Integration of Circuit Design Automation and Genome-scale Modeling

Linh Huynh, Minseung Kim and Ilias Tagkopoulos
Department of Computer Science
& UC Davis Genome Center
University of California, Davis
{huynh,msgkim,itagkopoulos}@ucdavis.edu

1. INTRODUCTION

The design of a functional component that is part of a larger, interconnected ensemble, requires the following fundamental principles: First, the availability of characterized fundamental blocks, may it be transistors and capacitors for electronic design or promoters and coding regions for biological engineering, that can be assembled together into a functional entity. Second, the development of predictive models that are accurate enough to capture the dynamic behavior of the design component and its effect at a systems-level. Third, access to optimization tools that can provide optimal solutions, given user-defined constraints and objective functions, by utilizing the former (parts and model predictions) as its inputs. In this abstract, we present our work towards an efficient circuit optimization strategy and a data-driven, genome-scale host model that can be integrated to any circuit design platform. The resulting framework accounts for host-related and secondary effects, which are generally ignored but can have substantial effect on the host and circuit behavior.

	Running time (seconds)	
Design	DP	BB
`2-cascade`	1.8e-1	2.0e-2
`3-cascade`	2.1e-1	2.0e-2
`4-cascade`	2.5e-1	4.0e-2
`band-detector`	3.4e-1	8.0e-2
`feed-forward`	6.3e-1	7.0e-2
`2-not-and`	6.4e-1	4.0e-2
`3-input-and`	2.3	1.2e-1
`3-not-and`	3.6e1	1.5e-1
`2-to-1-mux`	1.1e2	1.6e-1
`D1`	1.2e3	9.8e-1
`D2`	2.0e3	4.5e-1

Table 1: A comparison between the running time of the dynamic programming (DP) approach [1] and the branch and bound (BB) approach.

2. METHODS AND RESULTS

2.1 An efficient optimization method with optimality guaranty

The carriers of information in gene circuits are chemical molecules within the cell (broadcast) instead of electrons within an isolated wire (unicast) as it is the case in electrical circuits. This adds two more necessary constraints for gene circuit design, namely the absence of cross-talk effects and connection compatibility between sub-circuits (i.e. the output molecules of a sub-circuit match the input molecules of all connected downstream sub-circuits). These constraints make the problem of selecting parts/modules to build a circuit become very difficult. Recently, an exact approach [1] was introduced to complement previous heuristic efforts [2] for that selection problem. This approach is based on a dynamic programming paradigm to explore the solution space by enumerating all sub-solutions, albeit at a large computational cost that can be prohibiting, in very large part database and circuit sizes. To address this, we devised a branch-and-bound method that estimated the bound of the solution cost (i.e. the bound on the number of parts and modules used in a solution) by relaxing the constraints of cross-talk absence and connection compatibility. With this bound information, we repeat the search for a complete solution (i.e. a solution that satisfies both the constraints) of a cost value from the lower bound to the upper bound until such a complete solution is found. By that way, we can skip the enumeration of all sub-solutions, which results in significant computational performance improvement. In a benchmark with 11 circuits that span several functional domains and a part library with 75 parts and 271 experimentally constructed modules, this method resulted in a remarkable improvement in running time (see table 1) when compared to the original approach in [1].

2.2 A multi-layer, genome-scale model for phenotypic predictions

The simulation of a host that has been genetically engineered is an important step towards design automation. For this reason, we developed an integrated genome-scale model [3] for phenotypic predictions of natural and engineering *E. coli* strains in several laboratory environments (Figure 1). We first constructed a normalized dataset that contains the expression of 4189 genes in 2262 conditions, including data for 31 strains and over 15 different media. We then created an integrative model that contains three submodels that bridge the transcriptional, signal transduction and metabolic layers. This model covers 3704 regulatory interactions, 151 instances of signal transduction systems and 2251 metabolic reactions. Parameters in the transcriptional sub-model were determined by fitting the gene expression level of 328 transcription factors over four sets of constraints (phenomenological, capacity, environmental and

