

Molecules as Automata

Representing Biochemical Systems as
Collectives of Interacting Automata

Luca Cardelli

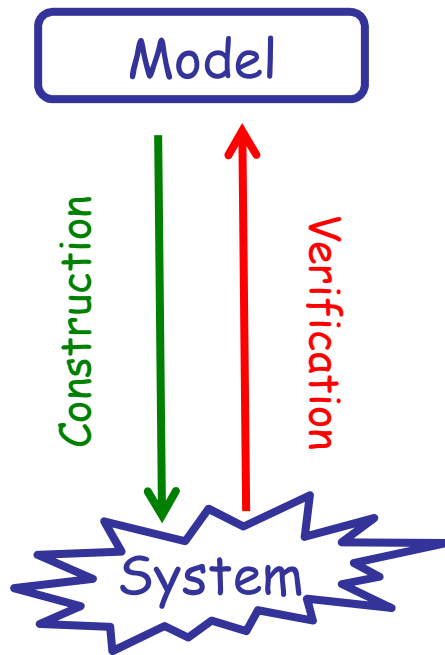
Microsoft Research

International Summer School on
Natural Computation, BNC'08
Bertinoro, 2008-09-21

<http://LucaCardelli.name>

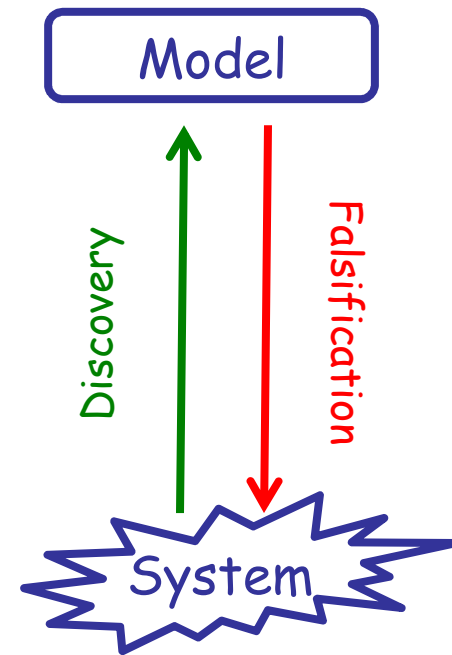
Scientific Method vs. Engineering Method

Engineering Method



Direct Engineering
(Synthetic Biology)

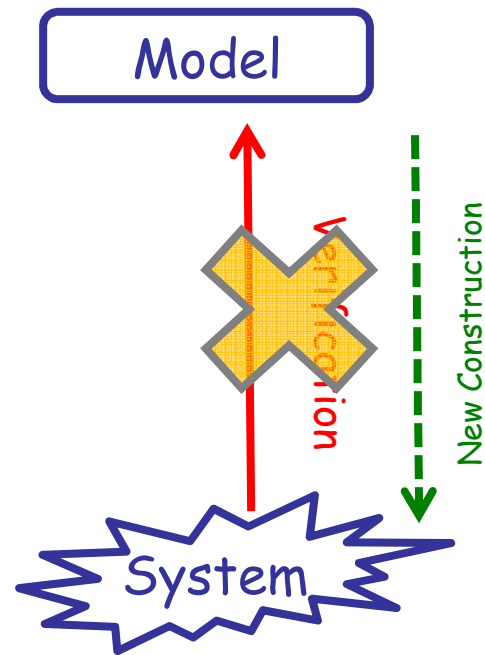
Scientific Method



Reverse Engineering
(Systems Biology)

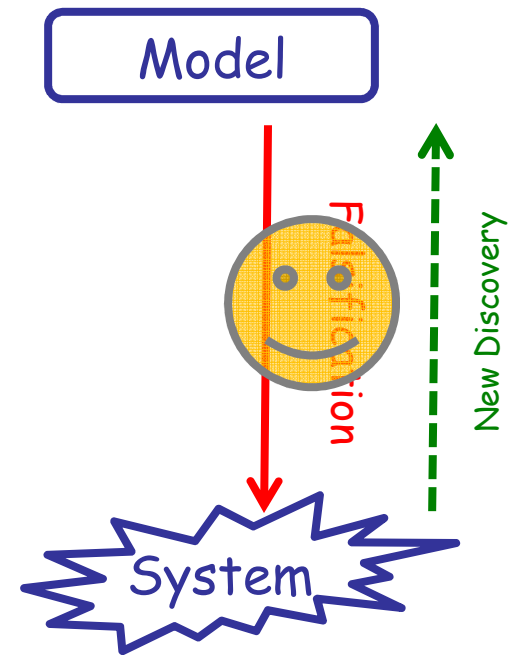
Scientific Method vs. Engineering Method

Engineering Method



Direct Engineering

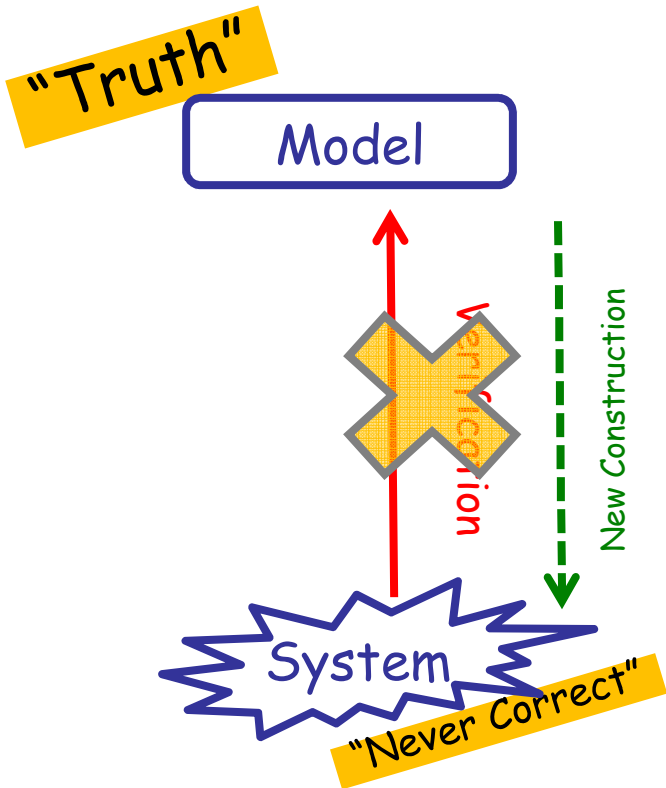
Scientific Method



Reverse Engineering

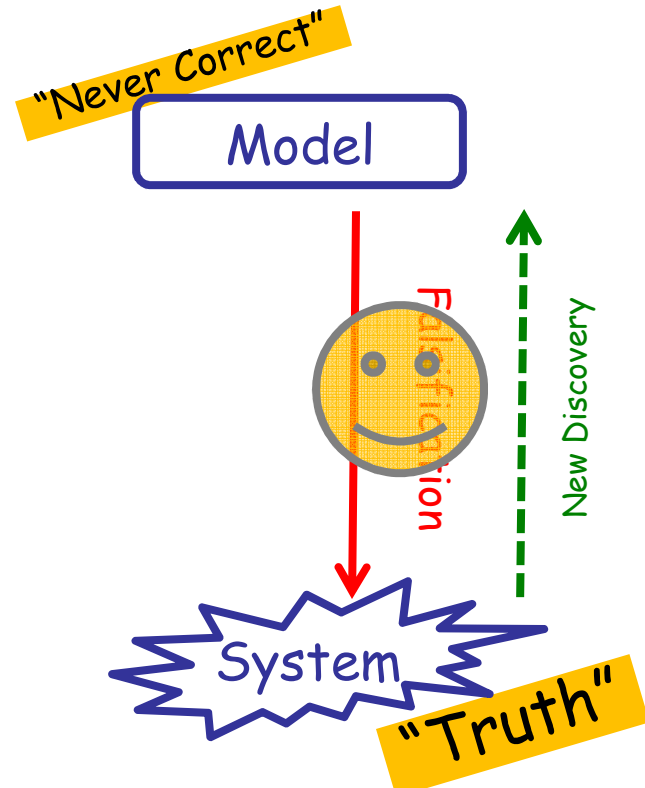
Scientific Method vs. Engineering Method

Engineering Method



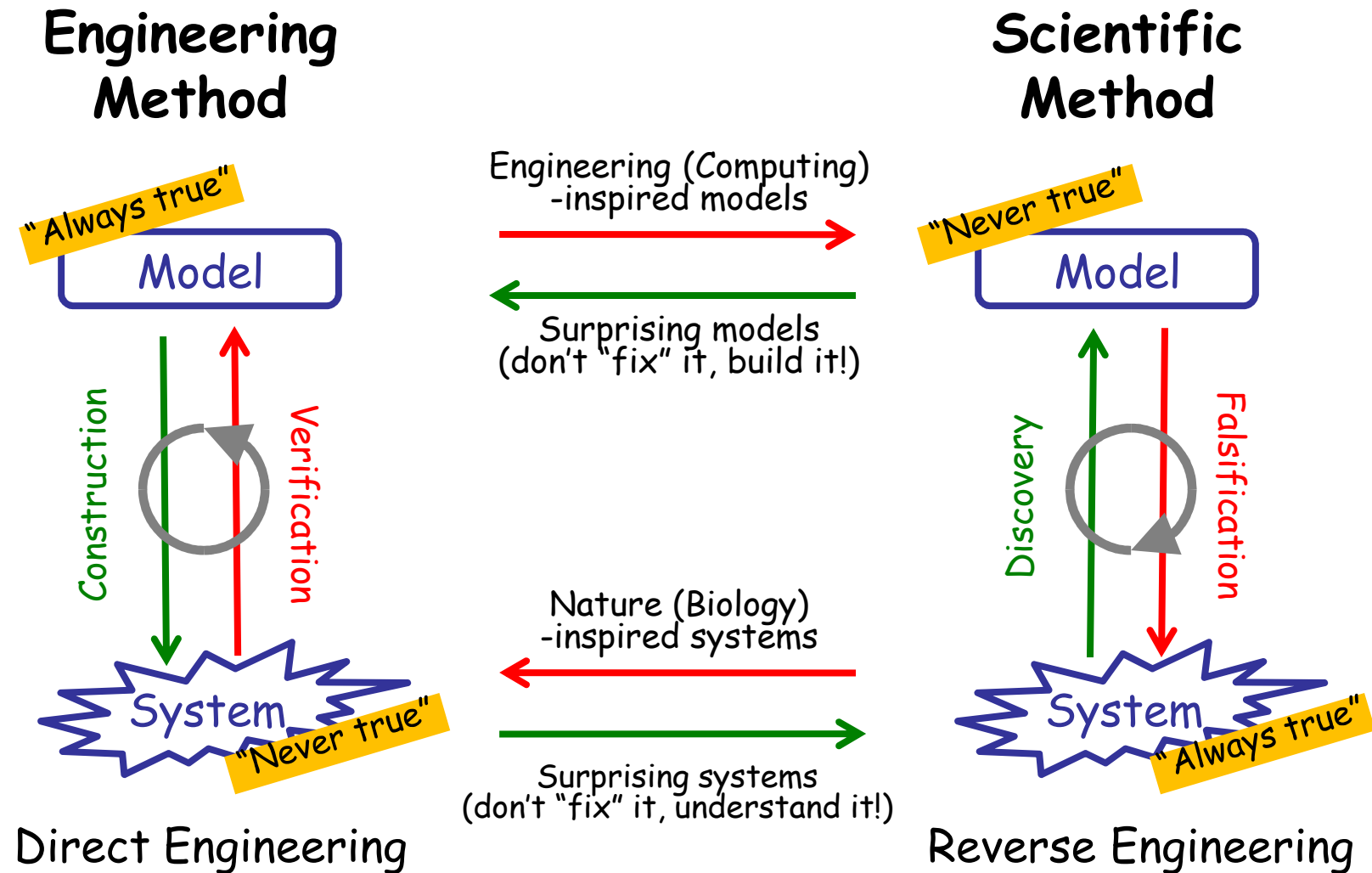
Direct Engineering

Scientific Method



Reverse Engineering

Scientific Method vs. Engineering Method

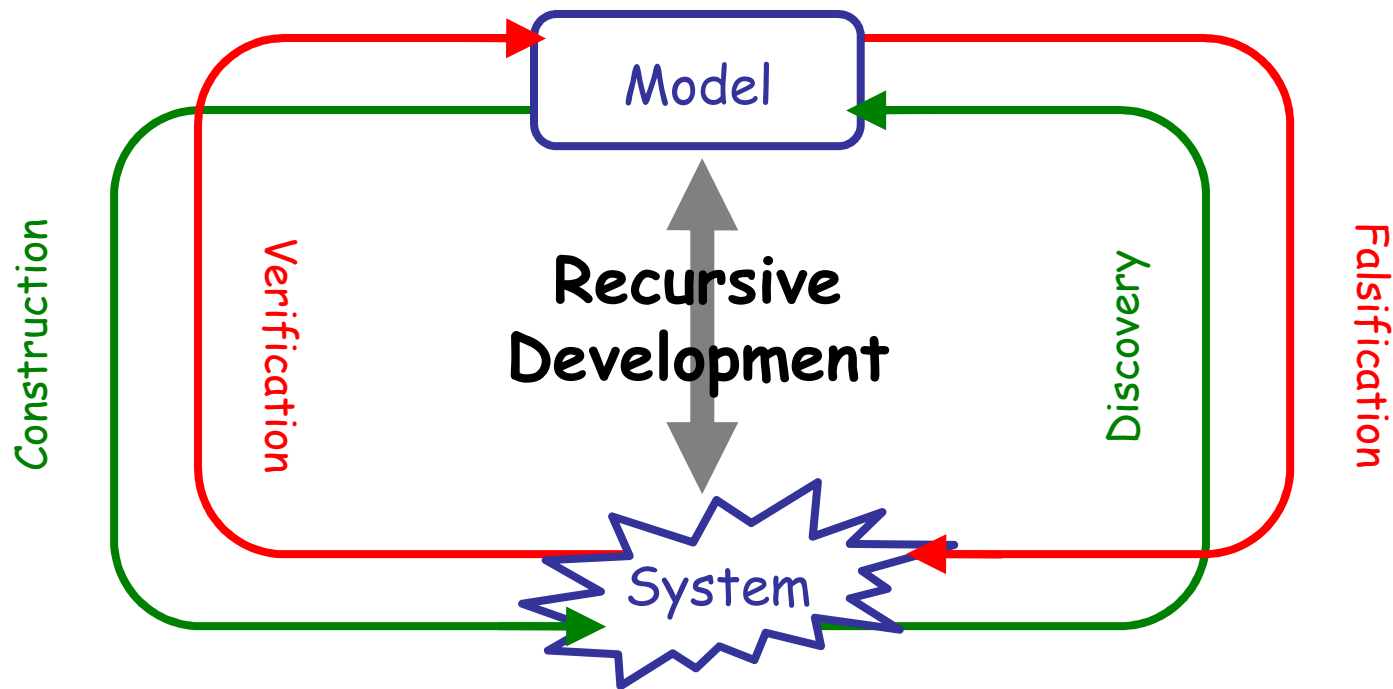


Scientific Method vs. Engineering Method

When the models and the systems are *both* too complex to *either* be the full Truth

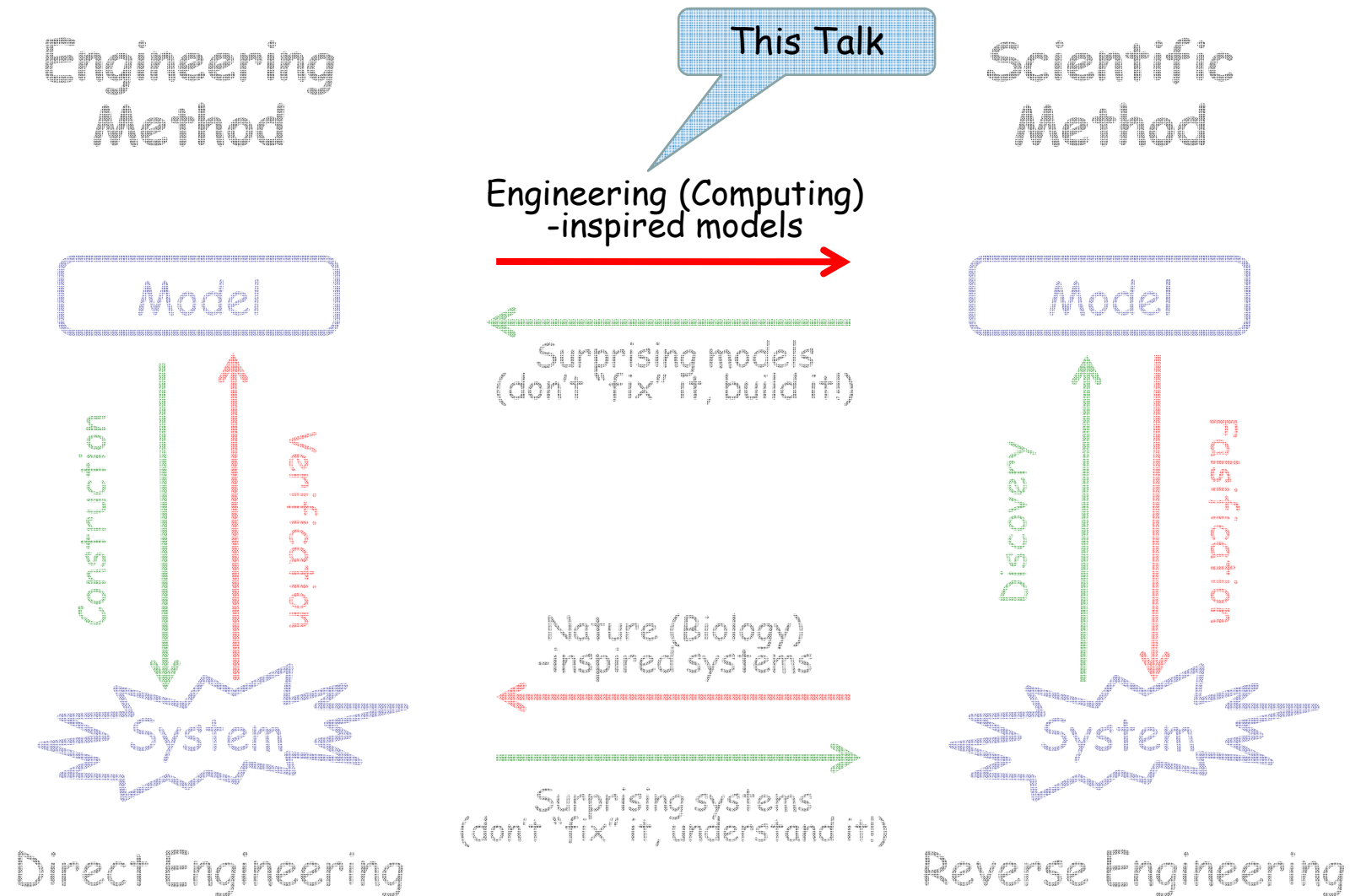
Combined Method

The models that we discover should be suitable for construction



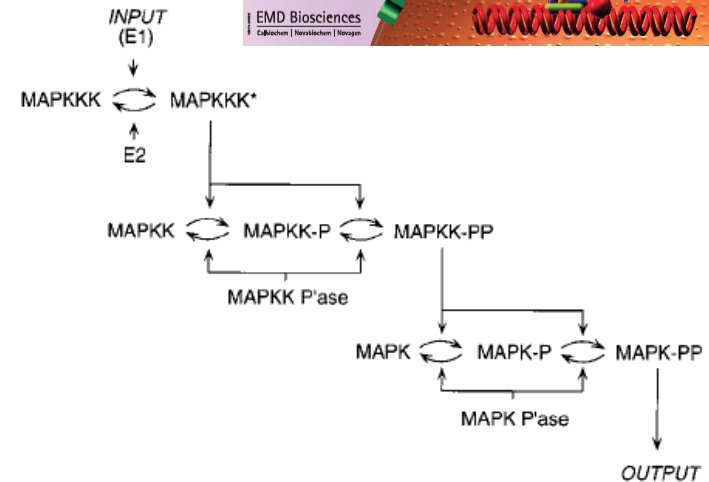
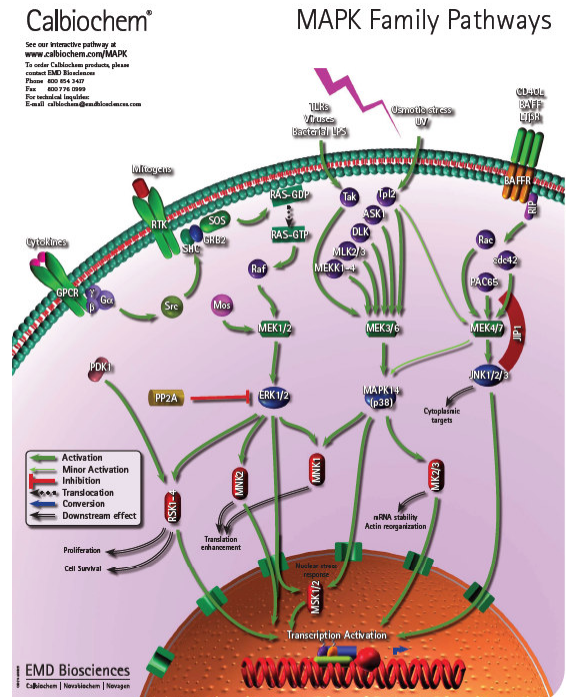
The systems that we build should be suitable for discovery

