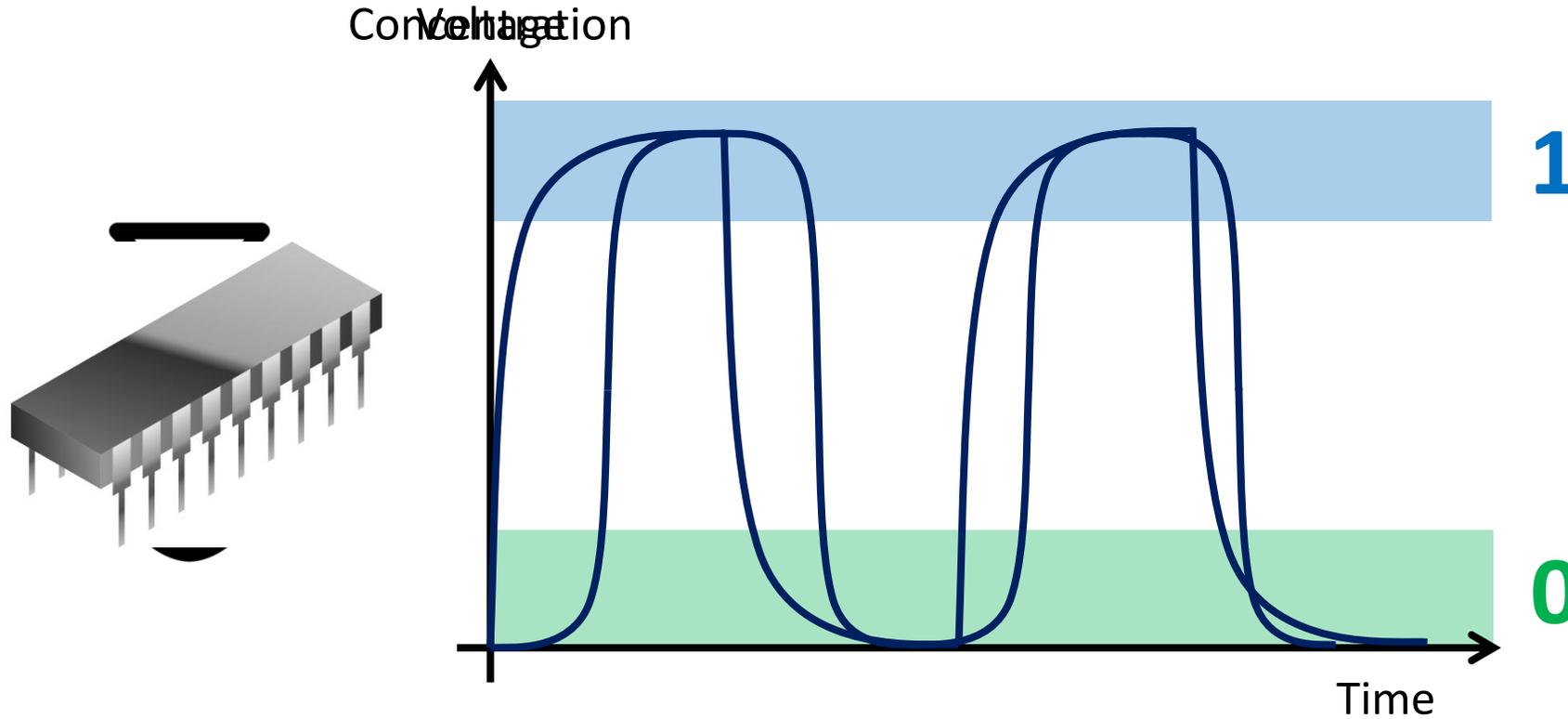


Thermodynamic binding networks

slides © 2021, David Haley and David Doty

ECS 232: Theory of Molecular Computation, UC Davis

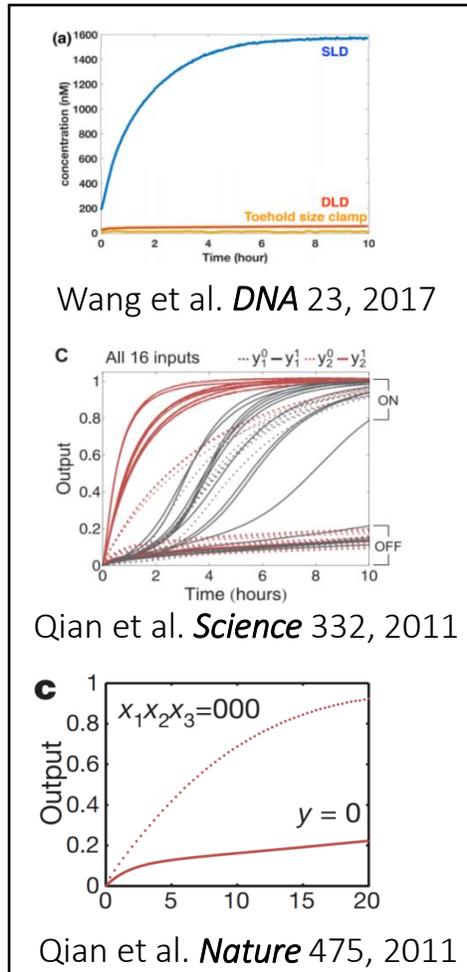
Representing Information with Molecules



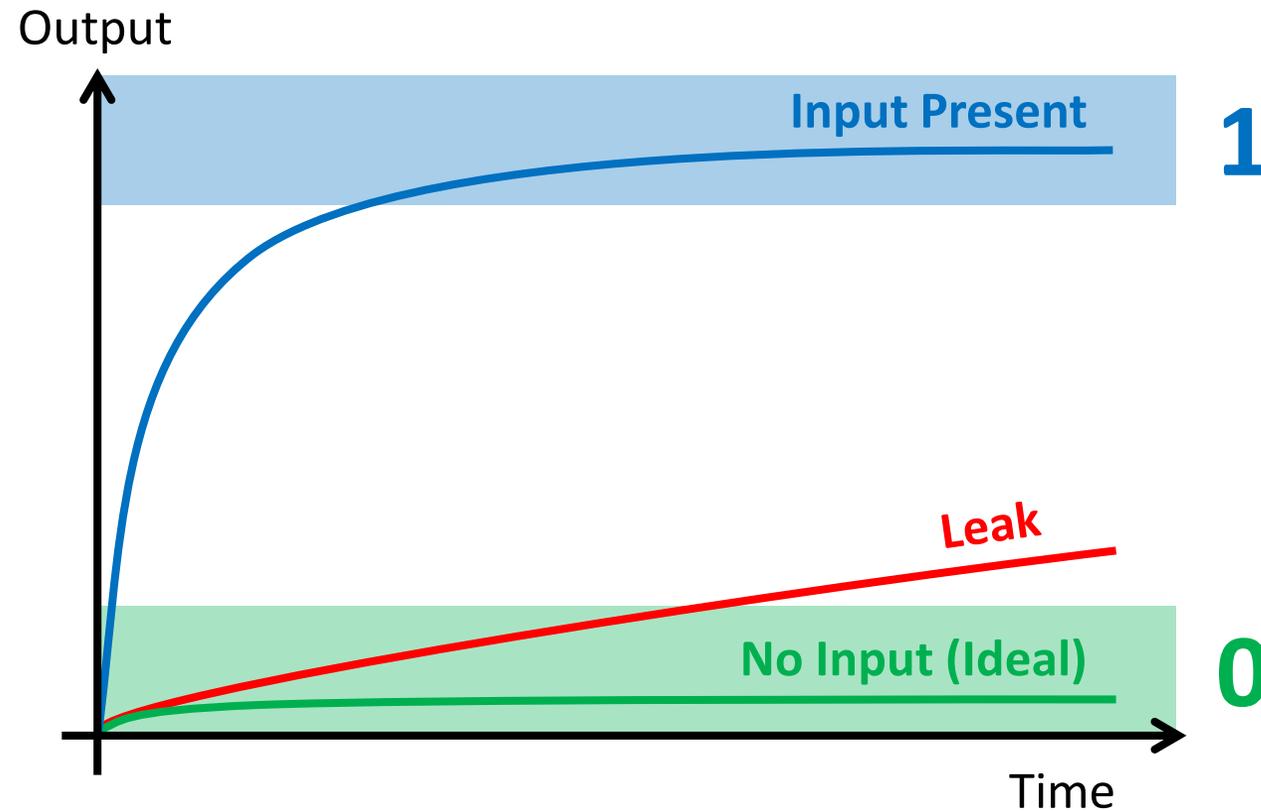
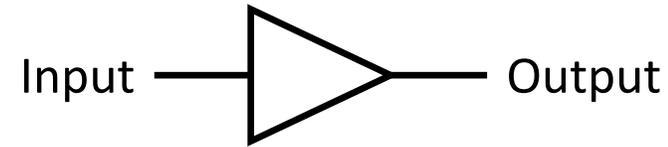
In a well-mixed solution,
concentration can represent Boolean input

Chemical Identity Gate: Idealized vs. Actual Behavior

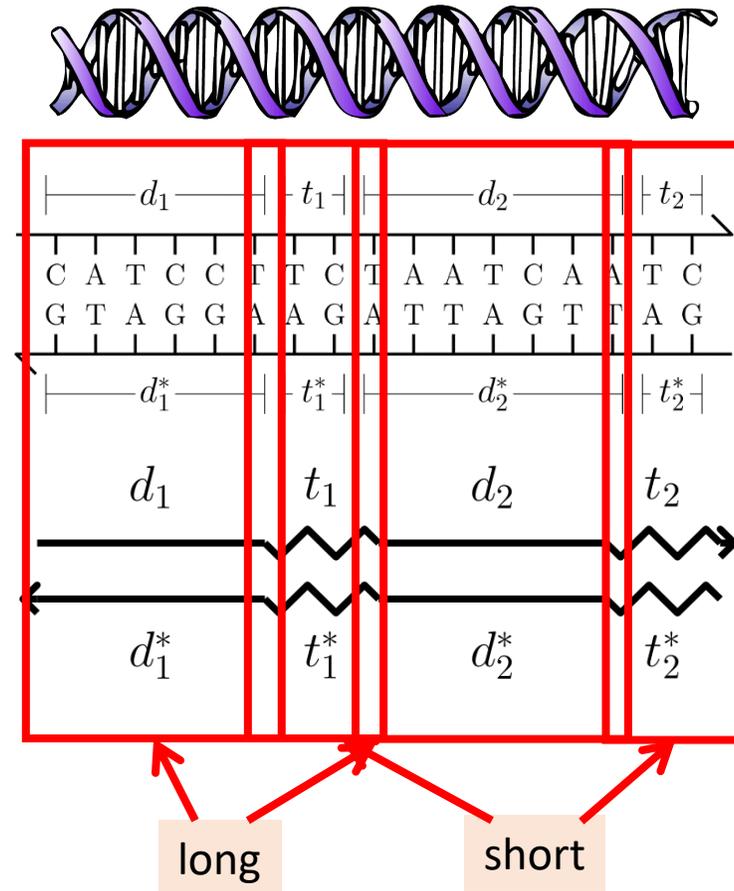
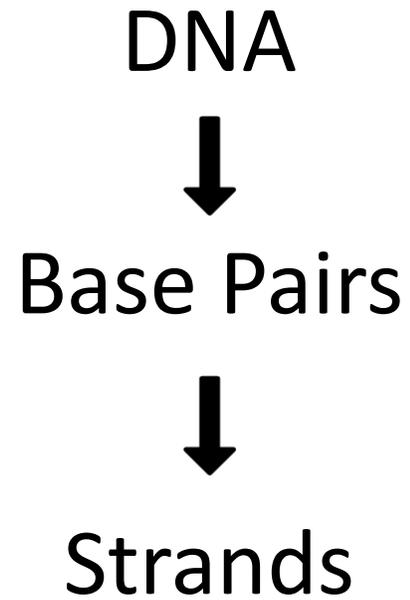
Experimental Implementation of Chemical Logic



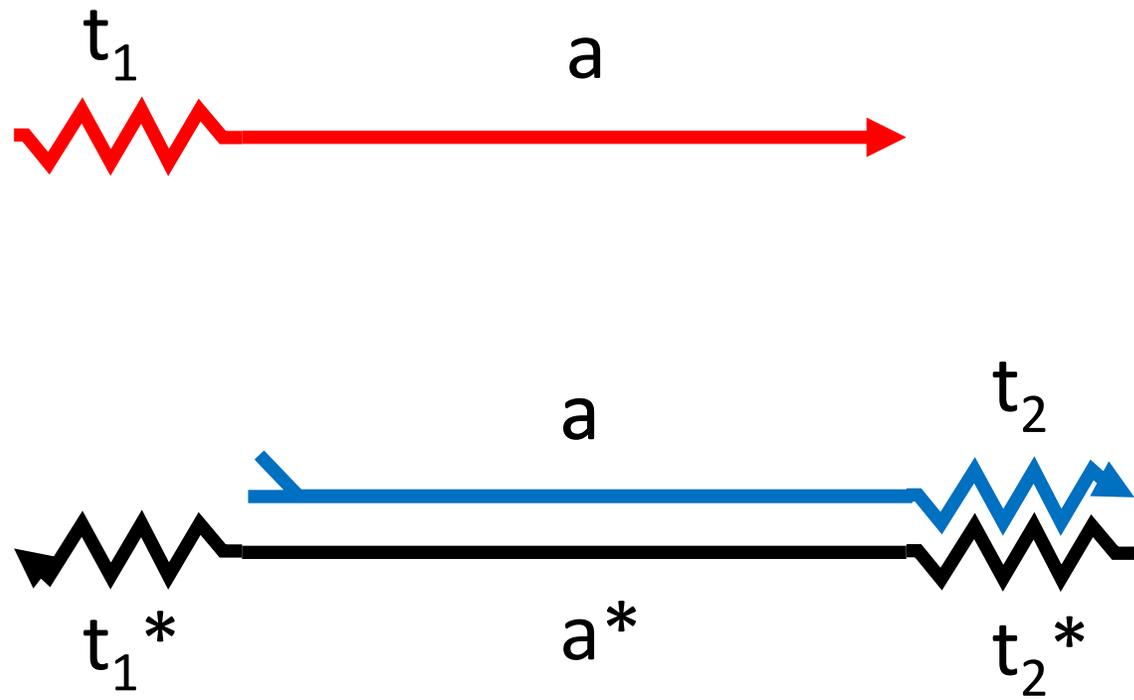
(Don't worry about the details in the pictures above)



