Doctoral Thesis Proposal

Autonomous Interactive Agents for Population Ecology

Aureliano Rama

Computer Science Department, University of Pisa

April, 14th 2010

Outline

- Introduction
- Population Ecology
- State of the art
- My proposal

Introduction

In the last few years many formalisms originally developed by theoretical computer scientits with the aim of describing concurrent interactive systems have been adapted to describe biological systems (in particular biochemical and cellular systems in the framework of systems biology). Some examples are:

- Petri Nets
- Pi-calculus
- PEPA
- Mobile Ambients
- Concurrent Constraint Programming
- Hybrid systems

Introduction

Moreover, some new formalisms have been developed with the specific aim of describing biological systems. A few examples are:

- k-calculus
- Beta-binders
- Calculus of Looping Sequences

Formal modelling allows systems to be described without ambiguities, and allows formal analysis techniques and tools to be developed.

Introduction

Nonetheless, no formalism was created to specifically address the description of complex ecosystem or population ecology.

Population ecology studies dynamics of often large group of individual animals (of one or more species) that comes from interactions between them and with the surronding environment.

Interesting problems arise in this field:

- extintion rates and causes
- reintroduction issues
- control of parasite populations
- genetic and evolutionary studies

Introduction

At the moment, works and studies that address these problems do so by means of mathematical models (classical approach) or simple ad-hoc simulations.

My goal is to develop a formalism that allows to describe such systems in a simple and non ambiguous way and to apply to them formal analisys technics, like model checking (besides the simulations).

Idea: define a formalism based on automata.

Automata are simple to understand also for non-computer science researchers and they are easy to study with formal tools (like model checkers).

Population Ecology

Population Ecology

Naturalists and biologists need to model and simulate large populations of interactive animals.

Two possible roads
- classical: differential equations
- modern: individual-based models

In any case we need to confront with motivation.

Pragmatic motivation: use the model as a tool

vs

Paradigmatic motivation: search for the overall meaning and the theory

Classical approach

- No reference to behavior
- Everything is reduced to population growth

Successful in devising and demostrating concepts such as density dependence and intra- and inter-specific competition.

Limited success in being predictive: they can produce understanding of the observations but the predictions are not testable

Individual-based models

Bottom-up approaches that start at the individual level
Can determine what individual properties and what elements are essential to the overall population dynamics.
Can include spatial dynamics and abiotic factors

Difficulties in the determine the right resolution and to have the same one in the whole model
Only partial knowledge is available at small resolutions

Individual-based models

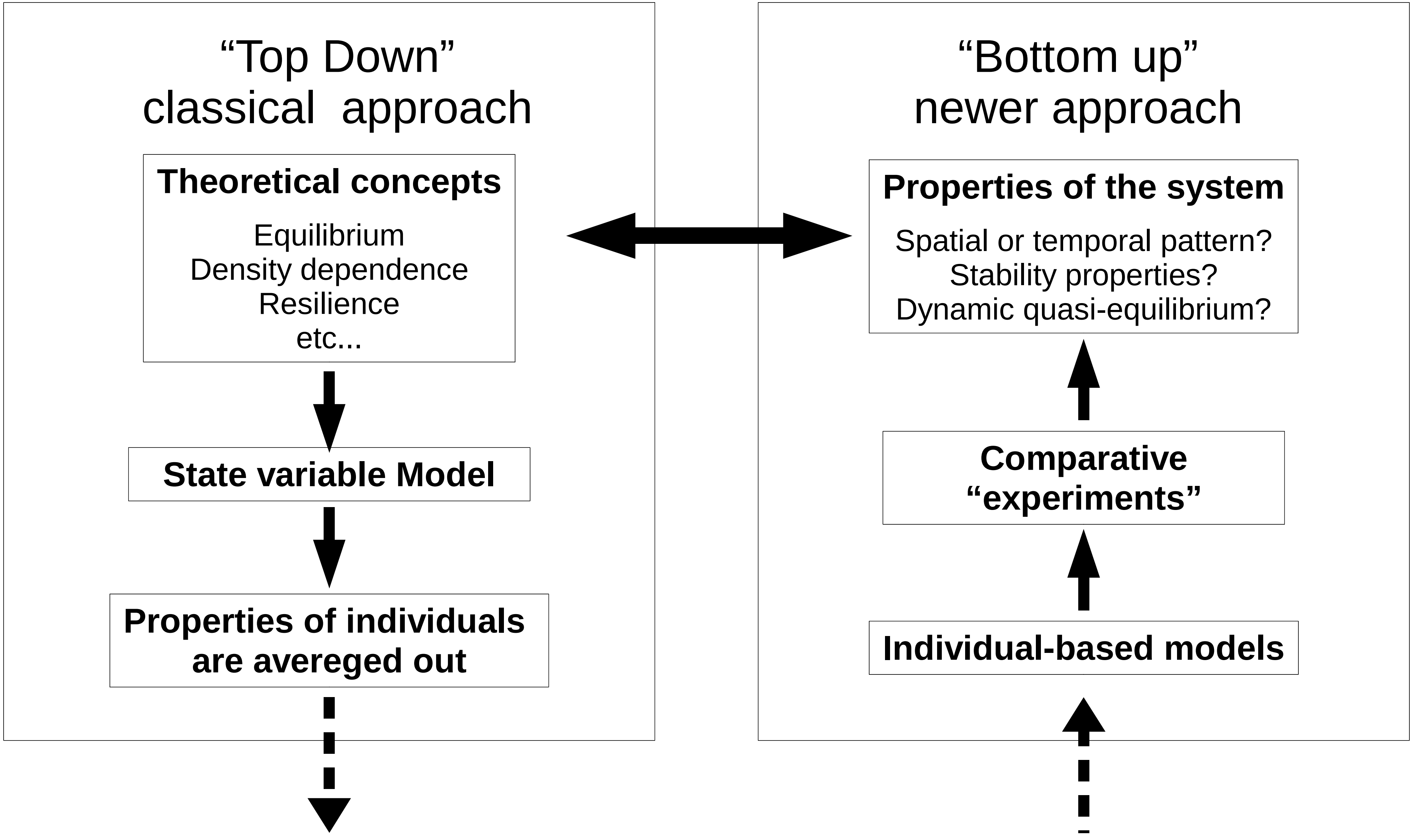
Supposed to be more testable as closer to reality.

But closer to reality means more parameters.

Which in turn means more effort to determine them.

Moreover, the search for the theory behind the model is more difficult since overall properties must emerge (and be recognized) from the details.

Approaches in ecological modelling



“Top Down”
classical approach

Theoretical concepts

Equilibrium
Density dependence
Resilience
etc...

State variable Model

Properties of individuals are avereged out

“Bottom up”
newer approach

Properties of the system

Spatial or temporal pattern?
Stability properties?
Dynamic quasi-equilibrium?

Comparative “experiments”

Individual-based models

Individuals: complex life cycles, individual variability, local interactions, resources, heterogeneous habitat, spatiality, etc...

Population Dynamics

As said, population ecology studies dynamics of large group of individual animals.

More often than not, the group species are intermixed, with different roles in the ecosystem: prey/predators, competitors for resources, costructive and destructive interactions, etc...
The environment itself is an important factor in these dynamics.

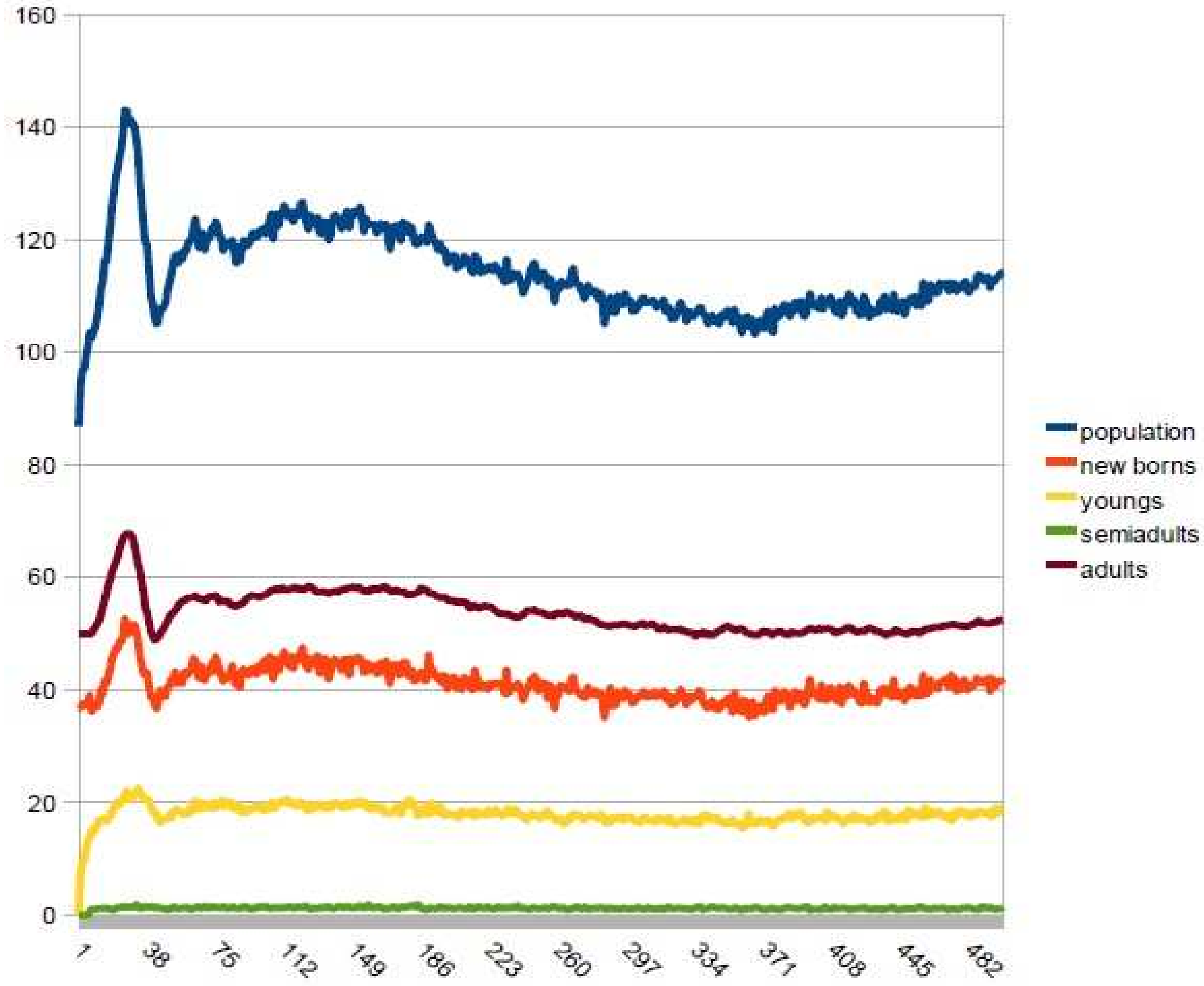
Biologists and zoologists are interested in simulations and predictions and have often quite large data sets that need automatic tools to analyze and with which validate models.

Case Study:
Emys orbicularis (pond turtle)

In a join work with Marco Zuffi (Museum of Natural History and Territory, University of Pisa) we developed a simulator to estimate population trends in groups of Emys, given the conditions and survival rates found in central Italy.

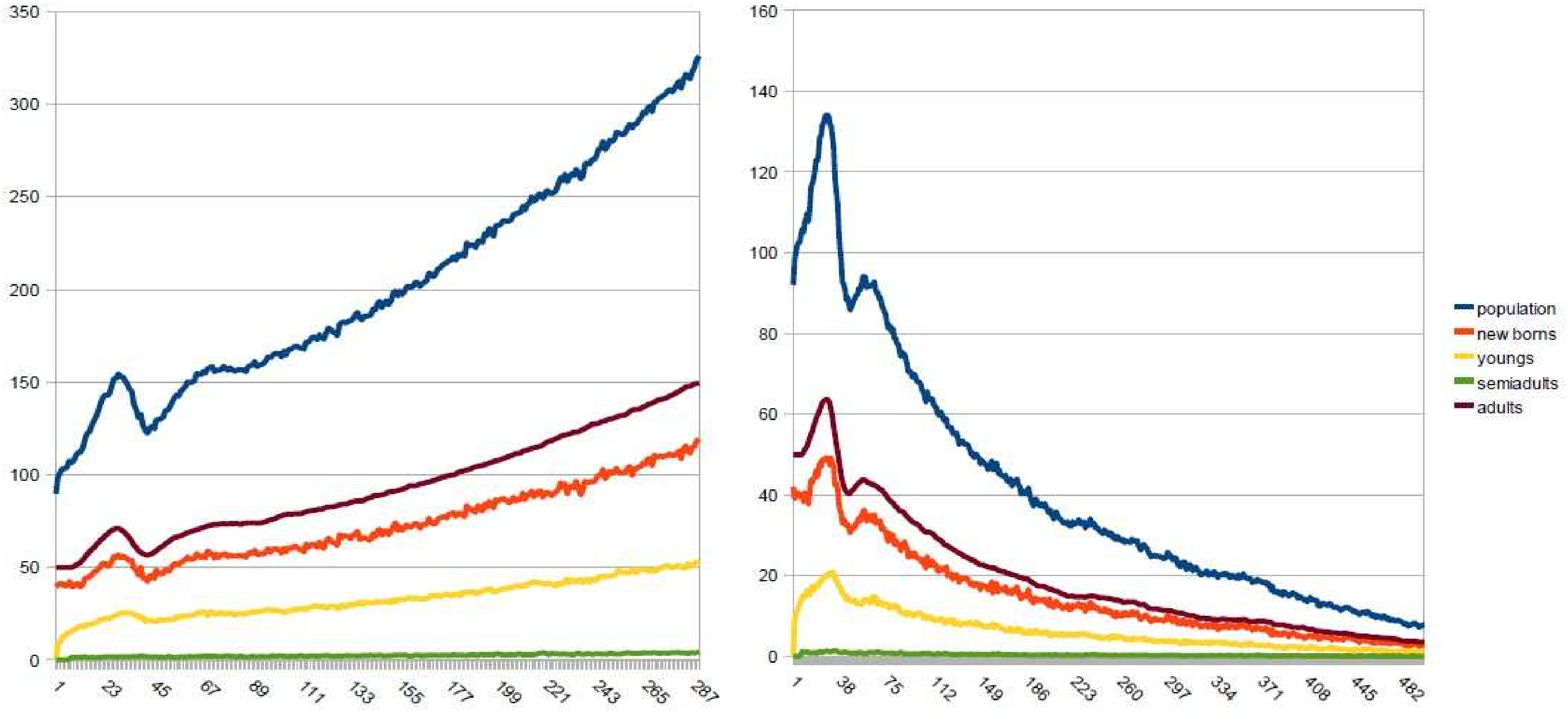
“*Headstarting in the European pond turtle, Emys orbicularis: a computational approach and a proposed model for management plans*”
in Seventh Annual Symposium on Conservation and Biology of Tortoises and Freshwater Turtles