Scientific Method vs. Engineering Method



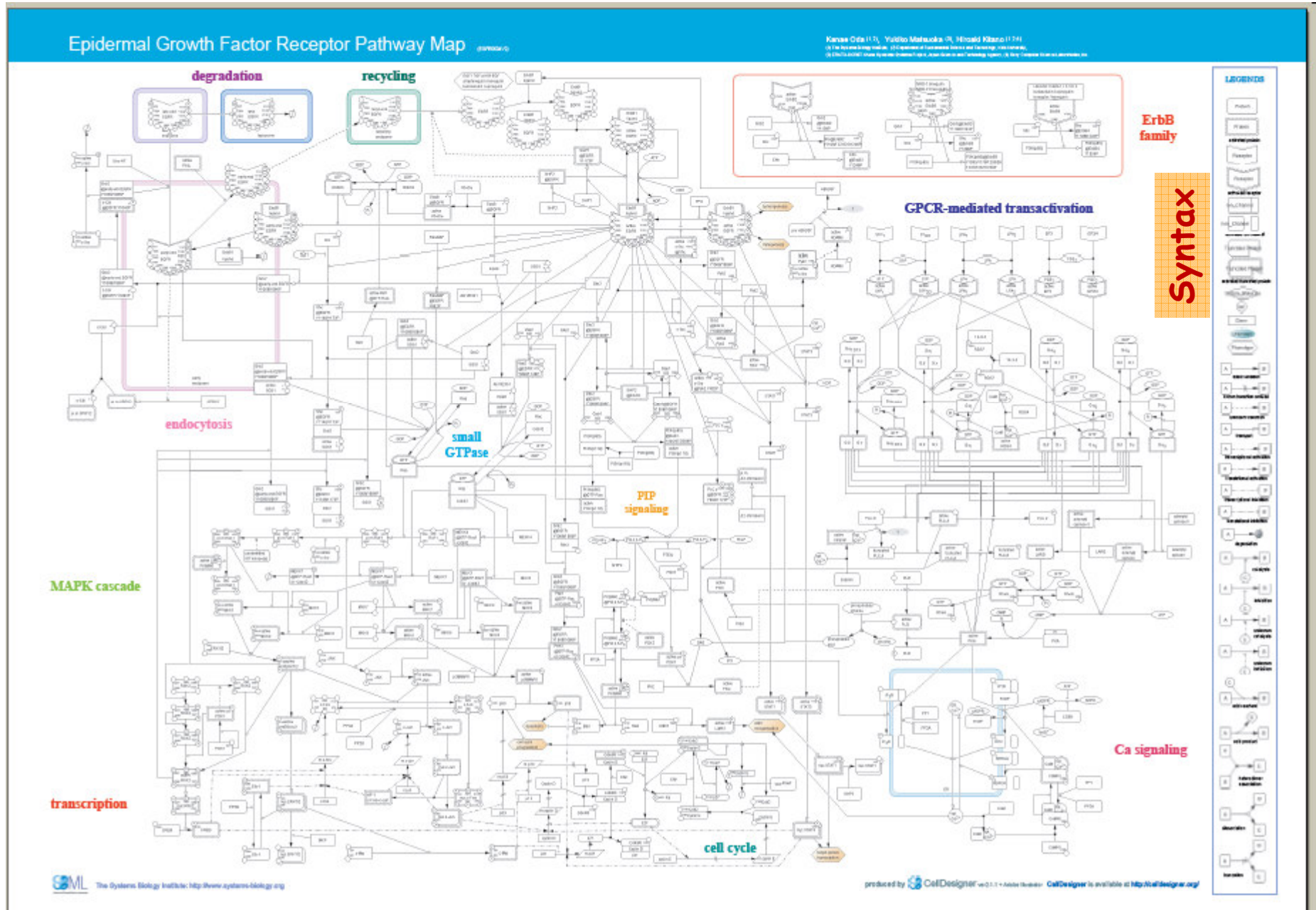
Motivation: Cells Compute

- No survival without computation!
 - Finding food
 - Avoiding predators
- How do they compute?
 - Unusual computational paradigms.
 - Proteins: do they work like electronic circuits?
 - Genes: what kind of software is that?
- Signaling networks
 - Clearly "information processing"
 - They are "just chemistry": molecule interactions
 - But what are their principles and algorithms?
- Complex, higher-order interactions
 - MAPKKK = MAP Kinase Kinase Kinase: that which operates on that which operates on that which operates on protein.
- General models of biological computation
 - What are the appropriate ones?



Ultrasensitivity in the mitogen-activated protein cascade,
Chi-Ying F. Huang and James E. Ferrell, Jr., 1996, *Proc. Natl. Acad. Sci. USA*, 93, 10078-10083.

Discrete State Transitions



Modeling Approach

- We believe that {petri nets, process algebra, term rewriting, multiagent systems} are {better, complementary} for modeling biological systems than {SBML, Kohn charts, chemical reactions, ODEs}.
- We take a paper from the literature (usually ODEs or chemical reactions) and “code it up” in e.g. Petri nets.
- How do we know that’s the “same system” ? How do we convince mathematical biologists that we are doing the “right thing”?

(Macro-) Molecules as (Interacting) Automata

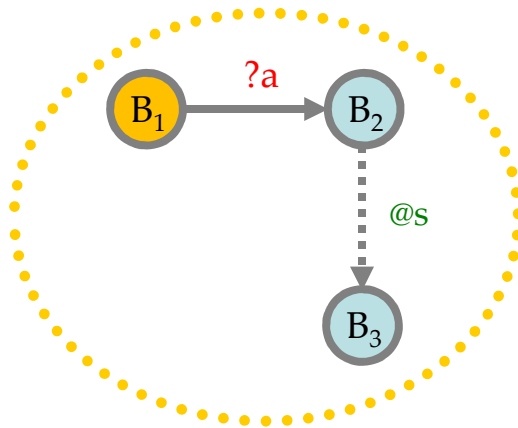
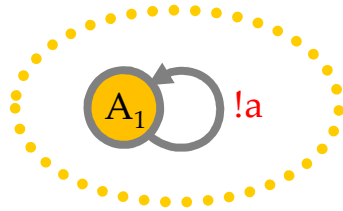
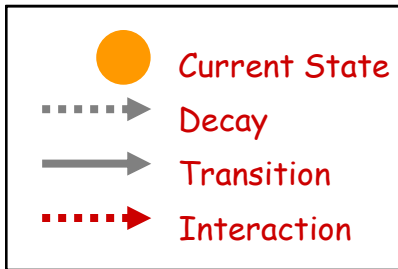
Process Algebra

[Hoare, Milner, Pnueli, etc.]

- **Reactive systems** (living organisms, computer networks, operating systems, ...)
 - Math is based on *entities that react/interact with their environment* ("*processes*"), not on *functions* from domains to codomains.
- **Concurrent**
 - **Events** (reactions/interactions) happen concurrently and asynchronously, not sequentially like in function composition.
- **Stochastic**
 - Or probabilistic, or nondeterministic, but is never about deterministic system evolution.
- **Stateful**
 - Each concurrent activity ("process") maintains its own local state, as opposed to stateless functions from inputs to outputs.
- **Discrete**
 - Evolution through **discrete transitions** between **discrete states**, not incremental changes of continuous quantities.
- **Kinetics of interaction**
 - An "**interaction**" is anything that moves a system from one state to another.

Interacting Automata

Legend



A_1 is a *state*

a is a *channel* i.e. a named *interaction interface* (e.g. a surface patch)

$?a, !a$ indicate any *complementarity* of interaction (e.g. charge)

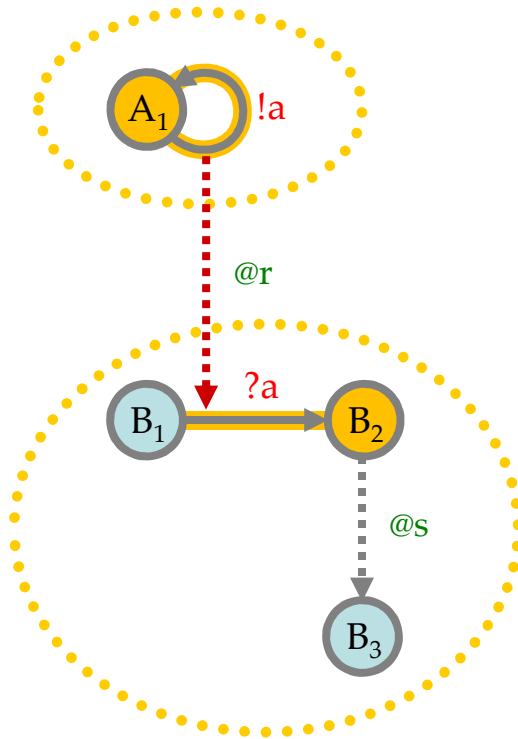
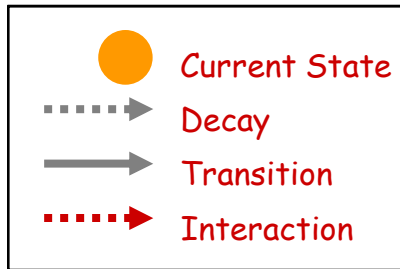
$?a, !a$ indicate *complementary actions*,

$@r, @s$ are rates

Kinetic laws:

Interacting Automata

Legend

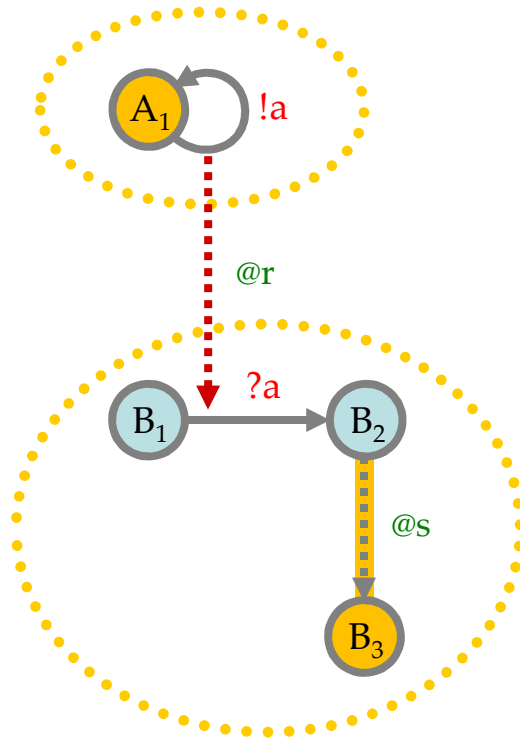
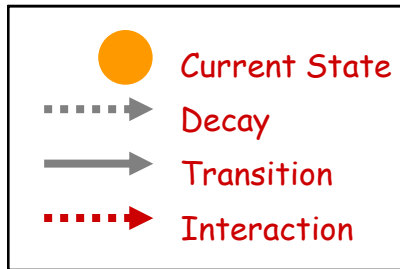


- A_1 is a *state*
- a is a *channel* i.e. a named *interaction interface* (e.g. a surface patch)
- $?,!$ indicate any *complementarity* of interaction (e.g. charge, shape)
- $?a, !a$ indicate *complementary actions*, joined by an interaction arrow - - - - ->
- $@r, @s$ are rates

Kinetic laws: ***Two complementary actions may result in an interaction.***

Interacting Automata

Legend



A_1 is a state

a is a channel i.e. a named interaction interface (e.g. a surface patch)

$?,!$ indicate any complementarity of interaction (e.g. charge)

$?a, !a$ indicate complementary actions, joined by an interaction arrow - - - - ->

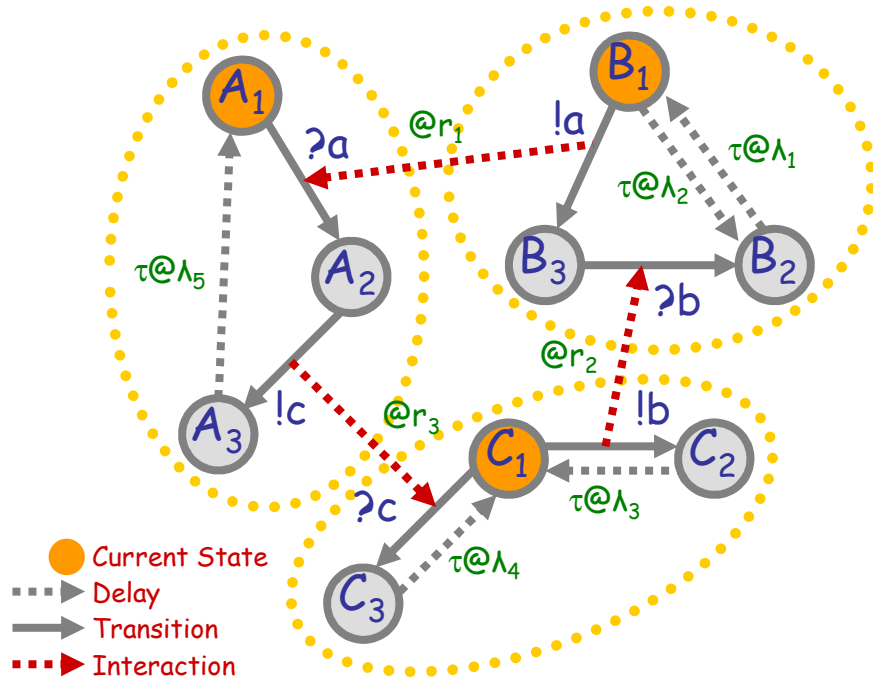
$@r, @s$ are rates

Kinetic laws:

Two complementary actions may result in an interaction.

A decay may happen spontaneously.

Interacting Automata



Interactions have rates. Actions DO NOT have rates.

The equivalent process algebra model

```

new a@r1
new b@r2
new c@r3
    } Communication channels

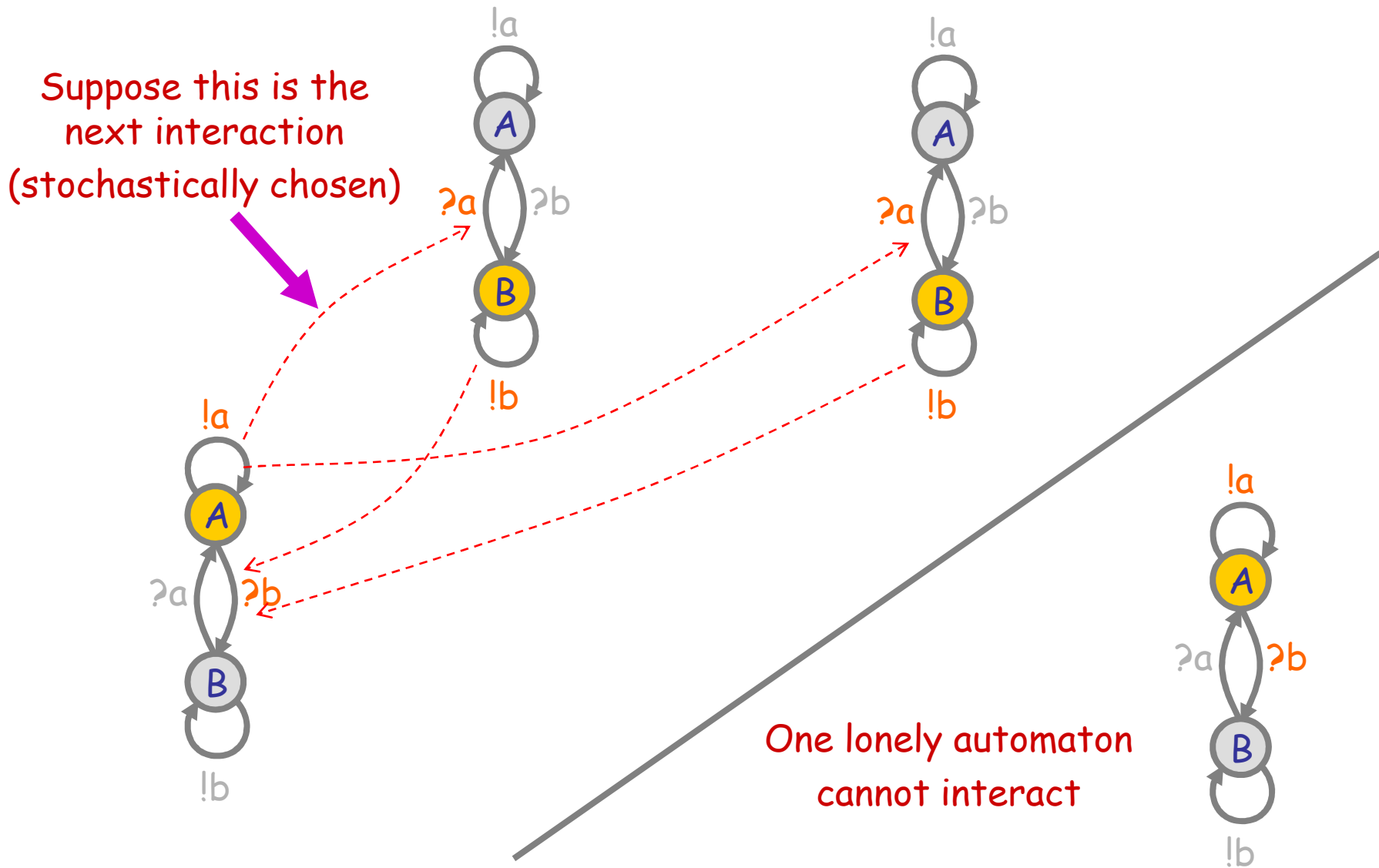
A1 = ?a; A2
A2 = !c; A3
A3 = τ@λ5; A1
    } Automata

B1 = τ@λ2; B2 + !a; B3
B2 = τ@λ1; B1
B3 = ?b; B2

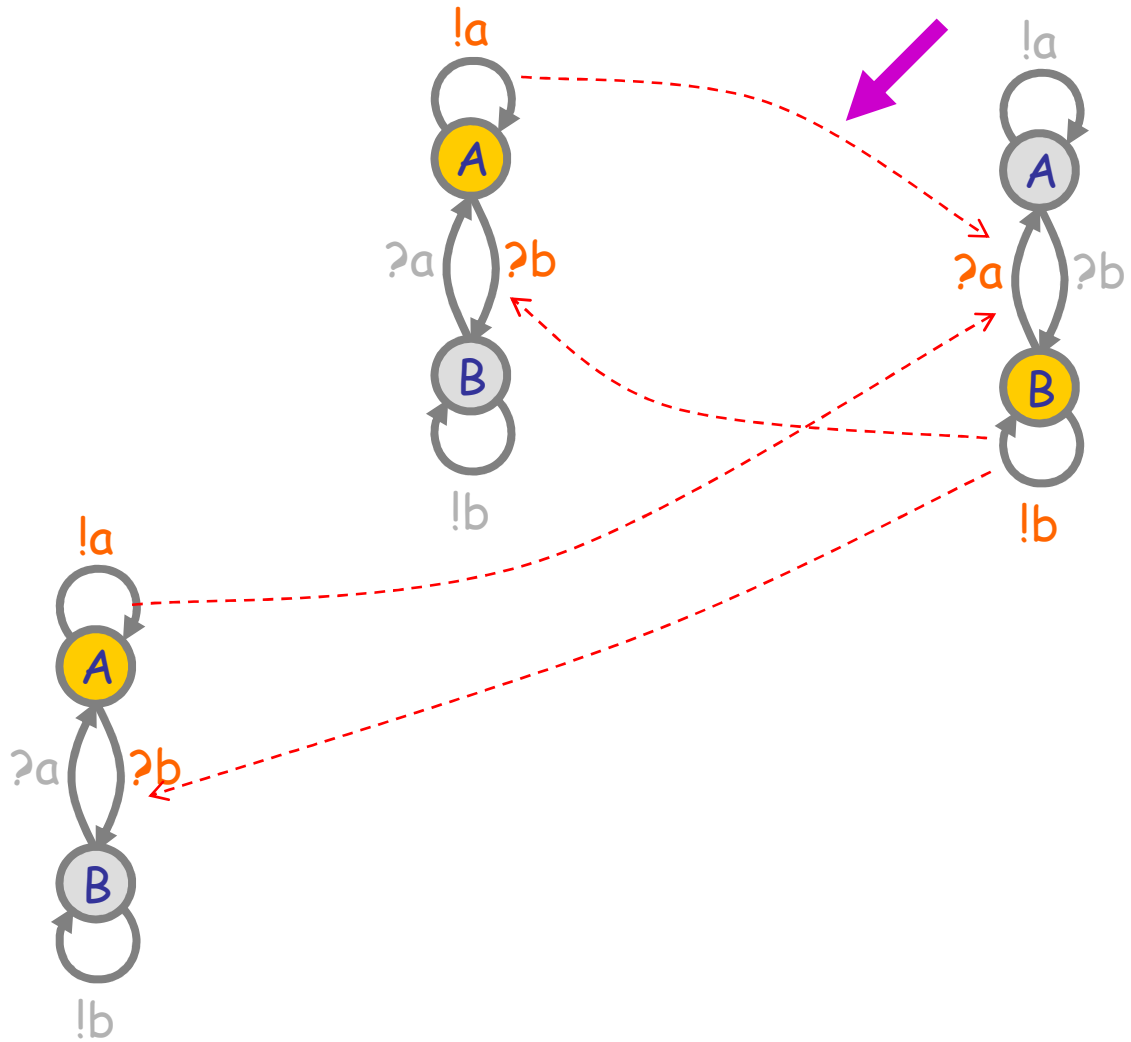
C1 = !b; C2 + ?c; C3
C2 = τ@λ3; C1
C3 = τ@λ4; C2

A1 | B1 | C1
    } The system and initial state
    
```

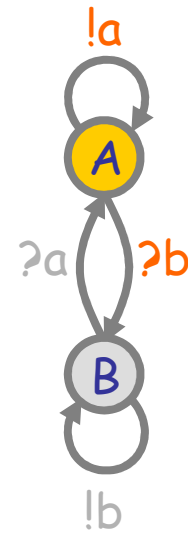
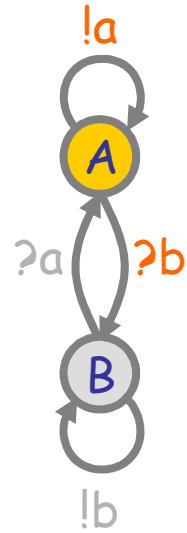
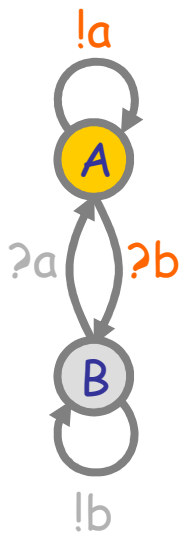

