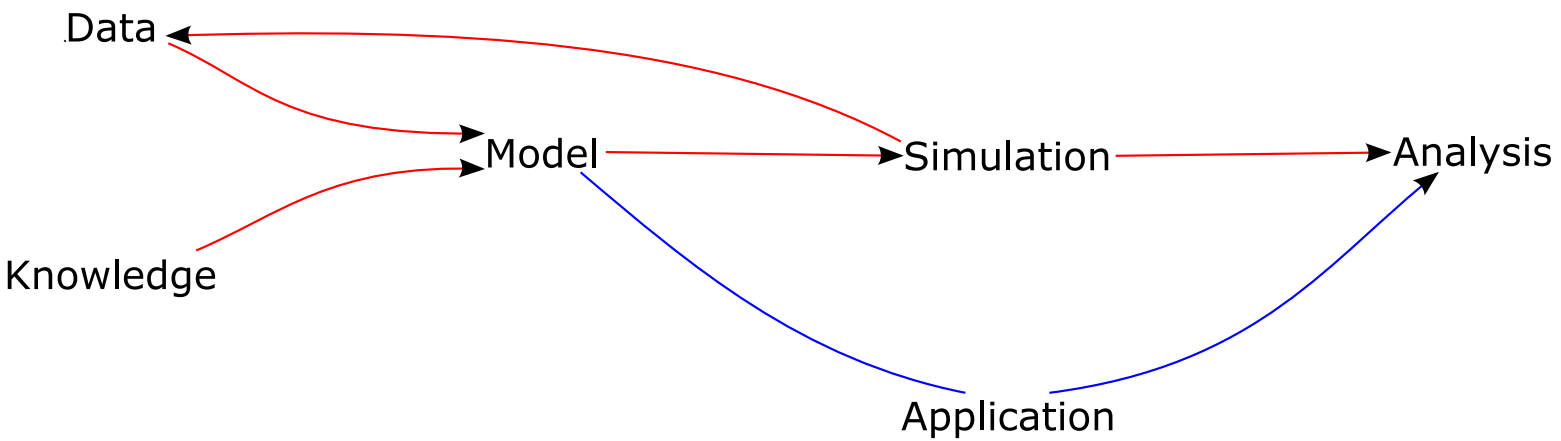
Modelling, simulation and analysis of biochemical systems

Paolo Cazzaniga

Dip. di Informatica, Sistemistica e Comunicazione (DISCo)
Università di Milano-Bicocca
cazzaniga@disco.unimib.it

Outline

Systems Biology



Outline

Experimental Evidences of Noise in Biology

- Many experimental evidences of stochasticity in living systems:

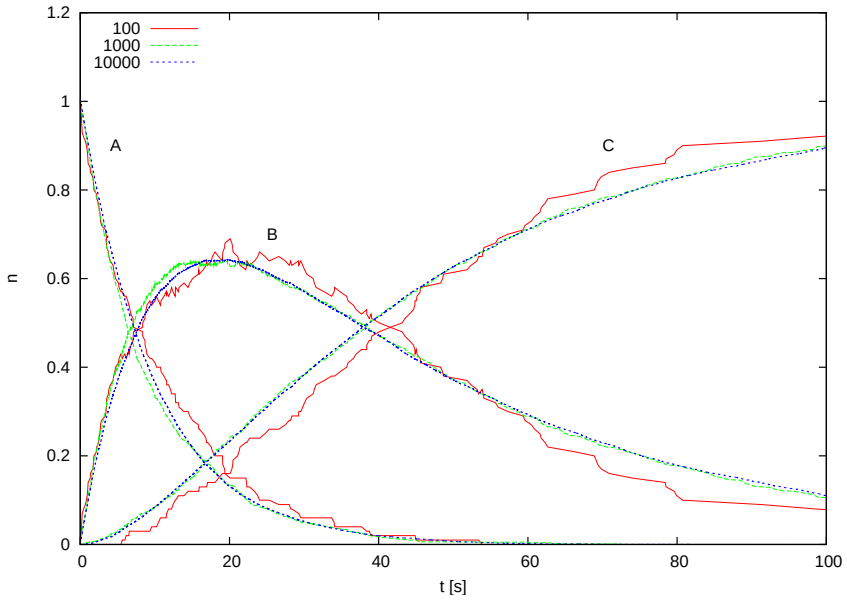
  transcription and translation [Abkowitz, 1996; Ozbudak, 2002], mRNA production is quantal [Hume, 2000] and in random pulses [Ross, 1994; Walters, 1995], the protein production occurs in short bursts and at random time intervals [Yarchuk, 1992; Chapon, 1982], in the λ phage the same starting conditions lead the system to two different kinds of evolutions (lysis/lysogeny) [Oppenheim,2005 (review)]...

- 2 kinds of cellular noise:
  - intrinsic noise - due to the inherent nature of the biochemical interactions
  - extrinsic noise - due to the external environmental conditions

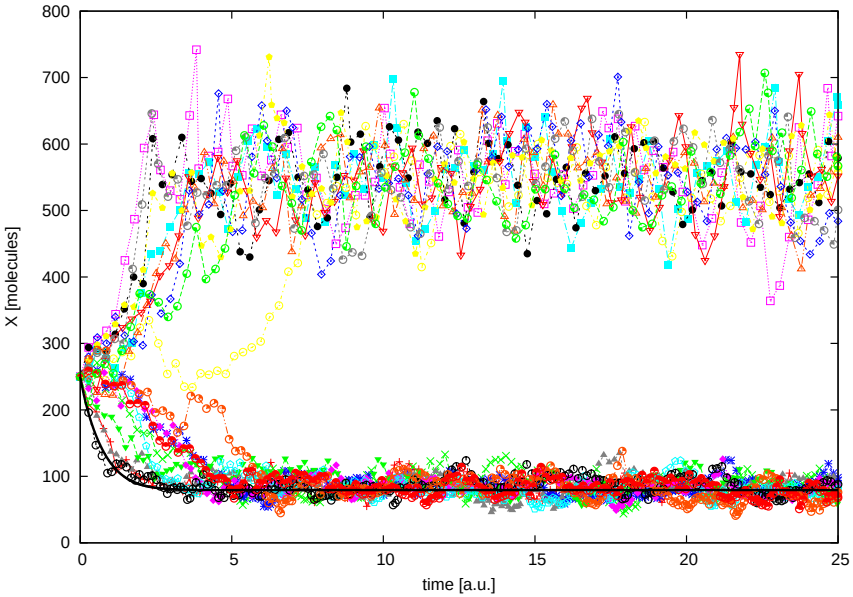
- Biological systems can be extremely non-linear and often exhibit many steady states, bifurcations or chaotic behavior

- Stochastic simulation is the probe to access the different evolutions

Experimental Evidences of Noise in Biology



noise influence $\propto \frac{1}{\sqrt{n}}$



same initial conditions $\neq$ evolutions

Outline

Deterministic Vs Stochastic Approaches

The deterministic approach - Ordinary Differential Equations:

$$\begin{aligned} \frac{dx_1}{dt} &= f_1(x_1, \ldots, x_n) \\ &\vdots \\ \frac{dx_n}{dt} &= f_n(x_1, \ldots, x_n) \end{aligned}$$

- molecular and environmental interactions are described by means of an equation for each molecular species x_i
- it requires the simultaneous solving of all the equations
- it captures the ensemble features of the system (global)

Deterministic Vs Stochastic Approaches

The stochastic approach:

- molecular interactions are described by means of probability distributions
- the probability distributions are dynamic. They evolve according to the system state
- it exploits a scattering perspective of chemical reactions
- it captures individuals behavior of the molecules (local)

Deterministic Vs Stochastic Approaches

Statistical physics argumentation shows that the stochastic approach:

- is always valid when deterministic is
- may be valid when ordinary deterministic is not, i.e. in a nonlinear system in the neighborhood of a chemical instability
- fully accounts for inherent statistical correlations and fluctuations neglected by the deterministic
- never approximates infinitesimal time increments by dt but uses finite time steps Δt

Stochastic approach requires HUGE computational time for big number of molecules!

$\Rightarrow$ It is not the definitive solution but an appropriate tool under "certain" conditions.

Outline

SSA

Stochastic Simulation Algorithm (SSA - Gillespie '76)

General Problem:

- V = fixed volume (well stirred, fixed experimental conditions)
- $\{S_i\}_{i=1,\ldots,N}$ = set of N chemical species in V
- $\{X_i\}_{i=1,\ldots,N}$ = current number of molecules in V
- $\{R_\mu\}_{\mu=1,\ldots,M}$ = set of M chemical reactions
- $\{c_\mu\}_{\mu=1,\ldots,M}$ = set of M reactions parameters

How does it evolve?

Fundamental Hypothesis:

$c_\mu \delta t$ = average probability that a particular combination of R_μ reactant molecules will react accordingly in the next time interval δt.

SSA

How does it work?

$\forall$ dynamical step the algorithm answers 2 questions:

- when the next reaction will occur? τ
- which one the next reaction will be? μ

this is done exploiting the Fundamental Hypothesis to get

$$P(\tau, \mu)\delta\tau = \begin{array}{l}\text{probability that the next reaction in V will occur} \\ \text{in the differential time interval } (t+\tau, t+\tau+d\tau) \\ \text{and will be } R_\mu\end{array}$$

$P(\mu, \tau) \propto a_\mu = c_\mu h_\mu$ propensity function

c_μ summarizes chemical and physical properties

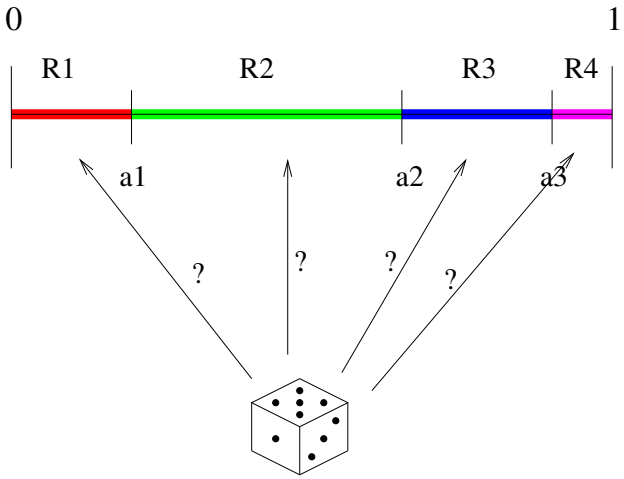
$$h_\mu = \prod_{1}^{N}{}_i \binom{X_i}{\alpha_i} \qquad \text{combinatorics}$$

$$a_0 = \sum_{1}^{M}{}_\mu a_\mu \qquad \text{normalization}$$

SSA: The Algorithm

for each step of the dynamics:

- compute the probability distribution
- toss r_1



- toss r_2 and compute $\tau = \frac{1}{a_0} \ln \frac{1}{r_2}$
- then apply the tossed rule to modify the number of molecules involved by that rule
- then update the evolution time $t += \tau$

Outline

τ Leaping

Tau leaping (Gillespie '06)

- Method used to speed up stochastic simulations firing more than one reaction per step.
- Idea: given a time increment τ find the exact probability distribution of rules application.
  ! as hard to solve as the CME !
- Solution: Approximate the exact behavior
- To obtain a good approximation, the changes in propensity functions are bounded $\Longrightarrow$ Leap Condition

$$\tau \text{ small enough s.t. } \Delta a_\mu \ll \varepsilon \text{ in } [t, t+\tau)$$

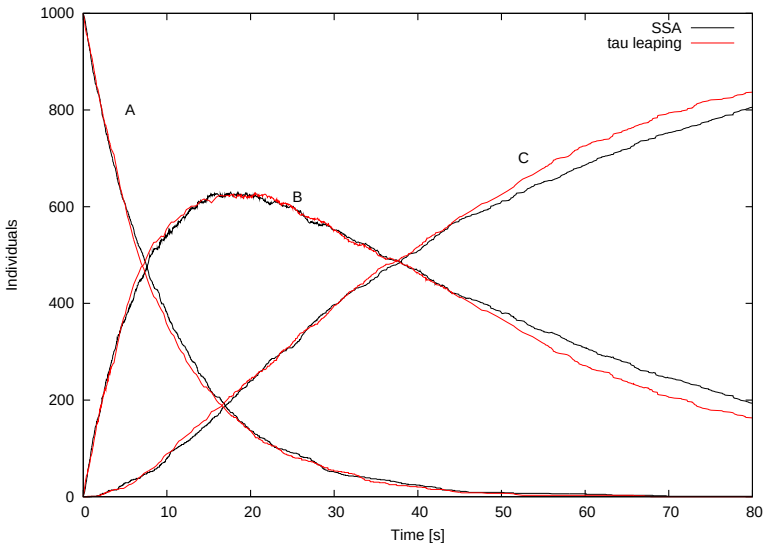
- Toss the reactions sampling a Poissonian distribution $P(a_\mu, \tau)$ with mean and variance $a_\mu \tau$.

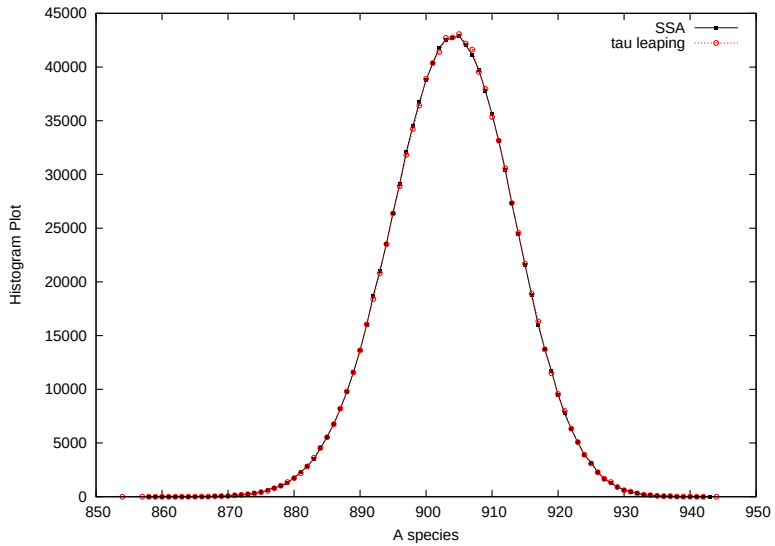
D.T. Gillespie et al., *Journ. Phys. Chem.*, 124:044109 (2006)

τ Leaping

The accuracy of the τ Leaping

- Consecutive reactions system:
  $A \xrightarrow{c_1} B$
  $B \xrightarrow{c_2} C$
- Test case to check the tau leaping method:



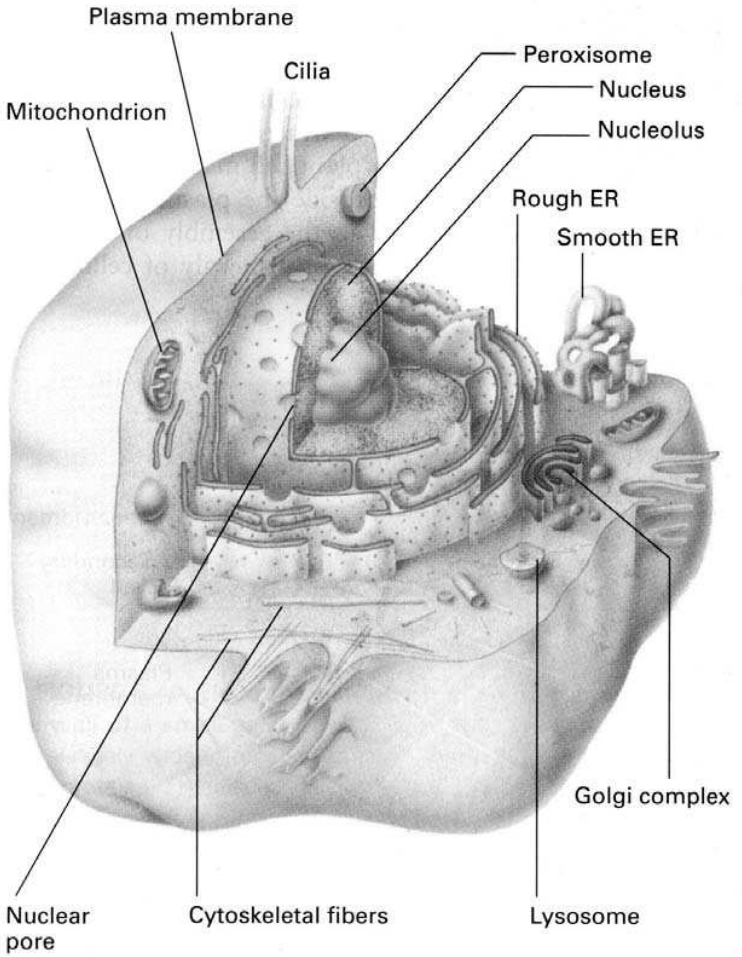


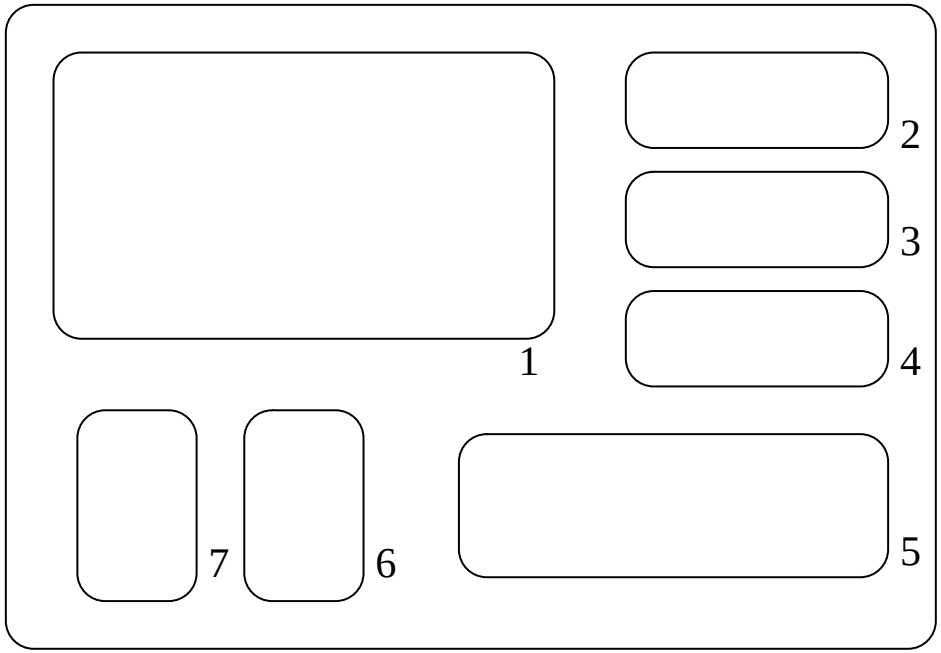
Outline

Membrane systems

- P systems (G.Paun, 1998): *calculus model* inspired by the cell
- *nondeterministic maximally parallel discrete* models for cellular process
- *essential features* of a cell captured by a P system:
  - cellular structure
  - biochemical substances
  - chemical reactions
  - communication / transport

