

Master SVS – PARCOURS Bio-informatique et Biologie Computationnelle (BBC)

year 2024–2025



Formal methods for discrete modelling

Jean-Paul Comet

Université Côte d'Azur

24 September 2024

Teaching organization

- Lectures : 2 sessions of 2 hours each
- Tutorial : 2 sessions of 2 hours each
- Teacher : Jean-Paul Comet Jean-Paul.Comet@univ-cotedazur.fr

	Sessions	schedule	teacher	Lecture/tuto
1	24 September 2024	8h-12h	JPC	lecture+tuto
2	1 October 2022	8h-12h	JPC	lecture+tuto

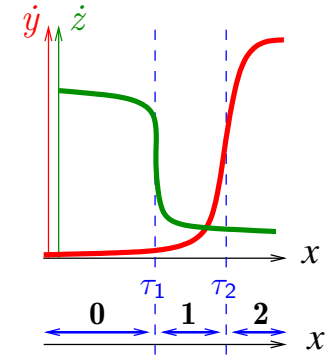
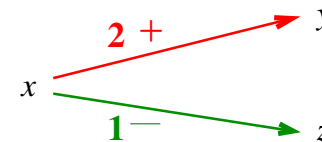
- Evaluation : A 2-hour exam on 8 October 2024 from 9 to 11 a.m.
- Course material + tutorials : <https://www.i3s.unice.fr/~comet/SUPPORTS/>

Plan

- 1 Discrete models for gene networks according to René Thomas
- 2 CTL
- 3 Techniques of software testing
- 4 General Schema for BRN
- 5 Genetically modified Hoare logic, and examples
 - Hoare Logic
 - Examples

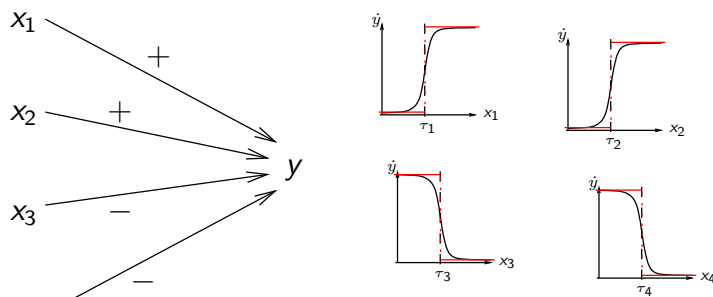
Multivalued Regulatory Graphs

Derivatives are sigmoids w.r.t. the source gene



First simplification : piecewise linear

Approximate sigmoids as step functions :



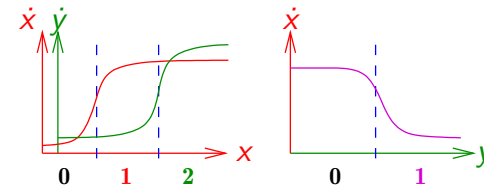
Presence of an activator = Absence of an inhibitor

$$\frac{dy}{dt} = k_0 + k_1 \cdot \mathbb{1}_{x_1 \geq \tau_1} + k_2 \cdot \mathbb{1}_{x_2 \geq \tau_2} + k_3 \cdot \mathbb{1}_{x_3 < \tau_3} + k_4 \cdot \mathbb{1}_{x_4 < \tau_4} - \gamma \cdot y$$

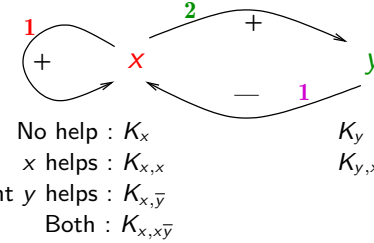
Solutions of the form $Ce^{-\gamma t} + \frac{\sum \mathbb{1}k_i}{\gamma}$ whose $\lim_{t \rightarrow \infty}$ is $\frac{\sum \mathbb{1}k_i}{\gamma}$

As many such equations as genes in the interaction graph
In each hypercube, all the trajectories have a unique attractive point, which can be outside de hypercube

Discrete Gene Networks (Thomas & Snoussi)



In each state, a variable v tries to go toward the interval numbered $K_{v,\omega}$: the one containing $\frac{\sum \mathbb{1}k_i}{\gamma}$

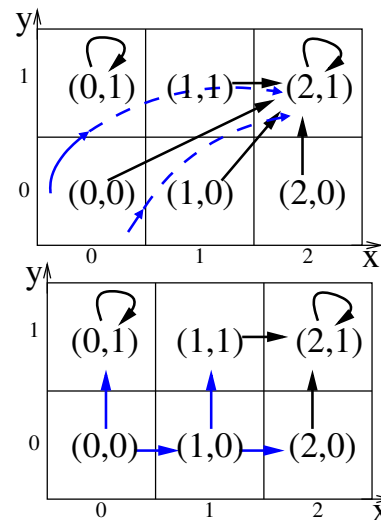


(x,y)	Focal Point
(0,0)	$(K_{x,\bar{y}}, K_y)$
(0,1)	(K_x, K_y)
(1,0)	$(K_{x,x\bar{y}}, K_y)$
(1,1)	$(K_{x,x}, K_y)$
(2,0)	$(K_{x,x\bar{y}}, K_{y,x})$
(2,1)	$(K_{x,x}, K_{y,x})$

Presence of an activator = Absence of an inhibitor = **A resource**

State Graphs

(x,y)	Focal Point
(0,0)	$(K_{x,\bar{y}}, K_y)=(2,1)$
(0,1)	$(K_x, K_y)=(0,1)$
(1,0)	$(K_{x,x\bar{y}}, K_y)=(2,1)$
(1,1)	$(K_{x,x}, K_y)=(2,1)$
(2,0)	$(K_{x,x\bar{y}}, K_{y,x})=(2,1)$
(2,1)	$(K_{x,x}, K_{y,x})=(2,1)$



