Molecules as Automata

Representing Biochemical Systems as Collectives of Interacting Automata

Luca Cardelli

Microsoft Research

Leicester, 2007-11-30

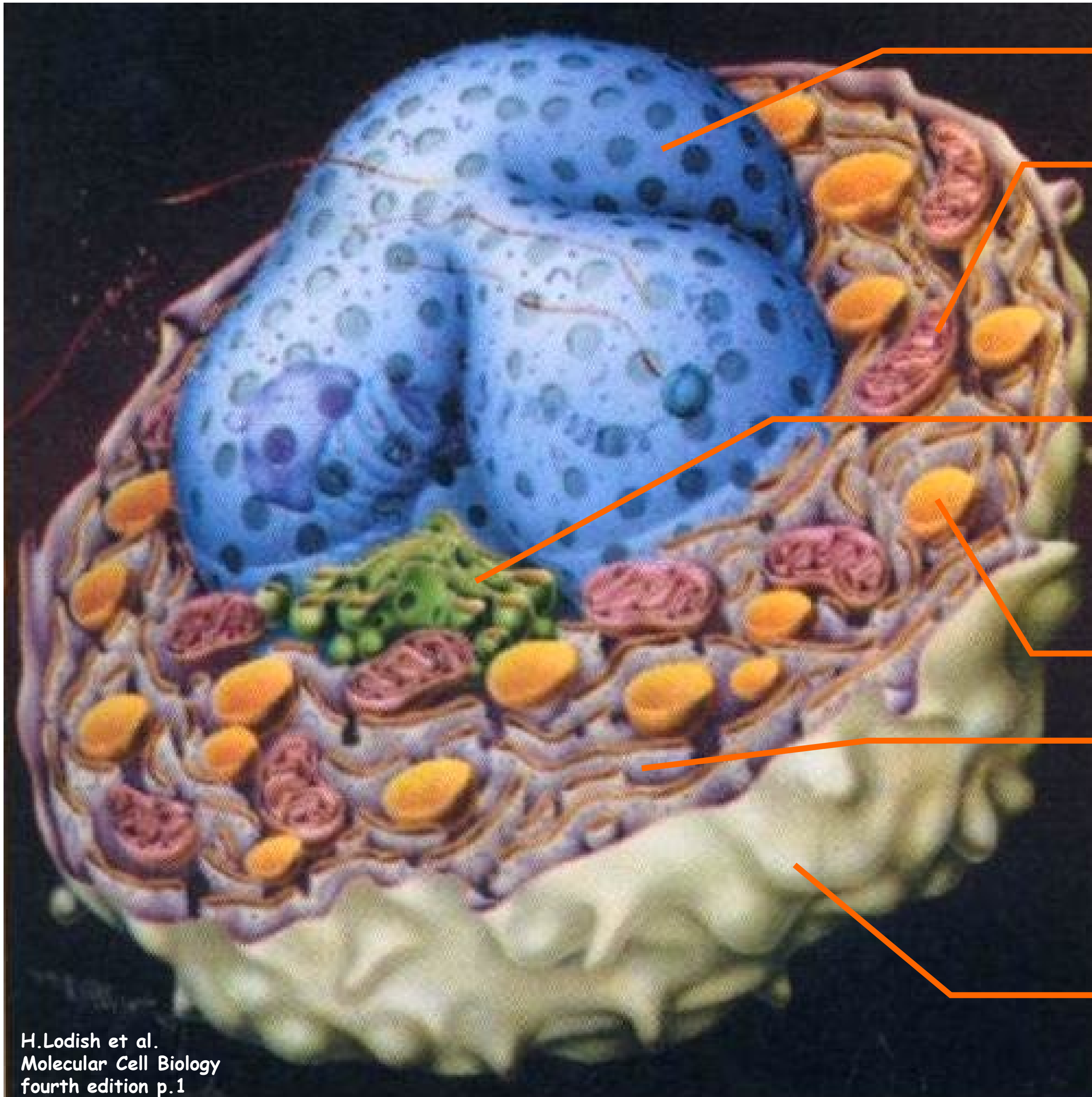
http://LucaCardelli.name

Structural Architecture

Eukaryotic Cell

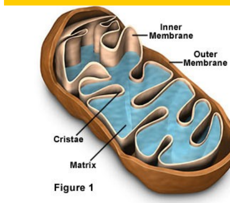
(10~100 trillion in human body)

Membranes everywhere



Nuclear membrane

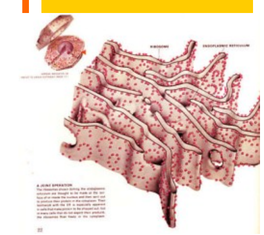
Mitochondria



Golgi

Vesicles

E.R.



Plasma membrane (<10% of all membranes)

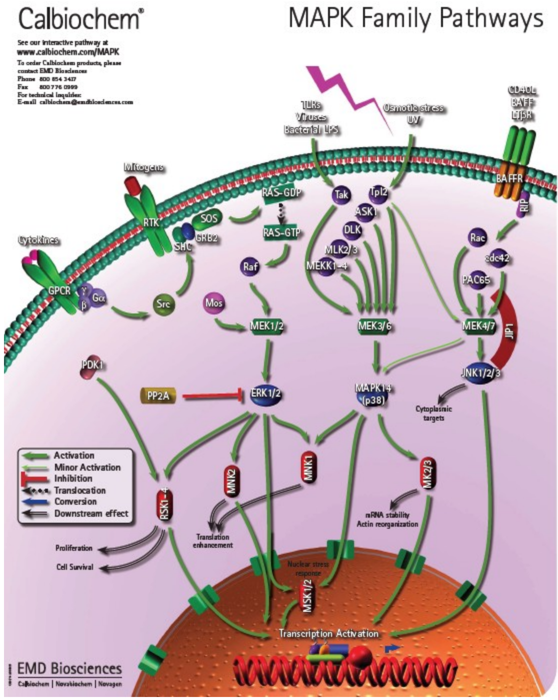
Cells Compute

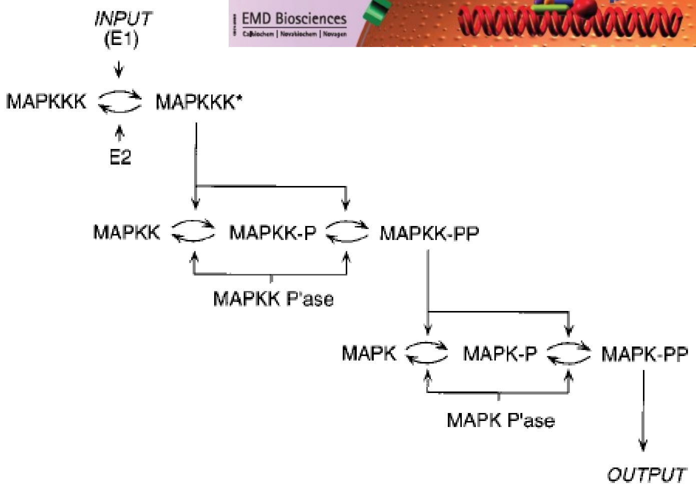
- No survival without computation!
  - Finding food
  - Avoiding predators

- How do they compute?
  - Unusual computational paradigms.
  - Proteins: do they work like electronic circuits? or process algebra?
  - 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.



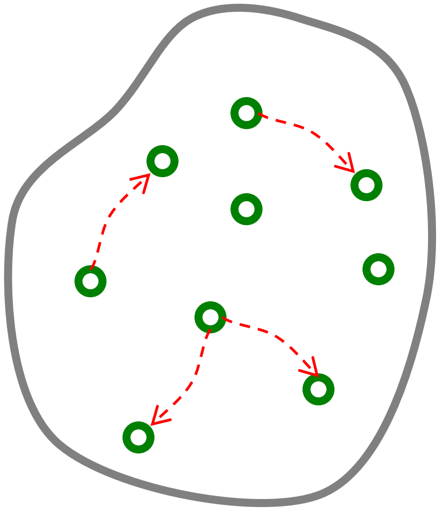


Ultrasensitivity in the mitogen-activated protein cascade, Chi-Ying F. Huang and James E. Ferrell, Jr., 1996, *Proc. Natl. Acad. Sci. USA*, 93, 10078-10083.

Stochastic Collectives

Stochastic Collectives

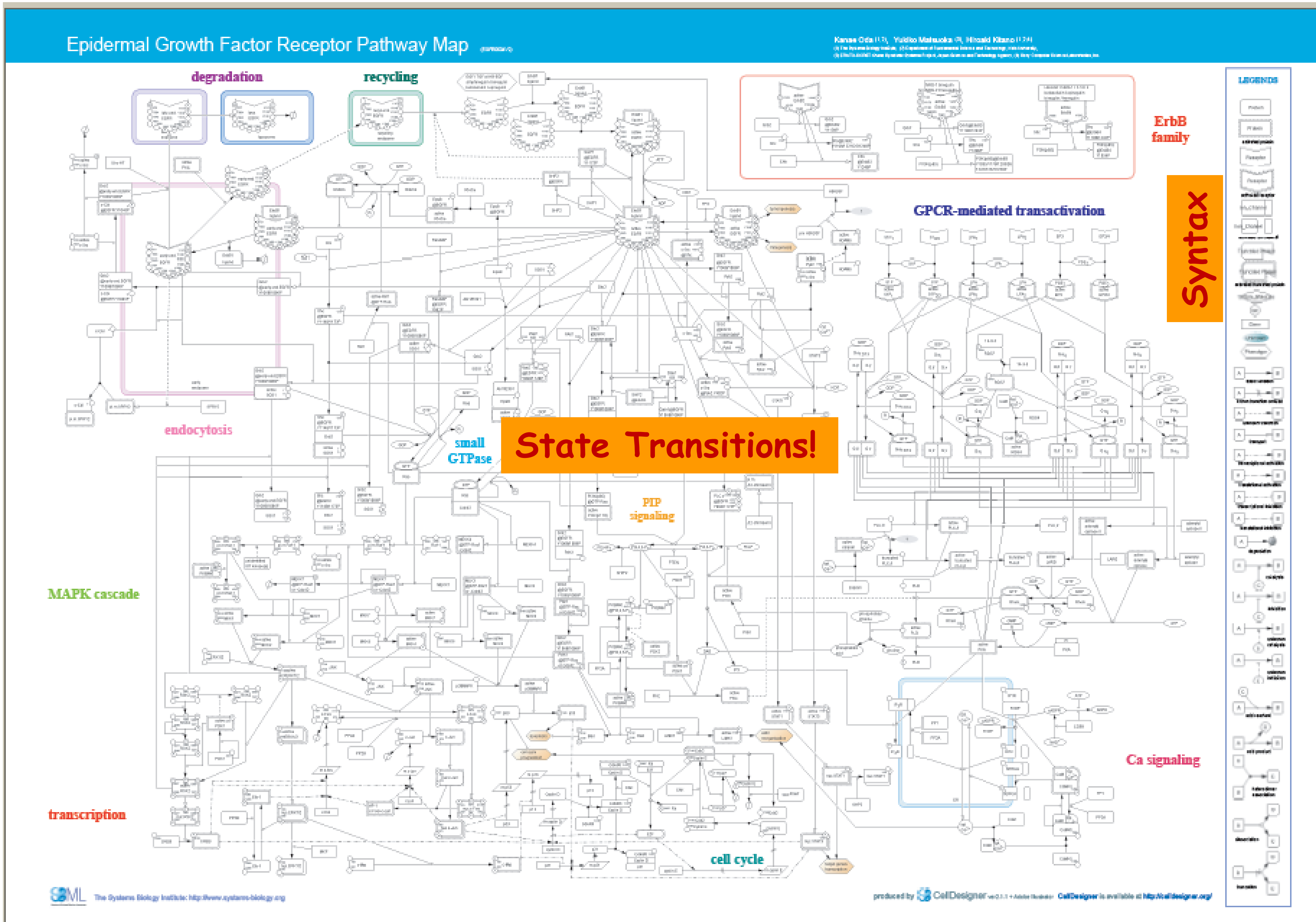
- "Collective":
  - A large set of interacting finite state automata:
    - Not quite language automata ("large set")
    - Not quite cellular automata ("interacting" but not on a grid)
    - Not quite process algebra ("collective behavior")
    - Cf. multi-agent systems and swarm intelligence



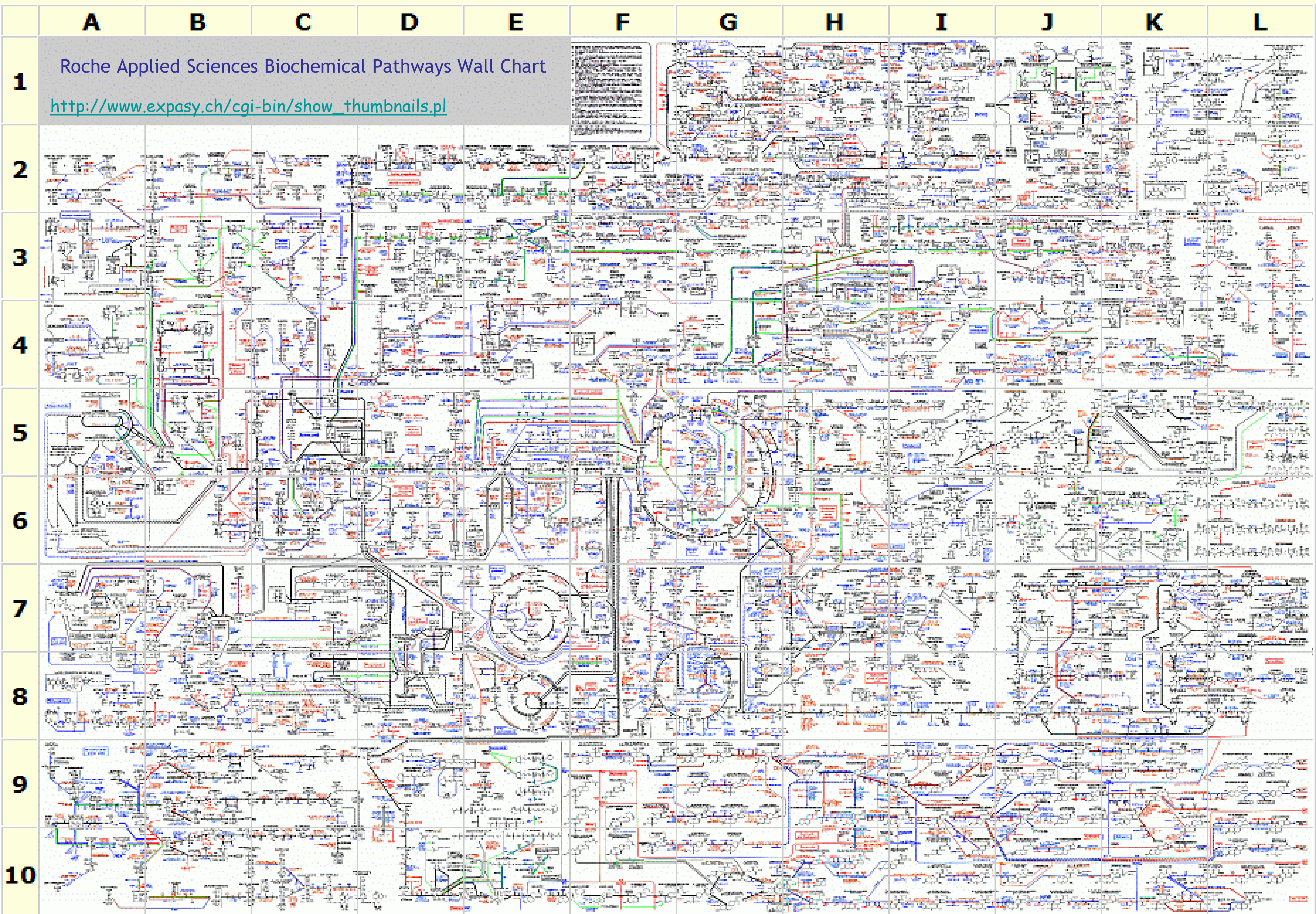
- "Stochastic":
  - Interactions have *rates*
    - Not quite discrete (hundreds or thousands of components)
    - Not quite continuous (non-trivial stochastic effects)
    - Not quite hybrid (no "switching" between regimes)

- Very much like biochemistry
  - Which is a large set of stochastically interacting molecules/proteins
  - Are proteins finite state and subject to automata-like transitions?
    - Let's say they are, at least because:
    - Much of the knowledge being accumulated in Systems Biology is described as state transition diagrams [Kitano].

