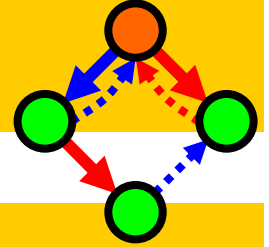


Our understanding of life will derive in large measure from *how* we understand chemistry.
Walter Fontana and Leo W. Buss, 1996

Artificial
Biochemistry



Chemistry and Processes

Luca Cardelli

Microsoft Research

The Microsoft Research - University of Trento
Centre for Computational and Systems Biology

Trento, 2006-05-22..26

www.luca.demon.co.uk/ArtificialBiochemistry.htm

From Chemistry to ODEs

Chemical Reactions

$A \rightarrow^r B_1 + \dots + B_n$	Degradation	$[A]^{\bullet} = -r[A]$	Exponential Decay
$A_1 + A_2 \rightarrow^r B_1 + \dots + B_n$	Asymmetric Collision	$[A_i]^{\bullet} = -r[A_1][A_2]$	Mass Action Law
$A + A \rightarrow^r B_1 + \dots + B_n$	Symmetric Collision	$[A]^{\bullet} = -r[A]([A]-1)$	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

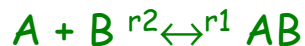
www.chemguide.co.uk

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!

Trimolecular reactions:



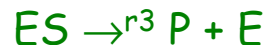
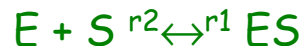
the measured "r" is an (imperfect) aggregate of e.g.:



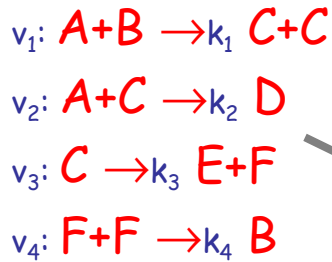
Enzymatic reactions:



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



From Reactions to ODE's

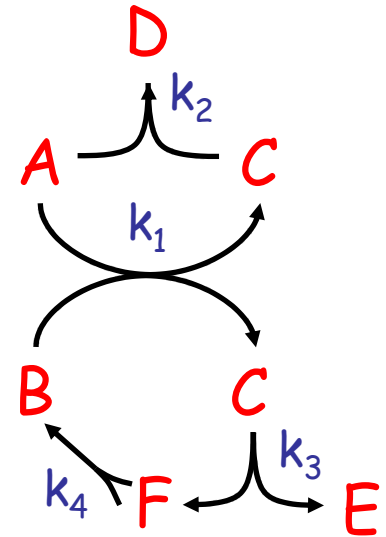


Write the coefficients by columns

		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

Stoichiometric Matrix



Quantity changes

Stoichiometric matrix

Rate laws

$$[X]^\bullet = N \cdot I$$

Read the concentration changes from the rows

Set a rate law for each reaction (Degradation/Asymmetric/Symmetric)

$$\begin{aligned}
 [A]^\bullet &= -I_1 - I_2 \\
 [B]^\bullet &= -I_1 + I_4 \\
 [C]^\bullet &= 2I_1 - I_2 - I_3 \\
 [D]^\bullet &= I_2 \\
 [E]^\bullet &= I_3 \\
 [F]^\bullet &= I_3 - 2I_4
 \end{aligned}$$

E.g. $[A]^\bullet = -k_1[A][B] - k_2[A][C]$

I	
I ₁	$k_1[A][B]$
I ₂	$k_2[A][C]$
I ₃	$k_3[C]$
I ₄	$k_4[F]([F]-1)/2$

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

Rate Law of Symmetric Reactions

The proper rate law I_4 of the symmetric reaction is $k_4[F]([F]-1)/2$, because that is the number of possible collisions between $[F]$ particles.

But the v_4 contribution to $[F]^\bullet$ is $-k_4[F]([F]-1)$ because 2 F are consumed in that reaction!

Compare with the contribution of v_4 to $[B]^\bullet$, which is $k_4[F]([F]-1)/2$.

That's why we said earlier that



Instead, e.g.:



$$[A]^\bullet = -I_1 - I_2$$

$$[B]^\bullet = -I_1 + I_4$$

$$[C]^\bullet = 2I_1 - I_2 - I_3$$

$$[D]^\bullet = I_2$$

$$[E]^\bullet = I_3$$

$$[F]^\bullet = I_3 - 2I_4$$

I	
I_1	$k_1[A][B]$
I_2	$k_2[A][C]$
I_3	$k_3[C]$
I_4	$k_4[F]([F]-1)/2$

$$[A]^\bullet = -k_1[A][B] - k_2[A][C]$$

$$[B]^\bullet = -k_1[A][B] + k_4[F]([F]-1)/2$$

$$[C]^\bullet = 2k_1[A][B] - k_2[A][C] - k_3[C]$$

$$[D]^\bullet = k_2[A][C]$$

$$[E]^\bullet = k_3[C]$$

$$[F]^\bullet = k_3[C] - k_4[F]([F]-1)$$

From Chemistry to Processes

Chemical Ground Form (CGF)

$E ::= X_1=M_1, \dots, X_n=M_n$

Definitions ($n \geq 0$)

$M ::= \pi_1;P_1 \oplus \dots \oplus \pi_n;P_n$

Molecules ($n \geq 0$)

$P ::= X_1 \mid \dots \mid X_n$

Solutions ($n \geq 0$)

$\pi ::= \tau_r \ ?n_{(r)} \ !n_{(r)}$

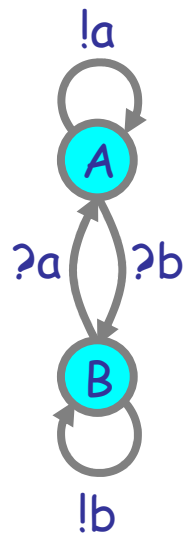
Interactions (delay, input, output)

CGF ::= E,P

Definitions with Initial Conditions

(To translate chemistry back to processes we need a bit more than simple 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 ($P \mid O = O \mid P = P$)
 and null molecule ($M \oplus O = O \oplus M = M$) ($\tau_0;P = O$)
 X_i are distinct in E
 Each name n is assigned a fixed rate r : $n_{(r)}$



Ex: interacting automata
 (which are CGFs using "|" only in initial conditions):

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

Automaton in state A

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

Automaton in state B

$A \mid A \mid B \mid B$

Initial conditions:
 2A and 2B

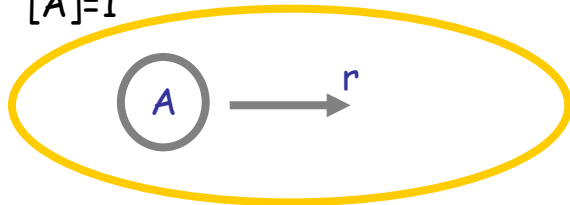
Coding of Degradation Reactions ($A \xrightarrow{r} A'$)

Particles

Rate Law

Processes

$[A]=1$



1 degradation at rate r

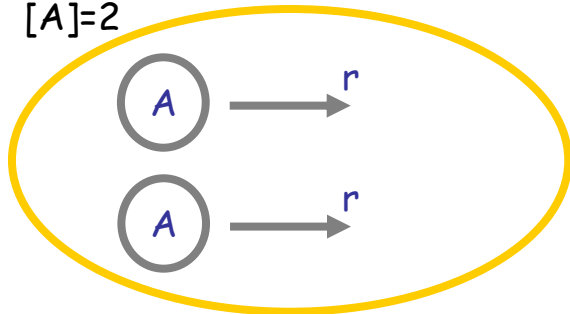
$$[A]^{\bullet} = -r[A]$$

$$= -r$$

$$A = \tau_r : A'$$

1 delay at rate r
(sampled from exponential distribution)

$[A]=2$



2 degradations at rate r

$$[A]^{\bullet} = -r[A]$$

$$= -2r$$

(constant half-life!)

$$A = \tau_r : A'$$

$$A = \tau_r : A'$$

2 delays at rate r
(sampled from exponential distribution)

(or rather, the CGF:

$$A = \tau_r : A'$$

$$A | A$$

)

Coding of Asymmetric Reactions ($A+B \xrightarrow{r} A'+B'$)