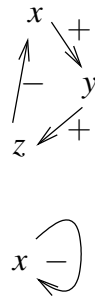
"desynchronization" → by units of Manhattan distance

Multistationarity vs. positive cycles

- A cycle in the interaction graph is *positive* if it contains an *even* number of inhibitions
- Theorem** : if the state graph exhibits several attraction basins then there is at least one positive cycle in the interaction graph
- Was a conjecture from the 70's to 2004 ; proved by Adrien Richard (and by Christophe Soulé for the continuous case)

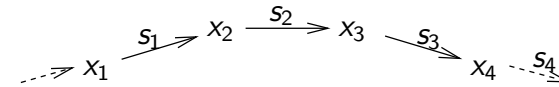


- A cycle in the interaction graph is *negative* if it contains a *odd* number of inhibitions
- **Thomas conjecture** : if the state graph exhibits an homeostasy (stable oscillations) then there is at least one negative cycle in the interaction graph
- Was a conjecture from the 70's to ≈2010. **Counter-examples** have been found (A. Richard, J.-P. Comet, P. Ruet)



Nonetheless it remains a very useful tip in practice when modelling biological examples!

Helps characterizing the saddle point (resp. center of the oscillations) of the behaviour “driven” by a positive (resp. negative) cycle.



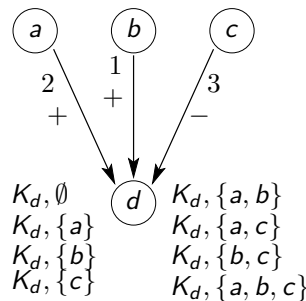
$$x_i = \text{threshold} \\ s_i - 1 \mid s_i$$

Whatever the sign of $x_i \rightarrow x_{i+1}$, for some set of resources ω one should have $K_{x_{i+1}, \omega} < s_{i+1} \leq K_{x_i, \omega x_i}$, all along the cycle

but it remains a heuristic, at least for negative cycles. . .

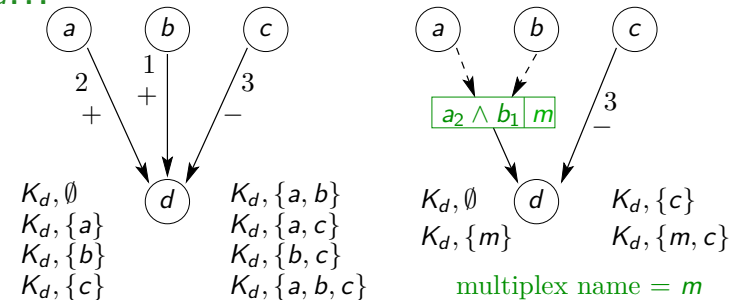
2^i parameters where i is the in-degree of the gene

$\prod_{\text{genes}} (o + 1)^{2^i}$ possible parameter values where o is the out degree of each gene



Yeast ≈ 7000 genes Human ≈ 25000 genes Rice ≈ 40000 genes

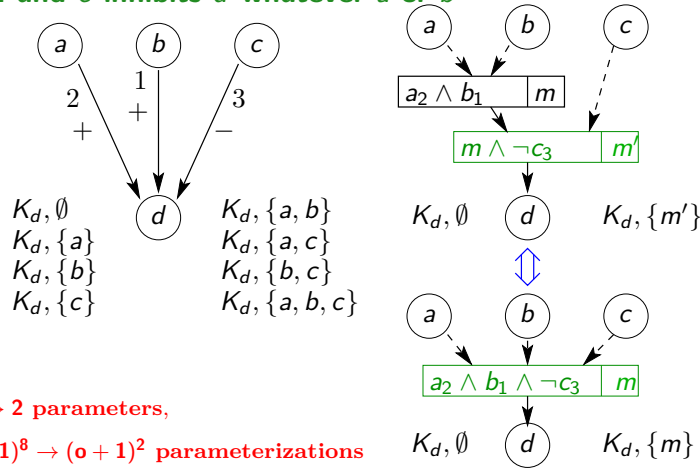
“Proteins of a and b form a complex before acting on d ...”



multiplex name = m
multiplex formula $\equiv a_2 \wedge b_1$
abbreviation :
 $v_i \equiv (v \geq i)$

8 → 4 parameters

"... and c inhibits d whatever a or b "



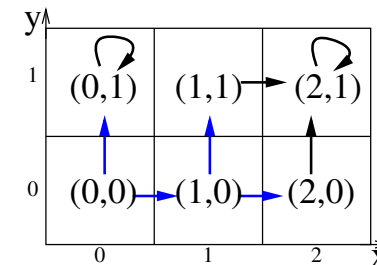
$8 \rightarrow 2$ parameters,
 $(o+1)^8 \rightarrow (o+1)^2$ parameterizations

Exhaustively identify the sets of (integer) parameters that cope with known behaviours from biological experiments

Solution = perform reverse engineering via formal logic

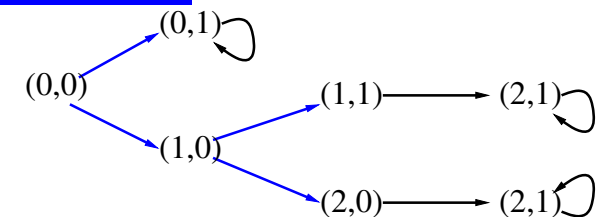
- 2003 : enumeration + CTL + model checking (Bernot, Comet, Pérès, Richard)
- 2005 : path derivatives + model checking (Batt, De Jong)
- 2005 : PROLOG with constraints (Trilling, Corblin, Fanchon)
- 2007 : symbolic execution + LTL (Mateus, Le Gall, Comet)
- 2011 : traces + enumeration + CTL + model checking (Siebert, Bockmayr)
- 2015 : genetically modified Hoare logic + constraint solving (Bernot, Comet, Roux, Khalis, Richard)

