

Transforming Cells into Automata

Based On The Following Papers:

1. Engineering Life: Building a FAB for Biology – David Baker, George Church, Jim Collins, Drew Endy, Joseph Jacobson, Jay Keasling, Paul Modrich, Christina Smolke and Ron Weiss
2. Genetic Circuit Building Blocks for Cellular Computation, Communications, and Signal Processing – Ron Weiss, Subhayu Basu, Sara Hooshangi, Abigail Kalmbach, David Karig, Rishabh Mehreja and Ilka Netravali

Additional References:

1. Lecture slides by Ravi Tiruvury
2. <http://cnx.org/content/m12383/latest/>
3. <http://en.wikipedia.com>

Outline

1. Background

- 1.1 Revisiting Logic Gates
- 1.2 Defining Signal Processing

2. Gene Networks

- 2.1 Definition
- 2.2 Need for Gene Networks

3. Generic Circuit

- 3.1 Electrical Circuit vs. Genetic Circuit
- 3.2 Building Genetic Circuits
- 3.3 Genetic Circuit Building Block

4. Circuit Design

- 4.1 Rational Design
- 4.2 Direct Evolution

5. Cell-Cell Communication

6. Signal Processing

7. Conclusions




1 Background

Genetic engineering with recombinant DNA is a powerful and widespread technology that enables biologists to redesign life forms by modifying or extending their DNA. Advances in this domain allow us to gain insight into the operating principles that govern living organisms, and can also be applied to a variety of fields including human therapeutics, synthesis of pharmaceutical products, and molecular fabrication of biomaterials, crops and livestock engineering.



Constructing DNA fragments that consist of almost any gene sequence is not a difficult task. However the behavior of the resulting genetic constructs is not easy to predict. To address this issue, it is important to develop an engineering methodology for creating synthetic gene networks that will allow us to engineer cells with the same ease and capability with which we currently program computers and robots. The first step in making programmed cell behavior a practical and useful engineering discipline is to assemble a component library of genetic circuit building blocks. These building blocks perform computation and communications using DNA-binding proteins, small inducer molecules that interact with these proteins. A component library of cellular gates can be defined that implement several digital logic functions.

1.1 Revisiting Logic Gates

Basic Logic Gates:

Type	Distinctive shape	Truth Table															
AND		<table border="1"> <thead> <tr> <th>A</th> <th>B</th> <th>out</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	A	B	out	1	1	1	1	0	0	0	1	0	0	0	0
A	B	out															
1	1	1															
1	0	0															
0	1	0															
0	0	0															
OR		<table border="1"> <thead> <tr> <th>A</th> <th>B</th> <th>out</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>1</td> </tr> <tr> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>0</td> <td>1</td> <td>1</td> </tr> <tr> <td>0</td> <td>0</td> <td>0</td> </tr> </tbody> </table>	A	B	out	1	1	1	1	0	1	0	1	1	0	0	0
A	B	out															
1	1	1															
1	0	1															
0	1	1															
0	0	0															
NOT		<table border="1"> <thead> <tr> <th>A</th> <th>out</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> </tr> </tbody> </table>	A	out	1	0	0	1									
A	out																
1	0																
0	1																

Universal Logic Gates:

Type	Distinctive shape	Truth Table															
NAND		<table border="1"> <thead> <tr> <th>A</th> <th>B</th> <th>out</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td>0</td> <td>1</td> </tr> <tr> <td>0</td> <td>1</td> <td>1</td> </tr> <tr> <td>0</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	A	B	out	1	1	0	1	0	1	0	1	1	0	0	1
A	B	out															
1	1	0															
1	0	1															
0	1	1															
0	0	1															
NOR		<table border="1"> <thead> <tr> <th>A</th> <th>B</th> <th>out</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>1</td> <td>0</td> </tr> <tr> <td>1</td> <td>0</td> <td>0</td> </tr> <tr> <td>0</td> <td>1</td> <td>0</td> </tr> <tr> <td>0</td> <td>0</td> <td>1</td> </tr> </tbody> </table>	A	B	out	1	1	0	1	0	0	0	1	0	0	0	1
A	B	out															
1	1	0															
1	0	0															
0	1	0															
0	0	1															

1.2 Defining Signal Processing

Signal processing is the processing, amplification and interpretation of signals (either analog or digital), and deals with the analysis and manipulation of signals. Signals of interest include sound, images, and biological signals such as ECG, radar signals, and many others. Processing of such signals includes storage and reconstruction, separation of information from noise (e.g., aircraft identification by radar), compression (e.g., image compression), and feature extraction (e.g., speech-to-text conversion).

