Ph.D. Proposal

Pedro Fontanarrosa, UID: 01143732

January 2020

Specific Aims

Write about the proposed work touching the research gap you are [1].

1 Automatic model generation and analysis of genetic circuits

generalizing modeling generation

The iBioSim tool [2], developed by Chris's Myers research group, is a computer-aided design (CAD) tool for modeling, analysis, and design of *genetic regulatory networks* (GRNs). This tool can, amongst other things, automatically generate models for GRNs and obtain predictions from their simulation. However, there are a plethora of different analysis methods that can be implemented to further expand what iBioSim is capable of and generalize model generation. Some of these improvements that can be implemented are:

- **Add road-blocking effects**

  Interference between two tandem promoters with the same orientation in the expression of a gene is a known phenomenon [3]. Mathematical modeling methods have been described to describe this roadblocking effect of tandem promoters [4] and could be accounted for in the automatic model generator in iBioSim.

- **Parameter estimation & optimization**

Systems biology has developed new *-omics* tools that offer the potential to take a "snapshot" of the inner workings of a circuit in a single experiment [5]. Some very popular transcriptomic methods of these are *RNA sequencing* (RNA-seq), which enables one to quantify the mRNA levels of each gate of a circuit with nucleotide resolution [6], and *Ribosome profiling* (Ribo-seq), which uses specialized mRNA sequencing to determine which mRNAs are being actively translated. These high-throughput experiments allow for a complete analysis of a genetic circuit and its impact on the organism, the transfer functions of each genetic gate or part when not in isolation, maximum and minimum levels of transcription, and many other transcriptome analyses [7], and can be used to estimate model parameters. A parameter estimation process could be instantiated in iBioSim to automatically estimate model parameters using optimization methods.

- **CRISPR/dCas9 circuit modeling**

  Recently, programmable and orthogonal CRISPR/dCas9 transcription factors have been developed [8, 9, 10, 11, 12]. These transcription factors can be used to build genetic circuits with dCas9-mediated repression. The CRISPR/dCas9 transcription factor system can yield orthogonal genetic gates with low variability and that show minimal retroactivity or effects on cell growth [8]. This makes the gates relatively easy to combine into Boolean logic circuits since the different guide-RNA's have very high specificity, and therefore there is no gate interference and/or leakage. Genetic circuits constructed this way are amongst the largest ever built in any organism [8].

  The automatic model generator of iBioSim can easily be adapted to model CRISPR gates and create automatic models for circuits that implement them. Estimation of the parameters for these gates could be done by analysis of results or the automatic estimation of parameters proposed earlier. This would expand the automated model generator of iBioSim to not only work with Cello composed and parameterized gates, but also CRISPR/dCas9 mediated repression gates, and have another method to generate models and simulations for circuits in yeast.

- **Circuit Hazard Identification**

  As of now, to analyze a circuit's function and logic hazards using the dynamic model generator is not an automatic process, and a certain grasp of the topic is needed in order to perform this analysis. To make this analysis available to any user without the need of any understanding

on how to do so, it would prove very useful to implement an analysis [13, 14] that can automatically detect function and logic hazards on a circuit and report them back to the user. This report can also include a list of input changes that may cause a glitching behavior. If a stochastic model of the circuit is available (aim 2), the probability of a glitch can also be reported. This automated analysis can be integrated in the automated model generator in iBioSim, or as a stand-alone application that can be exported to other software tools, like Cello [1].

- **Circuit performance and complexity scores**

  Circuit complexity is measured by the size and depth of a circuit. Complexity scores for electronic circuits essentially depend only on number of gates and can be easily calculated from the truth table or Karnaugh map of a circuit. However, deriving a complexity score for synthetic biology is not as simple since firstly, genetic circuits for a given truth table differ in both gene numbers and quantity of regulation factors and levels, and secondly, circuit performance is not guaranteed due to nonlinear interactions present when the genetic circuit is introduced into a host cell [15].

  Larger genetic circuits have a larger metabolic burden or load on the host cell and are harder/more costly to build [7]. Therefore, a complexity score may help a designer to set an upper limit on the complexity of a circuit if the known limit for the organism intended is known. Deriving or adopting a complexity score suitable for genetic circuits and developing an algorithm that can calculate such a score would be indeed be a valuable asset for the automated model generator.

  Circuit performance can be thought of as the capacity of a circuit to reproduce faithfully or successfully the truth table. There are two essential factors that determine the performance of a circuit, which are the extent that the *high* and *low* output signal of each genetic gate can be practically distinguished, and the transient dynamics after changes in the inputs that may produce incorrect results [15, 16].

