OPTIMIZATION-DRIVEN DESIGN OF SYNTHETIC GENETIC CIRCUITS USING BIOBRICKS

A Thesis in
Industrial Engineering
by
Ali Reza Zomorrodi

© 2011 Ali Reza Zomorrodi

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

December 2011
The thesis of Ali Reza Zomorrodi was reviewed and approved* by the following:

Costas D. Maranas  
Donald B. Broughton Professor of Chemical Engineering  
Thesis Co-Adviser

Soundar R.T. Kumara  
Allen E. Pearce/Allen M. Pearce Chaired Professor of Industrial and Manufacturing Engineering  
Thesis Co-Adviser

Paul Griffin  
Peter and Angela Dal Pezzo Department Head Chair  
Head of the Department of Industrial and Manufacturing Engineering

*Signatures are on file in the Graduate School.
ABSTRACT

The comprehensive compilation of the building blocks of synthetic genetic circuits in the MIT registry of standard biological parts (http://parts.mit.edu/) has provided a repository of spare parts to rationally create devices and systems with desired properties. The performance and interactions of the biological components in this database are primarily of a qualitative nature complicating their effective utilization for circuit design. Modeling approaches capable of harnessing the qualitative knowledge contained in this database are thus timely. Here, we introduce a computational framework that relies on the available qualitative information in the MIT registry to automatically identify the circuit components and connectivity for a desired response to the presence/absence of input signals. The promoters and ribosome binding sites are categorized as high, medium, and low efficiency and the protein expressions in the circuit are described using piecewise linear differential equations. The desired function of the circuit is also mathematically described as the maximization/minimization of an appropriately constrained objective function. We applied this framework to a variety of applications including design of a genetic toggle switch, a genetic decoder and a genetic half adder unit. The identified designs are consistent with previously constructed circuit configurations and in some cases point at completely new architectures. The non-intuitive circuit structures hint at the role of ribosome binding sites and relative protein abundance levels as controlling factors in circuit design. Our results demonstrate the value of the qualitative information in the MIT registry for coarse-grained circuit design and simulation in the absence of detailed quantitative and/or kinetic information.
# Table of Contents

List of Figures .................................................. v
List of Tables .................................................... vi
Acknowledgement ............................................... vii

Chapter 1 - Introduction ...................................... 1
  1.1. Background and literature review .................. 1
  1.2. Motivation .............................................. 3
  1.3. Objective .............................................. 4

Chapter 2 - Methods ........................................... 5
  2.1. Parameters definition for the parts in the MIT registry 5
  2.2. Development of the circuit design equations .......... 10
  2.3. Modeling the combinatorial effect of multiple regulatory proteins 16
  2.4. Circuit response modeling ................................ 18
  2.5. Optimization model .................................... 19

Chapter 3 - Results ........................................... 22
  3.1. Design of a genetic toggle switch ................. 23
  3.2. Design of a 2-4 bit genetic decoder ............... 25
  3.3. Design of a 1-bit genetic half adder ............... 29

Chapter 4 - Summary and conclusion ...................... 35

Bibliography ................................................... 38
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Effect of the regulatory proteins and ligands on transcription for different cases. (A)-(I): Normal state of the promoter is OFF. (A’)-(I’): Normal state of the promoter is ON. Refer to the definition of parameters $I_{pj}$ and $I'_{jl}$ for details.</td>
<td>6</td>
</tr>
<tr>
<td>3.1</td>
<td>Alternative circuit configurations suggested by the proposed framework when allowing for (A) two, (B) three (C) four and (D) five promoter-RBS-protein triples.</td>
<td>24</td>
</tr>
<tr>
<td>3.2</td>
<td>Seed circuit configuration producing YFP in the absence of glucose and presence of L-arabinose. This structure (without considering RBSs) was used by Dasika and Maranas [41] as a starting point for design of the genetic decoder using kinetic information. Here, we assume that all RBSs contributing to the expression of proteins in this circuit are high efficiency RBSs.</td>
<td>26</td>
</tr>
<tr>
<td>3.3</td>
<td>(A) An identified circuit structure for the genetic decoder starting from the seed circuit in Figure 3.2 when allowing for as many as twelve promoter-RBS-protein triplets. Ignoring the RBSs, this circuit has exactly the same structure as that obtained using kinetic description of interactions [41]. (B) Alternative circuit design of the same complexity without starting from the seed circuit. (C) A non-intuitive and more complex circuit design with up to 13 promoter-RBS-proteins.</td>
<td>27</td>
</tr>
<tr>
<td>3.4</td>
<td>Logic diagram for a half adder unit.</td>
<td>30</td>
</tr>
<tr>
<td>3.5</td>
<td>A non-intuitive identified circuit for the genetic half adder with eight promoter-RBS-protein triplets.</td>
<td>30</td>
</tr>
<tr>
<td>3.6</td>
<td>(A) A circuit structure for the genetic half adder with nine promoter-RBS-protein triplets, which has a global steady state and is consistent with a binary logic for regulatory protein-promoter interaction. (B) The corresponding logic circuit demonstrating how the identified design in (A) simulates the behavior of AND and XOR gates with Glucose and L-arabinose as inputs and YFP and GFP as outputs.</td>
<td>33</td>
</tr>
</tbody>
</table>
Summary of the available qualitative information in the MIT registry of standard biological parts for promoters, protein coding regions and RBSs.

Definition of the parameter $I'_{jl}$, representing the effect of a ligand $l$ on the function of a regulatory protein $j$. Here a positive effect implies that the ligand activates an inactive regulatory protein, whereas a negative effect means that the ligand disables an active regulatory protein.

Qualitative interpretation of the relative protein abundance levels based on the promoter strength and RBS efficiency, and their corresponding quantitative values (in parentheses) based on the values of the parameters $S_p$ and $h_r$. The basic assumption here is that the protein production level is governed by the additive effects of the promoter strength and RBS efficiency, i.e., protein production level (in the absence of active regulatory proteins and from a promoter whose normal state is ON) $\propto S_p + h_r$.

Effect of the presence/absence of active regulatory proteins on the transcription and, as a result, translation of a protein $j$ from a promoter $p$ and RBS $r$ (see Figure 2.1), and the corresponding values that variables $E_{pl}^{\text{Transcription}}$ and $E_{pl}^{\text{Translation}}$ should take. Note that here ‘no effect’ for translation means that the regulatory protein does not directly increase/decrease the translation rate using RBS $r$.

List of the promoters chosen by the proposed design framework for toggle switch, genetic decoder and genetic half adder.

Extracted information from the MIT registry regarding the interactions between ligands and regulatory proteins that appeared in the identified circuit structures for the toggle-switch, genetic decoder and genetic half adder.

Three typical low, medium and high efficiency RBSs used in the identified structures of the toggle switch, genetic decoder and genetic half adder. The reason why three typical RBSs was used is that our design framework accounts for only the efficiency of RBSs.

Truth tables for a 2-4 bit (A) binary decoder and (B) the equivalent genetic decoder. The genetic circuit should produce four different reporter proteins in response to presence/absence of two sugars glucose and L-arabinose. F0-F1 in (A) represent the outputs of the binary decoder, whereas BFP, GFP, YFP and RFP in (B) stand for blue, green, yellow and red fluorescent proteins, respectively.