2 Gene Networks**2.1 Definition**

A gene network (also called a Gene Regulatory Network (GRN) or genetic regulatory network,) is a collection of DNA segments in a cell which interact with each other and with other substances in the cell, thereby governing the rates at which genes are transcribed into mRNA. Genes can be viewed as nodes in such a network, with input being proteins such as transcription factors, and outputs being the level of gene expression. The node itself can also be viewed as a function which can be obtained by combining basic functions upon the inputs (in the Boolean network these are Boolean functions or gates computed using the basic AND OR and NOT gates in electronics). These functions have been interpreted as performing kind information processing within cell which determines cellular behavior. The basic drivers within cells are levels of some proteins, which determine both spatial (tissue related) and temporal (developmental stage) co-ordinates of the cell, as a kind of "cellular memory". The gene networks are

only beginning to be understood, and it is a next step for biology to attempt to deduce the functions for each gene "node", to assist in modeling behavior of a cell.

Gene networks act as analog biochemical computers to specify the identity and level of expression of groups of target genes. Central to this computation are DNA recognition sequences with which transcription factors associate. When active transcription factors associate with the promontory region of target genes, they can function to specifically repress (down-regulate) or induce (up-regulate) synthesis of the corresponding RNA. The immediate molecular output of a gene regulatory network is the constellation of RNAs and proteins encoded by network target genes. The resulting cellular outputs are changes in the structure, metabolic capacity, or behavior of the cell mediated by new expression of up-regulated proteins and elimination of down-regulated proteins.

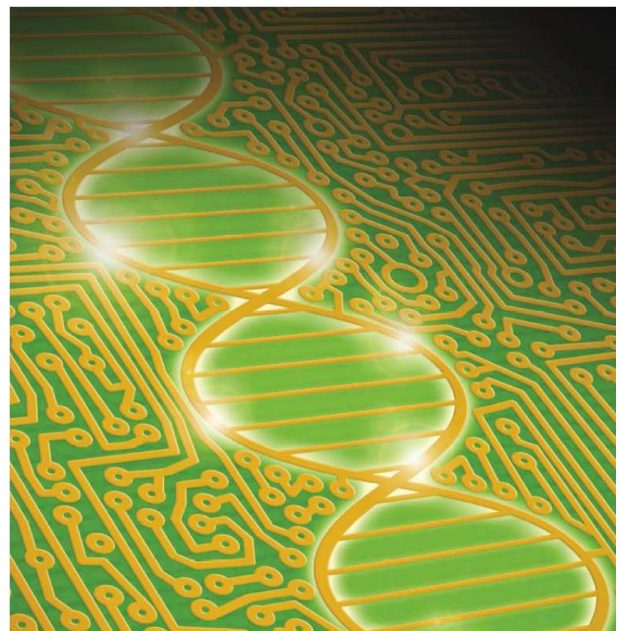
2.2 Need for Gene Networks

A central focus of genomic research concerns understanding the manner in which cells execute and control the enormous number of operations required for their function. Biological systems behave in an exceedingly parallel and extraordinarily integrated fashion. Feedback and damping are routine even for the most common activities. Thus, in this area of genomic biology, single gene perspectives are becoming increasingly limited for gaining insight into biological processes. Network models (Gene network) are becoming increasingly important for making progress in our understanding of the manner in which genes and molecules collectively form a biological system and harnessing this understanding in educated intervention for correcting human diseases.

3 Genetic Circuit

3.1 Definition

Genetic circuit is an approach to model genetic networks using boolean constructs such as AND, OR, NOT, NAND. These cellular gates include component for intracellular computation (i.e. NOT and NAND) and devices for external communication (i.e. IMPLIES and AND). The building blocks have already been assembled into several prototype genetic circuits in *Escherichia coli* bacteria cells, with up to three logic gates per cell. These genetic elements can also be configured to process environmental and internal biochemical analog signals.



3.2 Electrical Circuit vs. Genetic circuits

Electrical Circuits	Genetic Circuits
<ul style="list-style-type: none"> • Basic component of an Electrical Circuit: <i>Transistor</i> • Binary “1” => “high” voltage output • Binary “0” => “low” voltage output • Communication occurs in a fixed, closed environment (like a wire) • Outcomes are <i>deterministic</i> 	<ul style="list-style-type: none"> • Basic component of a Genetic Circuit: <i>Gene</i> • Binary “1” => “high” protein concentration • Binary “0” => “low” protein concentration • Communication occurs in an open environment with the signal possibly received by other than intended recipients • Outcomes are <i>stochastic</i>

3.3 Building Genetic Circuits

Step 1: Build a Genetic Component Library

- Biochemical Inverter
- IMPLIES Gate
- NAND Gate
- AND Gate

Step 2: Assemble them into a Biocircuit

Step 3: Tweak/tune the circuit and its components till the desired output is reached

Step 4: Check output by using a fluorescent protein as a reporter (*for illustrative purposes*)

3.4 Genetic Circuit Building Blocks

The first step in programming cells and controlling their behavior is to establish a library of well-defined components that serve as the building blocks of more complex systems.

