Syntax-guided optimal synthesis for chemical reaction networks

(published in CAV 2017)

Nicola Paoletti

Department of Computer Science, Stony Brook University

Static Analysis for Systems Biology Workshop, NYU

29 August 2017

Joint work with:

Luca Cardelli, Milan Ceska, Martin Franzle, Marta Kwiatkowska, Luca Laurenti, Max Whitby

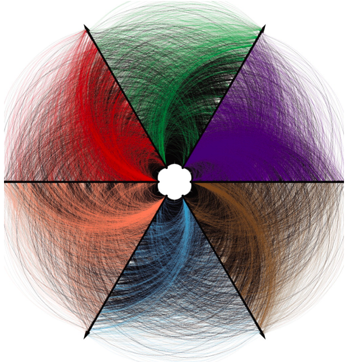
motivation: uncertainty in biological models

"SYSTEMS" UNCERTAINTY: partial knowledge of the system (e.g. unknown parameters, interactions)

AIM: understand life, fill knowledge gaps to derive predictive models consistent with observations

EXAMPLES: network reconstruction, parameter estimation

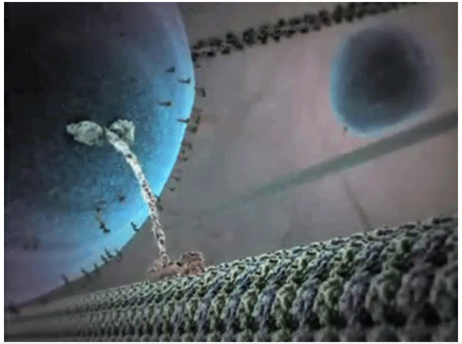
Human Transcription Factor Regulatory Network [Neph et al, Cell 150(6), 2012]



"SYNTHETIC" UNCERTAINTY: how to engineer living organisms to achieve specific functions?

AIM: automated design of correct-by-construction, optimal, biological processes/devices

EXAMPLES: synthetic biology, DNA programming, molecular computing



Segment from "Inner Life of the Cell," ©Harvard University

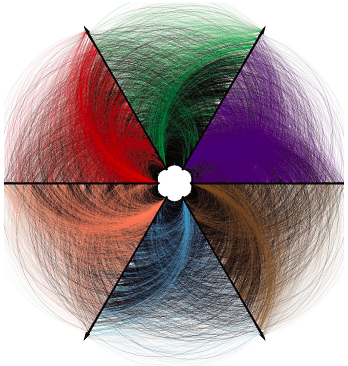
motivation: uncertainty in biological models

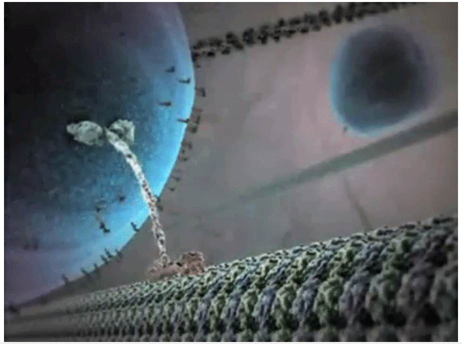
"SYSTEMS" UNCERTAINTY: partial knowledge of the system (e.g. unknown parameters, interactions)

"SYNTHETIC" UNCERTAINTY: how to engineer living organisms to achieve specific functions?

NEED FOR A MODELLING LANGUAGE ABLE TO CAPTURE UNCERTAINTIES AND AUTOMATED ANALYSIS METHODS TO RESOLVE THEM

Human Transcription Factor Regulatory Network [Neph et al, Cell 150(6), 2012]