Truth tables for a 1-bit (A) binary half adder and (B) the equivalent genetic half adder units. Here, YFP serves as the binary sum and GFP as the carry out.
Acknowledgement

My utmost gratitude goes to my adviser, Costas D. Maranas, for his continuous support, guidance and encouragement from the inception of project to completion of this thesis. His broad and deep scientific knowledge combined with his scholarly way of student advising has been exceptionally inspiring all his advisees including me. Indeed, I am indebted to him for teaching me how to perform high quality research. I also gratefully acknowledge Soundar R.T. Kumara for his understanding and help rendered to carry this research at such an inter-disciplinary level. Many thanks go to my group members, especially Madhukar Dasika, for his help and invaluable discussions and suggestions. I would like to convey my deepest gratitude to my wife Assieh whose dedication, love, encouragement and persistent confidence in me rendered my academic life smooth and manageable. Finally, I must thank my parents Fatemeh and Hossein for their love, continuous support and prays throughout my life and in all my pursuits.
Chapter 1- Introduction

1.1. Background and literature review

The goal of synthetic biology is to combine engineering principles and mathematical modeling for design and construction of genetic circuits with novel functions. The efforts in this direction started with design and experimental construction of two fundamental synthetic genetic circuits including a toggle switch [1] and the so-called repressilator [2]. The genetic toggle switch is a synthetic bistable gene-regulatory network [1], whereas the repressilator is an oscillating network that periodically induces the synthesis of a fluorescent protein [2]. Several other researchers used these foundations to construct similar toggle switches in *E. coli* [3] and mammalian cells [4], synchronized relaxation oscillators [5], a synthetic gene metabolic oscillator [6] and many others [3, 7-14]. There have been also other efforts for constructing cell-cell communication circuits [8, 15-24]. Synthetic genetic circuits were shown to have potential applications in many different areas ranging from biotechnology and bio-sensing [20, 25-30] to drug discovery and medicine [31-37]. For example, Farmer and Liao [25] as well as Chen and Wiess [20] demonstrated that the design and engineering of a synthetic regulatory circuit in *E. coli* and a synthetic cell-cell communication system in yeast can eliminate the need for monitoring of batch cultures and the control of gene expression through addition of expensive inducers [38]. The applications of synthetic biology in biotechnology is usually involved with integrating the native and *de novo* pathways to redirect the carbon flow towards the desired product. As an example, Martin *et al* [26] engineered *E. coli* to produce high-value terpenoid compounds such as the antimalarial drug artemisinin.
through introducing a synthetic mevalonate isoprenoid pathway from *S. cerevisiae*. In a more recent study, Shen and Liao [28] constructed a synthetic iterative pathway in *E. coli* for 2-ketoacid elongation. Potential application of synthetic biology in medicine was also demonstrated by using a synthetic genetic circuit to engineer replicating adenoviruses to selectively destroy tumor cells [31]. In another effort, Anderson *et al* [32] engineered the interaction between bacteria and mammalian cells so as *E. coli* was able to invade the specific cancer tumor cells.

In line with the rapid developments in the design and applications of synthetic genetic circuits the research community has taken a significant step towards the compilation and standardization of the building blocks of these circuits, by creating the MIT Registry of Standard Biological Parts (http://parts.mit.edu/). This registry is a collection of various genetic components (i.e., *biobricks*) such as promoters, protein coding regions, ribosome binding sites (RBSs) and devices such as logic gates. The registry has been continuously growing since its inception and it currently contains about 400 promoters, 350 protein coding sequences, 55 RBSs and 1250 composite parts. The development of this registry has enabled the design and construction of new integrated genetic circuits through the assembly of standardized interchangeable biological parts in analogy to the use of parts in electronic circuits. This has also spurred new computational approaches to assist the design process and verify the behavior of the genetic circuits prior to its experimental implementation. For example, Marchisio and Stelling [39] proposed a framework to simulate the behavior of synthetic genetic circuits with composable parts using ordinary differential equations, where different parts can be inter-connected through exchange of common signal carriers such as polymerases per second (PoPS) and ribosomes per
second (RiPS). In a subsequent study, the same researchers combined this approach with the so-called Karanaugh map method for design of electrical circuits to automate the design of digital biological circuits [40]. Alternatively, in a previous work from our group Dasika and Maranas [41] introduced OptCircuit, a computational framework that relies on the availability of kinetic parameters describing the function and interaction of circuit components to identify the required circuit elements and the structure that fulfills a prespecified task. More recently, Weeding et al [42] introduced a software tool called SynBioSS designer for quantitative simulation of biological circuits, which takes as input the list of components in a given circuit (e.g., biobricks from the MIT registry) and automatically generates a kinetic model of biological interactions involved.

1.2. Motivation

Despite these efforts, a key roadblock in comprehensive utilization of the basic parts available in the MIT registry is that, in contrast to electronic circuits, biological components in this database largely lack quantitative information characterizing their function and interactions. Instead, only a qualitative description of basic parameters of the biobricks and the interactions among them is currently available. This is due to not only the imprecise understanding of their function but more fundamentally due to the context-specific nature of their action. For example, only a qualitative description (low/medium/high) of promoter strength or efficiency of RBS as well as descriptive information about the interaction between regulatory proteins and promoters, or ligands and regulatory proteins is available in the MIT registry. Given this qualitative description of basic part characteristics, a key question here is how to systematically assemble these
components into complex circuits with desired responses to various ligand-specific stimuli. The computational approaches to address this problem are still nonexistent.

### 1.3. Objective

The objective of this study is to make use of the available qualitative information in the MIT registry of biological parts for the coarse-grained design of synthetic genetic circuits in the absence of detailed kinetic information characterizing the functions and interactions of the biobricks. To this end, we introduce an optimization-based framework combined with an adapted form of the piecewise-linear differential equations, which uses only a qualitative description of the efficacy of basic circuit components and interactions among them to suggest the circuit components from a list and alternative network structures to meet a desired functionality. We applied this framework to a variety of applications including design of a toggle switch, a 2-4 bit genetic decoder (a circuit that detects the presence/absence of two ligands) and a 1-bit genetic half-adder unit. The obtained results reveal the capability of this framework for designing circuits with simple and complex ligand-specific responses, using a qualitative description of biobricks available in MIT registry of biological parts.
Chapter 2 - Methods

2.1. Parameters definition for the parts in the MIT registry

The basic constitutive components of a genetic circuit in the MIT registry of standard biological parts are promoters (or regulatory regions), ribosome binding sites (RBSs) and protein coding regions. Qualitative information about the efficacy of each of these biobricks and interactions among them is available in the MIT registry (see Table 2.1), which can be potentially used to qualitatively predict the protein expressions levels in a circuit under a given condition. For example, promoter strength and RBS efficiency determine the level of transcription and translation, respectively, and thus have direct impact on the protein expression levels. The ON/OFF state of promoters in their normal

Table 2.1. Summary of the available qualitative information in the MIT registry of standard biological parts for promoters, protein coding regions and RBSs.

<table>
<thead>
<tr>
<th>Parts</th>
<th>Summary of the available information in the MIT registry</th>
<th>Related defined parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Promoters</td>
<td>Qualitative description of promoter strength (low/medium/high)</td>
<td>$S_p$</td>
</tr>
<tr>
<td></td>
<td>Normal state of the promoter (ON/OFF)</td>
<td>$Z_p$</td>
</tr>
<tr>
<td></td>
<td>Control information (how they are regulated, which proteins regulate them)</td>
<td>$I_{pj}$</td>
</tr>
<tr>
<td>Protein coding regions</td>
<td>Promoters under the control of regulatory proteins</td>
<td>$I_{pj}$</td>
</tr>
<tr>
<td></td>
<td>Ligands affecting (i.e., disabling/activating) the function of regulatory proteins</td>
<td>$I'_{jl}$</td>
</tr>
<tr>
<td>Ribosome binding sites (RBSs)</td>
<td>Qualitative information on the efficiency of RBSs (low/medium/high)</td>
<td>$\eta_r$</td>
</tr>
</tbody>
</table>
Figure 2.1. Effect of the regulatory proteins and ligands on transcription for different cases. (A)-(I): Normal state of the promoter is OFF. (A')-(I’): Normal state of the promoter is ON. Refer to the definition of parameters $I_{pj}$ and $I'_{jl}$ for details.
state (i.e., in the absence of any active regulatory proteins), the interaction of regulatory proteins and promoters, and finally the interaction of ligands and regulatory proteins are other pieces of information available in the MIT registry that can directly or indirectly impact the protein expression levels. The effect of the three latter parameters could be to turn on, increase or turn off transcription from a promoter. All these cases are summarized in Figure 2.1. Notably, this figure is used as a basis throughout this study for developing the design equations. In the following we define a set of parameters to mathematically capture the impact of each type of the biobrick and the interactions among them on the protein expression in a circuit. To this end, we first define the following sets:

\[
\begin{align*}
 P &= \{p \mid p = 1, 2, \ldots, N_P\} = \text{Set of the promoters} \\
 J &= \{j \mid j = 1, 2, \ldots, N_J\} = \text{Set of the protein coding regions} \\
 R &= \{r \mid r = 1, 2, \ldots, N_R\} = \text{Set of the RBSs} \\
 L &= \{l \mid l = 1, 2, \ldots, N_L\} = \text{Set of the ligands present in the medium} \\
 T &= \{t \mid t = 0, 1, \ldots, T_s\} = \text{Set of the time points}
\end{align*}
\]

Now, we define the following parameters for biobricks according to Table 2.1.

