BioSimWare: A P Systems-based Simulation Environment for Biological Systems

Daniela Besozzi[1], Paolo Cazzaniga[2], Giancarlo Mauri[2], Dario Pescini[2]

Università degli Studi di Milano
Dipartimento di Informatica e Comunicazione
Via Comelico 39, 20135 Milano, Italy
besozzi@dico.unimi.it

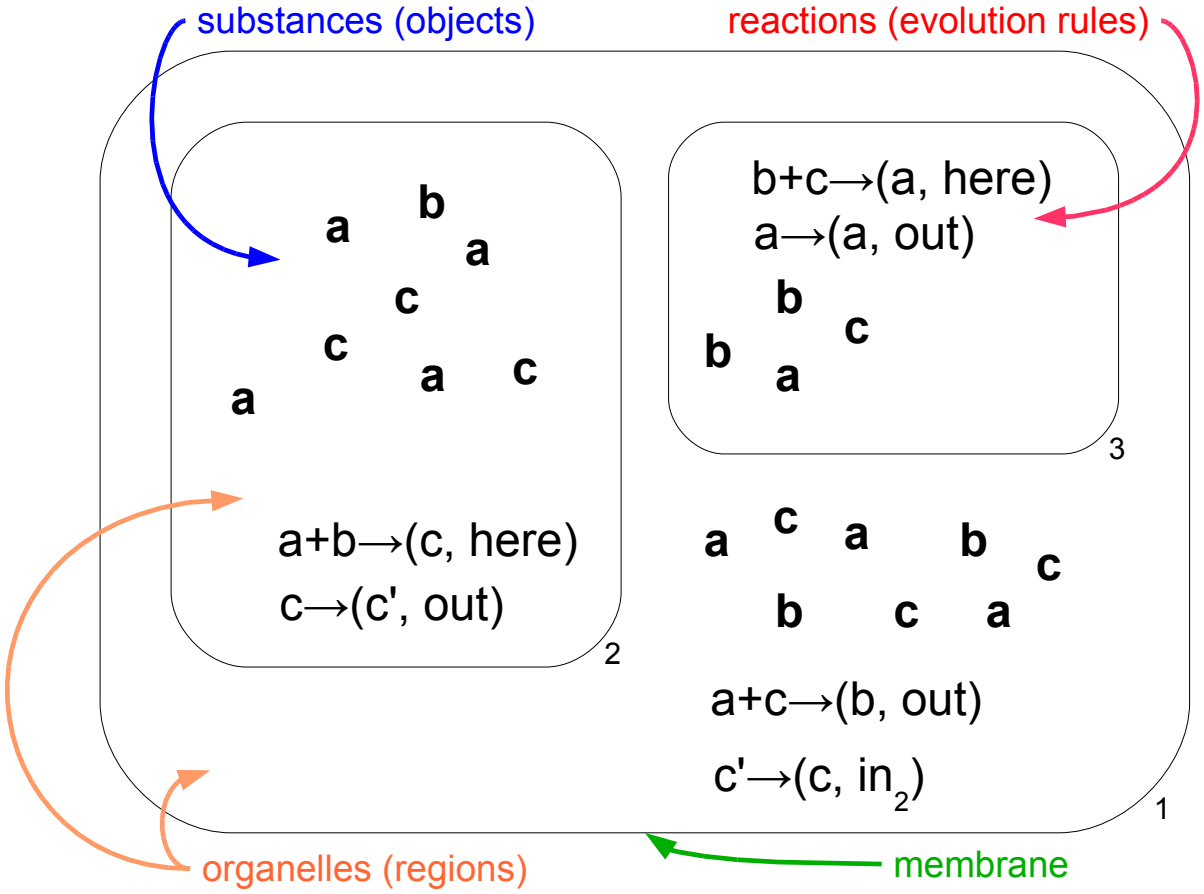
Università degli Studi di Milano-Bicocca
Dipartimento di Informatica, Sistemistica e Comunicazione
Viale Sarca 336, 20126 Milano, Italy
cazzaniga/mauri/pescini@disco.unimib.it

Outline

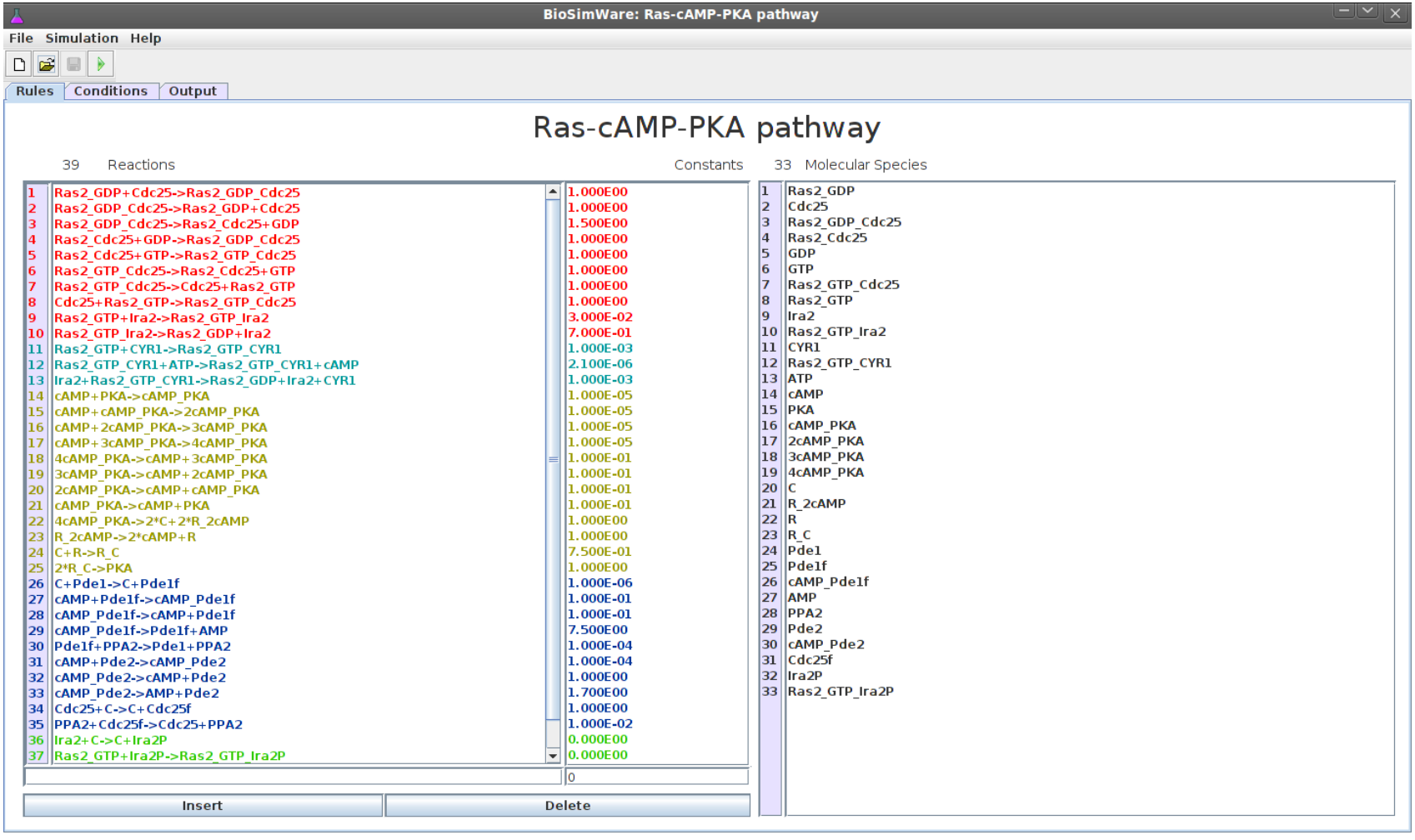
Outline

Introduction

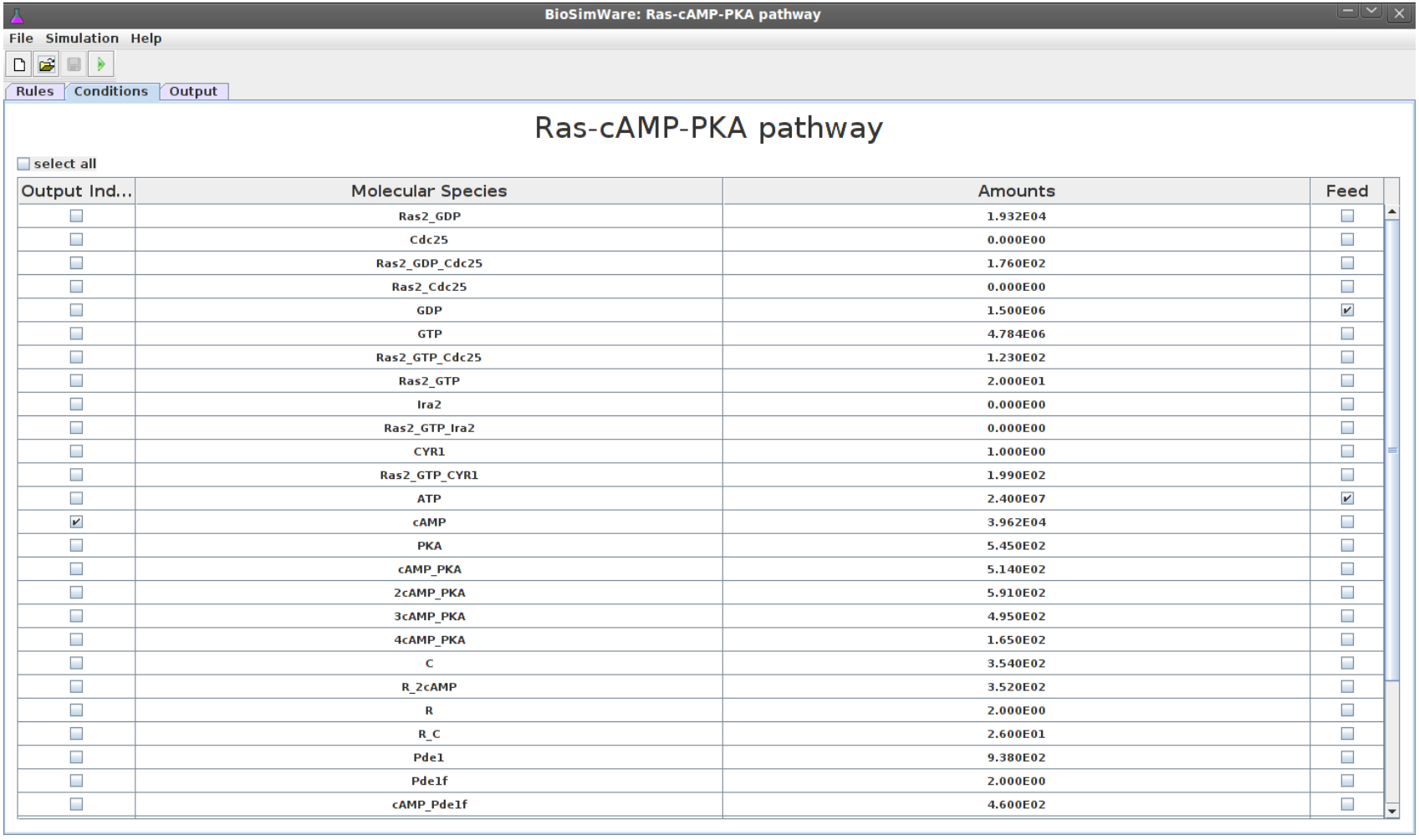
BioSimWare is a simulation environment based on P systems, that provides a user-friendly framework for the modeling of biological systems ranging from cellular processes to population phenomena.