Biochemical Inverter

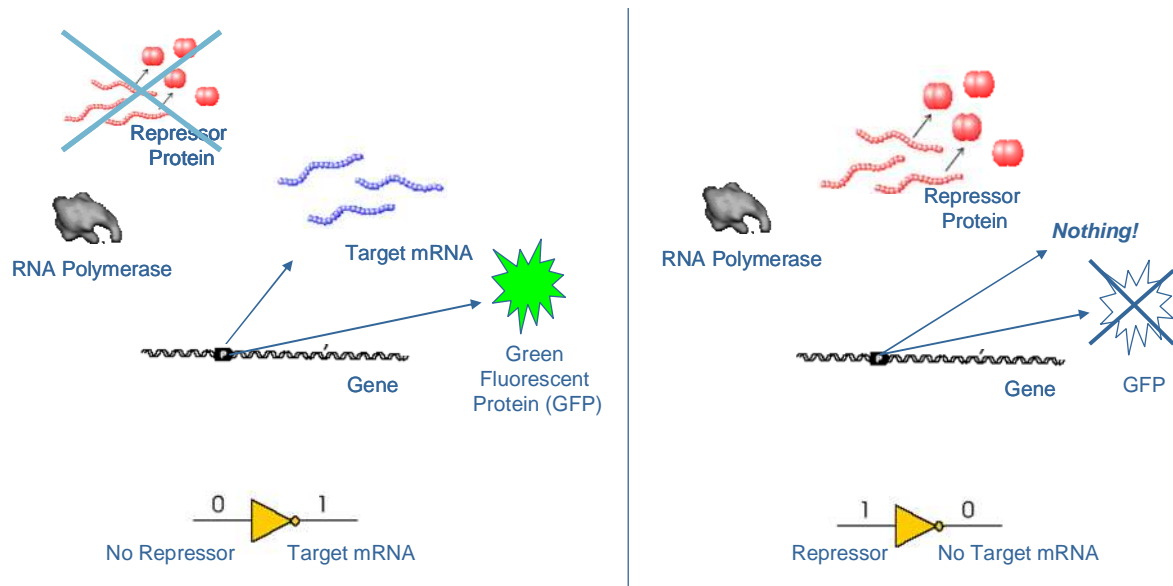


Figure: Biochemical Inverter

In the above figure, the presence or absence of Repressor Protein determines the two possible outputs. In the absence of Repressor Protein (input: 0) there is a formation of Target mRNA (output: 1) as indicated by Green Fluorescent Protein (GFP). When the Repressor Protein is present (input: 1), there is no Target mRNA (output: 0). The above phenomenon can be modeled by an inverter circuit and this is an example of Biochemical Inverter.

The below figure depicts a functional model of the inverter derived from its biochemical reaction phases. The first phase in inversion is the translation stage. The input signal to this stage, and thus the inverter, corresponds to the concentration level of the input mRNA. In the second phase, input protein monomers combine to form polymers that bind the operator, and subsequently repress the transcription of the output gene. In the final stage of the inverter, the transcription stage, RNA polymerase transcribes the regulated gene and the input signal is inverted.