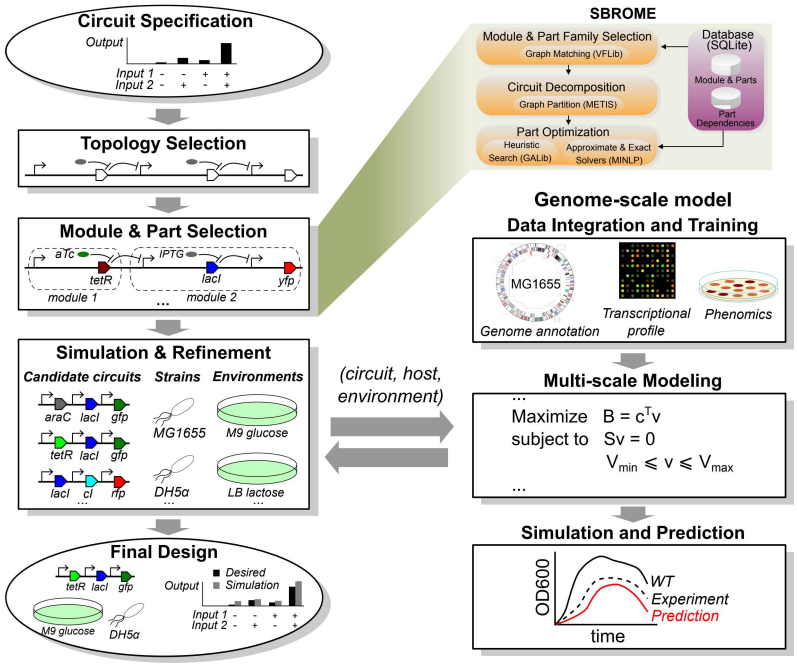

Figure 1: An overview on a computational framework to design and simulate synthetic gene circuits

genetic constraints). The integrated model was evaluated by performing cross-validation on the various datasets for growth and gene expression prediction, as well as predicting de novo experimentally measured data on the growth rate of 10 single-gene knock-outs for *E. coli* strains over different environments (28 genotype-phenotype combinations in total). Results show that our model can predict growth rates with 0.6 to 0.8 pearson correlation coefficient between the experimentally measured and computationally-derived predictions, which is significantly higher than M models and on par to other ME models so far [4]. Furthermore, the constructed model can sense environmental changes and translated them to changes in gene expression and growth, which is a significant step forward for bioengineering efforts.

2.3 Integration of a computer-aided design optimization platform with a genome-scale simulator

As shown in Figure 1, a list of top-ranked candidate circuits from the optimization framework, act as inputs to the genome-scale simulator, in order to predict the dynamic behavior within a specific host strain and environmental conditions. All possible triplets of circuit, host strain and environmental conditions are considered and their information is used to update the transcriptional sub-models by modifying the gene regulatory network and signal transduction sub-models by adding/removing related equations. Additionally, information about the environmental conditions affects the input of signal transduction and the metabolism sub-models, by setting the bound of related fluxes and components that are implicated in signal transduction pathways. The resulting benchmark on the triplet information serves as a decision support for laboratory construction and testing.

3. DISCUSSION

In this abstract, we present our results in an optimization method for part selection and an approach to integrate a design workflow and a genome-scale simulator. Although the incorporation of a genome-scale model to a design pipeline seems straight-forward, once each of these two frameworks are in place, there are a number of points to be considered. First, genome-scale and circuit models use a very different approach to modeling, with the former relying in a very small set of parameters that usually map statistical associations and not biophysical phenomena. In contrast, circuit models try to capture, in detail, the biophysical dynamics and use the appropriate kinetic constants (e.g. dissociation constants k_D, degradation rates k_{deg}) and modeling frameworks to do so. Merging these two worlds under a unifying framework that increases the model's predictive ability is quite challenging. As both fields move forward, data availability and coordinated efforts in both disciplines will be instrumental to close this gap.

4. REFERENCES

[1] L. Huynh and I. Tagkopoulos, "Optimal part and module selection for synthetic gene circuit design automation," *ACS Synth. Biol.*, 2014.

[2] L. Huynh, A. Tsoukalas, M. Köppe, and I. Tagkopoulos, "Sbrome: A scalable optimization and module matching framework for automated biosystems design," *ACS Synth. Biol.*, vol. 2, no. 5, pp. 263–273, 2013.

[3] Carrera, J., et al, "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction," *Mol. Syst. Biol.*, vol. 10:735, doi:10.15252/msb.145108, 2014.

[4] O'Brien, Edward J., et al, "Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction," *Mol. Syst. Biol.*, vol. 9, no. 1, 2013.