User Interface: rules specification

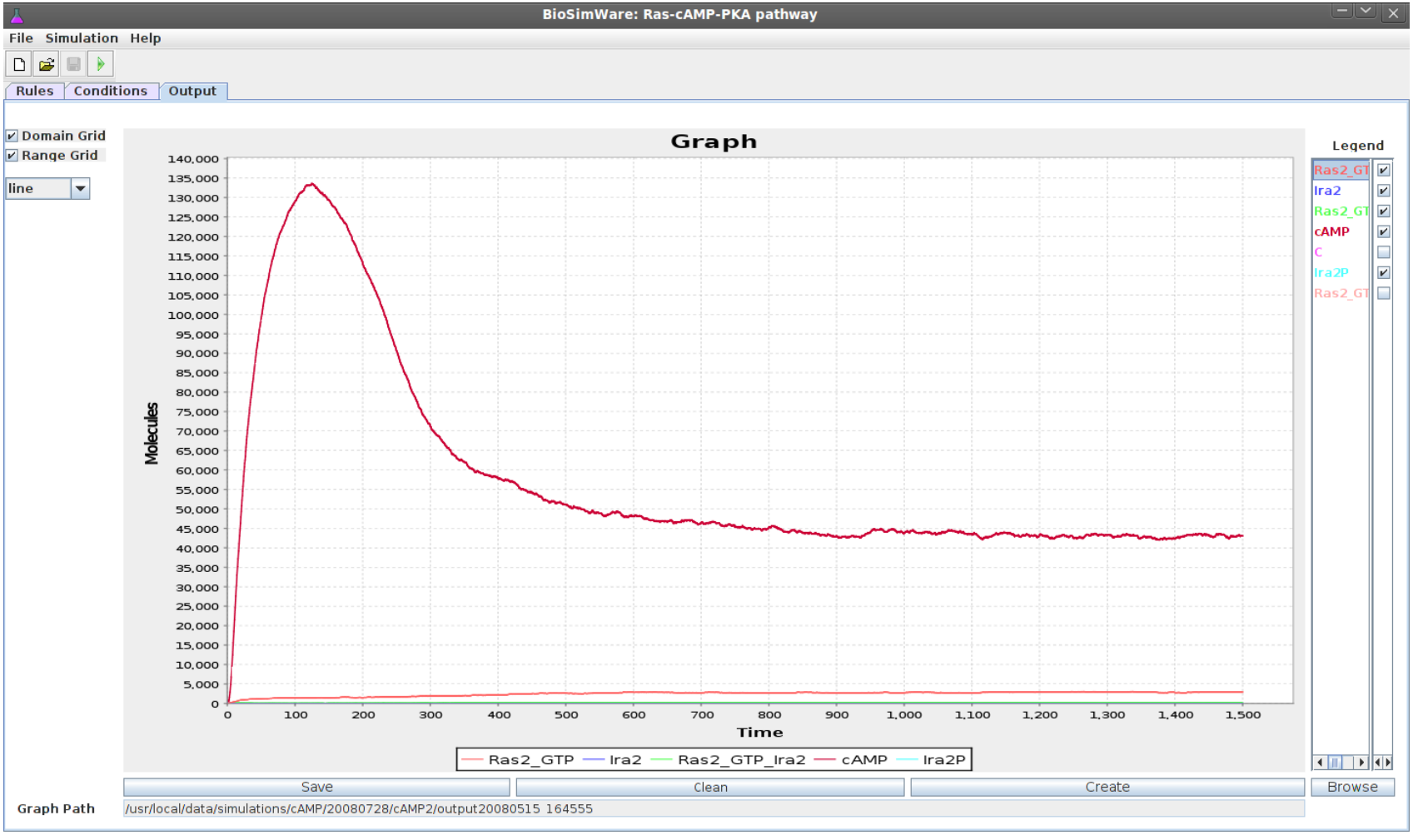


User Interface: system conditions specification



Output Ind...	Molecular Species	Amounts	Feed
☐	Ras2_GDP	1.932E04	☐
☐	Cdc25	0.000E00	☐
☐	Ras2_GDP_Cdc25	1.760E02	☐
☐	Ras2_Cdc25	0.000E00	☐
☐	GDP	1.500E06	☑
☐	GTP	4.784E06	☐
☐	Ras2_GTP_Cdc25	1.230E02	☐
☐	Ras2_GTP	2.000E01	☐
☐	Ira2	0.000E00	☐
☐	Ras2_GTP_Ira2	0.000E00	☐
☐	CYR1	1.000E00	☐
☐	Ras2_GTP_CYR1	1.990E02	☐
☐	ATP	2.400E07	☑
☑	cAMP	3.962E04	☐
☐	PKA	5.450E02	☐
☐	cAMP_PKA	5.140E02	☐
☐	2cAMP_PKA	5.910E02	☐
☐	3cAMP_PKA	4.950E02	☐
☐	4cAMP_PKA	1.650E02	☐
☐	C	3.540E02	☐
☐	R_2cAMP	3.520E02	☐
☐	R	2.000E00	☐
☐	R_C	2.600E01	☐
☐	Pde1	9.380E02	☐
☐	Pde1f	2.000E00	☐
☐	cAMP_Pde1f	4.600E02	☐

User Interface: plot of the dynamics



Outline

Available simualtion algorithm

Single Volume

- SSA
- tau leaping
- adaptive tau leaping
- average tau leaping dynamics
- ODE integrator

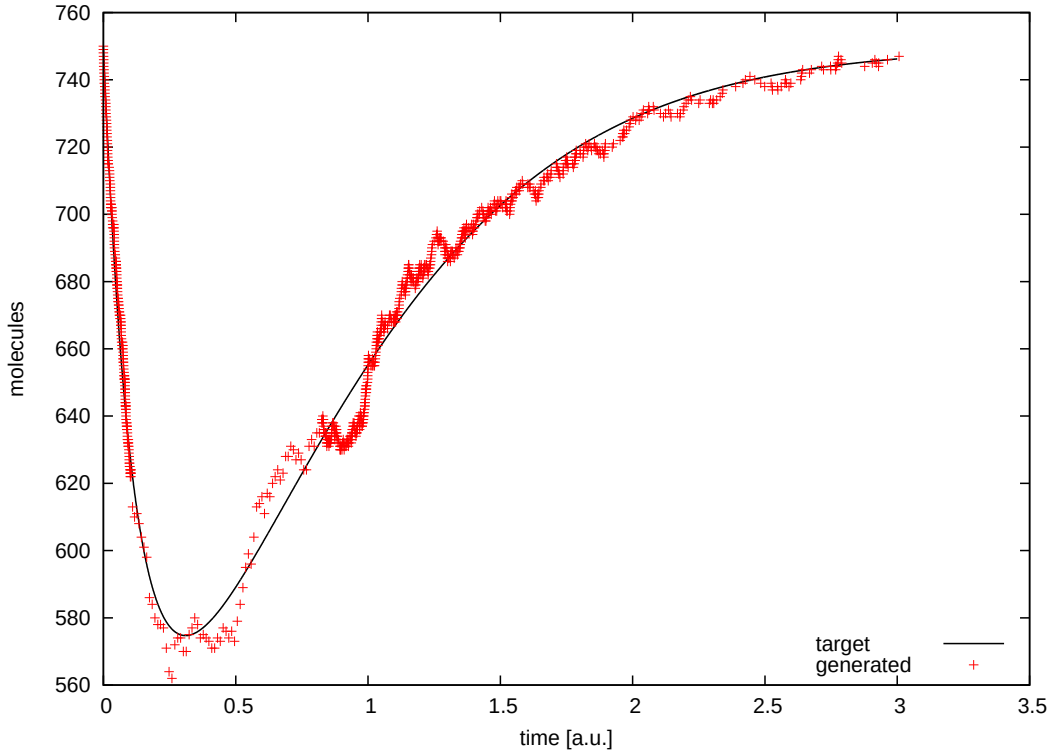
Multi Volume

- DPP
- tau-DPP
- Stau-DPP

Outline

PE: the goal

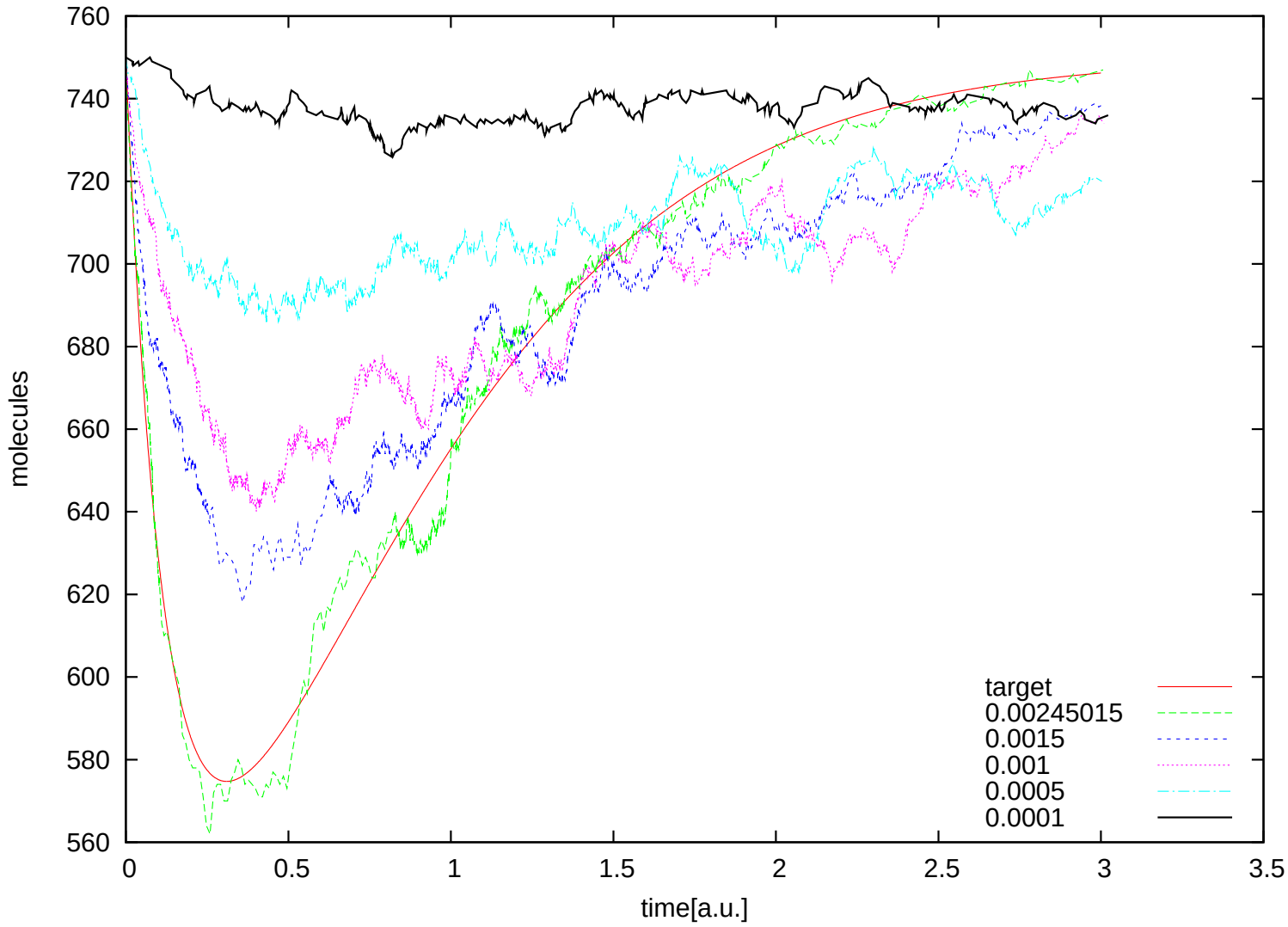
We would like to reproduce the target dynamics via stochastic simulations:



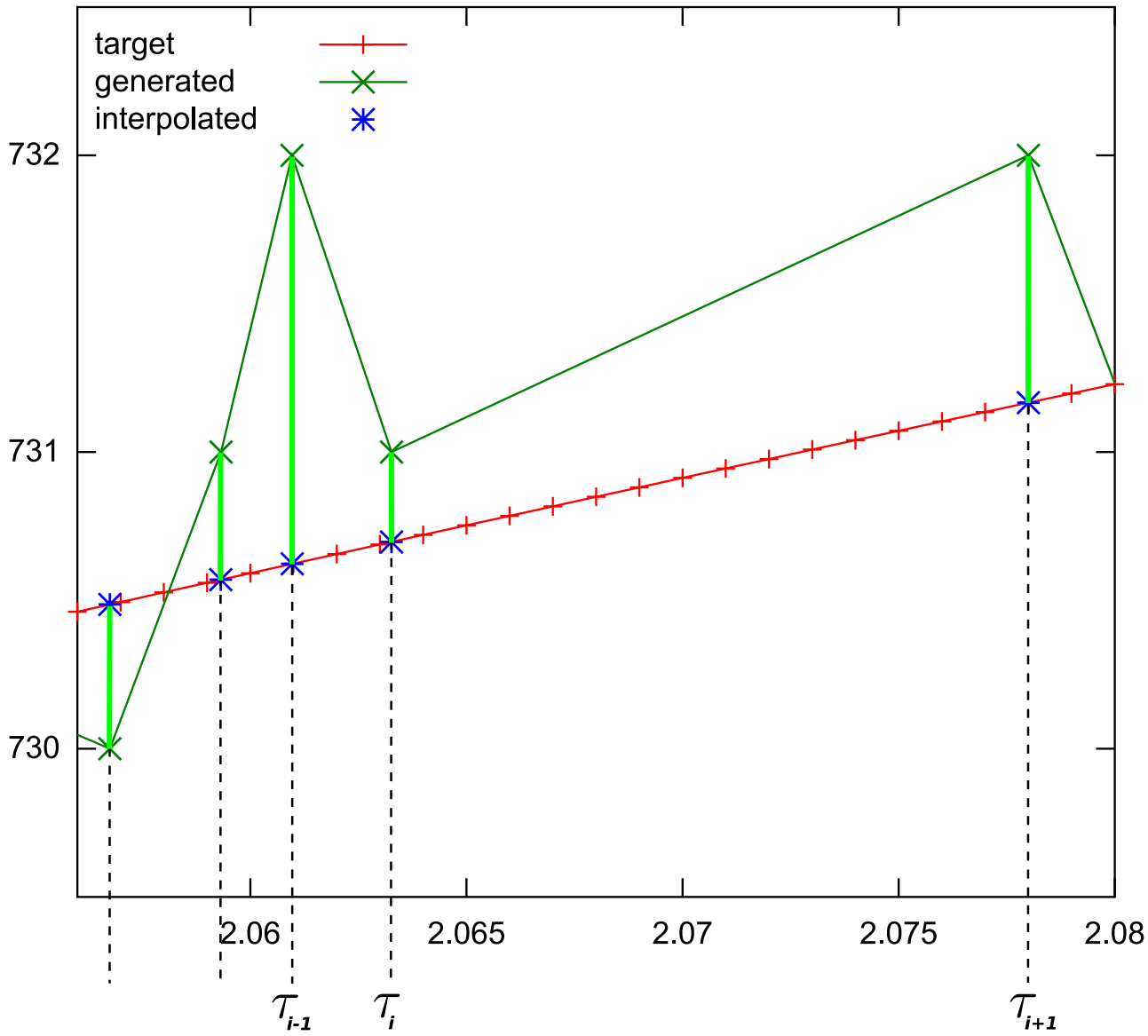
not necessarily using the same set of constants.

PE: the problem

Quantify the “distance” among the dynamics and find the closest one



The fitness: Standard Distance

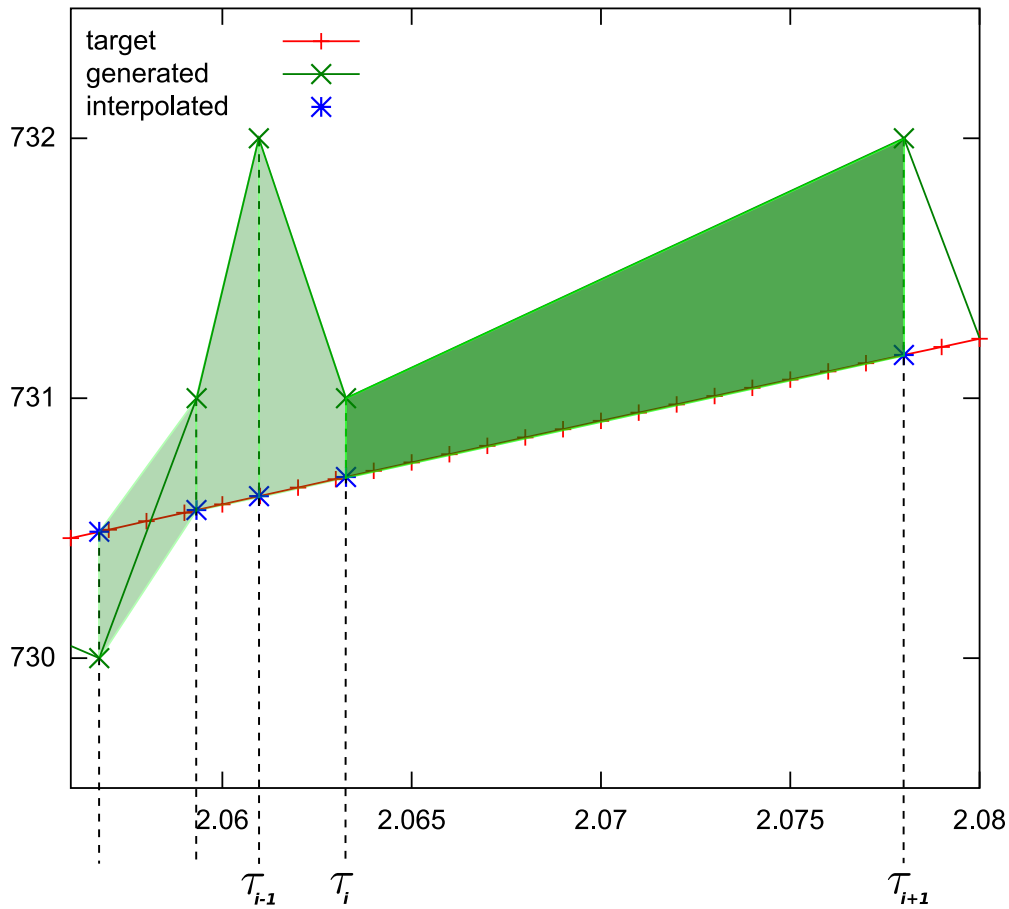


The fitness: tricks

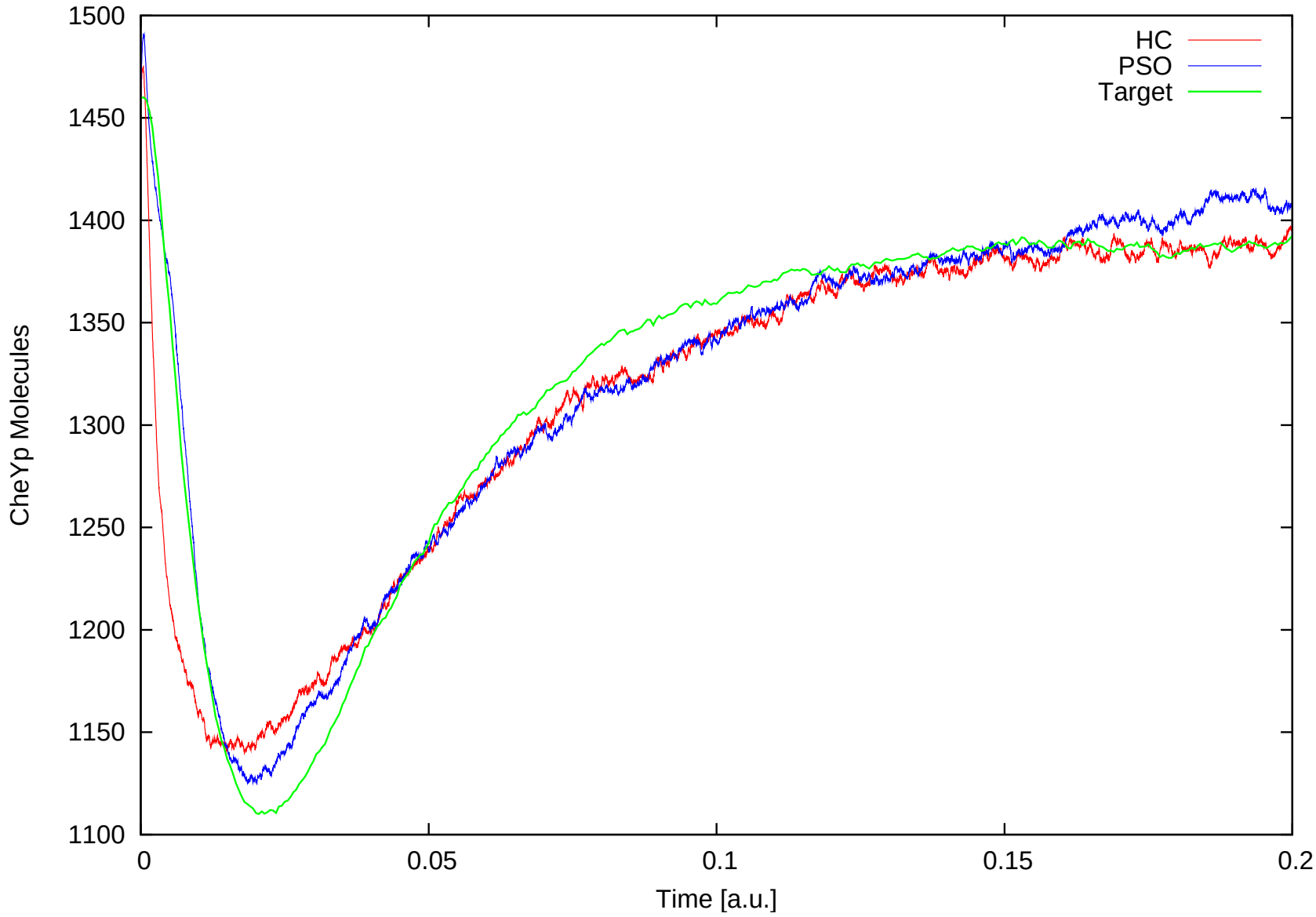
Facts related to intrinsic noise that should not be neglected:

- each outcome j is quantitatively different $\{X(\tau_i)\}_j$
  - not evenly sampled $\{\tau_i\}_j$
  - not same number of points
  - $\{\neq X(\tau_i)\}_j$
- in the thermodynamic limit $\{X(\tau_i)\} \to [X](t_i)$
  - exploit ensemble behavior (may flatten oscillations, not in phase outcome)
  - $\langle\, F\,(X(\tau_i))\,\rangle \stackrel{?}{=} F\,(\,\langle X(\tau_i)\rangle\,)$
- same parameters, possibly (almost always), generate different values of the fitness function ("weak convergence")

The fitness: “Area” distance



Reconstructed dynamics



Parameter sweep applications (PSA)

A PSA consists in a repeated execution of an application (usually performed a large number of times), where each execution is achieved using a different parametrisation.
In the context of biochemical stochastic models a PSA could be viewed as $PSA = (P; D; M)$, where

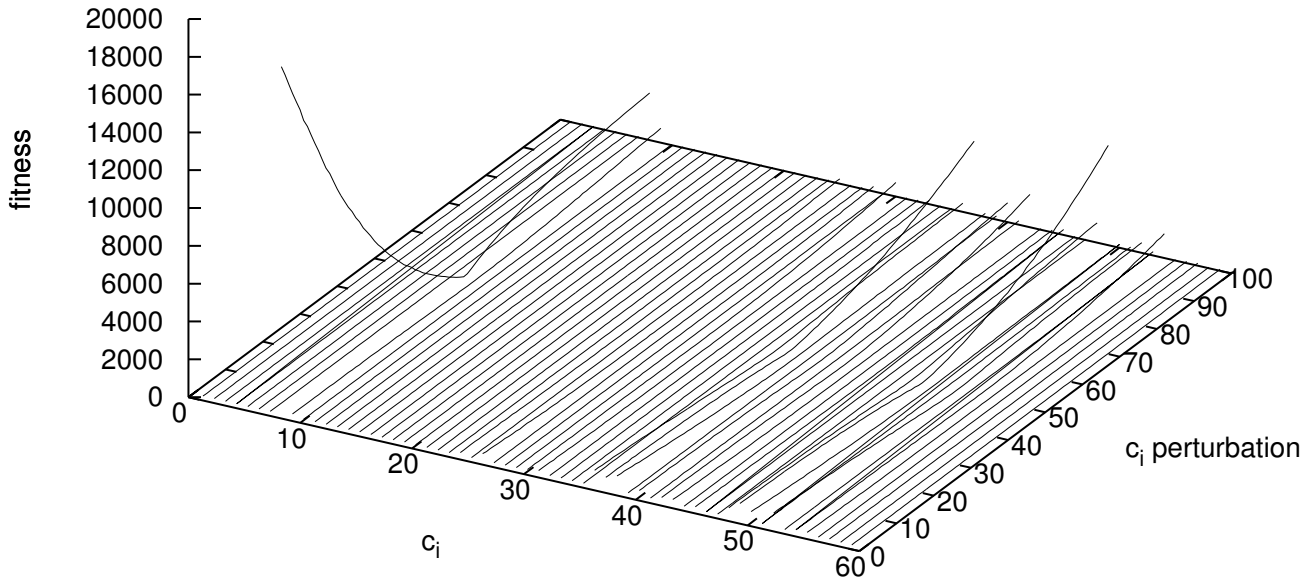
- $P = (p_1; \ldots; p_n)$ is the set of parametrisations with $p_i = (x_0^i; c^i)$
- $D = (d_1; \ldots; d_n)$ $\quad$ $d_i = (x_i(0); \ldots; x_i(t))$ where t is the halting time of the simulation i
- M is the biochemical model that provides the map $p_i \rightarrow d_i$

