Reconciling qualitative and abstract (and scalable) reasoning with Boolean networks

Loïc Paulevé[1,2], Thomas Chatain[3], Stefan Haar[3]

[1] CNRS, LaBRI, Bordeaux, France

[2] CNRS, LRI, Univ Paris-Sud, Univ Paris-Saclay, France

[3] LSV, ENS Paris-Saclay, Inria Saclay, France

Overview

Discrete Dynamical Systems
(Boolean Networks)

+

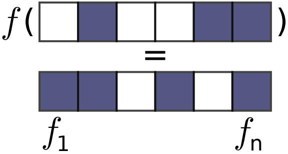
Systems Biology
(Signalling/regulation networks)

Concurrency Theory
(Semantics)

Introduction

Boolean Network (BN) $f : \mathbb{B}^n \to \mathbb{B}^n$

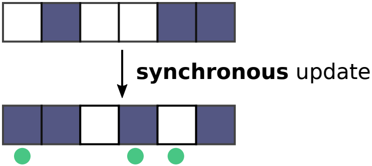
Configuration: $x \in \mathbb{B}^n$

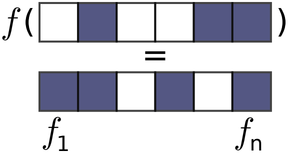


Introduction

Boolean Network (BN) $f : \mathbb{B}^n \to \mathbb{B}^n$

Configuration: $x \in \mathbb{B}^n$

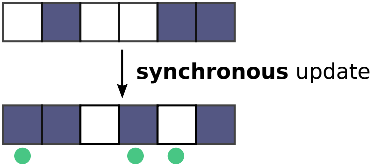


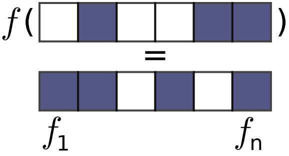


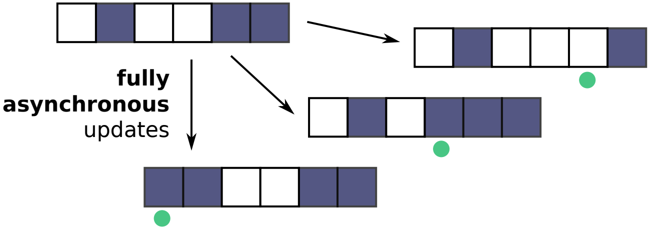
Introduction

Boolean Network (BN) $f : \mathbb{B}^n \to \mathbb{B}^n$

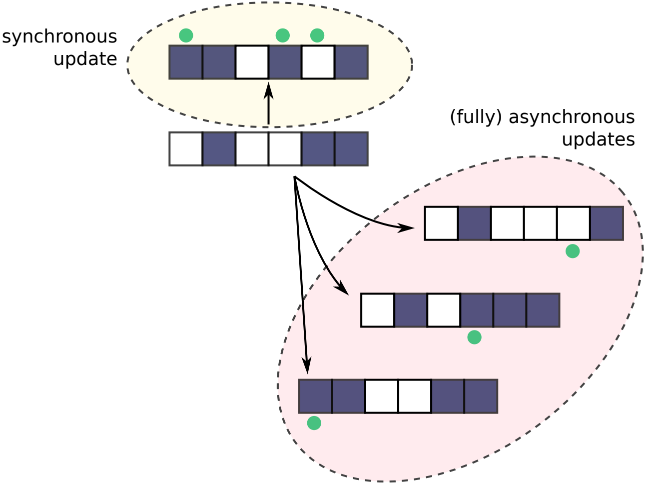
Configuration: $x \in \mathbb{B}^n$



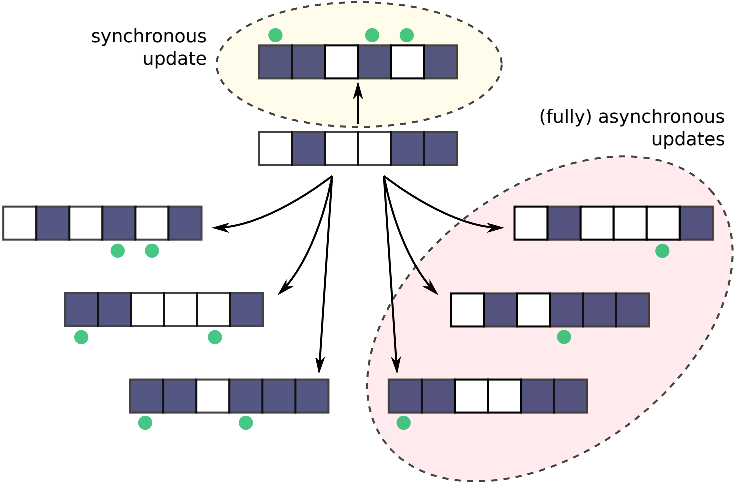




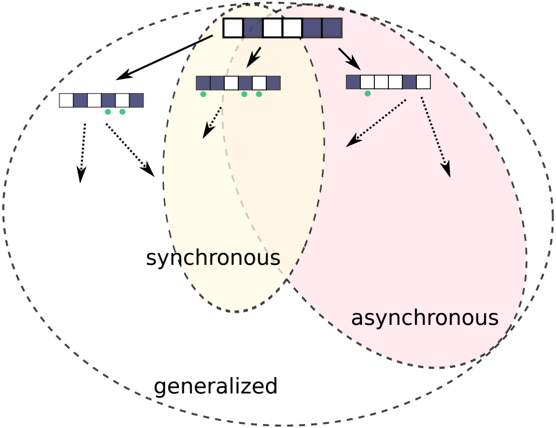
Generalized Asynchronicity



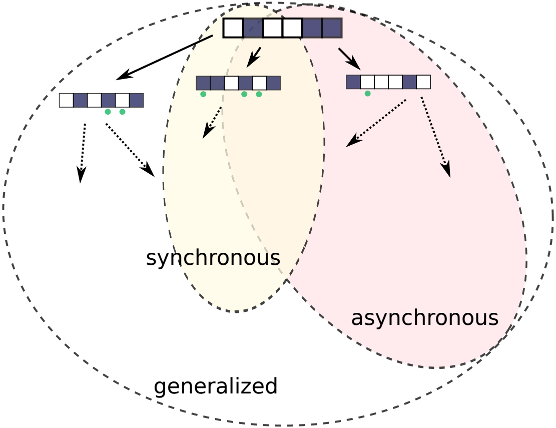
Generalized Asynchronicity



