# Artificial Biochemistry <br> Combining Stochastic Collectives 

## Luca Cardelli

Microsoft Research MSR-UniTN CC\&SB

Trento 2006-04-03
www. Iuca.demon.co.uk

## 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 ("finite state" and "collective")
- Not quite calculus (rate of change of "automata"??)
- Cf. "multi-agent systems" and "swarm intelligence"
- "Stochastic":
- Interactions have rates
- 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 accumated in Systems Biology is described as state transition diagrams [Kitano].


## State Transitions



## Even More State Transitions



## Reverse Engineering Nature

- That's what Systems Biology is up agains $\dagger$
- Exemplified by a technological analogy:
- Tamagotchi: a technological organism
- Has inputs (buttons) and outputs (screen/sound)
- It has state: happy or needy (or hungry, sick, dead...)
- Has to be petted at a certain rate (or gets needy)
- Each one has a slightly different behavior
- Reverse Engineering Tamagotchi
- Running experiments that elucidate their behavior
- Building models that explain the experiments
- Applications
- Engineering: Can we build our own Tamagotchi? (Sadly, no longer made.)
- Maintenance: Can we fix a broken Tamagotchi?


## Understanding T.Nipponensis

- Tamagotchi Nipponensis: a stochastic interactive automata
- 40 million sold worldwide; discontinued in 1998
- Still found "in the wild" in Akihabara
- Traditional scientific investigations fail

- Design-driven understanding fails
- We cannot read the manual (Japanese)
- What does a Tamagotchi "compute"? What is its "purpose"?
- Why does it have 3 buttons?
- Mechanistic understanding fails
- Few moving parts. Removing components mostly ineffective or "lethal"
- The "tamagotchi folding problem" (sequence of manufacturing steps) is too hard and gives little insight on function
- Behavioral understanding fails
- Subjecting to extreme conditions reveals little and may void warranty
- Does not answer consistently to individual stimuli, nor to sequences of stimuli
- There are stochastic variations between individuals
- Ecological understanding fails
- Difficult to observe in its native environment (kids' hands)
- Mass produced in little-understood automated factories
- It evolved by competing with other products in the baffling Japanese market
- Mathematical understanding fails
- What differential equations does it obey? (Uh?)



## A New Approach

- "Systems Technology" of T. Nipponensis
- High-throughput experiments (get all the information you possibly can)
- Decode the entire software and hardware
- Take sequences of tamagotchi screen dumps under different conditions
- Put 300 in a basket and shake them; make statistics of final state
- Modeling (organize all the information you got)
- Ignore the "folding" (manufacturing) problem
- Ignore materials (it's just something with buttons, display, and a program.)
- Abstract until you find a conceptual model (ah-ha: it's a stochastic automata).
- Do we understand what stochastic automata collectives can do?



## Automata Collectives

## Interacting Automata



Communicating automata: a graphical FSA-like notation for "finite state restriction-free $\pi$ calculus processes". Interacting automata do no $\dagger$ 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 automata" to obtain the underlying finite Markov transition system. [Buchholz]

## Interacting Automata Transition Rules



## Groupies and Celebrities



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)

$$
\begin{aligned}
& \text { directive sample } 5.01000 \\
& \text { directive plot } A() ; B() \\
& \text { new } a @ 1.0 \text { :chan( }) \\
& \text { new } b @ 1.0 \text { :chan( } \\
& \text { let } A()=\text { do !a; } A() \text { or ?b; } B() \\
& \text { and } B()=\text { do !b; } B() \text { or ?a; } A() \\
& \text { run } 100 \text { of }(A() \mid B())
\end{aligned}
$$

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 tiny bit of "noise" can make a huge difference



! b

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


```
directive sample 10.0 1000
directive plot Ga();Gb();Ca();Cb()
new a@1.0:chan()
new b@1.0:chan()
let Ca() = do !a;Ca() or ?a; Cb()
and }\textrm{Cb}()=do!b;Cb() or ?b;Ca(
let Ga() = do !a;Ga() or ?b;Gb()
and Gb()=do!b;Gb() or ?a;Ga()
run 1 of (Ca()|Cb())
run 100 of (Ga()|Gb())
```



## Doped Groupies

A similar way to break the deadlocks: destabilize the groupies by a small perturbation.
?a ?b