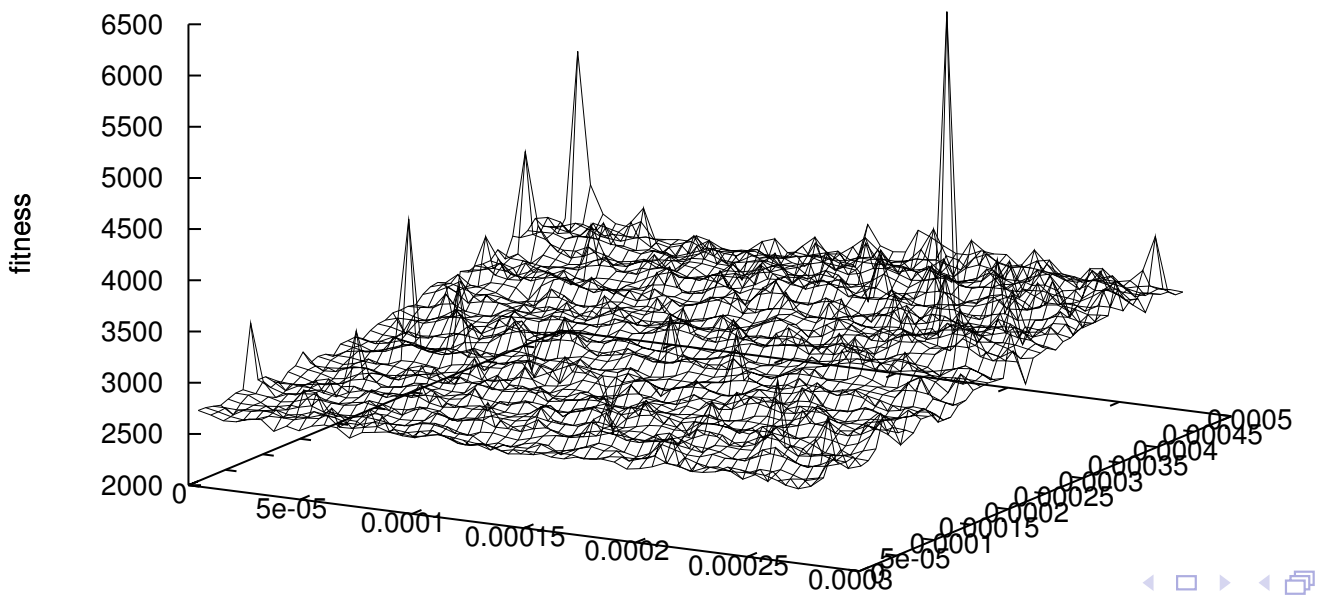
PSA: EGEE Grid

EGEE project infrastructure, a wide area grid platform for scientific applications, composed of thousands of CPUs, which implements the Virtual Organisation (VO) paradigm.

- More than 90 partners in 32 countries, organised in 13 Federations
- A Grid infrastructure spanning almost 240 sites across 45 countries
- An infrastructure of 41,000 CPU available to users 24 hours a day, 7 days a week
- More than 5 Petabytes (5 million Gigabytes) of storage
- Sustained and regular workloads of 30K jobs/day, reaching up to 98K jobs/day

Perturbed parameters fitness





Outline

The Schlögl System

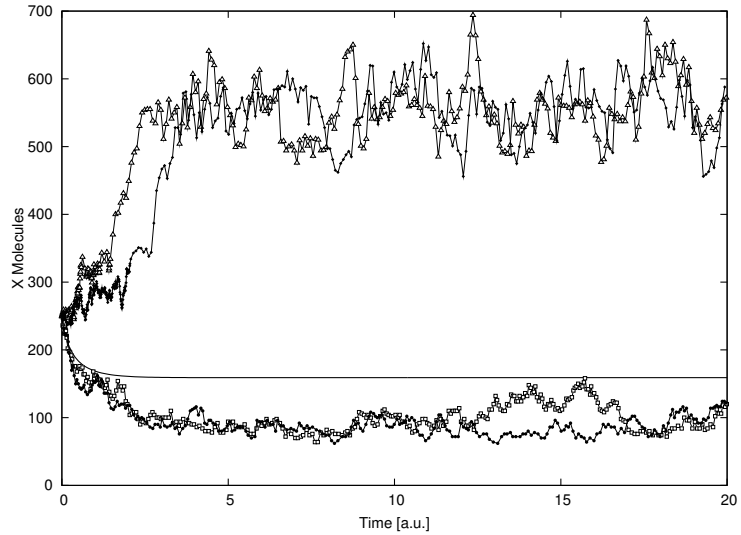
$$r_1 : A + 2X \xrightarrow{3 \cdot 10^{-7}} 3X$$

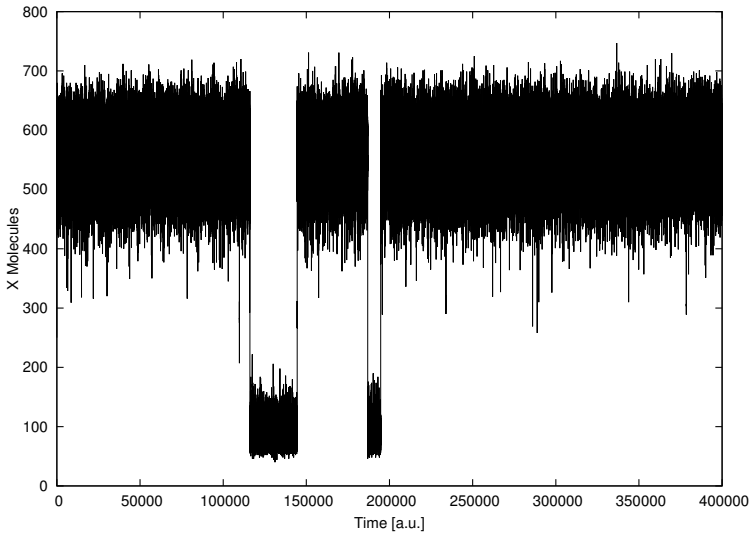
$$r_2 : 3X \xrightarrow{10^{-4}} A + 2X$$

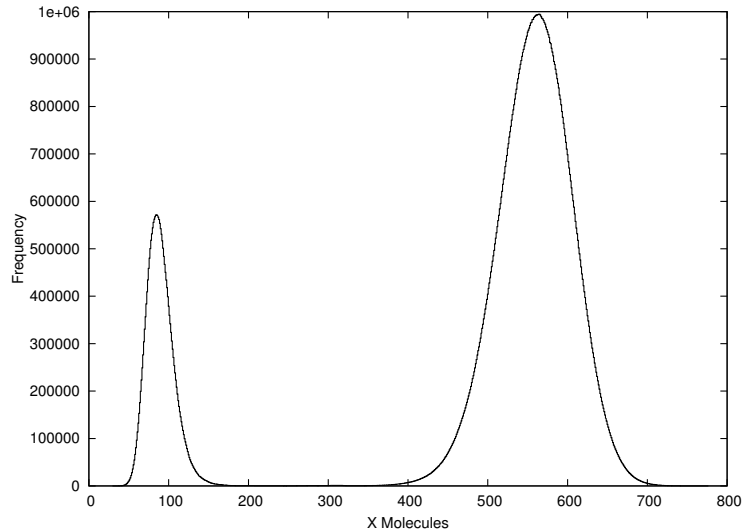
$$r_3 : B \xrightarrow{10^{-3}} 3X$$

$$r_4 : X \xrightarrow{3.5} B$$

$$A = 1 \cdot 10^5 \quad B = 2 \cdot 10^5 \quad X = 250$$







The Brusselator

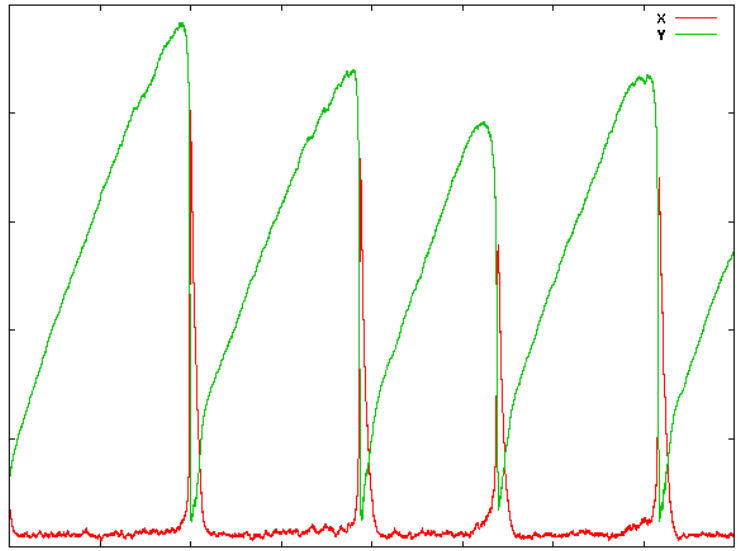
$$r_1 : A \xrightarrow{1} X$$

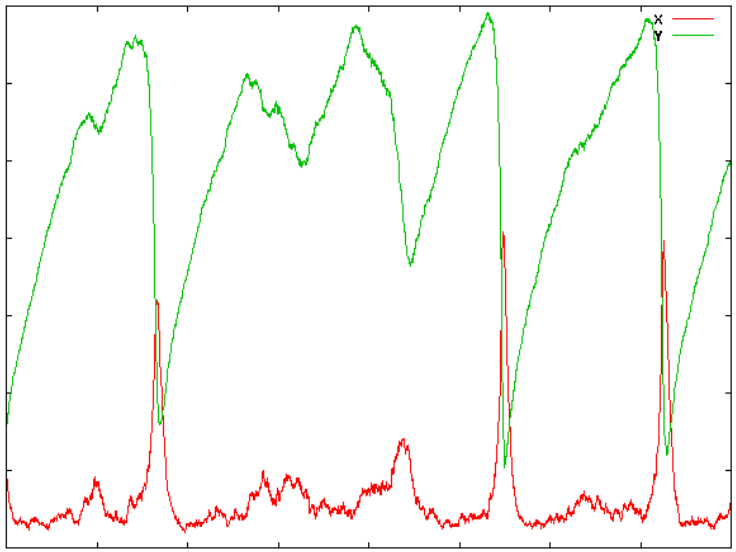
$$r_2 : B + X \xrightarrow{5 \cdot 10^{-3}} Y$$

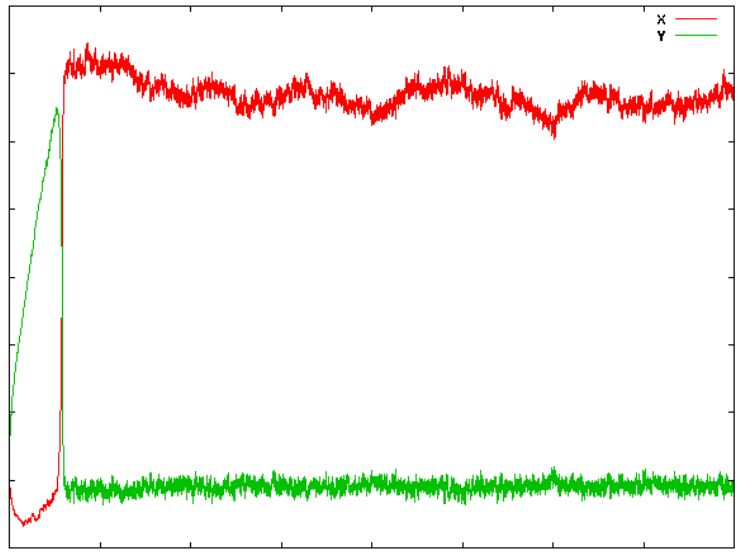
$$r_3 : 2X + Y \xrightarrow{2.5 \cdot 10^{-5}} 3X$$

