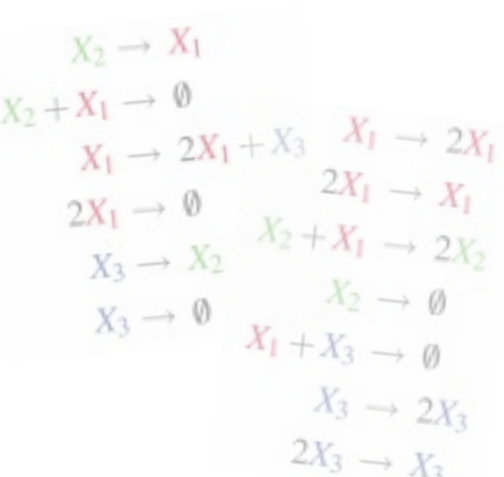
DNA 17 Tutorial 2:
The Programming Language of Chemical Kinetics, and How To Discipline Your DNA Molecules with Strand Displacement Cascades

David Soloveichik

(starting at UCSF)

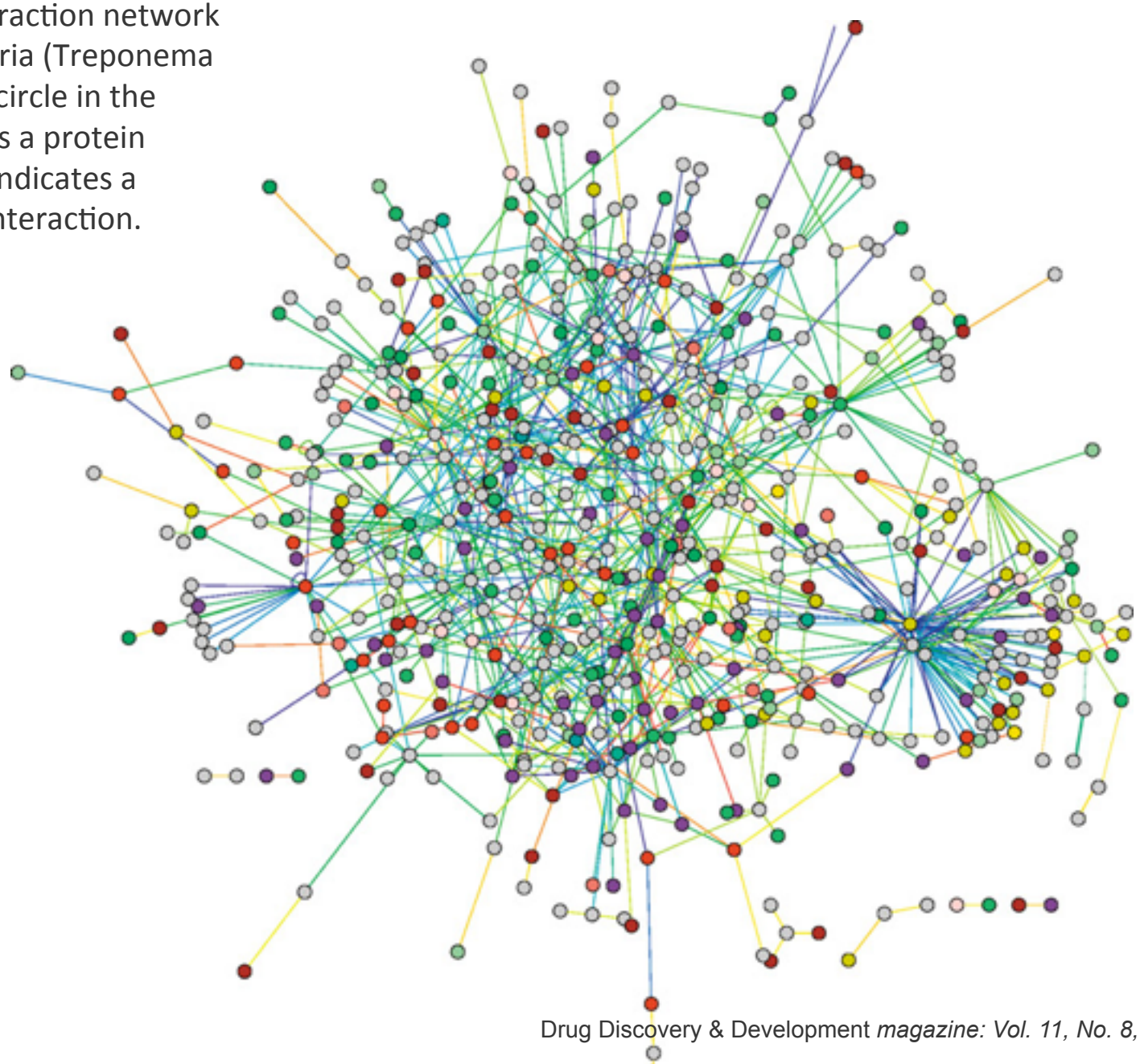
UNIVERSITY of WASHINGTON

CALIFORNIA INSTITUTE OF TECHNOLOGY 1891



Living cells are controlled by the most complex signaling networks. Can we (1) understand the function of these networks, (2) be able to engineer new ones with desired function?

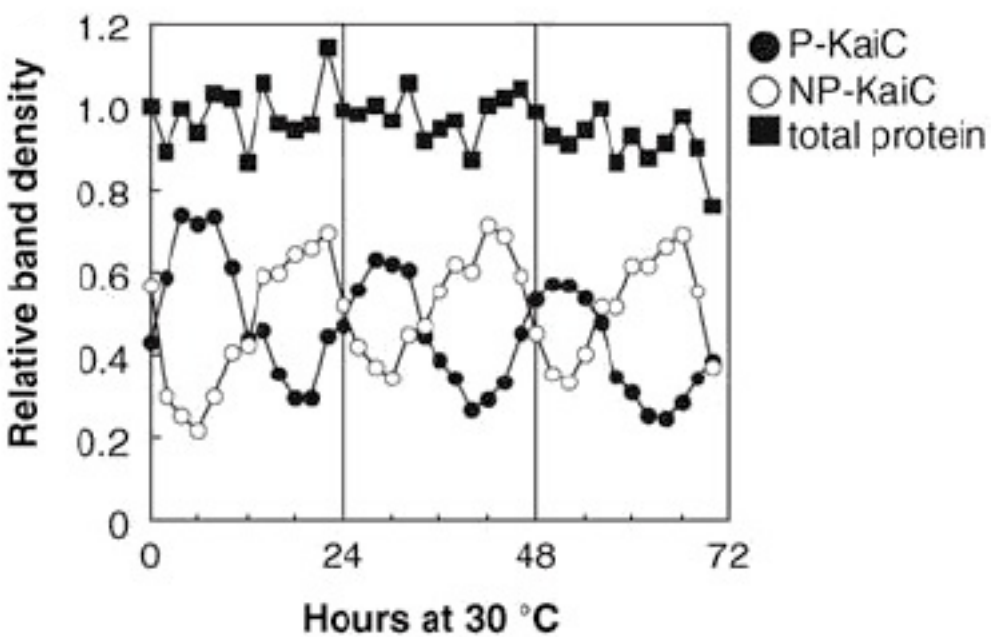
The protein interaction network of Syphilis bacteria (Treponema pallidum). Each circle in the figure represents a protein while each line indicates a direct physical interaction.



Drug Discovery & Development *magazine: Vol. 11, No. 8, August, 2008, pp. 22-25.*

Living cells are controlled by the most complex signaling networks. Can we (1) understand the function of these networks, (2) be able to engineer new ones with desired function?

in vitro data



model of phosphorylation reactions between A=KaiA, B=KaiB, C=KaiC

$$\mathrm{C}_i \underset{b_i}{\overset{f_i}{\rightleftarrows}} \widetilde{\mathrm{C}}_i,\ \mathrm{C}_i + \mathrm{A} \underset{k_i^{\mathrm{Ab}}}{\overset{k_i^{\mathrm{Af}}}{\rightleftarrows}} \mathrm{AC}_i \overset{k_{\mathrm{pf}}}{\rightarrow} \mathrm{C}_{i+1} + \mathrm{A} \qquad (1)$$

$$\widetilde{\mathrm{C}}_i + \mathrm{B} \underset{\tilde{k}_i^{\mathrm{Bb}}}{\overset{2\tilde{k}_i^{\mathrm{Bf}}}{\rightleftarrows}} \mathrm{B}\widetilde{\mathrm{C}}_i, \mathrm{B}\widetilde{\mathrm{C}}_i + \mathrm{B} \underset{2\tilde{k}_i^{\mathrm{Bb}}}{\overset{\tilde{k}_i^{\mathrm{Bf}}}{\rightleftarrows}} \mathrm{B}_2\widetilde{\mathrm{C}}_i \qquad (2)$$

$$\mathrm{B}_x\widetilde{\mathrm{C}}_i + \mathrm{A} \underset{\tilde{k}_i^{\mathrm{Ab}}}{\overset{x\tilde{k}_i^{\mathrm{Af}}}{\rightleftarrows}} \mathrm{AB}_x\widetilde{\mathrm{C}}_i, \mathrm{AB}_2\widetilde{\mathrm{C}}_i + \mathrm{A} \underset{2\tilde{k}_i^{\mathrm{Ab}}}{\overset{\tilde{k}_i^{\mathrm{Af}}}{\rightleftarrows}} \mathrm{A}_2\mathrm{B}_2\widetilde{\mathrm{C}}_i \qquad (3)$$

$$\mathrm{C}_i \underset{k_{\mathrm{dps}}}{\overset{k_{\mathrm{ps}}}{\rightleftarrows}} \mathrm{C}_{i+1}, \widetilde{\mathrm{C}}_i \underset{\tilde{k}_{\mathrm{dps}}}{\overset{\tilde{k}_{\mathrm{ps}}}{\rightleftarrows}} \widetilde{\mathrm{C}}_{i+1} \qquad (4)$$

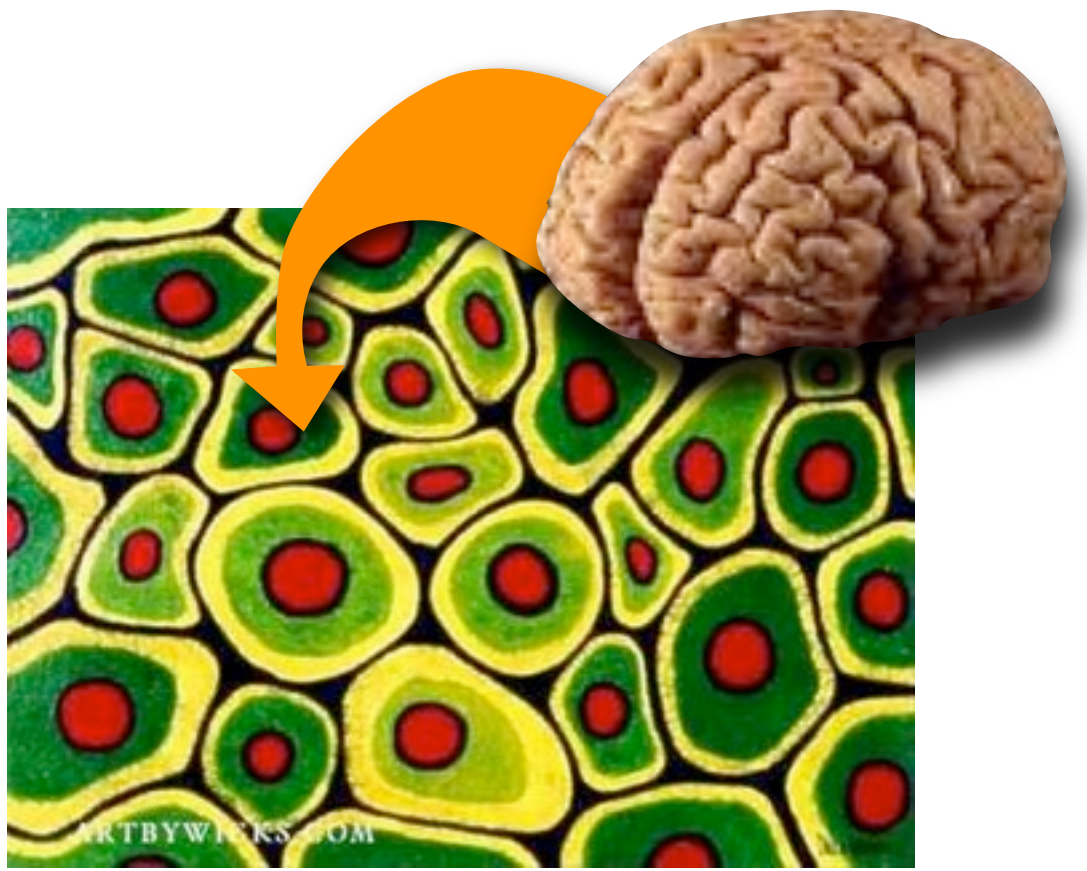
$$\mathrm{B}_x\widetilde{\mathrm{C}}_i \underset{\tilde{k}_{\mathrm{dps}}}{\overset{\tilde{k}_{\mathrm{ps}}}{\rightleftarrows}} \mathrm{B}_x\widetilde{\mathrm{C}}_{i+1}, \mathrm{A}_y\mathrm{B}_x\widetilde{\mathrm{C}}_i \underset{\tilde{k}_{\mathrm{dps}}}{\overset{\tilde{k}_{\mathrm{ps}}}{\rightleftarrows}} \mathrm{A}_y\mathrm{B}_x\widetilde{\mathrm{C}}_{i+1}\,. \qquad (5)$$

Najajima et al, Science, 308, 414-415, 2005
van Zon et al, PNAS, 104: 7420-7425, 2007

Living cells are controlled by the most complex signaling networks. Can we (1) understand the function of these networks, (2) be able to engineer new ones with desired function?

Want to notice patterns in biological signaling/regulatory networks

Engineer embedded controllers for artificial biochemical systems, "wet robots", smart drugs, etc

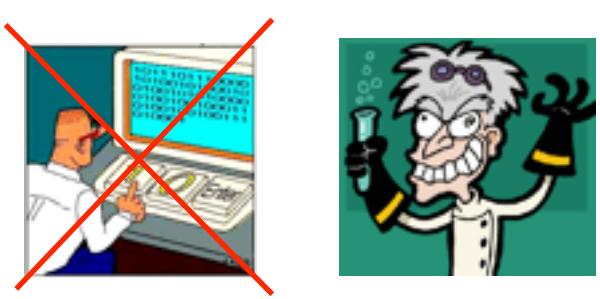


http://www.artbywicks.com

Part 1 of 2:

The Programming Language of Chemical Kinetics

Using the language of Chemical Reaction Networks (CRNs) prescriptively as a "programming language" rather than descriptively as a modeling language for existing systems



Real programmers code in CHEMISTRY

Outline of Part 1

- Stochastic CRN model
- Mass-action CRN model
- Stochastic CRNs:
    - illustrative examples: arithmetic operations
    - characterizing deterministic behavior
    - allowing error permits much more complex behavior (Turing universality)
- Mass-action CRNs:
    - dynamical systems
    - circuits

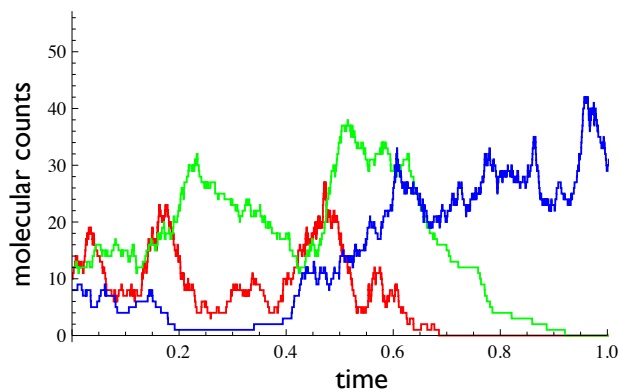
Chemical Reaction Networks (CRN)

syntax:

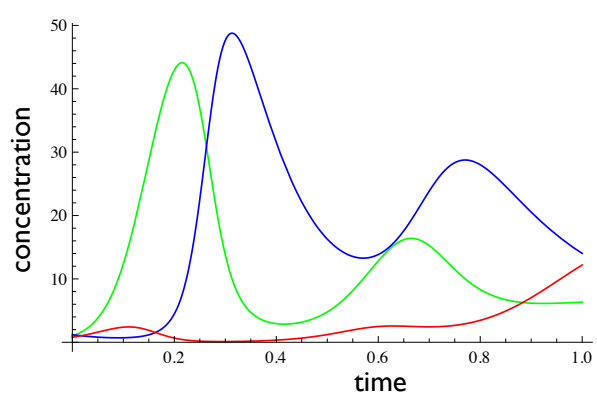
$$
\begin{aligned}
X_1 &\xrightarrow{30} 2X_1 \\
2X_1 &\xrightarrow{0.5} X_1 \\
X_2 + X_1 &\xrightarrow{1} 2X_2 \\
X_2 &\xrightarrow{10} \\
X_1 + X_3 &\xrightarrow{1} \\
X_3 &\xrightarrow{16.5} 2X_3 \\
2X_3 &\xrightarrow{0.5} X_3
\end{aligned}
$$

two possible semantics:

stochastic: discrete state space, continuous time Poisson process



mass-action: continuous ODEs



Stochastic CRNs model

- Finite set of **species** $\{A, B, C, D...\}$. A **state** is a vector of non-negative integers, specifying the molecular counts of each species. We also write molecular counts as $\#A, \#B...$