- 1 Discrete models for gene networks according to René Thomas
- 2 CTL
- 3 Techniques of software testing
- 4 General Schema for BRN
- 5 Genetically modified Hoare logic, and examples



As many possible state graphs as possible parameter sets... (huge number)

... from each initial state :



- **Atoms** = comparisons : $(x = 2), (y > 0) \dots$
Logical Connectives = $(\varphi_1 \wedge \varphi_2), (\varphi_1 \Rightarrow \varphi_2) \dots$
Temporal modalités = made de 2 characters :

first character	second character
A = for All path choices	X = neXt state
E = there Exists a choice	F = for some future state
	G = for all future state (Globally)
	U = Until

- **Examples :**
AX ($y=1$) : the concentration level of y belongs to the interval 1 in all states directly following the considered initial state.
EG ($x=0$) : there exists at least one path from the considered initial state where x always belongs to its lower interval.

neXt state :

- $EX\varphi$: φ can be satisfied in a next state
- $AX\varphi$: φ is always satisfied in the next states

eventually in the Future :

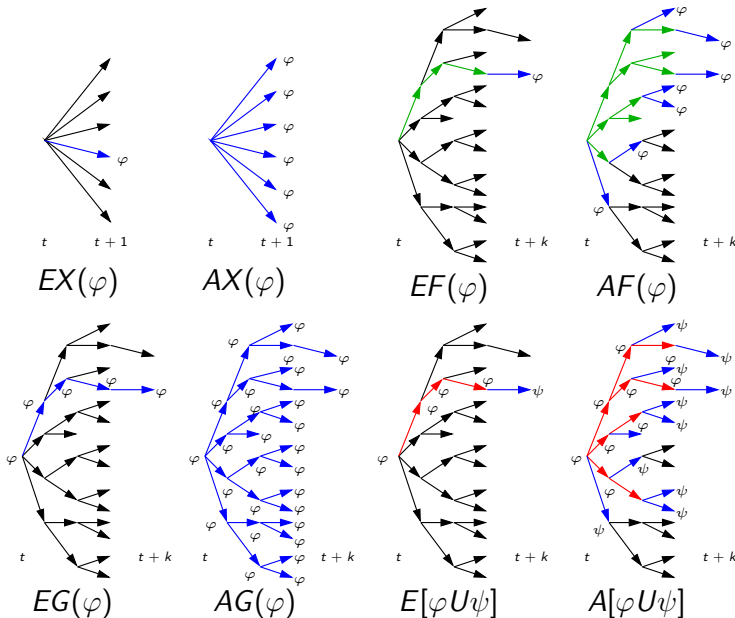
- $EF\varphi$: φ can be satisfied in the future
- $AF\varphi$: φ will be satisfied at some state in the future

Globally :

- $EG\varphi$: φ can be an invariant in the future
- $AG\varphi$: φ is necessarily an invariant in the future

Until :

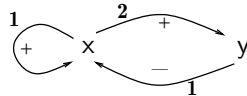
- $E[\psi U\varphi]$: there exist a path where ψ is satisfied until a state where φ is satisfied
- $A[\psi U\varphi]$: ψ is always satisfied until some state where φ is satisfied



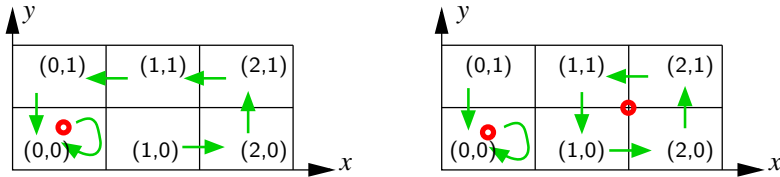
Let s_0 be a state. The CTL semantics is defined as follows

- $s_0 \models \top$ and $s_0 \not\models \perp$ $\forall p \in AP, s_0 \models p$ iff $p \in L(s_0)$,
- $s_0 \models \neg\varphi$ iff $s_0 \not\models \varphi$,
- $s_0 \models \varphi_1 \wedge \varphi_2$ (resp. $\varphi_1 \vee \varphi_2$) iff $s_0 \models \varphi_1$ and (resp. or) $s_0 \models \varphi_2$,
- $s_0 \models \varphi_1 \Rightarrow \varphi_2$ iff $s_0 \not\models \varphi_1$ or $s_0 \models \varphi_2$,
- $s_0 \models \varphi_1 \Leftrightarrow \varphi_2$ iff $s_0 \models (\varphi_1 \Rightarrow \varphi_2) \wedge (\varphi_2 \Rightarrow \varphi_1)$,
- $s_0 \models AX\varphi$ iff for all successors s_1 of s_0 , one has $s_1 \models \varphi$,
- $s_0 \models EX\varphi$ iff there exists a successor s_1 of s_0 such that $s_1 \models \varphi$,
- $s_0 \models AG\varphi$ iff $\forall s_i$ from any path $s_0s_1 \dots s_i \dots$, one has $s_i \models \varphi$,
- $s_0 \models EG\varphi$ iff \exists a path $s_0s_1 \dots s_i \dots$, s.t. $\forall s_i$, one has $s_i \models \varphi$,
- $s_0 \models AF\varphi$ iff \forall path $s_0s_1 \dots s_i \dots$, $\exists j$ s.t. $s_j \models \varphi$,
- $s_0 \models EF\varphi$ iff \exists a path $s_0s_1 \dots s_i \dots$, $\exists j$ s.t. $s_j \models \varphi$,
- $s_0 \models A[\varphi_1 U\varphi_2]$ iff \forall path $s_0s_1 \dots s_i \dots$, $\exists j$ s.t. $s_j \models \varphi_2$, and $\forall i < j, s_i \models \varphi_1$,
- $s_0 \models E[\varphi_1 U\varphi_2]$ iff \exists path $s_0s_1 \dots s_i \dots$, $\exists j$ s.t. $s_j \models \varphi_2$, and $\forall i < j, s_i \models \varphi_1$