  The automated model generator of iBioSim could be expanded to calculate an predicted circuit performance. This metric could be based solely on a statistical analysis of the difference between the predicted steady-state output and experimental results. Marchisio *et al.* [16, 15] propose the difference of the minimal and maximal output at steady-state for each gate as a main parameter to quantify gate and circuit performance. Whichever method is preferred, a report (along with complexity score) could be reported as soon as a model simulation of

3

the circuit is done, so as to help the user in the design process of a genetic circuit.

- **Add model generator for new model from Voigt's lab**

  I mean, since we are doing the work why not right?

2 Develop a dynamic stochastic model for genetic regulatory networks

The deterministic framework of ODE analysis is appropriate to describe the mean behavior of a system, with underlying assumptions such that the variables vary in a deterministic and continuous fashion. However, the stochastic nature of biochemical reactions, even at the single-gene level [17], generates significant intrinsic genetic noise to a system [18]. Furthermore, the underlying assumption of continuous-deterministic models that the number of molecules is high (or that the volume of the system is infinite) can be invalid for biochemical systems where there are very few transcription factors or only one copy DNA segment [19]. Since transcription factors, enzymes, and DNA copies can exist in systems at a low concentration such as a single molecule per cell, any realistic analysis of these systems must include stochastic effects, and therefore stochastic modeling and analysis [20].

In previous work [21], we analyzed function hazards and observed glitch behavior predicted using the deterministic-continuous approach of an ODE model analysis. This represents the average behavior expected for a system. However, with analysis of a stochastic model of the same circuits, we could compute the *probabilities* of each possible glitch and have the automated model generator report these probabilities. In this way, anyone using this model generator for design purposes can have an idea of the risks of this unwanted switching behavior happening and determine if they need to restrict the allowed input changes to the circuit to make certain that the glitching behavior does not occur. Furthermore, knowledge of these glitch probabilities can be used while exploring circuit design spaces with certain applications in mind (aim 4). A logic-synthesis method that uses hazard-preserving optimizations can be implemented as part of the genetic circuit design in iBioSim [2], or as a stand-alone application for use in other *genetic design automation* (GDA) tools like Cello [1]. The designer can determine/state which glitches should be avoided and the logic-synthesis method developed will provide a design that doesn't contain hazards that the user deemed critical. This can

4

be done by delimiting the design process using hazard-preserving optimizations to prevent likely function or logic hazards [22] .

A stochastic model would also allow to implement a sensitivity analysis of the kinetic model [23, 24, 25, 26, 27], which would allow one to systematically study the dependence of the quantities of interest on the parameters that define the model and the initial conditions in which it is simulated. Sensitivity analysis can be used to determine which parameters affect stability the most, and can be used to determine where in the circuit there needs to be feedback circuits, which connects to aim 4.

To be able to produce a stochastic model and analysis on circuits, a translation from RPU units to molecule count has to be developed. This translation would map RPU units to actual molecule count in a cell, and with it, the propensity of each reaction rate. Both of these mappings are needed to enable stochastic modeling and simulation. A stochastic analysis would not only allow for more precise models, but also for other types of analysis, which are described in the next sections. This will entail, amongst other things, the implementation of the following steps:

- Mapping method for RPU units and transcription rates to molecule counts for stochastic analysis

- Development of the automatic model generator of stochastic models for GRNs

- Characterization for stochastic analysis and modeling

- A *parametric sensitivity analysis* (PSA) analysis on GRNs

- Quantification of glitch propensities

- Re-consider synthesis using stochastic modeling

3 Develop consortium logic implementations, modeling and simulation

Microbial consortia can perform more complicated tasks and endure more changeable environments than monocultures can [28]. Moreover, if a genetic circuit is too large for a single organism to handle, the circuit can be subdivided into smaller genetic circuits for each strain of the consortium, since the

division of labor would make the burden of each circuit smaller. This makes synthetic biology for microbial consortium a new and important frontier for synthetic biology.

As biological research on bacteria consortium progresses, a modeling scheme has to be developed to accompany this progress. Mathematical models that describe a consortium of cells are becoming a necessary tool to drive, rather than supplement, design of these systems. From an engineering standpoint, the special considerations for modeling bacterial consortium has to be taken into account when building mathematical models that describe the interactions not only within each cell, but also between the different cells that compose the consortium. This will especially provide insights on unknown interactions (such as resource competition), when the predictions of the model can be compared to experimental results, and the information revealed can help genetic circuit designers to achieve desired symbiotic behaviors.

talk about parallel computing?

There are several attempts in the systems biology research discipline to develop models that describe the dynamic behavior of gut microbia in [29, 30, 31, 32, 33], to cite a few. However, modeling and analysis of well-defined synthetic genetic networks are more amenable to detailed mathematical modeling and analysis [34]. An automated model generator for consortium genetic circuits could be implemented in iBioSim using existing mathematical model for synthetic symbiotic ecosystems, which would enable to automatically create models for these systems without extensive knowledge on the topic.