Segment from "Inner Life of the Cell," ©Harvard University

chemical reaction networks (CRNs)

$$A + 2B \xrightarrow{k} C + D$$

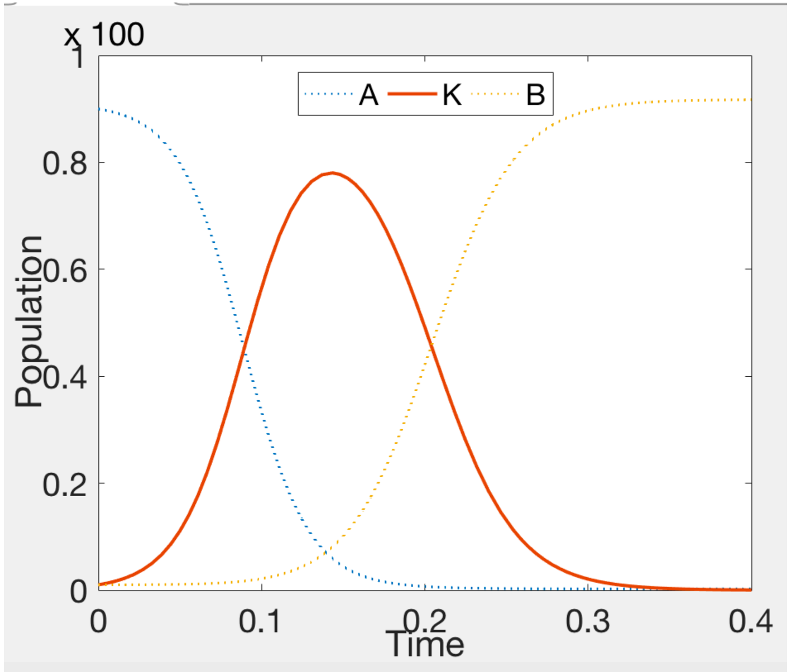
- CRNs are a **fundamental model for natural systems** (chemical, genetic, metabolic, ecological ...)
- **Fundamental computational structure** (equivalent to Petri Nets, Vector Addition Systems, ...)
- Can be "**compiled" into DNA**
- Biochemical interactions are **inherently stochastic** (CRN semantics typically described as continuous-time Markov chains)

problem: synthesis of CRNs

How to synthesize a network where

- Species K exhibits a bell-shaped profile (or an inflection point, local optimum, …)
- Variance of species B > variance of A
- Species B is monotonic
- …
- … and the network uses the least number of species/reactions (notion of cost/optimality)

What if both rate coefficients and network structure are unknown?



challenges of CRN synthesis

- We need quantitative reasoning while keeping stochasticity
- Classical CTMC semantics of CRNs is not scalable
- State-of-the-art approaches cannot deal with synthesis of both rate and structure

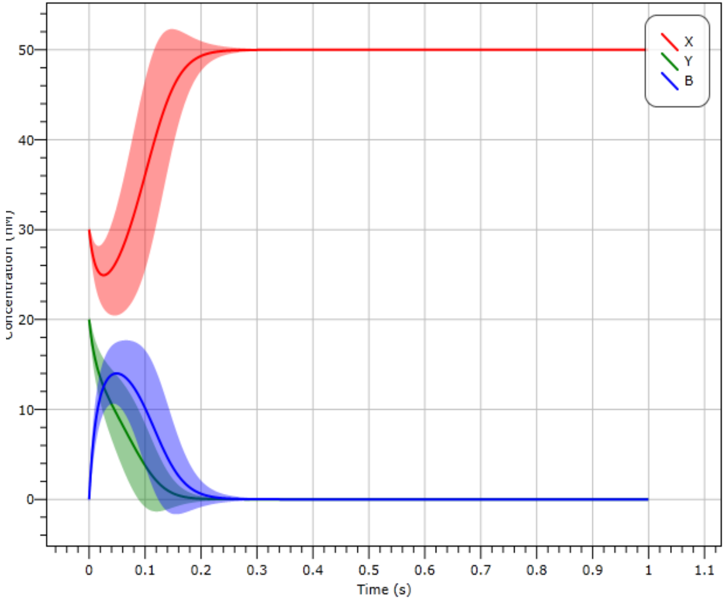
Related work

[Ceska et al., CMSB'14] CRN parameter synthesis as "parametric extension" of probabilistic model checking (can't synthesize structure)

[Dalchau et al., DNA, 2015] Structure and rates synthesized in two separate stages: inefficient, incomplete

idea 1: Linear Noise Approximation

- Stochastic semantics traditionally given by Chemical Master Equation of CTMC
- Linear Noise Approximation (LNA) produces ODEs for mean and covariances of species concentration over time
- Superior scalability, while keeping stochasticity



idea 2: syntax-guided program synthesis (SyGuS)