Emys Simulator Results



A good model for long term prediction (shown 500 years time frame), consistent with the known dynamics of similar populations

Emys Simulator Results



Extreme sensibility to survival rates:
less than 1% change in any of it is enough to loose the equilibrium

Case Study:
Tropheus Moorii (tropical cichlid)

These little colored fishes, natives of african central lakes (especially lake Tanganika) with over 40 different morphs distributed throughout the lake.

This form of peripatric speciation (speciation with loosely interconnected habitats) is quite difficult to explain given the existing models.

Tropheus Moorii

At the Museum of Natural History and Territory in Calci we are in the process of creating an experiment to measure the selection strenght of sexual preferences of individuals for fishes with same color skin.

State of the art

Formalisms for System Biology: Mobile Ambients

Mobile Ambients is a formalism created to describe concurrent and mobile systems in a network of computers and later adapted to describe biology entities (BioAmbients).

$$P \quad ::= \quad (\nu n)P \mid \mathbf{0} \mid P \mid Q \mid\ !P \mid n[P] \mid M.P$$

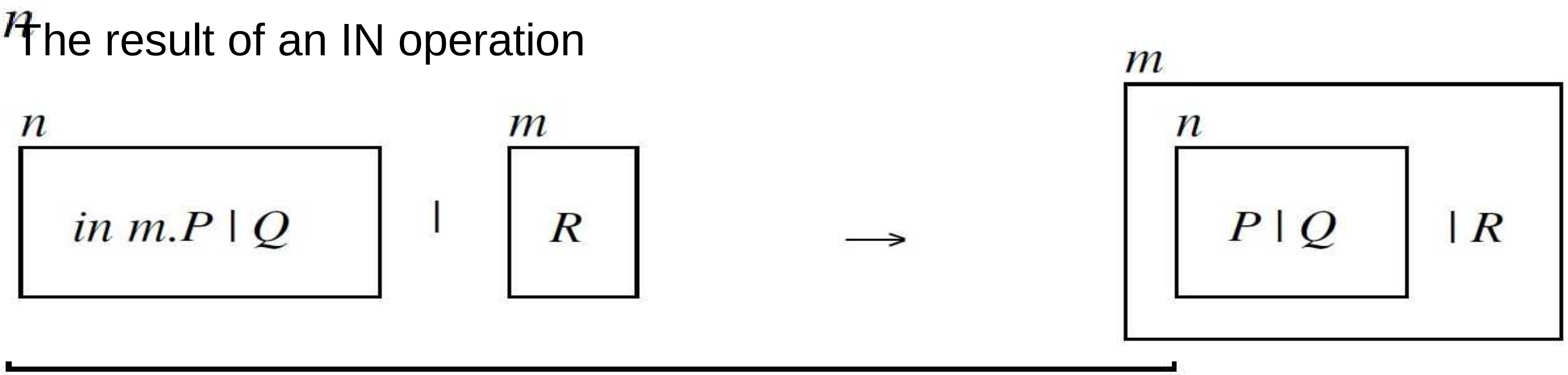
$$M \quad ::= \quad in\ n \mid out\ n \mid open\ n$$

Mobile Ambients

An ambient is :

- a bounded place where computation happens
- something that can be nested within other ambients
- something that can be moved as a whole

The result of an IN operation



BioAmbients

BioAmbients are an adaptation of the ambient formalism to depict biological systems.

Main differences:

- Compartments (ambients) are now nameless and are viewed as Membranes
- Communication channels are now either public or private
- Communication is not anymore only between peer ambients

BioAmbients

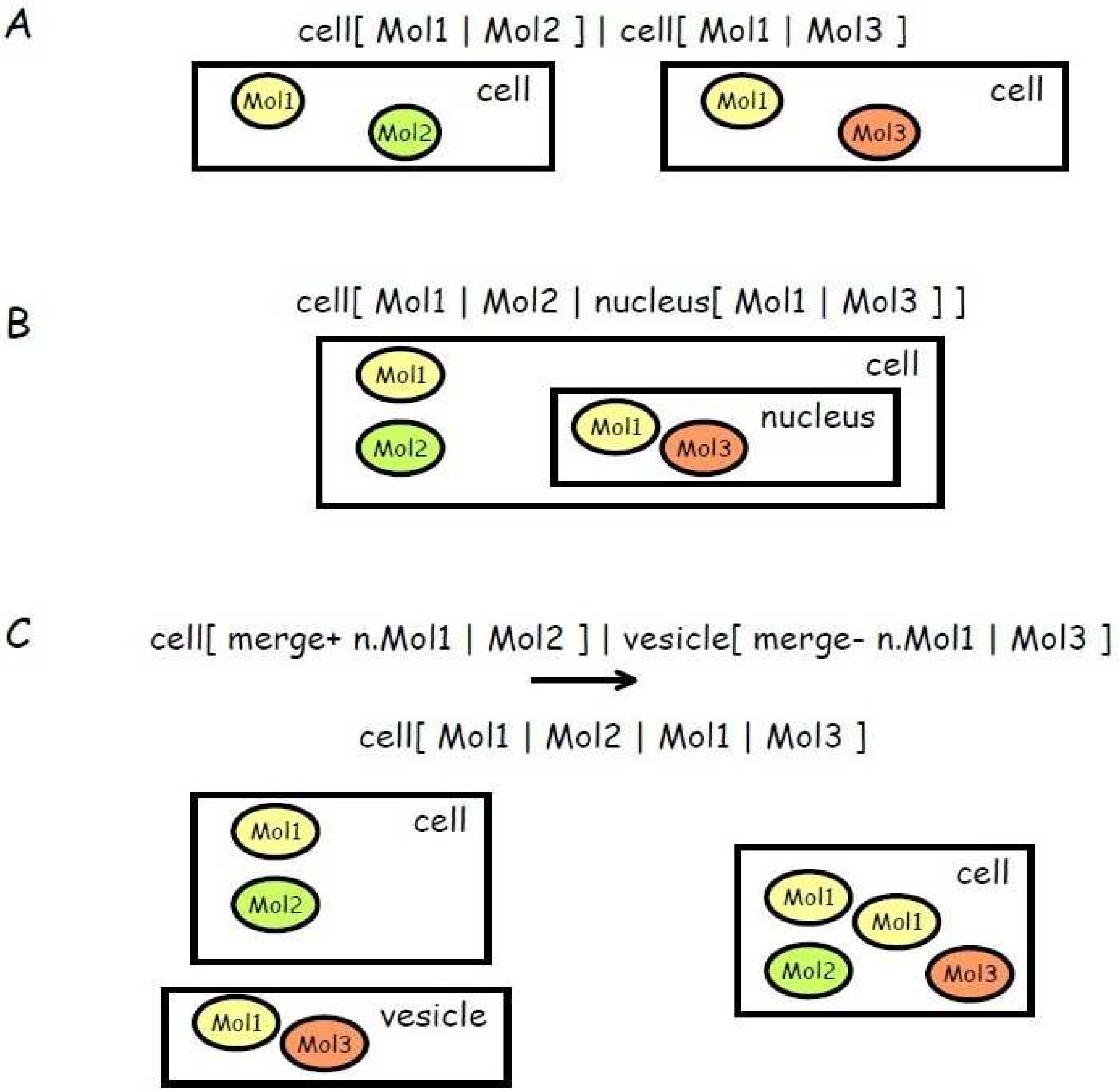
$$P \quad ::= \quad (\mathrm{new}\ n)P \mid P \mid Q \mid !P \mid [P] \mid \pi.P \mid M.P$$

$$M \quad ::= \quad \mathit{enter}\ n \mid \mathit{accept}\ n \mid \mathit{exit}\ n \mid \mathit{expel}\ n \mid \mathit{merge}{+}\ n \mid \mathit{merge}{-}\ n$$

$$\pi \quad ::= \quad \$n!\{m\} \mid \$n?\{m\}$$

$$\$ \quad ::= \mathit{local} \mid s2s \mid p2c \mid c2p$$

BioAmbients



A

cell[Mol1 | Mol2] | cell[Mol1 | Mol3]

B

cell[Mol1 | Mol2 | nucleus[Mol1 | Mol3]]

C

cell[merge+ n.Mol1 | Mol2] | vesicle[merge- n.Mol1 | Mol3]

$\longrightarrow$

cell[Mol1 | Mol2 | Mol1 | Mol3]

Calculus of Looping Sequences (CLS)

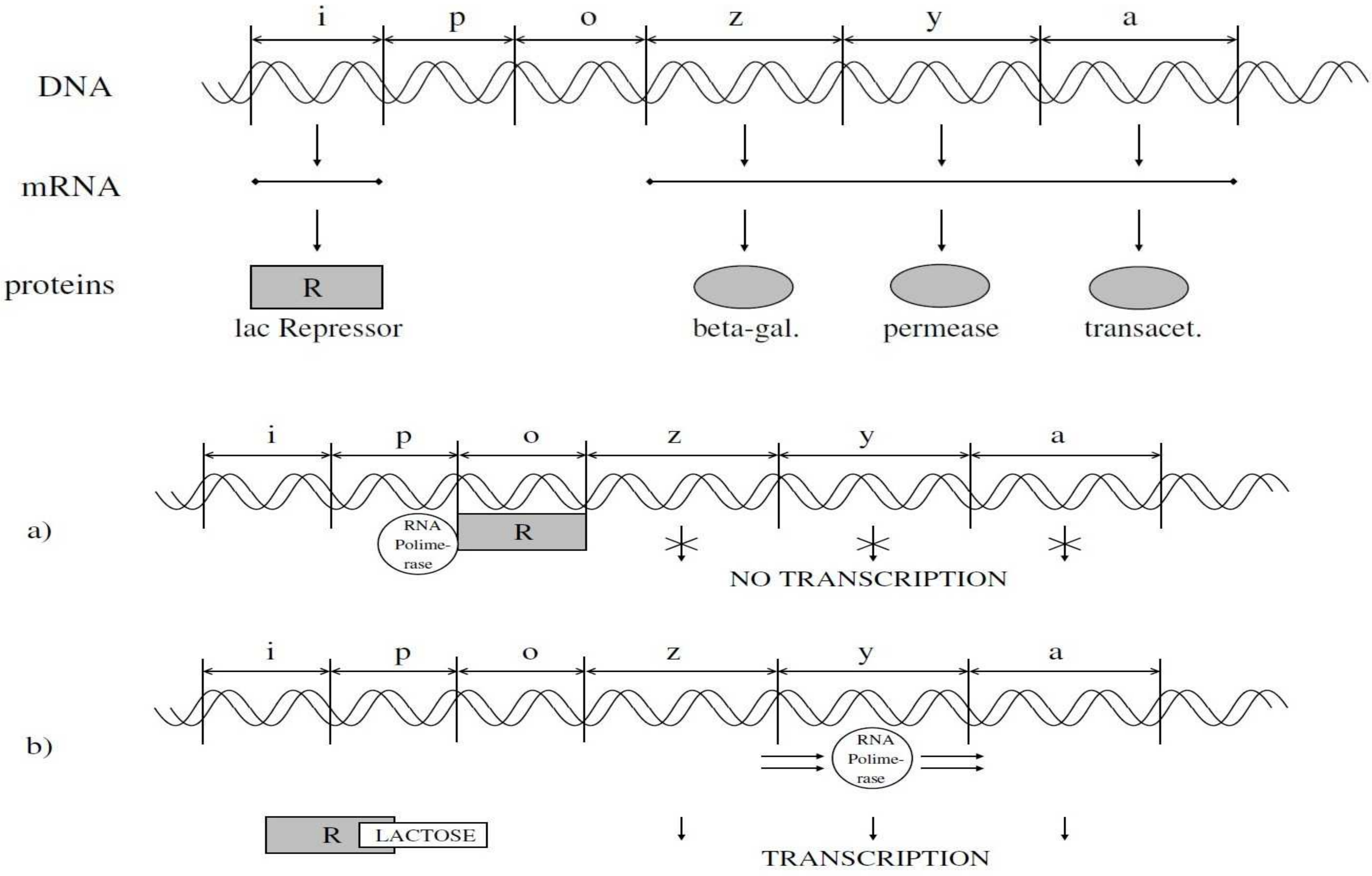
CLS are a formalism created to depict biological systems.