Common properties :
"functionality" of a sub-graph
Special role of "feedback loops"



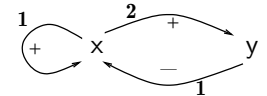
- positive : *multistationnarity* (even number of —)
- negative : *homeostasy* (odd number of —)



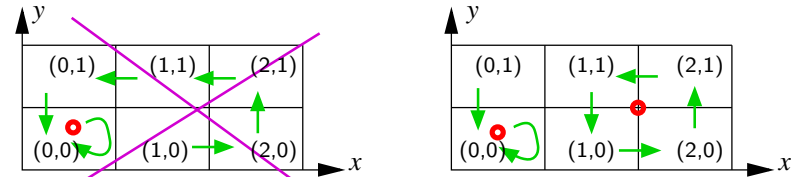
$$\text{Characteristic properties : } \begin{cases} (x = 2) \implies AG(\neg(x = 0)) \\ (x = 0) \implies AG(\neg(x = 2)) \end{cases}$$

They express "the positive feedback loop is functional"
(satisfaction of these formulas relies on the parameters $K_{..}$)

Common properties :
"functionality" of a sub-graph
Special role of "feedback loops"



- positive : *multistationnarity* (even number of —)
- negative : *homeostasy* (odd number of —)



$$\text{Characteristic properties : } \begin{cases} (x = 2) \implies AG(\neg(x = 0)) \\ (x = 0) \implies AG(\neg(x = 2)) \end{cases}$$

They express "the positive feedback loop is functional"
(satisfaction of these formulas relies on the parameters $K_{..}$)

- Efficiently computes all the states of a state graph which satisfy a given formula : $\{ \eta \mid M \models_{\eta} \varphi \}$.
- Efficiently select the models which globally satisfy a given formula.

Intensively used :

- to find the set of **all** possible discrete parameter values
- to check models under construction w.r.t. **known behaviours** (one often gets an empty set of parameter values!)
- and to prove the **consistency** of a biological **hypothesis**

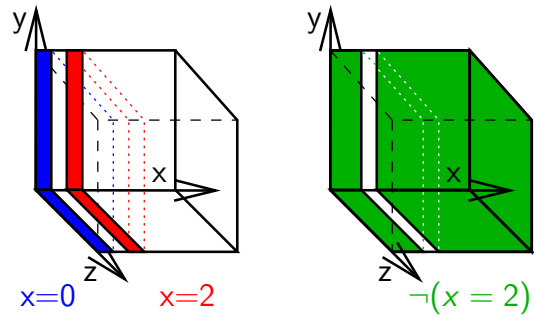
Computes all the states of a discrete state graph that satisfy a given formula : $\{ \eta \mid M \models_{\eta} \varphi \}$.

- Idea 1** : work on the state graph instead of the path trees.
- Idea 2** : check first the atoms of φ and then check the connectives of φ with a bottom-up computation strategy.
- Idea 3** : (computational optimization) group some cases together using BDDs (Binary Decision Diagrams).

Example : $(x = 0) \implies AG(\neg(x = 2))$

Obsession : *travel the state graph as less as possible*

$(x = 0) \implies AG(\neg(x = 2))$



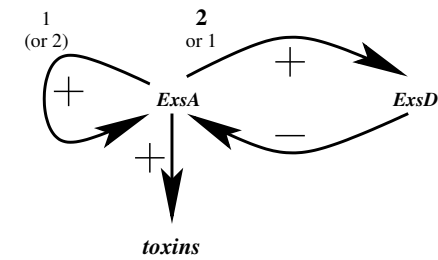
$AG(\neg(x = 2))?$

... one should **travel all** the paths from any green box and check if successive boxes are green : *too many boxes to visit.*

Trick : $AG(\neg(x = 2))$ is equivalent to $\neg EF(x = 2)$
start from the red boxes and follow the transitions backward.

and

Consistency of the epigenetic hypothesis



- 2 possible stable states :
 - $(EXsA = 2) \implies AX AF(EXsA = 2)$
 - $(EXsA = 0) \implies AG(\neg(EXsA = 2))$
- **Question 1, consistency** : proved by Model checking
8 models among 648, automatically extracted.
- **Question 2, and in vivo?**

Formula = Model-Experiment Link

Formulas are valid or invalid in relation to a set of given traces starting from a given state.