Cellular Structure





Outline

τ-DPP

Quantitative (stochastic) simulation of complex systems

Features

- The P system framework is used to describe the system
- Chemical reactions as rewriting rules
- A modified tau-leaping procedure, placed inside every volume, is used to describe the behaviour of the system

Problems

- Complexity of the algorithm: $O(MN)$
- The molecules are uniformly distributed inside the volumes

τ-DPP: How it works

The iterative macrosteps of the algorithm are:

1. Compute the probabilities of the rules
2. Compute a candidate time increment
3. Select the smallest time increment among volumes
4. Select the set of reactions to execute
5. Update the system

P. Cazzaniga, D. Pescini, D. Besozzi, G. Mauri, **Tau leaping stochastic simulation method in P Systems**, Membrane Computing, 7th International Workshop, WMC 2006,(H.J.Hoogeboom, G. Paun, G. Rozenberg, A. Salomaa, eds.) LNCS 4361, 298–313, 2006.

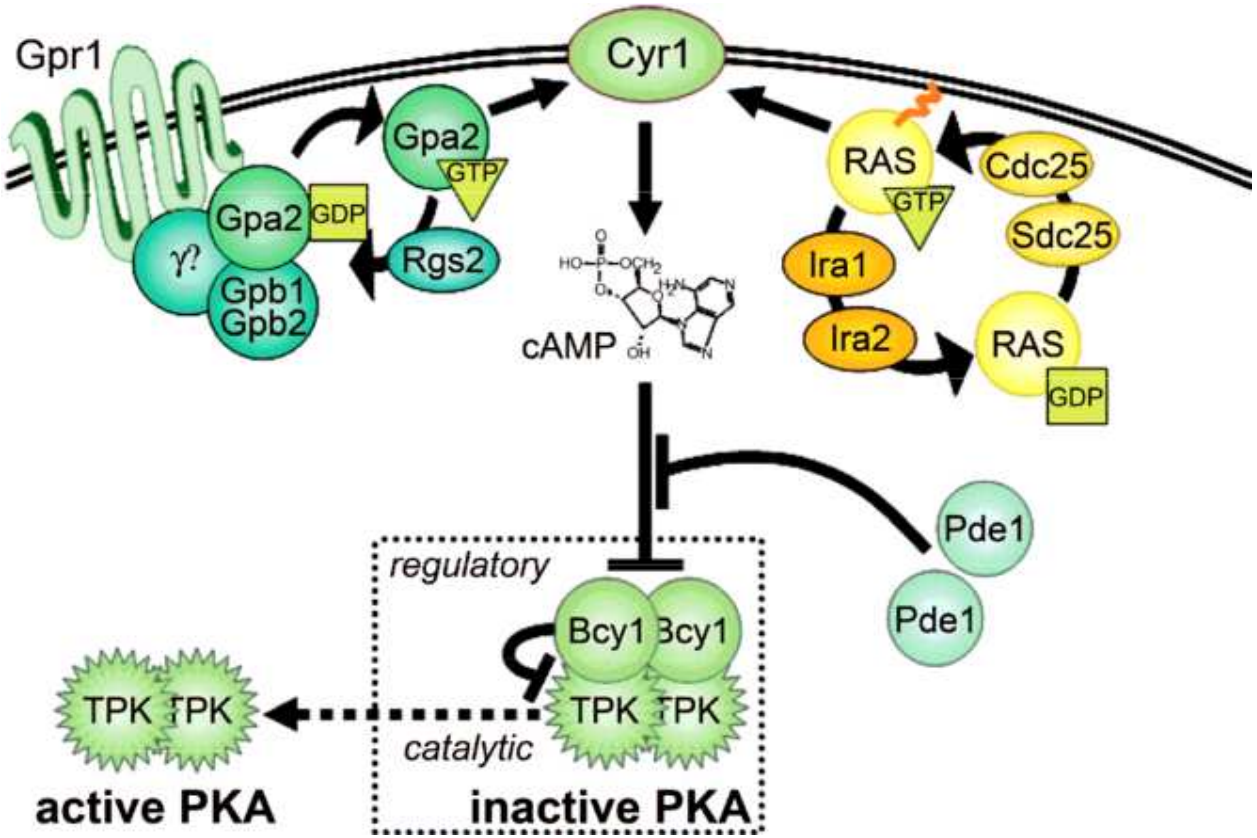
A τ-DPP variant: **Sτ-DPP**

- Hybrid structure: combines tissue and tree–like P systems
- It exploits dynamics description of τ-DPP
- The structure is independent from the communication channels between membranes. Two different graphs are used in the description: one to denote the membranes topology (i.e., membrane structure), and the other one the connections between membranes which allow the communication of objects
- Encompasses sizes of objects and membranes
- The set of reactions is enabled only if there is sufficient space

P. Cazzaniga, G. Mauri, L. Milanesi, E. Mosca, D. Pescini, **A novel variant of tissue P Systems for the modelling of biochemical systems**, Proceedings of the 10th International Workshop on Membrane Computing, WMC 2009 (G. Paun, M.J. Perez-Jimenez, A. Riscos-Nunez, G. Rozenberg, A. Salomaa, eds.), to appear in LNCS.

Outline

Ras/cAMP/PKA Pathway in the Yeast *S. cerevisiae*



Ras/cAMP/PKA Pathway in the Yeast *S. cerevisiae*

Reaction	Reagents	Products	Constant	Module
r_1	Ras2 · GDP + Cdc25	Ras2 · GDP · Cdc25	1.0	Ras2 switch cycle
r_2	Ras2 · GDP · Cdc25	Ras2 · GDP + Cdc25	1.0	
r_3	Ras2 · GDP · Cdc25	Ras2 · Cdc25 + GDP	1.5	
r_4	Ras2 · Cdc25 + GDP	Ras2 · GDP · Cdc25	1.0	
r_5	Ras2 · Cdc25 + GTP	Ras2 · GTP · Cdc25	1.0	
r_6	Ras2 · GTP · Cdc25	Ras2 · Cdc25 + GTP	1.0	
r_7	Ras2 · GTP · Cdc25	Ras2 · GTP + Cdc25	1.0	
r_8	Ras2 · GTP + Cdc25	Ras2 · GTP · Cdc25	1.0	
r_9	Ras2 · GTP + Ira2	Ras2 · GTP · Ira2	$3.0 \cdot 10^{-2}$	
r_{10}	Ras2 · GTP · Ira2	Ras2 · GDP + Ira2	$7.0 \cdot 10^{-1}$	
r_{11}	Ras2 · GTP + CYR1	Ras2 · GTP · CYR1	$1.0 \cdot 10^{-3}$	cAMP synthesis
r_{12}	Ras2 · GTP · CYR1 + ATP	Ras2 · GTP · CYR1 + cAMP	$1.0 \cdot 10^{-5}$	
r_{13}	Ras2 · GTP · CYR1 + Ira2	Ras2 · GDP + CYR1 + Ira2	$1.0 \cdot 10^{-3}$	
r_{14}	cAMP + PKA	cAMP · PKA	$1.0 \cdot 10^{-5}$	PKA activation
r_{15}	cAMP + cAMP · PKA	(2cAMP) · PKA	$1.0 \cdot 10^{-5}$	
r_{16}	cAMP + (2cAMP) · PKA	(3cAMP) · PKA	$1.0 \cdot 10^{-5}$	
r_{17}	cAMP + (3cAMP) · PKA	(4cAMP) · PKA	$1.0 \cdot 10^{-5}$	
r_{18}	(4cAMP) · PKA	cAMP + (3cAMP) · PKA	$1.0 \cdot 10^{-1}$	
r_{19}	(3cAMP) · PKA	cAMP + (2cAMP) · PKA	$1.0 \cdot 10^{-1}$	
r_{20}	(2cAMP) · PKA	cAMP + cAMP · PKA	$1.0 \cdot 10^{-1}$	
r_{21}	cAMP · PKA	cAMP + PKA	$1.0 \cdot 10^{-1}$	
r_{22}	(4cAMP) · PKA	2C + 2(R· 2cAMP)	1.0	
r_{23}	R · 2cAMP	R + 2cAMP	1.0	
r_{24}	R + 2	R · C	1.0	
r_{25}	2 * (R · C)	PKA	1.0	

Ras/cAMP/PKA Pathway in the Yeast *S. cerevisiae*

r_{26}	C + Pde1	C + $Pde1^{P}$	$1.0 \cdot 10^{-6}$	cAMP degradation
r_{27}	cAMP + $Pde1^{P}$	cAMP · $Pde1^{P}$	$1.0 \cdot 10^{-1}$	
r_{28}	cAMP · $Pde1^{P}$	cAMP + $Pde1^{P}$	$1.0 \cdot 10^{-1}$	
r_{29}	cAMP · $Pde1^{P}$	AMP + $Pde1^{P}$	7.5	
r_{30}	$Pde1^{P}$ + PPA2	Pde1 + PPA2	$1.0 \cdot 10^{-4}$	
r_{31}	cAMP + Pde2	cAMP · Pde2	$1.0 \cdot 10^{-4}$	
r_{32}	cAMP · Pde2	cAMP + Pde2	1.0	
r_{33}	cAMP · Pde2	AMP + Pde2	1.7	
r_{34}	C + Cdc25	C + $Cdc25^{P}$	10	
r_{35}	$Cdc25^{P}$ + PPA2	Cdc25 + PPA2	$1 \cdot 10^{-2}$	
r_{36}	Ira2 + C	$Ira2^{+}$ + C	$1 \cdot 10^{-2}$	
r_{37}	Ras2 · GTP + $Ira2^{+}$	Ras2 · GTP · $Ira2^{+}$	0.5	
r_{38}	Ras2 · GTP · $Ira2^{+}$	Ras2 · GTP + $Ira2^{+}$	1	
r_{39}	$Ira2^{+}$	Ira2	10	

The model involves

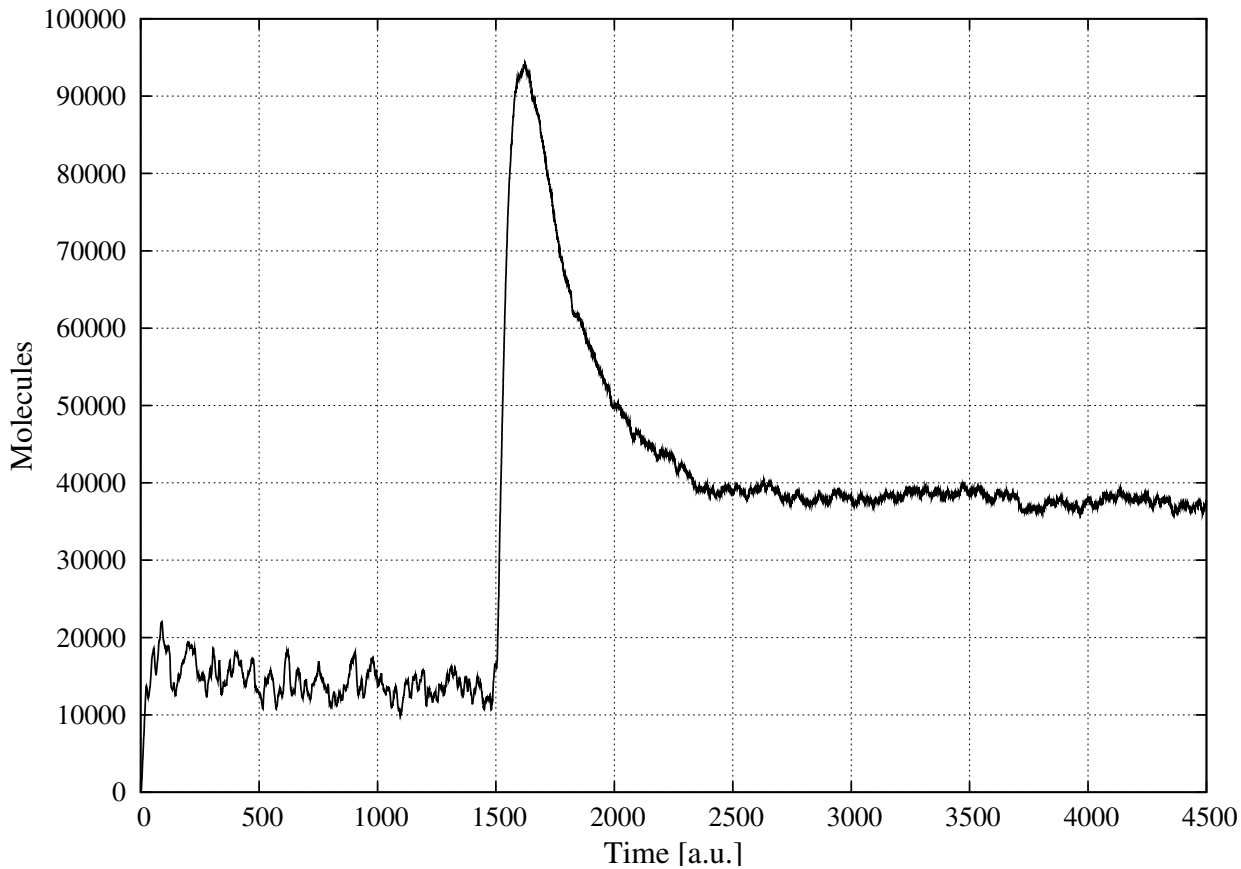
- 39 rules
- 30 molecular species
- 2 major feedback
- 2 Michaelis Menten schemes

P. Cazzaniga, D. Pescini, D. Besozzi, G. Mauri, S. Colombo, E. Martegani, **Modeling and stochastic simulation of the Ras/cAMP/PKA pathway in the yeast Saccharomyces cerevisiae evidences a key regulatory function for intracellular guanine nucleotides pools**, J. Biotechnology (2007)

Outline

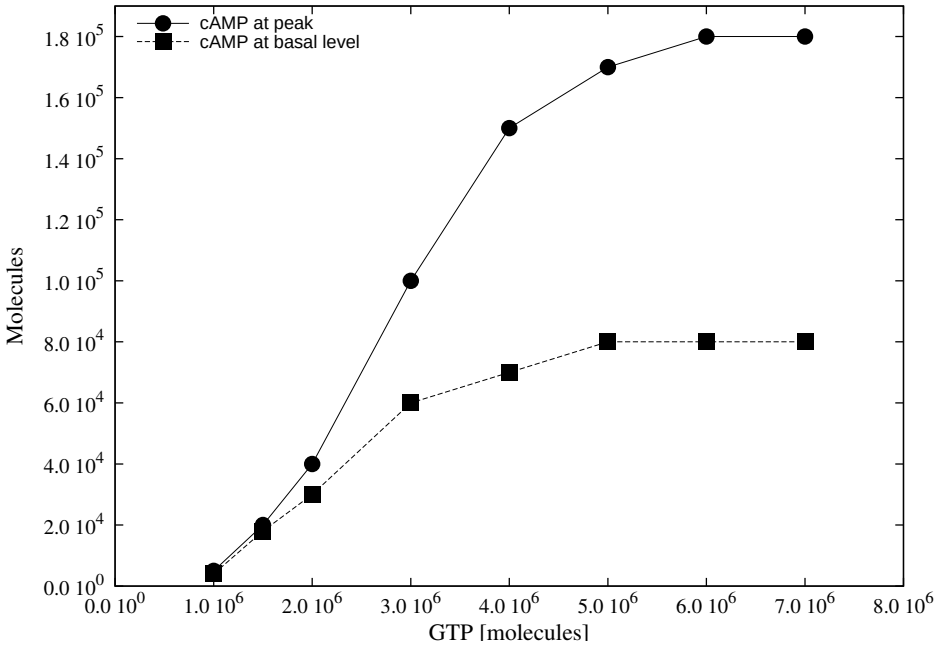
Ras/cAMP/PKA Pathway in the Yeast *S. cerevisiae*