Furthermore, we could implement a logic synthesis method in iBioSim for consortium circuits, to provide with an option to automatically design consortium synthetic circuits from a user-specified expected behavior. This method would split a circuit into independent ones for each organism in a consortium using the hazard-preserving method implemented in aim 2.

Paper idea? Avoiding glitches might even be easier in consortium synthetic biology. Since for **inter**-cellular interactions only the *average* behavior is important, glitches won't be as prominent. The designer only has to prevent function and logic hazards for **intra**-cellular interactions. So the more distributed a circuit is between modular circuits for a different organism, the glitching behavior is a problem.

6

This has to be logic-hazard free preserving. This can even be more logic (and function) hazard free as there can be organisms that hold memory (logic hazard) and change one input at a time (function hazard).
function or logic hazards won't matter for inter-cellular signals, because in the environment only the average matters. However, for intra-cellular signals there has to be logic and function preventing mechanisms.

4   Design genetic circuits with different applications in mind

Use the product of aims 1, 2, and 3 to help other researchers design novel genetic circuits. This means, spark collaboration with other research laboratories and ask about their specific design problems and use everything developed through the course of this Ph.D. to tackle those design goals. This aim is intended not only for designing novel circuits, but also re-designing circuits with known design problems or short-comings (like glitching behavior).

Delay circuit? Universal epigenetic circuit? Maybe researching feedback circuits, feedback loops and other stability circuits.

A *very* specific aim list:

- Stochastic modeling

- Hazard-free logic synthesis

- Consortium modeling

- Consortium circuit design automation

- Design circuits with specific applications in mind

References

[1] A. A. K. Nielsen, B. S. Der, J. Shin, P. Vaidyanathan, V. Paralanov, E. A. Strychalski, D. Ross, D. Densmore, and C. A. Voigt, "Genetic circuit design automation," *Science*, vol. 352, no. 6281, aac7341, 2016.

[2] L. Watanabe, T. Nguyen, M. Zhang, Z. Zundel, Z. Zhang, C. Madsen, N. Roehner, and C. Myers, "iBioSim 3: A tool for model-based genetic circuit design," *ACS Synth. Biol.*, 2018.

[3] A. Tamsir, J. J. Tabor, and C. A. Voigt, "Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'," *Nature*, vol. 469, no. 7329, pp. 212–215, 2011.

[4] K. Sneppen, I. B. Dodd, K. E. Shearwin, A. C. Palmer, R. A. Schubert, B. P. Callen, and J. B. Egan, "A Mathematical Model for Transcriptional Interference by RNA Polymerase Traffic in Escherichia coli," *Journal of Molecular Biology*, vol. 346, no. 2, pp. 399–409, 2005, 00092.

[5] T. E. Gorochowski, A. E. Borujeni, Y. Park, A. A. Nielsen, J. Zhang, B. S. Der, D. B. Gordon, and C. A. Voigt, "Genetic circuit characterization and debugging using RNA-seq," *Mol. Syst. Biol.*, vol. 13, no. 11, p. 952, 2017.

[6] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: A revolutionary tool for transcriptomics," *Nat. Rev. Genet.*, vol. 10, no. 1, pp. 57–63, 2009.

[7] Y. Xiang, N. Dalchau, and B. Wang, "Scaling up genetic circuit design for cellular computing: Advances and prospects," *Nat. Comput.*, vol. 17, no. 4, pp. 833–853, 2018.

[8] M. W. Gander, J. D. Vrana, W. E. Voje, J. M. Carothers, and E. Klavins, "Digital logic circuits in yeast with CRISPR-dCas9 NOR gates," *Nat. Commun.*, vol. 8, p. 15 459, 2017.

[9] A. Didovyk, B. Borek, L. Tsimring, and J. Hasty, "Transcriptional regulation with CRISPR-Cas9: Principles, advances, and applications," *Curr. Opin. Biotech.*, Tissue, Cell & Pathway Engineering (2016), vol. 40, pp. 177–184, 2016.

[10] L. A. Gilbert, M. H. Larson, L. Morsut, Z. Liu, G. A. Brar, S. E. Torres, N. Stern-Ginossar, O. Brandman, E. H. Whitehead, J. A. Doudna, W. A. Lim, J. S. Weissman, and L. S. Qi, "CRISPR-mediated modular RNA-guided regulation of transcription in Eukaryotes," *Cell*, vol. 154, no. 2, pp. 442–451, 2013.

[11] A. A. Nielsen and C. A. Voigt, "Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks," *Mol. Syst. Biol.*, vol. 10, no. 11, p. 763, 2014.