Reachable configurations



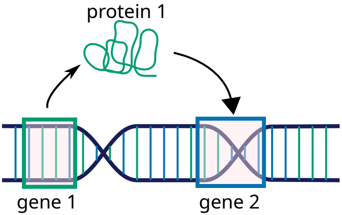
Reachable configurations



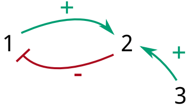
Should we reach configurations beyond generalized asynchronicity?

Boolean networks for biological processes

Example with gene regulation

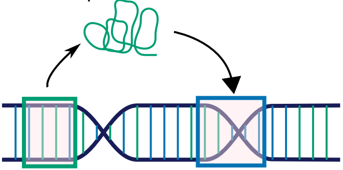


Influence graph

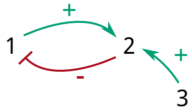


Boolean networks for biological processes

Example with gene regulation



Influence graph



Reachable configurations

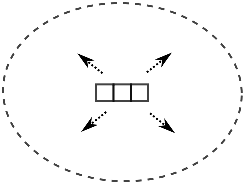
Boolean network

$$f_1(x) \triangleq \neg x_2$$

$$f_2(x) \triangleq x_1 \wedge x_3$$

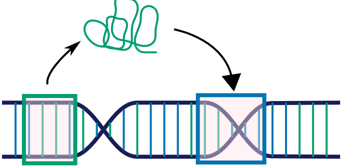
$$f_3(x) \triangleq \dots$$

\+ update mode =

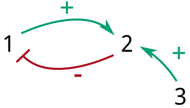


Boolean networks for biological processes

Example with gene regulation



Influence graph

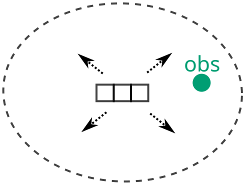


Reachable configurations

Boolean network

$$f_1(x) \triangleq \neg x_2$$
$$f_2(x) \triangleq x_1 \wedge x_3$$
$$f_3(x) \triangleq \ldots$$

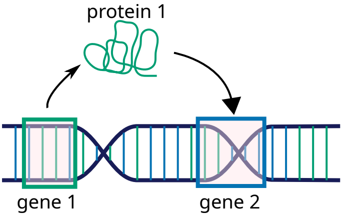
+ update mode =

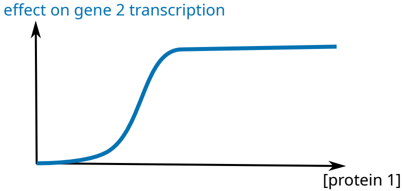


Validation w.r.t. observations (e.g. time series data)
⇒ we expect measurements match with reachable configurations

