Cellular automata and agent-based approaches for the modelling and simulation of biological systems: application to a simple genetic switch

Guillaume HUTZLER

IBiSc Lab. (Computer Science, Integrative Biology and Complex Systems)
FRE 3190 CNRS / Evry-Val d’Essonne University / Genopole
FRANCE

guillaume.hutzler@ibisc.univ-evry.fr

General context

- Aim = introducing space in the models
- Deterministic modelling → Stochastic modelling
  - Cellular automata
  - Agent-based modelling
- Desired properties
  - Modelling of local actions and interactions
  - Modelling of space
  - Discrete modelling
Why space does matter?

- Because some systems are the product of a morphogenetic process
  - in 2D (1D in space + t)

Why space does matter?

- Because some systems are the product of a morphogenetic process
  - in 3D (2D in space + t)
Why space does matter?

- Because some systems are the product of a morphogenetic process
  - in 3D

[R. Karlstrom & D. Dane, Development 123:461, 1996]
Why space does matter?

Because cells are crowded with molecules!

[ D. Goodsell ]

Why space does matter?

Because some phenomena occur in interconnected compartments
Why space does matter?

- Because the topology on which the reactions apply may matter

[ Mallavarapu et al. 2007 ]

Why space does matter?

- Because cells are polarized

[ M. Bornens ]
Why space does matter?

- Because some behaviours or structures are oriented by the environment
  - Chemotaxis
  - Wound healing

[ D. Bray ]

Why space does matter?

- Because the behaviour of cells is related to their dynamical mechanical structure (tensegrity)

[ D. Ingber ]
Why space does matter?

Because compartments are heterogeneous
- localized interactions vs. global computation can lead to qualitatively different dynamics
- “In conditions in which the continuum equations predict the population extinction, the individuals self-organize in spatio-temporally localized adaptive patches, which ensure their survival and development.” [Shnerb et al., PNAS 2000]

Why space does matter?

Because some phenomena are the result of a stigmergic process
- P.-P. Grassé « Stimulation of ant workers by the work already done »
- collective organisation principle thanks to the local modification of and perception of the environment of the individuals
Cellular Automata (CA): a first example

- **Conway’s Game of Life (70)**
  - the automaton is a regular grid of square cells that may be in two states: on and off
  - the evolution of a cell is determined by its own state and by the number of cells in the on state amongst the 8 adjacent cells $N_v$
    - a cell in the on state “dies”
      - if $N_v < 2$ (isolation)
      - or $N_v > 3$ (overcrowding)
    - a cell in the off state becomes “alive”
      - if $N_v = 3$ (reproduction)

Cellular Automata: a second example

- **1D automata**
  - the automaton is a one-dimensional array of cells that may be in two states: on and off
  - the evolution of a cell is determined by its own state and by the state of its two neighbours
  - e.g. rule 18
    - a cell in the on state “dies”
    - a cell in the off state becomes “alive” if it has exactly one neighbour in the on state
    - mimics a simple diffusion process [Meinhardt & Klingler 1987]
Cellular Automata: a general framework

- Introduced by J. von Neumann and S. Ulam as a way to study self-reproducing machines

- General properties
  - the CA space is a lattice of cells with a regular geometry
  - each cell is defined by its state, chosen from a limited range of values (e.g., 0 and 1)
  - the neighbourhood of the cells is defined in a uniform way as a finite set of indexes
  - time advances in discrete steps; all cells update synchronously, using the same updating rule or transition function (in uniform CA), depending only on local relations

- Applications for the modelling and simulation of complex systems in chemistry, physics, biology, ecology, economics, sociology

Cellular Automata: design choices

- Space dimension and lattice geometry
- Shape and size of the neighbourhood
- Boundary conditions
- Initial conditions
- State space
- Transition rules
A generic example:
The Greenberg-Hastings model (1978)

- A generic model of excitable medium
- Cells may be in three different states
  - idle (0)
  - excited (2)
  - refractory (1)
- Transition rules
  - if no neighbour in state
  - if at least one neighbour in state

The Greenberg-Hastings model
Space dimension and lattice geometry

- Periodic tiling of a $d$ dimensions space
  - the cells have to fill entirely the space
  - the lattice reproduces identically for translations in $d$ independent directions
- One-dimensional automaton
- Two-dimensional automata
  - Square lattice
  - Triangular lattice
  - Hexagonal lattice
- Three-dimensional automata
  - Square lattice
  - other geometries
**The Greenberg-Hastings model**

**Shape and size of the neighbourhood**

- **Neighbourhood**
  - set of cells with which this cell is able to interact
  - i.e. the set of cells whose state determine the transition of the cell, from one configuration to the next

