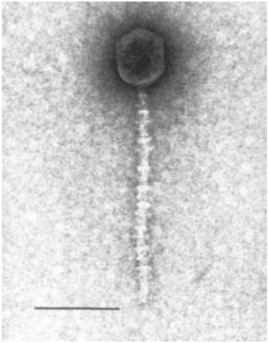
Genetic Design Automation:
Progress and Future Research Directions

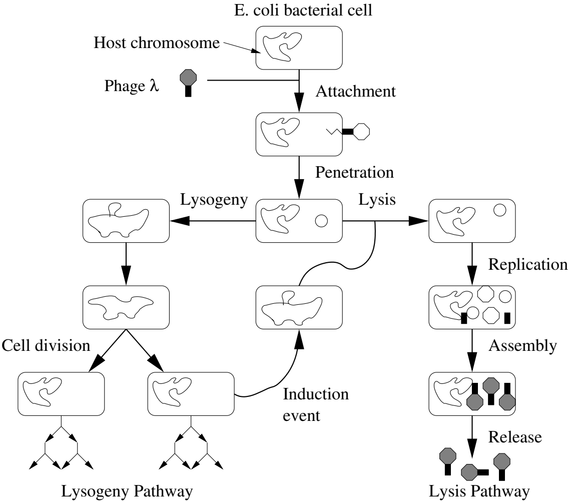
Chris Myers[1], Kevin Jones[1], Nathan Barker[2], Hiroyuki Kuwahara[3],
Curtis Madsen[1], Nam Nguyen[4], Chris Winstead[5]

[1]University of Utah
[2]Southern Utah University
[3]Carnegie Mellon University
[4]University of Texas at Austin
[5]Utah State University

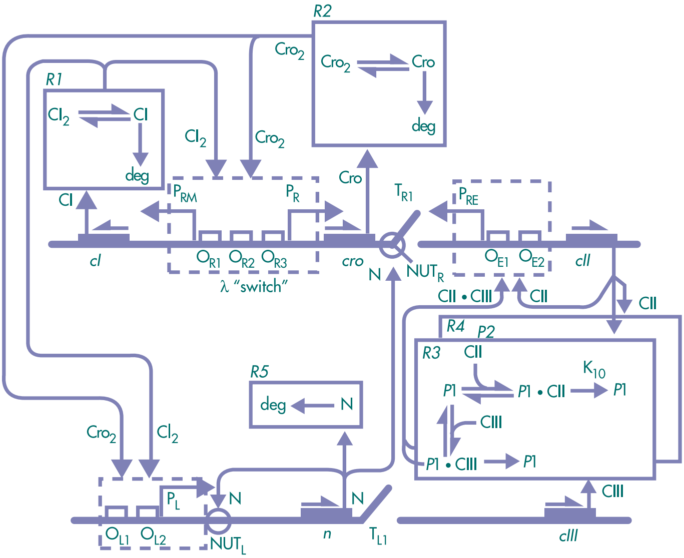
RoSBNet Synthetic Biology Workshop
September 16, 2009

Phage λ Virus

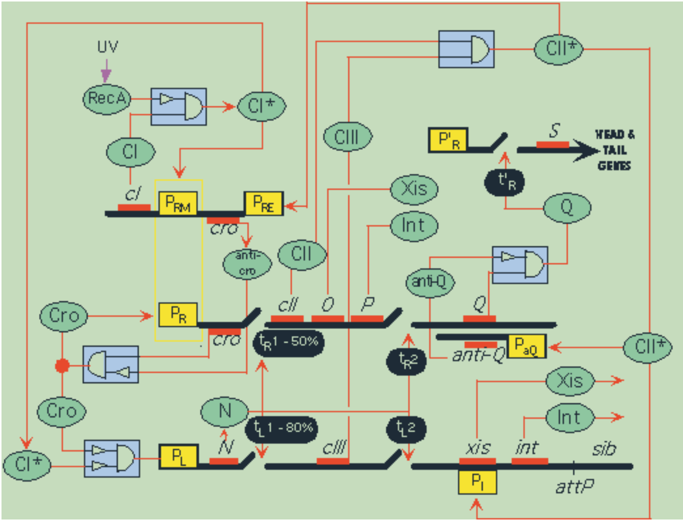




Phage λ Decision Circuit

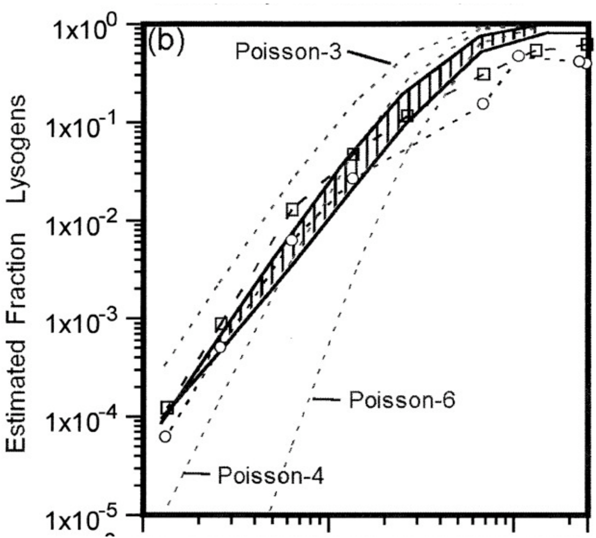


Asynchronous Circuit?



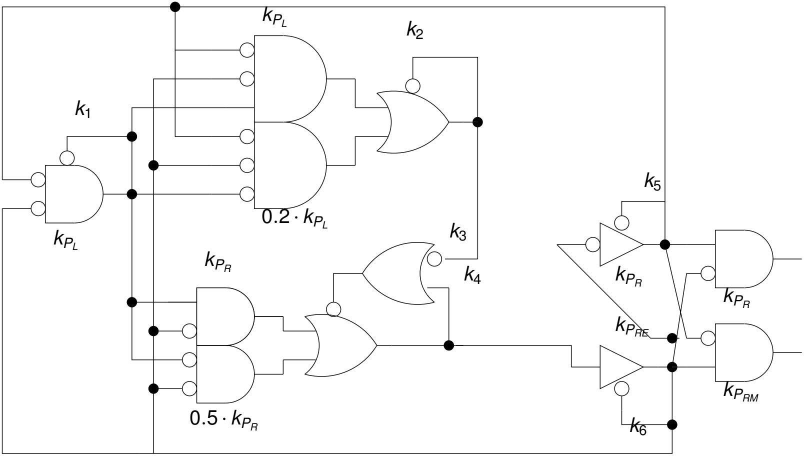
McAdams/Shapiro, Science (1995)

Stochastic Circuit?

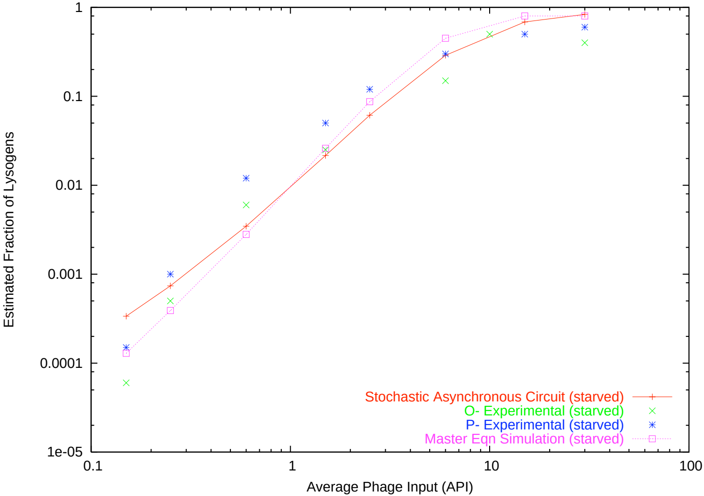


Arkin/Ross/McAdams, Genetics (1998)

Stochastic Asynchronous Circuit?



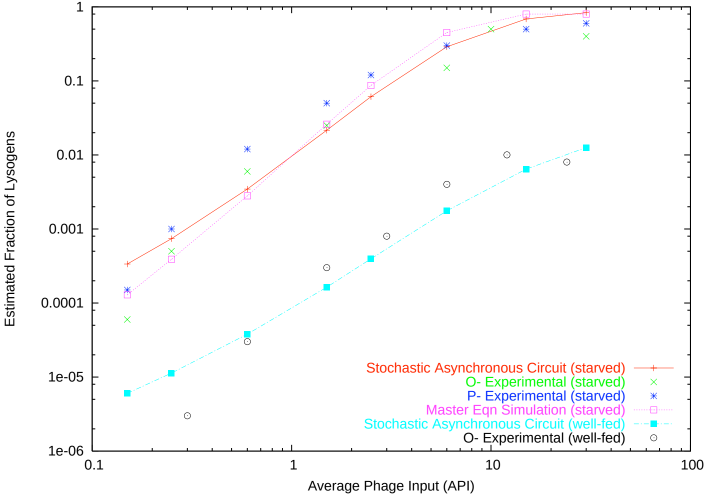
Stochastic Asynchronous Circuit Results



SAC results generated in only 7 minutes.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

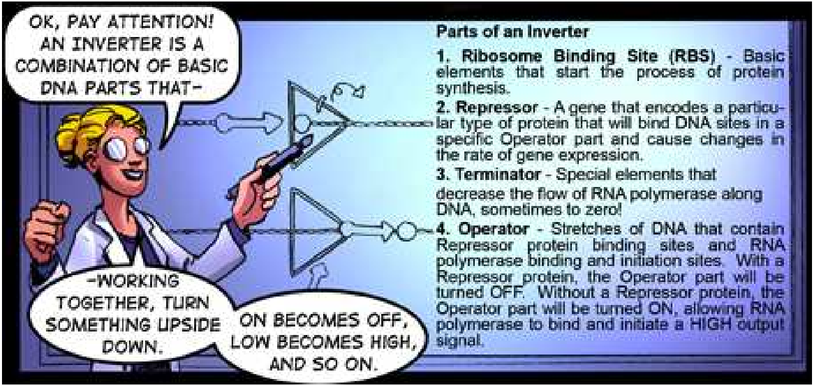
Stochastic Asynchronous Circuit Results



SAC results generated in only 7 minutes.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

Synthetic Biology



Parts of an Inverter

1. Ribosome Binding Site (RBS) - Basic elements that start the process of protein synthesis.

2. Repressor - A gene that encodes a particular type of protein that will bind DNA sites in a specific Operator part and cause changes in the rate of gene expression.

3. Terminator - Special elements that decrease the flow of RNA polymerase along DNA, sometimes to zero!

4. Operator - Stretches of DNA that contain Repressor protein binding sites and RNA polymerase binding and initiation sites. With a Repressor protein, the Operator part will be turned OFF. Without a Repressor protein, the Operator part will be turned ON, allowing RNA polymerase to bind and initiate a HIGH output signal.

(From “Adventures in Synthetic Biology” - Endy et al.)

Genetic Engineering vs. Synthetic Biology

- *Genetic engineering* (last 30 years):
  - *Recombinant DNA* - constructing artificial DNA through combinations.
  - *Polymerase Chain Reaction* (PCR) - making many copies of this new DNA.
  - *Automated sequencing* - checking the resulting DNA sequence.
- Synthetic biology adds:
  - *Standards* - create repositories of parts that can be easily composed.
  - *Abstraction* - high-level models to facilitate design.
  - *Automated construction* - separate design from construction.