**Promoter and RBS characteristics:** Based on the available information in the MIT registry as summarized in Table 2.1, the promoters are categorized as high, medium and low strength and similarly RBSs as high, medium and low efficiency. In accordance, the following parameters are defined to describe the function and efficacy of promoters and RBSs.
\[ S_p = \begin{cases} 1 & \text{if } p \text{ is a low strength promoter} \\ 2 & \text{if } p \text{ is a medium strength promoter} \\ 3 & \text{if } p \text{ is a high strength promoter} \end{cases} \]

\[ \eta_r = \begin{cases} 1 & \text{if the efficiency of RBS } r \text{ is low} \\ 2 & \text{if the efficiency of RBS } r \text{ is medium} \\ 3 & \text{if the efficiency of RBS } r \text{ is high} \end{cases} \]

Notably, the basic assumption in definition of parameters \( S_p \) (promoter strength) and \( \eta_r \) (RBS efficiency) is that all the relative values are measured with respect to the same reference state.

The normal state of promoter (i.e., in the absence of any active regulatory proteins) is also abstracted in the definition of the following parameter:

\[ Z_p = \begin{cases} 1 & \text{if promoter } p \text{ is ON in its normal state} \\ 0 & \text{otherwise} \end{cases} \]

**Regulatory protein-promoter interaction**: We assume that the regulatory proteins act only in the level of transcription and (post)translational regulation is not considered in our model. The interaction of a regulatory protein \( j \) and a promoter \( p \) can be captured by defining a parameter \( I_{pj} \) as following:

\[ I_{pj} = \begin{cases} -1 & \text{if protein } j \text{ acts as repressor for promoter } p \\ 0 & \text{if protein } j \text{ has no effect on promoter } p \\ 1 & \text{if protein } j \text{ acts as an activator for promoter } p \end{cases} \]

A simplifying assumption in the definition of this parameters is that the effect of an active repressor protein on a promoter \( p \) is to completely stop transcription from that promoter, i.e., we ignore the probable leakiness of the promoters (see Figures 2.1D, D’, I
Table 2.2. Definition of the parameter $I'_{jl}$, representing the effect of a ligand $l$ on the function of a regulatory protein $j$. Here a positive effect implies that the ligand activates an inactive regulatory protein, whereas a negative effect means that the ligand disables an active regulatory protein.

<table>
<thead>
<tr>
<th>Effect of the ligand on regulatory protein</th>
<th>Requires a ligand to be activated</th>
<th>Requires no ligand to be activated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Status of the regulatory protein</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td></td>
<td>Inactive</td>
<td>Active</td>
</tr>
<tr>
<td></td>
<td>Active</td>
<td>Inactive</td>
</tr>
<tr>
<td>$I'_{jl}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

and $I'_{jl}$). Alternatively, it is hypothesized that the effect of an active activator protein on a promoter $p$ whose normal state is ON is to increase transcription from that promoter (see Figures 2.1B' and G').

**Ligand-regulatory protein interaction:** There can be two different outcomes for the effect of ligands on regulatory proteins. In the first case, a regulatory protein can perform its biological function in the absence of a ligand (see Figures 2.1B-E and 2.1B'-E'). The presence of a ligand in this case may disable the regulatory protein (see Figures 2.1C, E, C' and 2.1E'). In the second case, the regulatory protein requires the presence of a ligand to be activated (through binding to it) so as to perform its regulatory role (see Figures 2.1F-I and 2.1F'-I'). All these cases for the interaction of a ligand $l$ and a regulatory protein $j$ are accounted for in the definition of parameter $I'_{jl}$, as given in Table 2.2.

Notably, a basic assumption here is that when a ligand is denoted as present, its
concentration is high enough to affect the function of the regulatory protein. In addition we assume that ligands can interact directly with regulatory proteins.

Considering different scenarios for the effect of ligands on regulatory proteins as mentioned above and represented in Figure 2.1, a regulatory protein is defined as active if (i) the protein is in a form (either bonded to an activating ligand or alone in the absence of any inhibiting ligands) such that it is capable of performing its biological function, (ii) its concentration is above a threshold necessary to exert its regulatory effect. This definition implies that, the combinatorial effect of ligands and regulatory proteins on transcription from a promoter under control is condensed into the concept of active regulatory proteins.

2.2. Development of the circuit design equations

A class of piecewise linear differential equations, which was originally proposed by Glass and Kauffman [43] to provide a coarse-grained description of gene regulatory networks is adapted here to describe interactions in synthetic gene circuits. This model combines logic rules for the regulation of genes and synthesis of their products with (linear) free decay [44]. For each node one continuous and one discrete (Boolean) variable is defined. The continuous variable, $C_{jt}$, denotes the concentration of gene product $j$ at time $t$, and the binary variable, $Y_{jt}$, represents whether an activation (or switching) threshold, $\theta$, is reached at time $t$:

$$Y_{jt} = \begin{cases} 
1 & \text{if } C_{jt} \geq \theta \\
0 & \text{if } C_{jt} < \theta
\end{cases}$$

The system of piecewise linear differential equations can be described as following:
\[
\frac{dC_j}{dt} = G_j(Y_{1t},Y_{2t},...,Y_{Njt}) - C_j \quad \forall \ j \in J , \ t \in T
\]  

(2.1)

where \( G_j \) representing the protein generation, is a logical function that depends only on the values of the Boolean inputs \( Y_{1t},Y_{2t},...,Y_{Njt} \). The steady state of this model is obtained by setting the right-hand side of equation (2.1) to zero:

\[
G_j(Y_{1ss},Y_{2ss},...,Y_{Nss}) = C_j^{ss} \quad \forall \ j \in J
\]  

(2.2)

where the superscript ‘ss’ represents the steady-state. By using a finite difference approximation for the left-hand side of equation (2.1) and choosing \( \Delta t = 1 \), this equation is recast as the synchronous Boolean model for dynamics of genetic networks [44]:

\[
C_{j_{t+1}} = G_j(Y_{1t},Y_{2t},...,Y_{Njt}) \quad \forall \ j \in J , \ t \in \{0,1,2,...,T_s-1\}
\]  

(2.3)

This class of piecewise linear differential equations was previously used to study the evolution of complex dynamics in electronic models of synthetic genetic circuits [45].

In order to design synthetic genetic circuits using the available qualitative information in the MIT registry, we modified equations (2.2) and (2.3) so as to determine the relative protein abundance levels in the presence of a ligand \( l \) considering all possible interactions summarized in Figure 2.1. To this end, all the relevant variables appeared in the piecewise linear differential equations were extended by addition of a subscript \( l \), representing the effect of a ligand \( l \). Furthermore, to account for the effect of promoters and RBSs, the respective variables are extended through addition of indices \( p \) and \( r \). In the following we present the basic ideas for developing a design equation under steady state with modifying equation (2.2), where the subscript \( t \) representing time points, drops out from all variables which are time dependent. Development of design equations for
The non-steady state case can be done exactly in the same way through modifying equation (2.3) and keeping the subscript \( t \) for all time dependent variables.