- **SyGuS**: correctness specification + syntactic template for the program [Alur et al. "Syntax-guided synthesis"]
- **Program sketching** [Solar-Lezama et al., PLDI'05]: programming with holes and variables (to model incomplete information), resolved using constraint solving (SMT)

idea 2: syntax-guided program synthesis (SyGuS)

- SyGuS: correctness specification + syntactic template for the program [Alur et al. "Syntax-guided synthesis"]
- Program sketching [Solar-Lezama et al., PLDI'05]: programming with holes and variables (to model incomplete information), resolved using constraint solving (SMT)

CRN sketch example

1) $A + 2B \xrightarrow{10} C + D$

2) $A + 2B \xrightarrow{(0,15]} C + ?$

3) $\{A, C\} + c_1 B \xrightarrow{(0,15]} c_2 B + C + ?$, with $c_1 > c_2$

our approach

- Linear noise approximation (LNA) semantics of CRNs
- SMT over non-linear reals and ODEs (SMT-ODE), iSAT(ODE) solver [Eggers et al., ATVA'08]

Contributions:

- First sketching language for synthesis of CRNs
- Specification language
- Novel optimal synthesis problem, encoded as an SMT-ODE solving problem through LNA
- Prototype + evaluation on 3 case studies

sketching language – Bell shape example

- Finite set of **species** (mandatory and optional)

Λ_m={K} and Λ_o= {A, B}

- **Declared variables** for species (λ_i), stochiometric coefficients (c_i), and rates (k_i)
- Variables express the uncertainty

λ_1, λ_2: {A,B},
$c_1,\ldots,c_4$: [0,2],

$k_1,\ldots,k_3$: [0,0.1]

- **Constraints** on initial state and variables

$K_0=1 \wedge A_0,B_0 \in [0,100]$

$\lambda_1 \neq \lambda_2 \wedge c_1 < c_2 \wedge c_3 > c_4$

sketching language – Bell shape example

Species

Λ_m={K} and Λ_o= {A, B}

Declared variables

λ_1, λ_2: {A,B},
$c_1,\ldots,c_4$: [0,2],
$k_1,\ldots,k_3$: [0,0.1]

Constraints

$K_0=1 \wedge A_0,B_0 \in [0,100]$

$\lambda_1 \neq \lambda_2 \wedge c_1 < c_2 \wedge c_3 > c_4$

Reaction sketch

$\tau_1 = \lambda_1 + c_1 K \rightarrow^{k1} c_2 K$

$\tau_2 = \{0,1\}\lambda_2 + c_3 K \rightarrow^{k2} ?\lambda_2 + c_4 K$

$\tau_3 = \varnothing \rightarrow^{k3} \{\lambda_2, [1,2]K\}$

Inline, implicit declarations ({0,1}, ?, [1,2])

Choice between λ_2 and $[1,2]K$

Example instantiation

$A + K \rightarrow^{0.01} 2K$

$B + K \rightarrow^{0.1} 2B$

$\varnothing \rightarrow^{0.001} K$

specification language

- Supports **constraints about the expected number and variance of molecules, and their derivatives**
- A property describes the required temporal profile as a sequence of phases

Bell shape (2 phases):

1

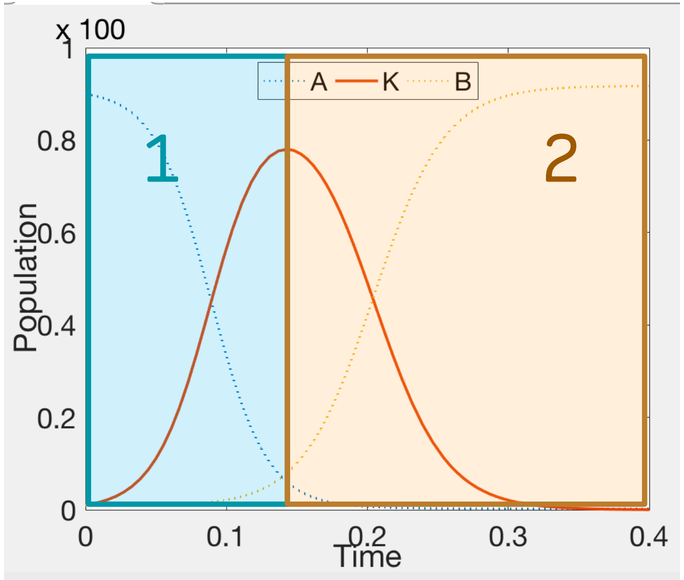
$$\text{inv}_1 \equiv E^{(1)}[K] \geq 0$$