(source: Drew Endy)

Genetic Design Automation (GDA)

- Standards, abstraction, and automated construction are the cornerstones of *Electronic Design Automation* (EDA).
- EDA facilitates the design of more complex integrated circuits each year.
- Crucial to the success of synthetic biology is an improvement in methods and tools for *Genetic Design Automation* (GDA).
- Experiences with EDA can jump start the development of GDA.

Current State of GDA (Standards)

- Registry of standard biological parts used to design synthetic genetic circuits (`http://partsregistry.org`).
- Adequate characterization of these parts is an ongoing effort.
- *Systems Biology Markup Language* (SBML) has been proposed as a standard representation for the simulation of biological systems.
- Many simulation tools have been developed that accept models in the SBML format (Copasi, Jarnac, CellDesigner, SimBiology, iBioSim, etc.).

Current State of GDA (Abstraction)

- Existing SBML-based GDA tools model biological systems at the molecular level.
- A typical SBML model is composed of a number of chemical *species* (i.e., proteins, genes, etc.) and *reactions* that transform these species.
- This is a very low level representation which is roughly equivalent to the layout level for electronic circuits.
- Designing and simulating genetic circuits at this level of detail is extremely tedious and time-consuming.

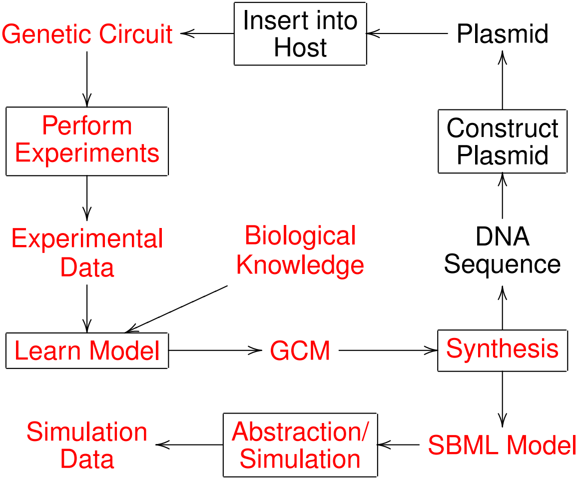
Current State of GDA (Automated Construction)

- Several companies have formed that will construct a plasmid from an arbitrary DNA sequence.
- It is still difficult, however, to separate design and construction issues.
- To achieve this, a GDA tool that supports higher-levels of abstraction for modeling, analysis, and design of genetic circuits is essential.

Overview

- This talk describes our research to develop a GDA tool that utilizes abstraction to improve the efficiency of analysis and design.
- The design of a quorum trigger circuit is presented as a case study.

Genetic Circuit Analysis



Genetic Circuit

Insert into Host

Plasmid

Perform Experiments

Construct Plasmid

Experimental Data

Biological Knowledge

DNA Sequence

Learn Model

GCM

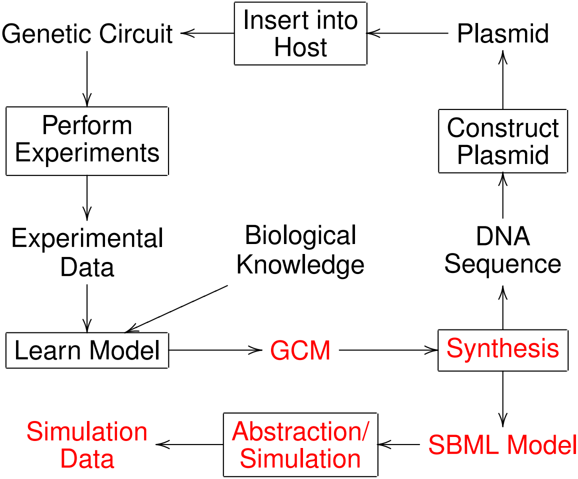
Synthesis

Simulation Data

Abstraction/ Simulation

SBML Model

Genetic Circuit Design



Genetic Circuit

Insert into Host

Plasmid

Perform Experiments

Construct Plasmid

Experimental Data

Biological Knowledge

DNA Sequence

Learn Model

GCM

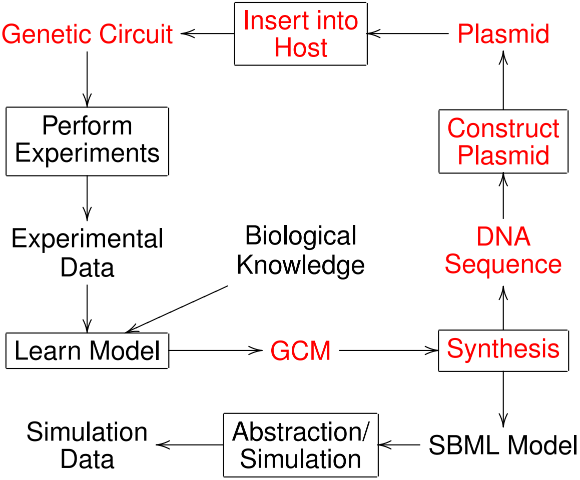
Synthesis

Simulation Data

Abstraction/ Simulation

SBML Model

Genetic Circuit Construction



Genetic Circuit

Insert into Host

Plasmid

Perform Experiments

Construct Plasmid

Experimental Data

Biological Knowledge

DNA Sequence

Learn Model

GCM

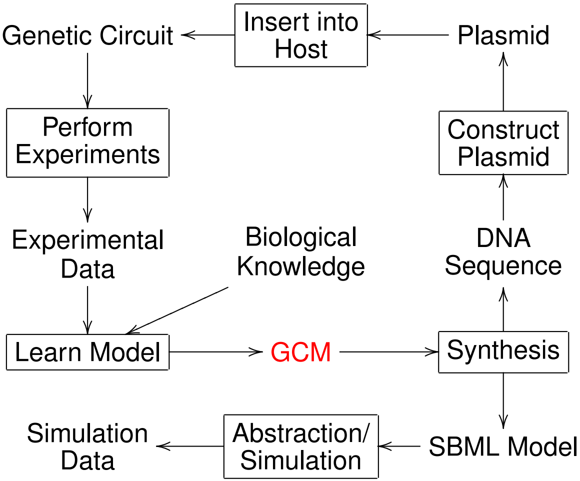
Synthesis

Simulation Data

Abstraction/ Simulation

SBML Model

Genetic Circuit Model (GCM)



Genetic Circuit

Insert into Host

Plasmid

Perform Experiments

Construct Plasmid

Experimental Data

Biological Knowledge

DNA Sequence

Learn Model

GCM

Synthesis

Simulation Data

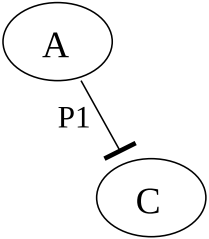
Abstraction/ Simulation

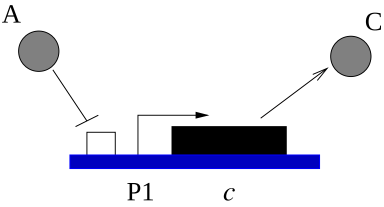
SBML Model

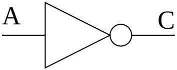
Genetic Circuit Model (GCM)

- Provides a higher level of abstraction than SBML.
- Includes only important species and their influences upon each other.
- GCMs also include structural constructs that allow us to connect GCMs for separate modules through species ports.

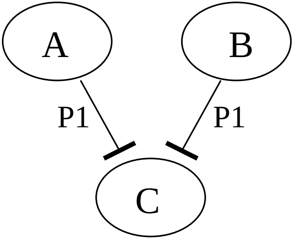
A Genetic Not Gate

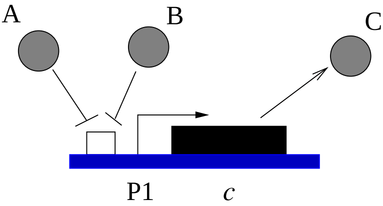


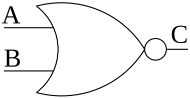




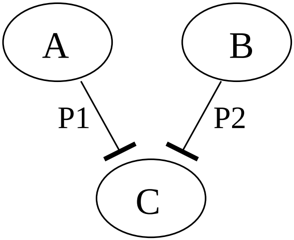
A Genetic Nor Gate

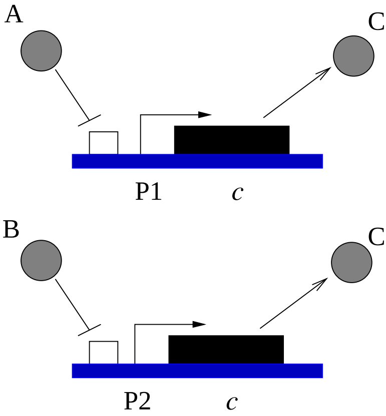


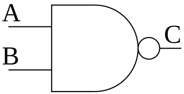




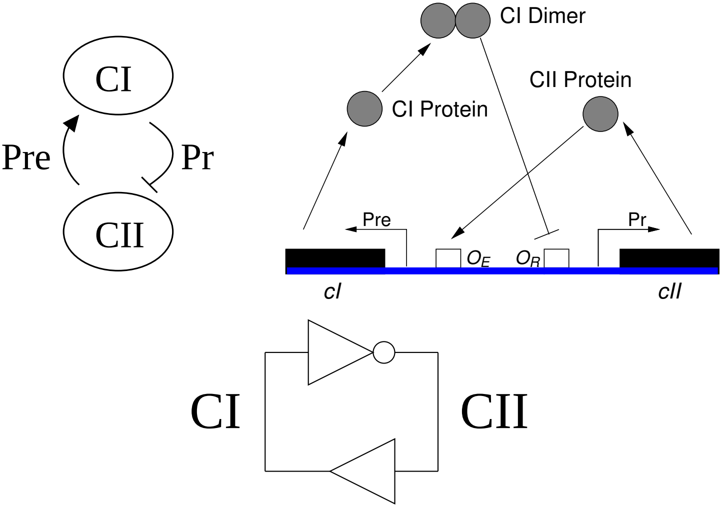
A Genetic Nand Gate



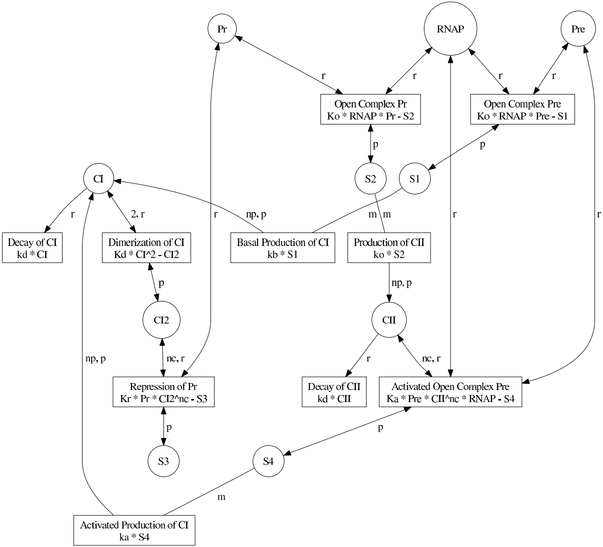




A Genetic Oscillator

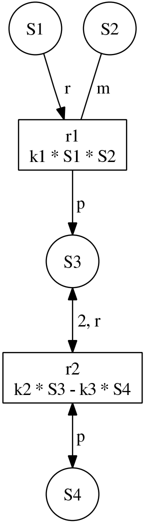


Molecular Representation



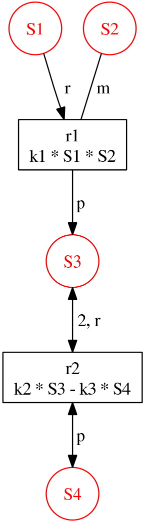
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