The steady state relative abundance level of protein \( j \) produced from different promoters and RBSs in presence of a ligand \( l \), \( C_{jl} \), can be determined by modifying equation (2.2) as following:

\[
C_{jl} = \sum_p \sum_r G_{jprl} \quad \forall \quad j \in J, \ l \in L \tag{2.4}
\]

where \( G_{jprl} \) is a modified form of \( G_j \), denoting the production level of protein \( j \) from promoter \( p \) and RBS \( r \) in the presence of a ligand \( l \) (we drop the superscript ‘ss’ representing steady state for the simplicity of notations). It is important to note that \( C_{jl} \) in this equation is no longer a protein concentration, but it is a proxy entry that reflects its relative abundance. Equivalently, equation (2.3) can be modified to account for the (qualitative) dynamics of protein expression.

The key assumption in defining \( G_{jprl} \) is that the protein expression level accruing from a particular promoter and RBS is governed by the additive effects of promoter strength and RBS efficiency (i.e., \( S_p + \eta_r \)). The values of the relative protein abundance levels derived from this customized protein generation function can take any non-negative integer value depending on the circuit structure. These values, which are not quantitative but rather indicative of their relative abundance, are interpreted as follows and based on Table 2.3:

\[
\text{Relative protein abundance level} = \begin{cases} 
\text{Very low} & \text{if } C_{jl} = 1,2 \\
\text{Low} & \text{if } C_{jl} = 3 \\
\text{Medium} & \text{if } C_{jl} = 4 \\
\text{High} & \text{if } C_{jl} = 5 \\
\text{Very high} & \text{if } C_{jl} \geq 6
\end{cases}
\]
Table 2.3. Qualitative interpretation of the relative protein abundance levels based on the promoter strength and RBS efficiency, and their corresponding quantitative values (in parentheses) based on the values of the parameters $S_p$ and $\eta_r$. The basic assumption here is that the protein production level is governed by the additive effects of the promoter strength and RBS efficiency, i.e., protein production level (in the absence of active regulatory proteins and from a promoter whose normal state is ON) $\propto S_p + \eta_r$.

<table>
<thead>
<tr>
<th>Promoter strength ($S_p$)</th>
<th>RBS efficiency ($\eta_r$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low (1)</td>
<td>Low (1)</td>
</tr>
<tr>
<td>Medium (2)</td>
<td>Medium (2)</td>
</tr>
<tr>
<td>High (3)</td>
<td>High (3)</td>
</tr>
<tr>
<td>Very low (2)</td>
<td>Low (3)</td>
</tr>
<tr>
<td>Low (3)</td>
<td>Medium (4)</td>
</tr>
<tr>
<td>Medium (4)</td>
<td>High (5)</td>
</tr>
<tr>
<td>High (5)</td>
<td>Very high (6)</td>
</tr>
</tbody>
</table>

A high relative abundance (which corresponds to a value of five) is assumed to be the activation threshold ($\theta$) for a regulatory protein to be capable of performing its biological function. To mathematically enforce this threshold, the binary variable $Y_j$, which was introduced previously, is redefined as following by addition of a subscript $l$:

$$Y_{jl} = \begin{cases} 1 & \text{if } C_{jl} \geq 5 \\ 0 & \text{otherwise} \end{cases}$$

This definition can be imposed by using the following constraint:

$$5Y_{jl} \leq C_{jl} \leq (U_{Cjl} - 4)Y_{jl} + 4 \quad \forall \ j \in J , \ l \in L$$  \hspace{1cm} (2.5)

where $U_{Cjl}$ is the upper bound on $C_{jl}$ (e.g., 100).

We define the function $G_{iprl}$ such that it captures all possible scenarios summarized in Figure 2.1 in line with the available qualitative information in the MIT registry. As stated before, in our modeling framework we ignore the probable leakiness of the promoters, i.e., it is hypothesized that in the presence of an active repressor protein, transcription from the repressed promoter is stopped completely irrespective of the strength of...
promoter and RBS or presence of active activator proteins (see Figures 2.1D, D’, 2.1 and I’). The protein generation function in equation (2.4) is thus defined as following:

$$G_{jprl} = E_{pl}^{\text{Transcription}} S_p + E_{pl}^{\text{Translation}} x_r \quad \forall \ j \in J , \ p \in P , \ r \in R , \ l \in L$$  \hspace{1cm} (2.6)

where $E_{pl}^{\text{Transcription}}$ and $E_{pl}^{\text{Translation}}$ are two variables defined to impose the effects of an active regulatory protein in the presence of ligand $l$ on transcription from promoter $p$ and, as a result, translation with RBR $r$. Considering all possibilities represented in Figure 2.1, these two variables should take values represented in Table 2.4. Note that in the presence of an active activator protein and absence of any repressors, $E_{pl}^{\text{Transcription}}$ would take a value of $1 + Z_p$, i.e., if the promoter is ON in its normal state ($Z_p = 1$) then $E_{pl}^{\text{Transcription}} = 2$. However, this does not necessarily imply that the transcription from promoter is doubled. Instead it is just the indicative of a relative increase in transcription levels. In general

![Table 2.4. Effect of the presence/absence of active regulatory proteins on the transcription and, as a result, translation of a protein $j$ from a promoter $p$ and RBS $r$ (see Figure 2.1), and the corresponding values that variables $E_{pl}^{\text{Transcription}}$ and $E_{pl}^{\text{Translation}}$ should take. Note that here ‘no effect’ for translation means that the regulatory protein does not directly increase/decrease the translation rate using RBS $r$.](image)
$E_{\text{pl}}^{\text{Transcription}}$ and $E_{\text{pl}}^{\text{Translation}}$ are functions of $Y_{jl}$, $Z_p$, $I_{pj}$ and $I'_{jl}$ and are mathematically defined as following so as to take the values given in Table 2.4:

$$E_{\text{pl}}^{\text{Transcription}} = \left[ 1 - \left( Y_{jl} I'_{jl} \right)_{I_{pj} = 1} \right] \left[ Z_p + \left( Y_{jl} I'_{jl} \right)_{I_{pj} = 1} \right] \quad \forall \ p \in P \ , \ l \in L \quad (2.7)$$

$$E_{\text{pl}}^{\text{Translation}} = \left[ 1 - \left( Y_{jl} I'_{jl} \right)_{I_{pj} = 1} \right] \left[ Z_p + \left( 1 - Z_p \right) \left( Y_{jl} I'_{jl} \right)_{I_{pj} = 1} \right] \quad \forall \ p \in P \ , \ l \in L \quad (2.8)$$

Note that in these equations the terms $\left( Y_{jl} I'_{jl} \right)_{I_{pj} = 1}$ and $\left( Y_{jl} I'_{jl} \right)_{I_{pj} = -1}$ represent whether or not an active regulatory protein for promoter $p$ in the presence of ligand $l$ exists, i.e., they would take a value of one if an active activator or repressor protein $j$ is present, respectively, and zero otherwise.

The key design variable in our formulation is a binary variable $M_{p_{ij}}$ determining the required circuit components as well as the circuit structure. This binary variable will assume a value of one if protein $j$ should be expressed from a promoter $p$ and with a RBS $r$ in the circuit and zero otherwise:

$$M_{p_{ij}} = \begin{cases} 
1 & \text{if protein } j \text{ is expressed from promoter } p \text{ and RBS } r \\
0 & \text{otherwise}
\end{cases}$$

The general form of the circuit design equation can now be written as following through extending equation (2.4) to account for the design variables $M_{p_{ij}}$:

$$C_{jl} = \sum_p \sum_r M_{p_{ij}} G_{jprl} \left( Y_{jl}, Y_{2l}, \ldots, Y_{N_{jl}}, S_p, \eta_r, Z_p, I_{pj}, I'_{jl} \right) \quad \forall \ j \in J \ \& \ l \in L \quad (2.9)$$

where the protein generation function $G_{jprl}$ (see equations 2.6 to 2.8) is a function of the activation thresholds ($Y_{jl}$), promoter strength ($S_p$), RBS efficiency ($\eta_r$), the normal state of the promoter ($Z_p$) and the interactions between regulatory proteins and promoters ($I_{pj}$) and between ligands and regulatory proteins ($I'_{jl}$). If one is interested in the dynamic
behavior of the protein production all equations above can be extended by incorporation of subscript $t$, representing evolution of variables at a specific time instant. However, it should be noted that this analysis provides only a qualitative (and not exact) interpretation of the dynamics of the circuit.

