

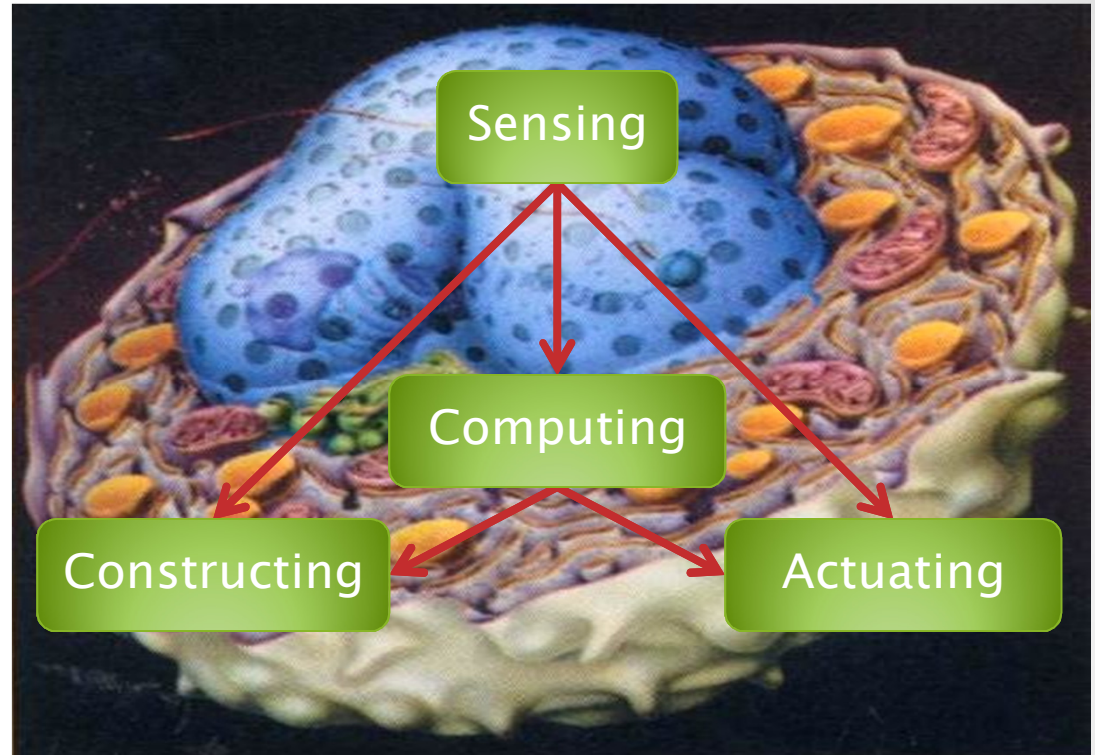
Two-Domain DNA Strand Displacement

Luca Cardelli
Microsoft Research

Dagstuhl, 2010-07-06
<http://lucacardelli.name>

Nanoscale Engineering

- Sensing
 - Reacting to forces
 - Binding to molecules
- Actuating
 - Releasing molecules
 - Producing forces
- Constructing
 - Chassis
 - Growth
- Computing
 - Signal Processing
 - Decision Making

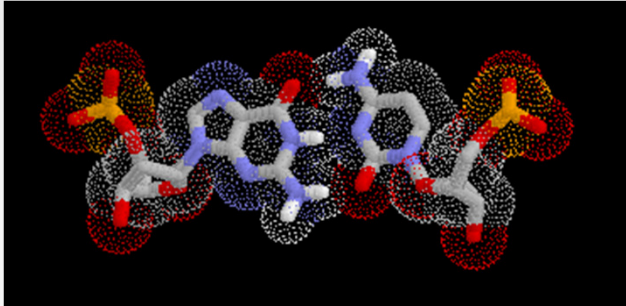


Nucleic Acids can do all this.
And interface to **biology**.
And are **programmable**.

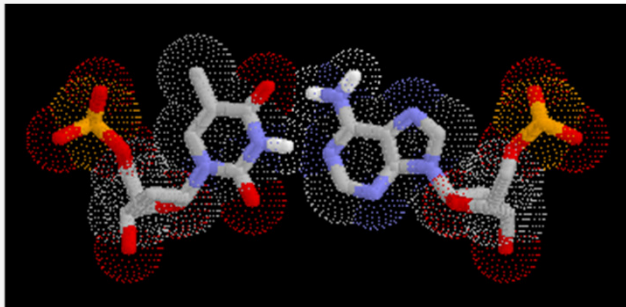
Strand Displacement Basics

...

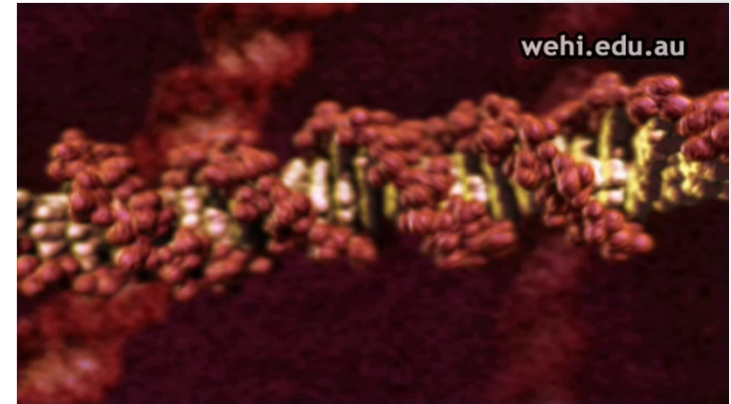
DNA



GC Base Pair
Guanine-Cytosine

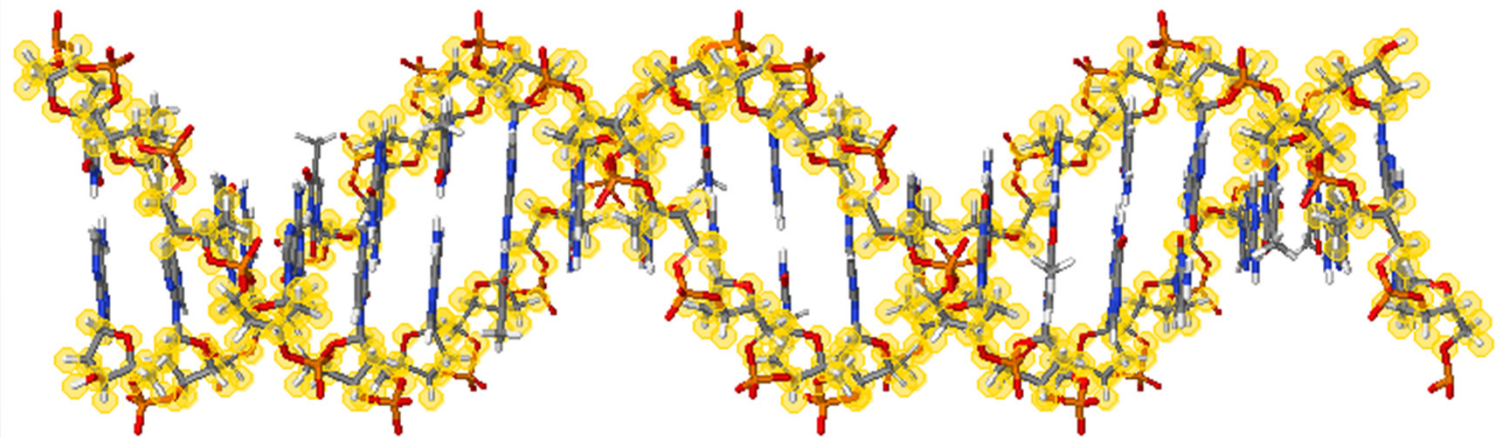


TA Base Pair
Thymine-Adenine



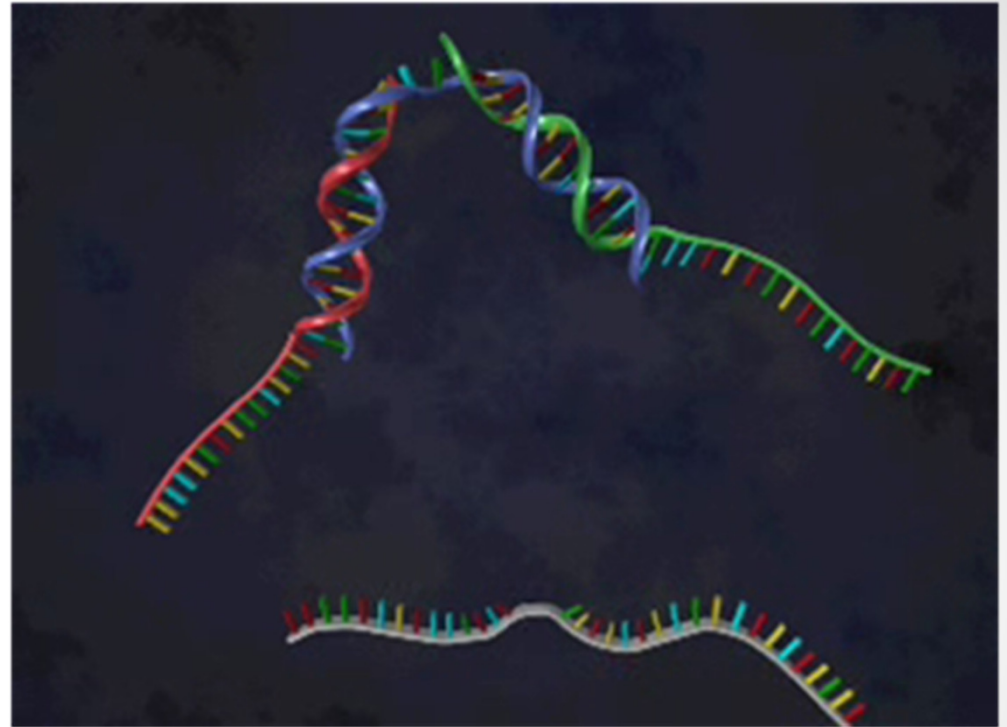
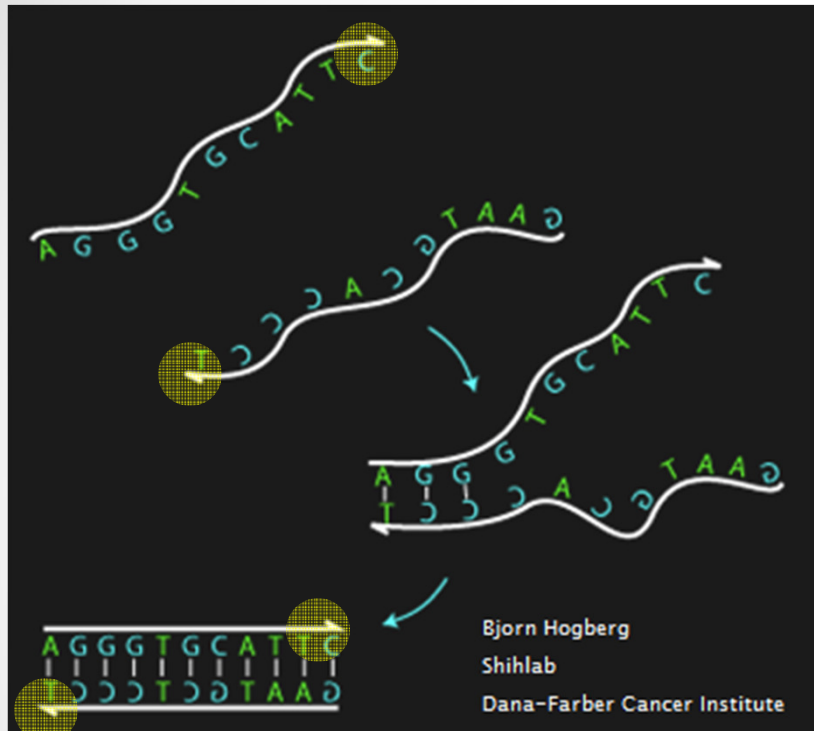
[Interactive DNA Tutorial](http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/dna.html)

(<http://www.biosciences.bham.ac.uk/labs/minchin/tutorials/dna.html>)



Sequence of Base Pairs (GACT alphabet)

Hybridization



- Strands with opposite orientation and complementary base pairs stick to each other (Watson-Crick duality).
- This is all we are going to use
 - We are not going to exploit DNA replication, transcription, translation, restriction and ligation enzymes, etc., which enable other classes of tricks.

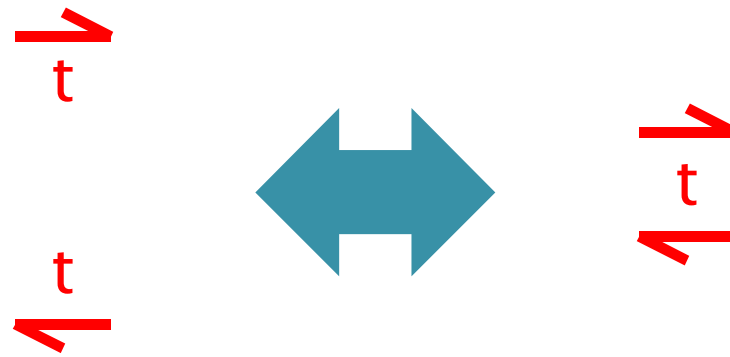
Domains

- Subsequences on a DNA strand are called **domains**.
- PROVIDED they are “independent” of each other.

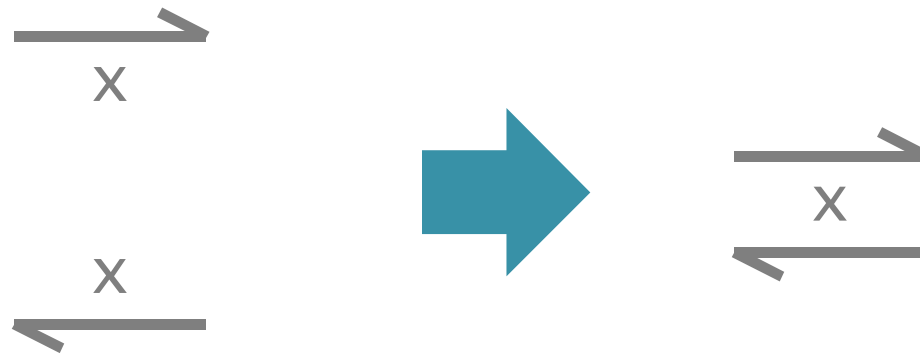


- I.e., differently named domains must not hybridize:
 - With each other
 - With each other's complement
 - With subsequences of each other
 - With concatenations of other domains (or their complements)
 - Etc.
- How to choose domains (subsequences) that are suitably independent is a tricky issue that is still somewhat of an open problem (with a vast literature). But it can work in practice.

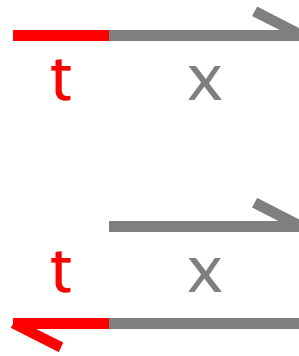
Short Domains



Long Domains

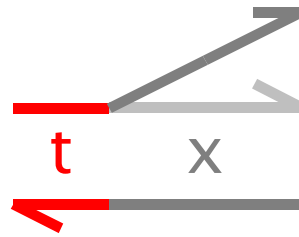


Strand Displacement



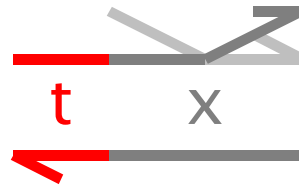
“Toehold Mediated”