[12] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim, "Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression," *Cell*, vol. 152, no. 5, pp. 1173–1183, 2013.

[13] J. A. Brzozowski, Z. Esik, and Y. Iland, "Algebras for hazard detection," in *Proceedings 31st IEEE International Symposium on Multiple-Valued Logic*, 2001, pp. 3–12.

[14] J. Brzozowski and Z. Ésik, "Hazard Algebras," *Formal Methods in System Design*, vol. 23, no. 3, pp. 223–256, 2003.

[15] M. A. Marchisio and J. Stelling, "Automatic design of digital synthetic gene circuits," *PLoS Comput. Biol.*, vol. 7, no. 2, e1001083, 2011.

[16] M. A. Marchisio, "Parts & pools: A framework for modular design of synthetic gene circuits," *Front. Bioeng. Biotechnol.*, vol. 2, 2014.

[17] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.

[18] P. E. M. Purnick and R. Weiss, "The second wave of synthetic biology: From modules to systems," *Nat. Rev. Mol. Cell Biol.*, vol. 10, no. 6, pp. 410–422, 2009.

[19] Y. N. Kaznessis, "Models for synthetic biology," *BMC Syst. Biol.*, vol. 1, no. 1, p. 47, 2007.

[20] N. Crook and H. S. Alper, "Model-based design of synthetic, biological systems," *Chem. Eng. Sci.*, Synthetic Biology, vol. 103, pp. 2–11, 2013.

[21] P. Fontanarrosa, H. Hosseini, A. Borujeni, Y. Dorfan, C. Voigt, and C. Myers, "Analyzing genetic circuits for hazards and glitches," in *11th International Workshop on Bio-Design Automation*, Cambridge, UK, 2019, pp. 32–33.

[22] C. J. Myers, *Asynchronous Circuit Design.* Salt Lake City, Utah, USA: John Wiley & Sons, 2001, 425 pp.

[23] D. K. Dacol and H. Rabitz, "Sensitivity analysis of stochastic kinetic models," *J. Math. Phys.*, vol. 25, no. 9, pp. 2716–2727, 1984.

[24] C. Damiani, A. Filisetti, A. Graudenzi, and P. Lecca, "Parameter sensitivity analysis of stochastic models: Application to catalytic reaction networks," *Comput. Biol. Chem.*, vol. 42, pp. 5–17, 2013.

[25] A. Gupta and M. Khammash, "An efficient and unbiased method for sensitivity analysis of stochastic reaction networks," *J. R. Soc. Interface*, vol. 11, no. 101, p. 20 140 979, 2014.

[26] A. D. Irving, "Stochastic sensitivity analysis," *Appl. Math. Model.*, vol. 16, no. 1, pp. 3–15, 1992.

[27] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *J. Theor. Biol.*, vol. 254, no. 1, pp. 178–196, 2008.

[28] K. Brenner, L. You, and F. H. Arnold, "Engineering microbial consortia: A new frontier in synthetic biology," *Trends Biotechnol.*, vol. 26, no. 9, pp. 483–489, 2008.

[29] K. Faust, "Microbial consortium design benefits from metabolic modeling," *Trends Biotechnol.*, vol. 37, no. 2, pp. 123–125, 2019.

[30] A. L. Gould, V. Zhang, L. Lamberti, E. W. Jones, B. Obadia, N. Korasidis, A. Gavryushkin, J. M. Carlson, N. Beerenwinkel, and W. B. Ludington, "Microbiome interactions shape host fitness," *PNAS*, vol. 115, no. 51, E11951–E11960, 2018.

[31] F. Pinto, D. A. Medina, J. R. Pérez-Correa, and D. Garrido, "Modeling metabolic interactions in a consortium of the infant gut microbiome," *Front. Microbiol.*, vol. 8, 2017.

[32] O. S. Venturelli, A. V. Carr, G. Fisher, R. H. Hsu, R. Lau, B. P. Bowen, S. Hromada, T. Northen, and A. P. Arkin, "Deciphering microbial interactions in synthetic human gut microbiome communities," *Mol. Syst. Biol.*, vol. 14, no. 6, e8157, 2018.

[33] R. J. White, G. C. Y. Peng, and S. S. Demir, "Multiscale modeling of biomedical, biological, and behavioral systems (Part 1) [Introduction to the special issue]," *IEEE Eng. Med. Biol. Mag.*, vol. 28, no. 2, pp. 12–13, 2009.

[34] P. K. R. Kambam, M. A. Henson, and L. Sun, "Design and mathematical modelling of a synthetic symbiotic ecosystem," *IET Syst. Biol.*, vol. 2, no. 1, pp. 33–38, 2008.