- **Square lattices**
  - von Neumann neighbourhood
    - radius 1: four adjacent cells
    - radius $r$: $N_{ij} = \{ (k, l) \in L \mid |i-k| < r \land |j-l| < r \}$
  - Moore neighbourhood
    - radius 1: eight adjacent cells
    - radius $r$: $N_{ij} = \{ (k, l) \in L \mid |i-k| < r \land |j-l| < r \}$

**Boundary conditions**

- CA are theoretically infinite but practically finite
- **Three types of boundaries**
  - periodic
  - reflective
  - fixed-value
The Greenberg-Hastings model

Initial conditions

- Observation
  - a line of excited cells above a line of refractory cells

- Different initial conditions
  - 2% of excited cells
  - 1 excited cell next to a refractory cell
  - 2% of excited cells and 2% of refractory cells
  - random initialisation

The Greenberg-Hastings model

State space

- Compromise between
  - keeping the number of states low...
    - to be able to explore in a systematic way all possible automata
    - to be able to specify the transition rules explicitly and to store them in a transition table
  - ...and increasing the number of states
    - to have a better precision
    - extreme case = continuous variable → coupled map lattice

- Multiple variables cells
  - the state space for a single cell is the Cartesian product of the state spaces of each variable
### The Greenberg-Hastings model

**Transition rules**

- **Conditioned by**
  - the geometry of the lattice
  - the neighbourhood
  - the state space

- **Specified**
  - explicitly:
    - enumerating exhaustively all the possible configurations
    - giving in each case the new state for the cell
    - e.g.: \( (0, 2, 0, 0, 0) \rightarrow 2 \)
  - implicitly:
    - expressing the rules as formulas
    - give the new state of a cell as a function of
      - its own state
      - the states of its neighbours

- **Probabilistic rules**
  - probabilities associated to the different possible states

### Reaction-diffusion systems

- **Invented by A. Turing (early 50’s)**
  - proposed as a cause of various embryological patterns

- **Combination of**
  - a reaction in each cell between chemicals that behave as activators and inhibitors
  - the diffusion of the chemicals between neighbouring cells

\[
\frac{\partial a}{\partial t} = \rho a^3 - \mu a + D_a \frac{\partial^2 a}{\partial x^2} + \rho_a
\]

\[
\frac{\partial h}{\partial t} = \rho h^3 - \mu h + D_h \frac{\partial^2 h}{\partial x^2}
\]
CA modelling of a genetic switch

- "With these characteristics, cellular automata provide rather general discrete models for homogeneous systems with local interactions. They may be considered as idealizations of partial differential equations, in which time and space are assumed discrete, and dependent variables taken on a finite set of possible values." [S. Wolfram]

**Proposition**

- CA = spatialized explicit Euler Scheme
- make the correspondence between deterministic modelling and CA modelling
  - discretisation of space
  - discretisation of time
  - continuous variables

CA modelling of the λ phage

**Discretisation of space**

- **In deterministic modelling, all the chemical species**
  - are present in a single compartment
  - are able to react with each other without any restriction

- **In the CA model**
  - the global compartment is divided into a collection of smaller compartments (cells)
  - the state of the cells is composed of variables giving the concentration of the chemical species (X, Y)
  - reactions occur in each single cell
  - chemicals are exchanged between the cells according to:
    \[
    dx = D_x \left( \frac{1}{N(x)} \sum_{n=1}^{N} a_n - x \right)
    \]
    - with \( D_x \) the proportion of \( x \) shared with the neighbouring cells
    - \( N(x) \) the neighbourhood of the cell
CA modelling of the $\lambda$ phage

Discretisation of time

- Differential equations express small variations of the concentration of the chemical specie $x$ on a small time-interval
  - $dx/dt = f(x)$
  - $x(t+dt) = x(t) + dt.f(x)$
- The general deterministic model
  - $\frac{dX}{dt} = \frac{\alpha_1}{1 + Y} - X$ (1)
  - $\frac{dY}{dt} = \frac{\alpha_2}{1 + X} - Y$ (2)
- The corresponding CA model
  - $\Delta X = \frac{\alpha_1}{1 + Y} - X$ and $X(t+1) = X(t) + \Delta X$
  - $\Delta Y = \frac{\alpha_2}{1 + X} - Y$ and $Y(t+1) = Y(t) + \Delta Y$

Initial parameterization

- In the deterministic model
  - Two stable states using the following parameterization:
    - $\lambda = \beta_1 = 1$
    - $\alpha_1 = \alpha_2 = 5$
    - $dt = 0.01$

![Graphs showing initial parameterization](image)
Adding heterogeneity

- the initial concentration of X and Y is not homogeneous in the cell
- to have a mean initial concentration of $X_0$, we choose randomly the initial value of each cell of the automata in the interval $[0; 2 \times X_0]$

Adding diffusion

- if $D_X = 0$, each cell of the automaton is an individual system
- if $D_Y = 1$, everything is mixed up and homogenized very quickly
- we chose an intermediate value of 0.1
Different possible behaviours

Playing with parameters ($dt$)
What do the transitions look like?