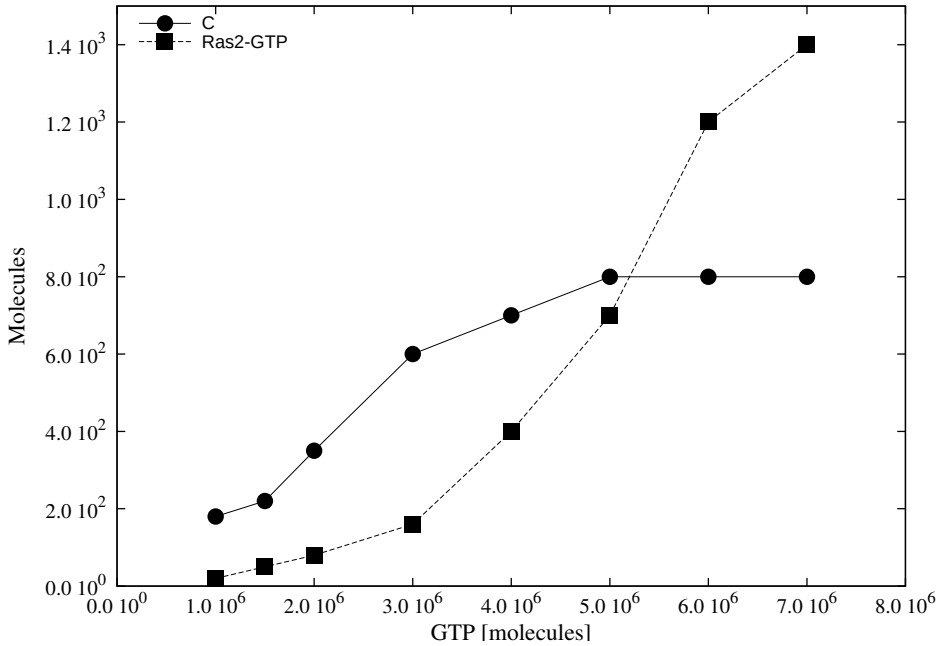
cAMP response to glucose addition at time 1500 [a.u.]:



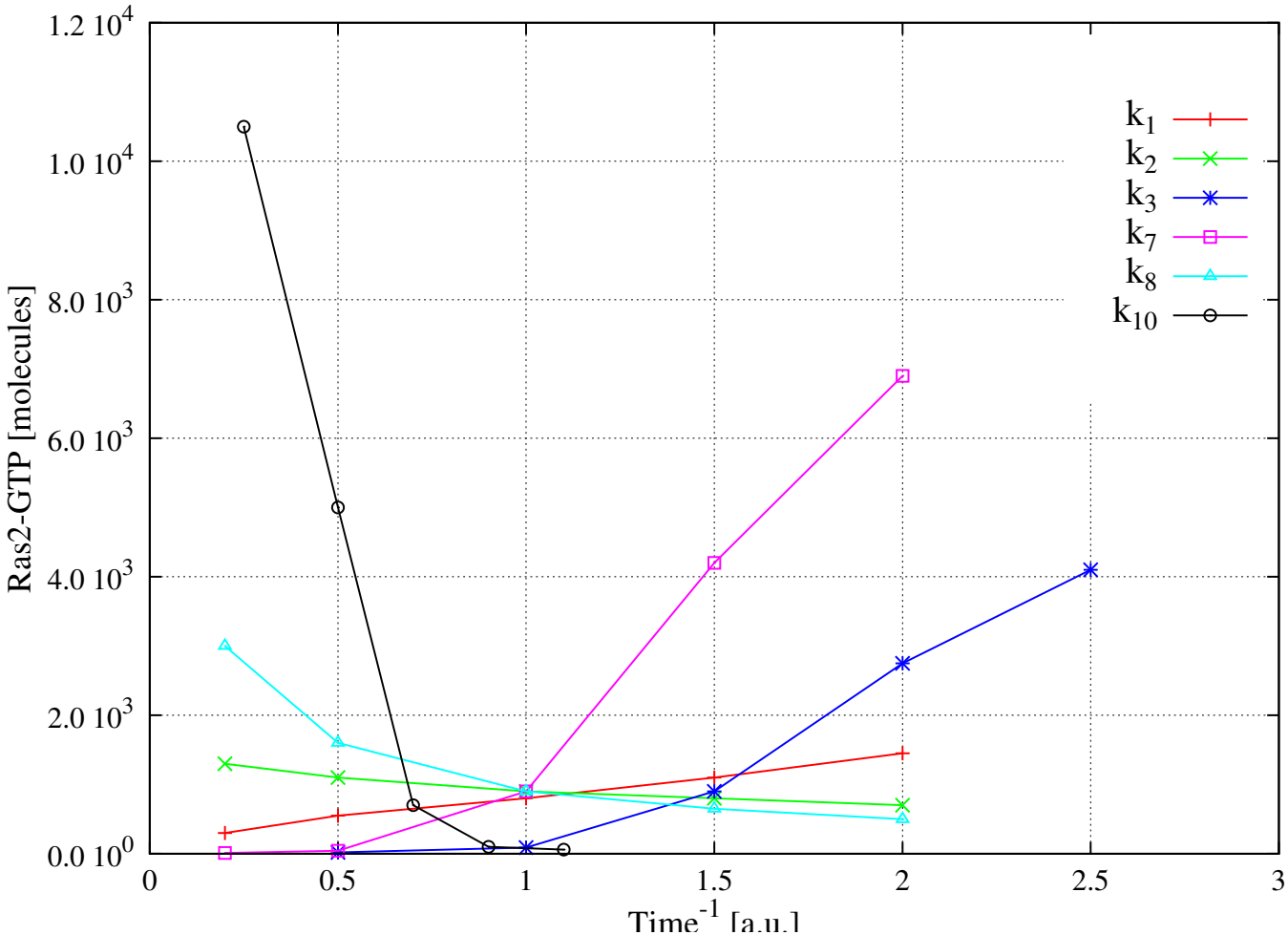
Ras/cAMP/PKA Pathway in the Yeast *S. cerevisiae*

Effect of different GTP input values on cAMP accumulation (left) and on Ras2·GTP and PKA activity (right)





Sensitivity of Ras2·GTP module

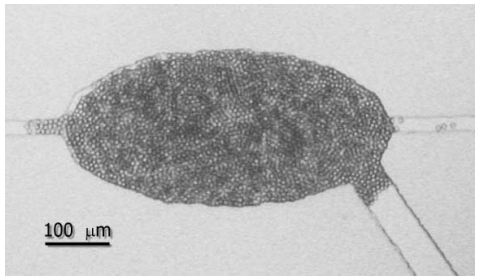


Outline

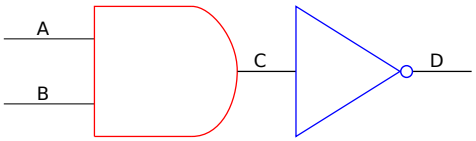
Microflow reactors

Microflow technology is an important tool to realize molecular computing:

- Topology: composed by several volumes
  reactors $\leftrightsquigarrow$ membranes
- Communication: objects flow among volumes
  channels $\leftrightsquigarrow$ communication rules
- Evolution:
  chemical reacting process $\leftrightsquigarrow$ multiset rewriting rules
- "Noise": stochastic evolution
  μ scale $\leftrightsquigarrow$ τ leaping



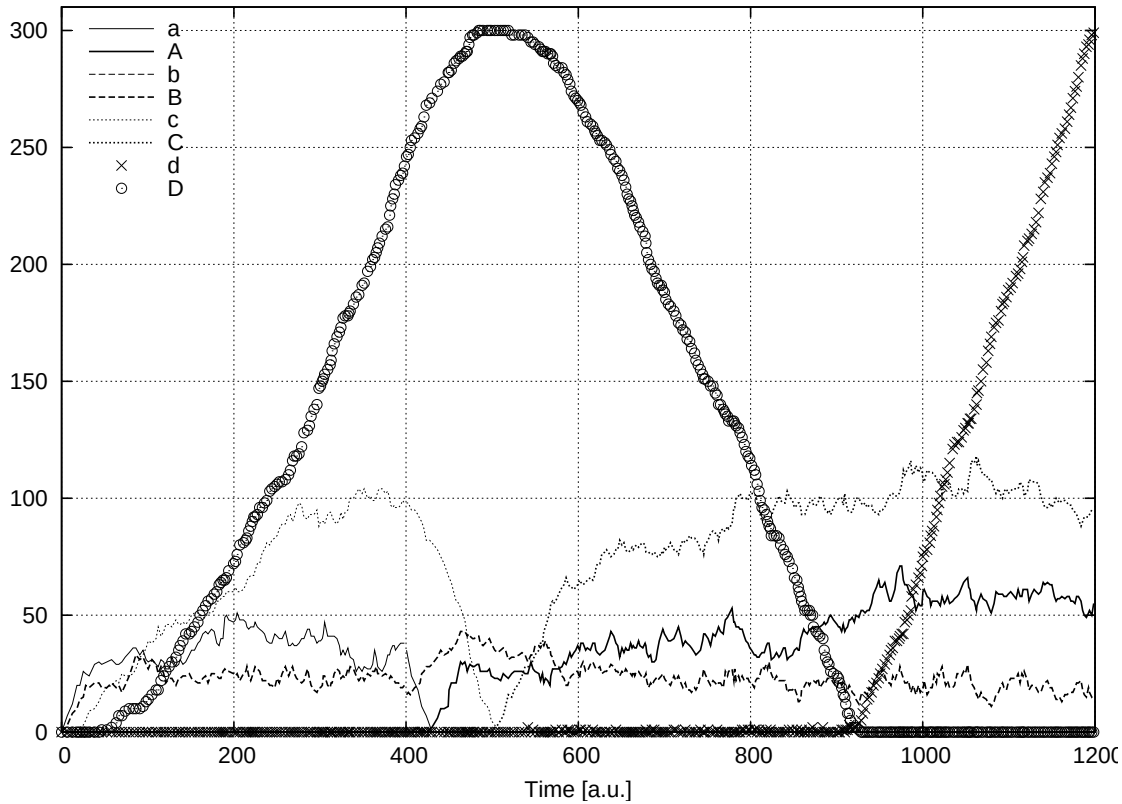
Boolean functions



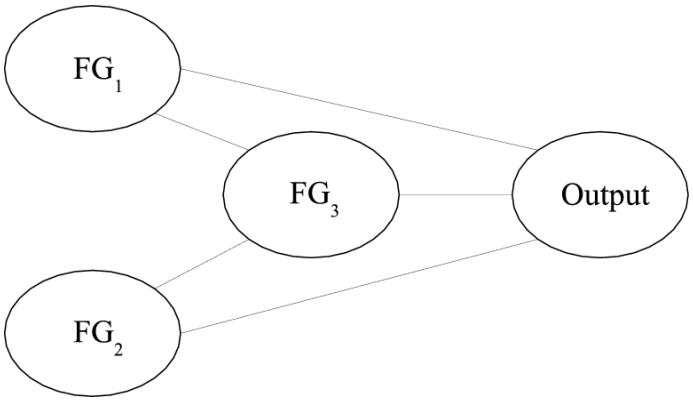
	B = 0 = b	B = 1 = B
A = 0 = a	D	D
A = 1 = A	D	d

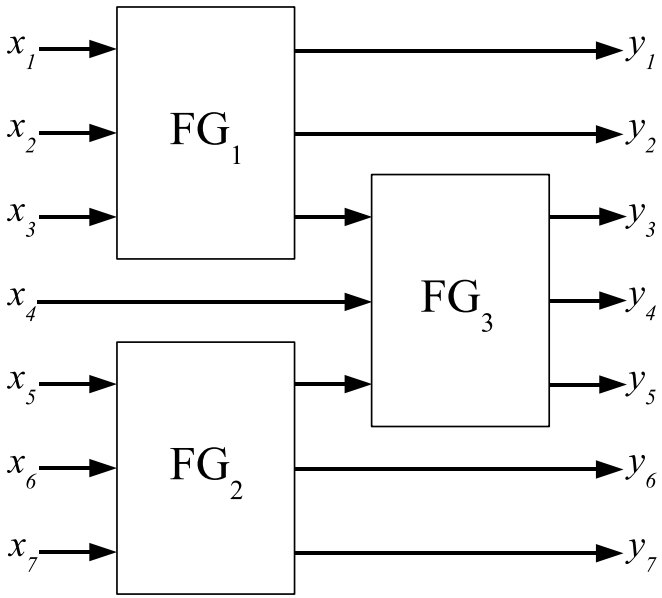
Reaction	Rule	Constant
r_1	$a + b \longrightarrow c$	$1 \cdot 10^{-3}$
r_2	$a + B \longrightarrow c$	$1 \cdot 10^{-3}$
r_3	$A + b \longrightarrow c$	$1 \cdot 10^{-3}$
r_4	$A + B \longrightarrow C$	$1 \cdot 10^{-3}$
r_5	$c \longrightarrow D$	$1 \cdot 10^{-2}$
r_6	$C \longrightarrow d$	$1 \cdot 10^{-2}$
r_7	$a + A \longrightarrow \lambda$	$1 \cdot 10^{-1}$
r_8	$b + B \longrightarrow \lambda$	$1 \cdot 10^{-1}$
r_9	$c + C \longrightarrow \lambda$	$1 \cdot 10^{-1}$
r_{10}	$d + D \longrightarrow \lambda$	$1 \cdot 10^{-1}$
r_{11}	$\lambda \longrightarrow a$	$c_{11} \in \{1, 0\}$
r_{12}	$\lambda \longrightarrow A$	$c_{12} \in \{1, 0\}$
r_{13}	$\lambda \longrightarrow b$	$c_{13} \in \{1, 0\}$
r_{14}	$\lambda \longrightarrow B$	$c_{14} \in \{1, 0\}$

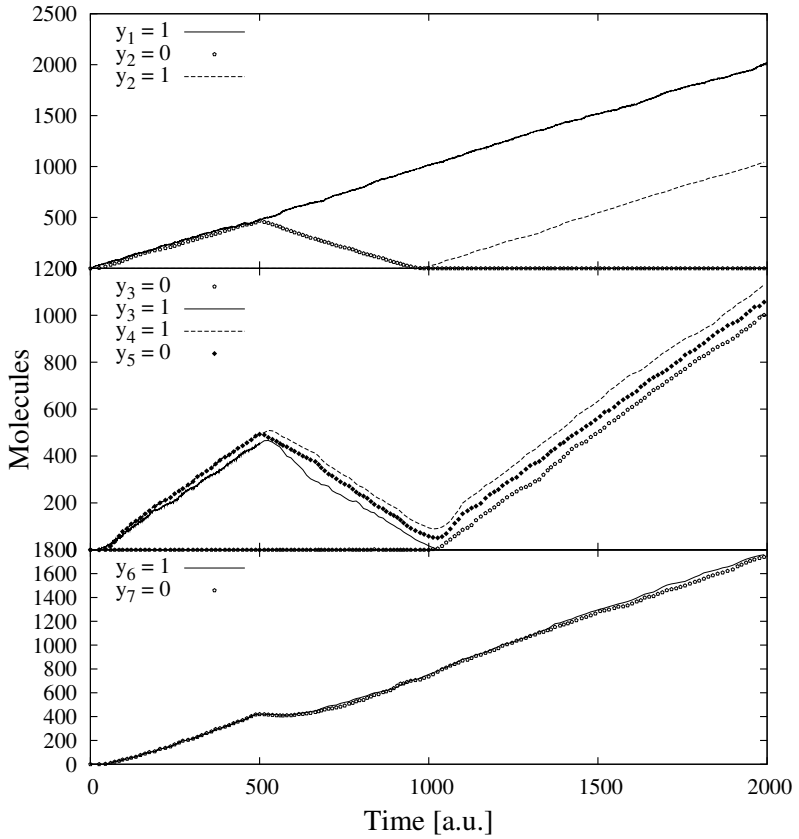
Inputs: $t = 0$ $\{a, B\}$ $t = 400$ $\{A, B\}$