2.3. Modeling the combinatorial effect of multiple regulatory proteins

In some cases combination of two or more regulatory proteins is needed to affect a specific promoter. This combinatorial effect of multiple regulatory proteins can be expressed by using the logic operations AND, OR or combinations thereof, exactly in the same the gene-protein-reaction associations are described in genome-scale metabolic models [46]. For example, the presence of both LuxR and LuxI is necessary to induce transcription from promoter pLux. This case represents an AND relationship between LuxR and LuxI to regulate pLux. All these cases can be readily incorporated in our formulation through replacing $Y_{jl}$ in equations (2.7) and (2.8) with a new binary variable and adding appropriate constraints to simulate the logic operations. In the following we describe each type of these interactions in details using ideas from genome-scale metabolic models [47]:

(i) Two regulatory proteins $j_1$ and $j_2$ are required to positively/negatively regulate a promoter $p$ through the formation of a single multi-protein complex. This case, which can be expressed as a logic AND relationship between $j_1$ and $j_2$ is enforced mathematically by replacing $Y_{jl}I_{jl}$ in equations (2.7) and (2.8) with a new binary variable $W_{pl}$ and adding the following set of constraints to the optimization problem:
\[
\begin{align*}
W_{pl} & = (Y_{j_1} I_{j_1} + Y_{j_2} I_{j_2}) - 1 \\
W_{pl} & \leq \frac{Y_{j_1} I_{j_1} + Y_{j_2} I_{j_2}}{2}
\end{align*}
\] (2.10)

Note that based on these equations, if at least one of \(Y_{j_1} I_{j_1}\) or \(Y_{j_2} I_{j_2}\) is zero, the binary variable \(W_{pl}\) is forced to zero. In addition, the new binary variable \(W_{pl}\) includes a subscript \(p\) (instead of \(j\)) to emphasize that it is exclusively defined for promoter \(p\). The above set of equations can be generalized for the case where more than two regulatory proteins are required to control transcription from a promoter \(p\), as following:

\[
\begin{align*}
W_{pl} & = \sum_{k=1}^{n_j} \left(Y_{j_k} I_{j_k}\right) - (n_j - 1) \\
W_{pl} & \leq \frac{\sum_{k=1}^{n_j} \left(Y_{j_k} I_{j_k}\right)}{n_j}
\end{align*}
\] (2.11)

where, \(n_j\) represents the total number of the regulatory proteins needed to control (positively or negatively) transcription from promoter \(p\).

(ii) Either of regulatory proteins \(j_1\) or \(j_2\) is sufficient to positively/negatively regulate transcription from promoter \(p\). This case, indicates a logic OR relationship between \(j_1\) and \(j_2\) and can be imposed mathematically by introducing a new binary variable \(W_{pl}\) and adding the following constraints to the problem:

\[
\begin{align*}
W_{pl} & \geq \frac{Y_{j_1} I_{j_1} + Y_{j_2} I_{j_2}}{2} \\
W_{pl} & \leq Y_{j_1} I_{j_1} + Y_{j_2} I_{j_2}
\end{align*}
\] (2.12)

Based on these equations, the binary variable \(W_{pl}\) is forced to zero only if both \(Y_{j_1} I_{j_1}\) and \(Y_{j_2} I_{j_2}\) are zero. The above set of equations can be also generalized for the case, where
more than two regulatory proteins (i.e., \( n_j > 2 \)) positively or negatively control transcription form a promoter \( p \) through OR relationships:

\[
W_{pl} = \frac{\sum_{k=1}^{n_j} (Y_{j_1,k} \cdot I_{j_1,k})}{n_j}
\]

\[
W_{pl} = \sum_{k=1}^{n_j} (Y_{j_2,k} \cdot I_{j_2,k})
\]

\[
(iii) \text{ More than two regulatory proteins with a combination of AND and OR relationships are required to positively/negatively control transcription from a promoter } p. \text{ This case can be simply simulated through combining equations (2.11) and (2.13). For instance, if the logic relationship among three regulatory proteins } j_1, j_2 \text{ and } j_3 \text{ regulating a promoter } p \text{ is expressed as } j_1 \text{ AND } (j_2 \text{ OR } j_3), \text{ the set of constraints simulating this relationship can be written as following:}
\]

\[
W'_{pl} \geq \frac{Y_{j_1,l} \cdot I_{j_1,l} + Y_{j_2,l} \cdot I_{j_2,l}}{2}
\]

\[
W'_{pl} \leq Y_{j_2,l} \cdot I_{j_2,l} + Y_{j_3,l} \cdot I_{j_3,l}
\]

\[
W_{pl} = (Y_{j_1,l} \cdot I_{j_1,l} + W'_{pl}) - 1
\]

\[
W_{pl} = \frac{Y_{j_1,l} \cdot I_{j_1,l} + W'_{pl}}{2}
\]

where \( W'_{pl} \) is a variable defined to simulates \((Y_{j_2,l} \cdot I_{j_2,l} \text{ OR } Y_{j_3,l} \cdot I_{j_3,l})\), using the first two equations. The third and fourth equations also define \((Y_{j_1,l} \cdot I_{j_1,l} \text{ AND } W'_{pl})\). More complex cases can be simulated in a similar way.

2.4. Circuit response modeling

Using an optimization-based platform for designing synthetic genetic circuits necessitates the selection of an objective function whose maximization/minimization captures the
desired circuit behavior in the presence or absence of biological signals (e.g., ligands). Typically, this is accomplished through the maximization of the sum of the relative abundance levels of a set of proteins (usually reporters) only in the presence of specific ligands:

\[
\text{Maximize} \quad z = \sum_{j_d \in J_d} \sum_{l_d \in L_d} C_{j_d l_d}
\]

where \( J_d \subseteq J \) and \( L_d \subseteq L \) are the set of desired proteins and ligands and \( C_{j_d l_d} \) is the steady state relative abundance level of the desired protein in the presence of a particular ligand \( l_d \). In most cases the desired circuit response also involves a zero (or low) concentration of some (reporter) proteins in the presence of specific ligands. This can be also enforced easily through adding appropriate constraints to the optimization problem. As an example if a circuit should produce only green fluorescent protein (GFP) in presence of ligand \( l_1 \) and only yellow fluorescent protein (YFP) in presence of ligand \( l_2 \), then this behavior can be modeled as following:

\[
\text{Maximize} \quad z = C_{\text{GFP}, l_1} + C_{\text{YFP}, l_2} \\
\text{s.t.} \quad C_{\text{YFP}, l_1} = 0 \\
\text{} \quad C_{\text{GFP}, l_2} = 0
\]

### 2.5. Optimization model

Based on the equations developed above, the problem of designing a synthetic genetic circuit using the qualitative information in the MIT registry about biobricks can be written as the following optimization problem:
Maximize \[ z = \sum_{j_d \in J_d} \sum_{l_d \in L_d} C_{j_d l_d} \] (2.14)

subject to

\[ C_{j_d l_d} = 0 \quad \forall \ j_d \in J_{undd} \ (J_{und} \subset J), \ l_d \in L_{d} \ (L_{d} \subset L) \] (2.15)

\[ C_{jl} = \sum_{j \in J} \sum_{l \in L} M_{prj} \left( S_p + E_{pl}^{\text{Translation}} Y_{jl} \right) \quad \forall \ j \in J, \ l \in L \] (2.16)

\[ E_{pl}^{\text{Transcription}} = \left[ 1 - \left( Y_{jl} I_{jl} \right) \right]_{I_{jl} = -1} \left[ Z_p + \left( Y_{jl} I_{jl} \right) \right]_{I_{jl} = -1} \quad \forall \ p \in P, \ l \in L \] (2.7)

\[ E_{pl}^{\text{Translation}} = \left[ 1 - \left( Y_{jl} I_{jl} \right) \right]_{I_{jl} = -1} \left[ Z_p + \left( 1 - Z_p \right) \left( Y_{jl} I_{jl} \right) \right]_{I_{jl} = -1} \quad \forall \ p \in P, \ l \in L \] (2.8)

\[ 5Y_{jl} \leq C_{jl} \leq (U^{C_p} - 4)Y_{jl} + 4 \quad \forall \ j \in J, \ l \in L \] (2.5)

\[ \sum_{r} \sum_{j} M_{prj} \leq N_{p}^{\max} \quad \forall \ p \in P \] (2.17)

\[ \sum_{p} \sum_{r} M_{prj} \leq N_{j}^{\max} \quad \forall \ j \in J \] (2.18)

\[ \sum_{p} \sum_{r} \sum_{j} M_{prj} \leq N_{\text{Circuit}}^{\max} \] (2.19)

\[ Y_{jl}, M_{prj} \in \{0, 1\} \]

\[ C_{jl}, E_{pl}^{\text{Transcription}}, E_{pl}^{\text{Translation}} \geq 0 \]

Equation (2.15) indicates that the concentration of all undesired (reporter) proteins in presence of a desired ligand should be zero (\( J_{und} \) is the set of all undesired proteins).

