Syntax-Guided Optimal Synthesis for Chemical Reaction Networks

Luca Cardelli[a,b], Milan Češka[c], Martin Fränzle[d], Marta Kwiatkowska[b], Luca Laurenti[b], Nicola Paoletti[e,2] and Max Whitby[b]

[a] *Microsoft Research Cambridge, UK*

[b] *Department of Computer Science, University of Oxford, UK*

[c] *Faculty of Information Technology, Brno University of Technology, Czech Republic*

[d] *Dept. of Computer Science, Carl von Ossietzky Universität Oldenburg, Germany*

[e] *Department of Computer Science, Stony Brook University, USA*

Extended abstract

Chemical Reaction Networks (CRNs) are a language widely used for modelling and analysis of biochemical systems and molecular devices [17,5], and represent a fundamental computational structure, equivalent to Petri nets [15], Vector Addition Systems [14] and distributed population protocols [2].

Driven by potential applications ranging from smart therapeutics to biosensors, the automated design, or *synthesis*, of CRNs that exhibit prescribed dynamics is a major goal of synthetic biology [9,6,19]. This problem is even more challenging when stochasticity must be considered, which is the case when low molecular counts are involved, as common for molecular computation [12]. Indeed, current approaches for the automated synthesis of *stochastic CRNs* are limited to the estimation or synthesis of rate parameters [8,20], which neglect the network structure, and often suffer from scalability issues [11].

In this extended abstract, we give an overview of our recently published method [7] for the syntax-guided synthesis of stochastic CRNs, which enables, for the first time, efficient synthesis of both topology/structure and rate parameters of the network. Our work borrows from the field of syntax-guided program synthesis [1], which

[1] This work has been partially supported by the Czech Grant Agency grant No. GA16-17538S (M. Češka), Royal Society professorship, EPSRC Programme on Mobile Autonomy (EP/M019918/1), and NFS CPS Frontier project CyberCardia (N. Paoletti)

[2] Email: nclpltt@gmail.com

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: www.elsevier.com/locate/entcs

is based on the idea of supplementing the correctness specification with a syntactic template that describes a high-level structure of the program and constrains the space of allowed programs. In particular, we focus on *program sketching* [16], where the template is a partial program with holes (incomplete information) that are automatically resolved using a constraint solver.

One contribution is the definition of the *first sketching language for CRNs*, which allows designers to capture syntactic constraints on the network topology and, at the same time, leave unspecified parts of the network for which only limited knowledge is available. A CRN sketch can be seen as a parametric CRN, where the parameters can be unknown species, rates or stoichiometric constants. For instance, the sketch

$$\{A,B\} \;+\; c_1 B \;\xrightarrow{(0,15]}\; c_2 B + C + ?$$

captures all reactions consuming one copy of either A or B and c_1 copies of B, where c_1 is an unknown stoichiometric constant. The reaction has a positive rate ranging not greater than 15, and produces c_2 copies of B (with c_2 unknown), one copy of C and an unknown amount of an unknown species (term ?).

Often not only the correctness of synthesized network is important, but also its optimality with respect to a given cost [3]. To this purpose, we define a cost function that captures the structural complexity of the CRNs and reflects the cost of physically implementing it using DNA [5].

To ensure computational feasibility of the synthesis process, we employ the *Linear Noise Approximation (LNA)* of the Chemical Master Equation (CME) [18]. The LNA describes the time evolution of expectation and variance of the species in terms of ODEs, thus capturing the stochasticity intrinsic in CRNs, but, in contrast to solving the CME (that has a number of ODEs equals to the state space size), scales well with respect to the molecular counts. It follows that the stochastic behaviour of a sketch can be represented as a set of parametric ODEs, which can be adequately solved as a *satisfiability modulo theories (SMT) problem over the reals with ODEs*, for which dedicated solvers [13] exist.

To specify correctness requirements, we consider a class of constraints describing a dynamical profile as a finite sequence of phases, where each phase is a formula over the expected number and variance of molecules, and, crucially, their derivatives over time. This allows us, for instance, to express that a given species must have a monotonic behaviour, or exhibit a specific number of oscillations. Similarly, we can require that one species has higher/lower variance than another species. This specification language clearly provides greater expressiveness compared to simple reachability specifications or temporal logic.

We therefore formulate and provide a solution to the following optimal synthesis problem:

Input: CRN sketch, correctness specification, and cost function.
Output: a sketch instantiation (a concrete CRN), if exists, that satisfies the specification and minimizes the cost.

The optimal solution for a given sketch is computed using the *meta-sketch* abstraction for CRNs inspired by [3]. It combines a representation of the syntactic

$\mathsf{Dec} = \{c_1, \ldots, c_4 : [0,2],\ k_1, k_2, k_3 : [0, 0.1],\ \lambda_1, \lambda_2 : \{A, B\}\},$
$\mathsf{Con} = \{\lambda_1 \neq \lambda_2,\ c_1 < c_2,\ c_3 > c_4\},$
$\mathsf{Ini} = \{K_0 = 1 \wedge A_0 \in [0, 100] \wedge B_0 \in [0, 100]\}$

$\lambda_1 + c_1 K \xrightarrow{k_1} c_2 K$
$\{0,1\}\lambda_2 + c_3 K \xrightarrow{k_2} ?\lambda_2 + c_4 K$
$\emptyset \xrightarrow{k_3} \{\lambda_2, [1,2]K\}$

$A + K \to^{56} 2K; \quad K + B \to^{43} 2B$

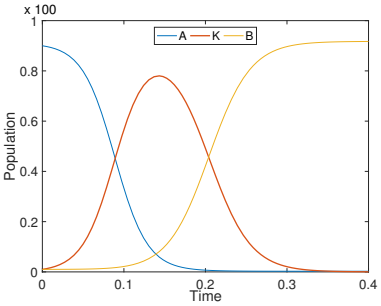

Fig. 1. **Left:** The sketch for bell-shape generator. Dec is the list of variable declarations: stoichiometric constants $c_1, \ldots, c_4$, rates k_1, k_2, k_3, and species λ_1, λ_2. Con and Ini are, respectively, constraints on variables and initial conditions. **Right:** CRN producing the bell-shape profile (species K) synthesized by our algorithm.

search space with the cost function and defines an ordered set of sketches. This cost-based ordering allows us to effectively prune the search space during the synthesis process and guide the search towards the minimal cost.

