AFRL-AFOSR-VA-TR-2020-0190

Theory-based Engineering of Biomolecular Circuits in Living Cells

Domitilla Del Vecchio
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

09/14/2020
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ RTA2
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
14-09-2020	Final Performance	01 Mar 2019 to 28 Feb 2020

4. TITLE AND SUBTITLE
Theory-based Engineering of Biomolecular Circuits in Living Cells

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-14-1-0060

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Domitilla Del Vecchio

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
77 Massachusetts Avenue, Cambridge, MA, USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2020-0190

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The objective of this research is to establish a data-driven theoretical framework based on mathematics to enable the robust design of interacting biomolecular circuits in living cells that perform complex decision-making. Microbiology as a platform has substantial advantages with respect to human-made hardware, including size, power, and high sensitivity/selectivity. While the latest advances in synthetic biology have rendered the creation of simple functional circuits in microbes possible, our ability of composing circuits that behave as expected is still missing. This hinders the possibility of designing robust complex decision-making circuits, such as those that recognize and classify chemical signatures and those that program degradation of pre-specied materials upon contact. Overcoming this bottleneck goes beyond the engineering of new parts or new assembly methods. By contrast, it requires a deep understanding of the dynamical interactionsamong synthetic modules and the cell machinery, a particularly hard task since dynamics arenonlinear, stochastic, and involve multiple scales of resolution both in time and space.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			LEVE, FREDERICK

Unclassified	Unclassified	Unclassified	UU	**PAGES**	**19b. TELEPHONE NUMBER** *(Include area code)*

Theory-Based Engineering of Biomolecular Circuits in Living Cells (Grant Number: FA9550-14-1-0060)

Domitilla Del Vecchio (PI), James J. Collins (co-PI), Richard M. Murray (co-PI), and Eduardo D. Sontag (co-PI)

Reporting period (NCE): March 1, 2019-February 28, 2020

Abstract. The objective of this research is to establish a *data-driven theoretical framework* based on mathematics to enable the robust design of interacting biomolecular circuits in living cells that perform complex decision-making. Microbiology as a platform has substantial advantages with respect to human-made hardware, including size, power, and high sensitivity/selectivity. While the latest advances in synthetic biology have rendered the creation of simple functional circuits in microbes possible, our ability of composing circuits that behave as expected is still missing. This hinders the possibility of designing robust complex decision-making circuits, such as those that recognize and classify chemical signatures and those that program degradation of pre-specified materials upon contact. Overcoming this bottleneck goes beyond the engineering of new parts or new assembly methods. By contrast, it requires a deep understanding of the dynamical interactions among synthetic modules and the cell machinery, a particularly hard task since dynamics are nonlinear, stochastic, and involve multiple scales of resolution both in time and space.

In this project, we propose an interdisciplinary approach merging mathematics, control and dynamical systems theory, electrical circuit theory, and synthetic biology, in order to tackle this problem. We propose to establish a design-oriented theoretical framework that explicitly accounts for interactions among circuits, between the circuits and the cell machinery, and provides engineering solutions to mitigate the undesirable effects of these interactions (compositionality effort). Within this framework, we will develop mathematical tools to quantify the propagation of stochasticity through the nonlinear dynamics of biological networks (stochasticity effort) and to incorporate spatial heterogeneity effects (spatial heterogeneity effort). These research efforts will be focused on solving concrete engineering problems on a prototype experimental system that integrates different sensors to classify a number of chemical signatures.

1 Introduction and summary of achievements in past year

The advancements of the past decade in DNA recombinant technology and measurement techniques have paved the way for the rising field of synthetic biology [1, 2]. Today, it is possible to fabricate simple circuits in living cells to implement a variety of functions, such as counter circuits [3], circuits with memory [4], and oscillators [5–7]. This leads the way to a number of exciting applications, including biosensing, classification and recognition of the environment's chemical composition, and programmed degradation of materials. Microbiology as a platform has substantial advantages with respect to human-made hardware, including size (devices are practically invisible), power (cells self-power taking nutrients from the environment), and high sensitivity/selectivity (nano-molar range). Complex decision making tasks, such as classification and recognition, require circuits

with many interacting components working in a predictable fashion. For example, classification or recognition of the environment chemical composition would require networks of biosensors in a common *E. coli* strain. These biosensors may be distributed across different cells, each of which sample the environment and provide a distributional response to allow fast classification of multi-species chemical signatures.