System dynamics is obtained through term rewriting rules.

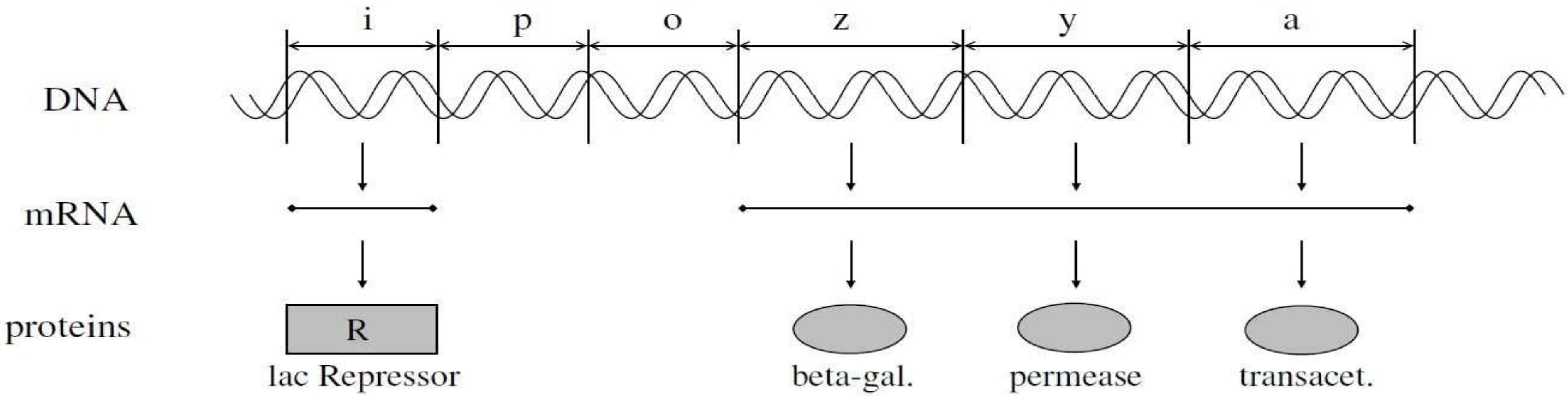
Each rule has a reaction rate associated and this rate is used to choose which rule apply to the term by the Gillespie algorithm (stochastic nondeterminism).

$$\begin{array}{lcl} T & ::= & S \quad \mid \quad (S)^{L} \rfloor T \quad \mid \quad T \mid T \\ S & ::= & \epsilon \quad \mid \quad a \quad \mid \quad S \cdot S \end{array}$$

CLS Example


i
p
o
z
y
a
DNA
mRNA
proteins
R
lac Repressor
beta-gal.
permease
transacet.
a)
RNA Polime-rase
R
NO TRANSCRIPTION
b)
RNA Polime-rase
R
LACTOSE
TRANSCRIPTION

CLS Example



$$Ecoli \ ::= \ (m)^{L} \rfloor (lacI \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym)$$

CLS Example

$$Ecoli \ ::= \ (m)^{L} \rfloor (lacI \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym)$$

$$\begin{aligned}
lacI \cdot \widetilde{x} &\mapsto lacI' \cdot \widetilde{x} \mid repr && \text{(R1)}\\
polym \mid \widetilde{x} \cdot lacP \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot PP \cdot \widetilde{y} && \text{(R2)}\\
\widetilde{x} \cdot PP \cdot lacO \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot lacP \cdot PO \cdot \widetilde{y} && \text{(R3)}\\
\widetilde{x} \cdot PO \cdot lacZ \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot lacO \cdot PZ \cdot \widetilde{y} && \text{(R4)}\\
\widetilde{x} \cdot PZ \cdot lacY \cdot \widetilde{y} &\mapsto \widetilde{x} \cdot lacZ \cdot PY \cdot \widetilde{y} \mid betagal && \text{(R5)}\\
\widetilde{x} \cdot PY \cdot lacA &\mapsto \widetilde{x} \cdot lacY \cdot PA \mid perm && \text{(R6)}\\
\widetilde{x} \cdot PA &\mapsto \widetilde{x} \cdot lacA \mid transac \mid polym && \text{(R7)}
\end{aligned}$$

CLS Example

$$Ecoli \ ::= \ (m)^L \rfloor (lacI \cdot lacP \cdot lacO \cdot lacZ \cdot lacY \cdot lacA \mid polym)$$

$$repr \mid \widetilde{x} \cdot lacO \cdot \widetilde{y} \ \mapsto \widetilde{x} \cdot RO \cdot \widetilde{y} \qquad \text{(R8)}$$

$$LACT \mid (m \cdot \widetilde{x})^L \rfloor X \ \mapsto (m \cdot \widetilde{x})^L \rfloor (X \mid LACT) \qquad \text{(R9)}$$

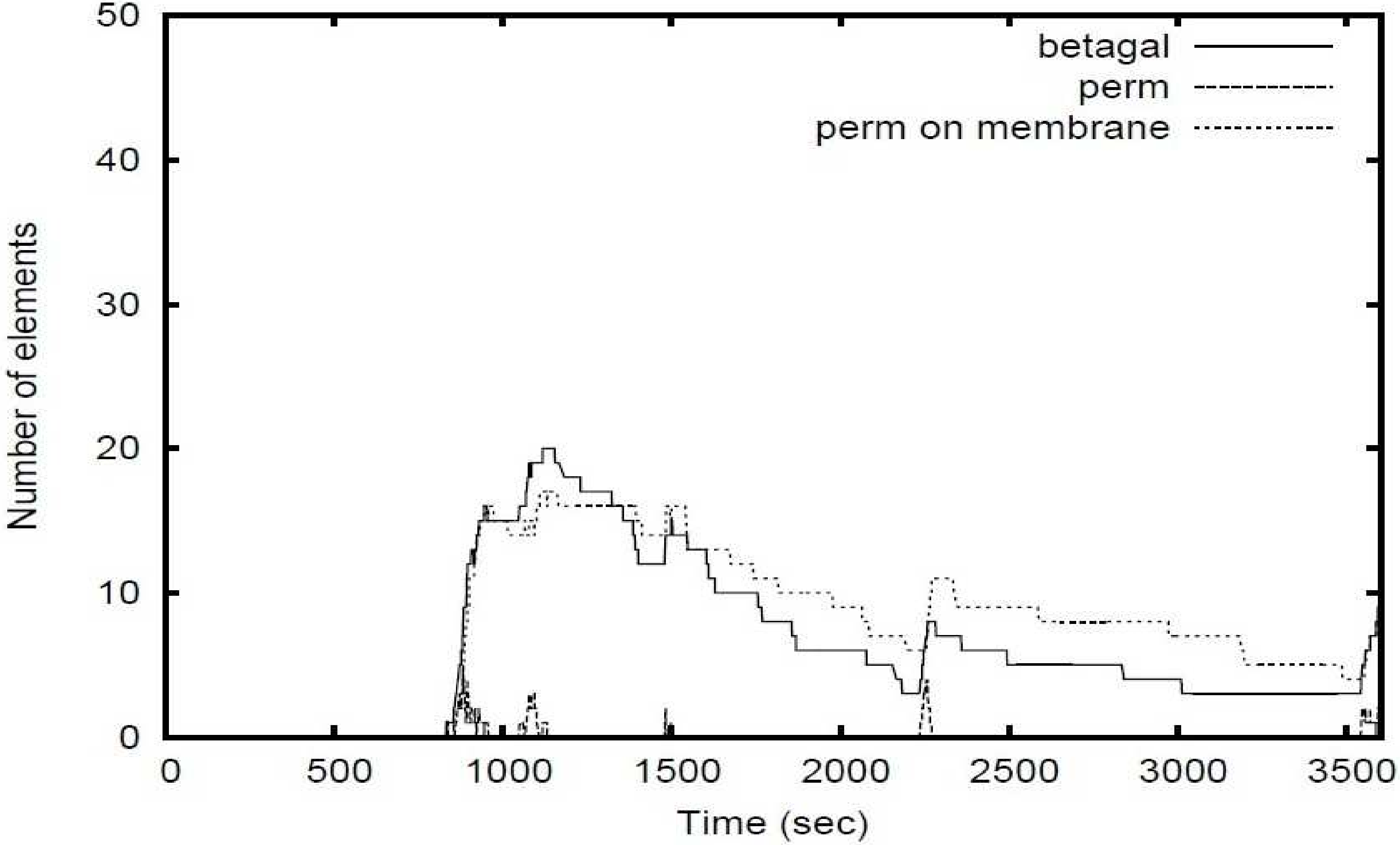
$$\widetilde{x} \cdot RO \cdot \widetilde{y} \mid LACT \ \mapsto \widetilde{x} \cdot lacO \cdot \widetilde{y} \mid RLACT \qquad \text{(R10)}$$

$$(\widetilde{x})^L \rfloor (perm \mid X) \ \mapsto (perm \cdot \widetilde{x})^L \rfloor X \qquad \text{(R11)}$$

$$LACT \mid (perm \cdot \widetilde{x})^L \rfloor X \ \mapsto (perm \cdot \widetilde{x})^L \rfloor (LACT \mid X) \qquad \text{(R12)}$$