- Finite set of **reactions**. Each reaction is specified in chemical notation; for example:

$$A + B \xrightarrow{k} C$$

indicates a reaction in which the counts of A and B are decreased by 1, and the count of C is increased by 1. Each reaction has an associated **rate constant** k. (Unimolecular rate constants have units of sec^{-1}, bimolecular rate constants have units of $liters \cdot molecules^{-1} \cdot sec^{-1}$)

continued ➠

McQuarrie 1967, van Kampen, Gillespie 1977, etc

Stochastic CRNs model

- The system evolves via a **continuous time Poisson process**:

reaction type			propensity a_j: the probability of reaction *j* in time instant *dt*
A	$\xrightarrow{k}$	$\ldots$	$k \cdot \#A$
$A+B$	$\xrightarrow{k}$	$\ldots$	$k \cdot \#A \cdot \#B / v$
$A+A$	$\xrightarrow{k}$	$\ldots$	$k \cdot \#A(\#A-1)/v$

time until next reaction is exponential random variable with rate $\sum a_j$

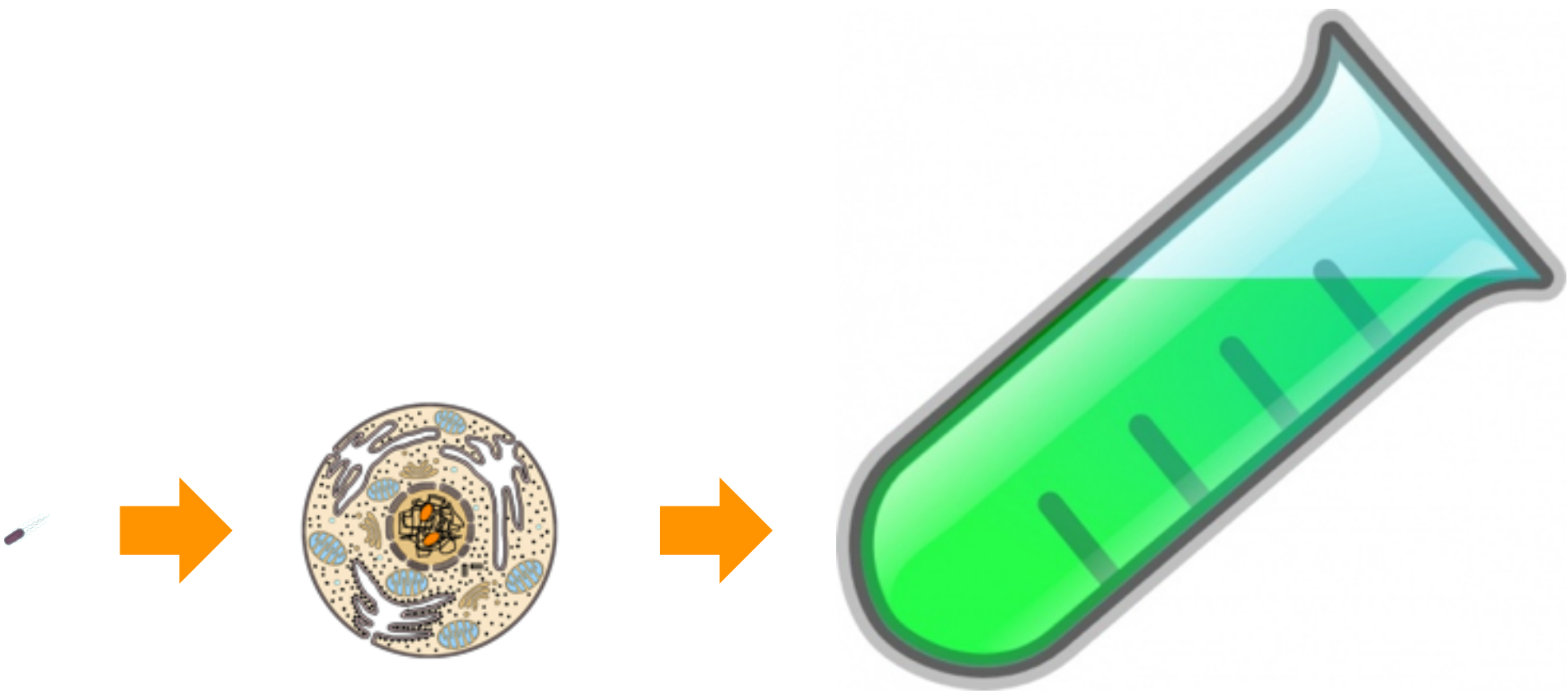
probability that next reaction is *j** is $a_{j^*} / \sum a_j$

McQuarrie 1967, van Kampen, Gillespie 1977, etc

Scaling from stochastic to mass-action

Increase solution volume v and the molecular counts of all species such that for each species $\#X/v$ stays constant. (Because we measure mass-action concentration in moles/liter, we have to multiply bimolecular rate constants by Avogadro's number.)

In this way we get mass-action regime in the limit $v \to \infty$.



Mass-action CRNs model

- Finite set of **species** $\{A, B, C, D...\}$. A **state** is a vector of non-negative real numbers, specifying the concentrations of each species. We also write concentrations as $[A]$, $[B]$...

- Finite set of **reactions**. Each reaction is specified in chemical notation; for example:

$$A + B \xrightarrow{k} C$$

  Each reaction has an associated **(mass-action) rate constant** *k*. (Unimolecular rate constants have units of sec^{-1}, bimolecular rate constants have units of $molar^{-1} \cdot sec^{-1}$)

continued ➠

Mass-action CRNs model

- In any state **reaction fluxes** are as follows:

reaction type			reaction flux a_j (moles · liter^{-1} · second^{-1})
A	$\xrightarrow{k}$	$\ldots$	$k \cdot [A]$
$A + B$	$\xrightarrow{k}$	$\ldots$	$k \cdot [A] \cdot [B]$
$A + A$	$\xrightarrow{k}$	$\ldots$	$k \cdot [A]^2$

- The system behaves according to the ODEs:

$$\frac{d}{dt}[X] = \sum_{\text{reaction } j} a_j \cdot [\text{net stoichiometry change of } X \text{ in reaction } i]$$

example:

$$\begin{aligned} A &\xrightarrow{k_1} B \\ B + B &\xrightarrow{k_2} C \end{aligned}$$

$$\begin{aligned} \frac{d[A]}{dt} &= -k_1[A] \\ \frac{d[B]}{dt} &= k_1[A] - 2k_2[B]^2 \\ \frac{d[C]}{dt} &= k_2[B]^2 \end{aligned}$$

general case formally:

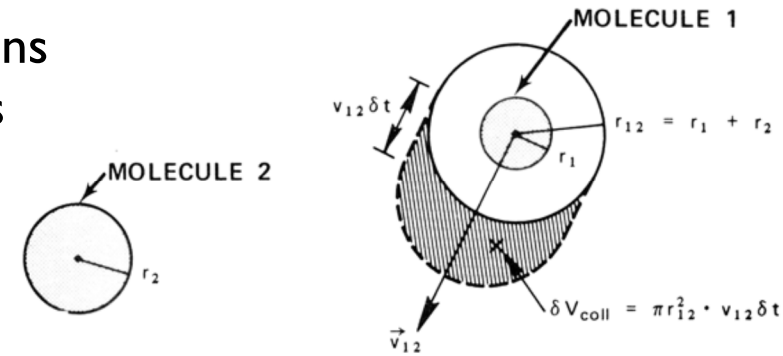
$$\nu^s_{1j} X_1 + \nu^s_{2j} X_2 + \nu^s_{3j} X_3 + \ldots \xrightarrow{k_j} \nu^f_{1j} X_1 + \nu^f_{2j} X_2 + \nu^f_{3j} X_3 + \ldots$$

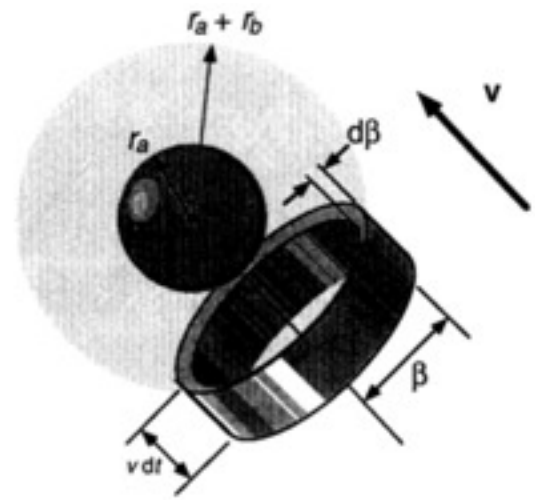
$$\frac{dx_i}{dt} = \sum_j (\nu^f_{ij} - \nu^s_{ij})\, k_j \prod_k x_k^{\nu^s_{kj}}$$

Physical justification for CRNs

Assumptions for stochastic:

- well-mixedness
- bouncing ball interactions
- instantaneous reactions





McQuarrie 1967, van Kampen, Gillespie 1977, 1992, etc

Extra assumptions for mass-action:

- large molecular counts of all species

Kurtz, "The relationship between stochastic and deterministic models for chemical reactions", *J Chem Phys* 57:2976, 1972

What about thermodynamics, conservation of mass?

- thermodynamics says all reactions must be reversible
- if a reaction can occur with a catalyst, it must occur without it (at a slower rate)
- closed system must satisfy Gibbs free energy
- molecules are made of atoms; must be consistent with atomic decomposition and conservation of mass

open systems, implicit energy and mass sources, effectively irreversible reactions, high energy barriers

Abstract CRNs have been extensively theoretically investigated

- simulation: accuracy, computation time
- equilibrium analysis: number of steady states, bistability?, oscillation?, limit cycle?, chaos?
- deviant behavior of stochastic compared to mass-action
- derived models: Michaelis-Menten, Hill functions, GRNs, S-Systems, NHCA, etc
- time-separation arguments
- network motifs: search and in-silico evolution

But designing molecular algorithms? Not so much...

Judging speed of chemical "algorithms"

- Can speed up any behavior by a factor of α if we multiply all rate constants by α. not meaningful

- Can speed up behavior by increasing molecular counts but keeping volume the same. This translates to increasing concentration in mass-action. impossible: must remain well-mixed, respect finite density of matter

Thus for meaningful analysis of speed of chemical "algorithms":

- fix largest rate constant (say $k=1$)

- asymptotically, volume v = O(total molecular count). This translates to bounding maximum concentrations in mass-action.

Programming exercises with stochastic CRNs

Programming exercise 1:

$\#B := 2 \cdot \#A$ $\quad$ n A's

Eventually produce the right number of *B*'s

$A \xrightarrow{1} B + B$

expected time of this algorithm: $\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{1} = \Theta(\log n)$ (fast!)

expected time for one instance of reaction whose propensity is #*A* = *n*

Programming exercise 2:

Detect a molecule of A among n molecules in solution volume $v \approx n$

Whole test-tube eventually "turns Yellow" if and only if there is at least one molecule of *A*

start with n X

$A + X \xrightarrow{1} A + Y$

$X + Y \xrightarrow{1} Y + Y$

expected time of this algorithm: $1 + \Theta(\log n) = \Theta(\log n)$ (fast!)

expected time for one instance of first reaction whose propensity is $(\#A \cdot \#X)/v = (1 \cdot n)/n = 1$

expected time for n instances of second reaction whose propensity is $(\#X \cdot \#Y) / v$

Programming exercises with stochastic CRNs

Programming exercise 3:

$\#A := 2 \cdot \#A$

n A's
volume $v \approx n$

Even in the limit $t \to \infty$ must be some probability of error.

Programming exercise 4:

Is $\#A$ even?

n A's
volume $v \approx n$

Eventually converge to the right answer.

start with $1T$ *and* n A

$A + T \xrightarrow{1} F$

$A + F \xrightarrow{1} T$

expected time of this algorithm: at least n (slow!)

Expected time for the last A to react with T or F.
Propensity of this reaction is $(\#A \cdot \#T,F) \,/\, v = (1 \cdot 1)/n$

How can stochastic CRNs compute "deterministically"?

Let's start with computing predicates (well-studied).

Species ***T*** and ***F*** representing output.
YES is absorbing set of states with #***T***>0, #***F***=0.
NO is absorbing set of states with #***T***=0, #***F***>0.
Everything else (not in **YES** or **NO**) can have arbitrary #***T***, #***F***.
YES is reachable from any state that is reachable from a yes-input, but is not reachable from any no-input.
NO is reachable from any state that is reachable from a no-input, but is not reachable from any yes-input.

(For simplicity assume finite state space for any input. Need to be more careful for infinite state spaces (eg. $A \rightarrow 2A$), but intuition is the same.)

Angluin, Aspnes, Eisenstat, Ruppert, "The Computational Power of Population Protocols," *Distributed Computing 20: 279-304* (2006)

How can stochastic CRNs compute "deterministically"?

Examples of computing a predicate in this way:

Predicate: $\#A > \#B$

start with $1F$ *and input amounts of A, B*

$A+F \rightarrow T$
$B+T \rightarrow F$

Predicate: $\#A$ is even

start with $1T$ *and input amount of A*

$A+T \rightarrow F$
$A+F \rightarrow T$

Predicate: $\#A > \#B$ and $\#A$ is even

start with $1\ F1, 1\ T2, 1\ F$ *and input amounts of A, B*

$A+T2 \rightarrow A'+F2$
$A+F2 \rightarrow A'+T2$
$A'+F1 \rightarrow T1$
$B+T1 \rightarrow F1$
$F1+T \rightarrow F1+F$
$F2+T \rightarrow F2+F$
$T1+T2 \leftrightarrow TT$
$TT+F \rightarrow TT+T$

How can stochastic CRNs compute "deterministically"?

Then the class of predicates that can be computed is well-characterized:

Boolean combination of **threshold predicates** and **modulo predicates:**

threshold predicate: $\{\boldsymbol{x} \mid \boldsymbol{x} \cdot \boldsymbol{v} \geq r\}$

modulo predicate: $\{\boldsymbol{x} \mid \boldsymbol{x} \cdot \boldsymbol{v} \equiv r \pmod m\}$

Stochastic CRNs are Turing Universal, allowing an (arbitrarily small, non-zero) error probability

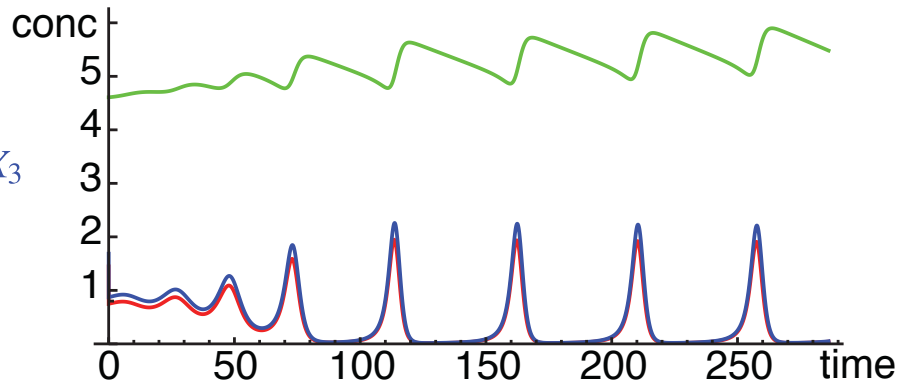
- Arbitrarily small, non-zero error probability over all time. Error probability controlled by initial molecular count of "accuracy species"
- Two kinds of constructions: Register Machine simulation, Turing Machine simulation
- Turing universal computation can be made fast: t=poly(number of TM steps).
- But: Register machine or Turing machine simulation just doesn't feel natural for CRNs. Information is stored in unary. Register machine: slow; Turing machine: too many reactions. Has a feeling of shoehorning existing paradigms to a very different system.

David Soloveichik, Matt Cook, Erik Winfree, Shuki Bruck, "Computation with Finite Stochastic Chemical Reaction Networks", *Natural Computing* 7:615-633 (2008)

Which dynamical systems can be implemented with mass-action CRNs?

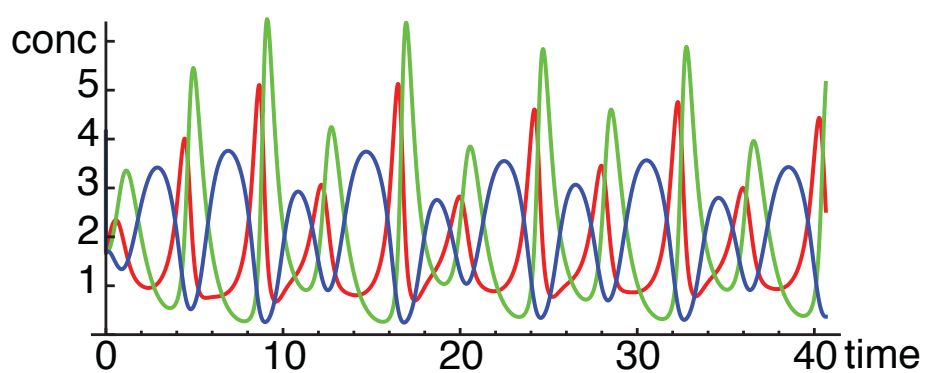
Oregonator (limit cycle oscillator)

$$X_2 \to X_1$$
$$X_2 + X_1 \to \emptyset$$
$$X_1 \to 2X_1 + X_3$$
$$2X_1 \to \emptyset$$
$$X_3 \to X_2$$
$$X_3 \to \emptyset$$



Rössler (chaotic)

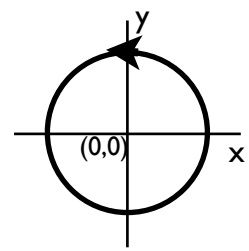
$$X_1 \to 2X_1$$
$$2X_1 \to X_1$$
$$X_2 + X_1 \to 2X_2$$
$$X_2 \to \emptyset$$
$$X_1 + X_3 \to \emptyset$$
$$X_3 \to 2X_3$$
$$2X_3 \to X_3$$



Which dynamical systems can be implemented with mass-action CRNs?

Some ODEs cannot be directly implemented

Example (linear oscillator): $\dot{x} = -y$, $\dot{y} = x$



Problems:
(1) Chemical concentrations cannot be negative.
(2) A species must appear as a reactant to be consumed. Thus the rate of its consumption must be proportional to its concentration.

Solution with representation change: $x = [Xp]-[Xn]$ $y=[Yp]-[Yn]$

$$Xp + Xn \xrightarrow{\text{fast}} \varnothing$$

$$\left.\begin{array}{l} Yp \xrightarrow{1} Xn + Yp \\ Yn \xrightarrow{1} Xp + Yn \end{array}\right\} \quad \dot{x} = [\dot{Xp}]-[\dot{Xn}] = [Yn]-[Yp] = -y$$

$$Xp \xrightarrow{1} Yp + Xp$$
$$Xn \xrightarrow{1} Yn + Xn$$

Kevin Oishi and Eric Klavins, "A biomolecular implementation of linear I/O systems", *IET Systems Biology*, 5: 252–260 (2011)

Also: see Kevin Oishi's poster at DNA17

Which dynamical systems can be implemented with mass-action CRNs?

Korzuhin's theorem:

It is always possible to construct a CRN in which the concentration of some species coincides with any desired accuracy for any desired period of time with the behavior of a given system of polynomial ODEs with non-negative integer powers.

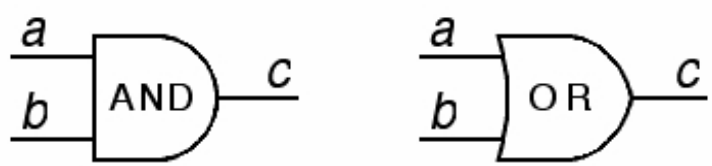
(Applies only to the behavior of the ODEs in the positive orthant.)

