthetic biology [5, 95]. In the context of biocomputing, bistability is particularly important in constructing

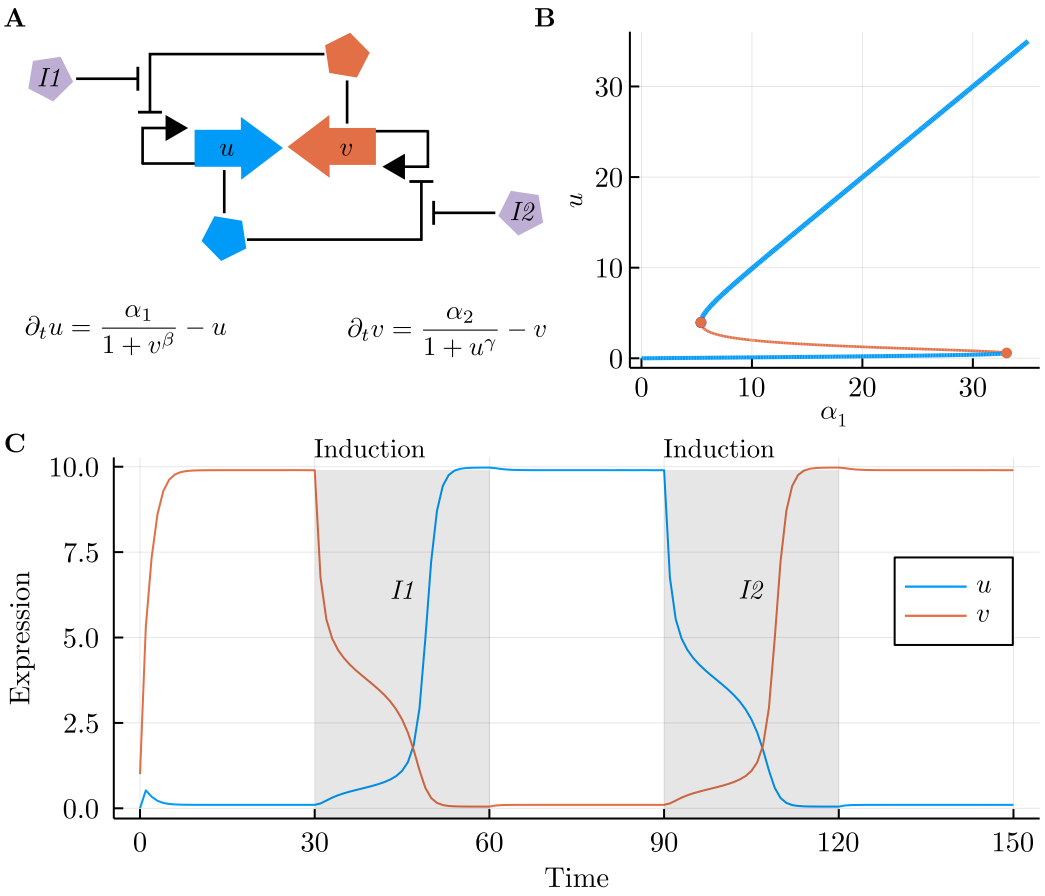


Figure 2.9: The genetic toggle switch is classic synthetic gene network. A schematic of the gene network in shown in **A**, where the mutual repression of the genes u and v produces bistability. In addition the repressive activity of u and v can be impaired by inducers I_1 and I_2, in order to flip the switch. **B** shows a bifurcation diagram of the system where parameter α_1 is varied and the fixed points of the system are plotted. **C** shows the toggle switch behaviour over time, including two inductions which flip the switch between the two stable states.

single-bit memories or transitions between discrete states of the system [18].

Genetic toggle switch

The genetic toggle switch was one of the first synthetic gene networks to be engineered in *E. coli* and is capable of exhibiting bistable dynamics. The networks is shown in Figure 2.9A and uses two genes that repress one another to achieve positive feedback and admit the possibility of bistability[27].

A key part of the study was a mathematical model of the switch consisting of two ODEs, the analysis of which predicts what kind of genetic components would be required for

bistability to emerge in the network. These two equations are used to simulate the time course operation of the switch in Figure 2.9.

A mathematical model is useful because not all gene networks with the topology of mutual inhibition are bistable. In order to save the effort of a trial and error approach to finding suitable networks, parameter values which produce bistability in the model can be identified *in silico*, and subsequently inform choices of suitable genetic components.

This can be done with a bifurcation analysis of the mathematical model, which is an analysis of how the system's dynamics change as a particular parameter is varied. A bifurcation diagram of the genetic toggle switch is shown in Figure 2.9B. The parameter that is varied is α_1, which is the maximal rate of expression of u. Plotted against α_1 are the branches of fixed points of the system and their stability. It can be seen that below $\alpha_1 \approx 5$ (left orange circle) the system has only one stable (thick lines) fixed point, so is monostable. Beyond this the dynamics of the system change to have 3 fixed points, two stable and on unstable (thin line). This bistability exists until $\alpha_1 \approx 33$, where two of the fixed points disappear and the system is monostable again. This kind of analysis has been used frequently to study the dynamics of synthetic biological networks [62], and I use this approach in Chapters 4 and 5.

The bistable region gives the genetic toggle switch its switching behaviour. In Figure 2.9C the expression levels of the two toggle genes u and v are plotted against time. The system rests in states where either u or v are highly expressed by the other not. The introduction of inducer I_1 when v is highly expressed moves the system to a resting state where u is highly expressed and 'switches off' v. This transition is reversible with the introduction of inducer I_2.

The genetic toggle switch and its mutual inhibition motif is foundational in synthetic biology. The topology of the gene network implementing the genetic toggle switch has also been embedded into larger networks to produce more complicated dynamics, to produce for example the AC-DC gene network [64], whose dynamics may be bistable or oscillatory or indeed both, depending on its parameters [65].

Since the switch's positive feedback motif is known to admit bistability, the main challenge for the design of synthetic biological switches is the identification of parameter sets in which bistability emerges [55]. There are strategies based on chemical reaction network theory that constrain these sets [63], and numerical techniques for detecting bistability (and in general multistability) in networks [71].

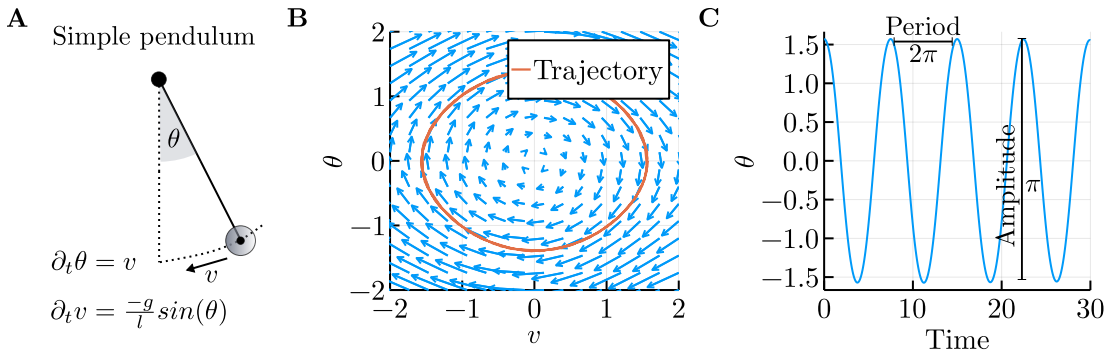


Figure 2.10: A simple pendulum without damping, an example of a oscillator. The pendulum in **A** can be modeled using two state variables, θ for the displacement angle and v for the pendulums velocity. The system can be described with two coupled ODEs. The vector field of these equations is shown in **B** which shows how trajectories flow (blue arrows) around the system. One example trajectory is shown in orange, and this trajectory is plotted over time in **C** for θ. θ oscillates with a specific period of 2π and amplitude of π.

Oscillation

It may be the case that a system never arrives at a fixed point. If this is the case then the system may display oscillatory behaviour, where the state variables move along a trajectory that repeats itself every period of time, and continues to do so indefinitely, never coming to rest. An example of an oscillator is an undamped pendulum, which swings back and forth periodically, and which without any damping force (for example due to air resistance or friction) will continue to do so forever. An illustration of this oscillating system is shown in Figure 2.10.

Repressilator

The repressilator was one of the first synthetic gene networks to be engineered. The networks consists of three transcriptional repressor genes connected in a negative feedback loop, and expression of the genes in the network can under certain conditions exhibit oscillations [23]. The schematic of this synthetic gene network is shown in Figure 2.11A.

As with the genetic toggle switch a mathematical model of the repressilator was developed using the ODEs in Equations 2.7-2.12, in order to inform the engineering of the gene network and predict which networks would be most likely to oscillate.

$$\partial_t m_1 = \frac{\alpha_m p_3^n}{K^n + p_3^n} - \gamma_m m_1 \tag{2.7}$$

$$\partial_t m_2 = \frac{\alpha_m p_1^n}{K^n + p_1^n} - \gamma_m m_2 \tag{2.8}$$

$$\partial_t m_3 = \frac{\alpha_m p_2^n}{K^n + p_2^n} - \gamma_m m_3 \tag{2.9}$$

$$\partial_t p_1 = \alpha_p m_1 - \gamma_p p_1 \tag{2.10}$$

$$\partial_t p_2 = \alpha_p m_2 - \gamma_p p_2 \tag{2.11}$$

$$\partial_t p_3 = \alpha_p m_3 - \gamma_p p_3 \tag{2.12}$$

where $m_{1,2,3}$ are the concentrations of mRNA for each repressor, $p_{1,2,3}$ are the repressor concentrations, α_m and α_p are the maximal transcription and translation rates, and γ_m and γ_p are the degradation/dilution rate of mRNA and repressor. The repressive interactions are modeled using Hill equations (Section 2.2) with half maximal repression concentration K and Hill coefficient n.

In the repressilator network, only one of p_1, p_2 or p_3 may be highly expressed at one time. They oscillate in sequence as shown in the timecourse simulation in Figure 2.11C. These oscillations emerge at Hopf bifurcation points which can be identified and plotted in bifurcation diagrams just as for the analysis of the genetic toggle switch, a bifurcation diagram for Equations 2.7-2.12 in plotted in Figure 2.11C.

I will use the same analysis, as well as timecourse simulations, in Chapter 5 for a electrogenetic oscillator, an adapted version of this synthetic network which incorporates an electrogenetic component, and which produces oscillations in electrical current.

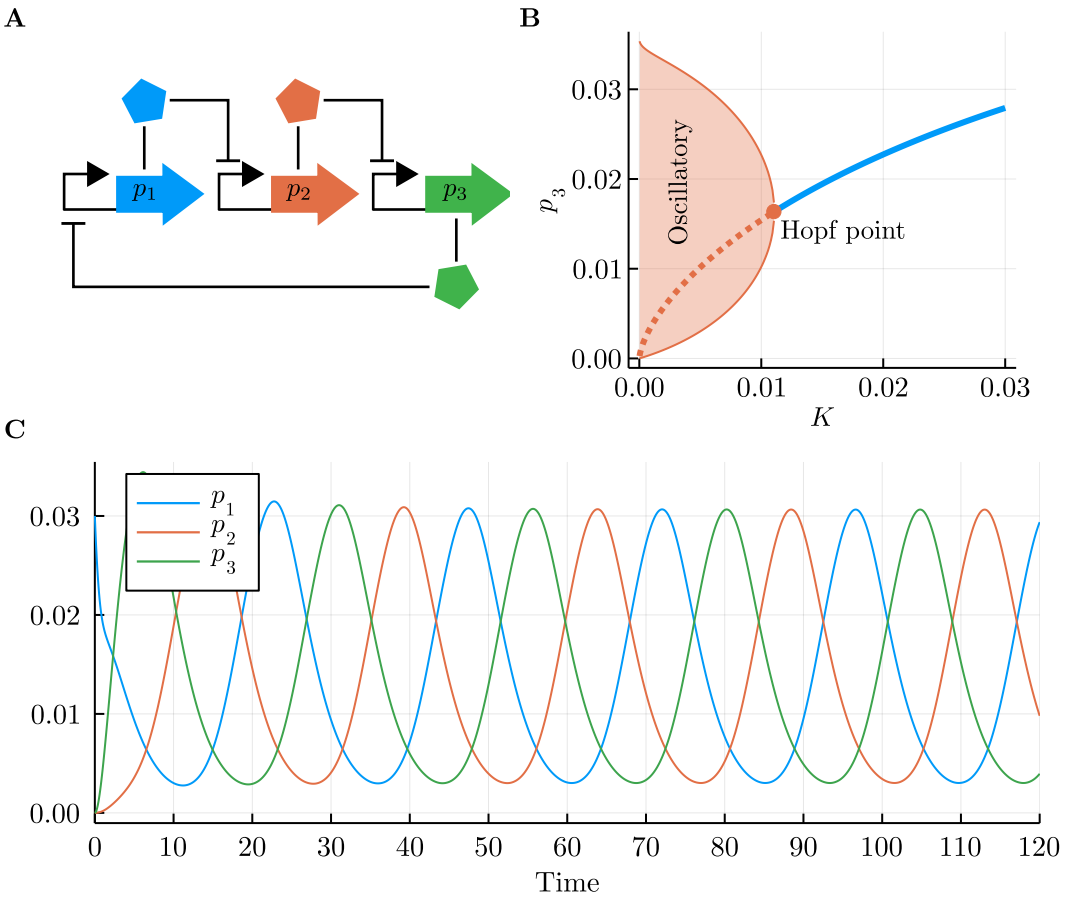


Figure 2.11: Numerical analysis of the repressilator. In **A** the synthetic gene network is shown, implemented with a negative feedback loop between three repressors p_1, p_2 and p_3. **B** shows the bifurcation diagram with the half maximal repression concentration K as the bifurcation parameter. A single Hopf bifurcation point exists at $K \approx 0.011$, below which oscillations emerge (orange shaded area). **C** shows the time course for the case of $K = 0.005$. All repressor concentrations exhibit sustained oscillations.

2.4 Context in gene networks

The construction of a perfectly closed synthetic biological system is neither possible nor desirable. Synthetic gene networks for example rely on co-option of the transcriptional and translational machinery of their host, and the usefulness of synthetic biological systems in general stems from their ability to interact with and affect their environment.

In the study of thermodynamics a distinction is drawn between the system under study and the *surroundings*. The system is the thing under study, and is usually imagined as spatially separate from the rest of the universe, its surroundings. Context in synthetic biology is a similar idea, though the line between the system and its surroundings is smudged considerably.

Here I try to sharpen the line again by defining context in synthetic biology in terms of the design stage of the DBTL cycle. Context is everything that the synthetic biologist is unwilling or unable to engineer. This may change depending on the system being designed, but for most synthetic gene networks, would include for example the availability of RNAP, ribosome [94], host or vector DNA nearby the engineered DNA [82], growth rate of the host organism [60, 44], growth media [8], or pH [99]. All these factors and their effects combined are the context of the synthetic biological system.

The development of a synthetic biological device can be a long and expensive process involving many iterations of the design, build, test and learn cycle. Ideally this optimisation process would be performed only once, producing a modular synthetic biological device that could be deployed any context. However the dependence of synthetic genetic circuits on contextual parameters such as bacterial host is widely cited, and frustrates the development of such perfectly modular systems [57].

One solution might be to include contextual effects in the mathematic modeling, but in general the whole of the intracellular and *extracellular* environment can be considered as context. Models that take into account some contextual dependencies for genetic circuits can be successfully developed [94]. However modeling contextual effects means models become larger and less tractable from the point of view of bottom-up design. Efforts to increase tractability through modularisation of contextual effects have been made using an "aspect-oriented" design paradigm[7]. However, ultimately a model must include some aspects of the context and exclude others to remain useful, with the consequence that model predictions for one context may not be valid for another context.

Another solution is to have a control strategy for the synthetic biological network which estimates the state of the system and provide correcting feedback if the state deviates from predefined setpoints. This kind of control strategy has previously been implemented *in silico*, where real time measurements are available and real time control signals can be provided [14, 56], but control strategies can also be built in to the synthetic biological network itself, for example using integral or proportional-integral-derivative feedback control [3, 11].

If the application allows it might be possible to eliminate the performance differences associated with contextual effects by standardising the context in which synthetic gene networks are deployed. Current work toward identifying and developing standard chassis organisms [20], including hosts that are designed and minimised to perform in specific environments [53], may help improve reproducability of research [83] and also generate insights into the impact of context.

Avoiding the use of a host organism altogether may also be used to eliminate or at least limit the effects of context. Synthetic biology using cell-free systems offers the opportunity for more control over environment and can remove some of the complexity that produces intricate contextual effects [85]. The degree to which these systems are open to their environment varies, and so also it is to be expected that the significance of context varies also, but both open [38] and closed [98] systems have impactful applications.

The broad range of contextual effects has previously been split into classes [13]. I present some classifications here with specific examples, some of which appear in the work of later Chapters of this thesis.

Genetic context

Synthetic gene networks can behave differently depending on the spatial organisation of their genetic material. This means that when composing synthetic gene networks, the choice of position in the DNA for individual components can their behaviour, this is known as 'compositional context' [13]. A good example is the 'roadblocking' phenomena observed when two promoter regions are in close proximity, where transcription rate from the promoters was seen to depend on their ordering [59]. In this case the phenomena could be measured and included as rules in the automated design processes to bring this otherwise contextual parameter under control. Another aspect of compositional context is

retroactivity, which has been shown to make modularity difficult in gene networks [22], and which is addressed in designs of synthetic gene networks using insulating components [59].

In bacterial synthetic biology, gene networks are often synthesised into plasmid vectors which contain other genetic components, such as antibiotic resistance genes and plasmid replication machinery [54]. These components provide a genetic context for synthetic gene networks and I show in Chapter 3 that this context can affect the performance of gene networks.

Host context

Synthetic gene networks usually rely on the machinery and resources provided by their host in order to express genes [8]. Of most direct importance for gene networks is the availability of RNAPs[39] and ribosomes[90], which both affect and are affected by the expression of exogenous genes in synthetic gene networks.

The dynamics of gene expression in general is also intimately linked with the growth rate of the host organism. The interaction of growth rate and synthetic gene networks has been predicted to affect the behaviour of simple gene regulatory motifs, as well as the emergence of bistability in genetic switches [44]

Other factors of the relationship between host and synthetic gene networks include issues of orthogonality of the synthetic genetic components and those belonging to the host. Synthetic genetic components that are orthogonal to one organism are not guaranteed to be orthogonal in another, leading to intended interactions between the host and the network that can lead to failure of the latter [92].

In Chapter 3 I measure the differences performance of synthetic gene networks in different hosts.

Spatial context

The intracellular space is often assumed to be homogeneous in synthetic biological design. This assumption is useful but it has been shown in some cases that spatial arrangement of biological networks within the cell can influence their behaviour. For example, the distance between transcription events of a transcription factor and its cognate promoter has been

suggested to affect transcriptional regulation [79], and in particular the noise profile of gene expression [35].

As mentioned previously, synthetic biological systems should be open to the environment in order to do useful work. But this often exposes the system to fluctuations or changes in the environment that might affect their performance. The range of environmental variables that constitute context is extremely broad. Temperature can affect the rates of transcription from promoter regions [69]. Growth medium has been found to affect the activity of promoter regions [13].

Some of these environmental variables may be homogeneous in space, particularly if the environment is well-mixed. In Chapters 4 and 5 however, I explore spatial effects in a heterogeneous and asymmetrical environment, the electroactive biofilm, and show that the extent of heterogeneity can change the dynamics of synthetic gene networks.

2.5 Microbial redox and electrochemistry

Bacteria can respire by coupling the oxidation of organic substrates to the reduction of inorganic ones to generate energy. Oxidation and reduction are redox reactions, where electrons are gained and lost, respectively. In aerobic respiration oxygen is the electron acceptor, but in general other acceptors are possible.

Exoelectrogenic bacteria are capable of anaerobic respiration using extracellular acceptors, of exporting electrons generated from the oxidation of substrate directly to the extracellular environment, in a process known as EET. This ability allows us to engineer systems in which the terminal electron acceptor is a solid electrode, effectively linking central metabolic pathways such as the TCA cycle to electrochemistry. These systems are MESs and can link bacteria to electronics.

Linking exoelectrogens to the electrode

The mechanisms by which bacterial EET is linked to the electrochemistry of an electrode can be divided into two categories.

- Direct EET (Figure 2.12A) can occur if the bacteria are in contact with or very close to (several nanometres) the electrode [47]. Proteins called cytochromes are displayed

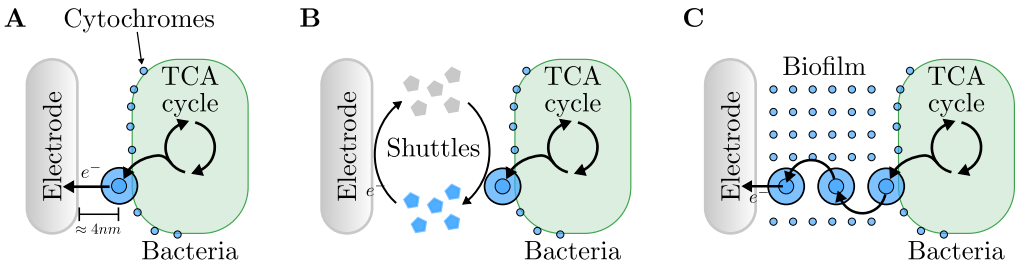


Figure 2.12: Redox and electrochemistry connects exoelectrogens to the electrode. In **A** close contact between outer membrane cytochromes (blue circles) and the electrode facilitates the transport of electrons generated by bacterial metabolism to the electrode. In **B** this scheme is extended for case where exoelectrogens secrete redox mediator molecules (hexagons) which diffuse between electrode and cytochrome, cycling between reduced and oxidised states to shuttle electrons. **C** illustrates cytochromes bound in the extracellular matrix of an electroactive biofilm. The matrix physically connects the electrode and the bacteria, and electron transport is achieved by successive exchanges of electrons between nearby cytochromes, forming a chain of 'hops' that end at the electrode interface.

on the outer membrane of the exoelectrogenic bacteria. These cytochromes typically have several redox centres that can be oxidised or reduced, and participate directly in redox reactions with the electrode [42, 88].

- Mediated EET can occur at longer ranges. Some exolectrogens produce and secrete redox shuttles, which diffuse back and forth between bacteria and electrode to transport electrons [52, 46] (Figure 2.12B). Alternatively, populations of exoelectrogens may form biofilms whose extracellular matrix contains cytochromes, and electrons can be transported by hopping between cytochromes that form a chain [47, 42] (Figure 2.12C). There is also evidence for conductive pili being produced in biofilms for the purpose of transporting electrons over greater distances [4, 51].

In later chapters the biofilm and electron transport by hopping is modeled. I use the term electroactive biofilm to mean biofilms comprised of exoelectrogens and an extracellular matrix of cytochromes capable of facilitating electron transport to an electrode.

Biofilm electrochemistry

At the interface of the electrode and the biofilm electrochemical reactions occur to exchange electrons between the electrode and cytochromes in contact with it. The movement of electrons across this interface can generate a measurable electrical current and the rate at which the reactions occur depend on the potential of the electrode. Thus there is a

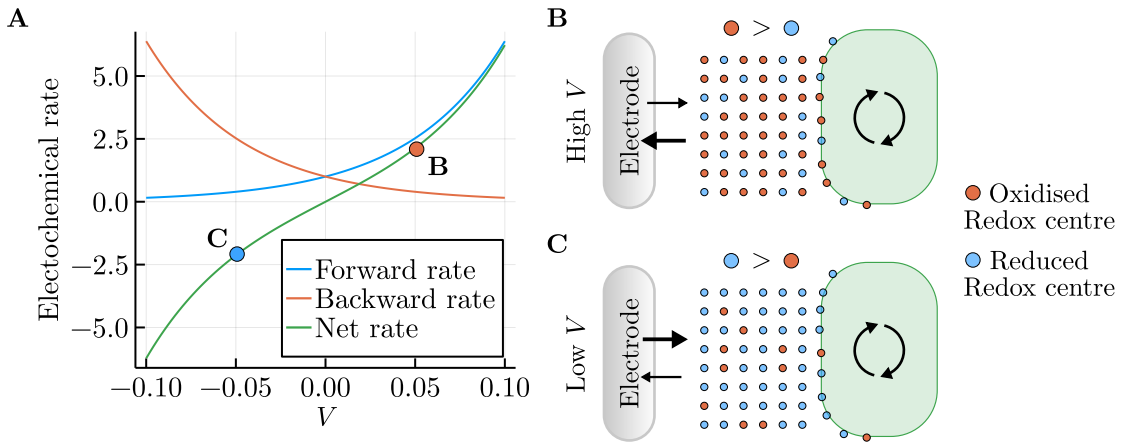


Figure 2.13: The effect of electrode potential V on the rates of electrochemical reactions and the redox environment in the biofilm. **A** shows a the well-known Butler-Volmer theory which relates V and the rate of electrochemical reactions. The rate of reduction of the electrode increases with V (blue line in **A**) and tends to extract electrons from the biofilm, increasing the proportion of oxidised redox centres (shown in **B** as orange circles). The rate of oxidation of the electrode decreases as V increases (orange line in **A**) and tends to push electrons into the biofilm, increasing the proportion of reduced redox centres (shown in **C** as blue circles). Current I is proportional to the net electrochemical rate, which is shown in green in **A**.

relationship between electrode potential and current.

Electrode potential is usually reported against a reference potential *. The extended Butler-Volmer relation describes the current as a function of electrode potential and the concentration of reduced and oxidised species at the electrode interface:

$$I = j_0 * \left(ue^{A(E-E_{eq})} - u^+e^{-B(E-E_{eq})}\right) \tag{2.13}$$

Where j_0 is the electron exchange rate at equilibrium, u and u^+ are the concentrations of reduced and oxidised species at the electrode interface, E is the electrode potential and E_{eq} is the equilibrium potential for the redox species participating in the reaction.