$$betagal \mid LACT \ \mapsto betagal \mid GLU \mid GAL \qquad \text{(R13)}$$

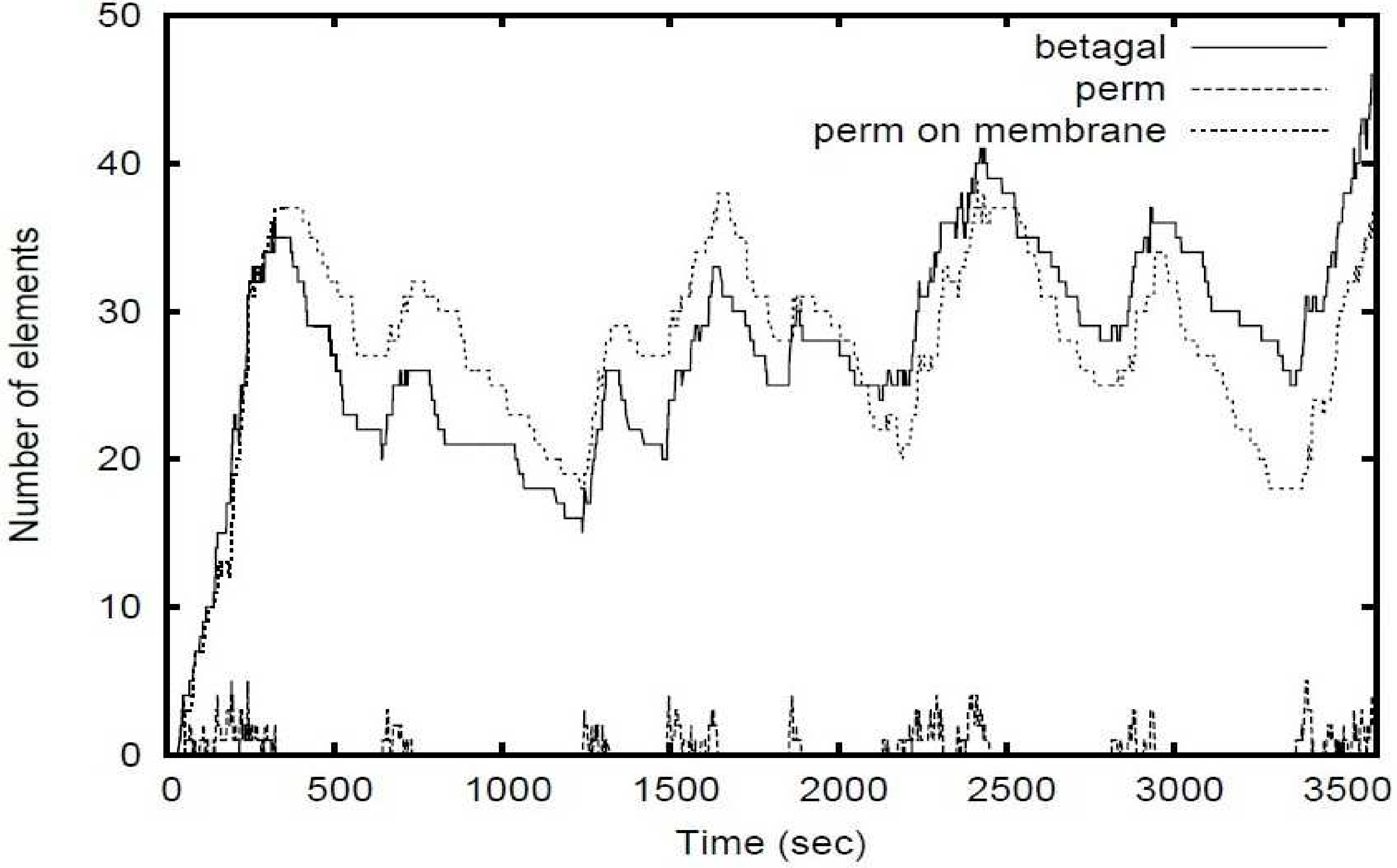
CLS Example Results



Production of enzymes in the absence of lactose

$$(m)^{L} \rfloor (lacI - A \mid 30 \times polym \mid 100 \times repr)$$

CLS Example Results



Production of enzymes in the presence of lactose

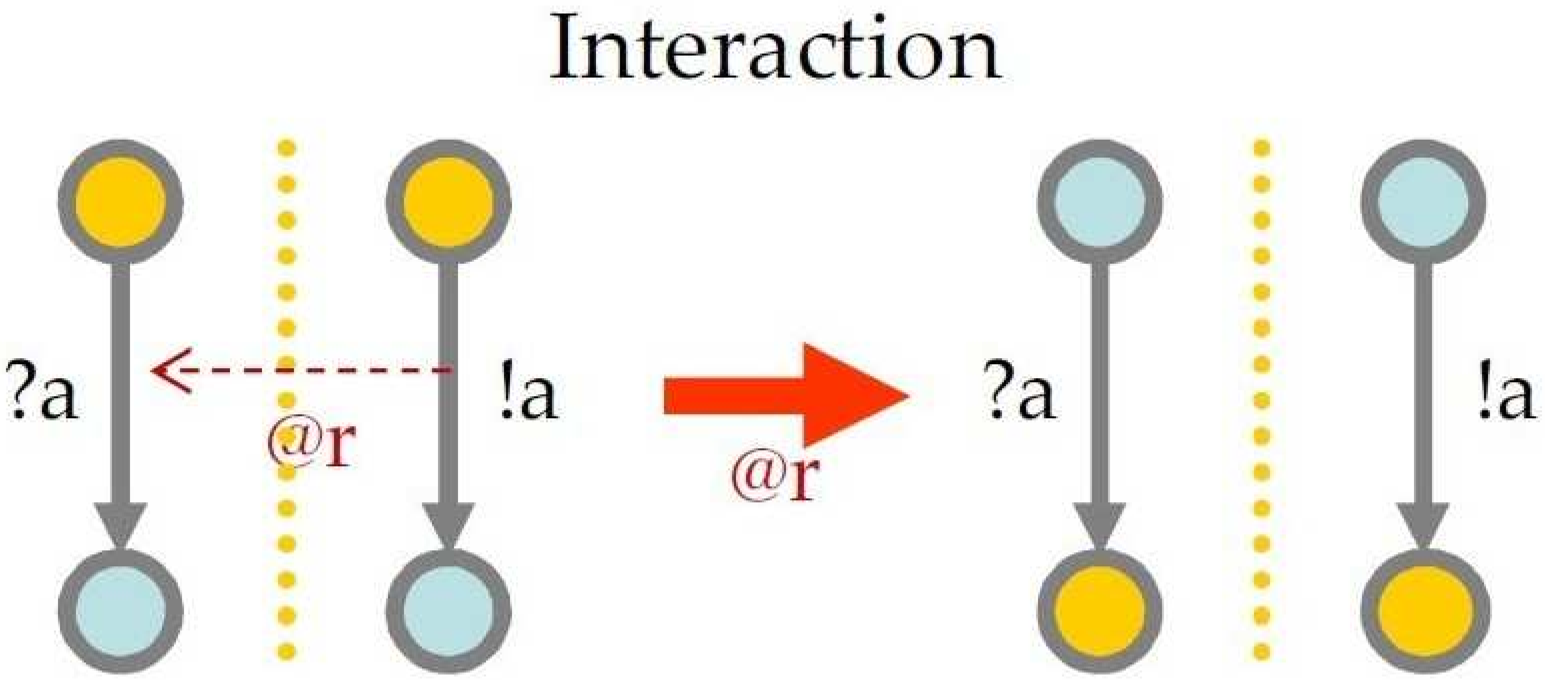
$100 \times LACT \mid (m)^{L} \rfloor (lacI - A \mid 30 \times polym \mid 100 \times repr)$

Artificial Biochemistry

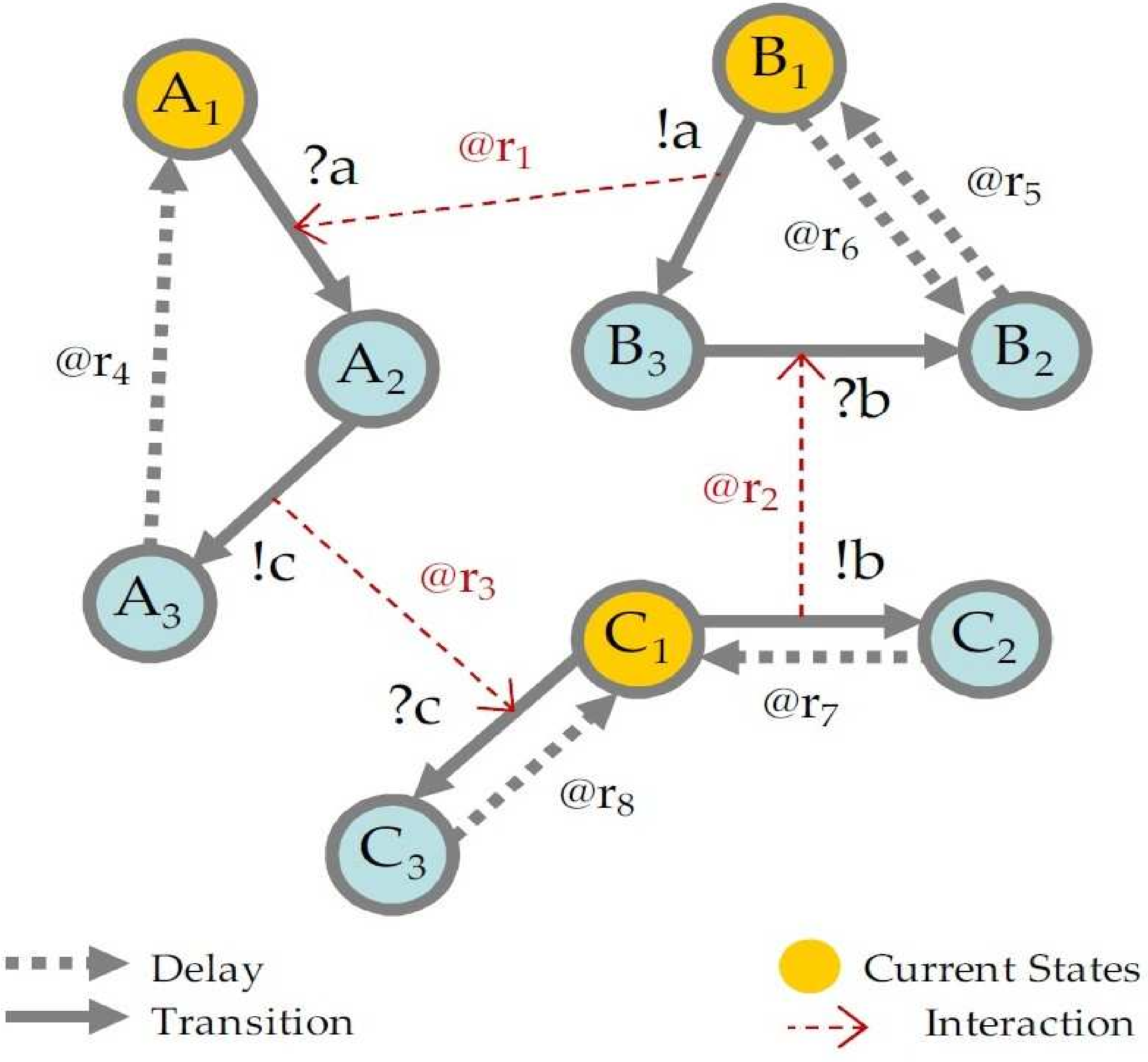
Molecules here are depicted as automata, with internal states and discrete transitions driven by external interactions.

This work is aimed at studying stochastic automata collective, a large number of *interactive*, *finite state* automata whose interactions (through communication channels) have *rates*.

