Proposal for Seed Grant for Collaborative Research Between NYU-Poly and NYU

Building Reliable Genetic Devices using Unreliable Ones

Elza Erkip (NYU-Poly), Dennis Shasha (Courant, NYU)

Start Date: June 1, 2013

Abstract

Synthetic biology attempts to build new biological systems by combining biological research and engineering. Progress in engineering disciplines critically depends on successful abstraction and standardization of complex devices and components. Similarly, synthetic biology has focused on the use of *BioBricks*, which are standard biological parts that can be combined together resulting in engineered devices that work within the cell to build smarter drugs and alternative treatments. However, as in other engineering disciplines, genetic devices are not always reliable and may fail in a stochastic fashion. This project addresses reliability of genetic devices using tools from electrical engineering. In contrast with the usual approach of trying to build more reliable genetic parts and devices, the proposed research recognizes that simple unreliable devices, which are available in abundance, can be potentially combined in *networks* to result in reliable end-to-end operation. Using the PIs' expertise in information theory and biological circuits, the proposed project builds the theoretical foundations and implementation details of such networks, potentially leading to a new paradigm in synthetic biology.

1 Overview

Synthetic biology is an engineering discipline that attempts to create devices that do not normally exist in nature [1, 2, 3]. Standardization of genetic parts, devices and systems, known as *BioBricks* [3, 4, 5], has made a considerable impact on the research of engineered circuits that work within the cell, resulting in applications such as smarter drugs that respond to external conditions adaptively, biofactories that produce cheaper organic materials and alternative cancer treatments [2, 6, 7, 8]. Among the many design methodologies for genetic circuits, a very popular one uses analogy with electrical circuits, particularly digital ones, which are easy to design, robust to cumulative errors and noise.

As in the case of electrical devices, genetic ones are not fully reliable and may fail in a stochastic fashion. For example, the experimental studies reported in [9, 10] show that for a large number of genetic switches working in parallel, the output exhibits a bimodal distribution due to cell-to-cell variations and the stochastic nature of biological systems. This effect can accumulate in larger networks, and can be detrimental in applications such as drugs and cancer treatments [7, 9] where being able to turn on or off a given set of genes accurately with respect to the concentration of single or multiple biochemical factors is crucial.

In this project, we propose to investigate novel ways of building reliable genetic devices. Consider for example the problem of detecting the presence of some combination of manmade toxic chemicals. Suppose we would like to determine whether proteins A and B are present together or proteins A and C are present together. Using the Biobricks setup, we can create *NOT* and *NOR* logic gates [11] whose promoters may be affected the presence of these proteins. Such a circuit might be expressed algebraically as NOR(NOT(A), NOR(B, C)), where the final NOR is attached to a green fluorescent protein (GFP), whose presence indicates an affirmative answer to the original question. Achieving greater reliability would ensure that the GFP would appear if and only if the proper combination of inputs appeared. More ambitious would be to deliver drugs according to a schedule if this triggering event occurred. To achieve time release, we would use a clock design based on a so-called repressilator [12]. At each clock cycle a certain amount of drug would be delivered.

One of the simplest components of the above system is a *genetic switch*, which results in a desired level of GFP based on the output of the final NOR gate. A genetic switch in general is a transcriptional network where the input is the concentration of a transcription factor or a regulator (for example a repressor or an activator), and the output has two distinct levels of some other protein concentration [3]. One of the best-known genetic switches is the λ -Phage switch [13], which manages a phage to enter into the lytic or the lysogenic cycles. Multimerization of the transcription factors and colloborative binding result in more reliable switches, by simply binding two repressors whose effect is larger than the sum of individual effects [1, 13, 14]. The efficiency of such systems is shown in Figure 1, by comparing this switch with a single repressor-operator system.

Most research on the reliability of genetic switches either re-engineers proteins and other regulatory elements or manipulates the network architecture among simple genetic parts [3, 9]. Our proposed approach, on the other hand, is to *design more reliable genetic switches by building networks of unreliable ones*. The motivation is that unreliable switches are easier and cheaper to construct and thus are available in larger quantities. Using the analogy with electrical circuits, we will investigate the approach of Shannon [15] where he proved that *arbitrarily* reliable relay[1]

[1] A type of electrical switch.

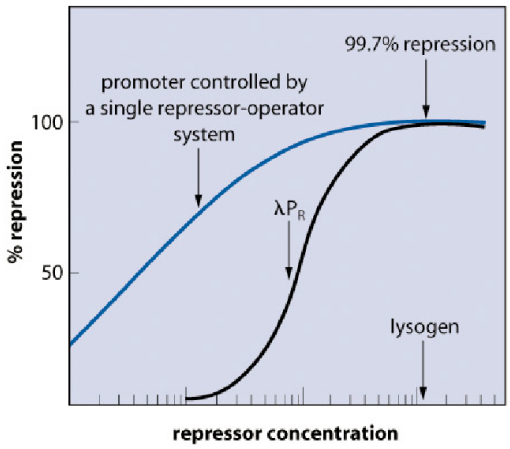

Figure 1: *Repression as a function of repressor concentration in two systems [13]. The black line represents a more reliable (steeper) system.*

operation can be accomplished by proper interconnection of a sufficient number of unreliable relays. We first briefly describe Shannon's framework before summarizing our proposed work.