Interactions in a Population



Interactions in a Population

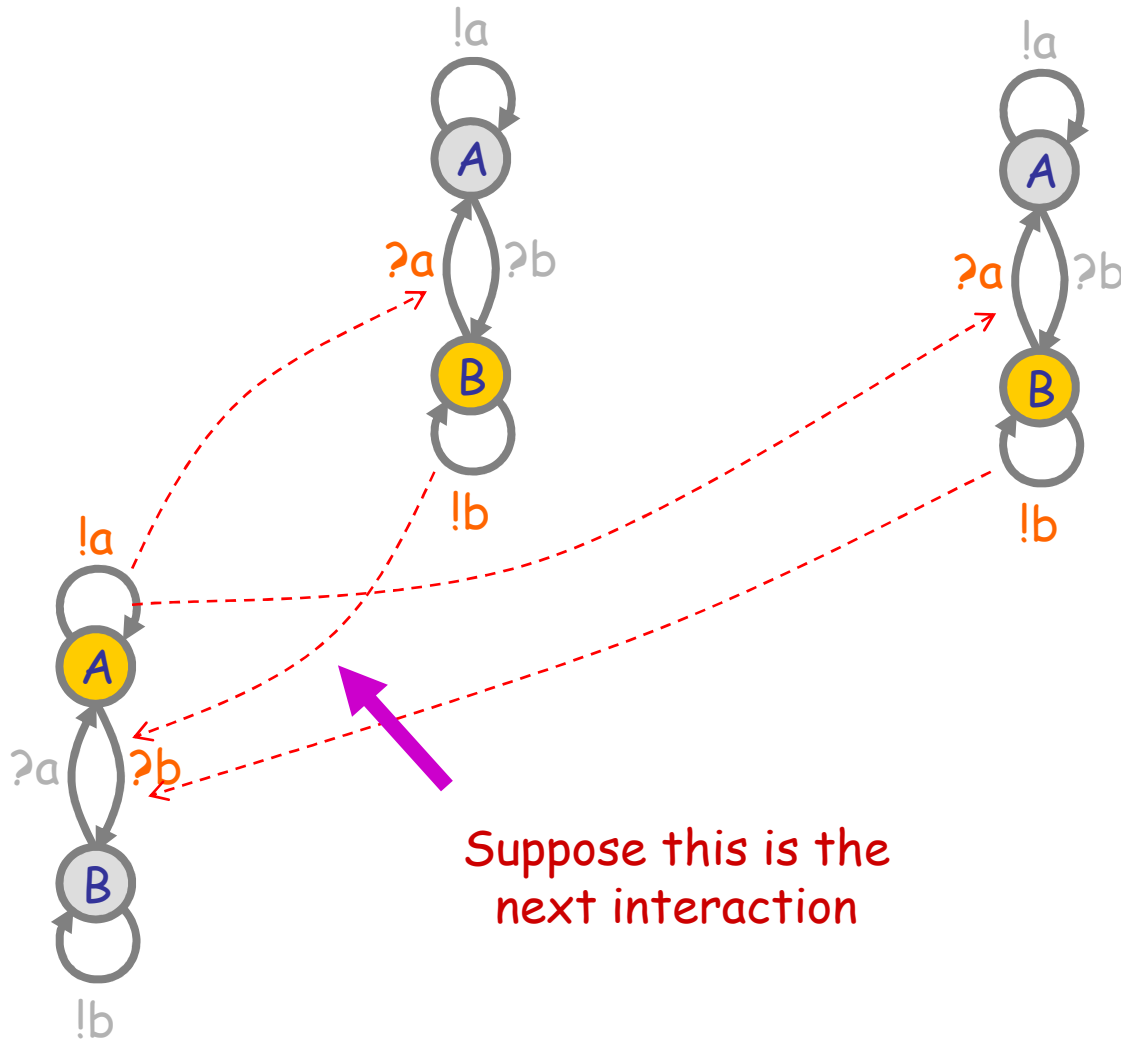


Interactions in a Population

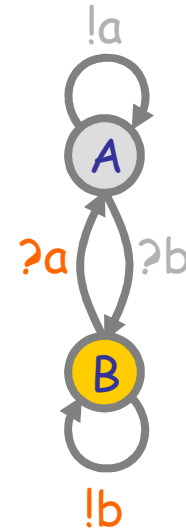
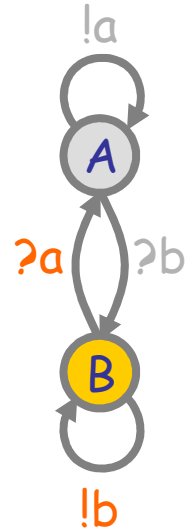
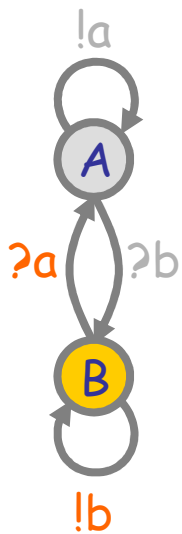


All-A stable population

Interactions in a Population (2)



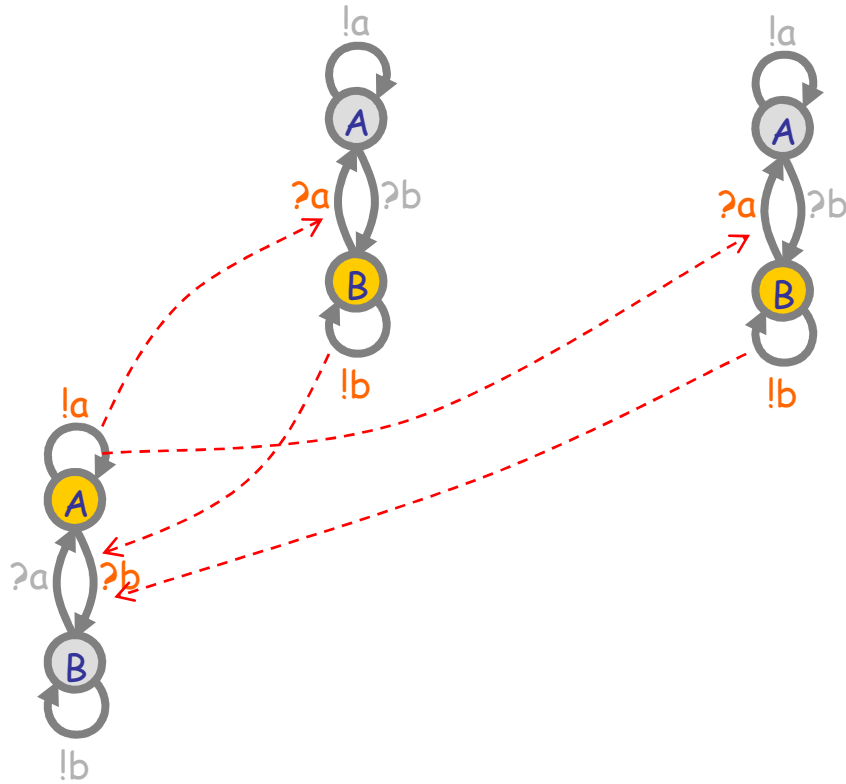
Interactions in a Population (2)



All-B stable population

Nondeterministic population behavior ("multistability")

CTMC Semantics



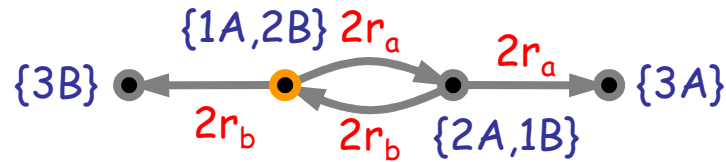
CTMC
(homogeneous) Continuous Time Markov Chain

- directed graph with no self loops
- nodes are system states
- arcs have transition rates

Probability of holding in state A:

$$\Pr(H_A > t) = e^{-rt}$$

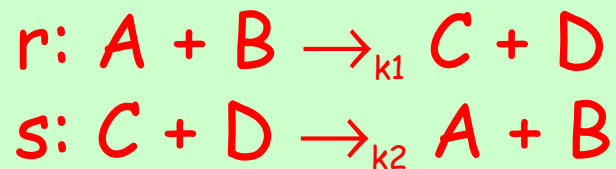
in general, $\Pr(H_A > t) = e^{-Rt}$ where R is the sum of all the exit rates from A



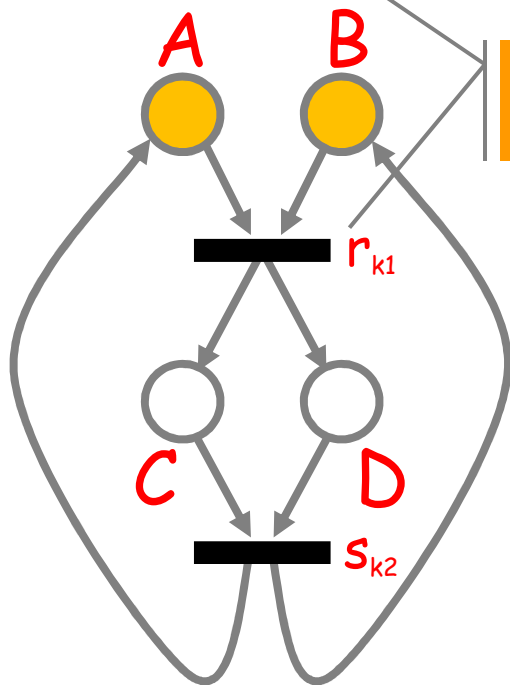
CTMC

Reactions vs. Components

Says what "A" does.



Does A become C or D?

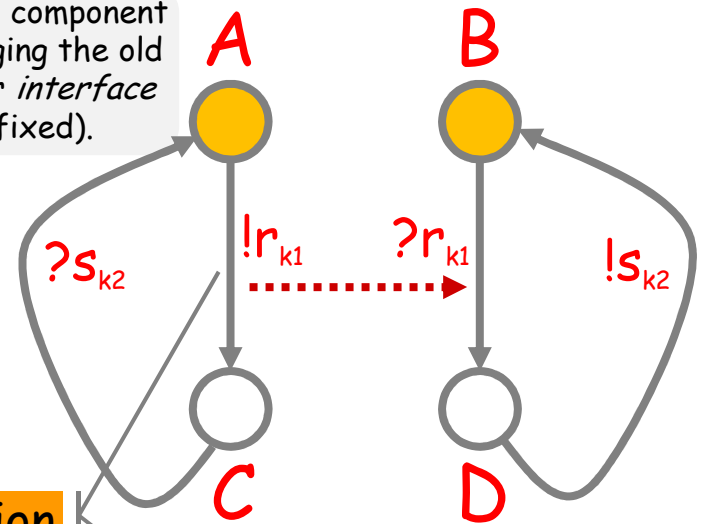


Reaction oriented

1 line per reaction

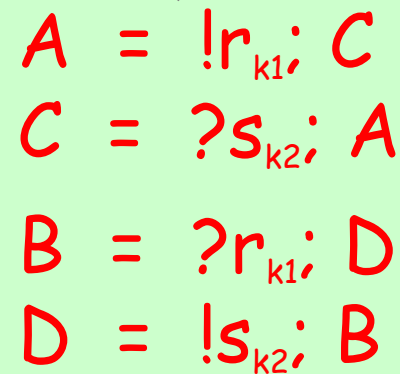
Says what "A" is.

Can add a new component without changing the old ones (if their interface remains fixed).



Interaction oriented

1 line per component



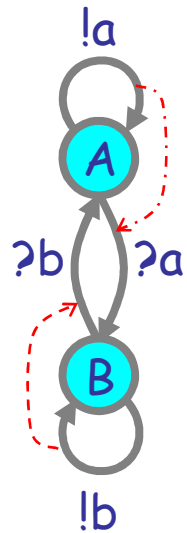
A becomes C not D!

The same "state space"

CTMC

Groupies and Celebrities

Groupies and Celebrities



Celebrity

(does not want to be like somebody else)

```
directive sample 1.0 1000
```

```
directive plot A(); B()
```

```
new a@1.0:chan()
```

```
new b@1.0:chan()
```

```
let A() = do !a; A() or ?a; B()
```

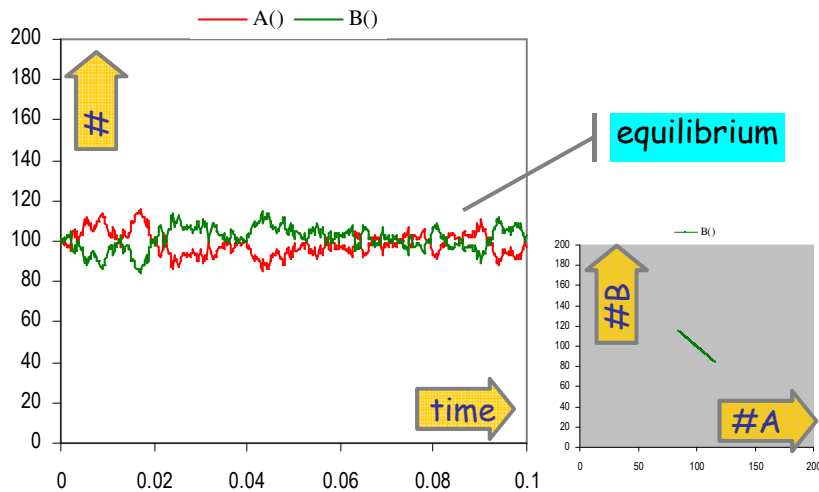
```
and B() = do !b; B() or ?b; A()
```

```
run 100 of (A() | B())
```

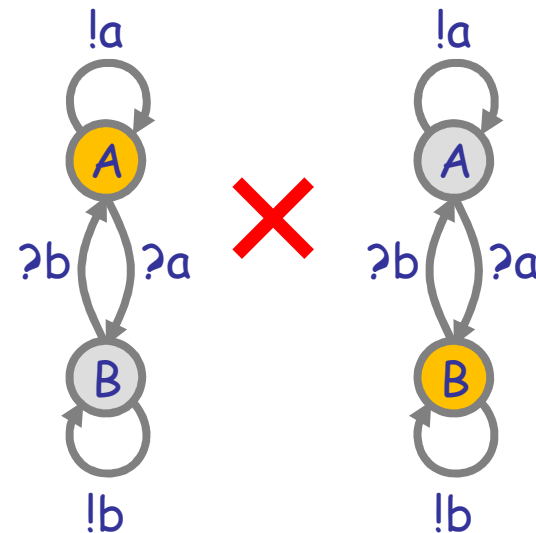
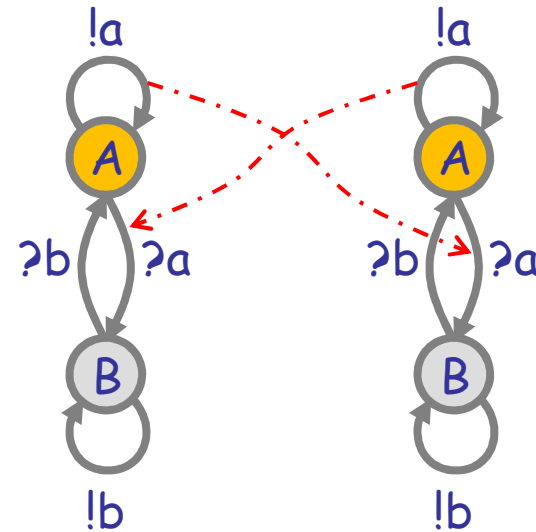
a@1.0

b@1.0

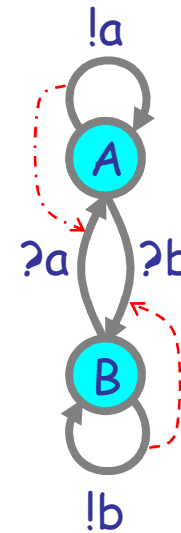
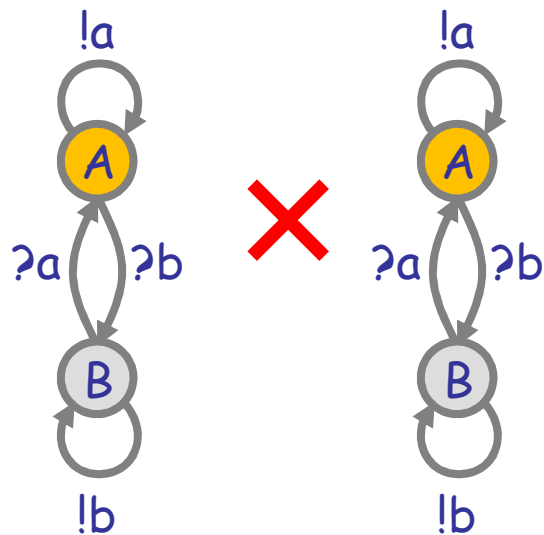
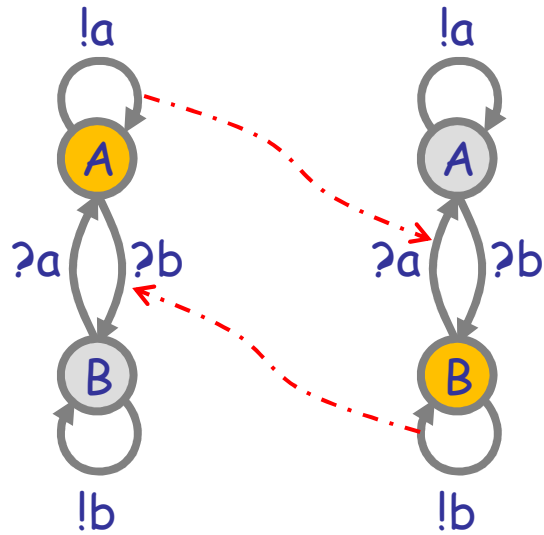
A stochastic collective of celebrities:



Stable because as soon as a A finds itself in the majority, it is more likely to find somebody in the same state, and hence change, so the majority is weakened.



Groupies and Celebrities



Groupie
(wants to be like somebody different)

```
directive sample 1.0 1000
directive plot A(); B()

new a@1.0:chan()
new b@1.0:chan()

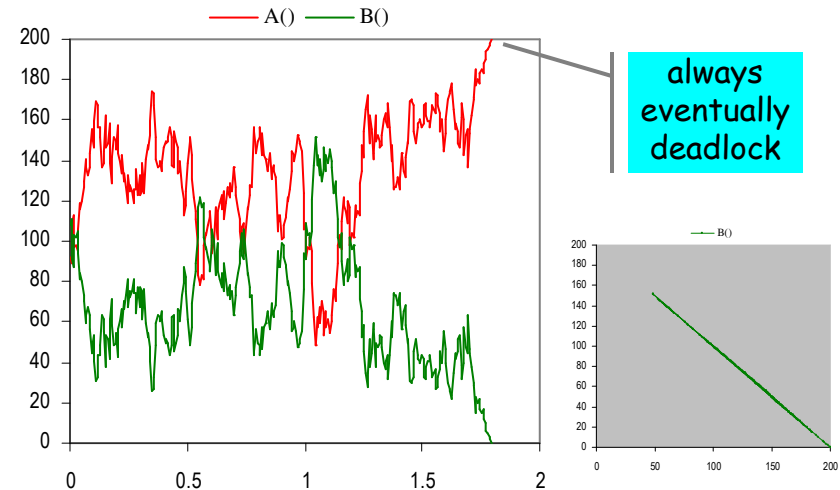
let A() = do !a; A() or ?b; B()
and B() = do !b; B() or ?a; A()

run 100 of (A() | B())
```

a@1.0

b@1.0

A stochastic collective of groupies:

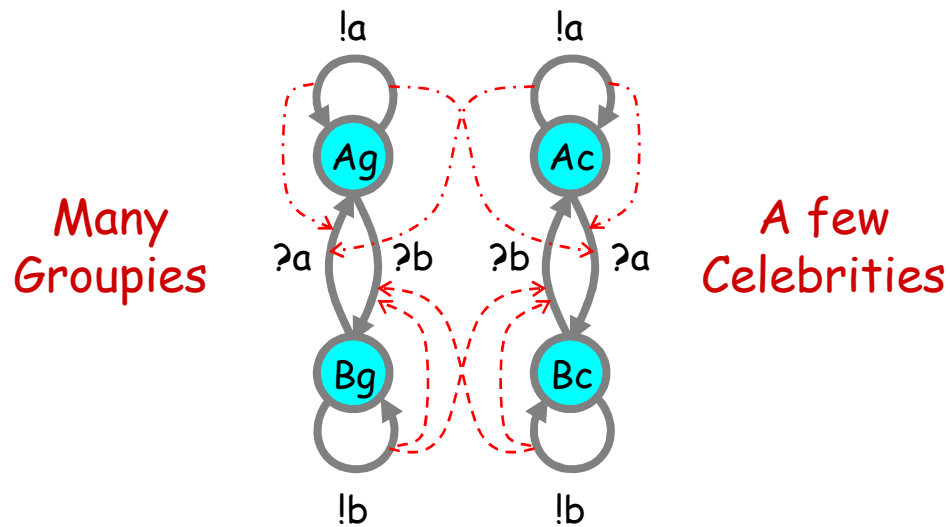


always eventually deadlock

Unstable because within an A majority, an A has difficulty finding a B to emulate, but the few B's have plenty of A's to emulate, so the majority may switch to B. Leads to deadlock when everybody is in the same state and there is nobody different to emulate.