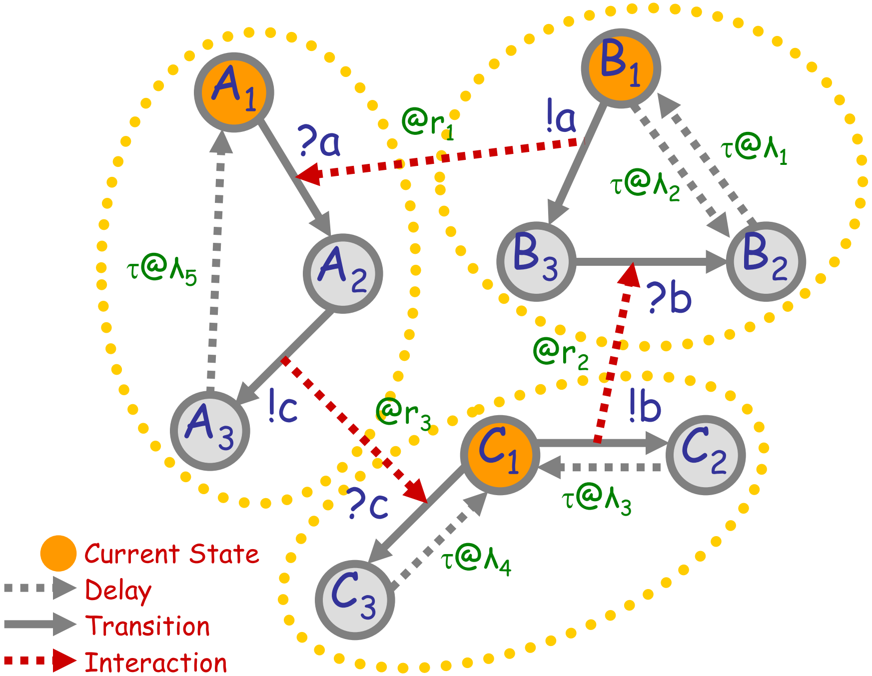
Towards Systems Biology



Compositionality (NOT!)



Interacting Automata



$\text{new } a@r_1$
$\text{new } b@r_2$
$\text{new } c@r_3$

Communication channels

$A_1 = ?a;\ A_2$
$A_2 = !c;\ A_3$
$A_3 = \tau@\lambda_5;\ A_1$

$B_1 = \tau@\lambda_2;\ B_2 + !a;\ B_3$
$B_2 = \tau@\lambda_1;\ B_1$
$B_3 = ?b;\ B_2$

$C_1 = !b;\ C_2 + ?c;\ C_3$
$C_2 = \tau@\lambda_3;\ C_1$
$C_3 = \tau@\lambda_4;\ C_2$

Automata

$A_1 \mid B_1 \mid C_1$

The system and initial state

Communicating automata: a graphical FSA-like notation for "finite state restriction-free π-calculus processes". *Interacting* automata do not even exchange values on communication.

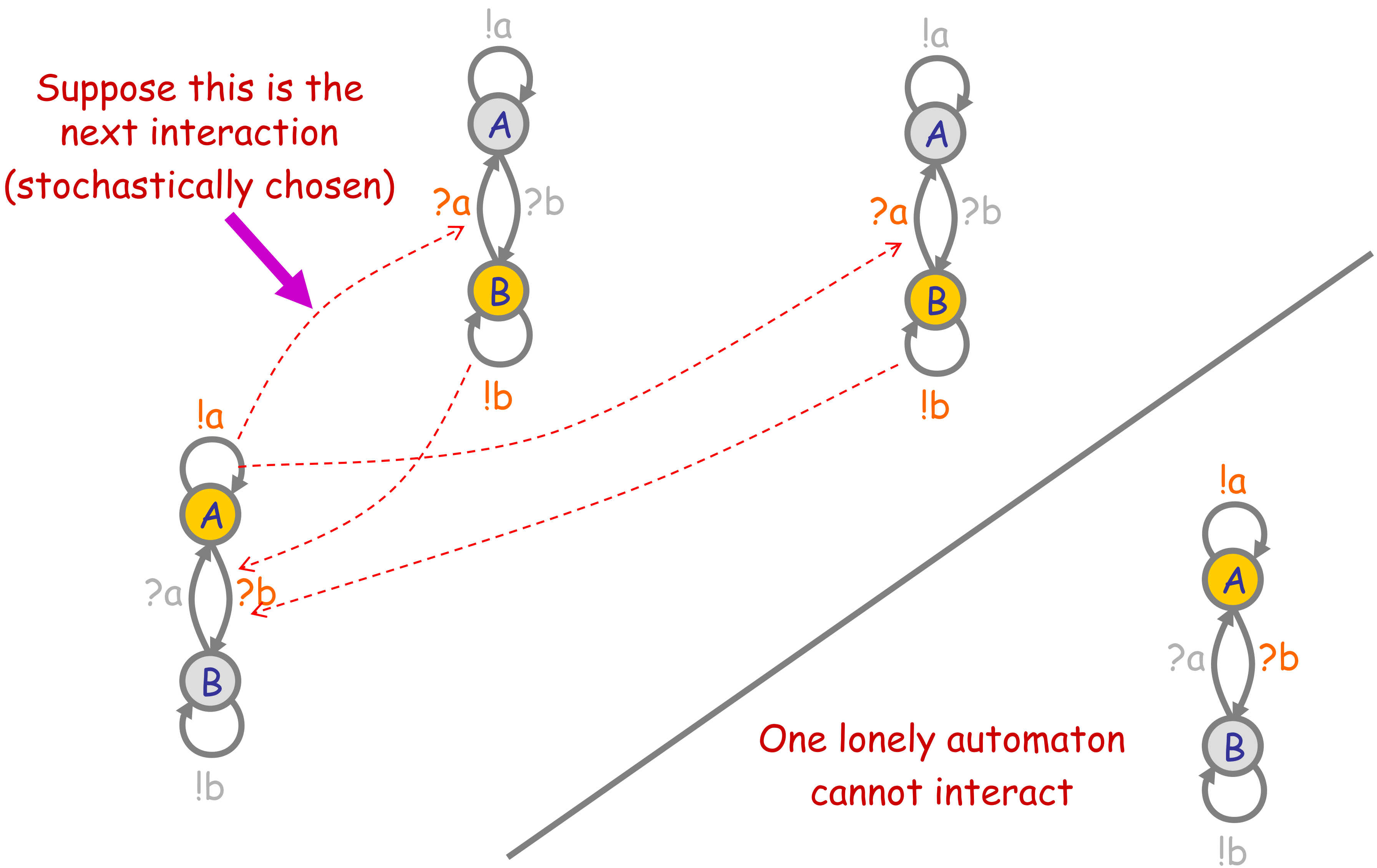
The stochastic version has *rates* on communications, and delays.

"Finite state" means: no composition or restriction inside recursion.

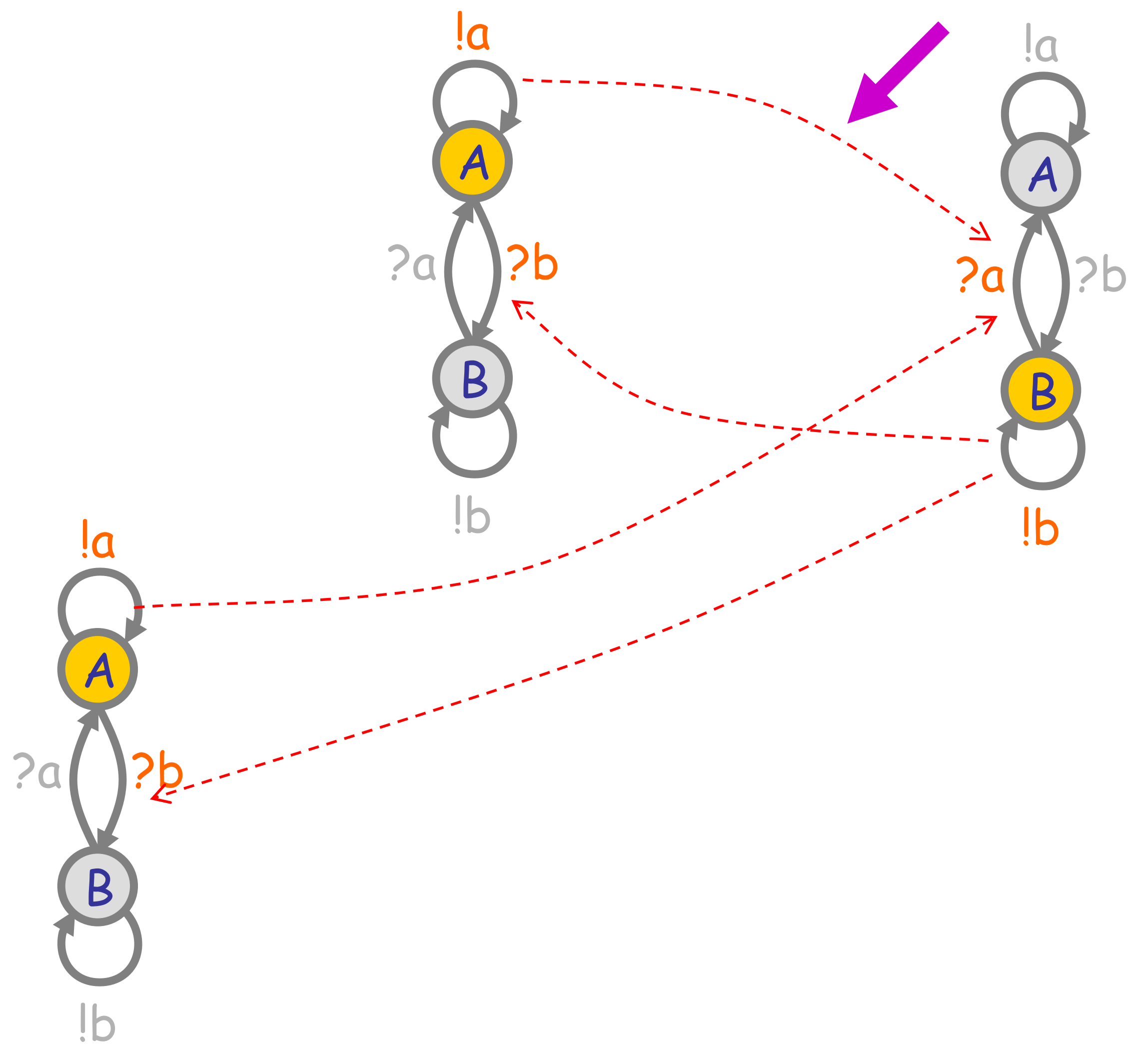
Analyzable by standard Markovian techniques, by first computing the "product automaton" to obtain the underlying finite Markov transition system. [Buchholz]

Interactions have rates. Actions DO NOT have rates.