Gene expression is not Boolean

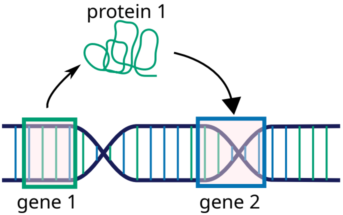
Qualitative modelling: Boolean vs multivalued networks

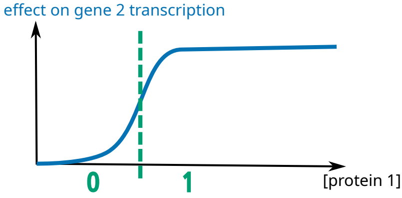




Gene expression is not Boolean

Qualitative modelling: Boolean vs multivalued networks



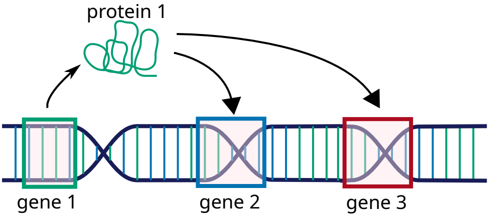


Boolean network

$$f_2(x) \triangleq x_1$$

Gene expression is not Boolean

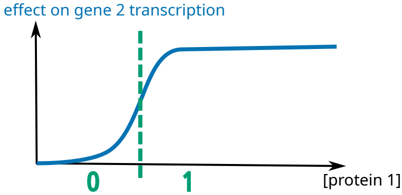
Qualitative modelling: Boolean vs multivalued networks



Boolean network

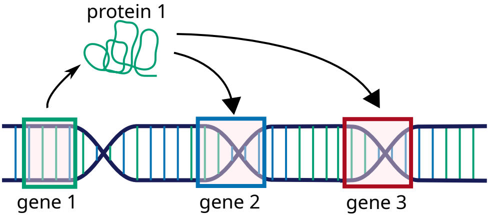
$$f_2(x) \triangleq x_1$$

$$f_3(x) \triangleq x_1$$



Gene expression is not Boolean

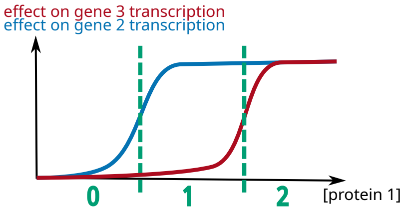
Qualitative modelling: Boolean vs multivalued networks



Boolean network

$$f_2(x) \triangleq x_1$$

$$f_3(x) \triangleq x_1$$



Multivalued network

$$f_2(x) \triangleq (x_1 \geq 1)$$

$$f_3(x) \triangleq (x_1 \geq 2)$$

Remark: Multivalued models can require different thresholds for each target

Properties of Boolean networks for biology

Given a Boolean network f of dimension n

Reachability (seq. of transitions from conf. x to y)
⇒ PSPACE-complete with update modes
Potential behaviours/capabilities of the cell

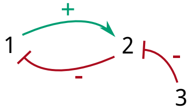
Fixpoints ($f(x) = x$)
⇒ NP-complete for sync/async/gasync
Steady states/phenotypes

Attractors (smallest sets of conf. closed by transitions)
⇒ PSPACE-complete with update modes
Steady states/phenotypes

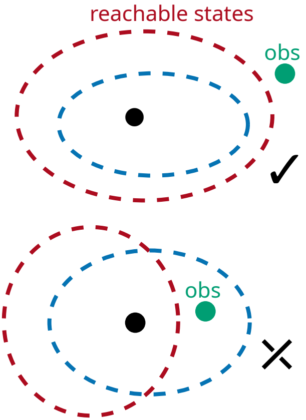
Qualitative vs abstract modelling

information

Influence graph



Consistency
analysis at Boolean level
transposable to multilevel?



Boolean network

- logic of activity w.r.t. regulators
- *update mode* (sync, async, etc.)

Multilevel network

+ define activation thresholds

Quantitative model

Update modes

of Boolean networks:

a **bug**…

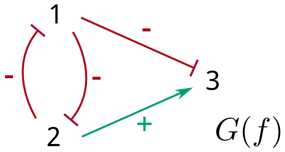
Motivating example

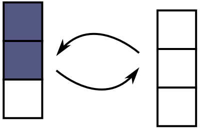
(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg x_2$$

$$f_2(x) \triangleq \neg x_1$$

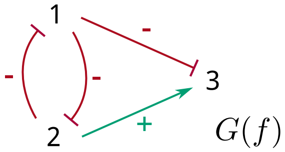
$$f_3(x) \triangleq \neg x_1 \wedge x_2$$