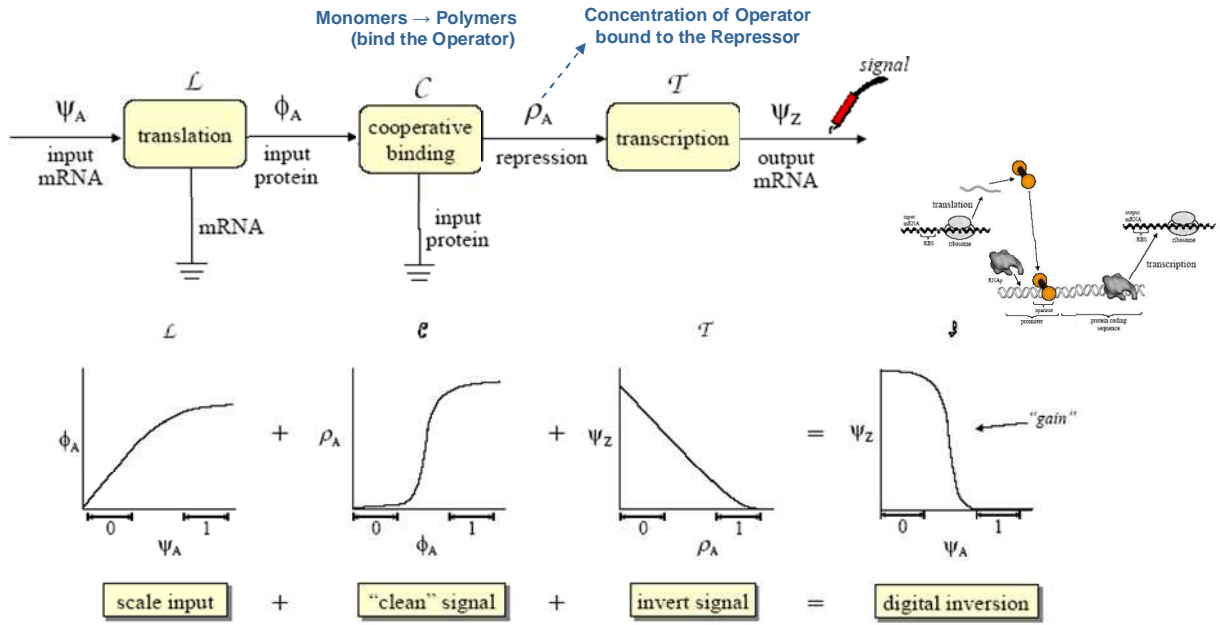


Figure: Inverter Functional Model

Biochemical NAND Logic Gate

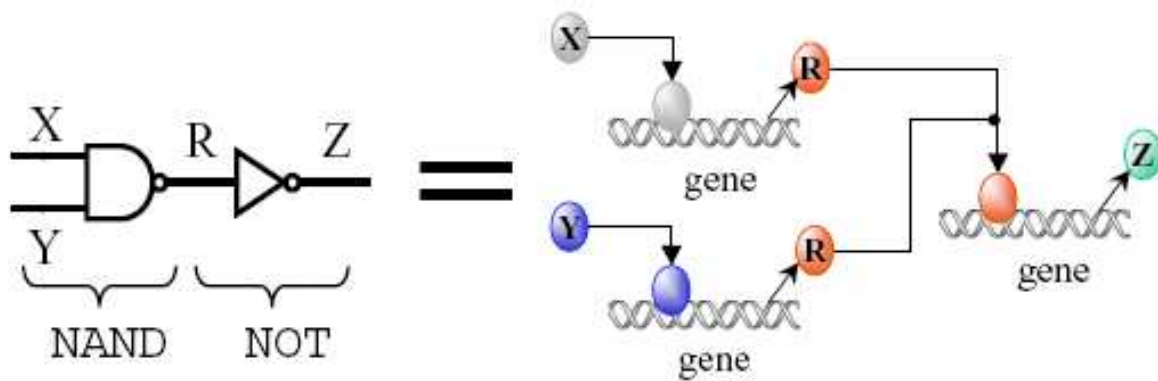


Figure: NAND Logic gate

Biochemical inverters are used to construct more sophisticated gates and logic circuits. For example, a NAND gate can be designed by “wiring OR” the outputs of two inverters by assigning them same output gene. The above figure depicts a circuit in which a NAND gate is connected to an inverter. The NAND gate is a universal logic element that can theoretically be used to wire any finite intracellular digital circuit.