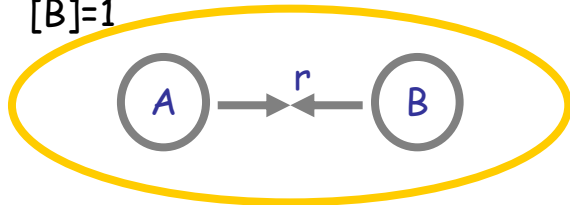
Particles

Rate Law

Processes

$$[A]=1$$

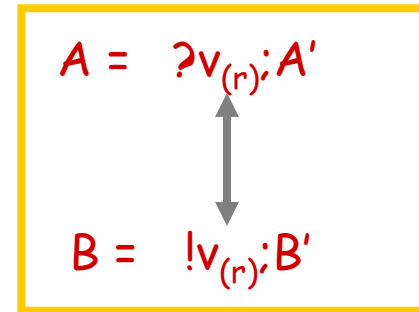
$$[B]=1$$



1 (productive) collision
between 1+1 particles, at rate r

$$[A]^\bullet = -r[A][B]$$

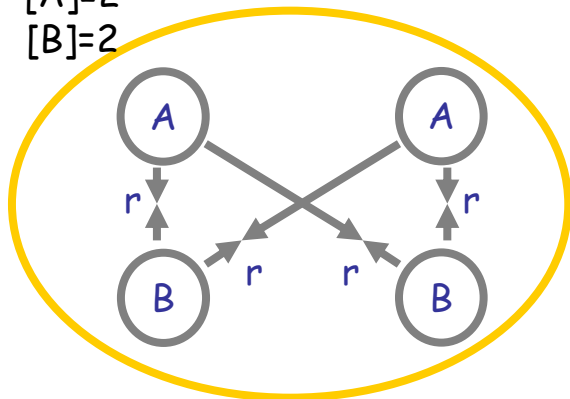
$$= -r$$



1 interactions between 1+1 processes
at rate r

$$[A]=2$$

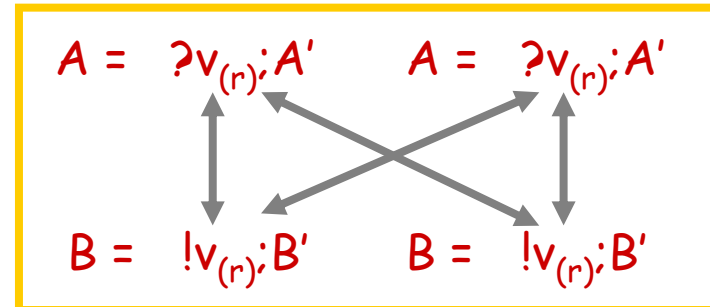
$$[B]=2$$



4 (productive) collisions
between 2+2 particles, at rate r

$$[A]^\bullet = -r[A][B]$$

$$= -4r$$



4 interactions between 2+2 processes
at rate r

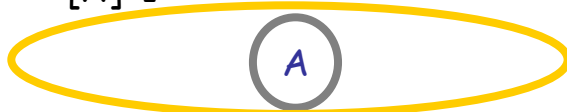
Coding of Symmetric Reactions ($A+A \rightarrow^r A'+A''$)

Particles

Rate Law

Processes

$[A]=1$



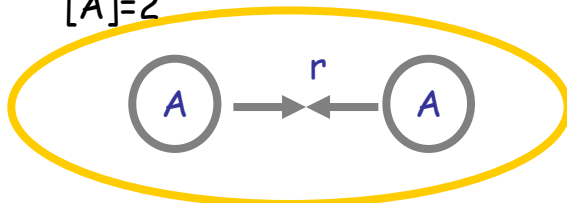
0 collisions between 1 particle

$$[A]^* = -r[A]([A]-1) = 0$$

$$A = ?v_{(r/2);A'} \oplus !v_{(r/2);A''}$$

0 interactions between 1 process

$[A]=2$



1 collision between 2 particles at rate r

$$[A]^* = -r[A]([A]-1) = -2r$$

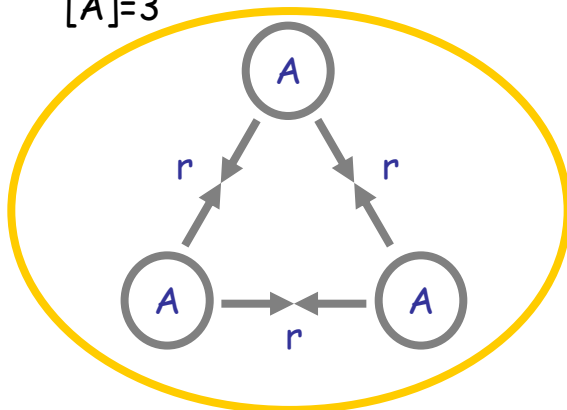
(remember: 2A are removed per collision assuming $A', A'' \neq A$)

$$A = ?v_{(r/2);A'} \oplus !v_{(r/2);A''}$$

$$A = ?v_{(r/2);A'} \oplus !v_{(r/2);A''}$$

2 interactions between 2 processes at rate r/2

$[A]=3$



3 collisions between 3 particles at rate r

$$[A]^* = -r[A]([A]-1) = -6r$$

3 collisions at rate r

$$A = ?v_{(r/2);A'} \oplus !v_{(r/2);A''}$$

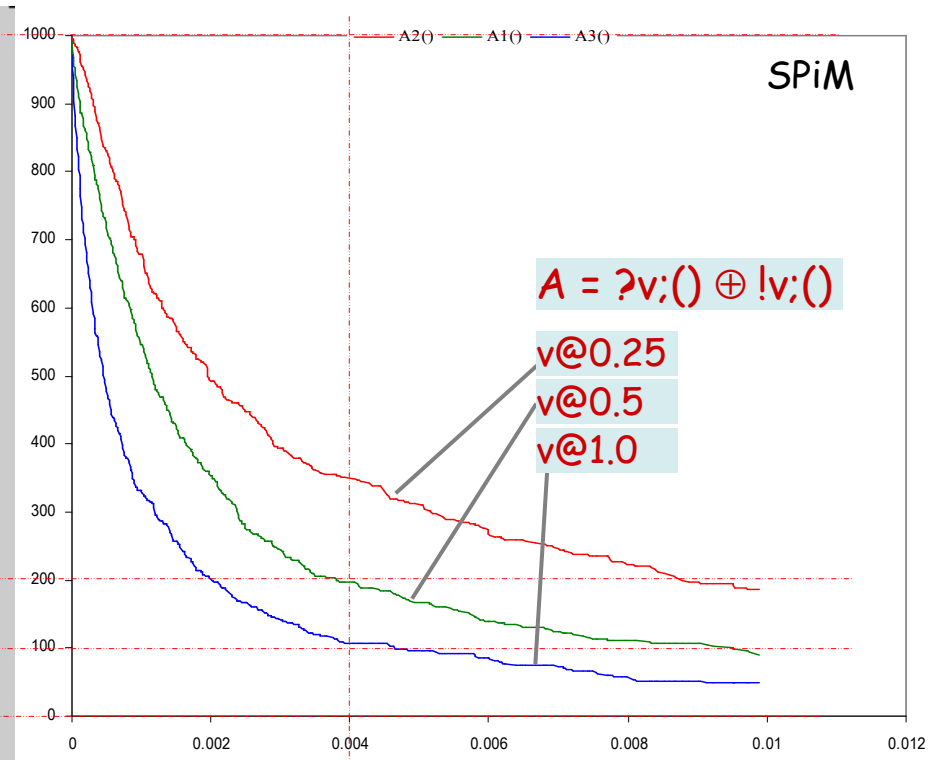
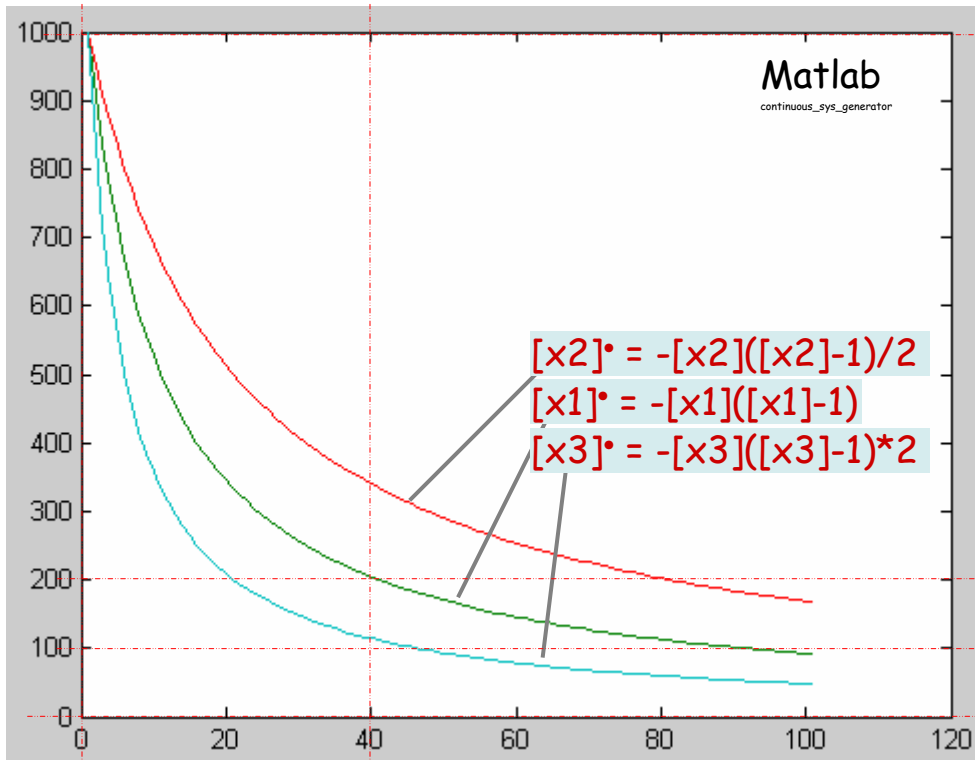
$$A = ?v_{(r/2);A'} \oplus !v_{(r/2);A''}$$