Strand Displacement



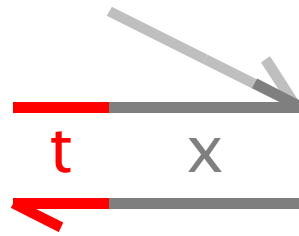
Toehold Binding

Strand Displacement



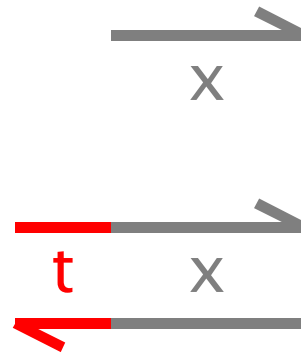
Branch Migration

Strand Displacement



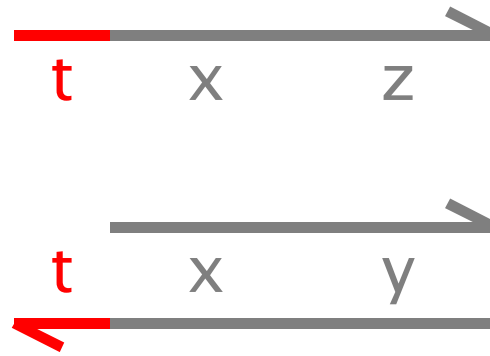
Displacement

Strand Displacement

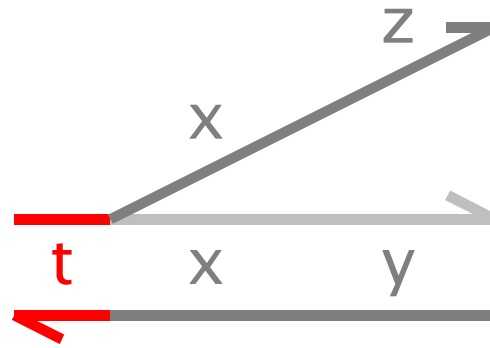


Irreversible

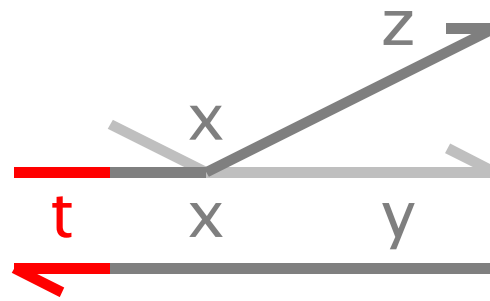
Bad Match



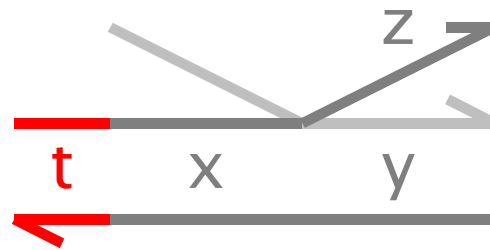
Bad Match



Bad Match

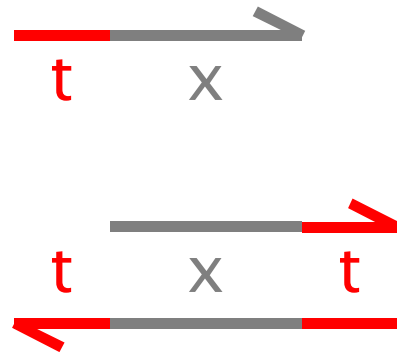


Bad Match

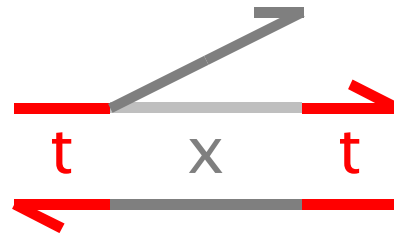


Cannot proceed
Hence will undo

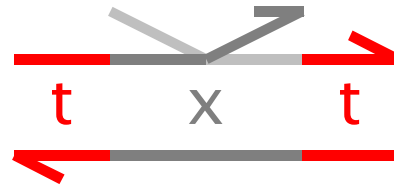
Toehold Exchange



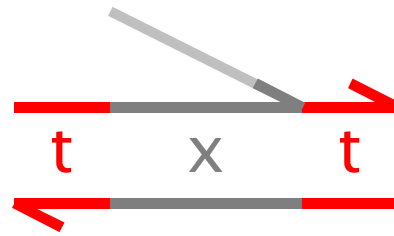
Toehold Exchange



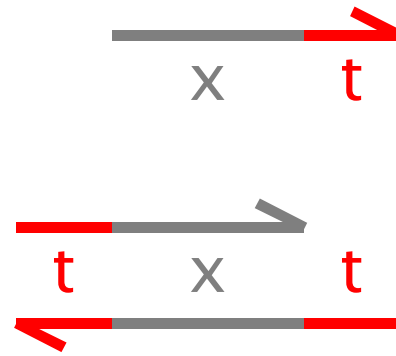
Toehold Exchange



Toehold Exchange

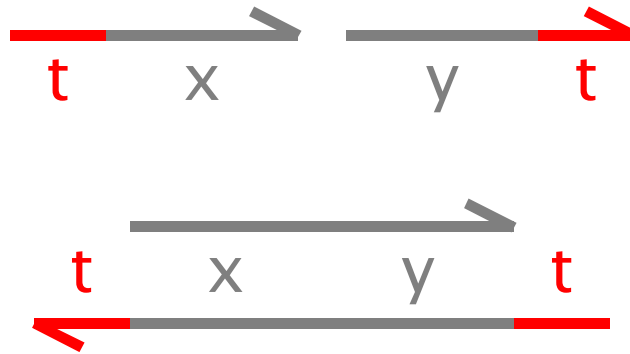


Toehold Exchange

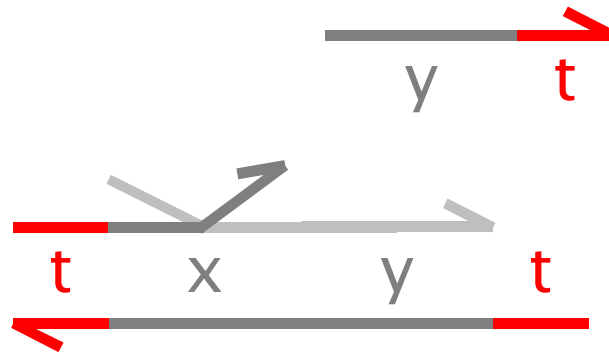


Reversible

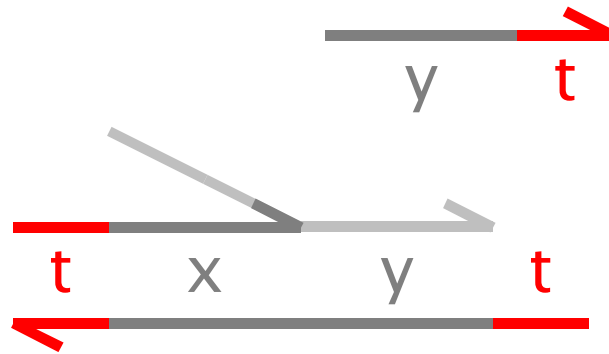
Cooperative Displacement



Cooperative Displacement

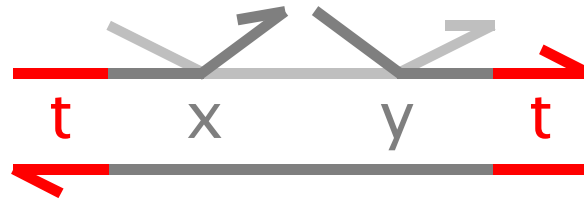


Cooperative Displacement

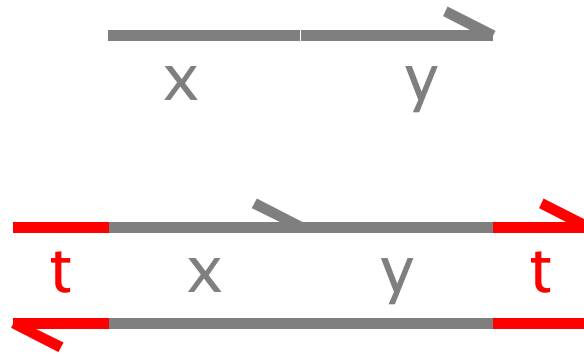


Single input
will reverse

Cooperative Displacement



Cooperative Displacement



Double input
is irreversible

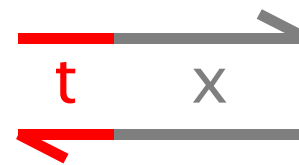
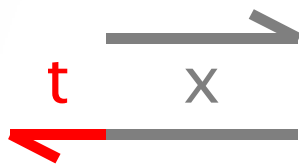
Summary (1)

Signal



SS Waste

Gate



DS Waste

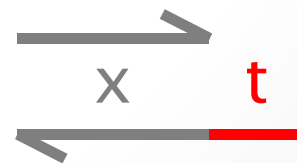
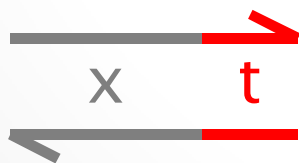


SS Waste



Co-Signal

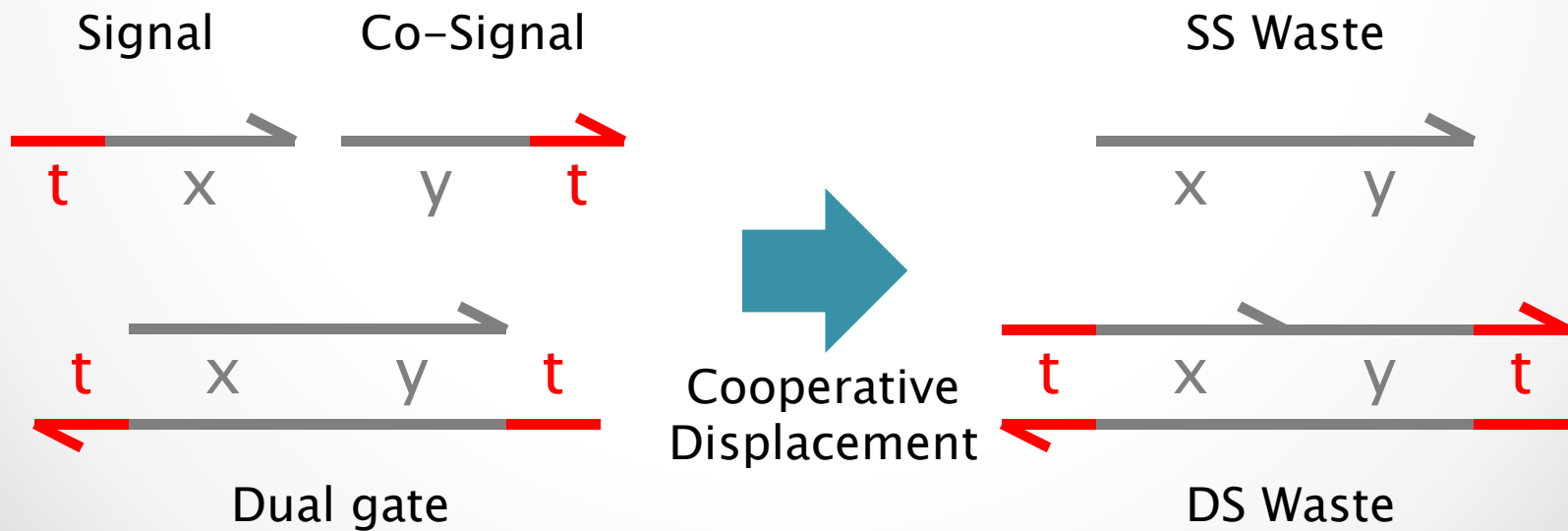
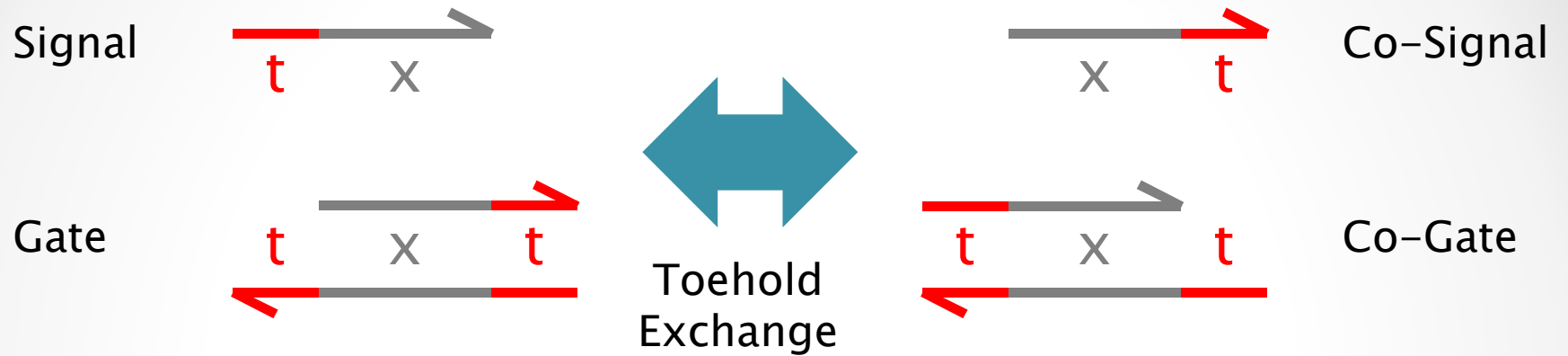
DS Waste



Co-Gate



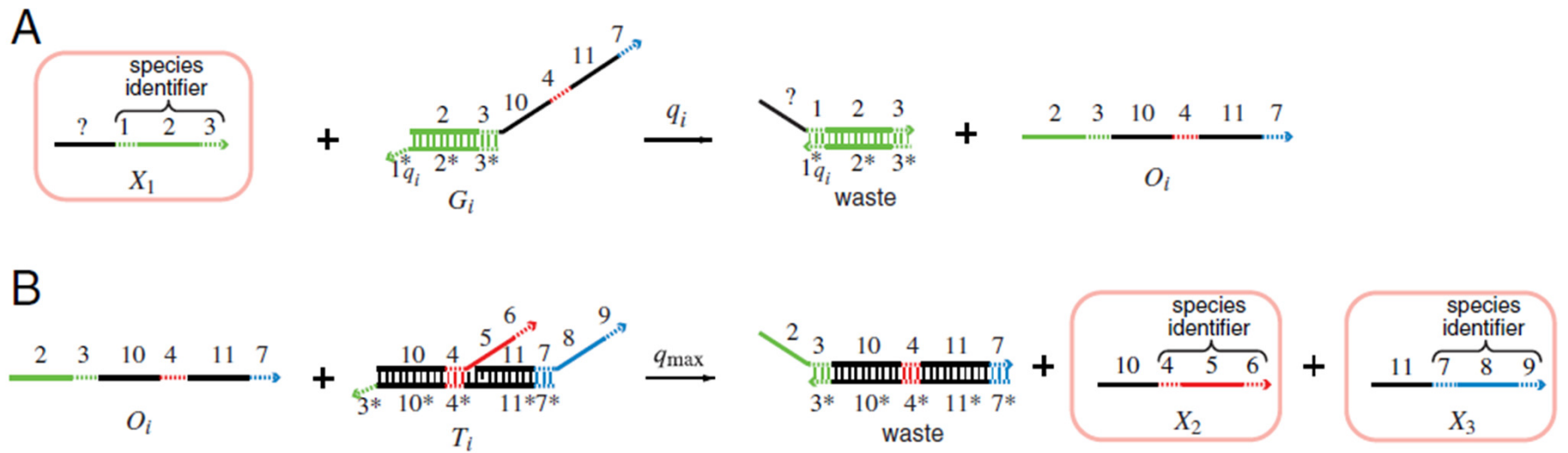
Summary (2)



Signals & Gates

...

Four-Domain Signals

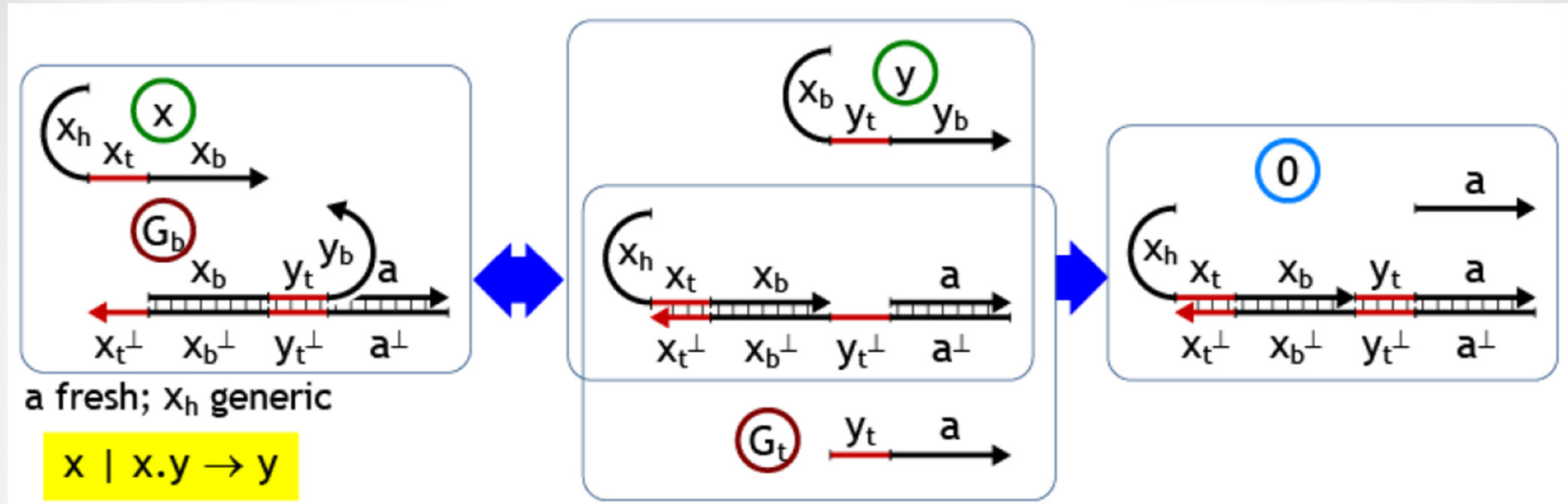


DNA as a universal substrate for chemical kinetics

David Soloveichik^{a,1}, Georg Seelig^{a,b,1}, and Erik Winfree^{c,1}

PNAS | March 23, 2010 | vol. 107 | no. 12 | 5393–5398

Three-Domain Signals



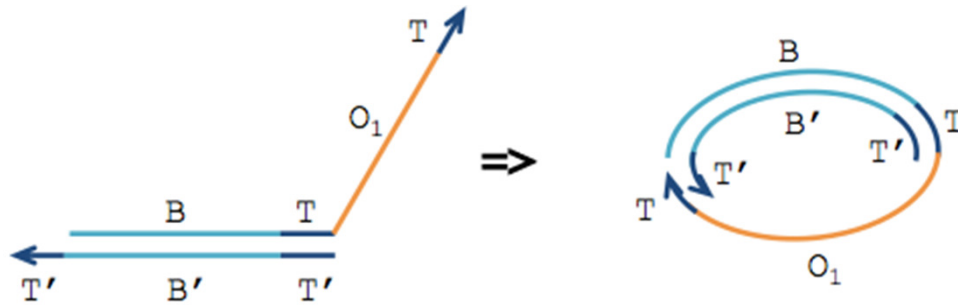
Strand Algebras for DNA Computing

Luca Cardelli

DNA Computing and Molecular Programming.