Both Together

A way to break the deadlocks: Groupies with just a few Celebrities



```
directive sample 10.0
directive plot Ag(); Bg(); Ac(); Bc()

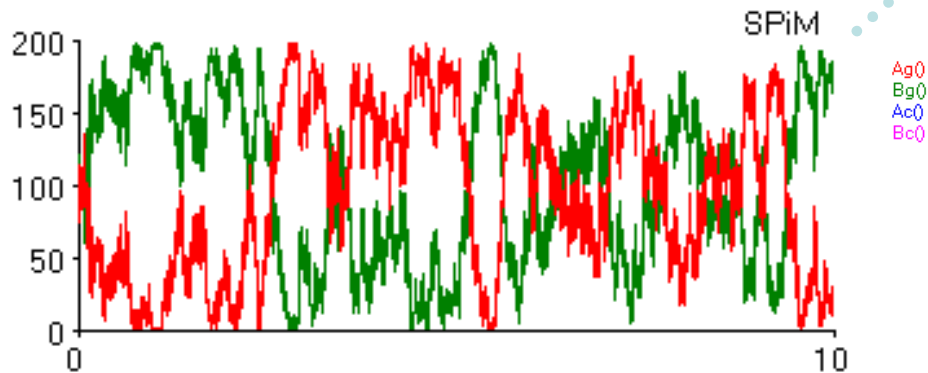
new a@1.0:chan()
new b@1.0:chan()

let Ac() = do !a; Ac() or ?a; Bc()
and Bc() = do !b; Bc() or ?b; Ac()

let Ag() = do !a; Ag() or ?b; Bg()
and Bg() = do !b; Bg() or ?a; Ag()

run 1 of Ac()
run 100 of (Ag() | Bg())
```

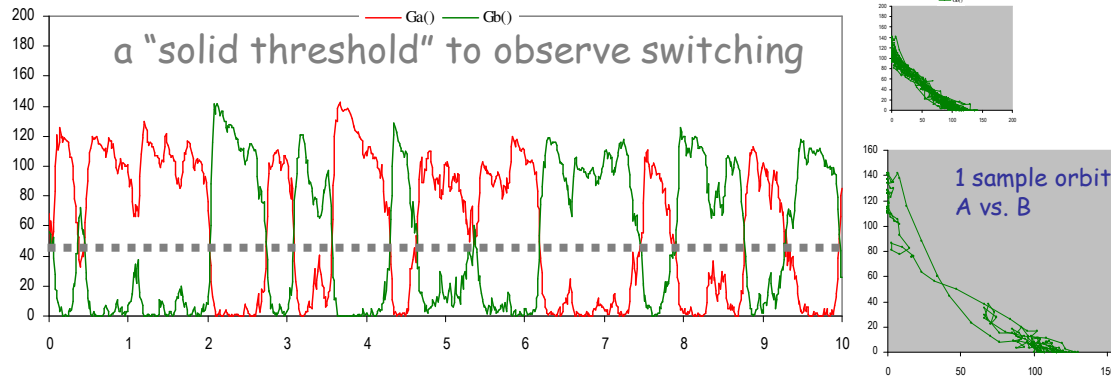
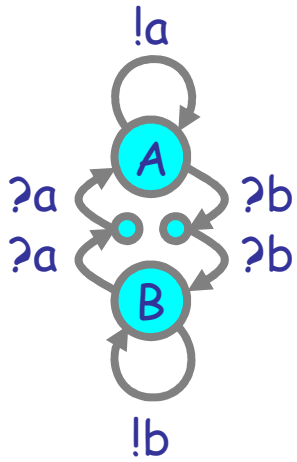
never
deadlock



A tiny bit of
"noise" can make a
huge difference

Hysteric Groupies

We can get more regular behavior from groupies if they "need more convincing", or "hysteresis" (history-dependence), to switch states.



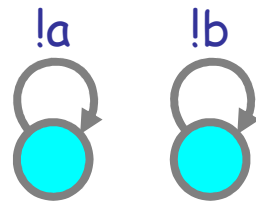
```
directive sample 10.0 1000
directive plot Ga(); Gb()

new a@1.0:chan()
new b@1.0:chan()

let Ga() = do !a; Ga() or ?b; ?b; Gb()
and Gb() = do !b; Gb() or ?a; ?a; Ga()

let Da() = !a; Da()
and Db() = !b; Db()

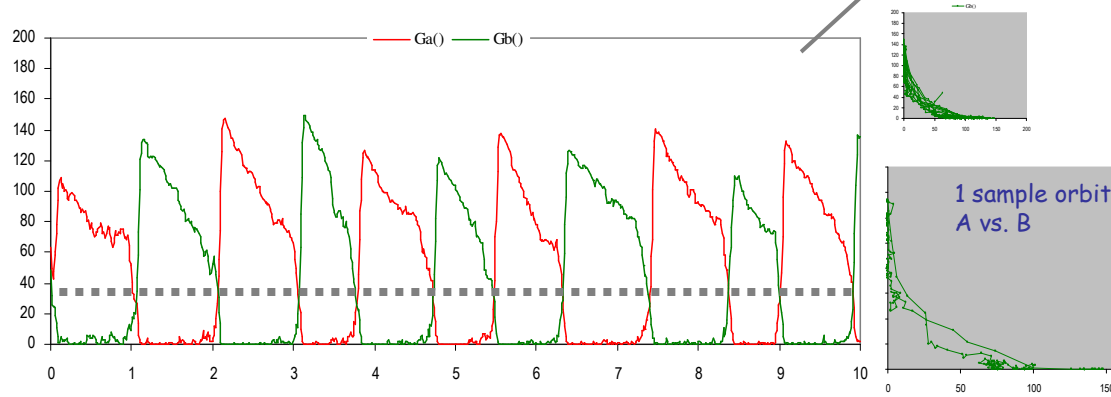
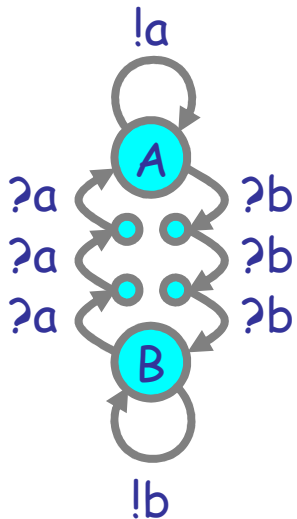
run 100 of (Ga() | Gb())
run 1 of (Da() | Db())
```



(With doping to break deadlocks)

N.B.: It will not oscillate without doping (noise)

"regular" oscillation



```
directive sample 10.0 1000
directive plot Ga(); Gb()

new a@1.0:chan()
new b@1.0:chan()

let Ga() = do !a; Ga() or ?b; ?b; ?b; Gb()
and Gb() = do !b; Gb() or ?a; ?a; ?a; Ga()

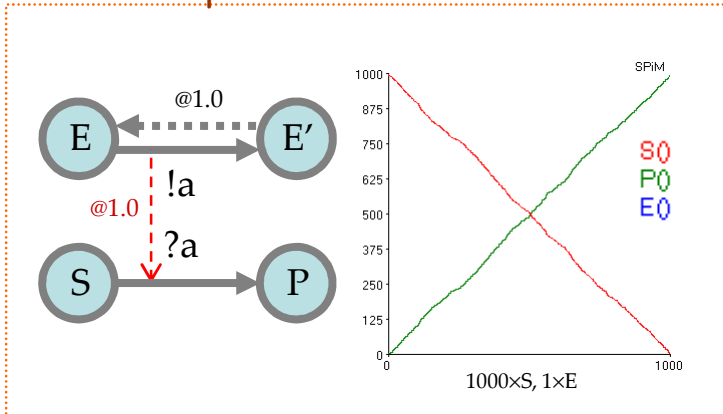
let Da() = !a; Da()
and Db() = !b; Db()

run 100 of (Ga() | Gb())
run 1 of (Da() | Db())
```

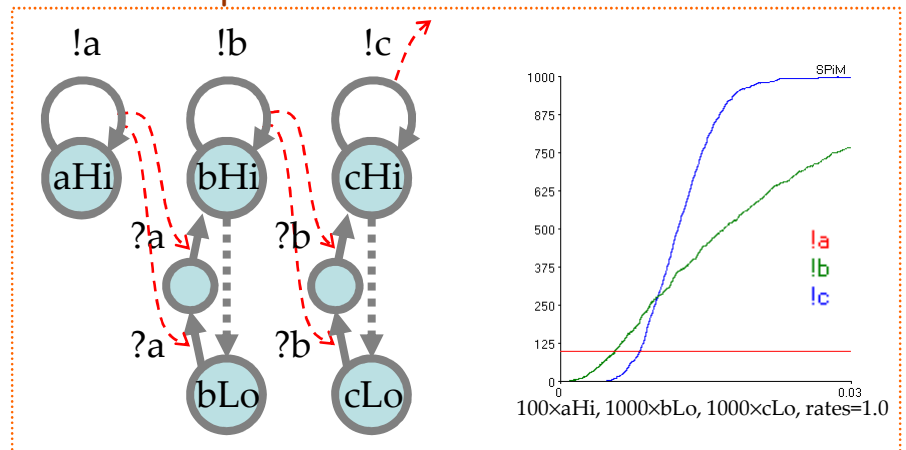


Some Devices

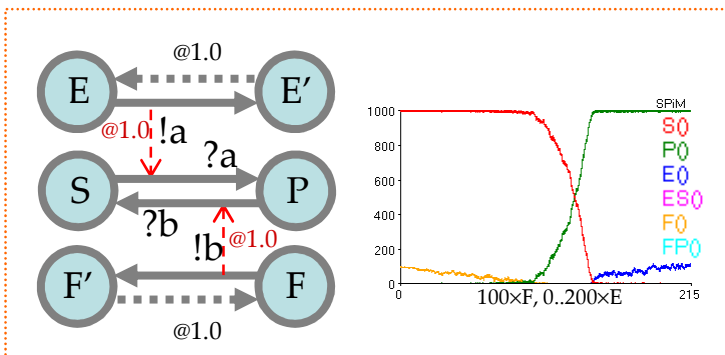
Linear Pump



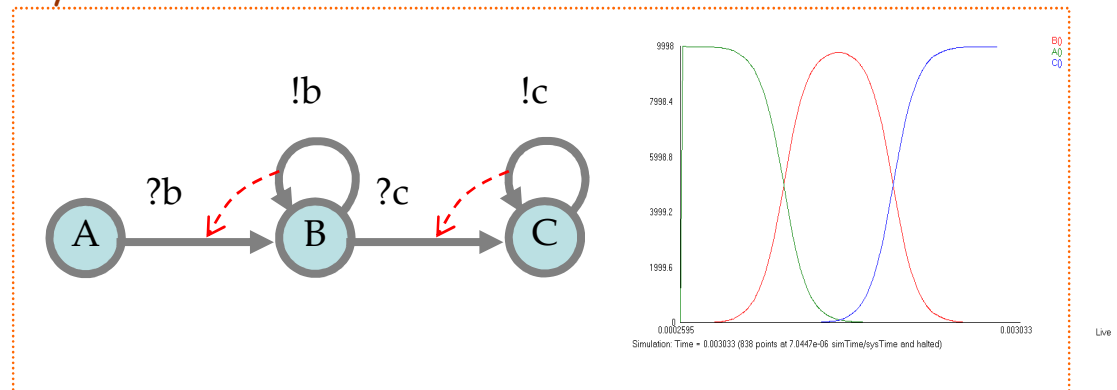
Cascade Amplifier



Ultrasensitive Switch

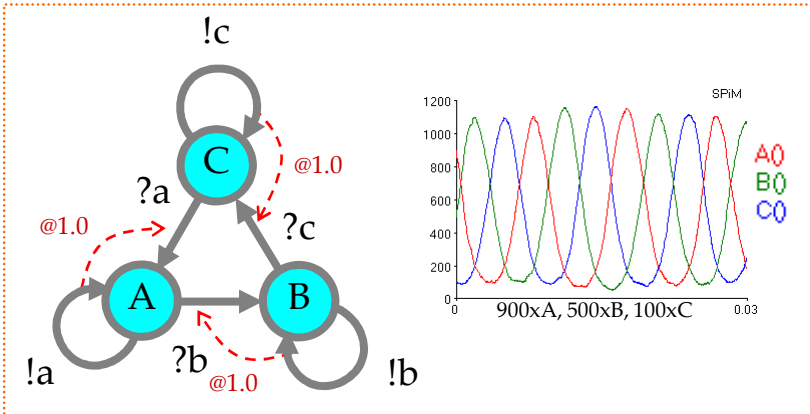


Symmetric Wave Generator

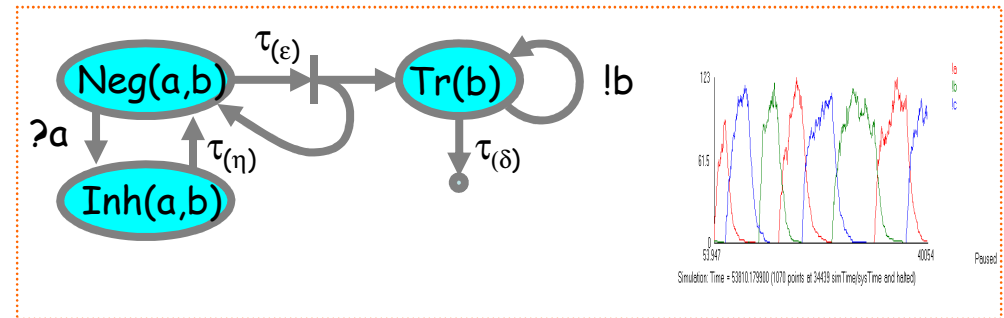


More Devices

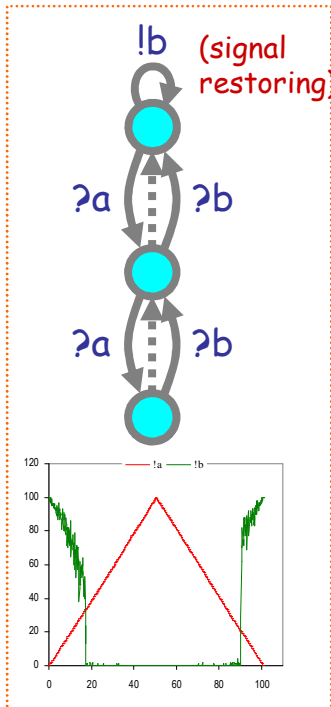
Oscillator



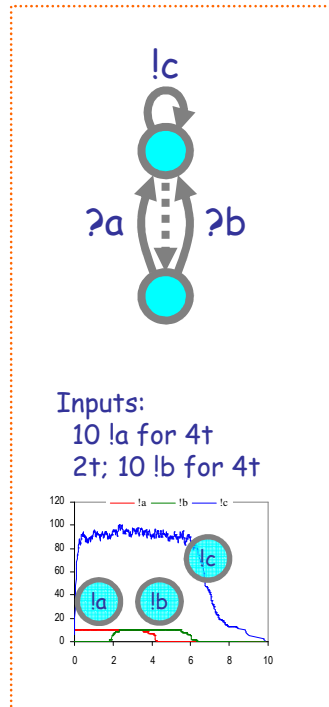
Repressilator (1 of 3 similar gates)



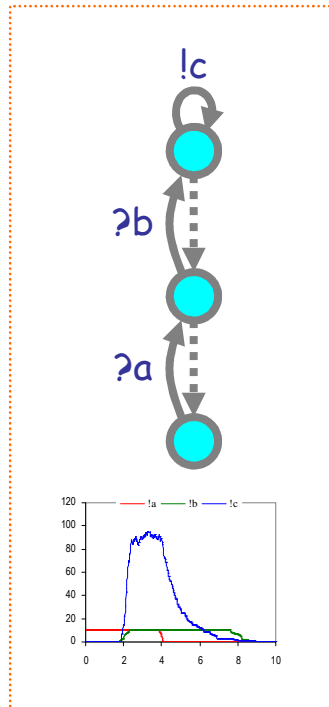
b = not a



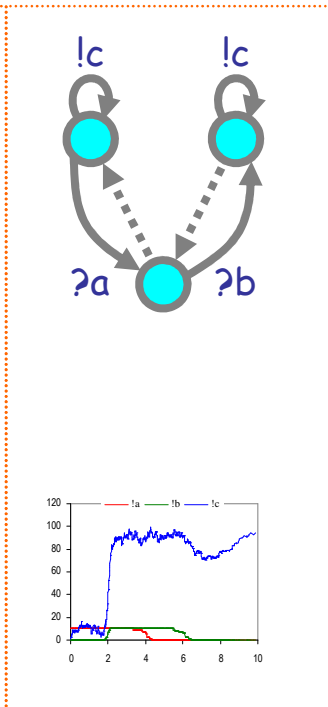
c = a or b



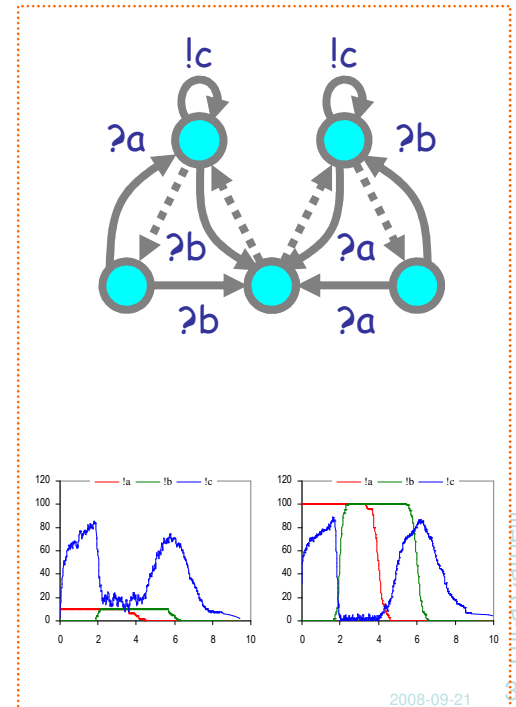
c = a and b



c = a imply b

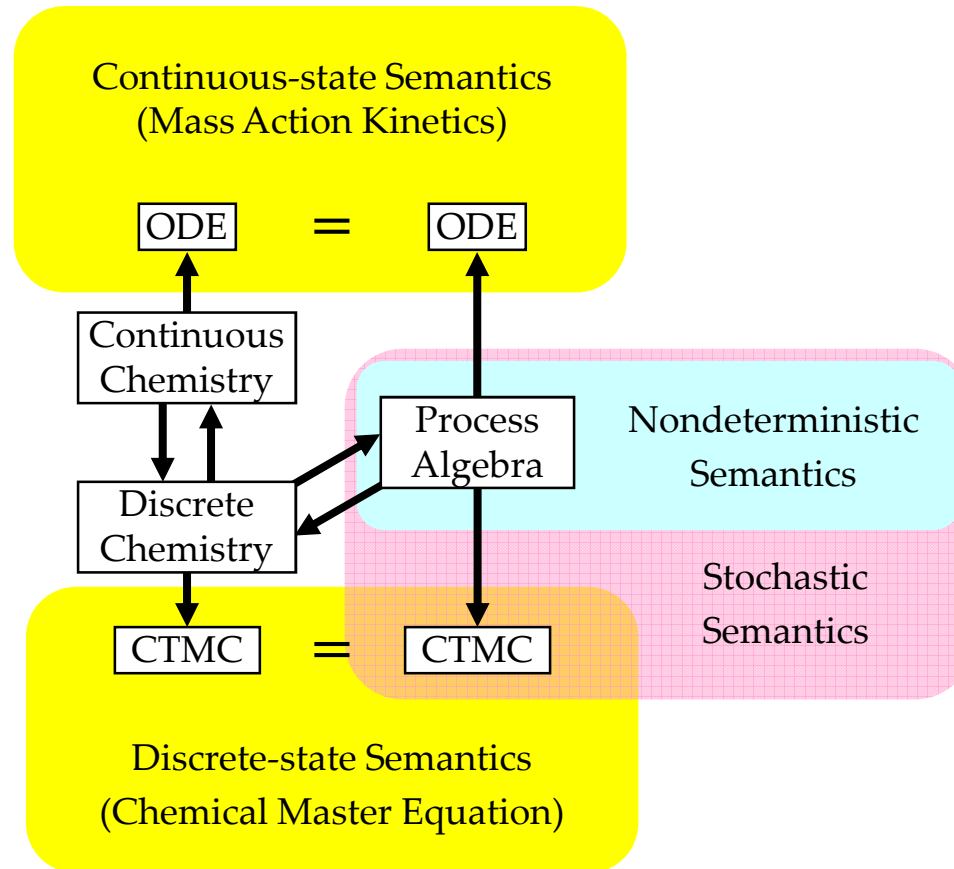


c = a xor b



Semantics of Collective Behavior

The Two Semantic Sides of Chemistry

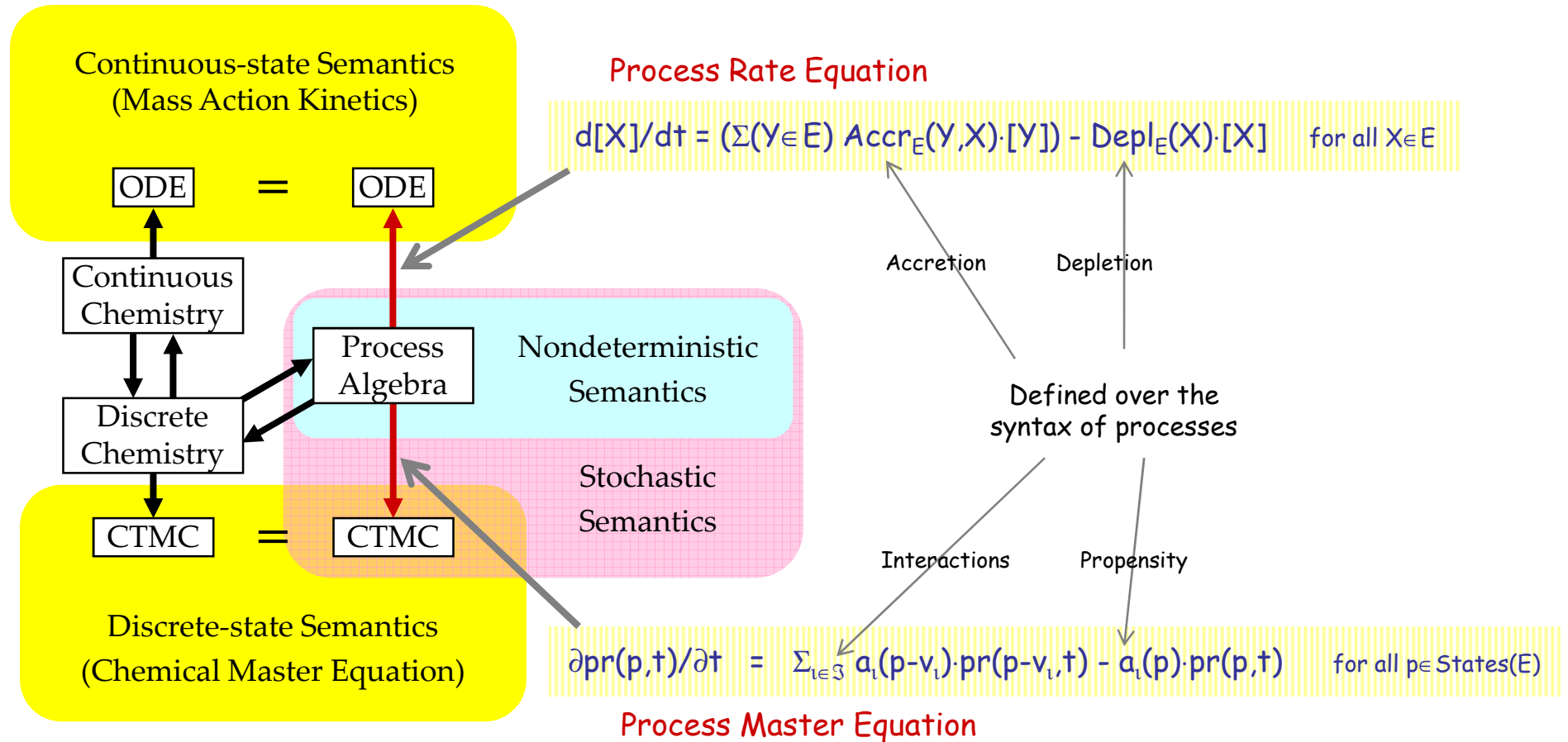


These diagrams commute via appropriate maps.