Unfortunately, our ability to compose circuits that behave as expected once interacting in the cellular environment is missing. The reality is that circuits, once integrated in the cell, do not perform any more as they would in isolation, and the sources of this problem are poorly understood. This problem is broadly referred to as *context dependence* [8] and has to do with the fact that parts and functional modules do not maintain their behavior in isolation when connected in larger systems and the cellular environment changes (see [9, 10] for recent reviews on this problem). This hinders the possibility of designing robust complex decision making devices, and limits the current spectrum of applications for synthetic biology, including biosensing, recognition, classification, and bioremediation. The origins of this problem cannot be understood only from the characterization of physical properties of parts, such as genes, promoter sequences, ribosome binding sites, proteins, etc., an area of focus in synthetic biology. Instead, tracing the roots of this problem requires a deep understanding of the dynamic interactions among synthetic parts and the cell machinery, which is particularly hard since dynamics are nonlinear, *stochastic*, and they involve *multiple scales of resolution both in time and space*. Control theory-based approaches have had initial and promising success in identifying some specific sources of this lack of modularity of genetic circuits and in finding effective engineering solutions that have been recently experimentally realized [11–13].

We believe that only a data-driven theory-based effort can tackle this problem. Therefore, we have assembled an interdisciplinary team that brings together expertise from synthetic biology, control and dynamical systems theory, electrical circuit theory, and mathematics. Our specific objectives are:

> **Objective 1:** (context-dependence) Establish a design-oriented theoretical framework that explicitly accounts for unwanted interactions among circuits, between the circuits and the cell machinery, and provides engineering solutions to mitigate the undesirable effects of these interactions. Overall, this will lead to synthetic genetic circuits that are robust to changes in their context.
>
> **Objective 2:** (stochasticity) Develop design-oriented analysis tools to quantify the effects of stochasticity on the nonlinear dynamics of biological networks.
>
> **Objective 3:** (spatial dynamics) Develop a quantitative methodology to incorporate spatial heterogeneity effects into the analysis and design framework.
>
> **Application:** (biosensing) Combine multiple sensors to develop robust and predictable chemical signature classifiers.

The ultimate outcome of this effort will be a concrete ability to design circuits that behave more robustly and predictably once interacting with each other in the cell environment, in the presence of stochastic effects and spatial dynamics, thus relieving a major bottleneck in synthetic biology. As a result, we will be able to apply synthetic biology for the creation of robust and more complex decision making tasks, including classification and recognition of biomarkers.

In the past year, we used the NCE to complete and submit several manuscripts. Major activities are described here.

Since mid 2019, Eduardo Sontag and his lab submitted final versions of various papers supported by this grant, and various papers submitted earlier were published, some of which in collaboration with Del Vecchio. These papers covered issues in control-theoretic aspects of synthetic biology as well as applications of the systems biology and mathematical tools developed during this research to areas outside synthetic biology. Among other results, we obtained an in vitro implementation of integral feedback, we characterized both the stochastic and deterministic dynamic landscapes of epigenetic regulation, we studied the dynamics of translation under various time-varying inputs, and we provided certificates of stability and safety for a large class of chemical interaction networks. Sontag gave lectures at various conferences, with acknowledgement to this grant, including the Opening Plenary at the 2020 European Control Conference (May 2020) and the closing keynote at a Harvard Workshop on Systems Biology (Nov 2019).

Related publications:

D.K. Agrawal, R. Marshall, V. Noireaux, and E.D. Sontag. In vitro implementation of robust gene regulation in a synthetic biomolecular integral controller. Nature Communications, 10:1-12, 2019.

M.A. Al-Radhawi and E.D. Sontag. Analysis of a reduced model of epithelial-mesenchymal fate determination in cancer metastasis as a singularly-perturbed monotone system. In C.A. Beattie, P. Benner, M. Embree, S. Gugercin, and S. Lefteriu, editors, Realization and Model Reduction of Dynamical Systems. Springer-Verlag, 2020. To appear. See preprint in arXiv:1910.11311.

S. Bruno, M.A. Al-Radhawi, E.D. Sontag, and D. Del Vecchio. Stochastic analysis of genetic feedback controllers to reprogram a pluripotency gene regulatory network. In Proc. 2019 Automatic Control Conference, 2019, pp. 5089-5096.

D. K. Agrawal, R. Marshall, M. Ali Al-Radhawi, V. Noireaux, and E. D. Sontag. Some remarks on robust gene regulation in a biomolecular integral controller. In Proc. 2019 IEEE Conf. Decision and Control, 2019, pp. 2820-2825

M.A. Al-Radhawi, D. Angeli, and E.D. Sontag. A computational framework for a Lyapunov-enabled analysis of biochemical reaction networks. PLoS Computational Biology, pp 16(2): e1007681.