AND Gate for Cell-Cell Communication

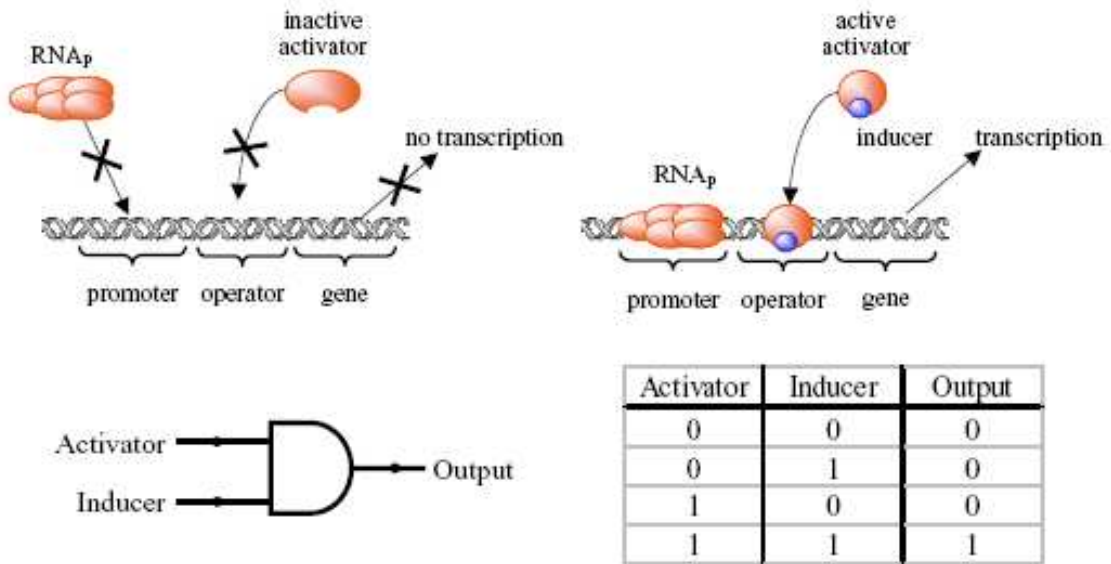


Figure: Detecting Cell-Cell Communication with AND Gate

The AND gate is utilized by cells to detect incoming messages sent by other neighboring cells. The biochemical reactions, the logic symbol, and the truth table for an intercellular gate that implements the AND function are illustrated in the above figure. In this construct, RNAp has a low affinity for the promoter and thus, basal transcription activity is minimal. It follows that in the absence of the activator and inducer, the logic output of the AND gate is LOW. When only the activator is present, the output is still LOW, since the activator has little affinity for the operator without its corresponding inducer. The output is HIGH only if both the activator and inducer are present. In this case, the inducer binds the activator and changes its conformation, yielding an activator/inducer complex that binds the promoter. This complex helps recruit RNAp to the promoter and initiate transcription to yield a HIGH output.

Ring Oscillator

Oscillators work on the principles of oscillation: a periodic fluctuation between two things based on changes in energy. The repressilator is an oscillatory gene network

constructed with three repressors that are not part of a natural biological clock network. The oscillation frequency of the genetic network is less than the cell division frequency, and as a result, the oscillations are propagated through the generations.

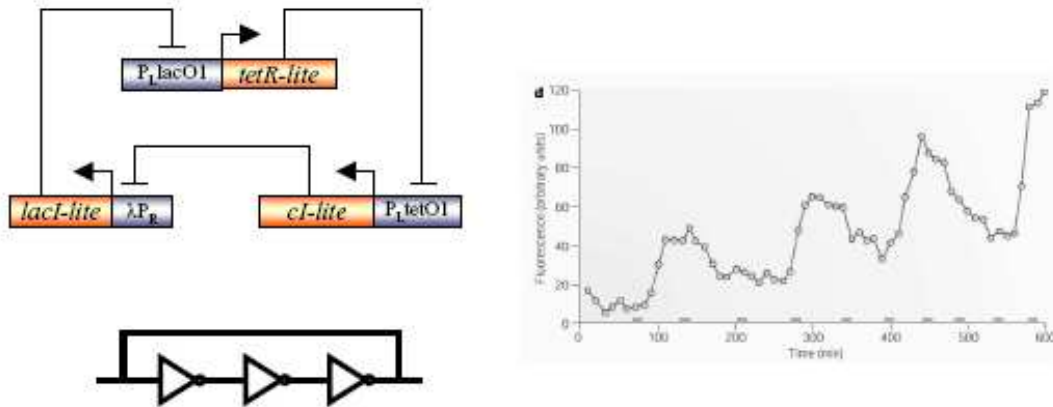


Figure: The repressilator circuit

In the above figure, the network comprises three fast-decaying versions of the repressors CI, LacI and tetR and their corresponding promoters. CI represses the expression of LacI, which in turn represses the transcription of tetR. The corresponding tetR Protein subsequently inhibits the expression of CI. When the kinetic characteristics are matched between the different gene components under the appropriate delay conditions, this system will oscillate with regular periodicity and amplitude.

4 Circuit Design

The goal is to design a DNA sequence that reliably implements a desired cellular function with quantitative precision. This becomes more and more important as the size of target synthetic gene network grows. There are two seemingly divergent approaches for genetic circuit design. The first methodology employs “rational design” in which an attempt is made to gain accurate knowledge of the genetic components and their compositions. The second approach, directed evolution uses large scale genetic mutations and combinatorial synthesis, combined with high throughput assays, to screen for genetic network variations that yield the desired behavior.

4.1 Rational Design

Modeling

Computational models are necessary for systematic circuit design and analysis. However, modeling a genetic circuit is more complicated:

- Interactions between circuit components (genes and proteins) are not fixed
- State transitions are rarely simultaneous
- Outcomes are not deterministic
- Gene networks tend to exhibit significant noise even in the simplest configurations

Depending on the requirement, deterministic and stochastic models are used.

Deterministic Model

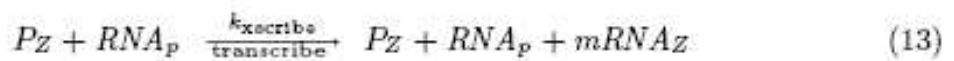
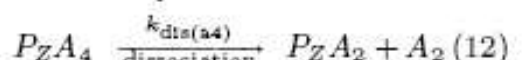
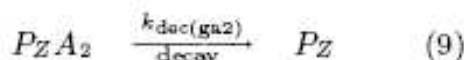
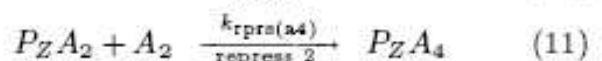
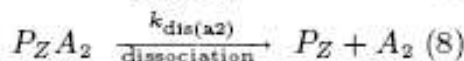
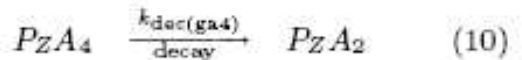
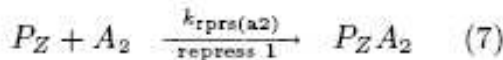
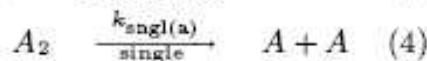
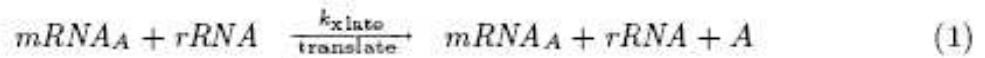
A common method for modeling biological circuits uses nonlinear ordinary differential equations:

- The circuit components, i.e. RNA, Protein and other molecule concentrations, are represented by time-dependent variables.
- Rate equations describe biochemical reactions as a function of concentrations of the circuit components. They are of the form:

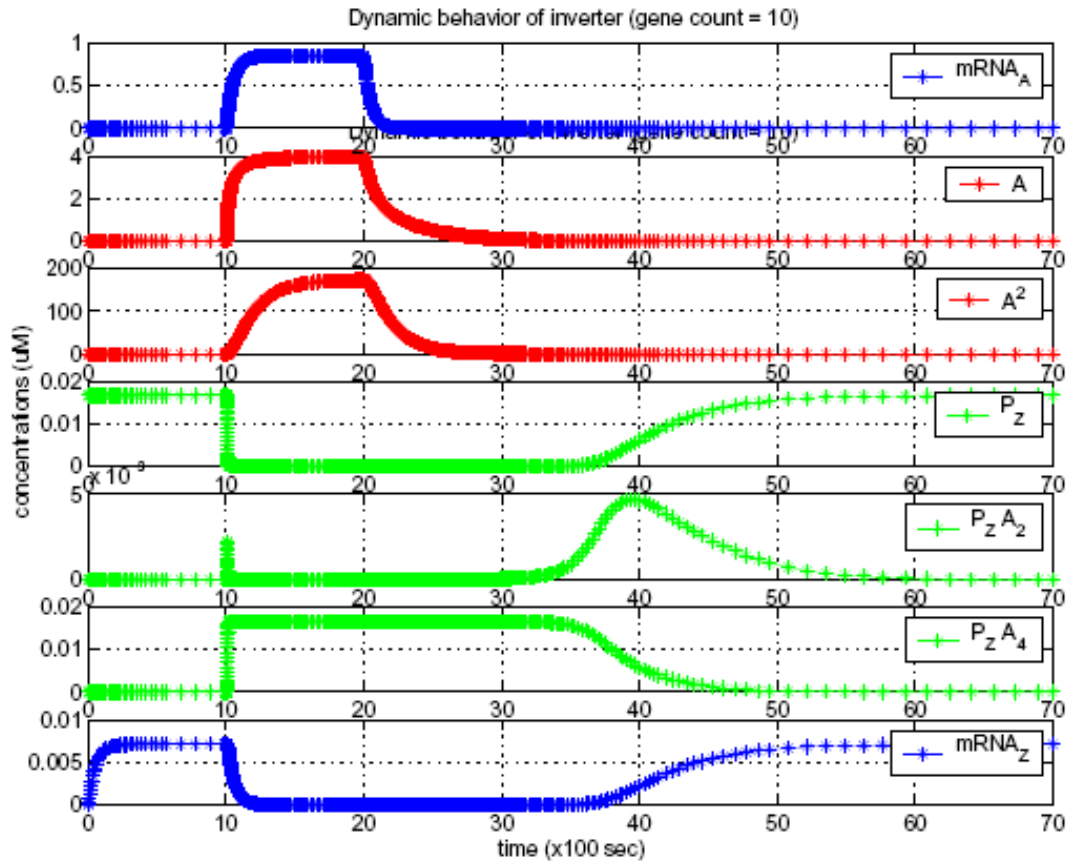
$$\frac{dx_i}{dt} = f_i(x), 1 < i < n$$

Where vector $x = [x_1, \dots, x_n]$ includes concentrations of proteins, mRNAs, other molecules and f_i is a nonlinear function

Modeling of an Inverter



The above table lists the chemical reactions that model the inverter. From the fourteen reactions, seven ordinary differential equations are derived, one for each molecular species in the system. Each differential equation describes the time-domain behavior of a particular molecular species.



Deterministic vs. Stochastic Models

ODE's are good for:

- Systems with large number of molecules for any given species
- Systems which are both continuous and deterministic.