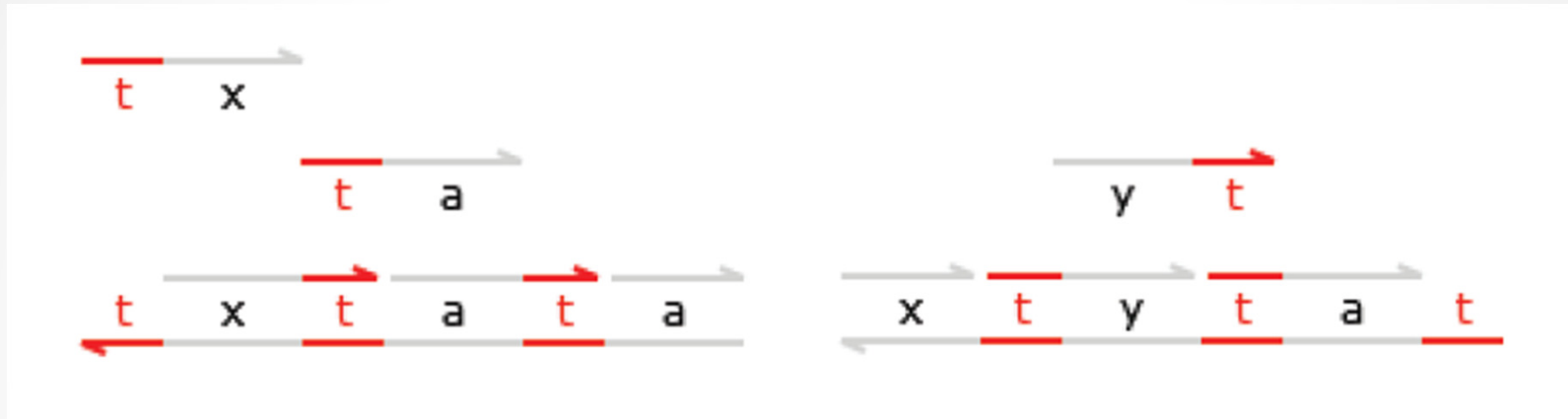
15th International Conference, DNA 15, LNCS 5877, Springer 2009, pp 12–24.

“Lulu’s Trouble”



(from D.Soloveichik)

Two-Domain Signals



Two-Domain DNA Strand Displacement

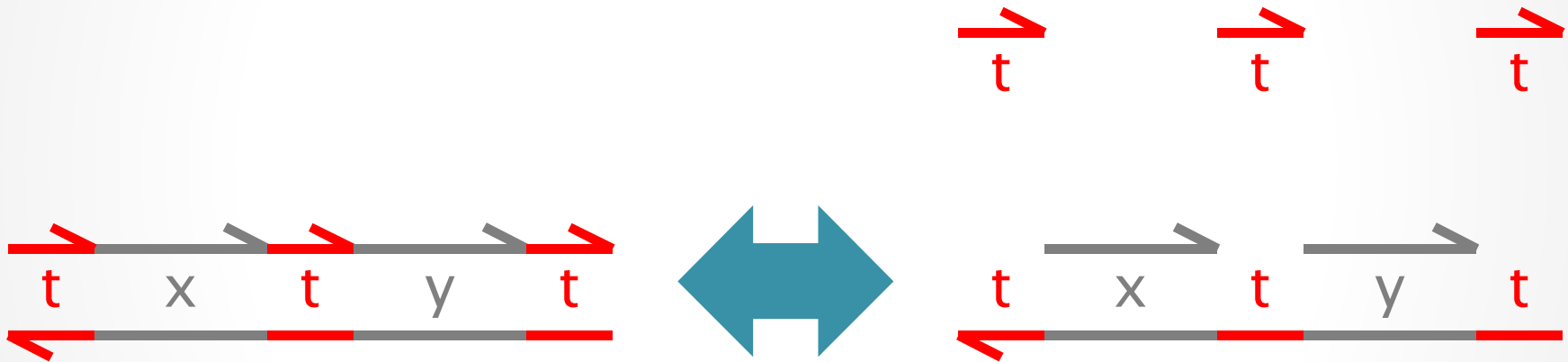
Luca Cardelli

In S. B. Cooper, E. Kashefi, P. Panangaden (Eds.):
Developments in Computational Models (DCM 2010).
EPTCS 25, 2010, pp. 33–47. May 2010.

Top-Nicked Double Strands

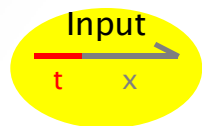
Signals have a simple structure: just two domains.

Gates have a simple structure:
'top-nicked' double-stranded DNA with no 'frills'.

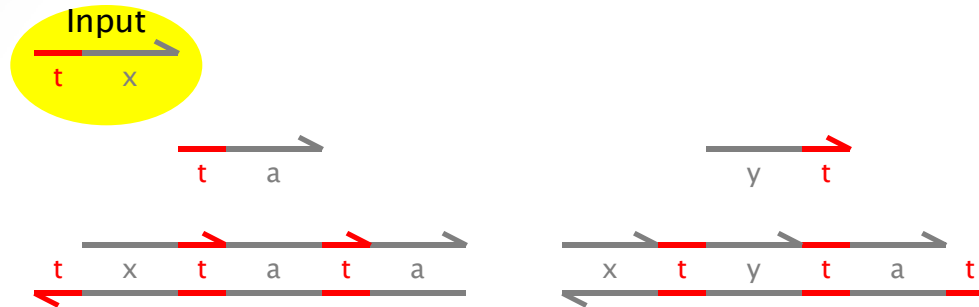


A top-nicked double-strand is 'equivalent'
to a double strand with open toeholds.
These situations shall not be distinguished.

Transducer $x \rightarrow y$

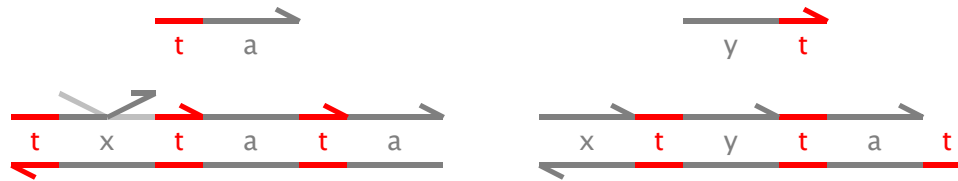


Transducer $x \rightarrow y$

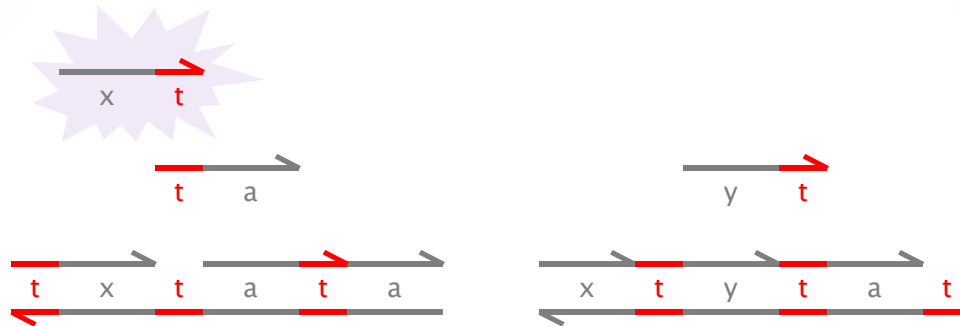


ta is a *private* signal (a different 'a' for each xy pair)

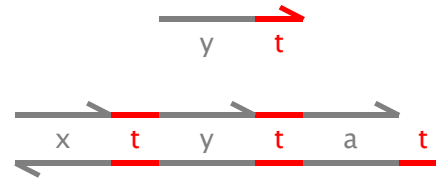
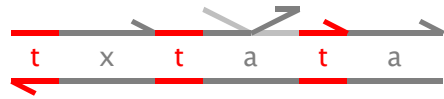
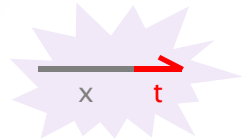
Transducer $x \rightarrow y$



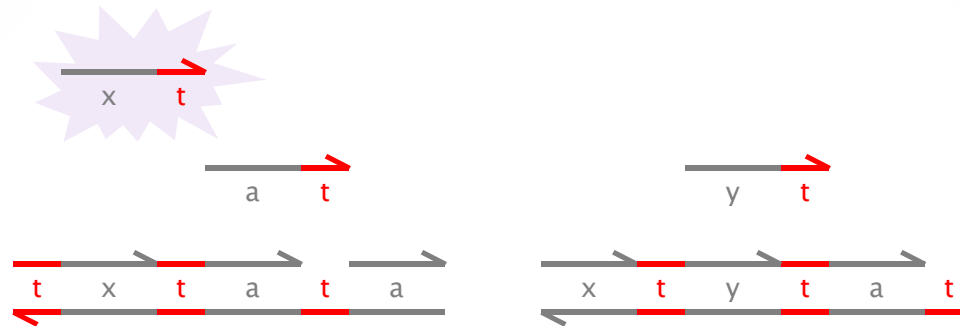
Transducer $x \rightarrow y$



Transducer $x \rightarrow y$

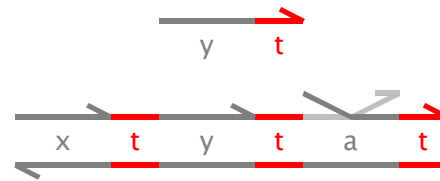
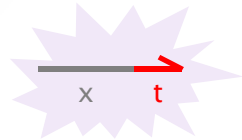


Transducer $x \rightarrow y$

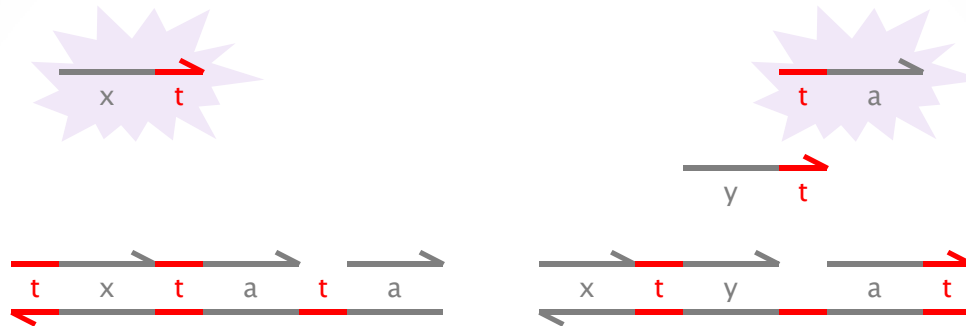


So far, a tx *signal* has produced an at *cosignal*.
But we want signals as output, not cosignals.

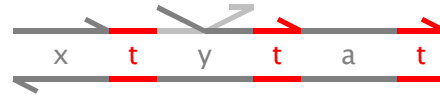
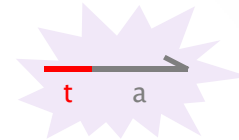
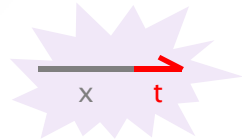
Transducer $x \rightarrow y$



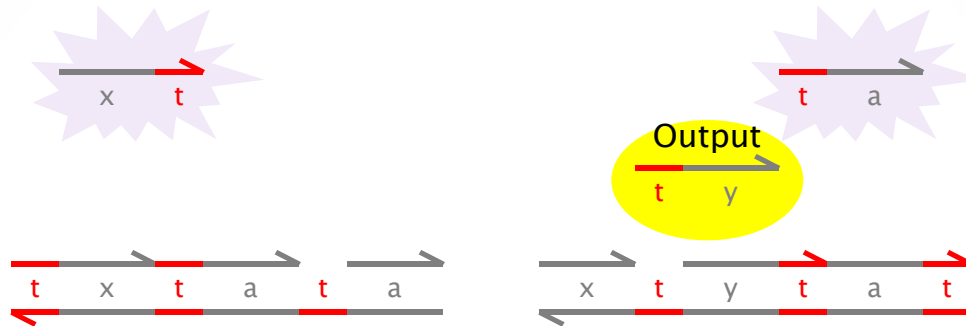
Transducer $x \rightarrow y$



Transducer $x \rightarrow y$



Transducer $x \rightarrow y$



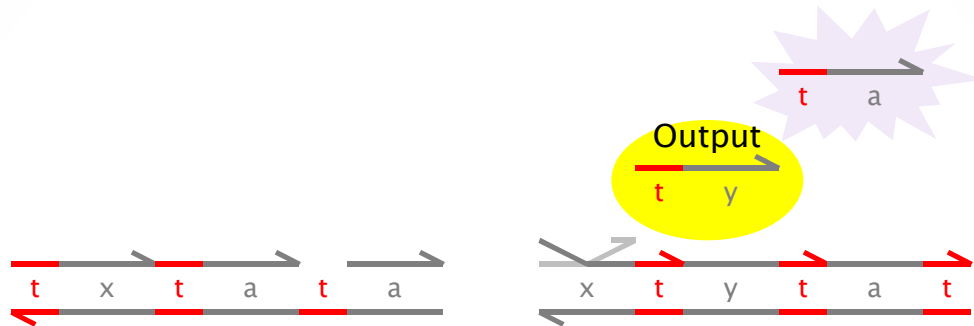
Here is our output **ty** *signal*.

But we are not done yet:

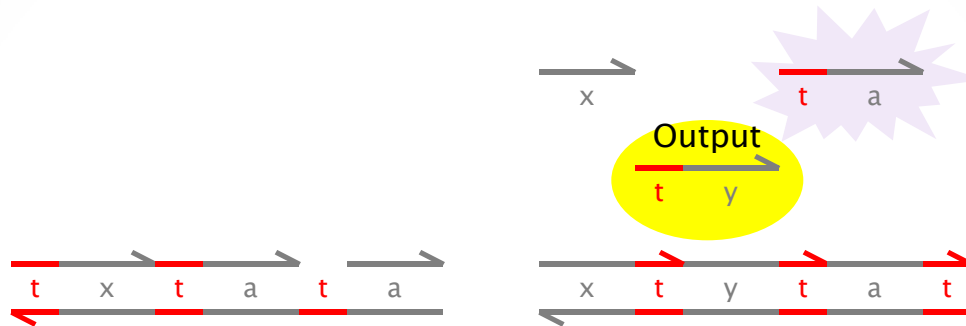
- 1) We need to make the output irreversible.
- 2) We need to remove the garbage.

We can use (2) to achieve (1).

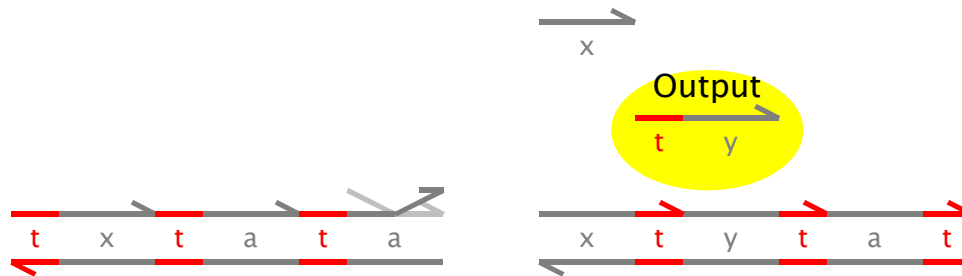
Transducer $x \rightarrow y$



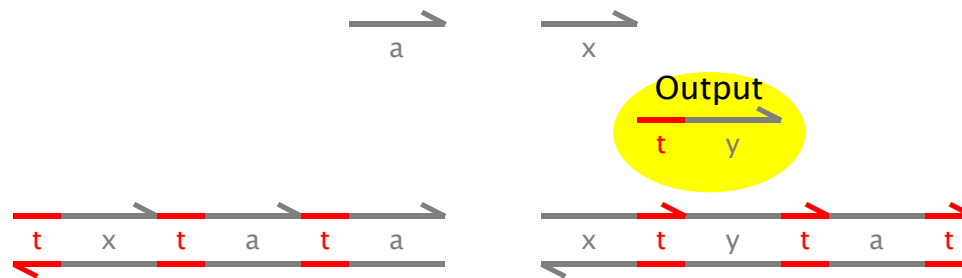
Transducer $x \rightarrow y$



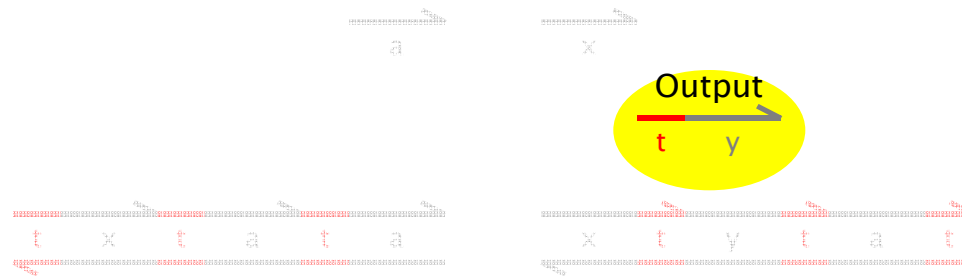
Transducer $x \rightarrow y$



Transducer $x \rightarrow y$



Transducer $x \rightarrow y$



Done.

Note the **tata** motif and how it helps in collection.