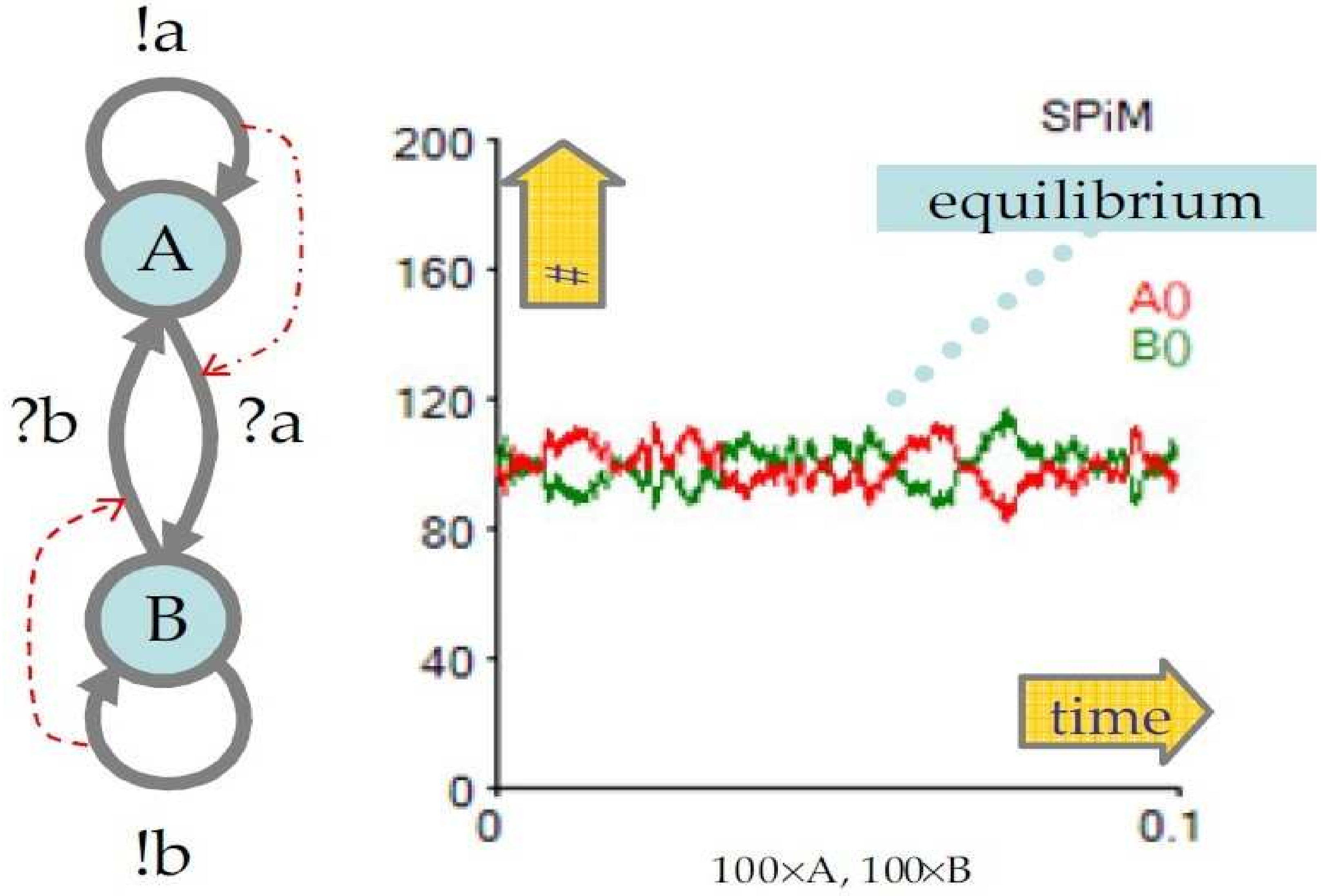
Artificial Biochemistry



Artificial Biochemistry

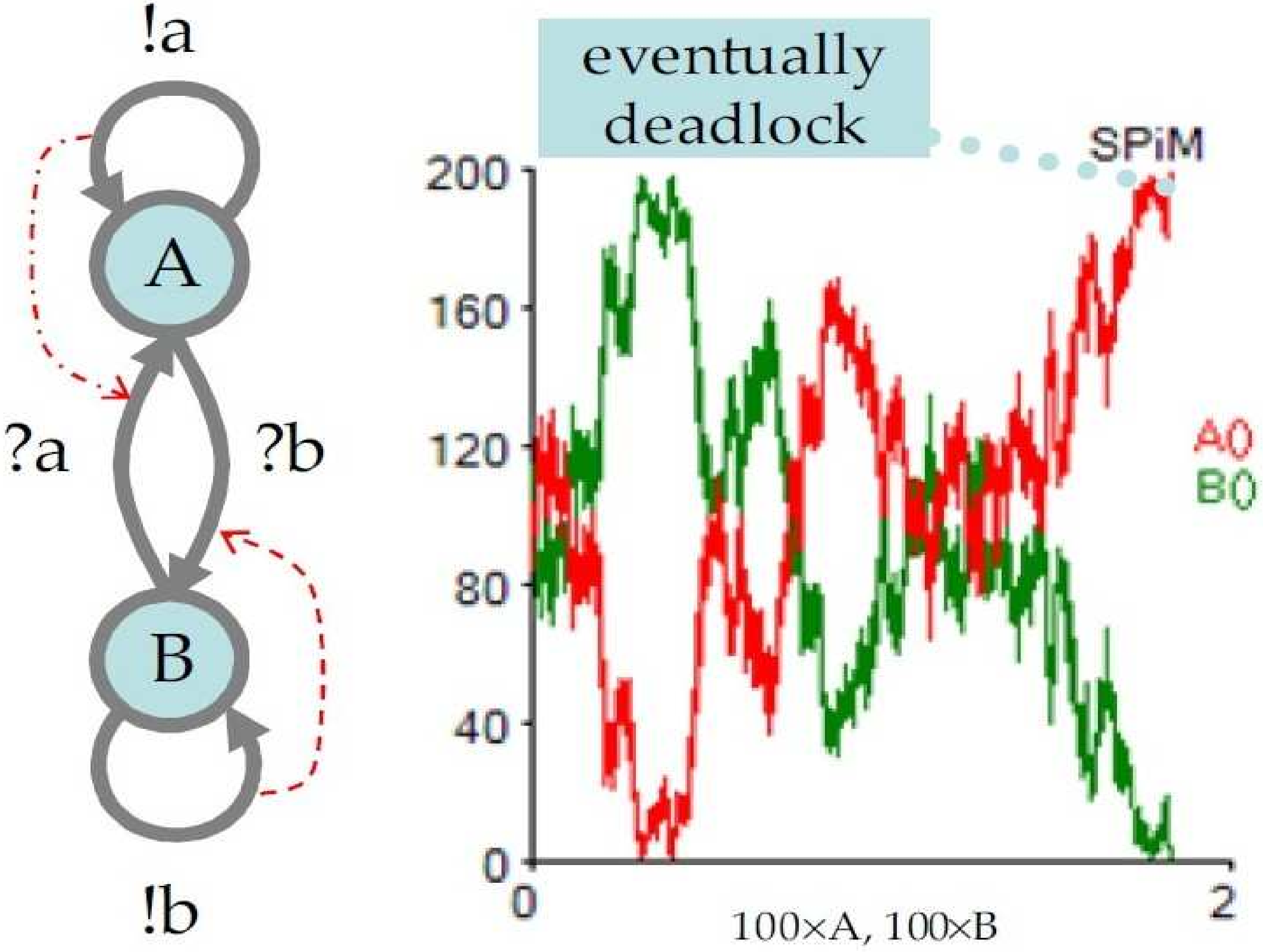


Artificial Biochemistry



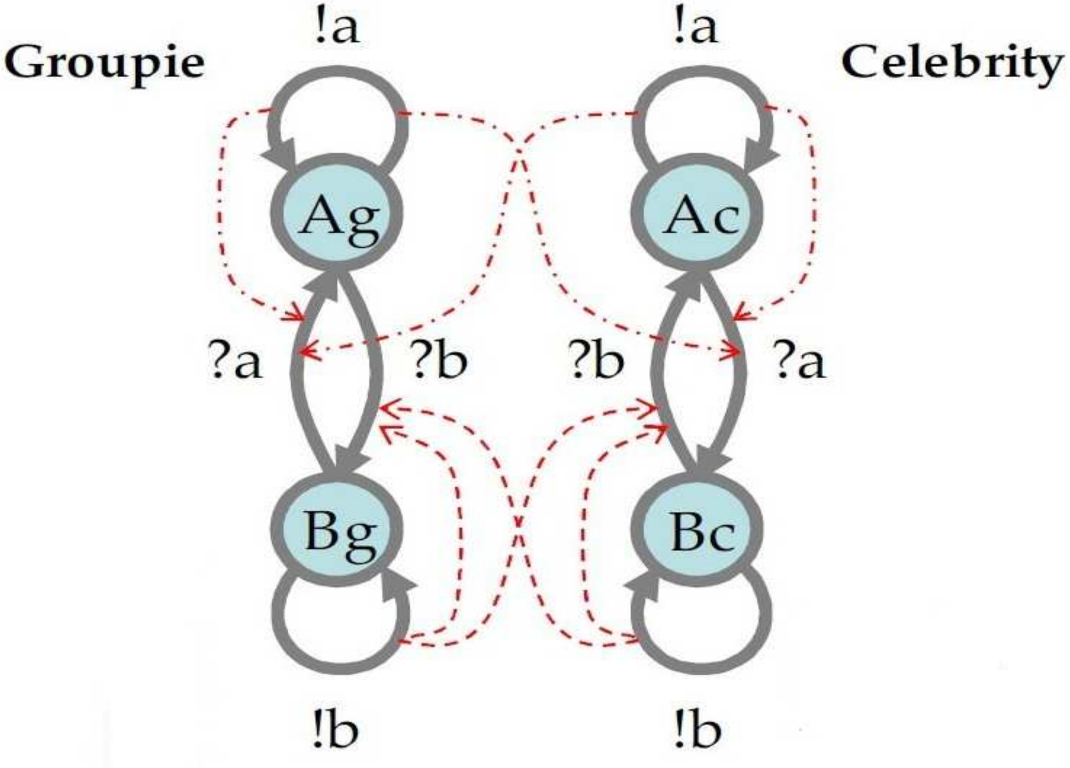
Celebrities : negative feedback

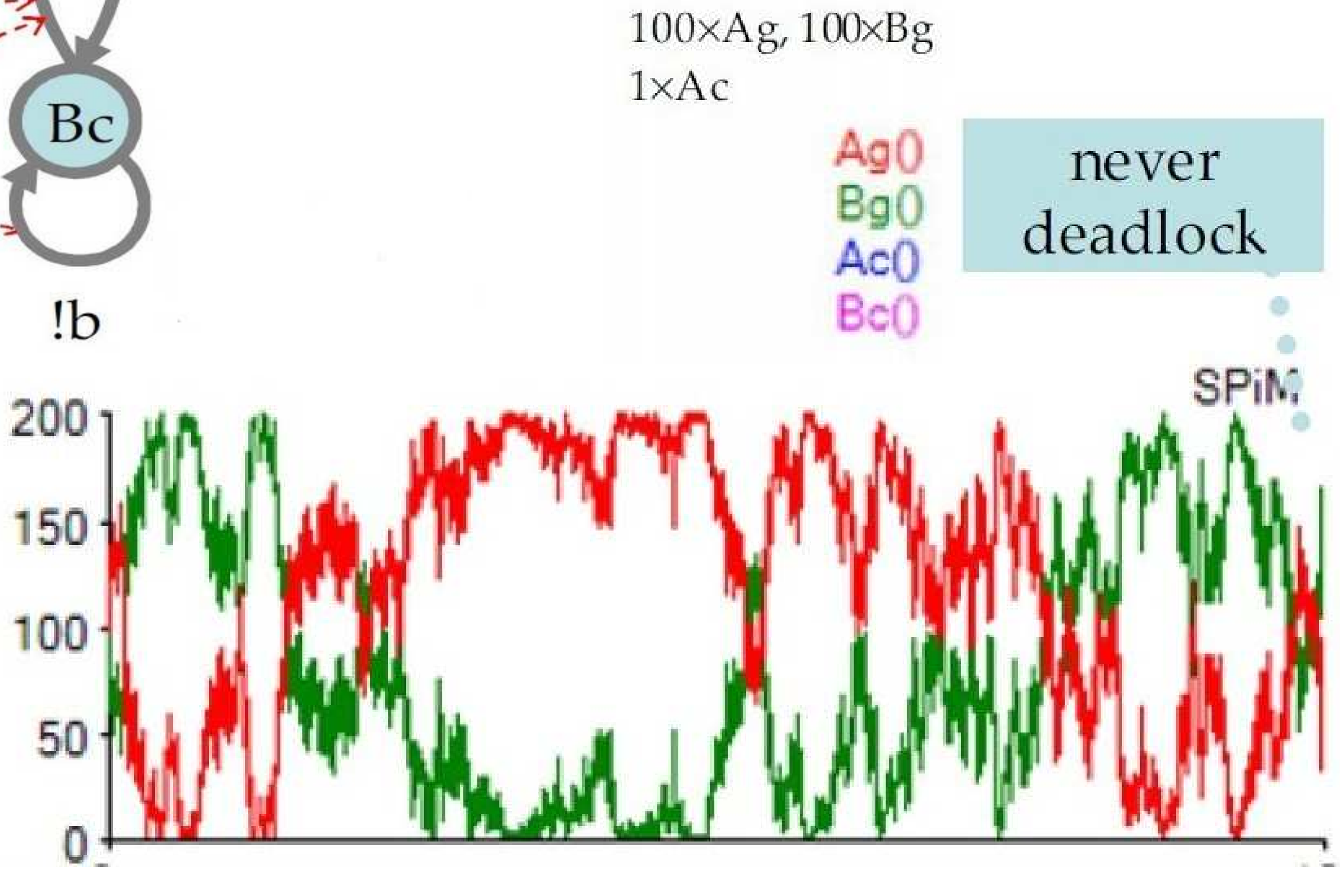
Artificial Biochemistry



Groupies : positive feedback

Artificial Biochemistry



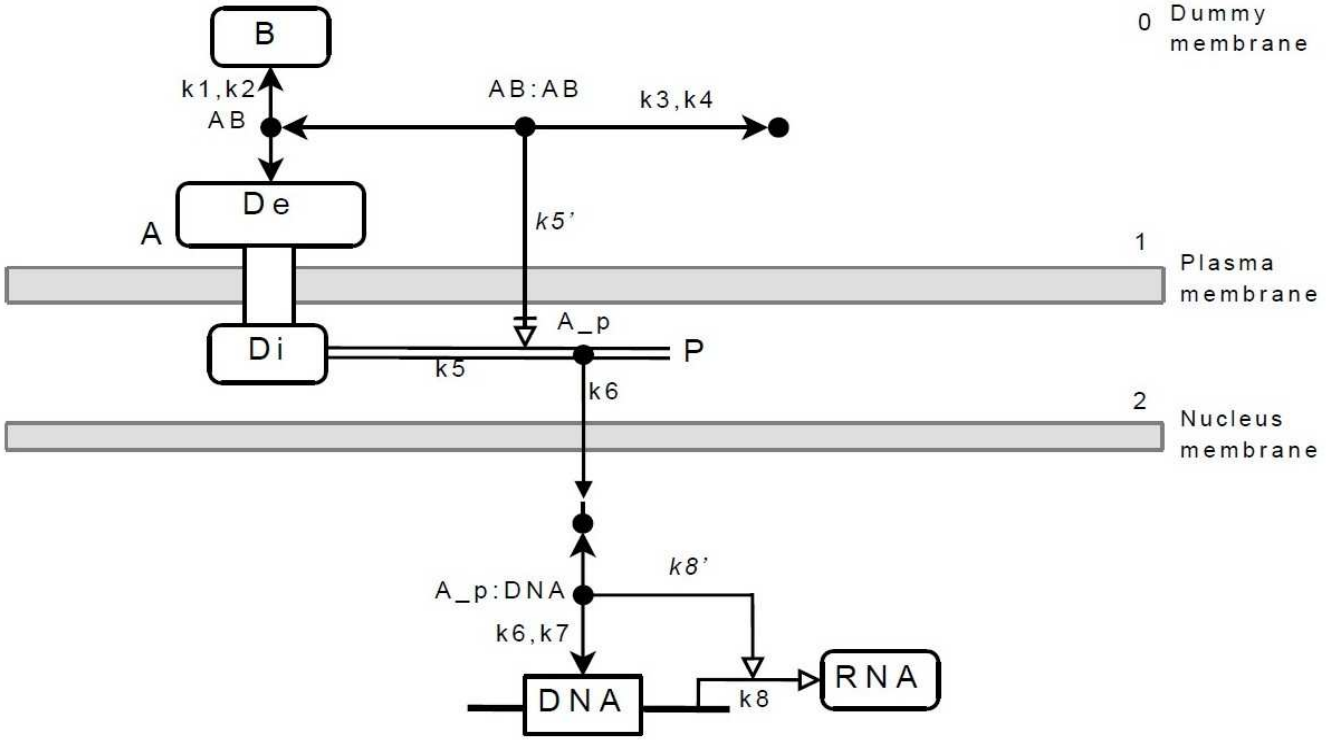


never deadlock

Celebrities +
Groupies:
oscillation

Molecular Interaction Maps (MIMs)

MIMs are a graphic notation for bioregulatory networks and depict possible interactions among species with different symbols to denote different reactions.