$$r_4 : X \xrightarrow{1.5} \lambda$$

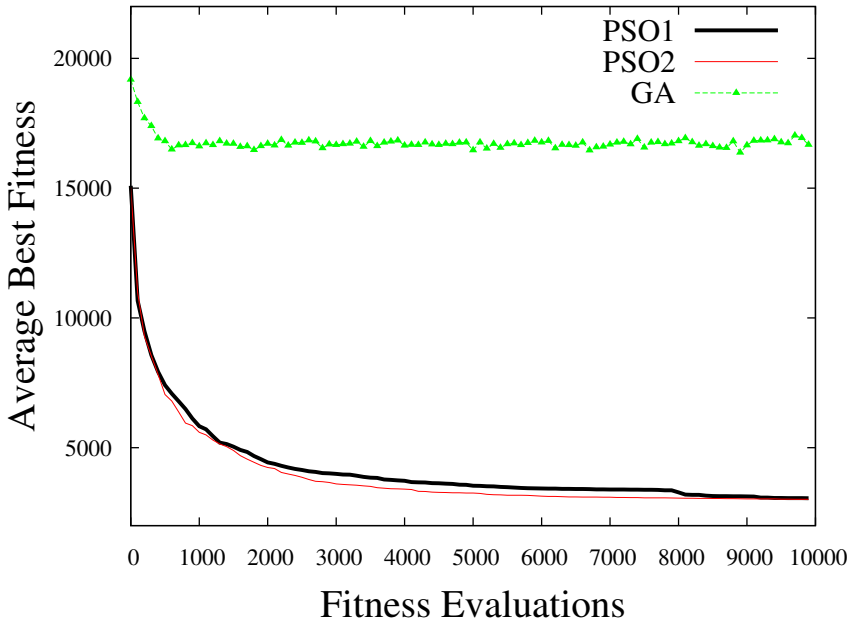
$$A = X = 200 \; B = 600 \; Y = 300$$

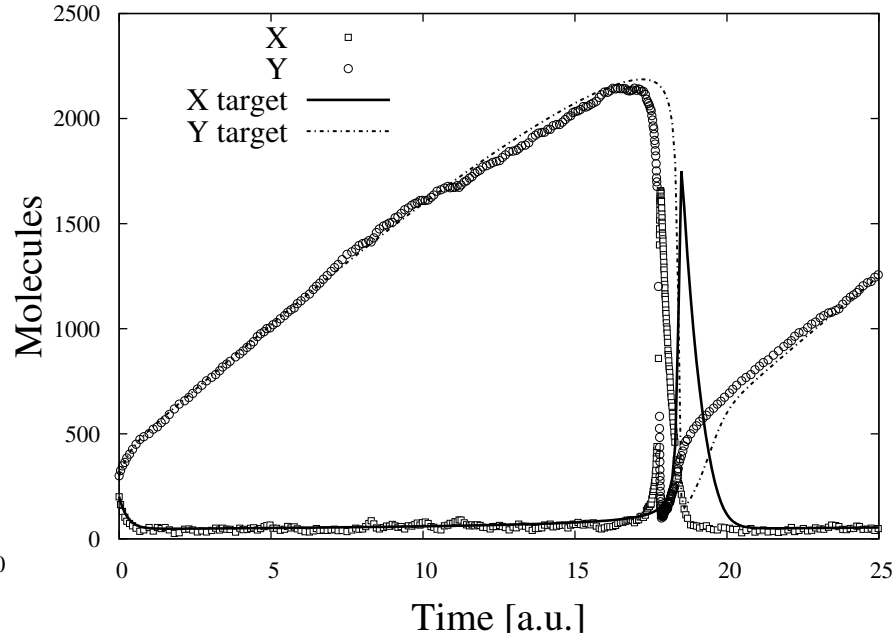


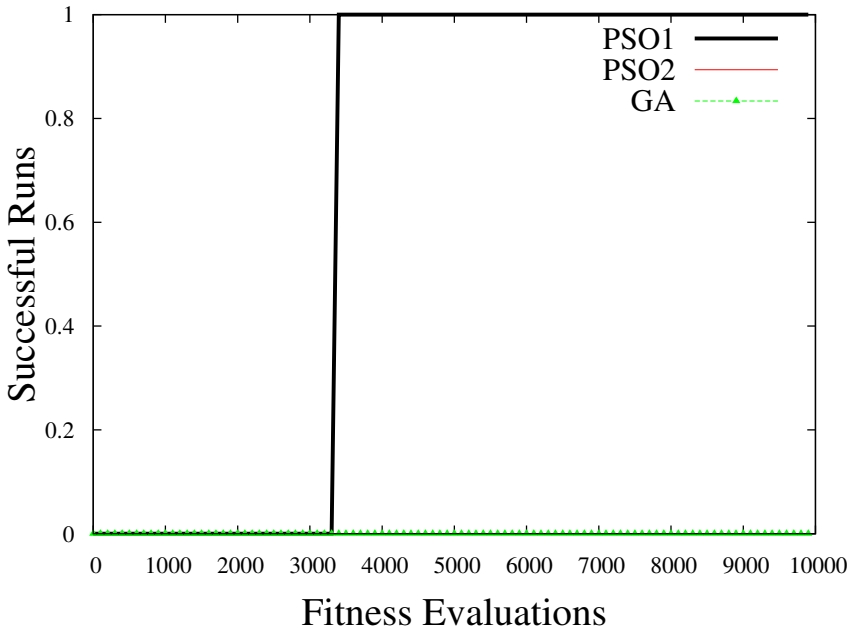




Comparison GA & PSO: Oscillating Brusselator

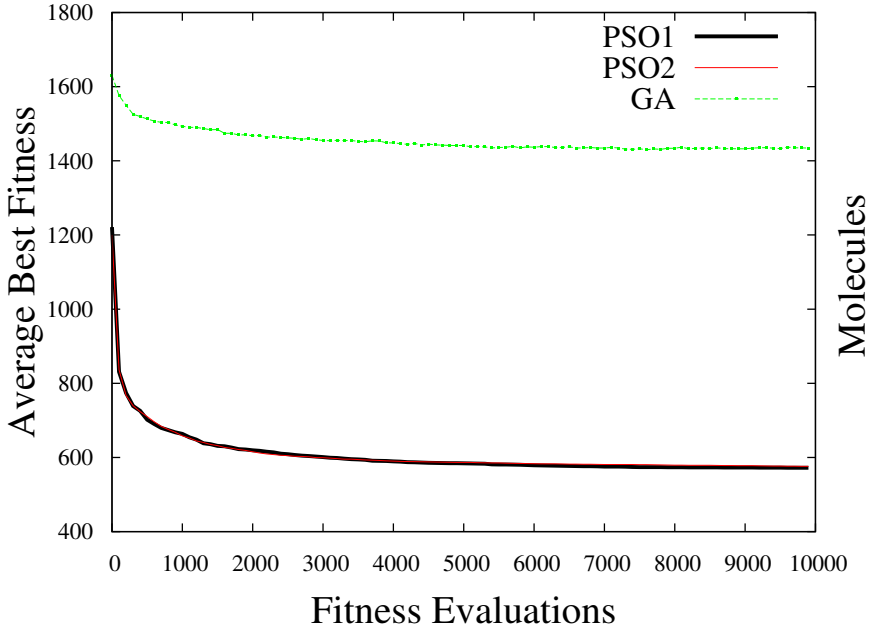


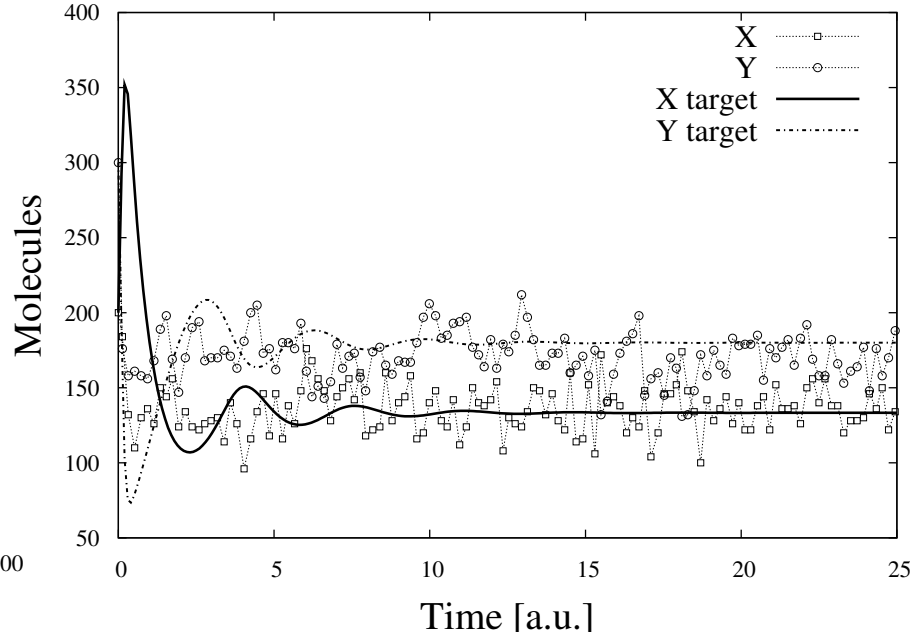


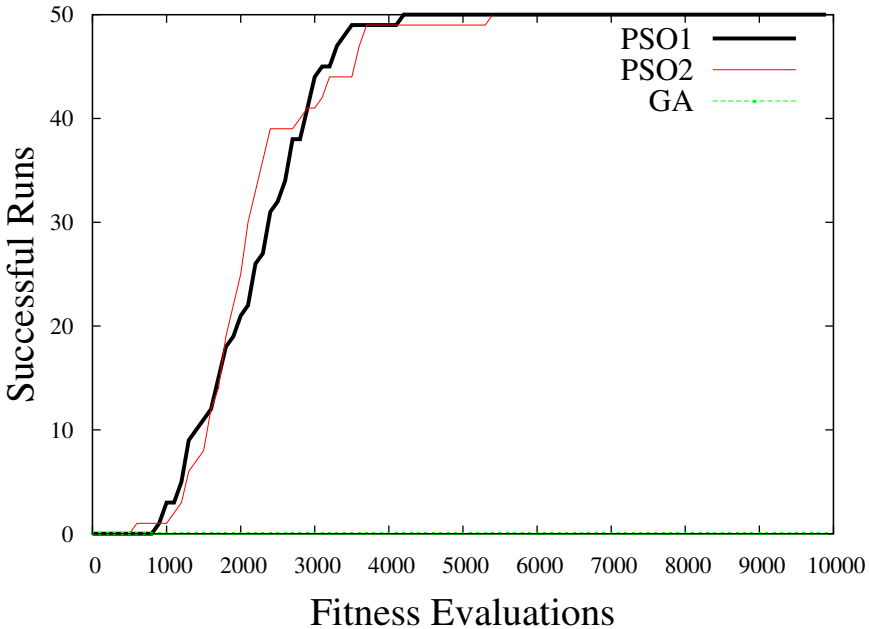


- PSO1/2 best performers
- GA worst average fitnesses
- dynamics faithfully reconstructed

Comparison GA & PSO: Damped Brusselator







- PSO1/2 best performers
- GA worst average fitnesses
- dynamics unfaithfully reconstructed

Stiff systems: an example

A system is said to be stiff if is characterized by well-separeted fast and slow dynamical modes, the fastest of which is stable.