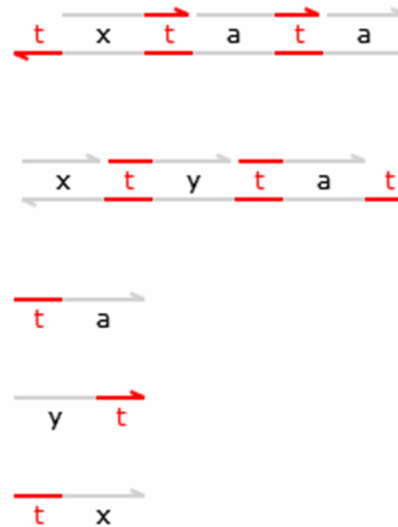
The Transducer in DSD

```
directive sample 50.0  
directive plot <t^ x>; <t^ y>  
directive scale 1.0  
new t@1.0,1.0
```

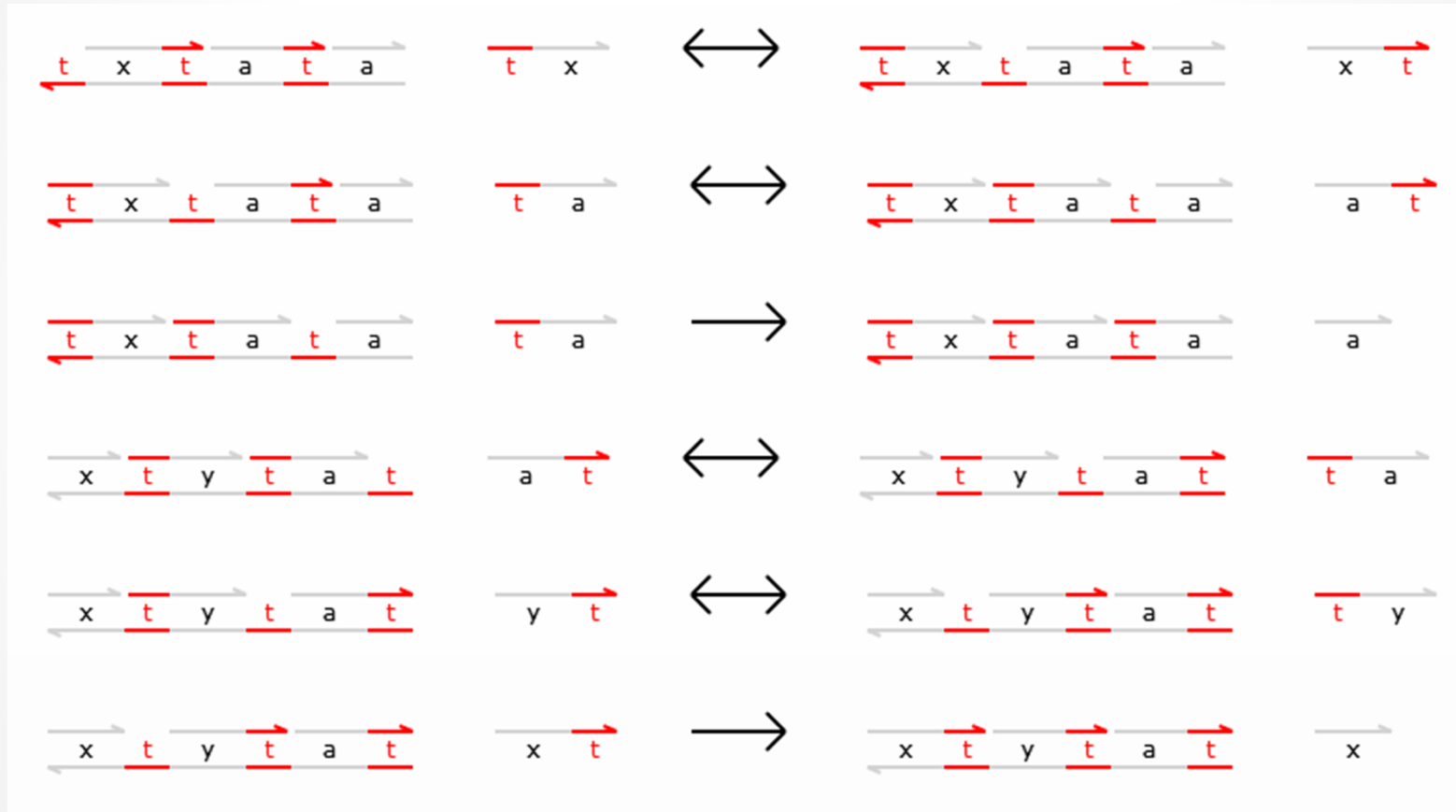
```
def Tr(N, x, y) =  
new a
```

```
( N* <t^ a>  
| N* <y t^>  
| N* t^:[x t^]:[a t^]:[a]  
| N* [x]:[t^ y]:[t^ a]:t^  
)
```

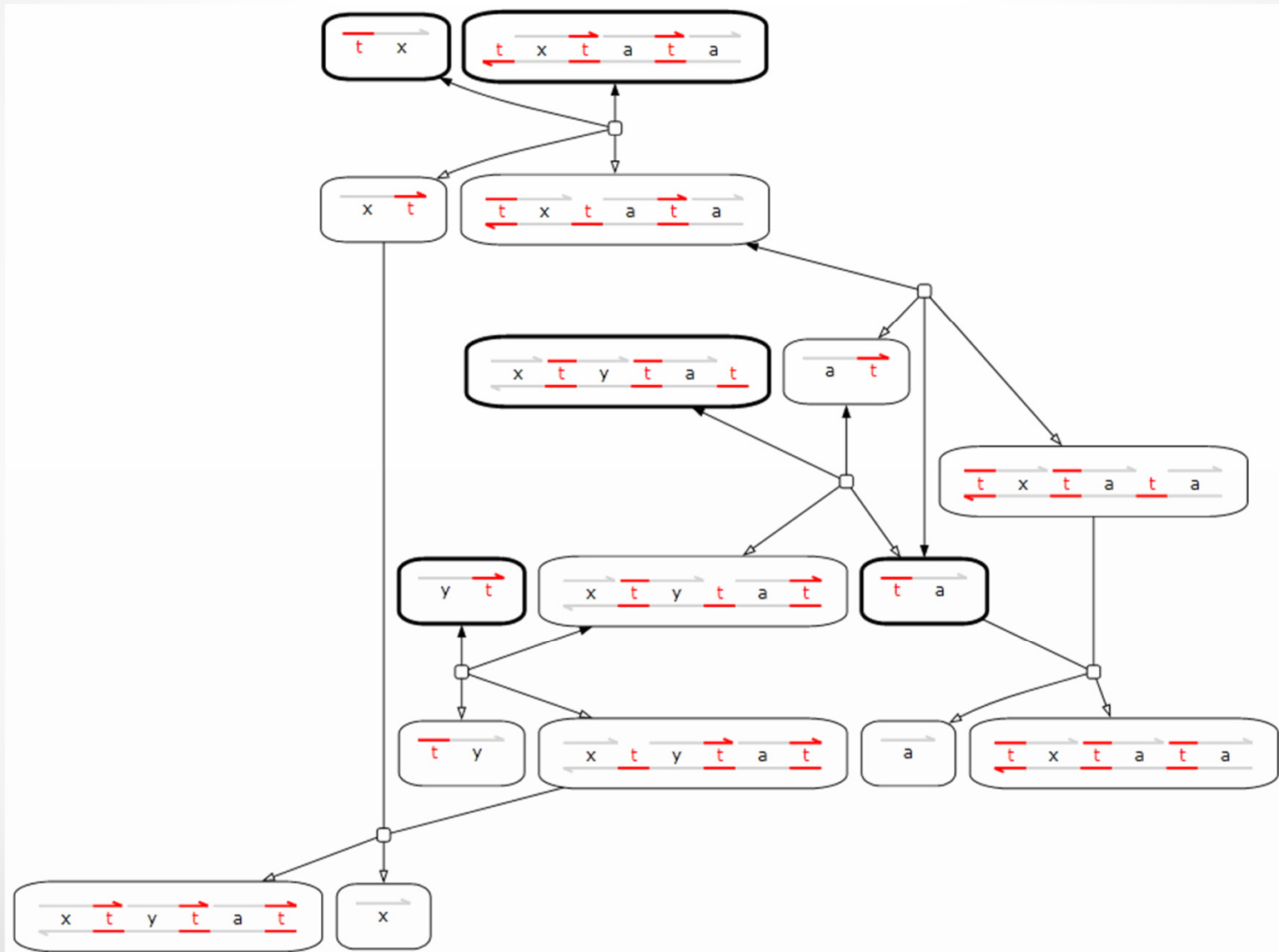
```
( Tr(10, x, y)  
| 1* <t^ x>  
)
```



Transducer Reactions



Transducer Reaction Graph



Transducer Simulation

Examples: Rules: Simulation: View options: Leaks: Domains: v0.12-20100408-1540

Code

Zoom

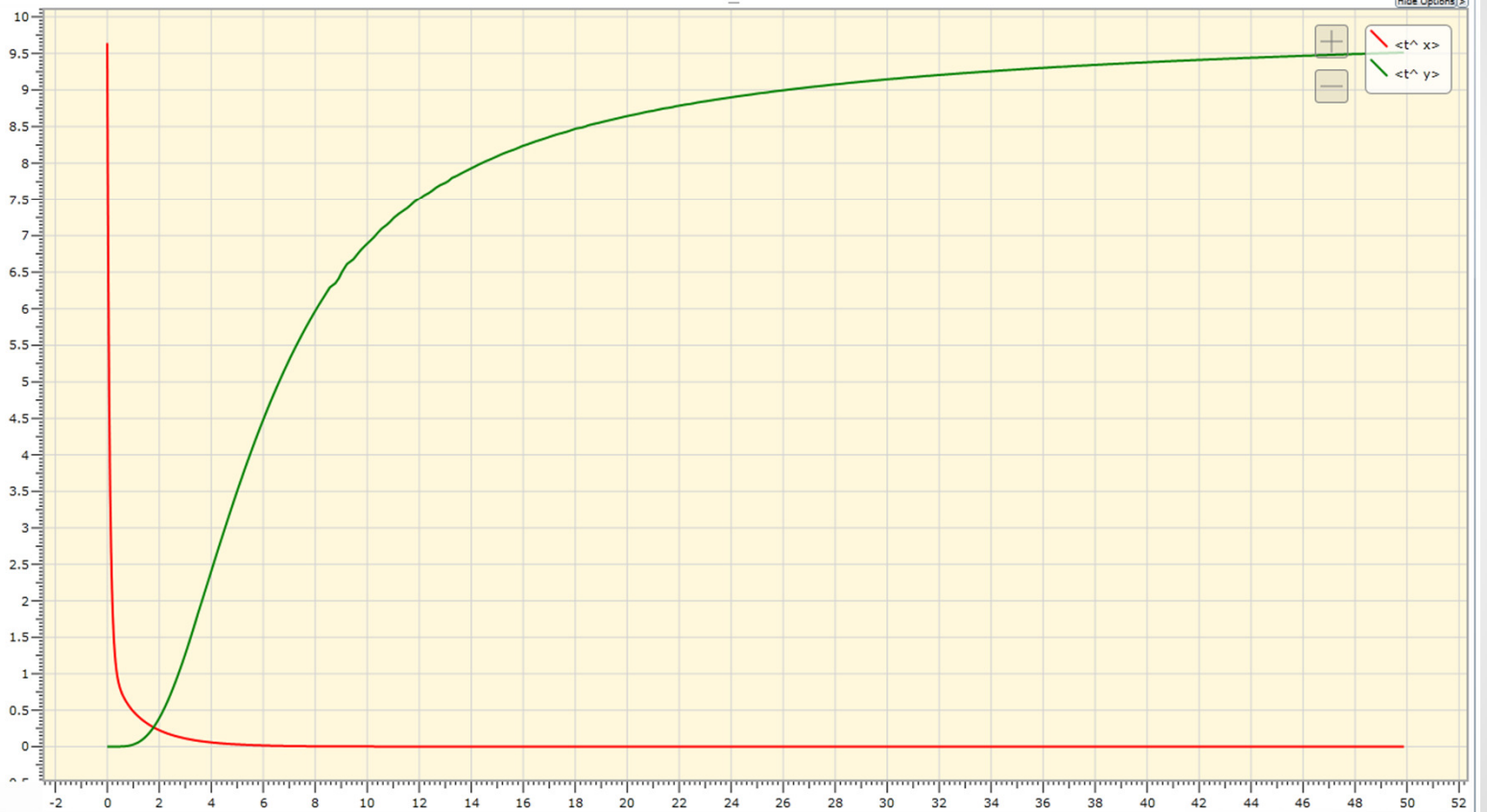
```
directive sample 50.0
directive plot <t^ x> <t^ y>
directive scale 1.0
new t@1.0,1.0

def Tr(N, x, y) =
  new a
  ( N* <t^ a>
  | N* <y t^>
  | N* t^:[x t^]:[a t^]:[a]
  | N* [x]:[t^ y]:[t^ a]:t^
  )

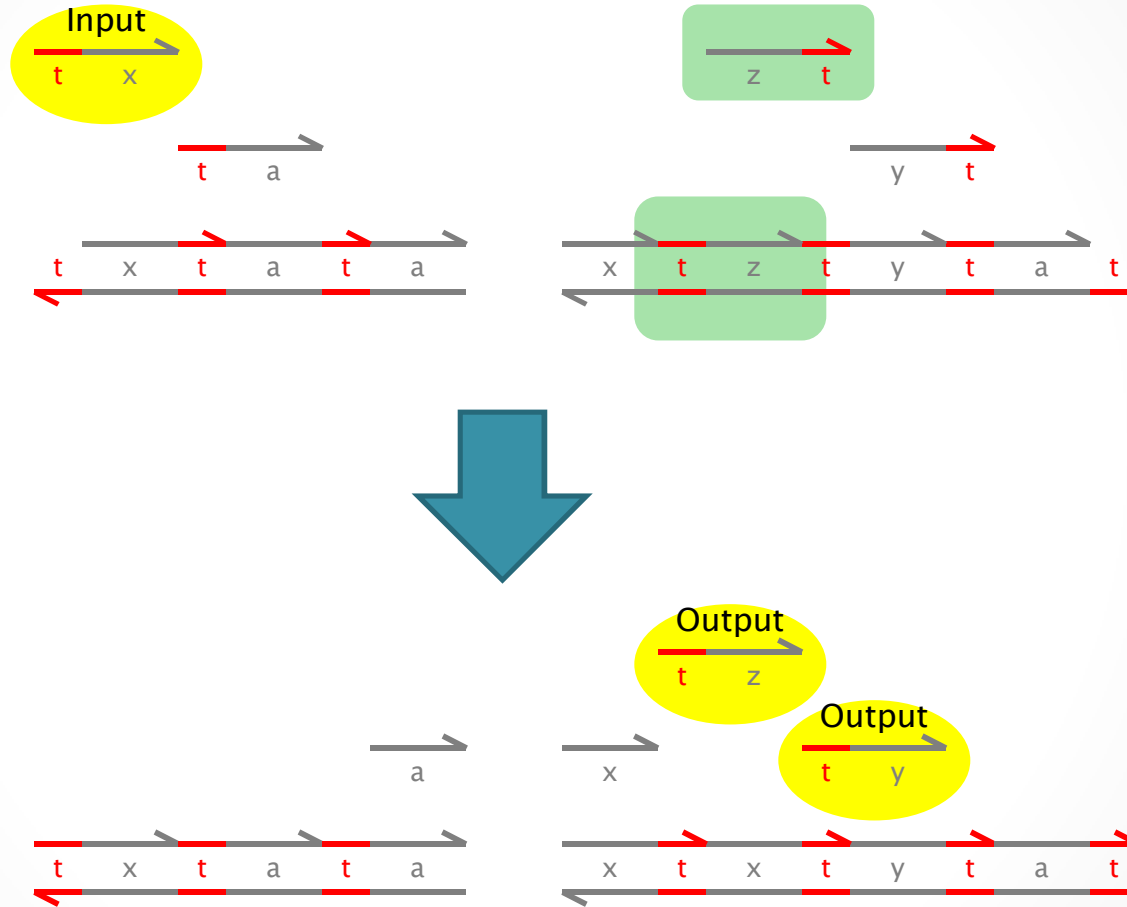
( Tr(10, x, y)
| 1* <t^ x>
)
```