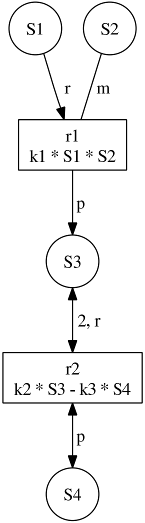
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



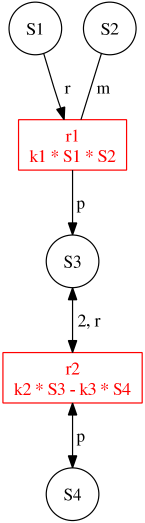
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



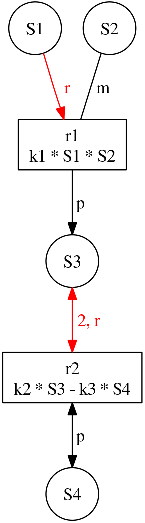
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



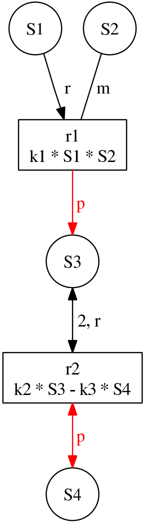
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



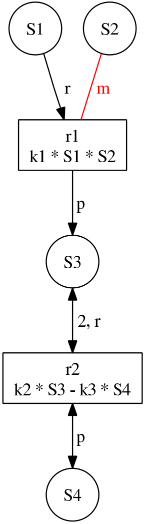
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



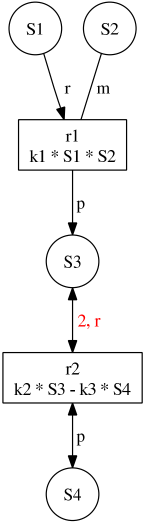
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



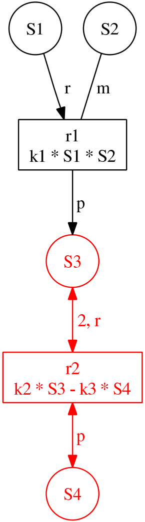
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



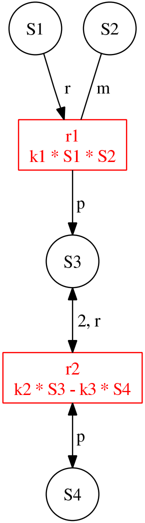
SBML: Main Elements

- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



SBML: Main Elements

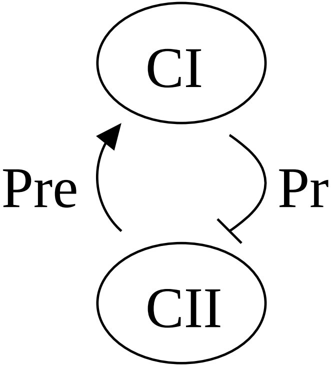
- Species
- Global parameters (ex. k1=0.1)
- Reactions
  - Reactants
  - Products
  - Modifiers
  - Stoichiometry
  - Reversible
  - Kinetic laws



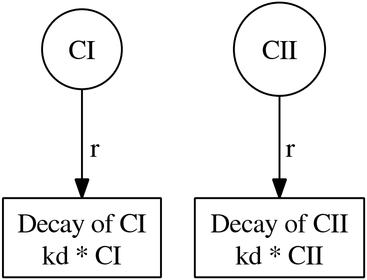
Synthesizing SBML from a GCM Representation

- Create degradation reactions
- Create open complex formation reactions
- Create dimerization reactions
- Create repression reactions
- Create activation reactions

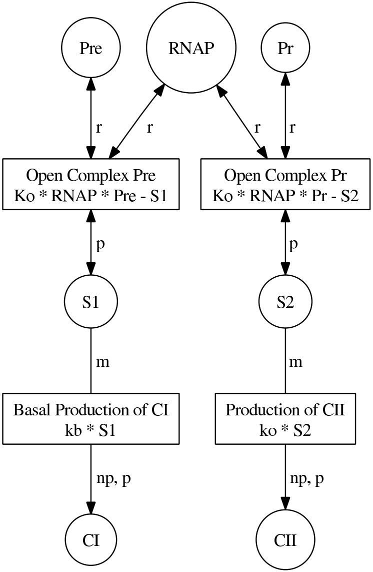
GCM Example



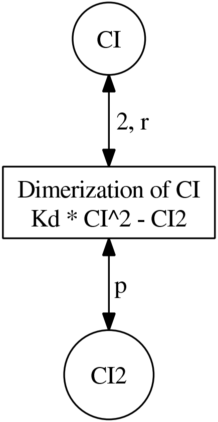
Degradation Reactions



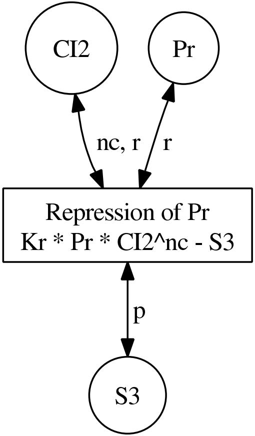
Open Complex Formation Reactions



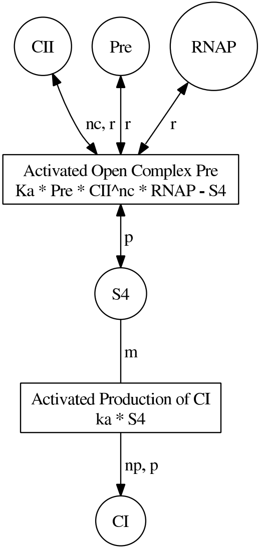
Dimerization Reactions



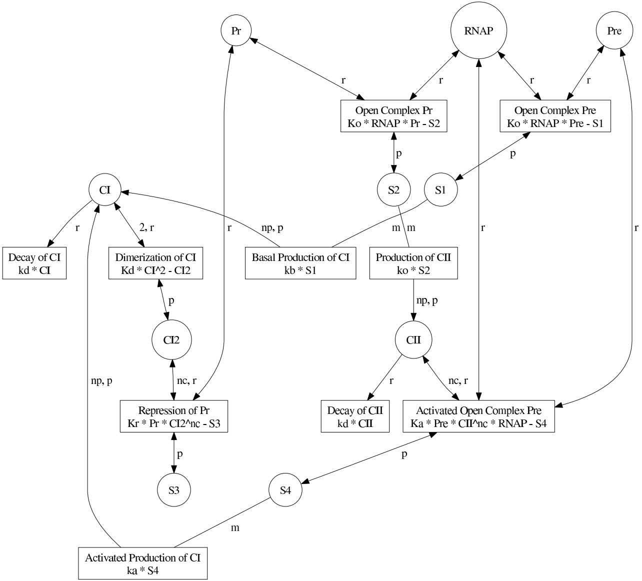
Repression Reactions



Activation Reactions



Complete SBML Model



Classical Chemical Kinetics

- Uses *ordinary differential equations* (ODE) to represent the system to be analyzed, and it assumes:
  - Molecule counts are high, so concentrations can be continuous variables.
  - Reactions occur continuously and deterministically.
- Genetic circuits have:
  - Small molecule counts which must be considered as discrete variables.
  - Gene expression reactions that occur sporadically.
- ODEs do not capture non-deterministic behavior.

NYTimes: Expressing Our Individuality, the Way E. Coli Do

Expressing Our Individuality, the Way E. Coli Do

Dr. Michael Elowitz

A colony of genetically identical E. coli is actually a mob of individuals. Under identical conditions, they behave in different ways.

By CARL ZIMMER
Published: April 22, 2008

Well
Tara Parker-Pope on Health

Socializing Appears to Delay Memory Problems
June 4, 2008, 12:34 PM

Colon Cancer in Family Predicts Better Survival
June 3, 2008

Jane Brody's New Knees
June 3, 2008

Brain Surgeons and Cellphones
June 2, 2008

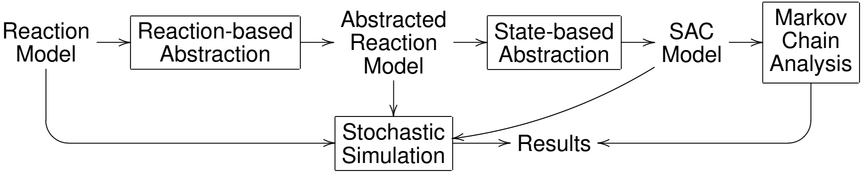
Cleaner Classrooms Mean Fewer Sick Kids
June 2, 2008

Rainbow and CC

Stochastic Chemical Kinetics

- To more accurately predict the temporal behavior of genetic circuits, *stochastic chemical kinetics* formalism can be used.
- Use Gillespie's *Stochastic Simulation Algorithm* which tracks the quantities of each molecular species and treats each reaction as a separate random event.
- Only practical for small systems with no major time-scale separations.
- Abstraction is essential for efficient analysis of any realistic system.