Constraint (2.16) is combination of equations (2.6) and (2.9) and describes the basic design equation. Equations (2.17) and (2.18) impose an upper limit on the number of times a promoter can be used and the number of times a protein can be expressed in a
circuit, respectively. Similarly, equation (2.19) determines the desired complexity of the circuit by imposing an upper limit on the total number of promoter-RBS-protein triplets in the circuit. We note that if \( E_{pl}^{\text{Transcription}} \) and \( E_{pl}^{\text{Translation}} \) in equation (2.16) are replaced by their corresponding expressions from equations (2.7) and (2.8), a nonlinear term (i.e., the product of two binary variables \( Y_{jl} \) and \( M_{prj} \)) appears in the formulation. This bilinear term can be linearized in the same way we simulate the AND relationship in equation (2.10), thereby converting our optimization problem to an integer linear program (ILP).
Chapter 3- Results

Here, we apply the proposed computational framework for the design of a variety of circuits including a toggle switch, a genetic decoder (a circuit that detects the absence/presence of multiple ligands) and finally a genetic half adder, which is able to add two one-digit genetic binary variables. All available promoters and regulatory protein coding regions in the MIT registry regarded as possible candidates for the design of these circuits, whereas only three typical (low, medium and high efficiency) RBSs are considered here as our design framework deals only with the efficiency of RBSs. The list of all components in the identified circuit structures along with their characteristics and interaction information extracted from the MIT registry are given in Tables 3.1-3.3.

Table 3.1. List of the promoters chosen by the proposed design framework for toggle switch, genetic decoder and genetic half adder.

<table>
<thead>
<tr>
<th>Promoters</th>
<th>MIT Part name</th>
<th>Normal State</th>
<th>Repressor</th>
<th>Activator</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_c</td>
<td>Bba_J23119</td>
<td>ON</td>
<td>-</td>
<td>-</td>
<td>High</td>
</tr>
<tr>
<td>P_Lac</td>
<td>BBa_R0010</td>
<td>OFF(^1)</td>
<td>LacI</td>
<td>CRP-cAMP(^1)</td>
<td>Medium(^2)</td>
</tr>
<tr>
<td>P_tet</td>
<td>BBa_R0040</td>
<td>ON</td>
<td>tetR</td>
<td>-</td>
<td>Medium</td>
</tr>
<tr>
<td>P_\lambda</td>
<td>BBa_R0051</td>
<td>ON</td>
<td>CI</td>
<td>-</td>
<td>High</td>
</tr>
<tr>
<td>P_BAD</td>
<td>BBa_I0500</td>
<td>OFF</td>
<td>AraC</td>
<td>CRP-cAMP(^1)</td>
<td>Medium</td>
</tr>
<tr>
<td>P_Ara</td>
<td>BBa_R0081(^4)</td>
<td>ON</td>
<td>AraC</td>
<td>-</td>
<td>Medium(^3)</td>
</tr>
</tbody>
</table>

\(^1\) For toggle switch example the normal state of this promoter is assumed to be ON and the positive regulation with CRP-cAMP was neglected to be able to compare our results with those of [1]. Furthermore, it is important to note that in the presence of glucose cAMP is not produced and vice versa.

\(^2\),\(^3\): There was no information available in MIT registry, assumed medium.

\(^4\) This part is not a real promoter. It should be connected to BBa_R0080, so as to be repressed by AraC. So, from P_Ara we in fact mean a promoter composed of BBa_R0081 and BBa_R0080.
Table 3.2. Extracted information from the MIT registry regarding the interactions between ligands and regulatory proteins that appeared in the identified circuit structures for the toggle-switch, genetic decoder and genetic half adder.

<table>
<thead>
<tr>
<th>Protein</th>
<th>MIT Part Name</th>
<th>Inhibitor</th>
<th>Activator</th>
</tr>
</thead>
<tbody>
<tr>
<td>LacI</td>
<td>BBa_C0012</td>
<td>IPTG</td>
<td>-</td>
</tr>
<tr>
<td>tetR</td>
<td>BBa_C0040</td>
<td>aTc</td>
<td>-</td>
</tr>
<tr>
<td>CRP</td>
<td>BBa_J58112</td>
<td>Glucose(^1)</td>
<td>cAMP</td>
</tr>
<tr>
<td>AraC</td>
<td>BBa_C0080</td>
<td>L_arabinose</td>
<td>-</td>
</tr>
</tbody>
</table>

\(^1\) No cAMP produced in the presence of glucose, thereby avoiding the CRP protein to be activated through binding with cAMP. This glucose implicitly suppresses the function of CRP, although it does not interact directly with it.

Table 3.3. Three typical low, medium and high efficiency RBSs used in the identified structures of the toggle switch, genetic decoder and genetic half adder. The reason why three typical RBSs was used is that our design framework accounts for only the efficiency of RBSs.

<table>
<thead>
<tr>
<th>RBS</th>
<th>RBS(_1)</th>
<th>RBS(_2)</th>
<th>RBS(_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Efficiency</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
</tr>
</tbody>
</table>

3.1. Design of a genetic toggle switch

In the first example we address the design of a toggle switch, which produces green fluorescent protein (GFP) in the presence of Isopropylthiogalactoside (IPTG) but not in the presence of anhydrotetracyclin (aTc). The identified network configurations with different complexities are shown in Figure 3.1. The simplest identified circuit (Figure 3.1A) contains only two promoter-RBS-protein triplets, where GFP is placed under the control of P\(_{Lac}\) and LacI is expressed from a (typical) strong constitutive promoter P\(_c\). When IPTG is present LacI is not able to repress transcription from PLac and GFP is thus freely expressed. On the other hand, when aTc is present, LacI can repress expression of GFP from PLac since aTc has no effect on the function of LacI. By allowing for up to
three promoter-RBS-protein triplets tetR is added to the circuit under the control of $P_{\text{Lac}}$, while LacI is expressed from the repressible promoter $P_{\text{Tet}}$, instead of a constitutive promoter (Figure 3.1B). Interestingly, this circuit configuration is the same as that of the well-studied toggle switch [1], and also that obtained by OptCircuit using kinetic equations [41].

The effect of increasing the complexity of the network by allowing for as many as four and five promoter-RBS-protein triplets is also shown in Figures 3.1C and D. As one can see from these figures, starting from the simplest circuit structure given in Figure 3.1A the effect of increasing the complexity of the network is to add extra regulatory proteins and promoters under their control in a cascade structure. The behavior of the cascades in genetic circuits has been investigated before by Hosshangi et al [48], where it was reported that the switching behavior becomes sharper for longer cascades. Therefore, increasing the complexity of this circuit provides the opportunity to get closer to a step-like behavior, which is desirable in most switching applications.

![Figure 3.1](image)

Figure 3.1. Alternative circuit configurations suggested by the proposed framework when allowing for (A) two, (B) three (C) four and (D) five promoter-RBS-protein triples.
We note that the majority of the RBSs that were identified in the suggested circuit structures were a typical high efficiency RBS (i.e., RBS$_3$), however in some cases a RBS with medium efficiency has been also used. This was expected because of the activation threshold for regulatory proteins, which was chosen to be the high relative abundance of protein, and since we restricted each protein to be expressed in the circuit only from a single promoter.