The decaying dimerization model:

$$r_1 : S_1 \xrightarrow{1} \lambda$$

$$r_2 : S_1 + S_1 \xrightarrow{10} S_2$$

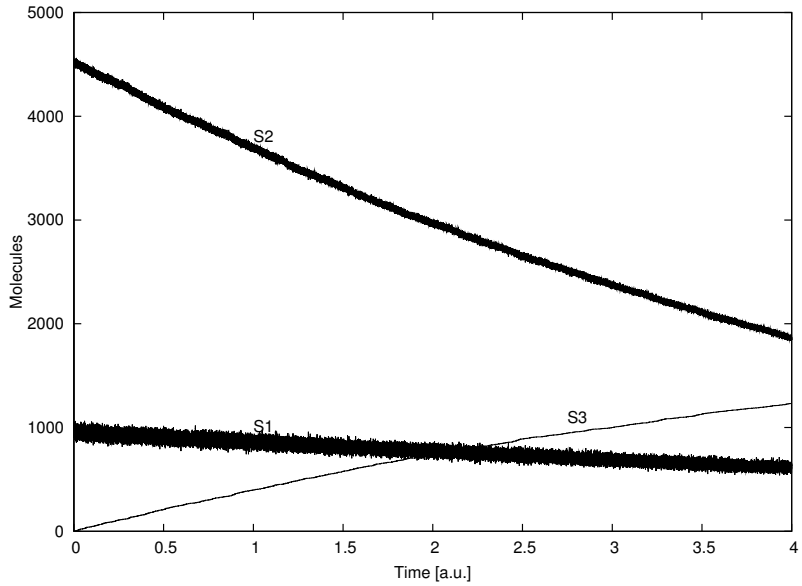
$$r_3 : S_2 \xrightarrow{10^3} S_1 + S_1$$

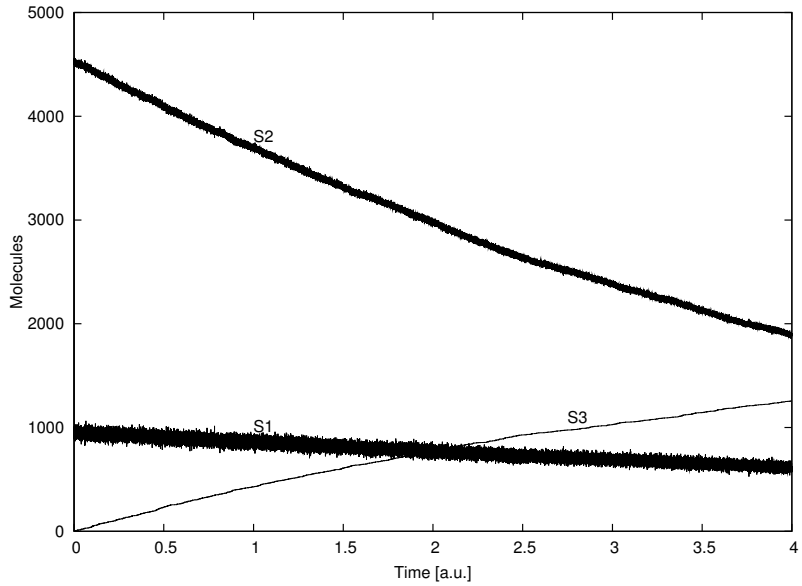
$$r_4 : S_2 \xrightarrow{0.1} S_3$$

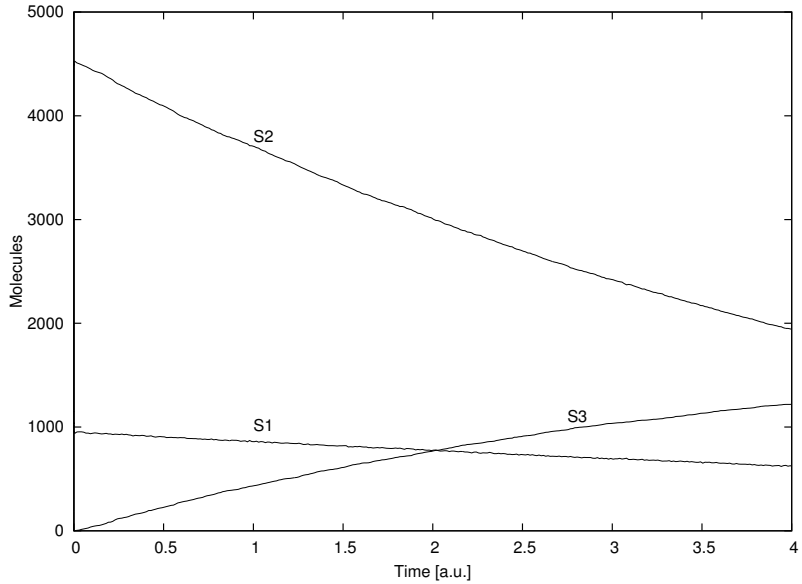
$$S_1 = 10^4 \; S_2 = S_3 = 0$$

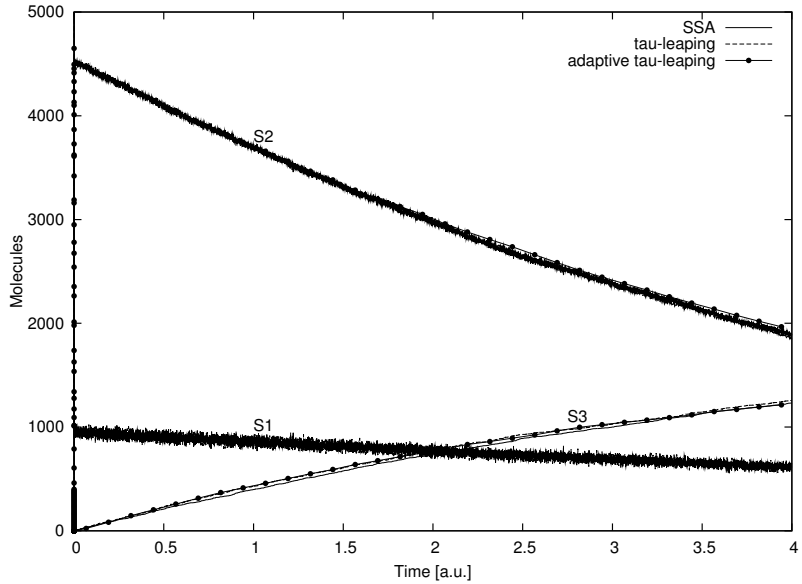
	SSA	**tau leaping**	**adaptive tau leaping**
average number of steps	$2.46 \cdot 10^7$	$1 \cdot 10^6$	945
total execution time	175h 1m	11h 47m	1m 51s

Decaying dimerization: algorithms comparison





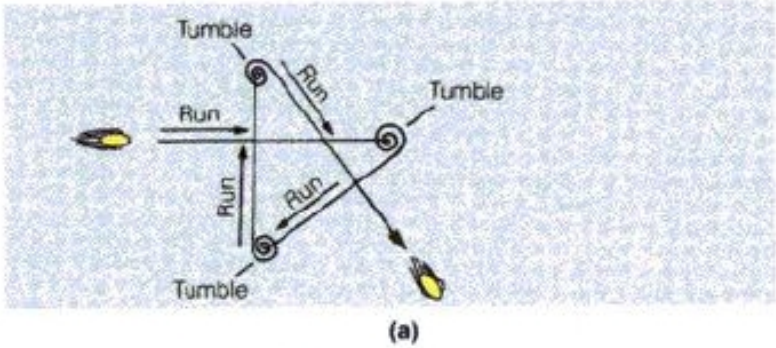


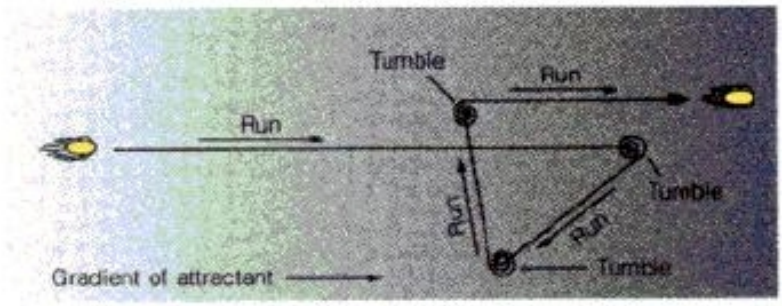


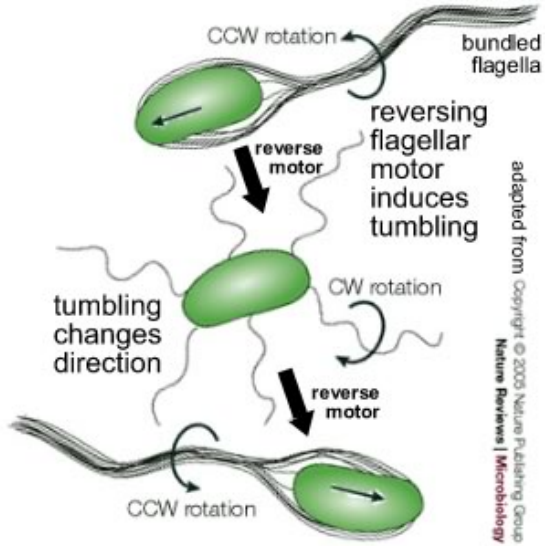
Bacterial chemotaxis

Chemotaxis allows bacteria to respond to ligand concentration gradients in their surroundings:

- random walk through an homogeneous environment $\rightarrow$ high switching frequency of flagellar rotation
- directional swimming in presence of ligand concentration gradient $\rightarrow$
  reduced switching frequency of flagellar rotation
- adaptation: if ligand concentration remains constant $\rightarrow$ switching frequency is reset to random walk level