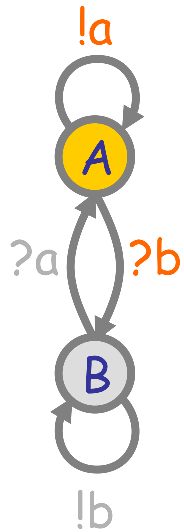
Interactions in a Population

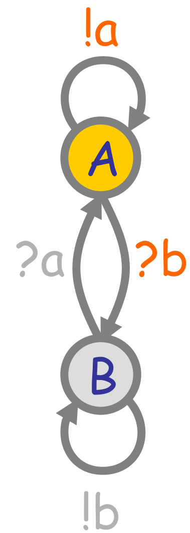


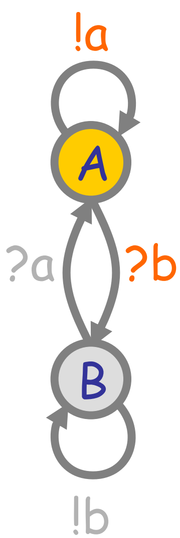
Interactions in a Population



Interactions in a Population

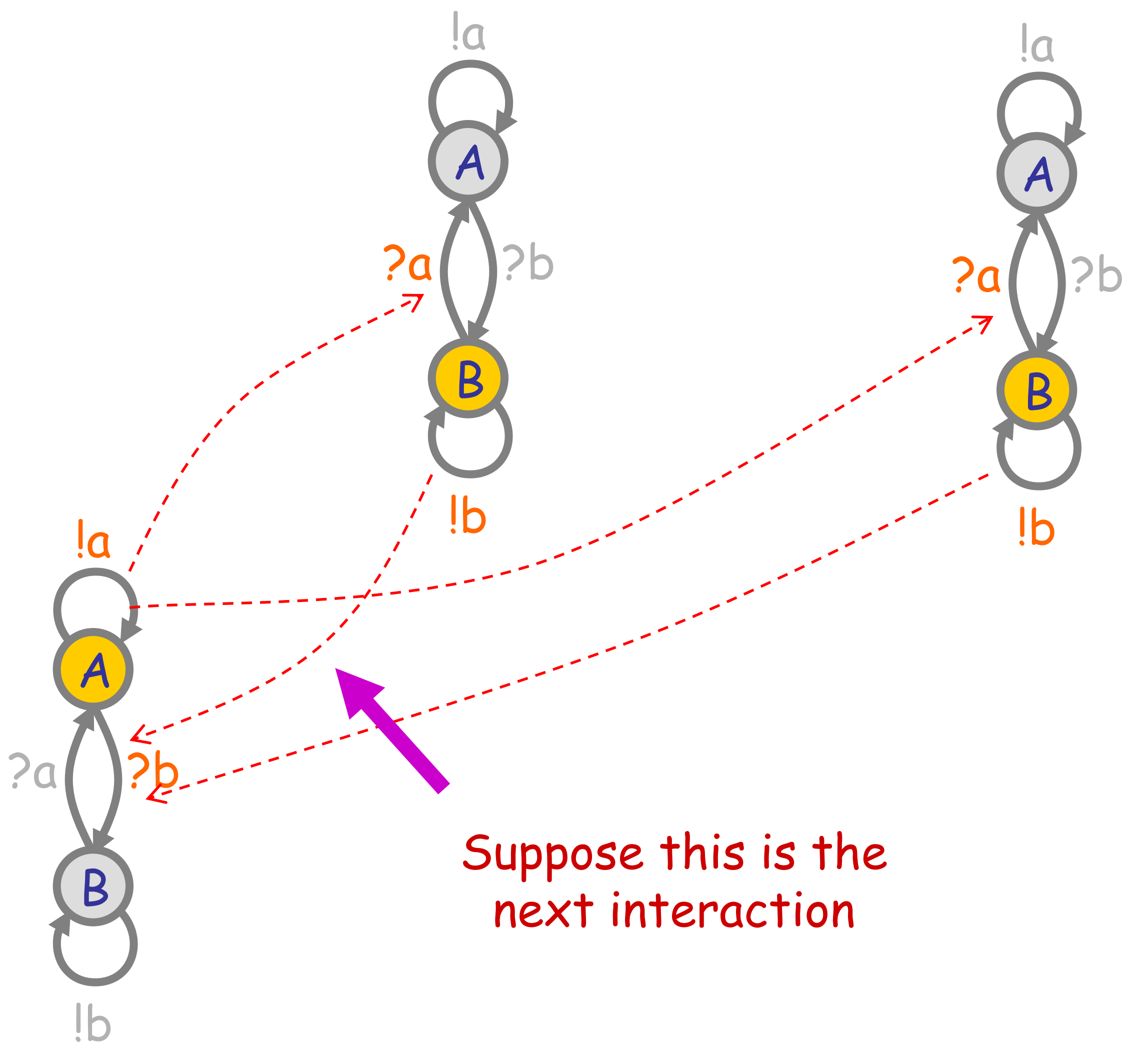






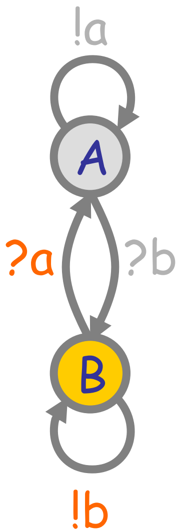
All-A stable population

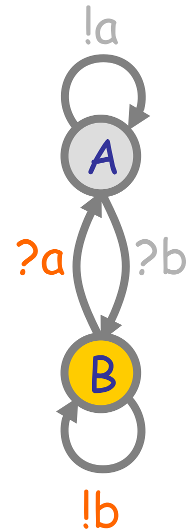
Interactions in a Population (2)

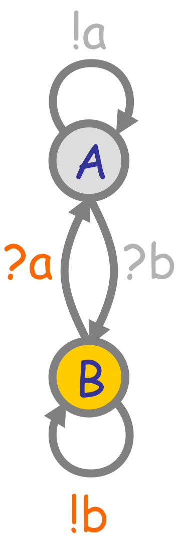


Suppose this is the next interaction

Interactions in a Population (2)



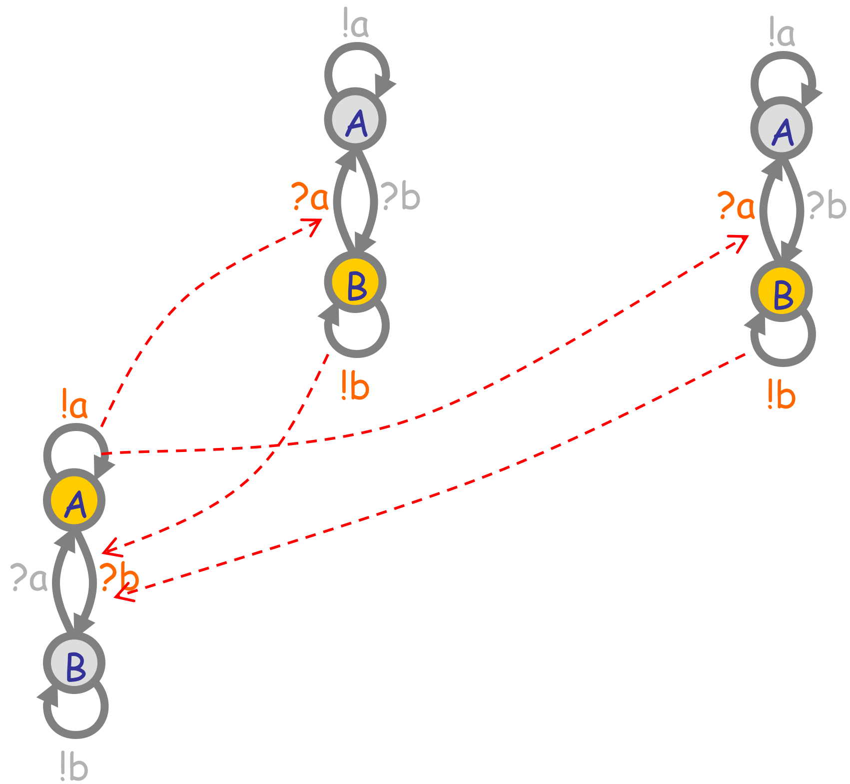


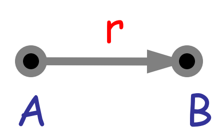


All-B stable population

Nondeterministic population behavior ("multistability")

CTMC Semantics



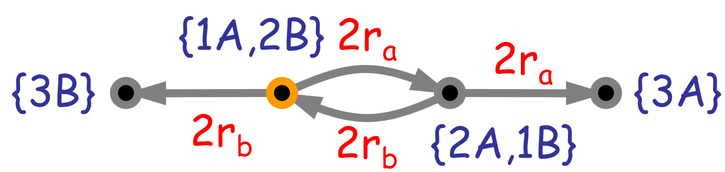


CTMC
(homogeneous) Continuous Time Markov Chain
- directed graph with no self loops
- nodes are system states
- arcs have transition rates

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

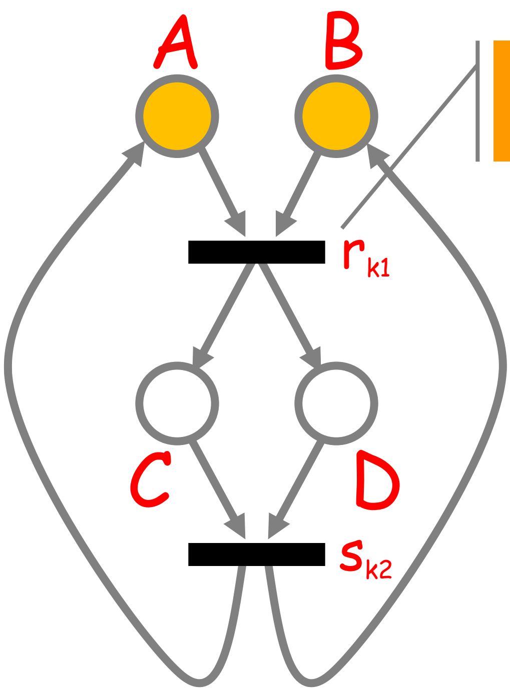
Chemistry vs. Automata

A process algebra (chemistry)

$r: A + B \rightarrow_{k1} C + D$

$s: C + D \rightarrow_{k2} A + B$

Does A become C or D?

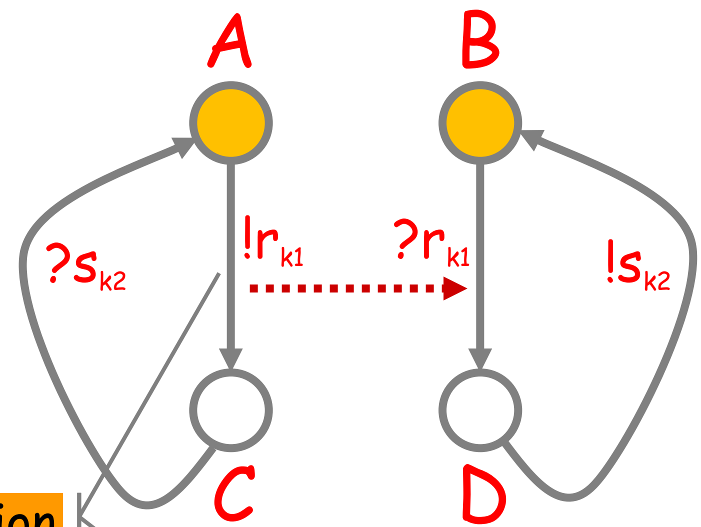


Reaction oriented

1 line per reaction

A Petri-Net-like representation. Precise and dynamic, but not modular, scalable, or maintainable.

A different process algebra (automata)



Interaction oriented

1 line per component

$A = !r_{k1}; C$

$C = ?s_{k2}; A$

$B = ?r_{k1}; D$

$D = !s_{k2}; B$

A becomes C not D!

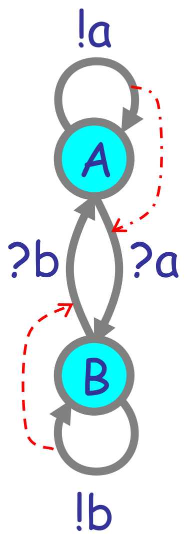
A compositional graphical representation (precise, dynamic *and* modular) and the corresponding calculus.

The same "model"

Maps to a CTMC	Maps to a CTMC

Groupies and Celebrities

Groupies and Celebrities



Celebrity

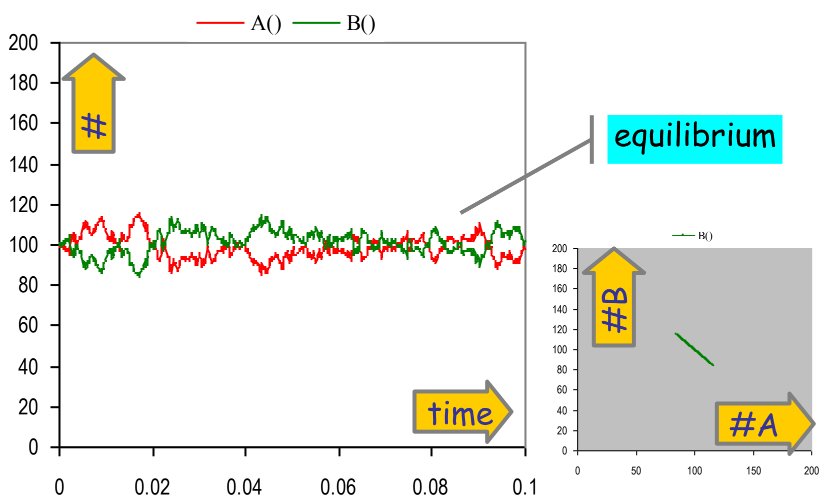
(does not want to be like somebody else)

```
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())
```

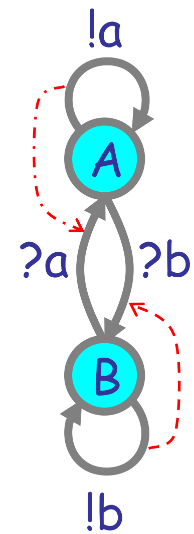
a@1.0

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.



Groupie

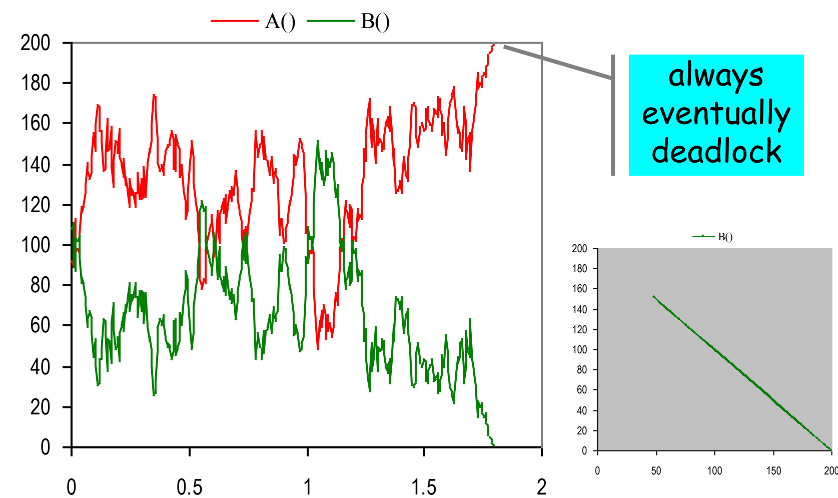
(wants to be like somebody different)

```
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:

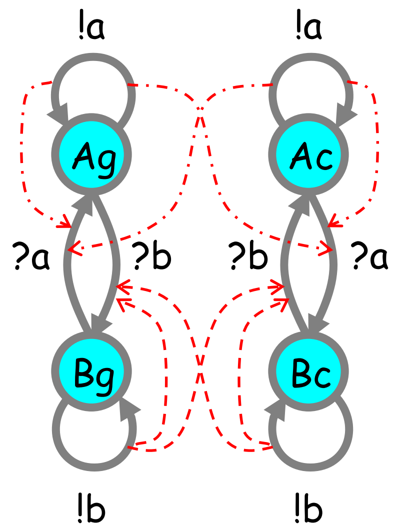


Unstable because within an A majority, an A has difficulty finding a B to emulate, but the few B's have plenty of A's to emulate, so the majority may switch to B. Leads to deadlock when everybody is in the same state and there is nobody different to emulate.

Both Together

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

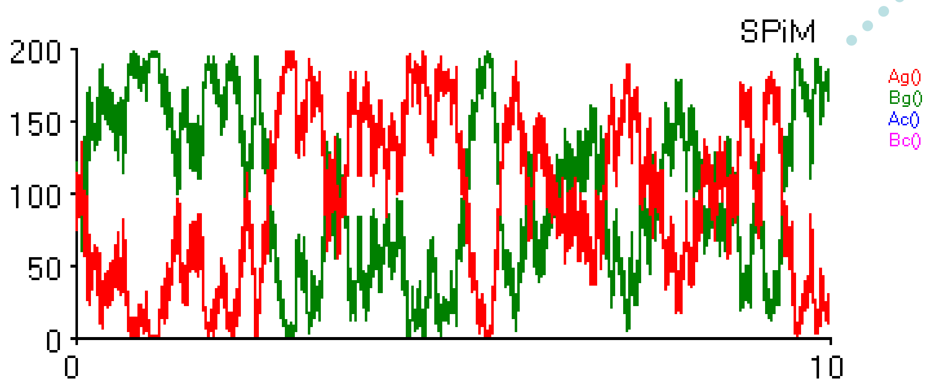
Many Groupies



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

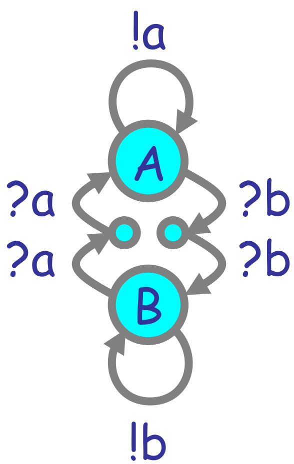


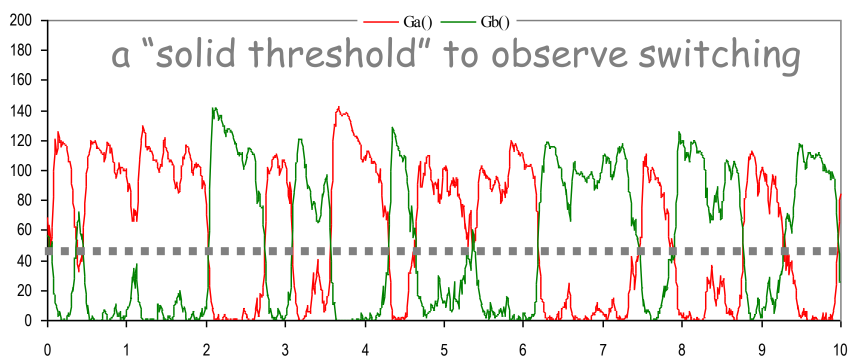
A tiny bit of "noise" can make a huge difference

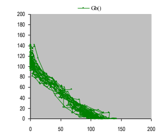
Regularity can arise not far from chaos

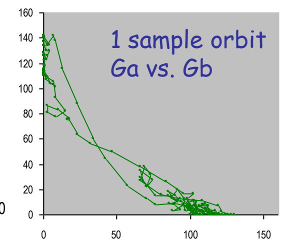
Hysteric Groupies

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

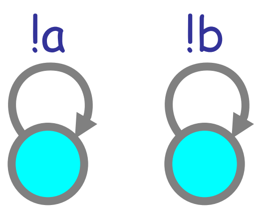








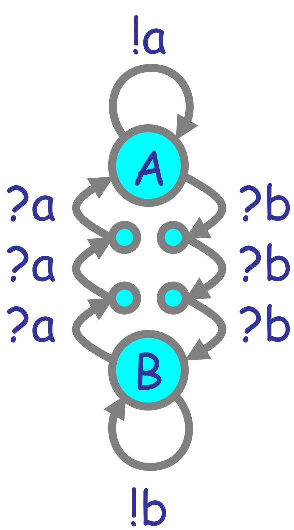