$$A = ?v_{(r/2);A'} \oplus !v_{(r/2);A''}$$

6 interactions between 3 processes at rate r/2

Symmetric Reactions: SPiM vs Matlab

$$[x1]^* = -r[x1]([x1]-1) \iff v@r/2$$



Continuous system generator

System info: Symmetric Reactions

IGNV*: internal gaussian noise variance
 IWNV*: internal white noise variance
 AGNV*: additive gaussian noise variance
 AWNV*: additive white noise variance

Solver: ode45 Additional Parameters:

Time: [0:0.0001:0.01] Transient time:

dX/dt	F(X,t)	IGNV*	IWNV*	AGNV*	AWNV*	Initial conditions
dx1/dt	-1.0*x1*(x1-1)	0	0	0	0	1000.0
dx2/dt	-1.0*x2*(x2-1)/2	0	0	0	0	1000.0
dx3/dt	-1.0*x3*(x3-1)*2	0	0	0	0	1000.0

Note: if IGNV or IWNV > 0 you should possibly use stochastic integrator instead this

Buttons: Add equation, Modify equation, Delete equation, Cancel, Ok

```

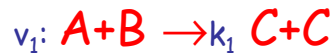
directive sample 0.01 1000
directive plot A2(); A1(); A3()

new v1@0.5:chan
new v2@0.25:chan
new v3@1.0:chan

let A1() = do ?v1;() or !v1;()
let A2() = do ?v2;() or !v2;()
let A3() = do ?v3;() or !v3;()

run 1000 of (A1() | A2() | A3())
    
```

Chemical Reactions to CGF



Interaction Matrix

channels and rates
(1 per reaction)

Half-rate for
symmetric
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		$!;0$	$\tau;(E F)$	
D				
E				
F				$?;B$ $!;0$

Fill the matrix by columns:

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

Asymmetric reaction $v_i: X+Y \rightarrow_k P_i$
add $?;P_i$ to $\langle X, v_i \rangle$ and $!;0$ to $\langle Y, v_i \rangle$

Symmetric reaction $v_i: X+X \rightarrow_k P_i$
add $?;P_i$ and $!;0$ to $\langle X, v_i \rangle$

Read out the resulting CGF by rows:

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

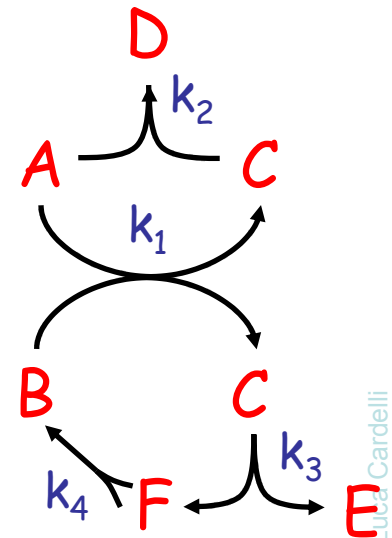
$$B = !v_{1(k1)};0$$

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

$$D = 0$$

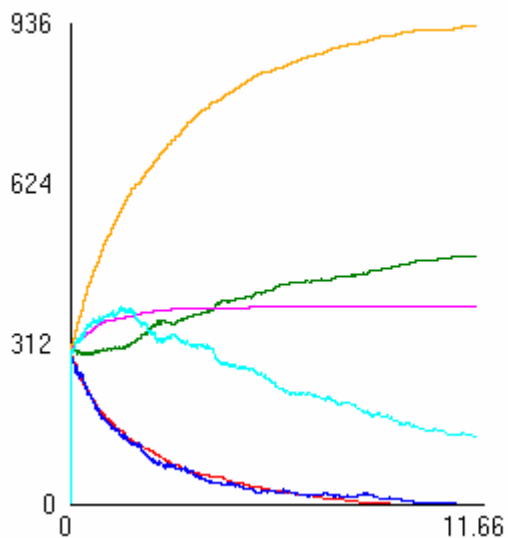
$$E = 0$$

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

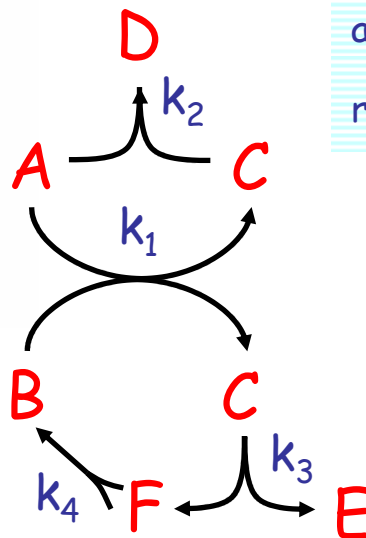


That CGF in SPiM

$A = ?v_{1(k_1)}:(C|C) \oplus ?v_{2(k_2)}:D$
 $B = !v_{1(k_1)}:O$
 $C = !v_{2(k_2)}:O \oplus \tau_{k_3}:(E|F)$
 $D = O$
 $E = O$
 $F = ?v_{4(k_4/2)}:B \oplus !v_{4(k_4/2)}:O$



A0
 B0
 C0
 D0
 E0
 F0



```

directive sample 10.0
directive plot A(); B(); C(); D(); E(); F()
  
```

```

val k1 = 0.001 new v1@k1:chan
val k2 = 0.001 new v2@k1:chan
val k3 = 1.0
val k4 = 0.001 new v4@k4/2.0:chan
  
```

```

let A() = do ?v1:(C())|C() or ?v2:D()
and B() = !v1
and C() = do !v2 or delay@k3:(E())|F()
and D() = ()
and E() = ()
and F() = do ?v4;B() or !v4

run 300 of (A())|B())|C())|D())|E())|F())
  
```

... in other words

From chemical reactions \mathcal{C} to a CGF $\text{Pi}(\mathcal{C})$:

$$\begin{aligned} \text{Pi}(\mathcal{C}) = \{ & (X = \oplus((v_i: X \xrightarrow{k} P) \in \mathcal{C}) \text{ of } (\tau_k; P) & \oplus \\ & \oplus((v_i: X+Y \xrightarrow{k} P) \in \mathcal{C} \text{ and } Y \neq X) \text{ of } (?v_{i(ki)}; P) & \oplus \\ & \oplus((v_i: Y+X \xrightarrow{k} P) \in \mathcal{C} \text{ and } Y \neq X) \text{ of } (!v_{i(ki)}; 0) & \oplus \\ & \oplus((v_i: X+X \xrightarrow{k} P) \in \mathcal{C}) \text{ of } (?v_{i(ki/2)}; P \oplus !v_{i(ki/2)}; 0) &) \\ & \text{s.t. } X \text{ is a species in } \mathcal{C} \end{aligned}$$

From Processes Directly to ODEs (hard)

"Micromodels": Continuous Time Markov Chains

- The underlying semantics of stochastic π -calculus (and stochastic interacting automata). Well established in many ways.
 - Automata with rates on transitions.
- "The" correct semantics for chemistry, executable.
 - Gillespie stochastic simulation algorithm
- Good description of "individual" behavior
- But does not give a good sense of "collective" properties.
 - Yes one can do simulation.
 - Yes one can do program analysis.
 - Yes one can do modelchecking.
 - But somewhat lacking in "predictive power".

"Macromodels": Ordinary Differential Equations

- Micromodels have lots of advantages
 - Compositional, compact, mechanistic, etc.
- But they always ask:
 - "Yes, but how does your automata model relate to the 75 ODE models in the literature?"
- From processes/automata to ODEs directly:
 - *In principle*: just write down the **Rate Equation**:
 - Determine the set of all possible *states* S of each process.
 - Determine the rates of the transitions between such states.
 - Let $[S]$ be the "number of processes in state S " as a function of time.
 - Define for each state S :
 $[S]^*$ = (rate of change of the number of processes in state S)
Cumulative rate of transitions from any state S' to state S , times $[S']$,
minus cumulative rate of transitions from S to any state S'' , times $[S]$.
 - Intuitive (rate = inflow minus outflow), but often clumsy to write down precisely.
- But why go directly from processes to ODE?
 - If we first convert processes to chemical reactions, then we can convert to ODEs by standard means!

