

# Programming Substrate-Independent Kinetic Barriers with Thermodynamic Binding Networks

Keenan Breik , Cameron Chalk , David Doty , David Haley , David Soloveichik

Engineering molecular systems that exhibit complex behavior requires the design of kinetic barriers. For example, an effective catalytic pathway must have a large barrier when the catalyst is absent. While programming such energy barriers seems to require knowledge of the specific molecular substrate, we develop a novel substrate-independent approach. We extend the recently-developed model known as thermodynamic binding networks, demonstrating programmable kinetic barriers that arise solely from the thermodynamic driving forces of bond formation and the configurational entropy of forming separate complexes. Our kinetic model makes relatively weak assumptions, which implies that energy barriers predicted by our model would exist in a wide variety of systems and conditions. We demonstrate that our model is robust by showing that several variations in its definition result in equivalent energy barriers. We apply this model to design catalytic systems with an arbitrarily large energy barrier to uncatalyzed reactions. Our results could yield robust amplifiers using DNA strand displacement, a popular technology for engineering synthetic reaction pathways, and suggest design strategies for preventing undesired kinetic behavior in a variety of molecular systems.



## 1 Introduction

Abstract mathematical models of molecular systems, such as chemical reaction networks, have long been useful in *natural* science to study the properties of natural molecules. With recent experimental advances in synthetic biology and DNA nanotechnology [1]–[3], [10], [11], such models have come to be viewed also as *programming languages* for describing the desired behavior of synthetic molecules.

We can describe a chemical program with abstract chemical reactions such as



In particular, a program may require (1) and forbid (2). But what remains hidden at this level of abstraction is a well-known chemical constraint: if (1) is possible, then (2) must also be, no matter the exact substances. Knowing this, we might try to slow (2) by ensuring  $B$  has high free energy. But then  $B + C$  must also have high free energy, so (1) slows in tandem. The only option to slow (2) but not (1) is to use a *kinetic barrier*: designing  $A$  so that, although it is possible for  $A$  to reconfigure into  $B$ , the system must traverse a higher energy (less favorable) intermediate in the absence of  $C$ .

It seems difficult to engineer kinetic energy barriers and catalysis in a way that is independent of the particular chemical substrate. For example, the development of novel protein enzymes requires the precise positioning of hydrophobic or electrostatic interactions, or otherwise chemically active sites, which are often hard to engineer from first principles. Catalysts based on DNA strand displacement reactions arguably promise the highest degree of programmability [?], [?]. Yet, kinetic

barriers here usually crucially depend on the specific toehold-sequestering mechanism. Importantly, such DNA-based catalysis suffer from a relatively large rate of the uncatalyzed reaction, and prior work lacks a method to arbitrarily increase the uncatalyzed energy barrier.

To develop a substrate-independent approach to engineering kinetic barriers we need to rely on a universal thermodynamic property that would be relevant in a wide variety of chemical systems. We focus on the entropic penalty of association (decreasing the number of separate complexes). Intuitively, the entropic penalty is due to decreasing the number of microstates corresponding to the independent three-dimensional positions of each complex (configurational entropy). This thermodynamic penalty can be made dominant compared with other factors by decreasing the concentration.

To formalize this entropic penalty, we use the *thermodynamic binding networks* (TBN) model [?], [5]. TBNs represent molecules as abstract *monomers* with binding sites that allow them to bind to other monomers. For a configuration  $\gamma$ , the TBN model defines  $H(\gamma)$  as the number of bonds formed, and  $S(\gamma)$  as the number of free complexes, and the *energy*  $E(\gamma) = -wH(\gamma) - S(\gamma)$  as a (negative) weighted sum of the two. To be applicable to a wide variety of chemical systems, the TBN model does not impose geometric constraints on bonding (monomers are simply multisets of binding sets). Implementation of TBNs requires choosing a concrete physical substrate and geometric arrangement that permits the desired configurations to form.

We augment the TBN model with a notion of kinetic paths (changes in configuration) due to merging of different complexes and splitting them up (and in Section 5 making, breaking, or exchanging bonds). This gives rise to a notion

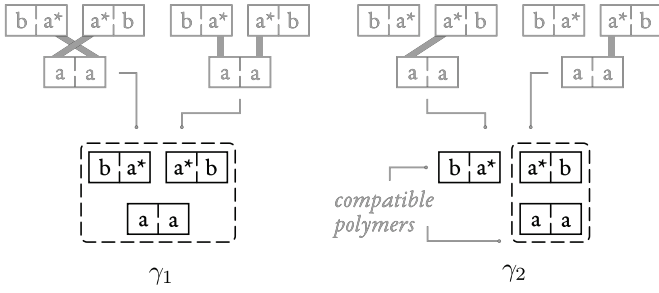


Fig. 1. Two configurations  $\gamma_1$  and  $\gamma_2$  of the TBN  $\mathcal{T} = \{\{a, a^*\}, \{a^*, b\}, \{a^*, b^*\}\}$ . Note that  $\mathcal{T}$  has 3 monomers but 2 monomer types and 6 sites but 3 site types. A dashed box indicates monomers that are part of the same polymer. A single configuration (bottom) can correspond to multiple ways of binding complementary sites (top), which are not distinguished in our model. In  $\gamma_2$  the polymer on the left has exposed sites  $\{b, a^*\}$  and the polymer on the right  $\{a, b\}$ ; they are thus compatible since the exposed site  $a^*$  of the left is complementary to exposed site  $a$  of the right. Since  $\gamma_2$  has compatible polymers it is not saturated, but  $\gamma_1$  is.

of *paths* of configurations, with different energies. Define the *height* of a path starting at  $\gamma$  as the maximum energy difference  $E(\delta) - E(\gamma)$  over all configurations  $\delta$  on the path. Then the kinetic energy *barrier* separating configuration  $\delta$  from configuration  $\gamma$  is the height of the minimum-height path from  $\gamma$  to  $\delta$ .

In Section 2 we introduce our main kinetic model. In Section 3 we give a sufficient condition for large kinetic barriers. In Section 4 we develop two constructions for catalytic systems. Both constructions yield families of TBNs parametrized by a complexity parameter  $n$  such that the uncatalyzed energy barrier scales linearly with  $n$ . The catalyzed energy barrier is always 1. We show a direct DNA strand displacement implementation of one of the constructions. Finally we show an autocatalytic TBN, with an arbitrarily large energy barrier to undesired triggering, that exponentially amplifies its input signal (Section 4.2.1). In Section 5 we show that our main model predicts the same barriers as a more complicated model, one that keeps track of individual bonds on top of whole polymers.

## 2 Kinetic model

Our kinetic models build on thermodynamic binding networks (TBN) [?, [5]. Intuitively, we model a chemical system as a collection of molecules, each of which has a collection of binding sites, which can bind if they are complementary. Although the TBN model is more general, DNA domains can be thought of as the prototypical example of binding sites. No geometry is enforced, which allows the model to handle topologically complex structures, such as pseudoknots.

### 2.1 TBN

(See Fig. 1.) Formally, a *TBN* is a multiset of monomer types. A *monomer type* is a multiset of site types. A *site type* is a formal symbol, such as  $a$ , and has a *complementary* type, denoted  $a^*$ . We call an instance of a monomer type a *monomer* and an instance of a site type a *site*.

### 2.2 Configuration

(See Fig. 1.) We may describe the configuration of a TBN at any moment in terms of which monomers are grouped into polymers. This way a *polymer* is a multiset of monomer types, and a *configuration* is a partition of the TBN into polymers.<sup>1</sup>

The *exposed sites* of a polymer is the multiset of site types that would remain if one were to remove as many complementary pairs of sites as possible. Each such pair is counted as a *bond*. Note that bonds are not specified as part of a configuration, and intuitively we think of polymers as being maximally bonded. Two polymers are *compatible* if they have some complementary exposed sites. A configuration is *saturated* if no two polymers are compatible. This is equivalent to having the maximum possible number of bonds.

### 2.3 Path

(See Fig. 2.) One configuration can change into another by a sequence of elementary steps. If  $\gamma$  can become  $\delta$  by replacing two polymers in  $\gamma$  with their union, then  $\gamma$  *merges* to  $\delta$  and  $\delta$  *splits* to  $\gamma$ , and we write  $\gamma \succ^1 \delta$ .<sup>2</sup> We denote by  $\succeq^1$ ,  $\succ$ ,  $\geq$  the reflexive, transitive, and reflexive transitive closures of  $\succ^1$ . A *path* is a nonempty sequence of configurations where each merges or splits to the next. Note that there is a path between any two configurations.<sup>3</sup>

We could imagine smaller steps that manipulate individual bonds. But surprisingly, such a bond-aware model leads to essentially equivalent kinetic barriers, which we prove in Section 5.

### 2.4 Energy

(See Fig. 2.) For a configuration  $\gamma$ , denote by  $H(\gamma)$  the number of bonds summed over all polymers. Denote by  $S(\gamma)$  the number of polymers.<sup>4</sup> Note that a saturated configuration has maximum  $H(\gamma)$ . The *energy* of  $\gamma$  is

$$E(\gamma) = -wH(\gamma) - S(\gamma),$$

<sup>1</sup>Consider swapping two monomers of the same type between two polymers in a configuration. We do not consider the result a different configuration. Note that monomers of the same type correspond to an entropy contribution that we ignore (see also footnote 4).

<sup>2</sup>Note that this would form a lattice of partitions if the configurations were sets instead of multisets.

<sup>3</sup>Our model allows incompatible polymers to be merged (i.e., two polymers merge without forming any new bonds). This represents spontaneous co-localization and comes with an energy penalty, as discussed later. For instance, to get from any configuration to any other, we can merge all initial polymers into one and then split into the desired end polymers; however, such a path could be very energetically unfavorable.

<sup>4</sup>The quantities  $H(\gamma)$  and  $S(\gamma)$  are meant to evoke the thermodynamic quantities of enthalpy and entropy, although the mapping is not exact. Indeed, the free energy contribution of forming additional bonds typically contains substantial enthalpic and entropic parts. Further, while  $S(\gamma)$  captures entropy due to independent positions of separate polymers, chemical free energy may consider a variety of other entropic contributions. These may include geometric configurations of a single polymer, as well as the entropy due to swapping indistinguishable monomers. We can justify focusing on  $S(\gamma)$  because its contribution arbitrarily predominates in taking the limit of large solution volume—that is, the low concentration regime [5].

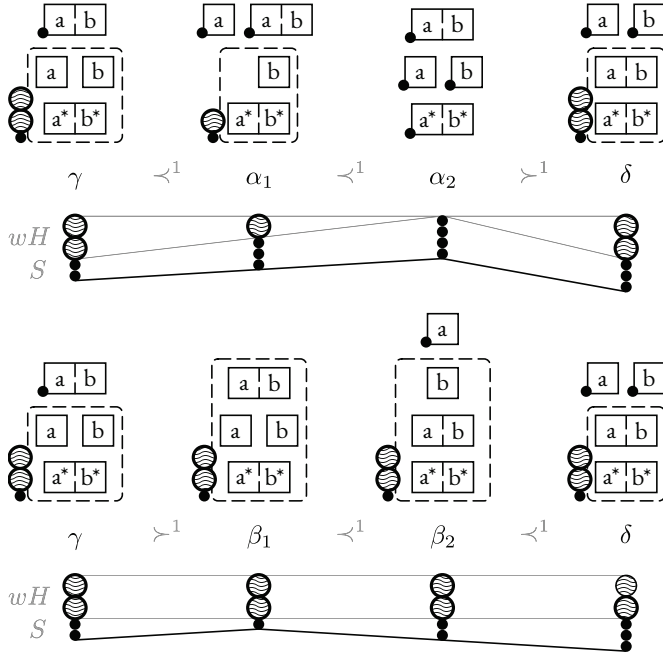


Fig. 2. A path  $p$  consisting of  $\gamma, \alpha_1, \alpha_2, \delta$  and a path  $p'$  consisting of  $\gamma, \beta_1, \beta_2, \delta$  in the TBN  $\mathcal{T} = \{\{a\}, \{b\}, \{a, b\}, \{a^*, b^*\}\}$ . The energy of each configuration is shown graphically below it. A large wavy disc indicates energy due to a bond. A small solid disc indicates energy due to a polymer. Lower is more favorable. Here bond strength  $w = 2$ , so a wavy disc is twice as tall as a solid disc. The height of  $p$  is  $h(p) = E(\alpha_2) - E(\gamma) = (-4) - (-2w - 2) = 2$ . The height of  $p'$  is  $h(p') = E(\beta_1) - E(\gamma) = (-2w - 1) - (-2w - 2) = 1$ .

where the bond strength  $w$  represents the benefit from gaining a bond relative to gaining a polymer.<sup>5</sup> Note that  $H(\gamma) \geq 0$  and  $S(\gamma) > 0$ , so  $E(\gamma) < 0$ , and that lower energy, which results from more bonds or more polymers, is more favorable. (The choice to make favorability correspond to lower energy, that is more negative, is motivated by consistency with the standard physical chemistry notion of free energy.) We call a minimum energy configuration *stable*.