```
directive sample 10.0 1000
directive plot Ga(); Gb()

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())
```

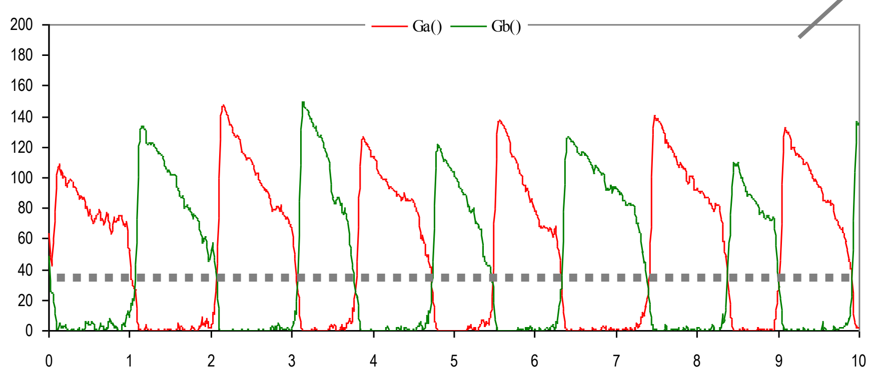


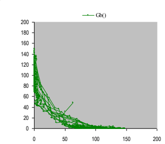
(With doping to break deadlocks)

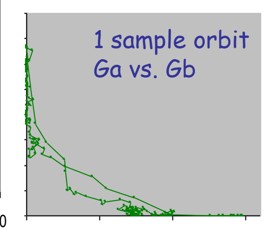
N.B.: It will not oscillate without doping (noise)

"regular" oscillation





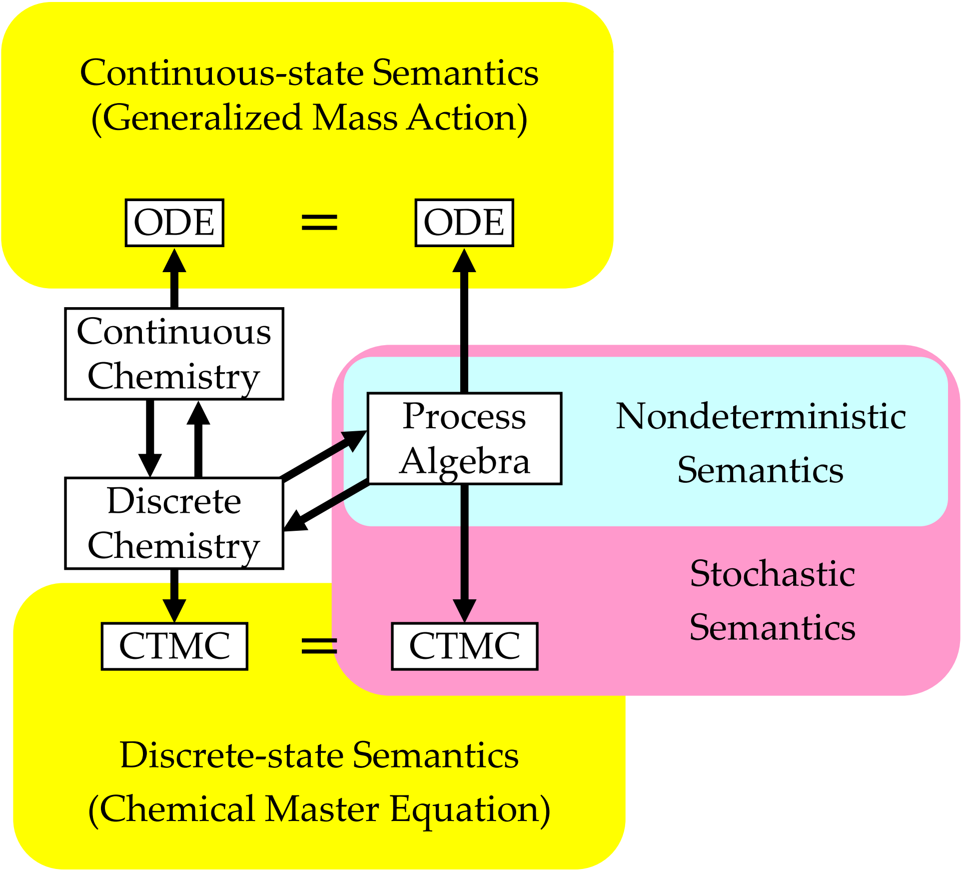





```
directive sample 10.0 1000
directive plot Ga(); Gb()