A and B are constants related to the thermodynamics of the electrode reaction, I assume $A = B = \frac{F}{2RT}$ from now on. I also define $V = E - E_{eq}$.

$$I = j_0 * \left(ue^{AV} - u^+e^{-AV}\right) \tag{2.14}$$

Equation 2.14 serves as a model for the rate of electrochemical reactions at the electrode

*Since potential is really the potential difference between the electrode and a reference electrode

that is used frequently in existing models of electrochemical reactions [45, 29, 67]. It links electrode potential V and current I as shown in Figure 2.13. It also describes a relationship between the V and the concentrations of oxidised and reduced redox centres in the biofilm. The ratio of these concentrations defines a specific reduction potential and redox environment inside the biofilm which can be thought of as a type of context.

Electrogenetics and electrogenetic synthetic biology

There exist in nature transcription factors whose activity depend on their redox state. A commonly cited example in *SoxR*, a transcription factor which is known to be active only when oxidised [30]. Such redox sensitive transcription factors can link the redox environment to transcriptional regulation in gene networks.

Interest in co-opting these redox-sensitive systems for synthetic biology is growing, partly owing to the implication that synthetic biology might help advance existing bioelectrochemical systems in bioremediation, biofuel production and power generation [40]. Recently the *SoxR* was used as the basis for the development of libraries of synthetic redox-sensitive promoters [48]. Engineered synthetic gene networks whose expression can be controlled using the potential of an electrode have been demonstrated [87, 84], as well as the integration of electrogenetics in a CRISPR system [6].

In Chapters 4 and 5 I will take these electrogenetic synthetic biology tools as given, and model how they link the synthetic gene networks described there to the redox context of an electroactive biofilm.

Chapter 3

Genetic logic gates library with genetic and host context

This chapter adapts the analysis first presented in Tas, Grozinger, Stoof, de Lorenzo and Goñi-Moreno [82]. I performed the analysis of data collected and preprocessed by Tas and Stoof. Many of the ideas used in the characterisation of the library, as well as the library itself, are adapted from Nielsen et al [59].

3.1 Introduction

Synthetic biology has developed sophisticated tools for engineering synthetic gene networks. These networks connect genes together using the same genetic regulatory processes that connect genes in naturally occurring networks. An example is transcriptional regulation, where the product of one gene either inhibits or activates the transcription of another gene (or itself).

Previous work has used synthetic networks of genes connected with transcriptional regulation to perform biocomputations, in particular to implement combinatorial logic circuits [59]. These synthetic networks were designed as modular units called genetic logic gates, which could be composed to construct more larger networks and implement many combinatorial logic circuits not included in the initial study. A software was developed which transformed descriptions of computations as Boolean formulae into specifications for synthetic gene networks which implemented these computations using a collection of genetic logic gates called a library. This approach allowed the automated design of genetic logic circuits using libraries of individual genetic logic gates.

Synthetic biological genetic logic gates

The individual logic gates making up the library are synthetic gene networks implementing NOT gates, whose topology is shown in Figure 2.3A. The expression levels of the input gene A and output gene Z are discretised into levels 1 and 0, corresponding to high and low expression, in order to represent the logical variables of combinatorial logic. The gene product of A is chosen so that it inhibits transcription from the promoter upstream of the output gene Z. Expression of A therefore represses the expression of Z, and so high expression levels of A lead to low expression levels of Z, and low expression levels of A allow Z to be expressed at high levels. The genetic NOT gate takes a single input and inverts it into the output using transcriptional regulation, transforming 0 into 1 and 1 into 0 as shown in the table of Figure 2.3A.

The NOT gate performs implements the simple logical negation operation on a single input. However this library of NOT gates was previously used to design and build circuits that implemented far more complex logical operations [59]. This was possible because 2 NOT gates can be composed together to build a NOR gate, as seen in Figure 2.3B. The NOR gate takes two inputs and transforms them to a single output as shown in the table

of 2.3B, while also a relatively simple operation, the NOR gate has the special property of functional completeness, meaning that any combinatorial logic circuit containing any variety of gates can be built by composing only NOR gates together. The property makes the genetic NOR gate an extremely powerful synthetic gene network from the perspective of biocomputing, and in turn makes the construction of well-characterised and modular libraries of NOT gates fundamental in pursuing the goal of engineering genetic logic circuits for a wide variety of biocomputing applications.

Challenges for genetic combinatorial logic circuits

Connected genetic NOR gates should in principle be able to perform any combinatorial logic computation desired due to their universality. In [59] it was shown that NOR gates could indeed be composed together in order to perform more complex combinatorial logic computations.

This has a number of implications for genetic logic circuits. First, the gates making up the circuit must use different molecules to represent their input and output signals, otherwise there will be significant cross-talk between gates because gates that should not be connected in the circuit will be connected by the signal molecules they have in common.

For this library of genetic NOT gates, this means using a different repressor molecule for each gate in a circuit. Unfortunately the differing properties of the repressor molecules and their cognate promoters means that the performance of each gate will differ also. As shown below these differences in performance must be taken into account in the design stage and affect the quality of the genetic NOT gate library.

Context is also a problem for libraries because it limits the ability for reuse. The large experimental effort expended on the characterisation (and optimisation) of the library's individual gates is inexorably tied into the experimental conditions that provide the context for the synthetic gene networks. Using the library in different host organisms is a large change in the context which might be expected to produce a large change in circuit performance. However other more subtle changes, such as the plasmid vector into which the circuit is inserted, will also affect performance. If genetic logic circuits are to be deployed for applications in the field, where context can vary and is not under experimental control, it will be necessary to understand the effects of context on synthetic gene networks and account for these in the design stages.

Objectives

1. Develop an implementation of a workflow for the characterisation of a library of genetic NOT gates. The workflow should take flow cytometry data and produce models of the measured synthetic gene networks that can later be used for combinatorial logic circuit design.

2. Apply the workflow to characterise a previously designed library of genetic NOT gates, in 7 different contexts, where each context is defined by the host and plasmid vector used.

3. Compare the characterisations from different contexts in order to measure the extent of the contextual effects.

4. Develop and apply metrics for assessing the quality of the genetic NOT gate library and compare these metrics across contexts.

5. Measure the qualities of libraries in merged contexts, libraries which make use of multiple plasmid vectors or host organisms.

3.2 Characterisation of library using flow cytometry data

The library used for this study is a collection of genetic NOT gates whose performance was previously optimised and characterised in the *Escherichia coli* strain *NEB10β* with plasmid vector *pAN* [59]. Here the same library is characterised using 7 different combinations of host organism and plasmid vectors which provide 7 different contexts for the genetic NOT gate library. These contexts are those listed in Table 3.1.

The library itself consists of 20 distinct NOT gates as synthetic gene networks with the structure shown in Figure 3.1A. Gates are implemented using different repressors (R) and their cognate promoters (P_R), as well as variations in a, which is a RBS. Accordingly NOT gates can be uniquely identified as R_a, and there are 20 such gates in the library, using 12 different repressors R.

The automated design workflow relies heavily on experimental data collected about the behaviour of the individual gates in the library [59]. Experimental data for the genetic

Host	Backbone (Plasmid vector)
KT2400	pSeva221
KT2400	pSeva231
KT2400	pSeva251
CC118λpir	pSeva221
CC118λpir	pSeva231
DH5α	pSeva221
DH5α	pAN

Table 3.1: The contexts used in the study. Seven different combinations of from three host organisms and four plasmid vectors were tested.

NOT gate library was previously collected using flow cytometry by replacing output gene of each gate by a gene encoding YFP. Each gate in the library was then inserted into a plasmid vector and subsequently transformed into a population of bacterial hosts. Measurements of the distribution of YFP expression across a large (on the order of 10^4) sample of the engineered cell population were made to determine the behaviour of each NOT gate in the presence of different concentrations of IPTG. These experiments were performed in each of the contexts listed in Table 3.1.

3.3 Results

Processing of flow cytometry data

Flow cytometry for each gate is performed at different levels of IPTG induction in order to obtain output gene expression distributions corresponding to different input gene expression levels. An example of how these distributions change in response to IPTG concentration is shown in Figure 3.1B. It can be seen that higher IPTG concentrations produce narrower distributions whose medians (dashed lines in Figure 3.1B) decrease as IPTG concentration increases. This is because IPTG binds to and inactivates LacI, preventing it from repressing transcription of R from the P_{Tac} promoter, turning on the expression of R. In turn R represses transcription of YFP from the P_R and turns off the expression of YFP. The distributions shown in Figure 3.1B, as well as those shown in the remaining figures in this Chapter, are based on a single flow cytometry experiment for each of the 12 IPTG concentrations tested.

The characterisation is completed by measuring the how expression from the P_{Tac} promoter

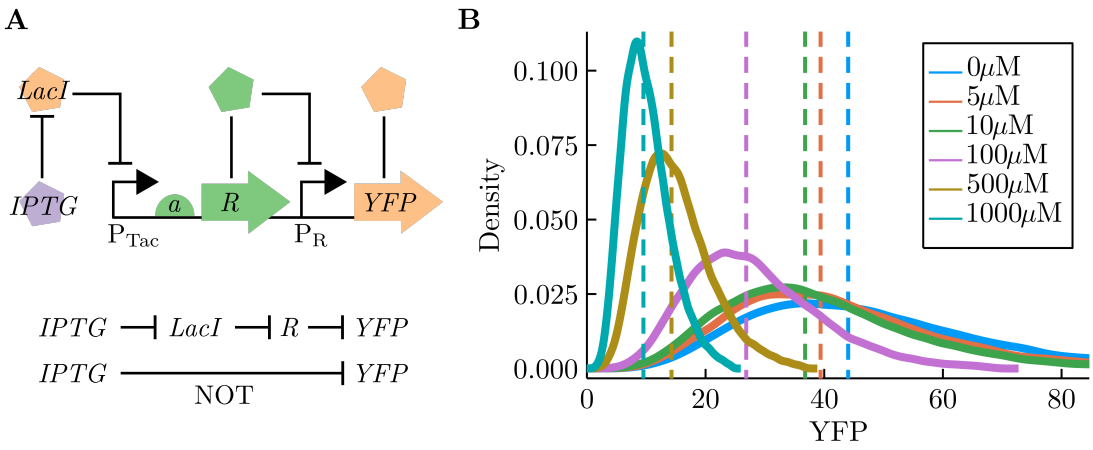


Figure 3.1: **A** shows the network used to characterise the genetic NOT gate with repressor *R*. In this network, IPTG binds to and inactivates *LacI*, relieving the inhibition of *R* expression by *LacI*. Expression of *R* inhibits the expression of YFP, which is controlled by the cognate promoter of *R*. **B** Flow cytometry measures the distributions of YFP expression in the presence of different concentrations of IPTG, the dashed lines indicate the median expression levels. The overall effect is that higher concentrations of IPTG lead to lower median expression of YFP. However, the distributions also overlap for each concentration of IPTG.

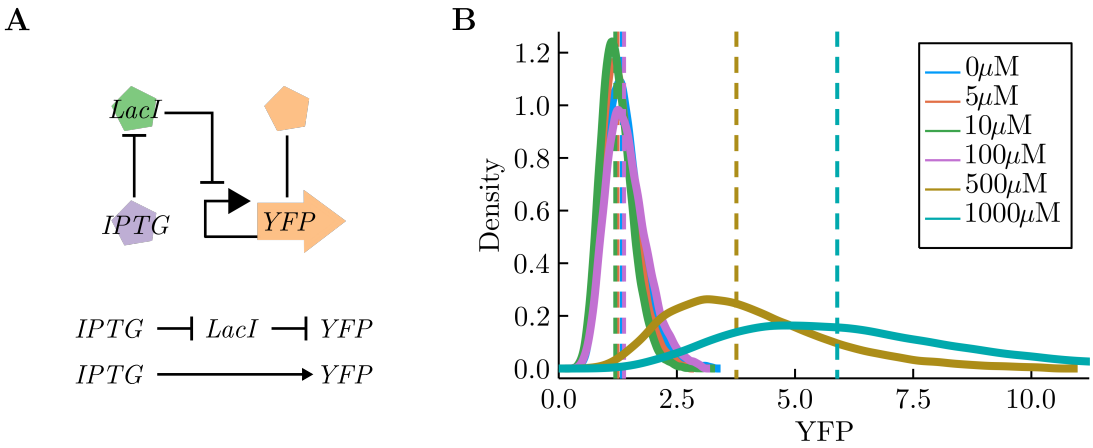


Figure 3.2: **A** shows the network used to characterise the inputs to the NOT gates. In this network, IPTG binds to and inactivates *LacI*, relieving the inhibition of YFP expression by *LacI*. **B** Flow cytometry measures the distributions of YFP expression in the presence of different concentrations of IPTG, the dashed lines indicate the median expression levels. The overall effect is that higher concentrations of IPTG tend to produce higher median expression of YFP. The networks appears fairly insensitive to IPTG and distributions overlap until $500\mu M$, where expression increases sharply and is more differentiated.

changes in response to IPTG concentration. This is done using flow cytometry of the network shown in Figure 3.2A, again at differing concentrations of IPTG to obtain distributions of YFP such as those shown in Figure 3.2B. It can be seen that higher IPTG concentrations produce broader distributions whose medians increase with IPTG concentration. These expression distributions are used as a proxy for the expression distribution of R in the network of Figure 3.1A. The combination of these two flow cytometry experiments at a given level of IPTG characterises the expression of YFP from the promoter P_R in response to the expression of R.

In total this characterisation was performed for 135 different gate–host–plasmid combinations, and at 12 different IPTG concentrations, using 1620 flow cytometry experiments.

Standardisation and conversion to relative units

The expression distributions obtained from flow cytometry were standardised using the FlowScatt software. FlowScatt aims to account for the effects of heterogeneous cell volume (perhaps due to different growth rates or phases of individual cells) by marginalising the expression distribution on the scattered light measurements for each sample [78]. The output is a distribution of expression that is conditional on the volume of a cell.

Additional flow cytometry was performed in each context with the autofluorescence (*1201* Figure 3.3A) and normalisation (*1717* Figure 3.4B) synthetic gene networks for the purposes of converting the results of the NOT gate cytometry into RPUs. This process further standardises gene expression distributions by subtracting background fluorescence measured with the *1201* network. It also normalises the expression distribution to the level of expression of YFP from the *1717* network. This process is designed to enable the comparison of experiments and account for differences in experimental setups and test conditions [43].

The conversion to RPUs was performed according to a previously described procedure [82] and can be summarised as follows:

1. Produce the histogram of the raw flow cytometry data.
2. Use FlowScatt [78] to produce a distribution conditioned on the mean cell volume.
3. Shift the distribution by subtracting the median of the corresponding autofluorescence (1201) experiment from all the samples.

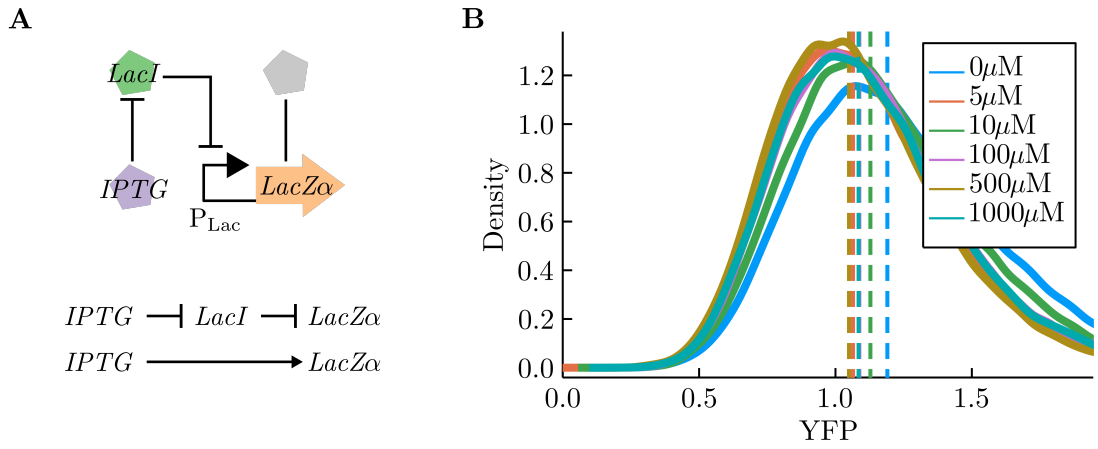


Figure 3.3: **A** shows the network used to measure the autofluorescence of the bacterial cells. In this network LacZα is expressed from the P_{Lac} at a level controlled by the concentration of IPTG. LacZα is not a fluorescent protein, so it is expected that any measured fluorescence is background. **B** shows the results of flow cytometry of this network in host organism KT2400 and with plasmid vector pSeva221. The flow cytometry is repeated at different concentrations of IPTG. As expected the results show very little response to IPTG.

4. Normalise the intensity values of the distribution by scaling them with the median of the corresponding normalisation (1717) experiment.

The analysis presented in the rest of this section uses RPUs to quantify gene expression, unless stated otherwise.

The response function model of the genetic NOT gate

Rational design of synthetic gene networks using the library requires a model for each genetic NOT gate which predicts how the input is transformed into the output. For the NOT gate in Figure 3.1A we require a model which describes the relationship between the expression levels of R and YFP.

I use a modified Hill equation, introduced in Section 2.2 as a simple model which gives a deterministic mapping between these input and output levels.

$$f_{YFP}(R) = \frac{(f_1 - f_0)k^n}{k^n + R^n} + f_0 \tag{3.1}$$

The Hill equation is modified so that it's minimum is at f_0, and the maximum is at f_1. The parameters f_0, f_1, k and n characterise the output gene expression f_{YFP}, given the

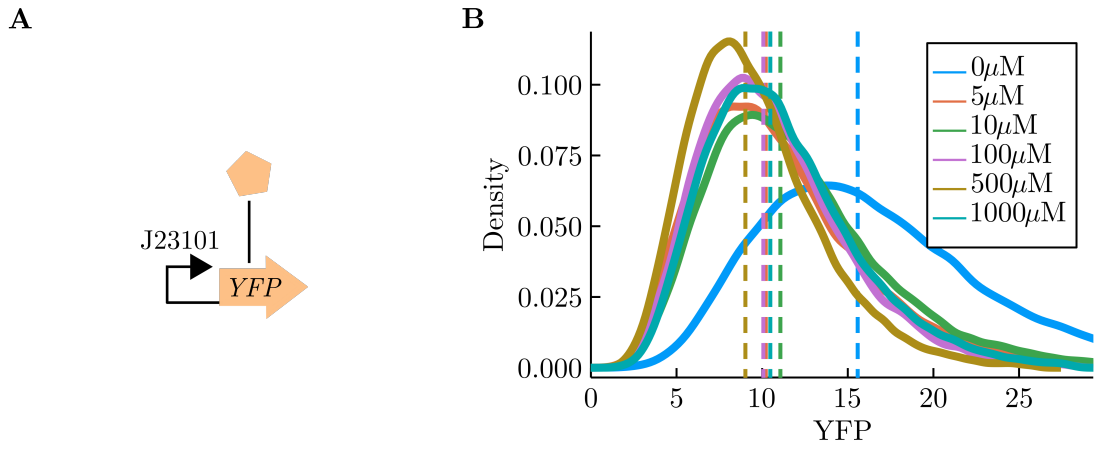


Figure 3.4: **A** shows the network used to measure the expression of YFP from the normalisation promoter *JS23101*. In this network YFP is expressed constitutively. **B** show the results of flow cytometry of this network in host organism KT2400 and with plasmid vector pSeva221, and at different concentrations of IPTG.

input gene expression of R. These parameters must be fitted to the experimental data for each genetic NOT gate, that is, fitted to experimentally obtained values for $f_{YFP}(R)$ and R. Adapting the approach previously taken to characterise the library in *E. coli* [59] these values are obtained as follows:

R is measured as described in the previous section with flow cytometry of the network in Figure 3.2. An example for the NOT gate *Lmra_N1* is shown in Figure 3.5A. For each concentration of IPTG the median of the distribution is taken as a point estimate of the expression level R_i. These can be plotted against IPTG concentration as in Figure 3.5C.

f_{YFP} is measured with flow cytometry of the network in Figure 3.1. Again the medians, shown as dashed lines in Figure 3.5B, are taken as point estimates for the expression level at each IPTG concentration. These are also plotted against IPTG concentration in Figure 3.5C.

$f_{YFP}(R_i)$ is then obtained by combining f_{YFP} and R_i according to IPTG concentration. The data points (circles) in Figure 3.5D plot this combination, and are used to fit Equation 3.1 and obtain the response function characterisation of the gate (solid line). In this work, least squares fitting (Levenberg–Marquardt algorithm) is used to find the parameters of Equation 3.1 and define $f_{YFP}(R)$ for all 135 genetic NOT gates. Table 3.2 lists the fitted parameters.

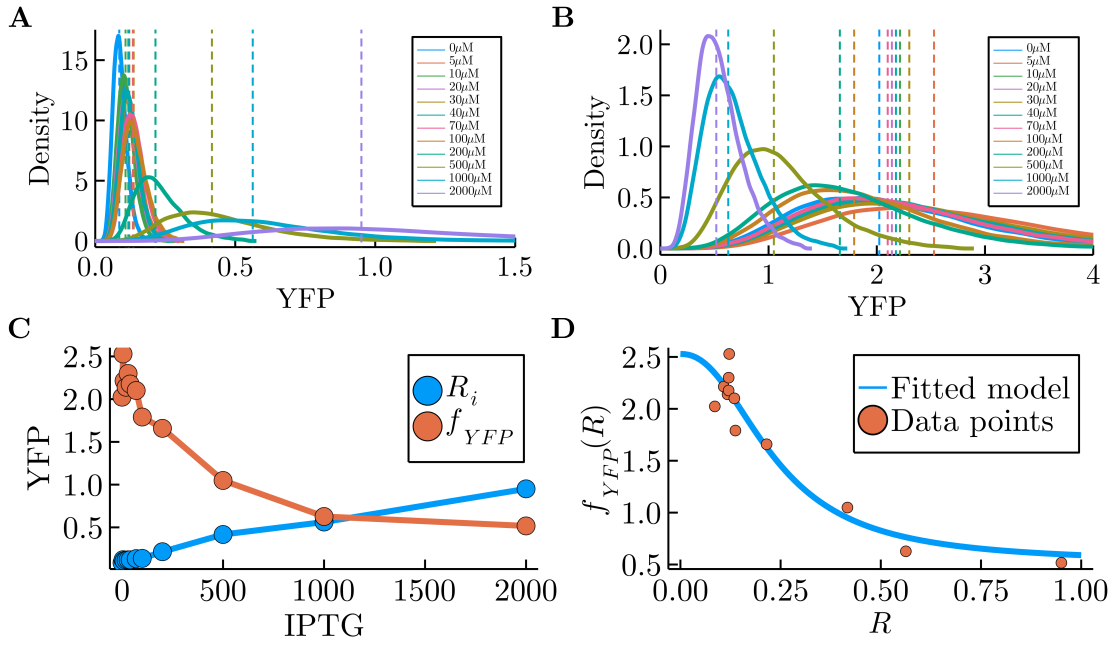


Figure 3.5: **A** show the expression distributions for the synthetic gene network of Figure 3.2A, at different concentrations of IPTG. In **B** the distributions for the synthetic gene network of Figure 3.1 for the same IPTG concentrations. The medians of these distributions are plotted against IPTG concentration in **C**. Plotting the medians of each set of distributions against one another produces the data points (circles) in **D**, which are used as data for the estimation of the parameters of Equation 3.1 to define the response function curve (solid line).

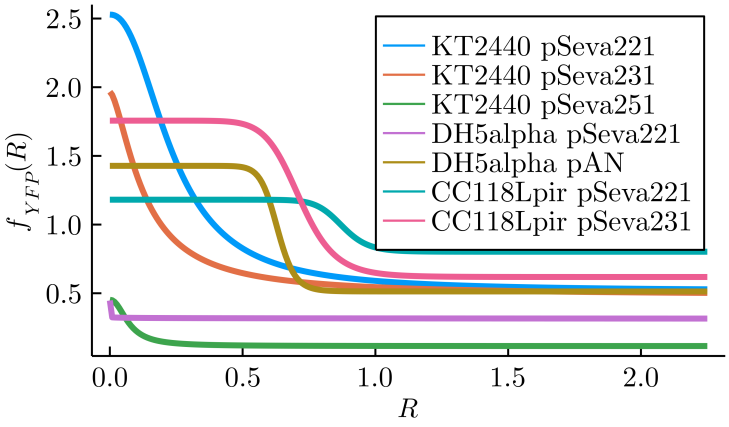


Figure 3.6: The response functions (Equation 3.1) for the gate *Lmra_N1* in all 7 contexts are plotted for comparison. There are significant quantitative and qualitative differences in the performance across the 7 contexts, despite the synthetic gene network being identical in all cases.

Context	f_0	f_1	k	n
KT2400 pSeva221	0.52	2.53	0.24	2.26
KT2400 pSeva231	0.49	1.97	0.13	1.55
KT2400 pSeva251	0.12	0.45	0.07	2.2
DH5α pSeva221	-	-	-	-
DH5α pAN	0.51	1.43	0.63	19.36
CC118λpir pSeva221	0.8	1.18	0.87	17.8
CC118λpir pSeva231	0.62	1.76	0.72	10.76

Table 3.2: The estimated best-fitting parameters for the *Lmra_N1* gate in each context.