However, in reality:

- Biochemical systems consist of few molecules for a given species
- They are usually discrete (reactions change population dynamics at irregular intervals) and stochastic (outcomes vary with order of reactions, environment, inter-component interactions)

•

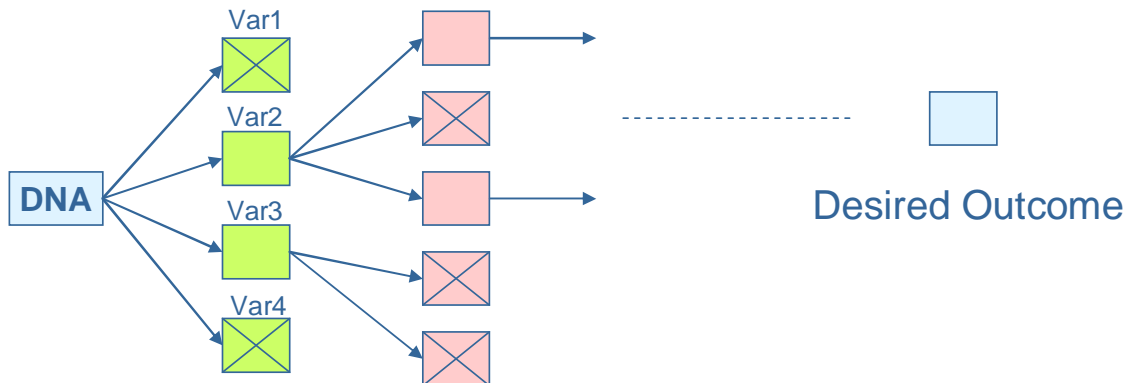
Tradeoff:

- Use Deterministic models if only average behavior needs to be modeled, and computational resources are limited.
- Use Stochastic models if accurate quantitative information about noise is available and large computational resources are available.

4.2 Directed Evolution

For this technique, one does not need to tackle with the issue of what DNA sites to mutate. Here is how it works:

- Library Creation: Mutate/recombine the gene (encoding the protein of interest) at random. Create a large library of variants.
- Variant Screening: Test how the variants perform and contribute to the overall response of the circuit.
- If favorable, screen those components, discard the rest, and proceed with mutating another component.



5 Cell-Cell Communication

Cell-cell communication involves a “chemical message” from a sender cell to a receiver cell, wherein subsequently a remote transcriptional response is activated.

The sender cell produces small signaling molecules using metabolic pathways. The molecules diffuse outside the membrane and into the environment. The signals then diffuse into the neighboring cells and they interact with proteins in receiver cells.