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())
```

Semantics of Collective Behavior

The Two Semantic Sides of Chemistry

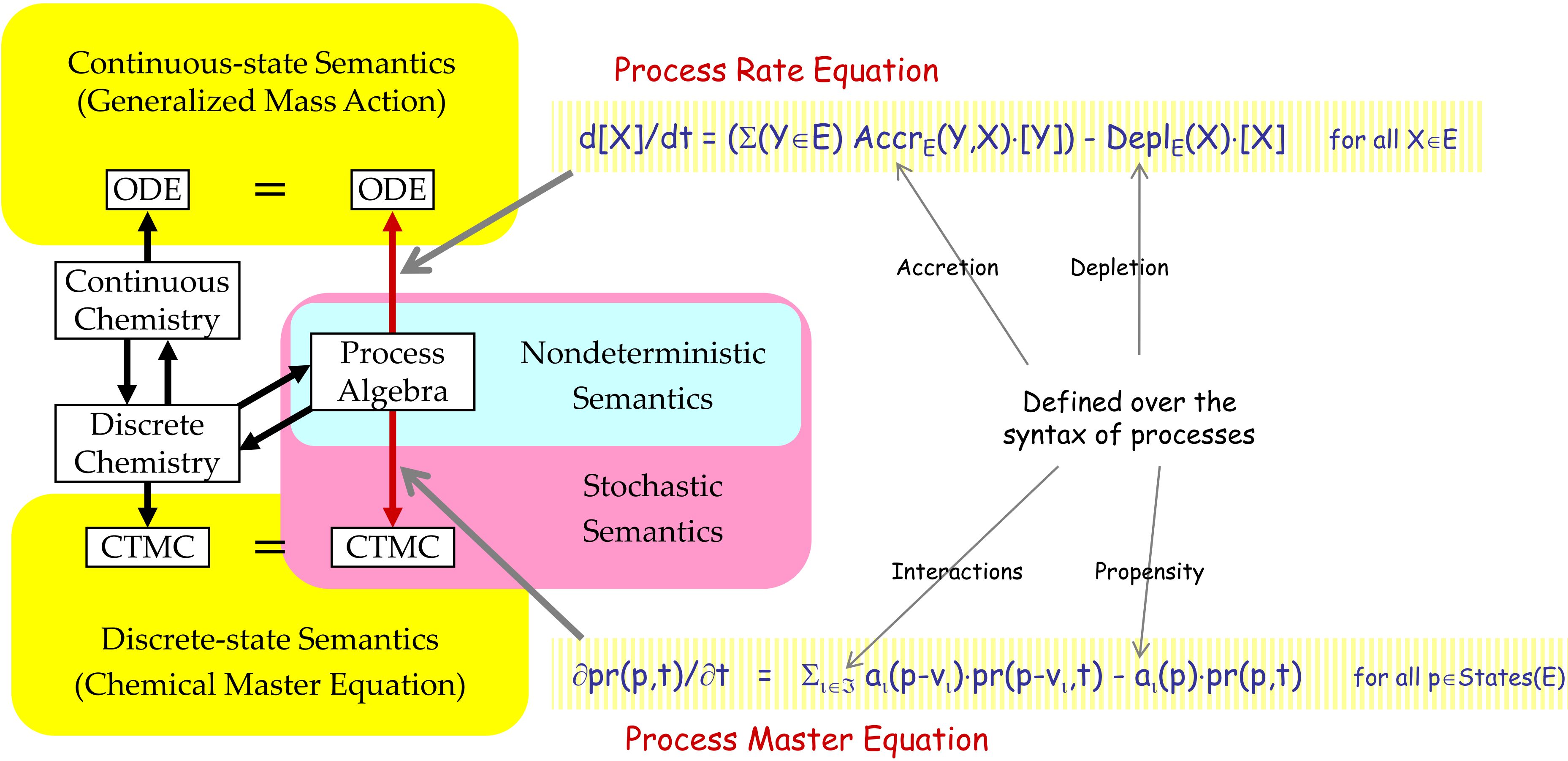


These diagrams commute via appropriate maps.

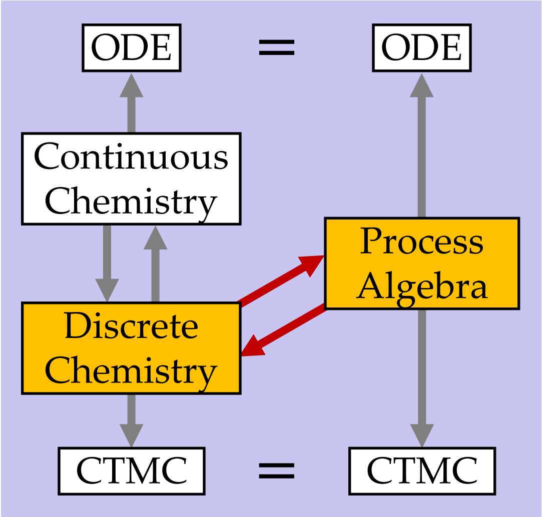
L. Cardelli: "On Process Rate Semantics" (TCS)

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

Quantitative Process Semantics



Stochastic Processes & Discrete Chemistry



Chemical Reactions

$A \to^r B_1 + \ldots + B_n$ (n≥0)	Unary Reaction	$d[A]/dt = -r[A]$	Exponential Decay
$A_1 + A_2 \to^r B_1 + \ldots + B_n$ (n≥0)	Hetero Reaction	$d[A_i]/dt = -r[A_1][A_2]$	Mass Action Law
$A + A \to^r B_1 + \ldots + B_n$ (n≥0)	Homeo Reaction	$d[A]/dt = -2r[A]^2$	Mass Action Law

(assuming $A \neq B_i \neq A_j$ for all i,j)

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!

Reactions have rates. Molecules *do not* have rates.

Trimolecular reactions:

$A + B + C \to^r D$

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

$A + B \leftrightarrow AB$

$AB + C \to D$

Enzymatic reactions:

$S \xrightarrow{E\ r} P$

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

$E + S \leftrightarrow ES$

$ES \to P + E$

Chemical Ground Form (CGF)

$E ::= 0 \;\vdots\; X{=}M, E$ — Reagents

$M ::= 0 \;\vdots\; \pi;P \oplus M$ — Molecules

$P ::= 0 \;\vdots\; X \mid P$ — Solutions

$\pi ::= \tau_{(r)} \;\vdots\; ?a_{(r)} \;\vdots\; !a_{(r)}$ — Actions (delay, input, output)

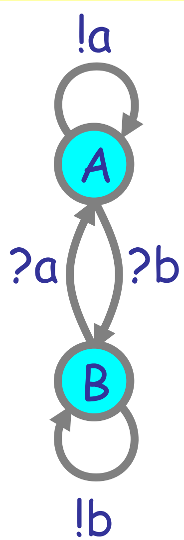
$CGF ::= E,P$ — Reagents plus Initial Conditions

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

Interacting Automata
+ dynamic forking

(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)
- 0 is the null solution ($P|0 = 0|P = P$) and null molecule ($M\oplus 0 = 0 \oplus M = M$)
- Each X in E is a distinct *species*
- Each name a is assigned a fixed rate r: $a_{(r)}$



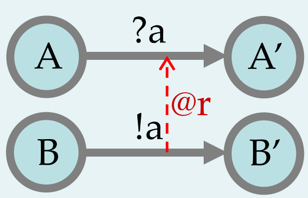
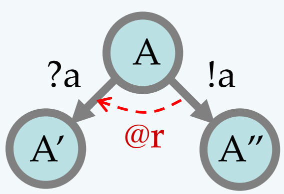
Ex: Interacting Automata
(= finite-control CGFs: they use "|" only in initial conditions):

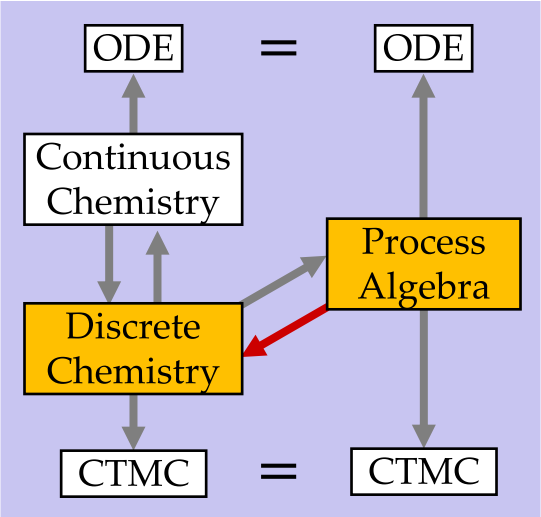
$A = !a;A \oplus ?b;B$ — Automaton in state A

$B = !b;B \oplus ?a;A$ — Automaton in state B

$A|A|B|B$ — Initial conditions: 2A and 2B

From Reagents to Reactions (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 Reagents to Reactions: Ch(E)

$E ::= 0 \;:\; X{=}M, E$ Reagents

$M ::= 0 \;:\; \pi;P \oplus M$ Molecules

$P ::= 0 \;:\; X \mid P$ Solutions

$\pi ::= \tau_{(r)} \;:\; ?a_{(r)} \;:\; !a_{(r)}$ Interactions (delay, input, output)

$CGF ::= E,P$ Reagents plus Initial Conditions

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

Chemical reactions for E,P: (N.B.: ‹...› are reaction tags to obtain multiplicity of reactions, and P is P with all the | changed to +)

Ch(E) :=

$\{(\langle X.i\rangle: X \to^{r} P)$ *s.t.* $E.X.i = \tau_{(r)};P\} \cup$

$\{(\langle X.i,Y.j\rangle: X + Y \to^{r} P + Q)$ *s.t.* $X \neq Y$, $E.X.i = ?a_{(r)};P$, $E.Y.j = !a_{(r)};Q\} \cup$

$\{(\langle X.i,X.j\rangle: X + X \to^{2r} P + Q)$ *s.t.* $E.X.i = ?a_{(r)};P$, $E.X.j = !a_{(r)};Q)\rangle \in E\}$

Initial conditions for P:

Ch(P) := P

From Reactions to Reagents (by example)

v_1: A+B $\rightarrow_{k_1}$ C+C

v_2: A+C $\rightarrow_{k_2}$ D

v_3: C $\rightarrow_{k_3}$ E+F

v_4: F+F $\rightarrow_{k_4}$ B

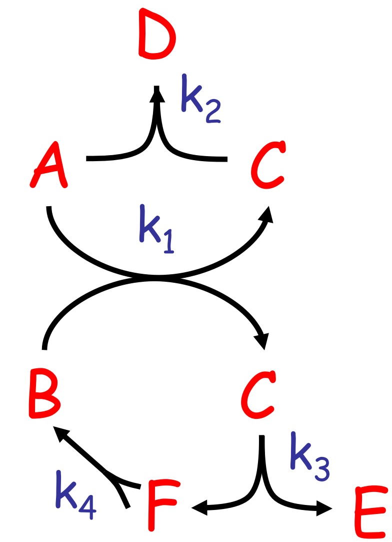
Interaction Matrix

channels and rates (1 per reaction)

Half-rate for homeo reactions

definitions (1 per species)

	$v_{1(k1)}$	$v_{2(k2)}$	$v_{3(k3)}$	$v_{4(k4/2)}$
A	?;(C|C)	?;D		
B	!;0			
C		!;0	τ;(E|F)	
D				
E				
F				?;B !;0



1: Fill the matrix by columns:

Degradation reaction v_i: X $\rightarrow_{k_i}$ P_i

add τ;P_i to ‹X,v_i›.

Hetero reaction v_i: X+Y $\rightarrow_{k_i}$ P_i

add ?;P_i to ‹X,v_i› and !;0 to ‹Y,v_i›

Homeo reaction v_i: X+X $\rightarrow_{k_i}$ P_i

add ?;P_i and !;0 to ‹X,v_i›

2: Read the result by rows:

A = ?$v_{1(k1)}$;(C|C) $\oplus$?$v_{2(k2)}$;D

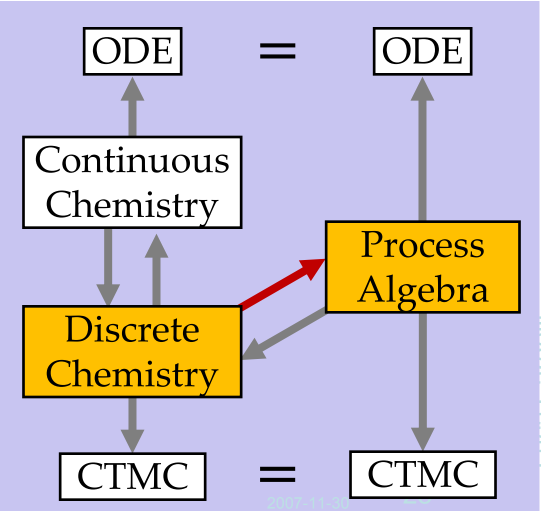
B = !$v_{1(k1)}$;0

C = !$v_{2(k2)}$;0 $\oplus$ τ_{k3};(E|F)

D = 0

E = 0

F = ?$v_{4(k4/2)}$;B $\oplus$!$v_{4(k4/2)}$;0

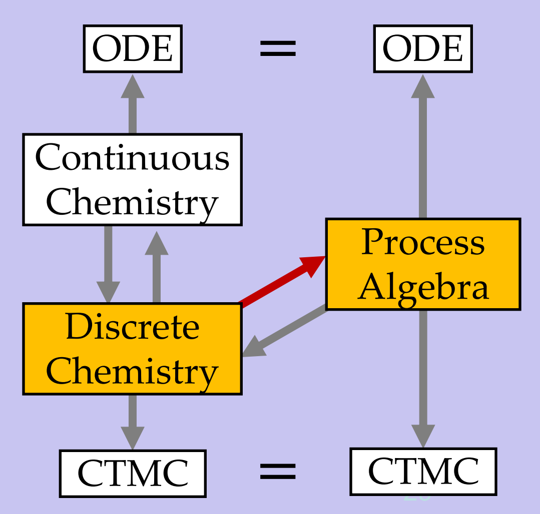


From Reactions to Reagents: Pi(C)

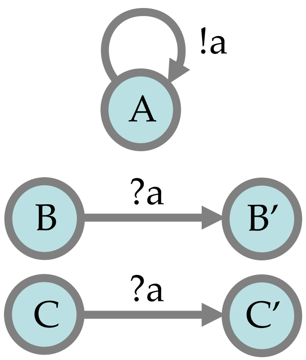
v:	$X \to^r Y_1 + \ldots + Y_n + 0$		Unary Reaction
v:	$X_1 + X_2 \to^r Y_1 + \ldots + Y_n + 0$		Binary Reaction

From uniquely-labeled (v:) chemical reactions *C* to a CGF Pi(*C*):

$$
\begin{aligned}
Pi(C) = \{(X = \; & \oplus((v{:}\ X \to^k P) \in C)\ \mathit{of}\ (\tau_{(k)};P) \quad \oplus \\
& \oplus((v{:}\ X+Y \to^k P) \in C \text{ and } Y \neq X)\ \mathit{of}\ (?v_{(k)};P) \quad \oplus \\
& \oplus((v{:}\ Y+X \to^k P) \in C \text{ and } Y \neq X)\ \mathit{of}\ (!v_{(k)};0) \quad \oplus \\
& \oplus((v{:}\ X+X \to^k P) \in C)\ \mathit{of}\ (?v_{(k/2)};P \oplus !v_{(k/2)};0) \quad) \\
& \mathit{s.t.}\ X \text{ is a species in } C\}
\end{aligned}
$$



Entangled vs Detangled



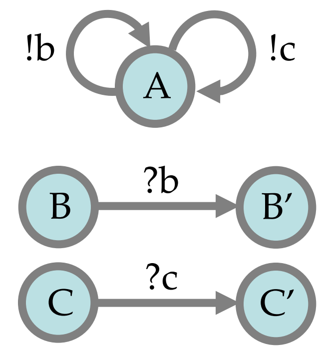
$a: A+B \to^{r} A+B'$

$a: A+C \to^{r} A+C'$

(a@r)

A = !a;A
B = ?a;B′
C = ?a;C′
B′ = 0
C′ = 0

Entangled: Two reactions
on one channel



$b: A+B \to^{r} A+B'$

$c: A+C \to^{r} A+C'$

(b@r)
(c@r)

A = !b;A ⊕ !c;A
B = ?b;B′
C = ?c;C′
B′ = 0
C′ = 0

Detangled: Two reactions
on two separate channels

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

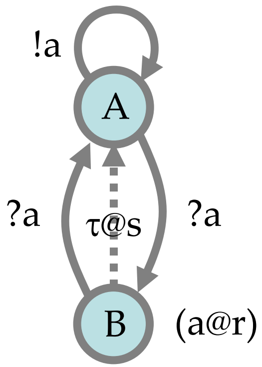
No process algebra equivalence is like this!

Detangled processes are in simple correspondence with chemistry.

Same Semantics

Could chemistry itself be that semantics?

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



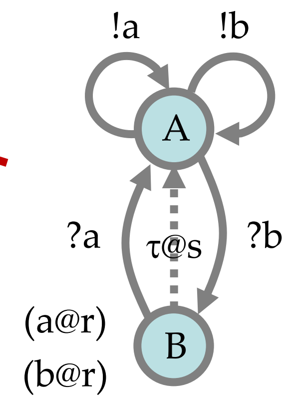
$A = !a;A \oplus ?a;B$

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

$B \to^{s} A$

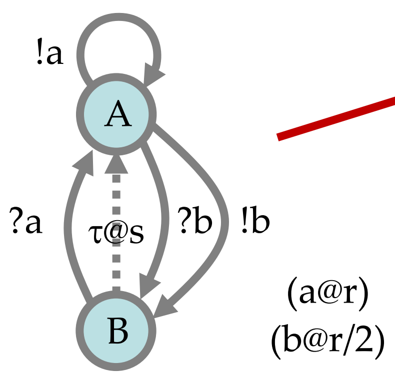
$A+B \to^{r} A+A$

$A+A \to^{2r} A+B$



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

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



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

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

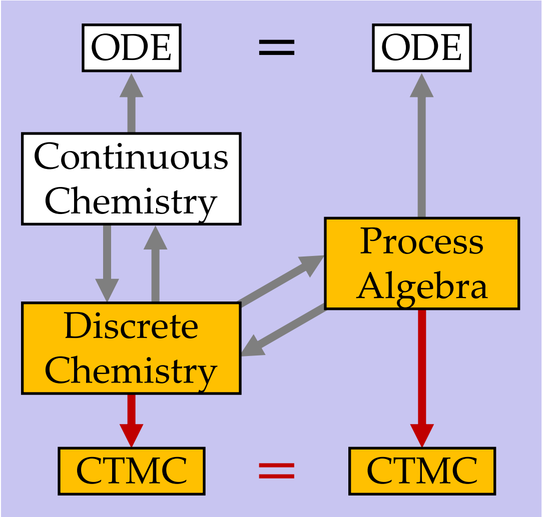
$B \to^{s} A$

$A+B \to^{r} A+A$

$A+A \to^{r} B+B$

Different reactions, but they induce the same ODEs

Discrete-State Semantics

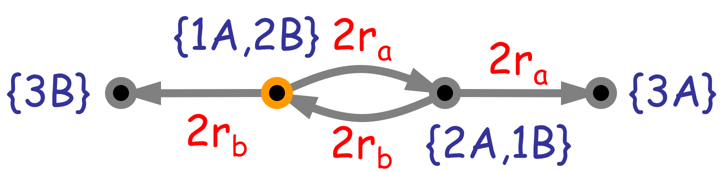


Discrete Semantics of Reactions

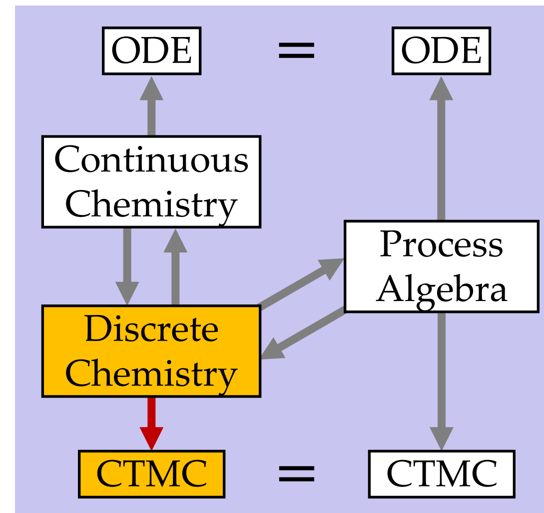
$A+B \to^{r} A+A$

$A+B \to^{r} B+B$

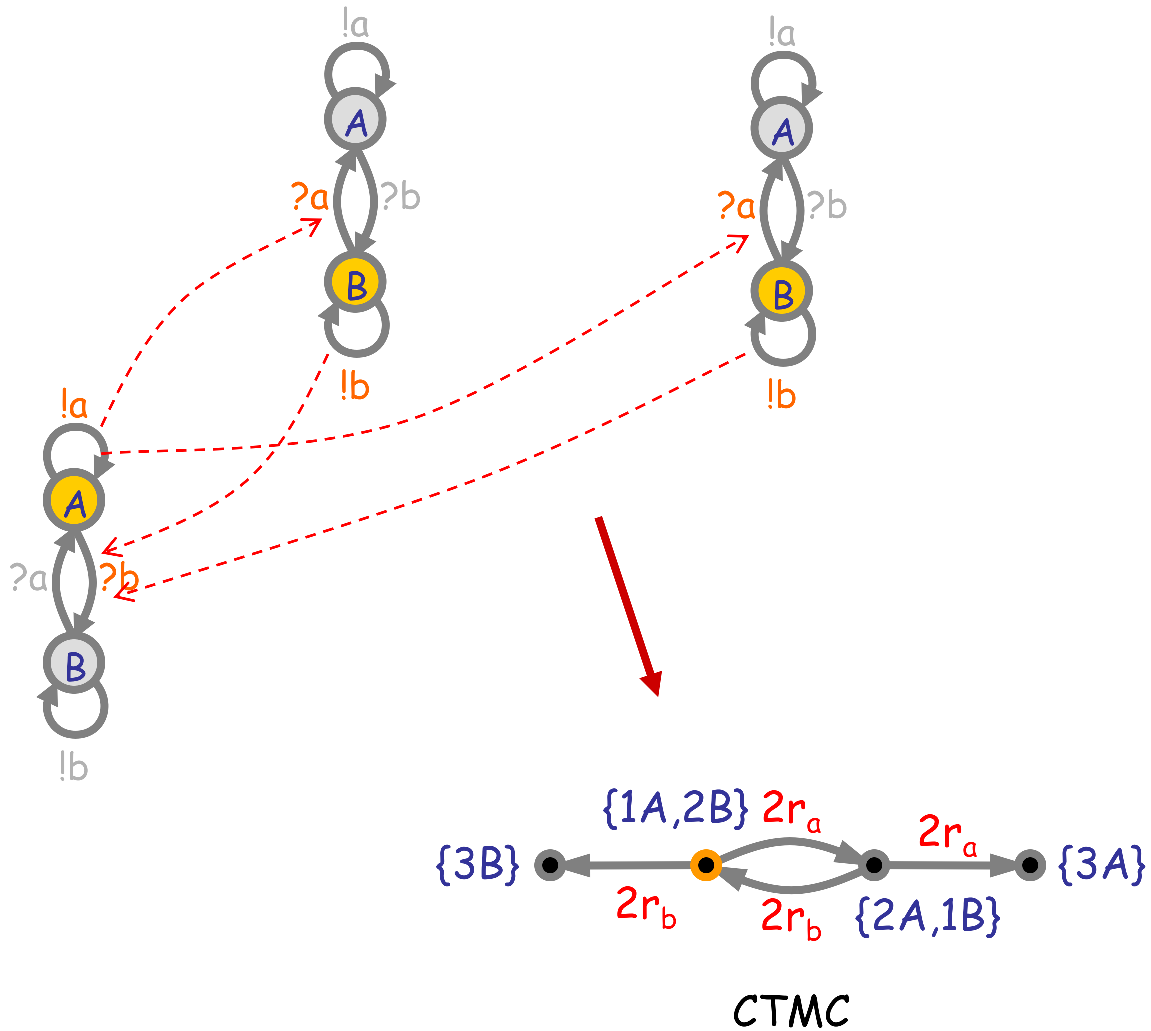
$A+B+B$

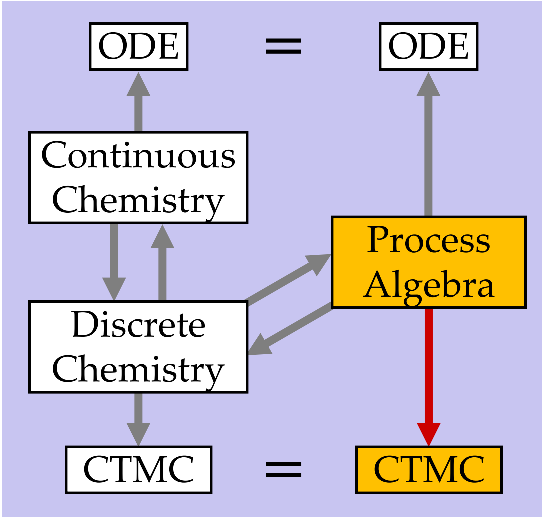


CTMC



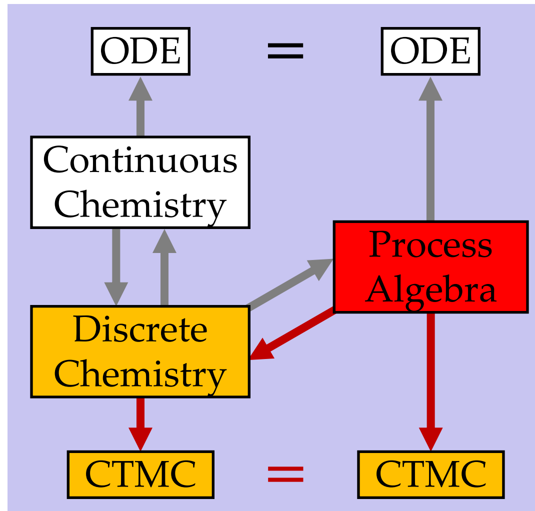
Discrete Semantics of Reagents

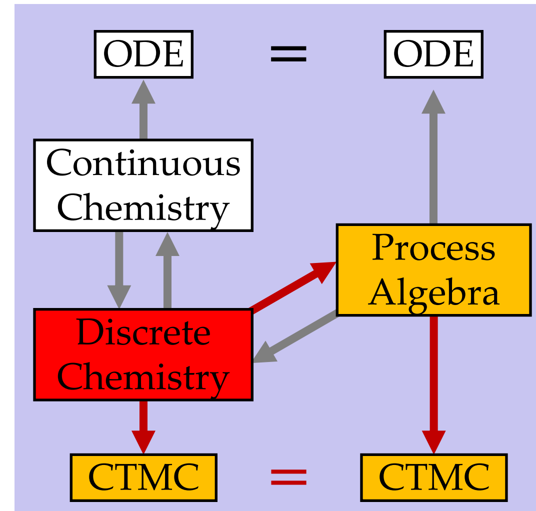




Discrete State Equivalence

- Def: ♒ is equivalent CTMC's (isomorphic graphs with same rates).
- Thm: E ♒ Ch(E)
- Thm: C ♒ Pi(C)





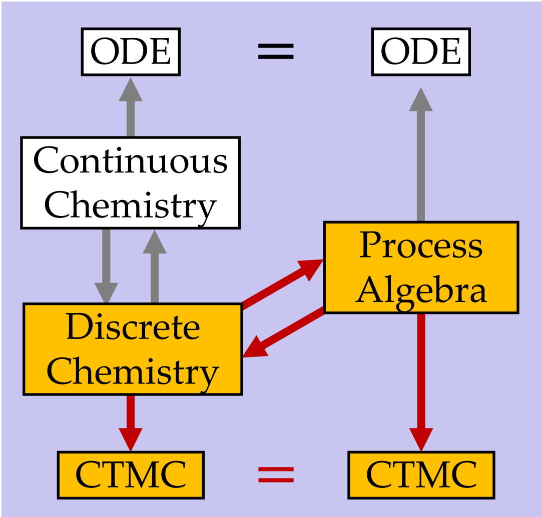
- For each E there is an E' ♒ E that is detangled (E' = Pi(Ch(E)))
- For each E in automata form there is an an E' ♒ E that is detangled and in automata form (E' = Detangle(E)).

Process Algebra = Discrete Chemistry

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

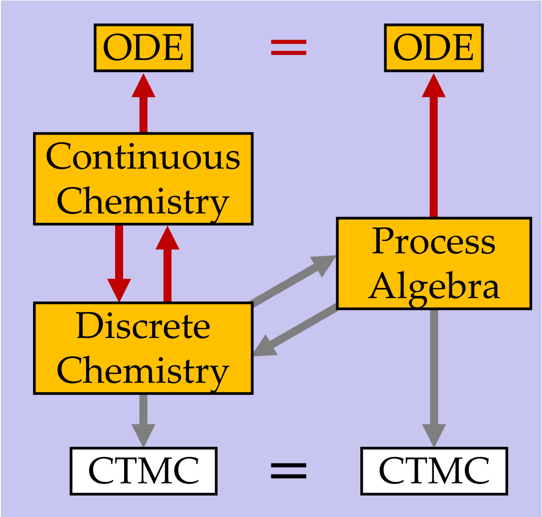
But CTMC are not the "ultimate semantics" because there are still questions of when two different CTMCs are actually equivalent (e.g. "lumping").

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



Continuous-State Semantics

(short version)



The Gillespie(?) Conversion

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

V = interaction volume

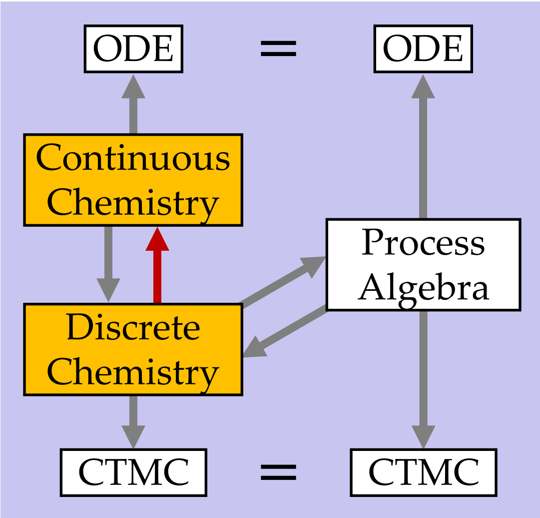
N_A = Avogadro's number

Think $\gamma = 1$

i.e. $V = 1/N_A$

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

molarity (concentration)



From Reactions to ODEs

v_1: $A+B \rightarrow_{k_1} C+C$

v_2: $A+C \rightarrow_{k_2} D$

v_3: $C \rightarrow_{k_3} E+F$

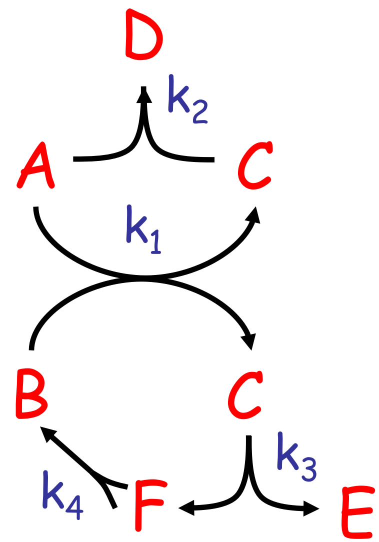
v_4: $F+F \rightarrow_{k_4} B$

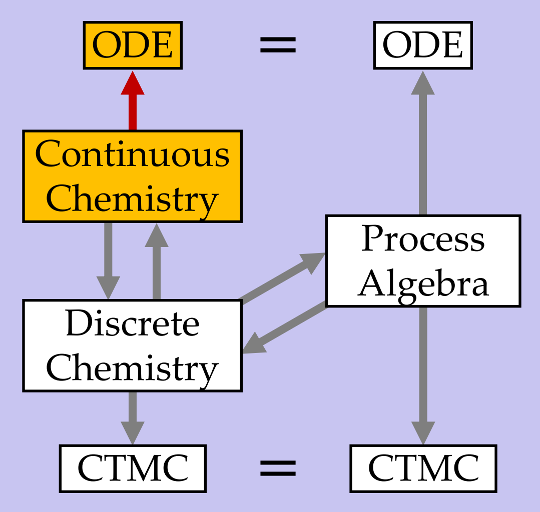
Write the coefficients by columns

Stoichiometric Matrix

reactions / species

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
X				





Quantity changes

Stoichiometric matrix

Rate laws

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

$$d[A]/dt = -l_1 - l_2$$
$$d[B]/dt = -l_1 + l_4$$
$$d[C]/dt = 2l_1 - l_2 - l_3$$
$$d[D]/dt = l_2$$
$$d[E]/dt = l_3$$
$$d[F]/dt = l_3 - 2l_4$$

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)

	l
l_1	$k_1[A][B]$
l_2	$k_2[A][C]$
l_3	$k_3[C]$
l_4	$k_4[F]^2$

X: chemical species

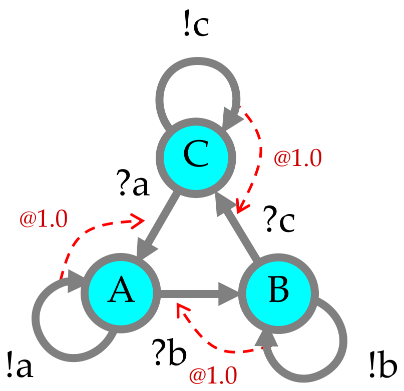
[-]: quantity of molecules

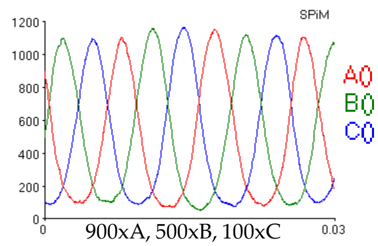
l: rate laws

k: kinetic parameters

N: stoichiometric matrix

From Processes to ODEs via Chemistry!



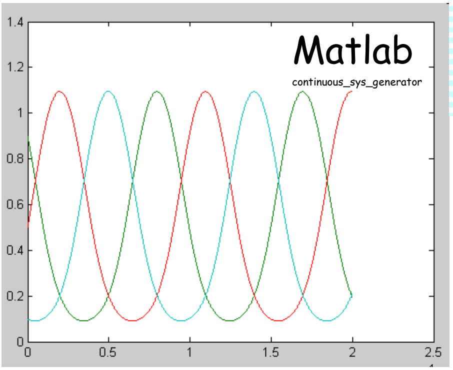