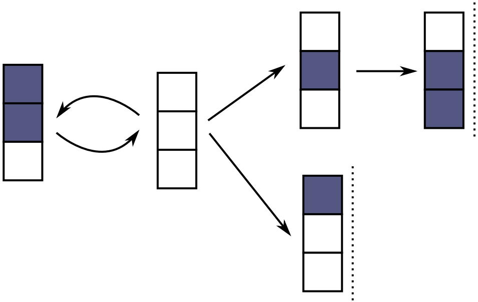




Motivating example
(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg x_2$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge x_2$$





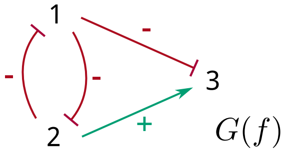
Motivating example

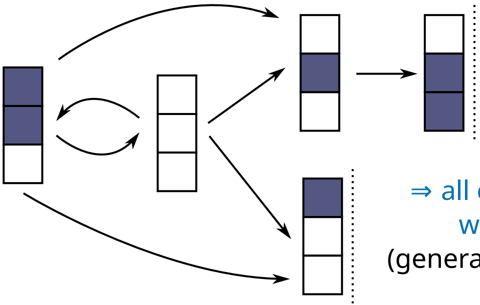
(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg x_2$$

$$f_2(x) \triangleq \neg x_1$$

$$f_3(x) \triangleq \neg x_1 \wedge x_2$$





⇒ all configurations reachable
with any update mode
(generalized) asynchronous mode

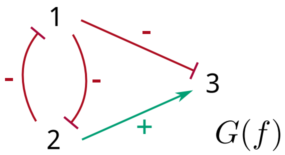
Motivating example

(embedded in many actual biological networks)

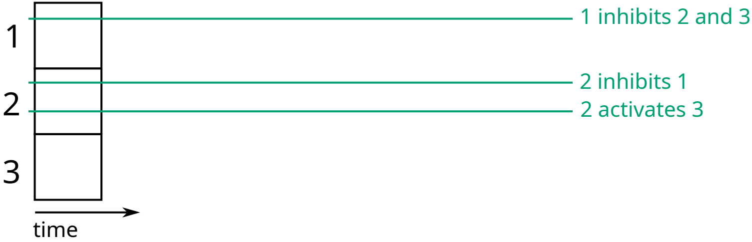
$$f_1(x) \triangleq \neg(x_2 \geq 2)$$

$$f_2(x) \triangleq \neg x_1$$

$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$



Compatible continuous/multilevel dynamics:



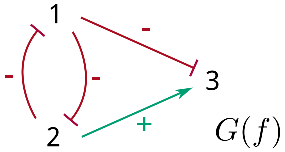
Motivating example

(embedded in many actual biological networks)

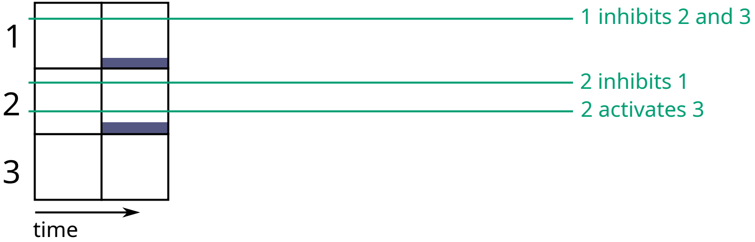
$$f_1(x) \triangleq \neg(x_2 \geq 2)$$

$$f_2(x) \triangleq \neg x_1$$

$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$

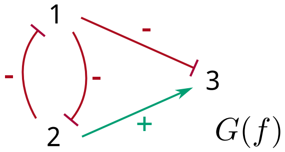


Compatible continuous/multilevel dynamics:

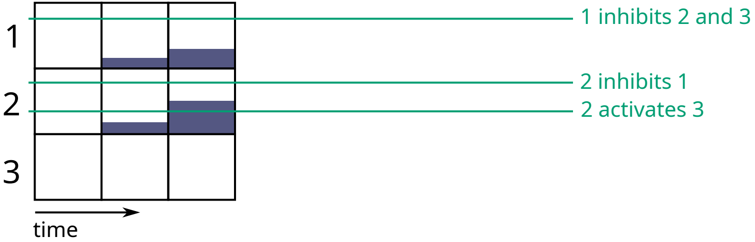


Motivating example
(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg(x_2 \geq 2)$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$



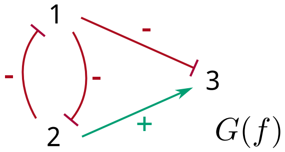
Compatible continuous/multilevel dynamics:



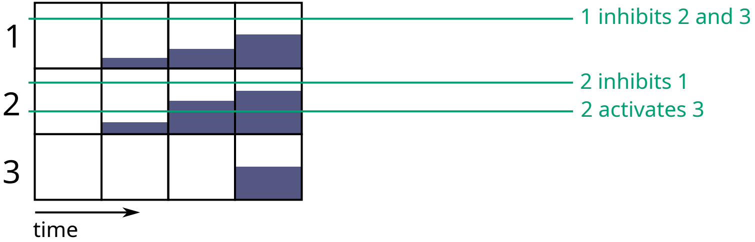
Motivating example

(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg(x_2 \geq 2)$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$



Compatible continuous/multilevel dynamics:

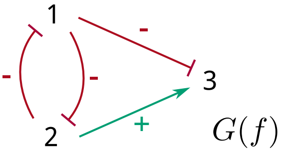


Motivating example
(embedded in many actual biological networks)

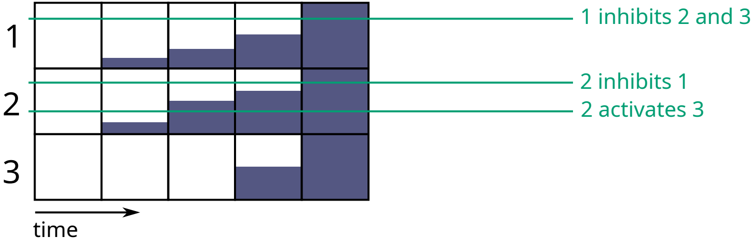
$$f_1(x) \triangleq \neg(x_2 \geq 2)$$

$$f_2(x) \triangleq \neg x_1$$

$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$

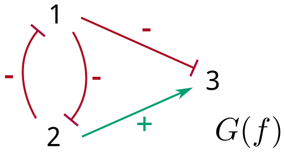


Compatible continuous/multilevel dynamics:

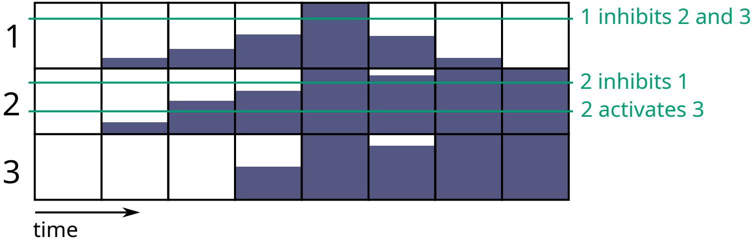


Motivating example
(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg(x_2 \geq 2)$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$

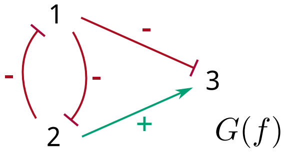


Compatible continuous/multilevel dynamics:

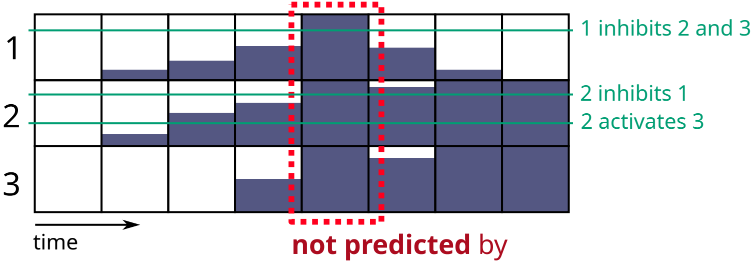


Motivating example
(embedded in many actual biological networks)

$$f_1(x) \triangleq \neg(x_2 \geq 2)$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge (x_2 \geq 1)$$



Compatible continuous/multilevel dynamics:



not predicted by
update modes in Boolean

Practical implications

Update modes can miss admissible transitions

Model synthesis from observations

⇒ Reject valid solutions (false negatives)

(wrongly concludes on reachability)

Prediction for reprogramming (control)

Find mutations such that

1. y (goal phenotype) is reachable from x

⇒ False negatives

2. z (bad phenotype) is not reachable from x

⇒ False positives

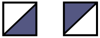
Most permissive semantics of Boolean networks