L. Cardelli: "On Process Rate Semantics" (TCS)

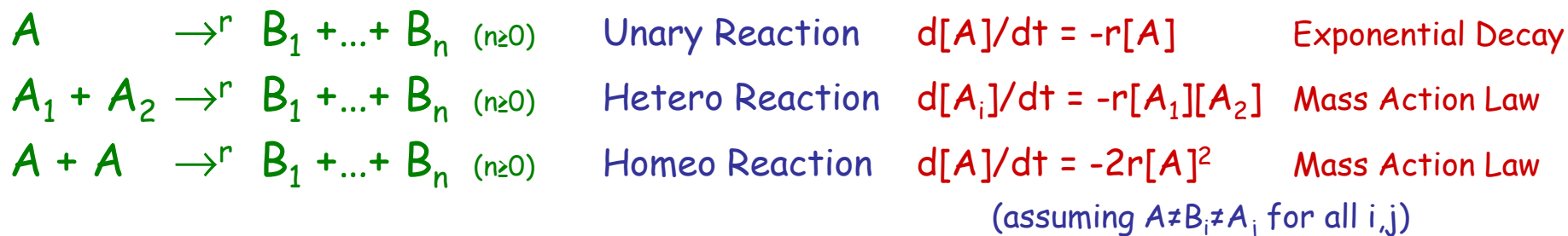
L. Cardelli: "A Process Algebra Master Equation" (QEST'07)

Quantitative Process Semantics



From CGF to Chemistry

Chemical Reactions



No other reactions!

JOURNAL OF CHEMICAL PHYSICS

VOLUME 113, NUMBER 1

The chemical Langevin equation

Daniel T. Gillespie^{a)}
 Research Department, Code 4T4100D, Naval Air Warfare Center, China Lake, California 93555

Genuinely *trimolecular* reactions do not physically occur in dilute fluids with any appreciable frequency. *Apparently* trimolecular reactions in a fluid are usually the combined result of two bimolecular reactions and one monomolecular reaction, and involve an additional short-lived species.

Chapter IV: Chemical Kinetics

[David A. Reckhow, CEE 572 Course]

... reactions may be either elementary or non-elementary. Elementary reactions are those reactions that occur exactly as they are written, without any intermediate steps. These reactions **almost always involve just one or two reactants**. ... Non-elementary reactions involve a series of two or more elementary reactions. Many complex environmental reactions are non-elementary. In general, **reactions with an overall reaction order greater than two, or reactions with some non-integer reaction order are non-elementary.**

THE COLLISION THEORY OF REACTION RATES

www.chemguide.co.uk

The chances of all this happening if your reaction needed a collision involving more than 2 particles are remote. All three (or more) particles would have to arrive at exactly the same point in space at the same time, with everything lined up exactly right, and having enough energy to react. That's not likely to happen very often!

Trimolecular reactions:



the measured "r" is an (imperfect) aggregate of e.g.:



Enzymatic reactions:



the "r" is given by Michaelis-Menten (approximated steady-state) laws:



Chemical Ground Form (CGF)

$E ::= O : X=M, E$

Reagents

$M ::= O : \pi; P \oplus M$

Molecules

$P ::= O : X | P$

Solutions

$\pi ::= \tau_{(r)} : ?a_{(r)} : !a_{(r)}$

Actions (delay, input, output)

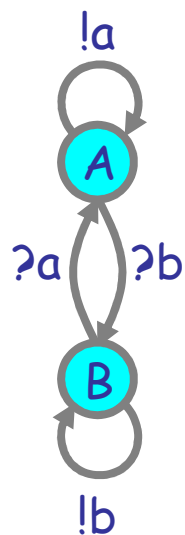
$CGF ::= E, P$

Reagents plus Initial Conditions

A stochastic subset of CCS
(no values, no restriction)

(To translate chemistry to processes we need a bit more than interacting automata: we may have "+" on the right of \rightarrow , that is we may need "|" after π .)

\oplus is stochastic choice (vs. + for chemical reactions)
 O is the null solution ($P|O = O|P = P$)
 and null molecule ($M \oplus O = O \oplus M = M$)
 Each X in E is a distinct *species*
 Each name a is assigned a fixed rate $r: a_{(r)}$



Ex: Interacting Automata

(= finite-control CGFs: they use "|" only in initial conditions):

$A = !a; A \oplus ?b; B$

$B = !b; B \oplus ?a; A$

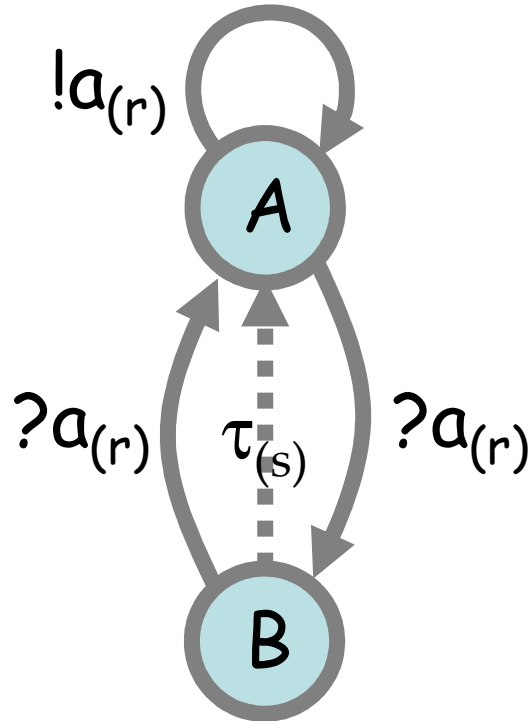
$A|A|B|B$

Automaton in state A

Automaton in state B

Initial conditions:
2A and 2B

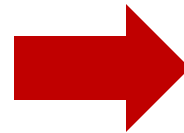
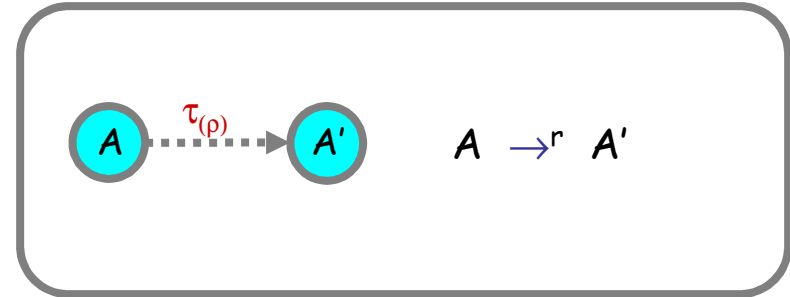
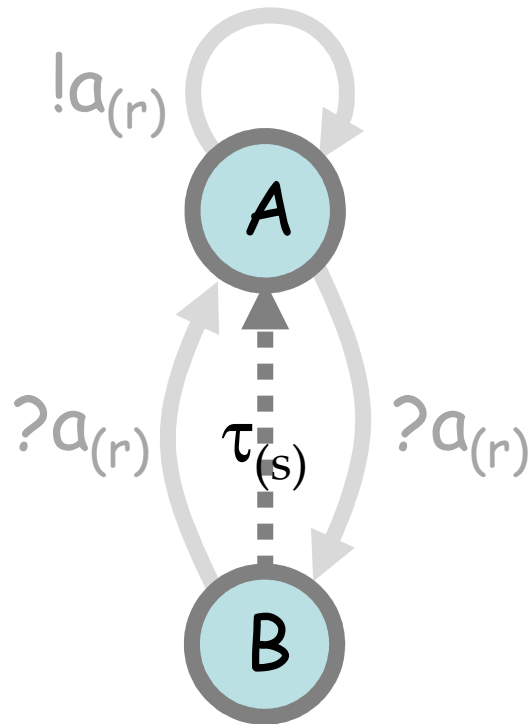
From CGF to Chemistry (by example)



$$A = !a_{(r)};A \oplus ?a_{(r)};B$$

$$B = ?a_{(r)};A \oplus \tau_{(s)};A$$

From CGF to Chemistry (by example)

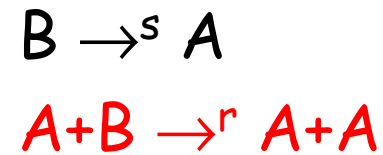
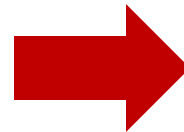
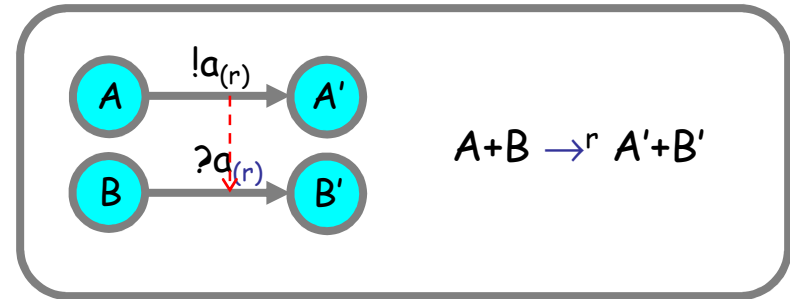
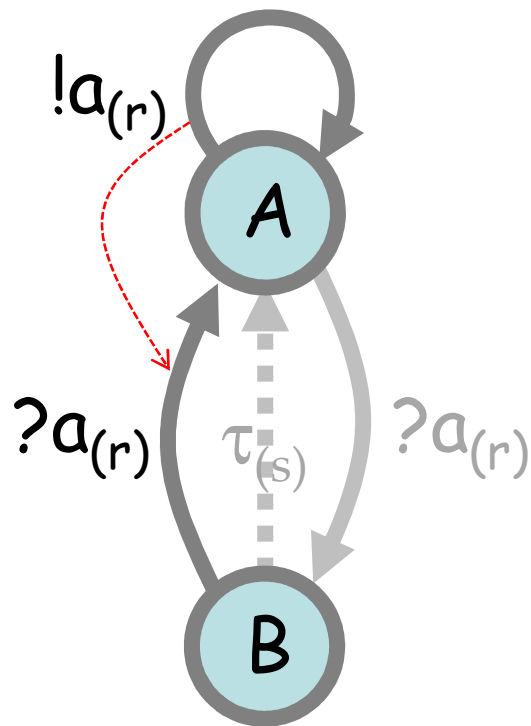


$$A = !a;A \oplus ?a;B$$

$$B = ?a;A \oplus$$

$$\tau_{(s)};A$$

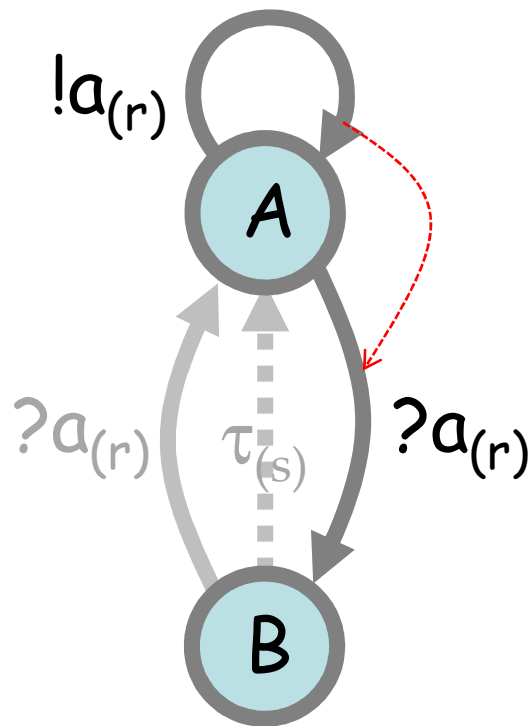
From CGF to Chemistry (by example)



$$A = !a;A \oplus ?a;B$$

$$B = ?a;A \oplus \tau_{(s)};A$$

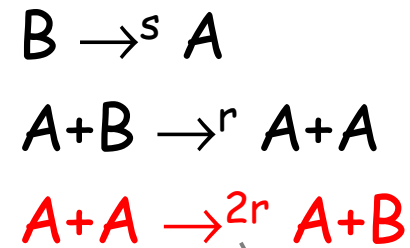
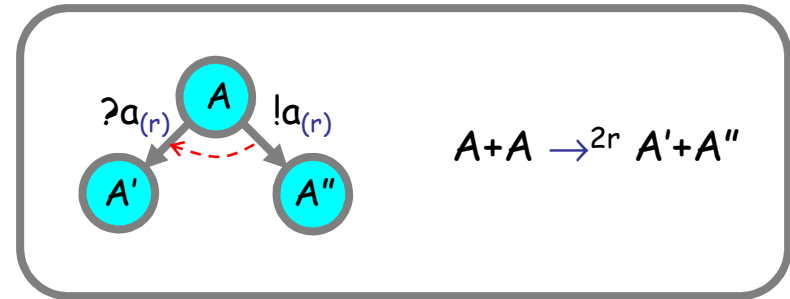
From CGF to Chemistry (by example)



$$A = !a;A \oplus ?a;B$$


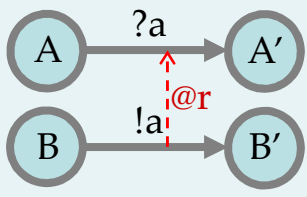
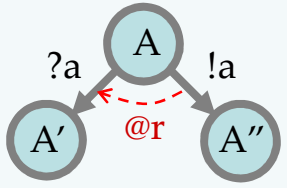
$$B = ?a;A \oplus$$

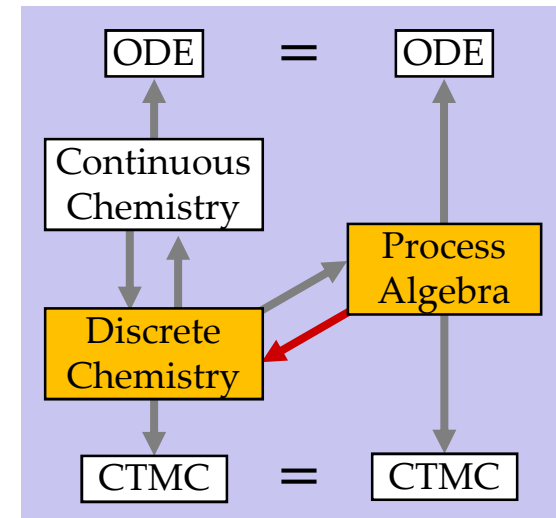
$$\tau_{(s)};A$$



Double rate for homeo reactions

From CGF to Chemistry (by example)

Interacting Automata	Discrete Chemistry
initial states $A \mid A \mid \dots \mid A$	initial quantities $\#A_0$
	$A \xrightarrow{r} A'$
	$A+B \xrightarrow{r} A'+B'$
	$A+A \xrightarrow{2r} A'+A''$



From CGF to Chemistry: Ch(E)

$E ::= O : X=M, E$	Reagents
$M ::= O : \pi; P \oplus M$	Molecules
$P ::= O : X P$	Solutions
$\pi ::= \tau_{(r)} : ?a_{(r)} : !a_{(r)}$	Interactions (delay, input, output)
$CGF ::= E, P$	Reagents plus Initial Conditions

$E.X.i \stackrel{\text{def}}{=} \text{the } i\text{-th } \oplus\text{-summand of the molecule } M \text{ associated with the } X \text{ reagent of } E$

Chemical reactions for E, P : (N.B.: $\langle \dots \rangle$ are reaction tags to obtain multiplicity of reactions, and P is P with all the $|$ changed to $+$)

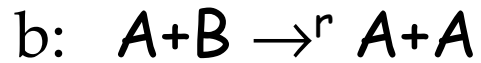
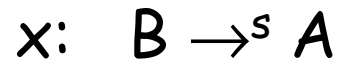
$Ch(E) :=$
 $\{ \langle X.i \rangle : X \xrightarrow{r} P \mid s.t. E.X.i = \tau_{(r)}; P \} \cup$
 $\{ \langle X.i, Y.j \rangle : X + Y \xrightarrow{r} P + Q \mid s.t. X \neq Y, E.X.i = ?a_{(r)}; P, E.Y.j = !a_{(r)}; Q \} \cup$
 $\{ \langle X.i, X.j \rangle : X + X \xrightarrow{2r} P + Q \mid s.t. E.X.i = ?a_{(r)}; P, E.X.j = !a_{(r)}; Q \} \in E$

Initial conditions for P :

$Ch(P) := P$

From Chemistry to CGF

From Chemistry to CGF (by example)

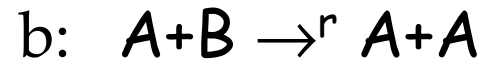


Unique reaction names

	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$	Reactions names
A				Half-rate for homeo reactions
B				

Species

From Chemistry to CGF (by example)

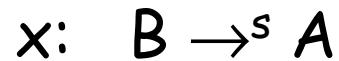


	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$
A			
B	$\tau;A$		

1: Fill the matrix by columns:

Degradation reaction $v_i: X \rightarrow k_i P_i$
add $\tau;P_i$ to $\langle X, v_i \rangle$.

From FSRN to CGF (by example)



	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$
A		?;A A	
B	$\tau;A$!;0	

1: Fill the matrix by columns:

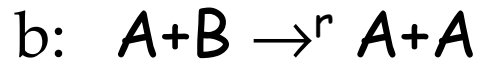
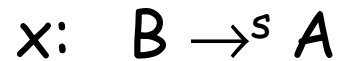
Degradation reaction $v_i: X \rightarrow k_i P_i$

add $\tau;P_i$ to $\langle X, v_i \rangle$.

Hetero reaction $v_i: X+Y \rightarrow k_i P_i$

add $?;P_i$ to $\langle X, v_i \rangle$ and $!;0$ to $\langle Y, v_i \rangle$

From FSRN to CGF (by example)



	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$
A		?;A A	?;A B !;0
B	$\tau;A$!;0	

1: Fill the matrix by columns:

Degradation reaction $v_i: X \rightarrow k_i P_i$

add $\tau;P_i$ to $\langle X, v_i \rangle$.

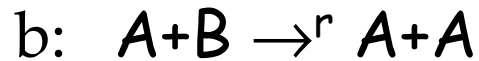
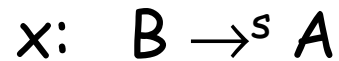
Hetero reaction $v_i: X+Y \rightarrow k_i P_i$

add $?;P_i$ to $\langle X, v_i \rangle$ and $!;0$ to $\langle Y, v_i \rangle$

Homeo reaction $v_i: X+X \rightarrow k_i P_i$

add $?;P_i$ and $!;0$ to $\langle X, v_i \rangle$

From FSRN to CGF (by example)



	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$
A		$?;A A$	$?;A B$ $!;0$
B	$\tau;A$	$!;0$	

1: Fill the matrix by columns:

Degradation reaction $v_i: X \rightarrow k_i P_i$
add $\tau;P_i$ to $\langle X, v_i \rangle$.

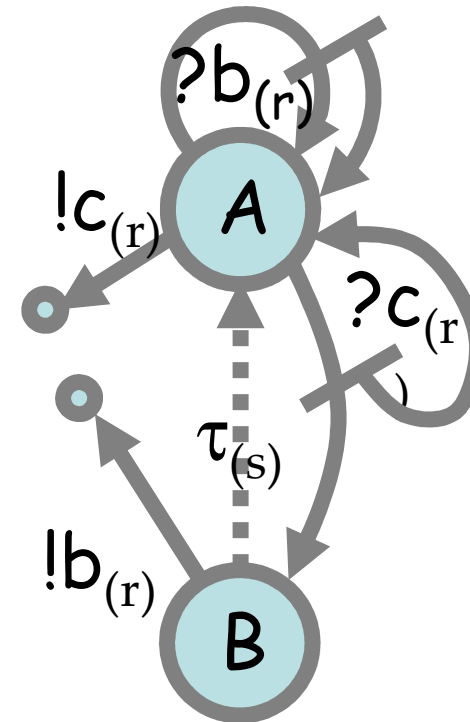
Hetero reaction $v_i: X+Y \rightarrow k_i P_i$
add $?;P_i$ to $\langle X, v_i \rangle$ and $!;0$ to $\langle Y, v_i \rangle$

Homeo reaction $v_i: X+X \rightarrow k_i P_i$
add $?;P_i$ and $!;0$ to $\langle X, v_i \rangle$

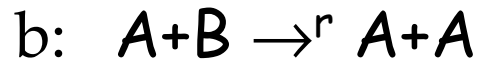
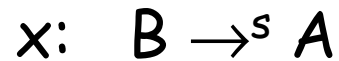
2: Read the result by rows:

$$A = ?b_{(r)}:(A|A) \oplus ?c_{(r)}:(A|B) \oplus !c_{(r)}:0$$

$$B = \tau_{(s)}:A \oplus !b_{(r)}:0$$



From FSRN to CGF (by example)



	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$
A		?;A	?;A B !;0
B	$\tau;A$!;A	

1: Fill the matrix by columns:

Degradation reaction $v_i: X \rightarrow k_i P_i$
add $t;P_i$ to $\langle X, v_i \rangle$.