```
directive sample 0.03 1000
directive plot A(); B(); C()

new a@1.0:chan new b@1.0:chan new c@1.0:chan
let A() = do !a;A() or ?b; B()
and B() = do !b;B() or ?c; C()
and C() = do !c;C() or ?a; A()

run (900 of A() | 500 of B() | 100 of C())
```

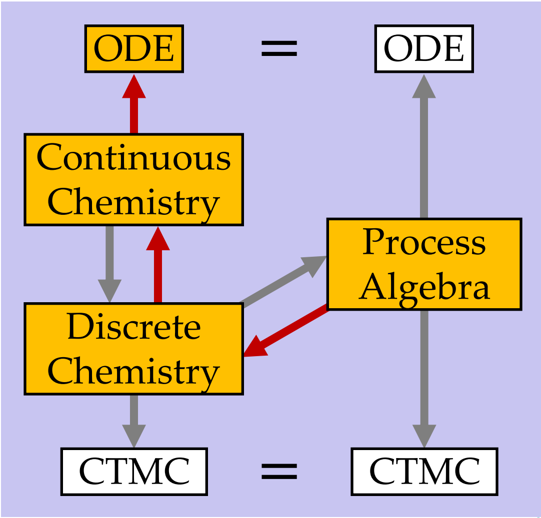
$A = !a_{(s)};A \oplus ?b_{(s)};B$
$B = !b_{(s)};B \oplus ?c_{(s)};C$
$C = !c_{(s)};C \oplus ?a_{(s)};A$

$A+B \rightarrow^{s} B+B$
$B+C \rightarrow^{s} C+C$
$C+A \rightarrow^{s} A+A$

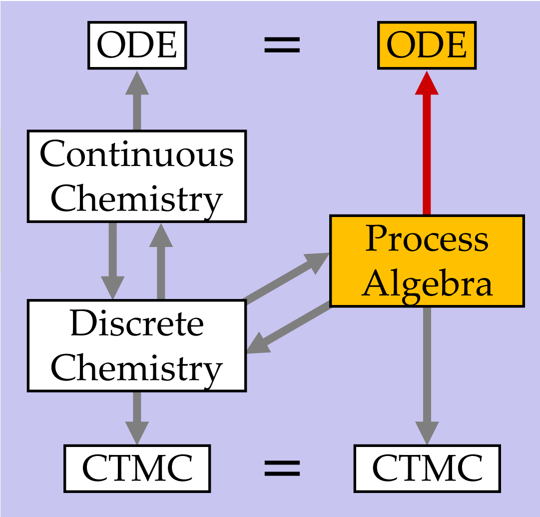
$d[A]/dt = -s[A][B]+s[C][A]$
$d[B]/dt = -s[B][C]+s[A][B]$
$d[C]/dt = -s[C][A]+s[B][C]$