The speed of interaction is proportional to the number of possible interactions.

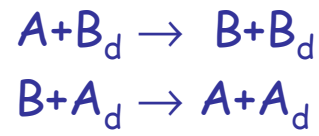
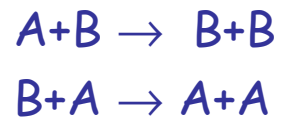
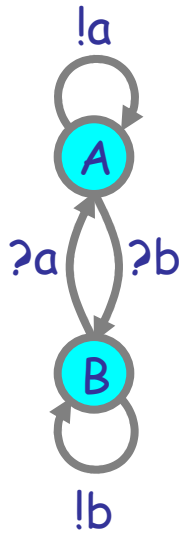
The rate of change of a quantity is the cumulative result of the (always changing) speed of the interactions that produce it and consume it. It is a **global** and **dynamic** property.



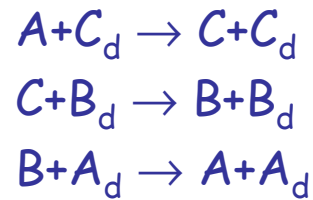
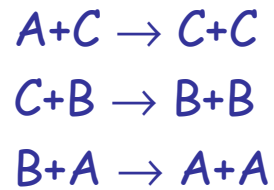
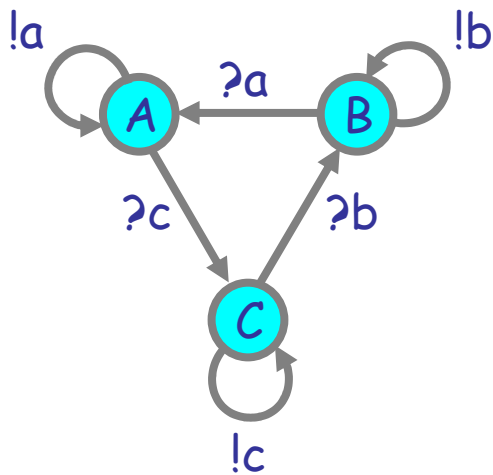
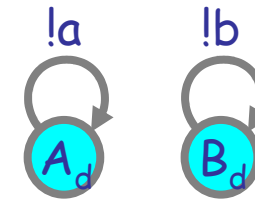
From Processes to Chemistry

Automata to Chemistry (by hand)

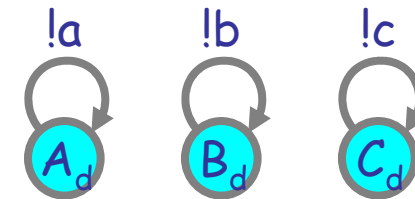
Automata to Chemistry



Doping

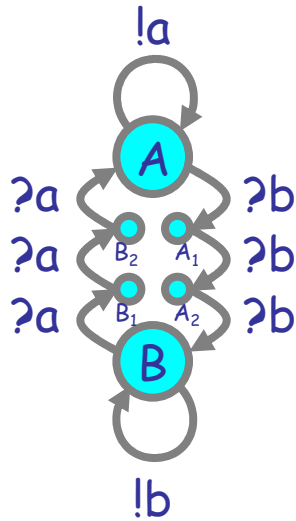


Doping

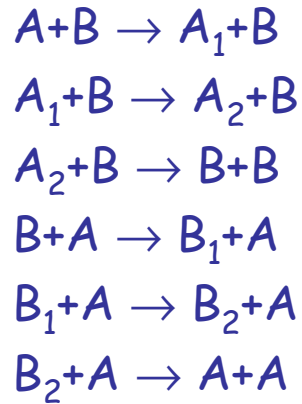


(Lack of) Compositionality

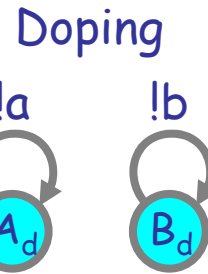
6 states



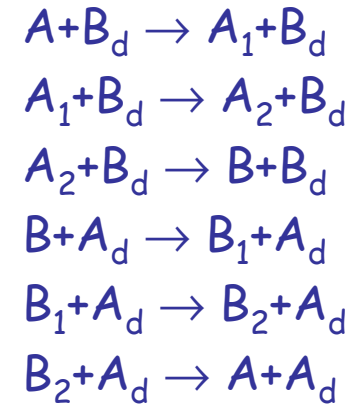
6 reactions



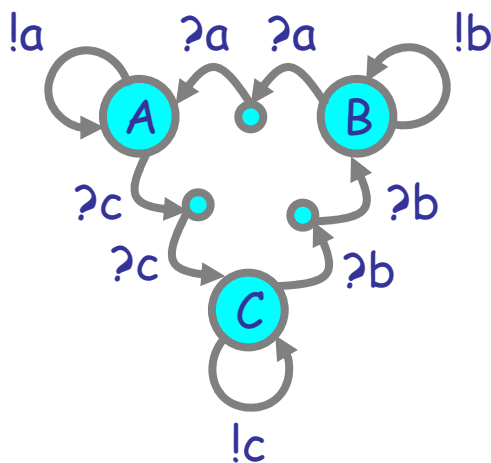
+2 states



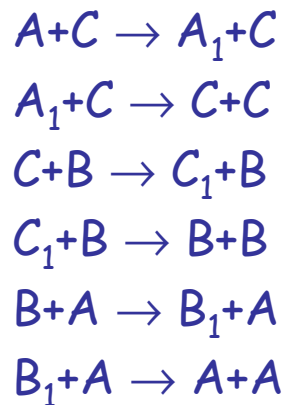
+6 reactions ??



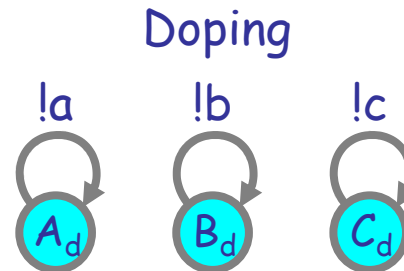
6 states



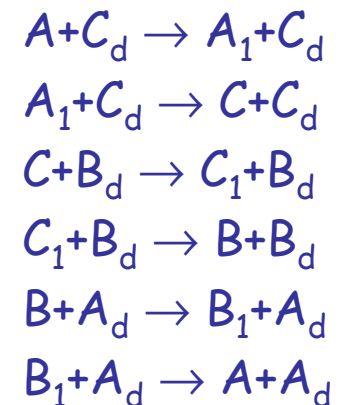
6 reactions



+3 states



+6 reactions ??

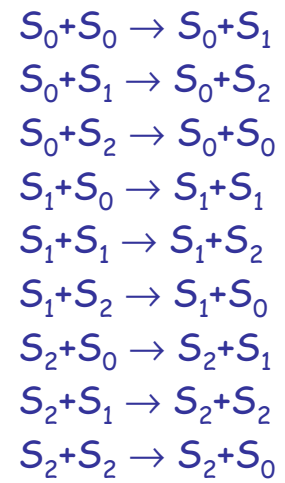
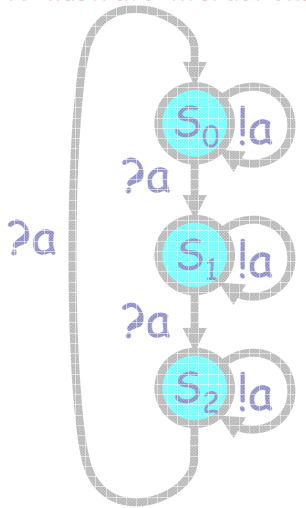
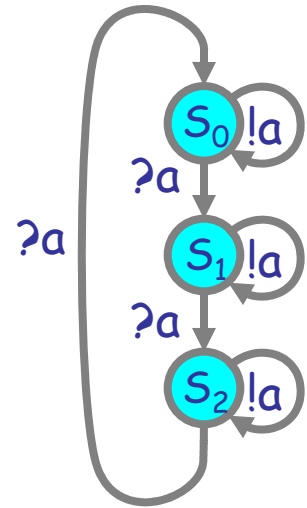