Fredkin Circuits



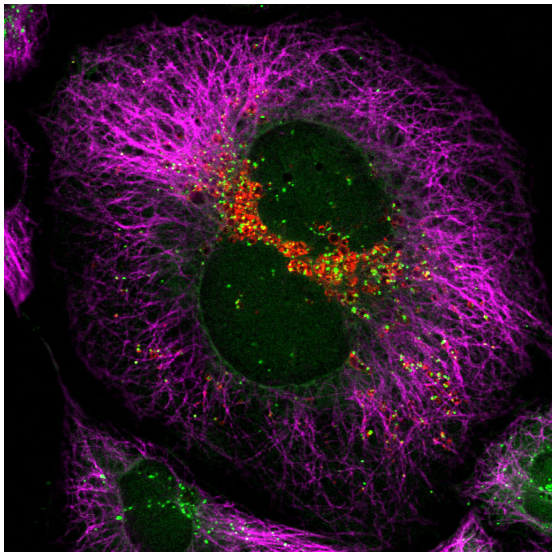


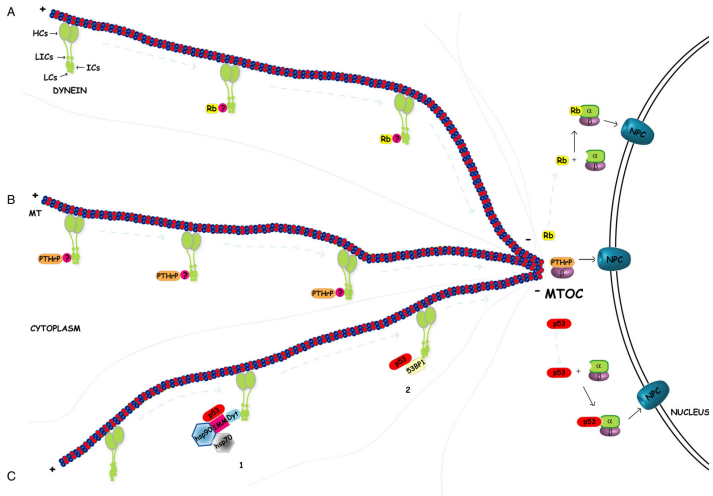


Outline

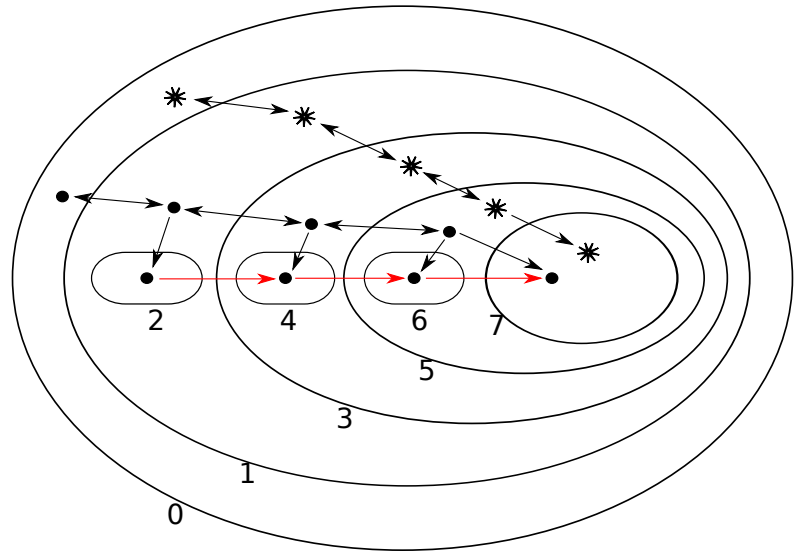
Microtubules

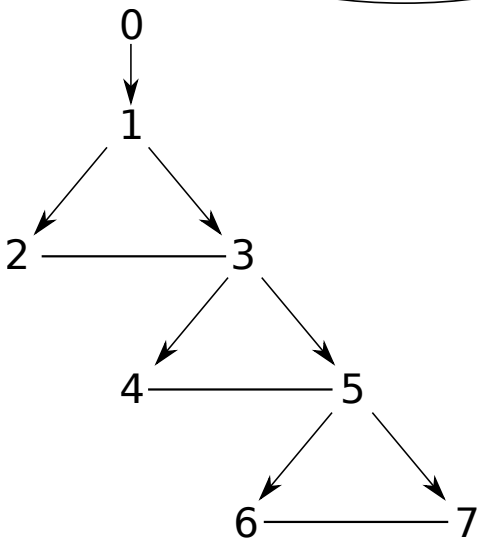
"Microtubules (MTs) can act as tracks to move cellular components based on their polarised filaments, which are organised in most cells with their minus ends located near the nucleus and their plus ends towards the cellular periphery." [Pouton et al.2006]

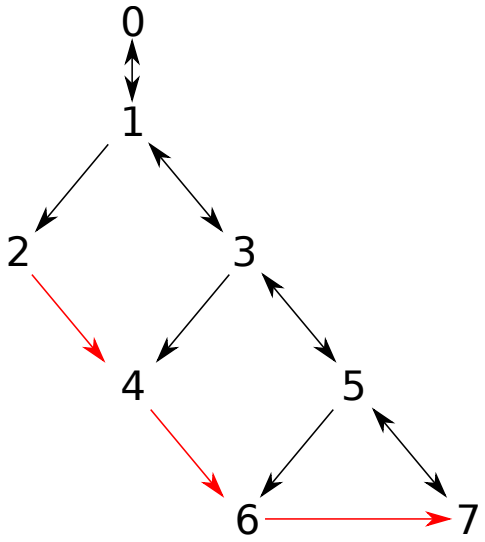




Microtubules

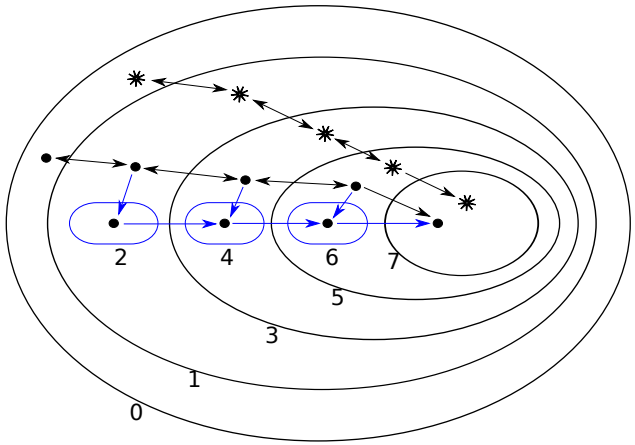


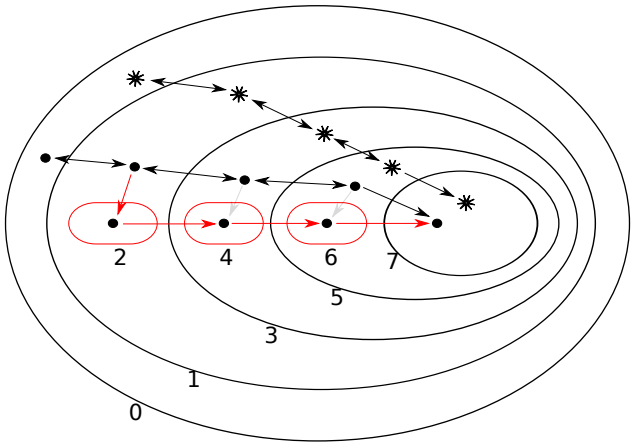




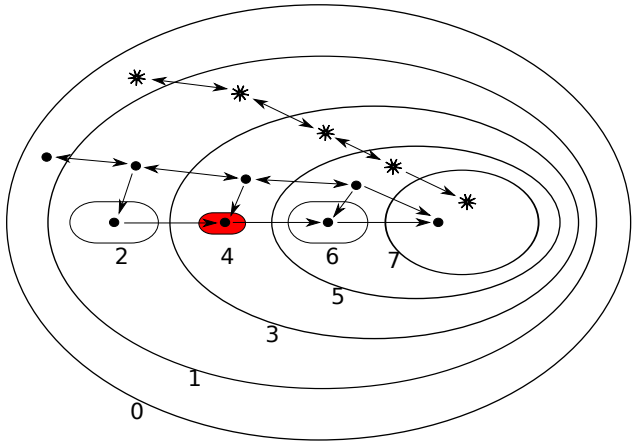
Preferred Communication and Sizes

- The influence of a preferred communication:
  pure diffusion vs whole microtubule vs sealed microtubule

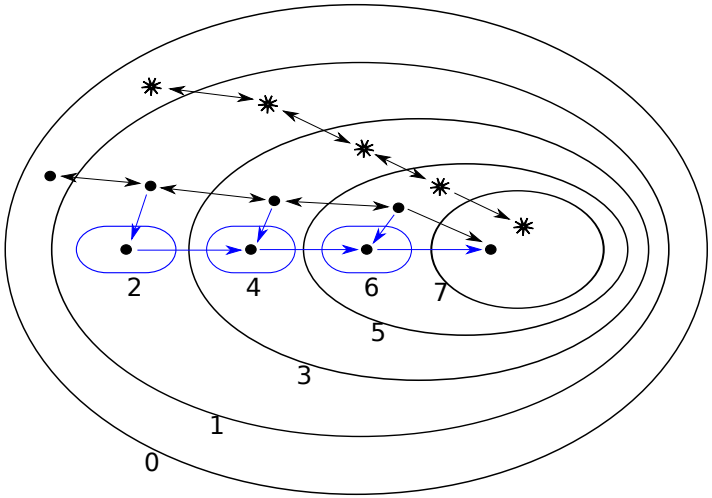


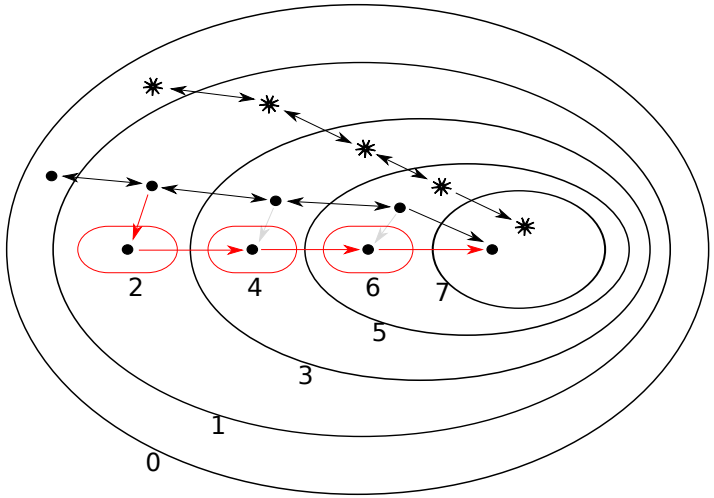


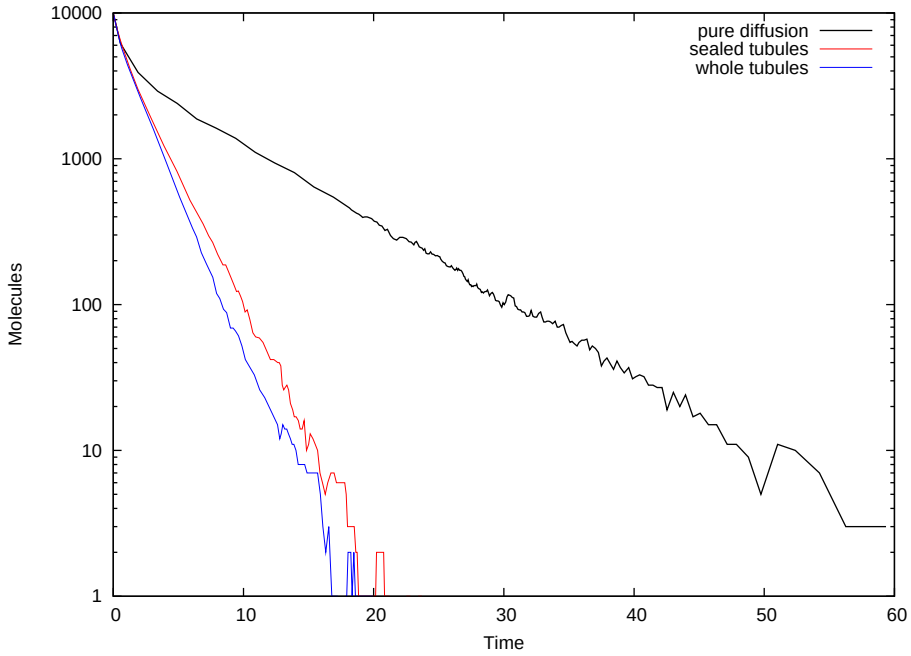
- The relevance of sizes:
  pure diffusion vs whole microtubule vs reduced microtubule