Constructed CRN has the following properties
-conservation of mass
-reactions of very simple form: at most two reactants and at most two products
-no autocatalysis

Korzuhin, Oscillatory Processes in Biological and Chemical Systems (Nauka, Moscow), (in Russian), pp 231–251 (1967)

Klonowski, "Simplifying principles for chemical and enzyme reaction kinetics," *Biophys Chem* 18:73–87 (1983)

Logic circuits with mass-action CRNs



$A + B \rightarrow C$

$A \rightarrow C$

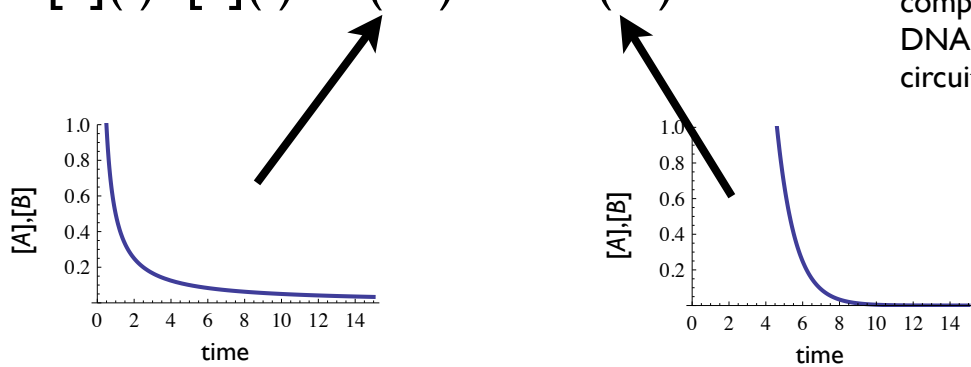
$B \rightarrow C$

Problems:

-no signal restoration

-not dynamic or reusable: can't change input values to get new output

-slow: for a single gate, if [*A*]=[*B*] then [*A*](t)=[*B*](t)=O(1/t) not O(e^{-t})



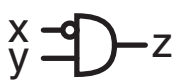
Seelig and Soloveichik, "Time complexity of multilayered DNA strand displacement circuits" in DNA 15

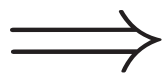
Dynamic logic circuits with mass-action CRNs

x	y	z
0	0	0
0	1	1
1	0	0
1	1	0

dual rail representation:

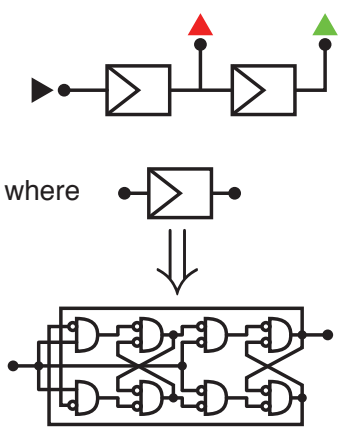
species x(0)	species x(1)	value x
high	low	0
low	high	1

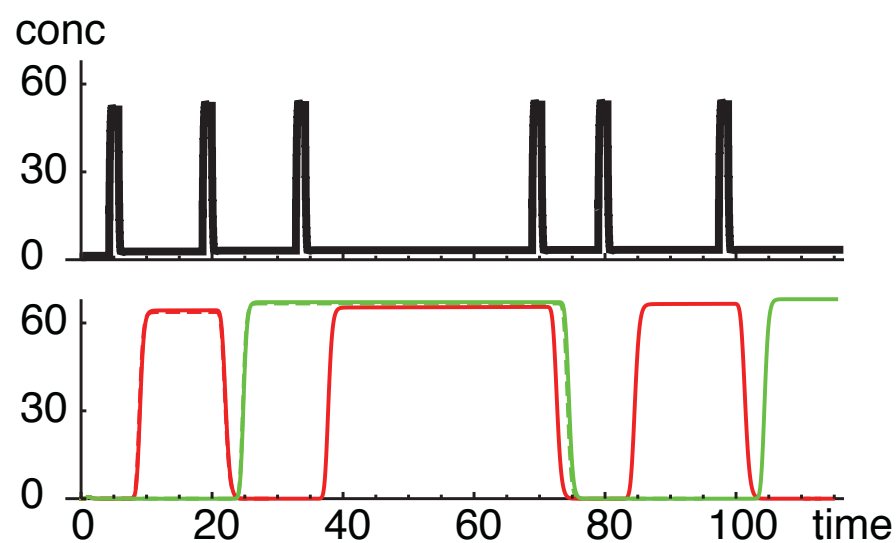




logic	thresholding
$x(0)+y(1)+w(0) \longrightarrow x(0)+y(1)+w(1)$	$w(0)+w(0)+z(1) \longrightarrow w(0)+w(0)+z(0)$
$x(1)+w(1) \longrightarrow x(1)+w(0)$	$w(1)+w(1)+z(0) \longrightarrow w(1)+w(1)+z(1)$
$y(0)+w(1) \longrightarrow y(0)+w(0)$	

2-bit pulse counter (digital circuit)



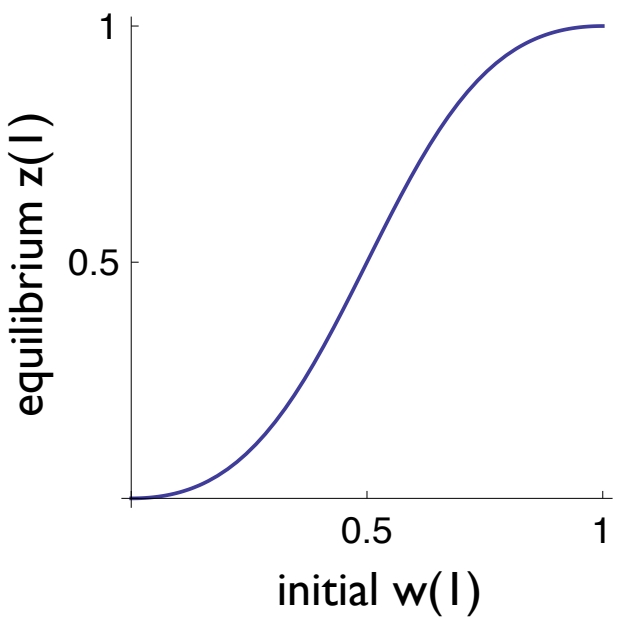


Soloveichik, Seelig, Winfree,"DNA as a universal substrate for chemical kinetics", *PNAS*, 107: 5393-5398 (2010)

Dynamic logic circuits with mass-action CRNs: signal restoration

Cooperativity

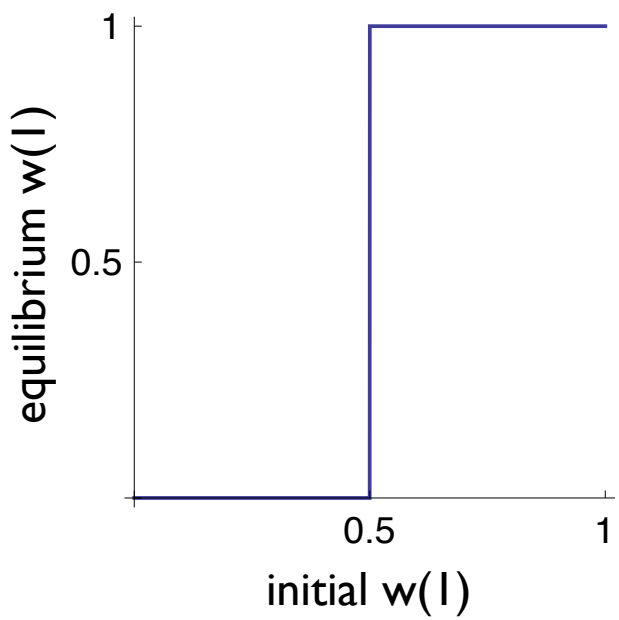
$$w(0)+w(0)+z(1) \xrightarrow{1} w(0)+w(0)+z(0)$$
$$w(1)+w(1)+z(0) \xrightarrow{1} w(1)+w(1)+z(1)$$



Magnasco, "Chemical kinetics is Turing universal", *Physical Review Letters* 78: 1190-1193 (1997)

Majority Algorithm

$$w(0)+w(1) \xrightarrow{1} sw$$
$$sw+w(0) \xrightarrow{1} 3\ w(0)$$
$$sw+w(1) \xrightarrow{1} 3\ w(1)$$



Jiang, Riedel and Parhi, "Digital Logic with Molecular Reactions," submitted.

Artificial Biochemistry

Luca Cardelli

Microsoft Research

Abstract

We model chemical and biochemical systems as collectives of interacting stochastic automata, with each automaton representing a molecule that undergoes state transitions. In this *artificial biochemistry*, automata interact by the equivalent of the law of mass action. We investigate several simple but intriguing automata collectives by stochastic simulation and by ODE analysis.

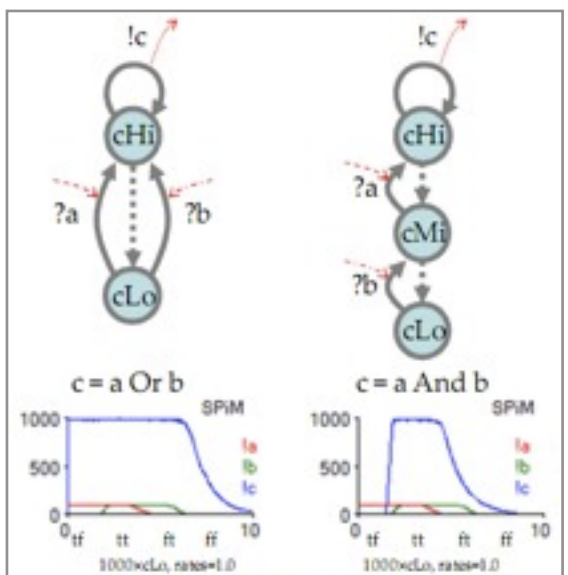


Figure 32 Or and And

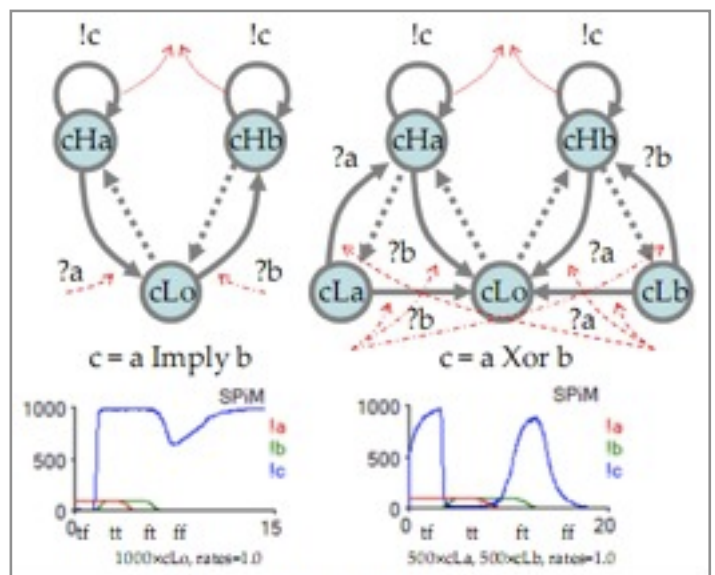


Figure 33 Imply and Xor

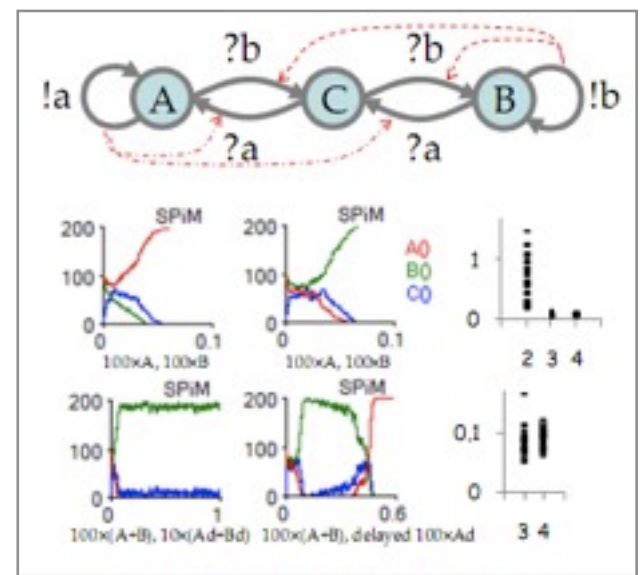


Figure 34 Memory Elements

Luca Cardelli, "Artificial Biochemistry", in Algorithmic Bioprocesses. Springer (2009)
(http://lucacardelli.name/Papers/Artificial%20Biochemistry.pdf)

Part 2 of 2:

How To Discipline Your DNA Molecules with Strand Displacement Cascades

Strand Displacement Cascades is a flexible technology for implementing complex nucleic-acid reaction networks in the laboratory

Engineering artificial signaling networks

Long-term goals:

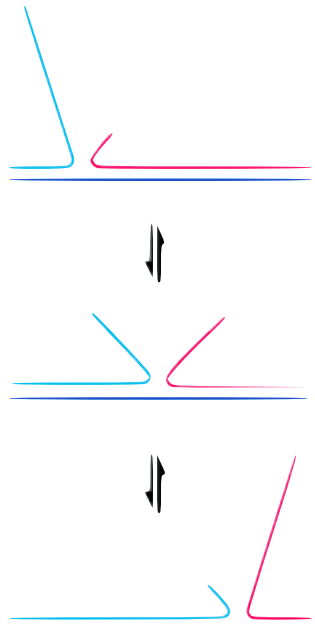
-Insert desired control module into cells? Medical applications? Smart drugs?

-Abiological systems: control modules for nanomotors, self-assembly, polymerization, other kind of chemistries? "Wet robot"?

-Develop clarity of thought for understanding biological signaling networks

Approach: No "alien" technology: only what we can understand and build

The strand displacement reaction



enzyme-free strand displacement aka branch migration

Green, C & Tibbetts, C. (1981) Nucleic Acids Research 9, 1905

Weinstock, P & Wetmur, J. (1990) Nucleic Acids Research 18, 4207

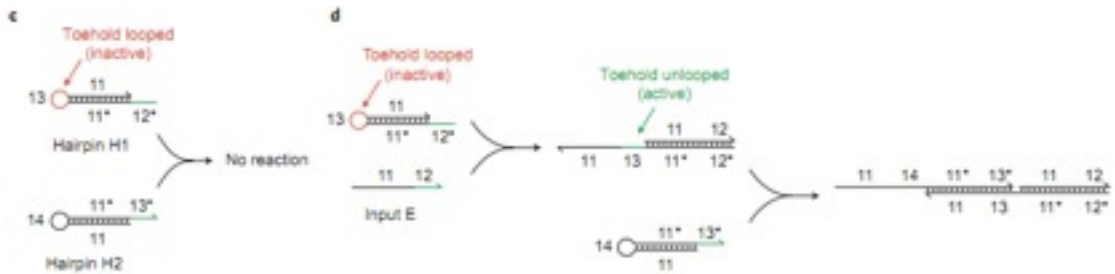
Panyutin, I & Hsieh, P. (1993) Journal of molecular biology 230, 413

...

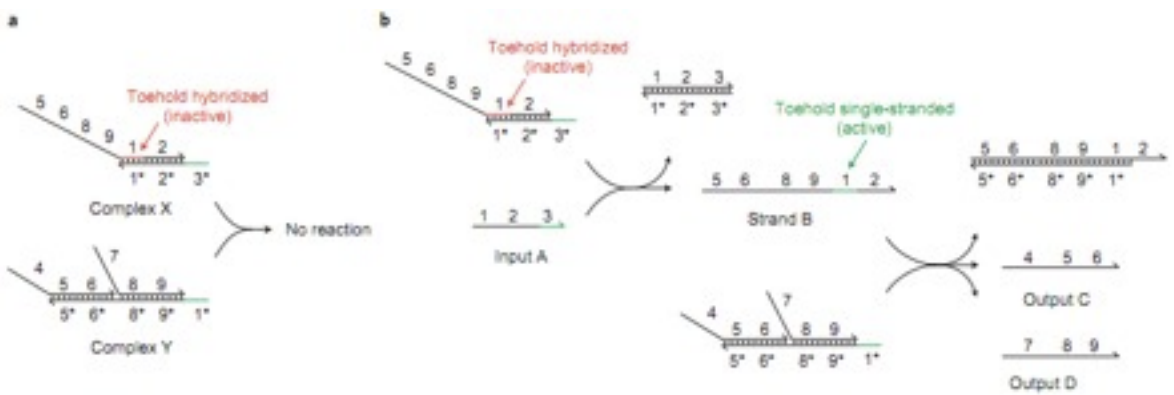
first systematic use in DNA nanotechnology

Yurke, B & Mills, A. P. (2003) Genetic Programming and Evolvable Machines 4, 111

Cascading of strand displacement reactions



Dirks, Pierce, “Triggered amplification by hybridization chain reaction,” *PNAS* 101, 15275 (2004).



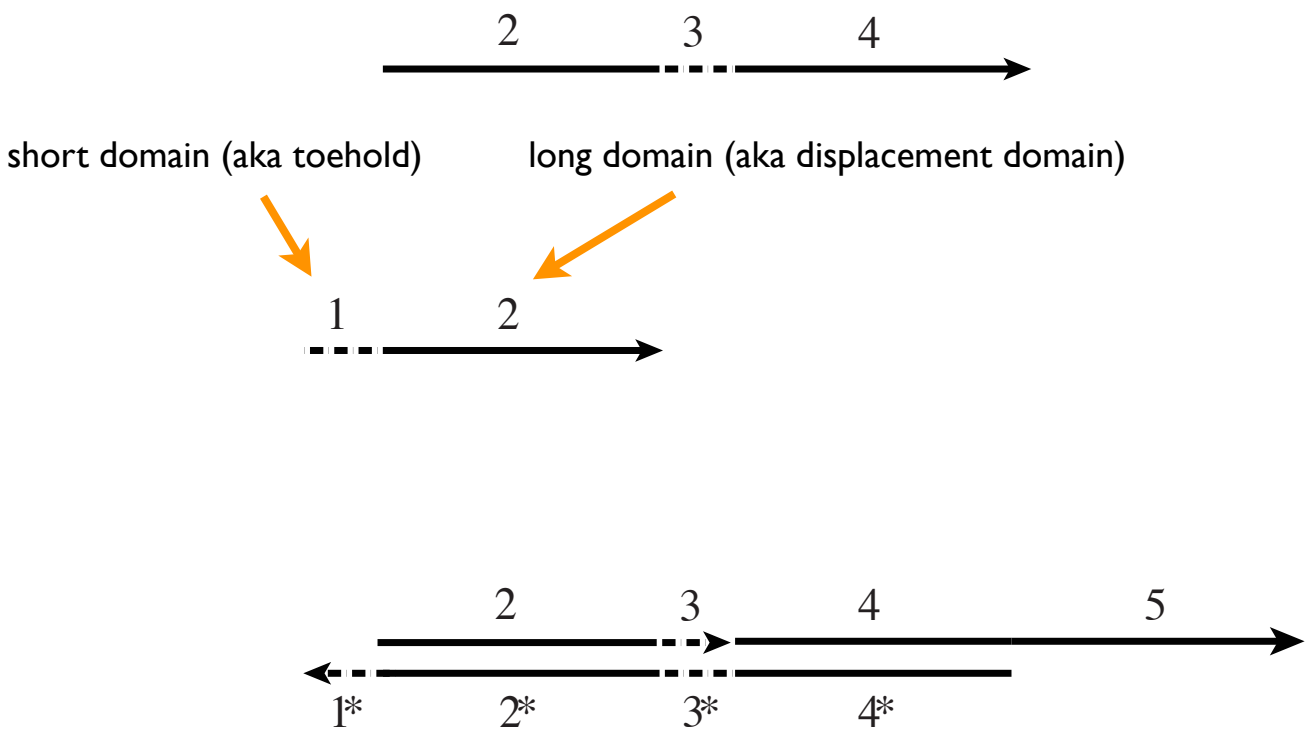
Seelig, Soloveichik, Zhang, Winfree, “Enzyme-free nucleic acid logic circuits,” *Science* 314, 1585 (2006).