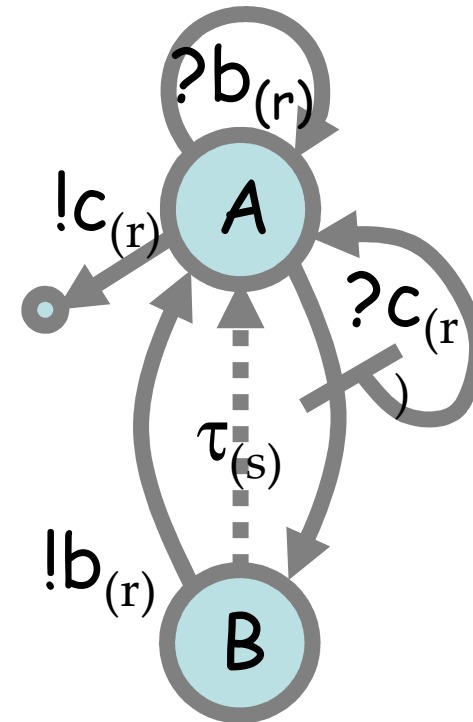
Hetero reaction $v_i: X+Y \rightarrow k_i P_i$
add $?;P_i$ to $\langle X, v_i \rangle$ and $!;0$ to $\langle Y, v_i \rangle$

Homeo reaction $v_i: X+X \rightarrow k_i P_i$
add $?;P_i$ and $!;0$ to $\langle X, v_i \rangle$

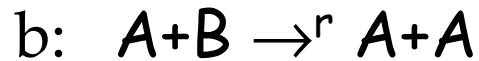
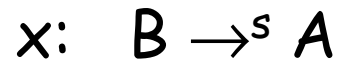
2: Read the result by rows:

$$A = ?b_{(r)};A \oplus ?c_{(r)};(A|B) \oplus !c_{(r)};0$$

$$B = t_{(s)};A \oplus !b_{(r)};A$$



From FSRN to CGF (by example)



	$x_{(s)}$	$b_{(r)}$	$c_{(r)}$
A		?;A	?;B !;A
B	τ ;A	!;A	

1: Fill the matrix by columns:

Degradation reaction $v_i: X \rightarrow k_i P_i$
 add $\dagger;P_i$ to $\langle X, v_i \rangle$.

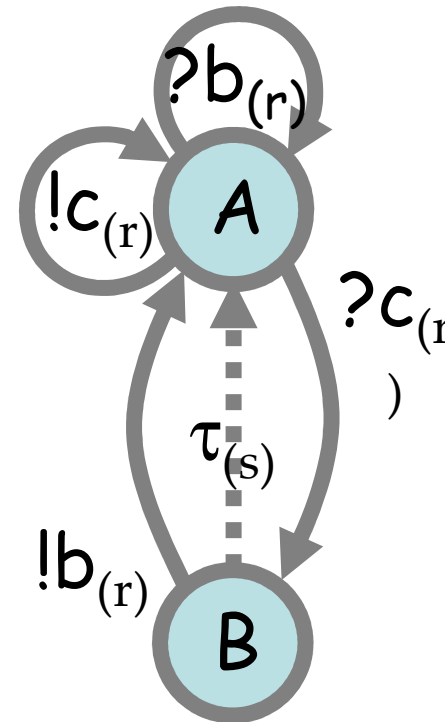
Hetero reaction $v_i: X+Y \rightarrow k_i P_i$
 add $?;P_i$ to $\langle X, v_i \rangle$ and $!;O$ to $\langle Y, v_i \rangle$

Homeo reaction $v_i: X+X \rightarrow k_i P_i$
 add $?;P_i$ and $!;O$ to $\langle X, v_i \rangle$

2: Read the result by rows:

$$A = ?b_{(r)};A \oplus ?c_{(r)};B \oplus !c_{(r)};A$$

$$B = \dagger_{(s)};A \oplus !b_{(r)};A$$



From Chemistry to CGF: $\text{Pi}(C)$

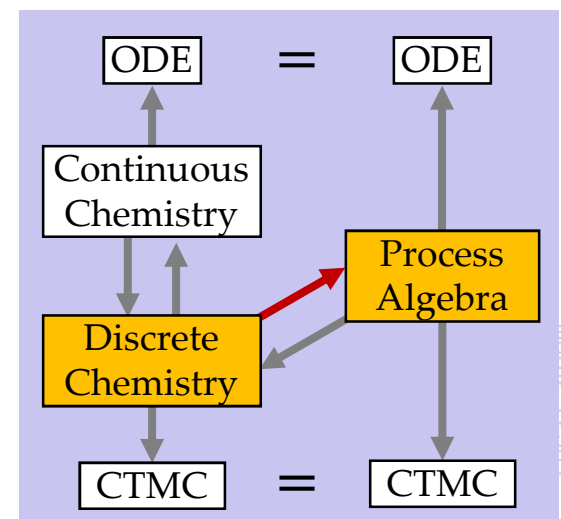
$v: X \xrightarrow{r} Y_1 + \dots + Y_n + 0$ Unary Reaction

$v: X_1 + X_2 \xrightarrow{r} Y_1 + \dots + Y_n + 0$ Binary Reaction

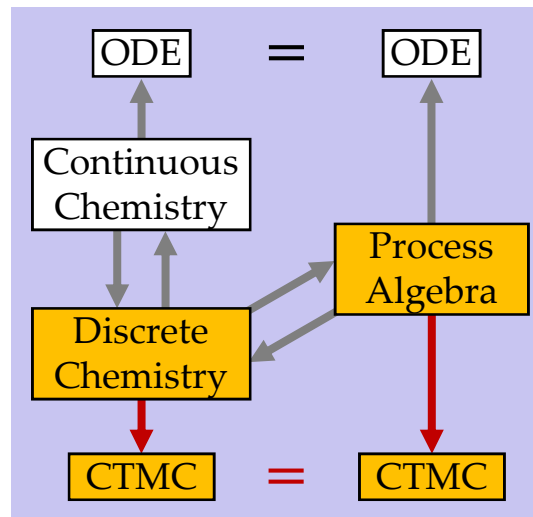
From uniquely-labeled $(v:)$ chemical reactions C to a CGF $\text{Pi}(C)$:

$$\text{Pi}(C) = \{ (X = \oplus ((v: X \xrightarrow{k} P) \in C) \text{ of } (\tau_{(k)}; P) \oplus ((v: X+Y \xrightarrow{k} P) \in C \text{ and } Y \neq X) \text{ of } (?v_{(k)}; P) \oplus ((v: Y+X \xrightarrow{k} P) \in C \text{ and } Y \neq X) \text{ of } (!v_{(k)}; 0) \oplus ((v: X+X \xrightarrow{k} P) \in C) \text{ of } (?v_{(k/2)}; P \oplus !v_{(k/2)}; 0)) \}$$

s.t. X is a species in C

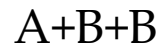
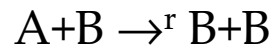
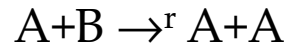


Discrete-State Semantics

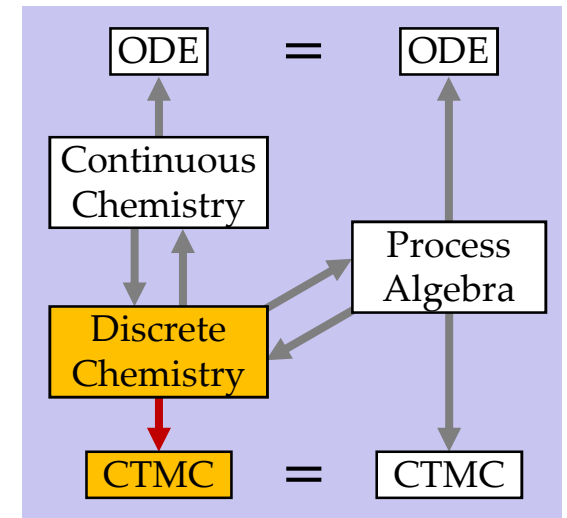
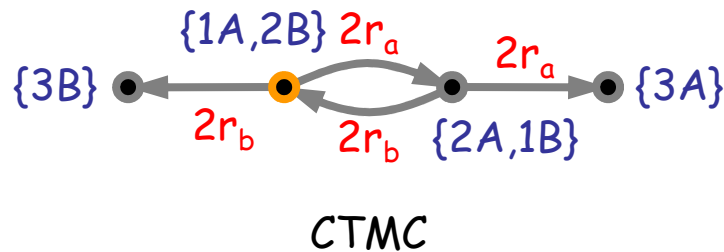


Discrete Semantics of Reactions

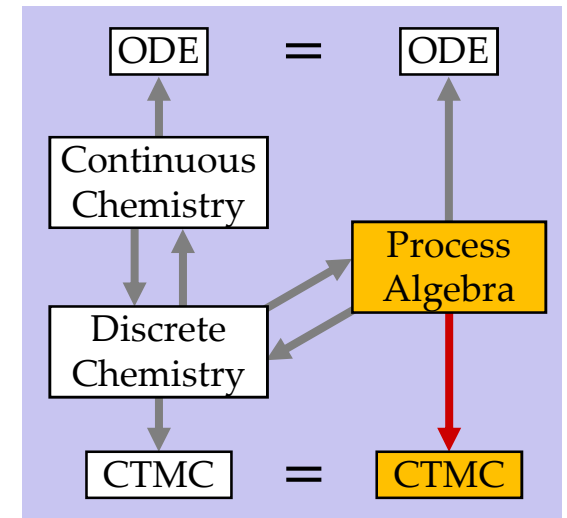
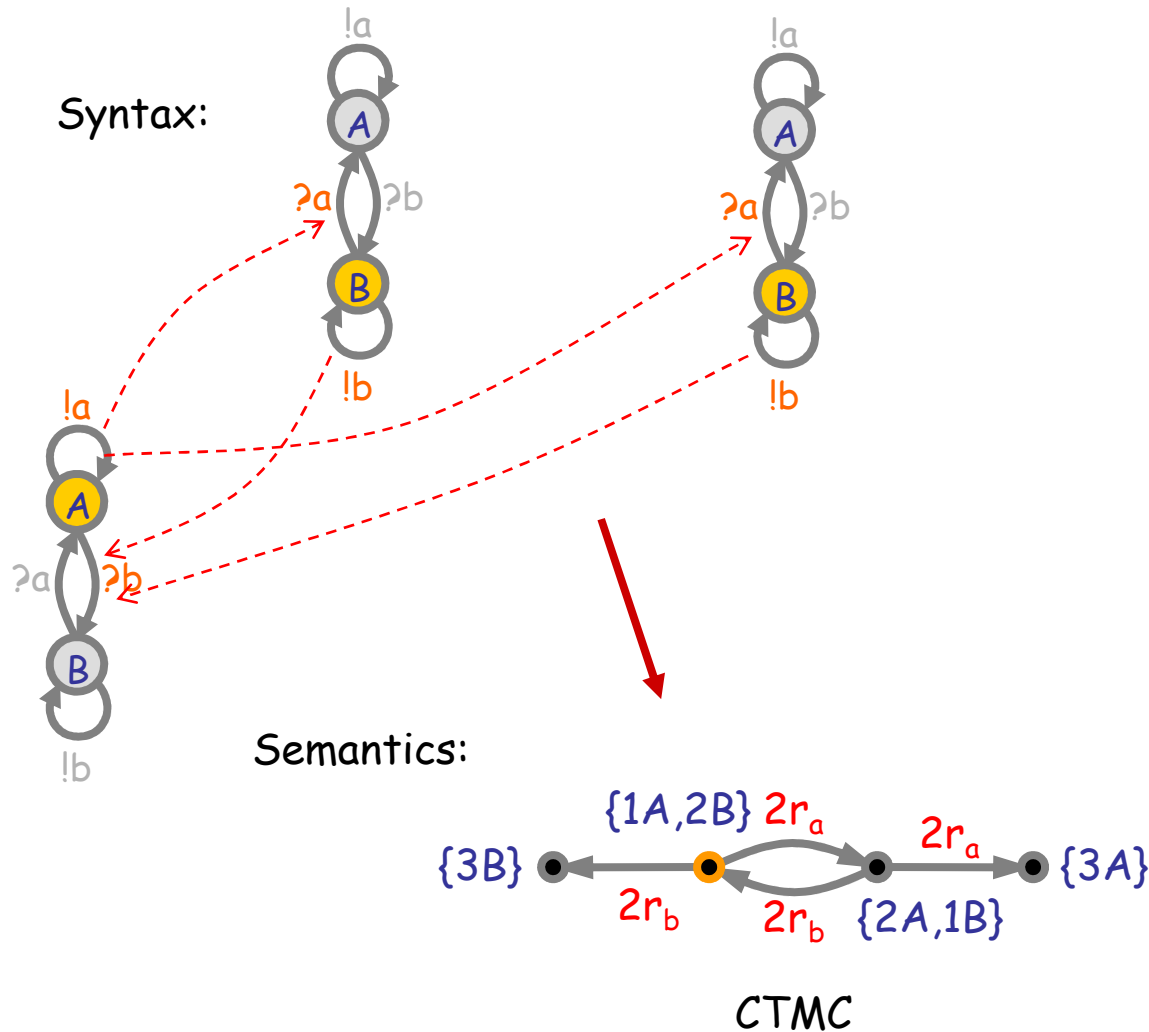
Syntax:



Semantics:



Discrete Semantics of Reagents

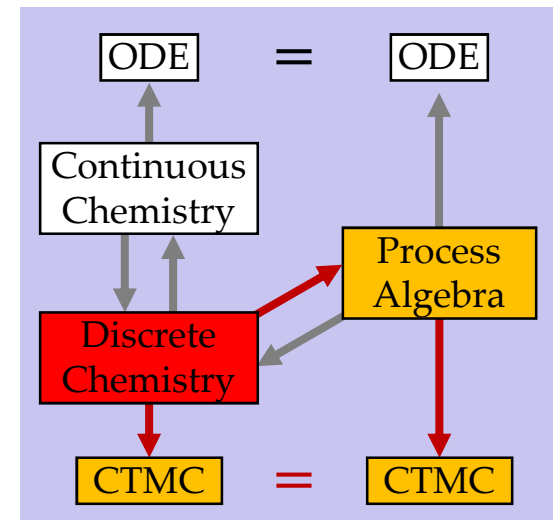
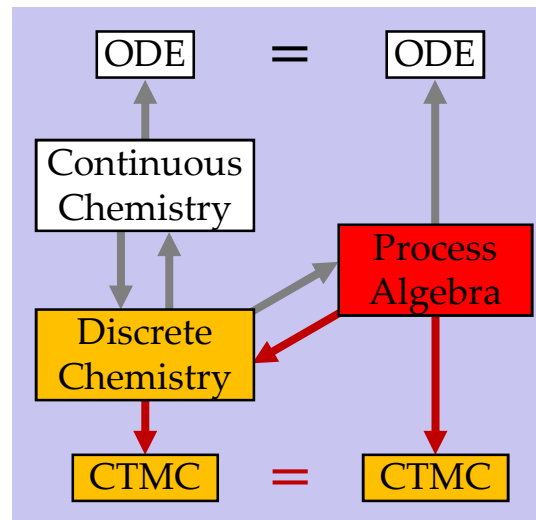


Discrete State Equivalence

- Def: \approx is equivalent CTMC's (isomorphic graphs with same rates).

- Thm: $E \approx \text{Ch}(E)$

- Thm: $C \approx \text{Pi}(C)$



- For each E there is an $E' \approx E$ that is detangled ($E' = \text{Pi}(\text{Ch}(E))$)

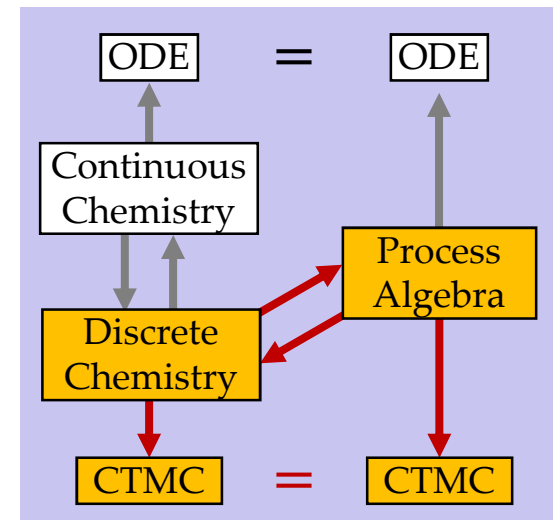
- For each E in automata form there is an $E' \approx E$ that is detangled and in automata form ($E' = \text{Detangle}(E)$).

Interacting Automata = Discrete Chemistry

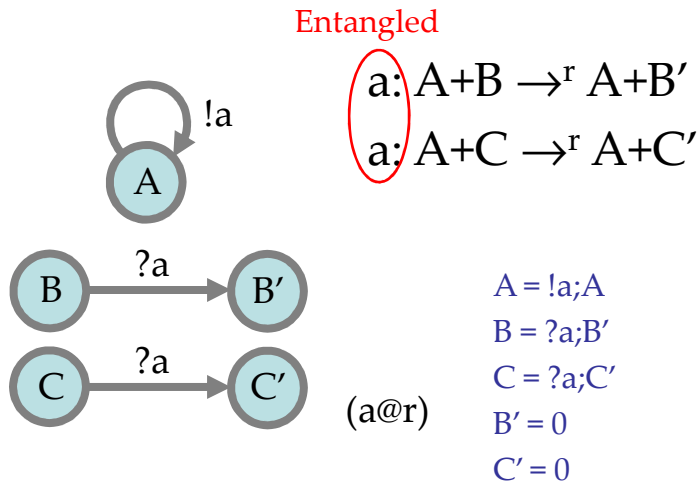
This is enough to establish that the process algebra is really faithful to the chemistry.

But CTMC are not the “ultimate semantics” because there are still questions of when two different CTMCs are actually equivalent (e.g. “lumping”).

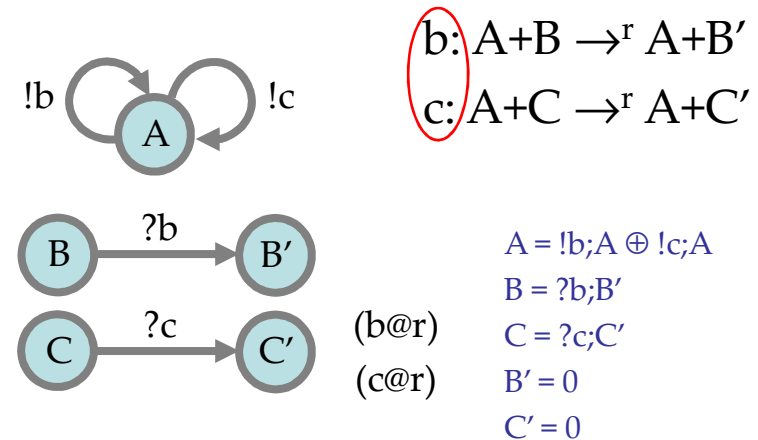
The “ultimate semantics” of chemistry is the *Chemical Master Equation* (derivable from the Chapman-Kolmogorov equation of the CTMC).



Entangled vs Detangled



Entangled: Two reactions on one channel



Detangled: Two reactions on two separate channels

We need a semantics of automata that identifies automata that have the "same chemistry".

No process algebra equivalence is like this!

Entangled automata lead to more compact models than in chemistry.

Detangled automata are in simple correspondence with chemistry.

From Discrete to Continuous Chemistry

The "Type System" of Chemistry

The International System of Units (SI) defines the following physical units, with related derived units and constants; note that *amount of substance* is a base unit in SI, like length and time:

mol	(a base unit)	mole, unit of <i>amount of substance</i>
m	(a base unit)	meter, unit of <i>length</i>
s	(a base unit)	second, unit of <i>time</i>
$L = 0.001 \cdot m^3$		liter (volume)
$M = mol \cdot L^{-1}$		molarity (concentration of substance)
$N_A : mol^{-1} \cong 6.022 \times 10^{23}$		Avogadro's number (number of particles per amount of substance)

For a substance $X: mol$, we write $[X]:M$ for the concentration of X , and $[X]':M \cdot s^{-1}$ for the time derivative of the concentration.

A **continuous chemical system** (C, V) is a system of chemical reactions C plus a vector of **initial concentrations** $V_X: M$, one for each species X .

The rates of unary reactions have dimension s^{-1} .

The rates of binary reactions have dimension $M^{-1}s^{-1}$.

(because in both cases the rhs of an ODE should have dimension $M \cdot s^{-1}$).

For a given volume of solution V , the volumetric factor γ of dimension M^{-1} is:

$$\gamma : M^{-1} = N_A V \quad \text{where } N_A : mol^{-1} \text{ and } V : L$$

$\#X / \gamma : M =$ concentration of X molecules

$\gamma \bullet [X] : 1 =$ total number of X molecules (rounded to an integer).