Levels of Abstraction



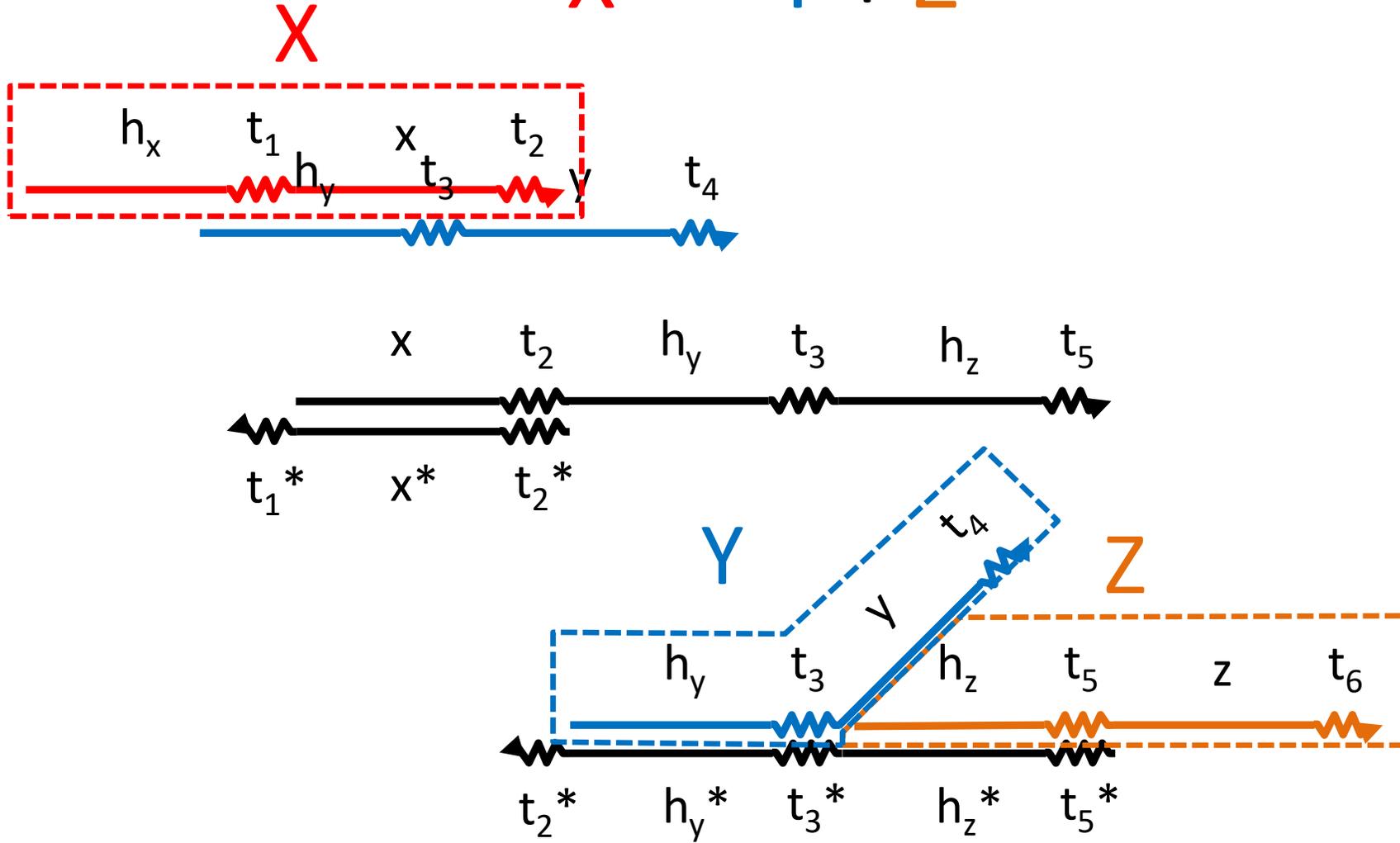
DNA strand displacement



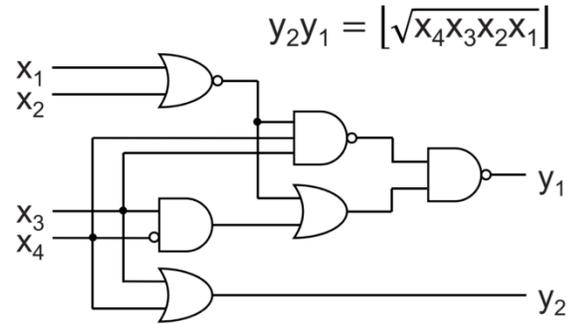
Bind

Displace

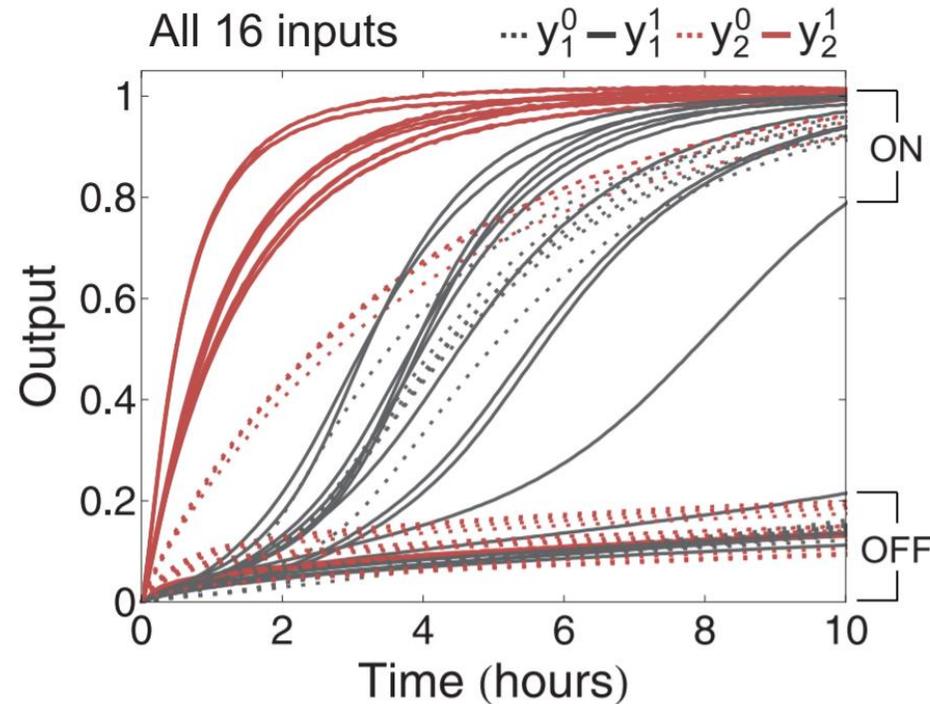
Release



Leak in strand displacement experiments

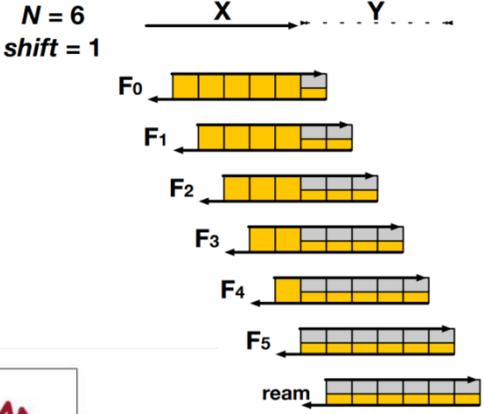


Source:
Lulu Qian, Erik Winfree.
Scaling Up Digital Circuit Computation
Science 332, 2011

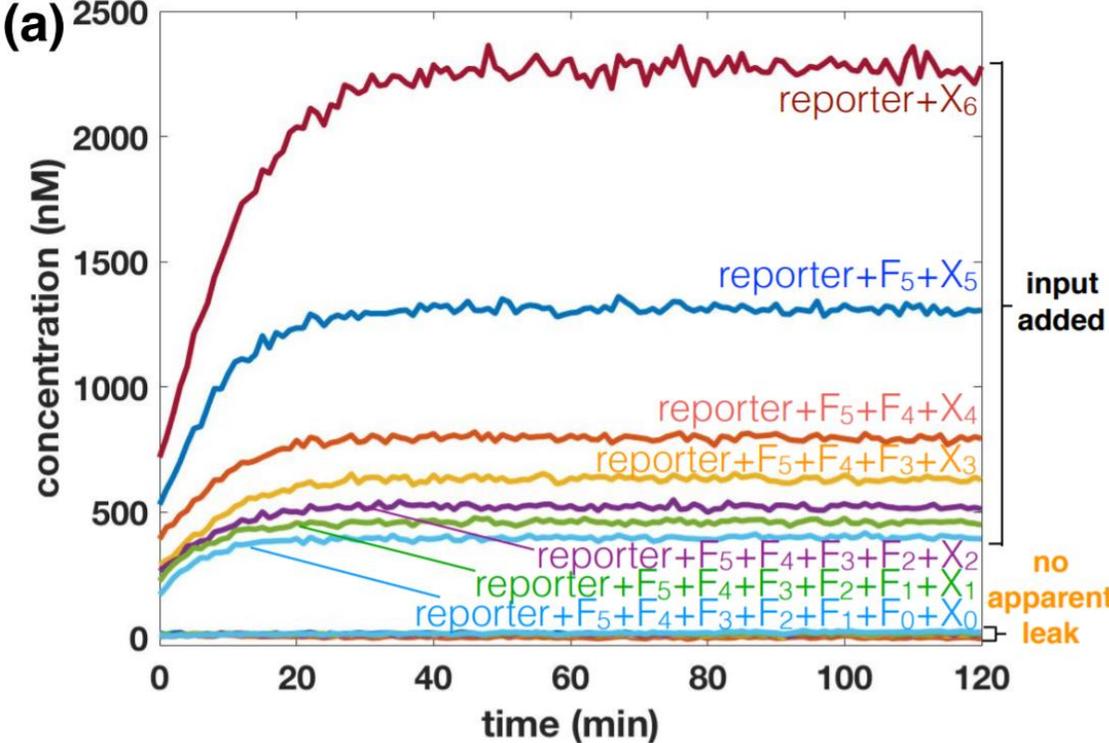


Reducing Leak

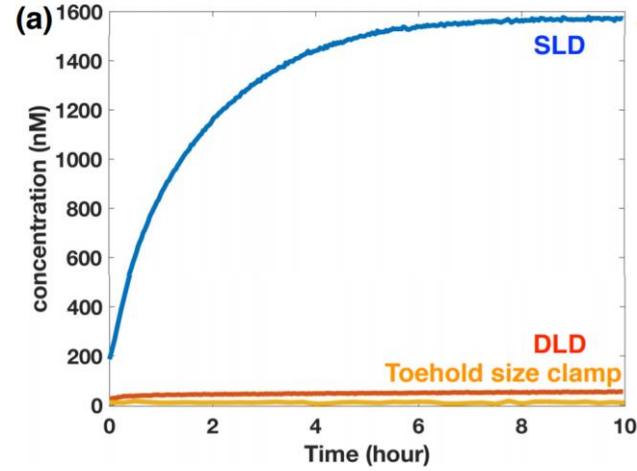
Intended:



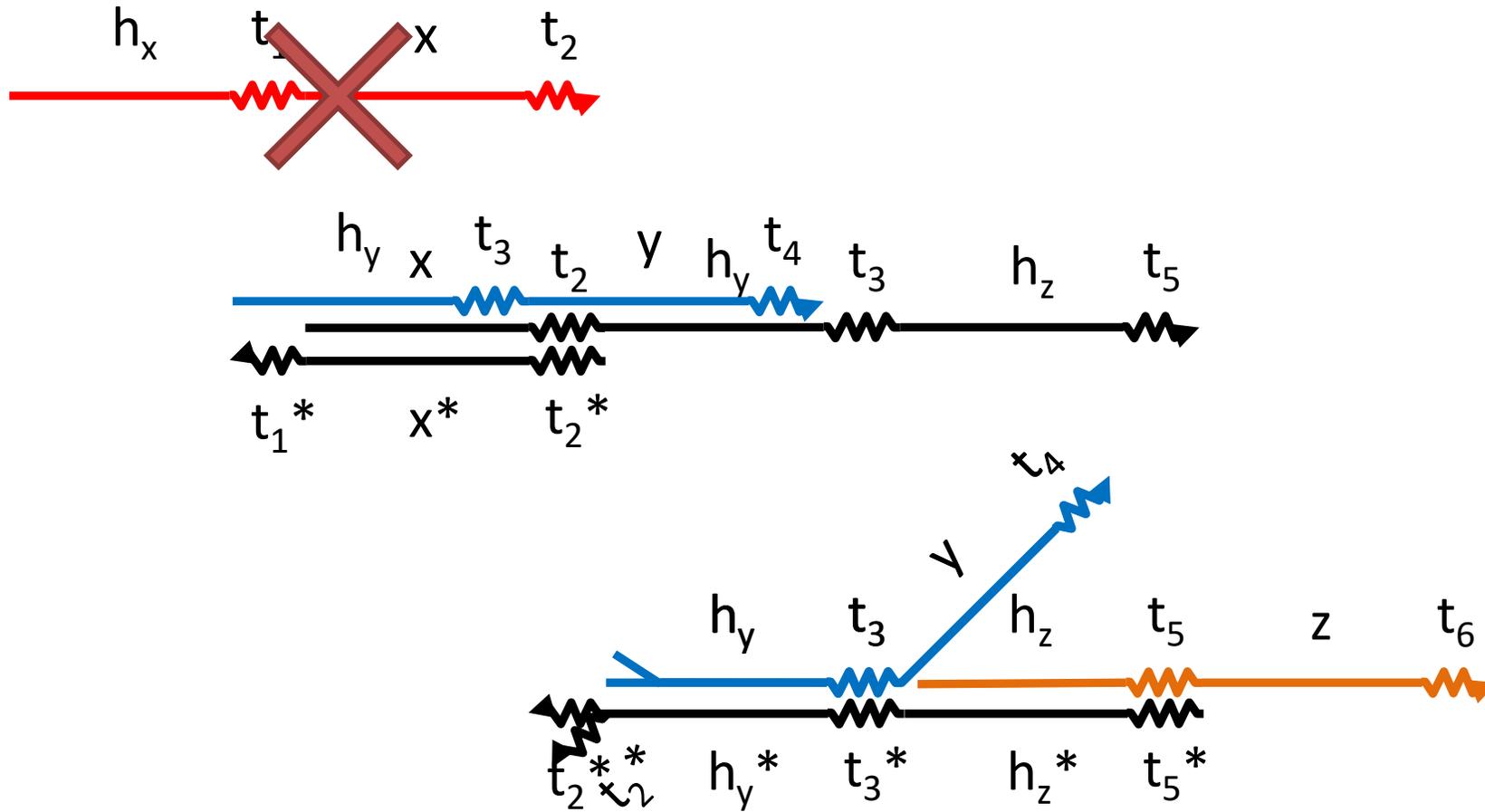
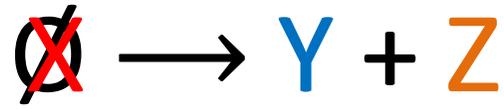
[Boya Wang, Chris Thachuk, Andrew Ellington, David Soloveichik. *The Design Space of Strand Displacement Cascades with Toehold-Size Clamps* DNA Computing Conference, 2017]



Leak:

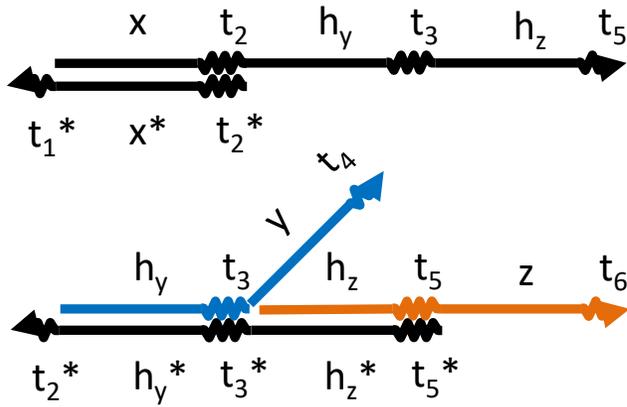


What causes leak
“kinetically”?