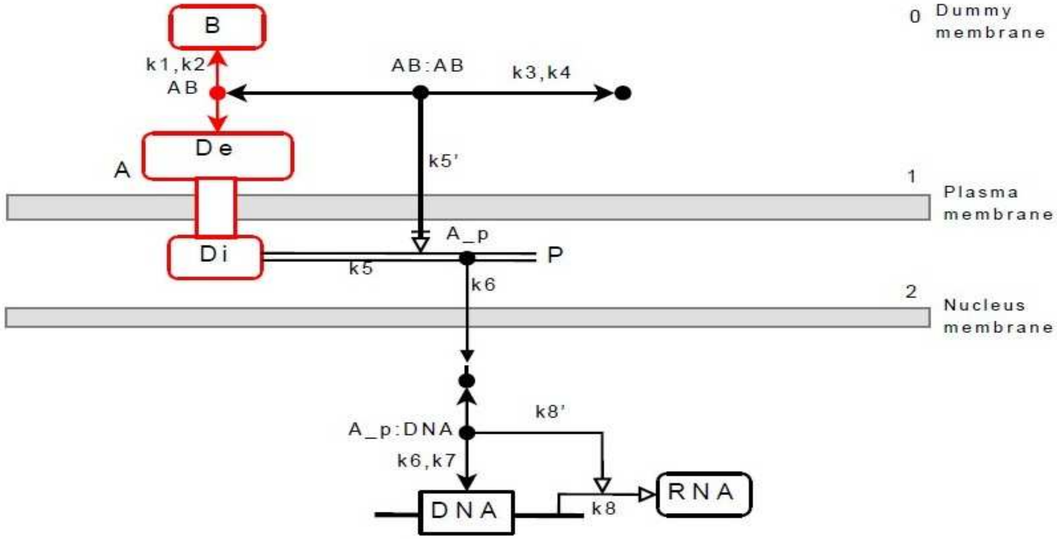
MIMs and CLS

- "*The Calculus of Molecular Interaction Maps*", Master Thesis, University of Pisa, 2008

- "*A Process Calculus for Molecular Interaction Maps*" in MeCBIC 2009 Proceedings, Electronic Proceedings in Theoretical Computer Science 11

- "*Simulation of Kohn's Molecular Interaction Maps Through Translation into Stochastic CLS*+" in PSI 2009 Proceedings, LNCS 5947

MIMs and CLS



$$(R_1)\quad (0\,|\,X_0)^L \rfloor \Big(B\,|\,\big((1\,|\,A\,|\,X_2)^L \rfloor X_3\big)\,|\,X_1\Big) \overset{f_1}{\longmapsto}$$

$$(0\,|\,X_0)^L \rfloor \Big(\big((1\,|\,AB\,|\,X_2)^L \rfloor X_3\big)\,|\,X_1\Big)$$

$$(R_2)\quad (0\,|\,X_0)^L \rfloor \Big(\big((1\,|\,AB\,|\,X_2)^L \rfloor X_3\big)\,|\,X_1\Big) \overset{f_2}{\longmapsto}$$

$$(0\,|\,X_0)^L \rfloor \Big(B\,|\,\big((1\,|\,A\,|\,X_2)^L \rfloor X_3\big)\,|\,X_1\Big)$$

Dynamic I/O Automata (DIOA)

DIOA model is a mathematical state-machine model for defining and analyzing dynamic systems of interacting components.

The systems considered are dynamic in two senses:

- components can be created and destroyed as computation proceeds, and
- the events in which the components may participate may change.

Timed P Automata (TPA)

We start with Paun's P System: a P system consists of a hierarchy of membranes, each of them containing a multiset of objects, representing molecules, a set of evolution rules, representing chemical reactions, and possibly other membranes.

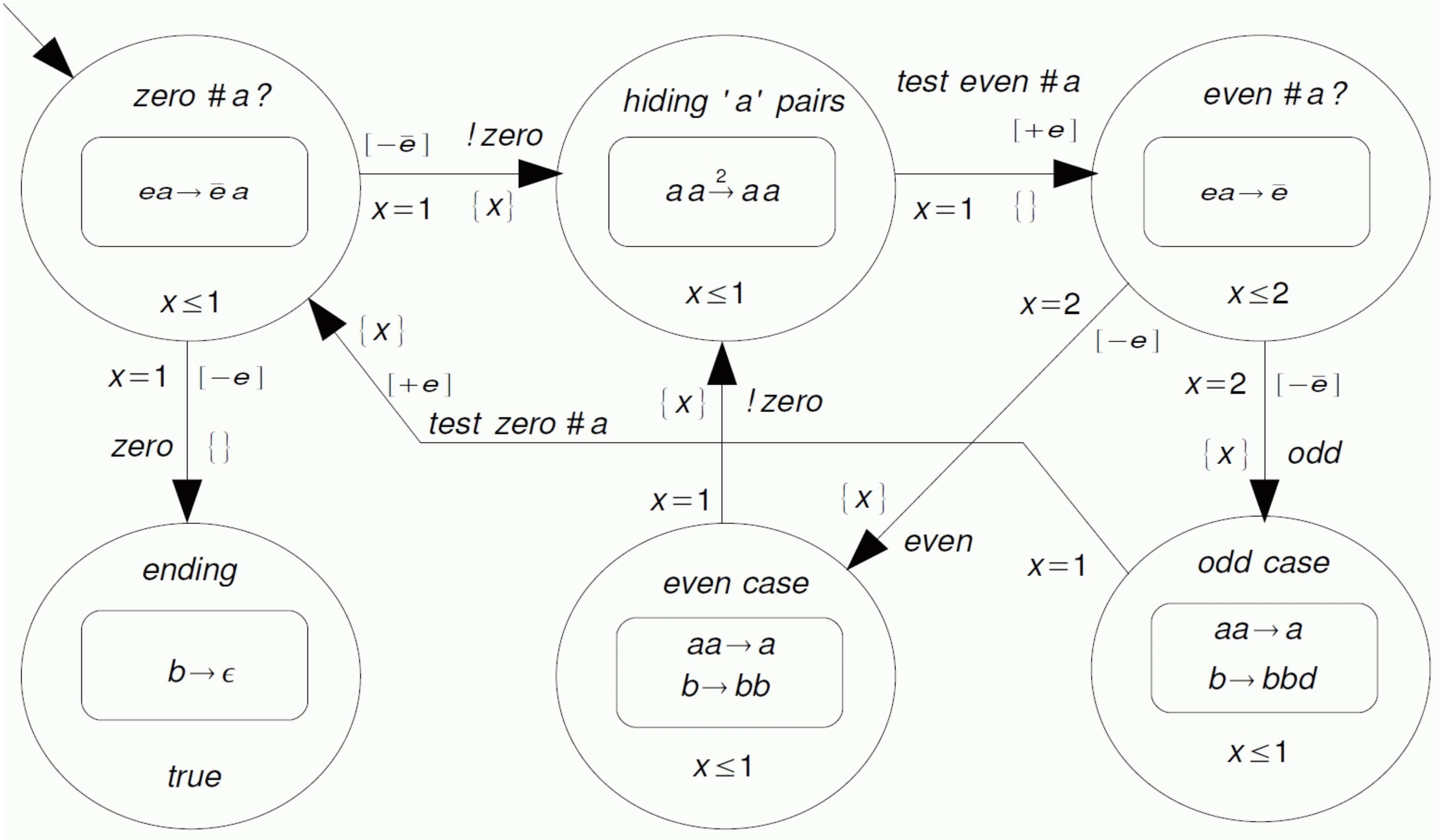
Then we define a timed P system as a P system in which evolution rules require a given number of time units to be completed.

Finally, a TPA consists of a timed automaton, with a discrete time domain, in which each state (location) is a TPS.

TPA

- Timed rules
  - Time (reagents get removed immediately and products get inserted after a certain delay)
  - Maximum parallelism
- Global clocks
- Timed transitions
  - Timed constraints
  - Clock resets
  - Deletions / Insertions of objects

TPA: Egypt Multiplication



TPA: Saddleback

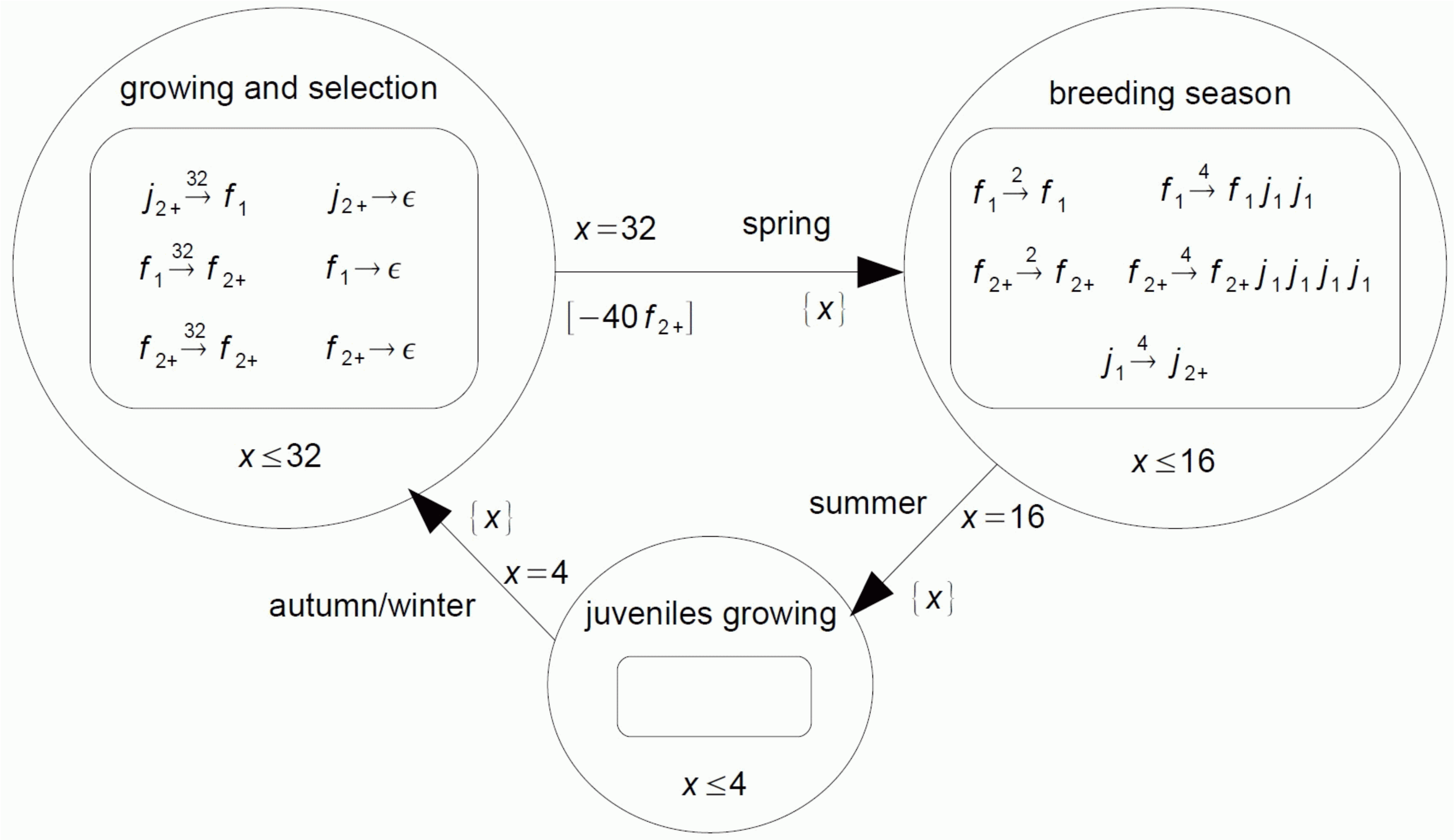
We have found a model for guiding the reintroduction of extirpated birds in New Zealand mainland.
The model is derived from the observation of the population of Saddleback birds (Philesturnus rufusater) on Mokoia Island

TPA: Saddleback

The model we have found:

- is a stochastic, discrete-time female-only model (the female-only
- approach assumes that there are sufficient males for all females to be paired)
- females are partitioned in two classes (first-year and older) with different fecundity rates (#edgings/season)
- an annual harvest of females is scheduled, with harvesting taking place at the start of breeding season.

TPA: Saddleback



Results are consistent with empirical observations.

Communicating Hierarchical State Machines (CHM)

In a hierarchical (nested) FSM, the states can be ordinary states or superstates, which are FSM themselves.

Superstates offer a convenient structuring mechanism

- stepwise refinement
- different granularity views
- sharing and reusing of component FSMs

CHMs are communicating hierarchical FSMs, that is concurrent FSMs that can synchronize on transition labels with a common alphabet and whose states can be other machines.