Merging incompatible polymers forms no additional bonds and so is unfavorable, since  $S(\gamma)$  drops without  $H(\gamma)$  rising. In contrast, when bond strength  $w > 1$ , merges between compatible polymers are energetically favorable. So every

<sup>5</sup>Our notion of energy idealizes the physical Gibbs free energy. In typical DNA nanotechnology applications, the Gibbs free energy is likewise a linear combination of  $H(\gamma)$  and  $S(\gamma)$ . We can estimate the Gibbs free energy  $\Delta G(\gamma)$  of a configuration  $\gamma$  as follows. Bonds correspond to domains of length  $l$  bases, and forming each base pair is favorable by  $\Delta G_{bp}^\circ$ . Thus, the contribution of  $H(\gamma)$  to  $\Delta G(\gamma)$  is  $(\Delta G_{bp}^\circ \cdot l)H(\gamma)$ . At 1 M concentration, the free energy penalty due to decreasing the number of separate complexes by one is  $\Delta G_{assoc}^\circ$ . At lower concentration  $c$  M  $< 1$  M, this penalty increases to  $\Delta G_{assoc}^\circ + RT \ln((1 \text{ M})/c)$ . The point of zero free energy is taken to be the configuration with no bonds, and all monomers separate. Thus, the contribution of  $S(\gamma)$  to  $\Delta G(\gamma)$  is  $(\Delta G_{assoc}^\circ + RT \ln((1 \text{ M})/c))(m - S(\gamma))$ , where  $m$  is the total number of monomers. To summarize,

$$\Delta G(\gamma) = (\Delta G_{bp}^\circ \cdot l)H(\gamma) + (\Delta G_{assoc}^\circ + RT \ln((1 \text{ M})/c))(m - S(\gamma)).$$

Note that, as expected, this is a linear combination of  $H(\gamma)$  and  $S(\gamma)$ , and that increasing the length of domains  $l$  weighs  $H(\gamma)$  more heavily, while decreasing the concentration  $c$  weighs  $S(\gamma)$  more heavily. Domains are routinely 15–25 bases long, and at 100 nM concentration at room temperature this corresponds to a relative bond strength  $w$  of 1.9–3.2.

stable (that is, minimum energy) configuration is saturated. This regime is typical of many real systems, and in particular, we can engineer DNA strand displacement systems [12] to have large bond strength  $w$  by increasing the length of domains.

## 2.5 Barrier

(See Fig. 2.) There are many paths from a start configuration  $\gamma$  to an end configuration  $\delta$ . The *height*  $h(p)$  of a particular such path  $p$  is the greatest energy difference  $E(\alpha) - E(\gamma)$  between any  $\alpha$  along  $p$  and  $\gamma$ . This measures the difficulty of traversing  $p$ . Notice that  $h(p) \geq E(\gamma) - E(\gamma) = 0$ .

Another reasonable definition of height is the greatest energy difference  $E(\beta) - E(\alpha)$  between any  $\alpha$  and later  $\beta$  on the path. We will be considering paths between stable (lowest energy) configurations, where the two definitions are equivalent.

Going from one configuration to another is difficult if each path has large height. The *barrier*  $b(\gamma, \delta)$  from  $\gamma$  to  $\delta$  is the least height of any path from  $\gamma$  to  $\delta$ . Notice that  $b(\gamma, \delta) \geq 0$  as well. Since paths are reversible, it is easy to show that if  $E(\gamma) = E(\delta)$  then  $b(\gamma, \delta) = b(\delta, \gamma)$ .

## 3 Saturated paths

Analyzing TBNs is simpler if we reason only about saturated configurations. Clearly, in the limit of large bond strength  $w$ , breaking a bond is so unfavorable that a least height path has only saturated configurations. The main result of this section, Corollary 8, shows that the barrier remains the same even if we consider paths that traverse only saturated configurations as long as  $w \geq 2$ . Such a threshold may be surprising: it might seem that breaking some bonds, even if locally unfavorable, might allow a path to bypass an otherwise large barrier elsewhere.

We prove Corollary 8 in Section 3.2 by showing how to transform an arbitrary path into a saturated path with height no greater. Section 3.1 defines a relation among configurations that allows us to keep track of changes in energy as we transform the path.

### 3.1 Bounds on energy change

When two polymers merge, knowing whether they are compatible makes the change in energy predictable. Recall that merging incompatible polymers results in no more bonds, so overall energy increases by 1. Merging compatible polymers results in at least one more bond, so overall energy decreases by at least  $w - 1$ . To make this precise, let  $\gamma \succ_\circ \delta$  (and let  $\gamma \succ_\bullet \delta$ ) mean that  $\gamma$  merges to  $\delta$  by combining two incompatible (compatible) polymers. Let  $\succeq_\circ$  ( $\succeq_\bullet$ ) be the reflexive, transitive closure of  $\succ_\circ$  ( $\succ_\bullet$ ).

**Observation 1.** If  $\gamma \succ_\circ \delta$ , then  $E(\delta) = E(\gamma) + 1$ . If  $\gamma \succ_\bullet \delta$ , then  $E(\delta) \leq E(\gamma) + 1 - w$ .

**Observation 2.** Let  $\Delta = S(\gamma) - S(\delta)$ . If  $\gamma \succeq_\circ \delta$ , then  $E(\delta) = E(\gamma) + \Delta$ . If  $\gamma \succeq_\bullet \delta$ , then  $E(\delta) \leq E(\gamma) + \Delta(1 - w)$ .

To apply these bounds to the general case  $\gamma \succeq \delta$ , we decompose  $\succeq$  into  $\succeq_\bullet$  and  $\succeq_\circ$ . This allows us to identify an intermediate configuration that has least energy (is most favorable).

**Theorem 3.** If  $\gamma \succeq \delta$ , then some  $\alpha$  has  $\gamma \succeq_\bullet \alpha \succeq_\circ \delta$ .

*Proof.* Let  $\gamma \succeq \delta$ , and let  $\alpha$  be a “most merged” configuration with  $\gamma \succeq \bullet \alpha \succeq \delta$  (no  $\alpha' \preceq \alpha$  has  $\gamma \succeq \bullet \alpha' \succeq \delta$ ). Then  $\alpha = \beta_0 \succ^1 \dots \succ^1 \beta_n = \delta$  for some configurations  $\beta_i$ . Consider the polymers  $P$  and  $Q$  in  $\beta_k$  merged by  $\beta_k \succ^1 \beta_{k+1}$ . We claim  $P$  and  $Q$  are incompatible.

To see so, note that  $P = P_1 \cup \dots \cup P_x$  and  $Q = Q_1 \cup \dots \cup Q_y$  for some polymers  $P_i$  and  $Q_j$  in  $\alpha$ . If any  $P_i$  and  $Q_j$  are compatible, then merging them in  $\alpha$  would produce a configuration that contradicts  $\alpha$  being most merged. So they are pairwise incompatible. Letting  $[X]$  denote the exposed sites of  $X$ , we have  $[P] \subseteq [P_1] \cup \dots \cup [P_x]$  and  $[Q] \subseteq [Q_1] \cup \dots \cup [Q_y]$ . So  $P$  and  $Q$  are incompatible.

So  $\beta_k \succ^1 \beta_{k+1}$  for each  $k$ , and  $\alpha = \beta_0 \succeq \bullet \beta_n = \delta$ .  $\square$

### 3.2 Saturated paths suffice

A *saturated path* is a path along which every configuration is saturated. For example, the bottom path  $p'$  in Fig. 2 is saturated. If  $\gamma$  and  $\delta$  are saturated, then let  $b_{\text{sat}}(\gamma, \delta)$  denote the barrier from  $\gamma$  to  $\delta$  when allowing only saturated paths. Since a saturated path is a path,  $b_{\text{sat}}(\gamma, \delta) \geq b(\gamma, \delta)$ . It turns out that if bond strength  $w \geq 2$ , then the reverse inequality also holds, so  $b_{\text{sat}}(\gamma, \delta) = b(\gamma, \delta)$ . And if  $w \geq 1$ , then the reverse inequality “almost” holds.

To show the reverse inequality, we turn an arbitrary path into a saturated path without increasing its height (much). We do this step by step by always merging just enough polymers to achieve saturation. To make this precise, let  $[\gamma]$  denote the set of saturated  $\gamma'$  with  $\gamma \succeq \bullet \gamma'$ , and let  $E_{\text{max}}(p)$  be the maximum energy of any configuration along the path  $p$ .

First we show how to saturate a split step.

**Lemma 4.** *Let  $w \geq 1$ . If  $\gamma \preceq^1 \delta$  and  $\gamma' \in [\gamma]$ , then some  $\delta' \in [\delta]$  and some saturated path  $p'$  from  $\gamma'$  to  $\delta'$  has  $E_{\text{max}}(p') \leq E(\gamma)$ .*

*Proof.* Let  $\gamma \preceq^1 \delta$  and  $\gamma' \in [\gamma]$ . Then  $\gamma \succeq \gamma'$ , so  $\delta \succeq \gamma'$ . So by Theorem 3, some  $\delta'$  has  $\delta \succeq \bullet \delta' \succeq \bullet \gamma'$ . By assumption,  $\gamma'$  is saturated, so  $\delta'$  is, so  $\delta' \in [\delta]$ . Now let  $p'$  be a path guaranteed by  $\gamma' \preceq \bullet \delta'$ . Then  $E_{\text{max}}(p') = E(\gamma') \leq E(\gamma)$ .  $\square$

To show how to saturate a merge step, we rely on being able to transfer a merge from one context to another.

**Lemma 5.** *If  $\gamma \succ^1 \delta$  and  $\gamma \succeq \gamma'$ , then some  $\alpha$  has  $\gamma' \succeq^1 \alpha$  and  $\delta \succeq \alpha$ .*

*Proof.* Let  $\gamma \succ^1 \delta$  and  $\gamma \succeq \gamma'$ . Let  $P$  be the polymer merged by  $\gamma \succ^1 \delta$ , and let  $\gamma^*$  be  $\gamma$  but with all polymers intersecting  $P$  merged. This way  $\gamma^* = \delta$  and  $\gamma' \succeq^1 \gamma^*$ .

Now  $\gamma = \beta_0 \succ^1 \dots \succ^1 \beta_n = \gamma'$  for some configurations  $\beta_i$ . So  $\delta = \gamma^* = \beta_0^* \succeq^1 \dots \succeq^1 \beta_n^* = \gamma^*$ . So choose  $\alpha = \gamma^*$ .  $\square$

Now we show how to saturate a merge step.

**Lemma 6.** *Let  $w \geq 1$ . If  $\gamma \succ^1 \delta$  and  $\gamma' \in [\gamma]$ , then some  $\delta' \in [\delta]$  and some saturated path  $p'$  from  $\gamma'$  to  $\delta'$  has  $E_{\text{max}}(p') \leq \max\{E(\gamma), E(\delta)\} + \max\{0, 2 - w\}$ .*

*Proof.* Let  $\gamma \succ^1 \delta$  and  $\gamma' \in [\gamma]$ . If  $\gamma' = \gamma$ , then let  $\delta' = \delta$ , and let  $p' = \gamma, \delta$ . Then  $E_{\text{max}}(p') = E(\delta)$ .

Otherwise  $\gamma' \neq \gamma$ . Now by Lemma 5, some  $\alpha$  has  $\gamma' \succeq^1 \alpha$  and  $\delta \succeq \alpha$ . So by Theorem 3, some  $\delta'$  has  $\delta \succeq \bullet \delta' \succeq \bullet \alpha$ . Since  $\gamma'$  is saturated,  $\alpha$  is, so  $\delta'$  is, so  $\delta' \in [\delta]$ . Now let  $p'$  be a path guaranteed by  $\gamma' \succeq^1 \alpha \preceq \bullet \delta'$ . Then  $p'$  is saturated. Also,

$E_{\text{max}}(p') = E(\alpha)$ . And  $\gamma' \succeq^1 \alpha$ , so  $E(\alpha) \leq E(\gamma') + 1$ . Since  $\gamma' \in [\gamma]$  and  $\gamma \neq \gamma'$ , we have  $\gamma \succ \bullet \gamma'$ , so  $E(\gamma') \leq E(\gamma) + 1 - w$ . So  $E_{\text{max}}(p') \leq E(\gamma) + 2 - w$ , and the result follows from the identity  $x \leq \max\{x, y\}$ .  $\square$

To saturate a full path, we saturate each step.