What causes leak “thermodynamically”?

Before:

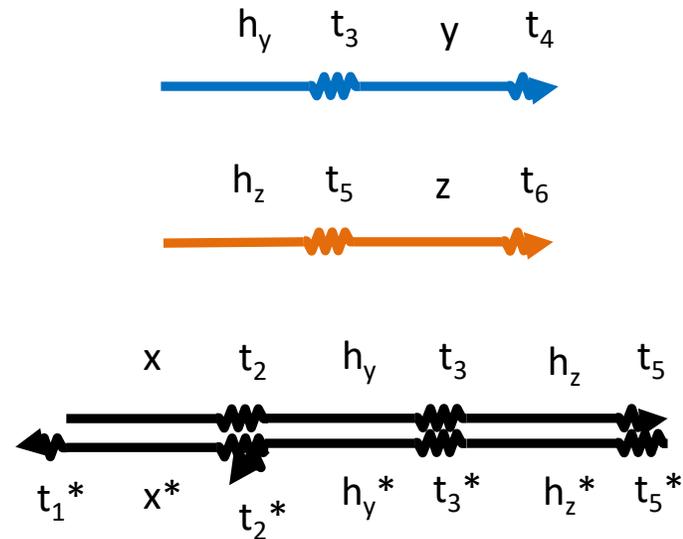


less favorable

slow

very slow

After:



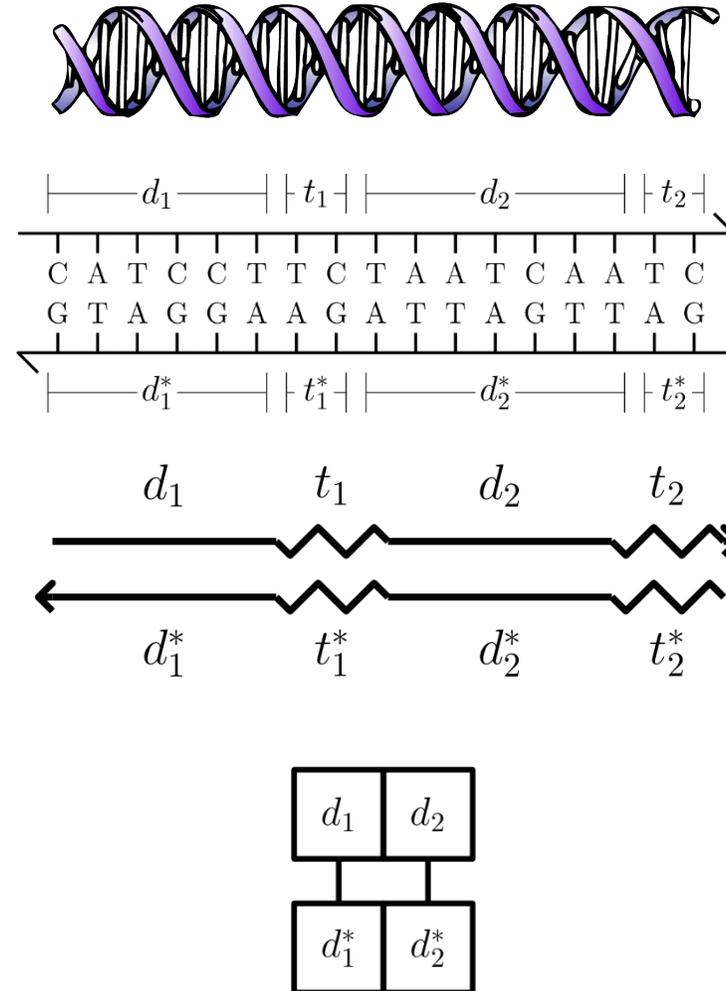
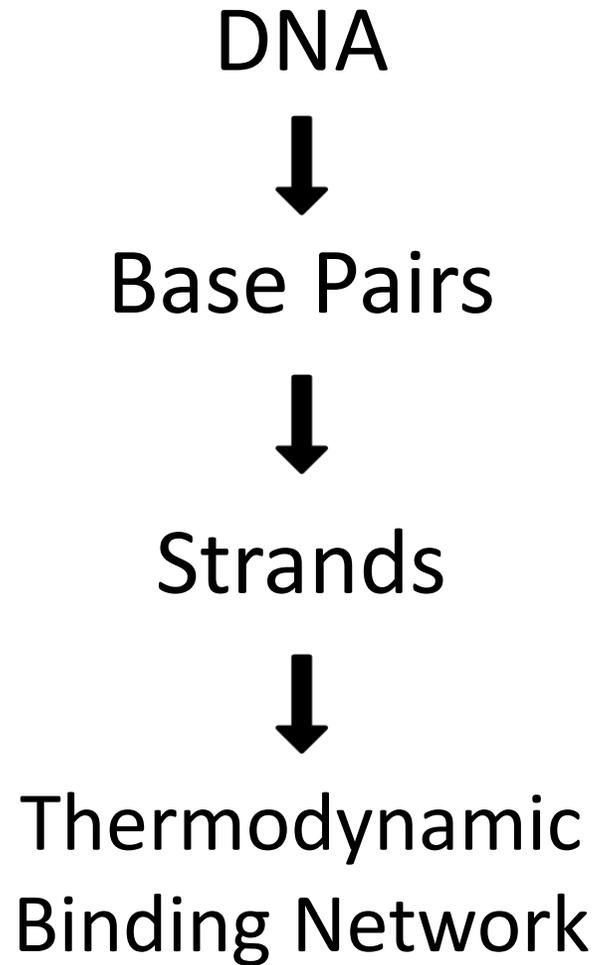
more favorable

Need a *kinetic* binding network model

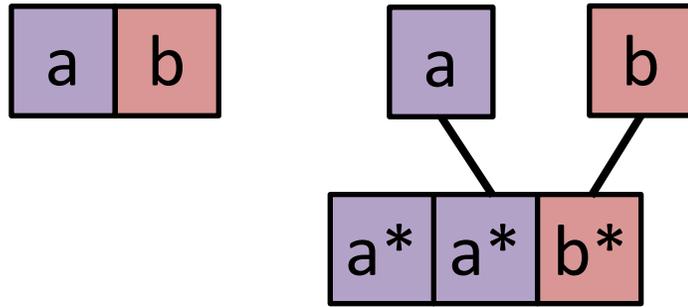
- Can we design pathways that maintain local stability?



Levels of Abstraction



Thermodynamic Binding Networks



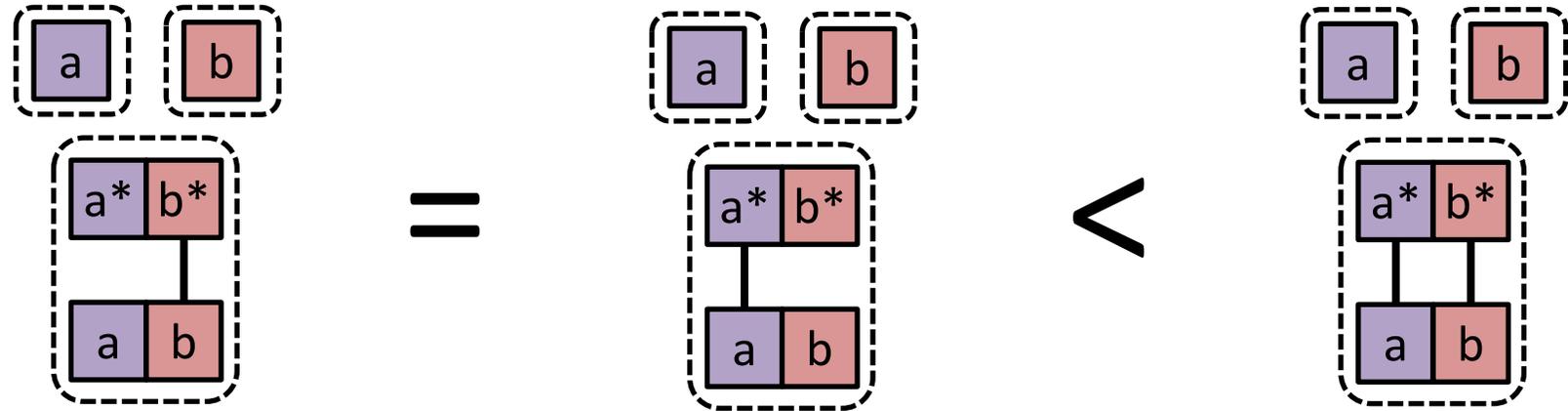
Geometry-Free Model:
The domains within a monomer are unordered

Monomer = collection of domains

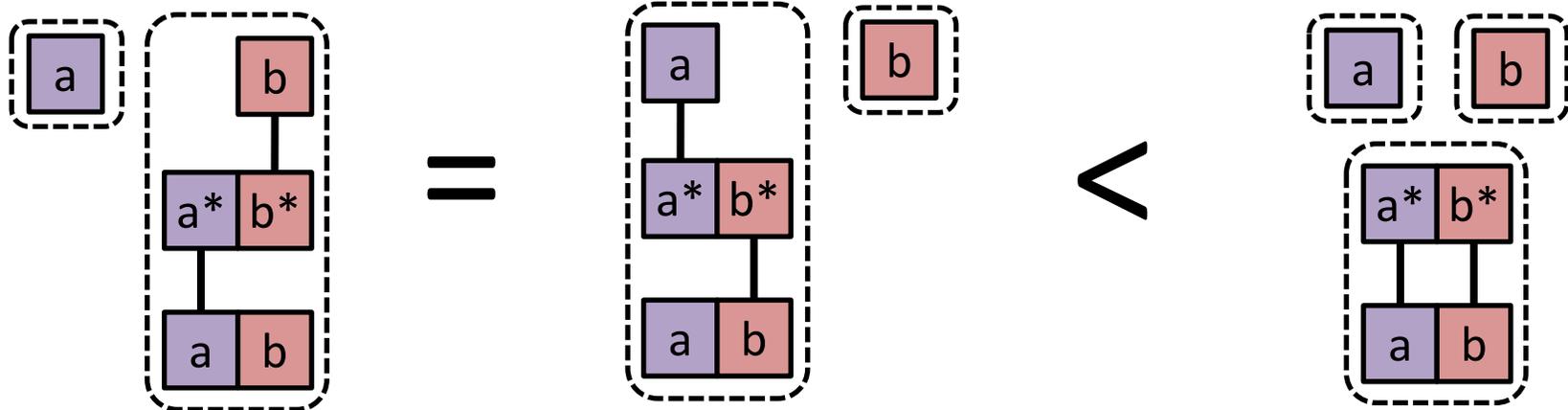
Configuration = how monomers are bound

Energetic favorability: Bonds and complexes

all else equal,
more bonds
= more favorable



all else equal,
more complexes
= more favorable



Tradeoff between #bonds and #complexes

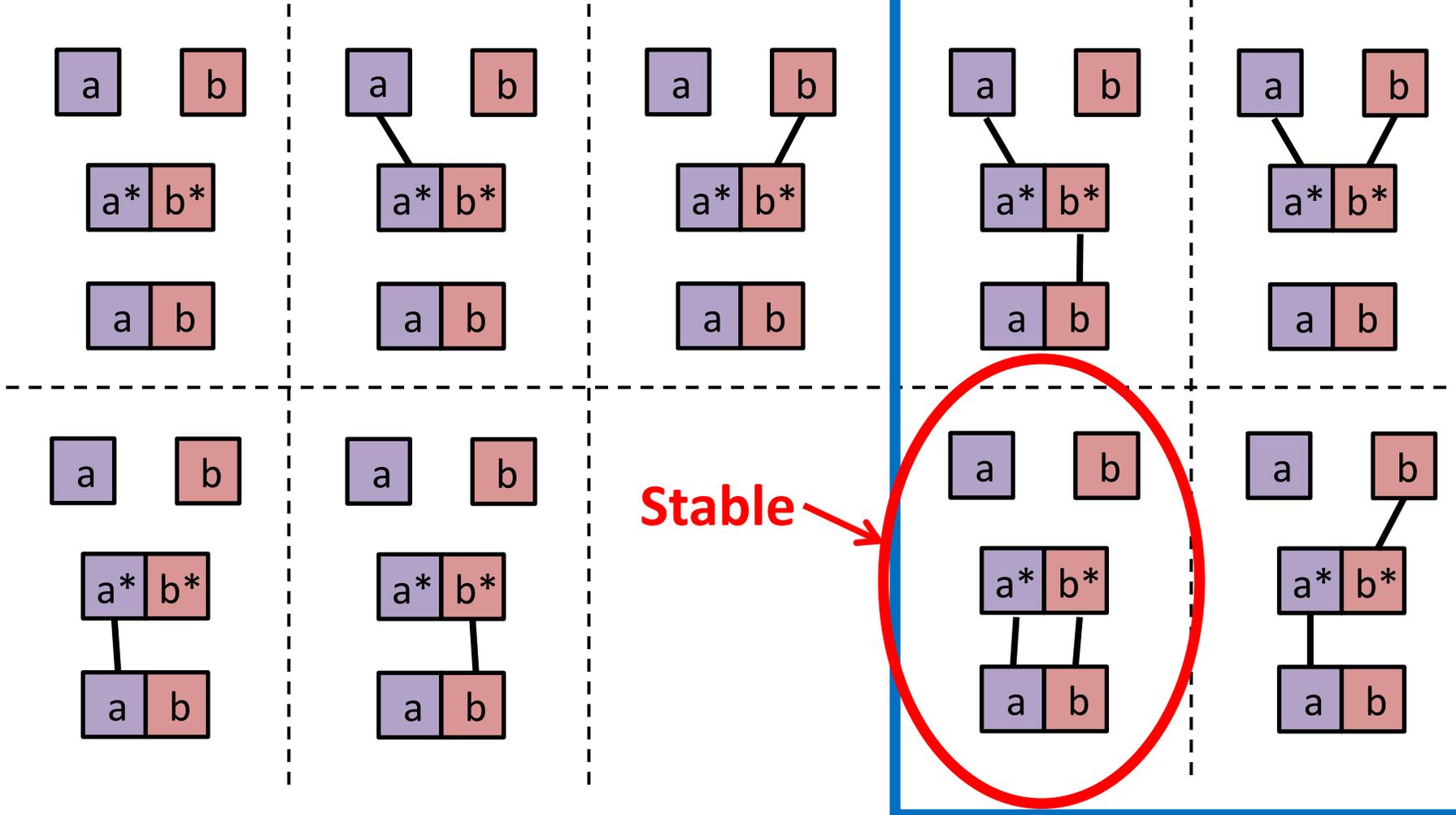
- in general, there's some weight parameter w :
energy = $w \cdot \#bonds + \#complexes$
(*physics notation: $\Delta G = \Delta H - T \cdot \Delta S$*)
- We often consider a natural limiting case:
 - favoring #bonds infinitely over #complexes
 - require maximal #bonds formed; use #complexes only as tiebreaker
 - Corresponds to bonds that are so strong they cannot spontaneously dissociate, but can exchange with each other to find configurations with more complexes