Doping ${ }^{(1)}$
directive sample 10.01000
directive plot $G a() ; G b() ; D a() ; D b()$

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

let $G a()=$ do !a; $G a()$ or ? $b ; G b()$
and $G b()=d o!b ; G b()$ or ? $a ; G a()$

$$
\text { let } D a()=\text { la; } D a()
$$

$$
\text { and } D b()=!b ; D b()
$$

$$
\text { run } 1 \text { of }(D a() \mid D b())
$$

$$
\text { run } 100 \text { of }(G a() \mid G b())
$$



## Hysteric Groupies

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

! b


(Still with doping)

directive sample 10.01000 directive plot $G a() ; G b()$
new a@1.0:chan() new b@1.0:chan()
let $G a()=$ do !a; $G a()$ or ?b; ?b; $G b()$ and $G b()=d o!b ; G b()$ or ?a; ?a; $G a()$
let $D a()=!a ; D a()$ and Db()$=!\mathrm{b} ; \mathrm{Db}()$
run 100 of $(\mathrm{Ga}() \mid G b())$
run 1 of $(\mathrm{Da}() \mid \mathrm{Db}())$
directive sample 10.01000
directive plot $G a() ; G b()$
new a@1.0:chan() new b@1.0:chan()
let $G a()=$ do !a; $G a()$ or ? $b ;$ ?b; ?b; $G b()$ and $G b()=d o!b ; G b()$ or ?a; ?a; ?a; $G a()$
let $D a()=!a ; D a()$
and Db()$=!\mathrm{b} ; \mathrm{Db}()$
run 100 of $(G a() \mid G b())$
run 1 of $(\mathrm{Da}() \mid \mathrm{Db}())$

## Hysteric 3-Way Groupies


directive sample 3.01000
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 ? $c ; ? c ; C()$
and $B()=d o!b ; B()$ or ? $a ; ? a ; A()$ and $C()=$ do !c; $C()$ or ?b; ?b; $B()$
let $D a()=!a ; D a()$ and Db()$=!\mathrm{b} ; \mathrm{Db}()$ and $D C()=!c ; D C()$
run 100 of $(A()|B()| C())$
run 1 of $(D a()|D b()| D c())$


## The Strength of Populations



At size $2 N$, on a shared channel, $\mu$ is $N$ times stronger than $\wedge$ : interaction easily wins over delay.
directive sample 0.011000
directive plot $B()$
$\mathrm{val} \mathrm{lam}=1000.0$
$\mathrm{val} \mathrm{mu}=1.0$
new a@mu:chan
let $A()=!a ; A()$
and $B()=? a ; C()$
and $C()=$ delay@lam; $B()$
run 1000 of $(A() \mid B())$


## Boolean Inverter Collectives

$b=n o t a$ b $b$
in presence of $a, b$ goes low in absence of $a, b$ goes high



$b=\operatorname{not} a$

the high b state reinforces itself (as a population)



$b=\operatorname{not} a$




[^0]new ele. 1. chan new bet. C .char



do
run 100 of Tm2 Lhi(a.b)











and










## Boolean Gate Collectives



$$
\sum A>A>\sum A>A
$$

Bidirectional
Polymerization
new c@ $\mu$ new stop@1.0

$$
\begin{aligned}
& A_{\text {free }}= \\
& \quad\left(\text { new rht@ } 1 \text {; !c(rht); } A_{\text {brht }}(r h t)\right) \\
& \quad+? c(I f t) ; A_{\text {blft }}(I f t) \\
& A_{\text {bff }}(I f t)= \\
& \left(\text { new rht@ } 1 ;!c(r h t) ; A_{\text {bound }}(I f t, r h t)\right) \\
& A_{\text {brht }}(r h t)= \\
& \text { ?c(lft); } A_{\text {bound }}(I f t, r h t) \\
& A_{\text {bound }}(I f t, r h t)=? \text { stop }
\end{aligned}
$$




Communicating Automata
Bound output !c ${ }^{( }{ }^{v} r$ r) and input ?ce) on automata transitions to model complexation

Monomer Automata


Bound both


## Bidirectional Polymerization

## Circular Polymer Lengths

Scanning and counting the size of the circular polymers (by a cheap trick).
Polymer formation is complete within 10t; then a different polymer is scanned every $100 t$.