**Theorem 7.** *Let bond strength  $w \geq 1$  and  $\gamma$  and  $\delta$  be saturated. Then*

$$b_{\text{sat}}(\gamma, \delta) \leq b(\gamma, \delta) + \max\{0, 2 - w\}.$$

*Proof.* Let  $\alpha_1$  and  $\alpha_n$  be saturated. Consider a path  $p = \alpha_1, \dots, \alpha_n$ . Let  $\alpha'_1 = \alpha_1$ . Then  $\alpha'_1 \in [\alpha_1]$ . So by Lemmas 4 and 6, for each  $i$  we get  $\alpha'_{i+1} \in [\alpha_{i+1}]$  and saturated  $p'_i$  from  $\alpha'_i$  to  $\alpha'_{i+1}$  with  $E_{\text{max}}(p'_i) \leq \max\{E(\alpha_i), E(\alpha_{i+1})\} + M_w$  where  $M_w = \max\{0, 2 - w\}$ . Now  $\alpha_n$  is saturated, so  $\alpha'_n = \alpha_n$ . So let  $p'$  be the concatenation of the  $p'_i$ . Then  $p'$  is a saturated path from  $\alpha_1$  to  $\alpha_n$ . And we have

$$\begin{aligned} E_{\text{max}}(p') &= \max_i E_{\text{max}}(p'_i) \\ &\leq \max_i \max\{E(\alpha_i), E(\alpha_{i+1})\} + M_w \\ &= \max_i E(\alpha_i) + M_w \\ &= E_{\text{max}}(p) + M_w. \end{aligned}$$

So  $h(p') \leq h(p) + M_w$ . So  $b_{\text{sat}}(\alpha_1, \alpha_n) \leq b(\alpha_1, \alpha_n) + M_w$ .  $\square$

Notice that we need bond strength  $w \geq 1$  in Theorem 7. If  $w < 1$ , then  $b_{\text{sat}}(\gamma, \delta)$  can be larger than  $b(\gamma, \delta)$  by an arbitrary amount.

Also notice that  $\max\{0, 2 - w\}$  is tight. To see this, the reader may check that  $b_{\text{sat}}(\gamma, \delta) = b(\gamma, \delta) + \max\{0, 2 - w\}$  for the following example:  $\gamma = \{\{m_1, m_2\}, \{m_3\}\}$  and  $\delta = \{\{m_1\}, \{m_2, m_3\}\}$  where  $m_1 = \{a, b\}$ ,  $m_2 = \{a^*\}$ ,  $m_3 = \{a, c\}$ .

Now since  $b_{\text{sat}}(\gamma, \delta) \geq b(\gamma, \delta)$ , we have the following corollary of Theorem 7, which is the main result of this section.

**Corollary 8.** *Let bond strength  $w \geq 2$  and  $\gamma$  and  $\delta$  be saturated. Then  $b_{\text{sat}}(\gamma, \delta) = b(\gamma, \delta)$ .*

## 4 TBNs with programmable energy barriers

In this section we present two constructions. Each is a family of TBNs indexed by an integer  $n$ . We call certain configurations of those TBNs *initial* and *triggered* and show an energy barrier between them. As  $n$  increases, the size of the energy barrier increases linearly. Each also has a catalyst, which reduces the energy barrier to 1 when added.

The first construction (translator cycle), discussed in Section 4.1, is based on a DNA strand displacement catalyst. Progress from the initial to triggered configurations with the catalyst can be physically implemented as a strand displacement cascade. Although this system has been previously proposed [12], [13], we rigorously prove an energy barrier for the first time.

The second construction (grid gate), discussed in Section 4.2, does not have an evident physical implementation (e.g., as a strand displacement system), but surpasses the translator cycle system in two ways. First, the grid gate can exhibit *autocatalysis*—that is, it can be modified so that the catalyst transforms the initial polymer into a polymer that has the same exposed binding sites as the catalyst, which can itself catalyze the transformation of additional initial polymers (leading to exponential amplification). Second, the grid gate

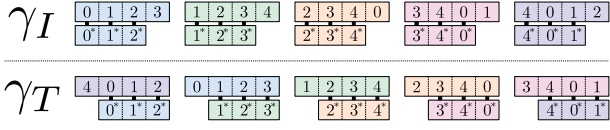


Fig. 3. The two stable configurations of a translator cycle with complex length  $z = 3$  and number of complex types  $c = 5$ . In place of binding site  $x_i$ , we write  $i$  for clarity.

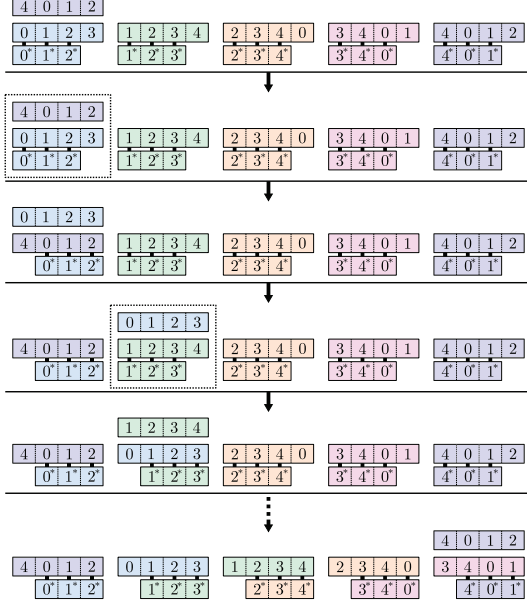


Fig. 4. A segment of the height 1 path which is possible because an extra copy of a top monomer,  $\{x_4, x_0, x_1, x_2\}$ , is present to act as a catalyst. The dotted arrow signifies a sequence of merge/split steps. In place of binding site  $x_i$ , we write  $i$  for clarity.

is *self-stabilizing*, which we define to mean that from any configuration, there is a barrier of zero to reach either an initial or triggered configuration. This intuitively ensures that the system cannot get stuck in an undesired local energy minimum.

Throughout both sections, we assume  $w \geq 2$ , so that by Corollary 8, we can determine energy barriers by considering only saturated paths. If we weaken this assumption to  $w \geq 1$ , then by Theorem 7 the barrier proved is within 1 of the barrier in the unrestricted pathway model (allowing unsaturated configurations). We believe that for these systems an  $\Omega(n)$  energy barrier exists even if  $w < 1$  but sufficiently large. However, studying the  $w < 1$  regime is left for future work (see Section 6).

The constructions demonstrate that catalysts and autocatalysts with arbitrarily high energy barriers can be engineered solely by reference to the general thermodynamic driving forces of binding and formation of separate complexes, which are captured in the TBN model.

#### 4.1 Translator cycle

Consider the TBN illustrated in Fig. 3. There are two particular configurations that interest us, an initial configuration  $\gamma_I$  and a triggered configuration  $\gamma_T$ . The two configurations are stable. In the presence of a catalyst monomer  $\{x_4, x_0, x_1, x_2\}$  (or an extra copy of any *top monomer*—any of the monomers with unstarred binding sites), a height one pathway exists to reach

$\gamma_T$ , illustrated by Fig. 4. If the catalyst is not present, we prove there is a barrier which can be made arbitrarily large by including more and longer monomer types. Further, this catalytic system is realizable as a DNA strand displacement cascade; more information about this connection can be found in Appendix A.1, and in [13]. In the case of many copies of each complex, since the catalyst is in fact any of the top monomers, the system may be used as an amplifier: at the end of the pathway shown in Fig. 4, another monomer with binding sites  $\{x_4, x_0, x_1, x_2\}$  becomes free which can catalyze another set of complexes which are in the initial configuration.

To program a large energy barrier, we give a formal definition for generalizing the translator cycle, parameterized by *complex length*  $z$  and *number of complex types*  $c$ . Given  $z \leq c$ , a  $(z, c)$ -translator cycle is a TBN with monomer types  $t_i$  (*top monomers*) and  $b_i$  (*bottom monomers*) for  $i \in \mathbb{Z}_c$ , where

$$t_i = \{x_i, x_{i+1 \pmod{c}}, \dots, x_{i+z \pmod{c}}\},$$

$$b_i = \{x_i^*, x_{i+1 \pmod{c}}^*, \dots, x_{i+z-1 \pmod{c}}^*\}.$$

A  $(z, c)$ -translator cycle may have any number of each monomer type as long as (1) for all  $i$ , the number of  $t_i$  is equal to the number of  $b_i$  and (2) there is at least one of each  $t_i$  and  $b_i$ . To justify constraint (1) note that including an extra top monomer can catalyze the cycle so the barrier disappears, while extra bottom monomers merge the two-monomer complexes to saturate, disrupting the desired initial and triggered configurations. Constraint (2) is required for the catalytic pathway (Fig. 4).

The *initial configuration* has each  $b_i$  in a polymer  $\{b_i, t_i\}$ , and a *triggered configuration*  $\gamma_T$  is any saturated configuration which contains a subset  $\{\{b_i, t_{i-1}\} \mid i \in \mathbb{Z}_c\}$  (at least one set of complexes in the triggered state). The rest of this section is dedicated to proving that the barrier between  $\gamma_I$  and  $\gamma_T$  depends on the complex length  $z$  and the number of complex types  $c$ , and can be made arbitrarily large: Formally, we prove that if  $z^2 = c$ , then  $b(\gamma_I, \gamma_T) > \frac{z}{2+z-1}$ .

To motivate our choice of  $z^2 = c$ , it is worth describing three available pathways that influence how the upper bound for the uncatalyzed barrier scales with  $z$  and  $c$ . For the first path, we can reach a configuration with a free top monomer, which can subsequently be used as a catalyst as in Fig. 4. For example, merging the three polymers with exposed sites  $x_4, x_0$ , and  $x_1$  with the polymer  $\{\{x_4, x_0, x_1, x_2\}, \{x_4^*, x_0^*, x_1^*\}\}$  allows the top monomer  $\{x_4, x_0, x_1, x_2\}$  to be split. In the general case, this requires merging  $z$  polymers and then following a height 1 path, and the path in total has height  $z$ . The second path brings all complexes together while reducing the number of polymers by  $c - 1$ , and then splits them into the triggered complexes, resulting in a height  $c - 1$  path. These paths show that the barrier is not larger than  $z$  or  $c - 1$ . Surprisingly, it is still not sufficient to set  $z = c - 1 = n$  to attain a barrier of  $n$ ; there is a complicated third path which has height  $\frac{2c}{z} - 1$  which is illustrated in the appendix in Fig. 10. The third path shows that  $c$  must be asymptotically larger than  $z$  to achieve a non-constant barrier. Thus we fix  $z = n$  and  $c = n^2$  for the remainder of the section.

Before we get into details, we give an overview of the proof that  $b(\gamma_I, \gamma_T) > \frac{n}{2+n-1}$ . First, we show that we can restrict our attention to configurations where polymers must have the same amount of top monomers as bottom monomers (denoted



as *normal form*), since other configurations have low polymer count. We think of pairing top and bottom monomers in normal form polymers. Initially, top and bottom monomers with the same indices are paired. In the triggered configuration, the top and bottom monomers are paired with different indices, notably  $t_i$  is bound to  $b_{i+1}$ ; we say the top index is offset “to the left” of the bottom index by one, or has offset “minus one”. We will formalize this notion of offset, and show that the sum of all offsets between pairs in the configuration, initially zero, does not change with merges and splits in paths of normal form configurations. In the single-copy case, this contradicts any path which reaches a triggered configuration, which must have a negative total offset.

In the multi-copy case, the negative offset of the triggered complexes can be canceled by positive offset elsewhere, resulting in zero total offset, so the argument is not as simple. We will show that polymers providing net positive offset have a size which grows along with the offset. The large size of these polymers then implies a barrier.

First, we provide an easy way to count how many complexes are in a configuration of a cycle. Note that since there is at least one unstarred binding site for each starred site, the starred sites *must* be bound in a saturated configuration. So we call the starred sites *limiting*. We can use this fact to argue about the number of separate polymers in saturated configurations. Since the starred binding sites are limiting and  $\gamma$  is saturated, each bottom monomer must be bound to at least one top monomer, so we can count  $S(\gamma)$  by counting the number of top monomers in separate polymers. This leads to:

**Observation 9.** *Given a  $(z, c)$ -translator cycle with  $k$  instances of top monomers, in a saturated configuration  $\gamma$ , if there is a polymer with  $m$  top monomers, then  $S(\gamma) \leq k - m + 1$ .*

Now we restrict the configurations and paths we must consider by describing a *normal form* for polymers.