Thermodynamic Binding Networks

saturated = maximum #bonds formed

stable = saturated, AND maximum #complexes

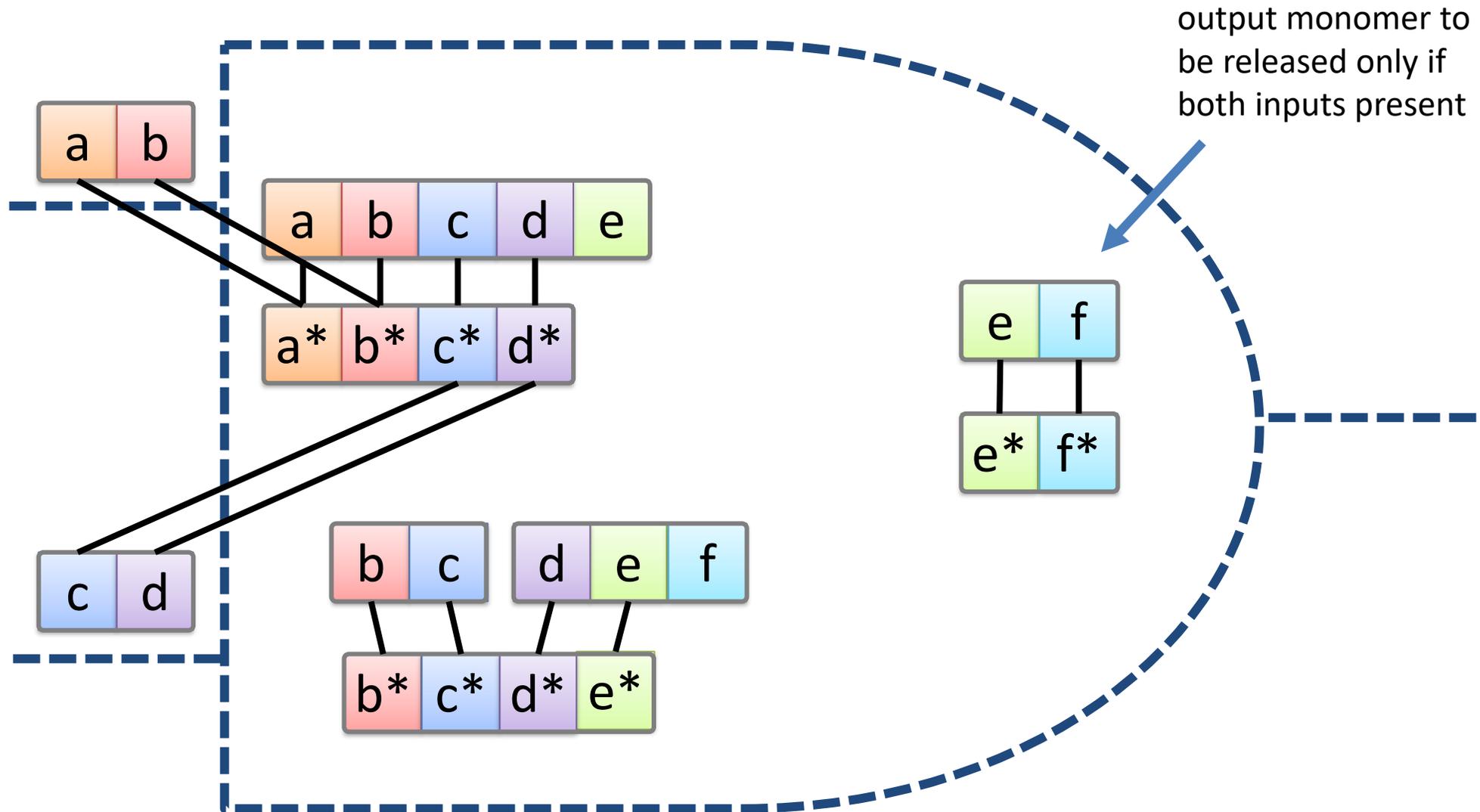
Saturated



If we're careful to make starred binding sites limiting, then *saturated = all starred sites are bound*

Computing via Thermodynamic Equilibrium

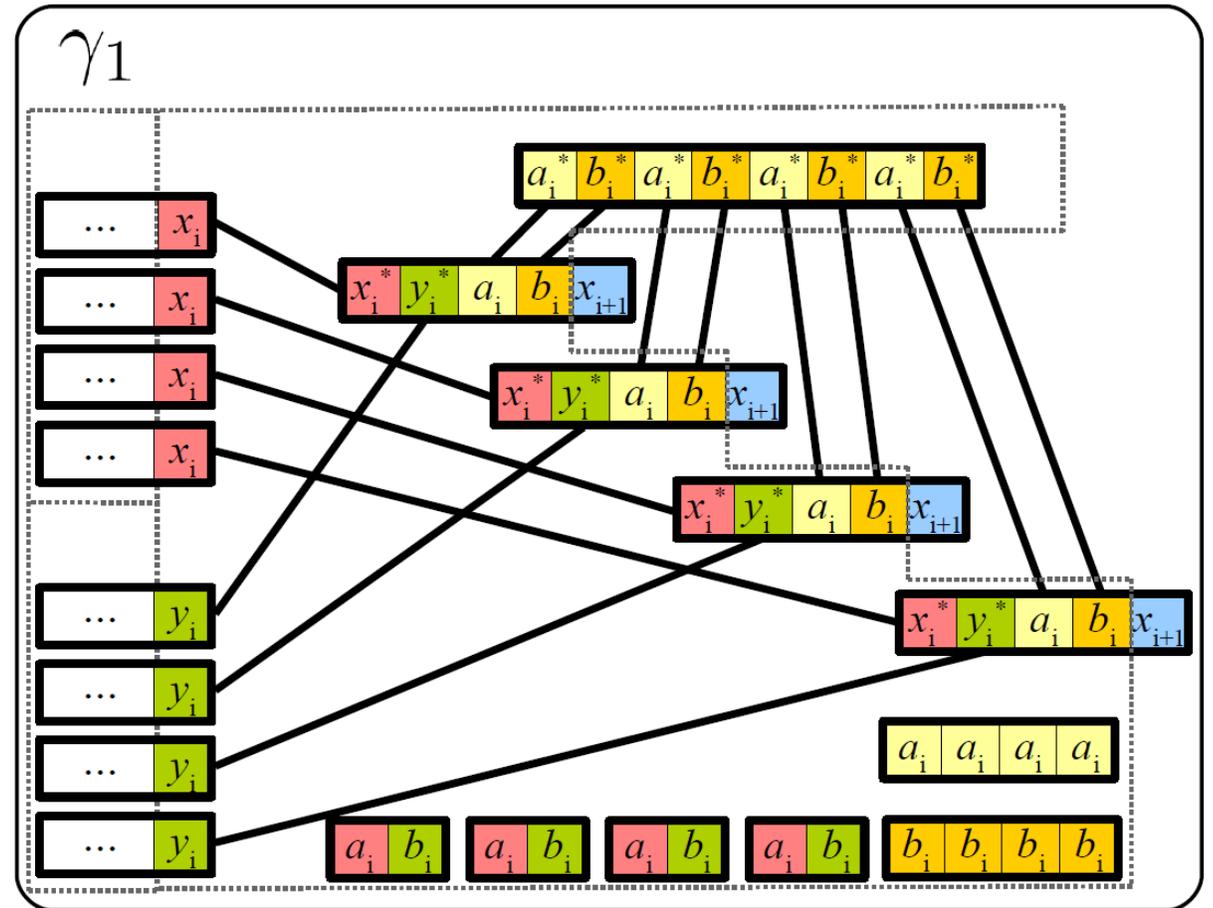
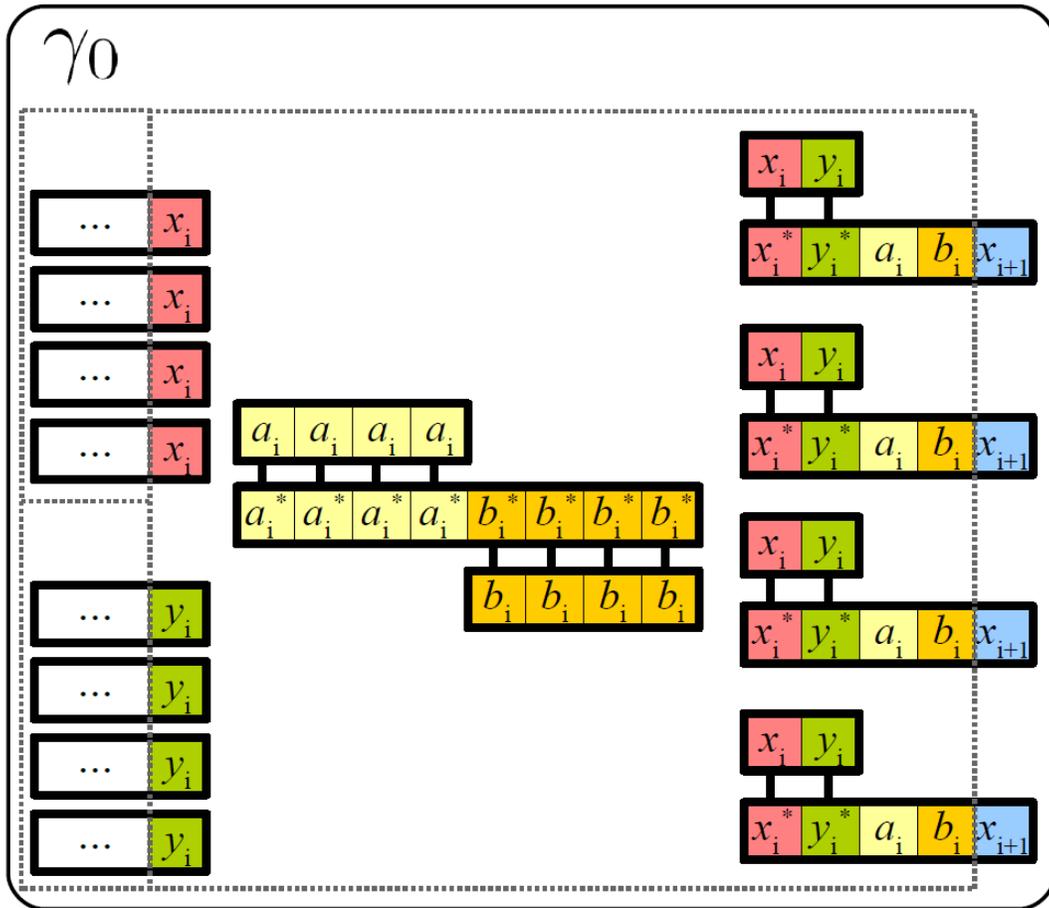
AND gate



Issues with Boolean logic

- How to compose?
 - We don't know how to prove the previous gate is composable, and used a more complex design in the paper
- Want “entropy gap”:
 - Need not merely that unwanted configurations are unstable (i.e., if saturated, they have lower entropy), but more strongly that they have much lower entropy.
 - We can use $O(n)$ domain/monomer types to achieve an entropy gap of n .
- Output convention?
 - Obvious one: “there's a unique stable configuration with the correct output”
 - It's problematic, so we have a one-sided convention:
 - if correct output is 0, unique stable configuration with correct answer
 - if correct output is 1, then both the “output=1” and “output=0” configurations are stable

Composable AND gate with entropy gap 3



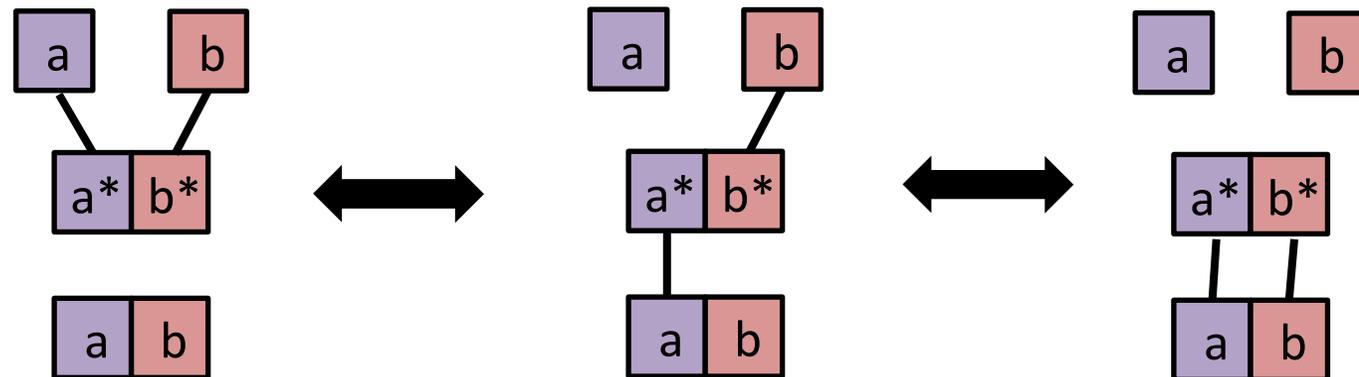
Rather than release a single output monomer, it suffices to gather all output domains on one complex.