NOT gate behaviour differs between contexts

The response functions for the same NOT gate in different contexts differ. Since the synthetic gene networks making up the NOT gates themselves are identical this suggests these differences arise from contextual effects on the synthetic gene networks. An example is shown in Figure 3.6, where the response functions of the gate *Lmra_N1* are shown in all seven contexts from Table 3.1. The response functions for genetically identical gates can differ significantly between contexts, and Table 3.2 shows how the parameters of Equation 3.1 change with context for this particular NOT gate. This result is found for across the entire library — the response function of a gate in one context does not indicate the response function of the gate in another.

The Fréchet distance was used to quantify the contextual effects on the NOT gates, by measuring the similarity between the response function curves characterised in different

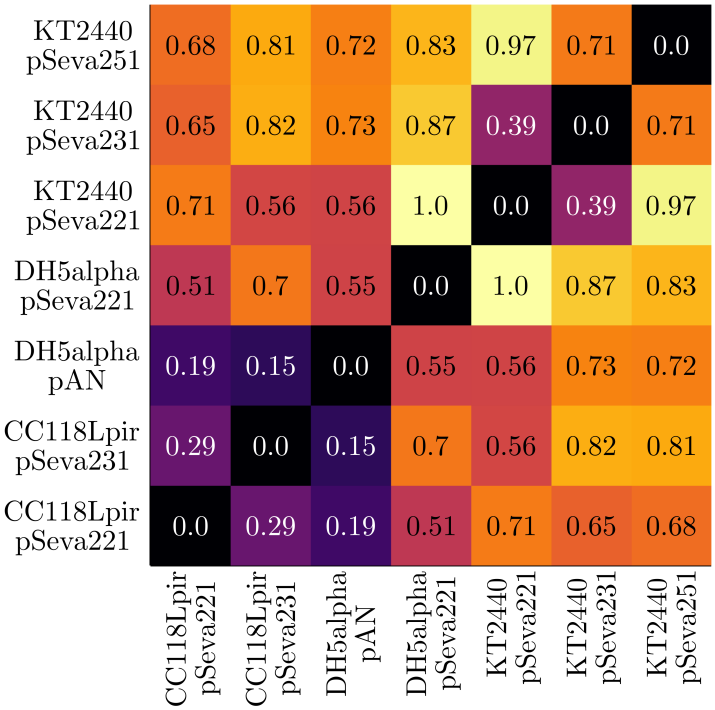


Figure 3.7: A heatmap showing the normalised Fréchet distances between the response function models for the NOT gate *Lmra_N1* in each context. Dark areas are those where the models are most similar, and lighter areas are those where the models are most different. It can be seen that the gate performance in the *E. Coli* strains is more consistent (bottom left corner) than in the *P. Putida* strain KT2400 (top right corner). For this gate, the contexts KT2400 with pSeva221 and DH5α with pSeva221 where found to have the greatest difference is performance according to Fr'echet distance.

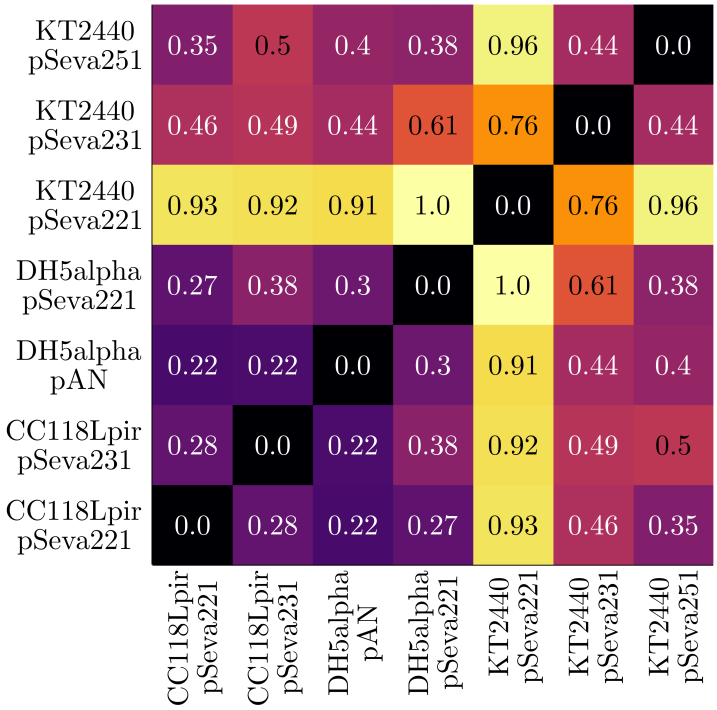


Figure 3.8: A heatmap presenting the normalised sum of Fréchet distances between the response function models for the all the NOT gates characterised in all 7 contexts. Dark areas are those where the models are most similar, and lighter areas are those where the models are most different. Overall performance varied more in *Pseudomonas putida* than *E. coli*, despite there being 2 different strains of *E. coli* being characterised and only a single *P. putida* strain. It is also notable that the KT2400 with pSeva221 context appears to be the most different from all others.

contexts. Figure 3.7 shows a matrix of pairwise Fréchet distances for the NOT gate *Lmra_N1*. The results suggest that variability between the *E. coli* strains tested is quite low, in particular between the two CC118λpir contexts and the DH5α with pAN context. Certainly, there appears to be more variability between contexts of different organisms than with different backbones. This is not surprising since synthetic gene networks interact with the host machinery in many different ways, including transcription, translation, protein folding, and indeed plasmid replication.

Figure 3.8 shows the similarity between contexts by summation of the Fréchet distances for all twenty NOT gates between each pair of contexts. Again, host organism appears to be the most important aspect of context in terms of variability in performance between contexts. These distances show that there is much more variability in the single *Psuedomonas putida* strain than in the 2 *E. coli* strains. Interestingly, the two most different contexts were KT2400 with pSeva221 and DH5α with pSeva221, despite having the same plasmid vector.

Assessing the quality of the library under the response function model

Since the performance of NOT gates varies between contexts it is possible to ask which is the best context in terms of NOT gate performance. A NOT gate should invert gene expression levels, transforming high input expression into low output expression and vice versa. Since Equation 3.1 is monotonically decreasing, higher input expressions always produce lower output expressions. However, in order to more easily interpret the outputs of the genetic NOT gates and to reduce failure due to measurement or process noise of the gates, it is desirable to separate the high and low output expression levels. This separation is achieved by applying thresholds to the output of the NOT gate which define high output, low output and ambiguous regions. In this work these thresholds are defined according to previous studies with the same library [59].

For the high output region:

$$f_{YFP}(R) > \frac{f_1}{2} \tag{3.2}$$

and for the low output region:

$$f_{YFP}(R) < 2f_0 \tag{3.3}$$

Which means that outputs that are greater than half the maximum possible output (f_1) are interpreted as high outputs, whilst those outputs that are less than twice the minimum possible output (f_0) are interpreted as low. These thresholds are marked for a *Lmra_N1* NOT gate in Figure 3.9 as horizontal dashed lines, which separate the output into high, low and ambiguous regions.

To ensure an unambiguous interpretation of the high and low outputs requires that:

$$\frac{f_1}{2} > 2f_0 \tag{3.4}$$

Meaning that there is a nonzero separation between the high and low regions. If there is no separation between these regions then the gate is considered nonfunctional. As such, a possible quantitative measure of the quality of a library in a particular context is the number of NOT gates which are functional according to this definition.

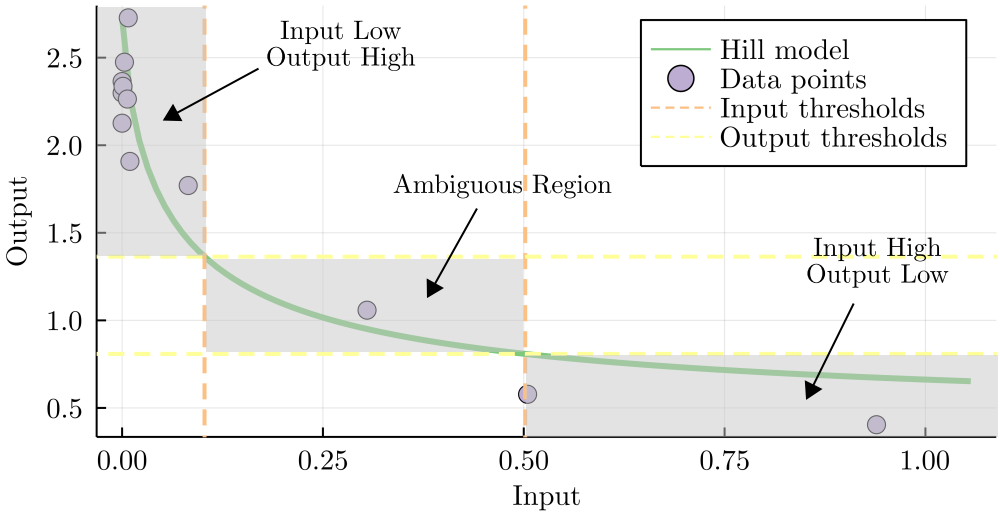


Figure 3.9: The response function model for a NOT gate. The x-axis is the RPU of expression from the gene network in Figure 3.2A which characterises the input to the NOT gate. The y-axis is the RPU of expression of YFP from the network in Figure 3.1A, in this particular case the NOT gate *Lmra_N1* in context KT2400 with pSeva221. The circles show the medians from each distribution and the solid line is the hill equation fitted to these data points. The dashed lines are the input and output thresholds that divide the response function into 3 regions, one with low input and high output (top left), one with high input and low output (bottom right) and an ambiguous region where the NOT gate is neither in a high or low state. The existence of the ambiguous region means the NOT gate is considered functional under the response function model.

Context	Functional NOT gates	Compatible Pairs
KT2400 pSeva221	6	1
KT2400 pSeva231	10	5
KT2400 pSeva251	4	0
DH5α pSeva221	18	70
DH5α pAN	20	38
CC118λpir pSeva221	18	6
CC118λpir pSeva231	20	24

Table 3.3: The number of functional NOT gates and the number of compatible pairs of gates in the library for each context.

Using this measure the library has the highest quality in those contexts with *E. coli* hosts and in fact there are two contexts in which all 20 NOT gates are functional. This is not surprising since the NOT gates were previously optimised for another *E. coli* strain [59].

It is clear that the quality of the library as measured by the number of functional NOT gates varies due to contextual effects. However it is also possible in some cases to improve the quality of the library by using more than one context. For example if both the pSeva221 and pSeva231 plasmid vectors are used in KT2400 to construct a library, 2 additional functional NOT gates are available, without need for the redesign and recharacterisation that would be necessary to introduce new gates into the library.

Compatibility of NOT gates as a quality score

In order to build circuits the individual NOT gates of the library must be connected in such a way that the output of one gate serves as the input of another. However it is essential that both the gates to be connected 'agree' on the levels of expression that represent high and low states, specifically that the output levels of the first gate are interpreted correctly as inputs to the second. Since more compatible pairs of gates in a library represents more degrees of freedom in the combinatorial design of circuits, the number of compatible pairs of NOT gates was calculated in order to measure the quality of the library.

This was done by calculating the input thresholds for each NOT gate, where the low input threshold is defined as the level of input expression r for which $f_{YFP}(r) = \frac{f_1}{2}$, and where the high input threshold is defined as the level of input expression r for which $f_{YFP}(r) = 2f_0$. As with the output thresholds, these input thresholds divide the input expression into 3 regions, the high input, low input and ambiguous regions, as shown in the example of Figure 3.9 (vertical dashed lines). The thresholds are calculated as follows:

$$IH = \left(\frac{k^n(f_1 - 2f_0)}{f_0}\right)^{\frac{1}{n}} \tag{3.5}$$

$$IL = \left(\frac{k^n f_1}{(f_1 - 2f_0)}\right)^{\frac{1}{n}} \tag{3.6}$$

where IH and IL are the high and low input thresholds. Figure 3.10 shows how compatibility is assessed for two example gates. In the case of this example where the first gate is

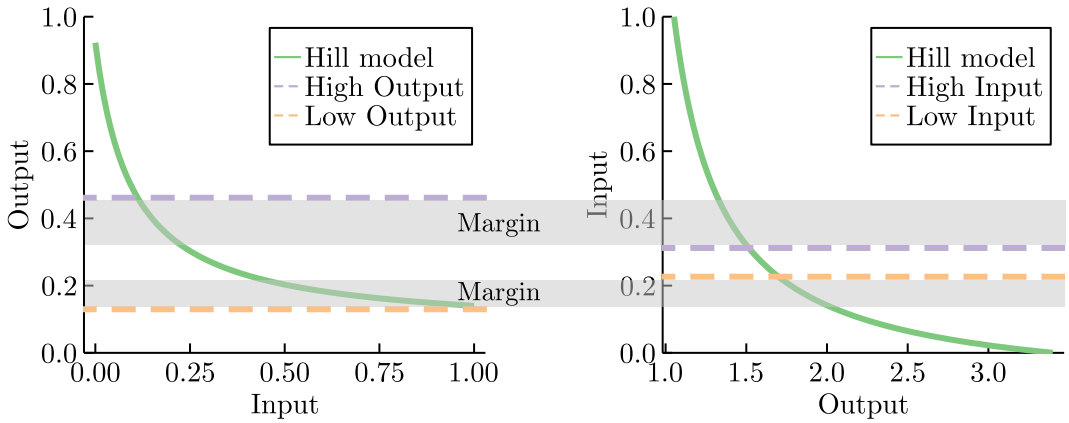


Figure 3.10: The response function model defines requirements for NOT gates to be connected. If the output of one gate (left) is to be connected to the input of a second (right), then the high(low) output threshold of the first gate must be greater than (less than) the high(low) input threshold of the second. This is meant to ensure that a high(low) output signal from the first gate is interpreted as a high(low) input by the second. Large margins between the thresholds (gray areas) are desirable to minimise the chances of misinterpretation of the signals between the two gates.

Srpr_S1 and the second gate is *Qacr_Q1*, both in the context of KT2400 with pSeva221 plasmid vector, the gates are defined as compatible, since the high output threshold of the first gate is greater than the high input threshold of the second (purple dashed lines), and the low output threshold of the first gate is lower than the low input threshold of the second (orange dashed lines).

A pairing of gates is defined as compatible in case that:

$$OH_A > IH_B \text{ and } OL_A < IL_B \tag{3.7}$$

where $OH = \frac{f_1}{2}$ is the high output threshold, $OL = 2f_0$ is the low output threshold, and IH and IL are defined as in Equations 3.5 and 3.6. The subscripts A and B refer to the first and second gates in the pair. An additional requirement is that the gates are orthogonal, which in this work requires only that each gate uses a different repressor molecule.

Not all pairs of gates in a library are compatible under this definition. As such the number of compatible pairs of NOT gates in a library can be used as a measure of the quality of the library. Figure 3.11 shows that the library in the context of KT2400 with pSeva221 plasmid vector, only 1 pair of NOT gates can be connected together. This is important since there

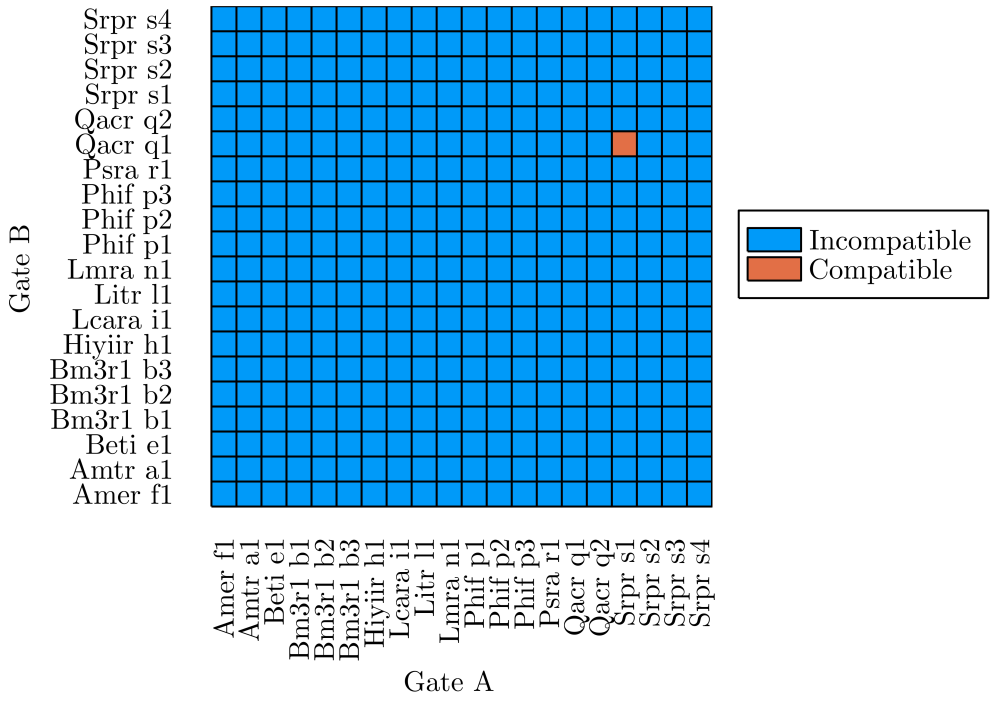


Figure 3.11: Grid showing the different connected pairs of NOT gates in the library with the output of gate A (x axis) serving as the input of gate B (y axis). Pairs which are compatible can be connected together correctly and are coloured in orange. This grid shows the results for the library characterised in the context KT2400 with the pSeva221 plasmid vector. There is only a single pair of compatible gates.

are 6 different functional NOT gates but only a single pair of those are compatible.

It is also notable that the quality of the library can be different when measured by the number of functional gates compared to the number of compatible pairs. In Table 3.3 for example, the library in the contexts DH5α with pSeva221 and CC118λpir with pSeva221 have the same number of functional gates. However, the number of compatible pairs in each of these contexts differ with 70 pairs in the DH5α context and only 6 in CC118λpir.

Contexts can be combined in order to increase the number of compatible pairs and the quality of the library. Figure 3.12 shows how allowing the freedom of choice in the plasmid vector for the library increases the number of compatible pairs in the host KT2400. The same library of 20 NOT gates can be placed in different contexts and connected together to increase the number of compatible pairs from 1 to 12.

This approach applied to context in general can improve the quality of the library further. Figure 3.13 shows that freedom in both plasmid vector and host organism can increase

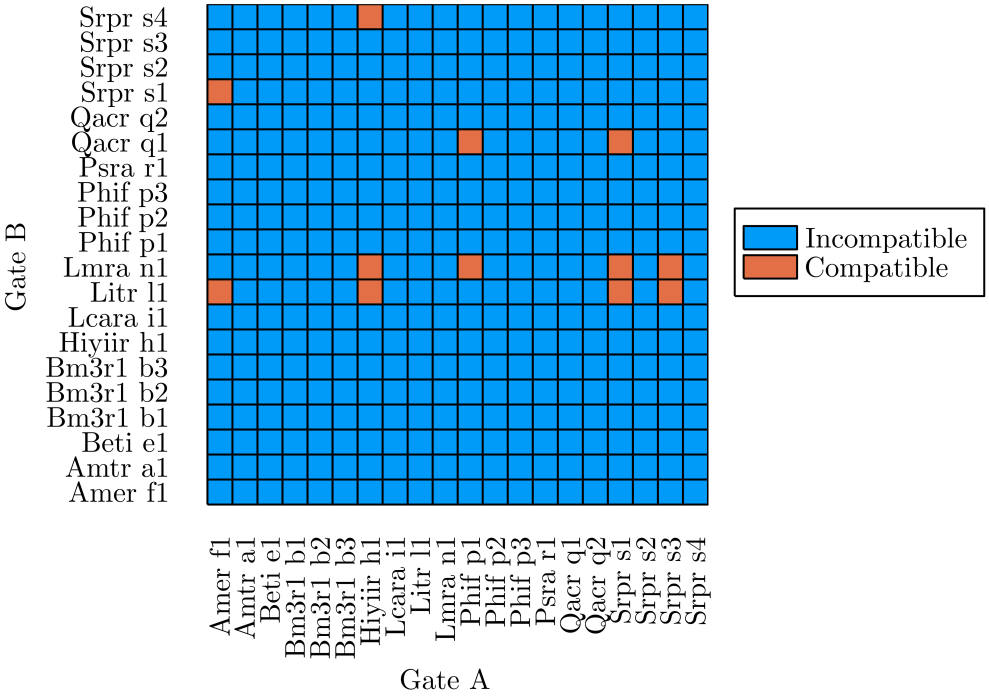


Figure 3.12: Grid showing the different connected pairs of NOT gates in the library with the output of gate A (x axis) serving as the input of gate B (y axis). Pairs which are compatible can be connected together correctly and are coloured in orange. This grid shows the results for the library in the host KT2400 but where gates can be placed in plasmid vectors pSeva221, pSeva231 or pSeva251. There are 12 pairs of compatible gates.

the number of pairing further, from 12 to 368. This is much greater than is possible in any single host (in DH5α 126 pairs, in CC118λpir 32 pairs and in KT2400 12 pairs).

Maximum chained length of compatible inverters as a quality score

The complexity of combinatorial logic circuits can be measured by their depth. Depth in a circuit is achieved by connecting the output of one gate to the input(s) of another. Here I interpret the maps of compatible pairs presented in the previous section as the adjacency matrix of a directed graph, in order to compute the longest possible chain of connected and compatible NOT gates for each library. I use the length of such a chain as an upper bound on the depth of circuits built using the library, and an upper bound on circuit complexity.

The maximum length of chain was computed for each individual context (host and plasmid

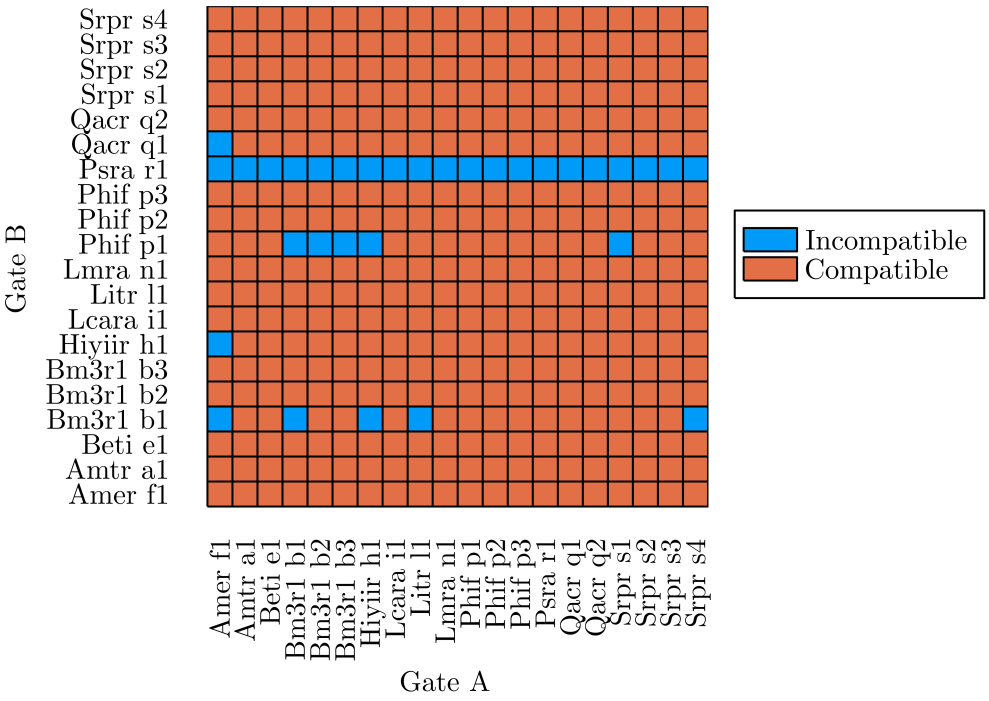


Figure 3.13: Grid showing the different connected pairs of NOT gates in the library with the output of gate A (x axis) serving as the input of gate B (y axis). Pairs which are compatible can be connected together correctly and are coloured in orange. This grid shows the results for the library where gates from any of the seven contexts can be connected. There are 368 pairs of compatible gates.

Strain	Plasmid	Maximum depth (number of NOTs)
KT2400	pSeva221	2
KT2400	pSeva231	2
KT2400	pSeva251	1
DH5α	pSeva221	3
DH5α	pAN	2
CC118λpir	pSeva221	2
CC118λpir	pSeva231	4
KT2400	pSeva221 or 231 or 251	3
DH5α	pSeva221 or pAN	4
CC118λpir	pSeva221 or 231	4
Any	Any	9

Table 3.4: A table of the depths, measured in number of sequentially connected NOT gates, for the single context libraries, libraries with variable plasmid vector, and the complete library using all hosts and plasmid vectors.