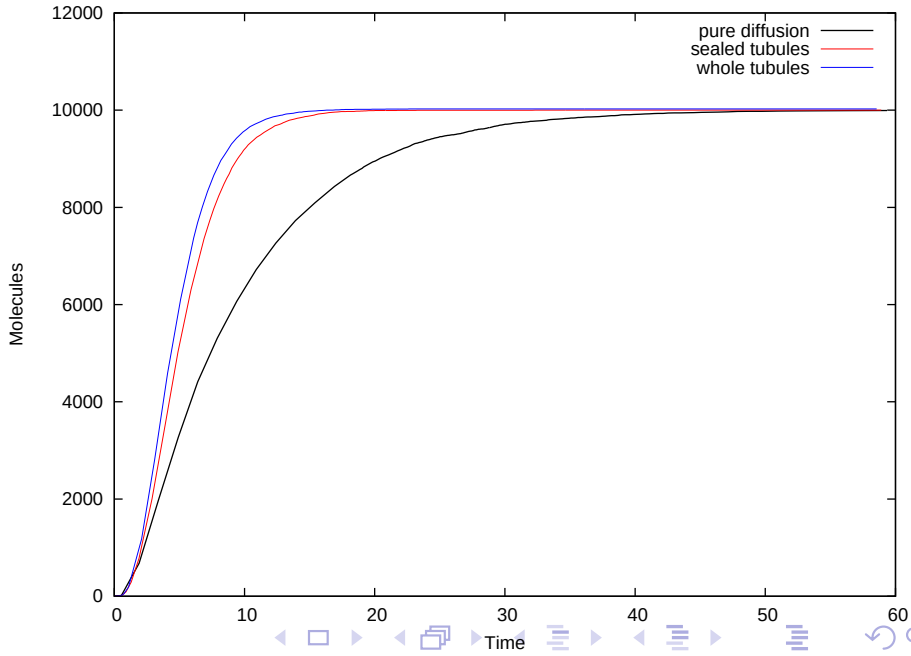


Preferred Communication - V_0, V_7

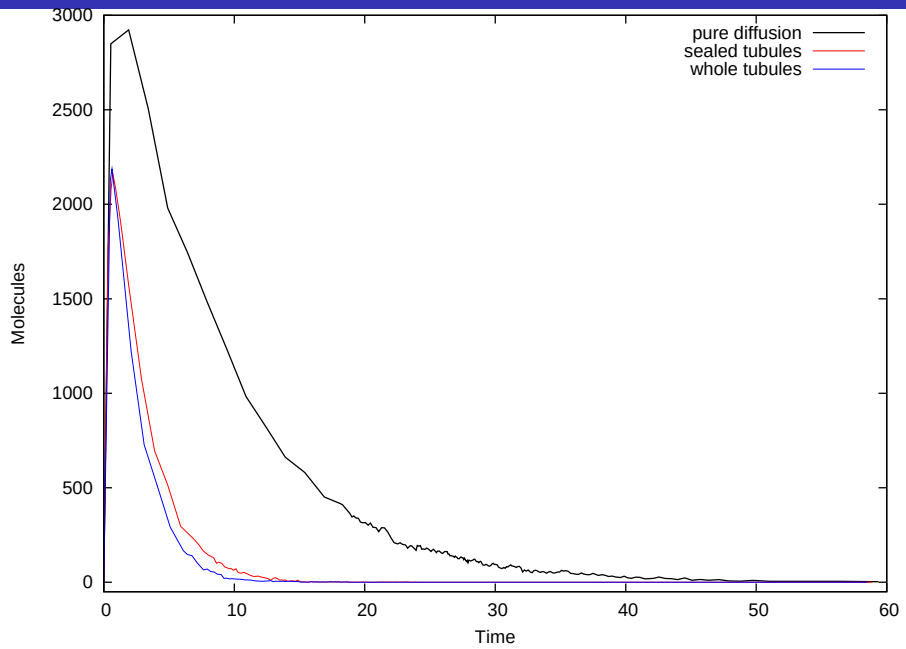


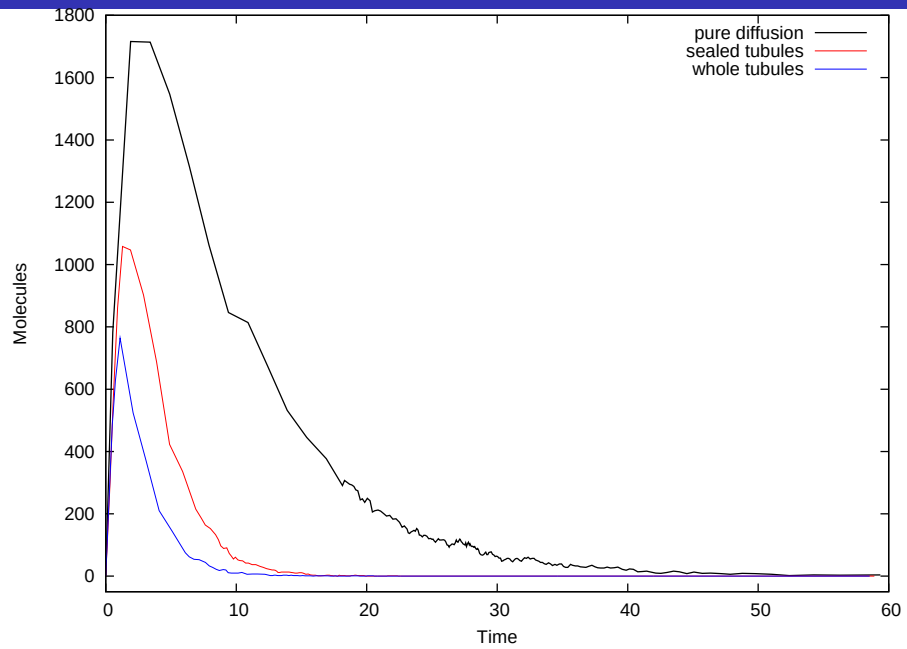


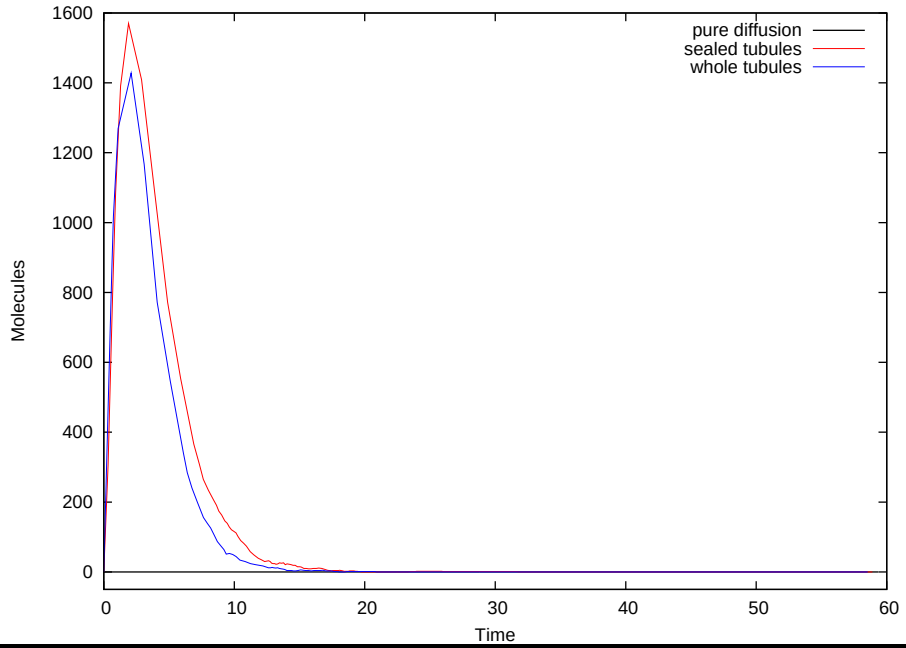


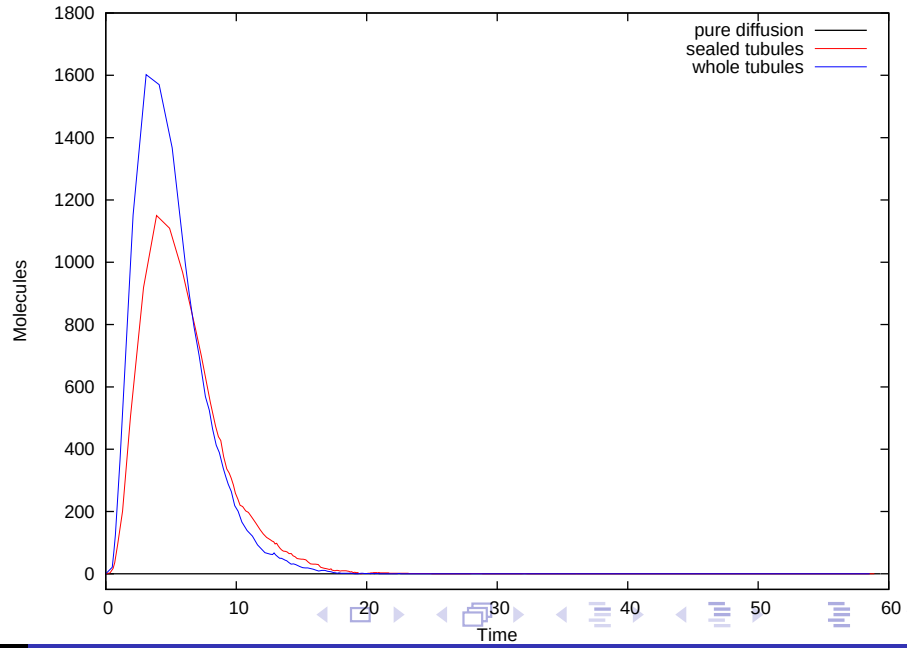


Preferred Communication - V_1, V_3, V_2, V_6

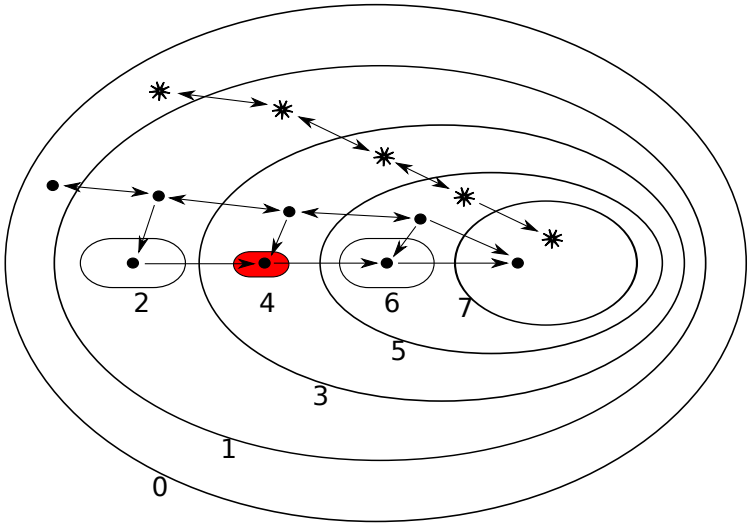


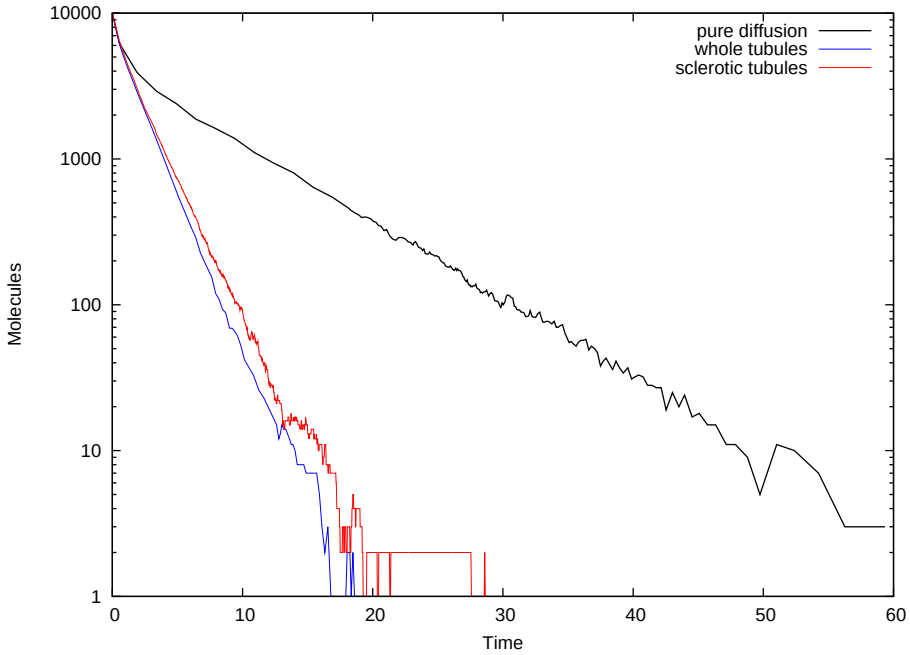


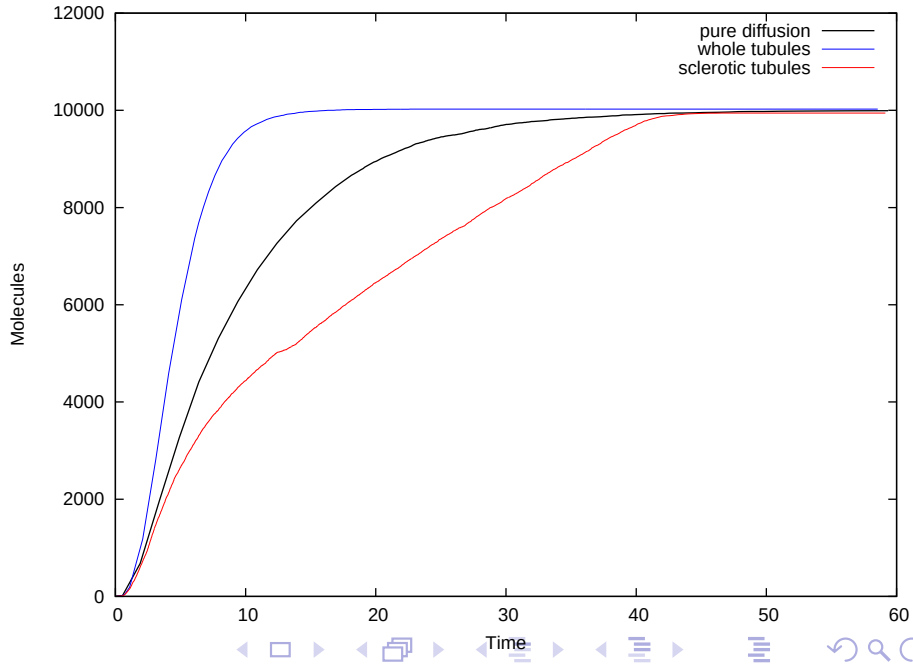




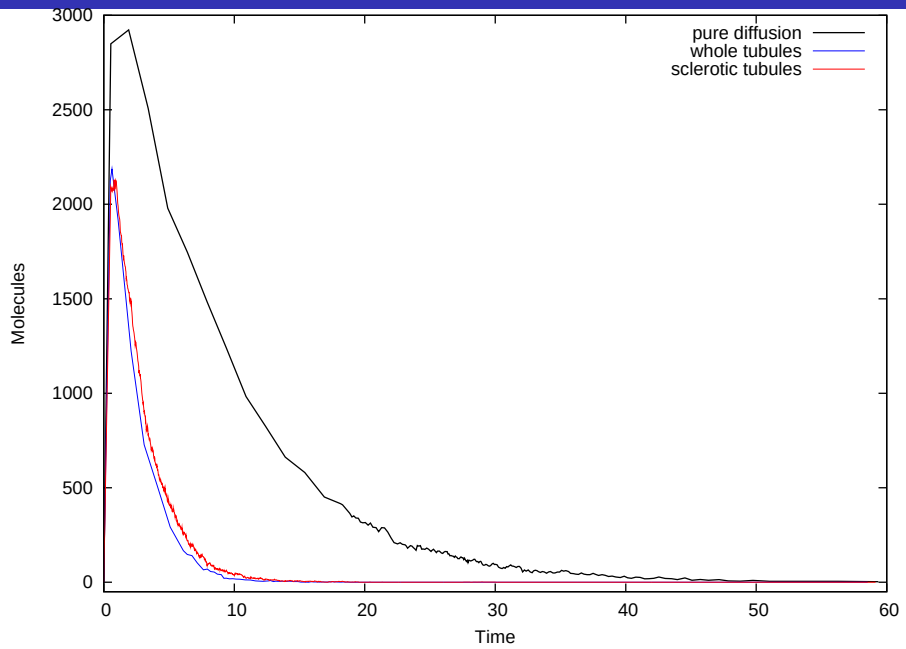
Reduced V_4 - V_0, V_7

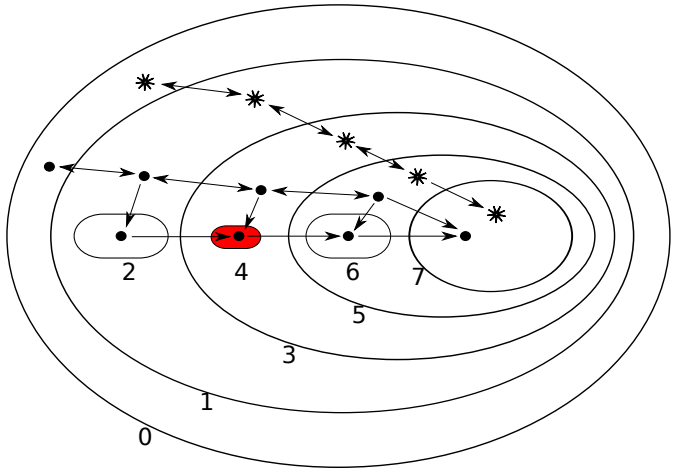


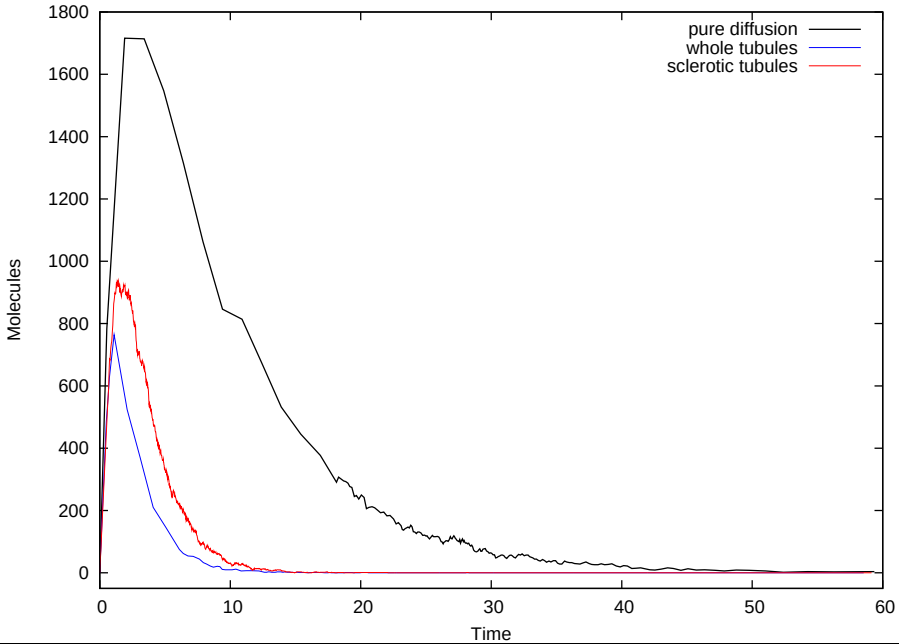


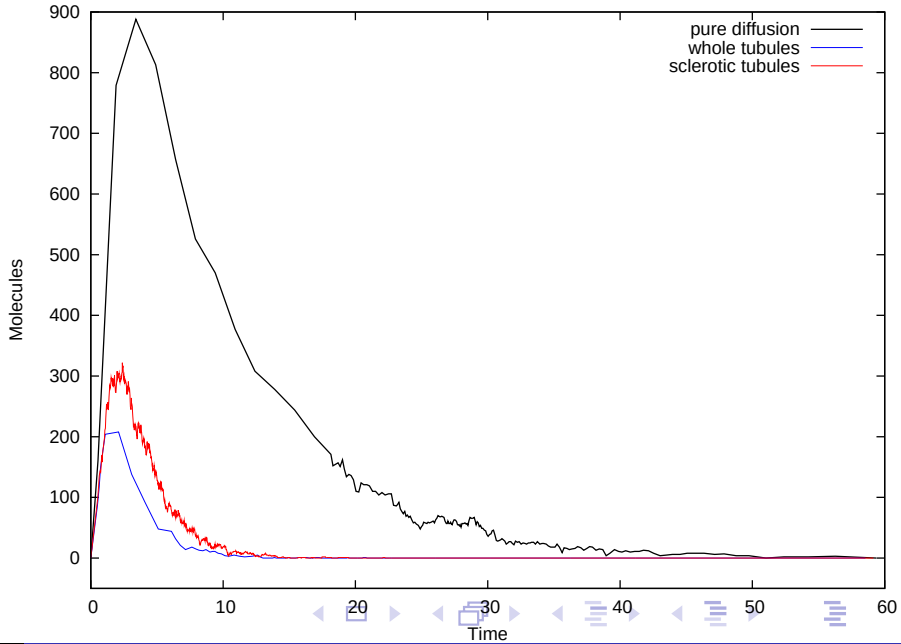


Reduced V_4 - V_1, V_3, V_5

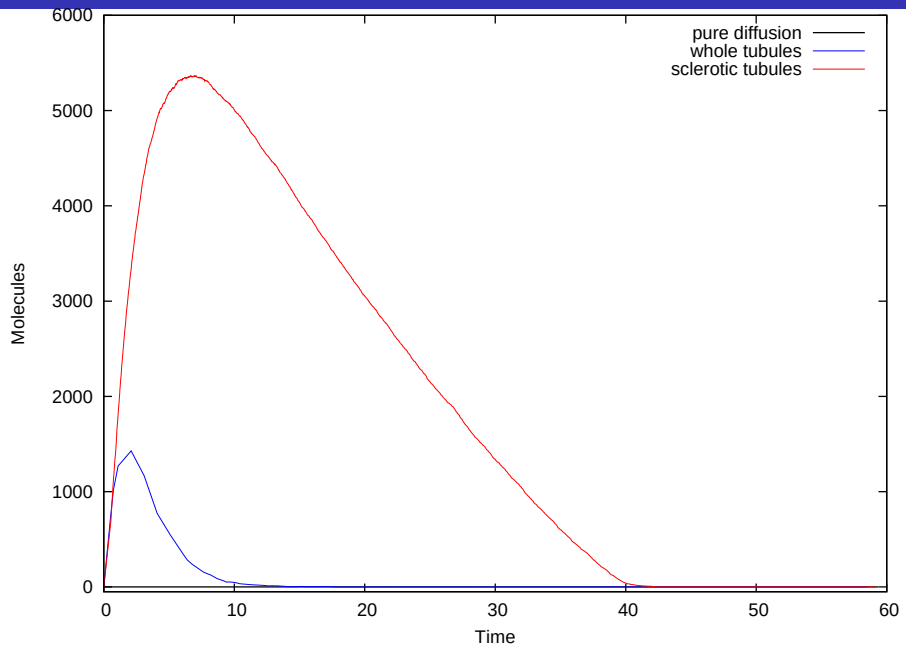


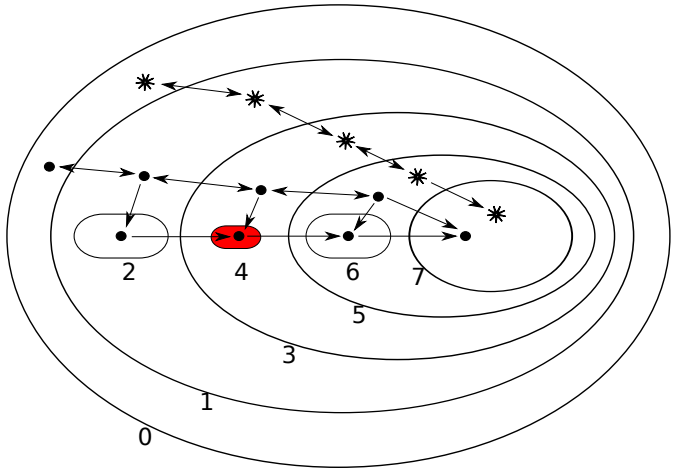


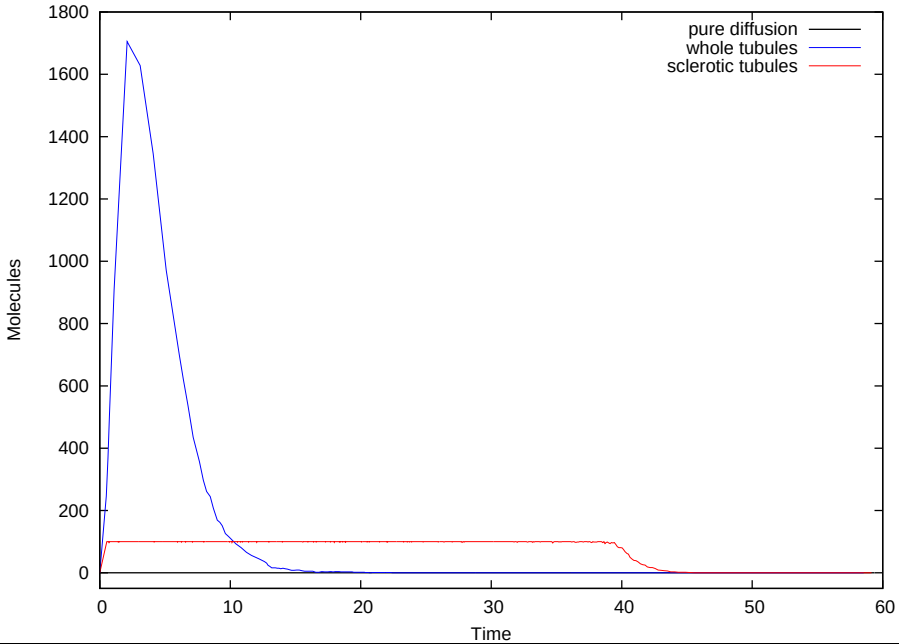


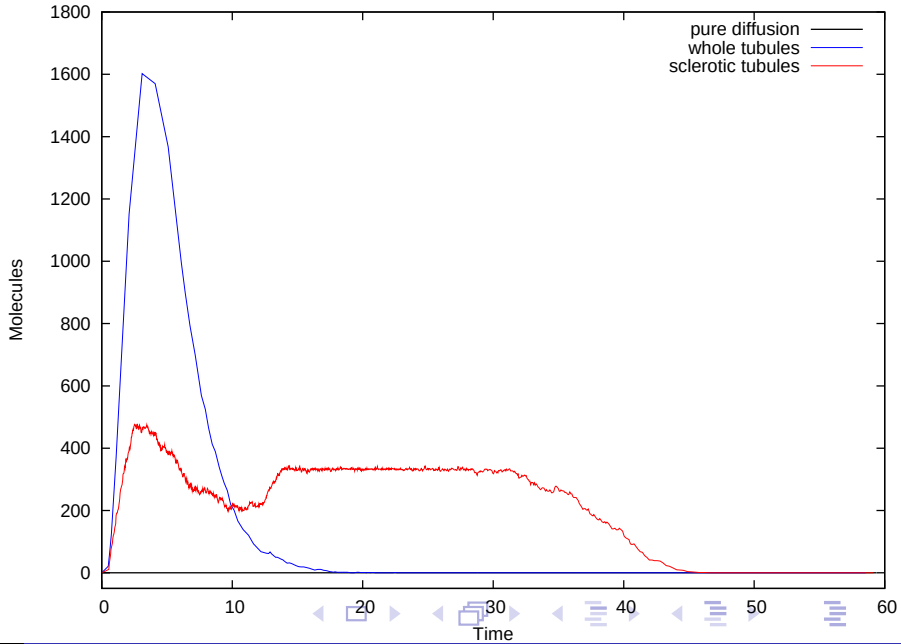


Reduced V_4 - V_2, V_4, V_6









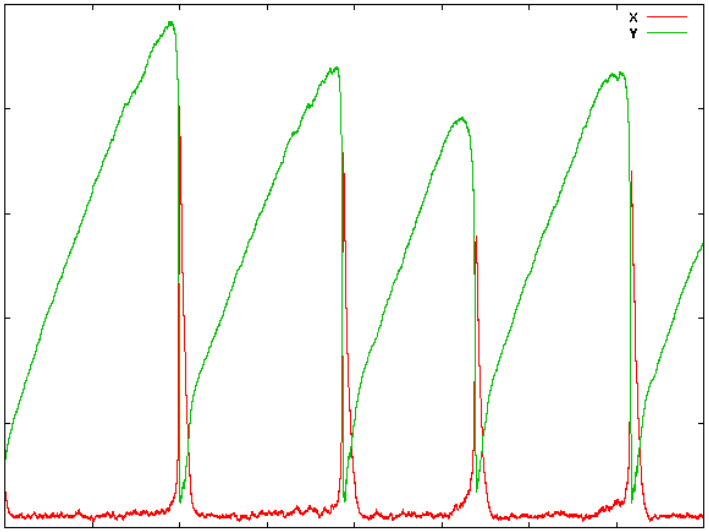
Outline

BZ Reaction & the Brusselator

The Brusselator(JCP 46 1695) is a theoretical extremely simplified version (though physically unrealistic) of the BZ (Belousov-Zhabotinskii) Reaction nowadays recognized as being the prototype for Chemical Oscillators:

$$A \longrightarrow X$$
$$B + X \longrightarrow Y + D$$
$$2X + Y \longrightarrow 3X$$
$$X \longrightarrow E$$

Parameters Influence

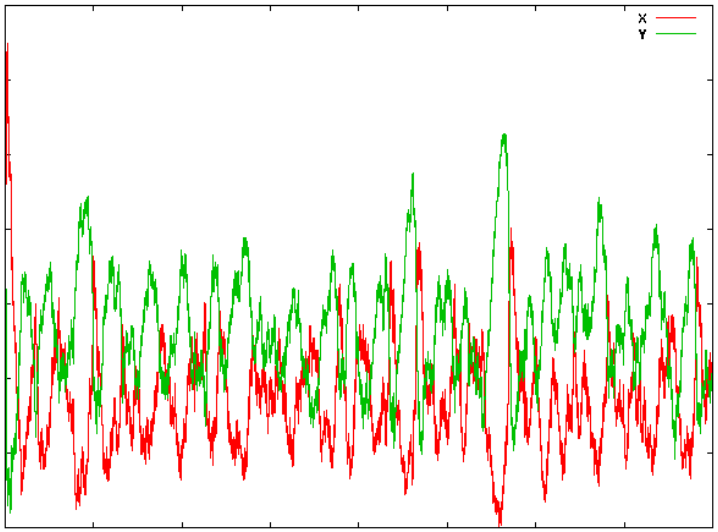


$c1 = 1$

$c2 = 5\ 10^{-3}$

$c3 = 2.5\ 10^{-5}$

$c4 = 1.5$



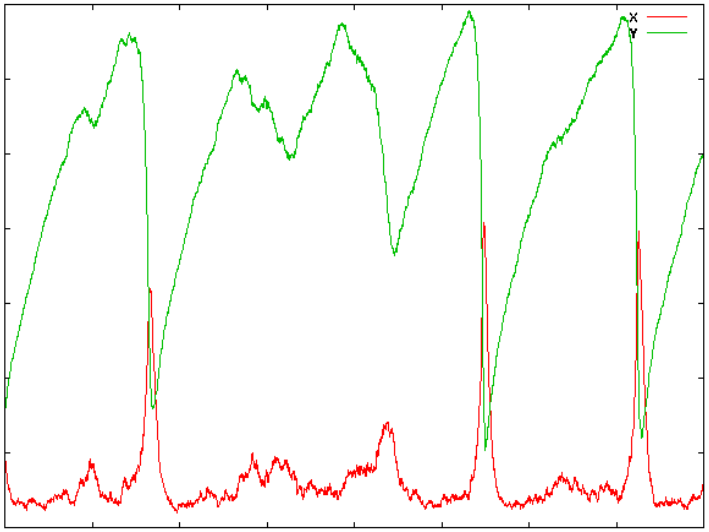
$c1 = 1$

$c2 = 5\ 10^{-3}$

$c3 = 2.5\ 10^{-4}$

$c4 = 1.5$

Parameters Influence

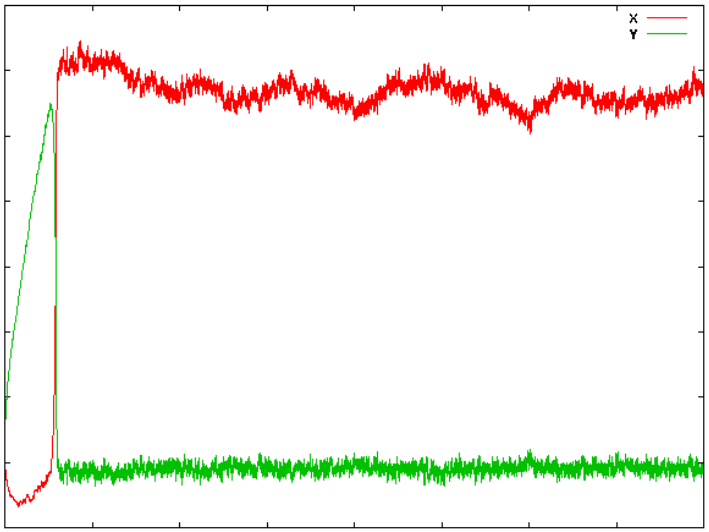


$c1 = 1$

$c2 = 3.25\ 10^{-3}$

$c3 = 2.5\ 10^{-5}$

$c4 = 1.5$



$c1 = 1$

$c2 = 5\ 10^{-3}$

$c3 = 2.5\ 10^{-5}$

$c4 = 1.5\ 10^{-1}$

Outline

Combinatorial Optimization Problems

- To solve a problem of combinatorial optimization means to find the "best solution" or "optimum" within a given set of alternatives solutions