$$\text{pre-post}_1 \equiv E^{(1)}[K]' = 0 \wedge E[K]' > 0$$

2

$$\text{inv}_2 \equiv E^{(1)}[K] \leq 0$$

$$\text{pre-post}_1 \equiv E[K]' \leq 1 \wedge T' = 1$$



optimal synthesis

PROBLEM: OPTIMAL SYNTHESIS OF CRNs

IN: Sketch S + Correctness specification ϕ + Cost function G

OUT: Instantiation I of S that satisfies ϕ (if exists) and is minimal w.r.t. G

Structural complexity cost:

COST = k_1*(num. of optional species in I) +
k_2*(total num. of reactants in I) +
k_3*(total num. of products in I)

Based on cost of implementation in DNA [Cardelli et al., MSCS ‘13]

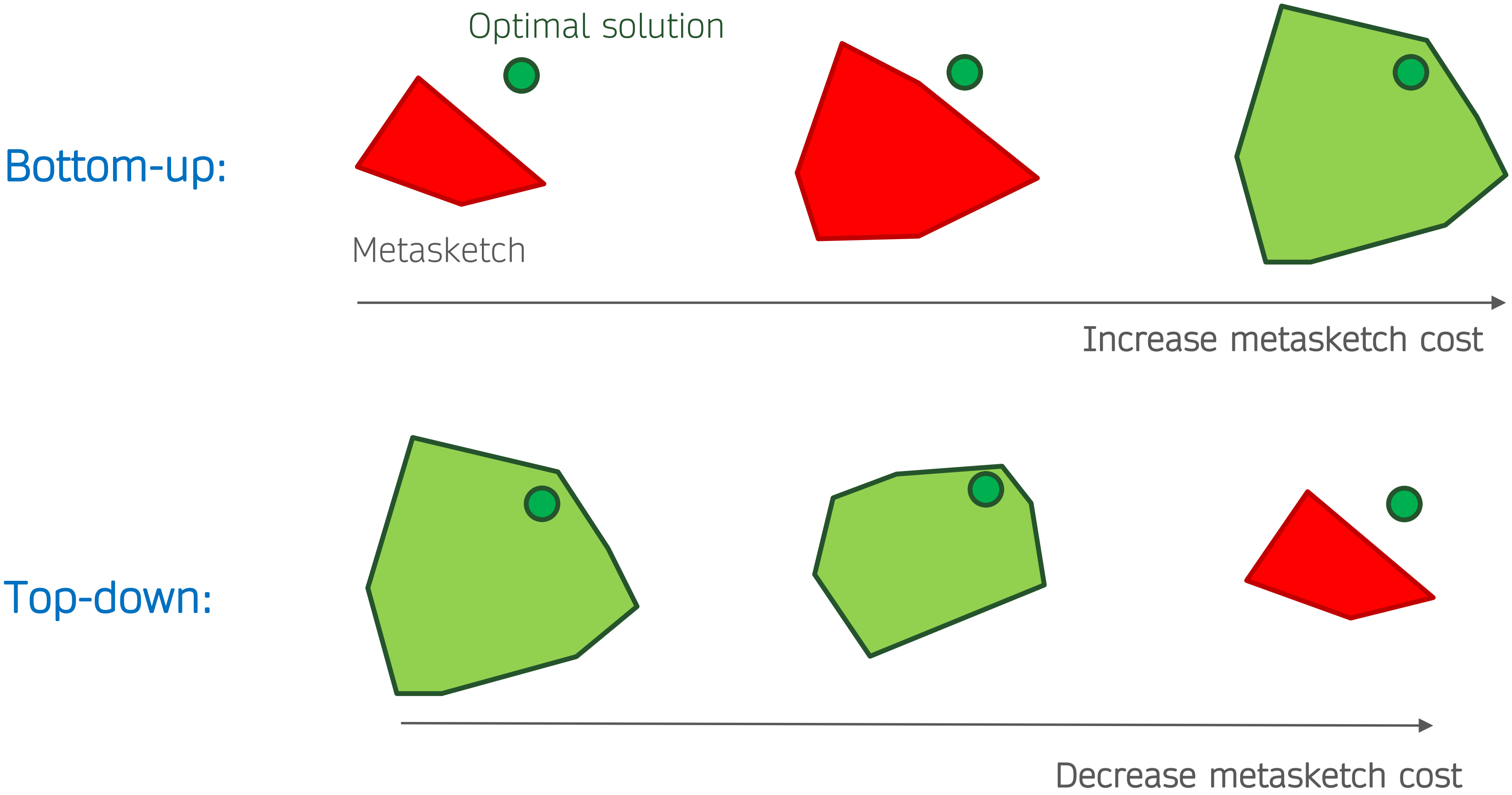
synthesis algorithm

- Builds on the notion of meta-sketch = sketch + cost constraints