They can be compared with

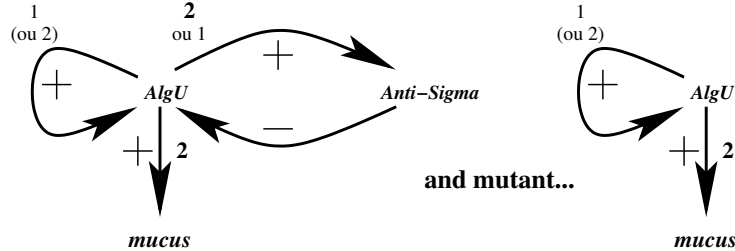
- all possible traces of the theoretical model
- all known experiments

\implies **They are therefore the link between models and biological objects.**

A simple example

- Phenotype modification, terminology :
 - **genetic modification** : heritable and irreversible (mutation)
 - **epigenetic modification** : heritable but reversible
 - **Adaptation** : non-heritable and reversible
- **Biological questions** :
 - Is cytotoxicity (and/or muco-toxicity) in the bacterium *Pseudomonas aeruginosa* epigenetic in nature?
 - [\implies Cystic Fibrosis]

wild pseudomonas aeruginosa :

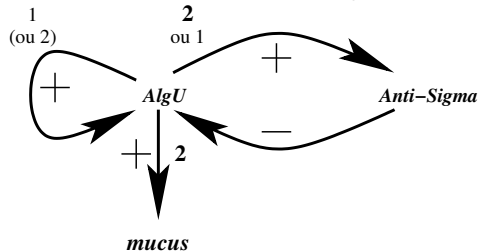


Epigenetic hypothesis (i.e. without mutation)

- The positive cycle is functional in spite of the negative cycle, with one state non-mucoid and the other mucoid.
- An external signal (produced by the diseased lung) could potentially switch AlgU from the low state to the high state.
- Selection pressure then favours mutants in the mucosal environment ⇒ New perspectives for therapy.

- Question 2 = Validate the stability of the two states in vivo.
- Non-mucoid state : $(AlgU = 0) \Rightarrow AG(AlgU < 2)$
A bacterium with its basal level of AlgU will not spontaneously become mucoid : validated daily
- Mucoid state : $(AlgU = 2) \Rightarrow AX AF(AlgU = 2)$
- Working hypothesis : AlgU can be brought to saturation, not measured
- Experimental design : pulse AlgU and then, after a transition period, test whether mucus production continues pulse de AlgU puis après une phase transitoire, tester si la production de mucus persiste (\Leftrightarrow check for hysteresis)
- Experimental designs can be generated automatically

- $AlgU = 2$ cannot be checked directly, but $mucus = 1$ can



- Lemma : $AG(AlgU = 2) \Leftrightarrow AF AG(mucus = 1)$
- (... Computer-aided proofing ...)

→ Experiment : $(AlgU = 2) \Rightarrow AF AG(mucus = 1)$

$A \Rightarrow B$	True	False
True	True	False
False	True	True

Karl Popper :

Validate = attempt to refute

So A false is useless

So start with a pulse...

Pulse allows to reach initial state $AlgU = 2$.

Otherwise we would have to establish a lemma :

$$(AlgU = 2) \Leftrightarrow (\text{something achievable})$$

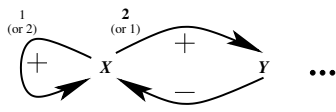
General form of a test :

$$(\text{something achievable}) \Rightarrow (\text{something observable})$$

- 1 Discrete models for gene networks according to René Thomas
- 2 CTL
- 3 Techniques of software testing
- 4 General Schema for BRN
- 5 Genetically modified Hoare logic, and examples

- Similar problem** = does the software meet its specification?
 Infinite number of possible test scenarios, select revealing ones
- Solution** = divide into scenario domains
 Behaviour assumed to be "uniform" within a domain
- Formula **unfolding** is used to divide the domains
 - **Probabilistic** approach : few unfolding, few (but large) domains, probabilistic drawing of many tests in each domain.
 - **Deterministic** approach : many unfoldings, small domains, selection of a single test per domain.

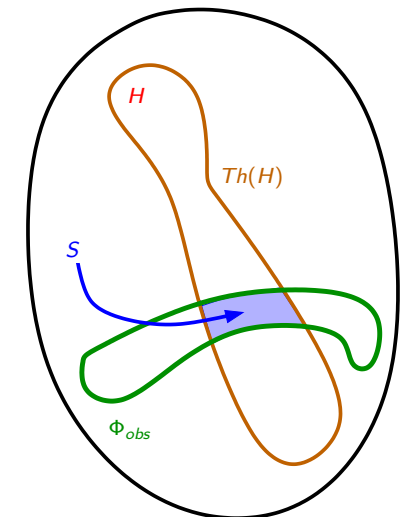
99% of defects can be detected automatically

- 
- $\Phi = \{\varphi_1, \varphi_2, \dots, \varphi_n\}$ and $M =$
 - **Set of formulae resulting from the hypothesis :**
 $Th(H) = \{\psi \mid \Phi, M \models \psi\}$
 - **Observable formulae :** Φ_{obs}
 $\{\psi \mid \psi \text{ of the form "achievable"} \Rightarrow \text{"observable"}\}$
 - **Problem :** $\Phi_{obs} \cap Th(H)$ is infinite
 \rightarrow Selecting "Revealers" in $\Phi_{obs} \cap Th(H)$
 - **P. aeruginosa :** By luck, there are 2 observable formulae $\psi_1, \psi_2 \in \Phi_{obs} \cap Th(H)$ such that $\{\psi_1, \psi_2\} \models \Phi$
 - **General computer science solution :** unfolding techniques (\simeq case-by-case reasoning) should make it possible to make explicit the assumptions made when limited to a fixed number of experiments.