2 Relay Networks

Shannon considered the scenario in which one has access to a large number of unreliable relays, and needs to build a single reliable one [15]. To model a relay, he viewed it as a channel with binary input X, and binary output, Y. He also defined the transition probabilities between the input and outputs as; $P(Y=1|X=1)=\alpha$, $P(Y=1|X=0)=\beta$ as shown in Figure 2. For a reliable relay, we need α to be close to 1, and β to be close to 0. Shannon defined the reliability function $h(p)$ as the probability that the output is 1 for a network of relays, where $p=P(Y=1)$ for an individual relay. Note that with this definition, a network serves as a more reliable relay than the original one if $h(p)\leq p$ for $p\leq p^*$ and $h(p)\geq p$ for $p\geq p^*$ for some $p^*\in(\beta,\alpha)$ with $h(p^*)=p^*$, simply indicating that by forming the network, the effective α has increased, while the effective β has decreased. This is illustrated in Figure 2. Using this approach, Shannon analyzed the interconnection of different networks (or network of networks) and proposed a design methodology leading to an arbitrarily reliable relay given enough number of unreliable relays.

3 Proposed Work

We first describe our proposed work within the context of a genetic switch, then discuss extensions to more complex components and devices. Our goal is to study whether a Shannon-type approach is applicable to building reliable genetic devices. We start with a mathematical formulation which will then prescribe practical implementation methodologies.

The first step is to mathematically model both the genetic switch and the interaction of multiple switches. Consider the following simple probabilistic model of a switch: The input to the switch is taken as the concentration of an repressor, the output of the switch is the concentration of another protein, for example a GFP as above. We let $X=1$ to denote *low* input repressor concentration, $X=0$ *high* concentration. Similarly, $Y=1$ denotes *high* output GFP concentration, while $Y=0$

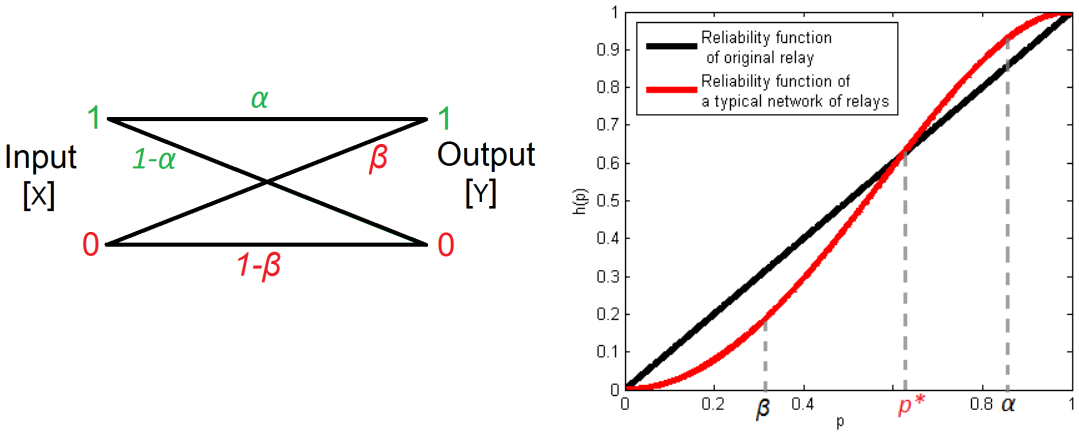

Figure 2: *Left: Representation of a relay. Right: Typical reliability functions.*

is *low* GFP concentration. Letting α be the probability that the output GFP concentration is *high* ($Y = 1$) while the input concentration is *low* ($X = 1$), and β be the probability that an output GFP concentration is *high* ($Y = 1$) when the input concentration is *high* ($X = 0$) we obtain the same model as in the relay scenario of Figure 2. In this case having a reliable genetic switch means α is close to 1, and β is close to 0, which is equivalent to having an ideal *inverter*.

With this simple genetic switch model, the next step would be to obtain a "calculus" of genetic switches that describes the input-output relationships of various switch connections as a function of the type of connection, as done by Shannon for relays, where three types of connections, *series*, *parallel*, and *composite*, were considered [16]. Once such models are obtained, one could modify Shannon's analysis for transcriptional switch networks and investigate general properties of such networks that form an effective reliable switch. An important question would be how physical restrictions on the size of the network would affect the overall switch reliability.

While this approach is promising, it has several shortcomings that need to be addressed for a more accurate model. Assuming constant α and β ignores self regulation of the genes and modeling transcription/translation chain with a binary channel may be an oversimplification. Furthermore, genetic switches do not work independently, unless the network is small [17].