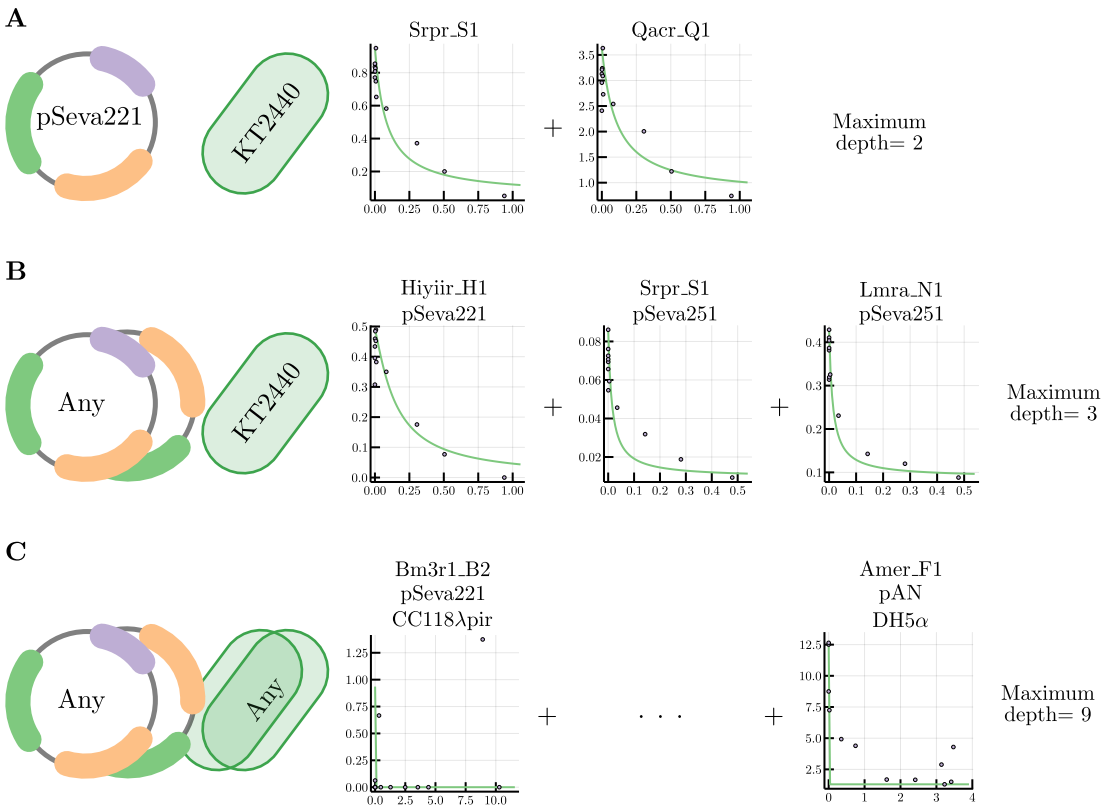


Figure 3.14: Connecting NOT gates in a sequence of compatible gates. In **A** the maximum length of such a chain when using just the KT2400 and pSeva221 context is 2. In **B** the maximum length increases to 3 when other plasmid vectors are allowed, and in **C** a chain of length 9 is made possible by allowing the use of all hosts and plasmid vectors.

vector pairing) as well as each strain with any available plasmid, and for the complete library including all contexts. There are again differences in this measure according to context as shown in Table 3.4. CC118λpir with pSeva231 was the best performing single context library, despite having fewer compatible pairs of gates than both DH5α contexts. The maximum length of chains increases when context gates from different contexts are added as expected. An specific example is shown in Figure 3.14, where a depth of 2 gates is possible using the single context library of KT2400 and pSeva221, adding different plasmids to the library increases possible depth to 3, and finally adding the option of using different hosts allows for a depth of 9.

Statistical model of the synthetic gene networks

The raw flow cytometry data produces histograms of the samples of each population. The response function model is useful in that it is very easy to work with and straightforward

to interpret, but the fact that it is based on point estimates means that some information obtained in the experiment is not included in the model [26]. In particular, the higher moments of the data, such as the variance and skewness (2^{nd} and 3^{rd} moments) are not represented in the reponse function model. These are important in the setting of synthetic biological design because they impact on the reliability of gene networks and in particular reflect the likelihood that noise in the network will result in incorrect interpretation of the input and output signals.

Functional gates with confidence intervals

This can be seen in Figure 3.1 where significant overlap exists between the distribution of expression at IPTG concentrations that should induce different output states of the NOT gate. This is more clearly shown by constructing the distribution of the differences between expression levels of gates at an IPTG concentration of 0 and of $2000\mu M$. The distribution of the differences is approximated by drawing samples from the distribution at 0 IPTG, and subtracting samples drawn from the distribution at $2000\mu M$ IPTG. For a correctly functioning NOT gate this difference is expected to be positive, since the presence of IPTG should be interpreted as a 1 signal and the output expression should be high to represent a 1 signal. In Figure 3.15A this distribution is shown for the gate *Bm3r1_B1*, but it can be seen that the expectation of the distribution is roughly 0. This means that the expression of YFP from the NOT gate for both levels of input is almost the same, making it very likely that the output is interpreted incorrectly.

In contrast Figure 3.15 shows a good example of the NOT gate *Lmra_N1*. The expected value of the difference distribution is positive. In addition 95% of samples from the difference distribution are above 0, meaning that output expression in response to low inputs are higher than output expression in response to high inputs with 95% probability, which is the intended function of the NOT gate.

This probability might be used to score the quality of NOT gates and as a basis for measuring the quality of a library of gates. Table 3.5 shows these probabilities for gates in the context of KT2400 with the pSeva221 plasmid vector. In order to interpret these scores as corresponding to functional or nonfunctional gates, a threshold or confidence interval can be chosen above which the gate is considered functional. Table 3.6 shows how many gates in each context are considered functional under two different confidence intervals, 95 and 66%. This is a measure of the quality of each library that takes into account the

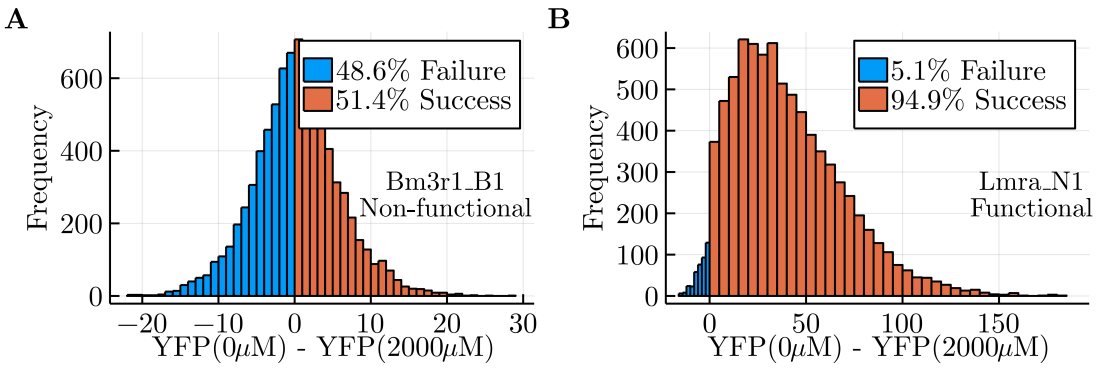


Figure 3.15: Two examples of difference distributions for genetic NOT gates. These are constructed by drawing samples from the flow cytometry data for the NOT gate at an IPTG concentration of 0, and subtracting samples drawn from the flow cytometry data collected at an IPTG concentration of $2000\mu M$. In **A** the difference distribution of the gate *Bm3r1_B1* in KT2400 with pSeva221. In this case it is clear that the NOT gate does not perform well, and in fact the output expression provides almost no information about the induction level. In **B** the difference distribution of the gate *Lmra_N1* in the same context. The NOT gate performs much better, where the uninduced state produces a higher output expression for almost 95% of samples.

information on the variability of expression given by the flow cytometry distributions of the NOT gates.

Comparison of these scores with those of Table 3.3 it can be seen that both measure agree that the library is of better quality in the *E.coli* hosts. At a 95% confidence interval the library in CC118λpir with the pSeva231 plasmid vector is significantly better than any other, but it is notable that even at a modest 66% interval, not all NOT gates are functional. It is also worth mentioning that in Table 3.6 the order of the different contexts changes depending on the size of the interval considered.

An appropriately chosen statistical model can capture higher moments in the data and model the entirety of the experimentally obtained histogram. The aim is to choose probability distributions for each synthetic gene network that reliably predicts the probability of finding a sample with a particular value for an experiment. That is we want a model $P(x;\theta)$, where:

$$\int_{-\infty}^{\infty} P(x;\theta) = 1 \tag{3.8}$$

and $P(x;\theta) = y$, where y is the probability of a cell sampled in the flow cytometry experiment having a YFP fluorescence signal of x. The parameters of the model are θ,

Host	Plasmid vector	Gate	Success percentage
KT2400	pSeva221	Amer_f1	0.63
KT2400	pSeva221	Amtr_a1	0.87
KT2400	pSeva221	Beti_e1	0.67
KT2400	pSeva221	Bm3r1_b1	0.55
KT2400	pSeva221	Bm3r1_b2	0.63
KT2400	pSeva221	Bm3r1_b3	0.51
KT2400	pSeva221	Hiyiir_h1	0.93
KT2400	pSeva221	Lcara_i1	0.96
KT2400	pSeva221	Litr_l1	0.92
KT2400	pSeva221	Lmra_n1	0.95
KT2400	pSeva221	Phif_p1	0.61
KT2400	pSeva221	Phif_p2	0.68
KT2400	pSeva221	Phif_p3	0.91
KT2400	pSeva221	Psra_r1	0.75
KT2400	pSeva221	Qacr_q1	0.93
KT2400	pSeva221	Qacr_q2	0.85
KT2400	pSeva221	Srpr_s1	0.95
KT2400	pSeva221	Srpr_s2	0.87
KT2400	pSeva221	Srpr_s3	0.93
KT2400	pSeva221	Srpr_s4	0.88

Table 3.5: The percentage of samples where uninduced expression is higher than induced expression in NOT gates in the KT2400 with pSeva221 context. Higher percentages indicate better NOT gates, since expression should be higher in the NOT gate in the uninduced condition.

Host	Plasmid vector	95% Functional	66% Functional
KT2400	pSeva221	1	15
KT2400	pSeva231	3	13
KT2400	pSeva251	1	8
CC118λpir	pSeva221	2	13
CC118λpir	pSeva231	10	17
DH5α	pSeva221	4	15
DH5α	pAN	4	17

Table 3.6: The number of NOT gates that are functional in each context. The number of functional gates is shown for two tolerances, 95% and 66%, where this percentage is the required percentage of samples for which the distribution of the difference between uninduced and induced expression is above zero.

and are to be estimated from the data. P is a probability density function, and the choice of P could be made for each of the four types of synthetic gene network characterised in this work, in order to reflect the different processes occurring in each to contribute to fluorescence output.

Statistical model for constitutive expression

The synthetic gene network used for normalisation (shown in Figure 3.4A) constitutively expresses YFP and is the simplest network expressing fluorescent protein for which data was available. It has been previously reported that a Gamma distribution can be used to model steady state concentrations of constitutively expressed proteins in populations of bacterial cells [25].

The Gamma distribution has the probability density function:

$$P(x; k, \theta) = \frac{x^{k-1} e^{\frac{-x}{\theta}}}{\Gamma(\alpha)\theta^k} \tag{3.9}$$

This probability density function is derived from the chemical master equation for a simple model of gene expression shown in Figure 3.16 that includes the processes of transcription and translation. For this reason the parameters of the distribution, k and θ, enjoy a physical interpretation and the results of parameter estimation for this model can to some extent be interpreted mechanistically. Specifically, the parameter k relates to the number of RNA bursts per cell cycle, produced as a result of transcription events. The parameter θ relates to the number of proteins produces per RNA, as a result of translation events.

The gamma distribution model captures the variance in gene expression levels due to the stochastic nature of the processes involved in transcription and translation. Here we adapt this model slightly into a hierarchical model to also capture heterogeneity in the bacterial population. The model is written as follows:

$$k_i \sim LogNormal(\mu_1, \sigma_1) \tag{3.10}$$

$$\theta_i \sim LogNormal(\mu_2, \sigma_2) \tag{3.11}$$

$$Y_i \sim Gamma(k_i, \theta_i) \tag{3.12}$$

where Y_i is the expression level of the i^{th} sample from the flow cytometry experiment, and k_i and θ_i are the k and θ parameters for the i^{th} sample. In this model each sample has its own parameters k and θ and the result of parameter estimation yield the distributions of k and θ over the sample population.

The model is written as a probabilistic program and Markov chain monte carlo methods are used to estimate the parameters $\mu_{1,2}$ and $\sigma_{1,2}$ from Equations 3.10. Figure 3.16 shows the results of the estimation for constitutive expression from the synthetic network shown in Figure 3.4A, in the context of KT2400 with pSeva221 plasmid vector in the absence of IPTG. k_i and $theta_i$ have a joint distribution shown in A. Visualising this joint distribution is important to understand that there is a correlation between these two variables, such that they should be sampled together, and is also useful for comparing the relative dispersion of both. The marginal distributions of k_i and θ_i across the population are shown in B and C. These marginal distributions show that k_i is more widely dispersed than θ_i. In D, a histogram of samples from the experimental data (blue bars) is plotted with the density (orange line) of the posterior predictive distribution for Y_i. The posterior predictive shows good agreement with the sample frequency density shown in the histogram.

The interpretation of the estimation is difficult for a single experiment in a single context. Although the models parameters have a physical meaning (in the sense that they determine the distributions of the variables k and θ, which have a mechanistic meaning as described above), the units of these parameters are not calibrated like the RPU units of expression used in the previous section. For example, the spread of the distributions in k and θ shown in Figures 3.16 B and C cannot be interpreted as being large or small in the absence of a suitable scale — k and θ are both dimensionless [25].

However, meaningful comparison can still be made between experiments or contexts. In order to solely identify the contextual effects of host and plasmid vector the posteriors for k and θ can be estimated and their difference compared. These differences should correspond to effects of the context, where they occur.

3.4 Discussion

Synthetic biological networks are dependent on the context in which they are placed at a fundamental level. For the case of the genetic logic gate library studied in this chapter, the results presented here clearly show that synthetic gene networks perform differently in

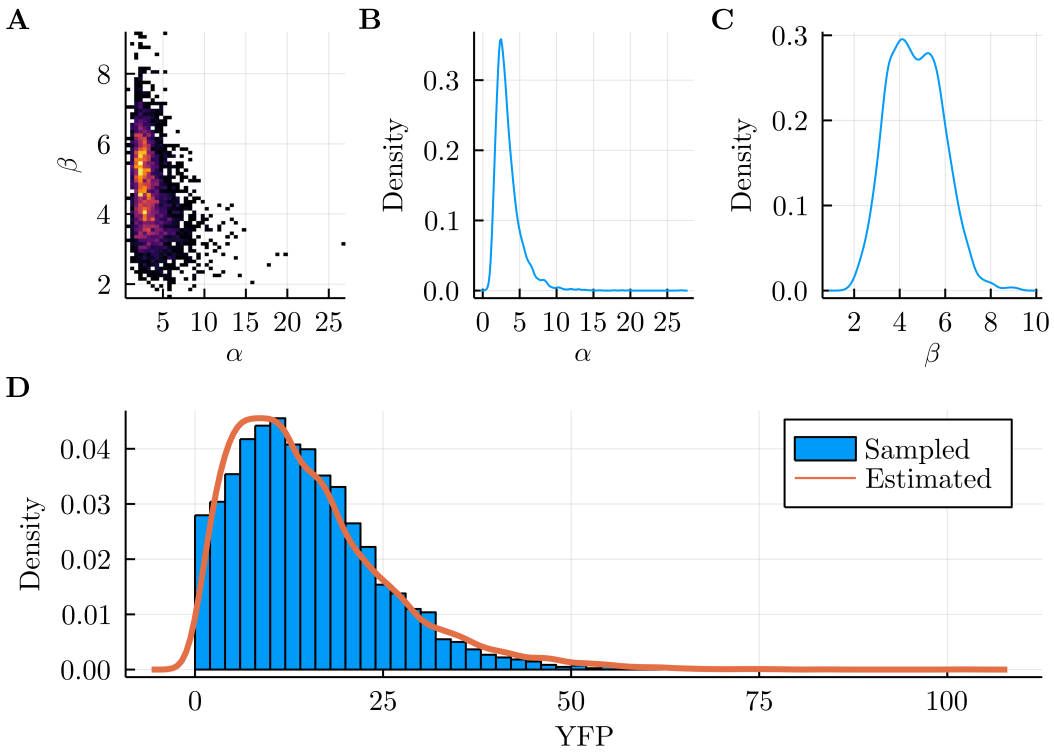


Figure 3.16: The results of using a Hamiltonian monte carlo markov chain algorithm to estimate the parameters of the model in Equations 3.10 from data sampled from flow cytometry for a constitutively expressed YFP in the context KT2400 with pSeva221. **A** shows the joint distribution of the variables k and θ sampled from the posterior. Their marginal distributions are shown in **B** and **C**. **D** shows the posterior distribution of YFP expression predicted by the model (orange line), against a frequency density histogram of another (testing) sample from the same population (flow cytometry experiment).

different contexts, despite being genetically identical.

As such the usefulness of the library for the design and construction of genetic logic circuits is also variable depending on the context in which the library is deployed. Most notably, quantitative and qualitative differences in the performance of the gates mean that the number of gates in the library which operate as intended in a particular context varies. A consequence of this variability is that the usefulness of the library can be improved by deploying different gates in different contexts, adding host organism and plasmid vector as a parameter in the design of genetic logic circuits.

This is an example of being able to use context as a parameter to engineer better synthetic biological networks. However this approach is based on differences between performance observed in experiments. Under the response function model it is difficult to interpret a physical meaning for the parameters and extract the information that would be required to understand why these differences in performance are observed. The model cannot be used to extrapolate to unseen contexts, or indeed to unseen network topologies, because this requires an understanding of the changes to the mechanisms which manifest as difference in performance between contexts.

Quantified differences between gate performance

A model of the gene networks can be used for the quantification of the differences in performance between contexts. Initially a response function model, an hill equation fitted to point estimates of expression data, was used to model the gene networks. Comparison of these models functions as a proxy for comparison of the performance of the gene networks themselves, at the cost of the comparison 'inheriting' the limitations of the model. For example, the response function model ignores variance and heterogeneity in gene expression, because it is based on point estimates of actual expression data.

Seven contexts were used in this study to highlight these differences. The two variable elements of the contexts were the plasmid vector into which the gene networks were synthesised, and the host organism into which these plasmids are transformed. Quantification of the differences in performace between contexts shows that the impact of these aspects of context vary. Figure 3.8 shows how host organism has more impact on performance than plasmid vector, with the differences between *E. coli* and *P. putida* contexts being greater than contexts in the same host with different plasmid vector. This is reasonable,

since the host provides all the machinery for genetic expression and regulation required by the genetic logic gates, as well as the machinery for replication of the plasmid vector.

At the level of individual gates, inspection of the response function models suggests that there is a complex interaction between both of the host organism, plasmid vector and gene network that produces the differences in performance. For example, there are cases where the type of transformation due to change of plasmid vector seen in one host is not reflected by the same change of plasmid vector in another. In Figure 3.6 this is seen for the case of a change of plasmid vector pSeva221 to pSeva231 — in the host KT2440, this results in a decrease in maximum expression, whereas in host CC118λpir, an increase in the maximum expression is observed.

Contextual differences as design parameters

The NOT gate should transform low expression levels into high expression levels and vice versa, but for some gates in some contexts this is not true. In these cases the gate is nonfunctional and the response function model can be used to determine (and actually define) this. The number of NOT gates in a particular library can be used to measure the library's quality, and this quality score is shown to depend on the context in which the library is used. Unsurprisingly the *E. coli* contexts, for which the library was previously optimised, scored most highly under this measure. Since some gates are functional in some contexts and not others, mixing contexts can improve this quality score. Synthesising different parts of a genetic circuit in different plasmid vectors for example might be considered, integrating context as a parameter in the design stage in order to exploit contextual effects on the individual gates.

The compatibility of pairs of NOT gates was also used to score the quality of the library for each context. This is important since circuits are built from composing individual gates, and gates can only be composed if they are compatible. Here the choice of model is important, and the model is used to define the notion of compatibility. Again it was seen that the quality of the library varied with context. Some contexts (such as KT2440 with pSeva251) had no compatible pairs of gates. However it can also be seen that combining different contexts, that is placing the pair of gates in two different contexts, can significantly increase the number of compatible pairs in the library. This can be seen as the increase of such compatible pairs in Figures 3.11, 3.12 and 3.13, suggesting that contextual parameters could be used to increase degrees of freedom in the combinatorial design of genetic logic

circuits.

Deterministic models based solely on point estimates of the flow cytometry distributions, such as the response function model of Equation 3.1, ignore information about population level heterogeneity that is contained in the experimental data. Since gene expression is an inherently noisy process [24] there is always a chance that a NOT gate will fail to perform as expected. The question really is how often the NOT gate can be expected to perform well. This question is only answered well by considering the entire distribution of gate performance as in Figure 3.15, where the difference distribution between the NOT gate in the 'on' and 'off' states is sampled. Assessment of the quality of the gate library using this approach can be significantly less optimistic than using the response function model.

Future work

Here we suggest how to use context as a parameter to engineer better synthetic biological networks using this library. However the approach is based on differences between performance observed in experiments. Under the response function model it is difficult to interpret a physical meaning for the parameters and extract the information that would be required to understand why these differences in performance are observed. The model cannot be used to extrapolate to unseen contexts, or indeed to unseen network topologies, because this requires an understanding of the changes to the mechanisms which manifest as difference in performance between contexts.

The use of probabilistic programming as part of the characterisation of the synthetic gene networks of this library is a promising in this respect. The parameters of the model presented in Equations 3.10 have a rough physical interpretation as relating to two processes fundamental to the behaviour of gene networks (transcription and translation). Potentially this could lead to understanding of mechanistic details of contextual effects.

Further since these programs (actually models) are composable they lend themselves to the combinatorial design process, and could possibly be used to extrapolate and make predictions about the performance of unseen gene networks in different contexts. In addition they capture the inherent heterogeneity of bacterial populations, a feature that is captured by flow cytometry experiments but is ignored by deterministic models based on point estimates (such as the response function model). Future work will develop further the probabilistic programs for the synthetic constructs in this library that build on the

model shown in Figure 3.16 to extract mechanistic details of performance that depend on context.

Chapter 4

Toggle switch in the spatial context of the electroactive biofilm

This chapter adapts the publication Grozinger, Heidrich and Goñi-Moreno [37] which has been accepted for publication in Microbial Biotechnology. I performed the design, modeling and numerical analysis of the switch, which is the focus of this chapter.

4.1 Introduction

It is described in Section 2.5 how exoelectrogenic bacteria can couple the oxidisation of a substrate to the reduction of an electrode to produce a MES. The balance of reduction and oxidisation reactions occurring at the electrode can be measured as an electrical current. The rate at which the bacteria in MESs metabolise substrate correlates with the electrical current that flows at the electrode. Ueki et al [89] used this principle to demonstrate genetic control of the electronic output of *Geobacter sulfurreducens*. By controlling the expression of genes required for acetate metabolism, such as *gltA* (citrate synthase) and *ato1* (acetyl-CoA transferase), the electronic output of the *Geobacter* could be switched on or off using a synthetic gene network.

The potential of the electrode is an important property of the MES that can be sensed in a variety of ways by different bacteria [40, 4, 49]. Tschirhart et al [87] used the redox sensitive promoter *pSoxR* to control rates of transcription in a synthetic gene network hosted by *Escherichia coli*. In that study redox mediator molecules were used to communicate the potential of an electrode to the synthetic gene network that controlled the expression of genes of interest. The electrode potential can be set using electronics, so this system demonstrates electronic control of gene expression.

A synthetic biological system which combines *both* the electronic control of gene expression and the genetic control of electronic output has not yet been presented in the literature. This Chapter first presents a design for a synthetic biological network which adapts the genetic toggle switch [27] such that it can:

- Be hosted by bacteria in a MES.
- Take electronic input in the form of electrode potential.
- Process this input using a synthetic gene network.
- Produce electronic output in the form of electrical current at the electrode.
- Retain the qualitative dynamics of a toggle switch (bistability).

A simple model of a MES which includes its important components and interactions between them is shown in Figure 4.1. Using this model it is possible to design a synthetic network which combines genetic control of substrate metabolism and feedback from the

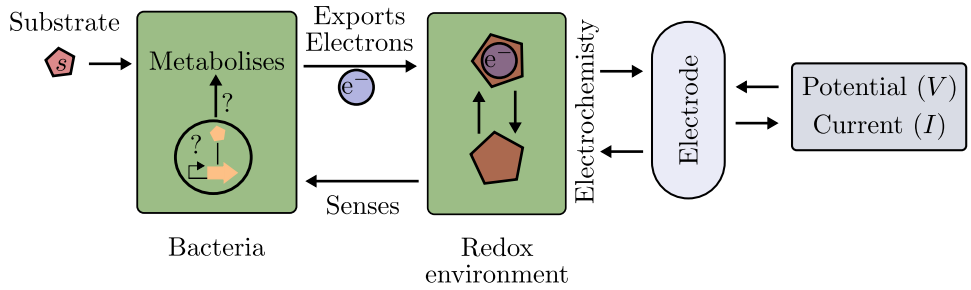


Figure 4.1: An abstract view of a MES where the overall reaction is substrate oxidation coupled with electrode reduction to produce electrical current. The substrate (left) is taken up by the electrogenic bacteria and metabolised. The resultant electrons are exported to electron acceptors in the redox environment surrounding the bacteria. The redox environment connects the bacteria to the terminal electron acceptor, the electrode, where electrochemical reactions can produce current (I). If the bacteria are engineered with synthetic gene network which controls their metabolic rate (marked as ?), then genetic control of current (I) can be achieved [89]. As well as the metabolic activity of the bacteria, the potential (V) of the electrode affects the redox environment by changing the ratio of oxidised to reduced electron acceptors at equilibrium. If synthetic gene networks can sense this ratio then electronic control of gene expression can be achieved [87].

redox environment to produce bistability. We call this network the 'electrogenetic toggle switch'.

However, problems arise because previous studies have presented evidence that show the redox environment cannot be assumed to be homogeneous [76]. The electroactive biofilm is actually an asymmetrical spatial domain with an electrode at one boundary that acts as a sink for electrons, and a bulk substrate solution at another boundary acting as an electron source.