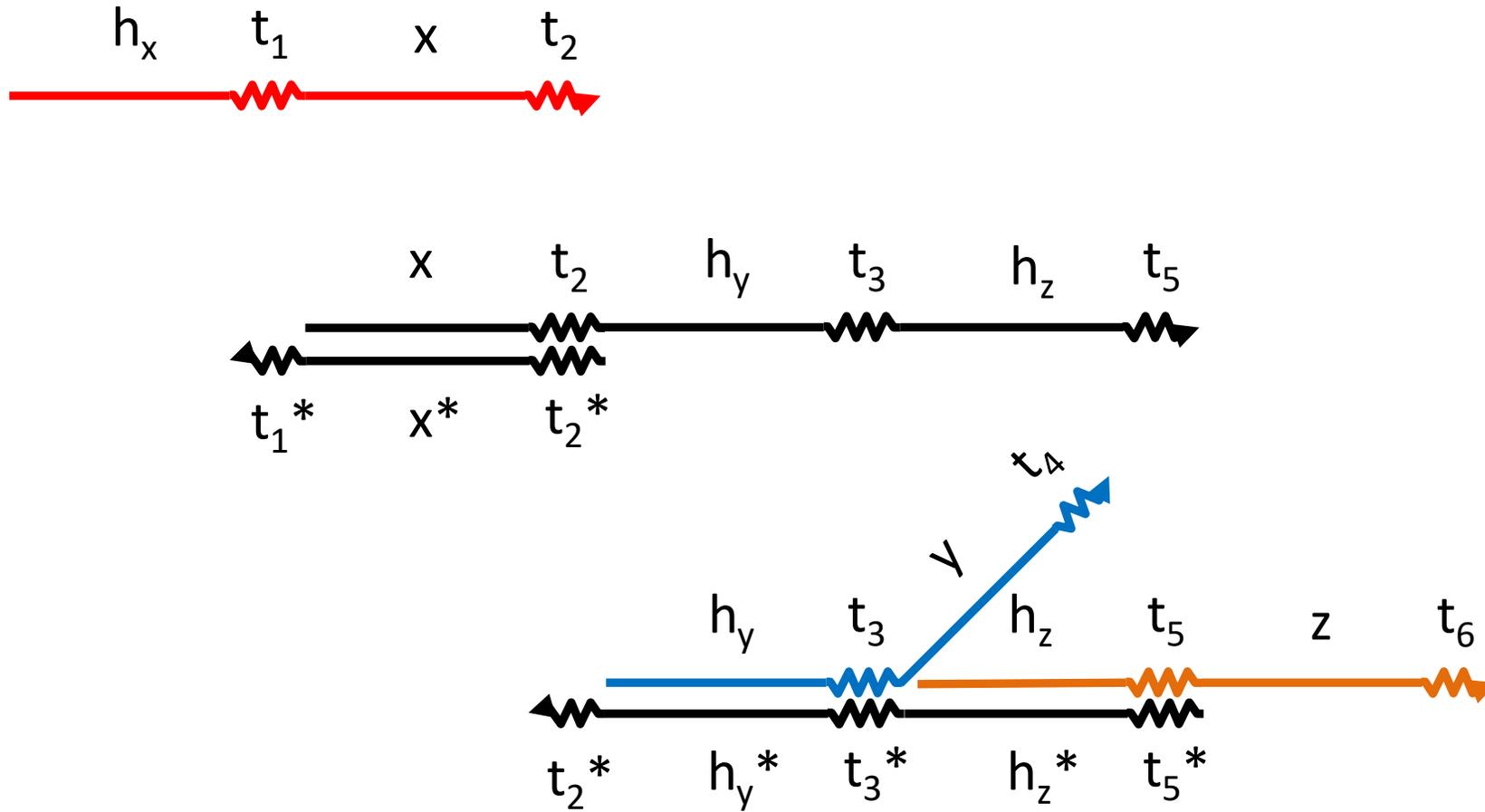
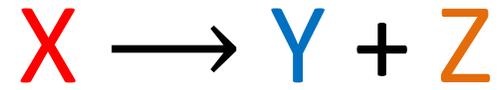
Kinetic pathways and energy barriers

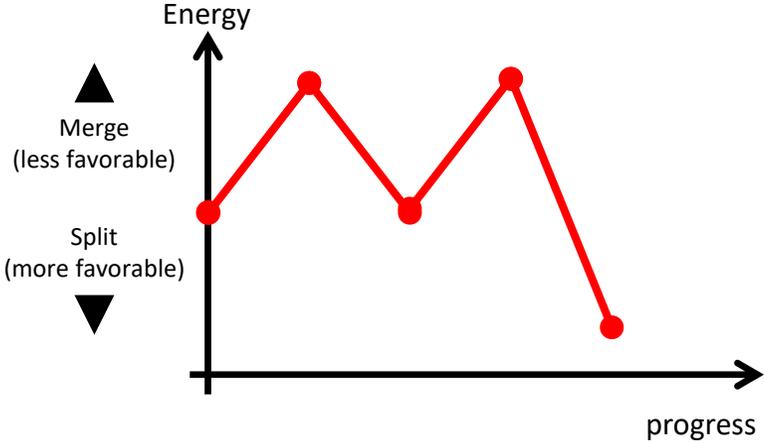
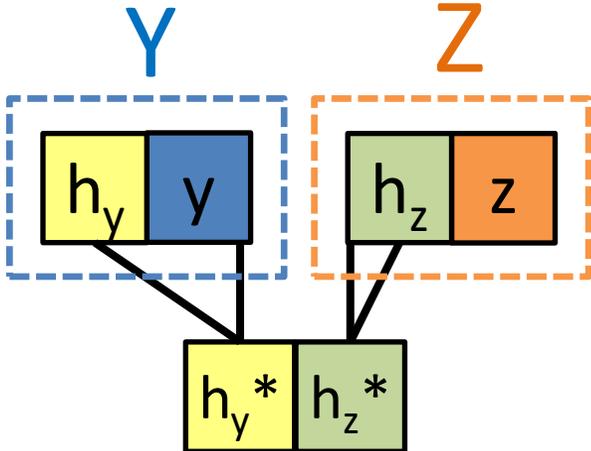
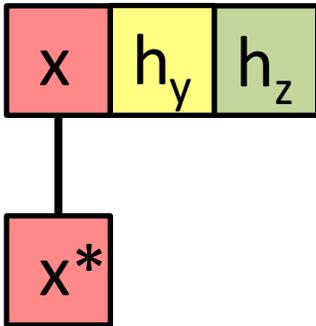
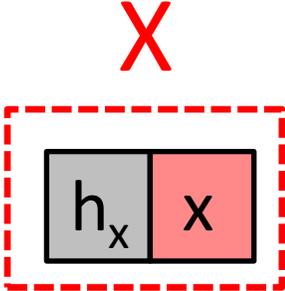
Pathways

Thermodynamics: Which configurations are energetically favorable

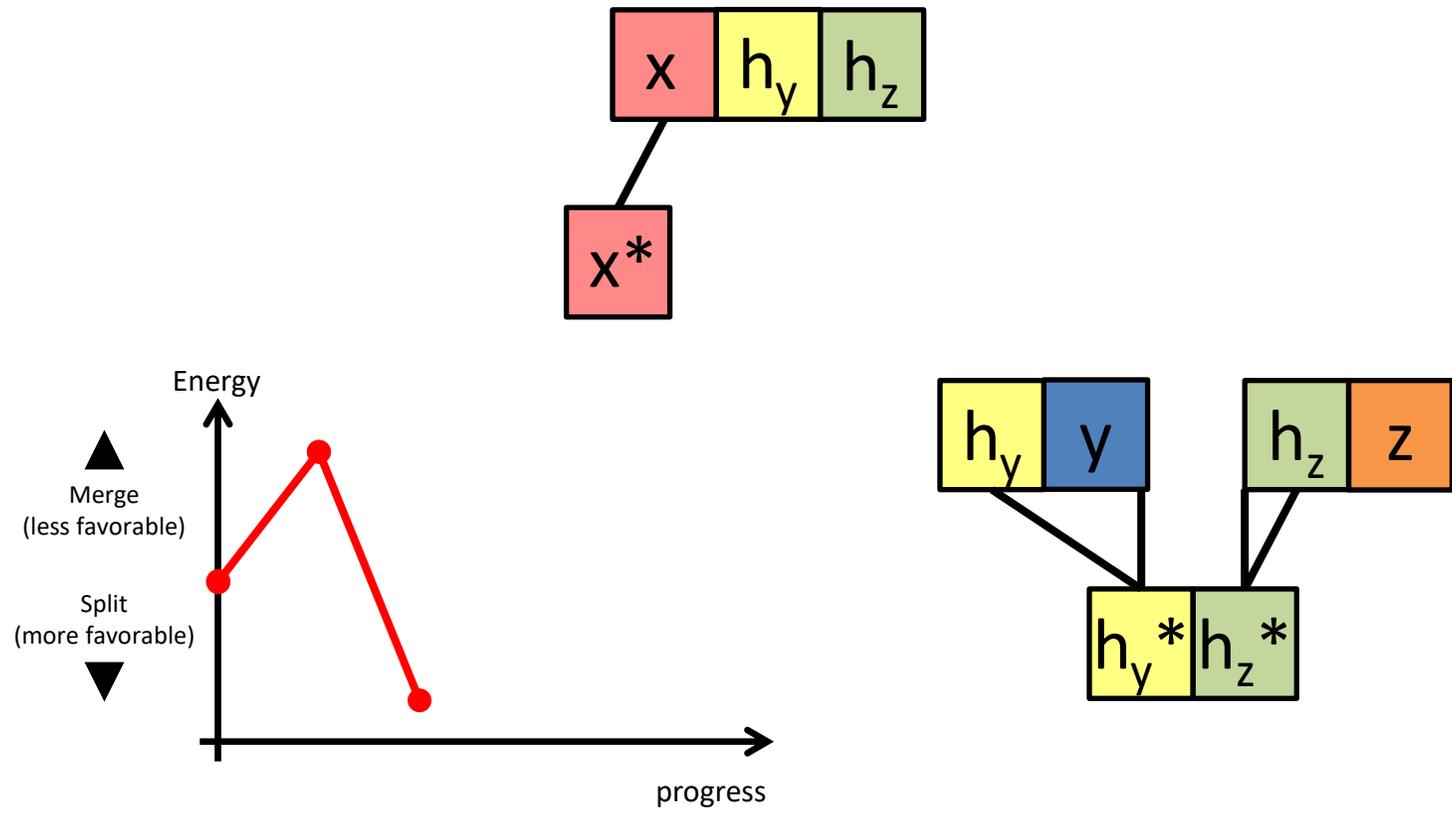
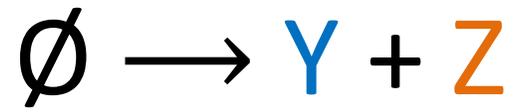
Kinetics: How a system moves between configurations over time







What causes leak
“kinetically”?



Kinetic Binding Networks

- Favorability is a combination of bond count and complex count

Weighted average:

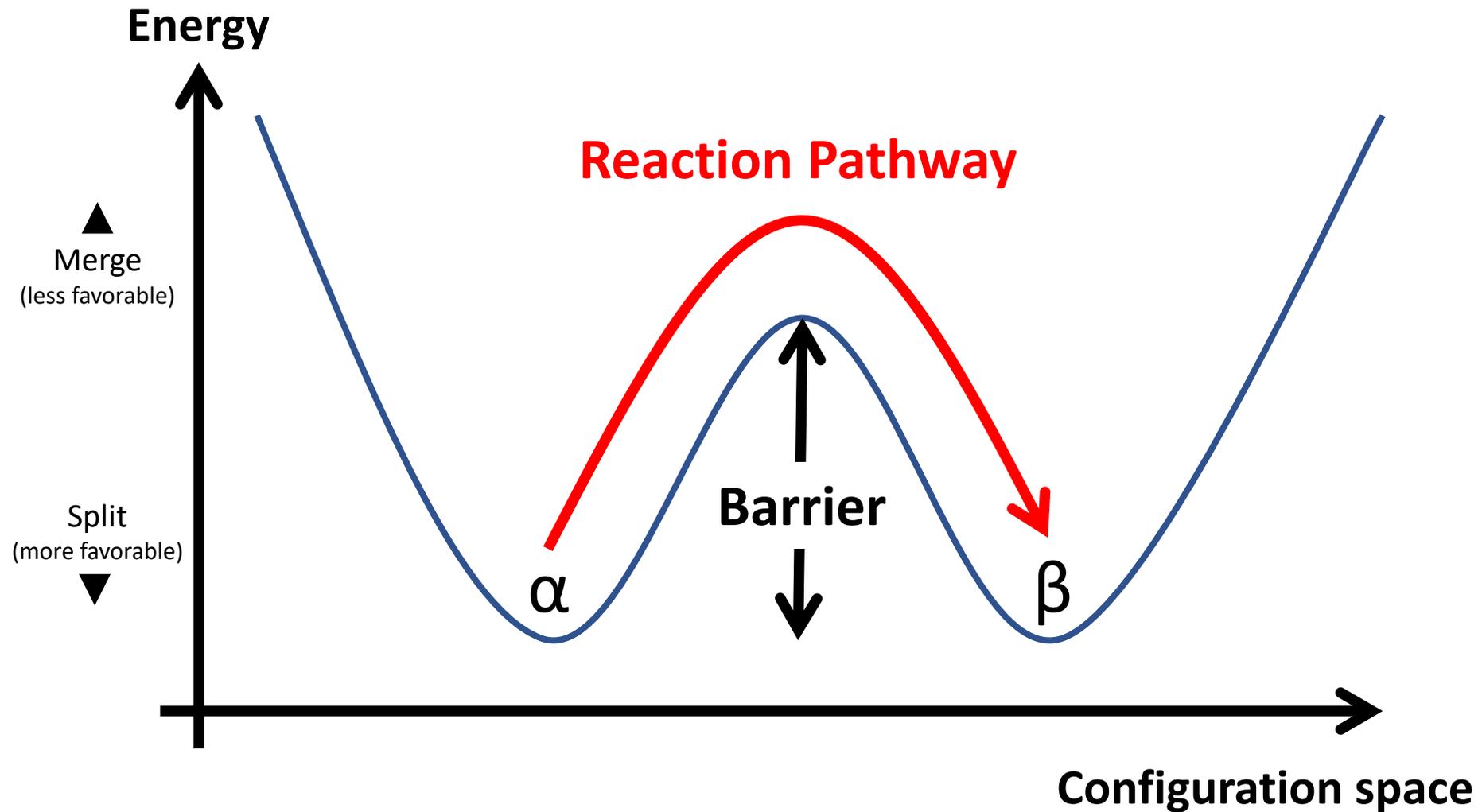
$$\text{Energy} := -w_H(\# \text{ bonds}) - (\# \text{ complexes})$$

- Define pathways to consist of merges and splits
- But for $w_H \geq 2$, only saturated pathways need be considered