The Gillespie Conversion

Discrete Chemistry	Continuous Chemistry	$\gamma = N_A V$	$:M^{-1}$
initial quantities $\#A_0$	initial concentrations $[A]_0$	with $[A]_0 = \#A_0/\gamma$	
$A \xrightarrow{r} A'$	$A \xrightarrow{k} A'$	with $k = r$	$:s^{-1}$
$A+B \xrightarrow{r} A'+B'$	$A+B \xrightarrow{k} A'+B'$	with $k = r\gamma$	$:M^{-1}s^{-1}$
$A+A \xrightarrow{r} A'+A''$	$A+A \xrightarrow{k} A'+A''$	with $k = r\gamma/2$	$:M^{-1}s^{-1}$

V = interaction volume

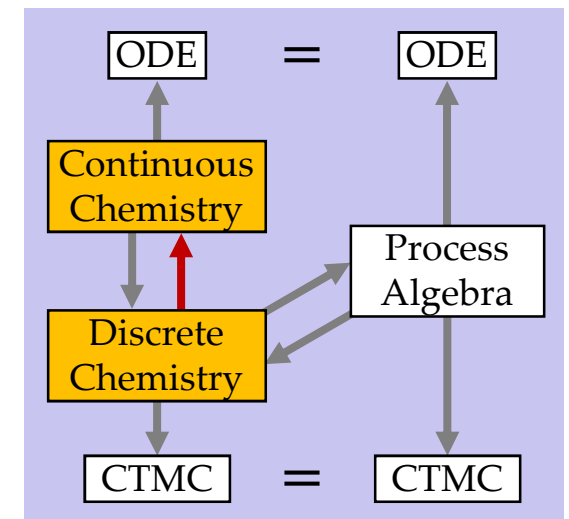
N_A = Avogadro's number

Think $\gamma = 1$

i.e. $V = 1/N_A$

$M = mol \cdot L^{-1}$

molarity (concentration)



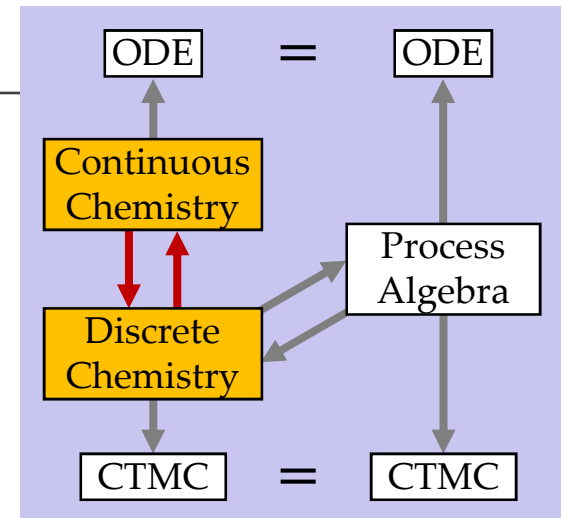
Cont_γ and Disc_γ

4.2-3 Definition: Cont_γ and Disc_γ

For a volumetric factor $\gamma: M^{-1}$, we define a translation $Cont_\gamma$ from a discrete chemical systems (C,P), with species X and initial molecule count $\#X_0 = \#X(P)$, to a continuous chemical systems (C,V) with initial concentration $[X]_0 = V_X$. The translation $Disc_\gamma$ is its inverse, up to a rounding error $\lceil \gamma[X]_0 \rceil$ in converting concentrations to molecule counts. Since γ is a global conversion constant, we later usually omit it as a subscript.

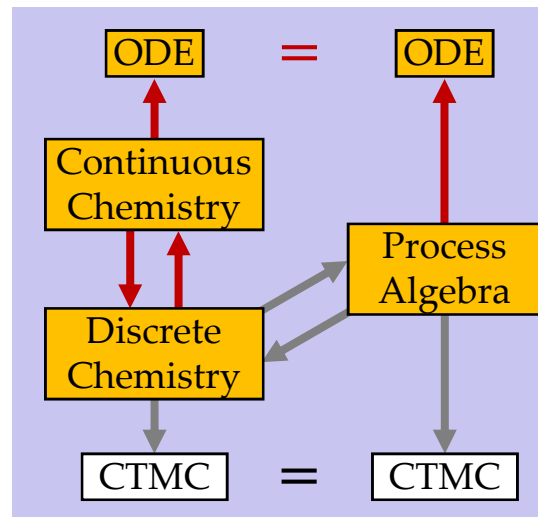
$Cont_\gamma(X \rightarrow^r P)$	$= X \rightarrow^k P$	with $k = r,$	$r:s^{-1}$	$k:s^{-1}$
$Cont_\gamma(X+Y \rightarrow^r P)$	$= X+Y \rightarrow^k P$	with $k = r\gamma$	$r:s^{-1}$	$k:M^{-1}s^{-1}$
$Cont_\gamma(X+X \rightarrow^r P)$	$= X+X \rightarrow^k P$	with $k = r\gamma/2$	$r:s^{-1}$	$k:M^{-1}s^{-1}$
$Cont_\gamma(\#X_0)$	$= [X]_0$	with $[X]_0 = \#X_0/\gamma$	$X_0:mol$	$[X]_0:M$
$Disc_\gamma(X \rightarrow^k P)$	$= X \rightarrow^r P$	with $r = k,$	$k:s^{-1}$	$r:s^{-1}$
$Disc_\gamma(X+Y \rightarrow^k P)$	$= X+Y \rightarrow^r P$	with $r = k/\gamma$	$k:M^{-1}s^{-1}$	$r:s^{-1}$
$Disc_\gamma(X+X \rightarrow^k P)$	$= X+X \rightarrow^r P$	with $r = 2k/\gamma$	$k:M^{-1}s^{-1}$	$r:s^{-1}$
$Disc_\gamma([X]_0)$	$= \#X_0$	with $\#X_0 = \lceil \gamma[X]_0 \rceil$	$[X]_0:M$	$X_0:mol$

$$Ch_\gamma := Cont_\gamma \circ Ch$$



Continuous-State Semantics

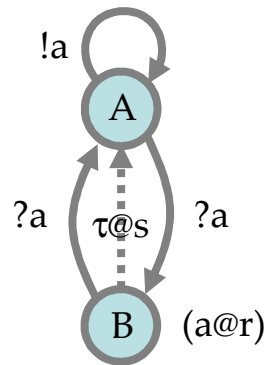
(summary)



Same Semantics

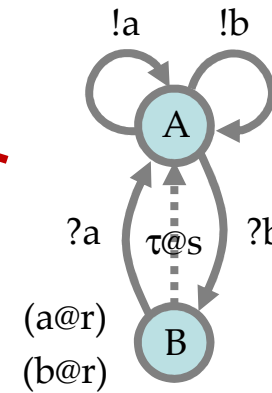
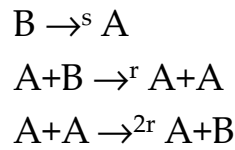
Could chemistry itself be that semantics?

No: different sets of reactions can have the same behavior!



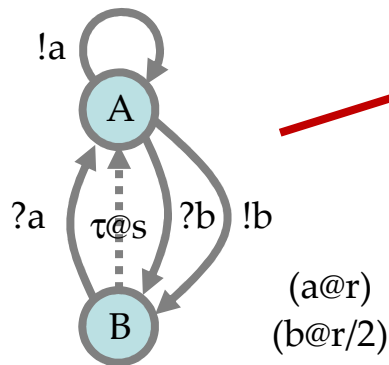
$$A = !a;A \oplus ?a;B$$

$$B = ?a;A \oplus \tau_{(s)};A$$



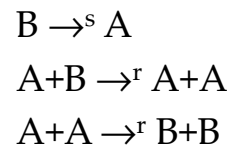
$$A = !a;A \oplus !b;A \oplus ?b;B$$

$$B = ?a;A \oplus \tau_{(s)};A$$



$$A = !a;A \oplus !b;B \oplus ?b;B$$

$$B = ?a;A \oplus \tau_{(s)};A$$



Different reactions,
but they induce the
same ODEs

From Reactions to ODEs (Law of Mass Action)



Write the coefficients by columns

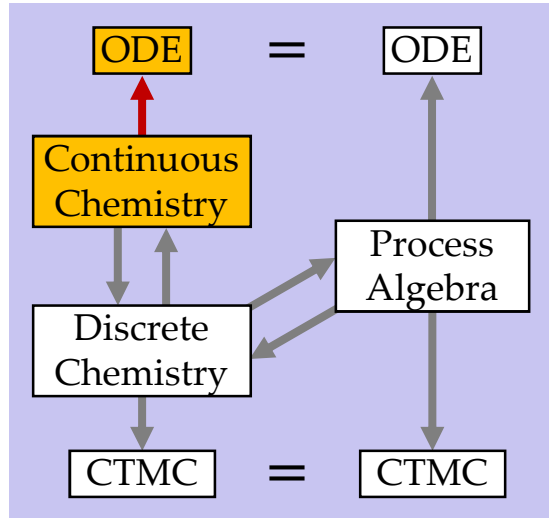
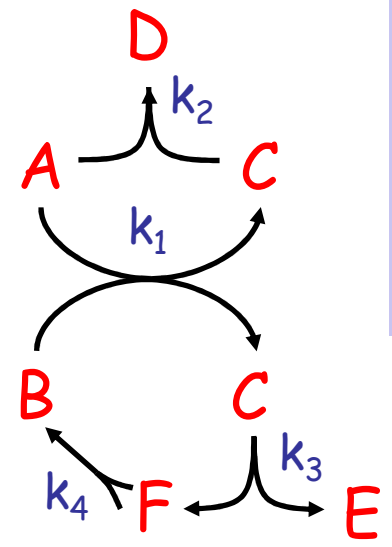
Stoichiometric Matrix

reactions

N	v_1	v_2	v_3	v_4
A	-1	-1		
B	-1			1
C	2	-1	-1	
D		1		
E			1	
F			1	-2

species

X



Quantity changes

Stoichiometric matrix

Rate laws

$d[X]/dt = N \cdot I$

$d[A]/dt = -I_1 - I_2$
 $d[B]/dt = -I_1 + I_4$
 $d[C]/dt = 2I_1 - I_2 - I_3$
 $d[D]/dt = I_2$
 $d[E]/dt = I_3$
 $d[F]/dt = I_3 - 2I_4$

Read the concentration changes from the rows

E.g. $d[A]/dt = -k_1[A][B] - k_2[A][C]$

Set a rate law for each reaction (Degradation/Hetero/Homeo)

I	
I_1	$k_1[A][B]$
I_2	$k_2[A][C]$
I_3	$k_3[C]$
I_4	$k_4[F]^2$

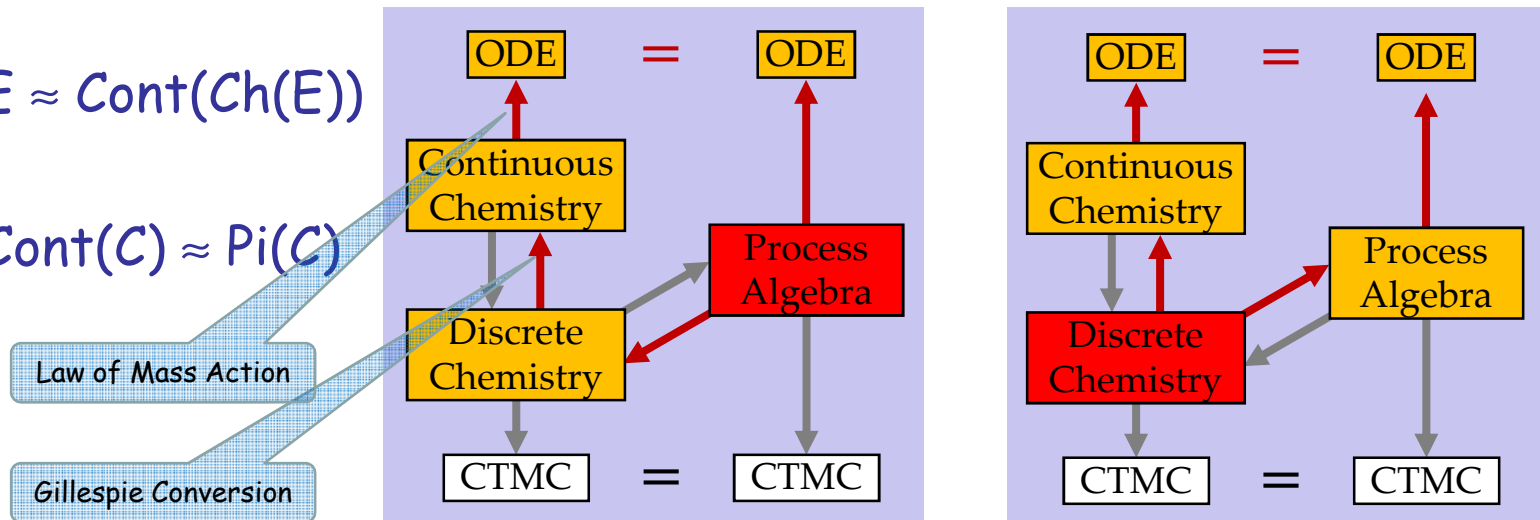
X: chemical species
[-]: quantity of molecules
I: rate laws
k: kinetic parameters
N: stoichiometric matrix

Continuous State Equivalence

- Def: \approx is equivalence of polynomials over the field of reals.

- Thm: $E \approx \text{Cont}(\text{Ch}(E))$

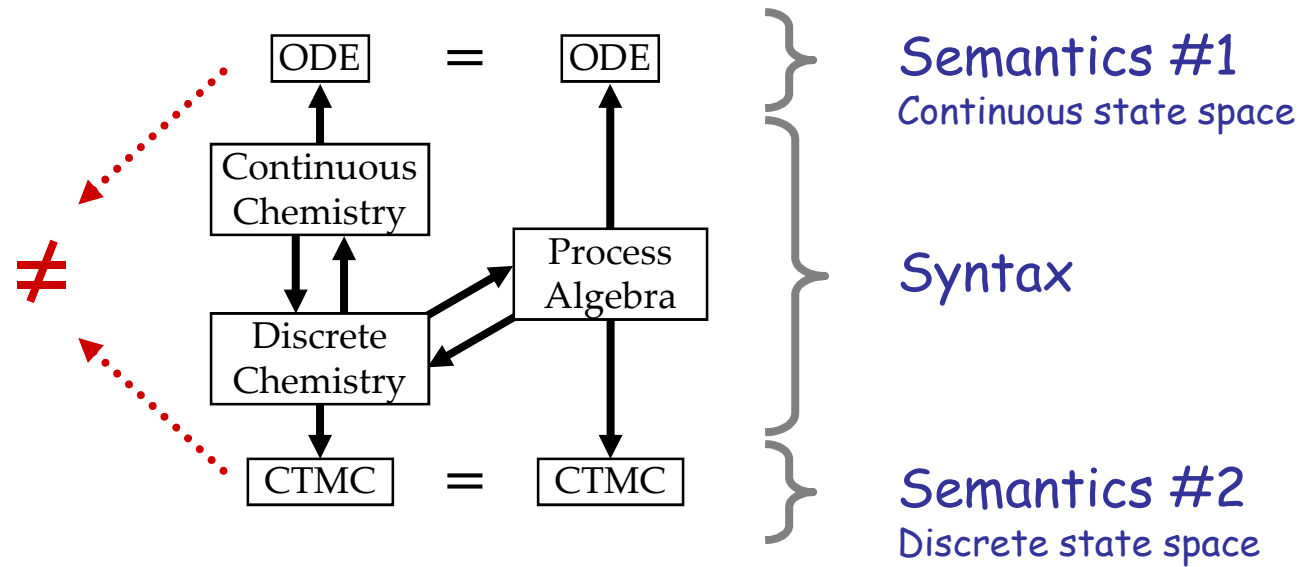
- Thm: $\text{Cont}(C) \approx \text{Pi}(C)$



- For each E there is an $E' \approx E$ that is detangled ($E' = \text{Pi}(\text{Ch}(E))$)

- For each E in automata form there is an $E' \approx E$ that is detangled and in automata form ($E' = \text{Detangle}(E)$).

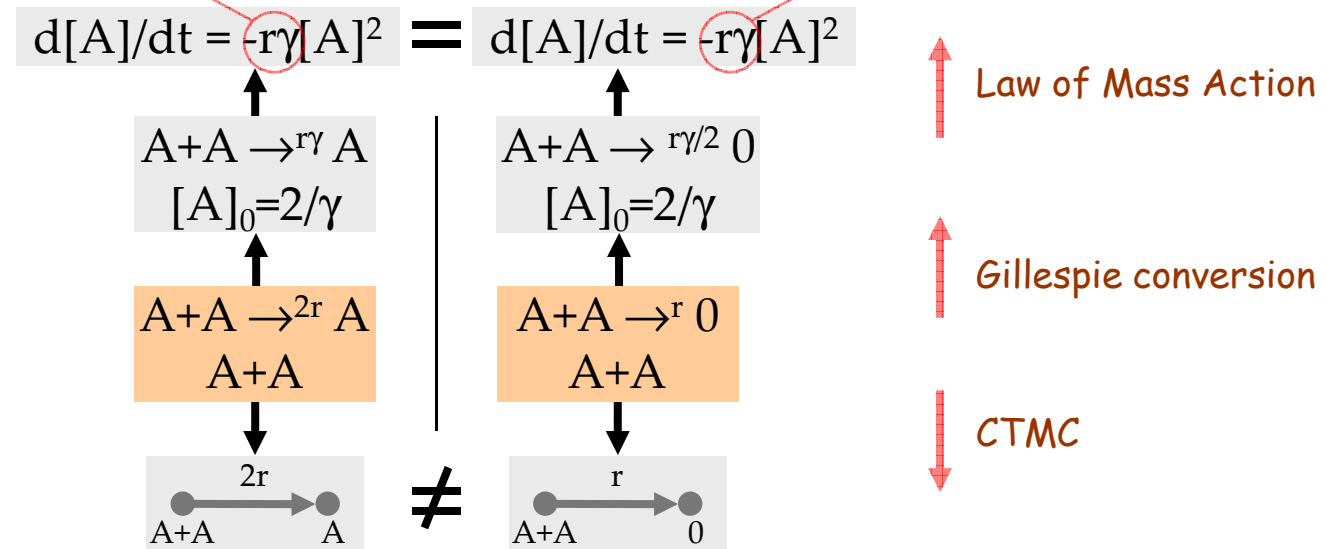
GMA \neq CME





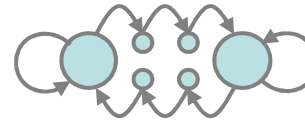
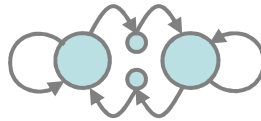
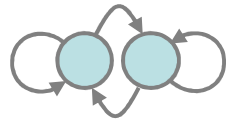
1*reaction rate $r\gamma$ because
1*A is lost in reaction.

2*reaction rate $r\gamma/2$ because
2*A are lost in reaction.

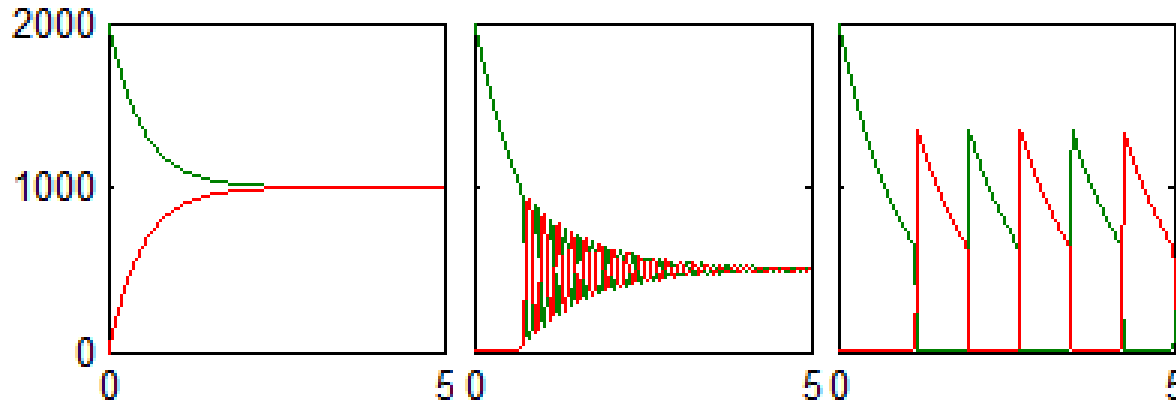


(For conservation of mass, consider instead $A+A \xrightarrow{2r} A+B$ vs. $A+A \xrightarrow{r} B+B$)

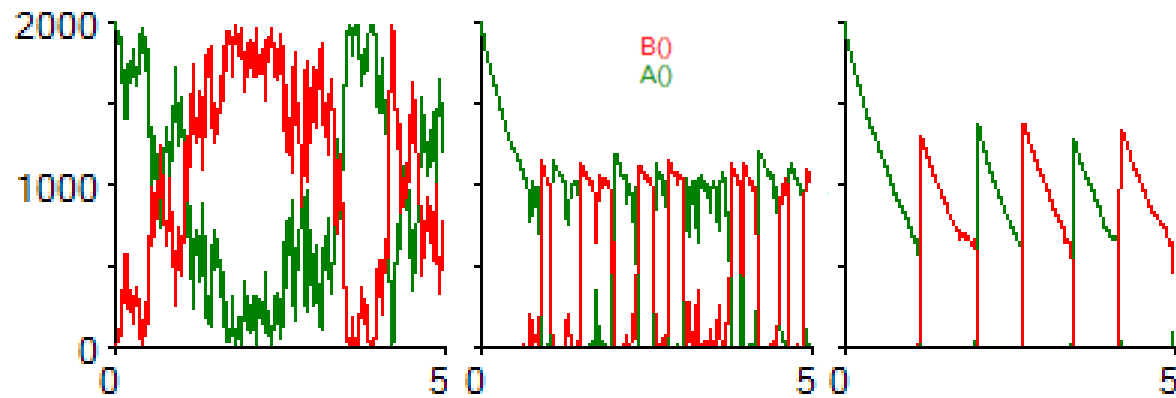
Continuous vs. Discrete Groupies



(all with doping)



Matlab



SPiM

$2000 \times A, 0 \times B, 1 \times A_d, 1 \times B_d, r = 1.0$

```
directive sample 5.0 1000
directive plot B0;A0
new a@1.0(chan)
new b@1.0(chan)
let A0 = do {a; A0} or {b; B0}
and B0 = do {b; B0} or {a; A0}
let Ad0 = {a; Ad0}
and Bd0 = {b; Bd0}
run 2000 of A0
run 1 of (Ad0 | Bd0)
```

```
directive sample 5.0 1000
directive plot B0;A0
new a@1.0(chan)
new b@1.0(chan)
let A0 = do {a; A0} or {b; B0}
and B0 = do {b; B0} or {a; A0}
let Ad0 = {a; Ad0}
and Bd0 = {b; Bd0}
run 2000 of A0
run 1 of (Ad0 | Bd0)
```

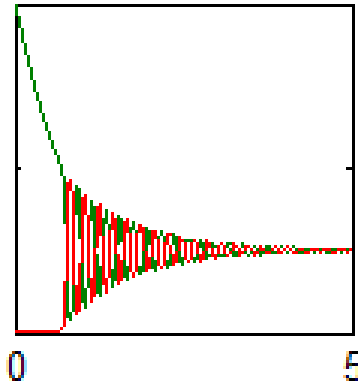
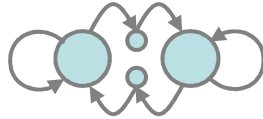
```
directive sample 5.0 1000
directive plot B0;A0
new a@1.0(chan)
new b@1.0(chan)
let A0 = do {a; A0} or {b; B0}
and B0 = do {b; B0} or {a; A0}
let Ad0 = {a; Ad0}
and Bd0 = {b; Bd0}
run 2000 of A0
run 1 of (Ad0 | Bd0)
```

```
Groupes ODEs - Groupies.mat
[0;0;0;5;0] r=1.0 k=1.0
A dx1/dt=x1*x4-x3*x1-x1*x4, 2000.0
A' dx1/dt=x3*x1-x3*x2-x1*x2, 0.0
B dx3/dt=x3*x2-x1*x3-x3*x2, 0.0
B' dx3/dt=x1*x3-x1*x3-x3*x1, 0.0
```

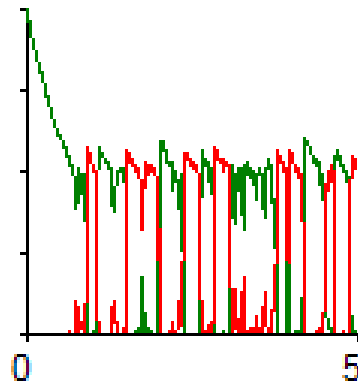
```
Groupes ODEs - Groupies Hysteric 1.mat
[0;0;0;5;0] r=1.0 k=1.0
A dx1/dt=x1*x4-x3*x1-x1*x4, 2000.0
A' dx1/dt=x3*x1-x3*x2-x1*x2, 0.0
B dx3/dt=x3*x2-x1*x3-x3*x2, 0.0
B' dx3/dt=x1*x3-x1*x3-x3*x1, 0.0
```

```
Groupes ODEs - Groupies Hysteric 2.mat
[0;0;0;5;0] r=1.0 k=1.0
A dx1/dt=x1*x4-x3*x1-x1*x4, 2000.0
A' dx1/dt=x3*x1-x3*x2-x1*x2, 0.0
A'' dx1/dt=x3*x1-x3*x2-x1*x2, 0.0
B dx3/dt=x3*x2-x1*x3-x3*x2, 0.0
B' dx3/dt=x1*x3-x1*x3-x3*x1, 0.0
B'' dx3/dt=x1*x3-x1*x3-x3*x1, 0.0
```

Scientific Predictions



After a while, all 4 states are almost equally occupied.