**Definition 10.** *Given a configuration of an  $(n, n^2)$ -translator cycle, a polymer is normal form if its number of top monomers is equal to its number of bottom monomers. A configuration is normal form if every polymer is normal form. A path is normal form if every configuration is normal form.*

Normal form paths are more restricted than arbitrary paths, and will be easier to reason about. To motivate them, we show that saturated paths from  $\gamma_I$  to  $\gamma_T$  that are not normal form must have a large height via a large polymer in some configuration, and so low height paths (if they existed) would be normal form.

The following lemma is a technical fact used in proofs of Lemmas 12 and 13. It gives properties for polymers in saturated configurations with  $x$  bottom monomers with  $n$  binding sites each, and  $y$  top monomers with  $(n + 1)$  sites each, with  $x > y$ . This will help us restrict to normal form polymers, which have  $x = y$ .

**Lemma 11.** *Assume  $x, y, n \in \mathbb{N}$ ,  $y(n + 1) \geq xn$ , and  $x > y$ . Then  $y \geq n$  and  $x \geq n + 1$ .*

*Proof.* If  $x > y$ , then  $x \geq y + 1$ , so  $y(n + 1) \geq xn \geq (y + 1)n$ . So  $yn + y \geq yn + n$ . So  $y \geq n$ . Since  $x > y$ ,  $x \geq n + 1$ .  $\square$

The next lemma shows the saturated configurations which are not normal form must have a large (size  $\Omega(n)$ ) polymer.

**Lemma 12.** *If a saturated configuration is not normal form, then some polymer has at least  $n$  top monomers.*

*Proof.* Since the configuration is not normal form, some polymer  $P$  has either fewer or more top than bottom monomers. If  $P$  has fewer, then let  $P' = P$ . If  $P$  has more, then some other polymer  $P'$  has fewer top monomers than bottom monomers. Let  $t$  (resp.,  $b$ ) be the number of top (resp., bottom) monomers in  $P'$ . The number of unstarred (resp., starred) binding sites in  $P'$  is  $t(n + 1)$  (resp.,  $bn$ ). Recall that the starred sites are limiting. So for each starred site in  $P'$ , there is at least one corresponding unstarred site, so  $bn \leq t(n + 1)$ . Since  $P'$  has  $b > t$ , Lemma 11 gives  $t \geq n$ , so  $P'$  has at least  $n$  top monomers.  $\square$

So saturated paths with low height must consist of normal form configurations, since otherwise they would have a polymer with many top monomers which implies a large height by Observation 9.

Now we formalize the offset of a pair of compatible monomers. (Recall that two monomers are compatible if they have complementary binding sites.) For  $k \in \mathbb{N}$  and  $a, b \in \mathbb{Z}_k$ , define the sequence

$$[a, b]_k = \langle a, a + 1, \dots, b \rangle \pmod{k}.$$

For example,  $[1, 3]_5 = \langle 1, 2, 3 \rangle$  and  $[3, 1]_5 = \langle 3, 4, 0, 1 \rangle$ . Also let  $\ell_S$  be the index of element  $\ell$  in sequence  $S$ . Then for monomers  $b_i$  and  $t_j$ , we define the *offset* to be  $f(b_i, t_j) = j_S - i_S$ , where  $S = [i - n, i + n - 1]_{n^2}$ .

We will define the offset of a normal form polymer in terms of compatible pairs of top and bottom monomers. To choose the pairs, we use the notion of *matchings* from graph theory. Given a normal form polymer  $P$ , let  $T$  be the set of top monomers and  $B$  be the set of bottom monomers. Define a bipartite graph  $G_P = \{T, B, E\}$  where  $\{b_i, t_j\} \in E$  if and only if  $b_i$  and  $t_j$  are compatible.

**Lemma 13.** *If a polymer  $P$  is in a normal form saturated configuration and has size  $|P| < 2n + 1$ , then there exists a perfect matching on  $G_P$ .*

The proof of Lemma 13 is in Appendix A.2.

For a perfect matching  $M$  on  $G_P$ , the matching offset is  $f(M) = \sum_{m \in M} f(m)$ . It will turn out that for small polymers, the offset of every perfect matching is the same. So we will use the matching offset to define the offset of a polymer. To do so, it will be useful to define a notion of “leftmost” and “rightmost” monomers in a polymer. For small polymers (of size less than about  $n$ ), these are intuitively well-defined since there are not enough monomers to “wrap-around” the entire cycle. We formalize this notion via a *cutoff* value for a polymer:

**Definition 14.** *Given a normal form polymer  $P$ , a cutoff value  $c_P \in \mathbb{Z}_{n^2}$  satisfies the following: let  $C$  be  $[c_P, c_P - 1]_{n^2}$ , then there is no edge  $\{b_i, t_j\} \in G_P$  such that  $i_C \leq n$  and  $j_C > n^2 - n$  or  $j_C \leq n$  and  $i_C > n^2 - n$  (recall that  $i_C$  denotes the index of  $i$  in  $C$ ).*

If a cutoff value exists, it means no possible bonds in the polymer—and equivalently no possible edges in any matching—cross the cutoff. Then the leftmost and rightmost pairs are easily defined with respect to the cutoff sequence  $[c_P, c_P - 1]_{n^2}$ . We prove a sufficient condition for there to exist a cutoff point:

**Lemma 15.** *For a normal form polymer  $P$  with size  $|P| < \frac{2n^2}{2n+1}$ , there exists a cutoff value  $c_P$ .*

*Proof.* First, we give the reasoning behind the choice of  $|P| < \frac{2n^2}{2n+1}$ . Intuitively, we want to choose a cutoff point at an index where the top monomer with that index is not compatible for any bottom monomer in the polymer. If there are  $k$  bottom monomers in a polymer, the union of the sets of compatible top monomers for those bottom polymers is at most  $k(2n+1)$ . If we have  $k(2n+1) < n^2$ , there will exist an index for a top monomer which is not compatible to any bottom in the polymer. This gives  $k < \frac{n^2}{2n+1}$ , and there are  $2k$  monomers in the polymer, so we must have  $|P| = 2k < \frac{2n^2}{2n+1}$ . Then let  $c_P$  be the index of that incompatible top monomer.

We now show that  $c_P$  is in fact a cutoff point as in Definition 14. Towards contradiction, let  $C$  be  $[c_P, c_P - 1]_{n^2}$  and assume there does exist an edge  $\{b_i, t_j\} \in G_P$  such that  $i_C \leq n$  and  $j_C > n^2 - n$  or  $j_C \leq n$  and  $i_C > n^2 - n$ . Then  $t_j$  is compatible for  $b_i$ , but  $c_P$  is in between  $i$  and  $j$ , so then  $t_{c_P}$  is a compatible top monomer for  $b_i$ , which contradicts our choice of  $c_P$  as an index of an incompatible top monomer for any bottom monomer in the polymer.  $\square$

To use the cutoff value in future lemmas, we consider configurations and paths with polymers' size restricted to less than  $\frac{2n^2}{2n+1}$ . We let  $n' = \frac{2n^2}{2n+1}$  and define the following:

**Definition 16.** *A polymer is  $n'$ -bounded if its size is less than  $n'$ . A configuration is  $n'$ -bounded if every polymer is  $n'$ -bounded. A path is  $n'$ -bounded if every configuration is  $n'$ -bounded.*

We can use a cutoff to show that the offset of every matching is the same.

**Lemma 17.** *For an  $n'$ -bounded polymer  $P$  in a normal form saturated configuration, for any two perfect matchings  $M$  and  $M'$  on  $G_P$ ,  $f(M) = f(M')$ .*

The proof of Lemma 17 is in Appendix A.2.

So we define the *polymer offset* of  $P$ ,  $f(P)$ , simply as the offset of any perfect matching on  $G_P$ . Given a configuration  $\gamma$ , we can define the *configuration offset* as  $\sum_{P \in \gamma} f(P)$ . We now show that under certain conditions, merges and splits do not change the configuration offset.

**Lemma 18.** *In a normal form saturated  $n'$ -bounded path, if  $\gamma \succ^{-1} \delta$ , then  $f(\gamma) = f(\delta)$ .*

*Proof.* Consider the two polymers which merge,  $P_1$  and  $P_2$ , and call the polymer which is their union  $P$ . Let  $M_{P_1}$  and  $M_{P_2}$  be any perfect matchings of  $G_{P_1}$  and  $G_{P_2}$ . Then  $M_P = M_{P_1} \cup M_{P_2}$  is a perfect matching on  $G_P$ , and  $f(M_P) = f(M_{P_1}) + f(M_{P_2})$ . Since the polymers are  $n'$ -bounded, we have by Lemma 17 that the polymer offsets equal the offsets of any matching, so  $f(P) = f(P_1) + f(P_2)$ . Then  $f(\gamma) = f(\delta)$  since their only different summands are  $f(P)$  and  $f(P_1), f(P_2)$ .  $\square$

From the above lemma, one can prove that in the case of one copy of each monomer type, the  $(n, n^2)$ -translator cycle has a barrier of  $\Omega(n)$  to trigger. An informal proof follows: any path which is not normal form or does not have small polymers must have a large height due to a polymer with many top monomers (see Observation 9). Otherwise, if we restrict paths to saturated normal form  $n'$ -bounded paths, Lemma 18 gives that from the initial configuration offset of zero, there

is no path which can change the configuration offset to  $-n^2$ , which is the offset of the triggered configuration. We leave out a formal statement of this proof for brevity, as we prove a more general theorem later—a barrier in the multi-copy case.

The above argument is not sufficient in the multi-copy case because the  $-n^2$  offset given by triggered polymers can be canceled out by positive  $n^2$  offset in other polymers. We will argue that having a large positive offset is (roughly) proportional to having a large polymer or several large polymers, so the  $n^2$  positive offset would require merging many complexes. To do so, we define and prove existence of a *sorted* matching on polymers. Intuitively, a sorted matching is a matching which has no crossing edges when the indices are sorted.

**Lemma 19.** *Given an  $n'$ -bounded polymer  $P$  with cutoff  $c_P$ , there exists a sorted matching  $M$  on  $G_P$  which satisfies that there does not exist  $\{b_{i_1}, t_{j_2}\}, \{b_{i_2}, t_{j_1}\} \in M$  with  $i_1 \leq i_2$  and  $j_1 \leq j_2$  with respect to the ordering given by the cutoff value sequence,  $[c_P, c_P - 1]_{n^2}$ .*

The proof of Lemma 19 is in Appendix A.2.

Using the sorted matching, we show that the maximum offset (if it is positive) of any one pair in a polymer is proportional to the size of the polymer. First we relate exposed sites to the size of a polymer, then relate the exposed sites to the maximum offset.

**Lemma 20.** *If a polymer  $P$  in a saturated normal form configuration has  $k$  exposed sites, it has size  $2k$ .*

*Proof.* Assume  $P$  is of size  $2s$  for some  $s$ . Then  $P$  has  $sn$  starred domains which must be bound in a saturated configuration.  $P$  has  $s(n+1)$  unstarred domains. So  $P$  has exactly  $s$  exposed sites for any  $s$ . Then to have  $k$  exposed sites, it must have size  $2k$ .  $\square$

**Lemma 21.** *Given a normal form  $n'$ -bounded polymer  $P$ , consider the sorted matching  $M$  of  $G_P$ . Let  $m$  be the value of the maximum offset of any pair in  $M$ , then  $|P| \geq 2(m+1)$ .*

*Proof.* We will show that  $P$  has at least  $m+1$  exposed sites, and thus by Lemma 20 is of size at least  $2(m+1)$ . Consider the pairs ordered by smallest bottom index to largest with respect to the cutoff  $c_P$  given by Lemma 15. Imagine constructing  $P$  by adding one pair at a time in order. We will show that when adding a pair, the number of exposed sites cannot decrease due to the order, and when we add the pair with offset  $m$ , the constructed polymer has  $m+1$  exposed sites.

First, consider adding a pair  $p$  with nonnegative offset  $f(p)$  to a polymer with  $k$  exposed unstarred sites. Note that the polymer containing only the two monomers in the pair  $p$  has  $f(p)+1$  exposed unstarred sites, and  $f(p)$  exposed starred sites. By the ordering, we know that the  $f(p)+1$  exposed unstarred sites cannot be bound by any bottom monomers in the polymer constructed thus far. Further, at most  $f(p)$  of the  $k$  exposed unstarred sites on the polymer prior to adding  $p$  can be bound after adding  $p$ , since only  $f(p)$  exposed starred sites are added. Thus the net change in exposed unstarred sites is plus one. Further, the number of exposed sites on the constructed polymer after adding  $p$  is at least  $f(p)+1$ .