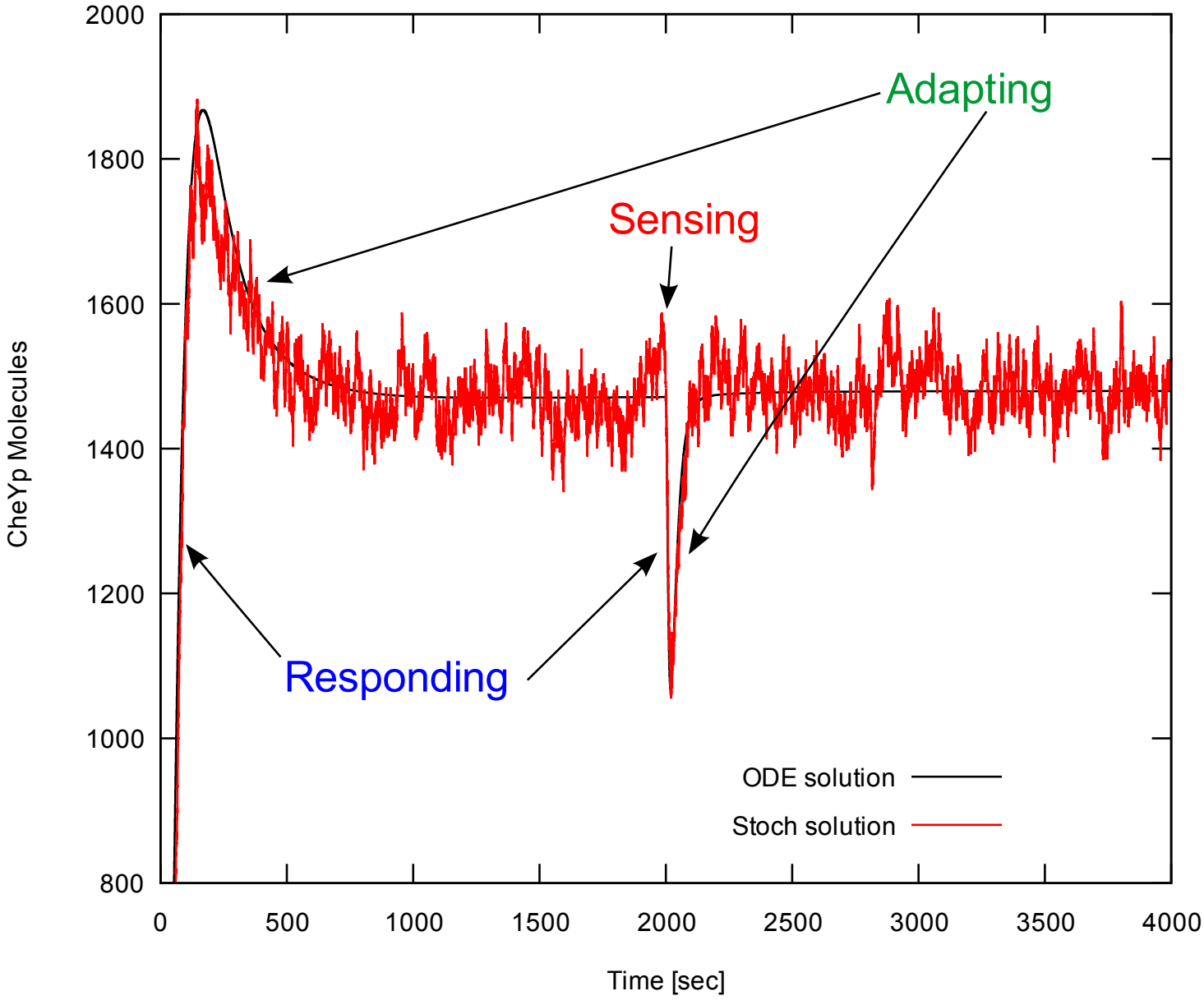


Sensing Responding Adapting

	Reagents	Products	Methyl. state
1	$2MCP^m$ + 2CheW	$2MCP^m$ ::2CheW	$m = 0$
2	$2MCP^m$ ::2CheW	$2MCP^m$ + 2CheW	$m = 0$
3	$2MCP^m$ ::2CheW + 2CheA	$2MCP^m$ ::2CheW::2CheA	$m = 0$
4	$2MCP^m$ ::2CheW::2CheA	$2MCP^m$ ::2CheW + 2CheA	$m = 0$
5-8	$2MCP^m$ ::2CheW::2CheA + CheR	$2MCP^{m+1}$::2CheW::2CheA + CheR	$m = 0, \ldots, 3$
9-12	$2MCP^m$ ::2CheW::2CheA + CheBp	$2MCP^{m-1}$::2CheW::2CheA + CheBp	$m = 1, \ldots, 4$
13-17	$2MCP^m$ ::2CheW::2CheA + ATP	$2MCP^m$ ::2CheW::2CheAp	$m = 0, \ldots, 4$
18-22	$2MCP^m$ ::2CheW::2CheAp + CheY	$2MCP^m$ ::2CheW::2CheA + CheYp	$m = 0, \ldots, 4$
23-27	$2MCP^m$ ::2CheW::2CheAp + CheB	$2MCP^m$ ::2CheW::2CheA + CheBp	$m = 0, \ldots, 4$
28-32	lig + $2MCP^m$ ::2CheW::2CheA	lig::$2MCP^m$ ::2CheW::2CheA	$m = 0, \ldots, 4$
33-37	lig::$2MCP^m$ ::2CheW::2CheA	lig + $2MCP^m$ ::2CheW::2CheA	$m = 0, \ldots, 4$
38-41	lig::$2MCP^m$ ::2CheW::2CheA + CheR	lig::$2MCP^{m+1}$::2CheW::2CheA + CheR	$m = 0, \ldots, 3$
42-45	lig::$2MCP^m$ ::2CheW::2CheA + CheBp	lig::$2MCP^{m-1}$::2CheW::2CheA + CheBp	$m = 1, \ldots, 4$
46-50	lig::$2MCP^m$ ::2CheW::2CheA + ATP	lig::$2MCP^m$ ::2CheW::2CheAp	$m = 0, \cdots, 4$
51-55	lig::$2MCP^m$ ::2CheW::2CheAp + CheY	lig::$2MCP^m$ ::2CheW::2CheA + CheYp	$m = 0, \ldots, 4$
56-60	lig::$2MCP^m$ ::2CheW::2CheAp + CheB	lig::$2MCP^m$ ::2CheW::2CheA + CheBp	$m = 0, \ldots, 4$
61	CheYp + CheZ	CheY + CheZ	
62	CheBp	CheB	

62 reactions, 32 molecular species

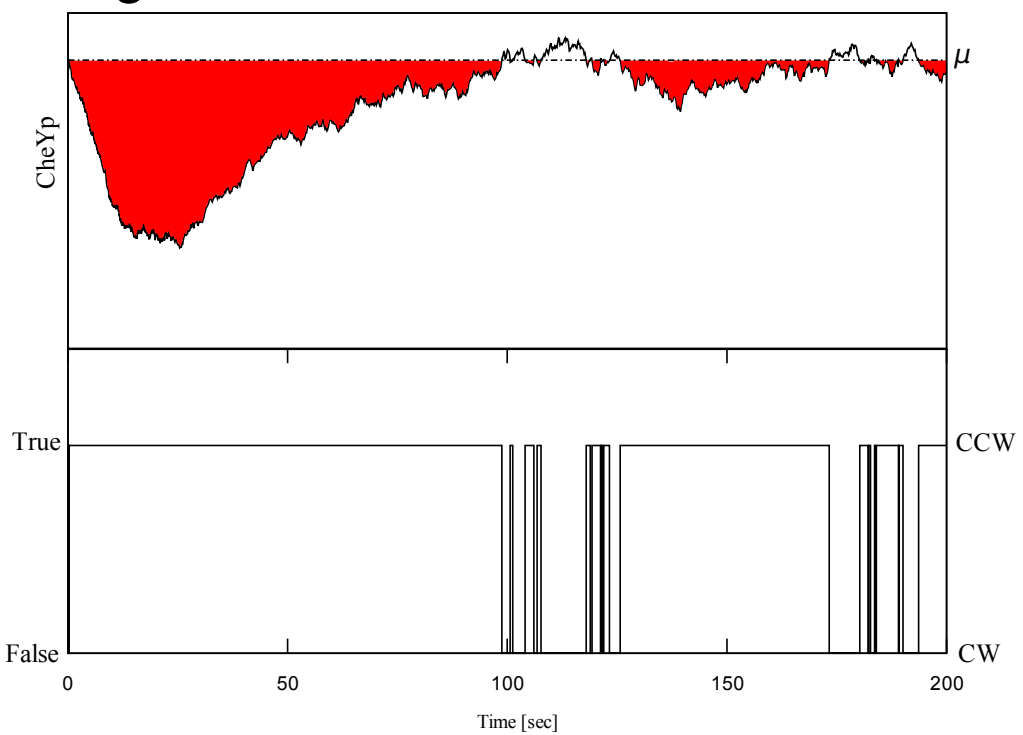
Sensing Responding Adapting



CheYp and the flagellar rotation

Our assumptions:

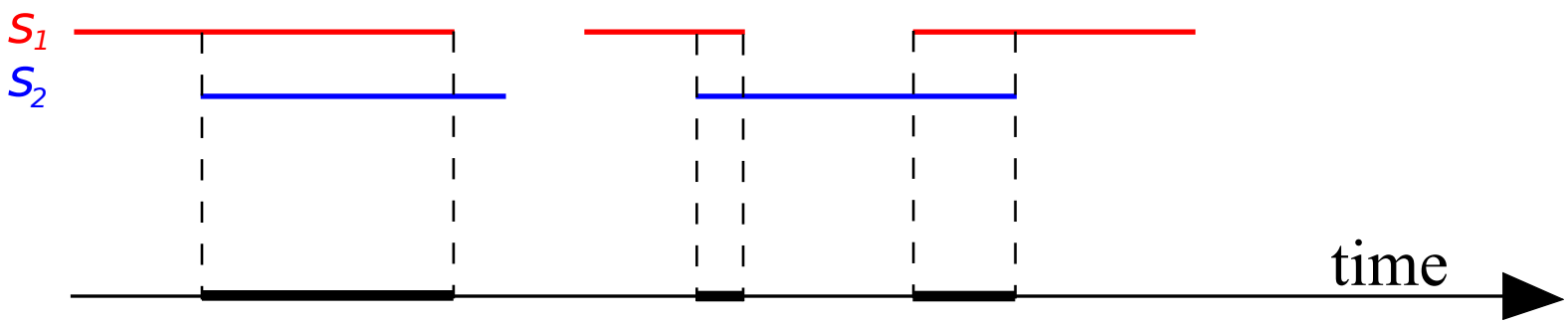
- the flagellar motor switch is sensitive to a threshold level of CheYp, that is hereby evaluated as the mean value of CheYp at steady state
- we make a one-to-one correspondence between the behavior of a single flagellum and one temporal evolution of CheYp generated by one run of the tau leaping algorithm.



Many flagella and running or tumbling

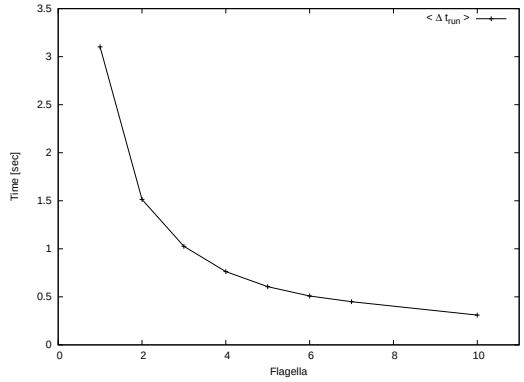
$$\mathcal{T}^{\mathbf{n}}_{\mathbf{sync}} = \{t \in \Delta t_{sim} \mid CCW_{s_i}(t) = true \text{ for all } i = 1, \ldots, n\}$$

$\mathcal{T}^{n}_{sync}$ is the set of all times during which *all* time series s_i are below the threshold μ



That corresponds to the running motion of the bacterium

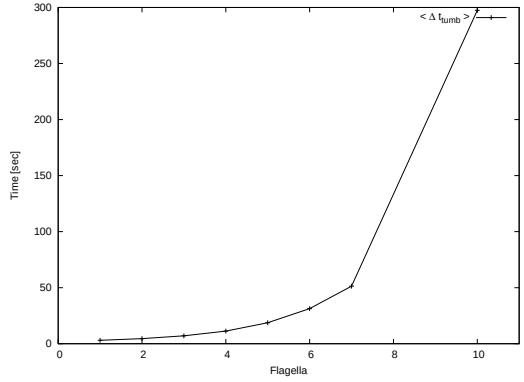
Many flagella and running or tumbling



Running

$< \Delta t_{run} >$

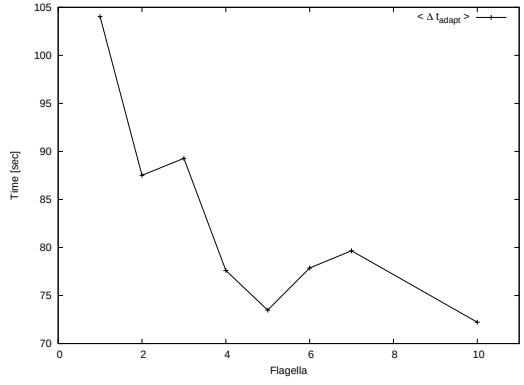
all flagella rotate CCW



Tumbling

$< \Delta t_{tumb} >$

some flagella rotate CW

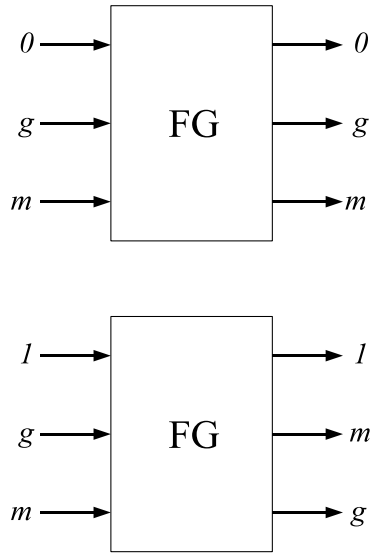


Adapting

$< \Delta t_{adapt} >$

length of the negative peak

Fredkin Gate: definition

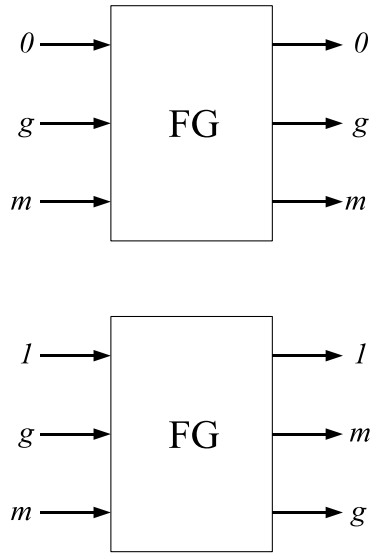


$\alpha_i\ \beta_i\ \gamma_i$	$\mapsto$	$\alpha_o\ \beta_o\ \gamma_o$
0 0 0		0 0 0
0 0 1		0 0 1
0 1 0		0 1 0
0 1 1		0 1 1
1 0 0		1 0 0
1 0 1		1 1 0
1 1 0		1 0 1
1 1 1		1 1 1

Reaction		Constant
r_1 :	$a + b \to a + d$	$c_1 = 1 \cdot 10^{-3}$
r_2 :	$a + c \to a + e$	$c_2 = 1 \cdot 10^{-3}$
r_3 :	$a + B \to a + D$	$c_3 = 1 \cdot 10^{-3}$
r_4 :	$a + C \to a + E$	$c_4 = 1 \cdot 10^{-3}$
r_5 :	$A + b \to A'$	$c_5 = 1 \cdot 10^{-3}$
r_6 :	$A' + c \to A + d + e$	$c_6 = 1 \cdot 10^{-1}$
r_7 :	$A' + C \to A + D + e$	$c_7 = 1 \cdot 10^{-1}$
r_8 :	$A + B \to A''$	$c_8 = 1 \cdot 10^{-3}$
r_9 :	$A'' + c \to A + d + E$	$c_9 = 1 \cdot 10^{-1}$
r_{10} :	$A'' + C \to A + D + E$	$c_{10} = 1 \cdot 10^{-1}$

Reaction		Constant
r_{11} :	$a + A \to \lambda$	$c_{11} = 1 \cdot 10^{-1}$
r_{12} :	$b + B \to \lambda$	$c_{12} = 1 \cdot 10^{-1}$
r_{13} :	$c + C \to \lambda$	$c_{13} = 1 \cdot 10^{-1}$
r_{14} :	$d + D \to \lambda$	$c_{14} = 1 \cdot 10^{-1}$
r_{15} :	$e + E \to \lambda$	$c_{15} = 1 \cdot 10^{-1}$
r_{16} :	$\lambda \to a$	$c_{16} \in \{0, 1\}$
r_{17} :	$\lambda \to A$	$c_{17} \in \{0, 1\}$
r_{18} :	$\lambda \to b$	$c_{18} \in \{0, 1\}$
r_{19} :	$\lambda \to B$	$c_{19} \in \{0, 1\}$
r_{20} :	$\lambda \to c$	$c_{20} \in \{0, 1\}$
r_{21} :	$\lambda \to C$	$c_{21} \in \{0, 1\}$