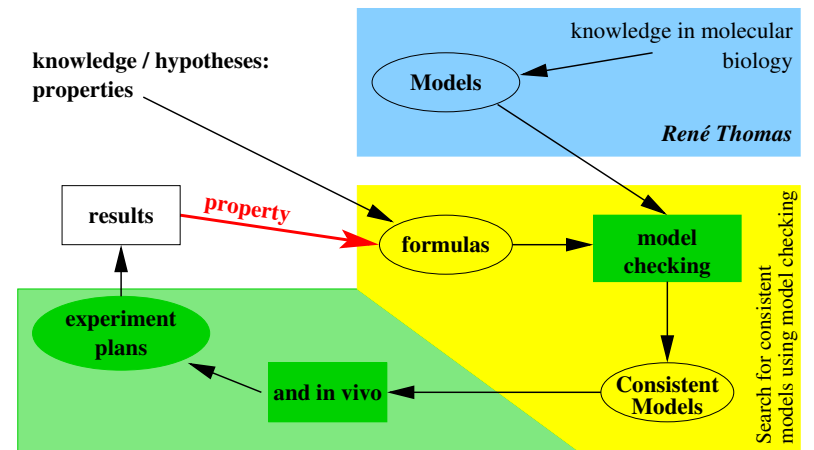
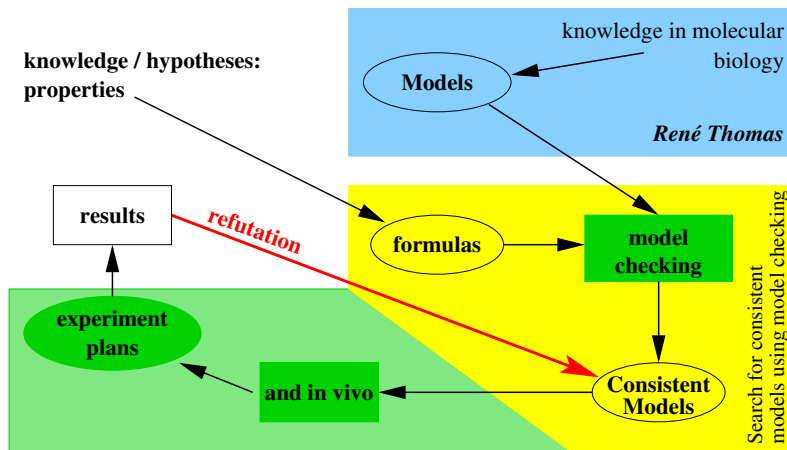
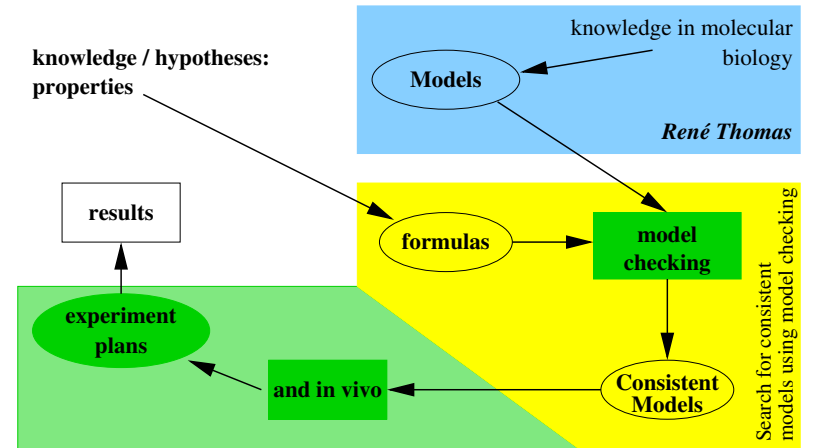
- **H** : hypothesis
- Φ_{obs} : possible experiments
- $Th(H)$: logical consequences of H
- **S** : experiments in relationship with H

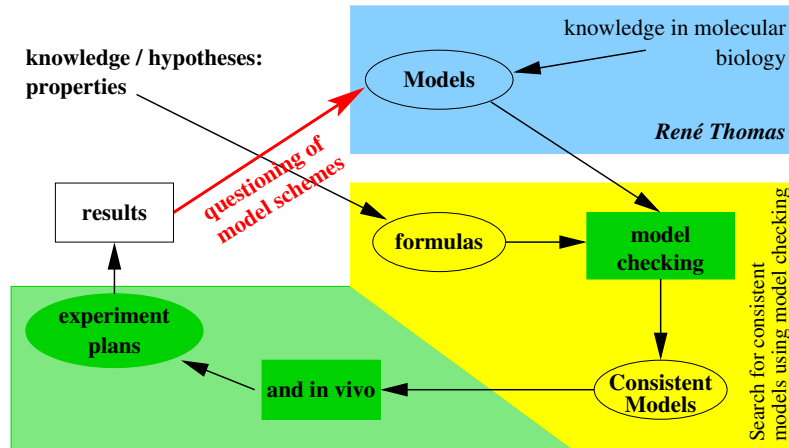
Refutability : $S \Rightarrow H$

The set S is infinite...
 Choice of experiments in S?
 ... optimizations



- 1 Discrete models for gene networks according to René Thomas
- 2 CTL
- 3 Techniques of software testing
- 4 General Schema for BRN
- 5 Genetically modified Hoare logic, and examples





- Given a set of models
- Given a set of models Given a set of possible experiments (in the form of formulas)
- Questions :
 - What experiment must be performed to reduce the set of consistent models? (equiprobable / non equiprobable models)
 - Ditto for n experiments (order, decision tree?)
 - Ditto with cost?