enabling new behaviours

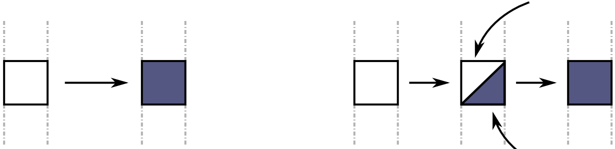
Most permissive semantics

- delay between firing and application of state change

  ⇒ allow interleaving other state changes

- in "intermediate" states 

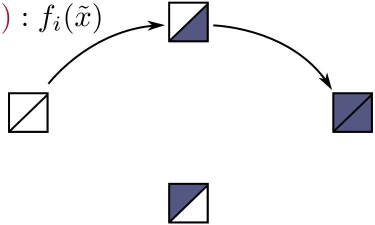
  other components choose what they see



Most permissive semantics

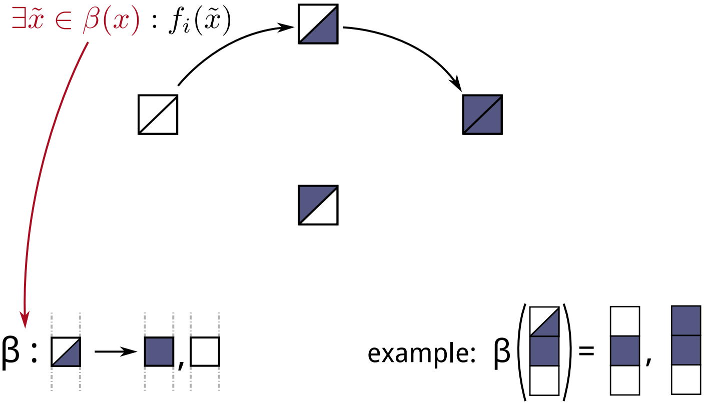
Rules for state of component i:

$\exists \tilde{x} \in \beta(x) : f_i(\tilde{x})$



Most permissive semantics

Rules for state of component i:

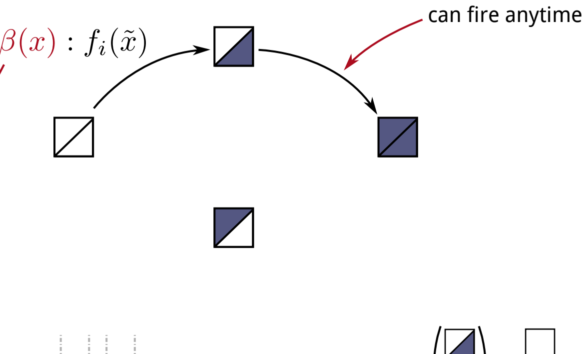


Choose value of "changing" components
(act as choosing an activation threshold)

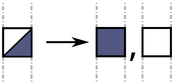
Most permissive semantics

Rules for state of component i:

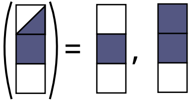
$\exists \tilde{x} \in \beta(x) : f_i(\tilde{x})$



β :



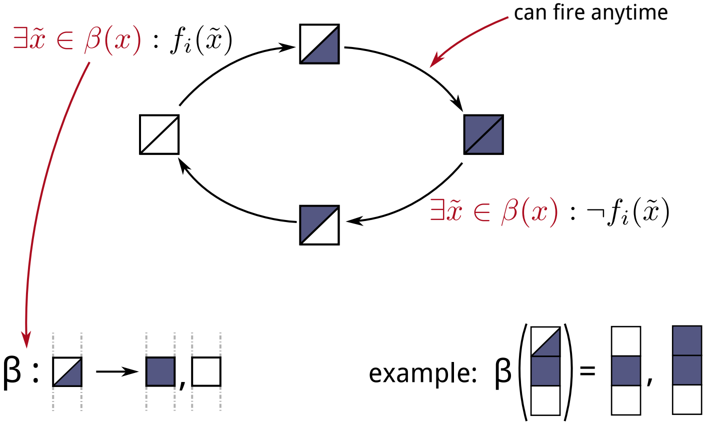
example: β



Choose value of "changing" components
(act as choosing an activation threshold)

Most permissive semantics

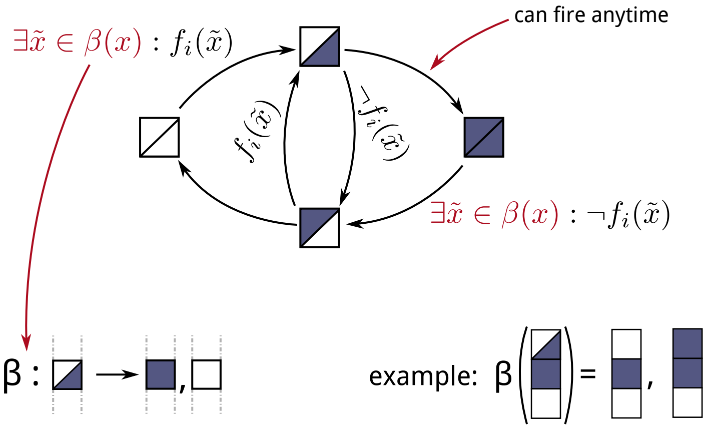
Rules for state of component i:



Choose value of "changing" components
(act as choosing an activation threshold)

Most permissive semantics

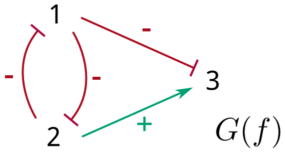
Rules for state of component i:

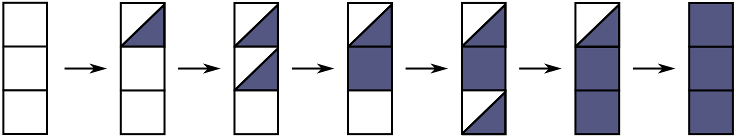


Choose value of "changing" components
(act as choosing an activation threshold)

Application to motivating example

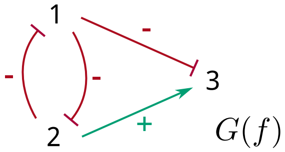
$$f_1(x) \triangleq \neg x_2$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge x_2$$

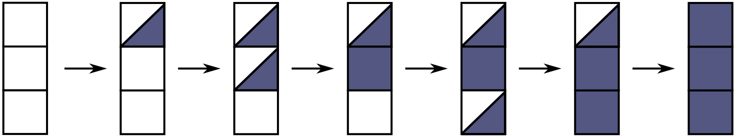




Application to motivating example

$$f_1(x) \triangleq \neg x_2$$
$$f_2(x) \triangleq \neg x_1$$
$$f_3(x) \triangleq \neg x_1 \wedge x_2$$





⇒ valid with respect to multivalued refinements

Properties of the most permissive semantics

Correct abstraction of multilevel/quantitative systems:

- includes all the transitions of every update mode
- multilevel refinements only remove behaviours

- Reachability can be decided in quadratic nb of transitions

(PTIME with locally-monotonic networks, or encoded as BDDs/Petri nets/...;
NP-complete otherwise; instead of PSPACE-complete with update modes)

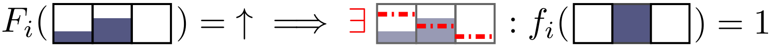
- Attractors are hypercubes (minimal trap spaces)
  - ⇒ finding attractors is NP-complete (instead of PSPACE-complete)
  - ⇒ fixpoints are the same as with update modes

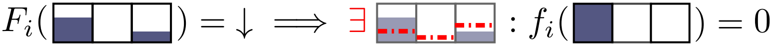
Refinements of Boolean Networks

A multivalued network

$$F : \mathbb{M}^n \to \{\uparrow, -, \downarrow\}$$

is a refinement of a Boolean network *f* iff



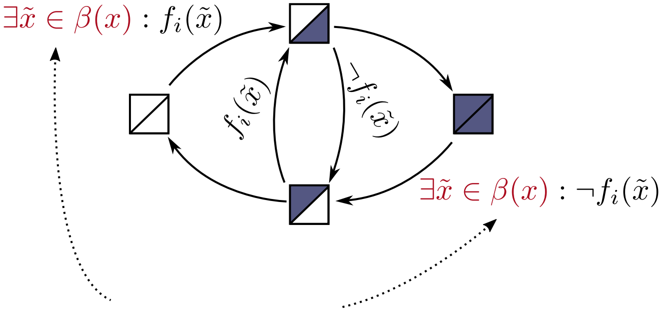


Most permissive semantics weakly simulates
any multivalued refinement with *any update mode*

(can be extended to ODEs)

Reachability with the most permissive semantics

Cost of one transition in component i



NP (SAT) in general;
Linear with
- locally-monotonic networks (f_i are monotone)
- when f is encoded as BDDs/Petri nets/...

Reachability with the most permissive semantics

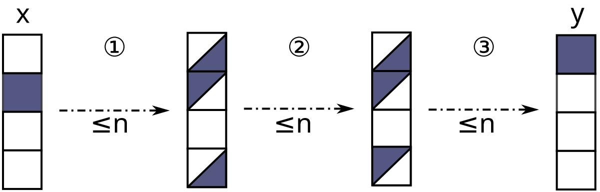
Deciding reachability requires quadratic nb of transitions

Main property:
y reachable from x ⇔ there exists a path of length ≤3n transitions

Reachability with the most permissive semantics

Deciding reachability requires quadratic nb of transitions

Main property:
y reachable from x ⇔ there exists a path of length ≤3n transitions



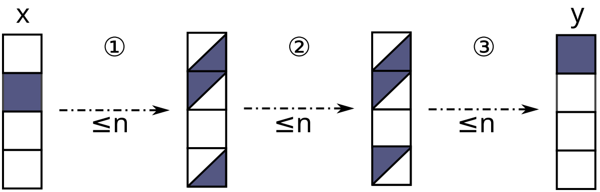
① - only transitions to "in-between" states*
② - orient towards final states
③ - converge to final states

*: some components must not be updated!