Diagrams from: Zhang, Seelig, “Dynamic DNA nanotechnology using strand-displacement reactions”, *Nature Chemistry* 3, 103 (2011).

Strand Displacement Cascades

possible moves:

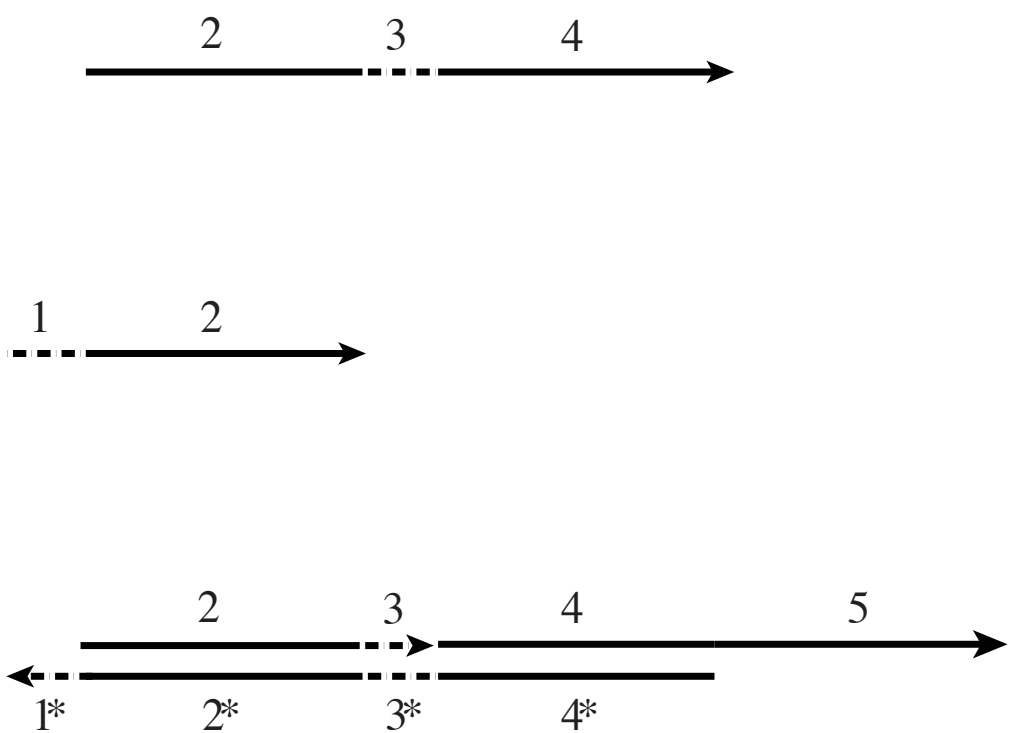
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

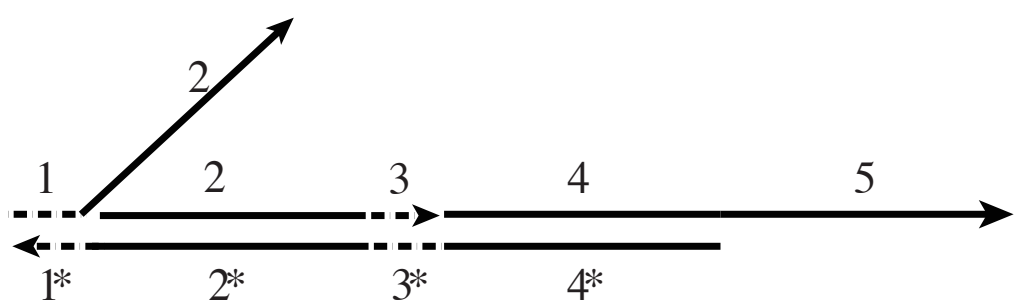
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

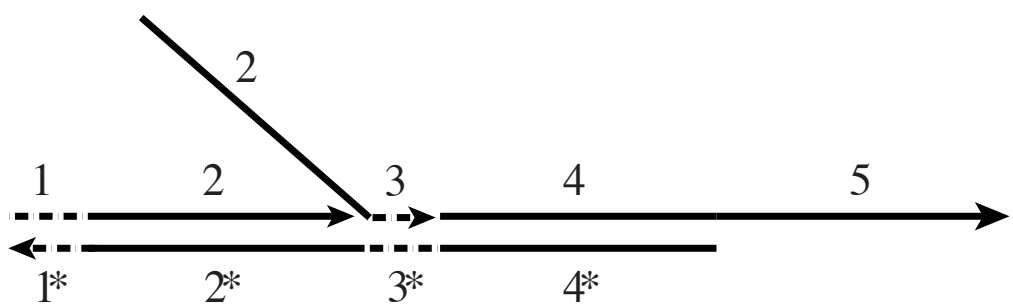
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

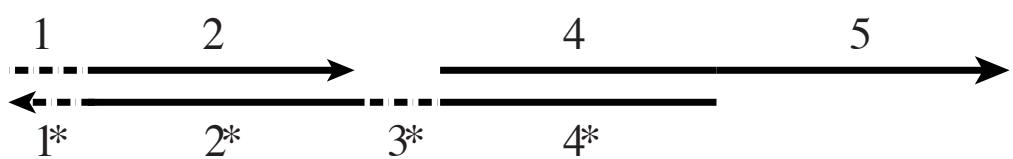
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

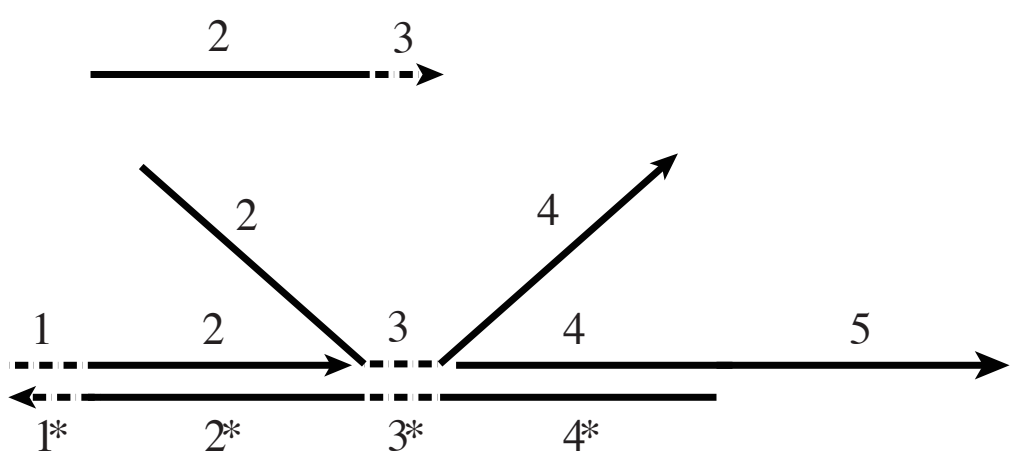
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

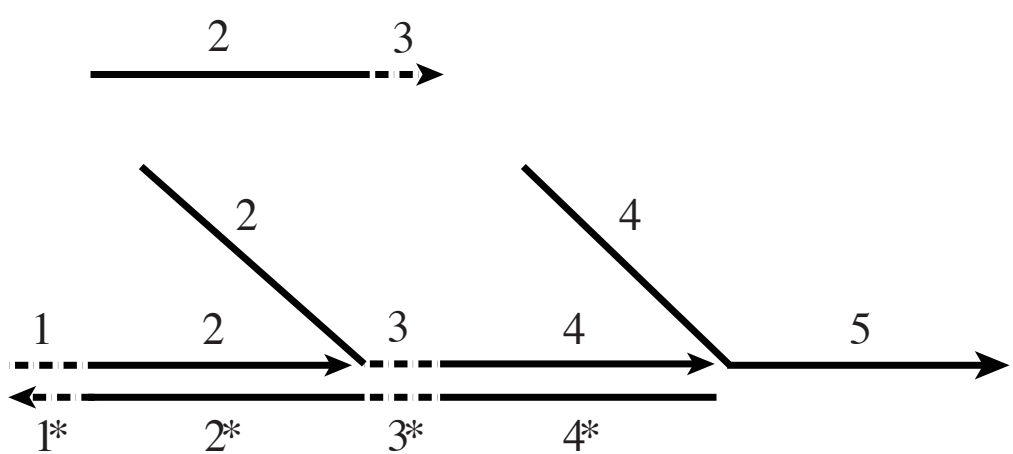
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

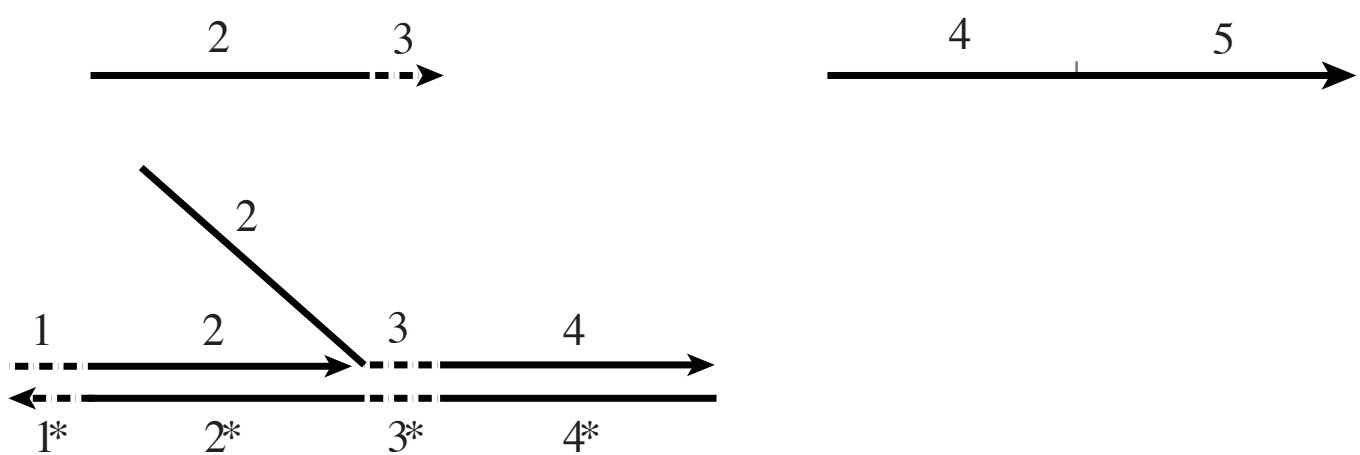
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

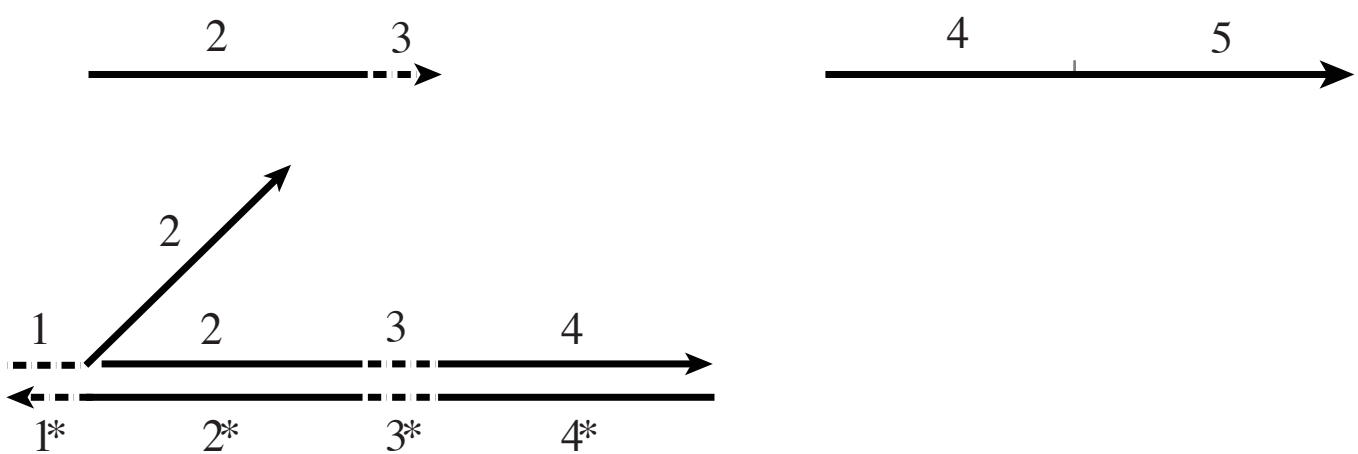
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



Strand Displacement Cascades

possible moves:

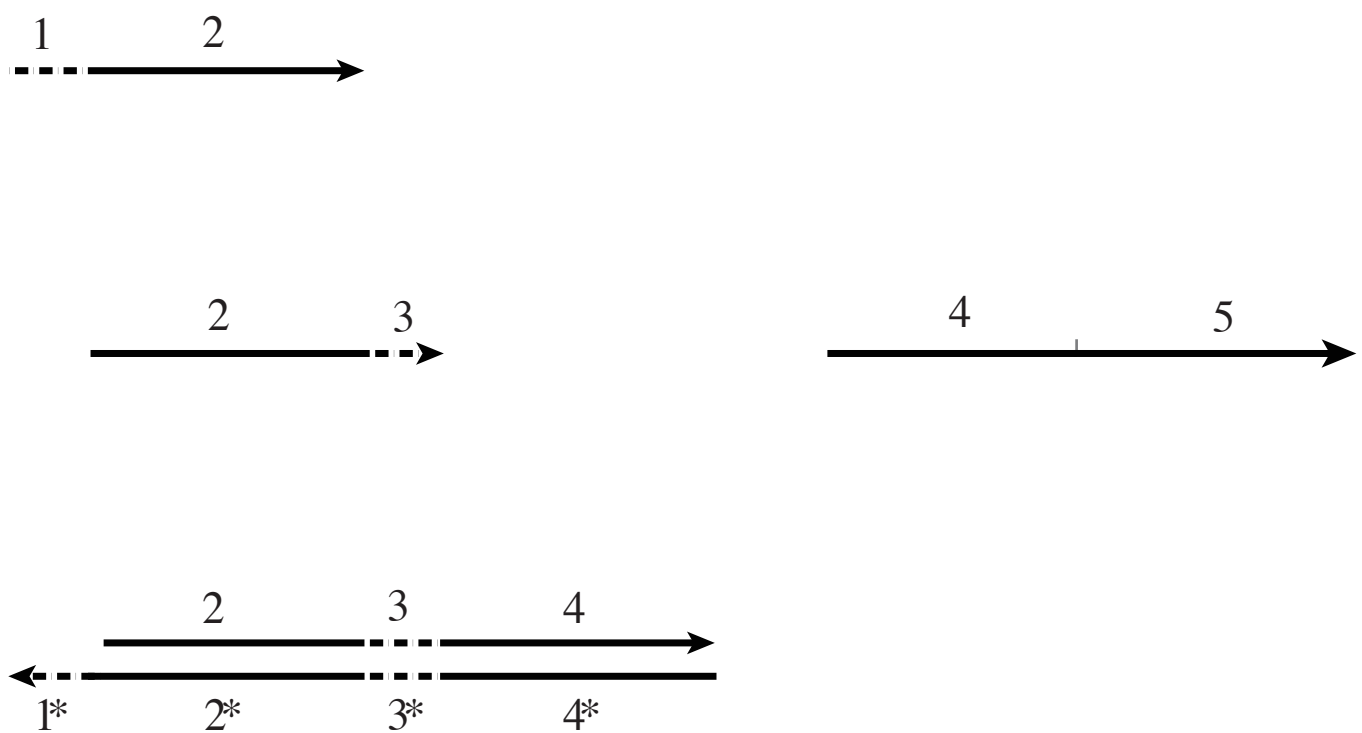
- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization



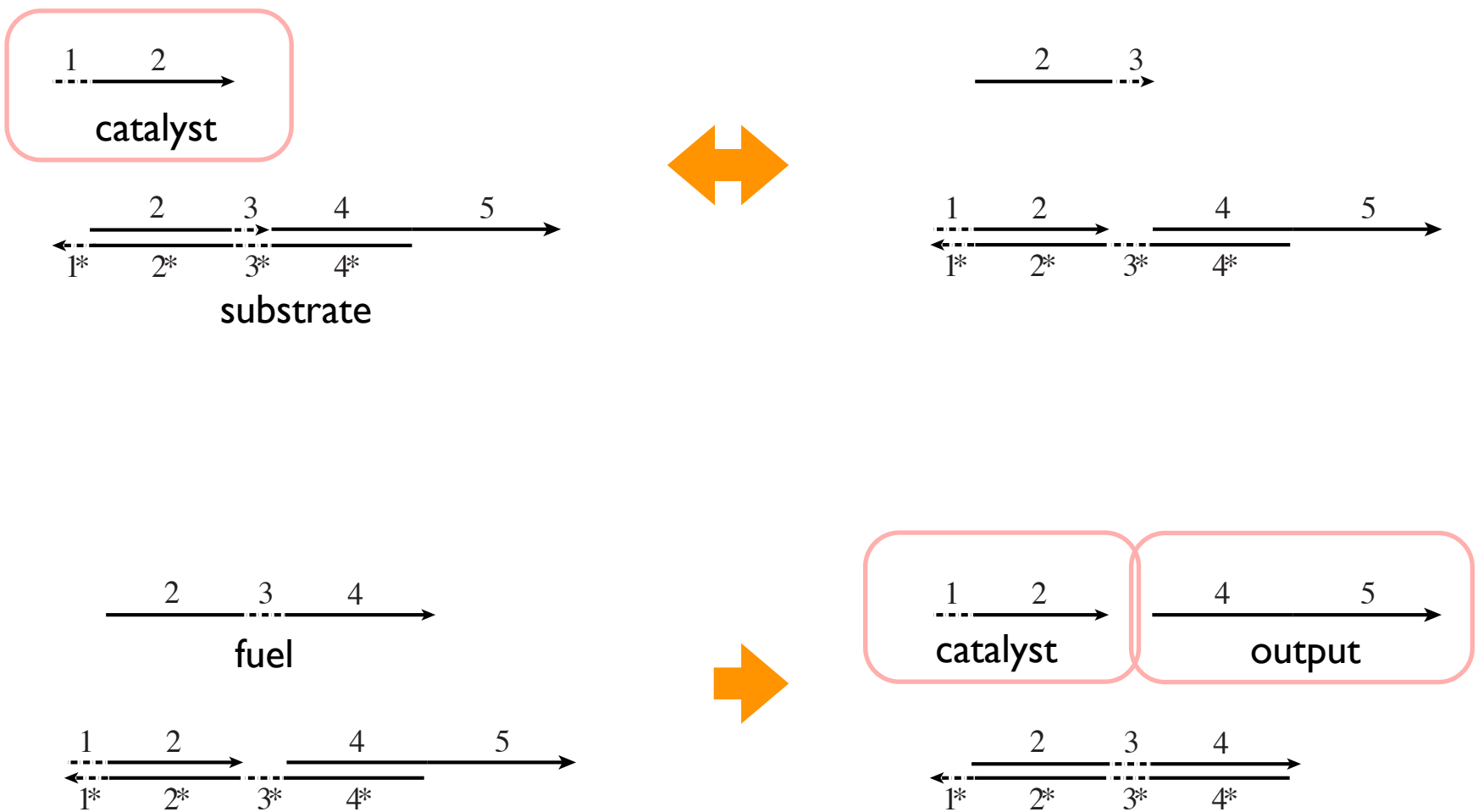
Strand Displacement Cascades

possible moves:

- **Bind** two complementary domains
- **Release** any strand held by only a short domain
- **Displace** a domain by an identical domain if this extends existing hybridization

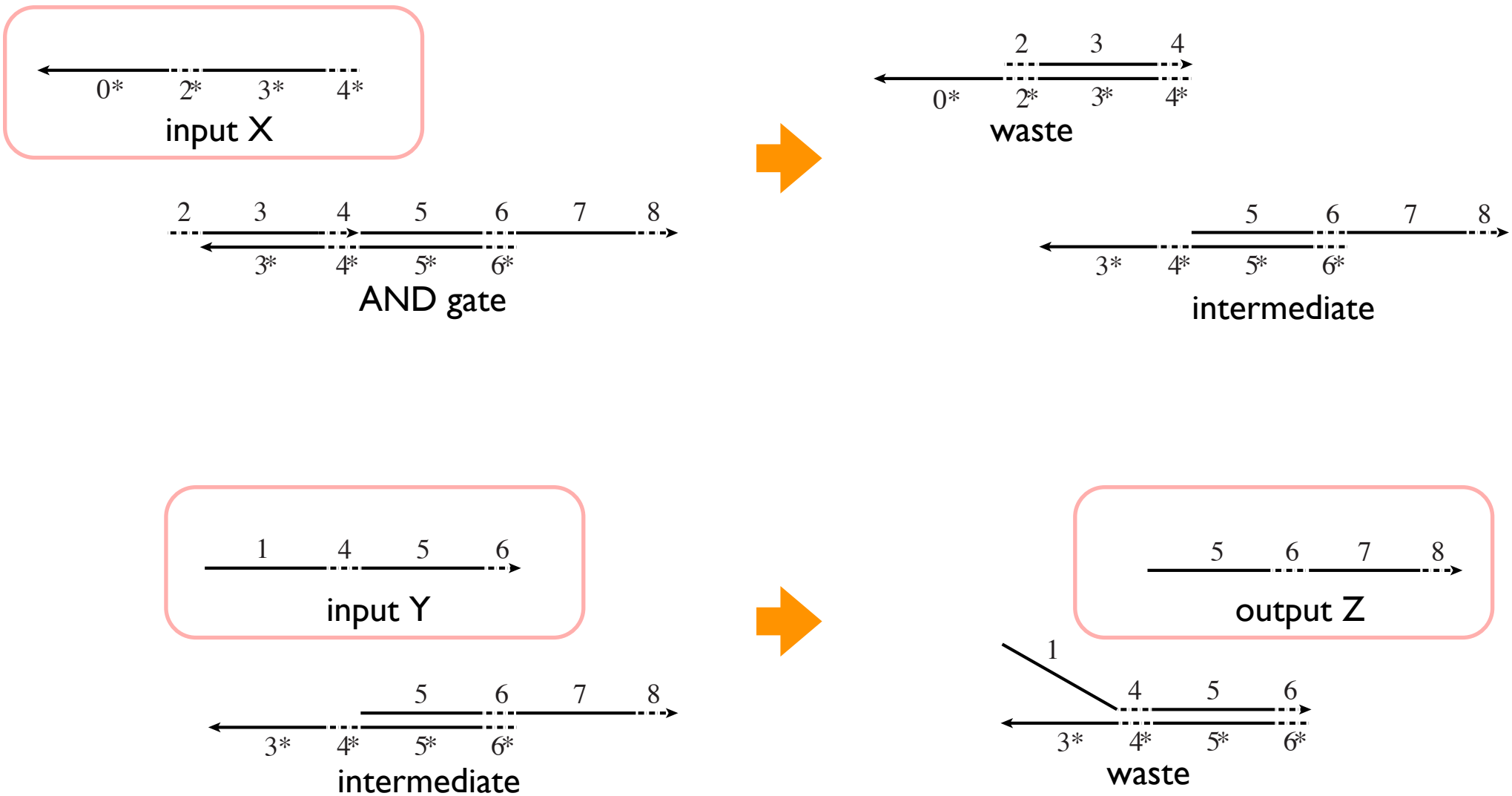


Strand Displacement Cascades
example: Catalyst



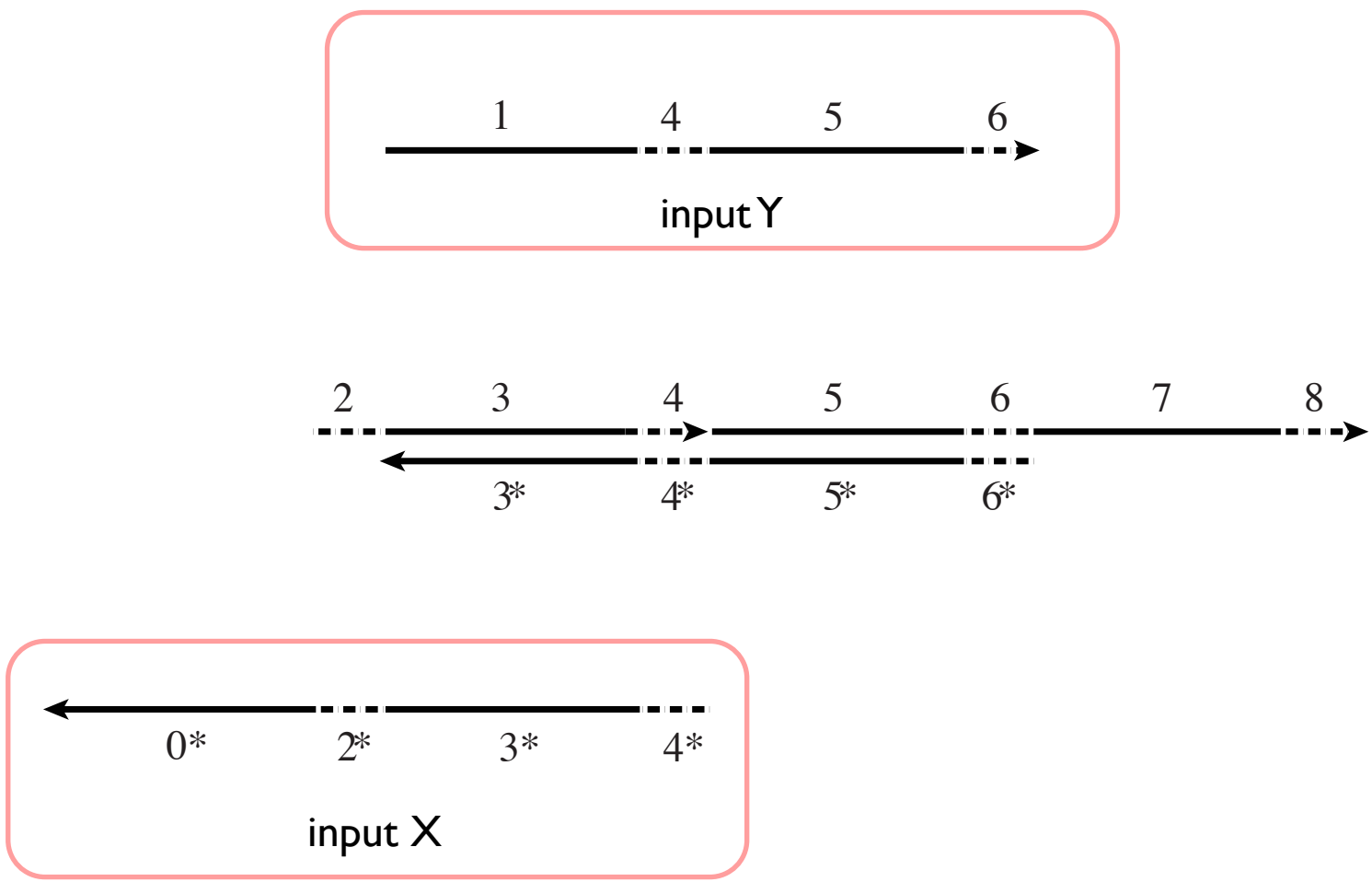
Zhang, Turberfield, Yurke, Winfree, *Science* 318: 1121-1125, 2007

AND Logic Gate



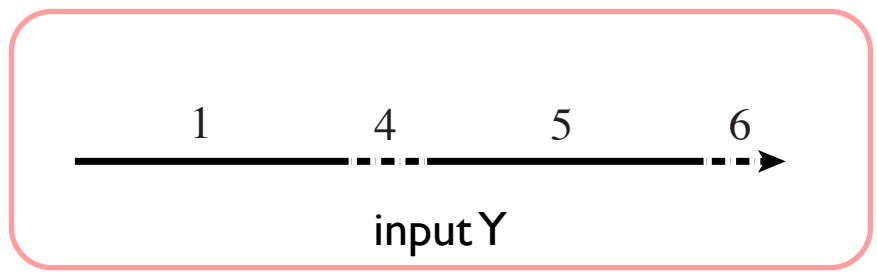
based on Seelig, Soloveichik, Zhang, Winfree, *Science* 314: 1585-1587 (2006)

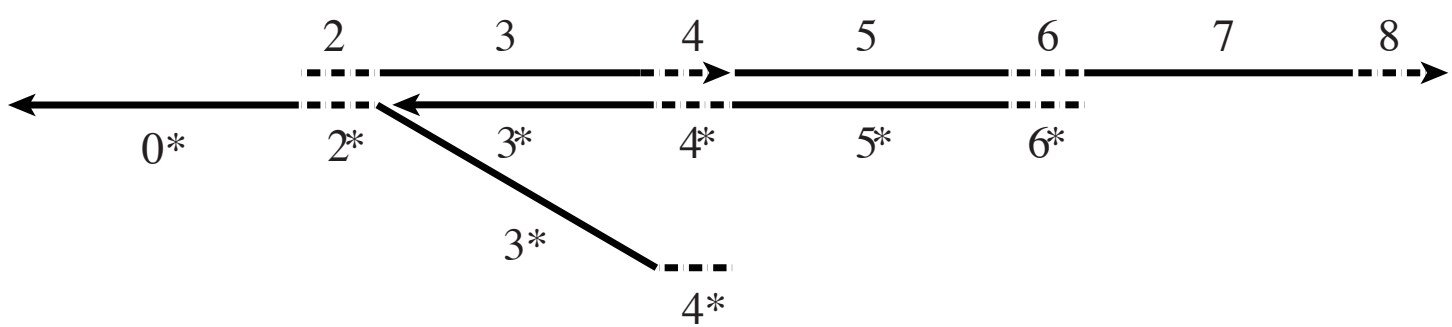
AND Logic Gate



based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

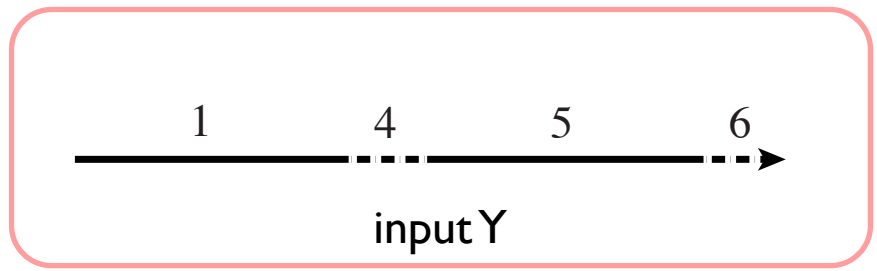
AND Logic Gate

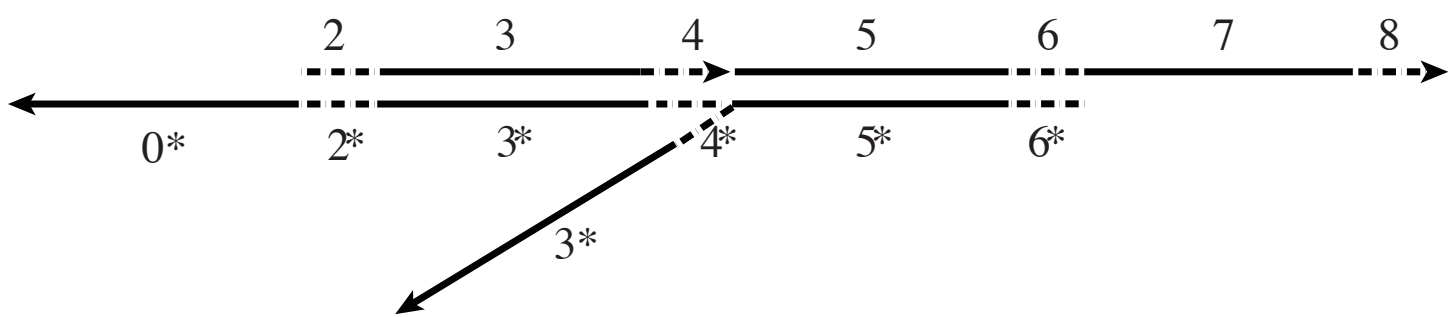




based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

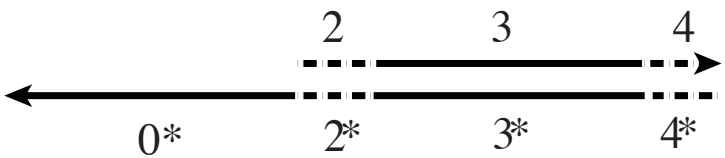
AND Logic Gate





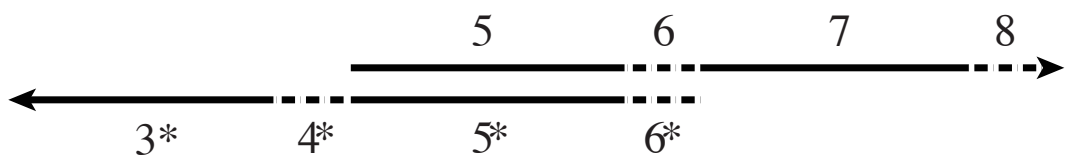
based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

AND Logic Gate



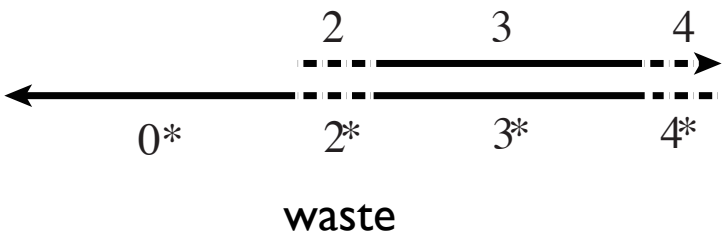
waste

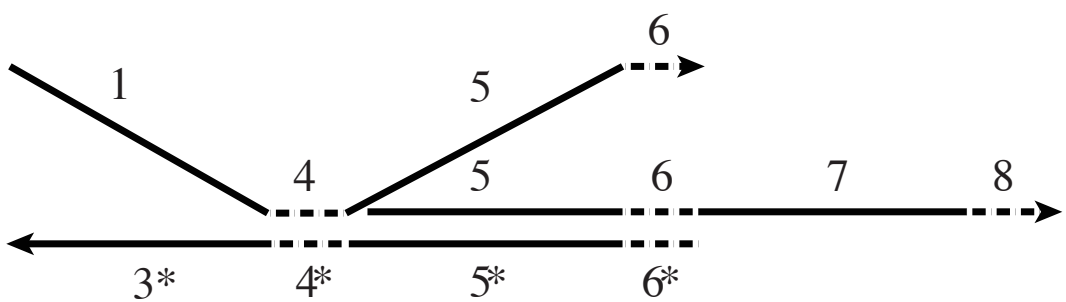
input Y



based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

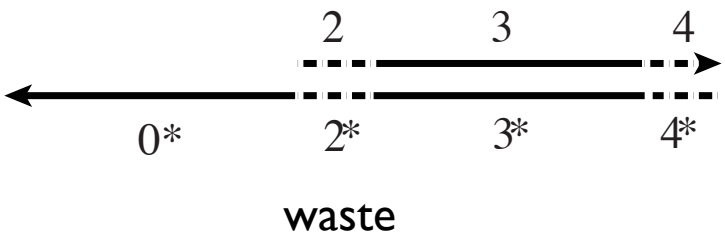
AND Logic Gate

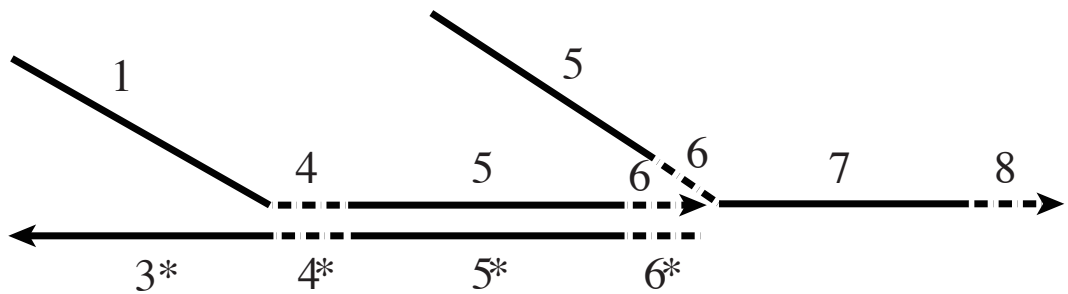




based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

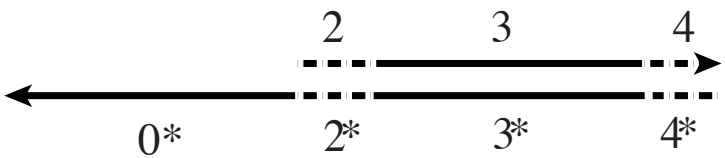
AND Logic Gate





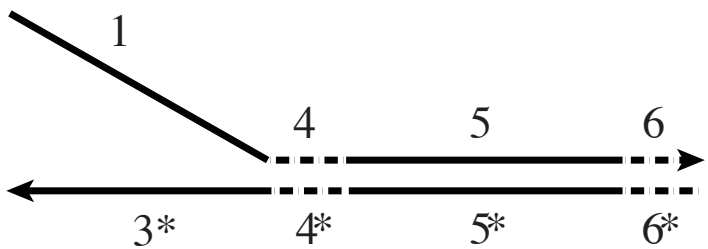
based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