Next, consider adding a pair  $p$  with negative offset to a polymer with  $k$  exposed unstarred sites. Note that due to the ordering, the bottom monomer in the pair has no domains in common with the unstarred exposed sites of the polymer

constructed thus far. So the number of exposed sites does not decrease.

In both cases, the number of exposed unstarred sites cannot decrease by adding polymers. Consider the point in this construction where we have just added the polymer with positive offset  $m$ . The constructed polymer thus far has the  $m + 1$  exposed unstarred sites given by the offset  $m$ , and the number of exposed sites cannot be reduced by adding the remaining polymers, so the final polymer  $P$  must have at least  $m + 1$  exposed sites. So by Lemma 20,  $|P| \geq 2(m + 1)$ .  $\square$

We prove two more lemmas before the proof of the barrier of the translator cycle. The first key lemma is that triggered polymers' negative offset must be canceled out by polymers with positive offset, but since positive offset results in large polymers (or many slightly larger polymers), such a configuration implies a large height for the path which contains it.

**Lemma 22.** *Given a normal form saturated  $n'$ -bounded configuration  $\gamma$ , if there exists a subset of polymers  $\mathcal{P} = \{P_1, \dots, P_k\}$  such that  $\sum_{P_i \in \mathcal{P}} f(P_i) \geq n^2$ , then  $S(\gamma_I) - S(\gamma) > 2n + 1$ .*

*Proof.* Consider  $\gamma$  and for each polymer  $P_i$  in  $\gamma$ , fix any sorted matching  $M_i$  on  $G_{P_i}$  given by Lemma 19 and denote the set of  $M_i$  by  $\mathcal{M}$ . Then for each  $P_i \in \mathcal{P}$ ,

$$f(P_i) \leq \frac{|P_i|}{2} \max_{p \in M_i} f(p),$$

since there are  $\frac{|P_i|}{2}$  pairs each with offset at most the max over the offsets. Since  $\gamma$  is  $n'$ -bounded,  $|P_i| < n'$ , so

$$f(P_i) < \frac{n'}{2} \max_{p \in M_i} f(p).$$

So we have the following:

$$\begin{aligned} \sum_{M_i \in \mathcal{M}} \frac{n'}{2} \max_{p \in M_i} f(p) &= \frac{n'}{2} \sum_{M_i \in \mathcal{M}} \max_{p \in M_i} f(p) \\ &> \sum_{P_i \in \mathcal{P}} f(P_i) \\ &\geq n^2, \end{aligned}$$

and further

$$\sum_{M_i \in \mathcal{M}} \max_{p \in M_i} f(p) > \frac{2n^2}{n'}. \quad (3)$$

Now we show that

$$S(\gamma_I) - S(\gamma) \geq \sum_{M_i \in \mathcal{M}} \max_{p \in M_i} f(p).$$

Consider  $\gamma'$ , the (unsaturated) configuration which is given by taking the polymers in  $\gamma$  and splitting them into pairs of top and bottom monomers based on the matchings  $M_i$ . Each bottom monomer is in a polymer with exactly one top monomer, so  $S(\gamma_I) = S(\gamma')$ . For each  $P_i$  with sorted matching  $M_i$  in  $\mathcal{P}$ , consider the pair  $p$  satisfying  $\max_{p \in M_i} f(p)$ . We know that in  $\gamma$ , each polymer  $P_i$  with  $p \in M_i$  must have size at least  $2(f(p) + 1)$  by Lemma 21. So  $P_i$  must be a polymer containing at least  $f(p)$  other pairs. So

$$S(\gamma') - S(\gamma) \geq \sum_{M_i \in \mathcal{M}} \max_{p \in M_i} f(p).$$

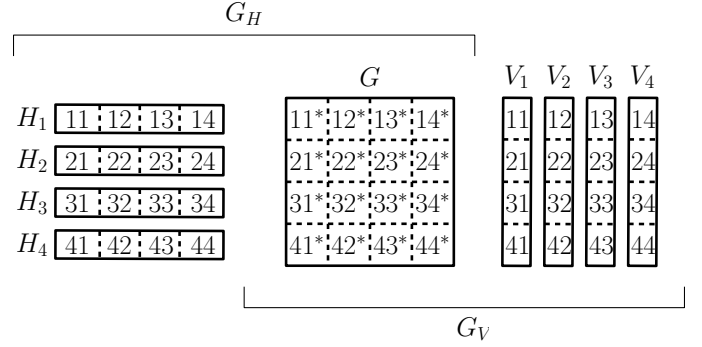


Fig. 5. The monomer types in the grid gate TBN for the case  $n = 4$ . In the figure, any two digit number  $ij$  represents domain  $x_{ij}$ , e.g.  $x_{23}^*$  is represented as  $23^*$ .

Since  $S(\gamma_I) = S(\gamma')$ , (3) gives us  $S(\gamma_I) - S(\gamma) > \frac{2n^2}{n'} = 2n + 1$ .  $\square$

**Theorem 23.** *Given an  $(n, n^2)$ -translator cycle,  $b(\gamma_I, \gamma_T) \geq \frac{n}{2+n-1}$ .*

*Proof.* We split the possible paths from  $\gamma_I$  to  $\gamma_T$  into three cases. Case 1: if a saturated path is normal form and  $n'$ -bounded, we have by Lemma 18 that for any  $\gamma_i$  in the path,  $f(\gamma_i) = f(\gamma_I) = 0$ . Consider the final configuration on the path, the triggered configuration  $\gamma_T$ . By definition, in  $\gamma_T$  there exists at least one pair of each  $b_i$  bound to  $t_{i-1} \pmod{n^2}$ . Each of these pairs has offset minus one, contributing minus  $n^2$  to the offset. However, we know that the configuration offset must be zero. So there must exist a subset of polymers  $\mathcal{P} = \{P_1, \dots, P_k\}$  such that  $\sum_{i=1}^k f(P_i) \geq n^2$ , so by Lemma 22,  $S(\gamma_I) - S(\gamma_T) > 2n + 1$ .

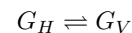
Case 2: if a path is not normal form, then by Lemma 12, there exists a configuration  $\gamma$  with a polymer with  $n$  top monomers. Let  $k$  be the total number of top monomers in the cycle and note that  $S(\gamma_I) = k$ . Then by Observation 9,  $S(\gamma) \leq k - n + 1$ . Then  $S(\gamma_I) - S(\gamma) \geq n - 1$ .

Case 3: if a path is normal form but is not  $n'$ -bounded, then by definition there exists a polymer of size at least  $n'$  with an equal number of top and bottom monomers; i.e., a polymer with at least  $\frac{n'}{2}$  top monomers. Let  $k$  be the total number of top monomers in the cycle and note that  $S(\gamma_I) = k$ . Then by Observation 9,  $S(\gamma) \leq k - \frac{n'}{2} + 1$ . Note that  $S(\gamma_I) = k$ . Then  $S(\gamma_I) - S(\gamma) \geq \frac{n'}{2} - 1 = \frac{n}{2n+1} - 1$ .

By Corollary 8, we restrict analysis to saturated paths. Then  $H(\gamma_I) = H(\gamma)$ , and so  $E(\gamma) - E(\gamma_I) = -(S(\gamma_I) - S(\gamma))$ . Among the three cases, the smallest lower bound on the height is  $\frac{n^2}{2n+1}$ , so the barrier is at least  $\frac{n^2}{2n+1} = \frac{n}{2+n-1}$ .  $\square$

## 4.2 Grid gate

Consider the TBN illustrated in Fig. 5. We focus on two polymer types  $G_H$  and  $G_V$  depicted in the figure, and show that there is a barrier  $n \in \mathbb{N}$  to convert  $G_H$  to  $G_V$  and vice versa. We parameterize the construction by  $n$  as follows. Define the following monomer types: “horizontal”  $H_i = \{x_{ij}\}_{j=1}^n$  for  $i \in \{1, \dots, n\}$ , “vertical”  $V_j = \{x_{ij}\}_{i=1}^n$  for  $j \in \{1, \dots, n\}$ , and “gate”  $G = \{x_{ij}^*\}_{i,j=1}^n$ . In the notation of chemical reaction networks, the net reaction





can occur in the presence of sufficiently many free  $H_i$ 's and  $V_j$ 's, but an energy barrier of  $n$  must be surmounted in order for this conversion to happen. In Section 4.2.1 we will show how this energy barrier can be reduced to 1 in the presence of a *catalyst* monomer, corresponding to the chemical notion of a catalyst reducing the activation energy required for a reaction to occur.

Note that throughout this section, the configurations considered are saturated, so that for two configurations  $\gamma$  and  $\delta$ , we have  $H(\gamma) = H(\delta)$ , and so  $E(\delta) - E(\gamma) = -(S(\delta) - S(\gamma))$ . That is, the energy difference between two configurations is the opposite of the difference in their polymer counts.

We fix a network  $\mathcal{T}_{gg}$  that contains any number of any of these monomer types, so long as there are enough of other monomers to completely bind all the  $G$  monomers (i.e., in saturated configurations there are no exposed starred sites). We define *base* configurations of the network to be those configurations that contain polymers of type  $G_H$  or  $G_V$ , with all other monomers in separate polymers by themselves. In Theorem 30 we show that these base configurations are stable (take  $m = 0$ ).

The following lemma establishes a necessary condition in any saturated configuration: that any  $G$  must be in a polymer with either all of the horizontal monomers or all of the vertical monomers.

**Lemma 24.** *In a saturated configuration of  $\mathcal{T}_{gg}$ , a polymer containing  $G$  also contains  $\{H_i\}_{i=1}^n$  or  $\{V_j\}_{j=1}^n$  as a subset.*

*Proof.* Suppose a polymer contains  $G$  but neither  $H_i$  nor  $V_j$  for some  $i$  and  $j$ . Then site  $x_{ij}^*$  on  $G$  is exposed, and so by definition of  $\mathcal{T}_{gg}$ , the configuration is not saturated.  $\square$

The following theorem then establishes that any saturated configuration in  $\mathcal{T}_{gg}$  is *self-stabilizing*, that is, it can reach a stable (base) configuration via a path with barrier 0 (i.e. using all splits).

**Theorem 25.** *For any saturated  $\gamma$  of  $\mathcal{T}_{gg}$ , some base  $\pi$  has  $\gamma \preceq_\circ \pi$ .*

*Proof.* Consider a saturated configuration  $\gamma$ . Suppose  $\gamma$  has a non-base polymer  $P$ . If  $P$  contains no  $G$ , then we can split into polymers of type  $H, V$ . Otherwise,  $P$  contains  $G$ , and by Lemma 24 we can split off a  $G_V$  or  $G_H$  polymer. The theorem holds by repeating this process.  $\square$

Note that since the base configurations are stable (this follows from Theorem 30 with  $m = 0$ ), and by Theorem 25 any other saturated configuration can reach a base configuration using only splits, the base configurations are also the only stable configurations in this network.

We now prove the desired energy barrier between different base configurations.

**Theorem 26.** *The barrier between different base configurations of  $\mathcal{T}_{gg}$  is  $n$ .*

*Proof.* Consider a saturated path  $p$  from a base configuration  $\gamma$  to another,  $\delta$ . Notice that  $\delta \not\preceq \gamma$ . So  $\delta \not\preceq_\circ \gamma$ . So some first  $\beta$  along  $p$  has  $\beta \not\preceq_\circ \gamma$ . But by Theorem 25, some other base  $\pi \neq \gamma$  does have  $\beta \preceq_\circ \pi$ .

Now take  $\alpha$  just before  $\beta$  along  $p$ . Then  $\alpha \preceq_\circ \gamma$  by definition of  $\beta$ . Since  $\alpha$  and  $\beta$  are adjacent on  $p$ , either  $\alpha \prec_\circ^1 \beta$  or  $\alpha \succ_\circ^1 \beta$ . The latter contradicts  $\beta \not\preceq_\circ \gamma$ . So  $\alpha \prec_\circ^1 \beta$ , implying  $\alpha \preceq_\circ \pi$ .