CHM

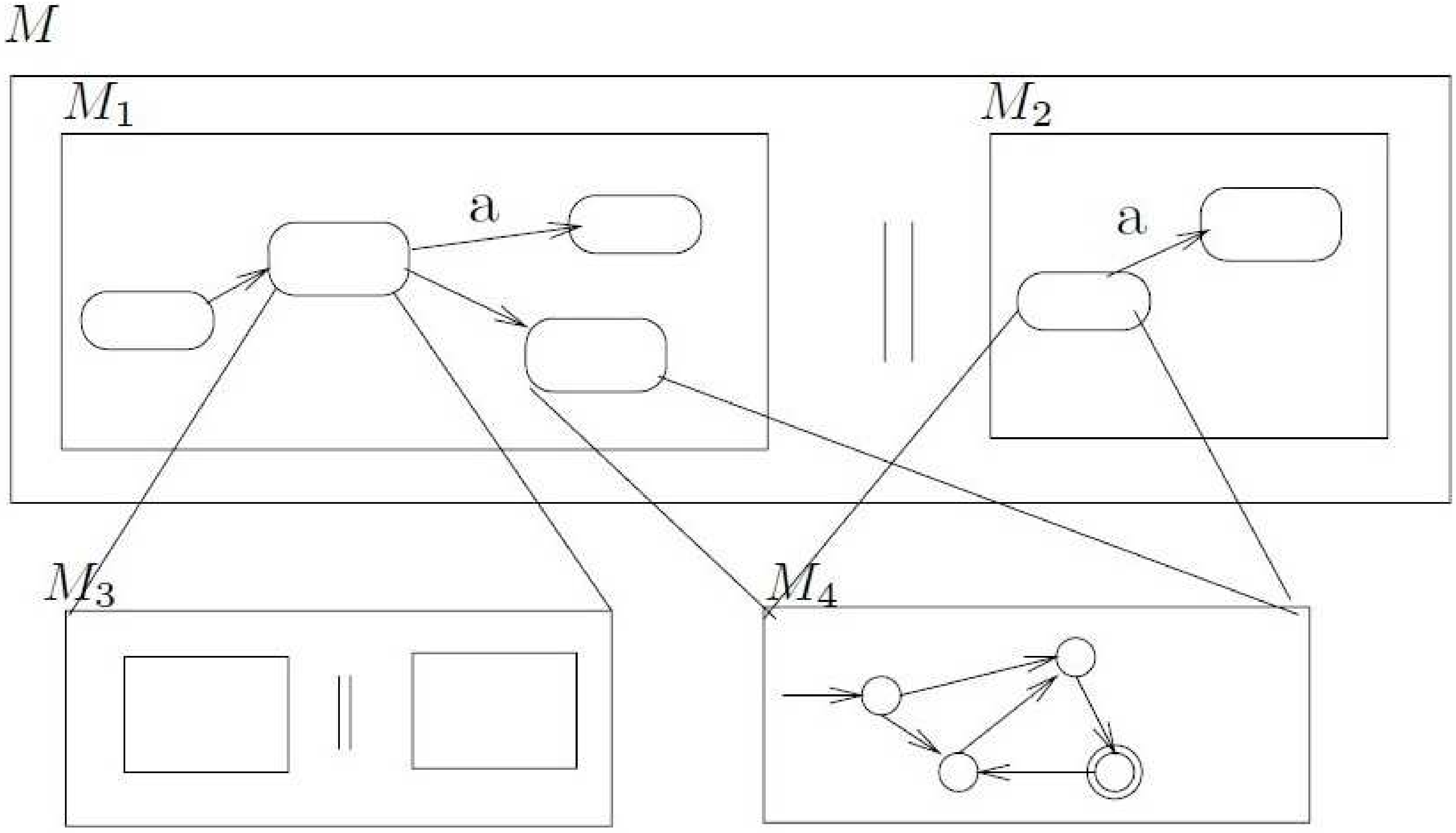
A CHM is one of the following:

- an ordinary FSM
- two CHMs in parallel $M_1 || M_2$
- an FSM with a labelling function μ that associates each state with a CHM from a finite set

In $M_1 || M_2$ both machines must accept the same symbol for their parallel composition to accept it.

In a hierarchical CHM, the external machine has a transition that link the underlying machines' final states and initial states.

A possible CHM



Hierarchy, parallel composition and component reuse

Dynamic Hierarchical Machines

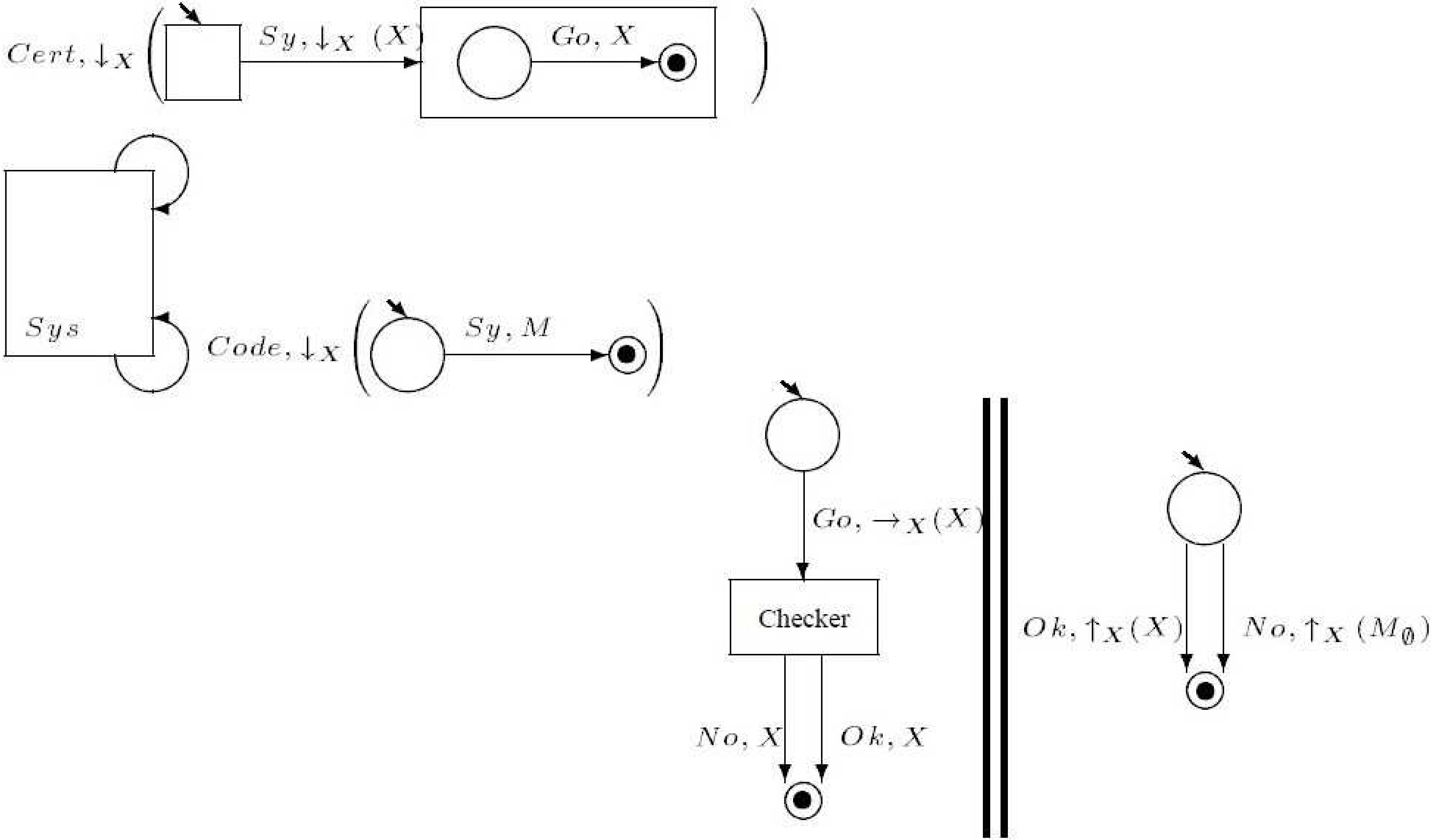
A DHM is a CHM whose transitions are decorated with the usual alphabetic trigger and

- either a ground DHM
- or a DHM along with a parameter name

In the first case firing the transition sends the ground DHM;

In the second case, the transition receives a DHM and substitutes it in the DHM indicated as an actual parameter in place of the formal parameter specified.

A possible DHM



This proposal

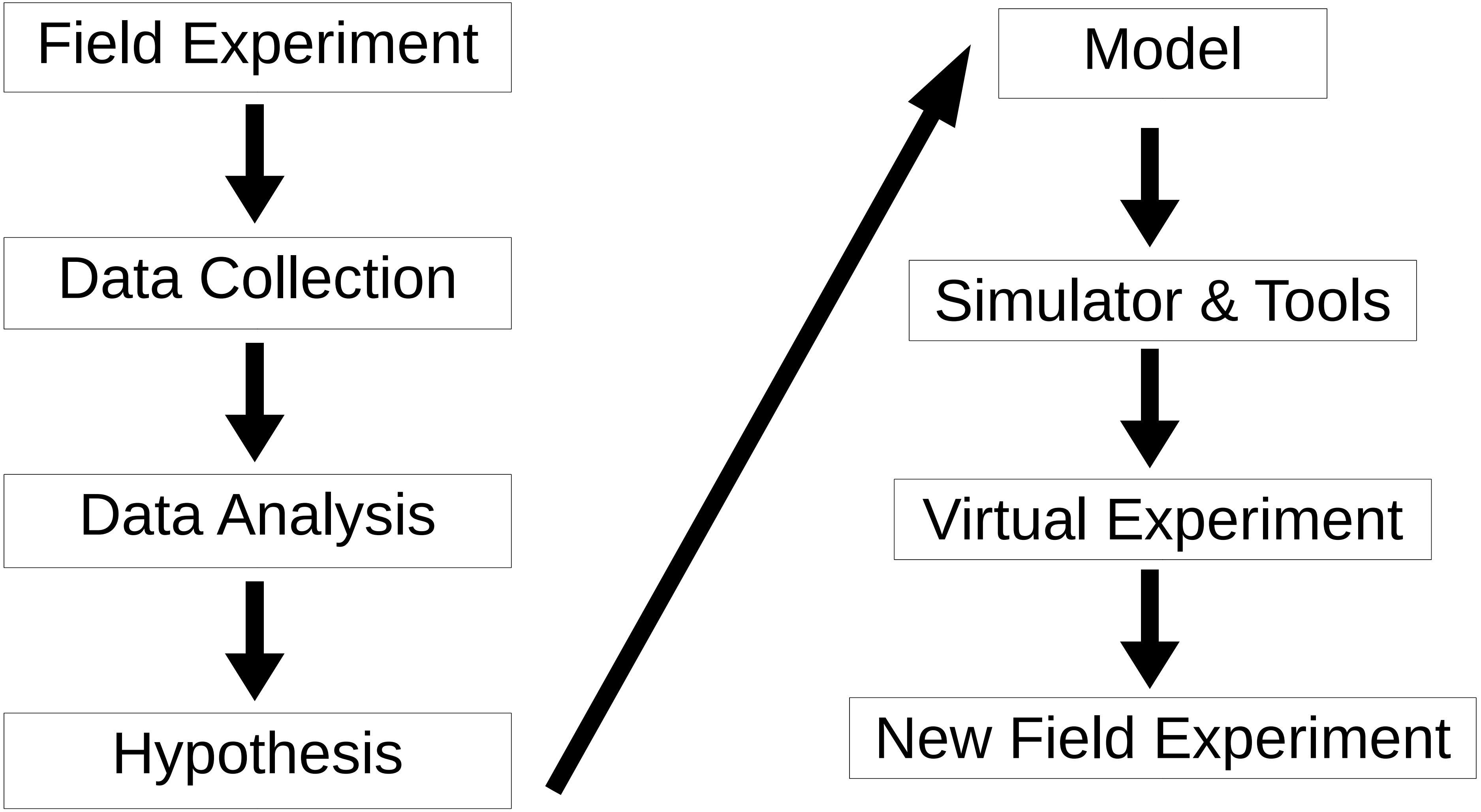
Goals

I aim to develop an easy to use formalism to describe large populations of individuals

- whose identity is unique
- who interact with each other
- who interact with the environment

On this formalism, we want to develop tools to do automatic analisys (i.e. model checkers) and to produce useful predictions (simulators) to be able to plan new field interventions (reintroduction or protection) or to predict where current ones are going.