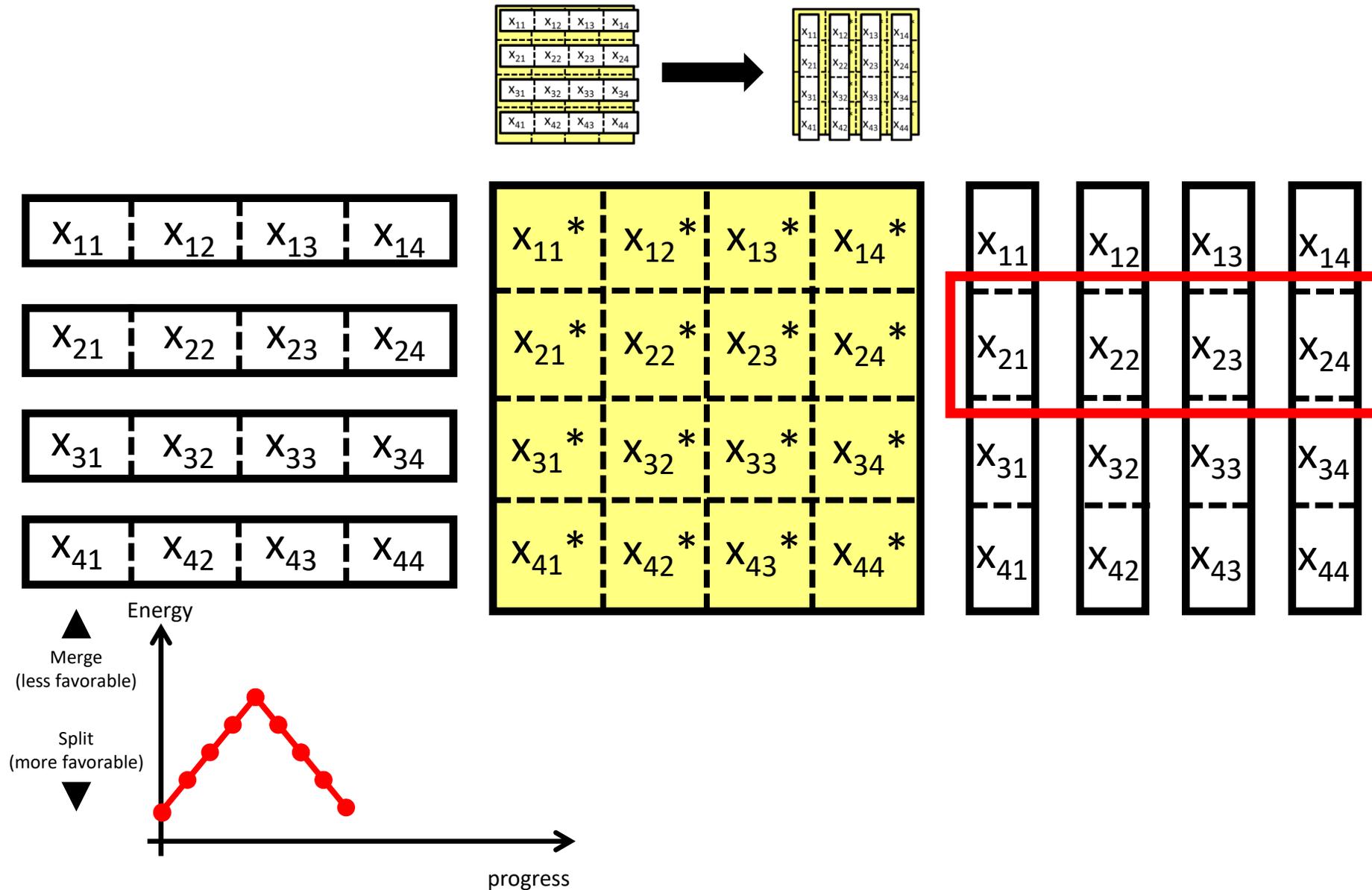
Since all saturated configurations have an equal number of bonds, we can focus solely on the number of complexes

[Keenan Breik, Cameron Chalk, David Doty, David Haley, David Soloveichik. *Programming Substrate-Independent Kinetic Barriers with Thermodynamic Binding Networks*. Computational Methods in Systems Biology 2018]

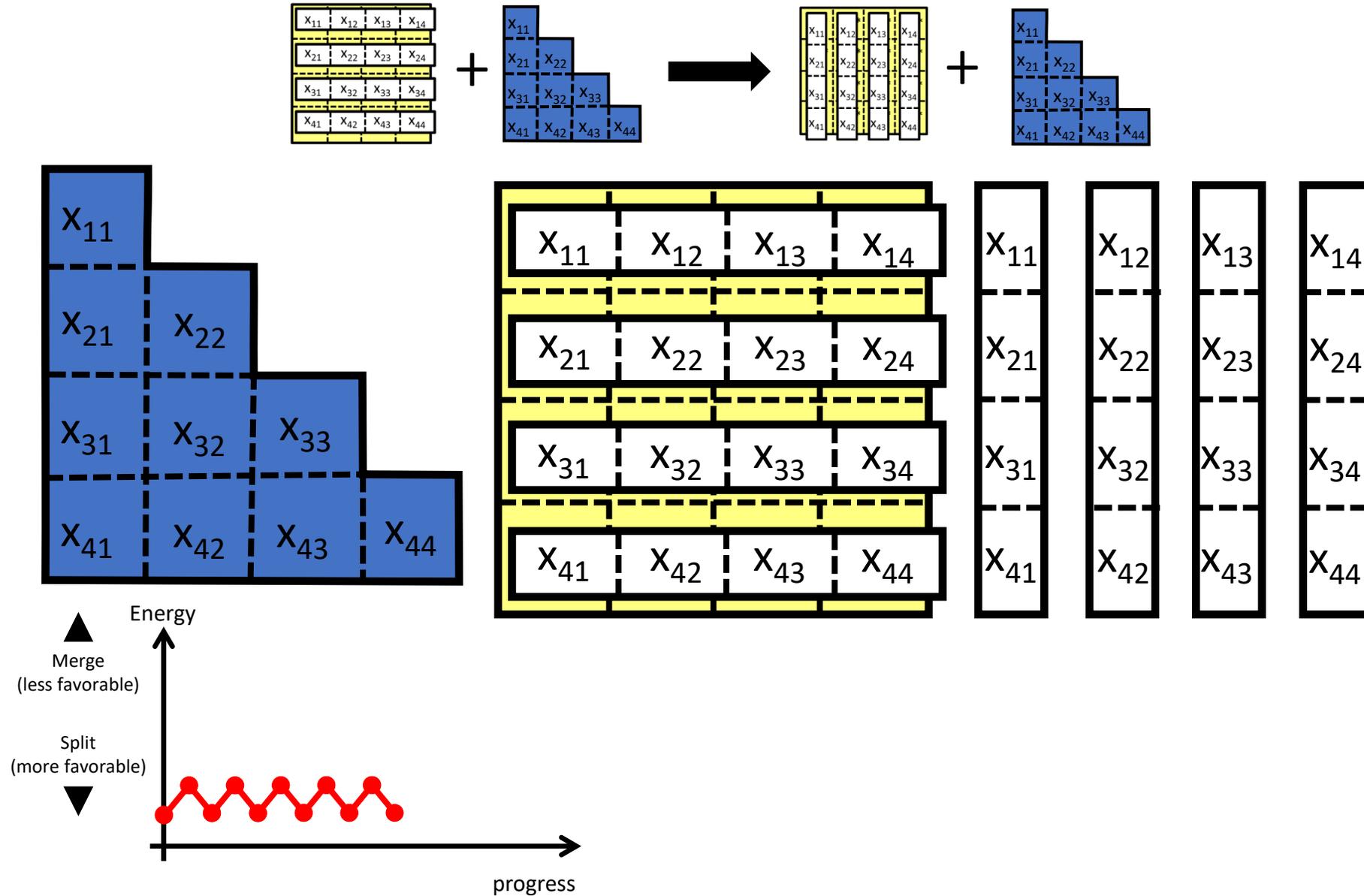
Large Energy Barriers



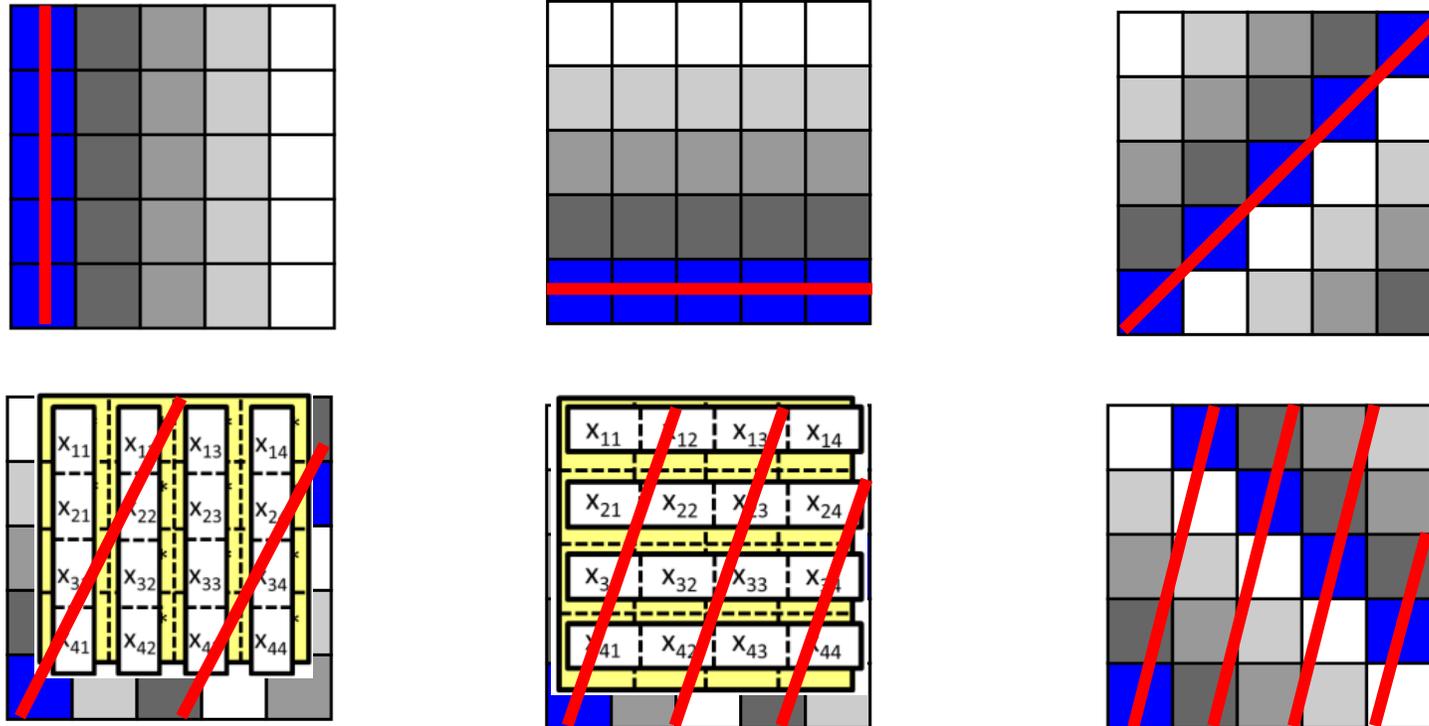
A Network with a Programmable Energy Barrier



Catalysis

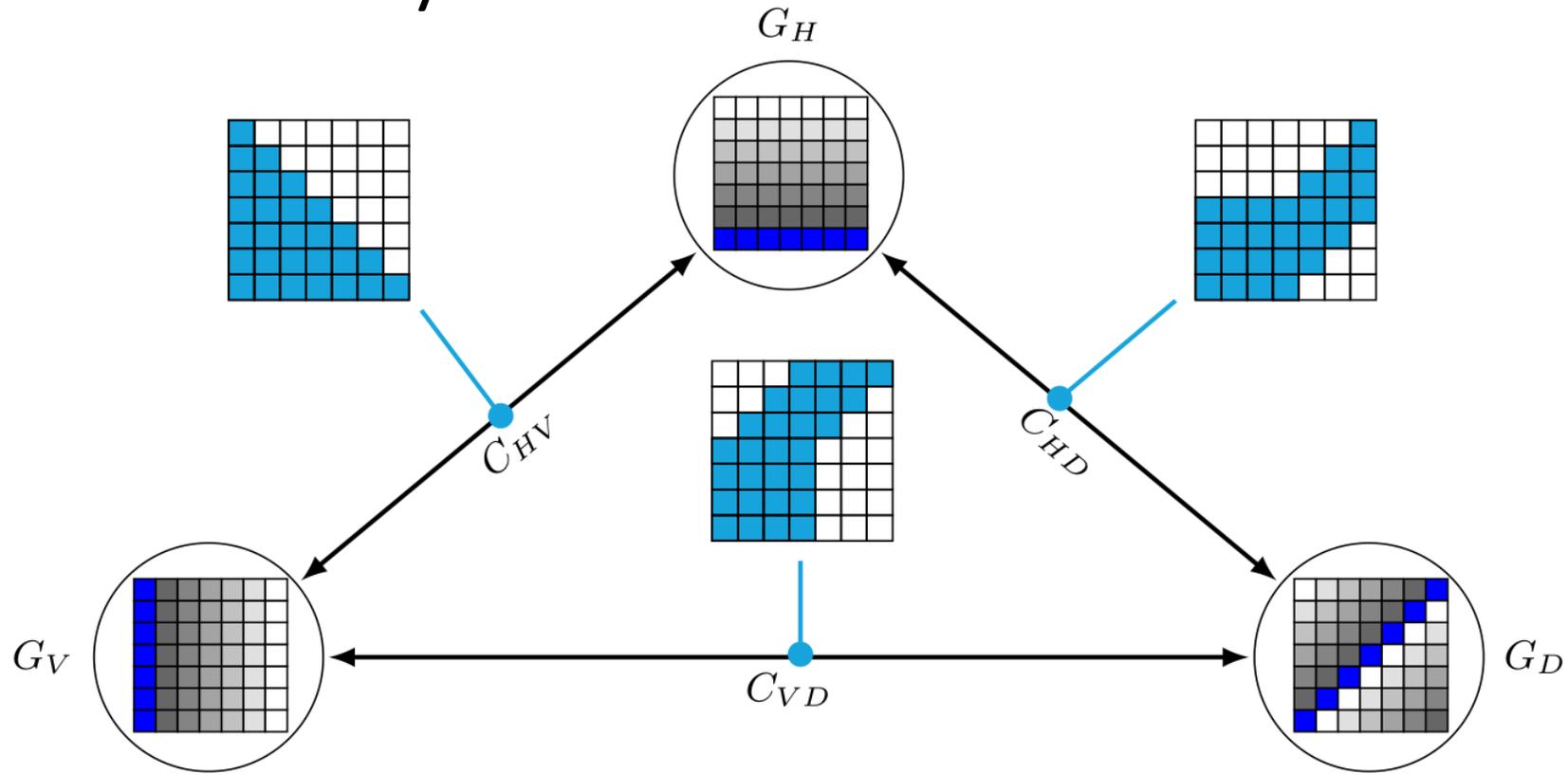


Multiple Stable Configurations



For a grid of prime size $n \times n$, there can be at most $n+1$ different stable configurations with barrier n to pass between any of them