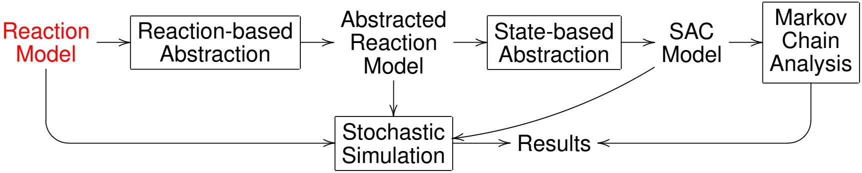
Automatic Abstraction



- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

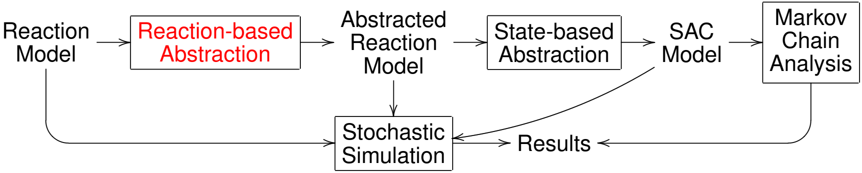
Automatic Abstraction



- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

Automatic Abstraction


Reaction Model

Reaction-based Abstraction

Abstracted Reaction Model

State-based Abstraction

SAC Model

Markov Chain Analysis

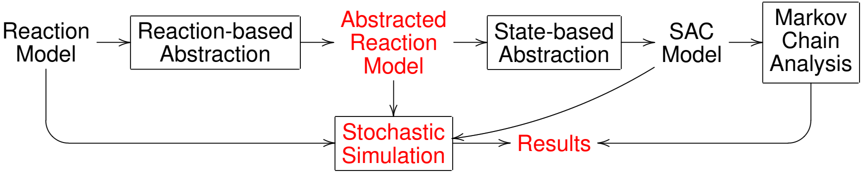
Stochastic Simulation

Results

- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

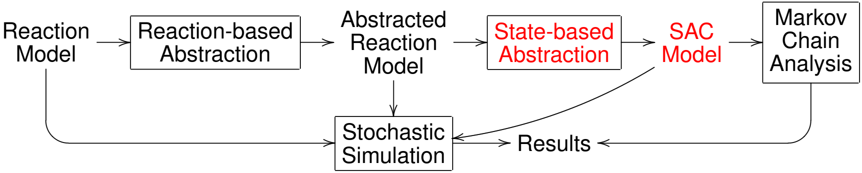
Automatic Abstraction



- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

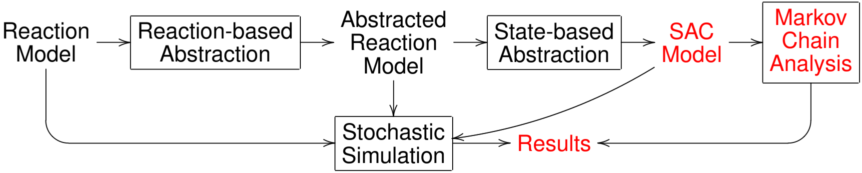
Automatic Abstraction



- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

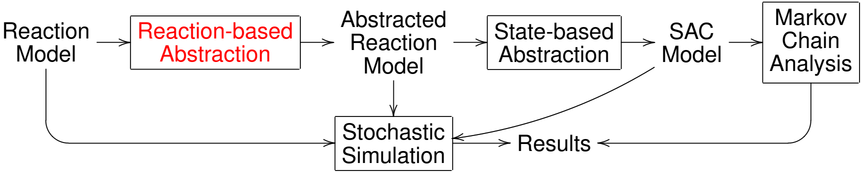
Automatic Abstraction



- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

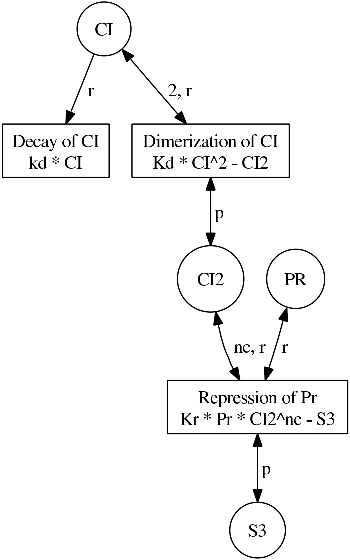
Automatic Abstraction



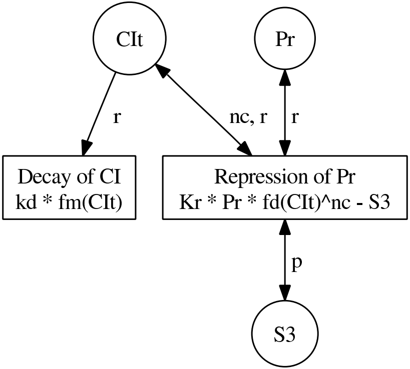
- Begins with a *reaction-based model* in SBML.
- Automatically abstracts this model leveraging the quasi-steady state assumption, whenever possible.
- Encodes chemical species concentrations into Boolean (or n-ary) levels to produce a *stochastic asynchronous circuit* (SAC) model.
- Can now utilize Markov chain analysis.

Kuwahara et al., Trans. on Comp. Sys. Bio. (2006)

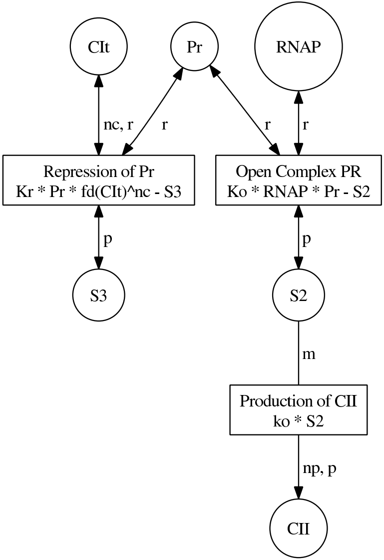
Dimerization Reduction



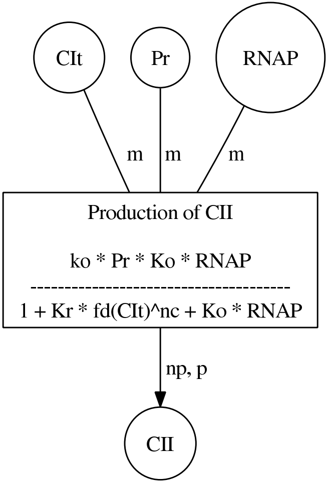
Dimerization Reduction



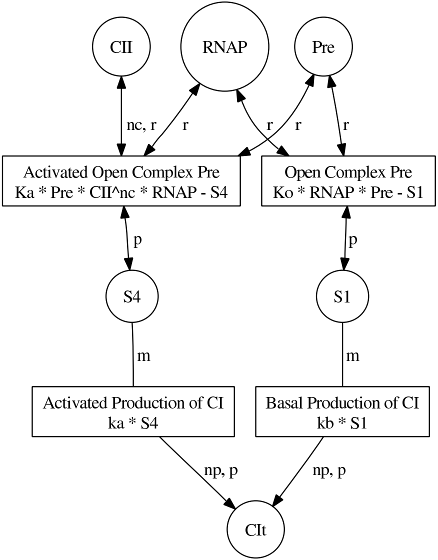
Operator Site Reduction (PR)



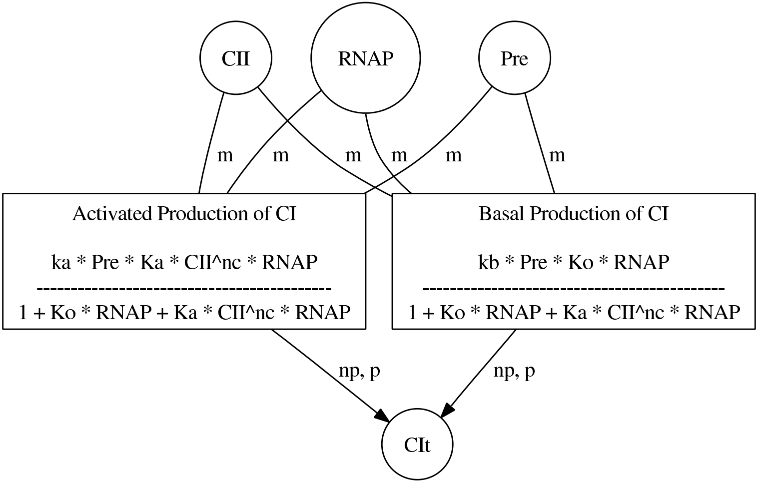
Operator Site Reduction (PR)



Operator Site Reduction (PRE)



Operator Site Reduction (PRE)



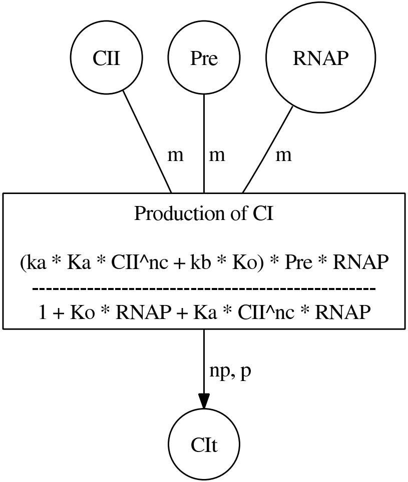
Activated Production of CI

ka * Pre * Ka * CII^nc * RNAP
----
1 + Ko * RNAP + Ka * CII^nc * RNAP

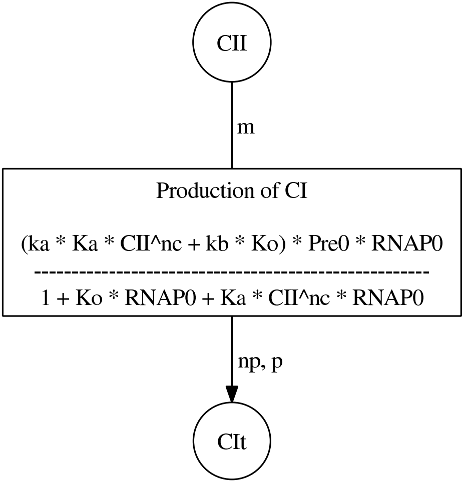
Basal Production of CI

kb * Pre * Ko * RNAP
----
1 + Ko * RNAP + Ka * CII^nc * RNAP

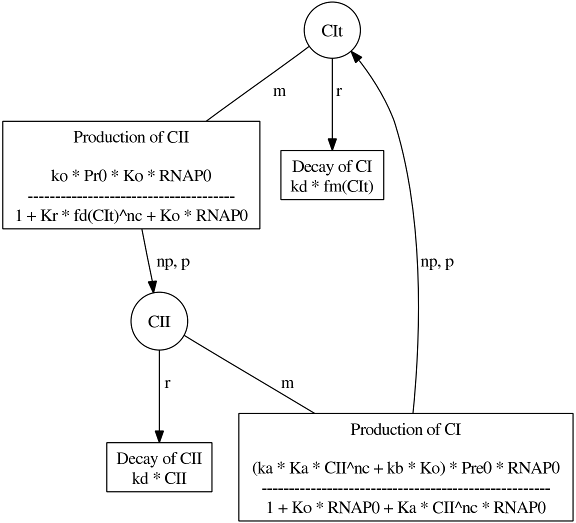
Similar Reaction Combination



Modifier Constant Propagation



Final SBML Model

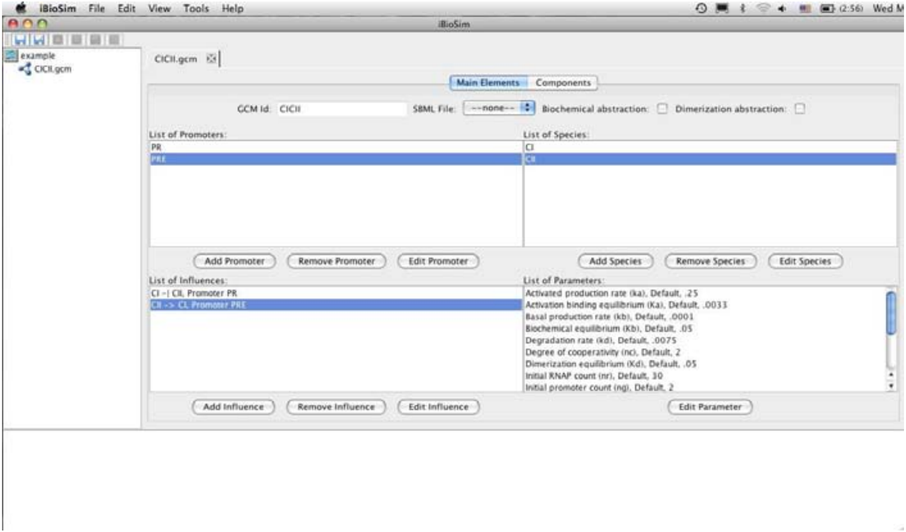


10 species and 10 reactions reduced to 2 species and 4 reactions

GCM Advantages

- Greatly increases the speed of model development and reduces the number of errors in the resulting models.
- Allows efficient exploration of the effects of parameter variation.
- Constrains SBML model such that it can be more easily abstracted resulting in substantial improvement in simulation time.