Single Automata Models are n^2 More Compact

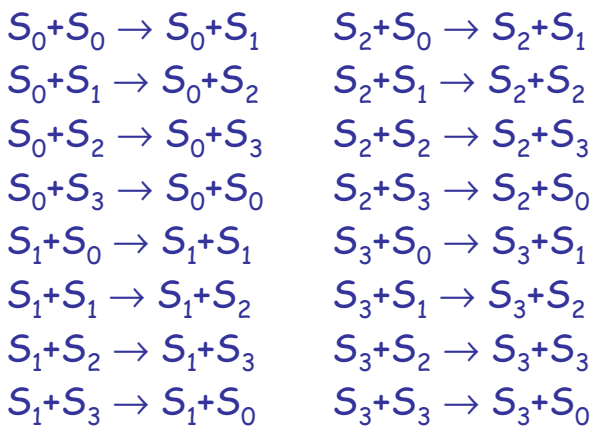
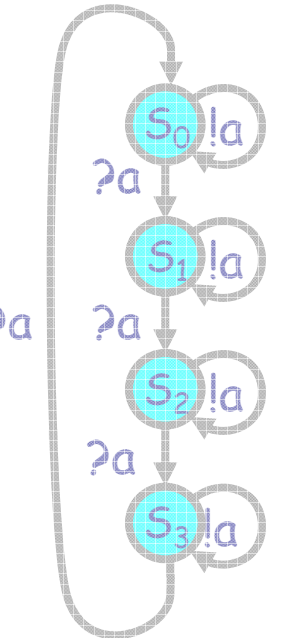
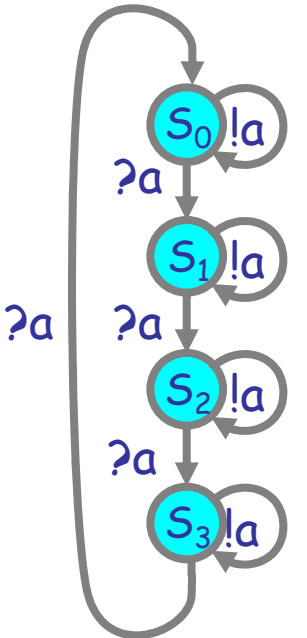
Automaton

A copy of the automaton to illustrate interactions

Chemistry



3 states
 (2*3 transitions)
 =
 3² reactions



4 states
 (2*4 transitions)
 =
 4² reactions

Process Normal Forms to Chemistry

Chemical Parametric Form (CPF)

$E ::= X_1(\mathbf{p}_1)=M_1, \dots, X_n(\mathbf{p}_n)=M_n$

$M ::= \pi_1:P_1 \oplus \dots \oplus \pi_n:P_n$

$P ::= X_1(\mathbf{p}_1) \mid \dots \mid X_n(\mathbf{p}_n)$

$\pi ::= \tau_r \ ?n(\mathbf{p}) \ !n(\mathbf{p})$

$CPF ::= E, P$

Definitions $(n \geq 0)$

Molecules $(n \geq 0)$

Solutions $(n \geq 0)$

Interactions

with initial conditions

CPF is NF without name generation.

CPF is CGF plus parameters.

\oplus is stochastic choice (vs. + for chemical reactions)

0 is the null solution ($P \mid 0 = 0 \mid P = P$)

and null molecule ($M \oplus 0 = 0 \oplus M = M$) ($\tau_0; P = 0$)

X_i are distinct in E , \mathbf{p} are vectors of names

\mathbf{p} are vectors of distinct names when in **binding position**

Each free name n in E is assigned a fixed rate r :

written either $n_{(r)}$, or $\rho_{CPF}(n)=r$.

Random example

$P(\mathbf{a}, \mathbf{b}) = ?a(\mathbf{c}); (P(\mathbf{c}, \mathbf{b}) \mid Q(\mathbf{c}))$

$Q(\mathbf{a}) = !a(\mathbf{a}); 0$

$P(x_{(r)}, y_{(s)}) \mid Q(x_{(r)})$

Repressilator

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

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

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

$Neg(\mathbf{x}, \mathbf{y}) \mid Neg(\mathbf{y}, \mathbf{z}) \mid Neg(\mathbf{z}, \mathbf{x})$

CPF to Chemistry: the CGF Case

CGF = The CPF subset with no communication, no parametric definitions.

$E ::= X_1=M_1, \dots, X_n=M_n$	Definitions ($n \geq 0$)
$M ::= \pi_1;P_1 \oplus \dots \oplus \pi_n;P_n$	Molecules ($n \geq 0$)
$P ::= X_1 \mid \dots \mid X_n$	Solutions ($n \geq 0$)
$\pi ::= \tau_r \ ?n_{(r)} \ !n_{(r)}$	Interactions (delay, input, output)
CGF ::= E,P	Definitions with Initial Conditions

Each X in E is seen as a separate *species*.

Chemical reactions for E : (N.B.: $\{\dots\}^m$ is a multiset, and P is P with all the \mid changed to $+$)

$$\begin{aligned} \text{Ch}_G(E) := & \{(X \xrightarrow{r} P) \text{ s.t. } (X \equiv \tau_r;P \oplus \dots) \in E\}^m \\ & \cup^m \{(X + Y \xrightarrow{r} P + Q) \text{ s.t. } X \neq Y, \langle (X \equiv ?n_{(r)};P \oplus \dots), (Y \equiv !n_{(r)};Q \oplus \dots) \rangle \in E^2\}^m \\ & \cup^m \{(X + X \xrightarrow{2r} P + Q) \text{ s.t. } (X \equiv ?n_{(r)};P \oplus \dots \equiv !n_{(r)};Q \oplus \dots) \in E\}^m \end{aligned}$$

Initial conditions for P :

$$\text{Ch}_G(P) := P$$

(Note on computing the multisets)

A multiset $M \in \text{Multiset}(S)$, where S is a set with equality, is a total function $S \rightarrow \text{Nat}$, which may also be written as a finite enumeration with repetitions: $\{\dots\}^m$.

Multiset binary union is the function $\cup^m(M, M') = \lambda s. M(s) + M'(s)$.

Multiset big union over a finite index set S is written: $\cup^m_{i \in S} \text{ of } M_i$.

- The shorthand: $\{(X \rightarrow^r P) \text{ s.t. } (X \equiv \tau_r; P \oplus \dots) \in E\}^m$

means: "for each $(X = \pi_1; P_1 \oplus \dots \oplus \pi_n; P_n) \in E$ and for each i such that $\pi_i = \tau_r$, return a copy of $(X \rightarrow^r P_i)$ ".

i.e., it is defined as the following finite union of singleton multisets:

$$\cup^m \{(X = \pi_1; P_1 \oplus \dots \oplus \pi_n; P_n) \in E\} \text{ of } (\cup^m \{i \text{ s.t. } \pi_i = \tau_r\} \text{ of } \{(X \rightarrow^r P_i)\}^m)$$

- The shorthand: $\{(X + Y \rightarrow^r P + Q) \text{ s.t. } X \neq Y, \langle (X \equiv ?n_{(r)}; P \oplus \dots), (Y \equiv !n_{(r)}; Q \oplus \dots) \rangle \in E^2\}^m$

means: "for each ordered pair $(X = \pi_1; P_1 \oplus \dots \oplus \pi_n; P_n), (Y = \sigma_1; Q_1 \oplus \dots \oplus \sigma_m; Q_m)$ from E^2 with $X \neq Y$, for each $n_{(r)}$, for each i such that $\pi_i = ?n_{(r)}$, and for each j such that $\sigma_j = !n_{(r)}$, return a copy of the reaction $(X + Y \rightarrow^r P_i + Q_j)$ ".

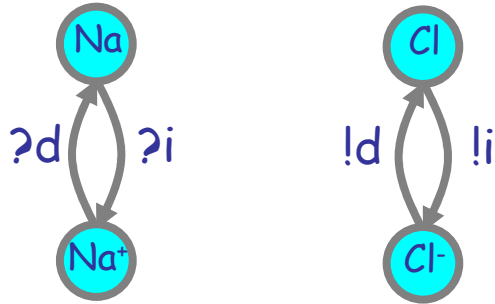
i.e., it is defined as the following finite union of singleton multisets:

$$\cup^m \langle (X = \pi_1; P_1 \oplus \dots \oplus \pi_n; P_n), (Y = \rho_1; Q_1 \oplus \dots \oplus \rho_m; Q_m) \rangle \in E^2 \text{ with } X \neq Y \\ \text{of } (\cup^m \{ \langle i, j \rangle \text{ s.t. } \pi_i = ?n_{(r)}, \rho_j = !n_{(r)} \} \text{ of } \{(X + Y \rightarrow^r P_i + Q_j)\}^m)$$

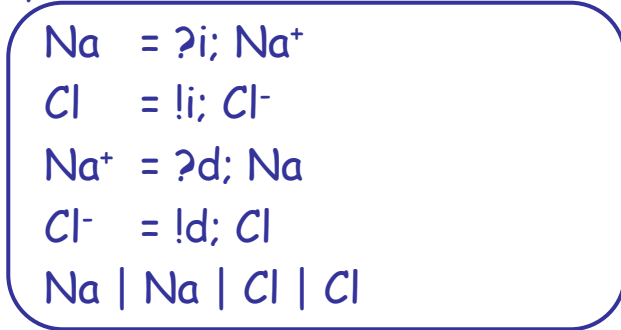
- The shorthand: $\{(X + X \rightarrow^{2r} P + Q) \text{ s.t. } (X \equiv ?n_{(r)}; P \oplus \dots \equiv !n_{(r)}; Q \oplus \dots) \in E\}^m$

means: "for each $(X = \pi_1; P_1 \oplus \dots \oplus \pi_n; P_n)$ in E , for each $n_{(r)}$, for each i such that $\pi_i = ?n_{(r)}$, and for each j such that $\pi_j = !n_{(r)}$, return a copy of the reaction $(X + X \rightarrow^{2r} P_i + P_j)$ ".