The above theoretical framework promises successful design methodologies as in the case of Shannon's relay networks, but in order to impact actual genetic devices, the developed theory should be tested and refined using simulations and implementation. For the simulation phase, we will make use of the various genetic simulation softwares (see [2] for a list of simulators). This will be followed by an implementation of the engineered genetic switch networks using parts from MIT Standard Registy [11, 18]. For the implementation stage we plan to establish cooperation with a molecular biology lab. Finally, the above network approach should be extended to build other reliable genetic devices such as oscillators, counters and detectors [2, 3], for which we will need to address the concepts of delay and feedback and include cyclic connections.

We expect that our approach, if successful, will provide a new paradigm for building reliable genetic devices and circuits such as the one in the toxic chemical detection example. It will also allow a new direction for information theory research, reinforcing recent emphasis on biological applications of information theory [19].

4 Project Team and Plans for External Support

The project team consists of Elza Erkip (NYU-Poly), Dennis Shasha (NYU, Courant), and their students who will be supported by this grant. Elza Erkip is an expert in information theory; she has won several awards for her work in network information theory and wireless communications. Dennis Shasha is a computer scientist who works closely with biologists to discover circuits of genetic causality from gene expression experiments and works on clocked circuits for DNA computation. Two PhD students (Anil Kocak and Farideh Rezagah) from NYU-Poly will be involved in this project. Anil carried out some of the initial work leading to this proposal. During Summer and Fall 2013, he will collaborate with Farideh (who is a more senior PhD student with extensive experience in information theory) in establishing a theoretical framework for genetic device networks and in carrying out simple simulations. During Spring 2014, Anil will continue to work on the theoretical aspects and will also initiate collaboration with two Masters students at NYU on the simulation and the implementation phases. Erkip and Shasha will be involved in the project at every step, ensuring accuracy and viability of the genetic and information theoretic models.

During the first year of the project, we plan to seek for external funding through NSF. This seed grant will allow us to devote time and resources (as detailed above) to obtain an initial set of results. Note that this is a completely new direction of research, so we will need to show some initial success to convince the grant agencies. We expect to complete some preliminary theoretical work within six months to submit a full proposal to NSF's Communication and Information Foundations (CIF) program which funds fundamental research in information theory (Deadline December 17, 2013). Specifically we will start with simple abstract models and illustrate feasibility by obtaining some theoretical results and by providing an initial set of simulations. The next six months of seed funding will allow us to refine the models, the analysis and simulations until external funding is obtained. Given the excitement in the information theory community in investigating biological applications [19], we expect significant interest from NSF's CIF program. Our longer term goal is to move the research in a more practical direction and attract funding from other agencies such as NIH.

References

[1] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*, volume 10. Chapman & Hall/CRC, 2006.

[2] W. Weber and M. Fussenegger. *Synthetic Gene Networks, Methods and Protocols*. Springer Protocols, Humana Press, 2012.

[3] Baldwin, Bayer, Dickinson, Ellis, Freemont, Kitney, Polizzi, and Stan. *Synthetic Biology - A Primer*. Imperial College Press, 2012.

[4] Reshma Shetty, Drew Endy, and Thomas Knight. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, 2(1), 2008.

[5] D. Endy. Foundations for engineering biology. *Nature*, 438(7067):449–453, 2005.

[6] E. Young and H. Alper. Synthetic biology: Tools to design, build, and optimize cellular processes. *Journal of Biomedicine and Biotechnology*, 2010, 2010.

[7] Z. Xie, L. Wroblewska, L. Prochazka, R. Weiss, and Y. Benenson. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science Signalling*, 333(6047):1307, 2011.

[8] J.W.H. Li and J.C. Vederas. Drug discovery and natural products: End of an era or an endless frontier? *Science*, 325(5937):161–165, 2009.

[9] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of a genetic toggle switch in escherichia coli. *Nature*, 403:339–342, 2000.

[10] T. Tian and K. Burrage. Stochastic models for regulatory networks of the genetic toggle switch. *Proceedings of the National Academy of Sciences*, 103(22):8372–8377, 2006.

[11] M. Galdzicki, C. Rodriguez, D. Chandran, H.M. Sauro, and J.H. Gennari. Standard biological parts knowledge base. *PloS one*, 6(2):e17005, 2011.

[12] J. Stricker, S. Cookson, M.R. Bennett, W.H. Mather, L.S. Tsimring, and J. Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.

[13] M. Ptashne and A.G. Switch. Phage lambda and higher organisms. *Cell Press & Blackwell Science*, 1992.

[14] M. Ptashne and A. Gann. *Genes & Signals*. CSHL Press, 2002.

[15] E.F. Moore and C.E. Shannon. Reliable circuits using less reliable relays I-II. *Journal of the Franklin Institute*, 262(3):191–208, 1956.

[16] C.E. Shannon. A symbolic analysis of relay and switching circuits. *Transactions of the American Institute of Electrical Engineers*, 57(3):713–723, 1938.

[17] B. Wang, R.I. Kitney, N. Joly, and M. Buck. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2:508, 2011.

[18] Registry of standard biological parts. `http://partsregistry.org`.

[19] O. Milenkovic at. al. (editors). *IEEE Transactions on Information Theory, Special Issue on Information Theory in Molecular Biology and Neuroscience*, 2010.