- $M = \{M_1, M_2, \dots, M_m\}$ and $F = \{F_1, F_2, \dots, F_f\}$

	F_1	F_2	...	F_f
M_1	1	1	...	0
M_2	1	0	...	0
...
M_m	0	1	...	0

using model checking :

- If the models are equi probable, we implement F_i which balances the 2 sets

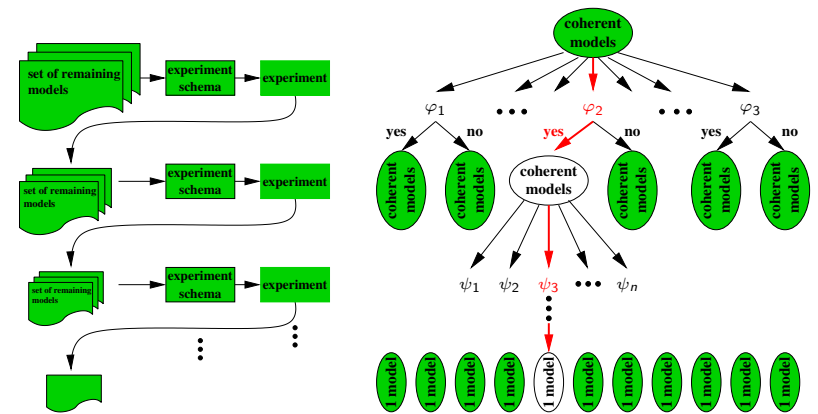
$$E_i = \{M_j | M_j \models F_i\} \quad \text{and} \quad \bar{E}_i = \{M_j | M_j \not\models F_i\}$$

- otherwise F_i which balances the 2 probabilities

$$p(\{M_j | M_j \models F_i\}) \quad \text{and} \quad p(\{M_j | M_j \not\models F_i\})$$

In fact, we're looking to minimize $E[\text{Size of set after exp.}]$

- $\min(|E_i| \times |E_i| + |\bar{E}_i| \times |\bar{E}_i|) = \min(|E_i|^2 + (N - |E_i|)^2)$
- $\min(N^2 - 2N|E_i| + 2|E_i|^2)$
- minimum in $N/2$



Choosing a complete strategy (2)

master BBC

Jean-Paul Comet

R. Thomas

CTL

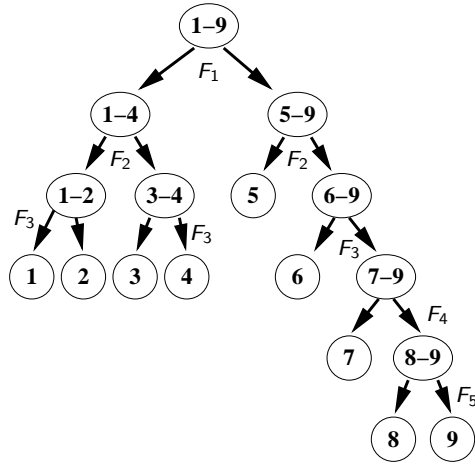
Software Engineering

General Schema

Hoare Logic

- The previous strategy doesn't give the minimum depth tree.
- Ex : 9 models ; 5 formulas, min. height = $\log_2(9) = 4$

	F ₁	F ₂	F ₃	F ₄	F ₅
M ₁	1	1	1	0	0
M ₂	1	1	0	1	1
M ₃	1	0	1	0	1
M ₄	1	0	0	1	0
M ₅	0	1	0	0	0
M ₆	0	0	1	0	0
M ₇	0	0	0	1	0
M ₈	0	0	0	0	1
M ₉	0	0	0	0	0
	4/5	3/6	3/6	3/6	3/6



Thanks to S. Vial for this example

Choosing a complete strategy (3)

master BBC

Jean-Paul Comet

R. Thomas

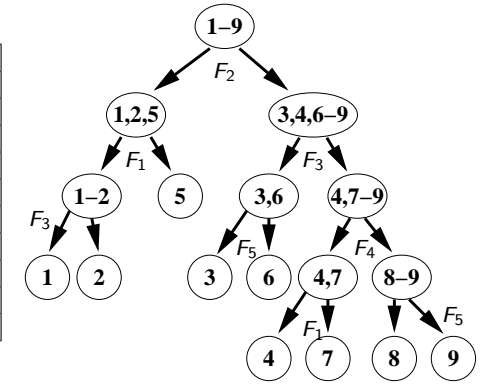
CTL

Software Engineering

General Schema

Hoare Logic

	F ₁	F ₂	F ₃	F ₄	F ₅
M ₁	1	1	1	0	0
M ₂	1	1	0	1	1
M ₃	1	0	1	0	1
M ₄	1	0	0	1	0
M ₅	0	1	0	0	0
M ₆	0	0	1	0	0
M ₇	0	0	0	1	0
M ₈	0	0	0	0	1
M ₉	0	0	0	0	0
	4/5	3/6	3/6	3/6	3/6



Choosing an optimal decision tree = NP-complete problem (reduction to the 3-DM problem, L. Hyafil and R.L. Rivest [1975])

Choosing a complete strategy (4)

master BBC

Jean-Paul Comet

R. Thomas

CTL

Software Engineering

General Schema

Hoare Logic

	Temporal formulas	Coherent models
1	$x = 0 \Rightarrow AXAF(x = 0)$	1, 3, 6, 7, 8, 9, 10
2	$x = 2 \Rightarrow AXAF(x = 2)$	1, 2, 3, 4, 5, 7, 10
3	$x = 1 \Rightarrow AXAF(x = 0)$	1, 3
4	$x = 1 \Rightarrow AXAF(x = 2)$	7, 10
5	$y = 0 \Rightarrow AXAF(y = 0)$	1, 2, 3, 6, 1, 2, 3, 6

If you don't want to :

- choose a discriminating formula at random
- choose a formula that is easy to implement in vivo (cost)
- adjust this choice according to intuition
- choose the formula that best cuts M