directive sample 1000.0 directive plot Abound(); ?count

## type Link = chan(chan)

type Barb = chan
val lam $=1000.0$ (* set high for better counting *)
val $m u=1.0$
new c@mu:chan(Link)
new enter@lam:chan(Barb)
new count@lam:Barb
let Afree () $=$
(new rht@lam:Link run
do !c(rht); Abrht(rht)
or ?c(lft); Ablft(lft))
and $\mathrm{Ablft}(\mid f t:$ Link $)=$
(new rht@lam:Link run
! $\mathrm{c}(\mathrm{rht})$ ) Abound( $(\mathrm{f} \dagger$, rht $)$ )
and Abrht(rht:Link) =
?c(lft); Abound(lft,rht)
and Abound(lft:Link, rht:Link) =
do ?enter(barb): (?barb |!rht(barb))
or ?|ft(barb): (?barb | ! rht(barb))
(* each Abound waits for a barb, exhibits it, and passes it to the right so we can plot number of Abound in a ring *)
let clock(t:float, tick:chan) = (* sends a tick every $\dagger$ time *) (val $+i=t / 1000.0$ val $d=1.0 /+i$
let $\operatorname{step}(n: i n t)=$
if $n<=0$ then !tick; clock(t,tick) else delay@d; step( $n-1$ )
run step(1000))
new tick:chan
let Scan() = ?tick; lenter(count); Scan()
run 100 of Afree()
run (clock(100.0, tick) | Scan())
$100 \times A_{\text {free, }}$ initially.
The height of each rising step is the size of a separate circular polymer.
(Unbiased sample of nine consecutive runs.)

new c@u
$A_{\text {free }}=$
(new Ift@へ; !c(lft); $\left.A_{b l f t}(I f t)\right)+$
?c $(r h t) ; A_{\text {brht }}(r h t)$
$A_{\text {blft }}(\mathrm{lft})=$
!Ift; $A_{\text {free }}+$
?c(rht); $A_{\text {bound }}(\mathrm{lft}, \mathrm{rht})$
$A_{\text {brht }}(r h t)=$
?rht; $A_{\text {free }}$
$A_{\text {bound }}(I f t, r h t)=$
! lft ; $A_{\text {brht }}(r h t)$
Poly/Depolymerization



Monomer
Automata


Free


Bound
both Free


## The Law of Mass Interaction

## Law of Mass Interaction

The speed of interaction ${ }^{\dagger}$ is proportional to the number of possible interactions.

Decay


## Exponential

 Decay lawRate of change proportional to number of possible decays.

Mass interaction


Interaction Law generalizes Decay Law

## Mass

Interaction law Rate of change proportional to number of possible interactions
${ }^{+}$speed of interaction (formally definable)
= number of interactions over time
not proportional to the number of interacting processes! $[P]$ is the number of processes $P$ (this is informal; it is only meaningful for a set of processes offering a given action, but a set of such processes can be counted and plotted)


Chemical Law of Mass Action http://en.wikipedia.org/wiki/Chemical_kinetics The speed of a chemical reaction is proportional to the activity of the reacting substances.
(Activity = concentration, for wellstirred aqueous medium)
(Concentration = number of moles per liter of solution)
(Mole $=6.022141 \times 10^{23}$ particles)


## Activity and Speed

stochastic algebras disagree!

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

c activity: 1 speed: $\wedge$
= The activity (= "concentration") on a channel is the number of possible interactions on that channel.

The speed of interaction on a channel, is the activity multiplied by the base rate of the channel.
directive sample 0.0110000 directive plot $A 1$ (): $A 2() ; A 3()$
new c1@1.0:chan
new c2@1.0:chan new c3@1.0:chan
let $A 1()=? c 1$
and $B 1()=!c 1$
let A 2()$=\mathrm{do}$ ? c2 or ?c2 and $B 2()=!c 2$
let A 3()$=$ do ?c3 or ?c3 and $B 3()=$ do !c3 or !c3
run 1000 of (A1() | B1()
$|A 2()| B 2()|A 3()| B 3())$