```
interval/step [0:0.001:20.0]
(A)    dx1/dt = - x1*x2 + x3*x1    0.9
(B)    dx2/dt = - x2*x3 + x1*x2    0.5
(C)    dx3/dt = - x3*x1 + x2*x3    0.1
```



Processes Rate Equation



Process Rate Equation for Reagents E in volume γ

$$d[X]/dt = (\Sigma(Y \in E)\ Accr_E(Y,X) \cdot [Y]) - Depl_E(X) \cdot [X]$$
$$\text{for all } X \in E$$

"The change in process concentration (!!) for X at time t is:
- the sum over all possible (kinds of) processes Y of:
  - the concentration at time t of Y
    - times the accretion from Y to X
  - minus the concentration at time t of X
    - times the depletion of X to some other Y"

$Depl_E(X) =$
- $\Sigma(i{:}\ E.X.i=\tau_{(r)};P)\ r\ +$
- $\Sigma(i{:}\ E.X.i=?a_{(r)};P)\ r\gamma \cdot OutsOn_E(a)\ +$
- $\Sigma(i{:}\ E.X.i=!a_{(r)};P)\ r\gamma \cdot InsOn_E(a)$

$Accr_E(Y, X) =$
- $\Sigma(i{:}\ E.Y.i=\tau_{(r)};P)\ \#X(P) \cdot r\ +$
- $\Sigma(i{:}\ E.Y.i=?a_{(r)};P)\ \#X(P) \cdot r\gamma \cdot OutsOn_E(a)\ +$
- $\Sigma(i{:}\ E.Y.i=!a_{(r)};P)\ \#X(P) \cdot r\gamma \cdot InsOn_E(a)$

$InsOn_E(a) = \Sigma(Y \in E)\ \#\{Y.i \mid E.Y.i=?a_{(r)};P\} \cdot [Y]$
$OutsOn_E(a) = \Sigma(Y \in E)\ \#\{Y.i \mid E.Y.i=!a_{(r)};P\} \cdot [Y]$

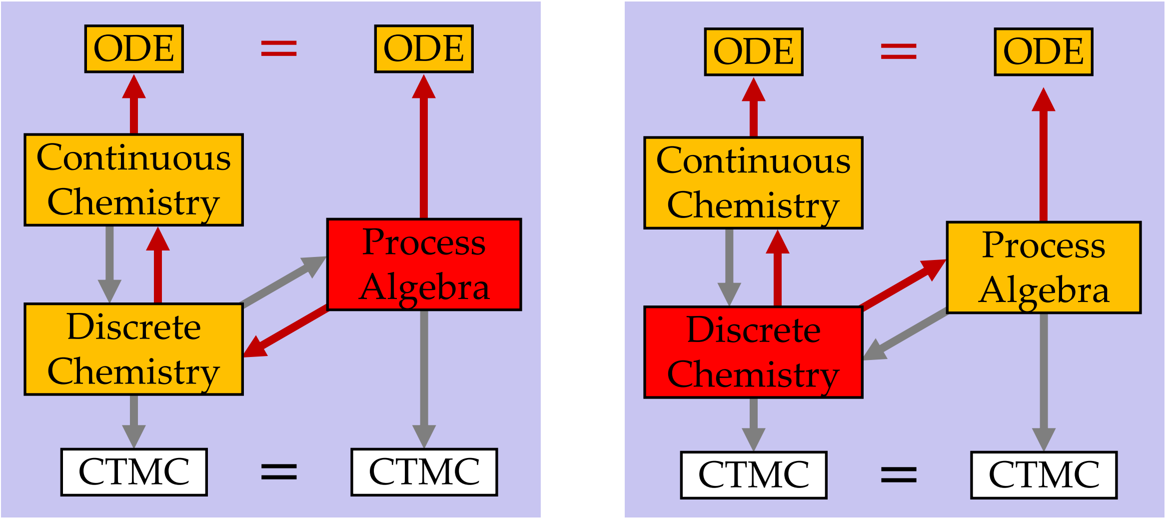
$X = \tau_{(r)};0 \quad \rightarrow \quad d[X]/dt = -r[X]$

$X = ?a_{(r)};0$, $Y = !a_{(r)};0 \quad \rightarrow \quad d[X]/dt = -r\gamma[X][Y]$, $d[Y]/dt = -r\gamma[X][Y]$

$X = ?a_{(r)};0 \oplus !a_{(r)};0 \quad \rightarrow \quad d[X]/dt = -2r\gamma[X]^2$

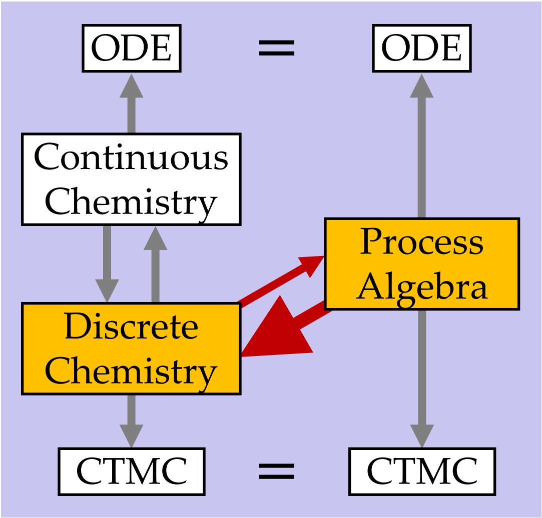
Continuous State Equivalence

- Def: ≈ is equivalence of polynomials over the field of reals.
- Thm: E ≈ Cont(Ch(E))
- Thm: Cont(C) ≈ Pi(C)

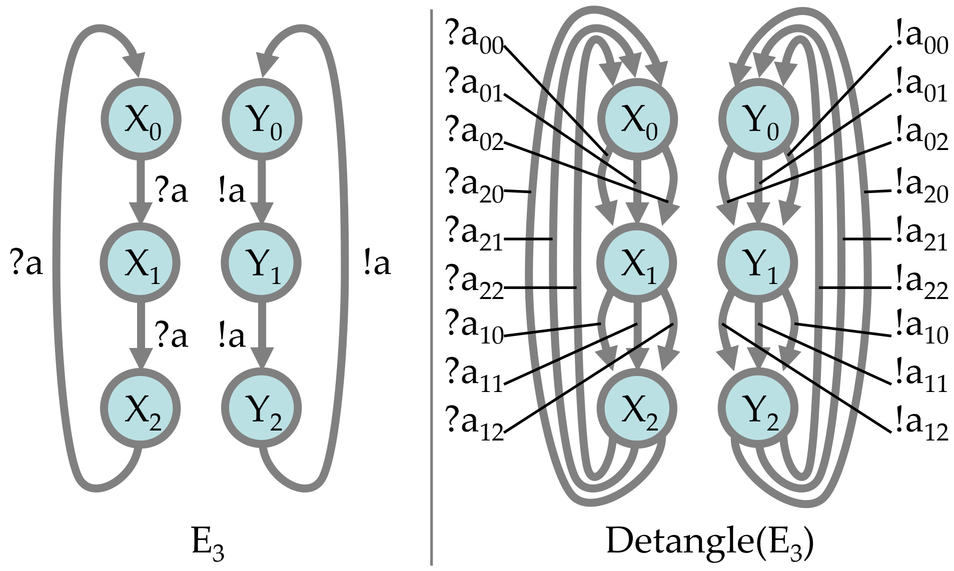


- For each E there is an E' ≈ E that is detangled (E' = Pi(Ch(E)))
- For each E in automata form there is an an E' ≈ E that is detangled and in automata form (E' = Detangle(E)).

Model Compactness



Entangled vs detangled



(closely related to Pi(Ch(E_3)))

n^2 Scaling Problems

- E_n has 2n variables (nodes) and 2n terms (arcs).
- $Ch(E_n)$ has 2n species and n^2 reactions.
- The stoichiometric matrix has size $2n \cdot n^2 = 2n^3$.
- The ODEs have 2n variables and $2n(n+n) = 4n^2$ terms
  (number of variables times number of accretions plus depletions when sums are distributed)

E_3

$X_0 = ?a_{(r)};X_1$
$X_1 = ?a_{(r)};X_2$
$X_2 = ?a_{(r)};X_0$
$Y_0 = !a_{(r)};Y_1$
$Y_1 = !a_{(r)};Y_2$
$Y_2 = !a_{(r)};Y_0$

$Ch(E_3)$

$a_{00}: X_0+Y_0 \rightarrow^r X_1+Y_1$
$a_{01}: X_0+Y_1 \rightarrow^r X_1+Y_2$
$a_{02}: X_0+Y_2 \rightarrow^r X_1+Y_0$
$a_{10}: X_1+Y_0 \rightarrow^r X_2+Y_1$
$a_{11}: X_1+Y_1 \rightarrow^r X_2+Y_2$
$a_{12}: X_1+Y_2 \rightarrow^r X_2+Y_0$
$a_{20}: X_2+Y_0 \rightarrow^r X_0+Y_1$
$a_{21}: X_2+Y_1 \rightarrow^r X_0+Y_2$
$a_{22}: X_2+Y_2 \rightarrow^r X_0+Y_0$

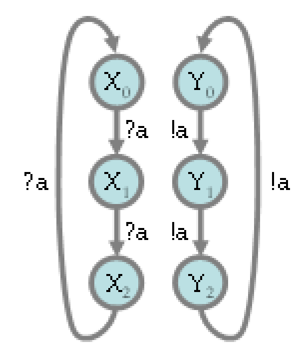
StoichiometricMatrix($Ch(E_3)$)

	a_{00}	a_{01}	a_{02}	a_{10}	a_{11}	a_{12}	a_{20}	a_{21}	a_{22}
X_0	-1	-1	-1				+1	+1	+1
X_1	+1	+1	+1	-1	-1	-1			
X_2				+1	+1	+1	-1	-1	-1
Y_0	-1		+1	-1		+1	-1		+1
Y_1	+1	-1		+1	-1		+1	-1	
Y_2		+1	-1		+1	-1		+1	-1

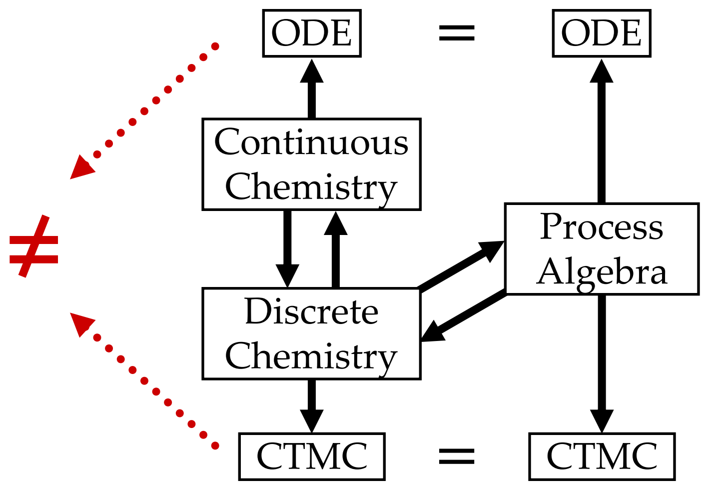
ODE(E_3)

$d[X_0]/dt = -r[X_0][Y_0] - r[X_0][Y_1] - r[X_0][Y_2] + r[X_2][Y_0] + r[X_2][Y_1] + r[X_2][Y_2]$
$d[X_1]/dt = -r[X_1][Y_0] - r[X_1][Y_1] - r[X_1][Y_2] + r[X_0][Y_0] + r[X_0][Y_1] + r[X_0][Y_2]$
$d[X_2]/dt = -r[X_2][Y_0] - r[X_2][Y_1] - r[X_2][Y_2] + r[X_1][Y_0] + r[X_1][Y_1] + r[X_1][Y_2]$
$d[Y_0]/dt = -r[X_0][Y_0] - r[X_1][Y_0] - r[X_2][Y_0] + r[X_0][Y_2] + r[X_1][Y_2] + r[X_2][Y_2]$
$d[Y_1]/dt = -r[X_0][Y_1] - r[X_1][Y_1] - r[X_2][Y_1] + r[X_0][Y_0] + r[X_1][Y_0] + r[X_2][Y_0]$
$d[Y_2]/dt = -r[X_0][Y_2] - r[X_1][Y_2] - r[X_2][Y_2] + r[X_0][Y_1] + r[X_1][Y_1] + r[X_2][Y_1]$

=

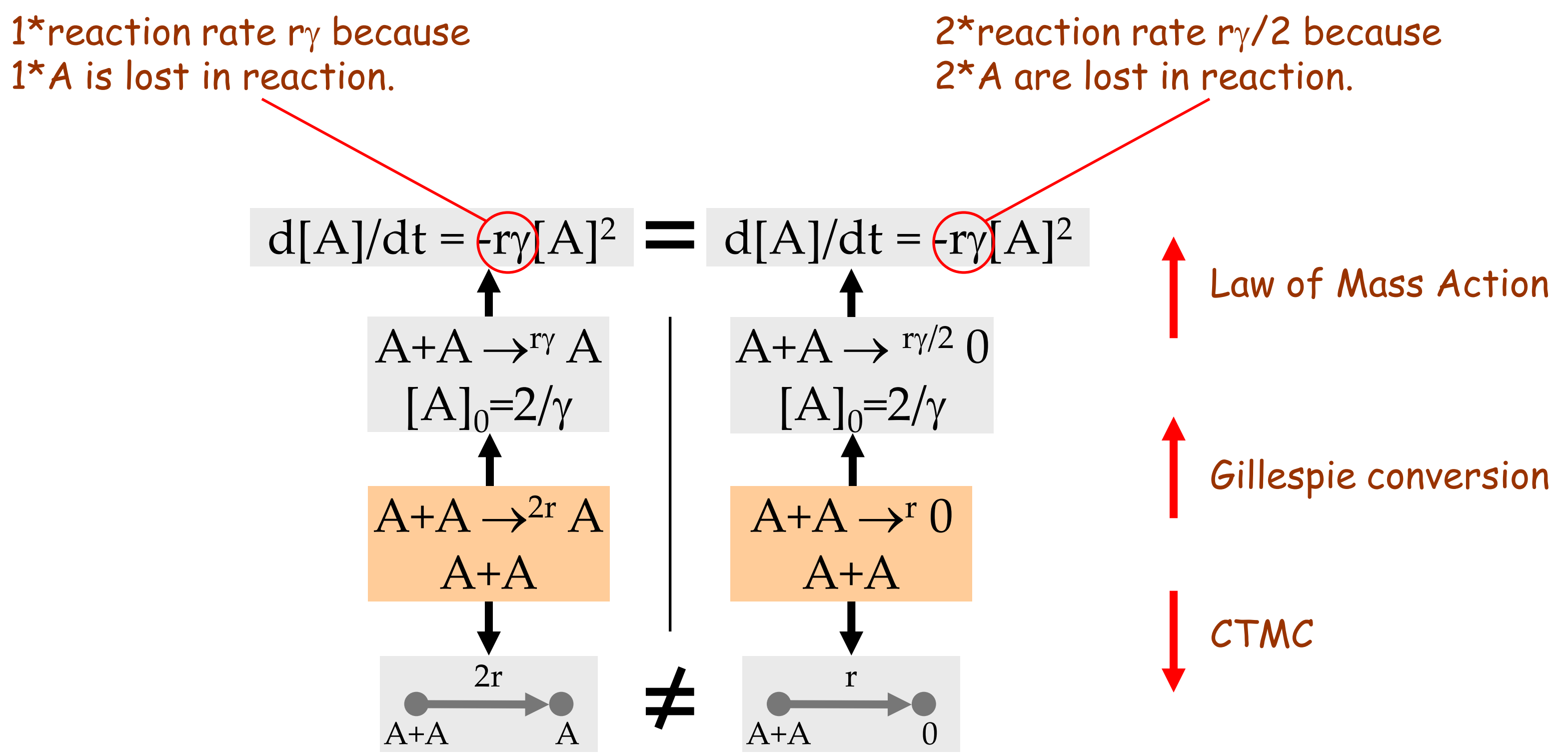


GMA ≠ CME

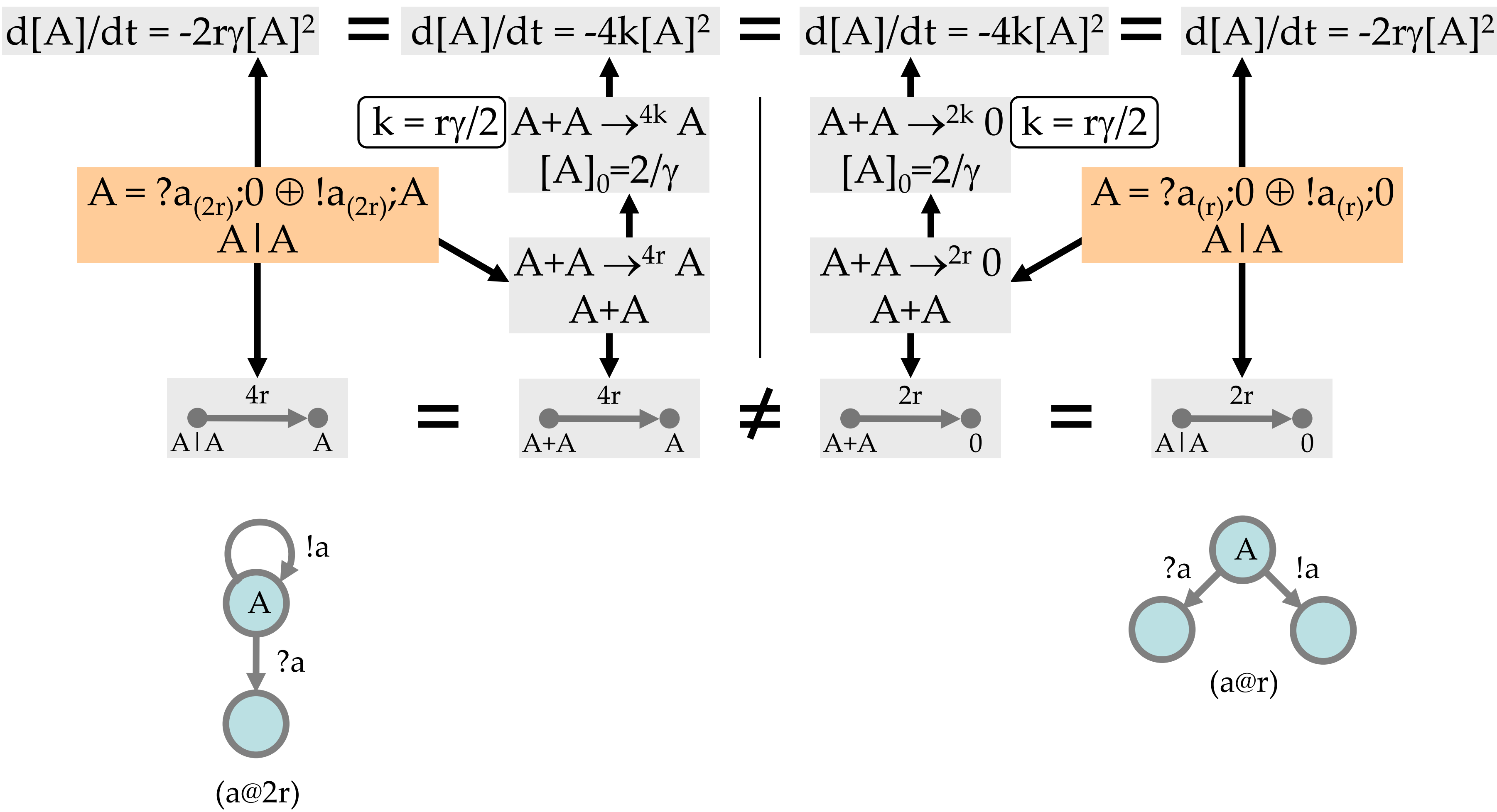


ODE
=
ODE
Continuous Chemistry
Process Algebra
Discrete Chemistry
CTMC
=
CTMC
≠

$A+A \to^{2r} A$ =? $A+A \to^{r} 0$

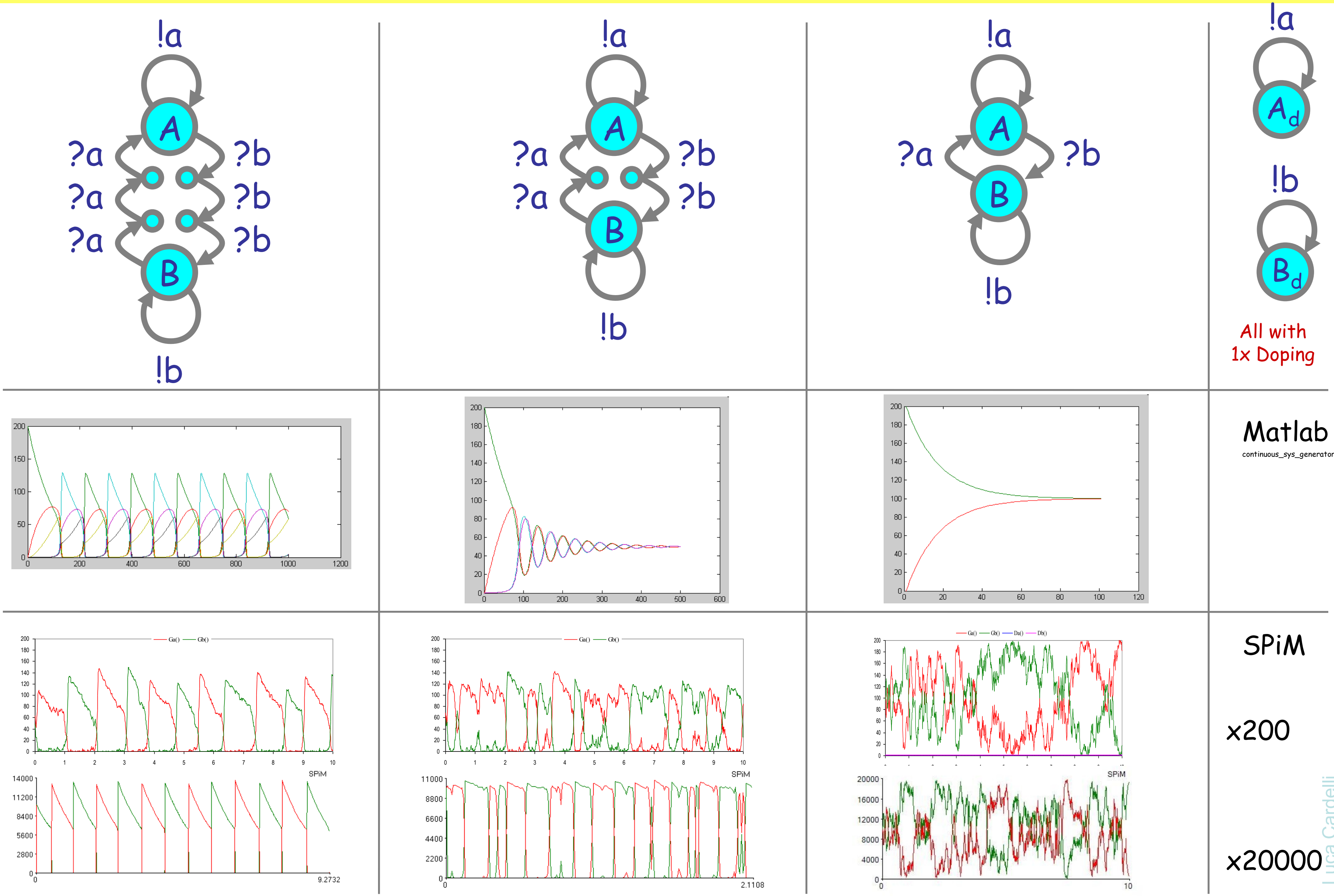


... as Automata



d[A]/dt = -2rγ[A]²
=
d[A]/dt = -4k[A]²
=
d[A]/dt = -4k[A]²
=
d[A]/dt = -2rγ[A]²
k = rγ/2
A+A →4k A
[A]0=2/γ
A+A →2k 0
[A]0=2/γ
k = rγ/2
A = ?a(2r);0 ⊕ !a(2r);A
A | A
A+A →4r A
A+A
A+A →2r 0
A+A
A = ?a(r);0 ⊕ !a(r);0
A | A
4r
A | A
A
=
4r
A+A
A
≠
2r
A+A
0
=
2r
A | A
0
!a
A
?a
(a@2r)
A
?a
!a
(a@r)

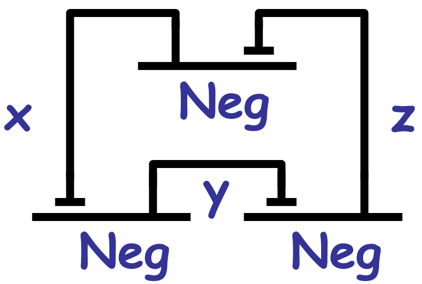
Continuous vs. Discrete Groupies

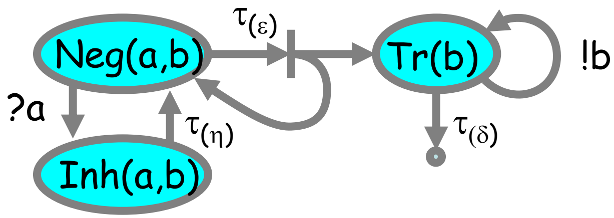


And Yet It Moves

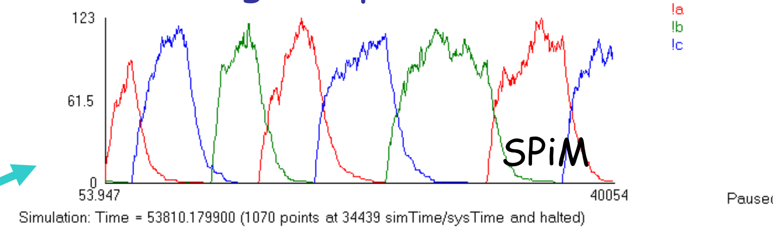
R.Blossey, L.Cardelli, A.Phillips: **Compositionality, Stochasticity and Cooperativity in Dynamic Models of Gene Regulation** (HFSP Journal)

The Repressilator





A fine stochastic oscillator over a wide range of parameters.