iBioSim: Genetic Circuit Editor



GCM Id: CICII

SBML File: --none--

Biochemical abstraction: ☐

Dimerization abstraction: ☐

List of Promoters:
- PR
- PRE

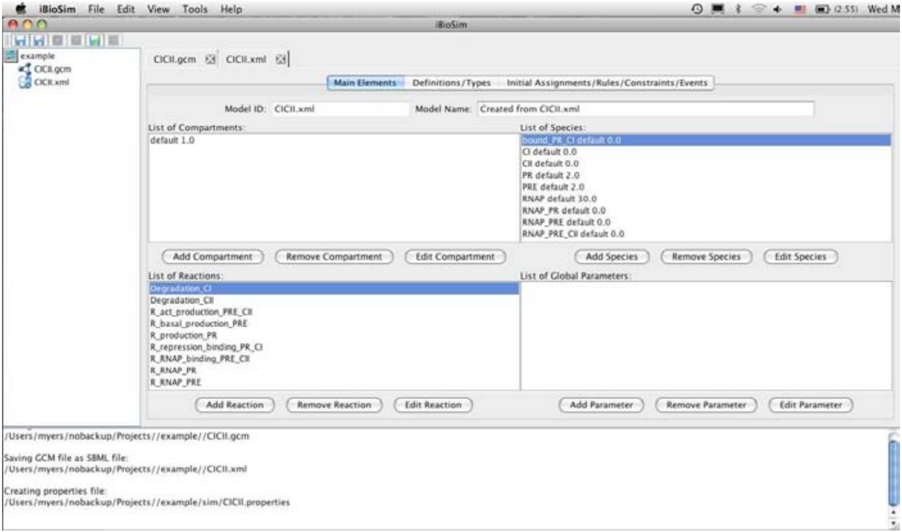
List of Species:
- CI
- CII

List of Influences:
- CI -| CII, Promoter PR
- CII -> CI, Promoter PRE

List of Parameters:
- Activated production rate (ka), Default, .25
- Activation binding equilibrium (Ka), Default, .0033
- Basal production rate (kb), Default, .0001
- Biochemical equilibrium (Kb), Default, .05
- Degradation rate (kd), Default, .0075
- Degree of cooperativity (nc), Default, 2
- Dimerization equilibrium (Kd), Default, .05
- Initial RNAP count (nr), Default, 30
- Initial promoter count (ng), Default, 2

Myers et al., Bioinformatics (2009)

iBioSim: SBML Editor



Model ID: CICII.xml
Model Name: Created from CICII.xml

List of Compartments:

default 1.0

List of Species:

bound_PR_CI default 0.0
CI default 0.0
CII default 0.0
PR default 2.0
PRE default 2.0
RNAP default 30.0
RNAP_PR default 0.0
RNAP_PRE default 0.0
RNAP_PRE_CII default 0.0

List of Reactions:

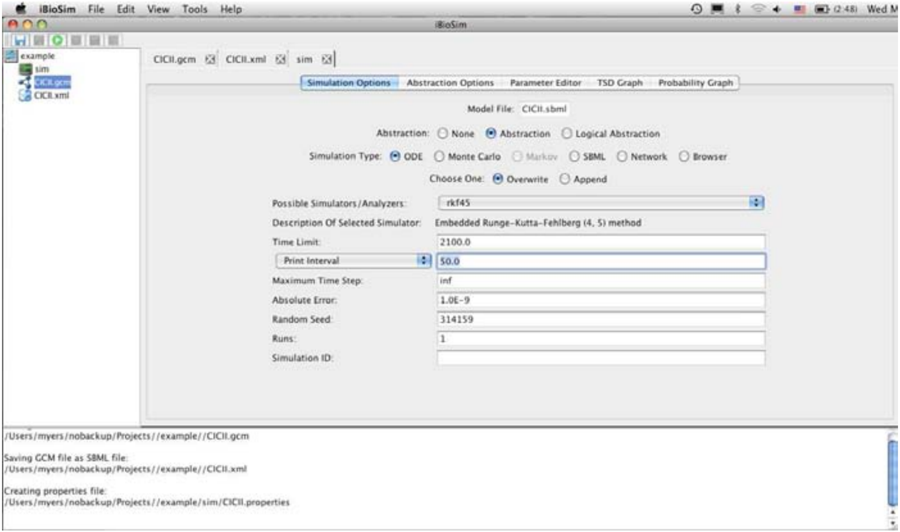
Degradation_CI
Degradation_CII
R_act_production_PRE_CII
R_basal_production_PRE
R_production_PR
R_repression_binding_PR_CI
R_RNAP_binding_PRE_CII
R_RNAP_PR
R_RNAP_PRE

List of Global Parameters:

```
/Users/myers/nobackup/Projects//example//CICII.gcm

Saving GCM file as SBML file:
/Users/myers/nobackup/Projects//example//CICII.xml

Creating properties file:
/Users/myers/nobackup/Projects//example/sim/CICII.properties
```

Myers et al., Bioinformatics (2009)

iBioSim: Analysis Engine



Model File: CICII.sbml

Abstraction: ○ None ● Abstraction ○ Logical Abstraction

Simulation Type: ● ODE ○ Monte Carlo ○ Markov ○ SBML ○ Network ○ Browser

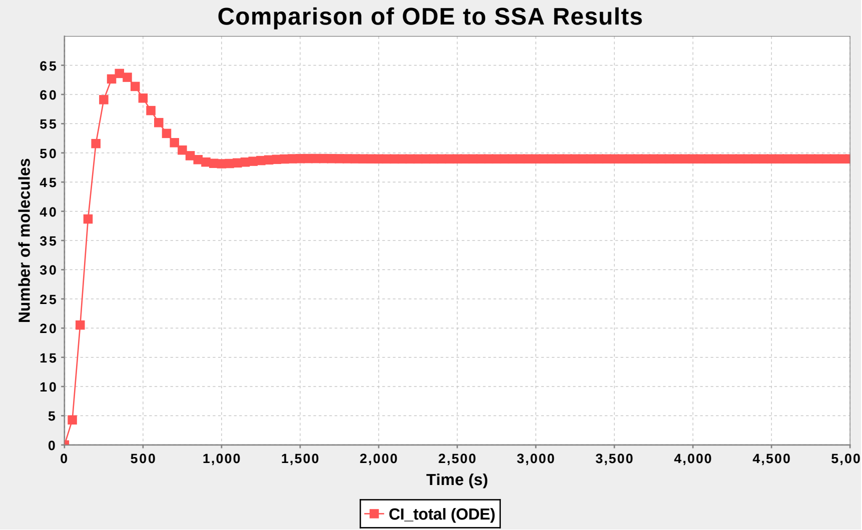
Choose One: ● Overwrite ○ Append

Possible Simulators/Analyzers:	rkf45
Description Of Selected Simulator:	Embedded Runge-Kutta-Fehlberg (4, 5) method
Time Limit:	2100.0
Print Interval	50.0
Maximum Time Step:	inf
Absolute Error:	1.0E-9
Random Seed:	314159
Runs:	1
Simulation ID:	

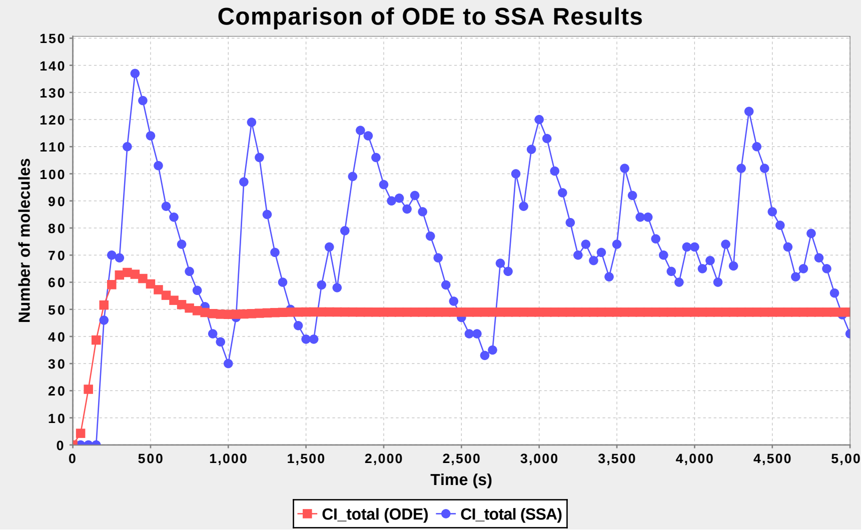
```
/Users/myers/nobackup/Projects//example//CICII.gcm

Saving GCM file as SBML file:
/Users/myers/nobackup/Projects//example//CICII.xml

Creating properties file:
/Users/myers/nobackup/Projects//example/sim/CICII.properties
```

Myers et al., Bioinformatics (2009)

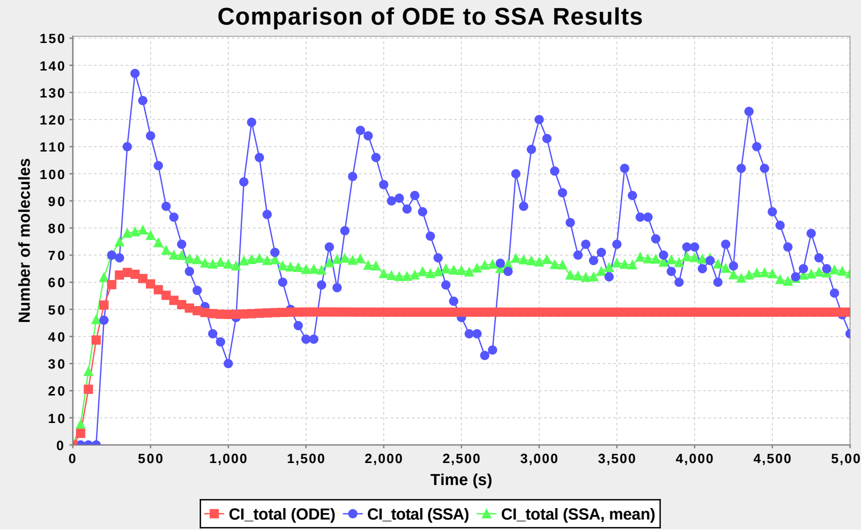
ODE Results for the Simple Genetic Oscillator



SSA Results for the Simple Genetic Oscillator



SSA Mean Results for the Simple Genetic Oscillator



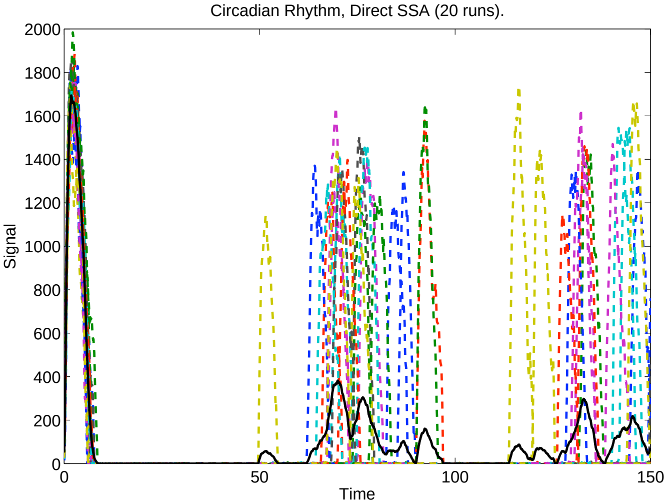
Marginal Probability Density Evolution

- The SSA predicts random behavior by generating sample paths.
- Species' statistics (mean/stdDev) are found by aggregating these paths.
- Complex systems switch states at numerous random times.
- Averaging of sample paths "washes out" meaningful behavior.
- Instead *marginal probability density evolution* (MPDE) method can be used to determine "typical" species statistics.