This asymmetry gives rise to gradients that mean bacteria close to the electrode are presented with a different redox environment than those that are farther from the electrode. Of course, when designing a synthetic biological network, we might expect that it behaves homogeneously

Considering this heterogeneous context it is not obvious that an electrogenetic toggle switch within the biofilm would admit the emergence of bistability under conditions that are physiological relevant. Here we develop a model combining elements of previous modelling studies of electroactive biofilms and a model of the electrogenetic toggle switch to answer this question.

4.2 Results

Model of wildtype electroactive biofilm

Exoelectrogens such as *Geobacter* form biofilms attached to electrodes. The biofilm provides the redox environment for the bacteria and connects the bacteria to the electrode. The bacteria oxidise substrate and export the resultant electrons to the biofilm which transports them to the electrode. A simple model for this process in a single step is as follows:

$$s + Q \xrightarrow{\alpha_q} q \tag{4.1}$$

s is the substrate. Q is free capacity for electrons in the biofilm. The electroactive biofilm has a finite capacity for electrons that limits the rate at which it can transport electrons, and consequently limits the rate at which the bacteria can export electrons (Figure 4.3D). q is the electrons that are exported to the biofilm and α_q is the rate of the overall reaction.

At the electrode, electrochemical reactions exchange electrons between the biofilm and electrode to convert q into Q and vice versa.

$$q \underset{k_r(V)}{\overset{k_f(V)}{\rightleftharpoons}} Q \tag{4.2}$$

k_f and k_r are the rates of the forward and reverse electrochemical reactions, which depend on the potential of the electrode V. The current that is observable at the electrode is proportional to the net rate * with proportionality constant F, the Faraday constant. The electrical current I therefore also depends on V, q and Q:

$$I(V) = F(qk_f(V) - Q_0k_r(V)) \tag{4.3}$$

$k_f(V)$ and $k_r(V)$ are defined as in other modeling frameworks using the Butler-Volmer relationship [45].

*The current is here defined as positive in the anodic direction, that is, the reduction of the electrode results in positive current and oxidation in negative current.

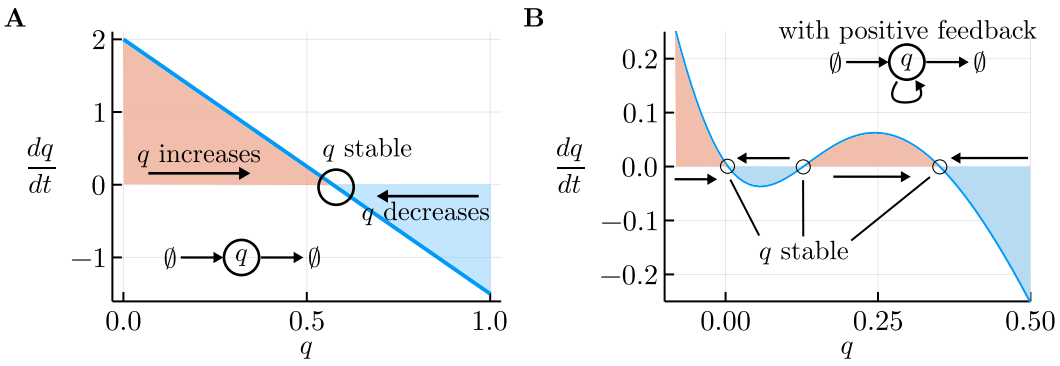


Figure 4.2: q plotted against its rate of change for two models of current production from an electroactive biofilm attached to an electrode. Fixed points of the systems are shown where the rate crosses 0. In **A** the simplified wild type model from Equation 4.6 and sketched in the inset. The only type of behaviour this system can exhibit is monostability, owing to the linear rate of change in q, negative slope and positive domain of q. In **B** the same simplified model but with positive feedback for q, written as in Equation 4.18. The positive feedback allows more varied dynamical behaviour, including bistability as shown in this example.

$$k_f(V) = j_0 e^{AV} \tag{4.4}$$

$$k_r(V) = j_0 e^{-AV} \tag{4.5}$$

where A is assumed to be a constant $A = \frac{F}{2RT}$ and j_0 is the rate at which electrons are exchanged with the electrode at equilibrium potential of the electrode and biofilm, so that $j_0 = k_f(0) = k_r(0)$.

The goal of the electrogenetic toggle switch is to design a system where $I(V)$ can come to rest at two different values for the same V. This is clearly only possible if q can come to rest at two different values for the same V.

With the assumptions of constant substrate s contributing to an overall electron export rate of α, constant finite biofilm capacity Q_0 and electrode potential V a single ODE equation can model the system's evolution over time, where $q(t)$ is the concentration of q at time t *.

$$\partial_t q(t) = (\alpha + j_0 e^{-AV})(Q_0 - q(t)) - q(t) j_0 e^{AV} \tag{4.6}$$

*In the following I will abreviate state variables $y(t)$ with y where it is clear enough from context that $y(t)$ is meant

In this Equation there are two terms, one for production of q and one for depletion of q. Where these terms are balanced there will be a fixed point of the system. An example is shown is Figure 4.2A where q is plotted against its rate of change $\partial_t q$. Since $\partial_t q$ is linear in q, there is a single fixed point. Furthermore, this fixed point is stable since for q above the fixed point, depletion is higher than production and q will decay back to the fixed point, and for q below the fixed point, the opposite is true. In fact, Figure 4.2A provides the geometric argument that there can only be exactly one stable fixed for this system. Therefore bistability, which requires three fixed points cannot exist in this system.

Adding limiting substrate

In Equations 4.6 and 4.18 substrate is assumed constant and not limiting. To relax this assumption I model substrate as entering the biofilm by diffusing through the interface between the bulk media and biofilm. The bulk media has a constant concentration of substrate S_0 and diffuses into the biofilm in the direction of the gradient across the interface at a rate D_s as shown in Figure 4.3B. This means that substrate can be depleted in the biofilm if it is consumed faster than it can enter.

Available substrate is consumed by the exoelectrogens in the biofilm at a rate modeled here using the Michaelis Menten equation. The model now must include a state variable $s(t)$ for substrate concentration.

$$\partial_t q = \left(\frac{\alpha_q s}{K_s + s} + j_0 e^{-AV} \right) (Q_0 - q) - q e^{AV} \tag{4.7}$$

$$\partial_t s = D_s (S_0 - s) - \left(\frac{\alpha_q s (Q_0 - q)}{K_s + s} \right) \tag{4.8}$$

Adding transport through the biofilm

The previous models do not include any spatial dimensions of the biofilm. If both substrate, s, and electrons, q, are transported infinitely fast there is no need to model space, since the concentrations of s and q will be homogeneous throughout the biofilm. However there is certainly a finite rate of transport in the biofilm, and there is good evidence that transport of q in the biofilm can be a limiting factor in current production.

In Equations 4.6 and 4.18 substrate is assumed constant and not limiting, and both

substrate and electrons can be transported through the biofilm instantaneously so that space does not need to be modeled. Now these assumptions are relaxed. Initially a model is considered which includes a single spatial dimension of the biofilm x, and the transport of both s and q through the biofilm. The electrode is placed at $x = 0$ and the bulk-biofilm interface at $x = L$. Typical values for L are in the tens of microns.

As described above, substrate enters the biofilm through the bulk-biofilm interface. Once in the biofilm the substrate diffuses at a rate of D_s as in shown in Figure 4.3B. Substrate cannot diffuse through the electrode-biofilm interface. This leads to the boundary conditions in Equations 4.14 and 4.17. The diffusion of substrate in x is described using Fick's second law to derive a PDE for $s(x,t)$, the concentration of substrate at position x at time t. $s(x,t)$ is modeled with the PDE in Equation 4.12.

Electron transport is modeled here as occurring by hopping between nearby redox centres that are bound to the extracellular matrix of the biofilm. This hopping process is illustrated in Figure 4.3A. Each redox centre in the model can be occupied by only a single electron, such that electrons cannot hop to centres that are already occupied (Figure 4.3C). The concentration of occupied redox centres at position x in the biofilm is $q(x,t)$ and the maximum $q(x,t)$ is the total capacity of the biofilm for electrons Q_0. The ratio of occupied to unoccupied redox centres defines the redox environment $\frac{q(x,t)}{Q_0+q(x,t)}$.

To model the transport of electrons by hopping the biofilm the following set of reactions are considered in a one dimensional biofilm discretised into n sections:

$$q_i + q_{i+1}^+ \xrightarrow{\text{k}} q_i^+ + q_{i+1} \quad \text{for } i \in [1, n)$$
$$q_i + q_{i-1}^+ \xrightarrow{\text{k}} q_i^+ + q_{i-1} \quad \text{for } i \in (1, n]$$

where q_i are occupied redox centres in the i^{th} section of the biofilm, q_i^+ are unoccupied redox centres in the i^{th} section of the biofilm, and k is the rate of electron hopping between redox centres. Using the assumptions of mass action kinetics to convert these reactions into an ODE gives:

$$\partial_t q_i = k\left(q_{i-1}q_i^+ + q_{i+1}q_i^+ - q_i q_{i-1}^+ - q_i q_{i+1}^+\right) \tag{4.9}$$

Since biofilm capacity is assumed finite and constant, $Q_0 = q_i + q_i^+$, and this ODE can be

written as:

$$\partial_t q_i = k\ Q_0\ (q_{i-1} - 2q_i + q_{i+1}) \tag{4.10}$$

$$= kQ_0\Delta x \left(\frac{q_{i-1} - 2q_i + q_{i+1}}{\Delta x^2} \right) \tag{4.11}$$

The right hand side of this equation is the central finite difference approximation of $D_q \partial_x^2 q(i,t)$, if $D_q = kQ_0\Delta x$, where Δx is the length of the discretised sections of the biofilm. Making this substitution allows for modeling electron hopping as diffusion, and I will use this approach to model electron transport in the biofilm.

Exchange of electrons between the electrode and the biofilm is really the exchange of electrons between the electrode and those redox centres close to or in contact with the electrode as shown in Figure 4.3B. In the spatial model of the biofilm this electron exchange that produces current is written using the boundary condition of Equation 4.15. Through the biofilm-bulk interface no electron transport occurs and the corresponding boundary condition is Equation 4.16.

The full biofilm-electrode model is as follows.

$$\partial_t s(x,t) = D_s \partial_x^2 s(x,t) - \frac{\alpha_q (Q_0 - q(x,t)) s(x,t)}{K_s + s(x,t)} \tag{4.12}$$

$$\partial_t q(x,t) = kQ_0\Delta x \partial_x^2 q(x,t) + \frac{\alpha_q (Q_0 - q(x,t)) s(x,t)}{K_s + s(x,t)} \tag{4.13}$$

$$\partial_x s(0,t) = 0 \tag{4.14}$$

$$\partial_x q(0,t) = j_0 \left(q(0,t) e^{AV} + q(0,t) e^{-AV} - Q_0 e^{-AV} \right) \tag{4.15}$$

$$\partial_x q(L,t) = 0 \tag{4.16}$$

$$s(L,t) = S_0 \tag{4.17}$$

Gradients form in the biofilm with this model

Equations 4.12-4.16 describe a model which permits the formation of gradients in both substrate and electrons in the biofilm. For example, substrate can be abundant in some

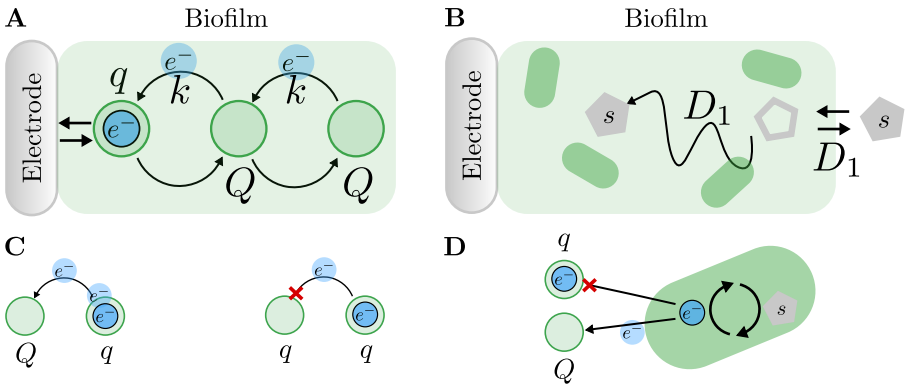


Figure 4.3: Illustration of the mechanisms of transport in the electroactive biofilm. **A** shows how electrons q are transported by hopping between neighboring redox centres. It is the electrons that move not the redox centres. Redox centres are the electrode interface exchange their electrons with the electrode and generate current. Hopping can only occur from occupied (q) to unoccupied (Q) redox centres, as shown in **C**. Substrate diffuses into and through the biofilm at rate D_s shown in **B**. Substrate s is converted into q and exported to the biofilm's unoccupied redox centres Q, shown in **D**.

regions of the biofilm (close to the biofilm-bulk interface) and depleted in others (close to the electrode). The magnitude of the gradients depends on the rates of biofilm transport, but also on the activity of the bacteria in exporting electrons.

This is seen in Figure 4.4A and B where substrate and electron concentration gradients are plotted from numerical simulations of Equations 4.12-4.16. Substrate tends to be more available closer to the biofilm-bulk interface than at the electrode, where it is depleted. Figure 4.4A predicts that substrate is almost completely absent at the electrode-biofilm interface if activity of the bacteria (α_q) is high, whereas it is almost fully available at the biofilm-bulk interface. The same is true for electron concentration in the biofilm in Figure 4.4B, meaning that unoccupied redox centres Q are most available at the biofilm-electrode interface, whereas the biofilm is almost at capacity close to the biofilm bulk interface.

Increased α_q also allows the biofilm to consume more substrate and produce more current density at the electrode as shown in Figure 4.4C. However it also produces a more heterogeneous context for the bacteria in the biofilm. In Figure 4.4D shows the contribution at each x position in the biofilm to the overall steady state current density. This shows the real consequence of gradients in the substrate and electron concentration — the bacteria rely on exporting electrons to derive energy from the substrate, so some bacteria will be able to grow while others starve. This is an aspect of the biofilm that I do not model here.

In general, the model predicts that increased activity of the bacteria in terms of increased α_q leads to larger magnitude gradients of all the state variables of the model and more

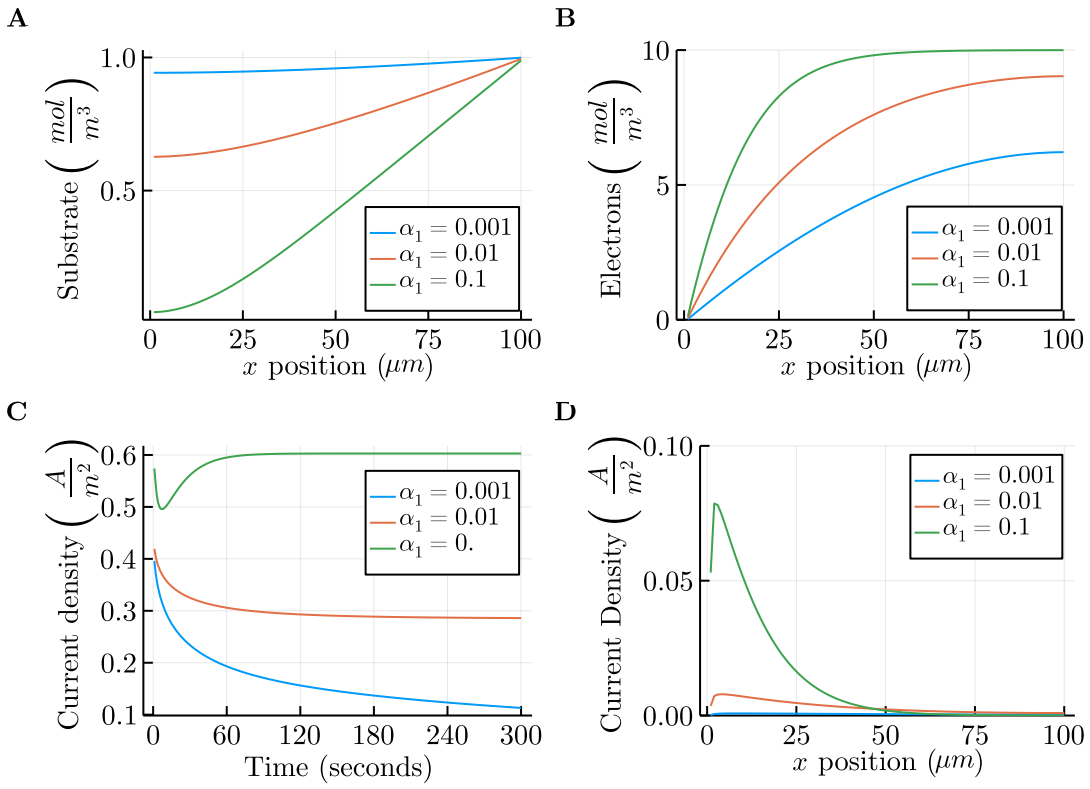


Figure 4.4: Plots of the gradients in electroactive biofilms and the current they produce, predicted using numerical simulation of Equations 4.12-4.17. **A** shows the gradient in substrate, which tends to be more available deeper in the biofilm and depleted close to the electrode. At high activities of the bacteria (α_q) this difference is the largest. The concentration of electrons q in the biofilm also exhibits this kind of gradient with a different shape shown in **B**. Again high α_q produces larger gradients. **C** shows the current density I over time at the electrode for the various α_q, and **D** shows the contributions to I from different depths in the biofilm.

heterogeneity in the biofilm.

Altogether the biofilm presented here offers a heterogeneous context where the environment in which any bacteria finds itself depends on its x position in the biofilm. This extends to any synthetic gene networks that the bacteria host.

Introducing bistability into the model

As described in Section 2.3, the toggle switch requires bistable dynamics. Figure 4.2A makes a geometric argument showing that bistability is impossible in the wild type system, and so modifications must be made to the wild type model in order to produce the bistability required for the electrogenetic toggle switch.

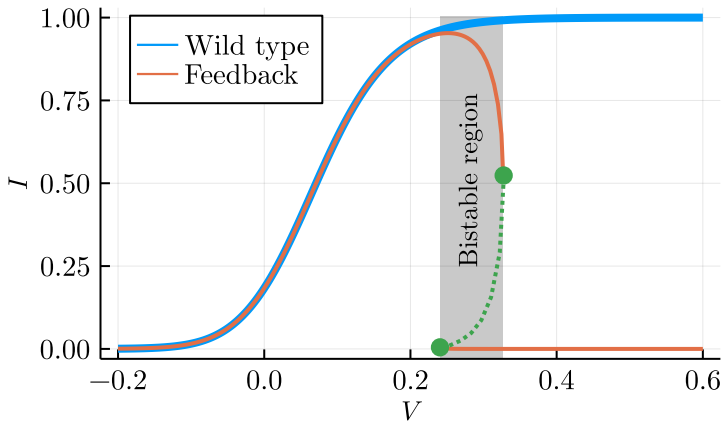


Figure 4.5: Bifurcation diagram for the wild type model of Equation 4.6 and the model with positive feedback of Equation 4.18. The bifurcation parameter is V and is plotted against the fixed point current outputs I. The solid lines are branches of stable fixed points (steady states). The wild type model (blue line) increases monotonically and resembles a sigmoid that approaches a current maximum asymptotically. The model with feedback is similar up to a $V \approx 0.22$ where bistability emerges. Within the bistable region there is a branch of unstable fixed points (green line) and two additional stable fixed points.

Since in this model the output current density I is a function of the state of the biofilm $q(x,t)$, bistability in $q(x,t)$ will result in bistability in I as well. This electronic output measures the state of the electrogenetic switch.

Both I and the state of the biofilm $q(x,t)$ also depend on the potential of the electrode V, another electronic input. The convenience of controlling the state of the biofilm electronically makes it attractive to use V to induce changes in the state of electrogenetic toggle switch that can be measured with I. For this reason the bifurcation analysis here will focus on V as a bifurcation parameter, and looks for bistability in I.

The modification that is required is positive feedback for q. Figure 4.2B uses Equation 4.18 to plot q against its rate of change when positive feedback is introduced. This positive feedback can be modeled using a modification to Equation 4.18.

$$\partial_t q = \left(\frac{\alpha q^2}{K^2 + q^2} + j_0 e^{-AV} \right) (Q_0 - q) - q e^{AV} \tag{4.18}$$

Where K is the value of q which produces half-maximal activation of q. This new ODE admits bistability. Figure 4.2B shows an example, where there are one unstable fixed point between two stable fixed points.

Figure 4.5 shows two examples I-V plots which I use for bifurcation analysis. The blue

line in Figure 4.5 plots the steady states of I against V for the model in Equation 4.6. For all values of V this equation has a single stable fixed point and so a single steady state I, indicating that the system is monostable. The orange line in Figure 4.5 plots the same analysis for the modified Equation 4.18, displaying the qualitatively different behaviour of bistability.

The strategy for obtaining bistability in the electrogenetic toggle switch will be to engineer this positive feedback for q. Just as in the genetic toggle switch [27] this can be done by introducing a second component and using mutual inhibition. In the following I model q's positive feedback by adding a hypothetical genetic network with a single gene a, whose expression both is repressed by and represses q, as shown in Figure 4.7A.

Equation 4.18 can be modified, and another ODE added for a, in order to model this system, *without taking into account spatial context.*

$$\partial_t q = \frac{\alpha_q a^{n_a}(Q_0 - q)}{{K_a}^{n_a} + a^{n_a}} + (Q_0 - q) j_0 e^{-AV} - q e^{AV} \tag{4.19}$$

$$\partial_t a = \frac{\alpha_a q^{n_q}}{{K_q}^{n_q} + q^{n_q}} - \gamma_a a \tag{4.20}$$

Where α_q is the maximum overall electron export rate, α_a is the maximum rate of a expression, γ_a is the degradation/dilution rate of a, K_a and K_q are the half maximal repression coefficients for a and q, and n_a and n_q are the Hill coefficients for the repression of q and a.

The existence of mutual inhibition between a and q provides the positive feedback for q that is needed to admit bistability. However, not all values for the parameters of the system will actually produce bistability, and an objective of the models presented here is to predict those which do. I am particularly interested in the parameters K_q, K_a, n_a, n_q, since they can be modified using synthetic gene networks. That is, they are not part of the context of the system and can be engineered using standard synthetic biological approaches.

In order to aid in the search for bistable parameter sets, I use some simplifying assumptions in order to produce a simplified nondimensional system. First an implicit assumption in Equation 4.18 is that substrate is not limiting and can be regarded as constant. Second a new assumption is introduced that the current density I is linear in q. This is approximately

true for small V. Equation 4.19 is modified to produce:

$$\partial_t q = \frac{\alpha_q a^{n_a}(Q_0 - q)}{{K_a}^{n_a} + a^{n_a}} - \gamma_1 q \tag{4.21}$$

where γ_1 is the proportionality constant of current density to q. The rate at which q 'degrades' in this simplified model.

The dimensionless variables $\tau = \frac{t}{t_c}$, $A = a_c a$ and $B = b_c q$ are introduced. I also let $n_a = n_q = 2$. Using the chain rule we obtain:

$$\partial_\tau A = t_c a_c \partial_t a = \frac{t_c a_c \alpha_a}{1 + \frac{B}{K_2 q_c}^2} - t_c \gamma_a A \tag{4.22}$$

$$\partial_\tau B = t_c b_c \partial_t q = \frac{t_c b_c Q_0 - t_c B}{1 + \frac{A}{K_1 a_c}^2} - t_c \gamma_1 B \tag{4.23}$$

The following substitutions were made for the characteristic constants, $t_c = \frac{K_q}{\alpha_q Q_0}$, $a_c = \frac{1}{K_a}$ and $b_c = \frac{1}{K_q}$ to obtain:

$$\frac{dA}{d\tau} = \frac{P_1}{1 + B^2} - P_2 A \tag{4.24}$$

$$\frac{dB}{d\tau} = \frac{1 - P_3 B}{1 + A^2} - P_4 B \tag{4.25}$$

where $P_1 = \frac{K_q \alpha_a}{\alpha_q K_a Q_0}$, $P_2 = \frac{K_q \gamma_a}{\alpha_q Q_0}$, $P_3 = \frac{K_q}{Q}$ and $P_4 = \frac{K_q \gamma_1}{\alpha_q Q_0}$.

The nullclines of Equations 4.24 and 4.25 are plotted in Figure 4.6A. I used inspection of these curves to adjust parameters $P_{1,2,3,4}$ so that there are three intersections of these nullclines and produce the bistability that is shown in Figure 4.6B. From this set of dimensionless parameters the following parameter relationships can be identified.

$$\frac{\alpha_a}{\gamma_a} = 10 K_a \tag{4.26}$$

$$Q_0 = 5 K_q \tag{4.27}$$

Using these relationships to set the values of the parameters K_a and K_q produces bistability

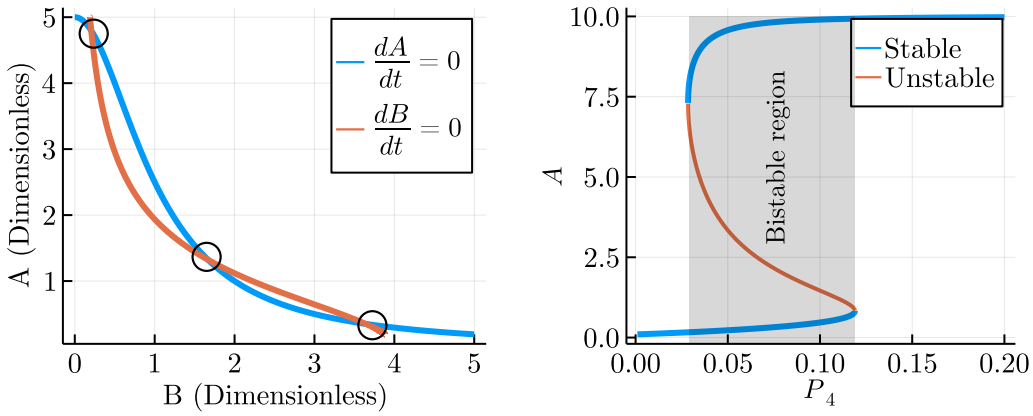


Figure 4.6: Bifurcation analysis of the dimensionless Equations 4.24 and 4.25. **A** plots the nullclines and shows that they can be made to have three intersections, marked with circles, which is a requirement for bistability. **B** shows a bifurcation diagram where A is plotted against the bifurcation parameter P_4. There is a shaded region of bistability from $P_4 \approx 0.025$ to $P_4 \approx 0.12$.

in the system of Equations 4.19 and 4.20. Figure 4.7B shows the bifurcation analysis for this case. The timecourse simulation in Figure 4.7E reveals that the bistability manifests as switching behaviour, where the switching between high and low current states can be achieved by transient steps in the electrode potential V (Figure 4.7C). This behaviour does not occur without the mutual inhibition shown in Figure 4.7A, and the timecourse in Figure 4.7D confirms that this wild type system is monostable, and that transient steps in V have only a transient effect on the current output I.