Playing with topology

\[ \frac{\text{dx}}{\text{t}} = 1 \]

\[ \frac{\text{dx}}{\text{t}} = 10 \]

\[ \frac{\text{dx}}{\text{t}} = 1, \text{ topologie cylindrique} \]
From CA to Agent-Based Modelling

Problems with CA
- the activity of several molecules is abstracted as a small region with concentrations
  - not really a model of a genetic switch
  - rather a simplified model of animal skins
- the geometry of the lattice constrains the geometry of the structures that are produced
  - horizontal or vertical stripes
- the exploration of the parameter space is very empirical

Switch to agent-based modelling
- individual modelling of the entities of the system
- modelling of the interactions between the entities
- modelling of the environment in which the entities "live"

The "philosophy" of agent-based models

Modelling entities of the real system as entities in the simulation
- integrating models and theories peculiar to an application domain

Modelling at a chosen level of abstraction (or granularity)
- mixing several levels of description in the same model
- expliciting the inter-relations between the different levels

Modelling the behaviour of the entities, not only their result at the population level
- reproducing the emergence of spatial and temporal structures

Accounting for local heterogeneities in the system
- because of a variability between the individuals
- because of spatial variations

Designing simulations that are kinds of virtual laboratories
- conducting experiments just in the same way as in the real life
- testing hypotheses about the dynamics of the system
ABM: main design issues

- Choosing the entities of the model
- Choosing a model for the representation of space and time
- Describing the behaviour of the agents and their interactions
- Choosing a computational model

Choosing the entities of the model

- Deciding which of the entities of the real system will be represented as agents
  - close to the reification problem in OOP
  - closely linked to the choice of the level of abstraction used for the description of the system

- Criteria
  - depends on the aim of the simulation
    - what hypotheses do the modeller wishes to test and validate?
  - compromise between the expected level of detail or realism, and the computing resources
  - strongly conditioned by
    - the available data about the real system
    - the ontologies used by the experts of the domain
  - multi-level modelling
The entities in the λ phage model

- The molecules that participate in the reactions
  - CI
  - CI2
  - Cro
  - Cro2
  - pRNA
  - recA

- The lambda operator
  - OR1, OR2, OR3
  - the genes that code for CI and Cro

Representation of space and time

- The spatial environment
  - the geometrical structure of the system
  - the diffusion medium for the signals
  - the support for the movements of agents

- The representation of time
  - as a succession of time-steps (discrete time simulation)
  - as a succession of events (discrete events simulation)

- The scheduling of agents
  - activation in turn in fixed order
  - activation in turn in a random order
  - activation in parallel
  - other activation schema...
Space and time in the \( \lambda \) phage model

**Space**
- regular lattice of square cells
  - structure: phage DNA
  - signals: molecules have to be in the same cell to be able to interact with each other
- movements in a continuous space

**Time**
- discrete-time simulation
- scheduling handled by the simulation platform

The behaviour of the agents

**Can be abstracted as three successive steps**
- perception
  - the agent retrieves information in its environment
- decision
  - the agent chooses which action to undertake (amongst a set of possible actions)
  - depends on its internal motivation(s) and its external perception(s)
- action
  - the agent executes the chosen action, thus modifying its environment, the physical one or the other agents

**Very general behavioural model**
- enables to take into account such different entities as atoms or molecules, cells, organs, animals, human beings, enterprises, etc.
The behaviour of the $\lambda$ agents

- **In the context of molecular biology**
  - molecules may
    - participate in reactions with other molecules
    - get spontaneously transformed
  - stochasticity:
    - reaction = when two molecules are close enough, they have some probability to interact with each other

- **Molecular behaviours**
  - perception: identifying the molecules with which the agent may react
  - "decision": random choice depending on the reaction kinetics
  - action: depends on the type of reaction
    - if the reaction is a catalysis, and the agent is an enzyme, then the action would be to transform the substrate into a product
    - if the reaction is a dimerisation, then the action would be to bind to the other molecule, etc.