**3.2. Design of a 2-4 bit genetic decoder**

Next, we investigate the capability of the proposed approach for design of a more complex genetic circuit, which is equivalent to a 2-4 bit digital decoder. A digital decoder is a multiple-input and output logic device that converts inputs signals into coded outputs (see Table 3.4A). The corresponding genetic decoder should produce different types of reporter proteins in response to presence/absence of two sugars (ligands) glucose and L-arabinose (see Table 3.4B). A more detailed description of this circuit and a number of designs using a kinetic description of interactions was suggested previously by Dasika and Maranas [41], where all circuit structures were identified by starting from a seed circuit that produces YFP in the absence of glucose and presence of L-arabinose in a feed-forward structure (see Figure 3.2).

Using the design framework presented in this study, we identified a number of designs involving a minimum of 12 promoter-RBS-protein triplets (see equation 2.19). One such circuit design starting from the seed circuit is depicted in Figure 3.3A. Interestingly, regardless of the type of RBSs used, this circuit has the same configuration as the one obtained using kinetic information [41]. However, in contrast to the earlier effort [41],
Table 3.4. Truth tables for a 2-4 bit (A) binary decoder and (B) the equivalent genetic decoder. The genetic circuit should produce four different reporter proteins in response to presence/absence of two sugars glucose and L-arabinose. F0-F1 in (A) represent the outputs of the binary decoder, whereas BFP, GFP, YFP and RFP in (B) stand for blue, green, yellow and red fluorescent proteins, respectively.

(A)  
<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>F0</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

(B)  
<table>
<thead>
<tr>
<th>Glucose</th>
<th>L-arabinose</th>
<th>BFP</th>
<th>GFP</th>
<th>YFP</th>
<th>RFP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 3.2. Seed circuit configuration producing YFP in the absence of glucose and presence of L-arabinose. This structure (without considering RBSs) was used by Dasika and Maranas [41] as a starting point for design of the genetic decoder using kinetic information. Here, we assume that all RBSs contributing to the expression of proteins in this circuit are high efficiency RBSs.

where the rest of the circuit elements were successively added to the seed circuit in a stepwise procedure to satisfy the desired responses, here we identify all the required circuit components in only one step. This is mainly due to the absence of nonlinearities associated with kinetic models in our design framework, which led to significant savings in the computational resources. The proposed framework was also able to suggest alternative designs without starting from the seed circuit as depicted in Figure 3.3B.
Figure 3.3. (A) An identified circuit structure for the genetic decoder starting from the seed circuit in Figure 3.2 when allowing for as many as twelve promoter-RBS-protein triplets. Ignoring the RBSs, this circuit has exactly the same structure as that obtained using kinetic description of interactions [41]. (B) Alternative circuit design of the same complexity without starting from the seed circuit. (C) A non-intuitive and more complex circuit design with up to 13 promoter-RBS-proteins.

In both suggested circuit structures, with and without starting from the seed circuit (Figures 3.3A and B), CRP and AraC, which are expressed from a typical strong constitutive promoter act as the input sensors of the circuit that are directly affected by the presence/absence of glucose and L-arabinose. The effect of glucose and L-arabinose is then propagated into the other parts of the circuit through interaction of CRP and AraC with $P_{BAD}$, $P_{Ara}$ and $P_{Lac}$ promoters such that the expression of a fluorescent protein in a specific condition is accompanied with the expression of appropriate repressor proteins to inhibit the production of the other three fluorescent proteins.

We note that the two suggested circuit structures given in Figures 3.3A and 3.3B happen to be completely consistent with a pure binary logic viewpoint of regulatory protein-
promoter interactions. However, this is not always the case, for example when the complexity of the network increases by allowing for as many as 13 promoter-RBS-protein triplets. As depicted in Figure 3.3C, in the identified circuit design (besides \( P_{\text{Lac}} \) and \( P_{\text{Ara}} \)) tetR is expressed from a typical strong constitutive promoter (\( P_c \)), and thus a pure binary description of regulation would imply that GFP, which is under the control of \( P_{\text{tet}} \), should never be expressed. Nonetheless this circuit structure deemed functional as our proposed algorithm accounts for not only protein-promoter interactions and presence/absence of a regulatory protein, but also for the efficiency of RBSs and the relative abundance levels of proteins. We note that, even though tetR is expressed from a strong constitutive promoter, \( P_c \), only a (typical) low efficiency RBS (i.e., RBS_1) contributes to the expression of tetR from this promoter. In addition to \( P_c \), tetR is also expressed from \( P_{\text{Ara}} \) and \( P_{\text{Lac}} \) in this circuit with a medium efficiency RBS (RBS_2). When glucose is present in the environment, tetR is not expressed from \( P_{\text{Ara}} \) and \( P_{\text{Lac}} \), (see Tables 3.1 and 3.2) and the relative abundance level of tetR (produced only from the constitutive promoter with a low efficiency RBS) does not meet the specified activation threshold (i.e., high relative protein abundance level) to repress \( P_{\text{tet}} \) and hence GFP can be freely expressed from \( P_{\text{tet}} \). Alternatively, when both glucose and L-arabinose are present, even though the expression of tetR from \( P_{\text{Lac}} \) is inhibited, the combined relative abundance levels of tetR produced from the constitutive (\( P_c \)) and \( P_{\text{Ara}} \) promoters are sufficient to satisfy the activation threshold and repress transcription of GFP from \( P_{\text{tet}} \). A similar scenario takes place when neither glucose nor L-arabinose is present. These non-intuitive network designs provide additional flexibility in the practical construction of a genetic circuit through choosing different types of RBSs or promoters that might be
interconnected more easily. Overall, we can see in this example that however the circuit response does not seem to be very complicated, the required network structure is complex and the proposed approach can suggest alternative designs for the desired functionality.

3.3. Design of a 1-bit genetic half adder

Here, we employ the proposed framework for the design of a genetic circuit, which is equivalent to a 1-bit half adder unit. Half adders are building blocks of full adders, which serve as essential components of Arithmetic Logic Units (ALU), one of the fundamental building blocks of any CPU in computers. A digital half adder takes two 1-digit binary inputs, adds them and generates a 2-bit output containing a binary sum and a carry out. The truth table for this circuit is given in Table 3.5A.

A half adder can be implemented simply by using two AND and XOR gates as depicted in Figure 3.4. The corresponding 1-bit genetic half adder postulated here accepts Glucose and L-arabinose as input signals such that the absence or presence of each of these two ligands can be interpreted as 0 or 1, respectively. This genetic circuit is expected to produce YFP when only one ligand is present and GFP when both of them are present. In

Table 3.5. Truth tables for a 1-bit (A) binary half adder and (B) the equivalent genetic half adder units.

Here, YFP serves as the binary sum and GFP as the carry out.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X</td>
<td>Y</td>
<td>Carry out</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>L-arabinose</td>
<td>GFP</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
addition, neither YFP nor GFP should be produced when both ligands are absent, i.e., YFP serves as the binary sum and GFP as the carry out. The truth table of this genetic circuit is given in Table 3.5B. This desired circuit response is imposed by using the following objective function and constraints:

\[
\text{Maximize} \quad z = C_{YFP, \text{Glucose}} + C_{YFP, \text{L}-\text{Arabinose}} + C_{\text{GFP, Glucose & L}-\text{Arabinose}} \\
\text{s.t.} \\
C_{YFP, \text{No Glucose & No L}-\text{Arabinose}} = 0 \\
C_{\text{GFP, No Glucose & No L}-\text{Arabinose}} = 0
\]

Upon mathematically describing the function of this circuit, we were able to get circuit designs of varying complexities. One of the suggested network configurations with eight promoter-RBS-protein triplets is demonstrated in Figure 3.5. By visual inspection of this circuit (and considering Tables 3.1 and 3.2), one can quickly realize that this structure contradicts the binary logic for protein-promoter interactions. For example, when only glucose is present, the expression of tetR and CI from \( P_{\text{Lac}} \) is turned off, however, this does not guarantee that \( P_{\text{tet}} \) and \( P_{\lambda} \) promoters can be freely transcribed. Observe that part
of the circuit containing $P_{tet}$ and $P_\lambda$ promoters resembles the structure of the well-known toggle switch [1] (see Figure 3.1B) where two promoters express the cognate repressor of each other (this is also the case for $P_\lambda$ and $P_{Lac}$).