Bifurcation analysis of a limited substrate model

In Equations 4.7 and 4.8 I introduced the possibility of limitation of substrate using Michaelis Menten kinetics. This is a standard approach to modeling enzyme kinetics with ODEs. These Equations can also be modified to include the positive feedback loop implemented with mutual inhibition shown in Figure 4.7A.

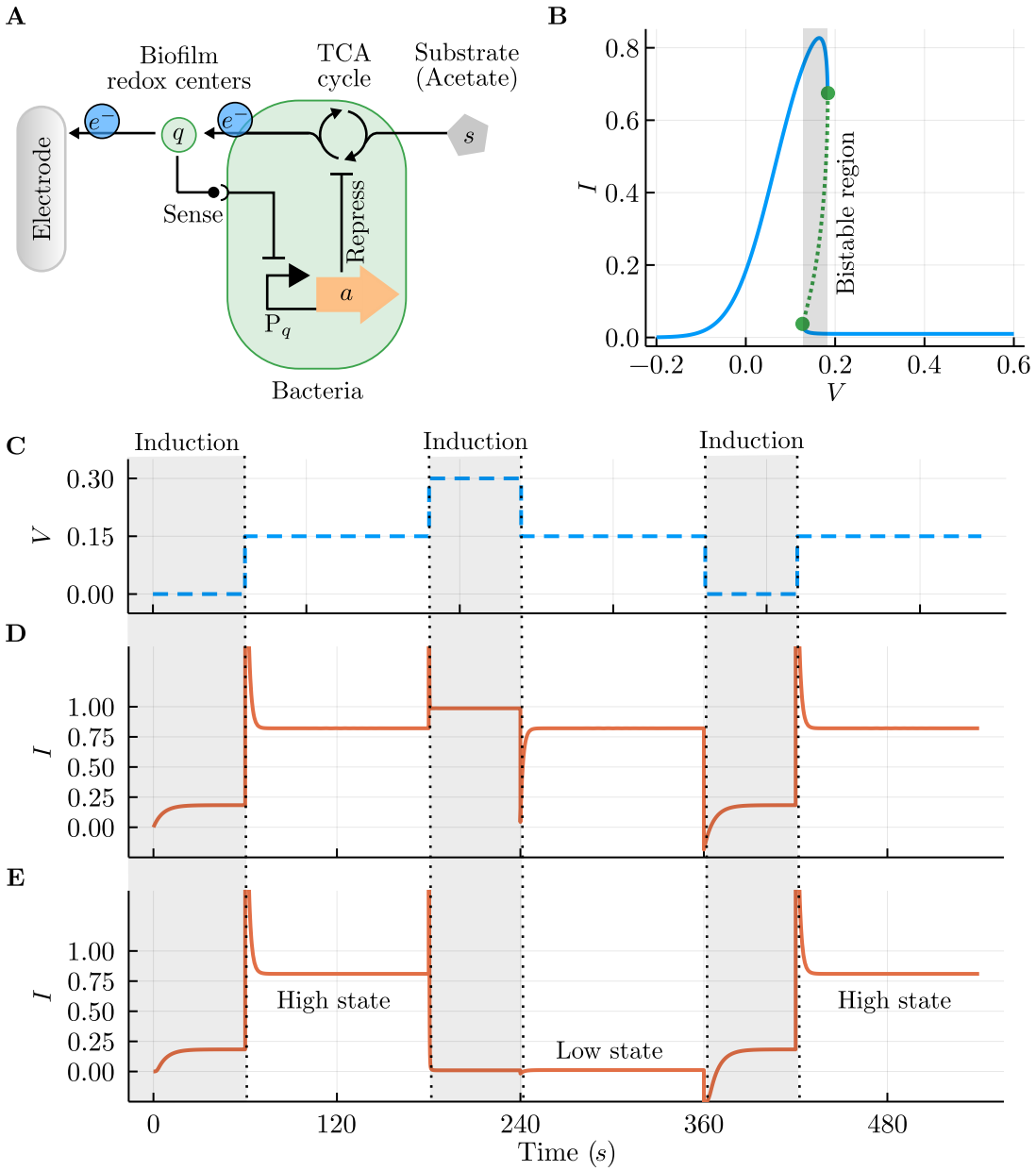


Figure 4.7: **A** shows an abstract model of the mutual inhibition of q and a, amounting to positive feedback for q. Substrate s, for example acetate, is converted to exported electrons via the exoelectrogen's TCA cycle at a rate which is repressed by a gene a. Exported electrons in the biofilm are q and are transported to the electrode to produce current. q can be sensed by the exoelectrogens and result in the repression of a, completing the positive feedback loop. In **B** this model is shown to produce bistability for a small range of electrode potentials V. V can be used as a signal as in **C**, to switch the system between the high and low current states in **E** by taking advantage of hysteresis in the system. **D** shows a timecourse for a wild type system without a feedback loop, and without bistable switching dynamics.

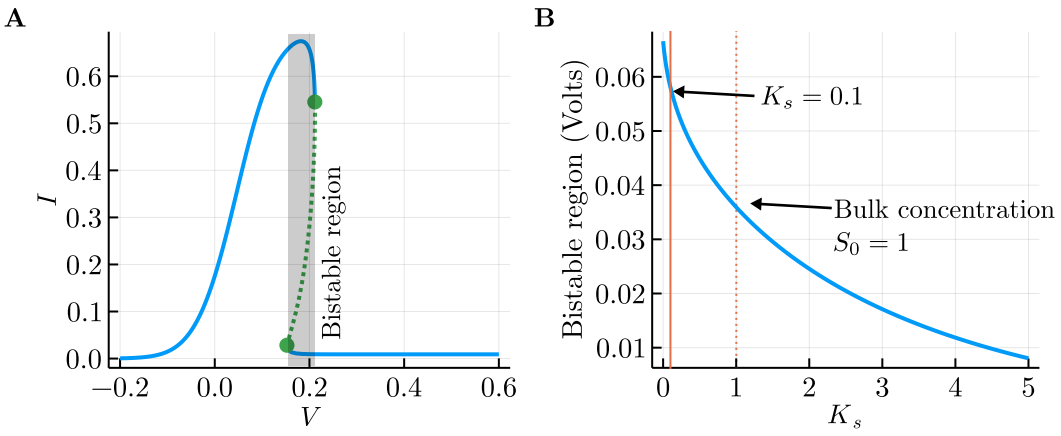


Figure 4.8: Bifurcation analysis for the model which includes substrate limitation, *but not spatial context*. The size of the bistable region is smaller than when substrate is assumed nonlimiting and constant, as shown in **A**. K_s is the parameter which relaxes these assumptions. Higher K_s increases the impact of substrate depletion on q production, and also reduces the size of the bistable region. Bistable region size is plotted against K_s in **B**. $K_s = 0.1$ is the value taken from the literature which I use in this thesis.

$$\partial_t q = \frac{\alpha_q a^{n_a}(Q_0 - q)s}{({K_a}^{n_a} + a^{n_a})(K_s + s)} + (Q_0 - q)j_0 e^{-AV} - qe^{AV} \tag{4.28}$$

$$\partial_t s = D_s(S_0 - s) - \frac{\alpha_q a^{n_a}(Q_0 - q)s}{({K_a}^{n_a} + a^{n_a})(K_s + s)} \tag{4.29}$$

$$\partial_t a = \frac{\alpha_a q^{n_q}}{{K_q}^{n_q} + q^{n_q}} - \gamma_a a \tag{4.30}$$

These new Equations allow substrate depletion to affect the rate at which q is produced. This system is still capable of producing bistability as is shown in Figure 4.8A. However, the bistable region is smaller as compared to the case presented in Figure 4.7B.

Substrate limitation therefore seems to reduce the robustness of the bistability in the model. Figure 4.8B shows how the bistable region shrinks as K_s increases. At $K_s = 0$ the model reduces to that without substrate limitation (the case shown in Figure 4.7B) and the bistable region is largest. With increasing K_s more substrate concentration is required to achieve high substrate conversion rates, and the bistable region shrinks. The value of K_s I use in the rest of this study is taken from [45] and is marked with a vertical line.

Bifurcation analysis of the full model

Until now bifurcation analysis has been performed on models where the spatial context, the position x in the biofilm, has been considered insignificant. This assumption is reasonable where transport of q and diffusion of s through the biofilm are quick enough compared to the rate of electron export α_q that gradients of q and s do not form. The biofilm can be considered as homogeneous and ODEs are sufficient to model the system mathematically.

As stated in Section 4.1 this is often not a good assumption [76]. In this case it is necessary to use a system of PDEs such as Equations 4.12-4.17 to include the spatial context in the model. I modify this set of Equations to include the gene a and the mutual inhibition of a and q, by first adding an Equation for $\partial_t a(x,t)$:

$$\partial_t a(x,t) = \frac{\alpha_a q(x,t)^{n_q}}{{K_q}^{n_q} + q(x,t)^{n_q}} - \gamma_a a(x,t) \tag{4.31}$$

where α_a is the maximum expression rate of a, n_q and K_q are the Hill coefficient and half maximal concentration for the repression of a by q, and γ_a is the degradation/dilution rate of a. a is modeled as unable to diffuse through the biofilm, it is intracellular only.

Additionally Equations 4.12 and 4.13 are modified to take into account the repression of electron export rate by a:

$$\partial_t s(x,t) = D_s \partial_x^2 s(x,t) - \frac{\alpha_q (Q_0 - q(x,t)) s(x,t) a(x,t)^{n_a}}{(K_s + s(x,t))({K_a}^{n_a} + a(x,t)^{n_a})} \tag{4.32}$$

$$\partial_t q(x,t) = k Q_0 \Delta x^2 \partial_x^2 q(x,t) + \frac{\alpha_q (Q_0 - q(x,t)) s(x,t) a(x,t)^{n_a}}{(K_s + s(x,t))({K_a}^{n_a} + a(x,t)^{n_a})} \tag{4.33}$$

The extent to which gradients form depend on the parameters D_s and k. D_s is the diffusion coefficient of the substrate and here I use a value obtained from the literature for acetate diffusion in a biofilm [45]. There are no explicit measurements of k in the literature, but instead I perform the analysis across a range of k spanning several orders of magnitude.

Figure 4.9A show how bistability can be lost as k decreases and biofilm gradients become more significant. For example the system exhibits bistability at $k = 10$, but an identical system where $k = 1$ displays only monostable behaviour. With all else being equal, bistability will emerge for certain values of k, as shown in Figure 4.9B, where for this

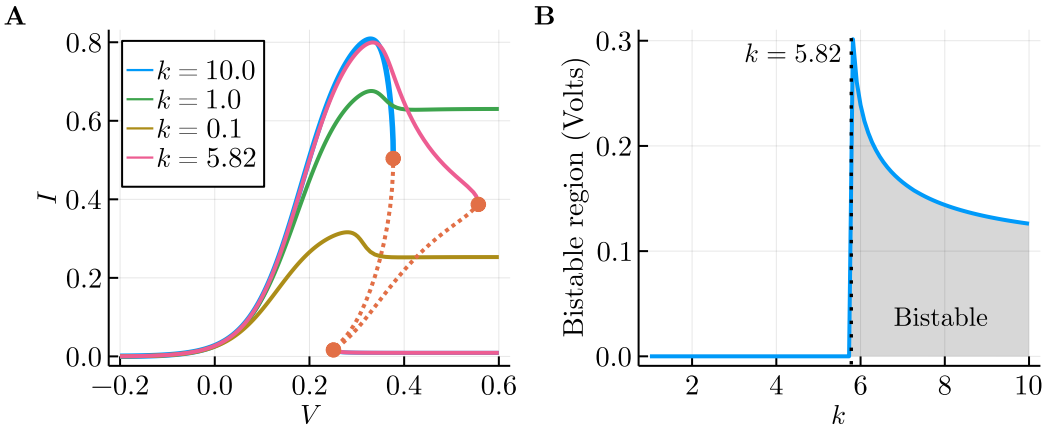


Figure 4.9: Bifurcation analysis for the full model, including spatial context of the biofilm, using numerical analysis of Equations 4.31, 4.32 and 4.33. **A** shows how the parameter k, which is the rate of q transport through the biofilm, can affect the emergence of bistability. The bistability that exists at high rates of transport disappears at lower rates, which correspond to higher magnitude gradients in the biofilm. **B** shows the relationship between k and the size of the bistable region. Bistability emerges at $k \approx 5.82$ and corresponds to the largest bistable region, as shown in **A** (pink branch). The size of the region drops off as k increases further, but levels off as $k > 10$.

example it is only values of k above 5.82 that exhibit bistability.[a]

The model presented here can be used to predict the affect of changes to the synthetic network making up the genetic part of the switch. This is useful for discovering changes to the parameters of the gene network which can recover bistable dynamics despite slow q transport. As an example Figure 4.10A performs a bifurcation analysis on the system with $k = 1$, which previously displayed only monostability. However, there are values of K_a ($K_q = 0.8$ and $K_q = 0.49$ in Figure 4.10A), which is a parameter related to the repression of a by q, for which bistability can be recovered. The range of values for K_a which produce bistable dynamics can be found numerically, and is shown in Figure 4.10B for this case to be between $K_q \approx 0.49$ and $K_q \approx 3.6$.

Parameter values for the wild type model

Physiologically reasonable parameter values can be extracted from the literature and are summarised in the following table, along with justification for their values. I divide these into parameters that are part of the synthetic gene network (Table 4.2 and so are potentially engineerable, and parameters that are part of the context (Table 4.1 and cannot be engineered with ease.

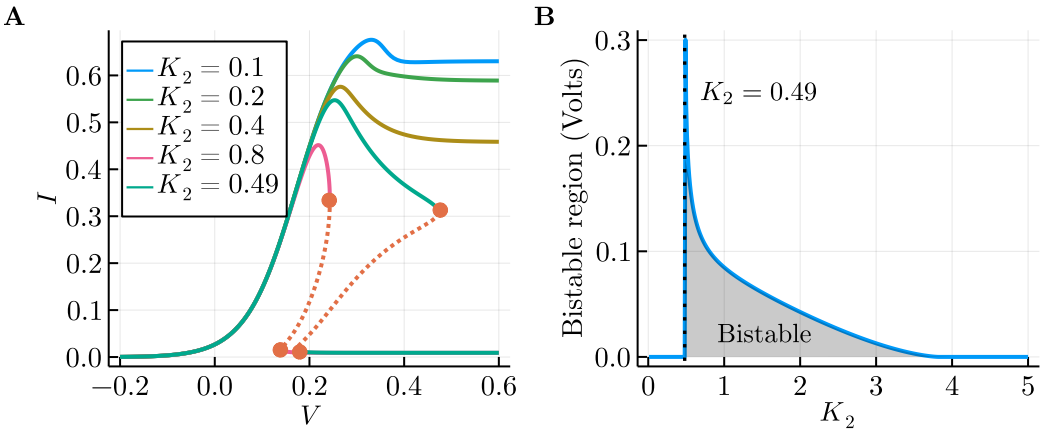


Figure 4.10: Bifurcation analysis for the full model with $k = 1$. **A** shows how bistability can be recovered by modifying the parameter K_a, which is a parameter related to the repression of a by q and can be engineered. Increasing K_a from 0.1 (which produces bistability in the nonspatial model but in the spatial one for $k = 1$) to 0.8 recovers the bistable dynamics. **B** shows the relationship between K_a and the size of the bistable region. Bistability emerges at $K_q \approx 0.49$ where the size is at a maximum. This is shown in **A** as the green branch. The size of bistable region decreases until it disappears at $K_q \approx 3.6$, beyond which the system in once again monostable.

Parameter	Value	Unit	Justification
j_0	3×10^{-2}	s^{-1}	Taken from [45]
Q_0	10	$mol\ m^{-3}$	Calculated from [73]
α_q	1.3×10^{-1}	s^{-1}	Chosen to produce $1\frac{A}{m^2}$ at $V = 0.6$ [73]
S_0	1	$mol\ m^{-3}$	Default value taken from [45]
K_s	0.1	$mol\ m^{-3}$	Taken from [45]
D_s	5.5×10^{-10}	$m^2 s^{-1}$	Taken from [45]
V	-	$volts$	Reasonable values between -1 and 1 volts
k	10	$m^3\ mol^{-1}\ s^{-1}$	Unknown variable
Δx	1×10^{-6}	m	Dependent on discretisation

Table 4.1: A table of parameters related to the biofilm, electrode or substrate. With the exception of electrode potential V these parameters are not easy to engineer.

Parameter	Typical Value	Unit	Justification
α_a	0.01	s^{-1}	Variable
γ_a	1	s^{-1}	Variable
K_q	1	$mol\ m^{-3}$	Variable
K_a	0.01	$mol\ m^{-3}$	Variable
n_q	2	-	The lowest necessary for bistability
n_a	2	-	The lowest necessary for bistability

Table 4.2: A table of parameters related to synthetic gene network of the electrogenetic toggle switch, in particular the level of gene expression of a and the parameters of the two mutual inhibitions. The parameters are to a greater extent amenable to engineering with synthetic biology tools.

4.3 Discussion

The analysis presented here suggest that an electrogenetic toggle switch could be engineered using a synthetic gene network. The electrogenetic toggle switch could be hosted by exoelectrogenetic bacteria, growing in a electrode-attached biofilm, and can produce bistable dynamics for parameters of the biofilm and electrode that are physiologically relevant. It is a hybrid device which takes an electronic input (electrode potential V), processes the input using a biological and gene network, and produces an electronic output (current density I). The input is in this case used to switch between high and low output states. This takes advantage of (engineered) bistability and hysteresis of the electroactive biofilm-electrode system.

Spatial homogeneity is a useful assumption for reducing the complexity of models that are used for the design of synthetic gene networks [79]. This is especially useful when the environment is well-mixed. However, a key prediction of the model is that the biofilm environment is heterogeneous, a phenomena also observed experimentally [76]. Since the electrogenetic toggle switch interacts with the biofilm in a fundamental way, this heterogeneity is important in the design of the switch since it changes the set of parameters for which bistability is present. I show that there are parameter sets for which bistability is predicted to emerge in a homogeneous biofilm (for example in Figure 4.7B) but for which bistability is lost when biofilm heterogeneity is taken into account (for example in Figure 4.9B. Despite this additional complexity, the model can also be used to predict the changes to the electrogenetic network that recover bistability in a number of the cases I present, and this just by engineering the synthetic gene network.

The results highlight some of the challenges involved in redesign of classic synthetic biolog-

ical networks as electrogenetic systems, particularly in the context of the heterogeneous environments that are typical in electroactive biofilms. The work presented here also provides a basis for future work in addressing these challenges and toward the rational design of hybrid synthetic biological device that can be measured and controlled both genetically and electronically.

The electrogenetic toggle switch in context

The exoelectrogens are connected to the electronic input and output of the switch by the redox state of the biofilm. This is a desirable property of the switch because it constitutes a closed-loop system that can monitored and controlled by electronics via the electrode. This offers opportunities for 'online' control strategies and real-time collection similar to the capabilities offered by optogenetics [14].

The rate of transport of electrons (and more generally, charge) through the biofilm is a contextual parameter that can affect the dynamics of the switch, but which cannot be eliminated from consideration since the bacteria depend on the biofilm for communication with the electrode. Instead it is necessary to engineer a synthetic gene network which can tolerate slower electrons transport rates and the gradients that form as a result. Fortunately the model I present here predicts that this is possible in a least some cases.

Possible implementations of the electrogenetic switch

The model I present here is abstract, only describing at a high level, and with lumped parameters, the repression interactions between electrons in the biofilm q and a hypothetical gene a which limits the rate at which these electrons are exported. Implementations of the electrogenetic toggle switch will have to identify suitable genes and mechanisms to create the required repressive interactions.

In *Geobacter* it has been shown that control of expression of *gltA* can be used to modulate their current output [89]. *gltA* encodes for a citrate synthase, which is a key enzyme for the metabolism of acetate in the TCA cycle. In the study expression of *gltA* was activated using IPTG, but this synthetic gene network could be also be modified or integrated into larger networks to be repressed. *gltA* seems like a good candidate for a in an implementation of the electrogenetic toggle switch.

It is also required that q can be sensed by the bacteria and that this redox signal can be transduced to repress a. Redox sensitive promoters such as *pSoxR* can be used to sense the redox state of the intracellular environment, but also rely on exogenous redox mediators to communicate with the electrode [87]. Perhaps more promising is the introduction of two component systems such as *Arc* systems into *Geobacter*. *Arc* systems have been shown to link the redox state of the periplasmic quinone pool to gene expression levels in *E. coli* [31]. This periplasmic quinone pool is linked with the redox state of the biofilm via outer membrane cytochromes such as *OmcB* and *OmcZ* in *Geobacter*, making such systems potentially useful for linking q to synthetic gene networks.

Future work

The model I present here should be expanded to include more details of the mechanisms implementing the interactions between q and a. This will help inform the choice of implementation, as described above. Once an implementation is built it can be used to generate experimental data for fitting the model and tuning it's parameters.

As discussed in Section 4.1, positive feedback for q is a requirement for bistability to emerge, and the original genetic toggle switch implemented this positive feedback using mutual inhibition, that is, using two repression interactions in a loop so that the overall effect is a positive feedback. This was necessary in order to be able to easily switch 'on' *and* 'off', since the inducers could only deactivate their target repressor molecules. This is not the case for the electrogenetic switch presented here, since the electrode potential can be used as an induction signal to both decrease (deactivate) or increase (activate) q, allowing the electrode to be used as an inducer in both 'directions'. There may therefore be an opportunity to implement the electrogenetic switch using a different network topology involving activations instead of repressions. Here the model could be useful to identify any advantages in terms of robustness for either topology, can perhaps open up more options for simpler physical implementations of the switch.

Additional spatial dimensions might also be included in future modeling work, the current model assumes that the only positional information that matters in the biofilm is depth, the distance from the electrode. Adding spatial dimensions and comparing the results should confirm this.

The biofilm is important in providing context for the electrogenetic toggle switch, but

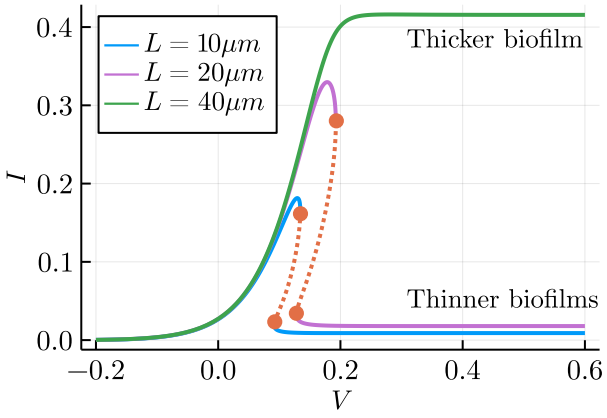


Figure 4.11: Bifurcation diagrams of biofilms of different depths, L. Numerical analysis of identical synthetic gene networks in the context of different biofilm depths reveals that bistable dynamics that are present in thinner biofilms (blue and purple lines) can be lost in thicker biofilms (green line)

the biofilm itself is a dynamical system that can grow and shrink over time. Figure 4.11 shows how biofilms can lose bistability as they become thicker, perhaps because they are more likely to have greater heterogeneity than thinner biofilms. In this simulation, biofilms of depth 10 and $20\mu m$ exhibit bistability whereas an indentical switch in a $40\mu m$ biofilm does not. Complicating matters is that the growth rate of the biofilm should be dependent on the activity of the bacteria, so that higher current densities lead to more growth. Therefore it is possible that the depth of the biofilm at a given time depends on the state history of the switch. Considering this complexity, it would be valuable to include growth in the model in order to aid the design of an electrogenetic toggle switch that can function correctly in a dynamic biofilm.

Chapter 5

Repressilator in the spatial context of the electroactive biofilm

This chapter adapts the model I developed in the previous chapter to examine a design of a electrogenetic oscillator. I adapt the synthetic biological network from the repressilator [23], and perform numerical analysis of the model.

5.1 Introduction

The previous Chapter analyses an electrogenetic gene network intended to produce bistable dynamics. In Section 2.3 I introduced oscillatory dynamics as another class of fundamental behaviours. In this Chapter I adapt the design of the well studied repressilator, a synthetic gene network oscillator, in a similar manner as the genetic toggle switch.

This electrogenetic repressilator should:

- Be hosted in an electroactive biofilm by bacteria in a MES.
- Include a synthetic gene network.
- Produce an electronic output in the form of electrical current at the electrode.
- Retain the qualitative dynamics of the repressilator (oscillations).