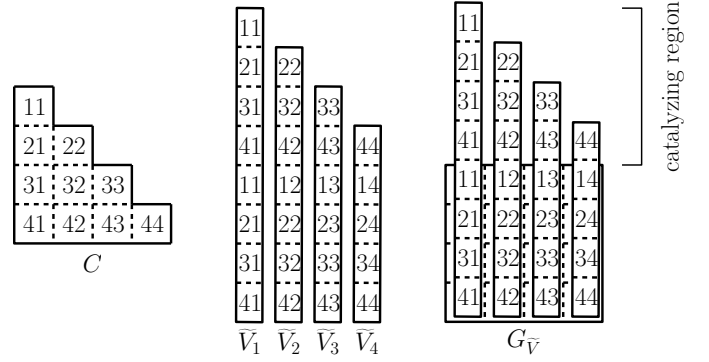


Fig. 6. Catalysts and autocatalysts in the grid gate TBN for the case  $n = 4$ . left:  $C$  is a single monomer that acts as a catalyst to convert between  $G_H$  and  $G_V$ . middle: Modified vertical monomers  $\{\tilde{V}_j\}_{j=1}^n$  with extra sites. right: After  $C$  converts  $G_H$  to  $G_V$  with modified vertical monomers,  $G_V$  has the same excess sites as  $C$  and acts as a catalyst itself (i.e. is “active” as a catalyst).

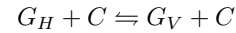
Let  $f(\gamma)$  count the  $H$  monomers with a  $G$  in  $\gamma$ . Since  $\gamma$  and  $\pi$  are different bases,  $wlog$ ,  $f(\pi) \geq n + f(\gamma)$ . Consider a path of  $k$  merges corresponding to  $\gamma \succeq_\circ \alpha$ . It can increase  $f(\cdot)$  by at most  $k$ . So  $f(\alpha) \leq f(\gamma) + k$ . A path of splits does not increase  $f(\cdot)$ , so  $\alpha \preceq_\circ \pi$  implies  $f(\alpha) \geq f(\pi)$ . So we get

$$\begin{aligned} k &\geq f(\alpha) - f(\gamma) \\ &\geq f(\pi) - f(\gamma) \geq n. \end{aligned}$$

So  $S(\gamma) - S(\alpha) = k \geq n$ .  $\square$

#### 4.2.1 Catalysis

The kinetic barrier shown for the grid gate can be disrupted by the presence of new monomer types. In fact, the model admits a *catalyst* monomer  $C$  that lowers the energy barrier from  $n$  to 1, i.e., in the presence of one or more  $C$ , a  $G_H$  can be converted into a  $G_V$ , and vice versa, with a sequence of merge-split pairs. In the notation of chemical reaction networks, this binding network implements the net reaction



with energy barrier 1, while maintaining a large energy barrier for the reaction  $G_H \rightleftharpoons G_V$ .

For the grid gate of size  $n \times n$ , we define a catalyst:  $C = \{x_{ij} \mid 1 \leq j \leq i \leq n\}$  illustrated in Fig. 6 (left).  $C$  is a monomer consisting of the “lower triangle” of the unstarred sites. The mechanism by which  $C$  can transform  $G_H$  to  $G_V$  with merge-split pairs is by an alternating processes of merges and splits shown in Fig. 7. Intuitively, in each step of the catalyzed reaction  $G_H + C \rightarrow G_V + C$ ,  $G$  switches its association with  $H_i$  (left) to its counterpart on  $V_j$  (right) by merging the evolving polymer (center) with  $V_j$  and then splitting off  $H_i$ .

Consider a network  $\{G, C\} \cup \{H_i\}_{i=1}^n \cup \{V_j\}_{j=1}^n$  which includes one instance of every monomer type that has been introduced, as well as the catalyst. As before, we shall be interested in net transitions between  $G_H$  and  $G_V$ , and so for this network we define the following configurations:  $\gamma_H^C = \{G_H, C\} \cup \{V_j\}_{j=1}^n$  and  $\gamma_V^C = \{G_V, C\} \cup \{H_i\}_{i=1}^n$ .

Theorem 27 states that transitions in the single-copy case, having arbitrarily large energy barriers according to

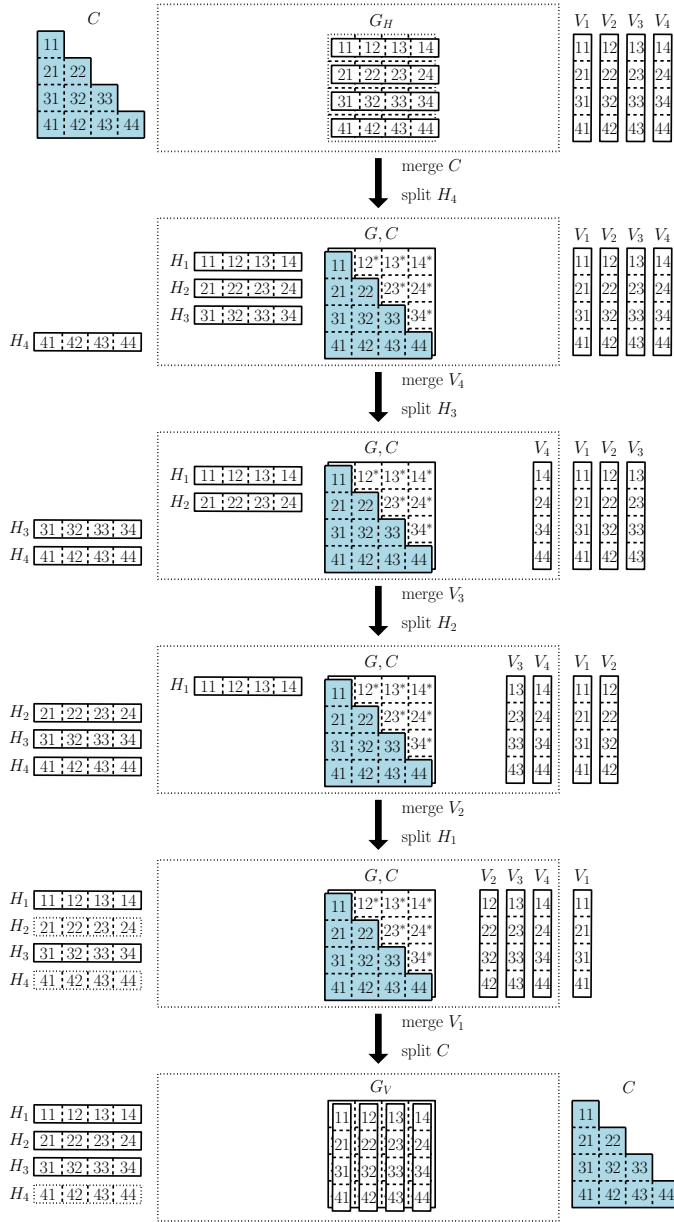


Fig. 7. Full pathway for reaction  $G_H + C \rightarrow G_V + C$ . In each stage, exactly one merge and one split occurs, and the center polymer remains saturated.

Theorem 26, in the presence of  $C$  have their barrier reduced to one.

**Theorem 27.**  $b(\gamma_H^C, \gamma_V^C) = b(\gamma_V^C, \gamma_H^C) = 1$ .

*Proof.* Consider the following saturated merge-split pathway that begins with configuration  $\gamma_H^C$  and ends with  $\gamma_V^C$  (illustrated in Figure 7).

- Merge  $G_H$  with  $C$ .
- Split  $H_n$  from the resulting polymer  $P_1$ .
- For  $1 \leq i \leq n-1$ , iteratively merge  $V_{i+1}$  with  $P_i$ , then split  $H_i$  from  $P_i$  to form  $P_{i+1}$ .
- Merge  $V_1$  with  $P_n$  and split off  $C$  to form  $G_V$ .

This path maintains saturation while never decreasing the polymer count by more than one, and so by Corollary 8 we have that  $b(\gamma_H^C, \gamma_V^C) \leq 1$ . As it is not possible to reach  $\gamma_V^C$  from

$\gamma_H^C$  in a saturated merge-split path that uses only splits, it will not be possible to have a zero barrier, giving  $b(\gamma_H^C, \gamma_V^C) = 1$ .

This merge-split path can be executed in the reverse fashion to show that  $b(\gamma_V^C, \gamma_H^C) = 1$ .  $\square$

To generalize the result to the multi-copy setting, we first observe that the height 1 pathway guaranteed in Theorem 27 still exists. What remains is to show that the base configurations, plus zero or more separate catalyst monomers, are stable.

In the arguments that follow, it will be useful to define the set  $D_{\uparrow} = \{x_{i,i+1}\}_{i=1}^{n-1} \cup \{x_{n,1}\}$ , which consists of the domains from the “shifted diagonal”. For  $m \in \mathbb{N}$ , and let  $\mathcal{T}_{gg}^m = \mathcal{T}_{gg} \cup \{m \cdot C\}$  denote the TBN  $\mathcal{T}_{gg}$  with  $m$  additional catalyst monomers.

The next lemma states that each  $G$  in a polymer in a saturated configuration must be accompanied by  $n$  additional monomers, thus giving a lower bound on the size of any polymer as a function of the number of  $G$ ’s contained within it.

**Lemma 28.** *If there are  $k$   $G$ ’s in a polymer  $P$  in a saturated configuration  $\delta$  of  $\mathcal{T}_{gg}^m$ , then  $|P| \geq k(n+1)$ .*

*Proof.* Consider the domains from  $D_{\uparrow}$ . As the starred versions of these domains are present on each  $G$ , to maintain saturation in  $\delta$ , it must be the case that each  $G$  in  $\delta$  is joined in  $P$  with a set of monomers that include these domains; however, no monomer in  $\{C, H_1, \dots, H_n, V_1, \dots, V_n\}$  has more than one of these domains. Thus, to be saturated, if there exist  $k$  instances of  $G$  in  $P$ , there must be at least  $kn$  additional monomers in  $P$  to bind the above chosen domains.  $\square$

The following lemma uses Lemma 28 to show that no saturated configuration has more polymers than are contained in the base configurations, even when some catalyst monomers are present.

**Lemma 29.** *If there are  $k$   $G$ ’s in  $\mathcal{T}_{gg}^m$ , then any saturated configuration  $\delta$  has  $S(\delta) \leq |\mathcal{T}_{gg}^m| - kn$ .*

*Proof.* Consider the polymers  $P_1, \dots, P_j$  in  $\delta$  containing all  $k$  copies of  $G$ , where  $P_i$  has  $k_i$  copies. Then by Lemma 28,  $\sum |P_i| \geq \sum k_i(n+1) = k(n+1)$ . So  $S(\delta) \leq j + |\mathcal{T}_{gg}^m| - k(n+1) \leq k + |\mathcal{T}_{gg}^m| - k(n+1) = |\mathcal{T}_{gg}^m| - kn$ .  $\square$

The next Theorem establishes that the base configurations, which were stable in the original network  $\mathcal{T}_{gg}$ , are still stable even if any number of catalysts should also be present in the network.

**Theorem 30.** *Let  $m \in \mathbb{N}$  and let  $\gamma$  be a base configuration of  $\mathcal{T}_{gg}$ . Then  $\gamma_m = \gamma \cup (m \cdot \{C\})$  is stable.*

*Proof.* The base configurations satisfy Lemma 29 with equality, thus they are stable.  $\square$

These results show that the catalyzed network is copy tolerant; that is, it behaves in the expected way even should the amounts of the constituent monomers (and catalyst) vary.

#### 4.2.2 Autocatalysis

The grid gate can also be modified to act in an *autocatalytic* manner. By modifying the vertical monomers it is possible for  $G_V$  to have a set of exposed monomers acting as a “catalyzing region”, which has the same structure and function as  $C$  (see Figure 6, middle and right).

To obtain an autocatalytic system, we modify the vertical monomers of the network to include additional sites that, when combined with  $G$ , form a catalyzing region that can act in the same manner as the catalyst  $C$ . See Figure 6 for an illustration.

Formally, we define the modified vertical monomers (see Fig. ) as:

$$\tilde{V}_j = \{x_{ij}\}_{i=1}^n \cup \{x_{ij}\}_{i=j}^n$$

We define  $G_{\tilde{V}} = \{G\} \cup \{\tilde{V}_j\}_{j=1}^n$  to be the version of  $G_V$  that uses the modified monomers. This polymer is the so-called auto-catalyst.

We now consider the network  $\tilde{T}_{gg} = \{2 \cdot G\} \cup \{H_i\}_{i=1}^n \cup \{2 \cdot \tilde{V}_j\}_{j=1}^n$  which includes enough monomers to create the autocatalyst as well as retain enough monomers to analyze transitions between  $G_H$  and  $G_{\tilde{V}}$ . The two configurations that we will be most interested in are:

$$\tilde{\gamma}_H = \{G_H, G_{\tilde{V}}\} \cup \{\tilde{V}_j\}_{j=1}^n$$

$$\tilde{\gamma}_V = \{2 \cdot G_{\tilde{V}}\} \cup \{H_i\}_{i=1}^n$$

The following theorem establishes that the autocatalyzed configurations  $\tilde{\gamma}_H$  and  $\tilde{\gamma}_V$  are stable, with a low energy barrier between them in the presence of the autocatalyst.