- It is mandatory to measure quantitatively the "quality" of each solution so that a comparison is possible.

Outline

PSO

- The particle swarm algorithm has been shown to optimize hard mathematical problems in continuous or binary space (Kennedy and Eberhart, 1995; Kennedy and Eberhart, 1997).
- Particles, defined as multidimensional points in space, adjust their trajectories toward their own previous best positions, and toward the previous best position found by any member of a topological neighborhood.
- The method has been applied to a wide range of testbed problems, as well as to many applications.

PSO

- A population of particles is initialized with random positions $\overrightarrow{x_i}$ and trajectories $\overrightarrow{v_i}$, such that $\overrightarrow{x_i}(t) = \overrightarrow{x_i}(t-1) + \overrightarrow{v_i}$. At each time-step, each particle is tested in an evaluation function (the fitness).
- If the present position is better than the previous best then the current position $\overrightarrow{x_i}$ is stored in a vector b_{id}. Thus, $\overrightarrow{b_{id}}$ is the best position found so far by a particle (individual best).
- As each particle is evaluated, the best-performing particle in its neighborhood is identified and its best position is stored in the vector $\overrightarrow{b_{gd}}$. Thus, $\overrightarrow{b_{gd}}$ is the best position found so far by the swarm (global best).

PSO

$$v_{id} = w\ v_{id} + c_1\ r_1(b_{id} - x_{id}) + c_2\ r_2(b_{gd} - x_{id})$$

- w inertia weight
- c_1 weight of the individual inheritance
- c_2 weight of the social inheritance

Outline

Genetic Algorithms

The starting point is a population of individuals (subset of the solution space) and a fitness function to assess the quality of each individual.

Iteratively, the "best" individuals are selected at each generation, and variation techniques are applied to them.

The aim is to get a better adapted set of individuals to the next generation.