Steps



Field Experiment

Data Collection

Data Analysis

Hypothesis

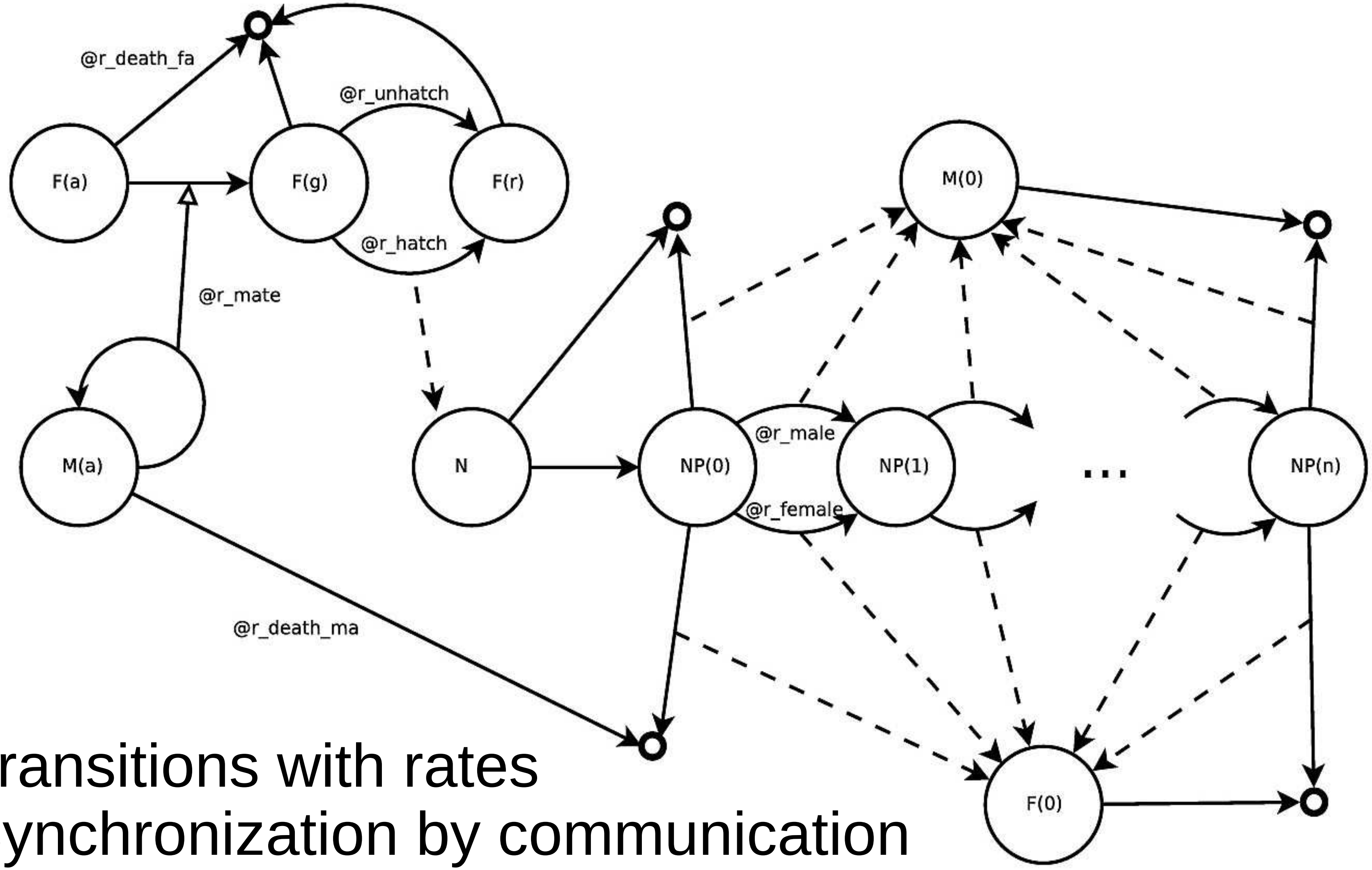
Model

Simulator & Tools

Virtual Experiment

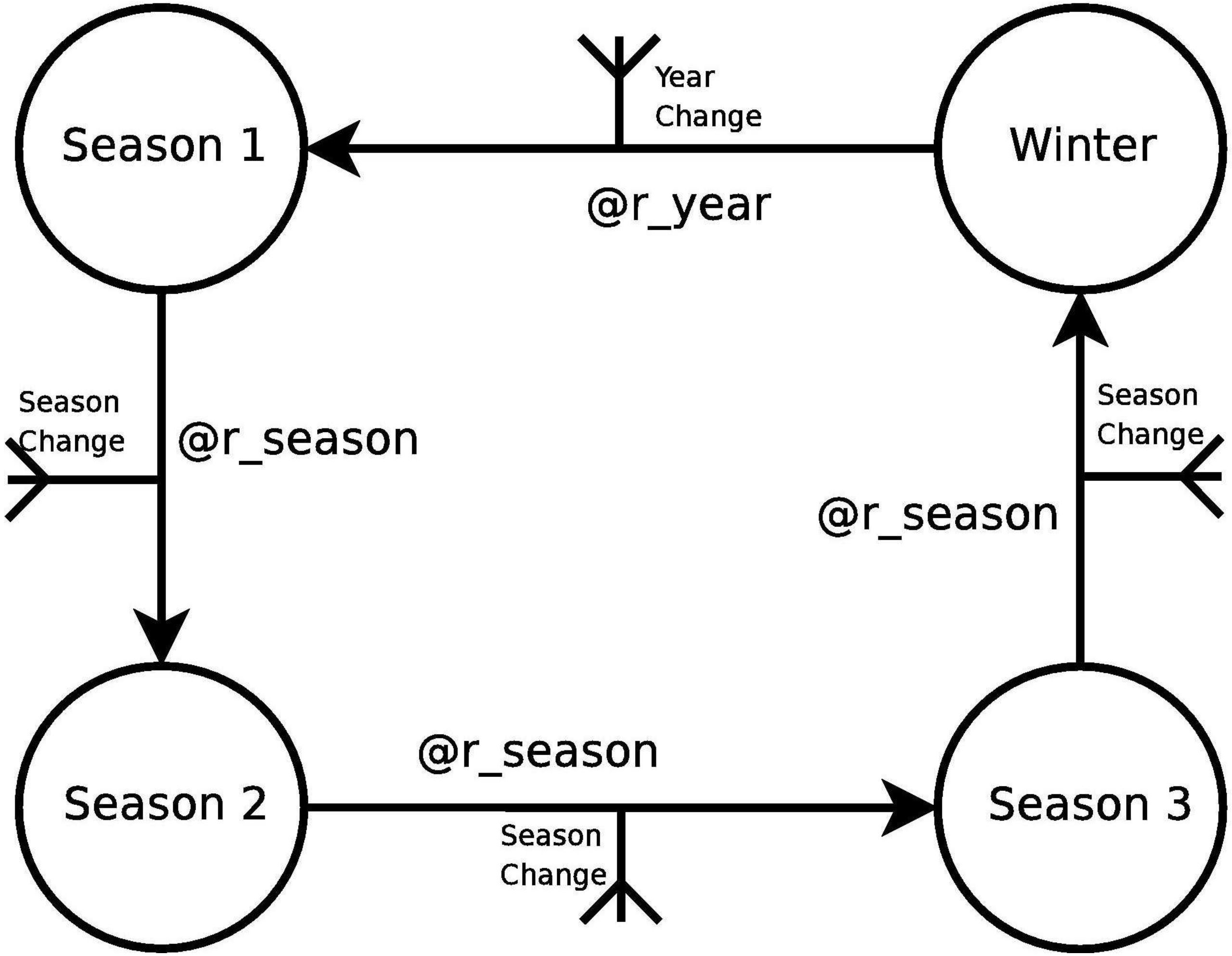
New Field Experiment

Emys FSM

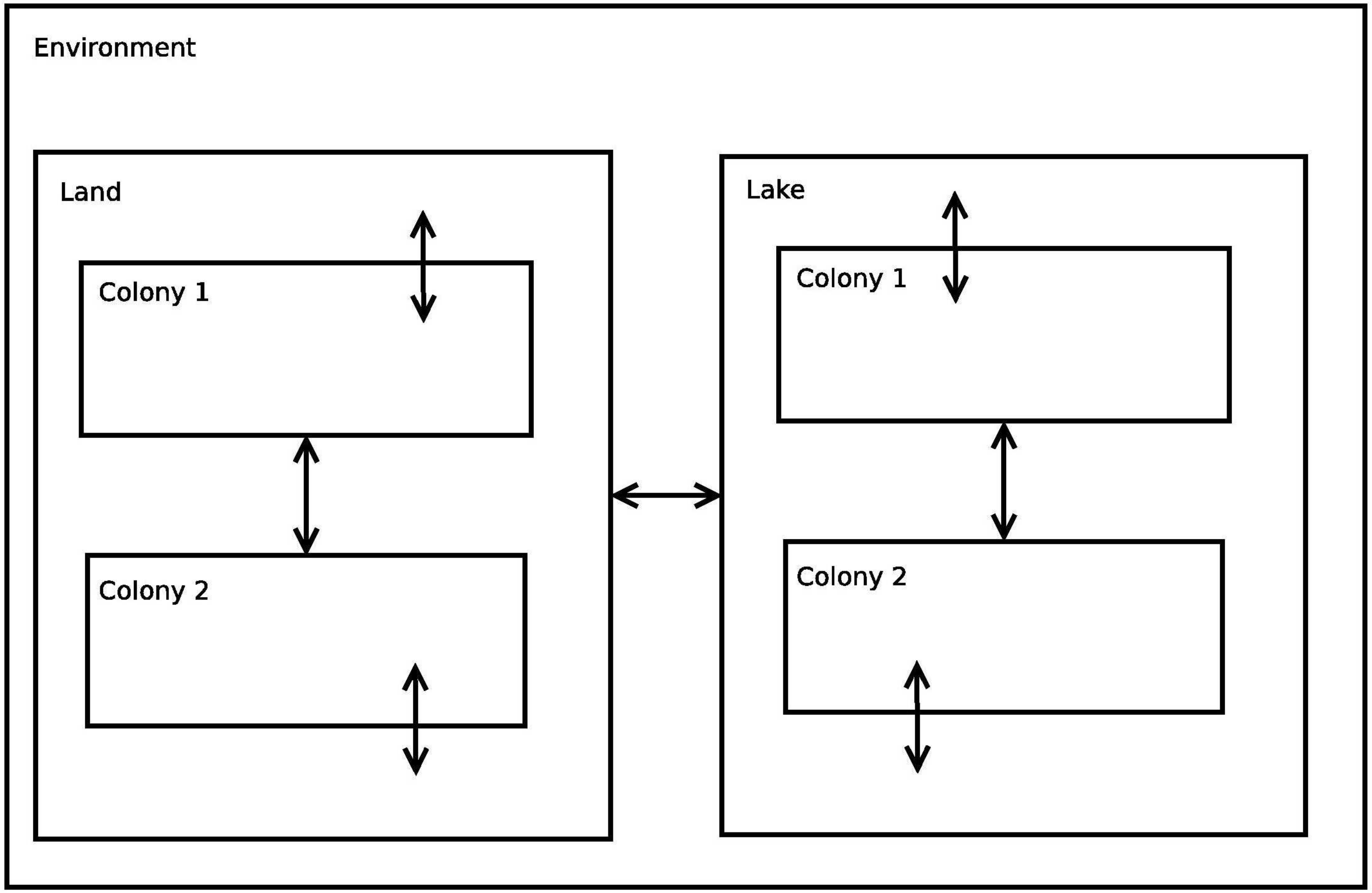


Transitions with rates
Synchronization by communication
Side effect: creation of new instances

Improvements: Broadcast Signals



Improvements: Automata Containers



Barriers for:
- synchronizations
- creations
- broadcast signals

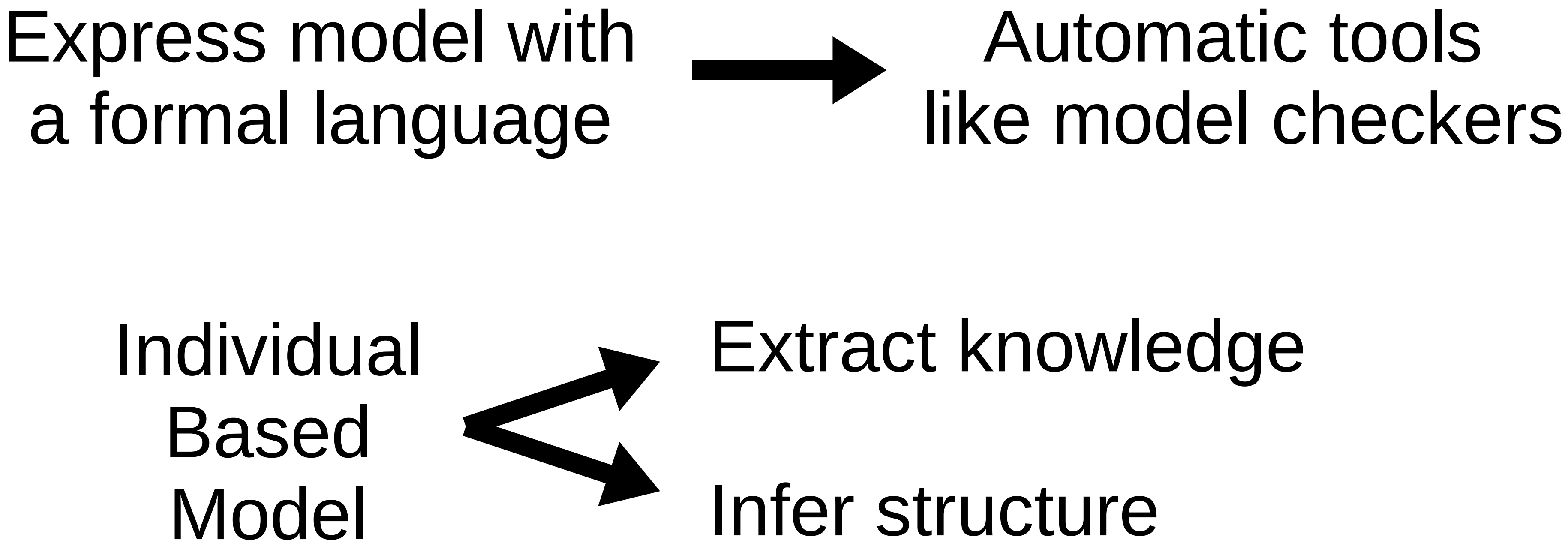
Impact

When considering modelling tools and simulators, many alternatives are available.

Yet, only a handful of natural scientists and biologists use these tools.

Working side by side with these researches I plan on crafting a mechanism that is adapted to their real needs while retaining formality and generality.

Impact



Express model with
a formal language

Automatic tools
like model checkers

Individual
Based
Model

Extract knowledge

Infer structure

Reach a synthesis between the “pragmatic” and
the “paradigmatic” aspects of simulations.

Work to do

- Identify principal characteristics of systems we want to model.
- Define a very simple ad-hoc formalism.
- Extend the formalism step by step, introducing probability (rates) and spatiality (to model populations distributed in an environment whose individuals can move through environments), broadcast signals, containers, etc...
- Develop analisys tools and techniques (simulators based on the formalism, translations into other languages to be able to use preexisting model checkers, etc...).
- Find other case studies.

Main References

Cardelli, Artificial biochemistry. 2009.

Grimm, Ten years of individual-based modelling in ecology. Ecological modelling, 1999.

Grimm et al., Pattern-Oriented Modeling of Agent-Based Complex Systems. Science, 2005.

Regev et al., Bioambients. Theor. Comput. Sci., 2004.

R. Alur at al., Communicating hierarchical state machines. Springer-Verlag, 1999.

R. Barbuti et al., Timed P Automata. Fundam. Inform., 2009.

Thank you.