Winstead et al., IWBDA (2009)

Example: Circadian Rhythms

- The VKBL circadian rhythm model from Vilar (2002) and Samad (2005):



Probability Density Evolution

- Iterative form of the Chemical Master Equation (CME):

$$\begin{aligned} p\left(\mathbf{x}'\right) &= \sum_{\Omega_k} \sum_{j} p\left(\mathbf{x}' \middle| \mathbf{x}, R_j\right) p\left(\mathbf{x}, R_j\right) \\ &= E_{\mathbf{x},R}\left[p\left(\mathbf{x}' \middle| \mathbf{x}, R_j\right)\right]. \end{aligned}$$

  where:

  - $\mathbf{x}$ is the system state at time t.
  - $\mathbf{x}'$ is the state at time $t + dt$.
  - Ω_k is the domain of $\mathbf{x}$.
  - R_j are the possible reactions (R_0 is no-reaction).

Conditional Independence Approximation

- Suppose the elements of $\mathbf{x}'$ are **conditionally independent**, given $\mathbf{x}$ and a **sequence** of reaction events $\mathbf{R}$, so that

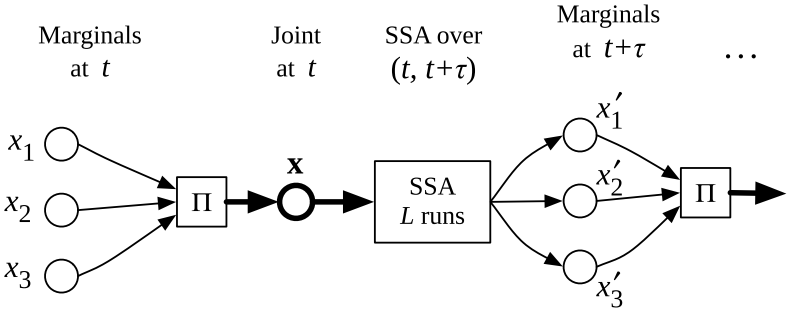
$$p\left(\mathbf{x}' \mid \mathbf{x}, \mathbf{R}\right) = \prod_{i=1}^{M} p\left(x_i' \mid \mathbf{x}, \mathbf{R}\right)$$

- Assuming that the covariances are small, then the updated joint probability density can be written as

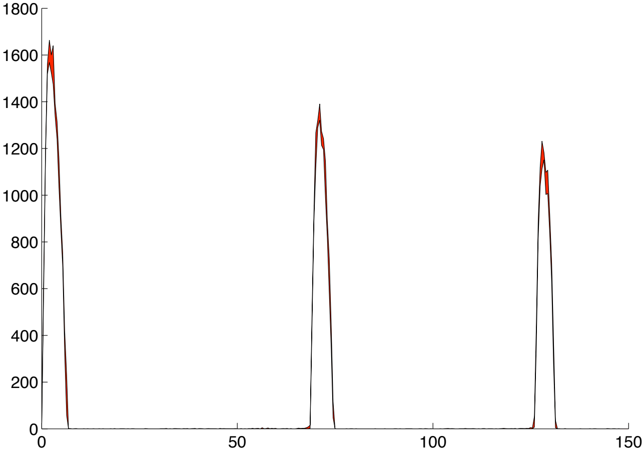
$$\begin{aligned} p\left(\mathbf{x}'\right) &= E_{\mathbf{x},\mathbf{R}}\left[\prod_{i=1}^{M} p\left(x_i' \mid \mathbf{x}, \mathbf{R}\right)\right] \\ &= \prod_{i=1}^{M} E_{\mathbf{x},\mathbf{R}}\left[p\left(x_i' \mid \mathbf{x}, \mathbf{R}\right)\right]. \end{aligned}$$

- This approximation allows evolving the **marginal distributions** for x_i', rather than the joint distribution for **x**.

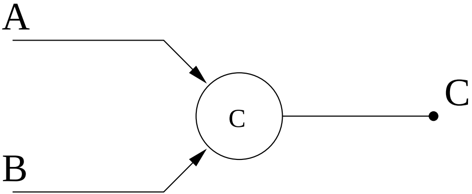
SSA-based MPDE



MPDE Results: Circadian Rhythm Example

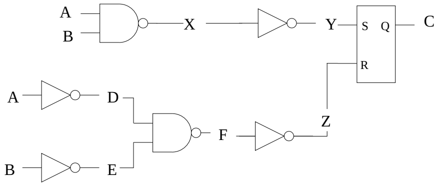


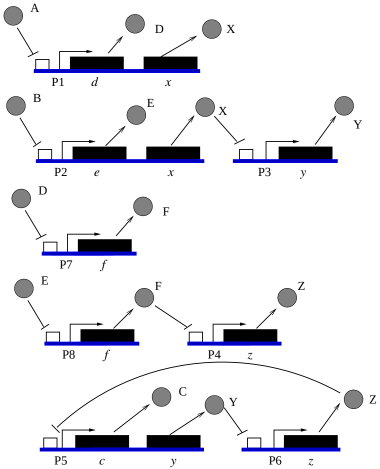
Genetic Muller C-Element



A	B	C’
0	0	0
0	1	C
1	0	C
1	1	1

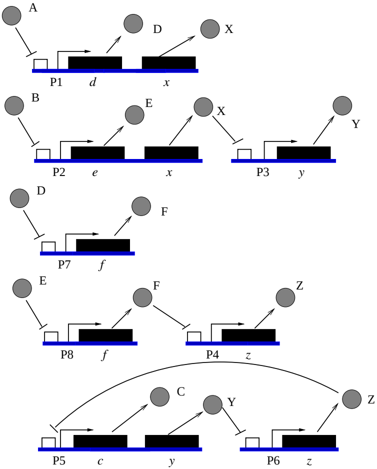
Toggle Switch C-Element (Genetic Circuit)

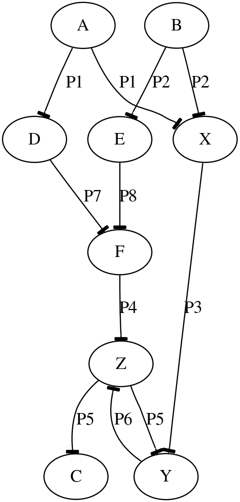




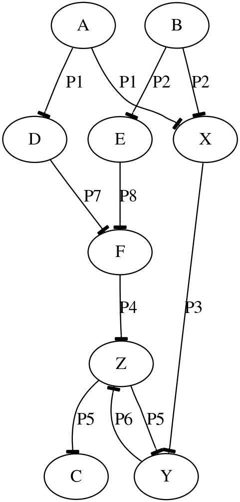
Nguyen et al., 13th Symposium on Async. Ckts. & Sys., 2007 (**best paper**)

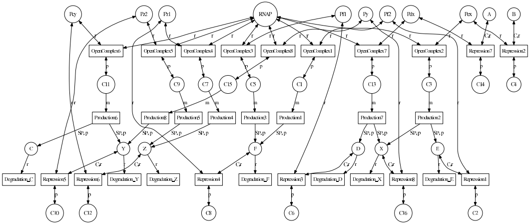
Toggle Switch C-Element (GCM)



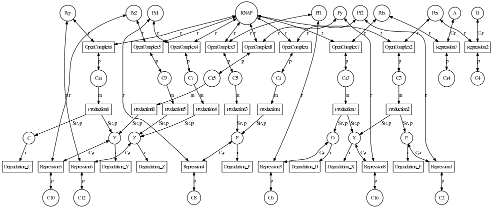


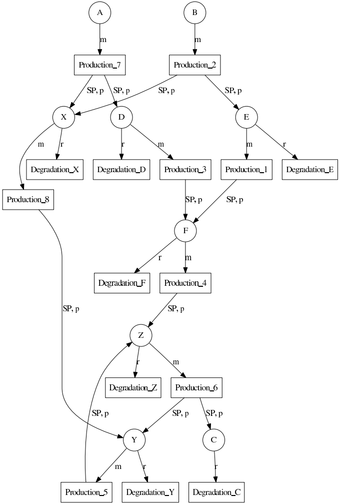
Toggle Switch C-Element (SBML)





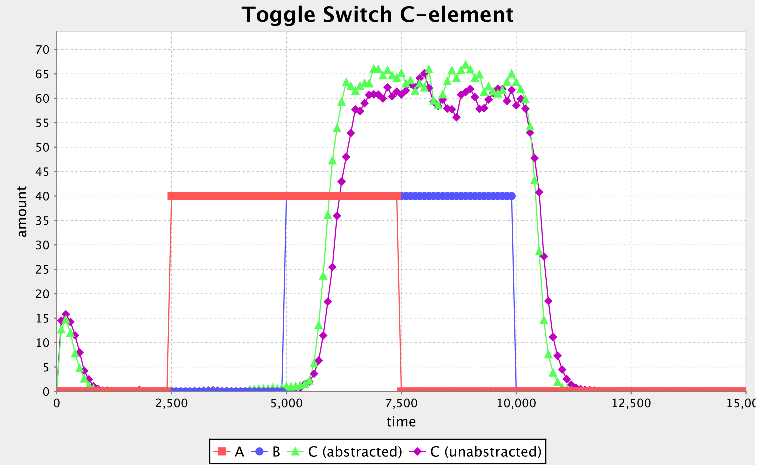
Toggle Switch C-Element (Abstracted)





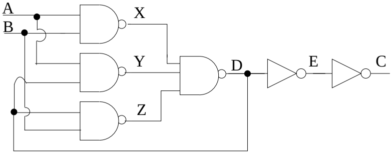
Reduced from 34 species and 31 reactions to 9 species and 15 reactions.