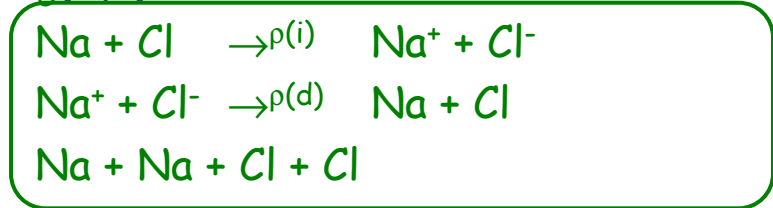
Example



$E, P:$



$\text{Ch}_G(E, P):$



$$\rho_E(n_{(r)}) = r$$

$\text{Ch}_G(E) :=$

$$\begin{aligned}
 &\dots \\
 &\cup^m \{(X + Y \xrightarrow{r} P + Q) \text{ s.t. } X \neq Y, \langle (X \equiv ?n_{(r)}; P \oplus \dots), (Y \equiv !n_{(r)}; Q \oplus \dots) \rangle \in E^2\}^m \\
 &\cup^m \dots
 \end{aligned}$$

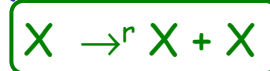
Subtler Examples

These are not *finite state systems*, but *finite species systems* are ok!

E:



C(E):

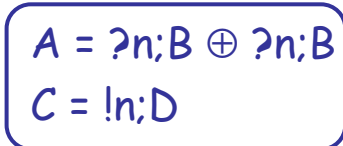


Unbounded state,
but only 1 species.
No problem!

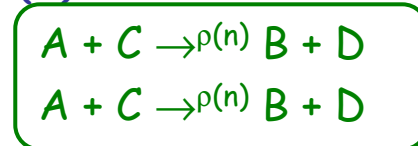
Multisets:

The same interaction can occur multiple times and must be taken into account:

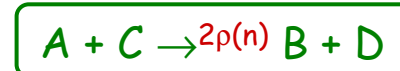
E:



C(E):



That is:

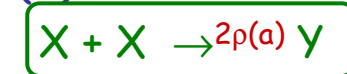


Symmetric reactions:

E:



C(E):



The rate of a was pre-halved and must be restored.

CPF to Chemistry: Handling Parameters

Consider first the CPF subset with no communication (pure $\tau_n, !n$).

Grounding (replace parameters with constants)

where X/p is a name in bijection with $\langle X, p \rangle$
(each X/p is seen as a separate *species*)

$$E ::= X_1(p_1)=M_1, \dots, X_n(p_n)=M_n$$

$$M ::= \pi_1;P_1 \oplus \dots \oplus \pi_n;P_n$$

$$P ::= X_1(p_1) \mid \dots \mid X_n(p_n)$$

$$\pi ::= \tau_r \quad \tau_n \quad !n$$

$$\begin{aligned} /(\pi_1;P_1 \oplus \dots \oplus \pi_n;P_n) &=_{\text{def}} \pi_1;/(P_1) \oplus \dots \oplus \pi_n;/(P_n) \\ /(X_1(p_1) \mid \dots \mid X_n(p_n)) &=_{\text{def}} X_1/p_1 \mid \dots \mid X_n/p_n \end{aligned}$$

Let N be the set of free names occurring in E (names not bound by definitions).

E_G is the **Parametric Explosion** of E (still a finite species system)
computed by replacing parameters with **all** combinations of free names in E

$$E_G := \{(X/q = /(M\{p \leftarrow q\})) \text{ s.t. } (X(p) = M) \in E \text{ and } q \in N^{\#p}\}$$

$$P_G := /P \quad (\text{simply ground the given initial conditions once})$$

E_G is a CGF! To obtain the chemical reactions $Ch_p(E)$, just compute $Ch_G(E_G)$

$$Ch_p(E) = Ch_G(E_G)$$

Example

E, P (with free names x, y):

$$\begin{aligned} P(a,b) &= ?a; (P(a,b) \mid Q(a)) \\ Q(a) &= !a; 0 \\ P(x,y) &\mid Q(x) \end{aligned}$$

Parametric Explosion

E_G, P_G :

$$\begin{aligned} Q/x &= !x; 0 \\ Q/y &= !y; 0 \\ P/x,y &= ?x; (P/x,y \mid Q/x) \\ P/y,x &= ?y; (P/y,x \mid Q/y) \\ P/x,x &= ?x; (P/x,x \mid Q/x) \\ P/y,y &= ?y; (P/y,y \mid Q/y) \\ P/x,y &\mid Q/x \end{aligned}$$

C :

$$\begin{aligned} Q/x + P/x,y &\xrightarrow{\rho(x)} P/x,y + Q/x \\ Q/x + P/x,x &\xrightarrow{\rho(x)} P/x,x + Q/x \\ Q/y + P/y,x &\xrightarrow{\rho(y)} P/y,x + Q/y \\ Q/y + P/y,y &\xrightarrow{\rho(y)} P/y,y + Q/y \\ P/x,y + Q/x & \end{aligned}$$

Chemical Reactions

Reduced C
(from initial conditions)

$$\begin{aligned} Q/x + P/x,y &\xrightarrow{\rho(x)} P/x,y + Q/x \\ P/x,y + Q/x & \end{aligned}$$

Note that E_G is not necessarily the "full" explosion (there may not be enough free variables in E, P to generate it). In particular, if E, P is closed (no free variables), then E_G contains only the parameterless definitions, which must then be closed and hence unrelated to any parametric definitions. One can add initial conditions $X_1(\mathbf{p}_1) \mid \dots \mid X_n(\mathbf{p}_n)$ where X_i are all the definitions in E and $\mathbf{p}_1 \dots \mathbf{p}_n$ are made of distinct (free) variables. Then E_G becomes the most general set of chemical reactions from E , out of any possible initial conditions.

Iterative Algorithm

The chemical reactions for E , computed directly from the parametric explosion E_G are highly redundant because they include all the parameter permutation symmetries. The following iterative algorithm, for the parametric case, computes a subset of E_G from the initial conditions of E, P . It produces a (usually) much smaller although not necessarily minimal set C .

Here:

C is the incrementally built set of chemical reactions

E_c (a CGF!) is the incrementally built set of grounded definitions from the original E .

initialization

$E_c := \{(X/q = / (M\{p \leftarrow q\})) \text{ s.t. } X(q) \text{ occurs in } P \text{ and } (X(p) = M) \in E\}$ (from initial conditions P)

iteration

$C := Ch_G(E_c)$

$E_c' := E_c \cup \{(X/q = / (M\{p \leftarrow q\})) \text{ s.t. } X/q \text{ occurs in } C \text{ and } (X(p) = M) \in E\}$

termination

if $E_c' = E_c$ then stop and return $(C, / (P))$ else $E_c := E_c'$ and iterate.

The algorithm terminates because E_c never shrinks and is always a subset of E_G , which is finite.