Species Reactions Graph Text SBML Domains Table Plot

Show all Hide all <t^ x> <t^ y>

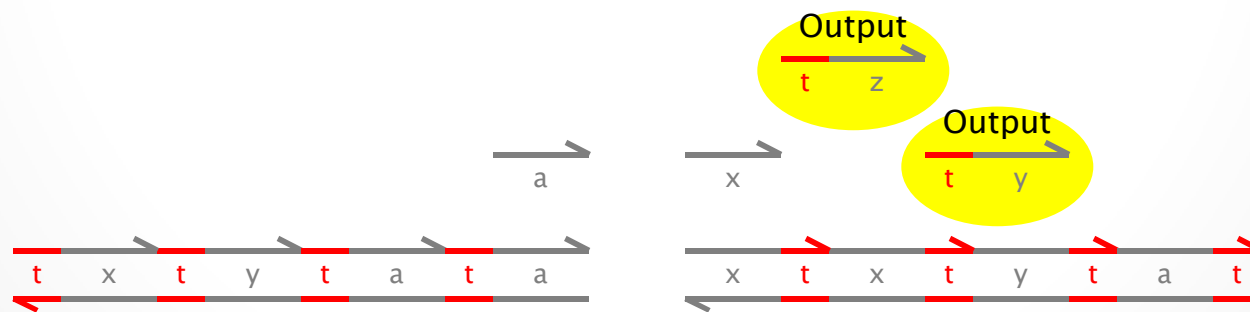
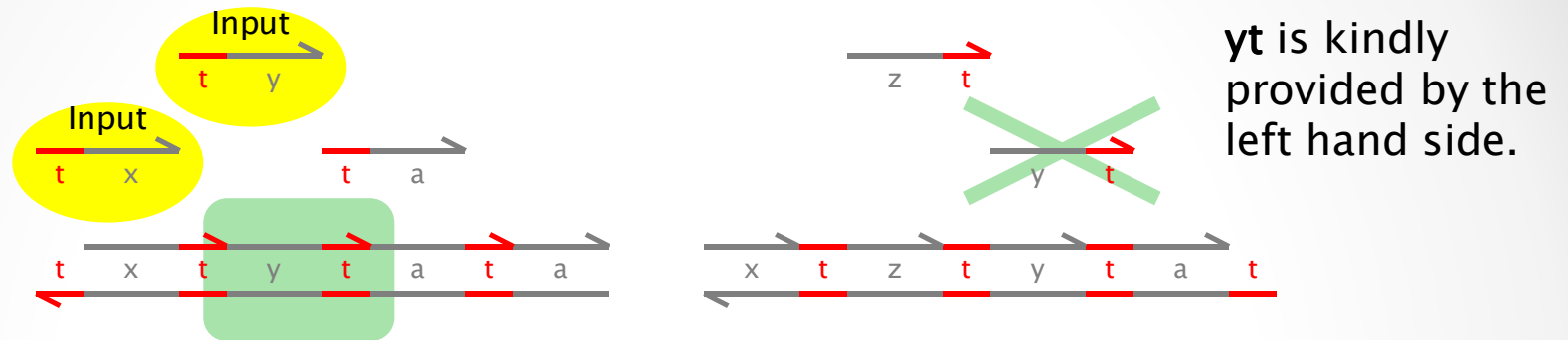


Fork $x \rightarrow y+z$



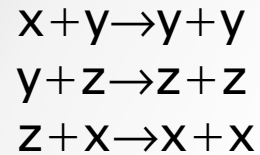
(Amplifier: $x \rightarrow x+x$)

Catalyst $x + y \rightarrow y + z$



(Autocatalyst: $x + y \rightarrow y + y$)

Autocatalytic Oscillator



```

directive sample 100.0 1000
directive plot <t^ x>; <t^ y>;
<t^ z>
(* directive scale 100.0 *)

```

```
new t@1.0,100.0
```

```

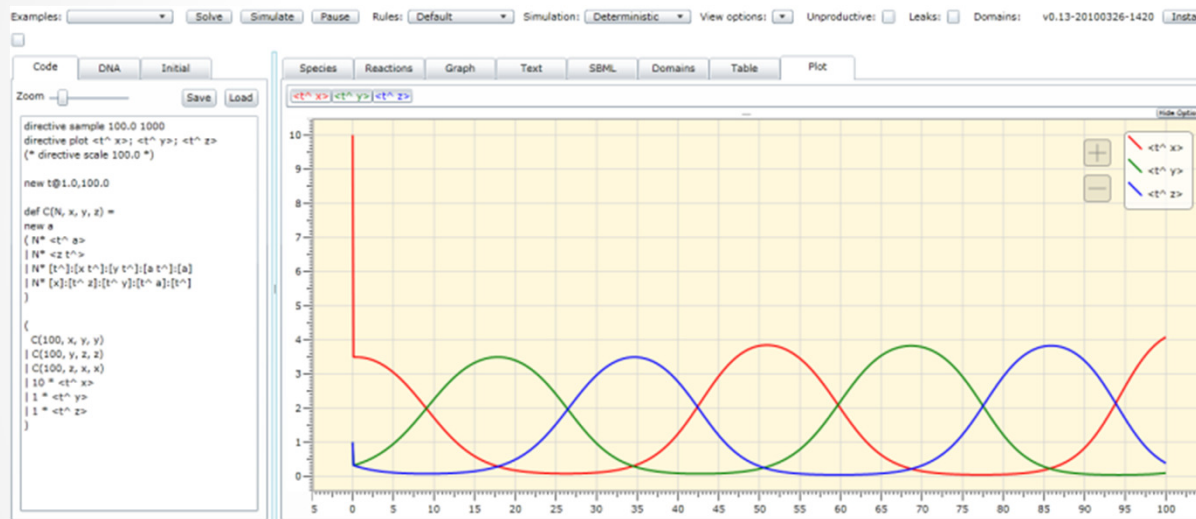
def C(N, x, y, z) =
new a
(N* <t^ a>
| N* <z t^>
| N* [t^]:[x t^]:[y t^]:[a t^]:[a]
| N* [x]:[t^ z]:[t^ y]:[t^ a]:[t^]
)

```

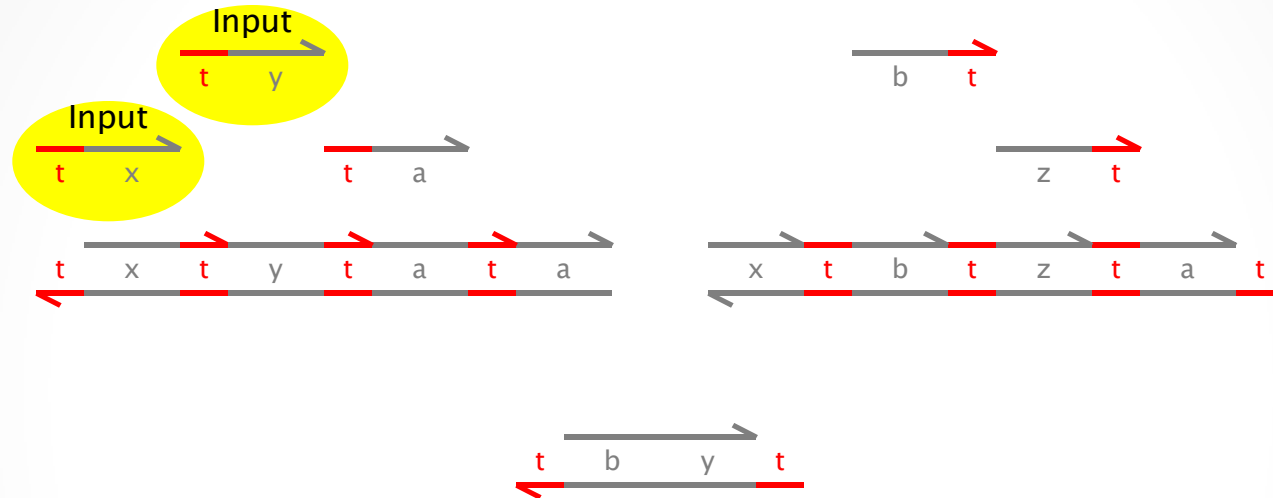
```

(
C(100, x, y, y)
| C(100, y, z, z)
| C(100, z, x, x)
| 10 * <t^ x>
| 1 * <t^ y>
| 1 * <t^ z>
)

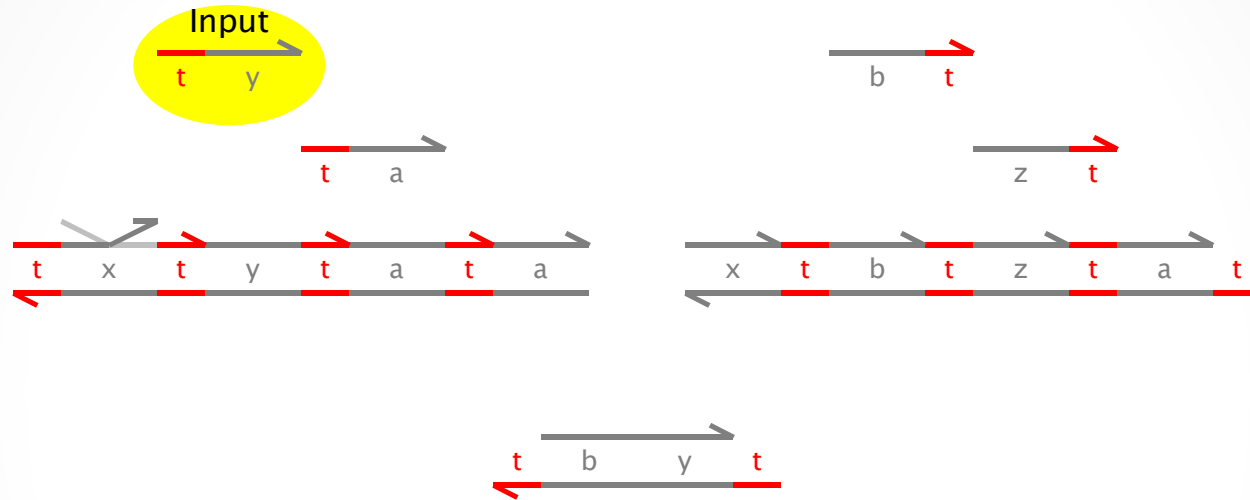
```



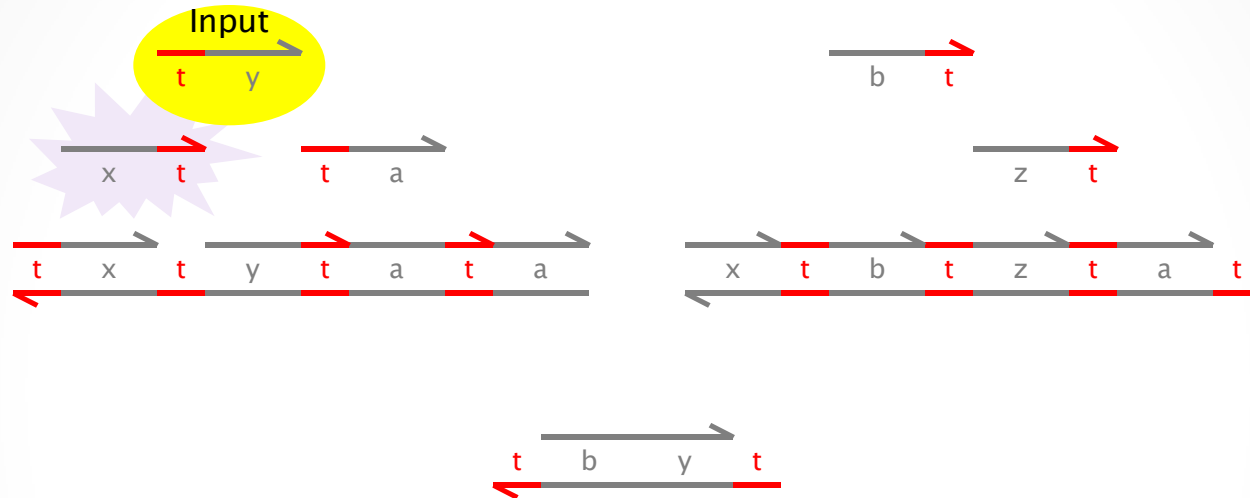
Join $x+y \rightarrow z$



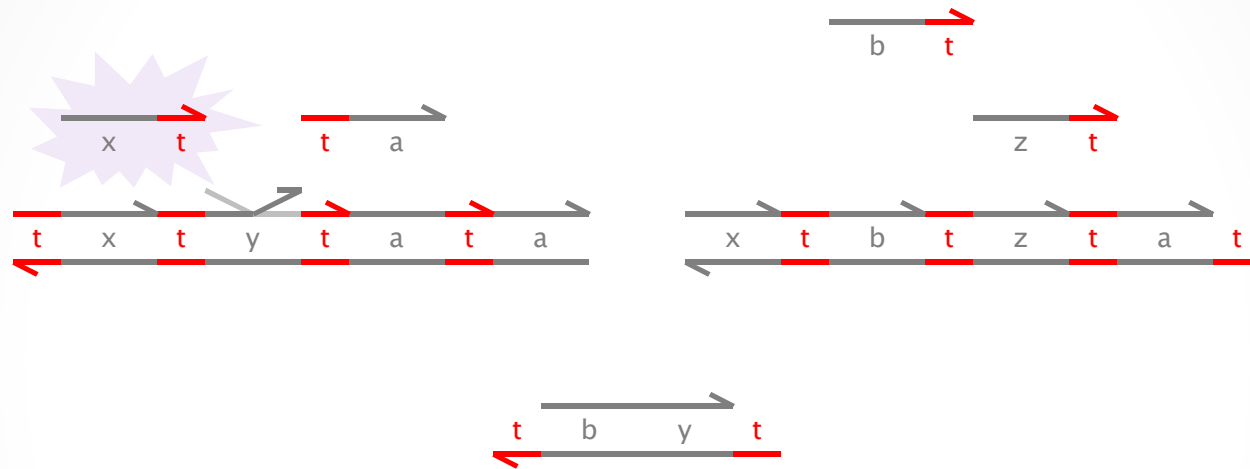
Join $x+y \rightarrow z$



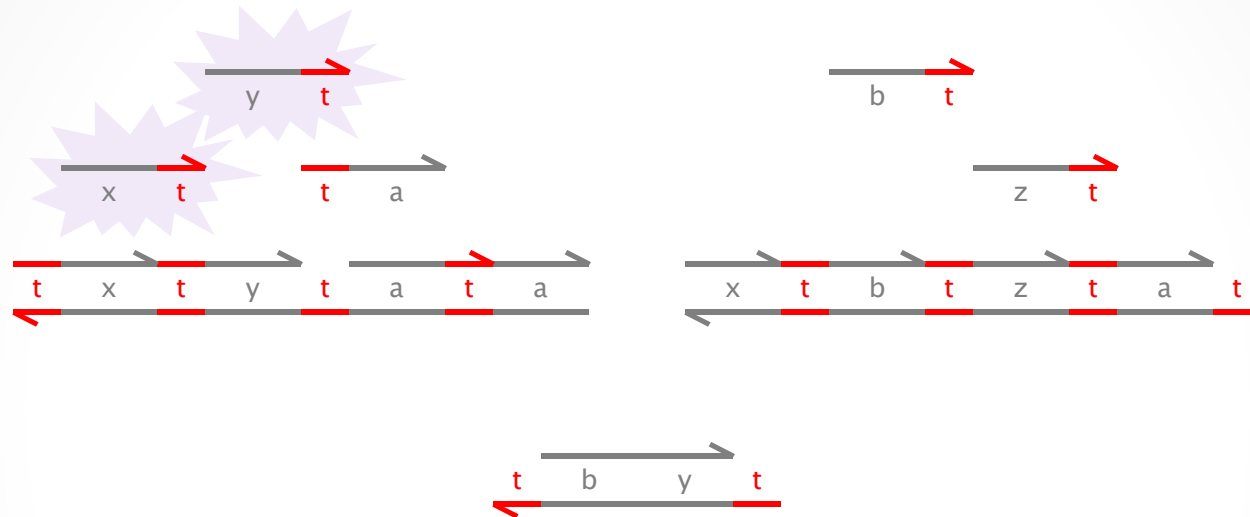
Join $x+y \rightarrow z$



Join $x+y \rightarrow z$



Join $x+y \rightarrow z$



We cannot have a collector just waiting for yt , because there may be innocent yt elsewhere in the system, like here!

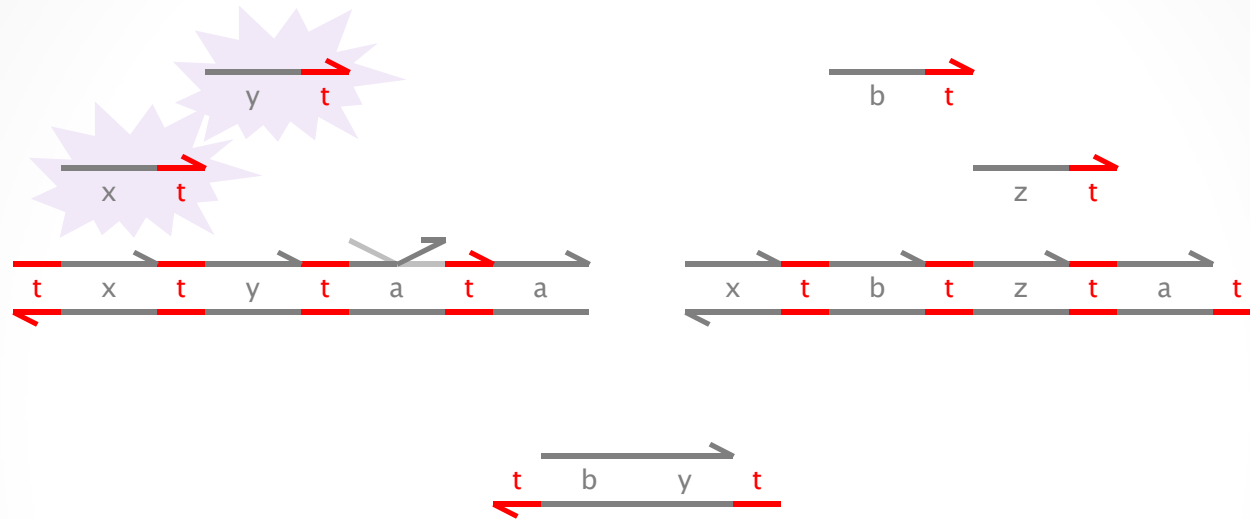


Transducer $x \rightarrow y$

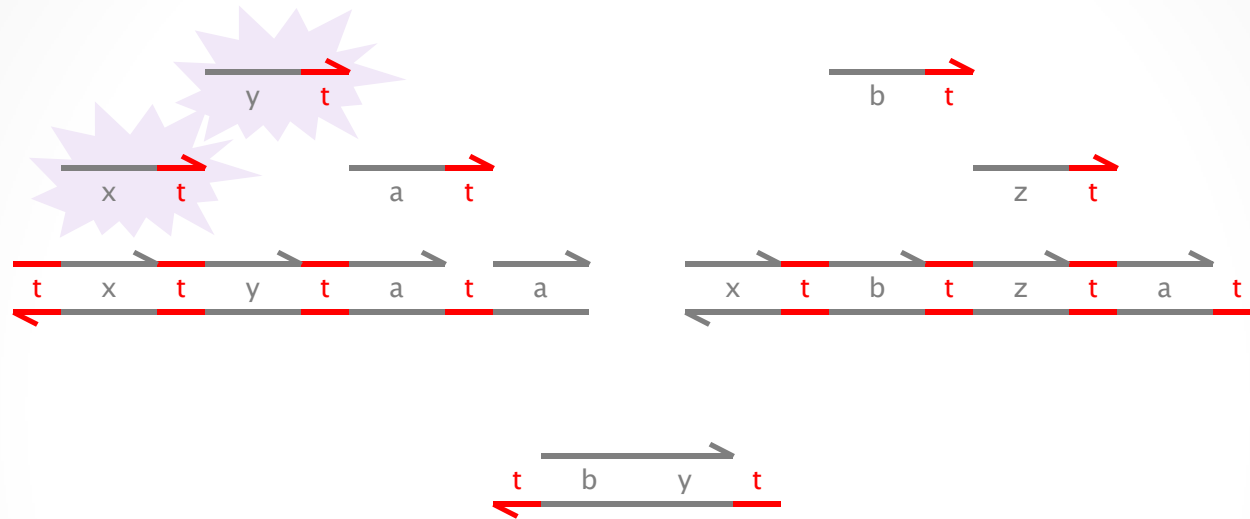
Instead, the collection of yt must be triggered only by a signal signifying that an $x+y \rightarrow z$ gate has fired. That signal is tb , which will trigger the collection of yt after output tz is produced.