Using the min-max algorithm to optimize selection :

- determine observable formulas
- limit tree depth (here, prof = 3)
- find the tree for which the cost is minimal

Choosing a complete strategy (4-b)

master BBC

Jean-Paul Comet

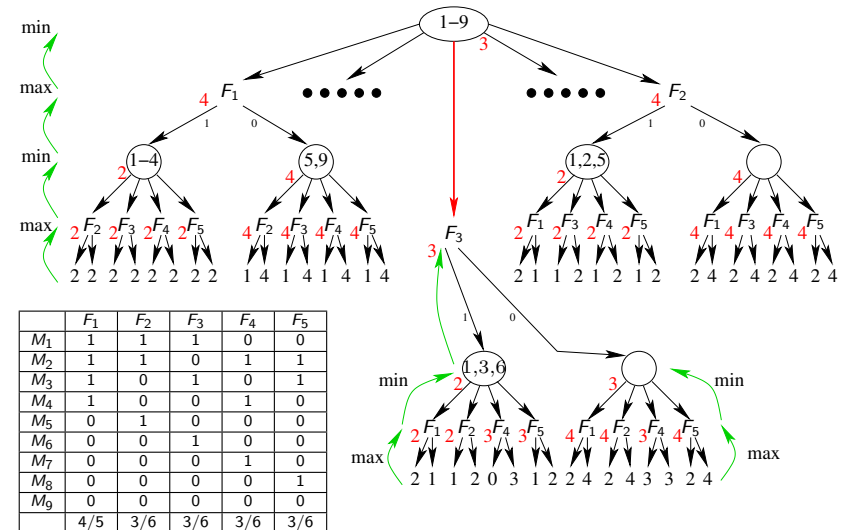
R. Thomas

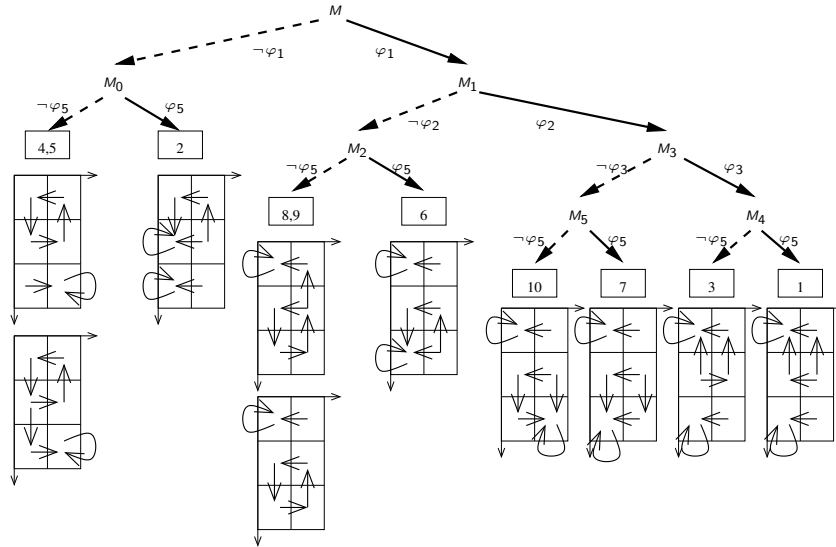
CTL

Software Engineering

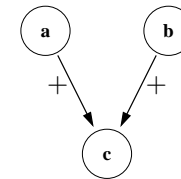
General Schema

Hoare Logic





Signs and parameters



K_c	=	1	Positive Actions of b
$K_{c,a}$	=	0	
$K_{c,b}$	=	2	
$K_{c,ab}$	=	1	
K_c	=	1	Negative Actions of a
$K_{c,a}$	=	0	
$K_{c,b}$	=	2	
$K_{c,ab}$	=	1	

Relationship between signs and parameters

- ODE system $\frac{dx_3}{dt} = (k + k_1 \cdot \mathbf{1}_{x_1 \rightarrow x_3} + k_2 \cdot \mathbf{1}_{x_2 \rightarrow x_3}) - \lambda \times x_3$
- Discretisation :

if neither x_1 nor x_2 acts on x_3 :	$d(k)$	K_{x_3}
if only x_1 acts on x_3 :	$d(k + k_1)$	K_{x_3, x_1}
if only x_2 acts on x_3 :	$d(k + k_2)$	K_{x_3, x_2}
if both x_1 and x_2 act on x_3 :	$d(k + k_1 + k_2)$	$K_{x_3, x_1 x_2}$
- Sum of positive numbers : Snoussi conditions :

$$\forall a \in \forall G^-(x), \forall \omega \subseteq G^-(x), K_{x,\omega} \leq K_{x,\omega \cup \{a\}}$$

Everywhere, the addition of a resource cannot reduce the the attractor

- Consequence : XOR is not possible

x_1	x_2	$x_3 = x_1 \text{ XOR } x_2$	
0	0	$K_{x_3} = 0$	
0	1	$K_{x_3, x_2} = 1$	
1	0	$K_{x_3, x_1} = 1$	
1	1	$K_{x_3, x_1 x_2} = 0$	

- Everywhere, the addition of a resource cannot reduce the the attractor
- There is a configuration where the addition of a resource creates an increase in the attractor

$$\forall a \in \forall G^-(x), \exists \omega \subseteq G^-(x), K_{x,\omega} < K_{x,\omega \cup \{a\}}$$

- The sign thus becomes a constraint on the parameters.
- Notation : $+_{obs}, -_{obs}$ to be distinguished from $+_{snoussi}, -_{snoussi}$