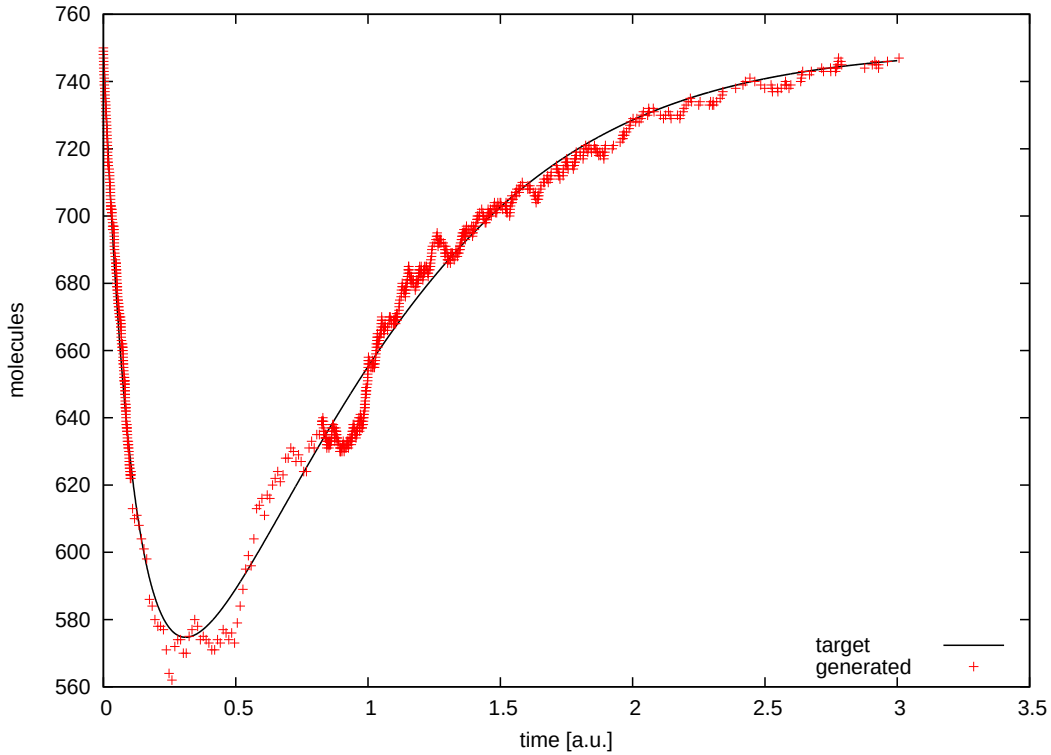
GA Operators

- One point average crossover: returns one offspring that contains at each position the average values of the parents chromosomes
- Elitism: applied to preserve the fittest individuals
- Gaussian mutation: perturbs an allele with a number drawn from a normal distribution with mean equal to the current value and σ fixed
- Range mutation: increments or decrements an allele of a prefixed quantity (that is a fraction of the size of the admissible range)
- Reinitialization: changes the value of an allele with an uniformly distributed random number in the admissible range

Outline

The Goal

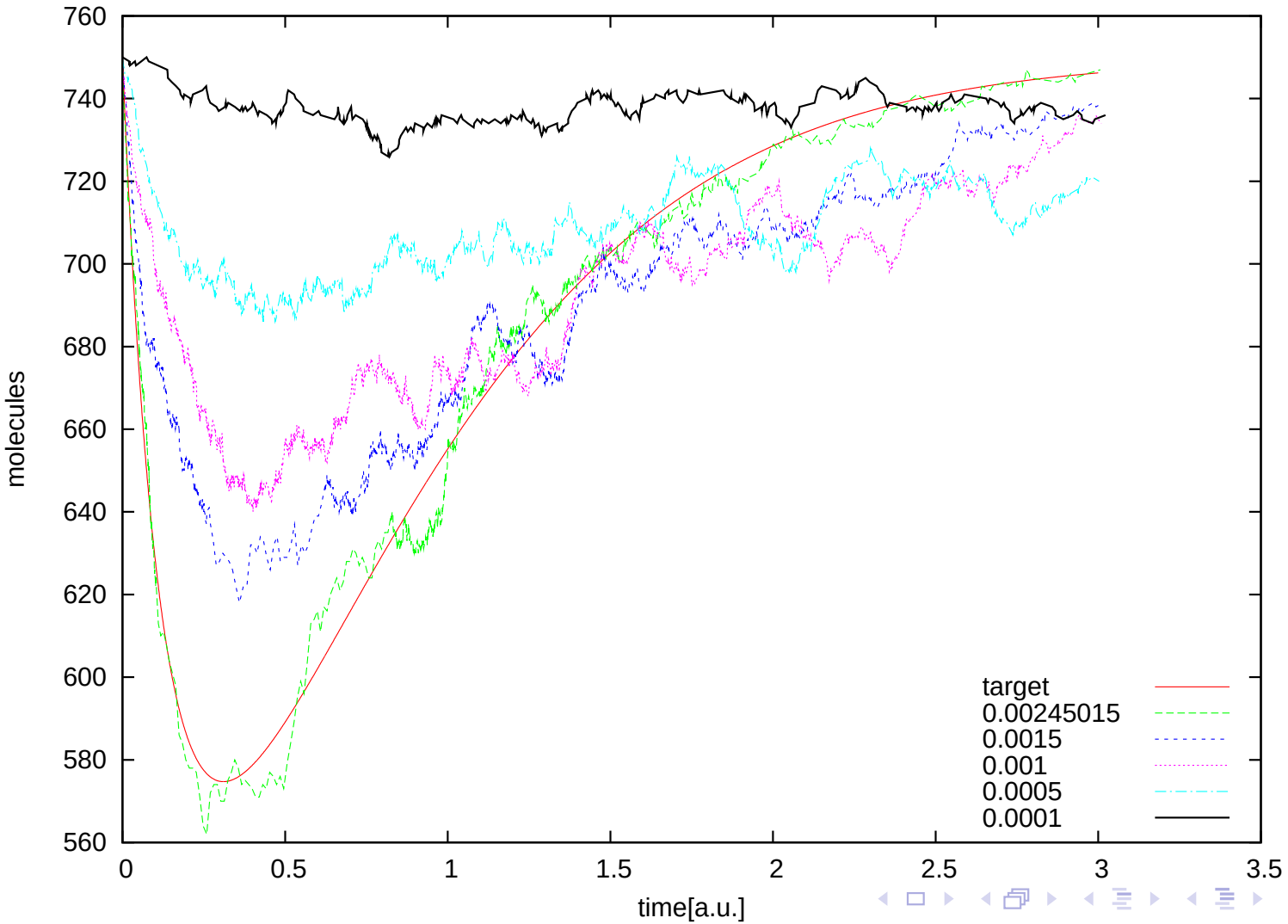
We would like to reproduce the target dynamics via stochastic simulations:



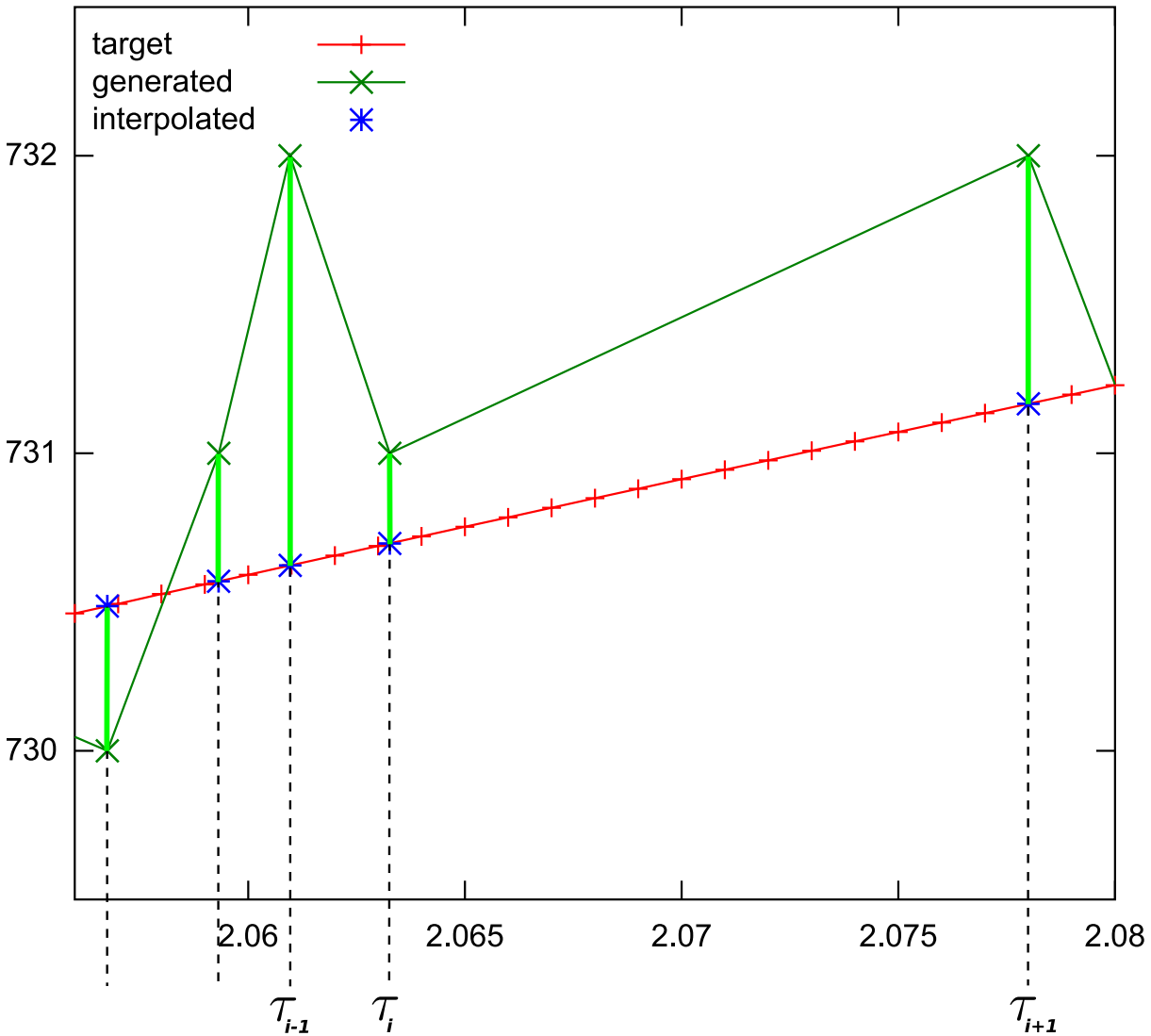
not necessarily using the same set of constants.

The Problem

Quantify the "distance" among the dynamics and find the closest one



Fitness: Standard Distance

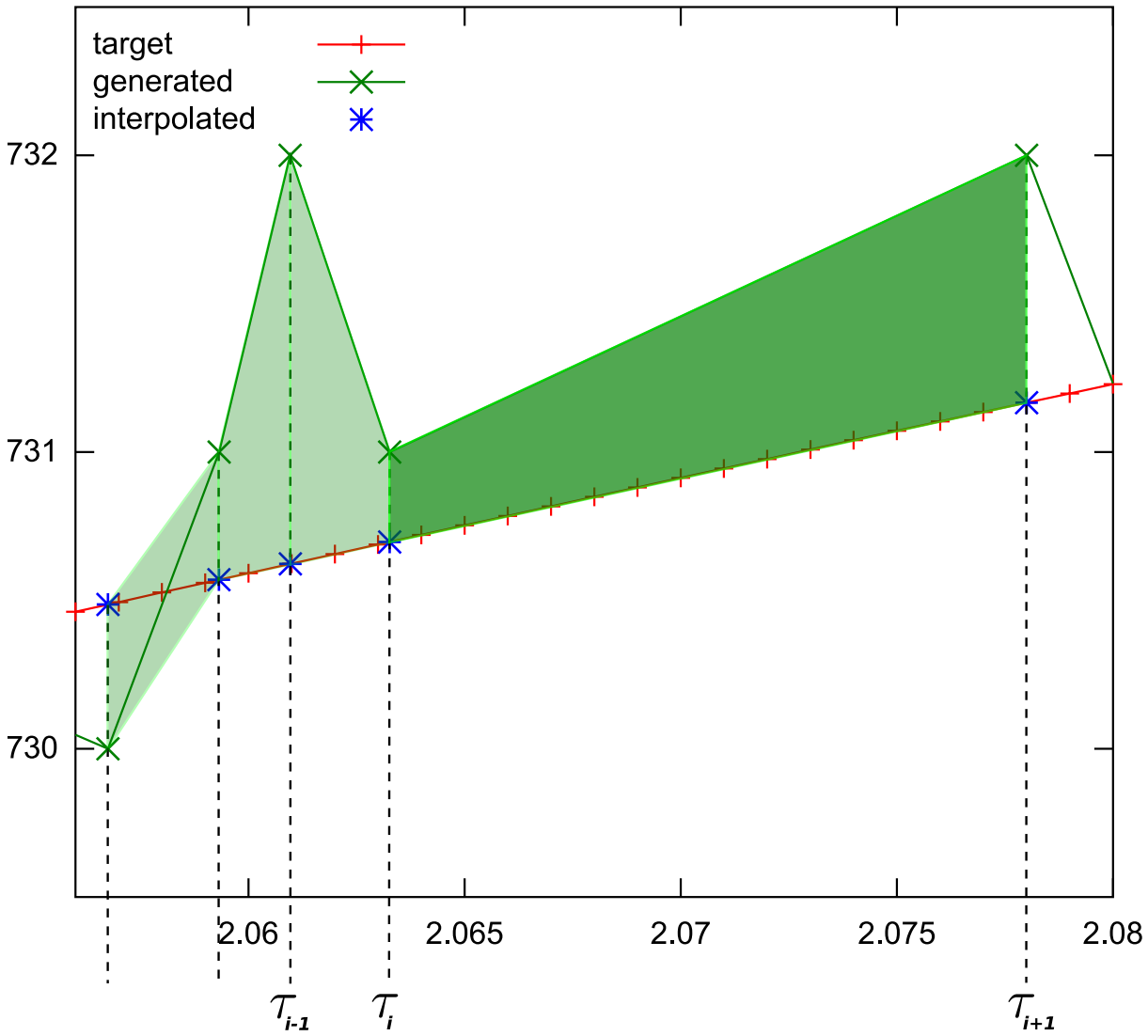


Fitness: Tricks

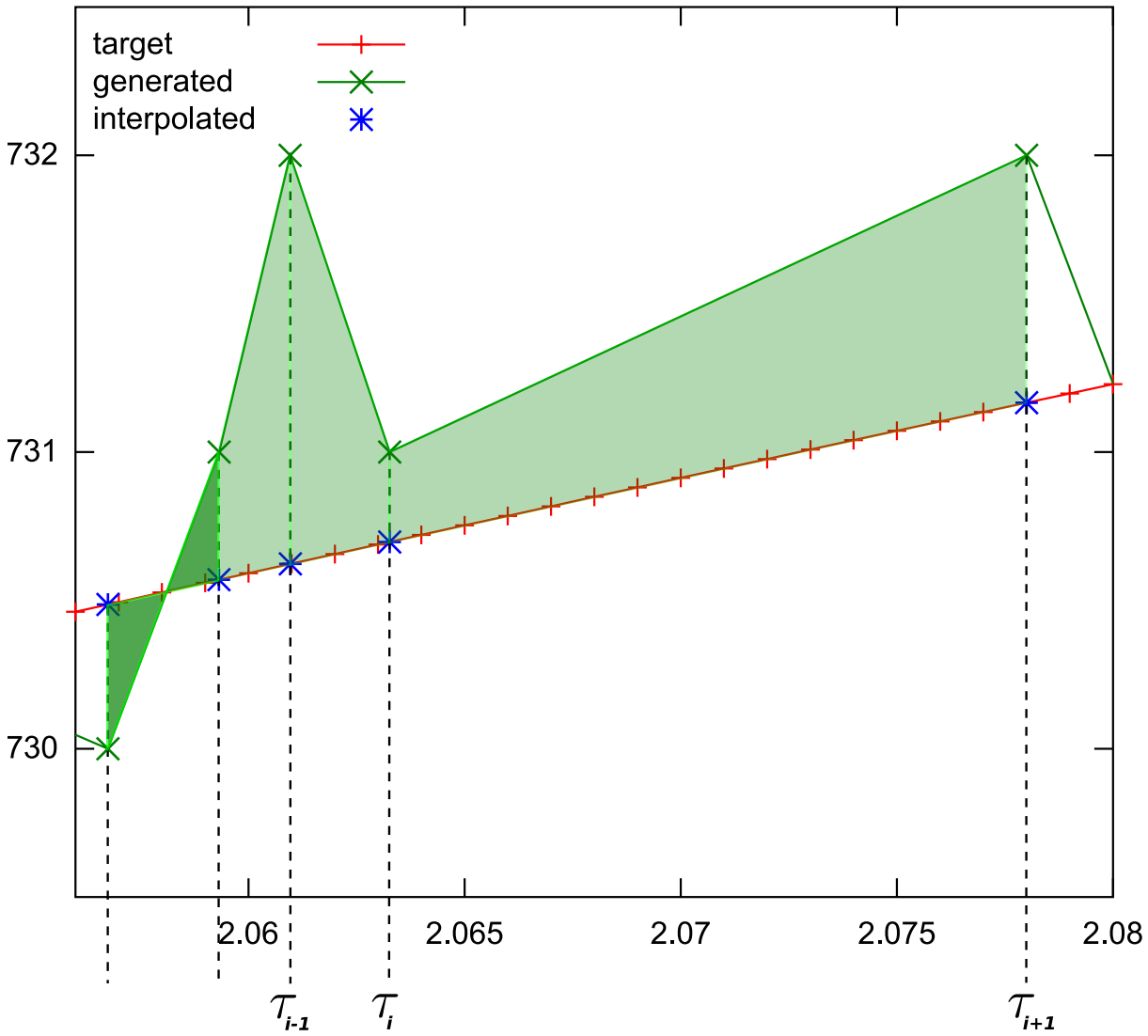
Facts related to intrinsic noise that should not be neglected:

- 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 $\{X(\tau_i)\} \rightarrow [X](t_i)$
  - exploit ensemble behavior (may flatten oscillations, not in phase outcome)
  - $\langle\ F\left(X(\tau_i)\right)\ \rangle \stackrel{?}{=} F\left(\ \langle X(\tau_i)\rangle\ \right)$
- same parameters, possibly (almost always), generate different values of the fitness function ("weak convergence")

Fitness: "Area" Distance



Fitness: "Area" Distance (refinement)