A key issue for oscillating gene networks is phase drift [68]. At the population level oscillatory dynamics are the result of contributions from individual synthetic gene networks hosted in individual cells. These gene networks are subject to both extrinsic and intrinsic noise that can cause the oscillations to lose coherence, such that the population as a whole no longer exhibits oscillations, even if their initial conditions are identical.

One well known solution that has received attention is the global coupling of the individual oscillators. Such a coupling was previously implemented for a population of genetic oscillators by using quorum sensing for intercellular communication, with the result of sustained, synchronised oscillations in a growing population of *Escherichia coli* [19].

By integrating the repressilator network with the biofilm, it is possible that biofilm electrons q might also constitute an intercellular communication network that amounts to a global coupling of individual oscillators. If so the synchronisation of oscillations might come 'for free' with the electrogenetic repressilator as a result of linking it to the redox context of the biofilm.

Here I make a preliminary analysis of the feasibility of this global coupling, as well as perform bifurcation analyses of the model.

5.2 Results

Figure 5.3A shows a schematic of the network of the proposed oscillator. There are two genes, p_1 and p_2, and the biofilm electron concentration q, arranged in a negative feedback loop. q is sensed by bacteria and represses the expression of p_1, p_1 represses the expression of p_2, and p_2 inhibits the rate of electron export to effectively repress the production of q, completing the loop.

Model formulation

Here I adapt the BioModels SBML model for the repressilator for modeling the synthetic gene network [61]. This includes state variables m_1 and m_2 for the mRNA encoding for p_1 and p_2. The PDEs for the gene network are as follows:

$$\partial_t m_1(x,t) = \frac{\alpha_m q(x,t)^{n_q}}{{K_q}^{n_q} + q(x,t)^{n_q}} - \gamma_m m_1(x,t) \tag{5.1}$$

$$\partial_t m_2(x,t) = \frac{\alpha_m p_1(x,t)^{n_p}}{{K_p}^{n_p} + p_1(x,t)^{n_p}} - \gamma_m m_2(x,t) \tag{5.2}$$

$$\partial_t p_1(x,t) = \alpha_p m_1(x,t) - \gamma_p p_1(x,t) \tag{5.3}$$

$$\partial_t p_2(x,t) = \alpha_p m_2(x,t) - \gamma_p p_2(x,t) \tag{5.4}$$

where α_m is the maximal transcription rate, γ_m is the mRNA degradation/dilution rate, α_p is the translation rate of the mRNA, γ_p is the degradation/dilution rate of protein, K_p is the half maximal repression concentration of transcription by the protein, and n_p is the Hill coefficient for the repression.

As with the electrogenetic toggle switch the mRNA and protein are intracellular species that do not diffuse in the biofilm.

The model is completed by the PDEs for q and s, and their boundary conditions, which are the same as those first presented in Chapter 4.

$$\partial_t s(x,t) = D_s \partial_x^2 s(x,t) - \frac{\alpha_q p_2(x,t)^{n_p}}{{K_p}^{n_p} + p_2(x,t)^{n_p}} \cdot \frac{(Q_0 - q(x,t))s(x,t)}{K_s + s(x,t)} \tag{5.5}$$

$$\partial_t q(x,t) = kQ_0 \Delta x \partial_x^2 q(x,t) + \frac{\alpha_q p_2(x,t)^{n_p}}{{K_p}^{n_p} + p_2(x,t)^{n_p}} \cdot \frac{(Q_0 - q(x,t))s(x,t)}{K_s + s(x,t)} \tag{5.6}$$

$$\partial_x s(0,t) = 0 \tag{5.7}$$

$$\partial_x q(0,t) = j_0 \left(q(0,t)e^{AV} + q(0,t)e^{-AV} - Q_0 e^{-AV} \right) \tag{5.8}$$

$$\partial_x q(L,t) = 0 \tag{5.9}$$

$$s(L,t) = S_0 \tag{5.10}$$

where the only difference from Equations 4.32 and 4.33 is that production of q is now repressed by p_2 instead of a. Parameters in common with the toggle switch and biofilm model have the same interpretation. Additional parameters are listed in Table 5.1 along with their typical values.

Parameter	Typical Value	Unit	Justification
α_m	3.011×10^{-5}	$mol\ m^{-3}\ s^{-1}$	Calculated from [61]
γ_m	$\frac{ln(2)}{120}$	s^{-1}	Taken from [61]
α_p	0.1155	s^{-1}	Calculated from [61]
γ_p	$\frac{ln(2)}{60}$	s^{-1}	Taken from [61]
K_p	5×10^{-3}	$mol\ m^{-3}$	Variable
n_q	3	-	Variable
n_p	3	-	Variable
K_q	0.2	$mol\ m^{-3}$	Variable

Table 5.1: A table of parameters related to synthetic gene network of the electrogenetic repressilator. The model shares some parameter with the previous model of the electrogenetic toggle switch, and these parameters are the same as listed in Tables 4.2 and 4.1 if not listed here.

Numerical timecourse simulation with these parameter values reveals oscillatory dynamics with a period of around 15 minutes as shown in Figure 5.1A. Oscillations are present in current output I as a result of oscillations in the concentration of electrons q at the electrode interface. In Figure 5.1B a heatmap of $q(x,t)$ is plotted where the y-axis represents position in the biofilm and the x-axis shows the evolution over time. The oscillations in q are present at all depths in the biofilm, with a slight gradient between the electode and bulk interfaces. This result shows that it is possible for oscillatory dynamics to emerge for this electrogenetic oscillator.

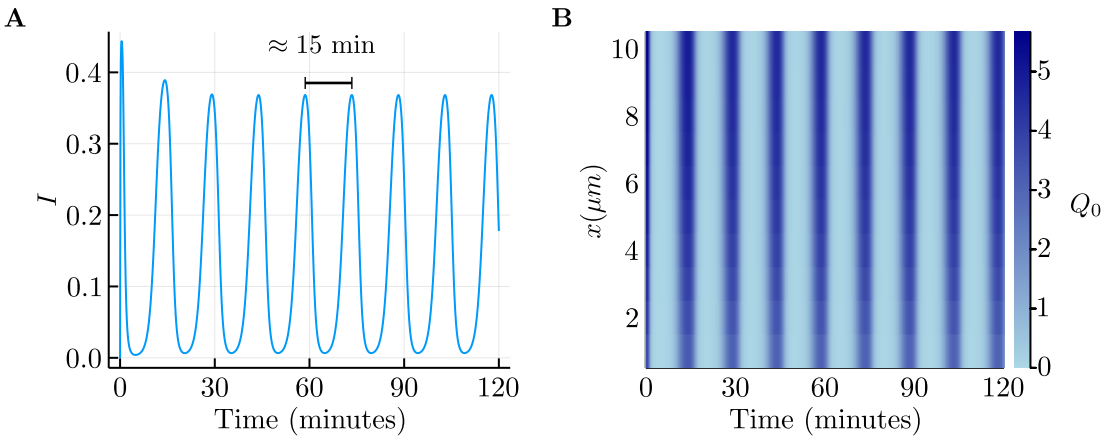


Figure 5.1: A timecourse simulation of the electrogenetic repressilator, with parameter values as listed in Table 5.1. In **A** the current density I is plotted over time and exhibits oscillations with a period of around 15 minutes. In **B** the electron (charge) concentration (z-axis) at each x position in the biofilm (y-axis) is shown against time (x-axis). There is a small gradient in the biofilm, and oscillation amplitude is seen to be greater at higher x, deeper in the biofilm.

Bifurcation analysis to identify oscillatory regions

The bifurcation analysis with parameter V is plotted in Figure 5.2A, shows a single Hopf bifurcation at $V = 0.11$ marking the emergence of oscillations. In contrast to the results from the electrogenetic toggle switch, the region for which the desirable dynamics are present is large. This analysis was performed for $k = 1$, which is a relatively slow rate of electron transport which produced significant gradients in the biofilm for the electrogenetic toggle switch and for which the synthetic gene network had to be adjusted to produce bistability.

This result might suggest that the electrogenetic oscillator might be in some sense easier to engineer, and more robust to electron transport rate, than the toggle switch. The analysis in Figure 5.2B supports this conclusion. Oscillations can emerge over at least 4 orders of magnitude for k, marked by the shaded region, with $1 - 3$ hopf bifurcations, and the system oscillates over wide ranges of V (though always for $V > 0.1$).

Oscillations can be switched on and off using the electrode potential to move the system between the oscillatory and nonoscillatory regions shown in Figure 5.2B. A timecourse simulation of this switching is shown in Figures 5.3B, C, and D. From an initial monostable state with $V = 0.07$, oscillations can be induced by stepping V to 0.2, and switched off again by stepping V back down to 0.07. The effect on $q(x, t)$ is as expected, shown in

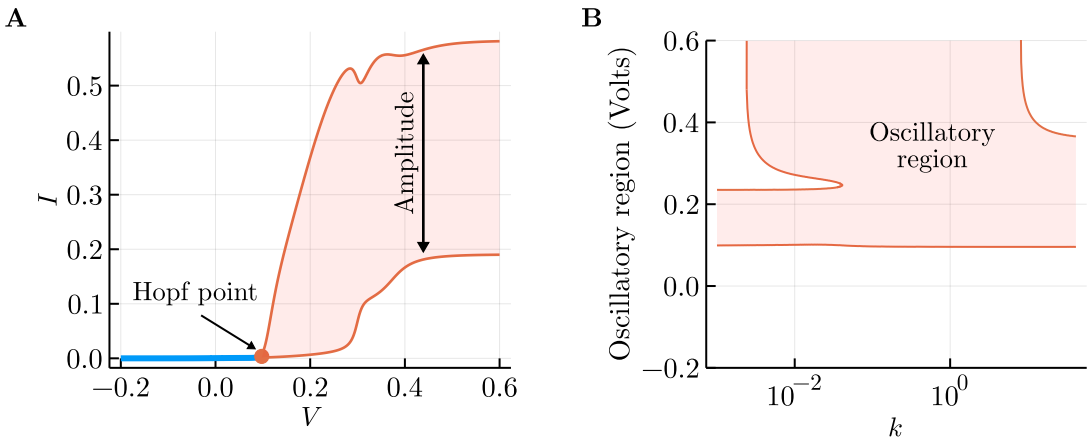


Figure 5.2: Bifurcation analysis of the Equations 5.1-5.10. In **A** the bifurcation parameter V is plotted against I, for the case where $k = 1$. A hopf bifurcation point marks the emergence of oscillatory dynamics at $V \approx 0.1$. The shaded orange region shows the minimum and maximum of the oscillations. **B** plots the existence and positions of hopf bifurcations as k is varied. At low k two hopf points (orange lines) bound a narrow oscillatory region (shaded). As k increases a third hopf point emerges and finally collides with the second to produce the large region of oscillations with a single hopf point that characterises the system dynamics until the reemergence of the second at $k \approx 6$.

Figure 5.3B.

Comparison with nonspatial model

In the limit of large k the electrons can be transported infinitely fast and the q is homogeneous. Here I use the ODEs 5.11-5.16 to model a homogeneous biofilm without a spatial dimension, and compare bifurcation analysis of this model with that presented in the previous section.

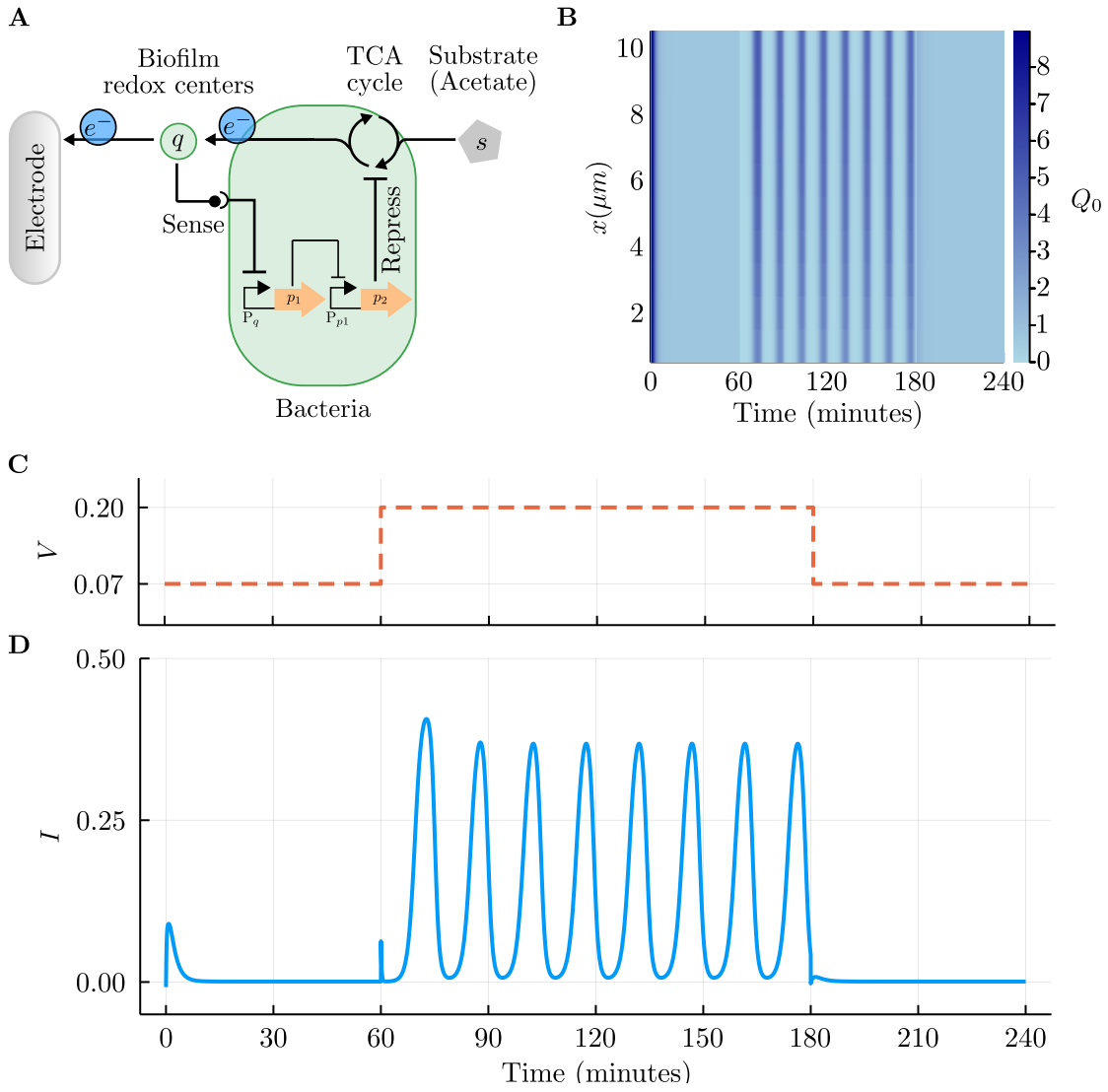


Figure 5.3: In **A** a diagram showing the interactions of the electrogenetic repressilator. q, p_1 and p_2 form a negative feedback loop just as in the orginal repressilator design, and under certain conditions their concentrations will oscillate, as will the consumption of substrate through the TCA cycle, which is repressed by p_2. **B**, **C** and **D** show the results of a timecourse simulation of Equations 5.5-5.10, with parameters from Table 5.1 and $k = 1$. Oscillatory dynamics can be switch on and off using V as shown in **C**, which produces the observed current density I shown in **D**. Oscillations in q correspond to those in I, and can similarly be switched on or off as shown in **B**.

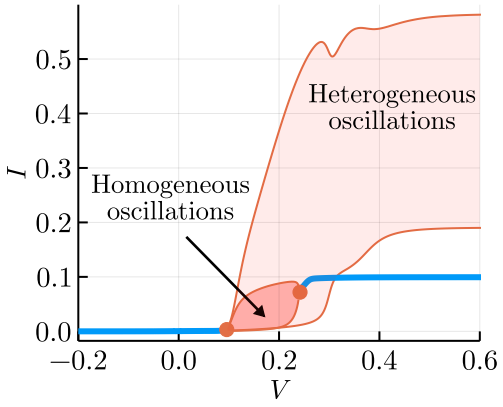


Figure 5.4: A bifurcation analysis comparing the homogeneous and heterogeneous model for bifurcation parameter V. The homogeneous model of Equations 5.11-5.16 does not include space and has two hopf bifurcation points bounding a relatively small region of bistability surrounded by monostable dynamics. The heterogeneous model of Equations 5.1-5.10 produce a single hopf point and much larger regions of oscillatory dynamics, in this example the case of $k = 1$ is shown.

$$\partial_t m_1 = \frac{\alpha_m q^{n_q}}{K_q{}^{n_q} + q^{n_q}} - \gamma_m m_1 \tag{5.11}$$

$$\partial_t m_2 = \frac{\alpha_m p_1^{n_p}}{K_p{}^{n_p} + p_1^{n_p}} - \gamma_m m_2 \tag{5.12}$$

$$\partial_t p_1 = \alpha_p m_1 - \gamma_p p_1 \tag{5.13}$$

$$\partial_t p_2 = \alpha_p m_2 - \gamma_p p_2 \tag{5.14}$$

$$\partial_t q = \frac{\alpha_q p_2^{n_p}}{K_p{}^{n_p} + p_2^{n_p}} \cdot \frac{(Q_0 - q)s}{K_s + s} - j_0 \left(q e^{AV} + q e^{-AV} - Q_0 e^{-AV} \right) \tag{5.15}$$

$$\partial_t s = D_s (S_0 - s) \frac{\alpha_q p_2^{n_p}}{K_p{}^{n_p} + p_2^{n_p}} \cdot \frac{(Q_0 - q)s}{K_s + s} \tag{5.16}$$

For the same parameter values and compared to the heterogeneous model, the region of oscillatory dynamics in this nonspatial model is much smaller, lying between $V = 0.11$ and $V = 0.23$. The addition of space into the model seems to allow the system to oscillate over a much greater range of V. Again this suggests that spatial context increases the robustness of the oscillator, whereas in Chapter 4 it seemed that robustness was degraded by this contextual effect.

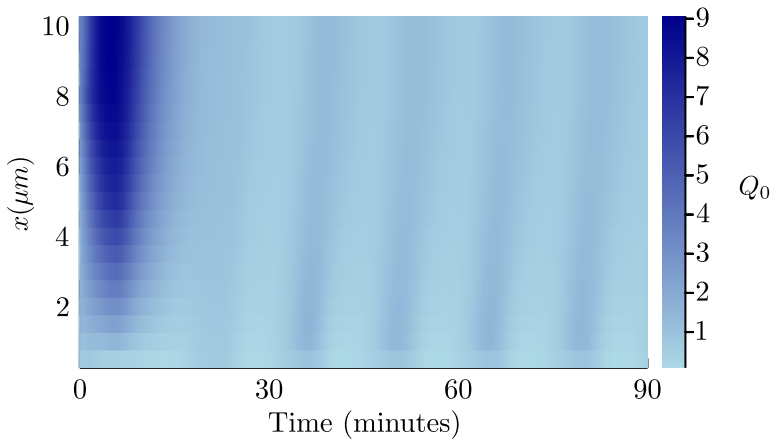


Figure 5.5: A heatmap of electron concentration q throughout the biofilm over time. In this simulation each region of the (discretised) biofilm has initial conditions which place the electrogenetic repressilator out of phase with one another. It is seen that after an initial transient period (until around 30 seconds), the coupling of these individual oscillators by the biofilm pulls them into phase with one another so that consistent oscillations begin to emerge.

Global coupling to prevent phase drift

I return to the heterogeneous model of Equations 5.5-5.10 to investigate the impact of space in the biofilm on phase drift in the electrogenetic oscillator. Figure 5.5 shows a heatmap of $q(x,t)$ in the biofilm. For this simulation the electron transport rate $k = 1$ and each discretised section of the biofilm is started in initial conditions that correspond to the oscillators being in randomly selected phases. That is, it is a simulation of individual, but identical, oscillators started at different times, and coupled together with the biofilm.

The results show that up to 30 minutes q increases substantially, particularly deeper in the biofilm. After this the biofilm exhibits oscillations with period of around 15 minutes, which could be expected as this matches the period seen in Figure 5.1. It seems as though the individual regions of the biofilm were able to synchronise their oscillations, perhaps through the global coupling facilitated by the electron transport and substrate diffusion processes.

5.3 Discussion

Numerical simulations suggest that the proposed electrogenetic oscillator could be built using a synthetic genetic network. The electrogenetic oscillator is hosted in an electroac-

tive biofilm by exoelectrogenic bacteria and produces oscillations that can be detected electronically as electrical current flowing through an electrode.

Bifurcation analysis of the model shows that this oscillator is more robust than the electrogenetic toggle switch to changes in the electron transport rate (k) in the biofilm (Figure 5.2). This is encouraging as this rate is a contextual effect that might differ between individual biofilms under different environmental conditions [96, 97, 66].

Another contrast with the electrogenetic toggle switch is the effect of spatial heterogeneity on the emergence of the desired dynamics. Interestingly this electrogenetic oscillator has a smaller region of oscillatory dynamics when spatial dimensions are neglected, and seemingly spatial distribution of the gene network encourages the emergence of oscillations. This might be a case then, where linking synthetic gene networks with their environmental context helps achieve the desired function.

Softer gradients in the electrogenetic oscillator

Even at relatively low electron transport rates ($k = 1$) the model predicts that the gradient of q in the biofilm will have a small magnitude, as shown in Figure 5.1B for example. An explanation for this might be that in the electrogenetic oscillator q has negative feedback (implemented using a loop of three repressive interactions), and that negative feedback is known to stabilise expression[28]. This might have important implications for dealing with spatial gradients in context for other synthetic gene networks and homogenising the electroactive biofilm.

Global coupling synchronises oscillations

In Figure 5.5 it is shown that a biofilm of oscillators starting in random phases tends toward synchronised oscillations. This reflects other results that have shown that intercellular coupling can solve the phase drift problem, for example by using diffusing quorum sensing molecules to couple synthetic gene oscillators [74]. It seems that the electroactive biofilm can also solve this problem, despite the asymmetry in the spatial domain and the overall tendency of electrons to flow toward the electrode.

Future work

As with the electrogenetic toggle switch, the model should be expanded to include the details of the interactions between q and p_1 and also between p_2 and the electron export rate that produces q. The goal is to engineer the synthetic gene network so that it can be tested and the model improved, and the model can help to inform the choice of implementation of these interactions.

It would also be interesting to explore the dependence of the period and amplitude of oscillations on the parameters of the model. This includes the dependence on V, where it would be interesting to see if identical gene networks can be induced to give quantitatively different oscillations in response to electrode potential. This has been a key aim for the engineering of synthetic biological oscillators for some time, and control of genetic oscillator's period or amplitude has previously been achieved by altering inducer levels [80] or by redesign of the gene network [86].

Although the numerical simulation in Figure 5.5 suggests that the dephasing problem could be solved by global coupling mediated by the biofilm, the results might not be convincing since I have used a deterministic model of the system. The root of the dephasing problem appears to be the intrinsic and extrinsic noise acting on the gene network, and the deterministic model fails to capture the effects of this noise. Future work might therefore require nondeterministic models using stochastic differential equations or Gillespie simulations, before concluding that the biofilm can help produce sustained coherent oscillations.

Chapter 6

Conclusion

This chapter summarises the work presented and offers some conclusions and final remarks.

In this thesis I have used mathematical modeling to investigate the impact of host, genetic and spatial context on the performance of three classes of synthetic gene network that are fundamental and widely used in biocomputing.

In Chapter 3 I used previously collected experimental data on a library of genetic logic gates to quantify the differences in the performance of the library due to changes in to contextual parameters, host organism and plasmid vector. Building circuits of genetic logic gates is a popular approach implemented the combinatorial logic model of computation for the purposes of biocomputation, but increasing the complexity of circuits relies on the modularity a reusability of the individual gates, both of which are impeded to some extent by contextual effects. My analysis reveals significant differences in gate performance due to context that seem difficult to predict *a priori*, since the performance of the gates changed not only in quantitative but also a qualitative sense.

I use metrics such as number of valid NOT gates (with given thresholds), number of compatible NOT gates, and an upper bound on maximum circuit depth to measure the usefulness of the library in each context, and in cross-context libraries. Although these metrics do not always agree which library is the most useful, in all cases usefulness is found to increase as more contexts are added to the library, suggesting the incorporation of context as a design parameter can be advantageous for combinatorial design of increasingly complex genetic logic circuits.

In Chapter 4 I propose a design for a synthetic biological network which is bistable and is intended to operate in an electroactive biofilm. Bistability is a fundamental type of dynamics in biocomputing and computing in general, in particular for providing a 'memory'. However, a biofilm provides a heterogeneous environment, presenting the same synthetic gene network with different environments at the same time, depending on their spatial position in the biofilm. The model I describe here predicts that this spatial and environmental context can degrade the performance of this electrogenetic toggle switch to the point that it loses bistable dynamics.

Fortunately the model also predicts that bistable dynamics can be restored by altering the genetic components that make up the synthetic gene network. In the case of the electrogenetic toggle switch the effects of context on the gene network are unavoidable, but the model can predict how they can be accounted for in order to preserve function.

Chapter 5 follows on from the electrogenetic toggle switch and I propose another redesign of a classical synthetic gene network, the repressilator, for integration into an electroactive

biofilm. This electrogenetic repressilator has oscillatory dynamics, another fundamental class of dynamics in computing which can be used as a clock to synchronise the transitions other computing operations, for example in sequential logic circuits*.

Again the biofilm provides a spatial context for the gene network in the oscillator. However in this case the model predictsthat this spatial context will improve robustness of system and expand the set of parameters for which the desired oscillatory dynamics can be observed. The reason for this is not clear, though it may have to do with the coupling of the gene networks that is mediated by the biofilm. The model also predicts that the same coupling will improve the coherence of oscillations in the system, and pull individual oscillators into phase with one other, even if they have initial conditions that are out of phase.