**Theorem 31.**  $\tilde{\gamma}_H$  and  $\tilde{\gamma}_V$  are stable.

*Proof.* Note that there are  $3n + 2$  total monomers in  $\tilde{T}_{gg}$ , and that  $S(\tilde{\gamma}_H) = S(\tilde{\gamma}_V) = n + 2$ . To show that these are stable, it suffices to show any other saturated configuration  $\delta$  obeys  $S(\delta) \leq n + 2$ . Consider the set of domains  $D_{\uparrow} = \{x_{i,i+1}\}_{i=1}^{n-1} \cup \{x_{n,1}\}$ , as in Lemma 28. Each monomer has at most one of each *type* of these domains, and except for  $x_{n,1}$ , has exactly one instance of each. The exception is  $\tilde{V}_1$ , which has two instances of  $x_{n,1}$ .

Let  $\delta$  be any saturated configuration of  $\tilde{T}_{gg}$ . We consider two cases: 1) that the  $G$ 's are in separate polymers, and 2) that the  $G$ 's are in the same polymer. In the case that the  $G$ 's are in separate polymers, as each  $G$  contains the starred versions of the  $n$  domain types in  $D_{\uparrow}$ , each polymer  $P$  with a single  $G$  must have  $n$  additional monomers to bind each of the starred versions of these domains. Note that in this case a single  $\tilde{V}_1$  cannot bind both instances of  $x_{n,1}^*$ , for this would result in both  $G$ 's being on the same polymer. This leaves at most  $(3n + 2) - 2(n + 1) = n$  remaining monomers. Thus  $S(\delta) \leq n + 2$ .

Now consider the case that the  $G$  are present in the same polymer. Let  $P$  be a polymer in  $\delta$  that contains two  $G$ 's. Then in  $P$ , the two instances of site  $x_{n,1}^*$  can be bound to one instance of  $\tilde{V}_1$ . (Note that if there were two  $\tilde{V}_1$  in this polymer, then one copy could be split into its own polymer while still maintaining saturation, increasing  $S$  and putting us in the first case). The remaining sites in  $D_{\uparrow}$  of these  $G$ 's must be bound to monomers containing  $2 \cdot \{x_{i,i+1}\}_{i=1}^{n-1}$ . No single monomer contains more than one site from this set, so this requirement must be satisfied by the inclusion of  $2(n - 1)$  additional non- $G$  monomers in  $P$ . Then there are at least  $2n + 1$  monomers in  $P$ , leaving at most  $(3n + 2) - (2n + 1) = n + 1$  additional monomers. Thus  $S(\delta) \leq 1 + (n + 1) = n + 2$ .

Since  $\delta$  was arbitrary, this shows  $\tilde{\gamma}_H$  and  $\tilde{\gamma}_V$  have maximal  $S$ , thus are stable.  $\square$

In particular, the desired feature of this network is that  $G_{\tilde{V}}$  acts as an autocatalyst.

**Theorem 32.**  $b(\tilde{\gamma}_H, \tilde{\gamma}_V) = 1$ .

*Proof.* Figure 6 shows that the exposed sites of  $G_{\tilde{V}}$  are exactly the sites of  $C$ . The proof then follows as in Theorem 27.  $\square$

Like the catalyzed network, the results for the auto-catalyzed network can be extended to the multi-copy case. The proof follows in straightforward fashion by an inductive argument.

## 5 Modeling bonds

The model, in Section 2, represents bonds implicitly. For example, as Fig. 1 shows, a single configuration can correspond to multiple ways of pairing up binding sites. This makes it easier to manipulate and reason about configurations.

But does this simplification of configurations affect the height of kinetic barriers? One might imagine that by having to manipulate individual bonds, one would need to overcome larger energy barriers than in our original model where all possible bonds are “automatically made”. Since bonds do change on an individual basis in a chemical system, this would mean that a barrier that exists in the bond-oblivious model is misleading. However, we show that the bond-aware model is essentially equivalent.

We now define the bond-aware model analogously to the definitions of Section 2.

### 5.1 Model

A *configuration*  $\gamma$  of a TBN is a matching among its complementary sites along with a partition of the components of that matching. A *polymer* of  $\gamma$  is a part of this partition. A *bond* is an edge in the matching. A configuration is *saturated* if the matching is maximal.

For a configuration  $\gamma$ , denote by  $H(\gamma)$  the total number of bonds. Let  $S(\gamma)$  and  $E(\gamma)$  be as before.

A *make* adds a missing bond in a polymer. A *break* removes an existing bond in a polymer. A three-way *swap* is changing one endpoint of a bond to another. A four-way swap is swapping the endpoint of one bond with the endpoint of another. A *path* is a sequence of configurations where each gets to the next by a merge, split, make, break, or swap.

$h(\cdot)$  is as before. Let  $b_{\text{bond}}(\gamma, \delta)$  be the barrier in the bond-aware model. Let  $b_{\text{sat-bond}}(\gamma, \delta)$  be over only paths with no break (and so no make).

### 5.2 Equivalence

We can view the original coarse kinetic model in Section 2 as a simplification of the more detailed model. To move between the two perspectives, we introduce some notation. For a polymer  $P$  of the bond-aware model, let  $\langle P \rangle$  be the collection of monomers in  $P$  (which is the corresponding polymer in the bond-oblivious model). For a configuration  $\gamma$ , let  $\langle \gamma \rangle$  be the collection of  $\langle P \rangle$  for each polymer  $P$  in  $\gamma$  (which is the corresponding configuration in the bond-oblivious model).

**Lemma 33.**  $E(\langle \alpha \rangle) \leq E(\alpha)$ .

*Proof.*  $S(\langle \alpha \rangle) = S(\alpha)$  and  $H(\langle \alpha \rangle) \geq H(\alpha)$ .  $\square$

The bond-aware model allows as much as the polymer model and more, so intuitively, a barrier in the bond-aware model should be no higher.

**Theorem 34.** *If  $E(\gamma) = E(\langle\gamma\rangle)$ , then  $b_{bond}(\gamma, \delta) \leq b(\langle\gamma\rangle, \langle\delta\rangle)$ .*

*Proof.* Consider a path  $p = \gamma_1, \dots, \gamma_n$ . Let  $\langle p \rangle = \langle\gamma_1\rangle, \dots, \langle\gamma_n\rangle$ , and let  $\langle\gamma_i\rangle$  have highest energy along  $\langle p \rangle$ . By Lemma 33,  $E(\langle\gamma_i\rangle) \leq E(\gamma_i)$ . So if  $E(\langle\gamma_1\rangle) = E(\gamma_1)$ , then  $h(\langle p \rangle) \leq h(p)$ .  $\square$

If  $\gamma$  is saturated, then  $\langle\gamma\rangle$  is saturated and  $E(\gamma) = E(\langle\gamma\rangle)$ . So this proof also proves the inequality for saturated paths in the two models.

**Lemma 35.** *If  $\gamma$  and  $\delta$  are saturated, then  $b_{sat-bond}(\gamma, \delta) \leq b_{sat}(\langle\gamma\rangle, \langle\delta\rangle)$ .*

What may be more surprising is that we can also establish a reversed inequality.

**Theorem 36.**  $b_{bond}(\gamma, \delta) + 1 \geq b(\langle\gamma\rangle, \langle\delta\rangle)$ .

*Proof.* Consider a path  $p$  from  $\langle\gamma\rangle$  to  $\langle\delta\rangle$ . Form a path  $p'$  from  $\gamma$  to  $\delta$  by adding makes, breaks, and swaps as follows. Before each split, swap enough to break as few bonds as possible. After each merge, make as many bonds as possible.  $\square$

A saturated path simply needs no makes or breaks, so this proof also proves the inequality for saturated paths in the two models.

**Corollary 37.**  $b_{sat-bond}(\gamma, \delta) + 1 \geq b_{sat}(\langle\gamma\rangle, \langle\delta\rangle)$ .

## 6 Future work

An important type of kinetic barrier is the barrier to nucleation in self-assembly processes. For example, in the abstract Tile Assembly Model (aTAM) [4], [6] a large nucleation barrier is necessary to ensure correct assembly of complex structures. Prior work demonstrated programmable nucleation barriers, both theoretically and experimentally [?], [8], [9]. The TBN model could be used to develop aTAM kinetic barriers that do not rely on the tile geometry. These barriers would be stronger, in the sense that they would exist even allowing lattice errors in assembled structures.

Ideally we not only want a large energy barrier to “bad” configurations, but we want to avoid getting stuck in local minima that keep us from getting to the good configurations. We can define “self-stabilizing” TBNs with the property that from *any* configuration, the TBN can reach *some* stable configuration with a low energy barrier path. This property is true for the grid gate (Theorem 25), but it is not true for the translator cycle. Is there a systematic way to test for self-stabilization, or ensure that a system is self-stabilizing by satisfying certain general properties?

Most of this paper considers the regime where forming a new bond is favorable even if it results a loss of a separate polymer (specifically,  $w \geq 1$  and  $w \geq 2$ ). However, we conjecture that both the translator cycle and the grid gate have  $\Omega(n)$  energy barriers even when bond strength is weak ( $w < 1$ ) as long as  $w = \Omega(1/n)$ . Showing that our constructions work in a wider range of experimental conditions would increase their practical applicability.

Can we use the definition of energy in TBNs to bootstrap a reasonable notion of probability of configurations or paths? For instance, in statistical thermodynamics it is common to consider the Boltzmann distribution induced by energy  $E$ , where

for each configuration  $\gamma$ ,  $\Pr[\gamma] = e^{-E(\gamma)} / (\sum_{\beta} e^{-E(\beta)})$ . This is the probability of seeing  $\gamma$  at thermodynamic equilibrium. One can also use the relative energy between two states to predict the relative rates of transition between them, which might allow defining a notion of path probability in the kinetic theory of TBNs.

A useful chemical module consists of the reaction  $X + X \rightarrow Y + Y$  (or more generally converting  $k > 1$  copies of  $X$  to  $Y$ ), which can act as a “threshold” to detect whether there are at least  $k$  copies of  $X$ . Analogous to a catalytic system, implementing the above reaction while forbidding  $X \rightarrow Y$  requires control of the energy barrier, and cannot be done simply by varying the energies of  $X$  and  $Y$ . Can we construct TBNs with arbitrarily large energy barriers in this case?

What is the computational complexity of deciding whether  $b(\gamma, \delta) \leq k$ , given two configurations  $\gamma, \delta$  and a threshold  $k$ ? This problem is decidable in polynomial space in the number of monomers in  $\gamma$ : any configuration can be written down in polynomial space, and guessing merges and splits yields a nondeterministic polynomial space algorithm (placing it in PSPACE by Savitch’s Theorem). However, low-height paths could be of exponential length, and thus it is not clear that the problem is in NP, since the obvious witness is a path with height  $\leq k$ . Is this problem possibly PSPACE-complete?

Both the grid gate and translator cycle use  $n^2$  unique site types to achieve an energy barrier of  $n$ . This can be reduced for the grid gate (for example, to use  $n$  domains, one can simply make each “row” of the grid the same site type), although we do not know how to make such a system work properly with a catalyst. Specifically, Theorem 30 fails with our initial attempts to create a catalyst for systems with  $O(n)$  site types. It is an interesting open question to design a catalytic system similarly to the grid gate, using only  $O(n)$  site types, that has a programmable energy barrier of  $n$  in the absence of the catalyst.

## Appendix A Translator cycle

### A.1 Strand displacement cascade

Fig. 8 shows the design scheme in which the monomers/polymers of the TBN studied in Section 4.1 are modeling an implementable substrate. The substrate are double-stranded DNA complexes. The complexes undergo a kinetic process called *strand displacement*, shown in Fig. 9, in which one strand attaches and displaces another, freeing a new strand which can then displace strands on other complexes in the system (hence the term *cascade*). Note that the displacement shown in Fig. 9 corresponds to one merge and one split of the pathway shown in TBN form in Fig. 4.

### A.2 Barrier of the translator cycle

Here we give proofs which were omitted in the main body.