Our PSO Implementation

$$v_{id} = w\ v_{id} + c_1 r_1 (b_{id} - x_{id}) + c_2 r_2 (b_{gd} - x_{id})$$

PSO1

- decreasing w inertia weight plus a Gaussian noise
- c_1, c_2 are independents

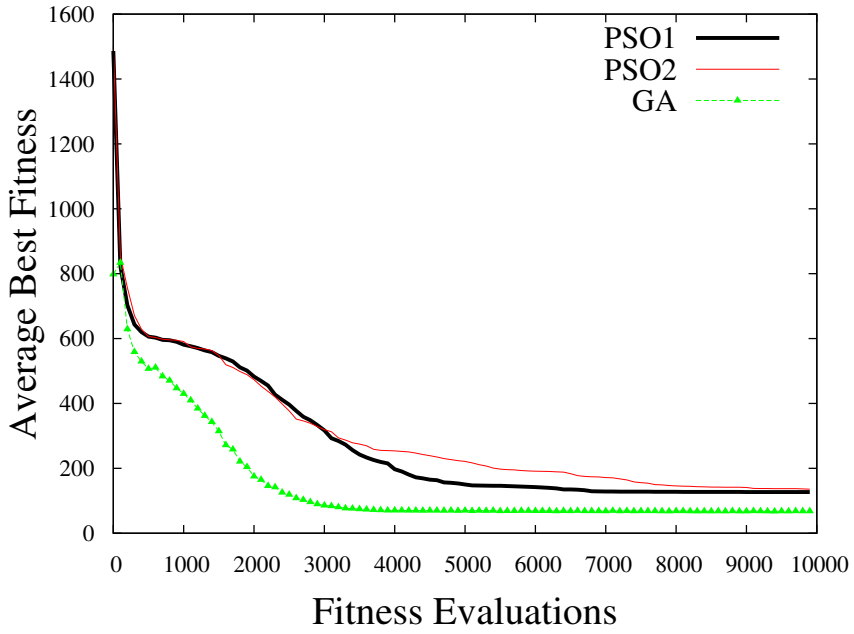
PSO2

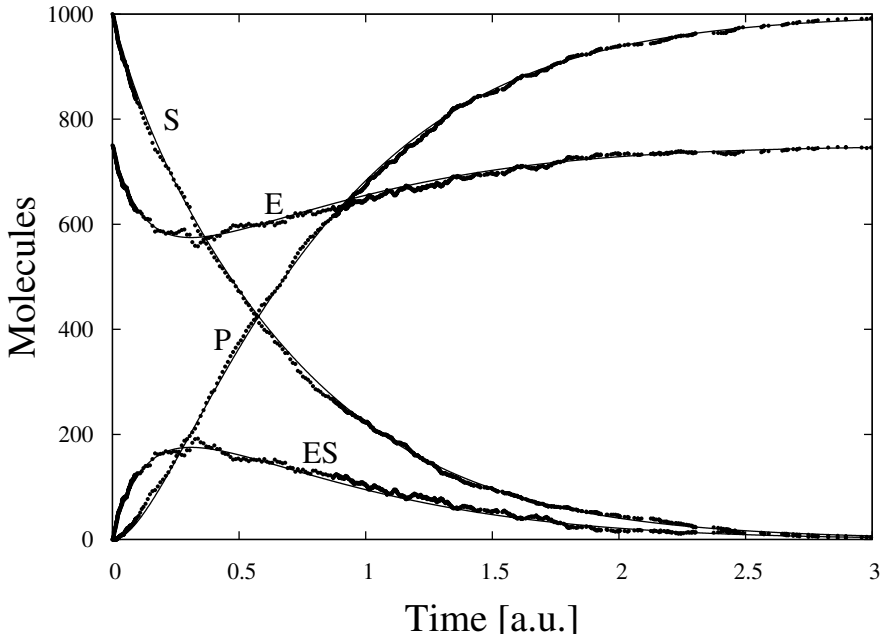
- decreasing w inertia weight plus a Gaussian noise
- c_1, c_2 co-evolve, modified with a Gaussian noise

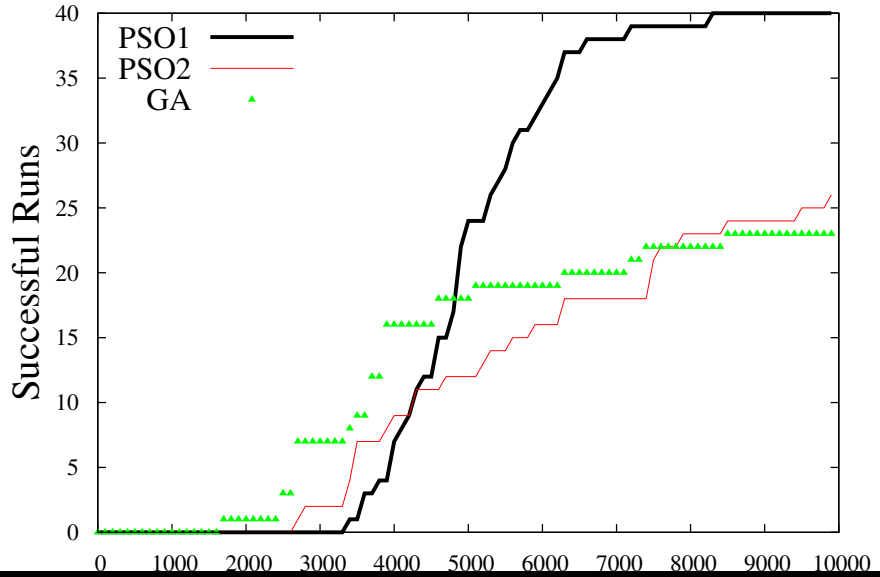
Our GA Implementation

- Elitism: 1% of the population
- Tournament strategy to select the offspring (selection pressure $= 5$)
- Crossover: gene average to generate the new individuals ($P_C = 0.95$)
- Mutation: uniform or Gaussian with $P_M \in [0.05, 0, 5]$

Comparison GA & PSO: Michaelis Menten

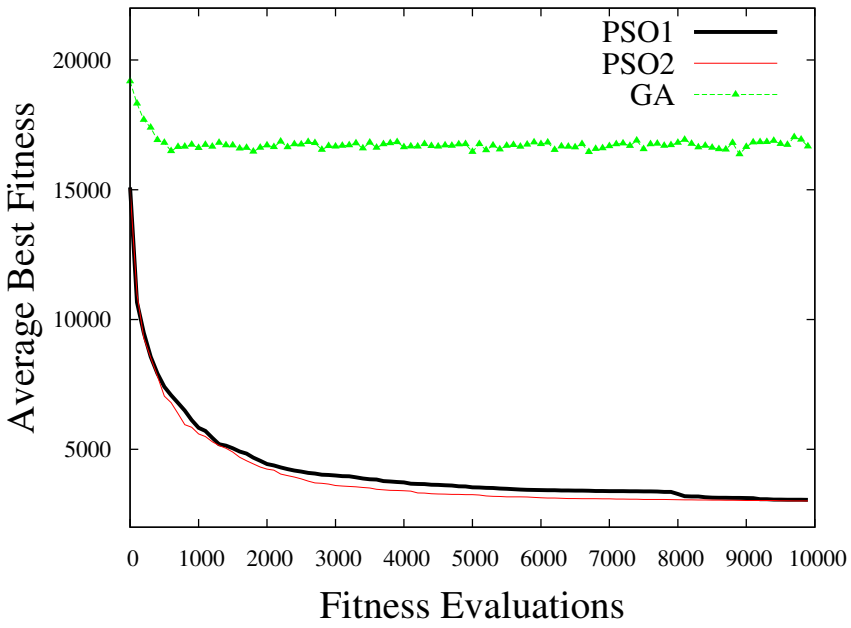


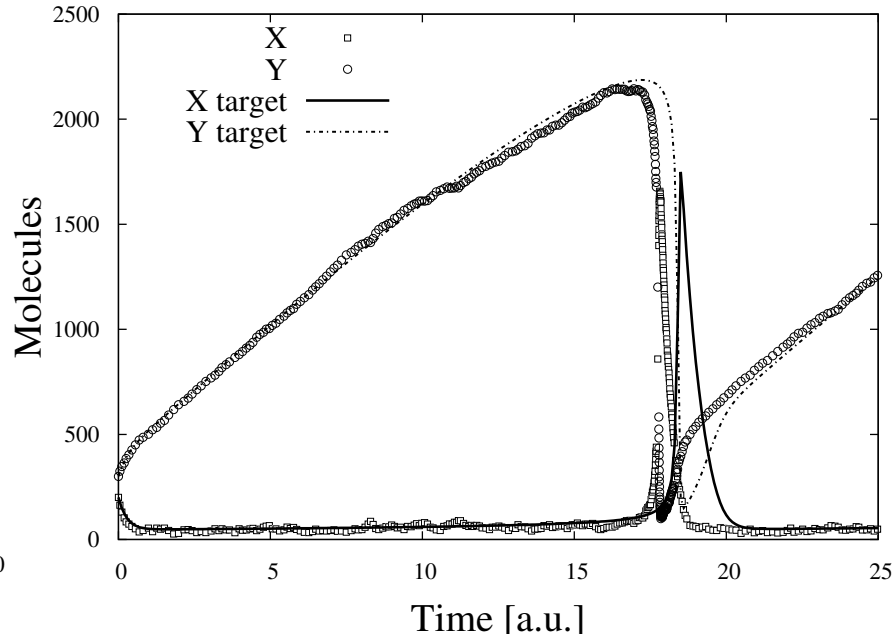


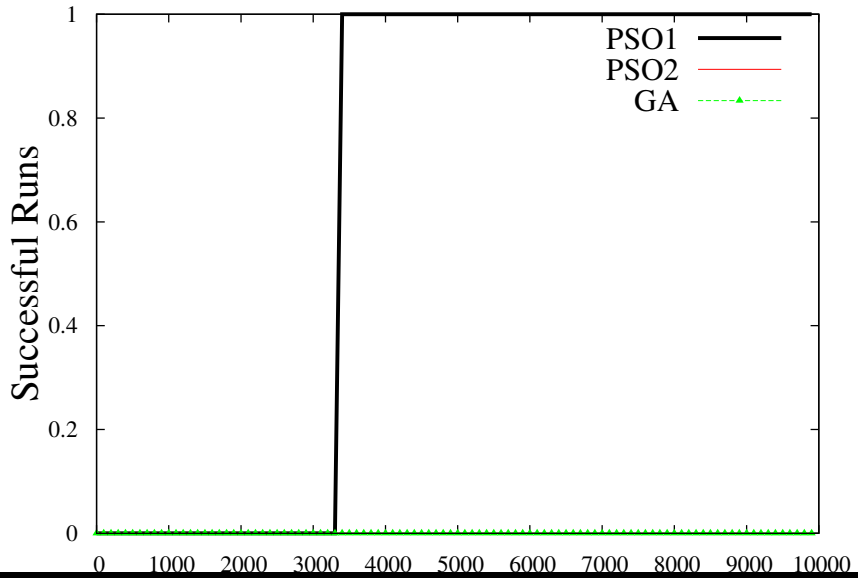


- PSO1 best performer
- similar average fitnesses
- dynamics faithfully reconstructed

Comparison GA & PSO: Oscillating Brusselator

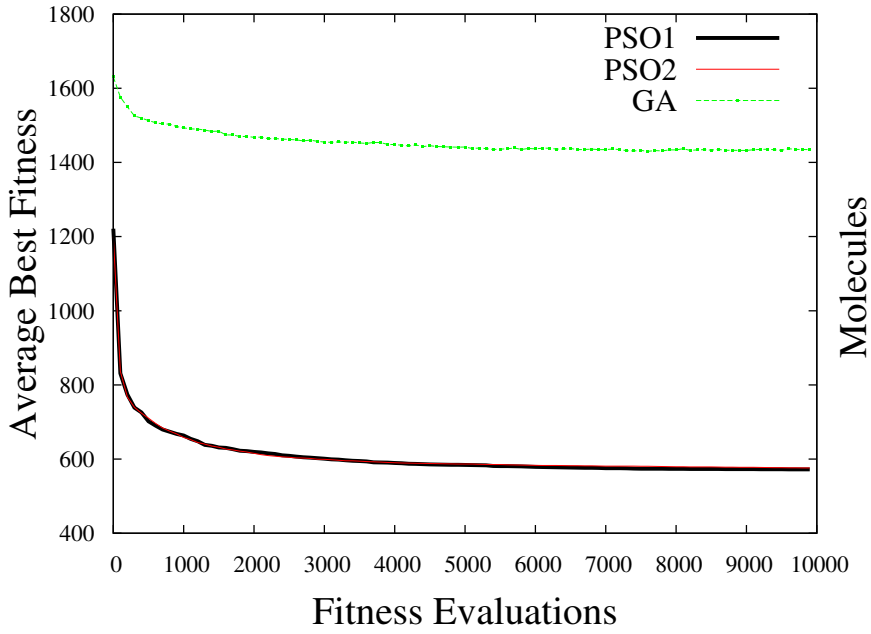


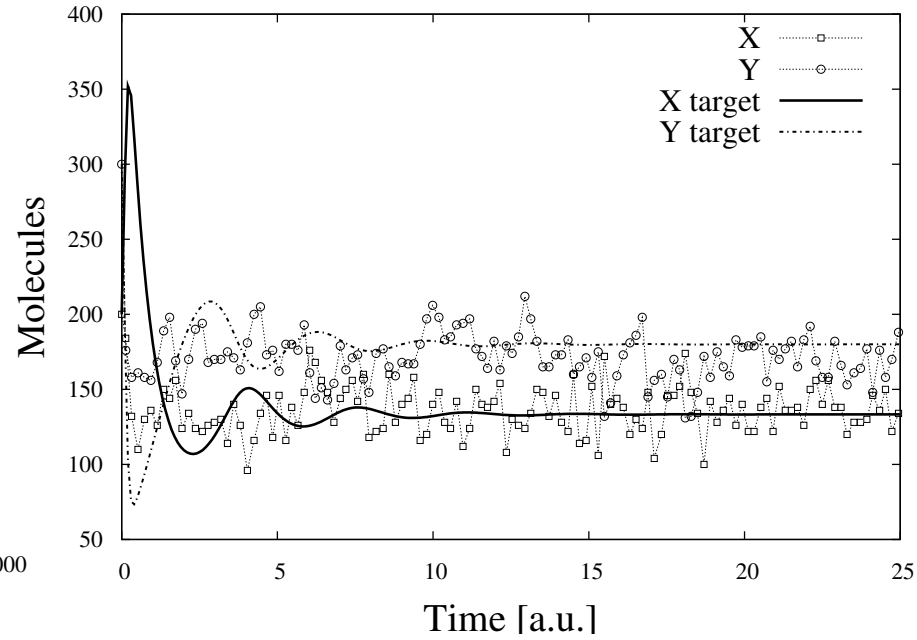


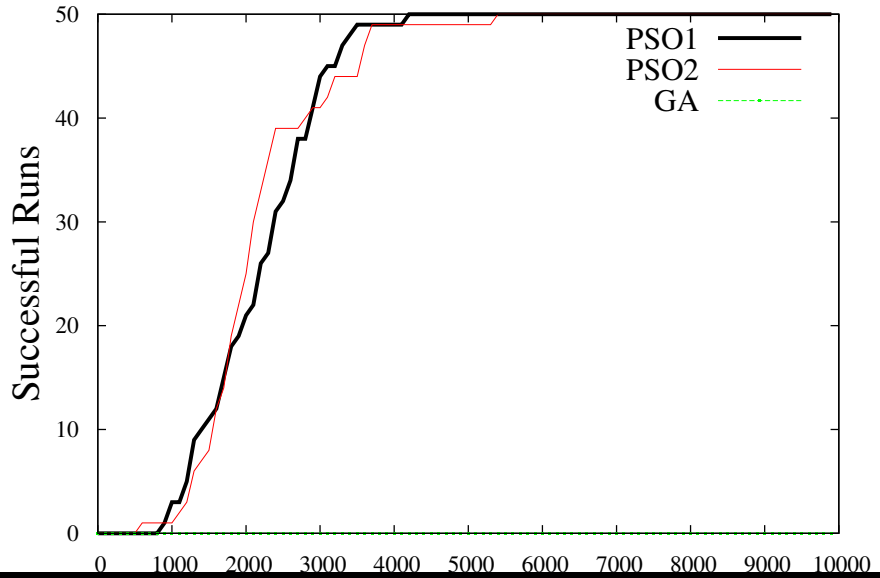


- PSO1/2 best performers
- GA worst average fitnesses
- dynamics faithfully reconstructed

Comparison GA & PSO: Dumped Brusselator







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

Ongoing works

Simulation tools

Integration of the notion of membrane potential in τ-DPP
Address parameter reconstruction issue: Genetic Algorithms, Particle Swarm Optimizer, etc.

Analysis

Tools for the analysis of the system dynamics (i.e. sensitivity analysis)
Analyse the properties of multi-stable systems
Role of "noise" in Molecular Dynamics-Computing

Modelling of biological systems

Ras/cAMP/PKA signalling pathway in yeast
Signal transduction in bacterial chemotaxis
Neuron cells and synaptic processes