Example: Neg(x, x)

E =

Neg(a,b) = ?a; Inh(a,b) \oplus τ_ϵ ; (Tr(b) | Neg(a,b))
 Inh(a,b) = τ_η ; Neg(a,b)
 Tr(b) = !b; Tr(b) \oplus τ_δ ; 0
 Neg(x,x)

----- initialization -----

$E_c := \{ \text{Neg}/x,x = ?x; \text{Inh}/x,x \oplus \tau_\epsilon; (\text{Tr}/x | \text{Neg}/x,x) \}$

----- iteration 1 -----

$C := \{ \text{Neg}/x,x \xrightarrow{\epsilon} \text{Tr}/x + \text{Neg}/x,x \}$

$E_c := \{ \text{Neg}/x,x = ?x; \text{Inh}/x,x \oplus \tau_\epsilon; (\text{Tr}/x | \text{Neg}/x,x) \}$
 $\text{Tr}/x = !x; \text{Tr}/x \oplus \tau_\delta; 0 \}$

----- iteration 2 -----

$C := \{ \text{Neg}/x,x \xrightarrow{\epsilon} \text{Tr}/x + \text{Neg}/x,x \}$
 $\text{Tr}/x \xrightarrow{\delta} 0$
 $\text{Tr}/x + \text{Neg}/x,x \xrightarrow{\rho(x)} \text{Tr}/x + \text{Inh}/x,x \}$

$E_c := \{ \text{Neg}/x,x = ?x; \text{Inh}/x,x \oplus \tau_\epsilon; (\text{Tr}/x | \text{Neg}/x,x) \}$
 $\text{Tr}/x = !x; \text{Tr}/x \oplus \tau_\delta; 0$
 $\text{Inh}/x,x = \tau_\eta; \text{Neg}/x,x \}$

----- iteration 3 -----

$C := \{ \text{Neg}/x,x \xrightarrow{\epsilon} \text{Tr}/x + \text{Neg}/x,x \}$
 $\text{Tr}/x \xrightarrow{\delta} 0$
 $\text{Tr}/x + \text{Neg}/x,x \xrightarrow{\rho(x)} \text{Tr}/x + \text{Inh}/x,x$
 $\text{Inh}/x,x \xrightarrow{\eta} \text{Neg}/x,x \}$

$E_c :=$ no change

----- termination -----

$\text{Neg}/x,x \xrightarrow{\epsilon} \text{Tr}/x + \text{Neg}/x,x$
 $\text{Tr}/x \xrightarrow{\delta} 0$
 $\text{Tr}/x + \text{Neg}/x,x \xrightarrow{\rho(x)} \text{Tr}/x + \text{Inh}/x,x$
 $\text{Inh}/x,x \xrightarrow{\eta} \text{Neg}/x,x$
 $\text{Neg}/x,x$

Example: $\text{Tr}(x) \mid \text{Neg}(x,y)$

$E =$

$$\begin{aligned} \text{Neg}(a,b) &= ?a; \text{Inh}(a,b) \oplus \tau_\varepsilon; (\text{Tr}(b) \mid \text{Neg}(a,b)) \\ \text{Inh}(a,b) &= \tau_\eta; \text{Neg}(a,b) \\ \text{Tr}(b) &= !b; \text{Tr}(b) \oplus \tau_\delta; 0 \\ \text{Tr}(x) \mid \text{Neg}(x,y) \end{aligned}$$

----- initialization -----

$$E^c := \{ \text{Tr}/x = !x; \text{Tr}/x \oplus \tau_\delta; 0 \\ \text{Neg}/x,y = ?x; \text{Inh}/x,y \oplus \tau_\varepsilon; (\text{Tr}/y \mid \text{Neg}/x,y) \}$$

----- iteration 1 -----

$$C := \{ \text{Tr}/x \rightarrow^\delta 0 \\ \text{Neg}/x,y \rightarrow^\varepsilon \text{Tr}/y + \text{Neg}/x,y \\ \text{Tr}/x + \text{Neg}/x,y \rightarrow^{\rho(x)} \text{Tr}/x + \text{Inh}/x,y \}$$

$$E^c := \{ \text{Tr}/x = !x; \text{Tr}/x \oplus \tau_\delta; 0 \\ \text{Neg}/x,y = ?x; \text{Inh}/x,y \oplus \tau_\varepsilon; (\text{Tr}/y \mid \text{Neg}/x,y) \\ \text{Tr}/y = !y; \text{Tr}/y \oplus \tau_\delta; 0 \\ \text{Inh}/x,y = \tau_\eta; \text{Neg}/x,y \}$$

----- iteration 2 -----

$$C := \{ \text{Tr}/x \rightarrow^\delta 0 \\ \text{Neg}/x,y \rightarrow^\varepsilon \text{Tr}/y + \text{Neg}/x,y \\ \text{Tr}/x + \text{Neg}/x,y \rightarrow^{\rho(x)} \text{Tr}/x + \text{Inh}/x,y \\ \text{Tr}/y \rightarrow^\delta 0 \\ \text{Inh}/x,y \rightarrow^\eta \text{Neg}/x,y \}$$

$E^c :=$ no change

----- termination -----

$$\begin{aligned} \text{Tr}/x &\rightarrow^\delta 0 \\ \text{Neg}/x,y &\rightarrow^\varepsilon \text{Tr}/y + \text{Neg}/x,y \\ \text{Tr}/x + \text{Neg}/x,y &\rightarrow^{\rho(x)} \text{Tr}/x + \text{Inh}/x,y \\ \text{Tr}/y &\rightarrow^\delta 0 \\ \text{Inh}/x,y &\rightarrow^\eta \text{Neg}/x,y \\ \text{Tr}/x + \text{Neg}/x,y \end{aligned}$$

CPF to Chemistry: Handling Communication

Grounding (replace parameters with constants)

just one main change: now also convert each input parameter into a ground choice of all possible inputs

N is the set of free names in E, P

$\#p$ is the length of p

n/p is a name in bijection with $\langle n, p \rangle$

X/p is a name in bijection with $\langle X, p \rangle$

(each X/p is seen as a separate *species*)

$$E ::= X_1(p_1)=M_1, \dots, X_n(p_n)=M_n$$

$$M ::= \pi_1;P_1 \oplus \dots \oplus \pi_n;P_n$$

$$P ::= X_1(p_1) \mid \dots \mid X_n(p_n)$$

$$\pi ::= \tau_r \quad ?n(p) \quad !n(p)$$

$$/N(\tau_r;P) = \tau_r; /N(P)$$

$$/N(!n_{(r)}(p);P) = !n/p_{(r)}; /N(P)$$

$$/N(?n_{(r)}(p);P) = \oplus_{(q \in N^{\#p})} \text{of } ?n/q_{(r)}; /N(P\{p \leftarrow q\})$$

$$/N(\pi_1;P_1 \oplus \dots \oplus \pi_n;P_n) = /N(\pi_1;P_1) \oplus \dots \oplus /N(\pi_n;P_n)$$

$$/N(X_1(p_1) \mid \dots \mid X_n(p_n)) = X_1/p_1 \mid \dots \mid X_n/p_n$$

E_G is again the **Parametric Explosion** of E

$$E_G := \{(X/q = /N(M\{p \leftarrow q\})) \text{ s.t. } (X(p) = M) \in E \text{ and } q \in N^{\#p}\}$$

$$P_G := /N(P) \quad (\text{simply ground the given initial conditions once})$$

$$\text{Ch}(E) = \text{Ch}_G(E_G) \quad E_G \text{ is again a CGF!}$$

Example

E (with free names x, y):

$$\begin{aligned} P(a,b) &= ?a(c); (P(c,b) \mid Q(c)) \\ Q(a) &= !a(a); 0 \\ P(x,y) &\mid Q(x) \end{aligned}$$

Parametric Explosion

E_G :

$$\begin{aligned} Q/x &= !x/x; 0 \\ Q/y &= !y/y; 0 \\ P/x,y &= ?x/x; (P/x,y \mid Q/x) \oplus \\ &\quad ?x/y; (P/y,y \mid Q/y) \\ P/y,x &= ?y/x; (P/x,x \mid Q/x) \oplus \\ &\quad ?y/y; (P/y,x \mid Q/y) \\ P/x,x &= ?x/x; (P/x,x \mid Q/x) \oplus \\ &\quad ?x/y; (P/y,x \mid Q/y) \\ P/y,y &= ?y/x; (P/x,y \mid Q/x) \oplus \\ &\quad ?y/y; (P/y,y \mid Q/y) \\ P/x,y &\mid Q/x \end{aligned}$$

C:

$$\begin{aligned} Q/x + P/x,y &\xrightarrow{\rho(x)} P/x,y + Q/x \\ Q/x + P/x,x &\xrightarrow{\rho(x)} P/x,x + Q/x \\ Q/y + P/y,x &\xrightarrow{\rho(y)} P/y,x + Q/y \\ Q/y + P/y,y &\xrightarrow{\rho(y)} P/y,y + Q/y \\ P/x,y + Q/x & \end{aligned}$$

Chemical Reactions

Reduced C
(from initial conditions)

$$\begin{aligned} Q/x + P/x,y &\xrightarrow{\rho(x)} P/x,y + Q/x \\ P/x,y + Q/x & \end{aligned}$$

Example

E (with free names x, y):

$$\begin{aligned} P(a,b) &= ?a(c); (P(c,b) \mid Q(c,c)) \\ Q(a,b) &= !a(a); 0 \oplus !b(b); 0 \\ P(x,y) &\mid Q(x,y) \end{aligned}$$