Directed Catalysis



Along a catalyzed pathway, the barrier is 1

Otherwise the barrier is $n/2$

Social Golfer Problem

- Can $25 (n^2)$ golfers play in 5-somes (n -somes) for $6 (n+1)$ days, so that no two golfers play together more than once?
- First studied by Euler.
- True if n is a prime power (2,3,4,5,7,8,9,11,13,...)
- False for smallest non-prime power $n=6$: can only play for 3 days!
[Gaston Tarry (1901). "Le Problème des 36 Officiers". *Compte Rendu de l'Association Française pour l'Avancement des Sciences*. Secrétariat de l'Association. 2: 170–203.]
- Unknown for next prime power $n=10$:
 - trivial upper bound is 11 days
 - best known lower bound is 3

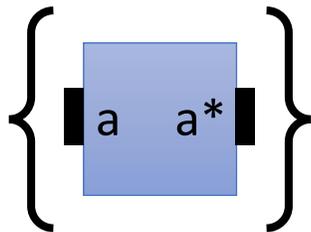
Thermodynamic self-assembly

Grafting the TBN model onto self-assembly

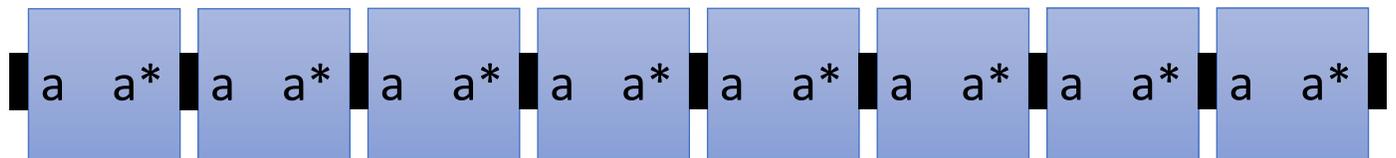
A modest goal

- **Informal:** Design monomers that self-assemble **arbitrarily large complexes**.
 - size of a complex = # monomers in the complex
- **Formal:** Design a set of monomer types so that, for all $S \in \mathbb{N}$, there is a stable complex of size at least S .
- Easy to do in Abstract Tile Assembly Model:

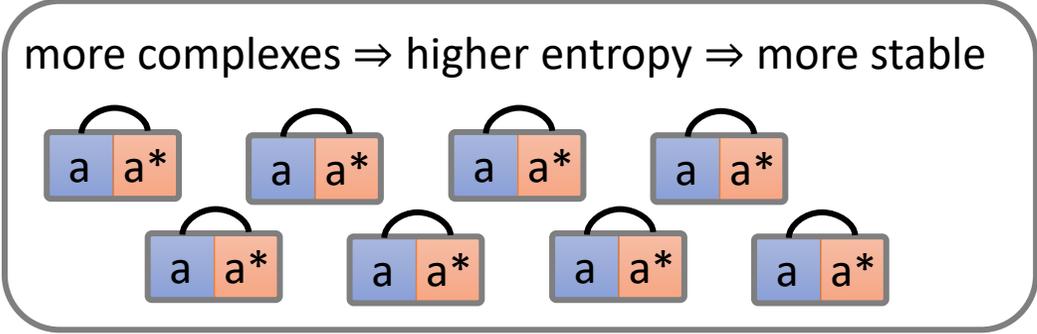
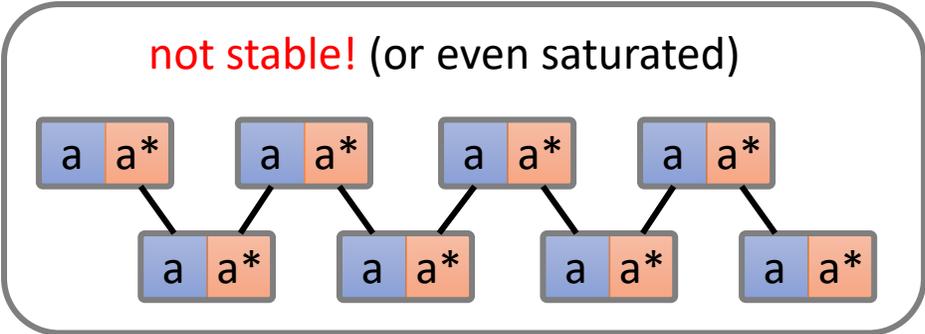
set of monomer types:



size-8 complex (assembly) formed
with 8 copies of monomer

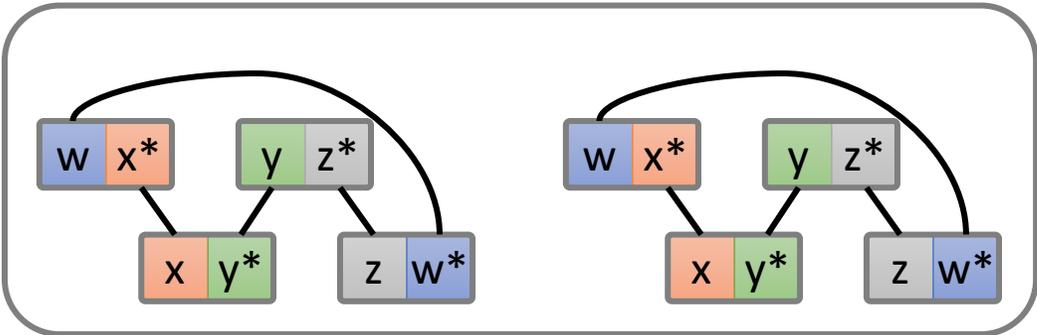
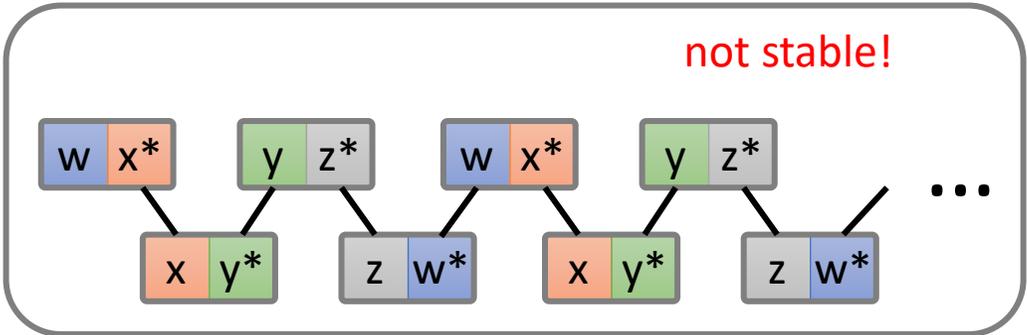


Difficulty of self-assembling large complexes



These have more complexes, and each is self-saturating (every domain can be bound *within* the complex)

attempt 2:



An even more modest goal

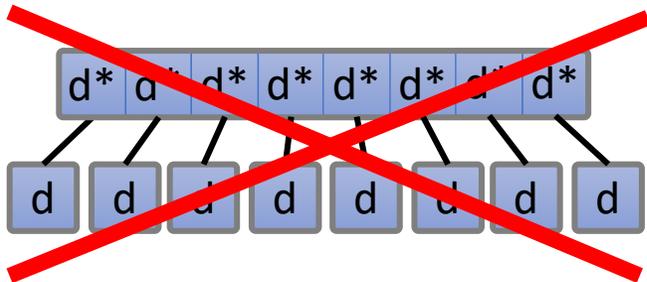
~~Original goal: Design a set of monomer types so that, for all $S \in \mathbb{N}$, there is a stable complex of size at least S .~~

and $O(1)$ domains per monomer

Re- Revised goal: For all $S \in \mathbb{N}$, design a set of M monomer types using D domain types with a stable complex of size at least S .

How large can we make S relative to D and M ?

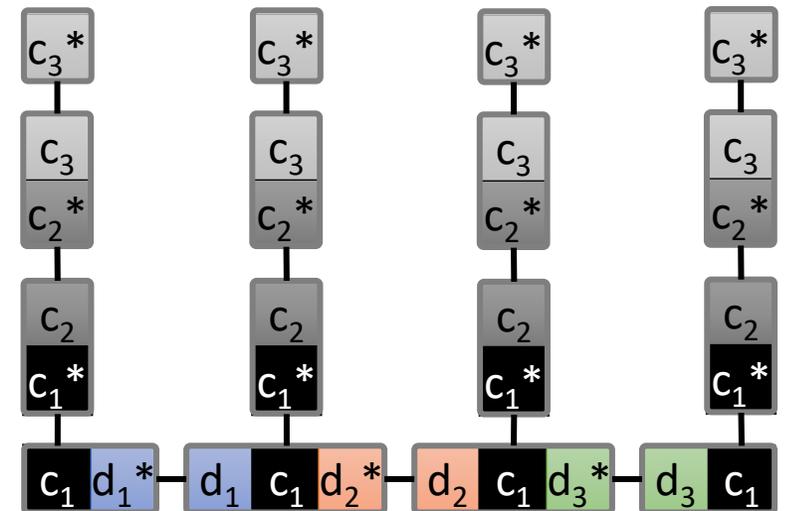
$D, M = O(1)$, $S = \text{arbitrarily large}$



$S \approx D$



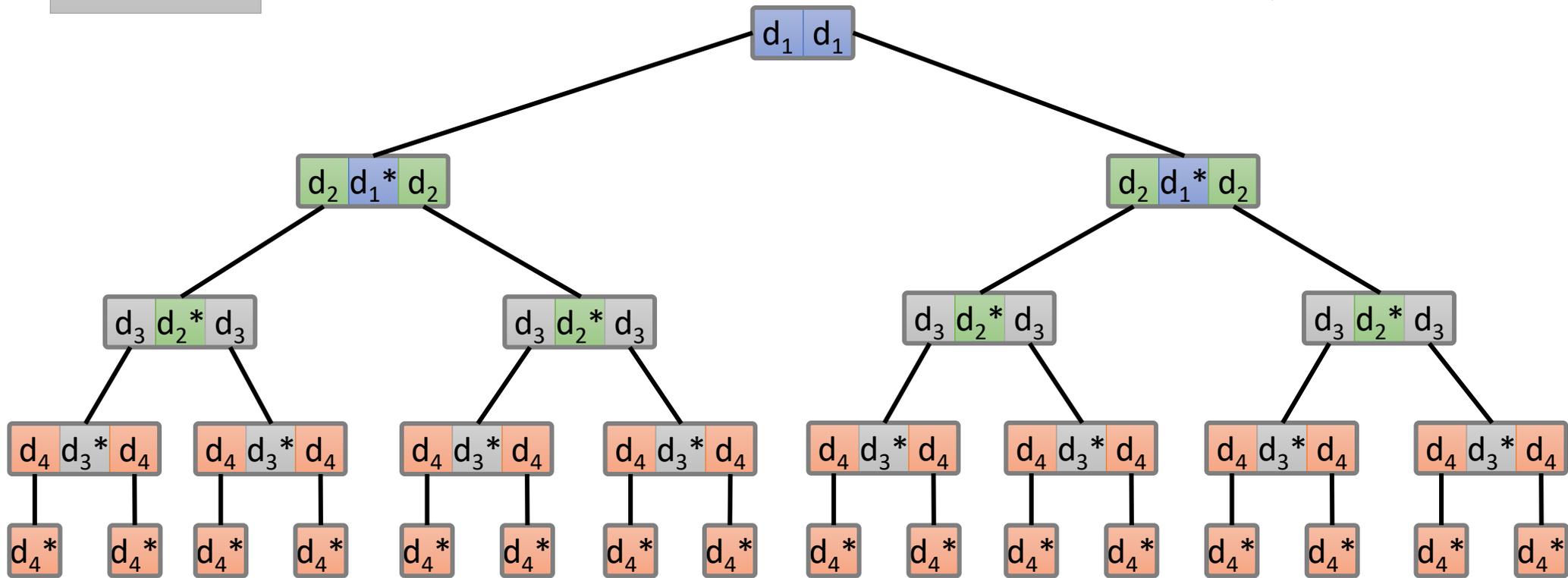
$S \approx D^2$



How large can we make S relative to D and M ?

$$S \approx 2^D?$$

~~$$S \approx 2^{2^D} ??$$~~



Stable complexes have at most exponential size

Theorem: Any thermodynamic binding network with

- D domain types,
- M monomer types,
- $\leq A$ domains per monomer type (note $D/A \leq M \leq A^{D+1}$)

Has stable complexes of size $\leq 2(M+D)(AD)^{2D+3} = \text{poly}(D^D)$ if $A = O(1)$

Monomers as vectors

- monomer {a, b*,b*, d,d,d,d,d*, e,e*} represented as (1,-2,0,3,0)
- sum of many monomers gives the number of excess domains in a fully bound (saturated) complex with those monomers
 - i.e., 2 copies of above monomer $2 \cdot (1,-2,0,3,0) = (2,-4,0,6,0)$ have an excess of **2** a's, **4** b*'s, **0** c's, **6** d's, **0** e's

Somewhat easy proof that unbounded size complexes cannot be assembled

Original goal: Design a set of monomer types so that, for all $S \in \mathbb{N}$, there is a stable complex P of size $\geq S$.

Theorem: Original goal is impossible.

Proof:

1. Suppose otherwise, let P_1, P_2, \dots in \mathbb{N}^m be an infinite sequence of stable complexes increasing in size.
 m is number of monomer types, $P_i(j) = \#$ monomers of type j in complex P_i .
2. Represent each monomer type as a vector in \mathbb{Z}^d as on previous slide.
 1. P_i is composed of monomers $m_{1i}, m_{2i}, \dots, m_{ki}$.
 2. Let $S_i = m_{1i} + m_{2i} + \dots + m_{ki}$. Note that there is a $m \times d$ matrix M such that $S_i = M \cdot P_i$.
3. Take several infinite subsequences:
 1. Since there are a finite number of domain types, some infinite subsequence of P_i 's agrees on which set of domain types are unbound.
 2. By Dickson's Lemma we may assume $P_1 < P_2 < \dots$ and $S_1 < S_2 < \dots$ i.e., each has all the monomers of the previous, plus some more, and each has all the unbound domains of the previous, plus some more.
4. Let $d = P_2 - P_1$. Then $M \cdot d = M \cdot P_2 - M \cdot P_1 = S_2 - S_1 \geq 0$.
5. i.e., $S_2 = S_1 + M \cdot d$ and all three are nonnegative,
6. i.e., we can split S_2 into 2 disjoint nonempty nonnegative subsets, S_1 and $M \cdot d$. **QED**

A digression into computational complexity

- INTEGER-PROGRAMMING problem

Given: integer matrix \mathbf{A} , integer vector \mathbf{b}

Question: is there a nonnegative integer vector \mathbf{x} such that $\mathbf{Ax} = \mathbf{b}$?