Fredkin Gate: definition

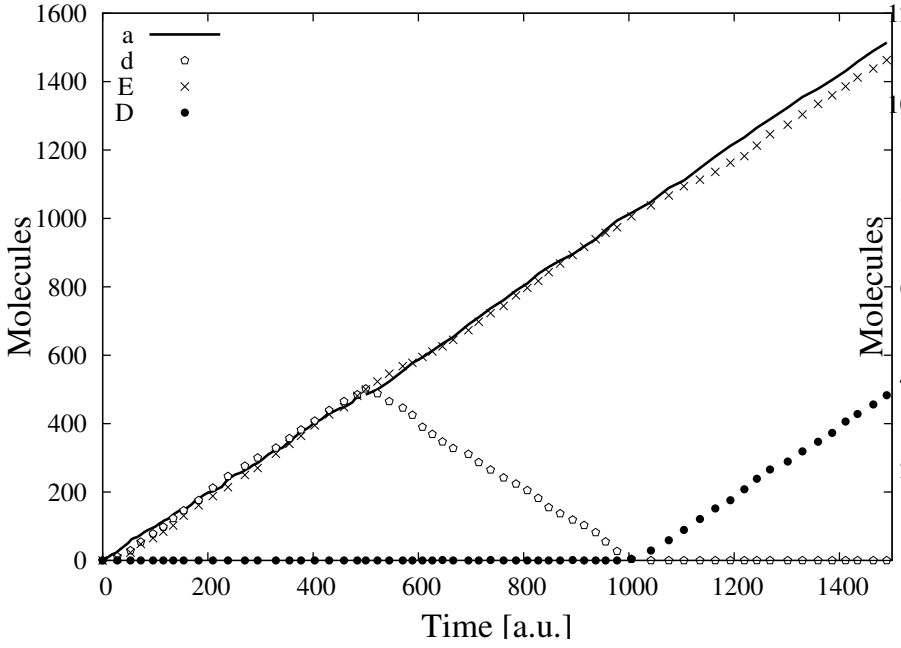


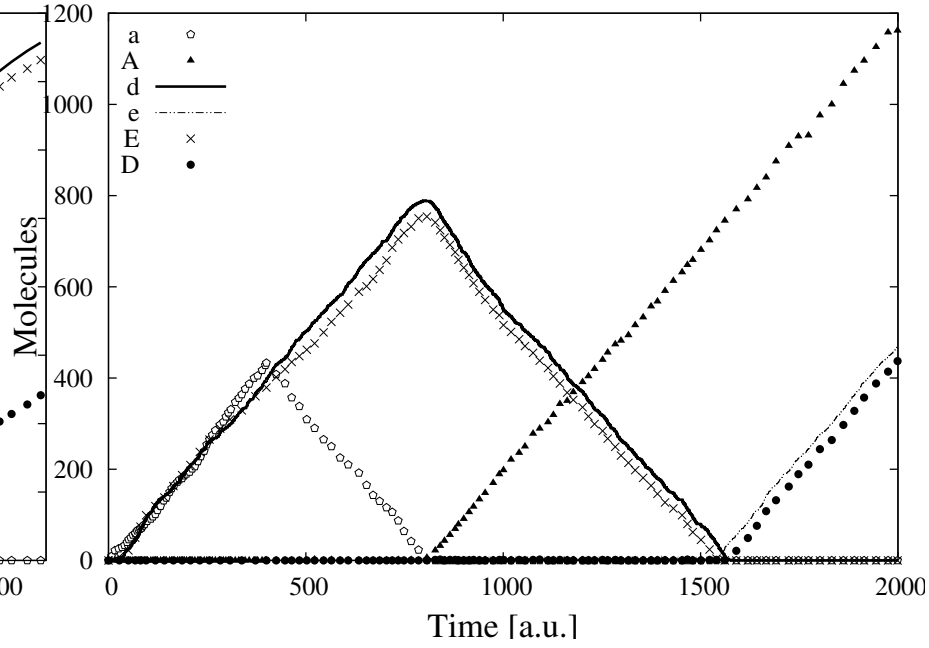
$\alpha_i\ \beta_i\ \gamma_i$	$\mapsto$	$\alpha_o\ \beta_o\ \gamma_o$
$a\ b\ c$		$a\ d\ e$
$a\ b\ C$		$a\ d\ E$
$a\ B\ c$		$a\ D\ e$
$a\ B\ C$		$a\ D\ E$
$A\ b\ c$		$A\ d\ e$
$A\ b\ C$		$A\ D\ e$
$A\ B\ c$		$A\ d\ E$
$A\ B\ C$		$A\ D\ E$

Reaction		Constant
r_1 :	$a + b \to a + d$	$c_1 = 1 \cdot 10^{-3}$
r_2 :	$a + c \to a + e$	$c_2 = 1 \cdot 10^{-3}$
r_3 :	$a + B \to a + D$	$c_3 = 1 \cdot 10^{-3}$
r_4 :	$a + C \to a + E$	$c_4 = 1 \cdot 10^{-3}$
r_5 :	$A + b \to A'$	$c_5 = 1 \cdot 10^{-3}$
r_6 :	$A' + c \to A + d + e$	$c_6 = 1 \cdot 10^{-1}$
r_7 :	$A' + C \to A + D + e$	$c_7 = 1 \cdot 10^{-1}$
r_8 :	$A + B \to A''$	$c_8 = 1 \cdot 10^{-3}$
r_9 :	$A'' + c \to A + d + E$	$c_9 = 1 \cdot 10^{-1}$
r_{10} :	$A'' + C \to A + D + E$	$c_{10} = 1 \cdot 10^{-1}$

Reaction		Constant
r_{11} :	$a + A \to \lambda$	$c_{11} = 1 \cdot 10^{-1}$
r_{12} :	$b + B \to \lambda$	$c_{12} = 1 \cdot 10^{-1}$
r_{13} :	$c + C \to \lambda$	$c_{13} = 1 \cdot 10^{-1}$
r_{14} :	$d + D \to \lambda$	$c_{14} = 1 \cdot 10^{-1}$
r_{15} :	$e + E \to \lambda$	$c_{15} = 1 \cdot 10^{-1}$
r_{16} :	$\lambda \to a$	$c_{16} \in \{0, 1\}$
r_{17} :	$\lambda \to A$	$c_{17} \in \{0, 1\}$
r_{18} :	$\lambda \to b$	$c_{18} \in \{0, 1\}$
r_{19} :	$\lambda \to B$	$c_{19} \in \{0, 1\}$
r_{20} :	$\lambda \to c$	$c_{20} \in \{0, 1\}$
r_{21} :	$\lambda \to C$	$c_{21} \in \{0, 1\}$

Fredkin Gate: simulations





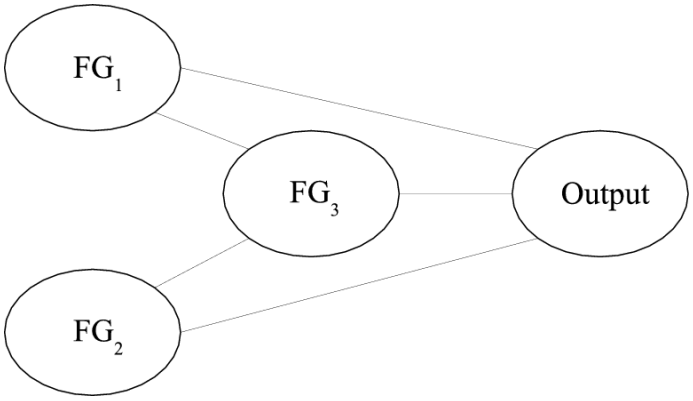
First input $(\alpha_i, \beta_i, \gamma_i) = (0, 0, 1)$ at $t = 0$
Second input $(\alpha_i, \beta_i, \gamma_i) = (0, 1, 1)$ at $t = 500$

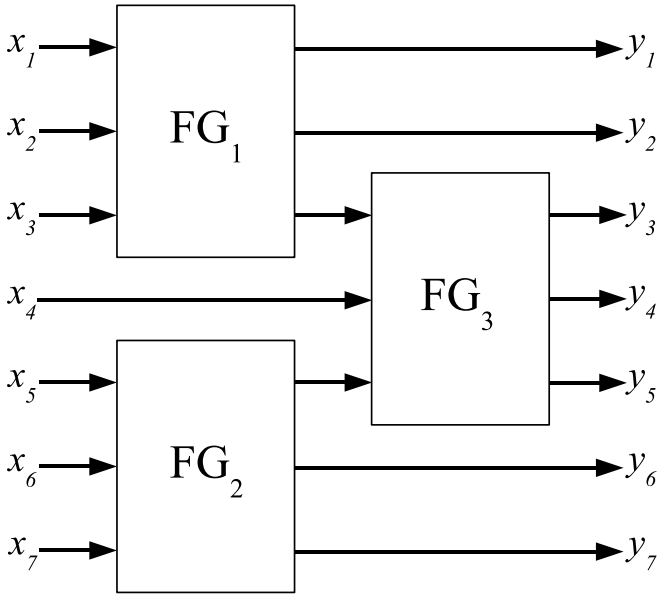
a b C $\rightarrow$ *a d E*
a B C $\rightarrow$ *a D E*

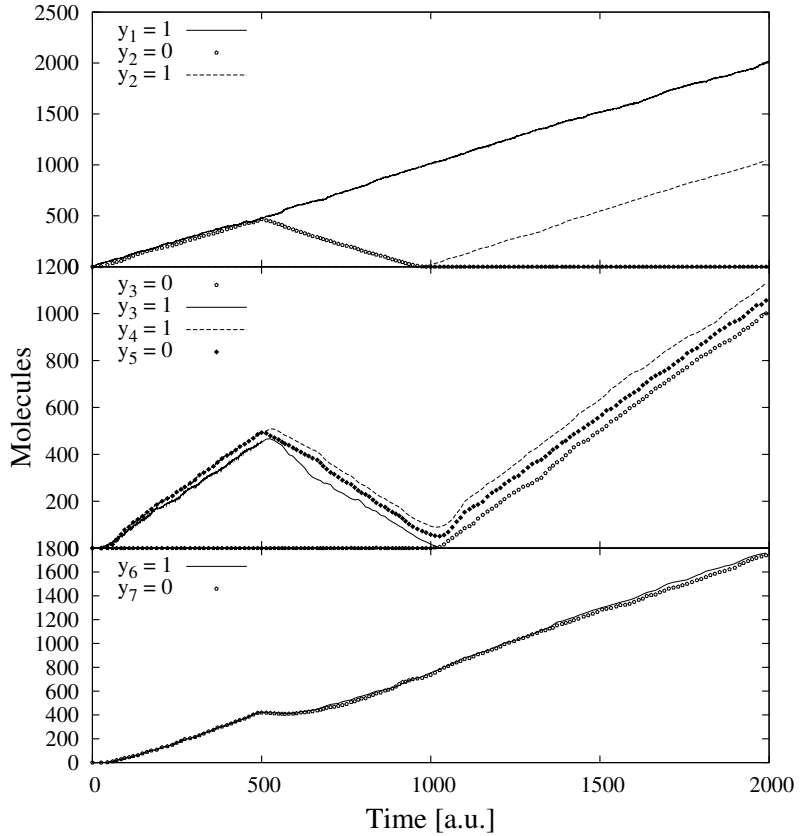
First input $(\alpha_i, \beta_i, \gamma_i) = (0, 0, 1)$ at $t = 0$
Second input $(\alpha_i, \beta_i, \gamma_i) = (1, 0, 1)$ at $t = 400$

a b C $\rightarrow$ *a d E*
A b C $\rightarrow$ *A D e*

Fredkin Circuits







Outline

Summary

BioSimWare is a simulation environments for the investigation of various biological systems that can range from cellular processes to population phenomena and to ecological systems.

Features

- Single/Multi volume simulations (stochastic and deterministic)
- Parameter Estimation (HC, GA, PSO)
- Parameter Sweep
- Analysis of the dynamics

Implementations

- Linux, Windows, Mac OS
- Single Processor, MPI, GRID
- SBML