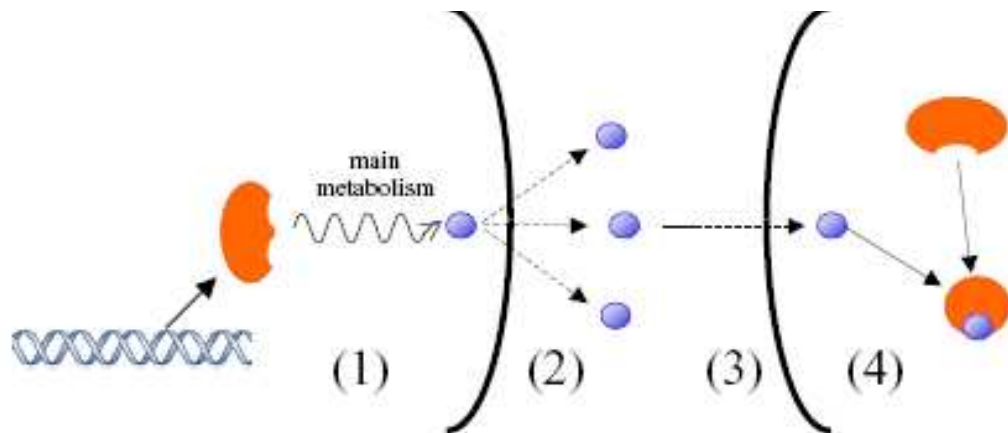


Figure: Cell-Cell Communication schematics

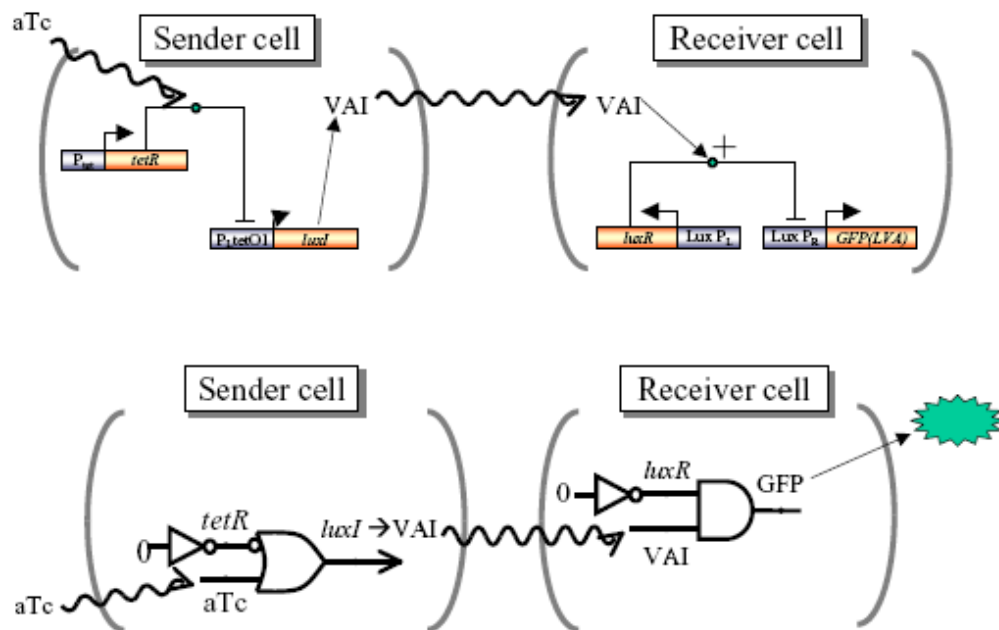


Figure: Genetic and corresponding logic circuits for cell-cell communications

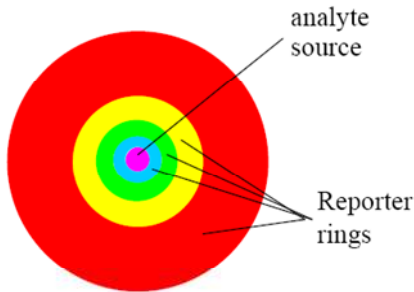
Notes:

- $tetR$ represses $luxI$. But inducer aTc overrides $tetR$ and induces $luxI$ production
- VAI \Rightarrow Vibrio Auto Inducer. Chemically, this is
- GFP \Rightarrow Green Fluorescent Protein, located downstream of $luxP_R$ promoter

6 Signal Processing

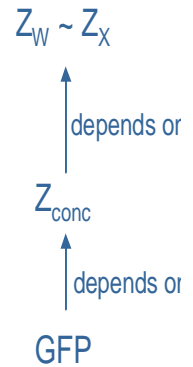
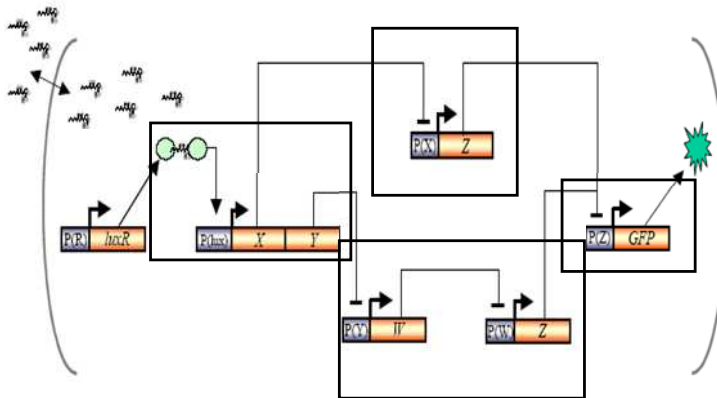
Consider a scenario where Cell A to respond to Cell B only if the signal sent by the sender falls in a particular concentration range. A Real-world Example: The retina generates electrical nerve signals in response to the photons detected by rhodopsin in retinal cells. Here, it's not just the "presence", but also the "strength" or "concentration" of the photons is important to generate an appropriate signal.

Analyte Source Detection is also an example of signal processing. Assume there is an analyte, which is a chemical secretion in a cellular grid. We want to know "where" the chemical is originating from. Intuitively, we can see that if a chemical is secreted from a point, its concentration is "highest" in the region around the center and decreases as we move away from the origin.



Source S (say HSL) is recognized by 4 Colored Reporters: BFP, GFP, YFP and RFP.
 BFP: $S_{conc} (1 - 0.8)$
 GFP: $S_{conc} (0.8 - 0.7)$
 YFP: $S_{conc} (0.7 - 0.5)$
 RFP: $S_{conc} (< 0.5)$

Analyte source detection



In the above circuit:

- Analyte Detection Component: Detects HSL presence and transcribes mRNA_{XY} to Proteins X and Y
- Low Threshold Component: Upon *high* HSL and *high* X input, Z gets suppressed.

- High Threshold Component: Upon *high* HSL, *high* Y and *low* W input, high Z O/P obtained
- Negating Component: The net difference of O/P concentrations of Z from Low and High Threshold components eventually determines the net concentration of Z and GFP.

7 Conclusions

The goal of this lecture was to illustrate an approach for creating synthetic gene networks for modifying and extending the behavior of living organisms. Researchers have characterized and assembled a genetic component library. There has also been successful implementation of prototype circuit. The primary challenge is to devise models and perform simulations that can accurately predict outcome of genetic networks.