As one might expect, simulations using our time-varying formulation (where all time dependent variables are extended by adding a subscript $t$) confirmed that (results not shown) a toggle switch structure can exhibit a periodic behavior similar to a repressilator [2]. This is possible if none of the inhibiting ligands for the repressor proteins are present and when the initial concentrations of the repressor proteins are zero. However, further investigation of the circuit behavior using the dynamic formulation revealed that this structure can also reach a steady state in the absence of inhibiting ligands, if the initial concentration of one of the repressor proteins is above the activation threshold (e.g., through production from an external source). For example, in the circuit of Figure 3.5, if the initial relative abundance level of tetR meets the activation threshold, the expression of CI and GFP from $P_{tet}$ promoter is blocked and $P_\lambda$ is free to express YFP in the presence of glucose. Similarly, the initial concentrations of both tetR and LacI in the presence of L-arabinose, and that of CI in the presence (as well as absence) of both glucose and L-arabinose, respectively, should be above the activation threshold so as to observe the desired responses. This circuit structure was identified by our algorithm because when using the steady state equations (see equations 2.2 and 2.16), any configuration that meets the desired behavior in steady state can be a valid solution (at least from a mathematical point of view) regardless of what the initial conditions are. Specifically, in this example it can be concluded that the zero relative protein abundance levels of repressor proteins at time $t = 0$ are not contained in the domain of attraction of
this steady state (using the network structure in Figure 3.5) implying that this circuit will not show the expected behavior if the initial protein concentrations are zero. These findings suggest that the identified circuit structures might depend on the initial conditions, if the system does not have a global steady state. Nonetheless, we might be rather interested in circuit configurations that show the desired behavior with zero (or any) initial protein concentrations. Although it is not possible to impose \textit{a priori} the desired initial conditions when designing a circuit using the steady state equations, one can test quickly, using our time-varying formulation, whether or not the identified design can exhibit the desired response when starting from zero (or any other) initial concentrations. To this end, one can fix the network topology (i.e., the binary variables $M_{\text{prj}}$) in the dynamic formulation and simulate the evolution of the protein abundance levels until they reach an attractor (i.e., steady state or limit cycle). We note that these non-intuitive circuit designs can also indicate that no circuit structures with the specified level of complexity (i.e., the number of gene-protein-RBS triplets) can satisfy the desired behavior for any (or a desired) initial conditions. Hence, one can address this problem by re-designing the circuit and allowing for higher complexities.

For the genetic half adder in this example we were able to identify an alternative circuit structure with a global steady state (i.e, reachable from any initial condition) by increasing the circuit complexity and allowing for as many as nine promoter-RBS-protein triplets. One of the suggested designs is depicted in Figure 3.6A. Notably, no toggle switch like structure can be identified in this circuit and the structure is consistent with the binary logic for protein-promoter interactions. Compared to the equivalent digital half
Figure 3.6. (A) A circuit structure for the genetic half adder with nine promoter-RBS-protein triplets, which has a global steady state and is consistent with a binary logic for regulatory protein-promoter interaction. (B) The corresponding logic circuit demonstrating how the identified design in (A) simulates the behavior of AND and XOR gates with Glucose and L-arabinose as inputs and YFP and GFP as outputs.
output, while \( P_{\text{Lac}} \), \( P_{\text{tet}} \) and \( P_{\lambda} \) model the behavior of an AND gate with Glucose and L-arabinose as inputs and GFP as the output.

Notably, design of a similar circuit with light and a chemical signal as inputs has been reported before [49], however in that study new parts were constructed to be used as biological XOR and AND gates with desired inputs. Our design framework on the other hand is capable of choosing and assembling the available basic components in the MIT registry into a structure that models the behavior of the required logic gates and meets the desired functionality. This feature provides the opportunity for using the available well-characterized parts in the MIT registry and alleviating the need for the experimental construction of new logic devices for a desired circuit function.
Chapter 4- Summary and conclusion

In this study we put forth a computational framework for the design and simulation of synthetic genetic circuits, which relies on a qualitative description of the basic parts’ characteristics and interactions among them available in the MIT registry to identify the required circuit elements from a list and alternative configurations to meet a desired functionality. This was achieved through extending the scope of the piecewise linear differential equations so as to account for various types of biological parts involved in a genetic circuit (i.e., promoters, RBSs, protein coding regions and ligands) and interactions thereof. Despite relying only on qualitative information, the identified designs by our approach were consistent with previously constructed genetic circuits. In addition, avoiding the nonlinearities associated with kinetic models in our framework led to significant computational savings as demonstrated for the genetic decoder example. The time-varying formulation suggested here enables one to simulate the relative temporal behavior of a given circuit under any conditions without a need for kinetic information. Although qualitative in nature, dynamic Boolean and piecewise linear models have been proven successful in gaining a rough insight into the dynamic behavior of biological systems (such as gene regulatory and signal transduction networks) when no detailed quantitative and/or kinetic information is available [44, 50-55].

In contrast to the previous computational design procedures for synthetic genetic circuits [41, 56, 57], which consider only promoters and protein coding regions, here we go one step further by taking into account the effect of ribosome binding sites and their efficiency on the final protein expression levels. In addition, we consider not only the presence or absence of regulatory proteins but also their relative concentrations in the
circuit through imposing an activation threshold below which it is unable to exert its regulatory effect. As demonstrated for the design of the genetic decoder, this led to the identification of non-intuitive circuit structures, which are not discernible by being contingent only on a binary logic for protein-promoter interactions.

It is important to note that the design framework proposed here is associated with a number of limitations owing to relying on only a qualitative description of the basic parts’ characteristics and their interactions as well as a number of simplifying assumptions that was needed for using a discrete model. For example, here we ignored the promoter’s leakiness, which might affect the circuit function under experimental conditions. In addition, we assumed that when a ligand is denoted as present, its concentration is high enough to affect the regulatory proteins. This implies that our framework is not able to capture the deviations of the circuit behavior from what is predicted in silico in response to variations in the ligand concentrations under experimental conditions. We also ignored in our design equations the translational regulation, which can be an important factor when constructing the circuit experimentally. Given these limitations, the alternative circuit structures suggested by our framework should be deemed as ‘coarse-grained’ designs, which might need further refinements. This implies that the identified circuit structures can be used as a starting point by experimentalists for more detailed designs, or by other computational approaches [58, 59] for further analysis and fine-tuning the kinetic parameters. For example, the circuit structures suggested by our framework could be used as the input for the global sensitivity analysis algorithm of Fong et al [58] that estimates the sensitivity of
the properties of a given circuit with respect to the circuit model parameters such as the rate constants, without knowing their precise values.

Therefore, despite the limitations stated above, the proposed approach establishes a starting framework for fully utilizing the list of components in the MIT registry that can be successively improved. Furthermore, our approach can afford additional adjustments to provide more accurate designs. For example, one can modify the design equations to consider the translational regulation. Salis et al [60] have also recently proposed an optimization algorithm that can predict the sequence of a synthetic ribosome binding site that provides a target translation initiation rate based on a proportional scale. This translation initiation rate is akin to the ribosome binding site efficiency that was used in our design framework. Therefore, the development of similar tools capable of determining transcription initiations rates (i.e., the promoter strength) could provide an opportunity to use the proposed framework for more detailed design of synthetic genetic circuits. Overall, this study ushers the computational means for harnessing the wealth of qualitative knowledge encoded in the MIT registry and provides a bird’s eye view on the coarse-grained design and simulation of synthetic genetic circuits using piecewise linear differential equations.
Bibliography