E.D. Sontag. Scale-invariance in biological sensing. In J. Baillieul and T. Samad, editors, Encyclopedia of Systems and Control. Springer-Verlag, 2020.

M. Margaliot and E.D. Sontag. Revisiting totally positive differential systems: A tutorial and new results. Automatica, 101:1-14, 2019.

M. Sadeghi, M.A. Al-Radhawi, M. Margaliot, and E.D. Sontag. No switching policy is optimal for a positive linear system with a bottleneck entrance. IEEE Control Systems Letters, 3:889-894, 2019.

E.D. Sontag. Scale-invariance in biological sensing. In J. Baillieul and T. Samad, editors, Encyclopedia of Systems and Control. Springer-Verlag, 2020.

Murray's team focused with student Andy Halleran on looking at random partitioning effects.

Specifically, plasmids are found across bacteria, archaea, and eukaryotes and play an important role in evolution. Plasmids exist at different copy numbers, the number of copies of the plasmid per cell, ranging from a single plasmid per cell to hundreds of plasmids per cell. This feature of a copy number greater than one can lead to a population of plasmids within a single cell that are not identical clones of one another, but rather have individual mutations that make a given plasmid unique. During cell division, this population of plasmids is partitioned into the two daughter cells, resulting in a random distribution of different plasmid variants in each daughter. We used stochastic simulations to investigate how random plasmid partitioning compares to a perfect partitioning model. Our simulation results demonstrate that random plasmid partitioning accelerates mutant allele fixation when the allele is beneficial and the selection is in an additive or recessive regime where increasing the copy number of the beneficial allele results in additional benefit for the host. This effect does not depend on the size of the benefit conferred or the mutation rate, but is magnified by increasing plasmid copy number.

Del Vecchio's group focused on submitting Carlos Barajas work on PDE models of genetic circuit dynamics within bacteria. Specifically, intracellular spatial heterogeneity is frequently observed in bacteria, where the chromosome occupies part of the cell's volume and a circuit's DNA often localizes within the cell. How this heterogeneity affects core processes and genetic circuits is still poorly understood. In fact, commonly used ordinary differential equation (ODE) models of genetic circuits assume a well-mixed ensemble of molecules and, as such, do not capture spatial aspects. Reaction-diffusion partial differential equation (PDE) models have been only occasionally used since they are difficult to integrate and do not provide mechanistic understanding of the effects of spatial heterogeneity. In this paper, we derive a reduced ODE model that captures spatial effects, yet has the same dimension as commonly used well-mixed models. In particular, the only difference with respect to a well-mixed ODE model is that the association rate constant of binding reactions is multiplied by a coefficient, which we refer to as the binding correction factor (BCF). The BCF depends on the size of interacting molecules and on their location when fixed in space and it is equal to unity in a well-mixed ODE model. The BCF can be used to investigate how spatial heterogeneity affects the behavior of core processes and genetic circuits. Specifically, our reduced model indicates that transcription and its regulation are more effective for genes located at the cell poles than for genes located on the chromosome. The extent of these effects depends on the value of the BCF, which we found to be close to unity. For translation, the value of the BCF is always greater than unity, it increases with mRNA size, and, with biologically relevant parameters, is substantially larger than unity. Our model has broad validity, has the same dimension as a well-mixed model, yet it incorporates spatial heterogeneity. This simple-to-use model can be used to both analyze and design genetic circuits while accounting for spatial intracellular effects.

Additional work was carried to complete and submit the works by Ukjin Kwan on stochastic model reduction. Specifically, the Chemical Master Equation (CME) is commonly used to describe the stochastic behavior of biomolecular systems. However, in general, the CME's dimension is very large or infinite, so analytical or even numerical solutions may be difficult to achieve. The truncation methods such as the Finite State Projection (FSP) algorithm alleviate this issue to some extent but not completely. To further resolve such a computational burden, we propose the Enhanced Finite State Projection (EFSP) algorithm, in which the ubiquitous time-scale separation is utilized to reduce the dimension of the CME. Our approach combines the original FSP algorithm and the model reduction technique that we developed, to approximate an infinite dimensional CME with a finite dimensional CME that contains the slow species only. Unlike other time-scale separation methods, which rely on the fast-species counts' stationary conditional probability distributions,

our model reduction technique relies on only the first few conditional moments of the fast-species counts. This is possible because we apply conditional moment closure to close the fast-species counts' dynamics. In addition, each iteration of the EFSP algorithm relies on the solution of the approximated CME that contains the slow species only, unlike the original FSP algorithm relies on the solution of the full CME. These two properties provide a significant computation advantage. The benefit of our algorithm is illustrated through a protein binding reaction example.