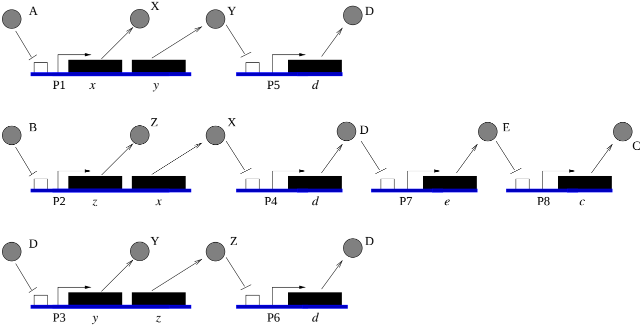
Toggle Switch C-Element (Simulation)



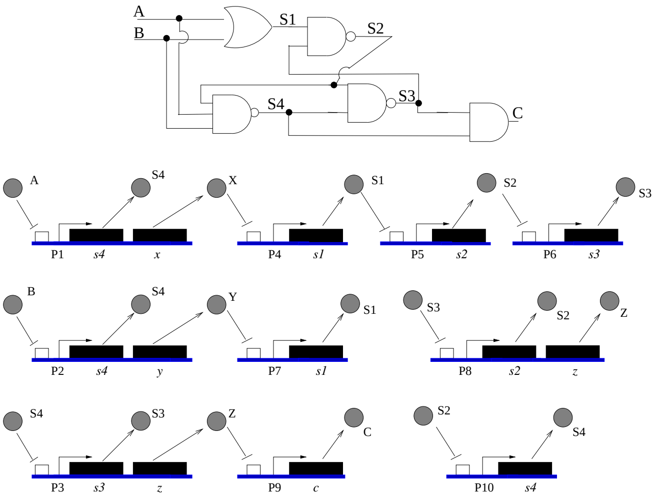
Simulation time improved from 312 seconds to 20 seconds.

Majority Gate C-Element (Genetic Circuit)

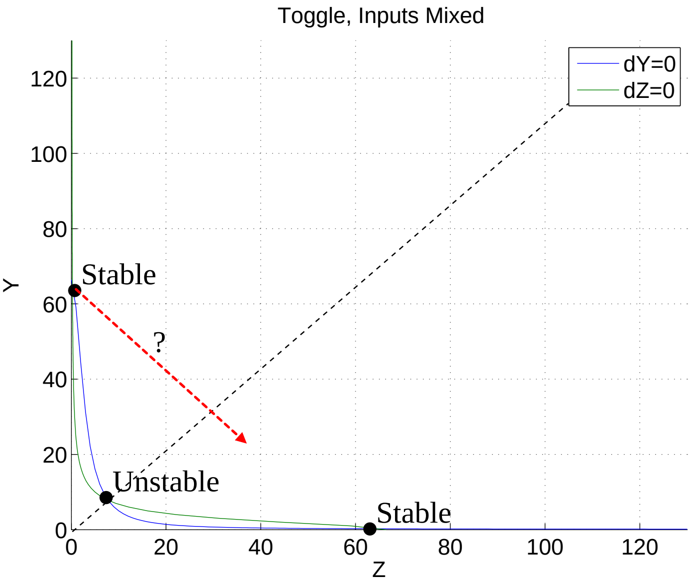




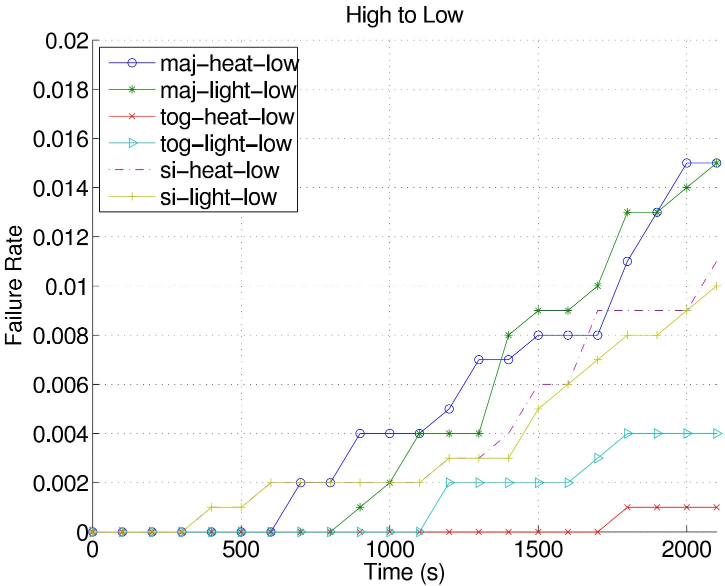
Speed-Independent C-Element (Genetic Circuit)



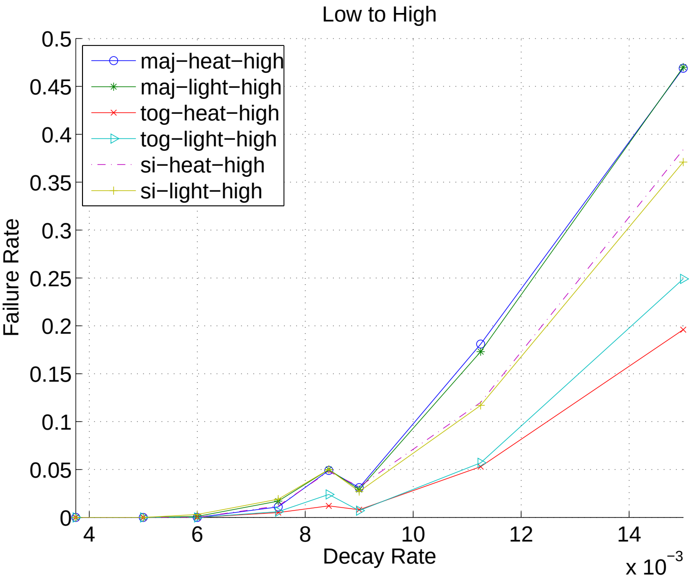
Nullclines and Probability of Failure



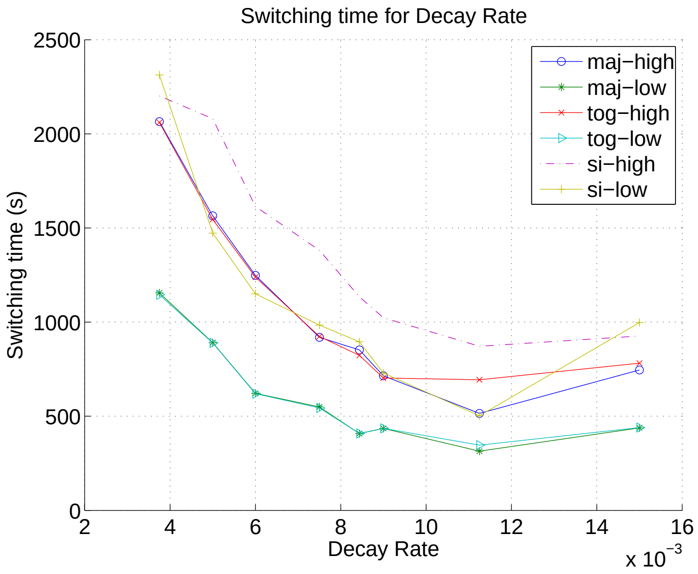
Comparison of Failure Rates for the C-element Designs



Effects of Decay Rates

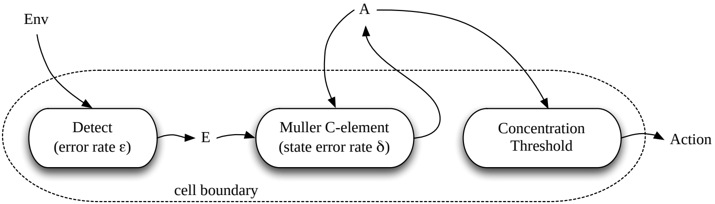


Effects of Decay Rates



Application: Bacterial Consensus

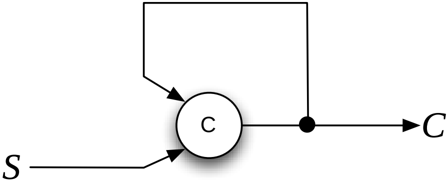
- One interesting application is designing bacteria that can hunt and kill tumor cells (Anderson et al.).
- Care must be taken in determining when to attack potential tumor cells.
- Can use a genetic Muller C-element and a bacterial consensus mechanism known as *quorum sensing*.
- C-element combines a noisy environmental trigger signal and a density dependent quorum sensing signal.
- Activated bacteria signal their neighbors to reach consensus.



Winstead et al., IBE Conference (2008)

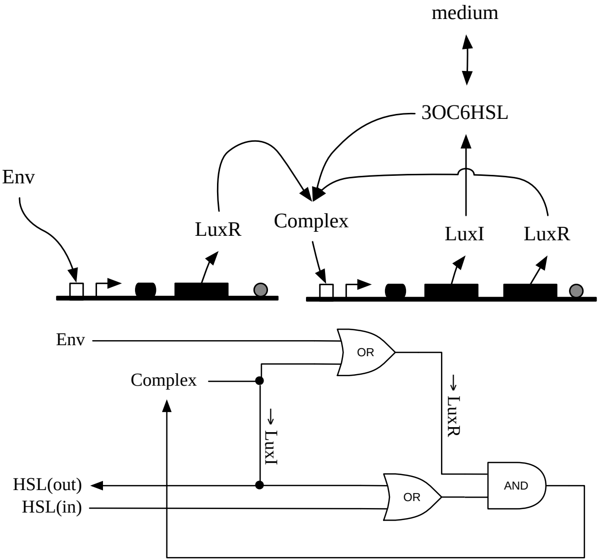
Confidence Amplifier

- A noisy C-element with a confidence-feedback loop:

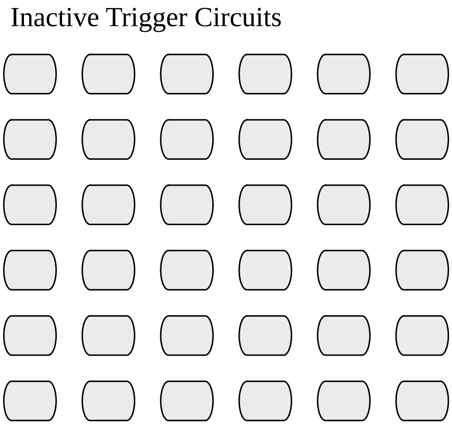


- The output “rails” to maximum confidence, even if *S* has low confidence.
- This configuration only works if the C-element is “noisy”. Otherwise, the circuit is permanently stuck in its initial state.