Reachability with the most permissive semantics

Deciding reachability requires quadratic nb of transitions

Main property:
y reachable from x ⇔ there exists a path of length ≤3n transitions



① - only transitions to "in-between" states*
② - orient towards final states
③ - converge to final states

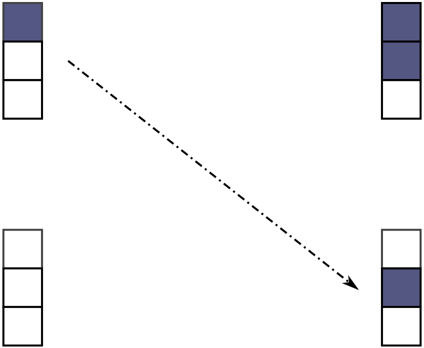
*: some components must not be updated!

NP in general
PTIME w/
locally-monotonic;
BDDs; Petri nets..

Attractors with the most permissive semantics

Attractor: smallest set of configurations closed by transitions

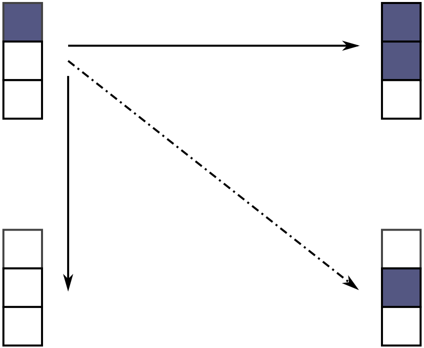
Attractors are hypercubes



Attractors with the most permissive semantics

Attractor: smallest set of configurations closed by transitions

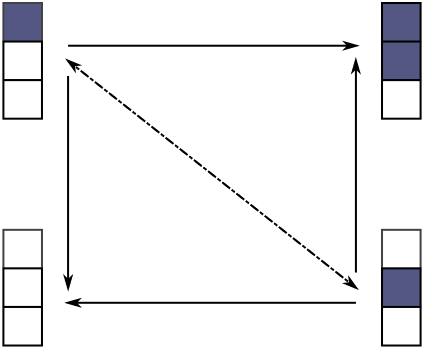
Attractors are hypercubes



Attractors with the most permissive semantics

Attractor: smallest set of configurations closed by transitions

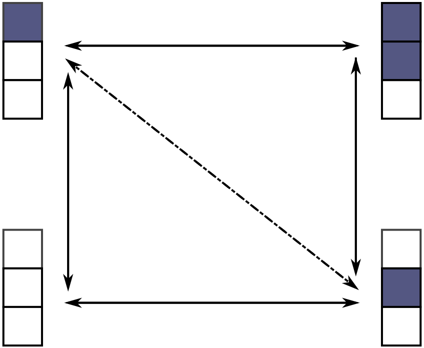
Attractors are hypercubes



Attractors with the most permissive semantics

Attractor: smallest set of configurations closed by transitions

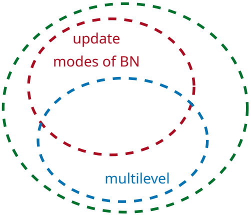
Attractors are hypercubes



Attractors of most permissive semantics = minimal trap spaces

Existence of attractor within hypercube is NP-complete

Is most permissive semantics restrictive?



Minimality of abstraction

to any "most permissive" transition,

there is corresponding multilevel transition

(work in progress w/ "most permissive" paths: non-minimal, but tricky counter-examples)

- fixpoints (stable states) are preserved (identical)
- trap spaces: known to be relevant for reasoning with attractors

  [Klarner et al in Nat. Comp. 2015] [Naldi in Front. Phys. 2018]

⇒ most permissive semantics seems still adequate to model differentiation processes !

Applications

Prototype python library + ASP (SAT) implementation
https://github.com/pauleve/mpbn

Boolean network synthesis from reachability properties
⇒ becomes NP
⇒ CaspoTS implements most permissive reachability (ASP)
https://github.com/bioasp/caspots

Computation of reachable attractors
⇒ In the order of ms for networks tested so far (~100 nodes)

WiP with most permissive semantics:

- model synthesis from differentiation data [Stéphanie Chevalier]
- prediction for cellular reprogramming

Conclusion

Update modes of Boolean networks (sync, async, etc.):

- difficult to justify (strong implications on dynamics)
- can miss important behaviours [CHP at AUTOMATA'18]

⇒ lead to reject valid models of biological systems...

- have limited tractability (model-checking, ...)

Most permissive semantics:

- correct abstraction: guarantees that adding information (multilevel, thresholds) will only remove behaviours
- simpler complexity: reachability PTIME, attractors NP

⇒ much higher tractability

Future work: most permissive for multilevel networks