---

Pseudo-brownian movement

- **Principle**
  - choose a totally random direction
  - move by a distance of 1 unit
Dimerisation and degradation

Monomers form dimers (and back)

- $\text{CI} + \text{CI} \rightarrow K_{\text{dim}} \text{CI}_2$
- $\text{CI}_2 \rightarrow K_{\text{dis}} \text{CI} + \text{CI}$
- $\text{CI} \rightarrow K_{\text{deg}} *$

$K_{\text{dim}}$ interpreted as the probability to dimerise when two CI molecules meet
- “meeting” = presence in the same cell

$K_{\text{dis}}$ interpreted as the probability that a dimer spontaneously dissociates at a given time-step (similarly for $K_{\text{deg}}$)

dimerisation may be seen as
- the creation of a binding between the two molecules
- or ...
- the creation of a new dimer molecule and the destruction of the two monomers

DNA modelling

Principle

- modelling of each operator site as an individual agent
- DNA molecule considered as static
- only the operator agents check if the binding can be done
  - when binding CI, Or2 has to check if Or1 is also bound to CI
- no hypothesis about which reaction occurs first
  - all reactions are concurrent and may occur
  - some are more probable than others
- when a dimer is bound
  - it inhibits its moving and dissociation behaviours
  - the operator inhibits its binding behaviour
Promoters & RNA polymerase modelling

- **Principle**
  - promoters $Pr$ and $Prm$ are not modelled as separate agents
    - $Or1 = Pr$
    - $Or3 = Prm$
  - only $Pr$ and $Prm$ check if RNAp molecules are present
  - when RNAp is bound
    - its heading is changed to be colinear with DNA
    - RNAp moves forward until it reaches the "stop" codon
    - RNAp is released and creates $n$ protein molecules

The model with CI only
The model with both CI and Cro

Sample lysogenic evolutions (~80%)
Sample lytic evolutions (~20%)

Distribution lysogeny vs. lyse
Triggering the lambda switch (1)

- **Spontaneous evolution sometimes chaotic**
  - may switch spontaneously in some cases from lysogeny to lyse

- **External activation of the switch**
  - mutation provoked by irradiation with UVs (or by temperature)
  - a new molecular type, recA, degrades CI in monomer state

---

Triggering the lambda switch (2)

- reverse mutation
- reverse mutation
Immunity

- At $t=100000$, a new phage DNA enters the bacterium

Crossing the levels of abstraction (1/2)

- **Switch from the molecular level to the cellular level**
  - either by enlarging the model to include several bacteria
  - ... or by modelling cells instead of molecules

- **Abstraction of the cellular behaviour**
  - model cells
    - bacteria
    - $\lambda$ phages
  - define cells’ behaviour
    - define the probability of lysis and lysogeny as a function of the initial conditions and of the parameters
    - include a cellular division behaviour
    - include a lysis behaviour
Interest of agent-based simulation

- **Characteristics**
  - explicit modelling of space
  - explicit modelling of entities behaviour
  - no need to simplify excessively the model to be able to study the system (incremental approach)
  - modelling at different levels of abstraction

- **Interests**
  - allows to study the spatial and temporal emergent phenomena in the system
  - allows to study the effects of stochasticity
  - allows to design artificial experiments that allow to make predictions
Problems with agent-based simulation

- **Difficult to tune the model (Pareto law)**
  - very fast to have a model that is approximately coherent
  - can be very long and tedious to tune the model precisely

- **Difficult to characterize the different possible dynamics of the system depending on the parameters**
  - has to be done by hand

**New research directions**
- automatic exploration of the parameter space of a model
- automatic characterisation of the emergent dynamics of the system

Back to general context

- **Aim = introducing space in the models**

```
Deterministic modelling                                Stochastic modelling
                           ↓                              ↓
  Cellular automata                                  Agent-based modelling
                           ↓                              ↓
      Games Theory Networks modelling
                        ↓                              ↓
                           ???
```
About NetLogo

The NetLogo home page is at http://ccl.northwestern.edu/netlogo/ for contact information, or to report a bug, visit http://ccl.northwestern.edu/netlogo/contact.shtml

NetLogo author: Uri Wilensky
Lead developer: Seth Tisue
Development team: Andrei Scheinkman, Craig Braszek, Esther Venceau, Josh Underman, Matthew Berland

We also thank the following for their substantial contributions: Emman McKenzie, Geoff Hulette, Chuck Hubert, Ian Sharpe, Daman Coella, Matt Goto, Brent Collins, Matt Heilige, James Newell

We also gratefully acknowledge the contributions of Dor Abrahamsson, Fernando Algoz, Jason Al, John Baker, Ronald Bailey, Stephen Rasmor, Rasha Bilbstein, Ritas