The mass interaction law [Buchholz] [Priami-Regev-Shapiro-Silverman] is compatible with chemistry [Gillespie] and incompatible with any other stochastic algebra in the literature! (including [Priami]; see [Hermanns])

$$
\text { c activity: } 4
$$

$$
\text { speed: } 4 \wedge
$$

Other algebras assign rates to actions, not channels, with speed laws:

```
\[
2 \Lambda^{*} 2 \Lambda=4 \Lambda^{2}
\]
\[
\max (2 \Lambda, 2 \lambda)=2 \Lambda[\text { Goetz }]
\]
\[
\min (2 \Lambda, 2 \Lambda)=2 \Lambda[\text { Priami }]
\]
```

$$
1 /(1 /(2 \Lambda)+1 /(2 \Lambda))=\Lambda[P E P A]
$$

$$
2 \wedge^{* 1}=2 \Lambda \text { (passive inputs) }
$$



```
c activity: 2
speed: 2^
```




## Possible Interactions

The speed of interaction is proportional to the number of possible interactions.
But a process cannot interact with itself.
Assume each process $P$ is in restricted-sum-normal-form. For each channel $x$ :
$\operatorname{In}(x, P)=$ Num of active $? x$ in $P$
$\operatorname{Out}(x, P)=$ Num of active ! $x$ in $P$
$\operatorname{Mix}(x, P)=\operatorname{In}(x, P) * \operatorname{Out}(x, P)$
\#interactions that cannot happen
in a given summation $P$
$\operatorname{In}(x)=\operatorname{Sum} P$ of $\operatorname{In}(x, P)$
$\operatorname{Out}(x)=\operatorname{Sum} P$ of $\operatorname{Out}(x, P)$
$\operatorname{Mix}(x)=$ Sum $P$ of $\operatorname{Mix}(x, P)$
total \#interactions that cannot happen
The global Activity on channel $x$ :

$$
\operatorname{Act}(x)=\left(\operatorname{In}(x)^{\star} \operatorname{Out}(x)\right)-\operatorname{Mix}(x)
$$

total cross product of inputs and outputs minus total \#interactions that cannot happen
The global speed of interaction on a channel $x$ :

$$
\operatorname{speed}(x)=\operatorname{Act}(x)^{\star} \operatorname{rate}(x)
$$



## Deriving Back Interaction Laws

The mass action law:

$[A]^{\prime}=-\operatorname{speed}(c)=-\lambda \operatorname{Act}(c)$ $\operatorname{Act}(c)=\left(\operatorname{In}(c)^{*} \operatorname{Out}(c)\right)-\operatorname{Mix}(c)$
$=\left([A]^{\star}[B]\right)-0$


The mixed interaction law:

$[A]^{\prime}=-$ speed $(c)=-\Lambda \operatorname{Act}(c)$ $\operatorname{Act}(c)=\left(\operatorname{In}(c)^{*} \operatorname{Out}(c)\right)-\operatorname{Mix}(c)$ $=\left([A]^{\star}[A]\right)-[A]=[A]^{\star}([A]-1)$ hence $[A]^{\prime}=\because \because \because[A]([A]-1):$

$$
\operatorname{Act}(x)=\left(\operatorname{In}(x)^{\star} \operatorname{Out}(x)\right)-\operatorname{Mix}(x)
$$

The decay law:


```
\([A]^{\prime}=\Sigma\left(c_{i}\right)-\operatorname{speed}\left(c_{i}\right)\)
    \(=\Sigma\left(c_{i}\right)-\lambda \operatorname{Act}\left(c_{i}\right)\)
\(\operatorname{Act}\left(c_{i}\right)=\left(\operatorname{In}\left(c_{i}\right)^{\star} \operatorname{Out}\left(c_{i}\right)\right)-\operatorname{Mix}\left(c_{i}\right)\)
    \(=(1 * 1)-0=1\)
```



## Conclusions

## Conclusions

- Stochastic Collectives
- Complex global behavior from simple components
- Emergence of collective functionality from "non-functional" components
- (C.f. "swarm intelligence": simple global behavior from complex components)
- Artificial Biochemistry
- Stochastic collectives with Law of Mass Interaction kinetics
- Connections to classical Markov theory, chemical Master Equation, and Rate Equation
- The agent/automata/process point of view
- "Individuals" that transition between states (vs. transmutation between "unrelated" chemical species)
- More appropriate for Systems Biology
- Stochastic $\pi$-calculus (SPiM) for investigating stochastic collectives
- Restriction+Communication $\Rightarrow$ Polymerization: FSA that "stick together"


[^0]:    directive sample 110.012
    diective poltac: 10