bt is a *private* signal (a different 'b' for each xyz triple)

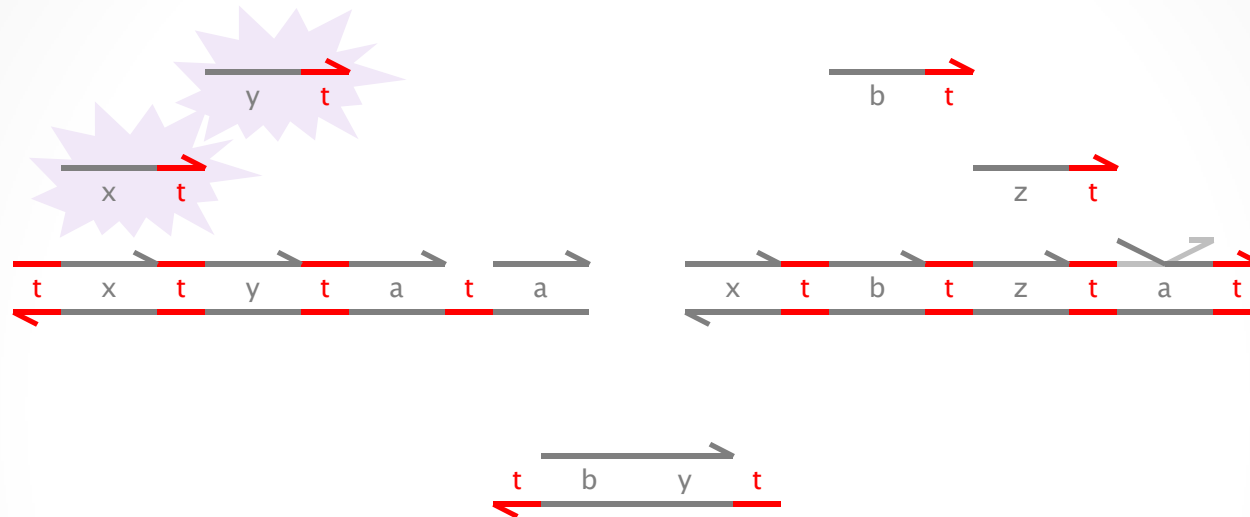
Join $x+y \rightarrow z$



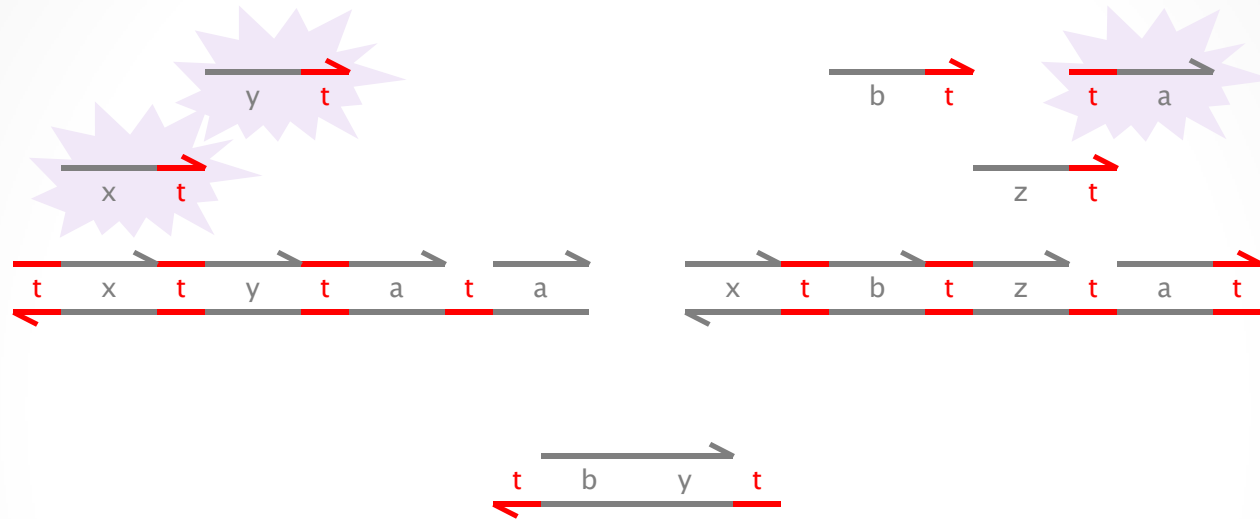
Join $x+y \rightarrow z$



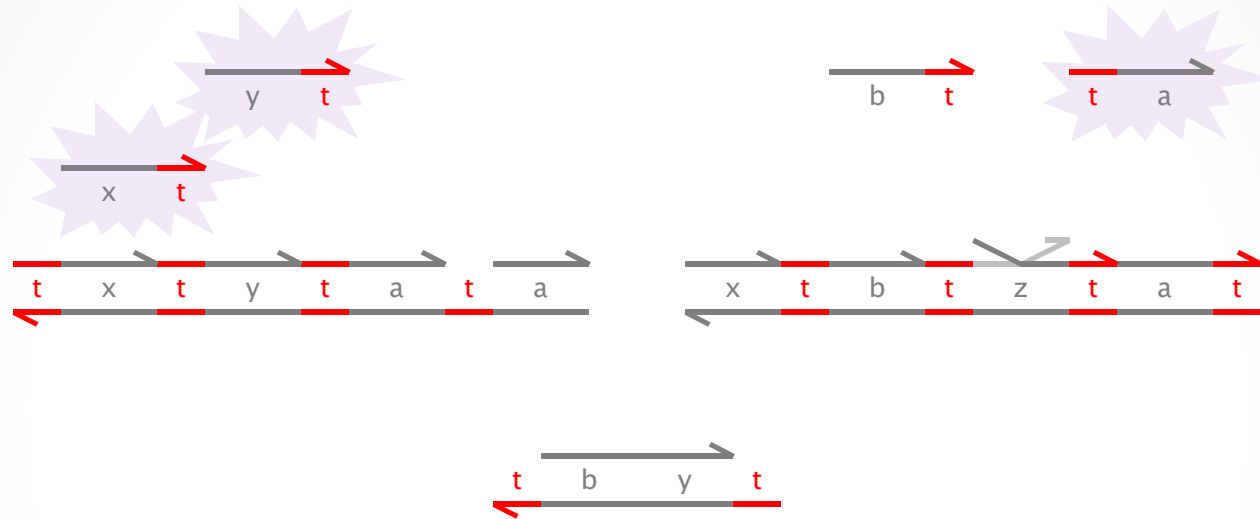
Join $x+y \rightarrow z$



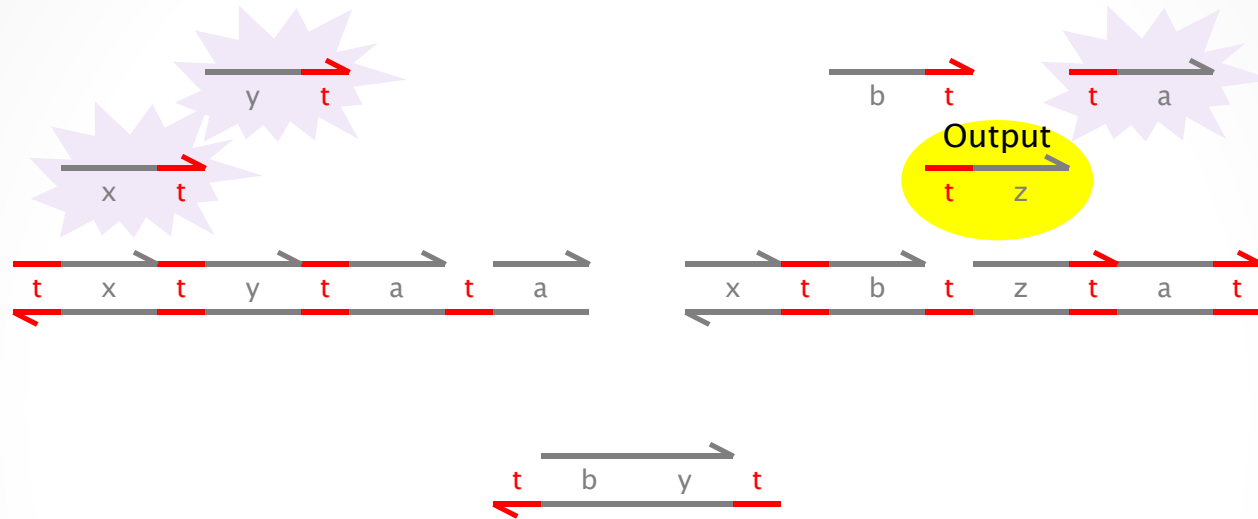
Join $x+y \rightarrow z$



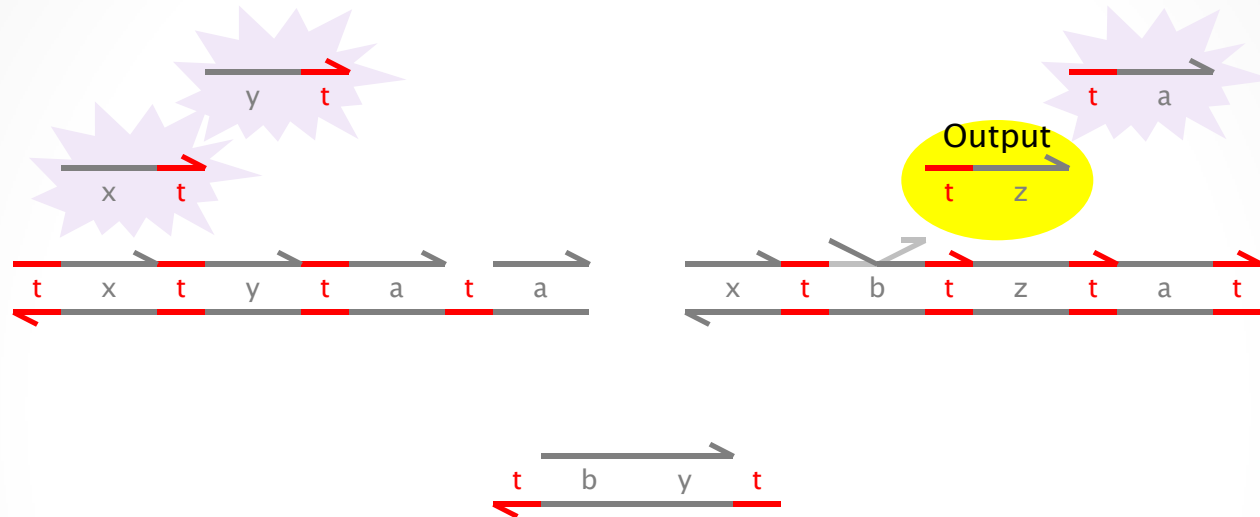
Join $x+y \rightarrow z$



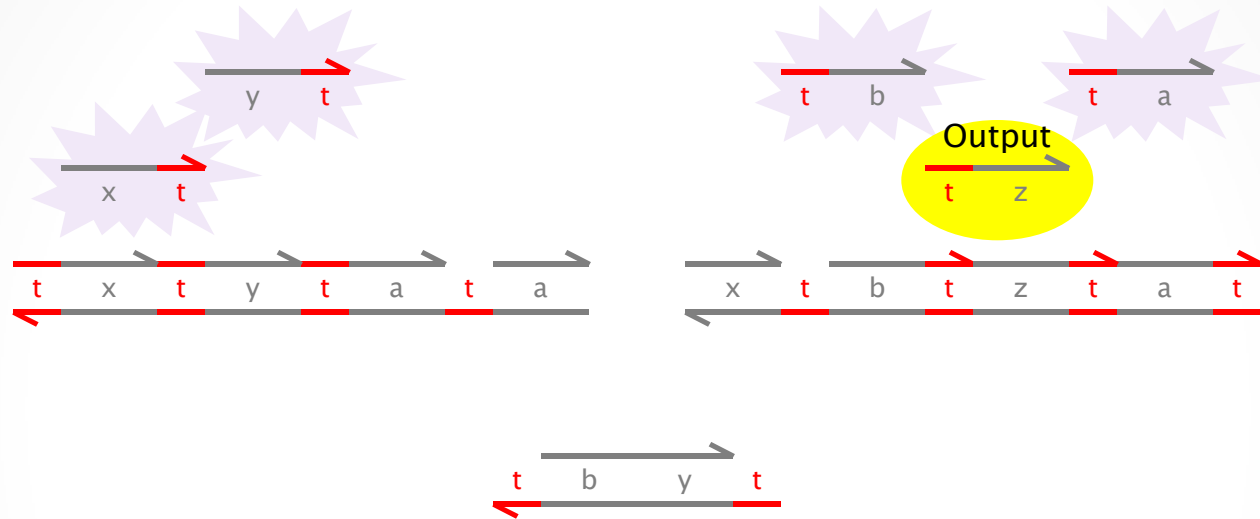
Join $x+y \rightarrow z$



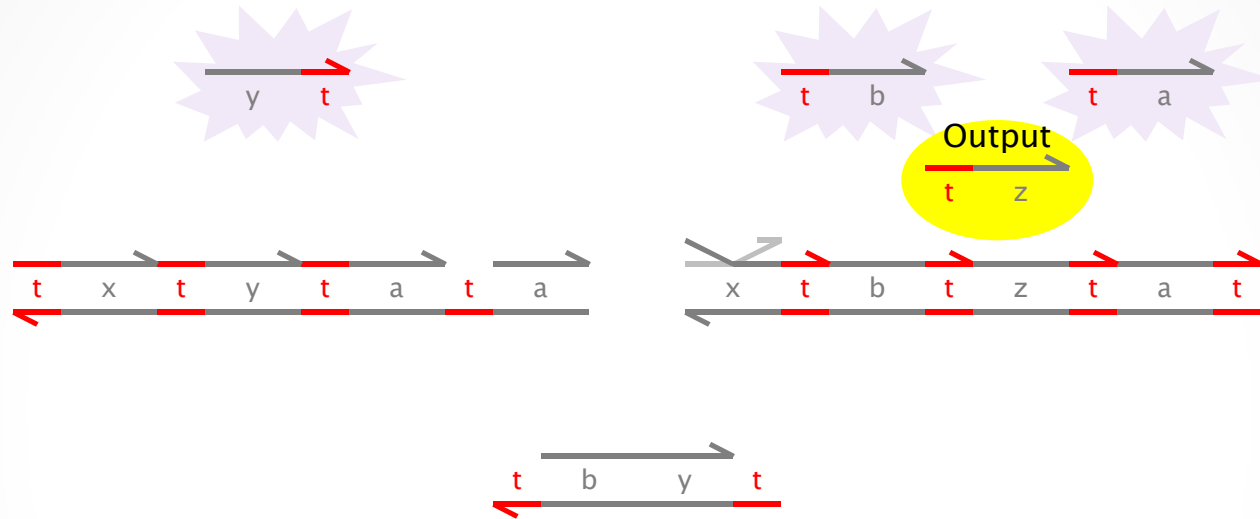
Join $x+y \rightarrow z$



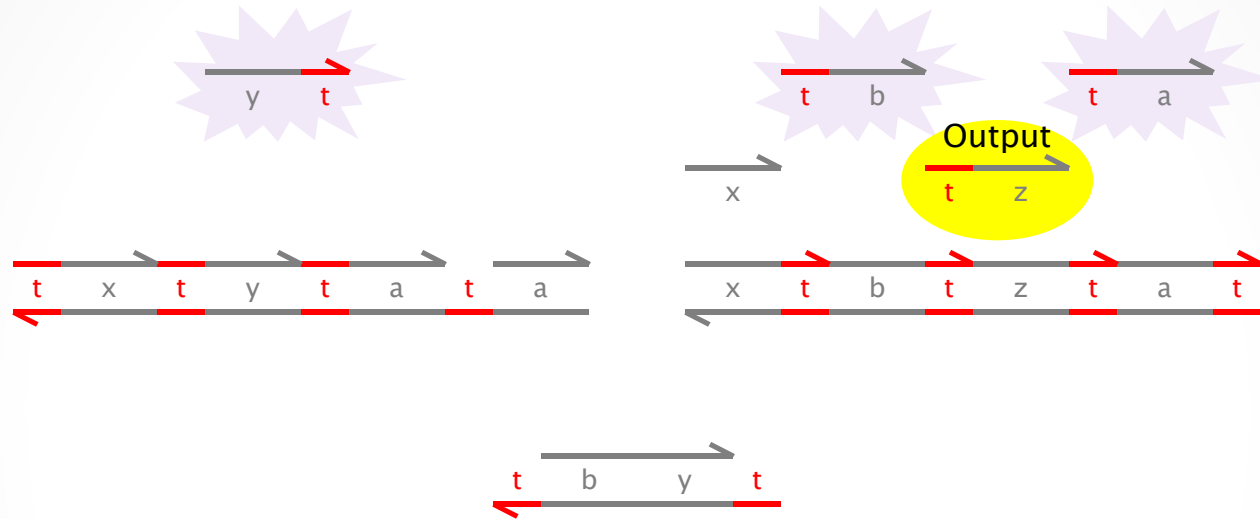
Join $x+y \rightarrow z$



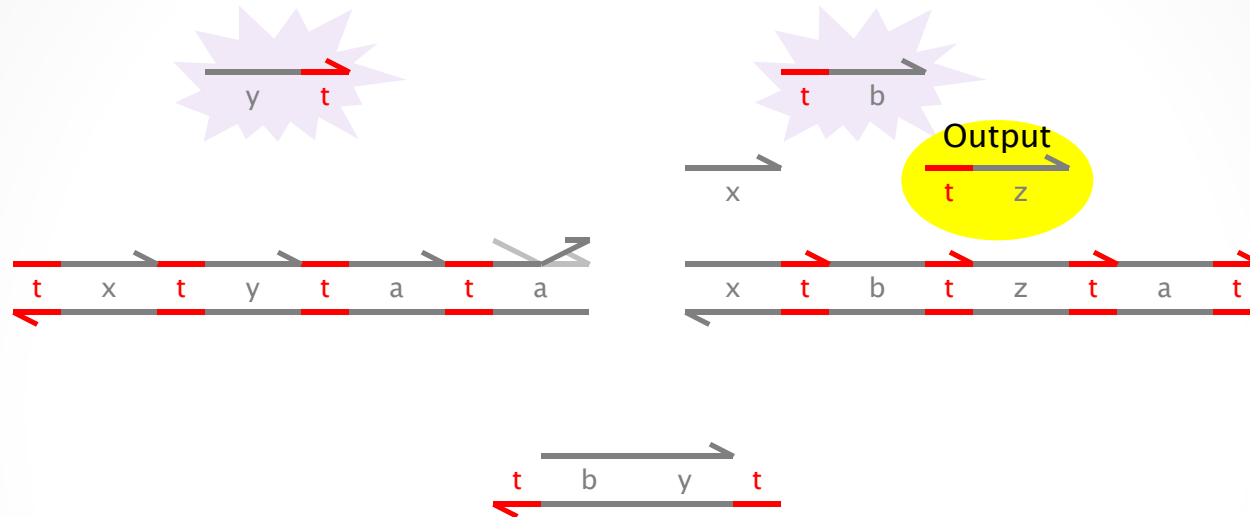
Join $x+y \rightarrow z$



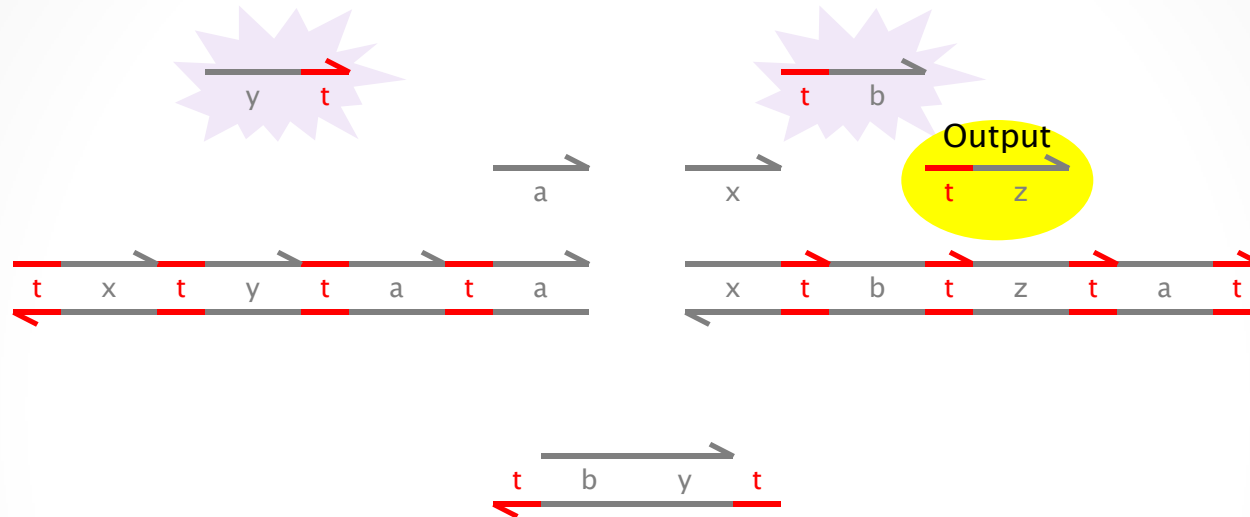
Join $x+y \rightarrow z$



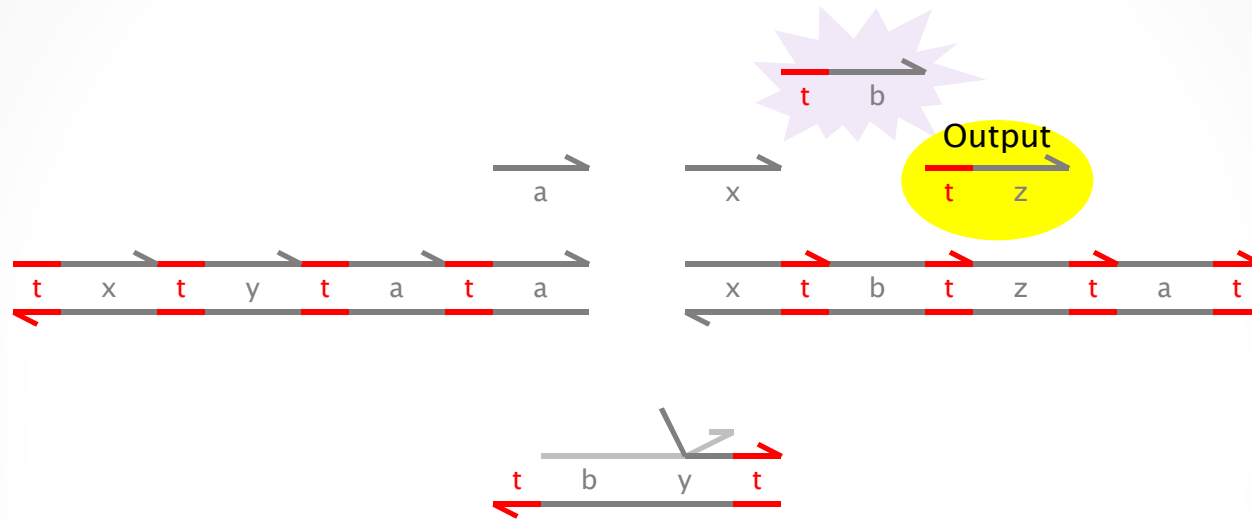
Join $x+y \rightarrow z$



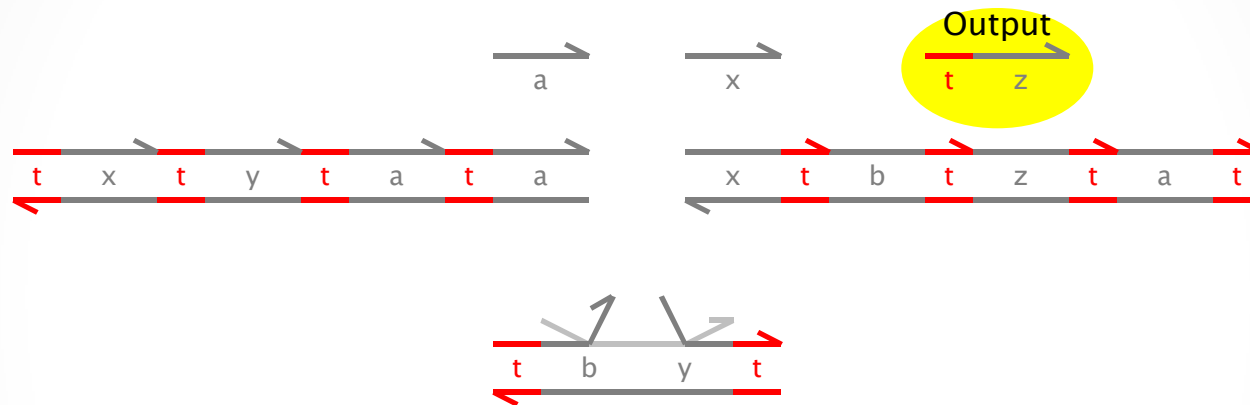
Join $x+y \rightarrow z$



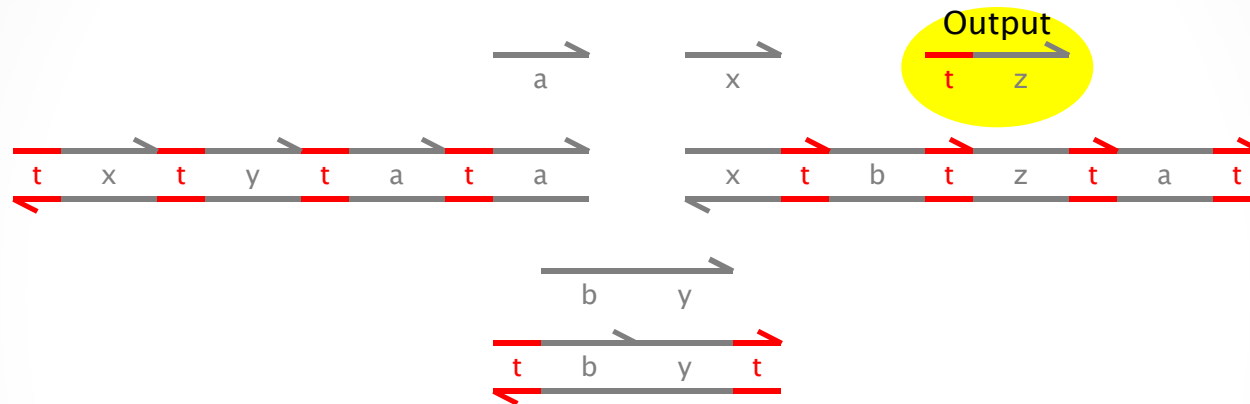
Join $x + y \rightarrow z$



Join $x+y \rightarrow z$



Join $x + y \rightarrow z$



General $n \times m$ Join-Fork

- Easily generalized to 3+ inputs (with 2+ collectors) etc.
- Easily generalized to 2+ outputs (like Fork) etc.

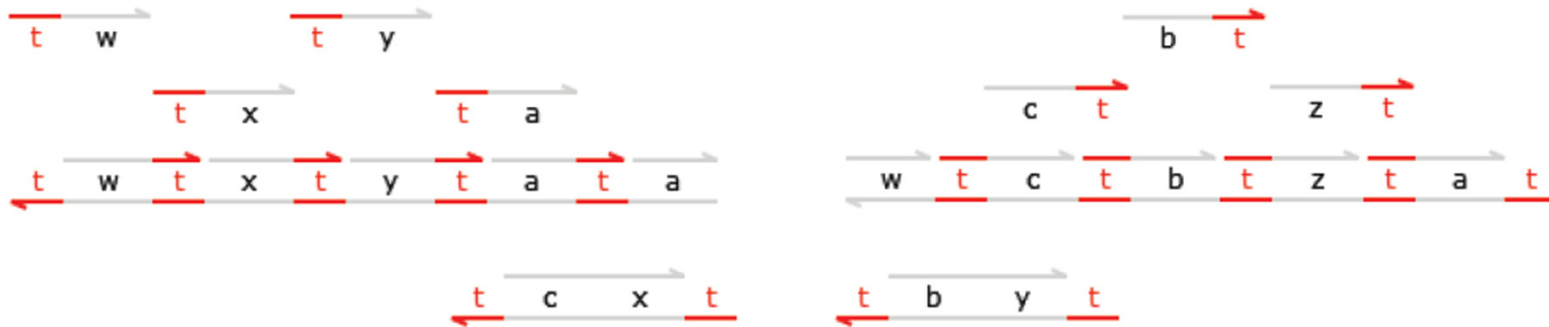


Figure 9: 3-Join $J_{wxyz} \mid tw \mid tx \mid ty \rightarrow tz$: initial state plus inputs tw, tx, ty .

Strand Algebra

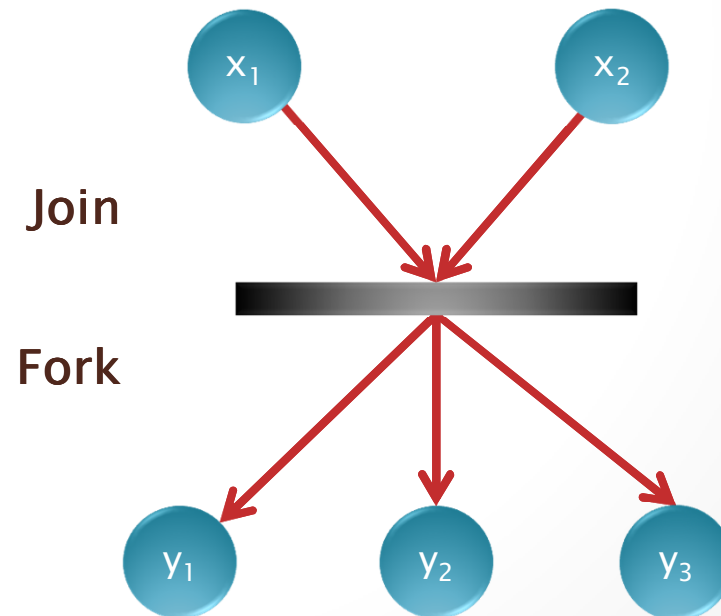
- An abstract description of signal–gate interactions:

$$x_1 \mid \dots \mid x_n \mid [x_1, \dots, x_n]. [y_1, \dots, y_m] \rightarrow y_1 \mid \dots \mid y_m$$

- Strand Algebra is an ‘intermediate language’