E_G :

$$\begin{aligned} Q/x,y &= !x/x; 0 \oplus !y/y; 0 \\ Q/y,x &= !y/y; 0 \oplus !x/x; 0 \\ Q/x,x &= !x/x; 0 \oplus !x/x; 0 \\ Q/y,y &= !y/y; 0 \oplus !y/y; 0 \\ P/x,y &= ?x/x; (P/x,y \mid Q/x,x) \oplus \\ &\quad ?x/y; (P/y,y \mid Q/y,y) \\ P/y,x &= ?y/x; (P/x,x \mid Q/x,x) \oplus \\ &\quad ?y/y; (P/y,x \mid Q/y,y) \\ P/x,x &= ?x/x; (P/x,x \mid Q/x,x) \oplus \\ &\quad ?x/y; (P/y,x \mid Q/y,y) \\ P/y,y &= ?y/x; (P/x,y \mid Q/x,x) \oplus \\ &\quad ?y/y; (P/y,y \mid Q/y,y) \\ P/x,y &\mid Q/x \end{aligned}$$

C:

$$\begin{aligned} Q/x,y + P/x,y &\rightarrow^{\rho(x)} P/x,y + Q/x,x \\ Q/x,y + P/x,x &\rightarrow^{\rho(x)} P/x,x + Q/x,x \\ Q/x,y + P/y,x &\rightarrow^{\rho(y)} P/y,x + Q/y,y \\ Q/x,y + P/y,y &\rightarrow^{\rho(y)} P/y,y + Q/y,y \\ \\ Q/y,x + P/y,x &\rightarrow^{\rho(y)} P/y,x + Q/y,y \\ Q/y,x + P/y,y &\rightarrow^{\rho(y)} P/y,y + Q/y,y \\ Q/y,x + P/x,y &\rightarrow^{\rho(x)} P/x,y + Q/x,x \\ Q/y,x + P/x,x &\rightarrow^{\rho(x)} P/x,x + Q/x,x \\ \\ Q/x,x + P/x,y &\rightarrow^{2\rho(x)} P/x,y + Q/x,x \\ Q/x,x + P/x,x &\rightarrow^{2\rho(x)} P/x,x + Q/x,x \\ \\ Q/y,y + P/y,x &\rightarrow^{2\rho(y)} P/y,x + Q/y,y \\ Q/y,y + P/y,y &\rightarrow^{2\rho(y)} P/y,y + Q/y,y \\ P/x,y + Q/x,y & \end{aligned}$$

reduced C (from initial conditions)

$$\begin{aligned} Q/x,y + P/x,y &\rightarrow^{\rho(x)} P/x,y + Q/x,x \\ Q/x,x + P/x,y &\rightarrow^{2\rho(x)} P/x,y + Q/x,x \\ P/x,y + Q/x,y & \end{aligned}$$

Exponential Explosion via Communication

Parametric form (size 3).

$$\begin{aligned}
 &X(t, f) = !t(f); X(f, t) \\
 &A(x_1, x_2, x_3) = \\
 &\quad ?x_1(y); A(y, x_2, x_3) \oplus \\
 &\quad ?x_2(y); A(x_1, y, x_3) \oplus \\
 &\quad ?x_3(y); A(x_1, x_2, y) \\
 &A(t_1, t_2, t_3) \mid \\
 &X(t_1, f_1) \mid X(t_2, f_2) \mid X(t_3, f_3)
 \end{aligned}$$

green = free names with associated rates

Exponentially larger ground form (size 8).

$$\begin{aligned}
 &X/t_1, f_1 = !t_1/f_1; X/f_1, t_1 \\
 &X/f_1, t_1 = !f_1/t_1; X/t_1, f_1 \\
 &X/t_2, f_2 = !t_2/f_2; X/f_2, t_2 \\
 &X/f_2, t_2 = !f_2/t_2; X/t_2, f_2 \\
 &X/t_3, f_3 = !t_3/f_3; X/f_3, t_3 \\
 &X/f_3, t_3 = !f_3/t_3; X/t_3, f_3 \\
 &A/t_1, t_2, t_3 = \\
 &\quad ?t_1/f_1; A/f_1, t_2, t_3 \oplus \\
 &\quad ?t_2/f_2; A/t_1, f_2, t_3 \oplus \\
 &\quad ?t_3/f_3; A/t_1, t_2, f_3 \\
 &A/f_1, t_2, t_3 = \\
 &\quad ?f_1/t_1; A/t_1, t_2, t_3 \oplus \\
 &\quad ?t_2/f_2; A/t_1, f_2, t_3 \oplus \\
 &\quad ?t_3/f_3; A/t_1, t_2, f_3 \\
 &\dots\text{etc for the other bit flips} \\
 &A/t_1, t_2, t_3 \mid \\
 &X/t_1, f_1 \mid X/t_2, f_2 \mid X/t_3, f_3
 \end{aligned}$$

We have not fully expanded the input summations in $A(\dots)$: only enough for the channels that are actually used: e.g., there would never be a communication on channel t_1/f_2 .

CPF Iterative Algorithm

The same as before, just using the new grounding $/_N$
(based on the set N of free names of E, P , for expanding inputs).

initialization

$$N = \text{fn}(E, P)$$

$$E_C := \{(X/q = /_N(M\{p \leftarrow q\})) \text{ s.t. } X(q) \text{ occurs in } P \text{ and } (X(p) = M) \in E\} \quad (\text{initial conditions})$$

iteration

$$C := \text{Ch}_G(E_C)$$

$$E_C' := E_C \cup \{(X/q = /_N(M\{p \leftarrow q\})) \text{ s.t. } X/q \text{ occurs in } C \text{ and } (X(p) = M) \in E\}$$

termination

if $E_C' = E_C$ then stop and return $(C, /_N(P))$ else $E_C := E_C'$ and iterate.

A General π -calculus Normal Form

$E ::= X_1(\mathbf{p}_1) = (\nu \mathbf{q}_1) M_1, \dots, X_n(\mathbf{p}_n) = (\nu \mathbf{q}_n) M_n$	Definitions	$(n \geq 0)$
$M ::= \pi_1; P_1 \oplus \dots \oplus \pi_n; P_n$	Molecules	$(n \geq 0)$
$P ::= X_1(\mathbf{p}_1) \mid \dots \mid X_n(\mathbf{p}_n)$	Solutions	$(n \geq 0)$
$\pi ::= \tau_r \ ?n(\mathbf{p}) \ !n(\mathbf{p})$	Interactions	
NF ::= E, P	with initial conditions	

Any π -calculus process can be written in this normal form.
 But not all of those can be translated to chemical reactions,
 because we need to restrict to a *finite number of species*.

Example (complexation):

$$\begin{aligned}
 P(\mathbf{a}) &= (\nu \mathbf{n}_{(d)}) \ !a(\mathbf{n}); P'(\mathbf{a}, \mathbf{n}) \\
 P'(\mathbf{a}, \mathbf{n}) &= \ !n; P(\mathbf{a}) \\
 Q(\mathbf{a}) &= \ ?a(\mathbf{n}); Q'(\mathbf{a}, \mathbf{n}) \\
 Q'(\mathbf{a}, \mathbf{n}) &= \ ?n; Q(\mathbf{a}) \\
 P(x_{(c)}) &\mid Q(x_{(c)})
 \end{aligned}$$

Example (enzymatic reaction):

$$\begin{aligned}
 E(\mathbf{a}) &= (\nu \mathbf{n}_{(d)}, \mathbf{m}_{(p)}) \ !a(\mathbf{n}, \mathbf{m}); E'(\mathbf{a}, \mathbf{n}, \mathbf{m}) \\
 E'(\mathbf{a}, \mathbf{n}, \mathbf{m}) &= \ !n; E(\mathbf{a}) \oplus \ !m; E(\mathbf{a}) \\
 S(\mathbf{a}) &= \ ?a(\mathbf{n}, \mathbf{m}); S'(\mathbf{a}, \mathbf{n}, \mathbf{m}) \\
 S'(\mathbf{a}, \mathbf{n}, \mathbf{m}) &= \ ?n; S(\mathbf{a}) \oplus \ ?m; P() \\
 P() &= \tau_0; 0 \\
 E(x_{(c)}) &\mid S(x_{(c)}) \mid S(x_{(c)})
 \end{aligned}$$

Summary

- From Chemistry to ODEs
 - Via the stoichiometric matrix
- From Chemistry to Processes
 - Via the interaction matrix
- From Processes to Chemistry
 - Ground processes: by analysis of normal form interactions
 - Parametric processes: by parametric explosion
- Compositionality
 - Processes are compositional in components (you just “add components”) and chemistry is compositional in reactions (you just “add reactions”). But:
 - Adding reactions has a *linear* effect on the number of components (adding one reaction adds one or two more process actions).
 - Adding components has a *quadratic* effect on the number of reactions (adding one species may add one reaction for each species).
 - ODEs grow quadratically along with the chemical reactions (stoichiometric matrix has one column per reaction).
 - Parameterization is available only on processes.

Q?