AND Logic Gate



waste

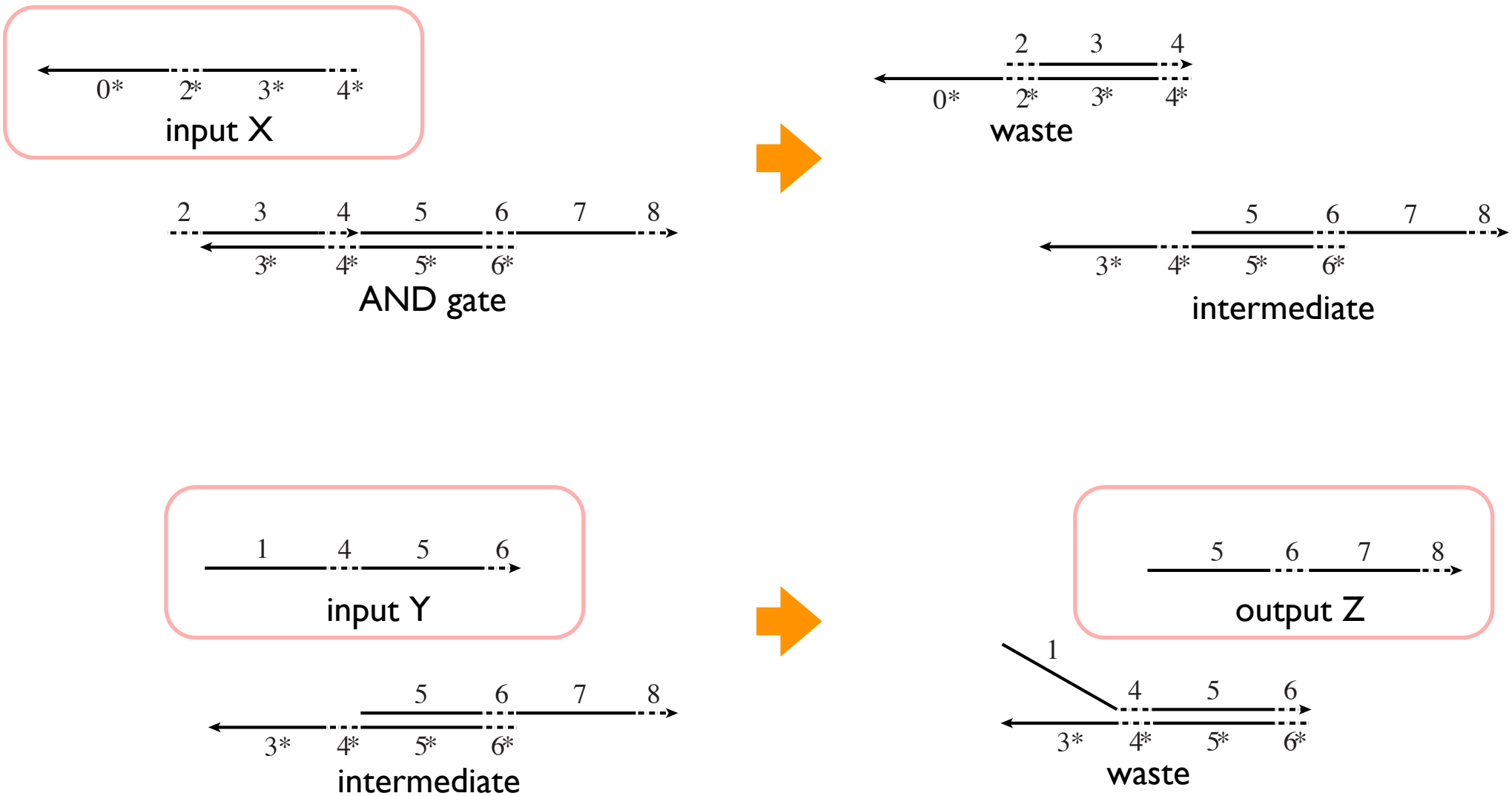
output Z



waste

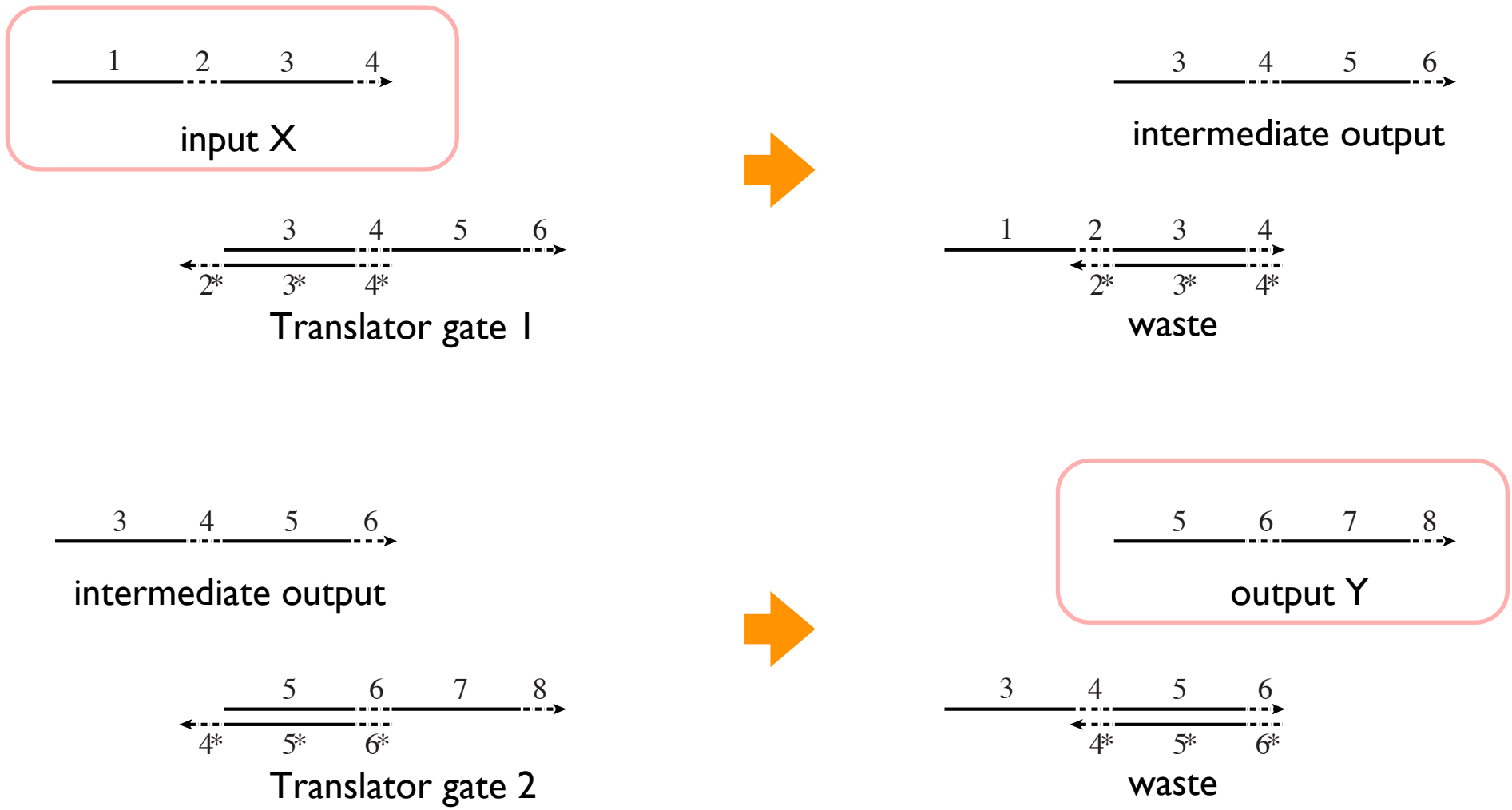
based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

AND Logic Gate



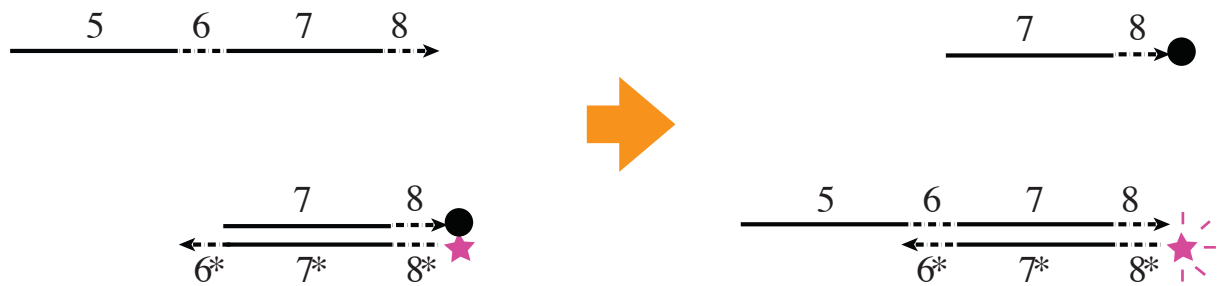
based on Seelig, Soloveichik, Zhang, Winfree, *Science* 314: 1585-1587 (2006)

Translator Gates: complete sequence independence

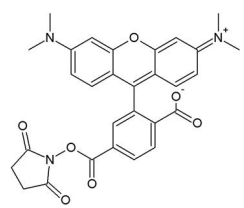


based on Seelig, Soloveichik, Zhang, Winfree, **Science** 314: 1585-1587, 2006

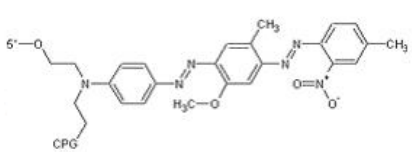
Experimental technique: Fluorescent readout



Fluorophore	Replaces	Dye-5′-T_{10} EX	EM	BHQ
BioSearch Blue		352	447	BHQ-0 λ_{MAX} 493 nm QR 430-520 nm
Acridine		362	462	
Coumarin		432	472	
FAM		495	520	
Rhodamine Green		503	528	
TET		521	536	BHQ-1 λ_{MAX} 534 nm QR 480-580 nm
CAL Fluor Gold 540	VIC/TET/JOE	522	544	
JOE		529	555	
VIC		538	554	
HEX		535	554	
CAL Fluor Orange 560	VIC/HEX/JOE	538	559	
Quasar 570	Cy3	548	566	
TAMRA		557	583	
Rhodamine Red		560	580	
CAL Fluor Red 590	TAMRA	569	591	BHQ-2 λ_{MAX} 579 nm QR 560-670 nm
Cy3.5		581	596	
ROX		575	602	
CAL Fluor Red 610	Texas Red	590	610	
CAL Fluor Red 635	LC Red 640	618	637	
Pulsar 650		460	650	BHQ - 3 λ_{MAX} 672 nm QR 620-730 nm
Quasar 670	Cy5	647	667	
Quasar 705	Cy5.5	690	705	

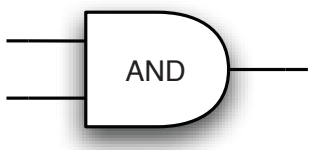


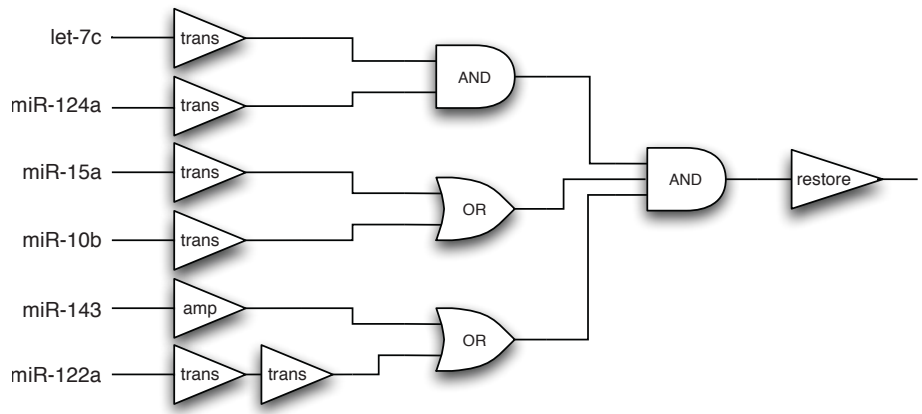
TAMRA:

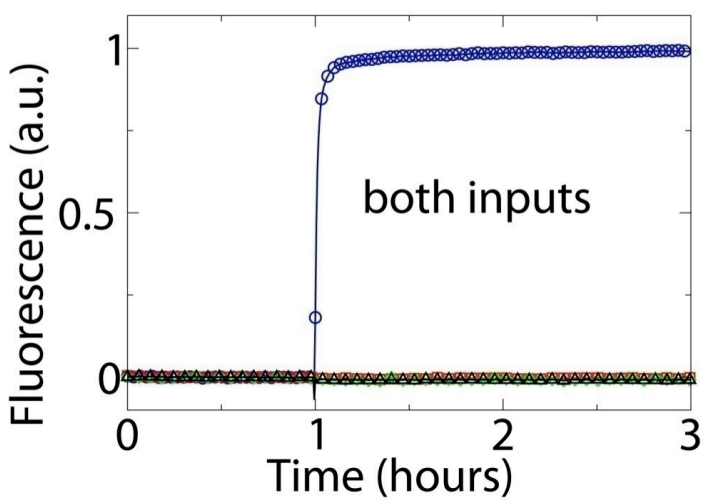


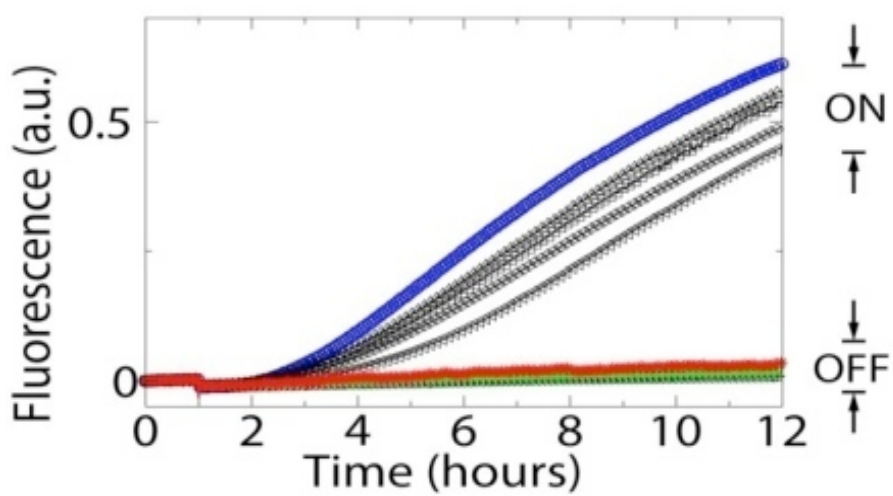
BHQ-1:

Experimental Data for One AND Gate and an 11-gate Logic Circuit



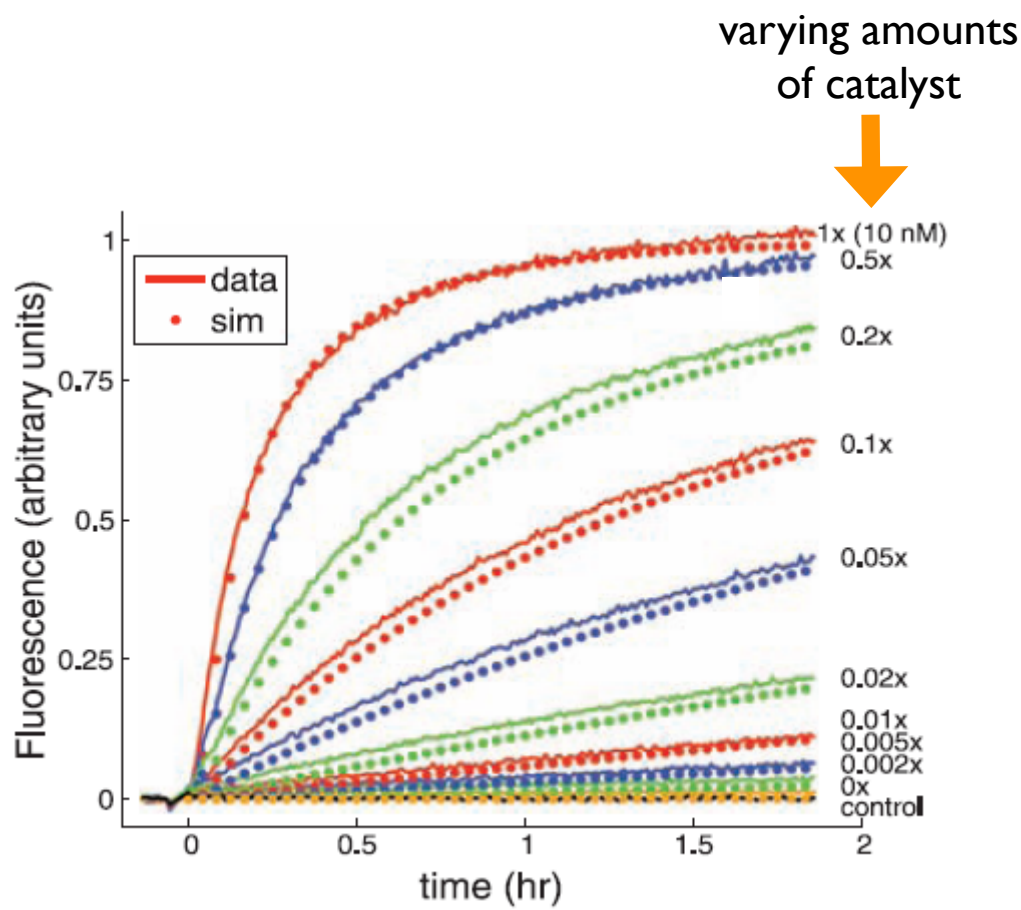






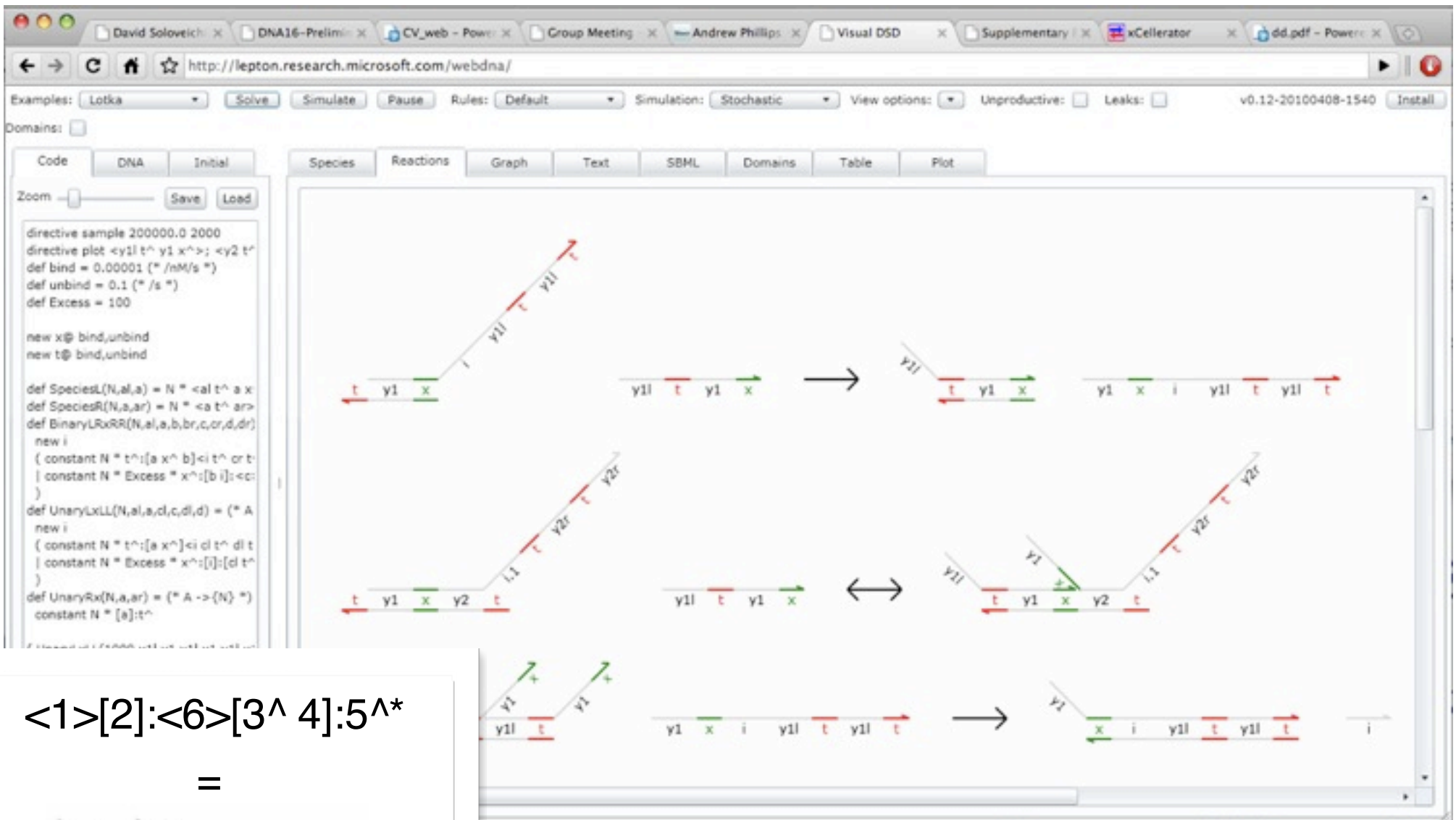
Seelig, Soloveichik, Zhang, Winfree, *Science* 314: 1585-1587, 2006

Experimental Data for Catalyst



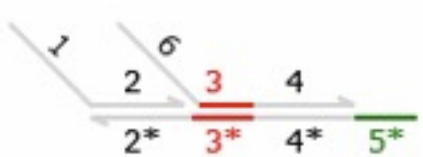
Zhang, Turberfield, Yurke, Winfree, *Science* 318: 1121-1125, 2007

VisualDSD: A formal language for describing and modeling strand displacement cascades



<1>[2]:<6>[3^ 4]:5^*

=



Lakin, Phillips, Cardelli at Microsoft Research
http://lepton.research.microsoft.com/webdna/

Rich Snider, Dmitry Danilov and Zoran Popovic
collaboration with Georg Seelig, David Baker

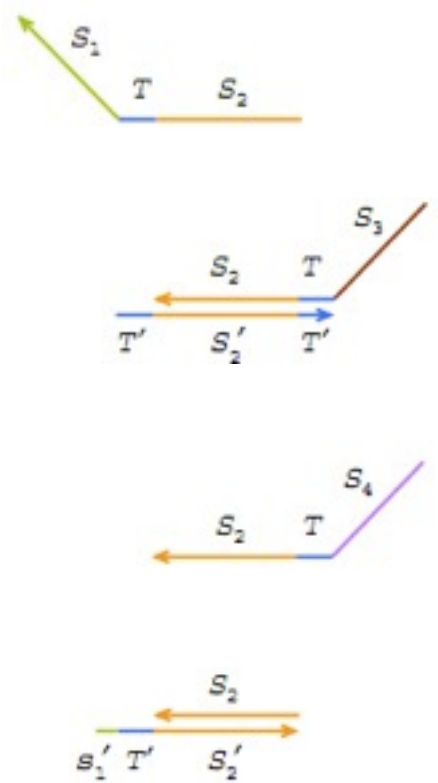
Flash Game Demo

You can play it too:
http://games.cs.washington.edu/DNA_Game/DNA.html

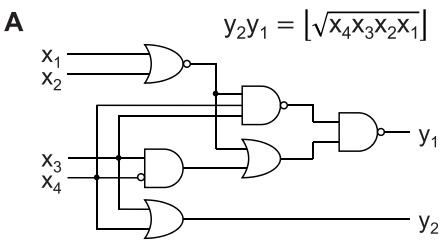
Play the introductory levels to get to the exponential amplifier challenge.
You can submit your solution to the challenge through the game website.

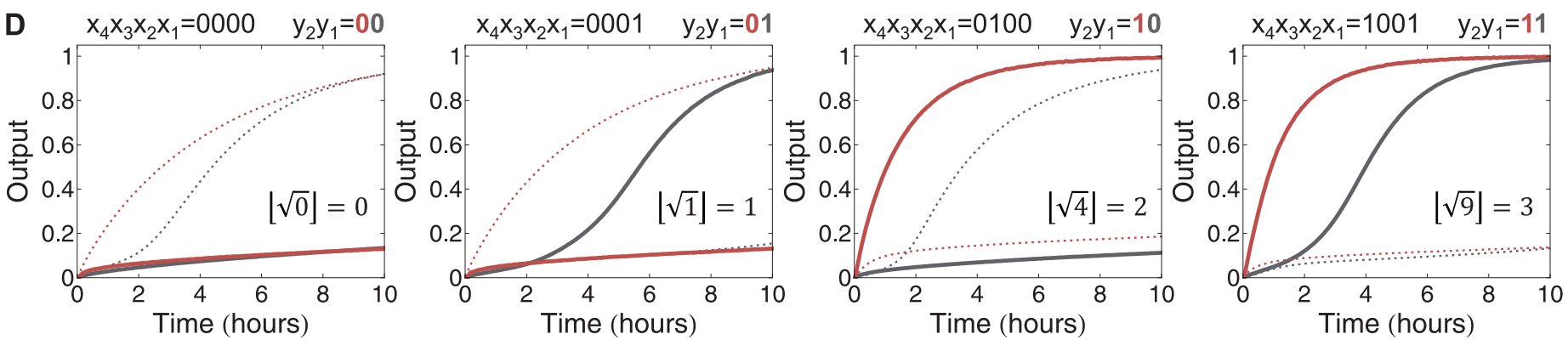
The largest strand displacement cascades implemented in the laboratory used "see-saw" gates

4 simple element types:



Logic circuits with see-saw gates

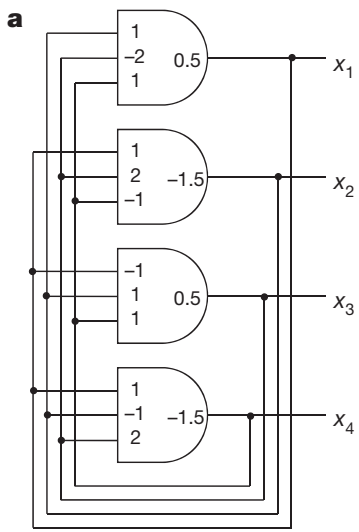




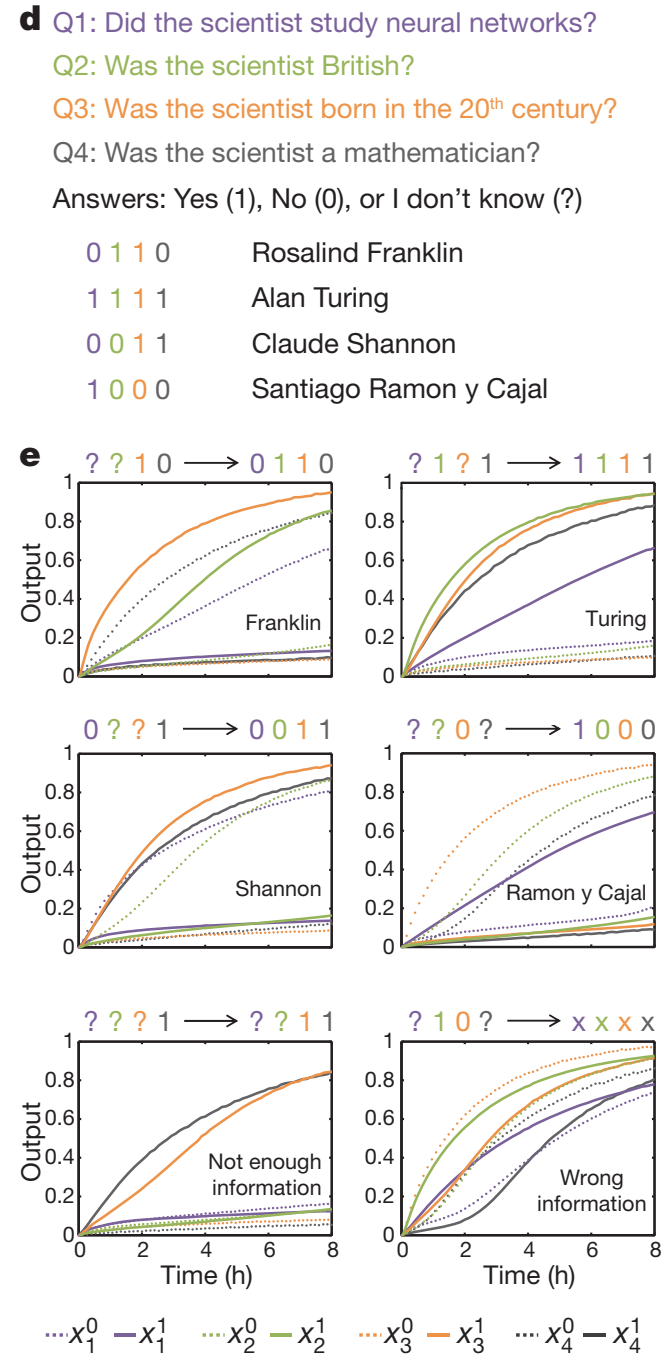
130 DNA strands
74 initial DNA species (excluding inputs)

Qian, Winfree, "Scaling up digital circuit computation with DNA strand displacement cascades," *Science* 332: 1196-1201 (2011)

Neural networks with see-saw gates



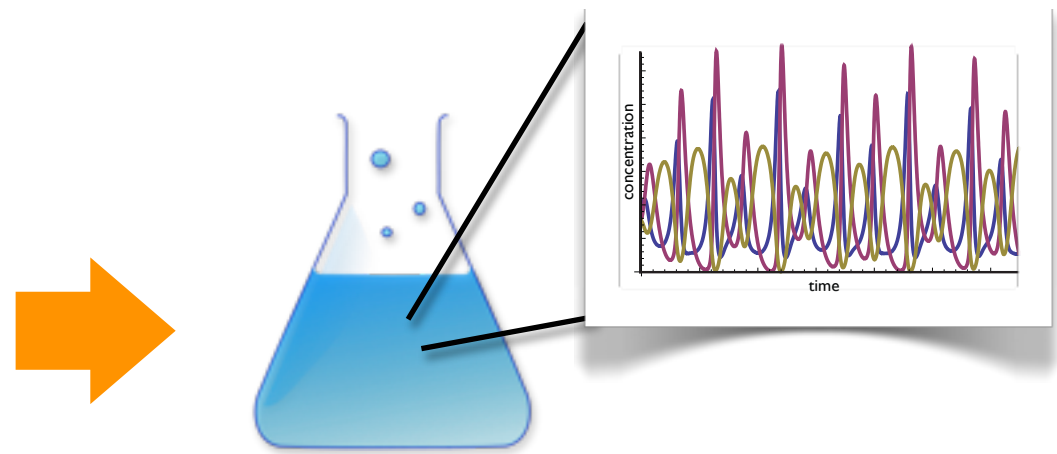
112 DNA strands
72 initial DNA species



Qian, Winfree, Bruck, "Neural network computation with DNA strand displacement cascades," *Nature* 475: 368-372 (2011)

Goal: Be able to take any mass-action CRN and implement it in the test tube

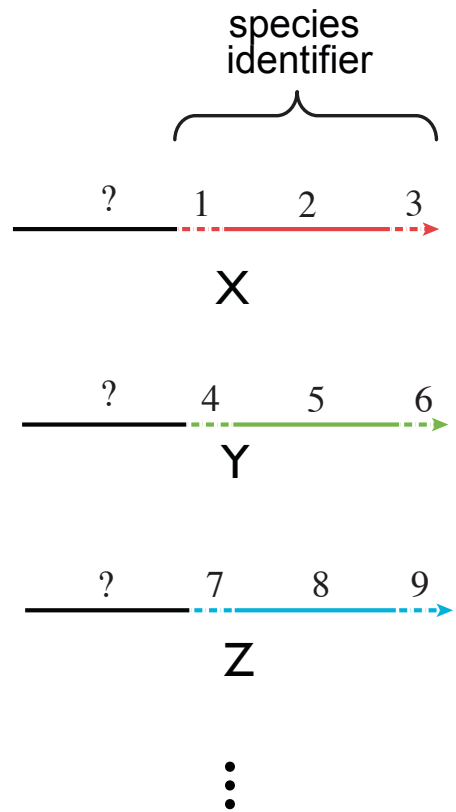
$$
\begin{aligned}
1: &\quad X_1 \xrightarrow{30} 2X_1 \\
2: &\quad 2X_1 \xrightarrow{0.5} X_1 \\
3: &\quad X_2 + X_1 \xrightarrow{1} 2X_2 \\
4: &\quad X_2 \xrightarrow{10} \\
5: &\quad X_1 + X_3 \xrightarrow{1} \\
6: &\quad X_3 \xrightarrow{16.5} 2X_3 \\
7: &\quad 2X_3 \xrightarrow{0.5} X_3
\end{aligned}
$$



- implicit energy/mass source (no conservation restrictions)
- can use auxiliary species to help mediate reactions
- desired behavior up to scaling in time and concentration
- allow degree of approximation: correct behavior in some limit

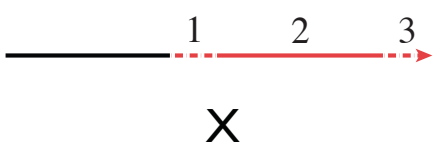
David Soloveichik, Georg Seelig, Erik Winfree, "DNA as a Universal Substrate for Chemical Kinetics", *PNAS* 107: 5393-5398, 2010