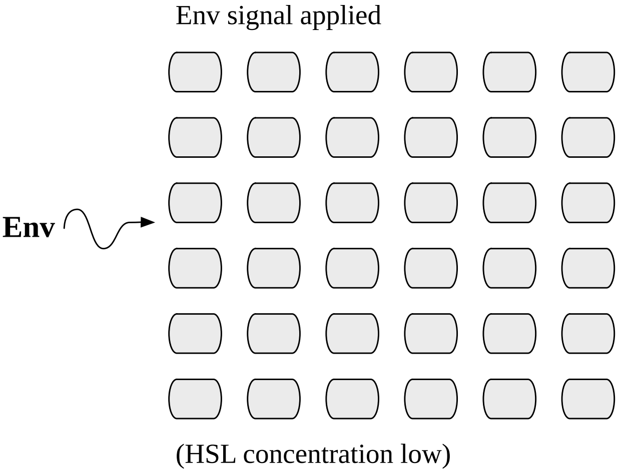
Quorum Trigger Circuit



Population Dynamics

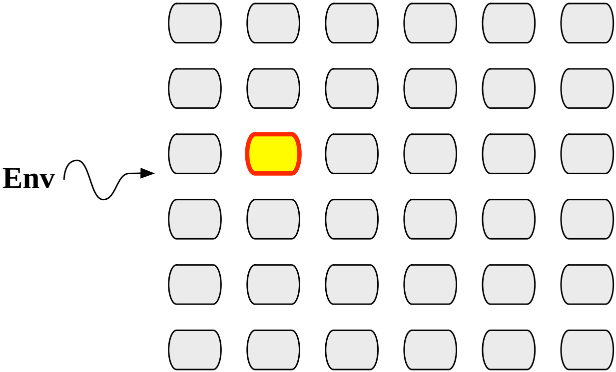


Population Dynamics



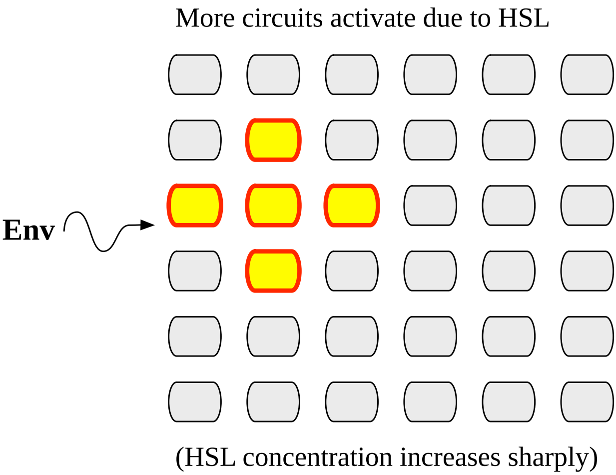
Population Dynamics

One circuit randomly activates



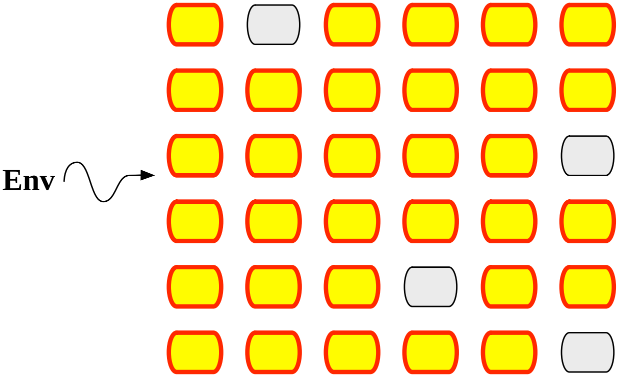
(HSL concentration increases)

Population Dynamics



Population Dynamics

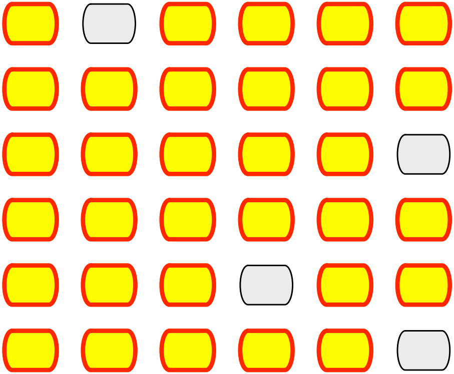
Avalanche effect: most cells activate



(HSL concentration saturates)

Population Dynamics

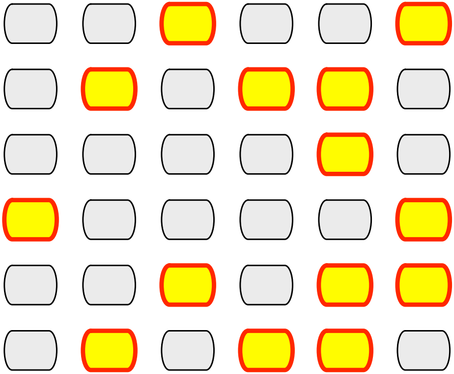
Env signal is removed.



(Circuits stay active)

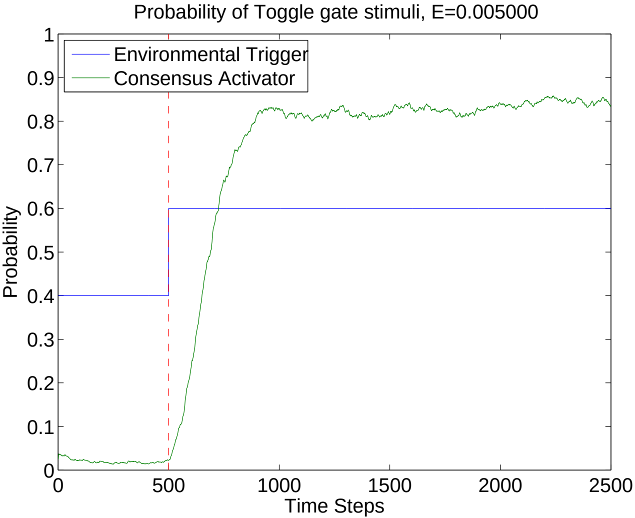
Population Dynamics

Time passes.

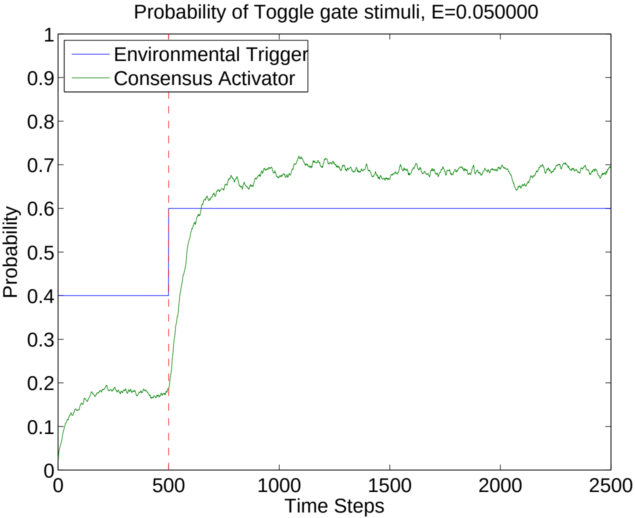


(Circuits randomly switch off)

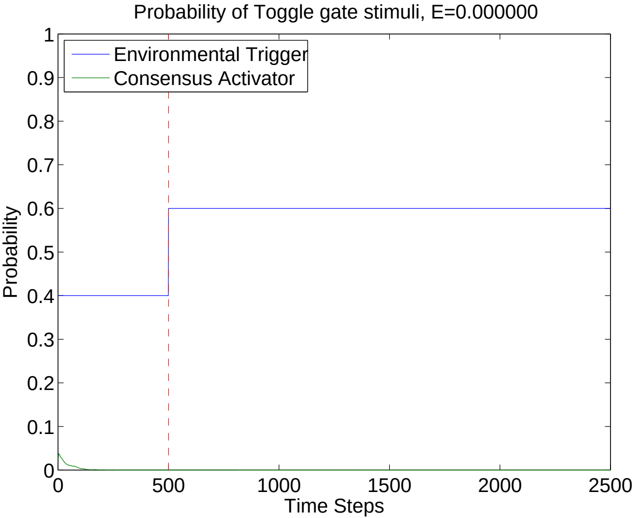
Simulation Results



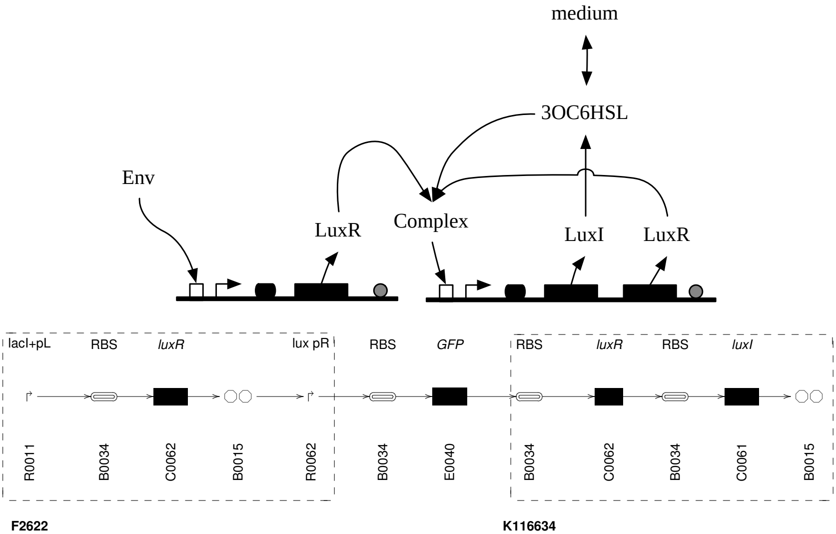
Simulation Results



Simulation Results



Quorum Trigger Design



Future GDA Research Directions

- Genetic circuits have no signal isolation.
- Circuit products may interfere with each other and host cell.
- Gates in a genetic circuit library usually can only be used once.
- Behavior of circuits are non-deterministic in nature.
- No global clock, so timing is difficult to characterize.
- To address these challenges, we are investigating soft logic models based on *factor graphs* and adapting asynchronous synthesis tools to a genetic circuit technology.

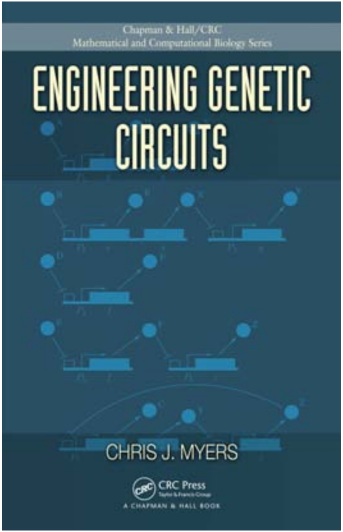
Biologically Inspired Circuit Design

- Human inner ear performs the equivalent of one billion floating point operations per second and consumes only 14 μW while a game console with similar performance burns about 50 W (Sarpeshkar, 2006).
- We believe this difference is due to over designing components in order to achieve an extremely low probability of failure in every device.
- Future silicon and nano-devices will be much less reliable.
- For Moore's law to continue, future design methods should support the design of reliable systems using unreliable components.
- Biological systems constructed from very noisy and unreliable devices.
- GDA tools may be useful for future integrated circuit technologies.

More Information

- Linux/Windows/Mac versions of `iBioSim` are freely available from:
  `http://www.async.ece.utah.edu/iBioSim/`
- Publications:
  `http://www.async.ece.utah.edu/publications/`
- Course materials:
  `http://www.async.ece.utah.edu/∼myers/ece6760/`
  `http://www.async.ece.utah.edu/∼myers/math6790/`

Engineering Genetic Circuits



Chapman & Hall/CRC
Mathematical and Computational Biology Series

ENGINEERING GENETIC CIRCUITS

CHRIS J. MYERS

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

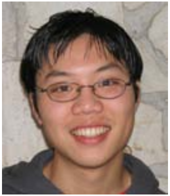
Acknowledgments


Nathan Barker


Keven Jones


Hiroyuki Kuwahara


Curtis Madsen


Nam Nguyen


Chris Winstead

This work is supported by the National Science Foundation under Grants No. 0331270, CCF-07377655, and CCF-0916042.