- 0/1-INTEGGER-PROGRAMMING is **NP**-complete (Karp 1972).
- Non-obvious fact: INTEGER-PROGRAMMING is in **NP**. (*independently due to [Borosh and Treybig 1976], [Gathen and Sieveking 1978], [Kannan and Monma 1978]*)
If $\mathbf{Ax} = \mathbf{b}$ has a solution, it has a “small” solution... $\max_i x_i \leq \exp(\max_{ij}(\mathbf{A}_{ij}, \mathbf{b}_j))$
- Papadimitriou’s proof: [*On the complexity of integer programming*. Papadimitriou, JACM 1981]
 - If \mathbf{x} is a *large enough* solution, there is $\mathbf{0} < \mathbf{y} < \mathbf{x}$, $\mathbf{y} \in \mathbb{N}^m$, such that $\mathbf{Ay} = \mathbf{0}$.
 - Defining $\mathbf{z} = \mathbf{x} - \mathbf{y}$, $\mathbf{Az} = \mathbf{A}(\mathbf{x} - \mathbf{y}) = \mathbf{Ax} - \mathbf{Ay} = \mathbf{Ax} - \mathbf{0} = \mathbf{b}$.
 - So \mathbf{z} is a strictly smaller solution than \mathbf{x} : \mathbf{x} cannot be the *smallest* solution.

Farkas' Lemma

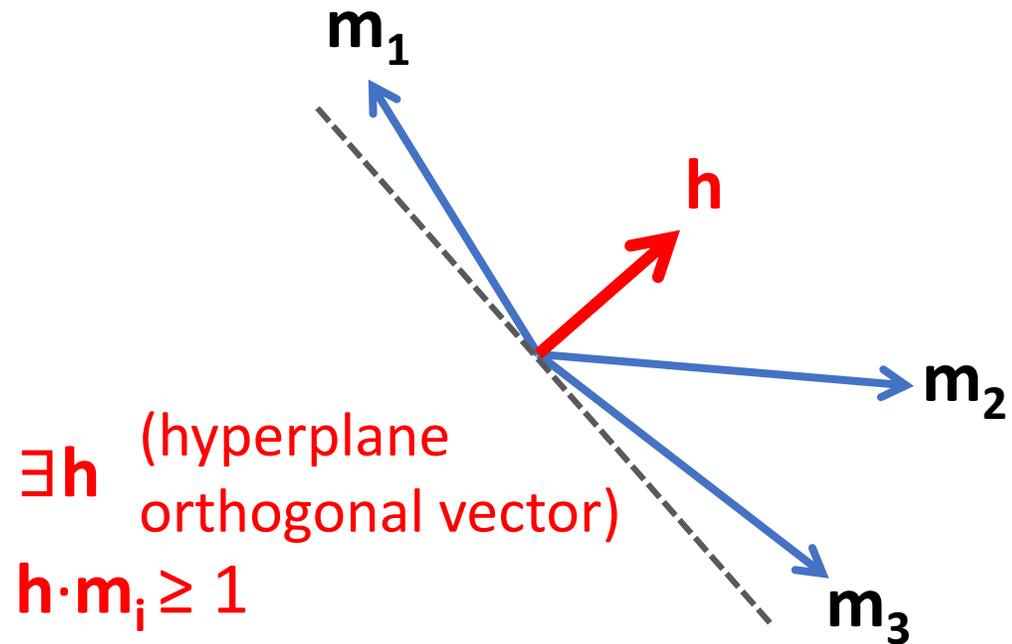
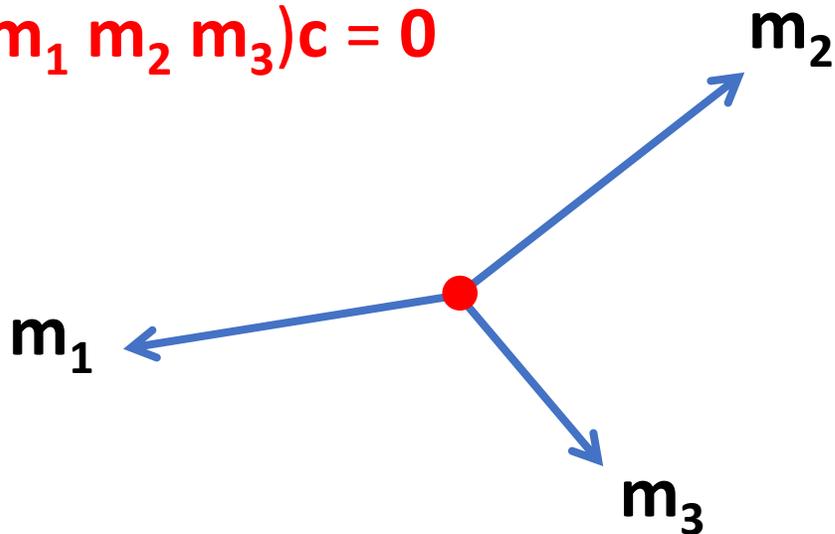
Given vectors $\mathbf{m}_1, \mathbf{m}_2, \dots$, they obey one of two constraints:

a) are directions of balanced forces

b) lie on one side of some hyperplane

$\exists \mathbf{c}$ (counts of monomers)

$$(\mathbf{m}_1 \ \mathbf{m}_2 \ \mathbf{m}_3) \mathbf{c} = 0$$

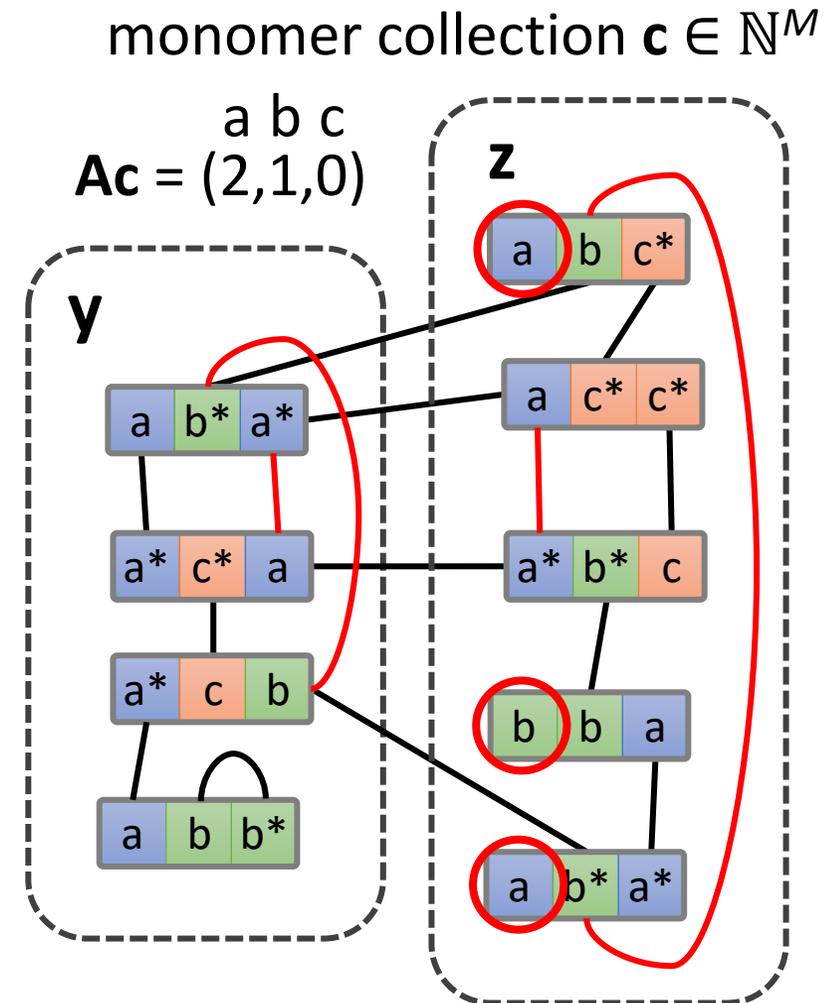


$\exists \mathbf{h}$ (hyperplane
orthogonal vector)

$$\mathbf{h} \cdot \mathbf{m}_i \geq 1$$

How to prove exponential complex size bound for complexes with cycles in binding graph?

- $\mathbf{A} = d \times m$ matrix: \mathbf{A}_{ij} = monomer \mathbf{m}_j 's excess of domain d_i over d_i^*
- If $\mathbf{Ac} = \mathbf{b}$, then \mathbf{b}_i = total # unbound d_i in any saturated configuration of \mathbf{c}
- If $|\mathbf{c}| >$ exponential in D , Papadimtriou's proof gives us subcollection $\mathbf{y} < \mathbf{c}$ such that $\mathbf{Ay} = \mathbf{0}$, (*Farkas' Lemma says that if this fails, then monomer vectors all lie on one side of a hyperplane, see next slide*)
- i.e., $\#d_i$ in $\mathbf{y} = \#d_i^*$ in \mathbf{y} , so \mathbf{y} is self-saturating.
- So whatever bonds were broken to separate \mathbf{y} can be re-bound within \mathbf{y} .
- By symmetry, the same bonds in $\mathbf{z} = \mathbf{c} - \mathbf{y}$ can be re-bound within \mathbf{z} .



If all monomer types lie on one side of hyperplane \mathbf{h} ...

- Consider “slack monomers” $\{d_1^*\}, \{d_2^*\}, \dots$, adding just enough to bind to all the excess d_i domains, so **saturated** (fully bound) == **all domains bound**
- If \mathbf{c} is count of all monomers including slack monomers ($\mathbf{c}(i)$ = count of \mathbf{m}_i), then $\mathbf{A}\mathbf{c} = \mathbf{0}$, where each column of \mathbf{A} represents a monomer (counts of domains).
- dot-product \mathbf{h} on both sides: $\mathbf{h}\cdot\mathbf{A}\mathbf{c} = \mathbf{h}\cdot\mathbf{0} = 0$, distribute through: $\sum_i(\mathbf{h}\cdot\mathbf{m}_i)\mathbf{c}(i) = 0$
- Let S be set of monomers with “small” counts, move them to one side:

$$-\sum_{i \in S}(\mathbf{h}\cdot\mathbf{m}_i)\mathbf{c}(i) = \sum_{i \notin S}(\mathbf{h}\cdot\mathbf{m}_i)\mathbf{c}(i)$$

- Then “small” $\geq -\sum_{i \in S}(\mathbf{h}\cdot\mathbf{m}_i)\mathbf{c}(i) = \sum_{i \notin S}(\mathbf{h}\cdot\mathbf{m}_i)\mathbf{c}(i) \geq \sum_{i \notin S}\mathbf{c}(i)$

$\mathbf{c}(i)$ (count of i 'th monomer) is small by definition, and $\mathbf{h}\cdot\mathbf{m}_i = O(1)$

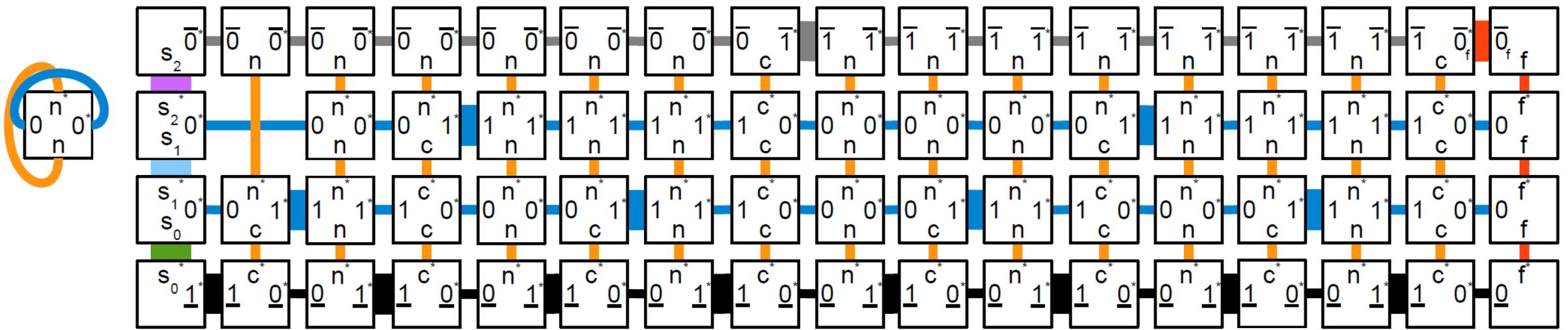
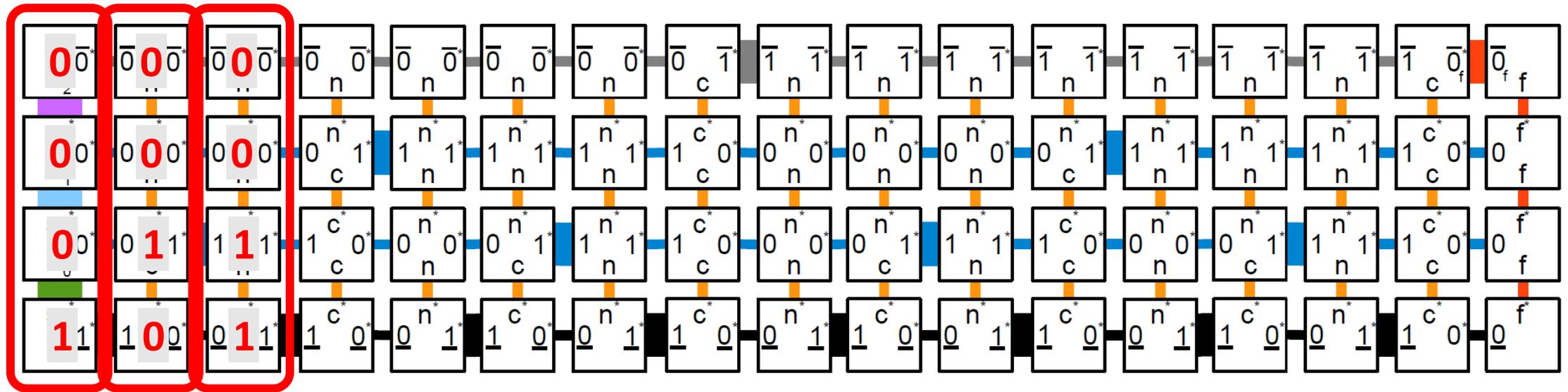
above

since $\mathbf{h}\cdot\mathbf{m}_i \geq 1$

Applying thermodynamic model to tile assembly

- Let's incorporate the thermodynamic binding network model into the abstract tile assembly model.
- How can we create a large assembly from a small number of tile types?

A thermodynamically **unstable** tile assembly counter



Conclusions

- Strong bonds (surprisingly) aren't sufficient to self-assemble large thermodynamically stable structures. *Geometry helps!*
- Kinetically self-assembling a thermodynamically stable structure has very strong guarantees on errors:
 - target structure eventually results despite arbitrary kinetic errors.
 - If it's the only stable structure, and free energy of other structures is much less, then it's the only result you'll see.
- Bad news: **NP**-complete to tell if a given configuration is unstable... even **NP**-hard to approximate entropy of stable configuration:
[Breik, Thachuk, Heule, Soloveichik, *Computing properties of stable configurations of thermodynamic binding networks*, Theoretical Computer Science 2019]