Format of formal species



invariant: species is active if species identifier is entirely single-stranded

Unimolecular reaction X→Y



X

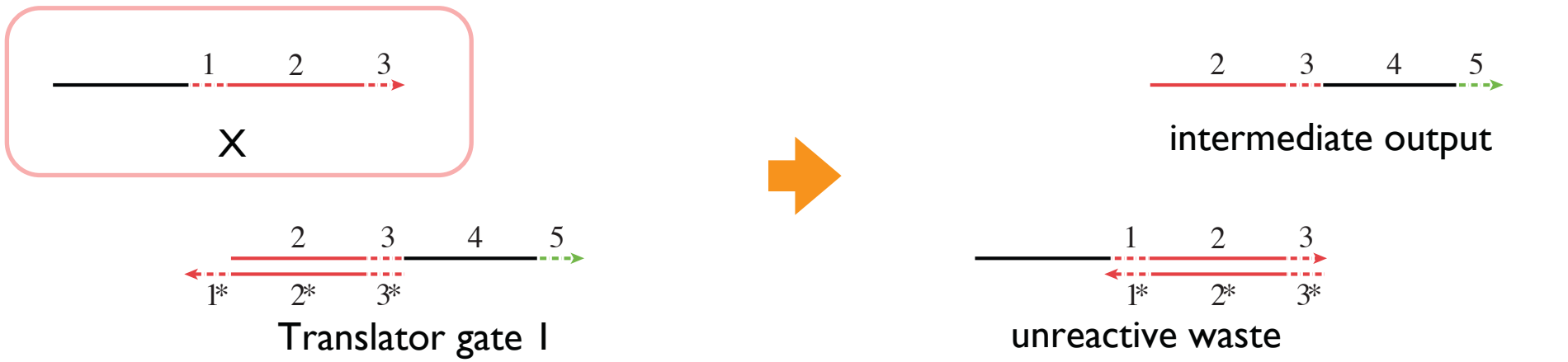
Y

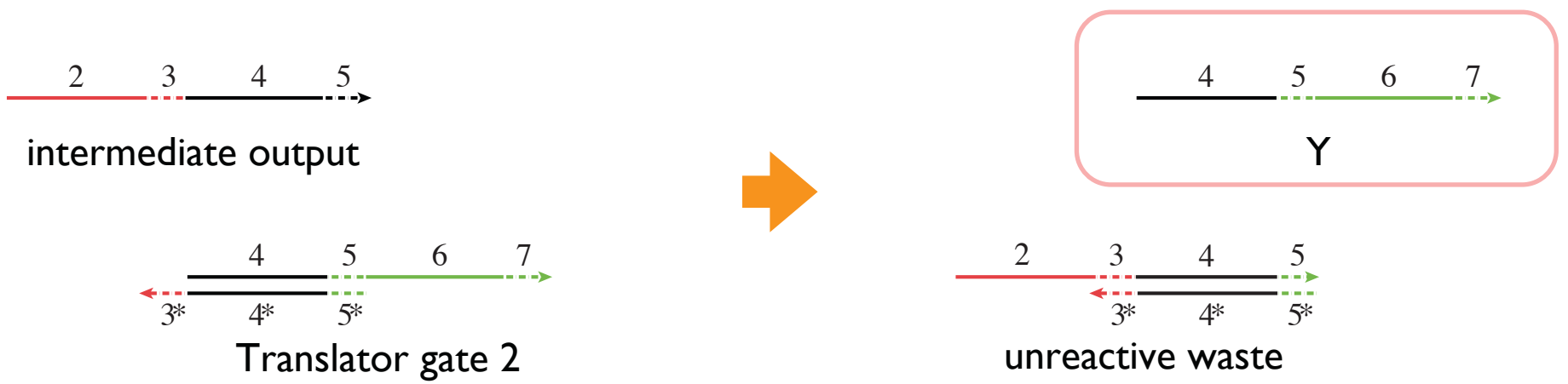
Translator gate 1

Translator gate 2

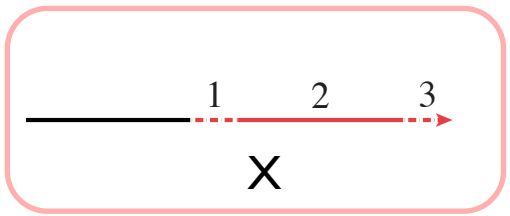
and unreactive waste

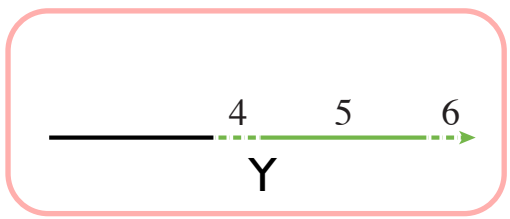
Unimolecular reaction $X \rightarrow Y$

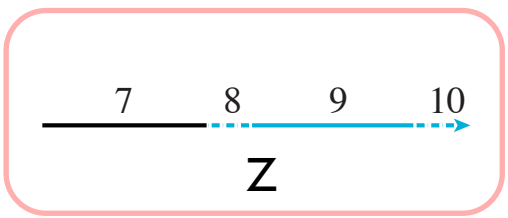


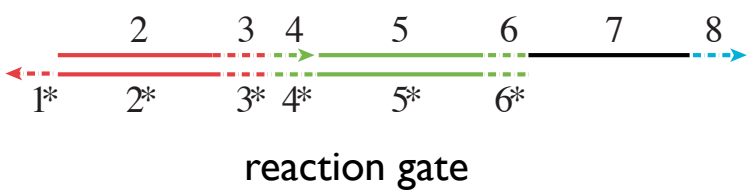


Bimolecular reaction X+Y→Z

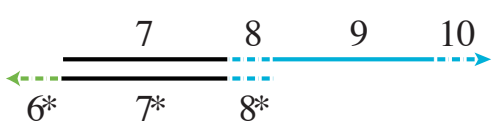








reaction gate

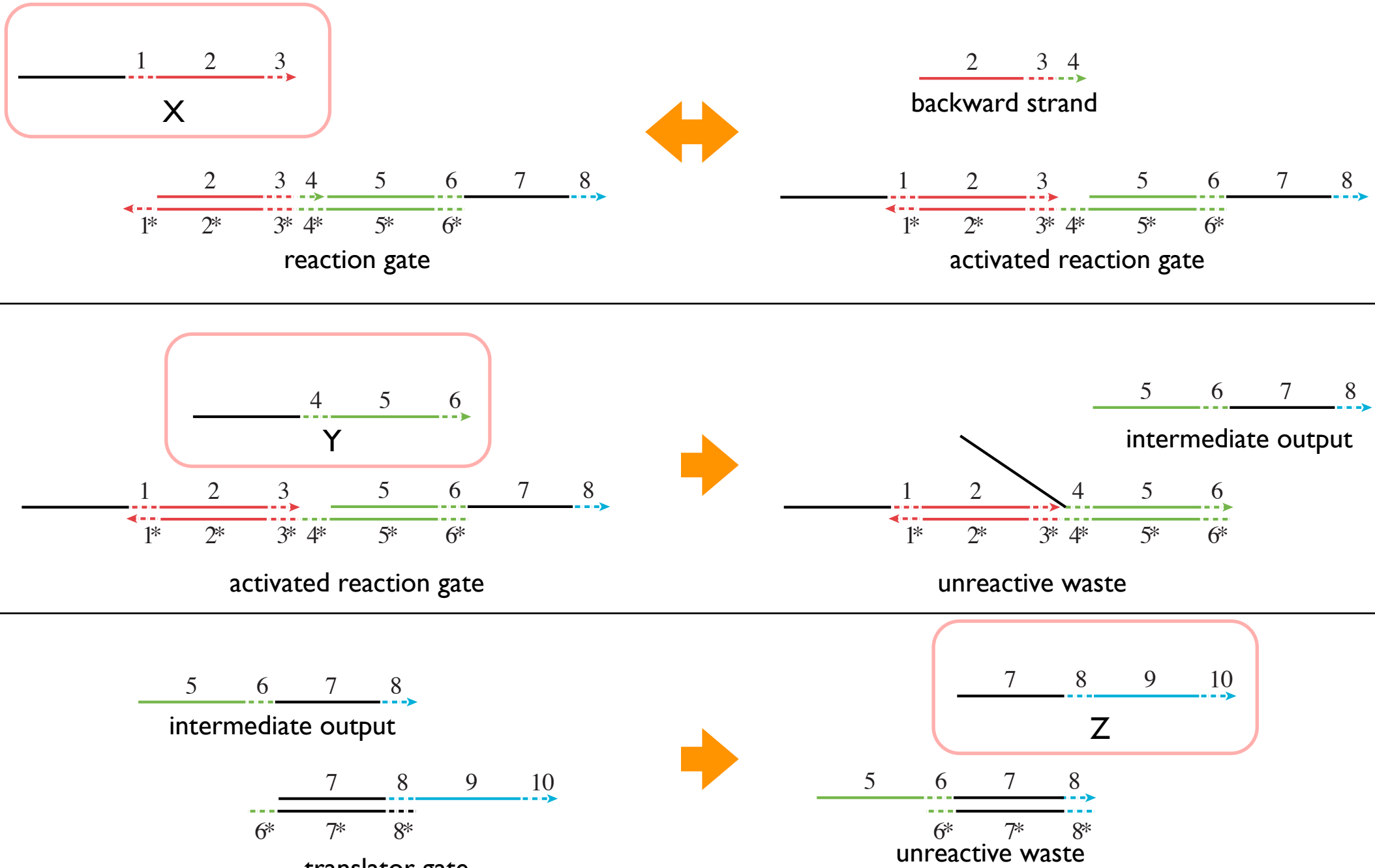


translator gate

backward strand

and unreactive waste

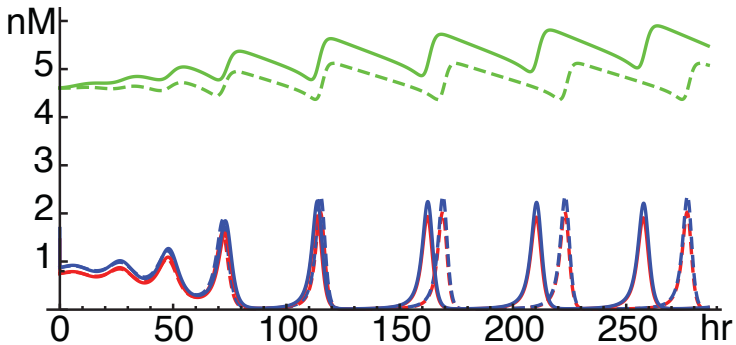
Bimolecular reaction X+Y→Z



Complex self-generated behavior with strand displacement cascades (simulations)

Oregonator (limit cycle oscillator)

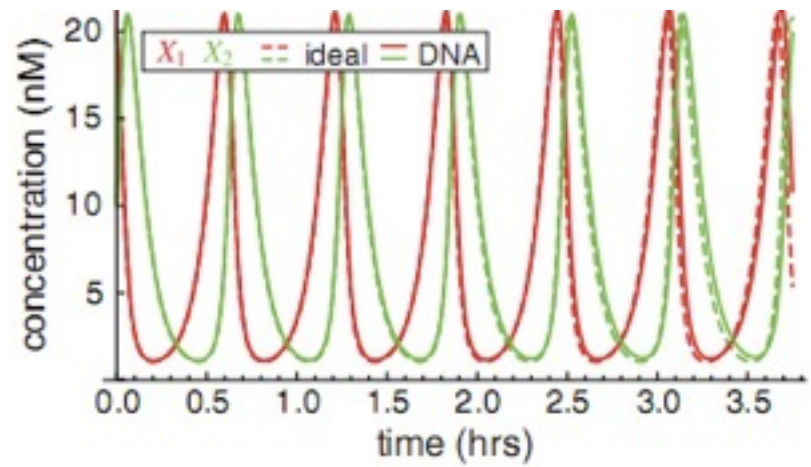
$$X_2 \rightarrow X_1$$
$$X_2 + X_1 \rightarrow \emptyset$$
$$X_1 \rightarrow 2X_1 + X_3$$
$$2X_1 \rightarrow \emptyset$$
$$X_3 \rightarrow X_2$$
$$X_3 \rightarrow \emptyset$$



Predator-prey

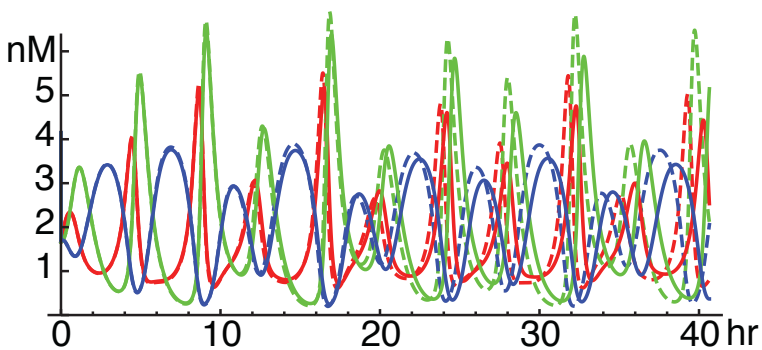
$$1: X_1 + X_2 \xrightarrow{k_1} 2X_2$$
$$2: X_1 \xrightarrow{k_2} 2X_1$$
$$3: X_2 \xrightarrow{k_3} \emptyset$$

	unscaled	scaled
k_1	1.5	$5 \cdot 10^5$ /M/s
k_2	1	1/300 /s
k_3	1	1/300 /s



Rössler (chaotic)

$$X_1 \rightarrow 2X_1$$
$$2X_1 \rightarrow X_1$$
$$X_2 + X_1 \rightarrow 2X_2$$
$$X_2 \rightarrow \emptyset$$
$$X_1 + X_3 \rightarrow \emptyset$$
$$X_3 \rightarrow 2X_3$$
$$2X_3 \rightarrow X_3$$

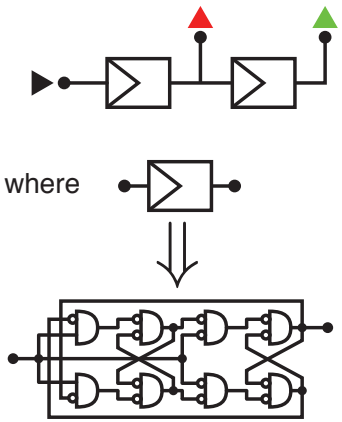


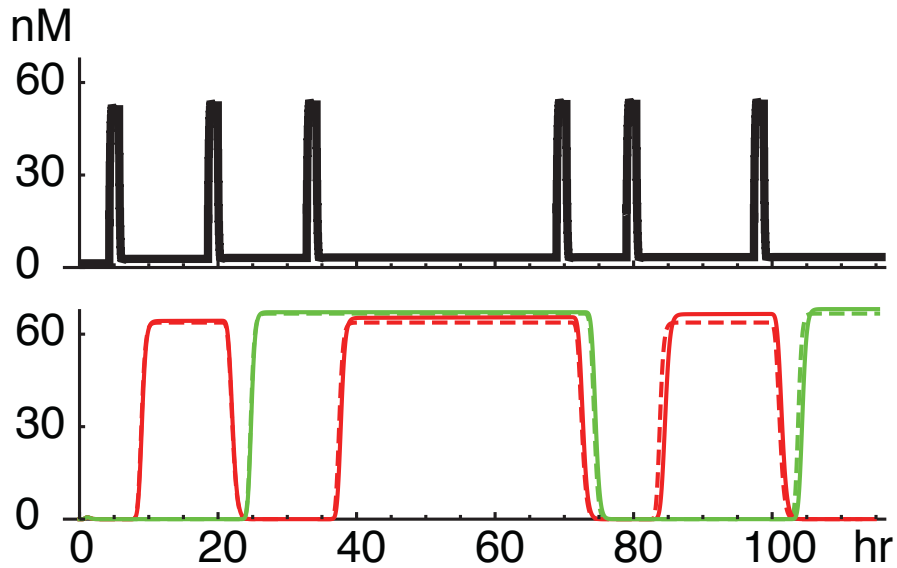
- - - simulation of ideal CRN

—— simulation of DNA implementation

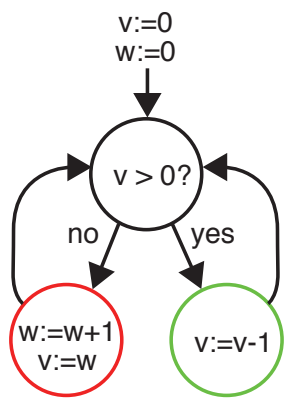
Dynamic logic circuits and state machines with strand displacement cascades (simulations)

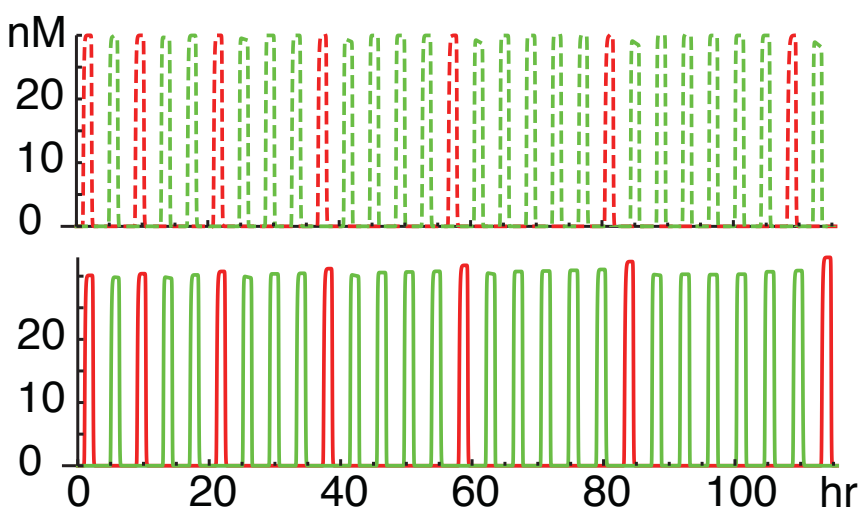
2-bit pulse counter
(digital circuit)





incrementor state machine





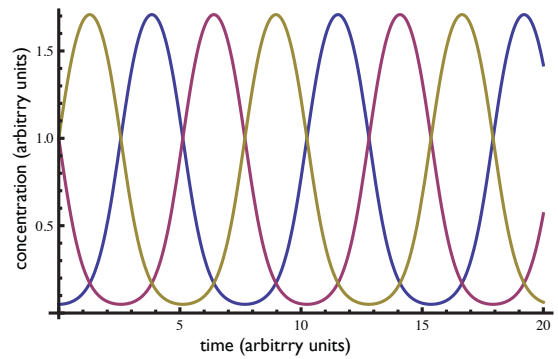
- - - simulation of ideal CRN

—— simulation of DNA implementation

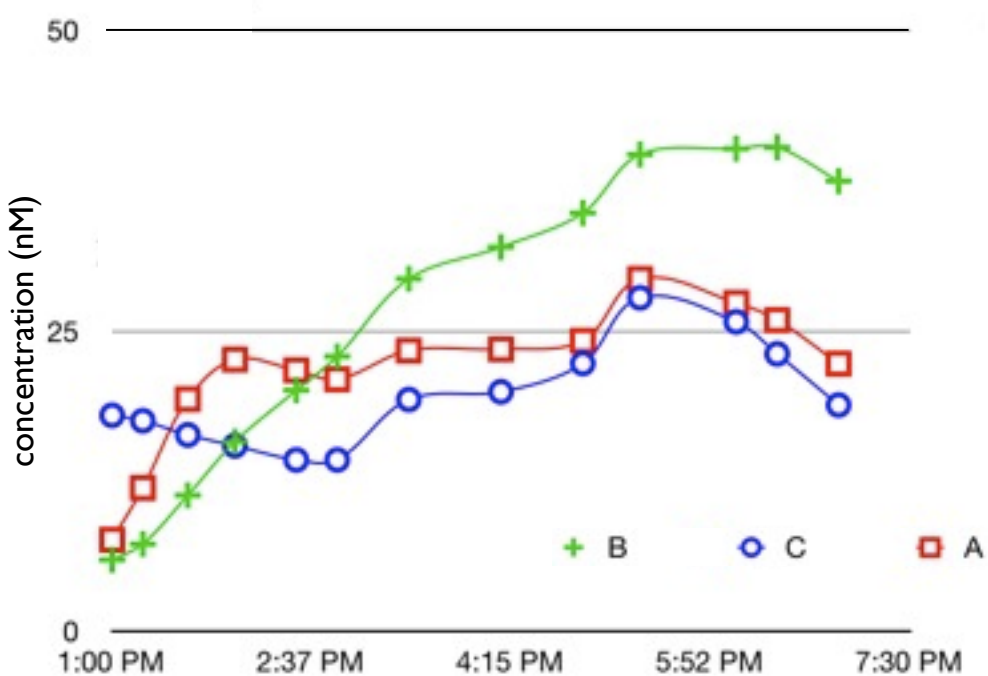
Toward laboratory implementation of CRN ⇒ strand displacement cascades (work in progress)

Ideal autocatalytic oscillator

$$A+B \xrightarrow{1} 2B$$
$$B+C \xrightarrow{1} 2C$$
$$C+A \xrightarrow{1} 2A$$



Recent laboratory experiment



The Programming Language of Chemical Kinetics

Using the language of Chemical Reaction Networks (CRNs) prescriptively as a "programming language" rather than descriptively as a modeling language for existing systems

stochastic and mass-action

different notions of input&output | fast/slow | possible/impossible | deterministic/allowing error

How To Discipline Your DNA Molecules with Strand Displacement Cascades

Strand Displacement Cascades is a flexible technology for implementing complex nucleic-acid reaction networks in the laboratory

artificial analogs of signaling networks | 3 rules description

catalytic amplifier, circuits, neural networks | theoretically can implement any CRNs

Acknowledgements

Jehoshua Bruck
Cezanne Camacho
Luca Cardelli
Ho-Lin Chen
Yuan-Jyue Chen
Matthew Cook
Eric Klavins
Kevin Oishi
Matthew Olson
Lulu Qian
Georg Seelig
Niranjan Srinivas
Erik Winfree
David Zhang

DNA 17
Winfree group
Rothemund group
Seelig group

NSF MPP grant

CI Fellows