**Lemma 13.** *If a polymer  $P$  is in a normal form saturated configuration and  $|P| < 2n + 1$ , then there exists a perfect matching on  $G_P$ .*

*Proof.* Given a set  $S$  of vertices, let  $N(S)$  be the set of vertices adjacent to a vertex in  $S$ . Hall’s condition states that a perfect matching exists on a bipartite graph  $\{V_1, V_2, E\}$  if and only if

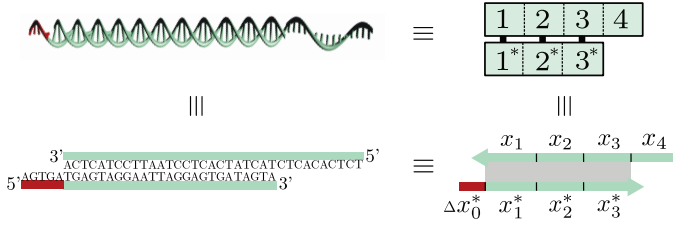


Fig. 8. DNA strand displacement implementation of the translator cycle. The schematic shows how the complexes in the translator cycle TBN correspond to DNA complexes.  $\Delta x_i$  denotes a truncated version of domain  $x_i$ . Note that these truncated domains, typically called *toeholds*, are not included in our TBN abstraction because they bind weakly and do not contribute to the energy barrier we consider.

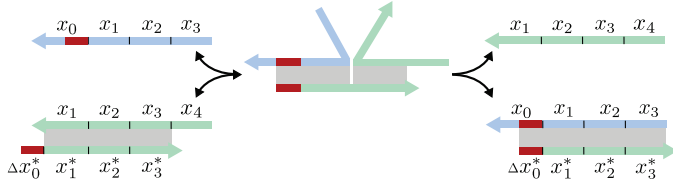


Fig. 9. The desired catalyzed pathway of the translator cycle consists of the strand displacement reactions shown. The displacement reaction initiates when the blue strand binds to the green complex at the toehold domain (red). The top strands compete for bonds in a random walk, eventually displacing the top green strand. Note that this displacement corresponds to one merge then split of the pathway shown in TBN form in Fig. 4.

for all subsets  $S \subseteq V_1$ ,  $|S| \leq |N(S)|$ . We prove this holds for  $G_P$ .

Consider any subset  $S \subseteq B$ . There are  $n|S|$  starred (limiting) binding sites. The number of sites on compatible top monomers for the set  $S$  is given by  $(n+1)|N(S)|$ . Since  $P$  is saturated, the  $n|S|$  starred sites must be bound to the  $(n+1)|N(S)|$  unstarred sites, so we have  $n|S| \leq (n+1)|N(S)|$ . If  $|S| > |N(S)|$ , Lemma 11 gives us that  $|S| > n+1$  and  $|N(S)| > n$ , so to avoid contradicting the assumption that  $|P| < 2n+1$ , it must be that  $|S| \leq |N(S)|$ .  $\square$

**Lemma 17.** For an  $n'$ -sized polymer  $P$  in a normal form saturated configuration, for any two perfect matchings  $M$  and  $M'$  on  $G_P$ ,  $f(M) = f(M')$ .

*Proof.* Intuitively, first we will shift the indices of the top and bottom monomers so that the leftmost index has value zero. Let  $c_P$  be a cutoff value given by Lemma 15. We rewrite each  $b_i$  or  $t_i$  as  $b_{i-c_P \bmod n^2}$  or  $t_{i-c_P \bmod n^2}$ . Since originally no edge crossed the cutoff, now no edge crosses zero. Note that this does not change the offset of any pair and thus does not change the offset of any matching. For each  $f(b_i, t_j)$ , since no edge  $\{b_i, t_j\}$  crosses zero we can think of the indices on a line, so we can rewrite the offset as  $f(b_i, t_j) = j - i$ . Then  $f(M) = \sum_k (j_k - i_k) = \sum_k j_k - \sum_k i_k$ . This expression is independent of the matching and only depends on the indices of the monomers in the polymer, so for any two matchings  $M$  and  $M'$ ,  $f(M) = f(M')$ .  $\square$

**Lemma 19.** Given an  $n'$ -sized polymer  $P$  with cutoff  $c_P$ , there exists a sorted matching  $M$  on  $G_P$  which satisfies that there does not exist  $\{b_{i_1}, t_{j_2}\}, \{b_{i_2}, t_{j_1}\} \in M$  with  $i_1 \leq i_2$  and  $j_1 \leq j_2$  with respect to the ordering given by the cutoff value,  $[c_P, c_P - 1]_{n^2}$ .

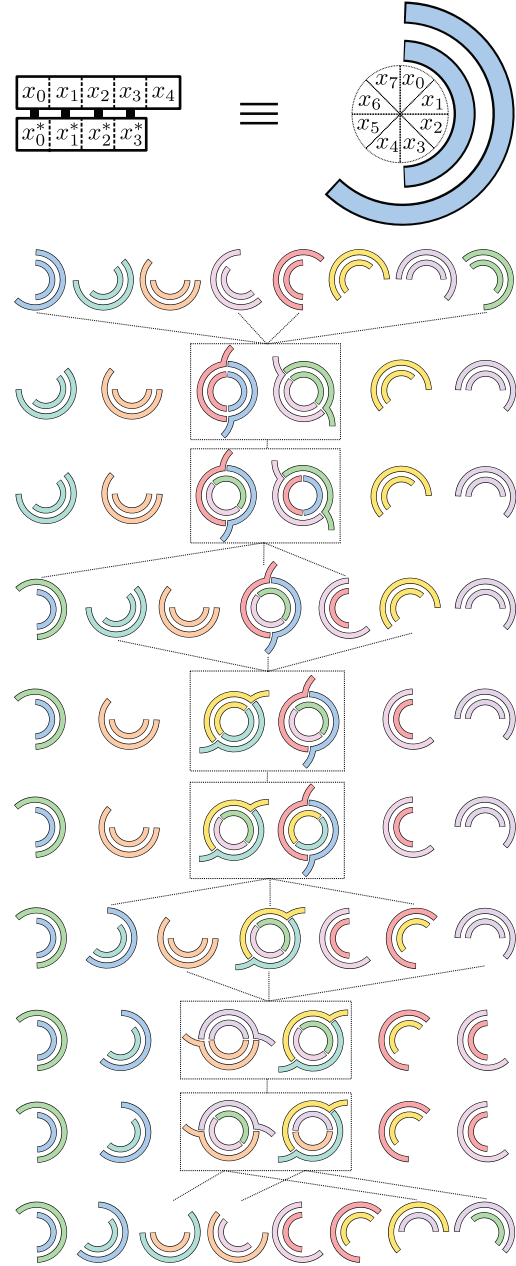


Fig. 10. A path from  $\gamma_I$  to  $\gamma_T$  with height  $\frac{2c}{z} - 1$ , showing that the complex length  $z$  must be asymptotically larger than the number of complex types  $c$  to achieve a super-constant barrier. This example has monomer size  $z = 4$  and number of complexes  $c = 8$ . The circular representation helps to visualize when a group of monomers covers all site types  $x_0, \dots, x_7$ . Each row of polymers denotes a different configuration on the path. Each is separated by several merges or splits. First, two initial complexes are brought together, chosen such that every starred binding site appears at least once. This size  $\frac{2c}{z}$  polymer acts as a catalyst in the following sense: another similar set of initial complexes can be merged into this polymer, and then split into their triggered complexes. Note in the fourth row, the large polymer is still saturated, although each bottom monomer is not in a polymer with its initial or triggered top monomer.

*Proof.* Given any matching  $M'$  which is not sorted, we show that we can swap the offending edges, resulting in a new matching which is in sorted order. Let  $S_{i_k}$  be  $[i_k - n, i_k + n - 1]_{n^2}$ , the sequence giving the indices of compatible top monomers for  $b_{i_k}$ . For any  $\{b_{i_1}, t_{j_2}\}, \{b_{i_2}, t_{j_1}\} \in M$  with  $i_1 \leq i_2$  and  $j_1 \leq j_2$ , note that  $j_2 \in S_{i_1}$  and  $j_1 \in S_{i_2}$ . The orderings



given by the sequences  $S_{i_1}$ ,  $S_{i_2}$  are the same orderings as given by  $[c_P, c_P - 1]_{n^2}$ , since  $S_{i_1}$ ,  $S_{i_2}$  are both subsequences of  $[c_P, c_P - 1]_{n^2}$ . Since  $i_1 \leq i_2$ ,  $S_{i_1}$  contains no elements greater than any of those in  $S_{i_2}$ , and  $S_{i_2}$  contains no elements less than any of those in  $S_{i_1}$ . So  $j_2 \in S_{i_1}$  and  $j_1 \in S_{i_2}$  with  $j_1 \leq j_2$  gives that both  $j_1, j_2 \in S_{i_1}$  and further  $j_1, j_2 \in S_{i_2}$ . So both  $t_{j_1}$  and  $t_{j_2}$  are compatible for both  $b_{i_1}$  and  $b_{i_2}$ . Since the edges of  $G_P$  are between bottom monomers and their compatible tops, we can swap  $\{b_{i_1}, t_{j_2}\}, \{b_{i_2}, t_{j_1}\}$  with  $\{b_{i_1}, t_{j_1}\}, \{b_{i_2}, t_{j_2}\}$  and the result is a matching on  $G_P$ . So given any perfect matching on  $G_P$ , we can construct a sorted matching by swapping the offending edges one-by-one.  $\square$

## Acknowledgements

DD and DH were supported by NSF grants CCF-1619343, CCF-1844976, and CCF-1900931. CC, KB, and DS were supported by NSF grants CCF-1618895 and CCF-1652824, and DARPA grant W911NF-18-2-0032.

## References

- [1] Alexandre Baccouche, Kevin Montagne, Adrien Padirac, Teruo Fujii, and Yannick Rondelez. Dynamic DNA-toolbox reaction circuits: a walkthrough. *Methods*, 67(2):234–249, 2014.
- [2] Luca Cardelli. Strand algebras for DNA computing. *Natural Computing*, 10(1):407–428, 2011.
- [3] Yuan-Jyue Chen, Neil Dalchau, Niranjan Srinivas, Andrew Phillips, Luca Cardelli, David Soloveichik, and Georg Seelig. Programmable chemical controllers made from DNA. *Nature Nanotechnology*, 8(10):755–762, 2013.
- [4] David Doty. Theory of algorithmic self-assembly. *Communications of the ACM*, 55(12):78–88, December 2012.
- [5] David Doty, Trent A. Rogers, David Soloveichik, Chris Thachuk, and Damien Woods. Thermodynamic binding networks. In Robert Brijder and Lulu Qian, editors, *DNA Computing and Molecular Programming: 23rd International Conference*, pages 249–266. Springer, 2017.
- [6] Matthew J. Patitz. An introduction to tile-based self-assembly and a survey of recent results. *Natural Computing*, 13(2):195–224, 2014.
- [7] John SantaLucia Jr and Donald Hicks. The thermodynamics of DNA structural motifs. *Annual Review of Biophysics and Biomolecular Structure*, 33:415–440, 2004.
- [8] Rebecca Schulman and Erik Winfree. Synthesis of crystals with a programmable kinetic barrier to nucleation. *Proceedings of the National Academy of Sciences*, 104(39):15236–15241, 2007.
- [9] Rebecca Schulman and Erik Winfree. Programmable control of nucleation for algorithmic self-assembly. *SIAM Journal on Computing*, 39(4):1581–1616, 2009. Preliminary version appeared in DNA 2004.
- [10] David Soloveichik, Georg Seelig, and Erik Winfree. DNA as a universal substrate for chemical kinetics. *Proceedings of the National Academy of Sciences*, 107(12):5393–5398, 2010.
- [11] Niranjan Srinivas, James Parkin, Georg Seelig, Erik Winfree, and David Soloveichik. Enzyme-free nucleic acid dynamical systems. *Science*, 358(6369):eaal2052, 2017.
- [12] Chris Thachuk, Erik Winfree, and David Soloveichik. Leakless DNA strand displacement systems. In *International Workshop on DNA-Based Computers*, pages 133–153. Springer, 2015.
- [13] Boya Wang, Chris Thachuk, Andrew D Ellington, Erik Winfree, and David Soloveichik. Effective design principles for leakless strand displacement systems. *Proceedings of the National Academy of Sciences*, 115(52):E12182–E12191, 2018.

**Cameron Chalk** received his B.S. and M.S. degrees from the University of Texas Rio Grande Valley, Edinburg, TX, USA, in 2015 and 2017 respectively. He is currently pursuing a Ph.D. degree from the University of Texas at Austin, TX, USA.



**David Doty** is an Assistant Professor of Computer Science with the University of California at Davis.



**David Haley** is a Ph.D. candidate in the Graduate Group in Applied Mathematics at the University of California at Davis.



**David Soloveichik** received the B.S. and M.S. degrees from Harvard University, Cambridge, MA, USA, in 2002 and the Ph.D. degree from the California Institute of Technology, Pasadena, CA, USA, in 2008. He is currently an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Texas at Austin, TX, USA.

**Keenan Breik** is a student at the University of Texas at Austin.