```
directive sample 50000.0 1000
directive plot !a; !b; !c

val dk = 0.001    (* Decay rate *)
val inh = 0.001   (* Inhibition rate *)
val cst = 0.1     (* Constitutive rate *)

let tr(p:chan()) = do !p; tr(p) or delay@dk

let neg(a:chan(), b:chan()) =
  do ?a; delay@inh; neg(a,b)
  or delay@cst; (tr(b) | neg(a,b))

val bnd = 1.0     (* Protein binding rate *)
new a@bnd:chan() new b@bnd:chan() new
c@bnd:chan()
run (neg(c,a) | neg(a,b) | neg(b,c))
```

Parametric representation

$Neg(a,b) = ?a;\ Inh(a,b) \oplus \tau_{\varepsilon};\ (Tr(b) \mid Neg(a,b))$

$Inh(a,b) = \tau_{\eta};\ Neg(a,b)$

$Tr(b) = !b;\ Tr(b) \oplus \tau_{\gamma};\ 0$

$Neg(x_{(r)},y_{(r)}) \mid Neg(y_{(r)},z_{(r)}) \mid Neg(z_{(r)},x_{(r)})$

$Neg/x,y \to^{\varepsilon} Tr/y + Neg/x,y$

$Neg/y,z \to^{\varepsilon} Tr/z + Neg/y,z$

$Neg/z,x \to^{\varepsilon} Tr/x + Neg/z,x$

$Tr/x + Neg/x,y \to^{r} Tr/x + Inh/x,y$

$Tr/y + Neg/y,z \to^{r} Tr/y + Inh/y,z$

$Tr/z + Neg/z,x \to^{r} Tr/z + Inh/z,x$

$Inh/x,y \to^{\eta} Neg/x,y$

$Inh/y,z \to^{\eta} Neg/y,z$

$Inh/z,x \to^{\eta} Neg/z,x$

$Tr/x \to^{\gamma} 0$

$Tr/y \to^{\gamma} 0$

$Tr/z \to^{\gamma} 0$

$Neg/x,y + Neg/y,z + Neg/z,x$

$d[Neg/x,y]/dt = -r[Tr/x][Neg/x,y] + \eta[Inh/x,y]$

$d[Neg/y,z]/dt = -r[Tr/y][Neg/y,z] + \eta[Inh/y,z]$

$d[Neg/z,x]/dt = -r[Tr/z][Neg/z,x] + \eta[Inh/z,x]$

$d[Inh/x,y]/dt = r[Tr/x][Neg/x,y] - \eta[Inh/x,y]$

$d[Inh/y,z]/dt = r[Tr/y][Neg/y,z] - \eta[Inh/y,z]$

$d[Inh/z,x]/dt = r[Tr/z][Neg/z,x] - \eta[Inh/z,x]$

$d[Tr/x]/dt = \varepsilon[Neg/z,x] - \gamma[Tr/x]$

$d[Tr/y]/dt = \varepsilon[Neg/x,y] - \gamma[Tr/y]$

$d[Tr/z]/dt = \varepsilon[Neg/y,z] - \gamma[Tr/z]$

simplifying (N is the quantity of each of the 3 gates)

$d[Neg/x,y]/dt = \eta N - (\eta + r[Tr/x])[Neg/x,y]$

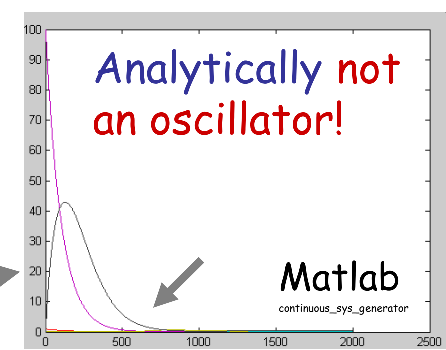
$d[Neg/y,z]/dt = \eta N - (\eta + r[Tr/y])[Neg/y,z]$

$d[Neg/z,x]/dt = \eta N - (\eta + r[Tr/z])[Neg/z,x]$

$d[Tr/x]/dt = \varepsilon[Neg/z,x] - \gamma[Tr/x]$

$d[Tr/y]/dt = \varepsilon[Neg/x,y] - \gamma[Tr/y]$

$d[Tr/z]/dt = \varepsilon[Neg/y,z] - \gamma[Tr/z]$



interval/step [0:10:20000] $N=1,\ r=1.0,\ \varepsilon=0.1,\ \eta=0.001,\ \gamma=0.001$

(Neg/x,y)	dx1/dt = 0.001 - (0.001 + x4)*x1	1.0
(Neg/x,y)	dx2/dt = 0.001 - (0.001 + x5)*x2	1.0
(Neg/x,y)	dx3/dt = 0.001 - (0.001 + x6)*x3	1.0
(Tr/x)	dx4/dt = 0.1*x3 - 0.001*x4	100.0
(Tr/y)	dx5/dt = 0.1*x1 - 0.001*x5	0
(Tr/z)	dx6/dt = 0.1*x2 - 0.001*x6	0

Conclusions

Conclusions

- Compositional models
  - Accurate (at the "appropriate" abstraction level).
  - Manageable (so we can scale them up by composition).
  - Executable (stochastic simulation).
- Analysis techniques
  - Mathematical techniques: Markov theory, Chemical Master Equation, and Rate Equation
  - Computing techniques: Abstraction and Refinement, Model Checking, Causality Analysis.
- Many lines of extensions
  - Parametric processes for model factorization
  - *Poly*automata for ***Bio***-Chemistry: complexation and polymerization
  - Ultimately, rich process-algebra based modeling languages.
- Quantitative techniques
  - Important in the "real sciences".