- Cost constraints reduce search space size for SMT solver and improve runtime

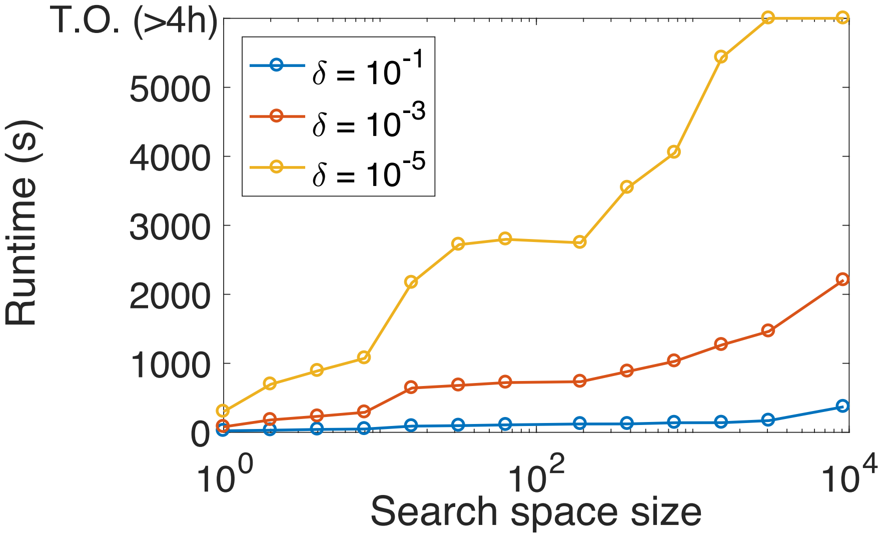
ALGORITHM

- Repeated calls to the solver under different cost constraints
- Generalized solving scheme:
  - Bottom-up: increase metasketch cost until SAT
  - Top-down: decrease metasketch cost until UNSAT
  - Binary search: use both SAT and UNSAT witnesses to bound optimal cost
- "Smart" UNSAT witness generation

synthesis algorithm



bell shape



- Bottom-up more effective (UNSAT instances are faster)
- Cost constraints reduce avg solving time for each call of between 37 and 83%

super Poisson

- **AIM**: synthesize "CRN implementation" of a stochastic process
- **Super-Poisson process** (variance > expectation)

Sketch:

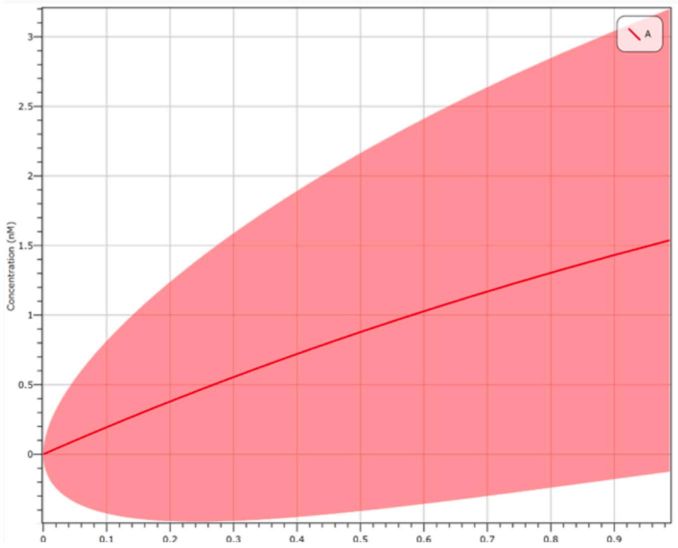
$$\boldsymbol{\tau_1} = \emptyset \rightarrow^{k1} c_1 A + c_2 \lambda_1$$
$$\boldsymbol{\tau_2} = A \rightarrow^{k2} c_3 \lambda_2$$

Specification:

$$inv_1 \equiv C[A] > E[A]$$
$$pre - post_1 \equiv T' = 1$$

Solution:

$$\emptyset \rightarrow^{23} 2A$$
$$A \rightarrow^{94} \emptyset$$



SOLVED IN 4 seconds!!!
Encoding size: 10 ODEs +
search space size 288

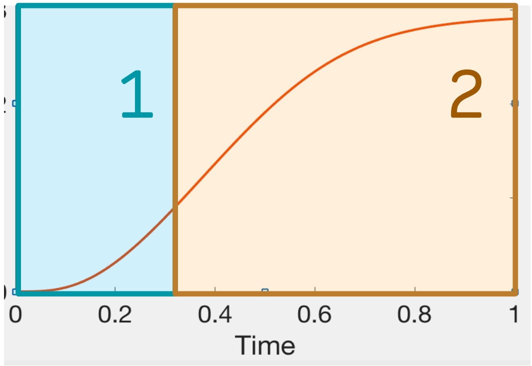
phosphorelay network

- Important signal transduction pathway [Csikász-Nagy et al., J. Royal Soc. Interface, 2011]
- **Aim**: find rate parameters s.t. output has switch-like profile

Sketch:

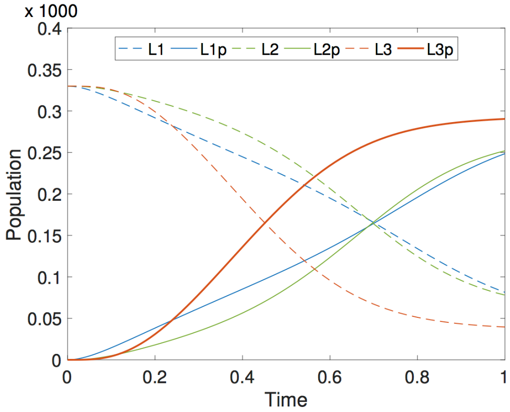
$$L1 + B \xrightarrow{k_1} B + L1p$$
$$L2 + L1p \xrightarrow{k_2} L1 + L2p$$
$$L2p + L3 \xrightarrow{k_3} L2 + L3p$$
$$L3p \xrightarrow{k_4} L3; \quad \emptyset \xrightarrow{1} B$$
$$k_1, \ldots, k_4 : (0, 100],$$
$$Li_0 = 330, Lip_0 = B_0 = 0$$

Specification (sigmoid in L3p):



Results

$$k_1 = 15, k_2 = 53, k_3 = 90, k_4 = 3$$



SOLVED IN 370 seconds
(9 ODEs, 7 for species, 2 for L3p derivatives)

CONCLUSIONS

- Method for synthesis of stochastic CRNs (both rates and structure)
- LNA semantics makes it as scalable as deterministic approximations
- First language for sketching CRNs
- Optimal synthesis algorithm based on SMT-ODE encoding

FUTURE WORK

- Explore multiple optimal solutions
- Combination of SMT solving and stochastic search
- Software tool