Related publications:

C. Barajas and D. Del Vecchio. Genetic Circuit-Host Ribosome Transactions: Diffusion-Reaction Model. Proc of American Control Conference, 2019

U. Kwon, M. Naghnaeian and D. Del Vecchio. Approximation of the Chemical Master Equation using conditional moment closure and time-scale separation. Proc of American Control Conference, 2019

D. Del Vecchio. Synthetic Biology. In Encyclopedia of Systems and Control, Baillieul, John, Samad, Tariq (Eds.), Springer March 2015 (Updated October 2019)

U. Kwan, M. Naghnaeian, and D. Del Vecchio. The Enhanced Finite State Projection algorithm using conditional moment closure and time-scale separation. Proc. IEEE Conf. on Decision and Control, 2020

C. Barajas and D. Del Vecchio. Effects of spatial heterogeneity on bacterial genetic circuits, PLOS Comput. Biol. Accepted, July 2020

Finally, with student Aaron Dy, Del Vecchio and Collins completed and submitted a journal article on using layered logic gates to detect viruses and bacteria using cell-free systems. Specifically, rapid and inexpensive detection of multiple nucleic acid biomarkers may enable point-of-care testing for panels of pathogens, monitoring of complex diseases, and assessing phenotypic information. Paper-based synthetic biology diagnostics have been shown to provide fast and affordable detection of nucleic acid targets, but these diagnostics have not yet combined sensors to perform flexible multiplexed detection. Here, we show that engineered biomolecular networks can integrate RNA toehold switch sensors to detect multiple nucleic acid sequences and provide a single readout. We demonstrate OR-gate logic with human papillomavirus subtype detection, four types of two-input circuits with sets of split proteins, two-input circuits for inflammatory mRNAs that produce a visual output, three- or four-input circuits using combinations of split protein outputs, and a three-input classifier for antibiotic susceptibility. Our results demonstrate that engineered biomolecular networks can process multiple sensor inputs to reduce cost, improve specificity, and build classifier circuits.

Related publications:

Aaron Dy. Cell-free synthetic biology for affordable, on-demand diagnostics. PhD Thesis, MIT August 2019

A. J. Dy, M. C. Kline, D. Del Vecchio, and J. J. Collins. Multi-input biomolecular circuits for smarter paper-based diagnostics. Submitted, 2020.

References

[1] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss. Synthetic biology: New engineering rules for an emerging discipline. *Molecular Systems Biology*, pages 1–14, 2006.

[2] D. Baker, G. Church, J. Collins, D. Endy, J. Jacobson, J. Keasling, P. Modrich, C. Smolke, and R. Weiss. ENGINEERING LIFE: Building a FAB for biology. *Scientific American*, pages 44–51, June 2006.

[3] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins. Synthetic gene networks that count. *Science*, 324:1199–1202, 2009.

[4] T.S. Gardner, C.R. Cantor, and J.J. Collins. Construction of the genetic toggle switch in *Escherichia Coli*. *Nature*, 403:339–342, 2000.

[5] M. B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:339–342, 2000.

[6] M. R. Atkinson, M. A. Savageau, J. T. Meyers, and A. J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 113:597–607, 2003.

[7] 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*, 456:516–519, 2008.

[8] S. Cardinale and A. P. Arkin. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, 7:856–866, 2012.

[9] D. Del Vecchio. Modularity, context-dependence, and insulation in engineered biological circuits. *Trends in Biotech (Cell Press)*, 33:111–119, 2015.

[10] D. Del Vecchio, Y. Qian, and A. Dy. Control theory meets synthetic biology. *J. Royal Society Interface*, 2016.

[11] S. Jayanthi and D. Del Vecchio. Retroactivity attenuation in bio-molecular systems based on timescale separation. *IEEE Trans. Aut. Control*, 56:748–761, 2011.

[12] D. Mishra, P. M. Rivera-Ortiz, A. Lin, D. Del Vecchio, and R. Weiss. A load driver device to engineer modularity in biological circuits. *Nat. Biotechnology, To Appear*, 2014.

[13] Kayzad Soli Nilgiriwala, José I. Jiménez, Phillip Michael Rivera, and Domitilla Del Vecchio. Synthetic tunable amplifying buffer circuit in *E. coli*. *ACS Synth. Biol.*, 4(5):577–584, 2015.