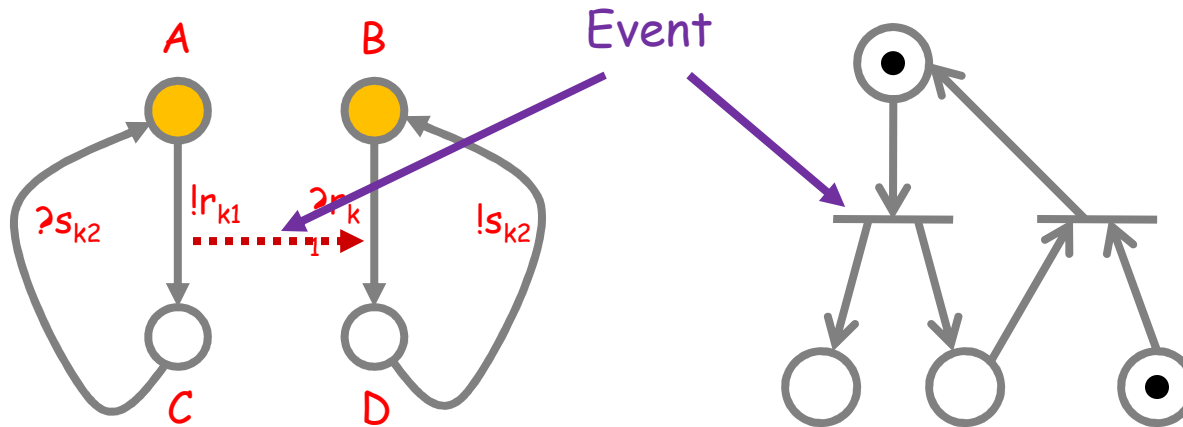


The 4 states are almost never equally occupied.

Discrete Analysis Techniques

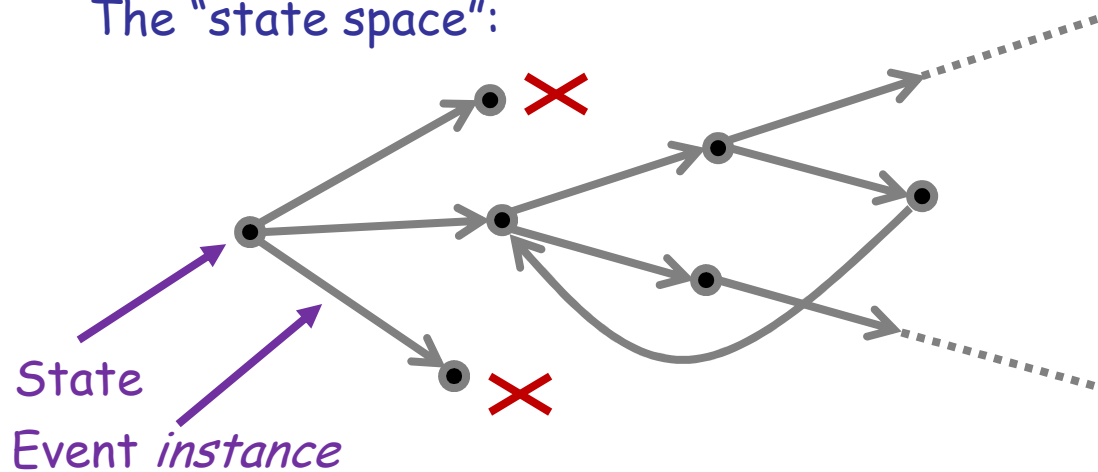
The Program vs. the State Space

The "program":



Finite

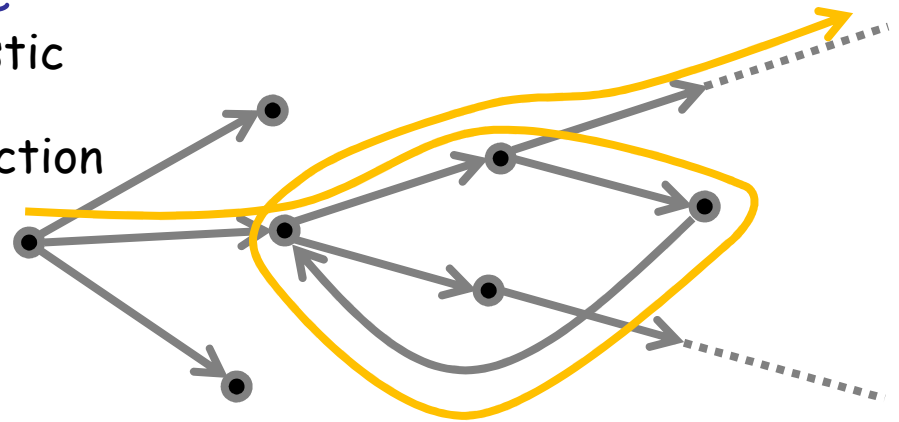
The "state space":



Potentially infinite

Simulation

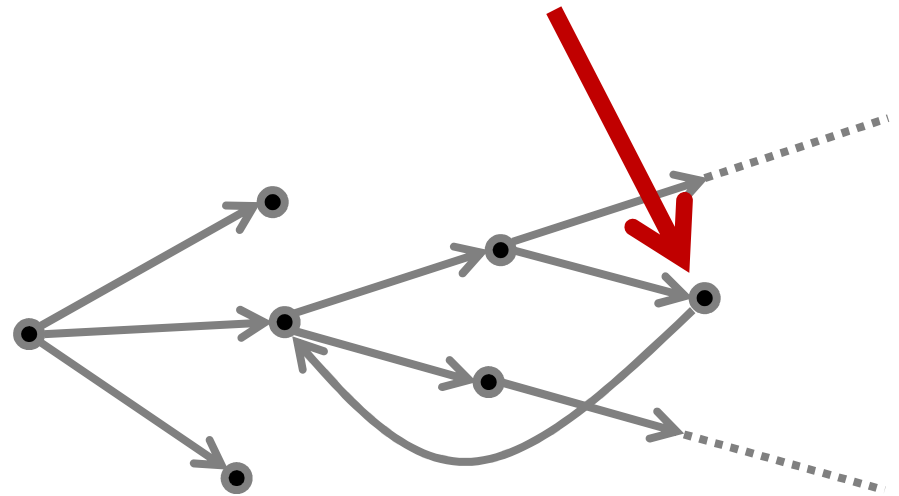
- Run “the program” through a walk in states space.
- Basic stochastic algorithm: Gillespie
 - Exact (i.e. based on physics) stochastic simulation of chemical kinetics.
 - Can compute concentrations and reaction times for biochemical networks.
- Stochastic Process Algebras
 - Now many [BioSPi, SPiM, BioPEPA, BetaBinders, ...]
- Hybrid approaches
 - Continuous + discrete/stochastic switching



Control Flow Analysis

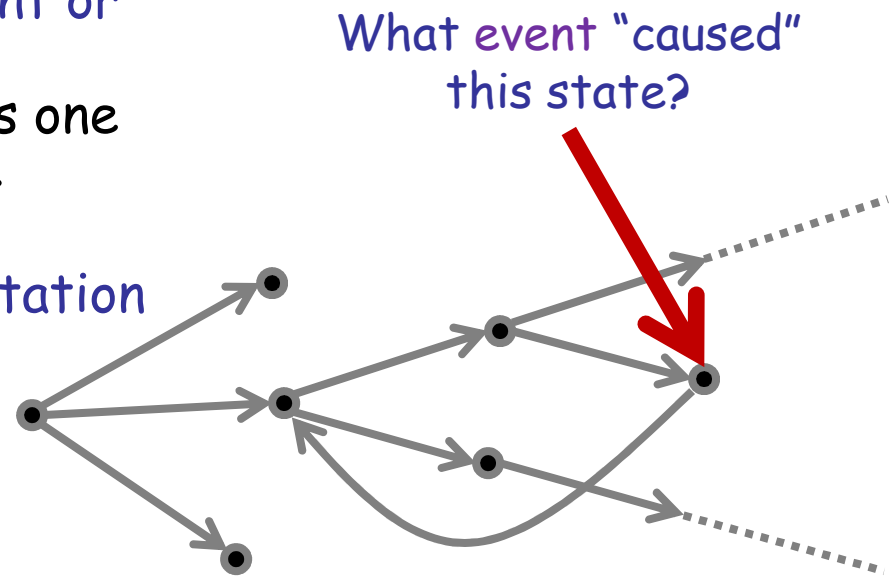
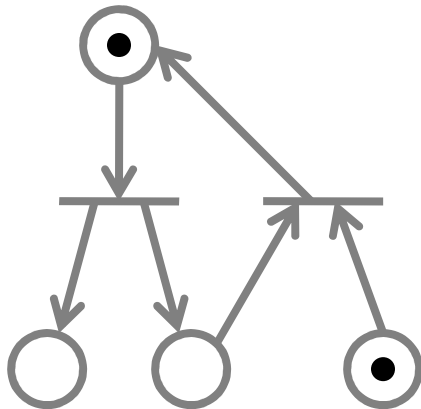
- Who may call who?
 - Overapproximation of behavior used to answer questions about what "cannot happen".

What *event* may (or may not) have been involved in reaching this state?



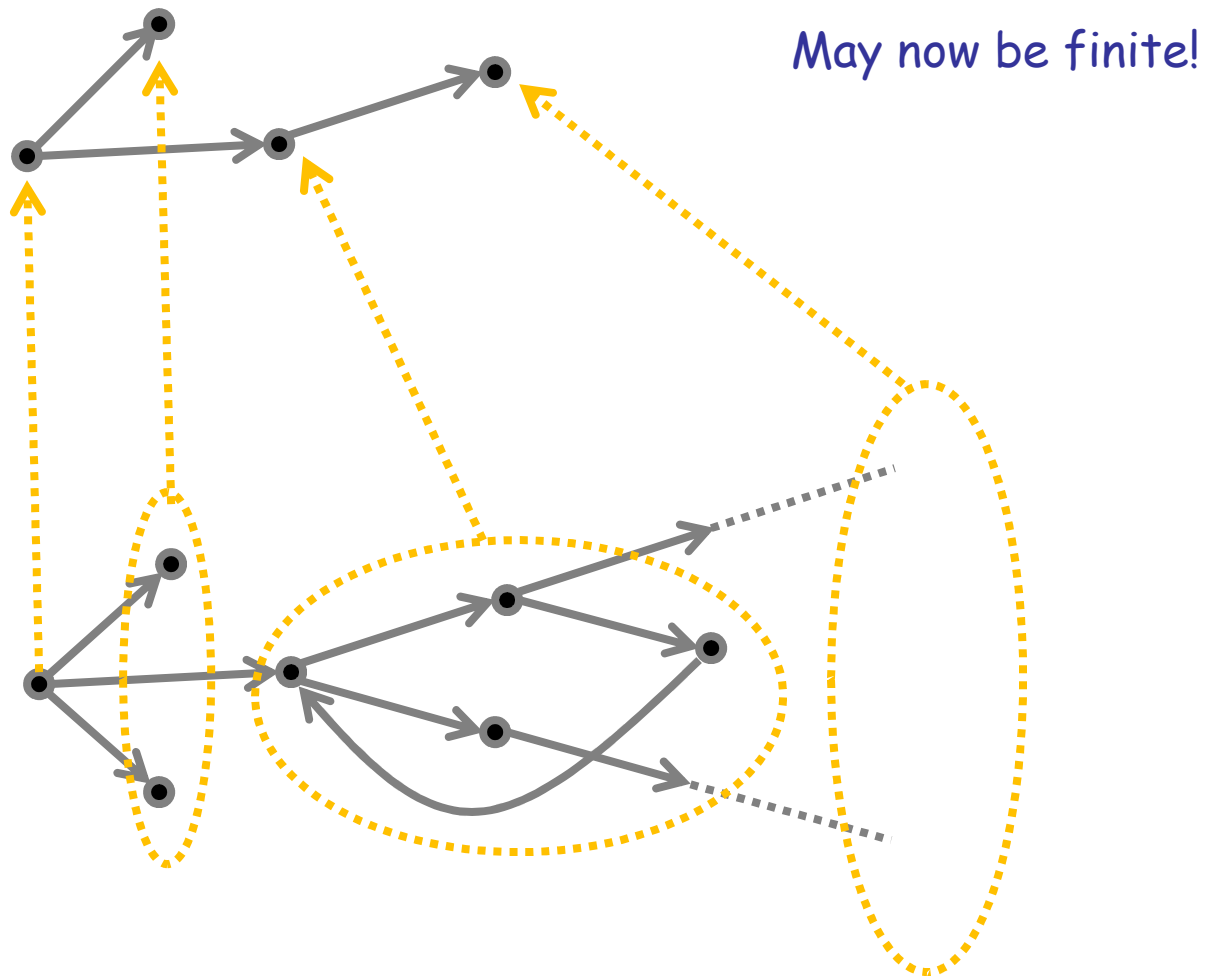
Causality Analysis

- What event caused what other event or state to happen?
 - E.g.: if in all possible executions one event always precedes another.
- Need a different level of representation (the "event space")
 - Petri Nets
 - Event Structures



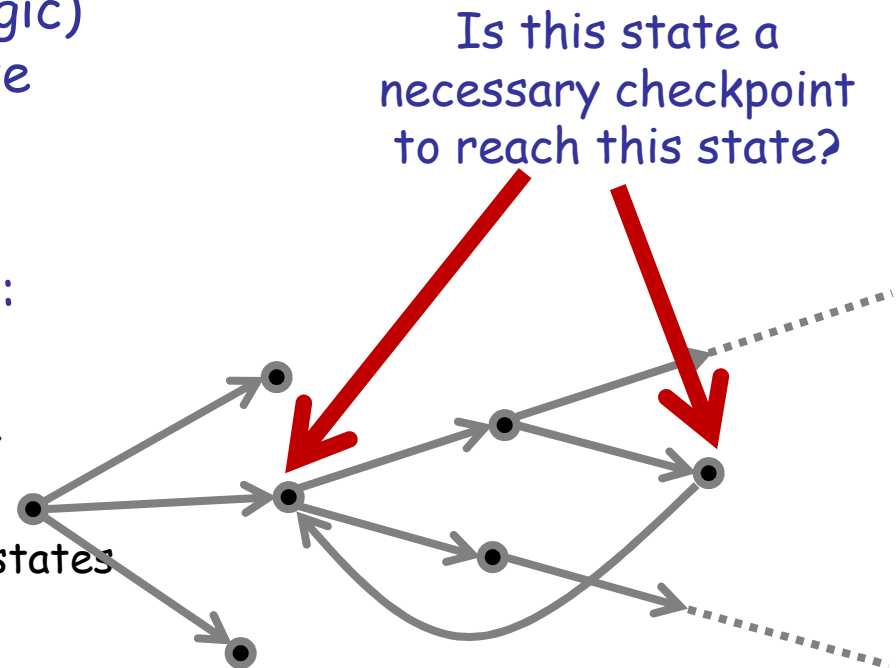
Abstract Interpretation

- Precisely relating abstract views to more concrete views of the system



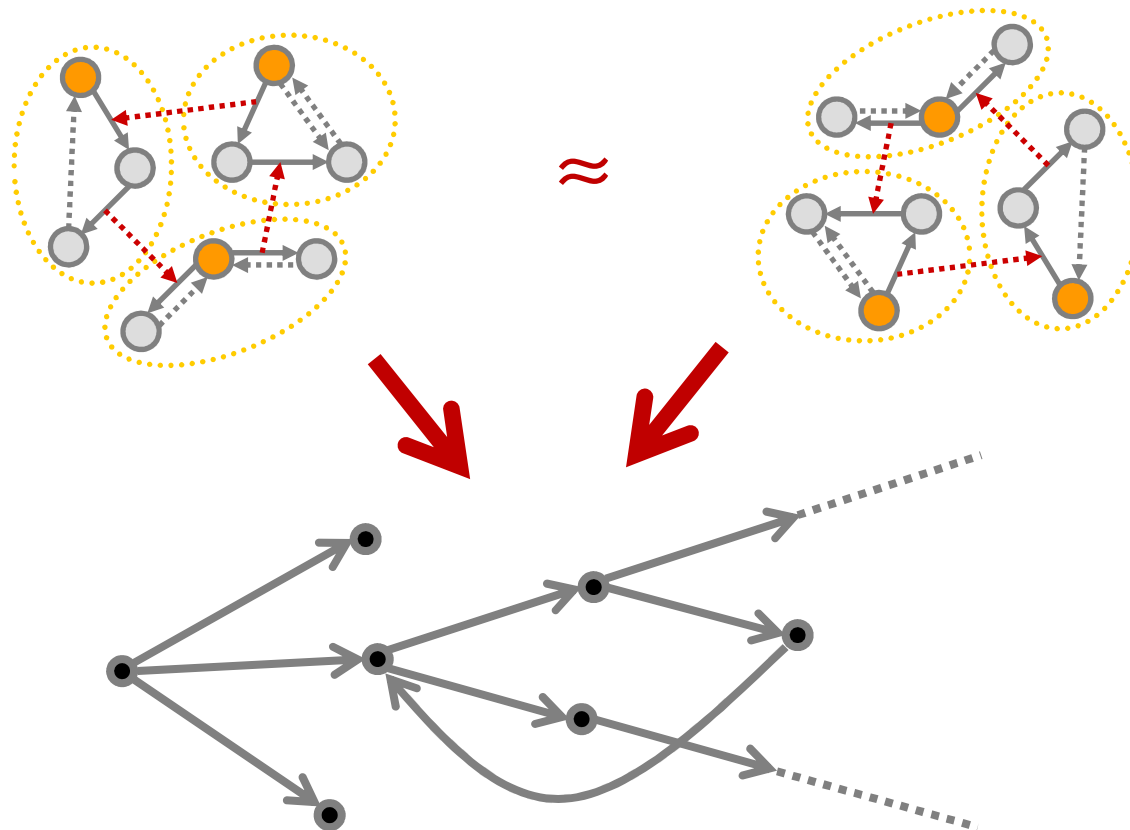
Modelchecking

- Asking questions (in Temporal Logic) about structure of a (finite) state space.
- Various flavors of modelchecking:
 - Temporal
 - About paths through state space
 - Quantitative
 - About quantitative measures of states
 - Probabilistic/Stochastic
 - About probabilities of reaching states.



Bisimulation

- Are two programs generating the same state space?
 - E.g.: Is a compact description of a system equivalent to a more detailed one in all possible environments?



Conclusions

Conclusions

- **Process Algebra**
 - An extension of automata theory to populations of interacting automata
 - Modeling the behavior of individuals in an arbitrary environment
 - Compositionality (combining models by juxtaposition)
- **Connections between modeling approaches**
 - Connecting the **discrete/concurrent/stochastic/molecular** approach
 - to the **continuous/sequential/deterministic/population** approach
- **Connecting syntax with semantics**
 - **Syntax** = model presentation (equations/programs/diagrams/blobs etc.)
 - **Semantics** = state space (generated by the syntax)
- **Ultimately, connections between analysis techniques**
 - We need (and sometimes have) good semantic techniques to analyze state spaces (e.g. calculus, but also increasingly modelchecking)
 - But we need equally good syntactic techniques to structure complex models (e.g. compositionality) and analyze them (e.g. process algebra)
- **A bright future for Computer Science and Logic in modern Biology**
 - Biology needs good analysis techniques for discrete systems analysis (modal logics, modelchecking, causality analysis, abstract interpretation, ...)