 - Four–three–two domain gates implement Strand Algebra.
 - Strand Algebra implements Boolean circuits, Petri Nets, FSA, Linear I/O Systems, Interacting Automata, etc.
- Two–domain gates implement Strand Algebra
 - N.B. this is a *conjecture*.

Petri Net Transitions

- Computing power equivalent to Petri Nets (not Turing complete).
- Not completely trivial: gates are consumed by activation, hence a persistent Petri net transition requires a stable population of gates.



Verification

...

Verification Issues

- Individual Components

- Reversible reactions (infinite traces)
- Interferences (deadlocks etc.) between copies of the same gate
- Interferences (deadlocks etc.) between copies of different gates
- Removal of active byproducts (garbage collection) is tricky

- Populations

- Gates come in (large) populations
- Each population *shares private domains* (technologically unavoidable)
- Correctness of populations means proofs with large state spaces
- Proofs about *arbitrary* population size?

- Environment

- The nano-environment is stochastic (noise, failures, etc.)
- Biology is messy
- But we should at least make sure our designs are *logically correct*

Correctness

- The spec of a transducer: $T_{xy} + tx \rightarrow ty$
 - Is it true at all?
 - Is it true *possibly*, or *necessarily*, or *probabilistically (measure 1)*?
 - Is it true in the context of other *identical transducers*?
 - Is it true *in all possible contexts*?
 - Is it *(more)* true for large populations?
 - Is it true for infinite populations (continuous limit)?

Nick Algebra

...