Final remark

Context is often framed as an unsolved problem in synthetic biology that prevents the scaling of biocomputational complexity, and the establishment of engineered cells as a dominant computing substrate for applications beyond simple biosensors or toy examples of logic circuits.

But it is no small irony that the very thing frustrating the emergence of this dominance also makes the proposition attractive in the first place. Context arises from living computational systems interacting with their environment, which is precisely what grants them potential in computational domains within which conventional computers are just not comfortable.

Synthetic biology should not hope to eliminate the effects of context but rather think about them as signposts on the way to computational domains — ones in which living systems will outperform nonliving ones. Growth, noise, evolution and spatial distribution are all things that conventional computers struggle with, but that biocomputers do not. The way to deal with context in synthetic biology is to consider it a first-class citizen that is included in the very beginning of the design stage of a synthetic biological system.

*Sequential logic circuits are similar to combinatorial logic circuits but can have feedback loops, meaning the output of computations can depend on previous as well as current inputs

Bibliography

[1] Martyn Amos and Angel Goñi-Moreno. Cellular computing and synthetic biology. In *Computational Matter*, pages 93–110. Springer, 2018.

[2] Ernesto Andrianantoandro, Subhayu Basu, David K Karig, and Ron Weiss. Synthetic biology: new engineering rules for an emerging discipline. *Molecular Systems Biology*, 2:2006.0028, May 2006.

[3] Stephanie K. Aoki, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature*, 570(7762):533–537, June 2019.

[4] Sarah E. Barchinger, Sahand Pirbadian, Christine Sambles, Carol S. Baker, Kar Man Leung, Nigel J. Burroughs, Mohamed Y. El-Naggar, and John H. Golbeck. Regulation of Gene Expression in Shewanella Oneidensis Mr-1 During Electron Acceptor Limitation and Bacterial Nanowire Formation. *Applied and Environmental Microbiology*, 82(17):5428–5443, 2016.

[5] Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, June 2000.

[6] Narendranath Bhokisham, Eric VanArsdale, Kristina T. Stephens, Pricila Hauk, Gregory F. Payne, and William E. Bentley. A redox-based electrogenetic CRISPR system to connect with and control biological information networks. *Nature Communications*, 11(1):2427, May 2020.

[7] Philipp Boeing, Miriam Leon, Darren N. Nesbeth, Anthony Finkelstein, and Chris P. Barnes. Towards an Aspect-Oriented Design and Modelling Framework for Synthetic Biology. *Processes*, 6(9):167, September 2018. Number: 9 Publisher: Multidisciplinary Digital Publishing Institute.

[8] Olivier Borkowski, Francesca Ceroni, Guy-Bart Stan, and Tom Ellis. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Current Opinion in Microbiology*, 33:123–130, October 2016.

[9] Dmitri Bratsun, Dmitri Volfson, Lev S. Tsimring, and Jeff Hasty. Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences*, 102(41):14593–14598, October 2005.

[10] Lukas Breitwieser, Ahmad Hesam, Jean de Montigny, Vasileios Vavourakis, Alexandros Iosif, Jack Jennings, Marcus Kaiser, Marco Manca, Alberto Di Meglio, Zaid Al-Ars, Fons Rademakers, Onur Mutlu, and Roman Bauer. BioDynaMo: a modular platform for high-performance agent-based simulation. *Bioinformatics*, 38(2):453–460, January 2022.

[11] Corentin Briat, Ankit Gupta, and Mustafa Khammash. Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. *Cell Systems*, 2(1):15–26, January 2016.

[12] Pablo Carbonell, Adrian J. Jervis, Christopher J. Robinson, Cunyu Yan, Mark Dunstan, Neil Swainston, Maria Vinaixa, Katherine A. Hollywood, Andrew Currin, Nicholas J. W. Rattray, Sandra Taylor, Reynard Spiess, Rehana Sung, Alan R. Williams, Donal Fellows, Natalie J. Stanford, Paul Mulherin, Rosalind Le Feuvre, Perdita Barran, Royston Goodacre, Nicholas J. Turner, Carole Goble, George Guoqiang Chen, Douglas B. Kell, Jason Micklefield, Rainer Breitling, Eriko Takano, Jean-Loup Faulon, and Nigel S. Scrutton. An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Communications Biology*, 1(1):1–10, June 2018.

[13] Stefano Cardinale and Adam Paul Arkin. Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, 7(7):856–866, July 2012.

[14] Remy Chait, Jakob Ruess, Tobias Bergmiller, Gašper Tkačik, and Călin C. Guet. Shaping bacterial population behavior through computer-interfaced control of individual cells. *Nature Communications*, 8(1):1535, November 2017.

[15] D. Chandran, W.B. Copeland, S.C. Sleight, and H.M. Sauro. Mathematical modeling and synthetic biology. *Drug discovery today. Disease models*, 5(4):299–309, 2008.

[16] Kiri Choi, Kyle J. Medley, and Herbert M. Sauro. Tellurium: An extensible python-based modeling environment for systems and synthetic biology - ScienceDirect, 2018.

[17] Victor Chubukov, Aindrila Mukhopadhyay, Christopher J. Petzold, Jay D. Keasling, and Héctor García Martín. Synthetic and systems biology for microbial production of commodity chemicals. *npj Systems Biology and Applications*, 2(1):1–11, April 2016.

[18] Neil Dalchau, Gregory Szép, Rosa Hernansaiz-Ballesteros, Chris P. Barnes, Luca Cardelli, Andrew Phillips, and Attila Csikász-Nagy. Computing with biological switches and clocks. *Natural Computing*, 17(4):761–779, December 2018.

[19] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326–330, January 2010.

[20] Víctor de Lorenzo, Natalio Krasnogor, and Markus Schmidt. For the sake of the Bioeconomy: define what a Synthetic Biology Chassis is! *New Biotechnology*, 60:44–51, January 2021.

[21] Víctor de Lorenzo, Kristala LJ Prather, Guo-Qiang Chen, Elizabeth O'Day, Conrad von Kameke, Diego A Oyarzún, Leticia Hosta-Rigau, Habiba Alsafar, Cong Cao, Weizhi Ji, Hideyuki Okano, Richard J Roberts, Mostafa Ronaghi, Karen Yeung, Feng Zhang, and Sang Yup Lee. The power of synthetic biology for bioproduction, remediation and pollution control. *EMBO Reports*, 19(4):e45658, April 2018.

[22] Domitilla Del Vecchio, Alexander J Ninfa, and Eduardo D Sontag. Modular cell biology: retroactivity and insulation. *Molecular Systems Biology*, 4(1):161, January 2008. Publisher: John Wiley & Sons, Ltd.

[23] Michael B Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335, 2000.

[24] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic Gene Expression in a Single Cell. *Science*, 297(5584):1183–1186, August 2002.

[25] Nir Friedman, Long Cai, and X. Sunney Xie. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Physical Review Letters*, 97(16):168302, October 2006.

[26] Luca Galbusera, Gwendoline Bellement-Theroue, Arantxa Urchueguia, Thomas Julou, and Erik van Nimwegen. Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria. *PloS One*, 15(10):e0240233, 2020.

[27] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339, January 2000.

[28] Timothy S. Gardner and James J. Collins. Neutralizing noise in gene networks. *Nature*, 405(6786):520–521, June 2000.

[29] Marcela N. Gatti and Rubén H. Milocco. A biofilm model of microbial fuel cells for engineering applications. *International Journal of Energy and Environmental Engineering*, 8(4):303–315, December 2017.

[30] P Gaudu and B Weiss. SoxR, a [2Fe-2S] transcription factor, is active only in its oxidized form. *Proceedings of the National Academy of Sciences of the United States of America*, 93(19):10094–10098, September 1996.

[31] D. Georgellis. Quinones As the Redox Signal for the Arc Two-Component System of Bacteria. *Science*, 292(5525):2314–2316, 2001.

[32] Rudolf Gesztelyi, Judit Zsuga, Adam Kemeny-Beke, Balazs Varga, Bela Juhasz, and Arpad Tosaki. The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences*, 66(4):427–438, July 2012.

[33] Daniel T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[34] Angel Goñi-Moreno and Martyn Amos. A reconfigurable NAND/NOR genetic logic gate. *BMC Systems Biology*, 6:126, September 2012.

[35] Ángel Goñi-Moreno, Ilaria Benedetti, Juhyun Kim, and Víctor de Lorenzo. Deconvolution of Gene Expression Noise into Spatial Dynamics of Transcription Factor–Promoter Interplay. *ACS Synthetic Biology*, 6(7):1359–1369, July 2017.

[36] Lewis Grozinger, Martyn Amos, Thomas E. Gorochowski, Pablo Carbonell, Diego A. Oyarzún, Ruud Stoof, Harold Fellermann, Paolo Zuliani, Huseyin Tas, and Angel Goñi-Moreno. Pathways to cellular supremacy in biocomputing. *Nature Communications*, 10(1):1–11, November 2019.

[37] Lewis Grozinger, Elizabeth Heidrich, and Ángel Goñi-Moreno. An electrogenetic toggle switch design, May 2022.

[38] Alexander Gräwe, Anna Dreyer, Tobias Vornholt, Ursela Barteczko, Luzia Buchholz, Gila Drews, Uyen Linh Ho, Marta Eva Jackowski, Melissa Kracht, Janina Lüders, Tore Bleckwehl, Lukas Rositzka, Matthias Ruwe, Manuel Wittchen, Petra Lutter, Kristian Müller, and Jörn Kalinowski. A paper-based, cell-free biosensor system for the detection of heavy metals and date rape drugs. *PLOS ONE*, 14(3):e0210940, March 2019. Publisher: Public Library of Science.

[39] Andras Gyorgy, José I. Jiménez, John Yazbek, Hsin-Ho Huang, Hattie Chung, Ron Weiss, and Domitilla Del Vecchio. Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophysical Journal*, 109(3):639–646, August 2015.

[40] Atsumi Hirose, Atsushi Kouzuma, and Kazuya Watanabe. Towards development of electrogenetics using electrochemically active bacteria. *Biotechnology Advances*, 37(6):107351, November 2019.

[41] Stefan Hoops, Sven Sahle, Ralph Gauges, Christine Lee, Jürgen Pahle, Natalia Simus, Mudita Singhal, Liang Xu, Pedro Mendes, and Ursula Kummer. COPASI—a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, December 2006.

[42] Kengo Inoue, Xinlei Qian, Leonor Morgado, Byoung-Chan Kim, Tünde Mester, Mounir Izallalen, Carlos A. Salgueiro, and Derek R. Lovley. Purification and Characterization of OmcZ, an Outer-Surface, Octaheme c-Type Cytochrome Essential for Optimal Current Production by Geobacter sulfurreducens. *Applied and Environmental Microbiology*, 76(12):3999–4007, June 2010.

[43] Jason R. Kelly, Adam J. Rubin, Joseph H. Davis, Caroline M. Ajo-Franklin, John Cumbers, Michael J. Czar, Kim de Mora, Aaron L. Glieberman, Dileep D. Monie, and Drew Endy. Measuring the activity of BioBrick promoters using an in vivo reference standard. *Journal of Biological Engineering*, 3(1):4, March 2009.

[44] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*, 139(7):1366–1375, December 2009.

[45] Benjamin Korth, Luis F. M. Rosa, Falk Harnisch, and Cristian Picioreanu. A framework for modeling electroactive microbial biofilms performing direct electron transfer. *Bioelectrochemistry*, 106:194–206, December 2015.

[46] Nicholas J. Kotloski and Jeffrey A. Gralnick. Flavin electron shuttles dominate extracellular electron transfer by Shewanella oneidensis. *mBio*, 4(1):e00553–12, January 2013.

[47] Ravinder Kumar, Lakhveer Singh, Zularisam A. Wahid, and Mohd Fadhil Md. Din. Exoelectrogens in microbial fuel cells toward bioelectricity generation: a review. *International Journal of Energy Research*, 39(8):1048–1067, 2015.

[48] Joshua M. Lawrence, Yutong Yin, Paolo Bombelli, Alberto Scarampi, Marko Storch, Laura T. Wey, Alicia Climent-Catala, PIXCELL IGEM TEAM, Geoff S. Baldwin, Danny O'Hare, Christopher J. Howe, Jenny Z. Zhang, Thomas E. Ouldridge, and Rodrigo Ledesma-Amaro. Synthetic biology and bioelectrochemical tools for electrogenetic system engineering. *Science Advances*, 8(18):eabm5091, May 2022.

[49] Caleb E. Levar, Colleen L. Hoffman, Aubrey J. Dunshee, Brandy M. Toner, and Daniel R. Bond. Redox potential as a master variable controlling pathways of metal reduction by *Geobacter* sulfurreducens. *The ISME Journal*, 11(3):741–752, March 2017.

[50] Diane Longo and Jeff Hasty. Dynamics of single-cell gene expression. *Molecular Systems Biology*, 2(1):64, January 2006.

[51] Derek R. Lovley and David J. F. Walker. Geobacter Protein Nanowires. *Frontiers in Microbiology*, 10, 2019.

[52] Enrico Marsili, Daniel B. Baron, Indraneel D. Shikhare, Dan Coursolle, Jeffrey A. Gralnick, and Daniel R. Bond. Shewanella secretes flavins that mediate extracellular electron transfer. *Proceedings of the National Academy of Sciences*, 105(10):3968–3973, March 2008.

[53] Esteban Martínez-García and Víctor de Lorenzo. The quest for the minimal bacterial genome. *Current Opinion in Biotechnology*, 42:216–224, December 2016.

[54] Esteban Martínez-García, Angel Goñi-Moreno, Bryan Bartley, James McLaughlin, Lucas Sánchez-Sampedro, Héctor Pascual del Pozo, Clara Prieto Hernández, Ada Serena Marletta, Davide De Lucrezia, Guzmán Sánchez-Fernández, Sofía Fraile, and Víctor de Lorenzo. SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic Acids Research*, 48(D1):D1164–D1170, January 2020.

[55] Lucia Marucci, David A. W. Barton, Irene Cantone, Maria Aurelia Ricci, Maria Pia Cosma, Stefania Santini, Diego di Bernardo, and Mario di Bernardo. How to Turn a Genetic Circuit into a Synthetic Tunable Oscillator, or a Bistable Switch. *PLOS ONE*, 4(12):e8083, December 2009.

[56] Andreas Milias-Argeitis, Marc Rullan, Stephanie K. Aoki, Peter Buchmann, and Mustafa Khammash. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nature Communications*, 7(1):12546, August 2016.

[57] Camillo Moschner, Charlie Wedd, and Somenath Bakshi. The context matrix: Navigating biological complexity for advanced biodesign. *Frontiers in Bioengineering and Biotechnology*, 10, 2022.

[58] Jonathan Naylor, Harold Fellermann, Yuchun Ding, Waleed K Mohammed, Nicholas S Jakubovics, Joy Mukherjee, Catherine A Biggs, Phillip C Wright, and Natalio Krasnogor. Simbiotics: A multiscale integrative platform for 3D modeling of bacterial populations. *ACS synthetic biology*, 6(7):1194–1210, 2017.

[59] 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. *Science*, 352(6281):aac7341–aac7341, 2016.

[60] Evangelos-Marios Nikolados, Andrea Y. Weiße, Francesca Ceroni, and Diego A. Oyarzún. Growth Defects and Loss-of-Function in Synthetic Gene Circuits. *ACS Synthetic Biology*, 8(6):1231–1240, June 2019.

[61] Nicolas Le Novère. Elowitz2000 - Repressilator \textbar BioModels.

[62] Irene Otero-Muras and Julio R. Banga. Synthetic Gene Circuit Analysis and Optimization. In Mario Andrea Marchisio, editor, *Computational Methods in Synthetic Biology*, Methods in Molecular Biology, pages 89–103. Springer US, New York, NY, 2021.

[63] Irene Otero-Muras, Julio R. Banga, and Antonio A. Alonso. Characterizing Multistationarity Regimes in Biochemical Reaction Networks. *PLOS ONE*, 7(7):e39194, July 2012.

[64] Jasmina Panovska-Griffiths, Karen M. Page, and James Briscoe. A gene regulatory motif that generates oscillatory or multiway switch outputs. *Journal of The Royal Society Interface*, 10(79):20120826, February 2013.

[65] Ruben Perez-Carrasco, Chris P. Barnes, Yolanda Schaerli, Mark Isalan, James Briscoe, and Karen M. Page. Combining a Toggle Switch and a Repressilator within the AC-DC Circuit Generates Distinct Dynamical Behaviors. *Cell Systems*, 6(4):521–530.e3, April 2018.

[66] Hung Phan, Matthew D. Yates, Nathan D. Kirchhofer, Guillermo C. Bazan, Leonard M. Tender, and Thuc-Quyen Nguyen. Biofilm as a redox conductor: a systematic study of the moisture and temperature dependence of its electrical properties. *Physical Chemistry Chemical Physics*, 18(27):17815–17821, July 2016.

[67] Cristian Picioreanu, Ian M. Head, Krishna P. Katuri, Mark C. M. van Loosdrecht, and Keith Scott. A computational model for biofilm-based microbial fuel cells. *Water Research*, 41(13):2921–2940, July 2007.

[68] Davit A. Potoyan and Peter G. Wolynes. On the dephasing of genetic oscillators. *Proceedings of the National Academy of Sciences*, 111(6):2391–2396, February 2014.

[69] José Pérez-Martıén and Manuel Espinosa. Correlation between DNA Bending and Transcriptional Activation at a Plasmid Promoter. *Journal of Molecular Biology*, 241(1):7–17, August 1994.

[70] Christopher Rackauckas and Qing Nie. DifferentialEquations.jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia. *Journal of open research software*, 5(1), 2017.

[71] Brandon C. Reyes, Irene Otero-Muras, and Vladislav A. Petyuk. A numerical approach for detecting switch-like bistability in mass action chemical reaction networks with conservation laws. *BMC Bioinformatics*, 23(1):1, January 2022.

[72] Daven Sanassy, Harold Fellermann, Natalio Krasnogor, Savas Konur, Laurentiu M. Mierla, Marian Gheorghe, Christophe Ladroue, and Sara Kalvala. Modelling and Stochastic Simulation of Synthetic Biological Boolean Gates. In *2014 IEEE Intl Conf on High Performance Computing and Communications, 2014 IEEE 6th Intl Symp on Cyberspace Safety and Security, 2014 IEEE 11th Intl Conf on Embedded Software and Syst (HPCC,CSS,ICESS)*, pages 404–408, August 2014.

[73] Germán D. Schrott, P. Sebastian Bonanni, Luciana Robuschi, Abraham Esteve-Nuñez, and Juan Pablo Busalmen. Electrochemical insight into the mechanism of electron transport in biofilms of Geobacter sulfurreducens. *Electrochimica Acta*, 56(28):10791–10795, December 2011.

[74] Spencer R. Scott and Jeff Hasty. Quorum Sensing Communication Modules for Microbial Consortia. *ACS Synthetic Biology*, 5(9):969–977, September 2016.

[75] Shimyn Slomovic, Keith Pardee, and James J. Collins. Synthetic biology devices for in vitro and in vivo diagnostics. *Proceedings of the National Academy of Sciences*, 112(47):14429–14435, November 2015.

[76] Rachel M. Snider, Sarah M. Strycharz-Glaven, Stanislav D. Tsoi, Jeffrey S. Erickson, and Leonard M. Tender. Long-range electron transport in Geobacter sulfurreducens biofilms is redox gradient-driven. *Proceedings of the National Academy of Sciences*, 109(38):15467–15472, September 2012.

[77] Brynne C. Stanton, Alec A.K. Nielsen, Alvin Tamsir, Kevin Clancy, Todd Peterson, and Christopher A. Voigt. Genomic Mining of Prokaryotic Repressors for Orthogonal Logic Gates. *Nature chemical biology*, 10(2):99–105, February 2014.

[78] Ruud Stoof, Lewis Grozinger, Huseyin Tas, and Ángel Goñi-Moreno. FlowScatt: enabling volume-independent flow cytometry data by decoupling fluorescence from scattering, July 2020.

[79] Ruud Stoof, Alexander Wood, and Ángel Goñi-Moreno. A Model for the Spatiotemporal Design of Gene Regulatory Circuits. *ACS Synthetic Biology*, 8(9):2007–2016, September 2019.

[80] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, November 2008.

[81] Alvin Tamsir, Jeffrey J. Tabor, and Christopher A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature*, 469(7329):212–215, January 2011.

[82] Huseyin Tas, Lewis Grozinger, Ruud Stoof, Victor de Lorenzo, and Ángel Goñi-Moreno. Contextual dependencies expand the re-usability of genetic inverters. *Nature*

Communications, 12(1):355, January 2021. Number: 1 Publisher: Nature Publishing Group.

[83] Jonathan Tellechea-Luzardo, Leanne Hobbs, Elena Velázquez, Lenka Pelechova, Simon Woods, Víctor de Lorenzo, and Natalio Krasnogor. Versioning biological cells for trustworthy cell engineering. *Nature Communications*, 13(1):765, February 2022.

[84] Jessica L. Terrell, Tanya Tschirhart, Justin P. Jahnke, Kristina Stephens, Yi Liu, Hong Dong, Margaret M. Hurley, Maria Pozo, Ryan McKay, Chen Yu Tsao, Hsuan-Chen Wu, Gary Vora, Gregory F. Payne, Dimitra N. Stratis-Cullum, and William E. Bentley. Bioelectronic control of a microbial community using surface-assembled electrogenetic cells to route signals. *Nature Nanotechnology*, 16(6):688–697, June 2021.

[85] Aidan Tinafar, Katariina Jaenes, and Keith Pardee. Synthetic Biology Goes Cell-Free. *BMC Biology*, 17(1):64, August 2019.

[86] Marios Tomazou, Mauricio Barahona, Karen M. Polizzi, and Guy-Bart Stan. Computational Re-design of Synthetic Genetic Oscillators for Independent Amplitude and Frequency Modulation. *Cell Systems*, 6(4):508–520.e5, April 2018.

[87] Tanya Tschirhart, Eunkyoung Kim, Ryan McKay, Hana Ueda, Hsuan-Chen Wu, Alex Eli Pottash, Amin Zargar, Alejandro Negrete, Joseph Shiloach, Gregory F. Payne, and William E. Bentley. Electronic Control of Gene Expression and Cell Behaviour in Escherichia Coli Through Redox Signalling. *Nature Communications*, 8(nil):14030, 2017.

[88] Toshiyuki Ueki. Cytochromes in Extracellular Electron Transfer in Geobacter. *Applied and Environmental Microbiology*, 87(10):e03109–20, April 2021.

[89] Toshiyuki Ueki, Kelly P. Nevin, Trevor L. Woodard, and Derek R. Lovley. Genetic Switches and Related Tools for Controlling Gene Expression and Electrical Outputs of Geobacter Sulfurreducens. *Journal of Industrial Microbiology & Biotechnology*, 43(11):1561–1575, 2016.

[90] Jesper Vind, Michael A. Sørensen, Michael D. Rasmussen, and Steen Pedersen. Synthesis of Proteins in Escherichia coli is Limited by the Concentration of Free Ribosomes: Expression from Reporter Genes does not always Reflect Functional mRNA Levels. *Journal of Molecular Biology*, 231(3):678–688, June 1993.

[91] Parisutham Vinuselvi, Seongyong Park, Minseok Kim, Jung Min Park, Taesung Kim, and Sung Kuk Lee. Microfluidic Technologies for Synthetic Biology. *International Journal of Molecular Sciences*, 12(6):3576–3593, June 2011.

[92] Baojun Wang, Richard I. Kitney, Nicolas Joly, and Martin Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2(1):508, October 2011.

[93] Leandro Watanabe, Tramy Nguyen, Michael Zhang, Zach Zundel, Zhen Zhang, Curtis Madsen, Nicholas Roehner, and Chris Myers. iBioSim 3: A Tool for Model-Based Genetic Circuit Design. *ACS Synthetic Biology*, 8(7):1560–1563, July 2019.

[94] Andrea Y. Weiße, Diego A. Oyarzún, Vincent Danos, and Peter S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, 112(9):E1038–E1047, March 2015.

[95] Thomas Wilhelm. The smallest chemical reaction system with bistability. *BMC Systems Biology*, 3(1):90, September 2009.

[96] Matthew D. Yates, Joel P. Golden, Jared Roy, Sarah M. Strycharz-Glaven, Stanislav Tsoi, Jeffrey S. Erickson, Mohamed Y. El-Naggar, Scott Calabrese Barton, and Leonard M. Tender. Thermally activated long range electron transport in living biofilms. *Physical Chemistry Chemical Physics*, 17(48):32564–32570, December 2015.

[97] Matthew D. Yates, Sarah M. Strycharz-Glaven, Joel P. Golden, Jared Roy, Stanislav Tsoi, Jeffrey S. Erickson, Mohamed Y. El-Naggar, Scott Calabrese Barton, and Leonard M. Tender. Measuring conductivity of living Geobacter sulfurreducens biofilms. *Nature Nanotechnology*, 11(11):910–913, November 2016.

[98] Tuval Ben Yehezkel, Arnaud Rival, Ofir Raz, Rafael Cohen, Zipora Marx, Miguel Camara, Jean-Frédéric Dubern, Birgit Koch, Stephan Heeb, Natalio Krasnogor, Cyril Delattre, and Ehud Shapiro. Synthesis and cell-free cloning of DNA libraries using programmable microfluidics. *Nucleic Acids Research*, 44(4):e35, February 2016.

[99] Lingchong You, Robert Sidney Cox, Ron Weiss, and Frances H. Arnold. Programmed population control by cell–cell communication and regulated killing. *Nature*, 428(6985):868–871, April 2004. Number: 6985 Publisher: Nature Publishing Group.