Building on the above cost-based ordering and the SMT solver iSAT(ODE) [13], we developed a synthesis algorithm and tool prototype for our method, which we evaluated on three case studies, representative of important problems studied in biology:

(i) **Bell-shape generator** (see Figure 1), where we synthesize CRNs with a bell-shape profile, a component occurring in signaling cascades;

(ii) **Super Poisson**, where we synthesize CRN implementations of stochastic processes with prescribed levels of process noise, and, in particular of processes having variance greater than its expectation; and

(iii) **Phosphorelay network**, where we synthesize CRNs exhibiting switch-like sigmoidal profiles, which is the biochemical mechanism underlying cellular decision-making, driving in turn development and differentiation.

While previous attempts [4,10] to solve the above problems relied on tedious and time-consuming manual tuning of CRN parameters, our method provides fully automated and efficient solutions. Able to synthesize challenging systems with up to 37 ODEs and $\sim$ 10K admissible network topologies, our method shows unprecedented scalability and paves the way for design automation for provably-correct molecular devices.

References

[1] R. Alur, R. Bodik, G. Juniwal, M. M. Martin, M. Raghothaman, S. A. Seshia, R. Singh, A. Solar-Lezama, E. Torlak, and A. Udupa. Syntax-guided synthesis. *Dependable Software Systems Engineering*, 40:1–25, 2015.

[2] D. Angluin, J. Aspnes, D. Eisenstat, and E. Ruppert. The computational power of population protocols. *Distributed Computing*, 20(4):279–304, 2007.

[3] J. Bornholt, E. Torlak, D. Grossman, and L. Ceze. Optimizing synthesis with metasketches. In *POPL'16*, pages 775–788. ACM, 2016.

[4] L. Cardelli. Artificial biochemistry. In *Algorithmic Bioprocesses*, pages 429–462. Springer, 2009.

[5] L. Cardelli. Two-domain DNA strand displacement. *Mathematical Structures in Computer Science*, 23(02):247–271, 2013.

[6] L. Cardelli, M. Kwiatkowska, and M. Whitby. Chemical reaction network designs for asynchronous logic circuits. In *DNA'16*, pages 67–81. Springer, 2016.

[7] L. Cardelli, M. Češka, M. Fränzle, M. Kwiatkowska, L. Laurenti, N. Paoletti, and M. Whitby. Syntax-guided optimal synthesis for chemical reaction networks. In *Computer Aided Verification*, 2017.

[8] M. Češka, F. Dannenberg, N. Paoletti, M. Kwiatkowska, and L. Brim. Precise parameter synthesis for stochastic biochemical systems. *Acta Informatica*, pages 1–35, 2016.

[9] H.-L. Chen, D. Doty, and D. Soloveichik. Rate-independent computation in continuous chemical reaction networks. In *ITCS'14*, pages 313–326. ACM, 2014.

[10] A. Csikász-Nagy, L. Cardelli, and O. S. Soyer. Response dynamics of phosphorelays suggest their potential utility in cell signalling. *J. R. Soc. Interface*, 8(57):480–488, 2011.

[11] N. Dalchau, N. Murphy, R. Petersen, and B. Yordanov. Synthesizing and tuning chemical reaction networks with specified behaviours. In *DNA'15*, pages 16–33. Springer, 2015.

[12] F. Dannenberg, M. Kwiatkowska, C. Thachuk, and A. J. Turberfield. Dna walker circuits: computational potential, design, and verification. *Natural Computing*, 14(2):195–211, 2015.

[13] A. Eggers, N. Ramdani, N. S. Nedialkov, and M. Fränzle. Improving the SAT modulo ODE approach to hybrid systems analysis by combining different enclosure methods. *Software and Systems Modeling*, 14(1):121–148, 2015.

[14] R. M. Karp and R. E. Miller. Parallel program schemata. *Journal of Computer and system Sciences*, 3(2):147–195, 1969.

[15] T. Murata. Petri nets: Properties, analysis and applications. *Proceedings of the IEEE*, 77(4):541–580, 1989.

[16] A. Solar-Lezama, R. Rabbah, R. Bodík, and K. Ebcioğlu. Programming by sketching for bit-streaming programs. In *PLDI'05*, pages 281–294. ACM, 2005.

[17] D. Soloveichik, G. Seelig, and E. Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences of the United States of America*, 107(12):5393–5398, 2010.

[18] N. G. Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[19] B. Yordanov, J. Kim, R. L. Petersen, A. Shudy, V. V. Kulkarni, and A. Phillips. Computational design of nucleic acid feedback control circuits. *ACS synthetic biology*, 3(8):600–616, 2014.

[20] C. Zimmer and S. Sahle. Parameter estimation for stochastic models of biochemical reactions. *J Comput Sci Syst Biol*, 6:011–021, 2012.