Nick Algebra

$S ::= t.x \mid x.t$

single strand

$\underline{D} ::= \emptyset \mid \underline{t} \mid \underline{x} \mid \underline{t.x} \mid \underline{x.t} \mid \underline{x.x} \mid \underline{D^+D}$

double strand

$U ::= S \mid \underline{D} \mid U|U \mid (vx)U$

soup

S



D



Algebraic Equality

$=$ is an equivalence relation,
and a congruence over the term syntax

$$\underline{D}_1 \dagger (\underline{D}_2 \dagger \underline{D}_3) = (\underline{D}_1 \dagger \underline{D}_2) \dagger \underline{D}_3$$

$$\emptyset \dagger \underline{D} = \underline{D} \dagger \emptyset = \underline{D}$$

$$U_1 | (U_2 | U_3) = (U_1 | U_2) | U_3$$

$$U_1 | U_2 = U_2 | U_1$$

$$\emptyset | U = U | \emptyset = U$$

$$(v x) U = (v y) (U \{y/x\}) \quad \text{if } y \notin \text{pd}(U)$$

$$(v x) \emptyset = \emptyset$$

$$(v x) (U_1 | U_2) = U_1 | (v x) U_2 \quad \text{if } x \notin \text{pd}(U_1)$$

$$(v x) (v y) U = (v y) (v x) U$$

Reduction

$$\underline{D}_1 \dagger \underline{t \dagger x t \dagger D}_2 \mid tx \leftrightarrow \underline{D}_1 \dagger \underline{tx \dagger t \dagger D}_2 \mid xt$$

exchange

$$\underline{D}_1 \dagger \underline{t \dagger x \dagger D}_2 \mid tx \rightarrow \underline{D}_1 \dagger \underline{tx \dagger D}_2$$

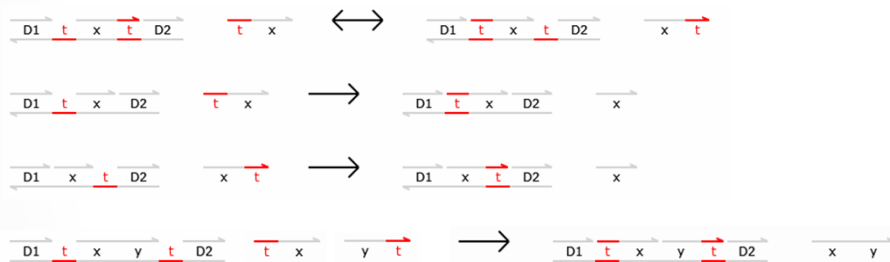
left coverage

$$\underline{D}_1 \dagger \underline{x \dagger t \dagger D}_2 \mid xt \rightarrow \underline{D}_1 \dagger \underline{xt \dagger D}_2$$

right coverage

$$\underline{D}_1 \dagger \underline{t \dagger xy \dagger t \dagger D}_2 \mid tx \mid yt \rightarrow \underline{D}_1 \dagger \underline{tx \dagger yt \dagger D}_2$$

cooperation



$$\underline{D} \rightarrow \emptyset \quad \text{if } \underline{D} \text{ not reactive}$$

waste

$$U_1 \rightarrow U_2 \Rightarrow U_1 \mid U \rightarrow U_2 \mid U$$

dilution

$$U_1 \rightarrow U_2 \Rightarrow (vx)U_1 \rightarrow (vx)U_2$$

isolation

$$U_1 = U_2, U_2 \rightarrow U_3, U_3 = U_4 \Rightarrow U_1 \rightarrow U_4$$

mixing

Reachability

- $U_1 \rightarrow^* U_2$ iff $U_1 \rightarrow \dots \rightarrow U_2$
 - That is, U_1 *may* reduce to U_2 .
- $U_1 \rightarrow^\forall U_2$ iff $\forall U, U_1 \rightarrow^* U \Rightarrow U \rightarrow^* U_2$
 - That is, U_1 *will* reduce to U_2 . (It cannot avoid the possibility of reducing to U_2).

Gate Definitions

- $T_{xay} = \underline{t^\dagger x t^\dagger a t^\dagger a} \mid ta \mid \underline{x^\dagger t y^\dagger t a^\dagger t} \mid yt$
- $T_{xy}^n = (va)((T_{xay})^n)$

- $F_{xayz} = \dots$
- $F_{xyz}^n = (va)((F_{xayz})^n)$

- $J_{xyaz} = \dots$
- $J_{xyz}^n = (va)((J_{xyaz})^n)$

Correctness

- Proposition: May-Correctness

$$T_{xy}^n | tx^n \rightarrow^* ty^n$$

$$F_{xyz}^n | tx^n \rightarrow^* ty^n | tz^n$$

$$J_{xyz}^n | tx^n | ty^n \rightarrow^* tz^n$$

- Easy case analysis and induction on n.

- Proposition: T_{xy}^1 Will-Correctness

$$T_{xy}^1 | tx \rightarrow^{\forall} ty$$

- Exhaustive case analysis enumerating all states of the system.
- Can be done by hand for T_{xy}^1 , and maybe T_{xy}^2 , but not really for T_{xy}^3 etc.
- Will-correctness for fork/join is harder (more states).
- Will-correctness for combinations of gates is harder (does not compose and requires analysis of joint state space).
- We are using modelchecking to verify some of these properties.
[Andrew Phillips & David Parker in PRISM]

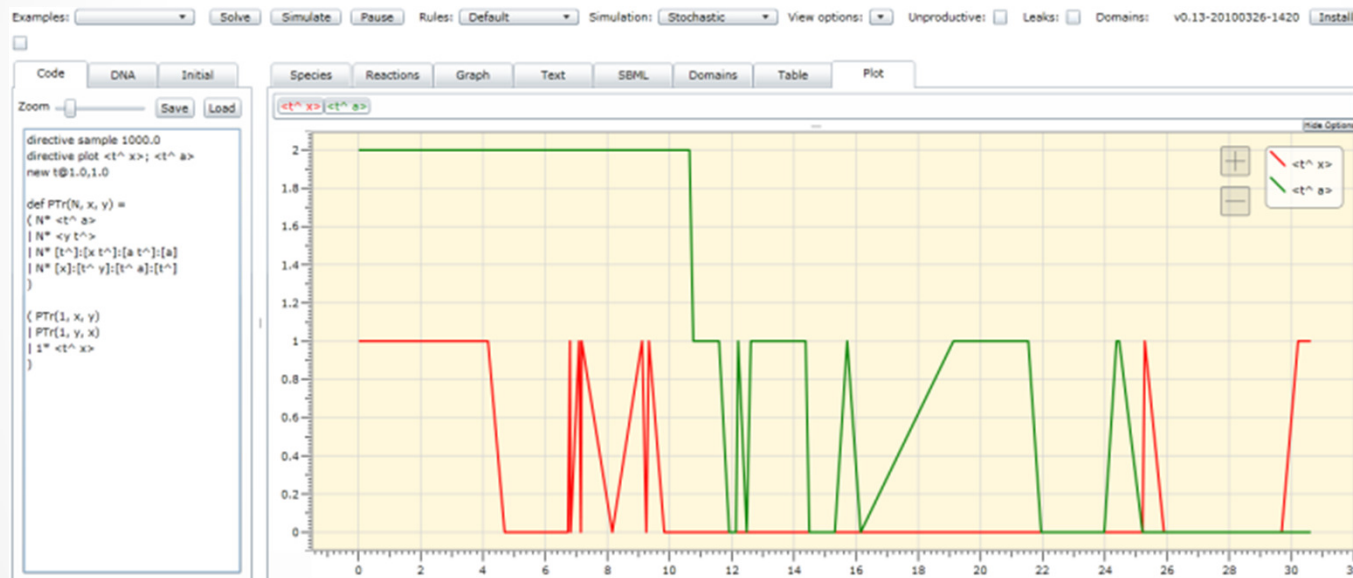
T^1_{xy} Will-Correctness

01. $(\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid y \neg}{\neg x}$
02. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid y \neg}{x \neg}$
03. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid y \neg}{x \neg} \mid a \neg$
04. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid y \neg}{x \neg} \mid \neg a$
05. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid y \neg}{x \neg}$
06. $\rightarrow (\forall a) \frac{x \neg y \neg a \neg}{y \neg} \mid x \neg$
07. $\leftrightarrow (\forall a) \frac{x \neg y \neg a \neg}{x \neg} \mid \neg y$
08. $\leftrightarrow (\forall a) \frac{x \neg y \neg a \neg}{\neg y}$
09. $\rightarrow \neg y$
10. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid x \neg}{\neg y} \rightarrow 07$
11. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid \neg y}{\neg y} \rightarrow 08$
12. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{\neg y}}{\neg y} \rightarrow 09$
13. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid x \neg}{\neg a} \mid \neg y \leftrightarrow 10$
14. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{x \neg y \neg a} \mid \neg a}{\neg y} \leftrightarrow 11$
15. $\leftrightarrow (\forall a) \frac{\frac{\frac{\neg x \neg a \neg a}{\neg a} \mid \neg a}{\neg a}}{\neg y} \leftrightarrow 12$

Interfering Transducers

• $T_{xay} \mid T_{yax}$ *sharing the same 'a'*

○ $T^1_{xay} \mid T^1_{yax} \mid tx^1$ **Correct Run**



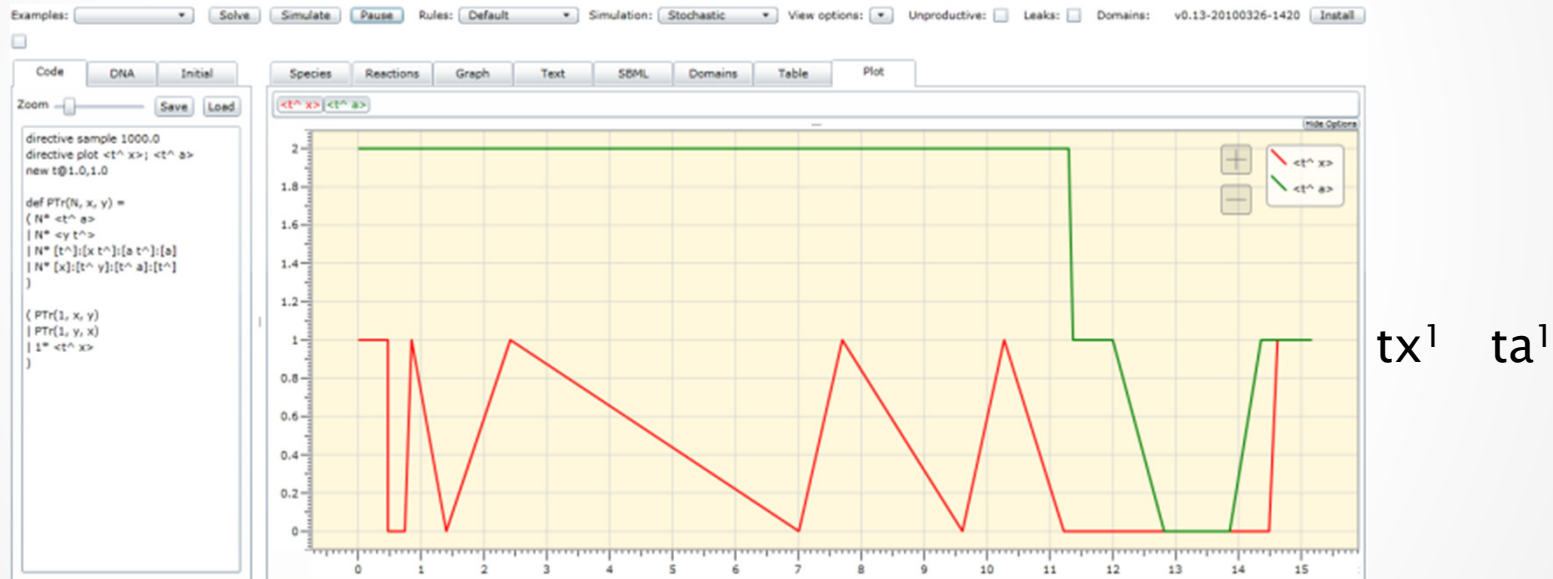
tx^1

ta^0

Interfering Transducers

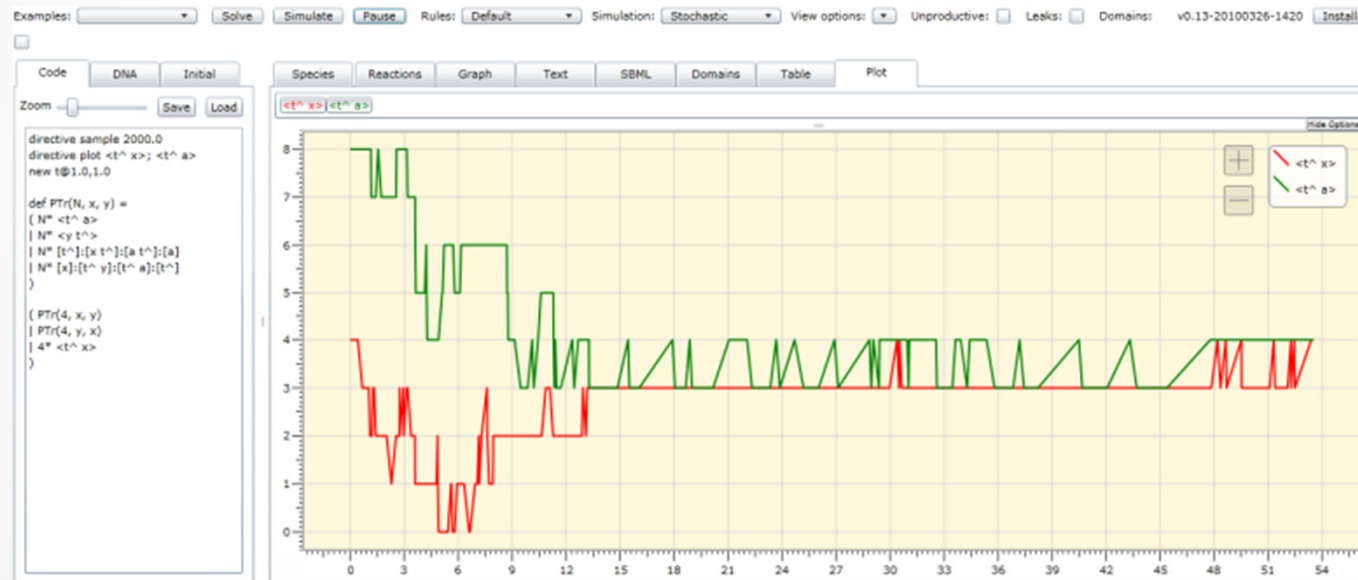
- $T_{xay} \mid T_{yax}$ *sharing the same 'a'*

- $T^1_{xay} \mid T^1_{yax} \mid tx^1$ **Incorrect Run**



Interfering Transducers

- $T_{xay} \mid T_{yax}$ *sharing the same 'a'*
 - $T^4_{xay} \mid T^4_{yax} \mid tx^4$ **4 copies all 'badly interfering' (rare case)**



tx^4 ta^4

Interfering Transducers

- $T_{xay} \mid T_{yax}$ *sharing the same 'a'*

- $T^{100}_{xay} \mid T^{100}_{yax} \mid tx^{100}$ **100 copies**



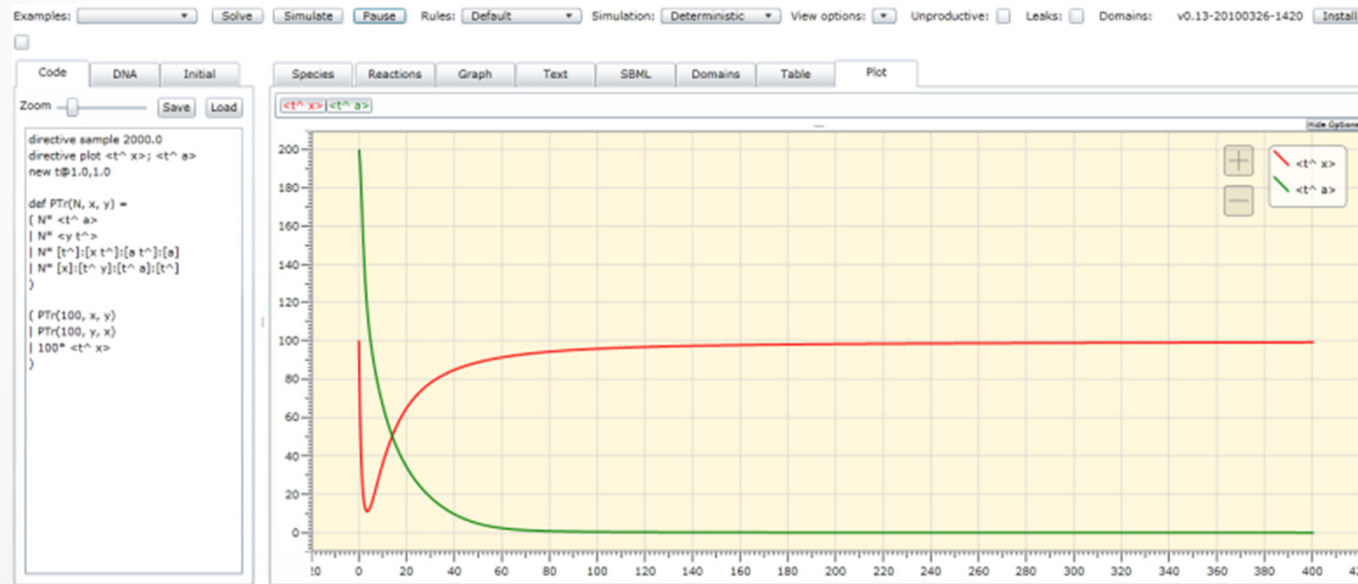
tx^{100}

ta^0

Interfering Transducers

- $T_{xay} \mid T_{yax}$ *sharing the same 'a'*

- $T^{100}_{xay} \mid T^{100}_{yax} \mid tx^{100}$ **continuous limit (ODE simulation)**



Interfering Transducers

- Although $T_{xay} \mid T_{yax} \mid tx \not\rightarrow^{\forall} tx$
- We have $T_{xay} \mid T_{yax} \mid tx \mid ty \rightarrow^{\forall} tx \mid ty$
- That means that a large population of such gates in practice does not deadlock easily: each pair of deadlocked gates can be unblocked by another pair correctly producing a ty as an intermediate product.
- **Wisdom of the masses**: individuals can be wrong, but the population is right. It is very unlikely that a significant fraction of gates ends up being deadlocked.

Conclusions

- A new architecture for general DNA gates
 - Simple signals, simple gate structures.
 - Self-cleaning: no garbage left by operation (except inert).
 - Enabling new ways of assembling gates.
 - Some experimental evidence that it works.
- A correspondingly simple algebra
 - For verifying gate designs mechanically.
- Verification issues
 - Verification techniques for gate populations.
 - Are the fork/join gates in Nick Algebra a correct implementation of (Strand Algebra and) Petri nets?